THE AI MUSIC PROBLEM

Why Machine Learning Conflicts With Musical Creativity

Christopher W. White

First published 2025

ISBN: 978-1-032-95976-4 (hbk)
ISBN: 978-1-032-95975-7 (pbk)
ISBN: 978-1-003-58741-5 (ebk)

Chapter 1

THE PROBLEMS FACING MUSICAL ARTIFICIAL INTELLIGENCE

DOI: 10.4324/9781003587415-1

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

1

THE PROBLEMS FACING MUSICAL ARTIFICIAL INTELLIGENCE

> "Of course… music is a great difficulty. You see, if one plays good music, people don't listen, and if one plays bad music people don't talk."
>
> *The Importance of Being Earnest*, Oscar Wilde

> "Rabbit's clever," said Pooh thoughtfully.
> "Yes," said Piglet, "Rabbit's clever."
> "And he has Brain."
> "Yes," said Piglet, "Rabbit has Brain."
> There was a long silence.
> "I suppose," said Pooh, "that that's why he never understands anything."
>
> *The House at Pooh Corner*, A. A. Milne

It's late at night, and a computer engineer is scribbling in her notebook. An academic journal has asked her to write about some recent innovations in Artificial Intelligence that she's been developing with her collaborator— what's being dubbed the *Analytical Engine*. Her draft outlines how the technology works, its capabilities, limitations, and potential applications. These notes touch on a huge range of topics, including how to program this new technology, how the machine retains information, and what tasks it's best suited for.

The journal expects a short commentary. Her notes are nearing a sprawling 20,000 words.

DOI: 10.4324/9781003587415-1

She starts writing about the machine's potential. She absentmindedly sips her tea, subconsciously searching for a caffeine jolt, and considers how someone might program the machine to generate new creative content. Her thoughts specifically drift toward whether this technology could write new music. She's played piano since she was young and is well acquainted with basic music theory. Her AI is good at processing well-defined mathematical problems, and if it reduced music composition to a series of formulas, it could certainly generate new music following those blueprints.

She starts writing. If compositional procedures were converted into some series of algorithms, then the computer:

> *might compose elaborate and scientific pieces of music of any degree of complexity or extent.*

She grimaces and puts down her tea. She worries her readers will get the impression that her machine might actually mimic a creative human musician. In her estimation, machines can only follow the pathways that engineers have programmed into them, and can only reshuffle the information they've been fed. Again, she starts writing:

> *The Analytical Engine has no pretentions to…*

She stops, searching for the right words. This AI could, in principle, produce an enormous amount of new music by following the formulas laid out its programming. She looks at her bookshelves full of poetry and novels, and at the paintings covering the walls above her desk. Given the right formulas, her programs will be able to generate enough words and images to overflow all of humanity's walls and bookshelves.

She shudders. This firehose of words and images wouldn't be the same as human-made creativity. She thinks back to music, and to the songs she loves to play and sing. This music isn't made by formulas. It sparks from human ingenuity, playfulness, and imagination. She loves performing and singing with other musicians, and she loves knowing that she's playing music made by other living, breathing humans. Even if her AI creates new and interesting melodies and harmonies, these computer-generated tunes won't really be *music*. They won't be *original*. They won't be *human*.

She finishes her sentence. This AI will never actually:

> *originate anything.*

***

Ada King, the Countess of Lovelace, would publish her notes in *Taylor's Scientific Memoirs* in 1843. Her work described the steam-powered invention of one of her close collaborators, Charles Babbage, an invention that served as a crucial predecessor to the digital computer.

Yet, Lovelace's sentiments feel like they could have been written today. As innovations in 21st century generative AI increasingly elbow their way into our daily lives, we find ourselves grappling with the same issues that furrowed her brow nearly two centuries ago. Can machines create new artistic, expressive, and creative content? If they can, how do they do it? And when do they do, is their resulting content actually *creative*?

It's no coincidence that Lovelace singles out *music* when she writes about computers' capacities and their limitations. In a lot of ways, music should be easy for AI to create. After all, most music follows sets of norms, structures, and formulas that can be easily adapted to a computer's algorithmic mind. But on the other hand, music can also be complex and unexpected. Music can be a social experience. Music can be felt in our bodies, and it can connect two people emotionally. These are all roles that are difficult—if not impossible—for a computer to step into.

This book is about musical AI, and how there's often a mismatch between what AI is good at and what makes music *music*. It's about the history of this field and how musical AI works. It's also about why it is so difficult for AIs to produce music that is compelling and well-made. It's about what we enjoy about music and why it's so hard for an AI to tap into that enjoyment. It's about why a technology that works so well for written language, voice replication, and even self-driving cars often stumbles when learning music's melodies, rhythms, phrases, and forms.

The AI Music Problem tells the story of musical AI by identifying five forces behind generative computer models: 1) the motivations and potential payoffs for creating a musical AI, 2) the availability of reliable musical examples to create data for these AIs, 3) the ways this data is represented to a computer, 4) the kinds of musical structures these machines can detect, and 5) how humans interpret music created by a computer. Throughout, I will provide a tour of contemporary AI's inner workings, comparing its engineering to the ways that humans make and listen to music, and will show how each of these five forces poses particularly difficult challenges for AIs that generate music. The book will also touch on some larger issues that increasingly powerful and intelligent machines pose—like how we define "creativity" and whether an AI can ever really "feel"—and does so through a musical lens.

The book isn't a technical manual. A little familiarity with music will be helpful, but no ability to read complex orchestral scores or fluently play an instrument is required, nor is writing code a prerequisite. But, I also

won't shy away from diving into the nitty gritty of what makes AI tick, or into the specific musical complexities that can prove so difficult for computers to grasp. In short, this book is your window into the music we love and AI's attempt to write it.

Five hurdles for musical AI

Figure 1.1 charts a course through five challenges for musical AI. The gambit begins with *Motivation*. Like any venture, AI's foray into music needs some reason to exist. It needs some obvious payoff, be it academic innovation or the lure of commercial success. It's this initial motivation that fuels the allocation of crucial resources like time, talent, and money. Next up, we hit the problem of *Examples*. AI machines are voracious learners that devour data. For a musical AI to learn to write music that actually sounds like *music*, we need a library of tunes so extensive that it satisfies this computational appetite. Compiling such a library is no small feat.

Once we have chosen and compiled our library of examples, we encounter the craggy landscape of musical *Representation*. Here, the issues become a bit more technical. Before AI can begin to understand musical examples, human engineers need to translate them into a language that computers speak—some format that a computer can read. Essentially, music needs to be cut up and parceled into packets that the computer can process. This is a process of translation, and it is fraught with choices. Will we teach the AI how audio waves work? Or will we ask it to learn the structure of written scores? How do we slice up melodies, describe the length of individual tones, or represent each momentary harmony? Each decision shapes the AI's understanding of music, carving out its abilities and limitations.

Just as written language builds sounds into words, words into sentences, and sentences into ideas and stories, music is organized according to its own levels of *Structure*. Each note and harmony don't behave as simple solitary sounds. Individual notes are like single threads in an expansive Persian rug. The sonic threads gather into themes and harmonies, which in turn weave into larger phrases. Phrases then fall into patterns that warp and weft into larger designs of repetition and contrast, like

FIGURE 1.1 Five factors behind musical AI's development.

verses, choruses, and refrains. Each of music's larger structures is made of smaller, interacting components that themselves are made of tinier atoms. For an AI to truly produce convincing music, it must somehow learn the logic of each of these multifaceted, interwoven levels.

Finally, any successful musical AI needs to produce music that people actually *like*. Music is a communal activity in which we hear and *Interpret* emotion and the human experience. The music we value most has some sort of shared social component. We like music because it's beautiful, but also because we hear something of our own lived experience within it. Far from simply assembling notes, a successful musical AI needs to learn to compose something *human*. This is a steep—if impossible—hill for a machine to climb.

Each of these hurdles does not stand alone. The series works like a row of dominos, with each item pushing forward into the next. Should the motivation behind AI research and development ignite with a roar of resources, we'd be met with the challenge of gathering enough musical examples for AI to learn from. Once we amassed a suitable library, we would need to master the translation of music into formats that AI can effectively digest. Supposing AI were to leap over this hurdle, it would then face that intricate, multi-layered puzzle of musical structure.

But let's imagine for a moment that AI sails over all these hurdles and starts producing music as well-constructed as any human could compose. We're then left with that final—and perhaps tallest—hurdle. Would we, as listeners, ever emotionally connect with AI-generated music? In my estimation, this series of challenges presents an enormous, compounding problem for successful musical AIs. In what follows, I explore each of these categories in a bit more depth and suggest ways that music provides a particularly thorny problem for AI at each juncture.

Motivation

I'll dive deeper into the programming and engineering behind musical AI in subsequent chapters, but to give a quick overview of a few important concepts, we can start with one the hottest buzzwords in contemporary AI research—the *Large Language Model*, or LLM. LLMs are digital engines that study vast arrays of text, images, or music to learn how to produce new content. These programs repeatedly pore over their training datasets, learning a bit more with each iteration, shaping their predictions, and refining their understanding of how words, images, or sounds are organized and arranged in their library of data. In this manner, LLMs build up expectations about what kinds of events follow other events, and what sorts of larger sequences occur before or after other sequences. This web of expectations, norms, and statistics forms the basis of the

model. Here, the word *model* means the combination of the computer program and its trained expectations, and it's in that sense I'll use the word "model" throughout this book. The data from which these models learn is often called *training data*, and it's from this training data that the model captures and internalizes norms, rules, and tendencies. Any model that learns from training data undertakes *machine learning*. In machine learning, the computer learns information directly from a dataset rather than being explicitly taught or preprogrammed by a human. Such a model can then create new content using the norms and tendencies learned from its training data. When commentators or engineers discuss *generative AI*, they are generally referring to some computer model that can create new content in this manner.

Deep learning is another buzzword of 21st-century AI. When we say that some model uses deep learning, we mean that its artificial mind employs several layers in its learning process. It's like learning the steps, hip action, rhythm, and upper-body movement of a dance all at once. Each individual gesture occupies one layer, and all the layers are united into one seamless dance move. To provide a textual example: it's not enough for a deep learning system to know that the word "cherry" often comes before the word "pie." To be effective, the system needs to understand the layers of possible meanings that govern how and where these words are used—categories like "fruit" and "dessert" and "edible items."

You might think of this process as the AI learning the dance steps of a dataset. Within written language, for instance, an AI identifies the linguistic choreography we perform unconsciously when writing words and sentences. Just as certain steps and turns characterize the subtle routines of a tango or salsa, the AI notices what words tend to occur together and the typical ordering of phrases. It then learns how these sequences can be varied and expanded—what words can substitute for others and the ways that words congeal into phrases that express larger ideas. Just as accomplished dancers learn to improvise and elaborate on basic dance steps, the AI eventually learns to write more expansive and varied sentences, and to draw from related topics. A musical AI works in much the same way, internalizing the kinds of melodies that synch with particular harmonies, and how complexes of sounds group together into phrases, forms, and fully composed songs.

In recent years, generative AI models have become increasingly good at creating new and useful content using deep-learning systems. Media headlines practically scream the praises of various types of deep learning every day, reporting how some new "transformer" or "convolutional neural network"—all types of deep learning systems—are creating alarmingly accurate images, texts, or voice replications.

Throughout this book, I'll be discussing models that learn tendencies, norms, and statistics from training data, internalize the data's statistical tendencies, and use them to create new musical content. My goal is to discuss how these models work and how they interact with music in general instead of focusing on the particulars of any specific engineering or implementation. I'll therefore tend towards referring to these models in their general forms like "machine learning" or "deep learning," as opposed to in their more specific incarnations like "LLMs," "transformers," "convolutional neural networks," and the like. In Chapter 2, I'll outline some of the basic programing and mathematics behind these models, focusing on the overall approaches that unite 21st-century computational AI. The chapter will then turn its attention to *why* companies and researchers pour time and resources into these types of models.

From one perspective, the *motivation* behind creating these sorts of models seems self-evident: to satisfy users' desires to create and consume new and interesting content. In 2024, for instance, the well-known chatbot ChatGPT had over 77 million monthly users in the United States relying on its technology to generate and hone text. In other words, roughly 1 in 5 people in the US were using this technology to produce new content every month. However, reading deeper into these headlines can also show how costly these models can be. The newest AI models often come with a hefty price tag, demanding millions, even billions, of dollars in computing power, data wrangling, and sheer electrical muscle to accomplish their tasks. The 2024 version of ChatGPT, for instance, was reported to have cost over $100 million to develop and train. The motivational calculus of generative AI, therefore, hinges not just on what it can possibly create, but how much it costs to develop and create this content.

From my vantage point, the financial allure of studying musical AI is far from evident. Put simply: music plays a less pervasive role in the capitalist marketplace than other forms of media. It therefore offers lesser incentive for the development of AI-driven applications.

Compare the user base and revenue streams of industry-leading companies like Microsoft Office, Adobe, and Ableton Live. Microsoft Office is a colossus in the realm of text and numerical documents, boasting billions in revenue and a user base stretching into the hundreds of millions. Adobe, a titan of the image-based documents sector, also draws revenues in the billions, with tens of millions of active users.

However, Ableton Live, perhaps the most popular digital music-writing program, plays to a much smaller audience. Its revenues are counted in the mere tens of millions of dollars, and it has a community of around a million users. By either metric, it is less than 10% the size of analogous companies in other media.

This imbalance naturally extends beyond money and into the allocation of human capital. Fewer engineers and researchers work on music AI, and large tech companies spend less time developing musical tools than projects in other media. Chapter 2 will explore this industry imbalance and the motivations shaping musical AI's funding, research, and development.

Examples

In the digital era, AI researchers find themselves in the middle of a veritable ocean of data. With every click of a mouse, they can open vast expanses of content neatly translated into the binary language of computers. This digital transformation ensures an almost endless supply of material for AI's learning and development.

Music is no exception to this digital bounty. At least in terms of sheer content, the online world is awash with music. It would take hundreds and hundreds of years to listen to the complete catalogue of a streaming service like Spotify, and websites like IMSLP.org and CPDL.org host musical scores that number in the hundreds of thousands. Yet, the way music is rendered into digital formats poses significant challenges for deep learning models. Digital music is hard for a computer to read, and hard to collate into *examples* for an AI's training data. While our eyes can easily recognize notes on a page and our ears effortlessly pluck notes and chords out of sound waves, these tasks are quite difficult for computers. In Chapter 3, I'll outline some reasons why musical scores pose such a challenge.

The problem of music notation

Consider the gulf between the relative simplicity of navigating a pdf of text and the complexity of interpreting a sheet of Western music. Reading a page of text requires knowledge of the alphabet and punctuation, and the ability to identify the beginning and end of letters, words, and lines. You'll need to be ready for intermittent paragraph indentations, and now and again you'll encounter an image, graph, or picture that you'll need to realize is *not* part of the written language.

Musical scores, by contrast, present a labyrinthine challenge. Working through any selection of music notation (Figures 1.3 and 1.6 in this chapter, for instance) gives an immediate sense of this warren. In Western notation, the five lines of a musical staff serve as the core. Time signatures can indicate how many beats are in a measure and where downbeats occur. The curved forms at the beginning of each staff are called "clefs," and are added to the staff as a key for deciphering exactly which notes

correspond to which lines and spaces. A five-line staff can appear alone, or it can be connected to other staves to indicate multiple musical lines being played concurrently. All the vertically aligned notes across all the connected staves will be played or sung at the same time.

Ovoid noteheads are placed onto the staff's five lines or in the spaces between. The closer notes are to the top of the staff, the higher they sound. To show how long each note lasts, noteheads can be filled in or hollow, and they can have stems—the vertical lines attached to the noteheads—or no stems. Stems can point upward or downward, and they may have flags or bars attached to them. Open noteheads indicate longer notes, while the notes with flags or bars are faster.

I'll stop here, even though there are many other notations connected to phrasing, loudness, instrumentation, and so on. There are many, *many* more aspects of music notation that any music reader must be ready to identify and interpret.

A page of music and a page of written text have many things in common. They are visual representations of suites of information, and we can extract this information by decoding the symbols of the page. But there are simply many more moving parts in musical notation than in text. This heightened complexity significantly amplifies the challenges for a computer attempting to recognize symbols on a score, multiplying the chances for errors during the information extraction process.

The problem of audio

Given the challenges that score-reading presents for computer recognition, why not turn to audio files instead? After all, digital music is readily available online in vast quantities. Unfortunately for musical AI, the task of plucking out individual notes, instruments, and other musical elements from sound waves, while easy for the human brain, proves to be a formidable challenge for computers.

The crux of the problem lies in what acousticians call the *overtone series*, or what musicians sometimes refer to as *partials*. If you sing a tone, the note gains its identity based on how fast your vocal cords make the air vibrate. When we sing a higher note, the sound waves vibrate faster. When we sing lower, the waves are slower. The same principles apply to piano strings, trumpets, drumheads, or any other instrument. Our ears hear different pitches according to how fast the instrument makes the air around it vibrate. The rate of vibration that makes a note sound like a particular pitch is its *fundamental frequency*.

But not all notes with the same fundamental frequency sound the same. Imagine the difference between singing a note with a warm, resonant

"aaaa" vowel and the same note sounded through a nasal, piercing "eeee." Any note with the same fundamental frequency sounds different on a piano, a violin, or a harp for the same reasons. These different tone qualities are rooted in the spectrum of *overtones*. Overtones are softer and higher pitches that color the sound of the fundamental pitch. The more nasal and sharp the sound, the more overtones are active in the tone. Warmer sounds like the theremin and violin have fewer overtones acting above the fundamental frequency, and harsher sounds like the oboe and harmonica have more.

From the computer's perspective, the problem is that overtones and fundamentals are both sound waves, and it's often hard to tell whether a particular sound wave is a fundamental or an overtone. Figure 1.2 illustrates this problem using *spectrograms*, visual representations of sound waves. The horizontal dimension on a spectrogram indicates time passing. The sound starts on the lefthand side of the image and continues as you move right. The vertical axis shows the speed at which sounds vibrate. Higher sounds will be on the top, and lower sounds will be on the bottom of the spectrogram. The lefthand pane in Figure 1.2 shows a single computer-generated version of a low C. Here, the computer is generating a pure pitch with no overtones (what an acoustician would call a sine waveform). Since there are no overtones present, one bright white band streaks horizontally across the bottom of the pane. The middle pane shows a piano playing that same low C. Now the spectrogram shows several overtones stacked on top of the fundamental. The white bands that parallel the low C are like floors of a skyscraper, each representing an individual overtone. In combination, the fundamental pitch and its overtones give the piano its distinctive sound. I've flagged one such overtone in the figure

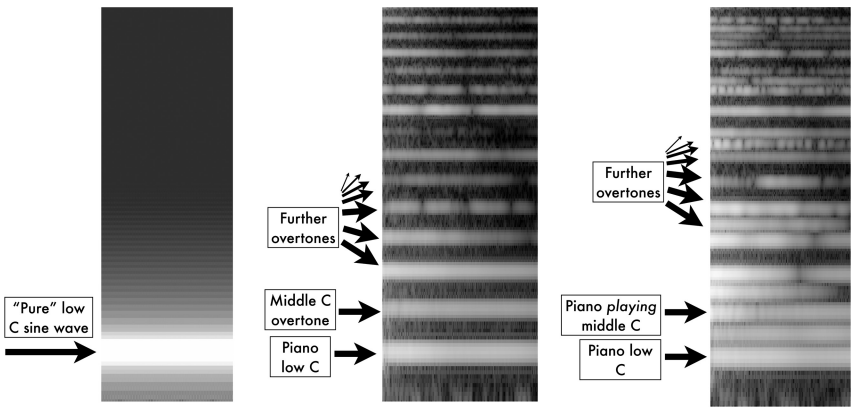

FIGURE 1.2 A spectrogram of the sine wave of low C, a piano playing a single low C, and a piano playing a chord with low C in it.

as the sound wave associated with middle C. When you play a low C on a piano, the overtones for middle C activate along with the fundamental, producing sound waves at the frequency of middle C which are folded into the overall sound. While our ears and brain perceive one single piano note, the envelope of sound waves that build the complex texture of the piano's low C includes this middle C overtone.

The pane at the right in the figure shows another spectrogram, but now of a chord with multiple notes. The lowest note of this chord is low C. You can see that fundamental pitch extending all the way through the graph. Middle C is also played in the chord, and the spectrogram shows a soundwave band at the same frequency we saw for middle C in the previous pane. However, this band now represents a second fundamental frequency, actually played above the low C, rather than an overtone arising from sympathetic vibrations activated by a singular low C. While our ears can easily hear the difference between these two musical situations, computers have a difficult time teasing out whether the middle C is an overtone or the result of an instrument actually playing middle C.

Certainly, the complexities of deciphering sound waves aren't insurmountable. There are tools that extract musical information from audio files. These programs have better success rates when they are primed to know the characteristics of the instruments involved, as this allows the program to anticipate specific overtone patterns. Moreover, some developers simply circumvent the challenge of isolating individual notes from a sound signal. Instead, they train their AI to identify patterns in how sounds are ordered and constructed. In these cases, the AIs learn the expected shapes and contours of a spectrogram rather than any direct information about musical specifics like pitches and rhythms. These strategies and their implications will be explored in Chapter 3.

The relative size of datasets

There's a barrier to entry into musical composition. You can't just pick up a guitar and start playing, and you can't download a digital audio workstation and become a proficient music producer in a single afternoon. Musical fluency typically demands extensive study and practice, and even those who achieve proficiency in playing an instrument may not necessarily excel at composing new music. This high barrier limits the pool of individuals creating music, resulting in musical datasets that tend to be more limited in size compared to other forms of media. For instance, while public domain music repositories like the International Music Score Library Project (IMSLP.org) or the Choral Public Domain Library (CPDL.org) boast impressive collections, the amount of music they have catalogued is dwarfed by the mountains of publicly accessible text offered by sites like

Google Books. Chapter 3 also dives deeper into this challenge, along with various other hurdles that musical AI encounters in its quest for data.

Representation

As an AI sifts through its data, it must parcel it into digestible bites. When engineers build models of visual art, they need to decide whether their AI will dissect the imagery into discrete objects or interpret the pictures pixel by pixel. Similarly, text-based AIs might process their data letter by letter, word by word, or phrase by phrase. A musical AI needs to know whether it's dividing its data into notes, chords, full measures, and so on.

This decision shapes the architecture of an AI's artificial "mind." An AI designed to interpret images pixel by pixel learns the norms and statistics of how those pixels are strung together. Its learning, memory, and generative processes—its *intelligence*—are all anchored in this pixelated perception of images. In contrast, a model that is trained to divide images into recognizable objects populates its intelligence with these entities and the web of relationships between them. These sorts of choices not only dictate how an AI processes information but also how it *represents* the medium in its computational memory.

When the popular chatbot ChatGPT processes text, it does not represent its data as a series of individual letters. Rather, it divides text in words and word components. Figure 1.3 uses vertical lines to show how ChatGPT would parse the sentence, "Generative Pretrained Transformers are super cool."[1] (A GPT, or a "Generative Pretrained Transformers" is the specific kind of deep learning underpinning the chatbot.) Notice how the last four words stand on their own. When ChatGPT learns or produces these words, it considers them individual items and remembers where and when they appear in sentences. The words "Generative" and "Pretrained," however, are split into component parts. The algorithm does not learn where and when to use the word "Pretrained." Instead, it learns how often and in what contexts "Pret" precedes "rained." A model like ChatGPT hones these choices through experimentation, with various tests determining that these *representations* lead the model toward efficient, coherent, and useful outputs.

Music grapples with these same issues. However, due to the many layers of information embedded in any musical moment, there are a multitude

FIGURE 1.3 A sentence divided into the chunks learned by ChatGPT and used in its text generation.

of options for how to divide a musical surface. Consider, for instance, the chord in Figure 1.4 and the veritable avalanche of ways to describe this simple passage. Here, I use two linked staves, indicating that all the notes aligning vertically occur at the same time. Notes that are horizontally displaced are read—and played or sounded—left to right through time. Notes on the upper staff sound higher than the notes in the lower staff. Here, there are four discrete notes in the music corresponding to the four separate parts in a typical choir. From the top of the higher "treble" clef staff to the bottom of the lower "bass" clef staff, the sopranos, altos, tenors, and basses all sing their own notes, with the soprano singing the top note with the upward-pointing stem, the alto singing the note with the downward-pointing stem directly underneath, and so on. The tenor line features a rhythm that is twice as quick as the other voices, as indicated by the bar joining its two notes. The moment begins with the pitches B, G, and D, with the tenor voice changing to an F while the other voices hold their pitches. The notes also occur in specific ranges. Working again from top to bottom, the highest note is a G above middle C in the soprano (top) voice, the alto has the D above middle C, and so on. I've also labeled the *intervals* between the lowest note and the other pitches in the chord. In musical parlance, the distance between B and G is a "6th," because those two letters span the distance of six notes. Western musical practice uses the first seven letters of the alphabet to name its pitches, and moving from B to G, traverses six letters: B→C→D→E→F→G. From this perspective, the chord involves a 6th, a 10th, and a 13th above the lowest note.

But there are yet more ways we can represent this chord. Musicians also often talk about the chord's notes as positions in a scale or key. This chord uses the notes of a C-major scale, which encompasses all the white notes of the piano. These correspond to the seven notes of the musical alphabet,

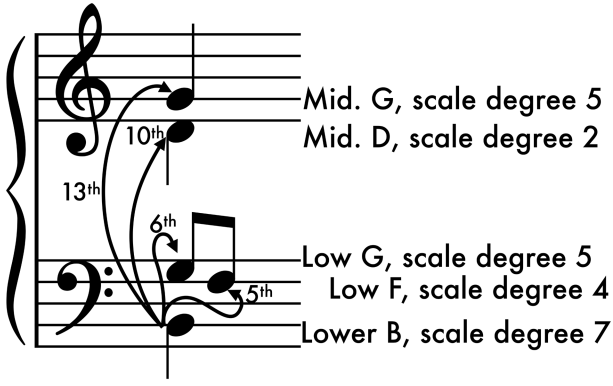

FIGURE 1.4 A chord labeled with several different ways of representing its constituent notes.

i.e., A, B, C, D, E, F, and G. We can also image this chord occurring in the *key* of C major, which means that the pitch C will serve as beginning and end of the scale. The seven letters of the scale are therefore ordered C, D, E, F, G, A, B, and back to C. And, since C serves as the point of origin for the scale, we call it "scale degree 1." At the second position in the scale, D will be scale degree 2. Thinking about the notes in our chords, F occupies the fourth position in the C-major scale, so it is scale degree 4. Similarly, G is scale degree 5, and B—as the last note of the scale—is scale degree 7. The example, therefore, consists of scale degrees 7, 5, 2, and 4.[2]

If all this information feels overwhelming, that's because it is. And there are so many more musical aspects of this moment, including the notes' distance from the previous or next chord, the harmony's relationship to the overarching key, not to mention the notes' rhythms. The problem, then, becomes choosing between this multitude of characteristics when making a concrete, computational representation of some musical chunk. Even with the handful of descriptions I've outlined, which would be the ideal method to represent the events of Figure 1.4 to a computer? Table 1.1 shows several options. Should an AI learn from the information within the "Pitch with height" representation or from the "Lowest pitch with intervals" information? From scale degrees or from note names? And, adding to the already compounding options, I've enclosed in parentheses the data that correspond to the second note sung by the quicker tenor voice. Should this note be integrated into the overall harmony, much like ChatGPT wraps the entire word "Transformer" into one event? Or should the harmony be divided into two separate moments, as the chatbot divides the word "Generative" into two separate components?

There are so many choices that an engineer must make when representing music, and in Chapter 4, I explore these issues in more depth. On the computational end, this large number of options sets up a difficult choice between more general information that is easier for an AI to learn and

TABLE 1.1 Various ways of representing the chord in Figure 1.4.

Characterization	*Representation*
Just note names:	G, B, D, (F)
Just scale degrees:	5, 7, 2, (4)
Names with repeats, ordered from bottom to top:	B, G, (F), D, G
SDs with repeats, ordered from bottom to top:	7, 5, (4), 2, 5
Pitch with height:	Low B, low G, (low F), middle D, middle G
Lowest pitch with intervals:	Low B, 6, (5), 10, 13

more specific, detailed data that is harder to learn but is easier to reassemble into new compositions. For instance, because it erases all information about specific pitches and note names, a scale degree representation is quite general and easy for a computer to learn.

The tradeoff, however, is that a general method of representation like this lacks specificity. Relying on scale degree information alone, for instance, strips away much information about key and exact note placement. If you only told a pianist to play scale degrees 5, 7, and 2, they wouldn't know what key to play in, let alone how many notes to play or how the notes should be arranged. Likewise, a model that trained only on scale degree information would not know how to translate its sophisticated musical knowledge into actual notes in a composition.

Conversely, descriptions like those in the "Pitch with height" or "Lowest pitch with intervals" categories are very specific. As we saw before, armed with this information, a pianist would know exactly which piano keys to push down, and a model equipped with representations of this sort would easily be able to generate notes. However, because these representations are so specific, they often get in the way of a model learning large-scale or general patterns. Engineers, then, need to balance the many options available to them while negotiating between general representations that help an AI learn the music's underlying grammar, rules, and effects with relative ease, and the specific information an AI needs to generate notes, melodies, and chords.

Structure

In April of 2022, OpenAI's Instagram account posted an image generated by Dall-E, the company's visual art LLM. The picture was the result of the prompt, "a teddy bear on a skateboard in Times Square." Dall-E did its job well, with the image realistically depicting a stuffed bear riding a skateboard in front of an out-of-focus midtown tableau. I've recreated the image using my own OpenAI account in Figure 1.5.

Deep-learning models are good at this sort of task, stitching together objects in a logical sequence with simple relationships. After observing many stuffed animals, the algorithm understands how to arrange pixels to represent plush fabric, and it understands how to generate these pixel patterns in the shape of a teddy bear. After similarly creating a skateboard, the algorithm knows how to arrange the two objects because it's seen plenty of things "standing on" other things, some of them skateboards. Additionally, it's seen enough photos backdropped by New York streets and foot traffic to know how to place those objects in Times Square.

Models that learn connections, trends, norms, and statistics within some training data—and this includes deep learning LLMs—tend to be

FIGURE 1.5 DALL-E's image generated from the prompt, "A teddy bear on a skateboard in Times Square."

excellent at identifying what I call *nested determined proximities*. Once the machine knows it's drawing a teddy bear, there are very few options for the kinds of pixels and patterns that occur next to one another. The sorts of pixels that appear *proximate* to one another are highly *determined*. There is a certain predictable regularity in how teddy bear fabric looks. Further decisions about both the fabric, the bear's shape, and its position are *nested*—they all grow out of the larger decision to make a "teddy bear" that is "standing." Similarly, shadows are added based on the position of the primary object relative to some off-screen light source. Throughout, the smallest components combine together to create larger and larger objects, with the global and local details all having clear relationships with one another.

The engineering of these sorts of models allows them to learn nested determined proximities easily. These programs repeatedly move through datasets to notice sequences and connections that happen often, identifying statistically frequent sequential patterns: what items follow what other items, what chunks of items occur adjacent to other chunks. It's simply easier for a program—or, for that matter, a human!—to notice things when they are near to each other than when they are far apart. Additionally, because their learning process is designed to compile the

norms and statistics of some dataset, these models and programs are predisposed to those things that happen a lot: things that are determined. After all, strongly practiced norms are easier to notice than seldom-followed preferences. Finally, these statistical models will favor situations in which they can nest several decisions together into one category. It is easier for their learning processes to chunk together several rules and norms under one umbrella than to remember each individual dictum. Pictures of teddy bears and New York backdrops exhibit all these characteristics. To be sure, many machine-learning models have procedures to deal with events that are far apart, infrequent occurrences, and hard-to-predict contingencies. But, nested, determined, proximities are still the easiest sorts of *structures* that a mathematical and statistical machine can learn.

Music is not like pictures of teddy bears. Figure 1.6 shows the melody of "Amazing Grace," annotated to highlight how the music is organized. The piece is written in the key/scale of C major. The entire hymn is divisible into four smaller phrases, as indicated by the grey boxes. The first three sections end on the note G, while the final section comes to rest on a C, scale degree 1 of the key. The second small phrase (labeled A^2) is a slight variation of the first (A^1). The fourth phrase begins as an exact repetition of the first, but it is cut short on the seventh note. Each "A" phrase contains the same melodic core, with the same notes and rhythms. I show these similarities in the solid black rectangles above the music. If you hum the tune, you'll immediately hear how similar each of these sections sounds.

More granularly, the A-section melody is saturated with rhythms that follow long notes with shorter notes. Every measure begins with a note that lasts two beats, as represented by the open noteheads with stems. These relatively long notes are followed in the third and final beat of each measure with a note lasting one beat (the filled-in noteheads with stems), or with two notes crammed into the one beat (the pairs of filled-in noteheads beamed together). While the end of A^2 departs from the melodic prototype of A^1, it also seems to be responding to the first phrase. As I show with the arrows above the staff, A^1 ends by descending to the low G on which the melody began. And, while A^2 also ends on a G, it does so by ascending to the higher G. This ascent ramps up the overall energy while remaining connected to the initial musical idea.

The B phrase uses new musical material, but material that still feels deeply connected with the preceding A phrases. For instance, the high and low Gs that ended A^1 and A^2 now seem to function as the melody's musical floor and ceiling, with the B phrase using these notes but never venturing past them. The rhythmic palette also expands with more variety while still retaining the long–short feel of the A sections. The first note of

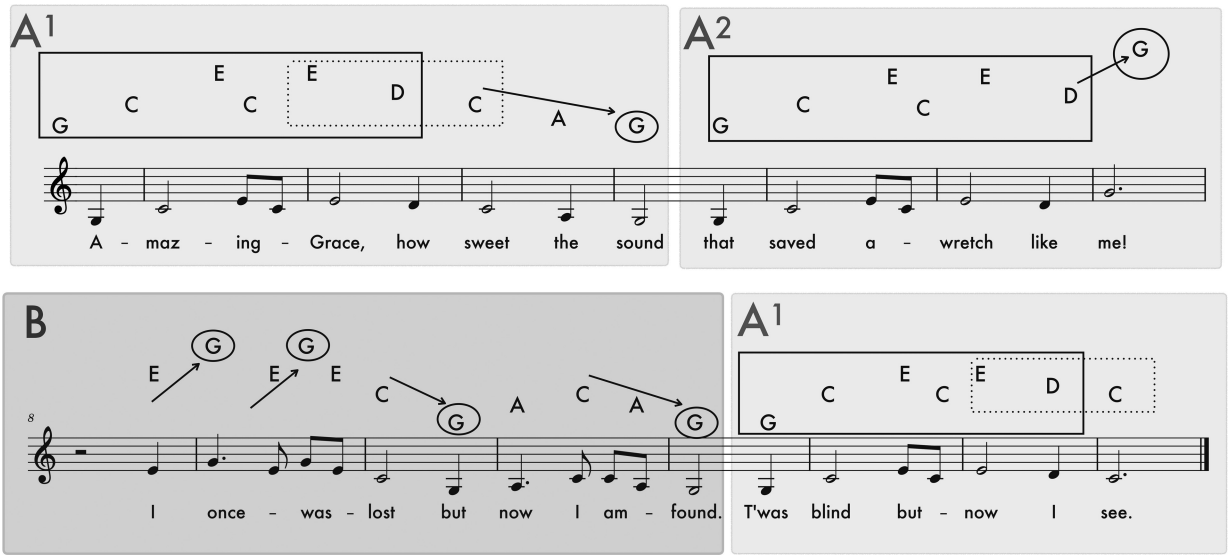

FIGURE 1.6 "Amazing Grace," by John Newton (composed 1772), annotated to show the melody's structure.

each measure remains the longest, though more rhythmic values dot the musical landscape.

The final A^1 section capitalizes on a short melodic idea that was almost hidden in the first phrase, namely, the E–D–C figure enclosed in a dotted box. In the first phrase, those notes were incidental pitches on the pathway to the section's end on low G. However, the last phrase reinterprets this melodic fragment and exploits the fact that it ends on C—scale degree 1 of the key and the first note of the melody's scale—to bring the entire melody to a close.

"Amazing Grace" is not a particularly complicated tune. And yet, very few of its characteristic structural and organizational elements involve immediately *proximate* events. The two uses of E–D–C, for instance, are separated by twelve measures. The low and high Gs that function as the tune's floor and ceiling occur sparsely throughout the hymn and tend to be separated by multiple measures. The return of A in the final measures references the tune's first moments, twelve measures earlier. Furthermore, these connections are not *determined*: just because one musical choice is made doesn't mean the next musical choice *must* be made. The use of a long–short rhythmic pattern in the opening measure didn't mandate its recurrence throughout the phrase, much less the entire tune. In fact, very few melodies replicate the same rhythm with the frequency of this hymn! Also, while many melodies organize their materials in an AABA pattern, there also are many other options for melodic organization. Additionally, most aspects of this musical organization are not *nested*. Unlike the rigid laws of physics that determine how an image's light source casts shadows across a picture, musical decisions are flexible and variable. For instance, some performers start the B section of "Amazing Grace" on a high A, careening through the high-G ceiling. Performing that high A doesn't break any laws or make the music sound incoherent, because the initial range of a melody doesn't dictate that subsequent sections must absolutely conform to the same range.

In other words, even a tune as simple as "Amazing Grace" is built on components that are neither adjacent in time, the result of one common cause, nor invariably present in other hymns. Any learning process favoring *nested determined proximities* would have a very difficult time noticing the structure and organization of this tune alone, let alone within the complexities of a large musical dataset.

In Chapter 5, I will address such issues of music's structural organization using a combination of music analysis, computational programming, and even research from music cognition to argue that musical structure poses special difficulties for AI. Overall, I'll suggest that any AI system

that relies on statistically frequent sequences will stumble when faced with music's variation, complexity, and unpredictability.

Interpretation

Music undisputedly makes people feel things. As a teenager, I was angsty, depressed, and closeted. I would frequently punctuate the end of a school day by squireling myself in my room with my headphones over my ears and blasting a mixture of 90s grunge-pop, Romantic piano concertos, and Andrew Lloyd Weber musicals as a sort of musical therapy. This music helped me process my maturation, sexuality, and interpersonal relationships, and would eventually help me get a handle on my depression.

As far as I can tell, I am far from alone in my relationship with music. People feel emotional connections with music, and music feels like it resonates with and supports the human experience. These are the connections that Ada Lovelace wrestled with when theorizing computational technology in the 19th century, and these relationships contribute to the deep-seated role music has in human society the world over.

For centuries, if not millennia, thinkers have grappled with the unique way that music conveys ideas, and their grapplings could—and do—fill volumes of books on their own. Some approaches, for instance, note that music can remind us of real-world sounds. Think of a descending melody that sounds like a human sadly sighing. These thinkers go on to argue that combinations of these references create webs of associations that result in a meaningful musical experience. Others will focus on bodily associations. Consider how descending melodies make us imagine moving physically downward.[3] But, regardless of the specific theories we use to explain this phenomenon—when we find meaning in music, it's about the human experience. We see ourselves, our bodies, our emotions, and our experiences reflected in the music's notes, melodies, and chords.

Of course, this is not to say that other media *never* rely on information about the human experience. A poem about lost love gains its meaning when the reader identifies with the poet's broken heart. A painting of a scream resonates with a viewer's own experience of screaming and its underlying emotions.

But music is an extreme case. In Chapter 6, we rejoin Ada Lovelace in her study, and use her and other historical figures' ideas to outline a distinction between what I call *associational* and *experiential* knowledge. Someone gains associational knowledge by connecting ideas together. If you read extensively about heartbreak, you will learn all the adjectives and turns of phrase that are most often used to describe that state of desolation,

and you will form associations between those linguistic elements and the idea of the experience. You will have associational knowledge.

On the other hand, if you get your heart broken, you will search for the adjectives and turns of phrase that capture your feelings. In the process, you will gain experiential knowledge. In both situations, you can write a poem about heartbreak, but only the latter will express a lived human experience.

Deep learning models learn by association. By churning through their enormous datasets, chatbots know the adjectives and phrases associated with the concept of heartbreak. By viewing datasets of labeled pictures, image-generating LLMs know the kinds of facial contortions that go into a gut-wrenching scream. But all this is associational knowledge. A chatbot's heart has never been broken, and no image-generating computer model has reacted to a situation by contorting its face into a scream.

In Chapter 6, I show that audiences deeply care about whether the art, literature, and music they consume is made from experiential knowledge. I outline historical sources and psychological experiments that point to the value humans place on content that reflects lived human experience. Because of this, even if AI produces content that looks exactly like something that a human could construct, we will still value it less if we know it wasn't made by a human. Just like the hypothetical poem penned by a person untouched by heartbreak, we will be skeptical of content that doesn't bubble out of experiential knowledge.

I argue that music exhibits an extreme version of this dynamic. Because musical meaning arises from gestures and metaphors connected to human experience, it is fundamentally disconcerting to have that meaning untethered from an actual life. If we love music for its reflection of our humanity, AI will never be able to provide music that we love simply by virtue of the fact that it is not human. Indeed, it's hard to imagine a teenager ever retreating to their room to process their sexuality through music made by an AI, regardless of the technical qualities of that music.

This book's approach

This book is for anyone interested in AI and its creative ability. Musicians worried about AI's capacity to generate content, college students acquainting themselves with the landscape of generative AI, computer programmers wanting to know more about music. It will also speak to anyone interested in the dramatic rise of AI in the last several years. All of these groups (and more!) comprise this book's intended audience. To maintain a clear focus on the arguments and concepts within the text, I consciously limit direct scholarly citations in the meat of the text, but I provide extensive

references at the end of each chapter. I also periodically use the endnotes to dialogue with readers from specialized backgrounds. For instance, if some piece of engineering will be of interest only to readers with coding knowledge, or if part of my music analysis references some important topic in musicology, I will relegate those technical asides to my endnotes.

Additionally, there will be several topics central to 21st-century AI that this book will not immediately address. Legal issues surrounding copyright and plagiarism are outside the bounds of these chapters. I also veer away from the nitty-gritty details of computer programming. While I outline the overall contours of various computational approaches, this book is not designed to teach the mathematics behind machine learning or how to code deep learning networks, nor is it even designed to engage specific types of AI models and their computational implementation. Rather, I aim to discuss broader and general trends in these approaches, and how the logic of statistics and mathematics might misfire when applied to musical composition and expression.

Some larger questions, and music's unique role in AI

Recent commentators have been quick to characterize deep learning LLMs as enigmatic and opaque. As a familiar refrain goes, these models are too complicated for even their programmers to understand. Because of their intricate architectures—the critiques claim—it's difficult to understand *why* these models work. And we're left only to marvel that they *do* work. To boot, media outlets and consumers alike have often been disproportionately focused on evaluating these modes' outputs. They stand agog at what the models are achieving instead of spending their time scrutinizing and understanding the operational mechanisms. Similarly, researchers have been so enamored with the power and potential of deep learning AI that they have spent their time, energy, and resources trying to create better and better results, to the exclusion of studying the models themselves and their roles in art and society.

Several scholars, engineers, and advocates have recognized this oversight and called for a more deliberate approach to AI development. Perhaps most famously, the Future of Life Institute published an open letter in 2023 titled "Pause Giant AI Experiments." It was signed by over 33,000 individuals, including many recognizable names from the tech and computing world. Such critiques advocate for a pivot away from the unyielding pursuit of AI progress towards a slower and deeper investigation into the inner workings of the models and their broader potential consequences for society. Yet, despite calls for caution, the rapid advancement of the

cutting-edge in AI continues at a breakneck pace, driven by the formidable economic rewards and incentives involved.

The idiosyncrasies of musical AI make it an invaluable case study for the larger field. Precisely because it's *already* slow and plagued by frequent hiccups, musical AI is an exemplary subject for scrutiny. After all, a slow-moving machine that constantly breaks down is easier to observe and analyze than a whirling, precise, and flawless apparatus. In a world searching for ways to study the broader ramifications of AI, music emerges as a compelling option.

Technology's sustained role in music and creativity

In 1935, Walter Benjamin wrote an influential essay titled "The Work of Art in the Age of Mechanical Reproduction." In it, he grappled with the then-new technologies of vinyl records and radio. These innovations marked a historic pivot, allowing musical performances to be captured and infinitely reproduced for the first time in human history. Prior to this technological leap, experiencing music was confined to ephemeral live performances. If you wanted to hear a piece of music, you needed a musician to perform it or you had to perform it yourself. By the time Benjamin wrote this essay, music played in London could not only be broadcast live to Moscow, San Francisco, and Kolkata, it could also be recorded, preserved to be heard decades later.

Despite almost a century having elapsed since the publication of Benjamin's essay and the incredible moves beyond the technological landscape it addresses, his commentary remains surprisingly relevant. Anticipating debates over nearly every subsequent technological innovation in music creation and distribution, Benjamin argued that there's an intrinsic value in the live transmission of music from performer to listener, a moment, he argued, could never be fully replicated by any mechanical means. Engaging with the innovations immediately surrounding him, Benjamin delved into fundamental issues of how music is constructed, how it functions in society, and how technological changes can run headlong into shifts in musical aesthetics. The enduring relevance of Benjamin's arguments in the 21st century shows how many of these topics transcend any specific era or context. The technology of music may have changed, but the issues Benjamin addressed continue to resonate.

I hope, in some small way, to undertake a similar project with this book. Here in the third decade of the 21st-century, AI—musical or otherwise—is dramatically new and current. I can easily imagine that, fifty years from now, the artistic worries, moral catastrophes, and intellectual hand-wringing that mark our current discussions of AI will seem quaint. But just as Benjamin's engagement with emerging technology fostered an enduring

analysis of the nature of music in society, so might thoughtful engagement with state-of-the-art AI have the potential to address larger questions about music. Worries about AI have pervaded creative discourse since the advent of the computer, and the effect of technology on music and musicians reaches back even to the invention of the printing press. This book's arguments are situated within the legacy of these conversations, and it aims to use the issues of our time to analyze music, in general, more deeply.

Notes

1 platform.openai.com/tokenizer.
2 I'm spilling a good bit of ink onto the concepts of key and scale degree because they are foundational to Western musical composition. For instance, scale degrees can quickly capture the basic identity of any given melody or harmony. Let's imagine you're singing some tune—let's say "Amazing Grace"—in some range that's comfortable for your voice—perhaps C major. If you ask your friend to sing the same tune but they have a higher voice than yours, they might sing the tune in some higher key—let's say F major. However, regardless of the key you and your friend use for your respective versions of "Amazing Grace," it will still be recognized as the same tune. This is because the relationships between the notes remain stable, and the melody consists of the same scale degrees. When talking about general ways that music behaves, composers and music critics often use scale degree information to describe its actions and emotive effects. Scale degree 1, for instance, acts as a key's most stable pitch and is often described as something of a "home base" for a melody. In contrast, scale degree 7 evokes instability, and it often feels like it is being pulled homeward toward scale degree 1.
3 For some examples, see Cook (2001), Cox (2016), Meyer (1956), Palfy (2022), Langer (1942), Kivy (1980), and Hatten (1994).

References and Further Reading

Allied Market Research. 2023. *Voice Cloning Market Size, Share, Competitive Landscape and Trend Analysis Report by Component, by Deployment Mode, by Application, by Industry Vertical: Global Opportunity Analysis and Industry Forecast, 2023-2032*. Report Code: A05513. July 2023. https://www.alliedmarketresearch.com/voice-cloning-market.

Agawu, V. K. 1991. *Playing with Signs: A Semiotic Interpretation of Classic Music*. Princeton: Princeton University Press.

Alfaro-Contreras, M., J. M. Iñesta, and J. Calvo-Zaragoza. 2023. "Optical Music Recognition for Homophonic Scores with Neural Networks and Synthetic Music Generation." *International Journal of Multimedia Information Retrieval* 12(12). https://doi.org/10.1007/s13735-023-00278-5

Bengio, Y., S. Russell, E. Musk, S. Wozniak, and Y. N. Harari. 2023. "Pause Giant AI Experiments: An Open Letter." Future of Life Institute. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Benjamin, W. 1968. "The Work of Art in the Age of Mechanical Reproduction." In *Illuminations*, edited by Hannah Arendt, 214–218. New York: Schocken Books. Original work published 1935.

Blair, E. 2023. "Grimes Says She's Created A 'Digital Voice' to Sing Her Songs so She Doesn't Have To." NPR, April 24, 2023. https://www.npr.org/2023/04/24/1171738670/grimes-ai-songs-voice.

Braguinski, N. 2022. *Mathematical Music: From Antiquity to Music AI*. New York: Routledge.

Christian, B. 2020. *The Alignment Problem: Machine Learning and Human Values*. New York: W. W. Norton.

Chu, Y., and P. Liu. 2023. "Public Aversion Against ChatGPT in Creative Fields?" *The Innovation* 4(4): 100449.

Clarke, L. 2023. "ChatGPT Is Pretty Bad at Poetry, According to Poets." *Vice*. https://www.vice.com/en/article/7kx9d9/chatgpt-is-pretty-bad-at-poetry-according-to-poets.

Cohn, R. 1992. "The Autonomy of Motives in Schenkerian Accounts of Tonal Music." *Music Theory Spectrum* 14(2): 150–170.

Coker, W. 1972. *Music and Meaning: A Theoretical Introduction to Musical Aesthetics*. New York: Free Press.

Cook, N. 2001. "Theorizing Musical Meaning." *Music Theory Spectrum* 23(2): 170–195.

Cosme-Clifford, N., J. Symons, K. Kapoor, and C. White. 2023. "Musicological Interpretability in Generative Transformers." *Proceedings of the 4th International Symposium on the Internet of Sounds in Pisa, Italy*. https://ieeexplore.ieee.org/xpl/conhome/10335168/proceeding.

Cox, A. 2016. *Music and Embodied Cognition: Listening, Moving, Feeling, and Thinking*. Bloomington: Indiana University Press. https://doi.org/10.2307/j.ctt200610s.

Drott, E. A. 2021. "Copyright, Compensation, and Commons in the Music AI Industry." *Creative Industries Journal* 14(2): 190–207.

Gertner, J. 2023. "Wikipedia's Moment of Truth." *New York Times*. https://www.nytimes.com/2023/07/18/magazine/wikipedia-ai-chatgpt.html.

Gotham, M. R. H., K. Song, N. Böhlefeld, and A. Elgammal. 2023. "Beethoven X: Es könnte sein! (It could be!)." In *Proceedings of the 3rd Conference on AI Music Creativity, AIMC*. https://aimusiccreativity.org/2022-aimc/.

Grand View Research. *AI Voice Cloning Market Size, Share & Trends Analysis Report By Component (Software, Service), By Deployment (On-premises, Cloud), By Application (Gaming, Advertising), By Vertical, By Region, And Segment Forecasts, 2023–2030*. Report ID: GVR-4-68040-083-1. Electronic (PDF), 100 pages.

Hanslick, E. (1854) 1986. *On the Musically Beautiful: A Contribution Towards the Revision of the Aesthetics of Music (Vom Musikalisch-Schönen)*. Translated by Geoffrey Payzant. Indianapolis: Hackett.

Hatten, R. S. 1994. *Musical Meaning in Beethoven: Markedness, Correlation, and Interpretation*. Bloomington: Indiana University Press.

Huang, C. A., C. Hawthorne, A. Roberts, M. Dinculescu, J. Wexler, L. Hong, and J. Howcroft. 2019. "The Bach Doodle: Approachable Music Composition with Machine Learning at Scale." In *Proceedings of the 20th International Society for Music Information Retrieval Conference*, 100–107. Delft, The Netherlands: ISMIR. https://arxiv.org/abs/1907.06637.

Karpathy, A. 2023. "Let's Build GPT: From Scratch, in Code, Spelled Out. [Video]. YouTube. https://www.youtube.com/watch?v=kCc8FmEb1nY.

Kivy, P. 1980. *The Corded Shell: Reflections on Musical Expression*. Princeton: Princeton University Press.

Krumhansl, C. L. 1990. *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.

Langer, S. 1942. *Philosophy in a New Key*. Cambridge, MA: Harvard University Press.
Long, H., and R. J. So. 2016. "Literary Pattern Recognition: Modernism Between Close Reading and Machine Learning." *Critical Inquiry* 42(2): 235–267.
Metz, R. 2024. "The AI Music Era Is Here. Not Everyone Is a Fan," *Bloomberg*, accessed June 4, 2024. https://www.bloomberg.com/articles/ai-music-era.
Meyer, L. B. 1956. *Emotion and Meaning in Music*. Chicago: University of Chicago Press.
Palfy, C. S. 2022. *Musical Agency and the Social Listener*. New York: Routledge.
Qian, J. 2022. "Research on Artificial Intelligence Technology of Virtual Reality Teaching Method in Digital Media Art Creation." *Journal of Internet Technology* 23(1): 125–132.
Rebelo, A., I. Fujinaga, F. Paszkiewicz, A. R. S. Marcal, C. Guedes, and J. S. Cardoso. 2012. "Optical Music Recognition: State-of-the-Art and Open Issues." *International Journal of Multimedia Information Retrieval* 1, 173–190. https://doi.org/10.1007/s13735-012-0004-6.
Roberts, A., Y. Mann, J. Engel, and C. Radebaugh. 2023. Magenta Studio. https://magenta.tensorflow.org/.
Rohrmeier, M. 2022. "On Creativity, Music's AI Completeness, and Four Challenges for Artificial Musical Creativity." *Transactions of the International Society for Music Information Retrieval* 5(1): 50–66. https://doi.org/10.5334/tismir.104.
Saint-Dizier, P. 2020. "Music and Artificial Intelligence." In *A Guided Tour of Artificial Intelligence Research*, edited by P. Marquis, O. Papini, and H. Prade. Cham: Springer. https://doi.org/10.1007/978-3-030-06170-8_16.
Samual, S. 2023. The Case for Slowing Down AI. *Vox*. https://www.vox.com/the-highlight/23621198/artificial-intelligence-chatgpt-openai-existential-risk-china-ai-safety-technology.
Sharma, G., K. Umapathy, and S. Krishnan. 2020. Trends in Audio Signal Feature Extraction Methods. *Applied Acoustics* 158, 107020. https://doi.org/10.1016/j.apacoust.2019.107020.
Shang, M., and H. Sun. 2020. Study on the New Models of Music Industry in the Era of AI and Blockchain. In *2020 3rd International Conference on Smart BlockChain (SmartBlock)*, 63–68. Zhengzhou, China.
Sörbom, G. 1994. "Aristotle on Music as Representation." *The Journal of Aesthetics and Art Criticism* 52(1): 37–46. https://doi.org/10.2307/431583.
Tegmark, M. 2017. *Life 3.0 Being Human in the Age of Artificial Intelligence*. New York: Knopf.
Tigre Moura, F., and C. Maw. 2021. "Artificial Intelligence Became Beethoven: How Do Listeners and Music Professionals Perceive Artificially Composed Music?" *Journal of Consumer Marketing* 38(2): 137–146.
Thompson, L., and D. Mimno. 2023. Humanities and Human-Centered Machine Learning. Working paper.
Tomlinson, Gary. 2015. *A Million Years of Music: The Emergence of Human Modernity*. New York: Zone Books.
Turing, A. M. 1950. "Computing Machinery and Intelligence." *Mind* 59(236): 433–460. https://doi.org/10.1093/mind/LIX.236.433.
Tymoczko, D. 2011. *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. New York: Oxford University Press.
Webster, P. 2002. "Historical Perspectives on Technology and Music." *Music Educators Journal* 89(1): 38–43. https://doi-org.silk.library.umass.edu/10.2307/3399883.

White, C. 2022. *The Music in the Data: Corpus Analysis, Music Analysis, and Tonal Traditions*. New York: Routledge.
White, C. 2023. Artificial Intelligence Can't Reproduce the Wonders of Original Human Creativity. *Chicago Tribune* [Op-Ed]. https://www.chicagotribune.com/opinion/commentary/ct-opinion-artificial-intelligence-human-creativity-chatgpt-20230112-mmoxqjgqtfaibgr663lsohtq34-story.html.
White, C., and M. Kozak. 2023. We Need AI Labels on Creative Content — But Not for the Reasons You Think. *Chicago Tribune* [Op-Ed]. https://www.chicagotribune.com/opinion/commentary/ct-opinions-artificial-intelligence-ai-creative-content-labels-20230603-ak26god46bhbfb2ookbqprnfce-story.html.
Wolfram, S. 2023. What Is ChatGPT Doing … and Why Does It Work? https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/.