Music HAI For Innate Human Musical Creativity And Instrumentality

Matthew Perez Gin
Cornell University
Ithaca, USA
mpg223@cornell.edu

Is it wrong, wanting to be at home with your record collection? It's not like collecting records is like collecting stamps, or beermats, or antique thimbles. There's a whole world in here, a nicer, dirtier, more violent, more peaceful, more colorful, sleazier, more dangerous, more loving world than the world I live in; there is history, and geography, and poetry, and countless other things I should have studied at school, including music.

– Nick Hornby, 1983

So tell me, do you have the answers for me?
I've spent my life in search of what I've got
And if I die before the trumpets call me
Will you tell my story for me?
Or at least the gist of what I'm fighting for

–Matthew Perez Gin

Author Keywords
Music HAI, Music Cognition, Songwriting, HCI

CSS Concepts
• Human-AI Interaction

INTRODUCTION

Music as a human phenomena is a vastly interdisciplinary concept, bridging global communities and cultures. The onset of artificial intelligence "hype" has exacerbated music research communities' efforts to provide novel, generalizable methods for audio technologies. Advancement in the music technology realm has transformed over recent years, integrating generalizable knowledge from other domains such as natural language processing (NLP), large language models (LLMs), Dynamic Time Warping (DTW), and Music Transformers, to name a few models.

Simultaneously, music cognition researchers have surmounted many of the complexities related to how humans perceive and process aspects of music, shedding light on major underpinnings such as enculturation, perception of musical events, and improvisation. State-of-the-art (SOTA) generative music models are also gaining the attention of researchers and major investment stakeholders (as well as negative attention from music rights organizations), taking startup form. But who's being left behind?

Music, of course, is nothing new–in fact, most music is recycled by humans during music creation and song-writing; however, artists and listeners are almost certainly going to bear the brunt of advancement in music technologies and are already facing artistic battles with generative music systems. In light of the vast array of problems in the music technology space, we focus our research on the *innate human human musical creativity and instrumentality*. Through human-AI interaction (HAI) approaches, we identify current shortcomings of SOTA music generation models, which lack controllability and effective explainability for musicians. In other words, current models do not provide interpretable feedback to users (if any at all) about their musical expressions. We also aim to remedy the data void within the Music Information Retrieval (MIR) and Music Cognition space, which has seen rippling effects in the training of music generations models. Musical training data that does not violate legal restrictions, and is representative of a wider scale of music communities–beyond Western music, is relatively sparse compared to other generative models such as LLMs. We have entered a new era of generated content, and music is no exception.

Our system leverages the principles of music cognition and music psychology–grounded in personal work in Fuzzy-Trace Theory and "The Gist of Music." We employ an HAI approach with Research through design (RtD) methods to propose a system that does not solely generate music but augments the creative processes humans undergo during music creation. The key to our HAI music system is emulate music creation as a temporally dynamic process–how a band creates music. With AI bandmates, artists will have the creative freedom to develop personal collaborators to interact with in a musical setting, while providing humans with musical context beyond current music AI capabilities.

RELATED WORK

Entrainment

In music, Entrainment refers to humans' ability to synchronize to music – beat, rhythm, pitch, or lyrical flow. In this research, I aim to understand the biomusicological basis of humans and music and to bridge the gap between the literature on musical experience and technological interaction that draws people to music technologies.

Start simple: Human-AI Co-creation in Songwriting

Huang et al. 2020 examine the collaborative relationship between musicians and developers, focusing on their needs, challenges, and approaches to leveraging the empirical characteristics of AI for music creation. Their research emphasizes aspects of "good" human-AI interaction, particularly in the context of generative music models. Huang poses the central question: "How might humans co-create with an open-ended set of deep generative models in a complex task setting such as songwriting?"

The study shows that teams composed of musicians and developers adopt a modular approach to human-AI co-creation, breaking down musical objectives into segments—such as melody, harmony, bassline, drums, multiple parts, structure, vocal synthesis, and instrument synthesis. Working incrementally and drawing on different music generation models, these teams build songs piece by piece.

Huang's work resonates with challenges identified in Gardner et al. (2022) regarding MT3 (Multi-Task Multitrack Music Transcription), where researchers grapple with documenting and recording multiple layers within a single song. The key insight from Huang et al.'s study is that, as in traditional music-making, creators prefer flexible, layered techniques over rigid, tool-specific approaches. Instead of following a strict "create X sound with Y tool" pattern, they blend and curate a variety of elements—X, Y, and Z sounds—into a final composition.

Connect the dots: AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition

Suh et al. (2021) take a similar approach to Huang's research on human-AI co-creation, positioning AI as a potential "glue" in music composition, particularly in social settings. Their focus is on how AI impacts creativity and influences human social dynamics during the creative process.

They identify five key roles of AI:

1. Implicitly establishing a common ground at the start of collaboration,
2. Acting as a psychological safety net for creative risk-taking;
3. Driving group progress;
4. Mitigating interpersonal stalling and friction; and
5. Shaping users' collaborative and creative roles.

Their findings draw on empirical research from other artistic domains, including creative writing, design ideation, visual metaphors, video game content generation, improvisation, public displays, and dance. This research offers a holistic, interdisciplinary perspective on music creation, emphasizing the social dimensions of human-AI collaboration that Huang and others touch on less extensively—particularly the human-to-human interactions that drive creativity.

When beginning this research, my goal was to understand the network of stakeholders within the music domain. This paper provides a clearer picture of how AI reshapes human roles in music composition. With AI tools readily available, creators are less siloed in their creative spaces and can engage in more open communication, particularly when taking creative risks.

Monkey see, Monkey do (respectfully): Music Creation by Example

Frid et al. establishes a user interface (UI) paradigm where an AI engine takes a song inputted by a human and enables them to interactively regenerate and mix their own AI-generated music (to mimic short music-video art, like in TikTok-form). They implement a user study using 104 video creators. The paradigm centers around "computational creativity" which collectivizes a ground-truth understanding of what humans deem as creative. In this sense, we can see a parallel (more so a contrast, though) between definitions of creativity in music generation, as described in Turetsky, 2003 (Ground-Truth Transcriptions of Real Music from Force-Aligned MIDI Syntheses), an older implementation of generative music where generation is based on symbolic transcriptions of music (force aligned via Dynamic Time Warping (DTW)) and takes essentially no human characteristics of ground-truth in mind.

The study found that users benefit from having access to large libraries of music, which is historically underrepresented in the music information retrieval (MIR) field, but is an increasingly demanded space–a problem highlighted in Turetsky and essentially all other state-of-the-art music generation literature. Frid establishes an interesting stratification between novice users and more musically-trained users, demonstrating that novice users may typically want to consider the high-level properties of music. This is important to note as my AI-tool will try to create a fairer, more representative form of music creation for users of all backgrounds. Additionally, like the notion of entrainment, this study supports that users benefit from having examples of music, moods, and melodic features that remind them of the music they are trying to create (like existing songs, though this is a largely debated legal and copyright domain).

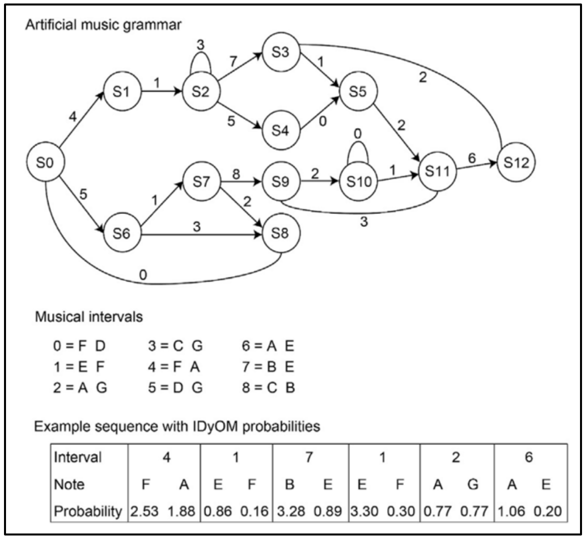
Building on prior work, such as Frid et al. (2020), Han et al. incorporate text-to-image representations of music and explore audio-video modes of music generation. A significant contribution of this research is its examination of the legal challenges surrounding music generation, particularly with the increasing popularity of short-video formats. These legal hurdles, rooted in complex copyright restrictions, pose significant difficulties for both developers and users in the music-AI domain, highlighting an urgent need for solutions in this evolving field.

Creativity. I'll be the judge of that: From learning to creativity: Identifying the behavioural and neural correlates of learning to predict human judgements of musical creativity

There is one major problem I see with both Human-AI interaction and Music Information Retrieval (MIR), and by extension their intersection in Music Generation: the literature often fails to address the psychological and cognitive aspects of music. My research aims to fill this gap urgently. As the use of music-generation tools continues to grow, I cannot accept advancements that overlook the deeply psychological and cognitive foundations of music.

Zioga et al. (2019) provide a foundational understanding of the behavioral and neural aspects of learning to better understand and predict human judgments related to music creation. They introduce the concept of artificial music grammar (AMG), which simulates real-life music learning by training nonmusicians on an unfamiliar grammar (Rohrmeier et al., 2011). Using behavioral and electroencephalogram (EEG) analyses, the researchers establish a paradigm for understanding music learning and creativity. They identify neural correlates of learning during the cognitive encoding of the artificial music grammar and find direct associations between behavioral and neural measures that reflect human judgments of creativity.

This cognitive understanding of human judgment is a key area I want my research to emphasize—something I find increasingly absent in state-of-the-art music-AI technologies. By integrating these insights, my research will strive to align music-generation tools more closely with the cognitive and psychological realities of how humans create and experience music.



Musical intervals

0 = F D	3 = C G	6 = A E
1 = E F	4 = F A	7 = B E
2 = A G	5 = D G	8 = C B

Example sequence with IDyOM probabilities

Interval	4		1		7		1		2		6	
Note	F	A	E	F	B	E	E	F	A	G	A	E
Probability	2.53	1.88	0.86	0.16	3.28	0.89	3.30	0.30	0.77	0.77	1.06	0.20

From Zioga et al., 2019. "This AMG is a finite-state grammar with tones belonging to the diatonic scale, constructed such that eight different tone pairs are combined under specific rules producing 18 different melodic sequences. An AMG was ideal for our purpose because it represents a completely novel musical style for all participants."

Music from the ~~heart~~brain, and it's underlying biological and psychological foundations: Music Perception

Pearce (2023) provides a comprehensive synthesis of current state-of-the-art approaches to understanding music perception and cognition. Building on the cognitive foundations of musical pitch, popularized by Cornellian Carol Krumhansl, Pearce explores the fundamental ways humans perceive and interact with music. His work establishes a crucial cross-disciplinary link between music and the human mind—both psychological and biological—that many music-AI studies fail to address.

One key consideration for the music generation research space is human emotion in music—how humans react to and anticipate music emotionally. While Pearce does not specifically examine human cognitive interaction with music technologies or music-AI, his literature synthesis plays a vital role in informing the development of music technologies that account for how humans cognitively engage with sound and audio.

Pearce also discusses the multimodal nature of cognition, often tied to visual elements such as dance, musical theater, opera, and video. This concept of multimodal interactions aligns well with Han et al. (2024)'s exploration of Understanding the Use of AI-Based Audio Generation Models by End-Users. Additionally, the persistent challenges of multimodal data collection and transcription in music technology could benefit from cognitive principles of human-music interaction. These principles provide a domain-specific framework for designing Human-AI Interaction (HAI) tools in music that are better suited to how people perceive and create music.

Transformers, and other powerful machines:

The current state-of-the-art generative music machines learning algorithms follow the transformer-based architecture, notably introduced by Vaswani et al., in 2017. However, previous methods for predictive music include the use of Hidden-Markov Models (HMM), Dynamic Time Warping (DTW), Convolutional and Recurrent Neural Networks (CNNs and RNNS). This section primarily focuses on the machine learning techniques used in music generation literature.

To compute, or not to compute: High Fidelity Neural Audio Compression

Défossez et al. (2022) introduce a state-of-the-art real-time, high-fidelity audio codec that uses neural networks to minimize audio information loss during the generation process. Their model takes existing audio samples as input and outputs audio snippets generated based on the model trained on those initial samples. While their research relies heavily on human-AI interaction methods for the music generation algorithm, they notably incorporate human evaluation in testing the performance of their generative model.

During evaluation, participants were presented with compressed audio codecs (both their own and those of competing methods) and asked to rate these codecs alongside the uncompressed ground-truth audio. This research represents a significant breakthrough in music generation by providing a more computationally efficient method for compressing audio. However, unlike Huang et al. (2020), Frid et al. (2020), and Suh et al. (2021), this paper does not primarily focus on human-AI interaction.

Despite this difference, Défossez et al. make groundbreaking contributions to music generation by introducing more computationally efficient architectures for researchers and developers. Their work on High-Fidelity Neural Audio Compression addresses a key challenge in music-AI: managing data efficiently while maintaining quality—an area that has historically posed significant difficulties for researchers in the field.

Who did it first?: Music Transformer

Following the notorious publication, “Attention Is All You Need,” Huang et al., 2018 and the team at Google Brain pushed their research on the implementation of transformer-based architectures for music generation. They leverage a self-attention method they consider to be well-suited for music modeling, tuning the model to encompass relative timing in music composition and performance. Huang posits that Shaw et al., 2018 (Self-Attention with Relative Position Representations) is computationally inefficient, especially with longer forms of audio generation. A notable method the team used relates to their use of Piano-e-competition data, specifically human-composed piano pieces. Human-generated content for model training is a major debate in the human-AI interaction aspects of music generation, and it is an important consideration to make when training generative music models. In other words, I pose the question: to what extent will we rely on human-created audio for model training versus synthetic, generative music? For now, it seems as though researchers face a problem with the quality of generative music – it’s not perfect right now. Huang performs a similar human evaluation compared to Défossez, finding that their model performed much better in listening tests. However, there was no mention of the types of folks evaluating the music, something that Pearce et al., 2020 do a good job of, as this is critical information to denote in generative music research.

Follow my lead, I might follow yours: MT3: Multi-Task Multitrack Music Transcription

Another challenge in generative music research is the transcription of existing music, known as transcriptions, which are crucial for modeling and training algorithms for audio creation. Gardner et al. (2022) and Google Brain introduce a multitrack music transcription method capable of transcribing audio from low-resource instruments, such as guitar, while maintaining strong performance on prominent auditory tracks like piano.
A significant connection to Human-AI interaction lies in Gardner et al.’s use of human-created and human-annotated Musical Instrument Digital Interface (MIDI) data. While much of Western music (e.g., classical music) is documented in physical sheet music—often handwritten—there is a scarcity of symbolic data representing the music of other cultures.

This disparity in representation poses a major challenge for fairness in AI-music tools, as many non-Western musical traditions lack formal documentation. These traditions are often rooted in living, temporally finite experiences rather than being preserved in written or symbolic form. Addressing the documentation disparity in music is essential for the design of my AI tool. Users, listeners, and musicians have diverse goals in music creation, and ensuring fairness and inclusivity in AI tools requires acknowledging and accommodating these differences.

Unaligned Supervision for Automatic Music Transcription In-the-Wild

differences. Maman et al. (2022) introduce NoteEM, a method for training an audio transcriber to automatically align musical scores with their corresponding performances. Their system employs an "unaligned supervision" design, using pseudolabels and pitch-shift augmentation to improve transcription accuracy for "in-the-wild" recordings. The motivation behind this research stems from the labor-intensive nature of generating training labels for music transcription, a process that could be partially automated to support the development of music-generation models.

This raises an important question: to what extent can we automate the transcription process without entirely eliminating human contributions to music data collection? While this work builds on similar efforts, such as Gardner et al. (2022), which explore automated labeling and ground-truth formation in music creation, it does not fully align (pun intended) with my vision. While music transcription remains a significant barrier to advancing generative music research, I believe it is crucial to involve stakeholders—such as domain experts, music creators, and listeners—in the curation of datasets for music model training. Documenting these processes, akin to "datasheets for datasets," is equally important as it enhances explainability, transparency, and intelligibility for a wide range of stakeholders, including users, researchers, developers, and designers.

The paper also highlights temporal and biological aspects of human-music interaction. Maman et al. echo Pearce (2023) in observing that the human ear is more sensitive to onset time than to sustain and offset time (the point when a sound ends). While I largely agree with this finding, it is also essential to consider other facets of music perception, as described by Pearce (2023).

These include fundamental frequency, spectral content, amplitude, amplitude envelope, and primary perceived musical attributes such as pitch, timbre, loudness, and timing. These broader considerations are vital for building systems that account for the richness of human music perception.

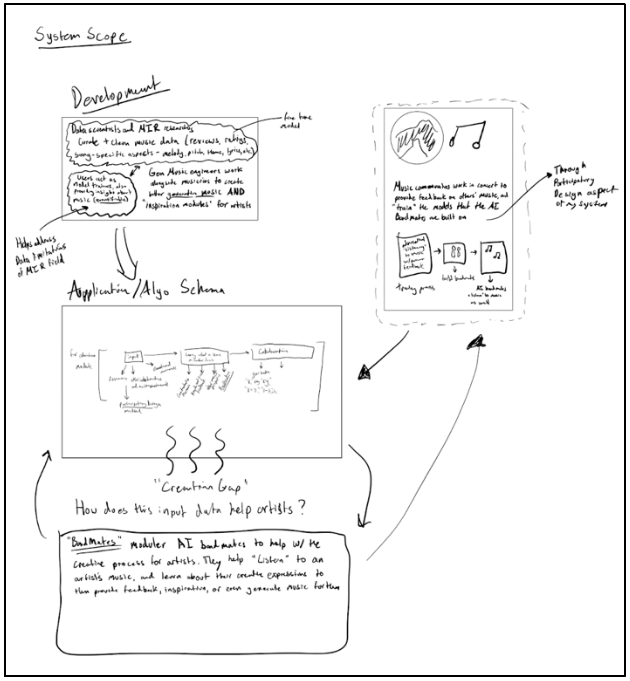
HAI for a Better World of Music – Starting With Data

I envision a collaborative generative music space in which humans–with any level of musical competence–can create music that reflects their individual and desired artistic expression while doing so in a way that can adapt to the users' creative flow in real-time. By doing so, the system should enhance the music creation experience for users and provide better, more engaging listening experiences for all audiences. However, the prerequisite to more engaging listening experiences is building a system that can learn from the artists themselves. The scope of the system is constrained to an individual artist; however, by design, my system aims to create a more resonant experience for listeners based on inputted feedback from the artist about their own reactions to the generated music.

AI music systems are constantly improving, but are highly "black-boxed," and confusing to users. The most advanced, commercially available generative music models to date are capable of taking inputs such as text descriptions, audio files, images, and video to generate music. However, these models have little to no interactive capability for music creation. My system leverages current multimodal machine learning and AI models and proposes the use of real-time interactive feedback–multimodal inputs–to learn from and enhance the creative flow of its users. The system draws on the theoretical understanding of "bandmates" through which my system will be able to learn about the artist's creative process and collaborate with the human artists to provide enhanced creative inspiration based on an initial target for the artist.

The primary stakeholders are end-users (artists of all musical talent and exposure who want to create music), and AI engineers and developers who collaborate with UX designers to understand the system, its training data, and ultimately its musical output. Indirect stakeholders may include external third-party listeners; though, in some ways, a listener can be the artist themself. Looking towards the future, my system will contribute to the collection of music data, a problem space in MIR, subsequently enhancing AI music models. Ultimately, my system will provide better, real-time input feedback from artists to the generative music model to enhance the creative musical experience of human artists.

TLDR: *Generative music systems should exist at the intersection of innate human musical creativity and instrumentality – they should intelligently learn from an artist's creative process while explaining how their musical predictions align with the user's desired expressions to best resonate with the intended listener.*


System Scope
Development
Data scientists and MIR researchers
Curate + clean music data (reviews, ratings, song-specific aspects - melody, pitch, theme, lyrics, etc)
fine tune model
Users act as model trainers, also providing insight about music (quantifiable)
Gen Music engineers work alongside musicians to create better generative music AND "inspiration modules" for artists
Helps address Data limitations of MIR field
Application/Algo Schema
Input
Collaboration
Music communities work in concert to provide feedback on others' music, and "train" the models that the AI Bandmates are built on
Through Participatory Design aspect of my system
Training process
build bandmates
AI bandmates "listen" to music as well
"Creation Gap"
How does this input data help artists?
"BandMates" moduler AI bandmates to help w/ the creative process for artists. They help "Listen" to an artist's music, and learn about their creative expressions to then provide feedback, inspiration, or even generate music for them

Design sketch for the scope of my HAI music system.

Development: Data Scientists and MIR researchers will collaborate to curate and clean the music data, overlapping with generative music engineers and musicians–as domain experts–to create better generative music and "inspiration modules" for artists–as AI bandmates.

Algorithm and Application

Based on current literature and SOTA (state-of-the-art) applications, our system will help alleviate one major problem we call "the creation gap." The creation gap stems from artists' cognitive and compositional struggles identified in the literature review and field study. We leverage transformer-based music generation models and identify critical issues in the data pipeline for music.

The Data Problem

Currently, SOTA large transformer-based music models are trained on relatively small data collections, many of which are subject to copyright. Our system creates a *musical data pipeline* to help address these data limitations. To do so, the design focuses on supporting human creativity, encouraging users to interact with the system as they would a human band.

The Gist – A Double Entendre

It's funny. The gist of this HAI music system centers around personal previous work in music psychology, which builds on Valerie Reyna's Fuzzy-Trace Theory (FTT), which provides a robust framework for addressing these issues by introducing a psychological dual-process model that accounts for both gist-based (global, relational, and abstract concepts) and verbatim-based (detail-oriented, numerical and fact-based cognition) processing. This HAI music system works to collect both granular, empirically quantifiably music data alongside more abstract or "gisty" music contexts.

The gist of the system is to address the shortcomings of data collection in MIR, and to provide artists with more freedom in terms of the use and even the model training of algorithm modules such as the AI bandmates we have proposed. Ultimately, addressing the data problem in music cognition and generation will allow for better model performance to help support musicians creatively.

(Diet) SOTA Music Models

So many beverages to choose from
I find myself confused
Such a simple task at hand…
I want the real sugar (real sugar)
I want the real sugar (real sugar)

–"Real Sugar," Minden

We now take a look at a SOTA symbolic music generation system from Thickstun et al., 2024: an Anticipatory Music Transformer (AMT) as a method for controllable music generation based on temporal point process (musical event process) representation. The model focuses on infilling control tasks, that generate music given a beginning and end musical event, and is trained on the "large and diverse Lakh MIDI music dataset" (Thickstun et al., 2024). The model bases its capabilities in musical "anticipation," which can be loosely defined as a computational version of anticipation and expectation in music cognition–the ability to anticipate, process, and adapt to future events (Dennett, 1991; Schultz et al., 1997, from Pearce, 2023). Thickstun echoes the gist of our work, identifying controllability over generative music models as a key issue that users and artists currently have: "Users value agency in human-AI collaborations, preferring to take an active role over more automated solutions."

Our system leverages text data as a critical input to training AI bandmates to contextualize and extrapolate musical information, personalized for an individual artist. Thickstun mentions that their team is intrigued by the potential to apply anticipation to generative symbolic music that is conditioned on localized text labels such as lyrics (Thickstun et al., 2024). We suggest that these labels could expand beyond just lyrics into other forms of textual data such as overall theme, target audience, genre, or other textually expressive aspects of music to train a generative music model that can provide textual representations of its interworking to artists through AI bandmates.

I Want The Real Sugar (Perez Gin, 2024)

Although substantial progress has been achieved separately in music cognition—exploring the psychological and biological underpinnings of human interaction with music—and in music technology—designing machine learning models for music creation—a notable divide remains between these fields. However, an HAI system at intersection of music psychology and human-AI interaction offers a path forward for advancing music generation technologies. Fuzzy Trace Theory (FTT) provides a valuable framework to connect these areas by aligning computational models with the cognitive mechanisms that drive music perception and creation.

To advance human-AI music generation, it is essential to bridge the gap between music cognition and music technology by developing AI models that incorporate cognitive principles. Our system FTT principles of emphasizing gist-based processing, which aligns with how humans perceive and create music. For example, novice users, as noted by Frid et al. (2020), benefit from engaging with high-level musical concepts, and AI systems that provide abstractions of musical ideas can enhance both creativity and accessibility. Additionally, incorporating cognitive processes such as expectation, memory, and emotional response—highlighted by Pearce (2023) and Zioga et al. (2019)—can make AI tools more intuitive and aligned with human musical cognition. This approach has the potential to create AI systems that not only generate music but also foster creativity, collaboration, and a deeper understanding of music as a profoundly psychological experience.

AMT emphasizes balancing simultaneous and sequential musical relationships while aligning computational processes with human cognition. Temporal decision-making in music requires managing discrete musical events alongside anticipated outcomes, a challenge addressed by frameworks such as FTT. By distinguishing between gist-based and verbatim representations, our system enables existing generative music models to prioritize high-level patterns like melody and rhythm (gist) while maintaining fidelity to detailed elements like note sequences and harmonic precision (verbatim).

We can leverage transformer music models like AMT by incorporating gist-based reasoning into anticipatory infilling tasks, capturing overarching structural patterns while preserving the fine-grained details essential for local coherence. Using the data collected with our music data pipeline, we can inform anticipation mechanisms by extending the model's context length dynamically, allowing for broader relational insights (feedback for artists from AI bandmates) while focusing on fine-grained accuracy during moments of high complexity (generative music suggestion from AI bandmates).

As outlined in the I/O flow, human evaluation metrics are also critical in assessing generated music. Evaluation may incorporate diverse contexts, such as live performances and group listening, to reflect the multifaceted nature of human music perception. By integrating FTT into AMT's representation learning, the model gains the ability to capture relational and temporal flows effectively, generating outputs that align with cognitive principles and enhance user experience through explainable and interpretable mechanisms. This data will also be interpretable through the datasheets to provide better transfer of music information from researchers to designers to developers to other stakeholders of our system such as music listeners.

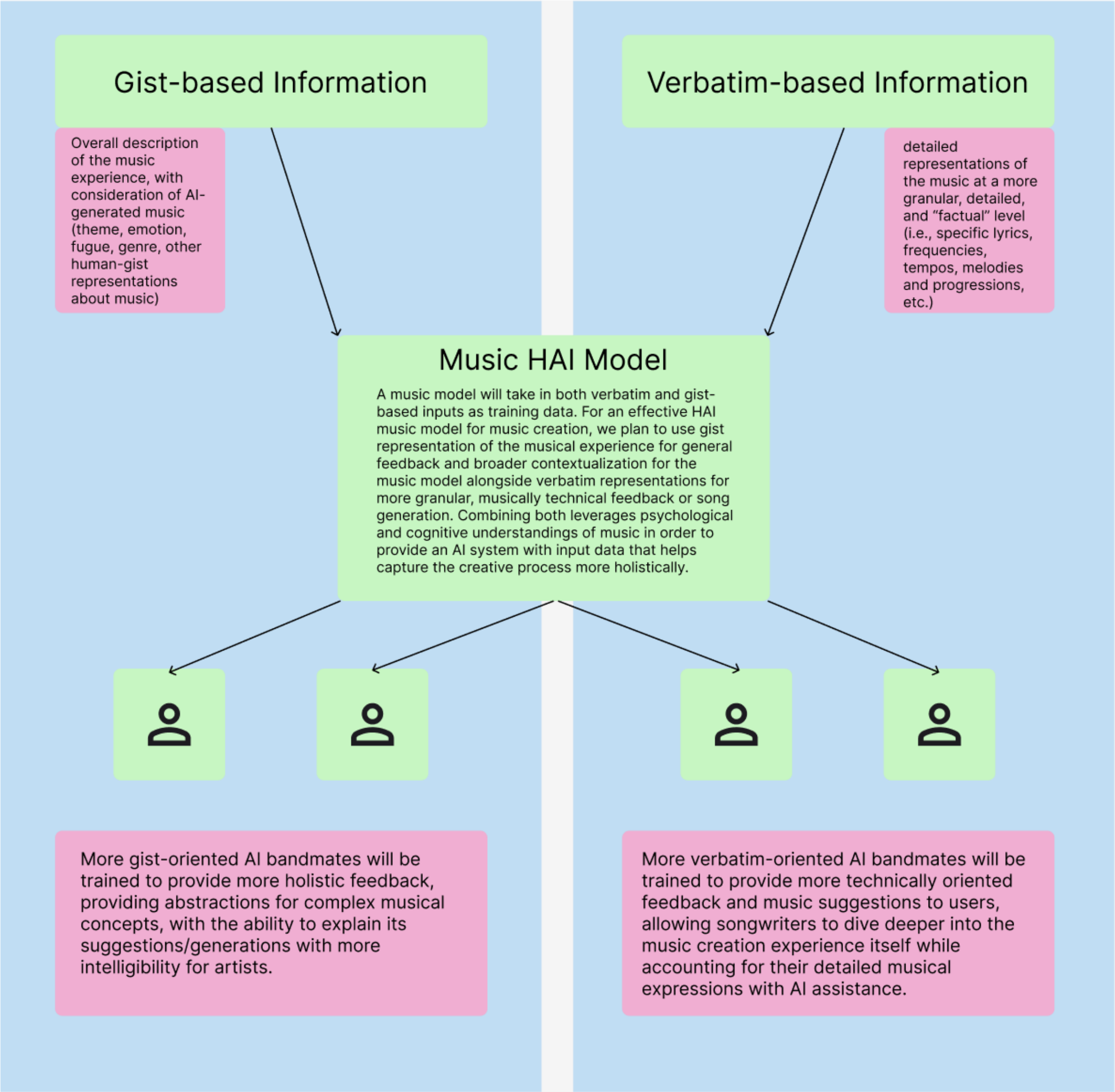
Healing the gash between music cognition and generative music models is a critical part of our design. By doing so, we can present a music system that provides better human-AI interaction, while providing psychological freedom in the creative process that is musically augmented by AI bandmates. An AI system that emulates a human bandmate will also enhance the artistic expression of users by creating a more personable environment. This is the "real sugar" in an effective AI songwriting assistant, not currently encompassed by other models.

Bridging the gap between music cognition and music technology – A New Music AI Model (Perez Gin, 2024)

In the context of music human-AI interaction, despite advancements in generative music models, there remains a significant gap in integrating psychological and cognitive aspects of music into AI-driven music technologies. This disconnect between music cognition and music technology arises from the latter's limited incorporation of the cognitive mechanisms that underlie human musical experience. Applying FTT (Fuzzy Trace Theory) to music generation highlights the need for AI systems to prioritize capturing the gist of music—its overarching structures, emotional narratives, and stylistic elements that resonate deeply with human listeners—over precise replication of musical details.

By focusing on gist-based representations, AI models can better align with how humans perceive and interact with music. For instance, as Frid et al. (2020) suggest, novice users often gravitate towards engaging with high-level musical concepts. AI systems that facilitate this interaction by offering abstractions of musical ideas can significantly enhance creativity and accessibility. Additionally, incorporating cognitive processes such as expectation, memory, and emotional response—emphasized by Pearce (2023) and Zioga et al. (2019)—can guide the design of AI tools that are not only more intuitive but also more attuned to human musical cognition.

To advance the field of human-AI music generation, it is crucial to bridge the gap between music cognition and music technology. Fuzzy Trace Theory offers a compelling framework by highlighting gist-based processing, which aligns with human cognitive approaches to music perception and creation. Our research focuses on developing AI models that integrate cognitive principles, enabling more natural and meaningful interactions between humans and AI in music generation. This approach has the potential to create AI systems that not only generate music but also enhance human creativity, foster collaboration, and deepen our understanding of music as a profoundly psychological and cultural phenomenon.



Gist-based Information

Overall description of the music experience, with consideration of AI-generated music (theme, emotion, fugue, genre, other human-gist representations about music)

Verbatim-based Information

detailed representations of the music at a more granular, detailed, and "factual" level (i.e., specific lyrics, frequencies, tempos, melodies and progressions, etc.)

Music HAI Model

A music model will take in both verbatim and gist-based inputs as training data. For an effective HAI music model for music creation, we plan to use gist representation of the musical experience for general feedback and broader contextualization for the music model alongside verbatim representations for more granular, musically technical feedback or song generation. Combining both leverages psychological and cognitive understandings of music in order to provide an AI system with input data that helps capture the creative process more holistically.

More gist-oriented AI bandmates will be trained to provide more holistic feedback, providing abstractions for complex musical concepts, with the ability to explain its suggestions/generations with more intelligibility for artists.

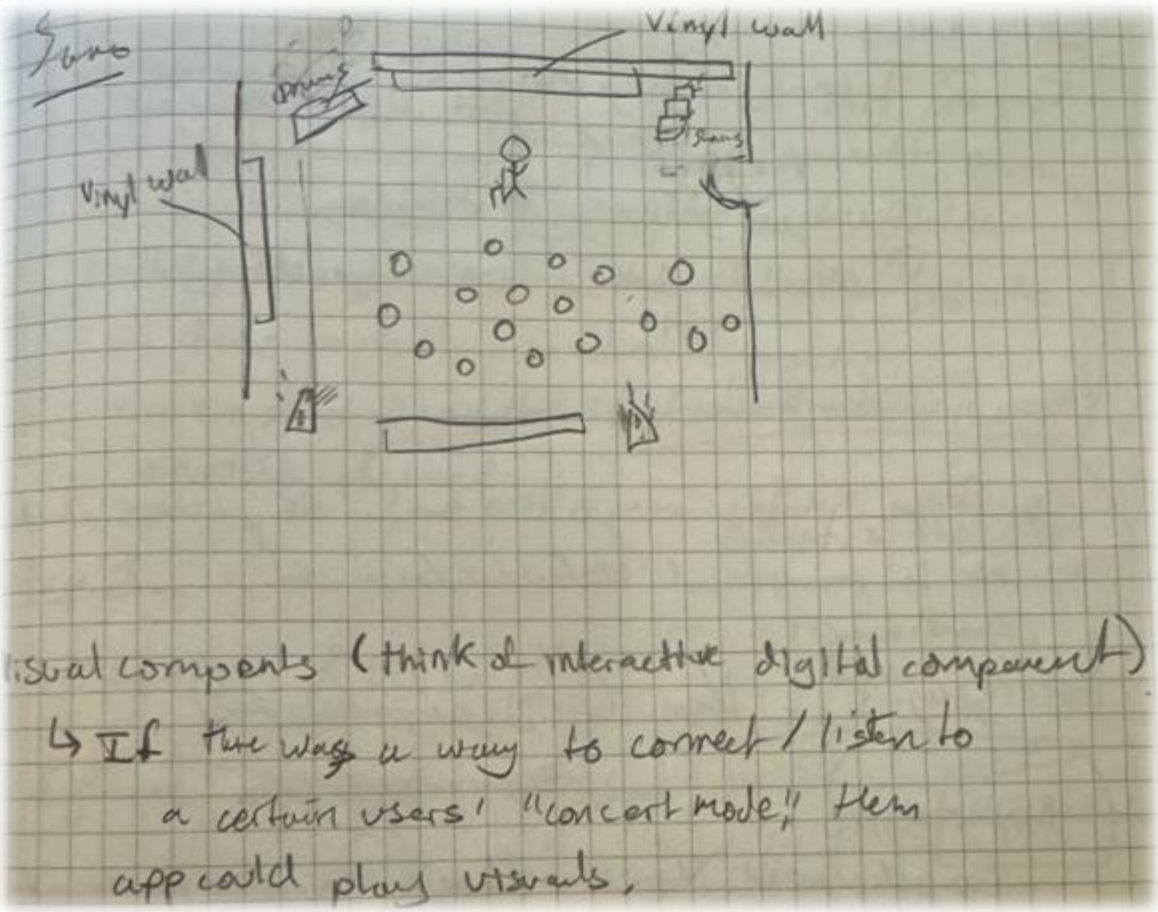
More verbatim-oriented AI bandmates will be trained to provide more technically oriented feedback and music suggestions to users, allowing songwriters to dive deeper into the music creation experience itself while accounting for their detailed musical expressions with AI assistance.

Taking the stage, hitting the dancefloor! An anthropologist's approach to music problem spaces.

A transformative yet originally overlooked aspect of music creation is music in a live setting. Below is an excerpt from a journal piece that I wrote during a "field study" using a SOTA music AI tool, Suno during an open mic event for Electric Buffalo Records at Cornell.

"Creation… What is that? As I slinked back to the recording booth in the Cornell Media Guild house, I could not help but stop for a moment to listen to the other performers practicing their songs. The laughter, the mistakes, and the redoes in their takes, brought me back to life. They are covering "FourFiveSeconds" by Rihanna, Kanye West, and Paul McCartney. These students, some of them my friends, sound amazing to me. In a live setting, there is much more life behind the scenes than what Suno's current features can compute. The buzzing imperfections of a practice session, the voice cracks in the singer's voice as she hits the high note spent weeks perfecting just to fall flat, seemed to be what people wanted. I wanted this too.

Creation is different with Suno. While I hoped that such an interaction–asking an audience for prompts of what they wanted to hear–would replicate that of a band as they grapple with a new song, there were a lot of missing components. These components, as Ben Camp states, are critical in the songwriting process: finding a melody that blends well with some lyrics, taking a chord progression and integrating it with multiple instruments, and focusing on each component with intention as it fits with the song's entirety."



isual compents (think of interactive digital component)
↳ If there was a way to connect / listen to
a certain users' "concert mode," then
app could play visuals,

A simple reflection on the complexities of live music was a breakthrough in our research process. While simple, breaking down live music performance into both artist and listener experiences is critical. Music is not a one-way street between an isolated listener and an isolated artist–rather, as American poet Henry Wadsworth Longfellow penned in 1835, "Music is the universal language of mankind." Akin to language, music is rooted in deeper social, cognitive, and psychological contexts that SOTA music AI systems do not account for, particularly because there is no empirical structure for mapping the flow of *creativity* information between artists–using generative music tools or not–and the listeners or audience. Like a conversation in linguistics, music represents a type of conversation or expression between the creator and the recipient of information. Current music AI systems also lack explainability, not being able to relay its algorithmic processes to human. Humans, specifically, experience individual musical events (pitch, timbre, loudness, and timing) and groups of events (chords, voices, phrases), that humans store as timescales and gist and verbatim representations over human development (Pearce, 2023 & Perez Gin, 2024). Our HAI music system addresses:

1. A key music cognition data limitation that both SOTA music generation systems and MIR as a whole face.
2. The lack of creative control of and feedback from music generation models, exacerbated by potential psychological strain of human judgement.

Research Through Design for HAI Music

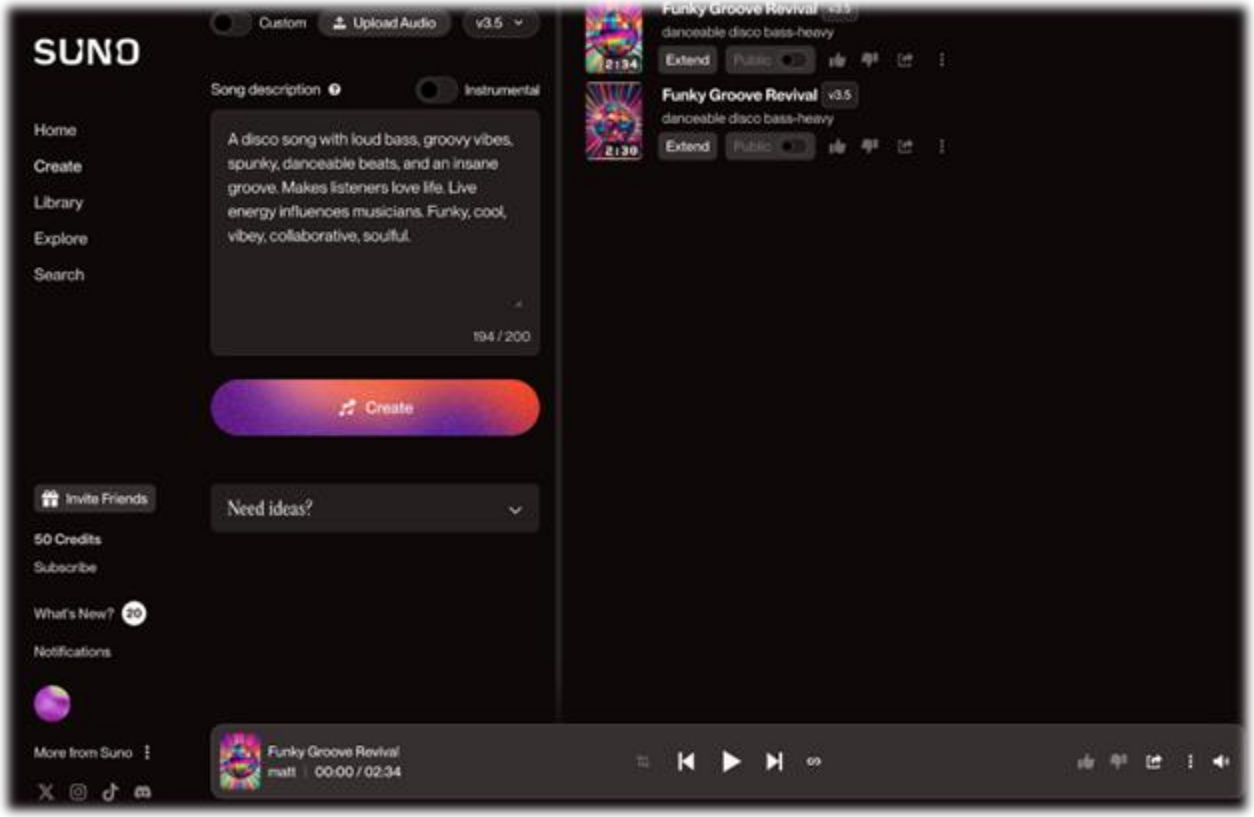
The field study represents the start of a Research Through Design method by immersing myself–an HAIR researcher–in the design process (or first step of the design process). In this context, we observed the interaction between musicians, their "instruments" (AI tools or traditional instruments like guitars, drums, voice, etc.), and listeners. The theoretical potential I focused on was a musician's or performer's ability to engage listeners in their performance. I've also categorized this notion as a *wicked problem*. In other words, there is no formal definition for the problem musicians face when engaging listeners. For one, music is a largely subjective, often "trendy," experience (both in its creation and reception) meaning that although there are ways musicians can anticipate being received well by listeners, we have no guaranteed solution.

We gleaned information about abstraction, as it relates to my system, taking the first step to understanding how to quantify the live musical performing and listening processes for better musical experiences. We can address uncertainty between a musician and an AI music tool through processes of abstraction. For example, a concept like artist-instrument-listener interaction is difficult to quantify. Leveraging the principles of music cognition, we can generalizable knowledge about the largely theoretical concept of engagement between artists and listeners (using AI music tools and not), ultimately to provide context and training data to an improved HAI music system.

> "Competent practitioners usually know more than they can say. They exhibit a kind of knowing-in-practice, most of which is tacit...Indeed, practitioners themselves often reveal a capacity for reflection on their intuitive knowing in the midst of action and sometimes use this capacity to cope with the unique, uncertain, and conflicted situations of practice."
>
> –Schön, Reflective Practitioners (1983).

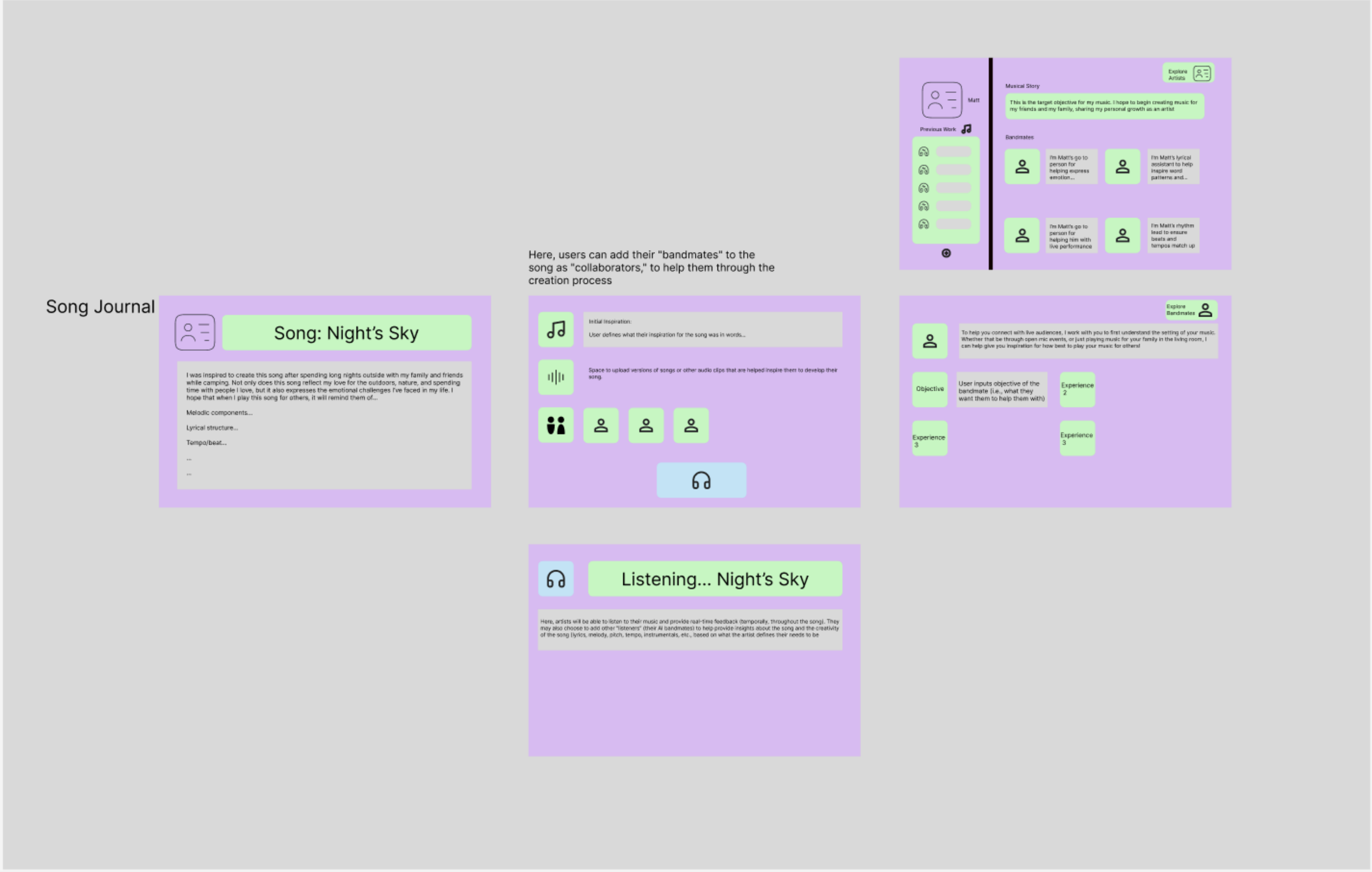
HCI practitioners possess varying levels of intuition about experiences musicians have in creating music. Leveraging tacit musical domain knowledge can help teams adapt to "uncertain, and conflicted situations of practice." In the case of the open mic performance, I was uncertain about the application of AI music tools. For example, during the Suno Field Study, I was uncertain about the listener's reception as well as the emotional experience I would undergo. I was also uncertain about how the AI music system would generate music based on the audience's input (for the song description prompt). Below, we see a snapshot of my Suno experience, where I tested a collective prompt from audience input to generate a song for a performance. While SOTA systems have received much praise–albeit with significant legal backlash from major music labels–for their improvements in recent months, we have identified some fundamental aspects of human-AI interaction that should be accounted for during research and design processes:

1. The reframing of musical interaction between human and AI agent as a relationship between members of a band (building trust with system).
2. The inclusion of artists as a "designer" of their music interaction model (increasing transparency).
3. The enculturation of designers, developers, and users in the music space for tacit music knowledge.



A disco song with loud bass, groovy vibes, spunky, danceable beats, and an insane groove. Makes listeners love life. Live energy influences musicians. Funky, cool, vibey, collaborative, soulful.

Design Artifact 1: AI Bandmates



I was inspired to create this song after spending long nights outside with my family and friends while camping. Not only does this song reflect my love for the outdoors, nature, and spending time with people I love, but it also expresses the emotional challenges I've faced in my life. I hope that when I play this song for others, it will remind them of...

Melodic components...

Lyrical structure...

Tempo/beat...

...

...

Here, users can add their "bandmates" to the song as "collaborators," to help them through the creation process

Initial Inspiration:

User defines what their inspiration for the song was in words...

Space to upload versions of songs or other audio clips that are helped inspire them to develop their song.

This is the target objective for my music. I hope to begin creating music for my friends and my family, sharing my personal growth as an artist

I'm Matt's go to person for helping express emotion...

I'm Matt's lyrical assistant to help inspire word patterns and...

I'm Matt's go to person for helping him with live performance

I'm Matt's rhythm lead to ensure beats and tempos match up

To help you connect with live audiences, I work with you to first understand the setting of your music. Whether that be through open mic events, or just playing music for your family in the living room, I can help give you inspiration for how best to play your music for others!

User inputs objective of the bandmate (i.e., what they want them to help them with)

Here, artists will be able to listen to their music and provide real-time feedback (temporally, throughout the song). They may also choose to add other "listeners" (their AI bandmates) to help provide insights about the song and the creativity of the song (lyrics, melody, pitch, tempo, instrumentals, etc., based on what the artist defines their needs to be

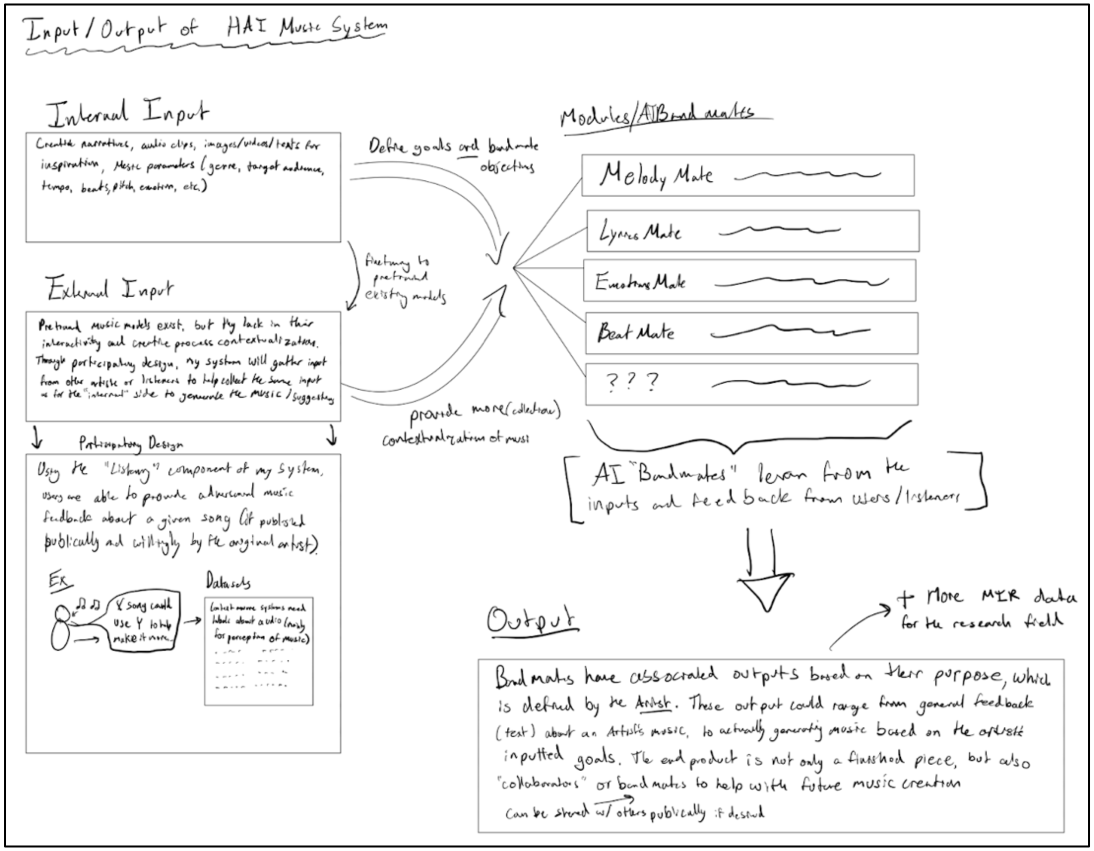
AI bandmates build on the *innate human musical creativity and instrumentality*, while providing a theoretical–through artificially intelligent agents–representation of collaboration and feedback on the creative process. Suh et al., 2024, identified psychological barriers that humans face when composing their own music. With this, our HAI system will provide artists with low-risk musical feedback from their trained AI bandmates, *conversing* with them in real-time to create music that adheres to their objectives and desired expressions. Such a setting enables artists to de-black box complex generative music models, forming a more personal, trust-based system with AI bandmates. Simultaneously, artists will be providing their AI bandmates with critical musical data (see FTT-Music Model Framework) to help the system become enculturated in the users' musical style, which is lacking in current music AI research methods and SOTA music generation models. This system addresses controllability of the model–for both feedback and music generation–and provides better data for designers and developers.

Human-Centered Music AI – Data I/O Design Artifact

When placing music in the context of AI, and vice versa, I have one primary human end-user in mind – the artist. The artists are at the front and the back of my ML pipeline (see I/O figure to the left). While the system is trained on existing available music data, there is a significant lack of music cognition datasets available for use, especially at the level of scalability necessary for effective AI systems. Artists, of course, have different levels of experience and existing musical output. Our HAI music system enables the users–a musical artist–to provide multimodal data inputs (audio clips, text, images, videos) to provide the learning model – in this case, "bandmates" – with specific tasks to help navigate the creative process from start to finish. The artists also play a pivotal role at the end of the ML pipeline. Essentially, artists may "train" their bandmates while also having the opportunity to critique the generative music suggestions after the model has already been provided context (i.e., "final" AI-assisted track or song, suggestions for performance, lyrics, etc.).

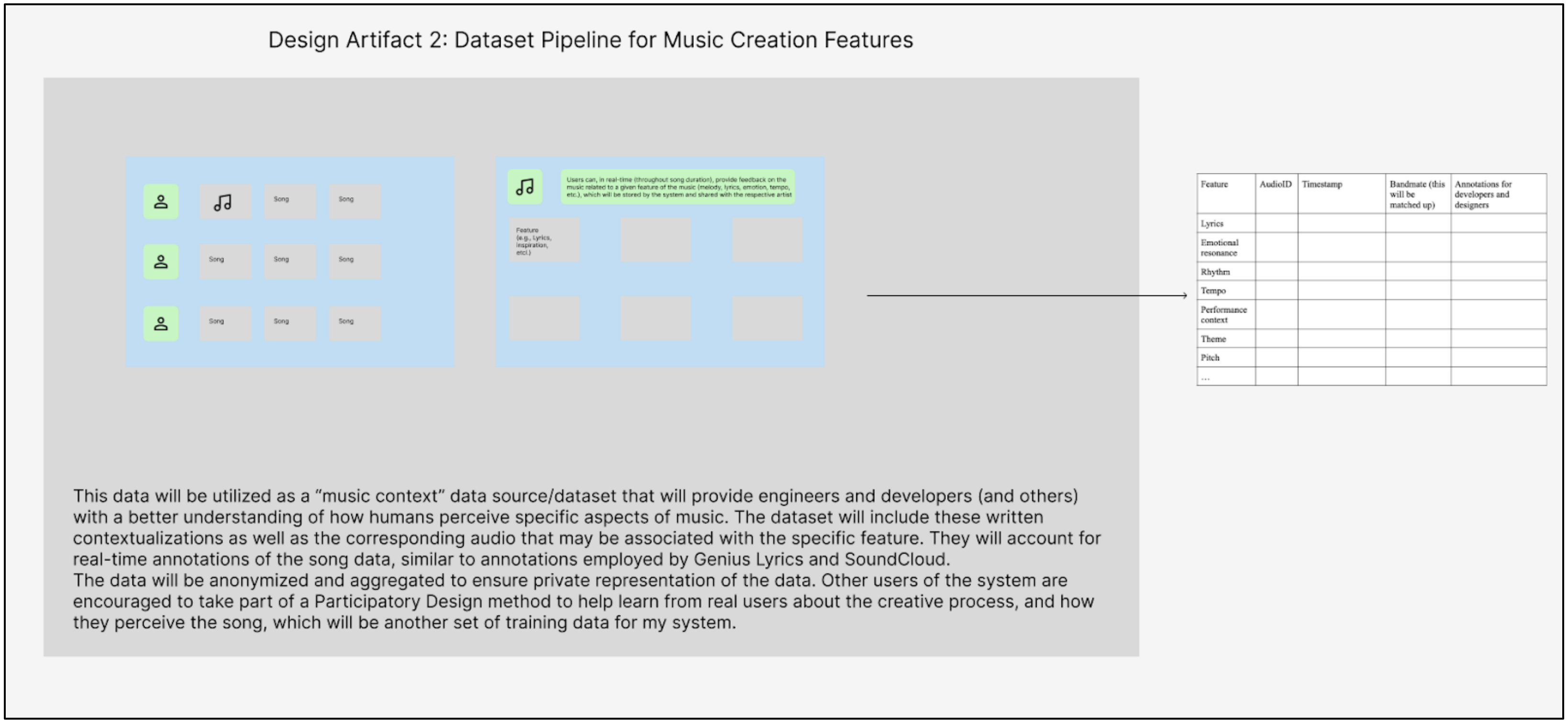
The other key humans in this design are the engineers, designers, and domain experts in music, who work together during the data collection and module building (and finetuning) for the musicians. They are the curators of the music data, ensuring proper measures are taken to provide detailed datasheets and effective design. They work with domain experts (music performance, music technology, etc.) to provide a better system foundation to help musical artists during their creative processes. They may possess some contextual vulnerability, as they may not be musicians themselves, however, the context provided by the users as aggregated data (audio, text descriptions, etc.), can help developers and designers form a better understanding of what end-users/creatives want to gain from the musical process.

In many ways, however, the artists themselves can help build the model for their AI bandmates by providing the input data, the goal of the model/bandmate for the specific, desired, musical objective, and to test the final output performance. This is not to say that listeners are external to our system; rather, they are human algorithm for musical experience to be transferred into quantifiable input data via our HAI music design. We define internal HAI model input as features such as creative narratives (text), audio clips, images, and videos for artist inspiration, and other music parameters such as (genre, target audience, tempo, pitch, emotion, etc.). Pretrained music models exists, but they lack in interactivity, controllability, and explainability of the creative process. Therefore, our model collects external input using participatory design methods to gather "listening" data, or adversarial feedback–as text or multimodal datatypes contextualized with text–from listeners who evaluate songs based on their musical features and cognitive perception.


Input / Output of HAI Music System
Internal Input
Creative narratives, audio clips, images/videos/texts for inspiration, Music parameters (genre, target audience, tempo, beats, pitch, emotion, etc.)
Define goals and bandmate objectives
finetuning to pretrained existing models
External Input
Pretrained music models exist, but they lack in their interactivity and creative process contextualization. Through participatory design, my system will gather input from other artists or listeners to help collect the same input as for the "internal" side to generate the music / suggestions
provide more (collective) contextualization of music
Participatory Design
Using the "Listening" component of my system, users are able to provide adversarial music feedback about a given song (if published publically and willingly by the original artist)
Ex
X song could use Y to help make it more...
Datasets
Context-aware systems need labels about audio (mainly for perception of music)
Modules/AI Bandmates
Melody Mate
Lyrics Mate
Emotions Mate
Beat Mate
? ? ?
AI "Bandmates" learn from the inputs and feedback from users / listeners
Output
+ More MIR data for the research field
Bandmates have associated outputs based on their purpose, which is defined by the Artist. These output could range from general feedback (text) about an Artist's music, to actually generating music based on the artists inputted goals. The end product is not only a finished piece, but also "collaborators" or bandmates to help with future music creation
Can be shared w/ others publically if desired

The input allows the model to learn the goals of the artist, tailoring the AI bandmate objectives while providing more collective contextualization of music. The data is then aligned with the layers or modules–AI Bandmates–of the model based on target features, training each bandmate to provide feedback as text output (like an LLM) or musical output for artists to use or reference in their creation.

We also leverage existing music generation methods such as Frid et al., 2020's work on music generation by example, which is one of the main goals of our system for AI bandmates to learn from artist examples of their music.



Design Artifact 2: Dataset Pipeline for Music Creation Features

Users can, in real-time (throughout song duration), provide feedback on the music related to a given feature of the music (melody, lyrics, emotion, tempo, etc.), which will be stored by the system and shared with the respective artist

Feature	AudioID	Timestamp	Bandmate (this will be matched up)	Annotations for developers and designers
Lyrics				
Emotional resonance				
Rhythm				
Tempo				
Performance context				
Theme				
Pitch				
...				

This data will be utilized as a "music context" data source/dataset that will provide engineers and developers (and others) with a better understanding of how humans perceive specific aspects of music. The dataset will include these written contextualizations as well as the corresponding audio that may be associated with the specific feature. They will account for real-time annotations of the song data, similar to annotations employed by Genius Lyrics and SoundCloud.
The data will be anonymized and aggregated to ensure private representation of the data. Other users of the system are encouraged to take part of a Participatory Design method to help learn from real users about the creative process, and how they perceive the song, which will be another set of training data for my system.

Design Artifact 2: Datasheets for developers and data scientists/HCI practitioners

The participatory design method will help collect additional data related to the contextualization of the music and audio–from a listener's perspective– which also serves as training data to help create the AI bandmates. Artists have a tremendous amount of power in that they can define what kind of bandmate they want to have to collaborate with, and how they would like the AI bandmates to assist them. In music, there is no objective or "true" musical output, and it is highly difficult–perhaps infeasible–to definitively quantify aspects of music such as taste. Therefore, there is no "expert" in the music creation process for this system other than the *artist themselves*. Though, musical educators and therapists can play a pivotal role in helping developers train an AI bandmate to provide easily understandable suggestions for the creative process.

Interaction designers will work with engineers to understand the technological opportunities and capabilities of the AI system. For engineers, this requires the understanding of generative music systems and transformers, as well as natural language processing capabilities to allow an AI bandmate to provide textual analysis, support, and recommendation to a user/artist.

To get a better understanding of music as an art, industry, and creative process, interaction designers may also will work with anthropologists to gain a better understanding of the current ways through which musicians create music and collaborate with other humans to create music. This will require the abstraction and distillation of things like live performances and live, in-person music creation with human band members.

Ensuring proper procedures for collecting music data will help interaction designers and researchers develop generalizable knowledge for music technologies in practice. We predict that music data for better music AI systems will also alleviate many limitations related to fairness in the music industry by re-popularizing music as a creative endeavor, rather than a popularity contest. Furthermore, effective datasheets–again, for users as well as designers and developers–to interpret musical data may help increase the trustworthiness of a HAI music system, allowing artists to take more authorship of their data just as they would a song. A system that allows for participatory design data collection will provide music AI systems with better contextualization and generalizability to broader communities of artists.

Muse it or lose it!

Measuring the effectiveness of a music AI system is challenging, especially one designed to integrate seamlessly into a musician's everyday life in an unremarkable manner. However, this system primarily aims to serve as a creative support, imposing little to no psychological strain—such as nerves or fear of criticism for their music—that might arise from collaborating with other humans or presenting their music to groups. With the help of AI bandmates, the system can shift between being the user's focus and functioning in the background. A user can utilize the system to document their inspiration (requiring little to no AI capabilities), but at any moment, they can introduce a "bandmate" to assist with creating, analyzing, and enhancing their music and creative process.

To make the HAI music system more effective, it is essential for developers and designers to align on music information retrieval data. Complex musical elements—such as emotional expression—can be better understood and conveyed through the use and creation of datasets and datasheets by domain experts (the training data that the bandmates rely on). In some ways, this approach is speculative, as it is not yet feasible to fully emulate a human bandmate's ability to curate and create music. However, incremental steps, such as improving the transfer of music data, can enhance current music generation and recommendation technologies, starting by focusing on the human creative process in music.

Ultimately, the system leverages the theoretical potential of music creation for individual solo artists. Initially, the vision was centered on the broader music community, considering many—if not all—stakeholders in both the live music industry and the digital music industry, particularly digital streaming platforms (DSPs). However, through research and design, we uncovered significant shortcomings in the controllability and feedback of state-of-the-art music generation models, as well as critical data gaps related to music cognition, particularly in live settings.

As a designer, I recognize that I hold subjective judgments about HAI music products. For instance, I have certain predispositions regarding the use of generative music in the creative process. I also realized that I have struggled with my own judgments, particularly concerns about the ethical, fair, and moral implications of generative music.

Re-An Anthropologist's Musical Digression

This is the final excerpt from my "field study" using Suno AI in a live setting, expressing my current state of mind regarding music technologies for human creativity.

"A soundtrack for my life sounds daunting without a person–or people–able to tune in and not just listen, but feel and interact more intimately than sending someone a playlist with "good songs." I find that through understanding the creative process of music in a live setting, tools like Suno may be able to better integrate and find a place in live music environments. Is this what people want right now? After tonight, I want to say maybe not. However, it is not an impossible task. CMG is a perfect setting to introduce Suno's potential to a community that loves live music, and I'm excited to dive deeper. For now, I need to take a step back–play my guitar, sing some songs to my friends on a lawn chair–and learn from other musicians and music lovers about what music truly means to them."

ACKNOWLEDGEMENTS

I thank Qian Yang and the INFO 6940 staff for the support of my research interests in music and dedication to providing an outstanding course in human-AI interaction. I also thank the Cornell community as a whole, including my many mentors and friends, as well as Cornell Media Guild, for their support in my research.

REFERENCES

Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. AI magazine, 35(4), 105-120.

Aramaki, E., Wakamiya, S., Yada, S., & Nakamura, Y. (2022). Natural Language Processing: From Bedside to Everywhere. Yearbook of Medical Informatics, 31(1), 243–253. https://doi.org/10.1055/s-0042-1742510

Bell, D. E., Raiffa, H., & Tversky, A. (1988). Descriptive, normative, and prescriptive interactions in decision making. In D. E. Bell, H. Raiffa, & A. Tversky (Eds.), Decision making: Descriptive, normative, and prescriptive interactions (pp. 9-30). Cambridge University Press.

Besson, M., & Schön, D. (2006). Comparison between language and music. Annals of the New York Academy of Sciences, 930(1), 232–258. https://doi.org/10.1111/j.1749-6632.2001.tb05736.x

Bian, W., Song, Y., Gu, N., Chan, T. Y., Lo, T. T., Li, T. S., ... & Trillo, R. A. (2023, June). MoMusic: A motion-driven human-AI collaborative music composition and performing system. In Proceedings of the AAAI conference on artificial intelligence (Vol. 37, No. 13, pp. 16057-16062).

Blanchard, L., Naseck, P., Egozy, E., & Paradiso, J. A. (2024). Developing Symbiotic Virtuosity: AI-Augmented Musical Instruments and Their Use in Live Music Performances.

Bly, S., & Churchill, E. F. (1999). Design through matchmaking: technology in search of users. interactions, 6(2), 23-31.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience. Advances in Child Development and Behavior, 28, 41-100.

Christiansen, M. H., & Chater, N. (2022). The language game. Basic Books.

Deng, J., Zhang, S., & Ma, J. (2024). Computational copyright: Towards a royalty model for music generative AI. University of Illinois Urbana-Champaign.

Gangopadhyay, N., & Pichler, A. (2024). Embodiment and agency in a digital world. Frontiers in Psychology, 15, 1392949.

Frid, E., Gomes, C., & Jin, Z. (2020, April). Music creation by example. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-13).

Gardner, J., Simon, I., Manilow, E., Hawthorne, C., & Engel, J. (2021). MT3: Multi-task multitrack music transcription. arXiv preprint arXiv:2111.03017.

Gingras, B., Pearce, M. T., Goodchild, M., Dean, R. T., Wiggins, G., & McAdams, S. (2016). Linking melodic expectation to expressive performance timing and perceived musical tension. Journal of Experimental Psychology: Human Perception and Performance, 42(4), 594.

Han, J., Yang, E., & Oh, U. (2024, May). Understanding the Use of AI-Based Audio Generation Models by End-Users. In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (pp. 1-7).

Hanson, D., Storm, F., Huang, W., Krisciunas, V., Darrow, T., Brown, A., ... & Pickrell, A. (2020). SophiaPop: Experiments in Human-AI Collaboration on Popular Music. arXiv preprint arXiv:2011.10363.

Hong, Z., Huang, R., Cheng, X., Wang, Y., Li, R., You, F., ... & Zhang, Z. (2024). Text-to-Song: Towards Controllable Music Generation Incorporating Vocals and Accompaniment. arXiv preprint arXiv:2404.09313.

Huang, C.-Z. A., Dinculescu, M., Vaswani, A., & Eck, D. (2018). Visualizing music self-attention. In Proceedings of the NeurIPS Workshop on Interpretability and Robustness in Audio, Speech, and Language (p. 1).

Huang, C. Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., ... & Eck, D. (2018). Music transformer. arXiv preprint arXiv:1809.04281.

Huang, C. Z. A., Koops, H. V., Newton-Rex, E., Dinculescu, M., & Cai, C. J. (2020). AI song contest: Human-AI co-creation in songwriting. arXiv preprint arXiv:2010.05388.

Huron, D. (2008). Sweet anticipation: Music and the psychology of expectation. MIT Press.

Klarlund, M., Brattico, E., Pearce, M., Wu, Y., Vuust, P., Overgaard, M., & Du, Y. (2023). Worlds apart? Testing the cultural distance hypothesis in music perception of Chinese and Western listeners. Cognition, 235, 105405.

Le, D.-V.-T., Bigo, L., Keller, M., & Herremans, D. (2024). Natural language processing methods for symbolic music generation and information retrieval: A survey. arXiv. https://doi.org/10.48550/arXiv.2402.17467

Maman, B., & Bermano, A. H. (2022, June). Unaligned supervision for automatic music transcription in the wild. In International Conference on Machine Learning (pp. 14918-14934). PMLR.

Odom, W., Zimmerman, J., Davidoff, S., Forlizzi, J., Dey, A. K., & Lee, M. K. (2012, June). A fieldwork of the future with user enactments. In Proceedings of the Designing Interactive Systems Conference (pp. 338-347).

People + AI Guidebook (Google PAIR)

Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. Annals of the New York Academy of Sciences, 1423(1), 378–395. https://doi.org/10.1111/nyas.13654

Pearce, M. (2023, May 24). Music perception. In Oxford Research Encyclopedia of Psychology. Retrieved September 15, 2024, from https://oxfordre.com/psychology/view/10.1093/acrefore/9780190236557.001.0001/acrefore-9780190236557-e-890

Perruchet, P., & Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. Psychonomic Bulletin & Review, 20(2), 310–317. https://doi.org/10.3758/s13423-012-0344-5

Reybrouck, M. (1989). Music and the higher functions of the brain. Journal of New Music Research, 18(1–2), 73–88.

Reyna, V. F. (2012). A new intuitionism: Meaning, memory, and development in fuzzy-trace theory. Judgment and Decision Making, 7(3), 332–359.

Sun, J., Yang, J., Zhou, G., Jin, Y., & Gong, J. (2024, May). Understanding human-AI collaboration in music therapy through co-design with therapists. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1–21).

Suh, M., Youngblom, E., Terry, M., & Cai, C. J. (2021, May). AI as social glue: Uncovering the roles of deep generative AI during social music composition. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1–11).

Sun, J., Yang, J., Zhou, G., Jin, Y., & Gong, J. (2024, May). Understanding Human-AI Collaboration in Music Therapy Through Co-Design with Therapists. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1-21).

Thickstun, J., Hall, D., Donahue, C., & Liang, P. (2023). Anticipatory music transformer. arXiv preprint arXiv:2306.08620.

Uncommon Measure: A Journey Through Music, Performance, and the Science of Time (Hodges, 2022)

Vaswani, A. (2017). Attention is all you need. Advances in Neural Information Processing Systems.

Ventura, M., & Toker, M. (2022). TRBLLmaker: Transformer reads between lyrics lines maker. arXiv preprint arXiv:2212.04917.

Xue, L., Song, K., Wu, D., Tan, X., Zhang, N. L., Qin, T., & Liu, T. Y. (2021). DeepRapper: Neural rap generation with rhyme and rhythm modeling. arXiv preprint arXiv:2107.01875.

Zioga, I., Harrison, P. M., Pearce, M. T., Bhattacharya, J., & Luft, C. D. B. (2020). From learning to creativity: Identifying the behavioural and neural correlates of learning to predict human judgements of musical creativity. NeuroImage, 206, 116311.