Artificial intelligence

Use of AI in the different disciplines of music production

By:
Daniel Heinemann (CS3)
Matr. no.: 44219

Lecturer: Prof. Dr. Koch - Aktuelle Themen (253504)

1 Introduction

The rapid advancement of Artificial Intelligence (AI) has permeated nearly every facet of our daily lives, revolutionizing various industries such as automotive, healthcare, and finance. Even in the realm of music production, an inherently creative and deeply human endeavor, AI's emergence has unlocked a lot of new opportunities and challenges.

This paper delves into the multifaceted applications of AI throughout the diverse stages of music production, like composition, arrangement, mixing, and mastering.

The following chapters are modeled in their order on a possible, but also widespread sequence of steps in music production. It starts with the step of creating a basic idea and instrumental for the track and ends with a complete and marketable song. Since music production consists of many different steps, the goal here is to give the reader an overview of the possible applications of artificial intelligence during the various steps.

2 Technical foundations

When it comes to developing AI-based music, a major challenge is translating music into a format that can be understood by machine learning models. Since these models perceive information as numerical vectors, the music must be represented as a series of numerical tokens that convey details about rhythm, notes, tempo, and other relevant factors that make up the song. These tokens serve as a representation of the music that can be processed by the AI system. [1]

One concrete approach of training the model is to use MIDI- files (*Musical Instrument Digital Interface*) as the "numerical vector", which are widely used in the electronic music industry for transferring music data between digital devices. MIDI files are structured files that contain ordered information about notes, rhythm changes, BPM (beats per minute), and other related data points. This data can be treated as a kind of natural language representation to train the model. [1]

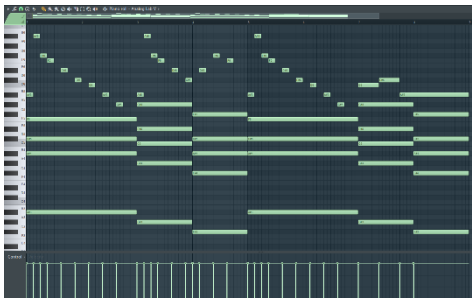

Figure 1: Visualization of a MIDI-File in FL Studio 20

Many machine learning algorithms also use raw audio data as an input at each time step. These inputs are typically represented as sequential input vectors, similar to those used in natural language processing (NLP). This allows the model to predict the next token in the sequence at each time step. But there is a limit to this technique because the audio-sequence of a full song can be very long. Even with CD-Quality (44kHz 16Bit) a normal length song would have millions of time steps. To actually work with those steps from start to finish of a song, the machine learning model would have to manage very long reaching dependencies, which would require a lot of computing power. [2]

One way to deal with this, is using Autoencoders which compress the input information and try to reproduce it correctly based on the reduced information in the output. Although this technique enables the neural network to process smaller compressed information, it often leads to a drawback with the

decoded output typically suffering from a reduction in resolution. [2]

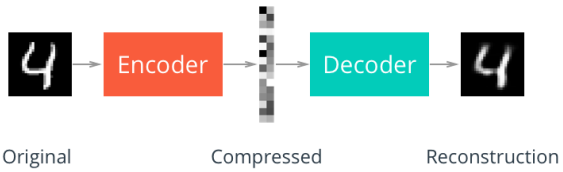

Figure 2: Basic functionality of an autoencoder

3 Initial idea (Melody)

When producing a song from scratch, you usually start by creating an instrumental. One of the essential tasks in the creating process is to think of a basic concept and create a melody, drum pattern and overall harmony.

When talking about assisting or even replacing this process with AI, it usually comes down to the generation of chord, melody or drum progressions based on the analysis of extensive musical data. This data can include existing songs, sheet music, midis or other musical recordings. The AI can extract information from this data about musical elements such as pitch, rhythm, harmony and melodic progressions. By learning this information, the AI-Systems are able to recognize patterns and reproduce them in some kind of new and different form. [3]

In general, there are several technical approaches to creating melodies (or musical patterns in general) with AI. One approach is the use of neural networks. A neural network is a computational model composed of interconnected nodes, or "neurons," arranged in layers. Each neuron receives input, applies a mathematical transformation, and passes the output to the next layer. By adjusting the connections and weights between neurons during training, neural networks can learn complex patterns, like there are in music and make predictions or decisions based on the input data. [3]

Generally, these networks can be trained with different architectures such as recurrent neural networks (RNNs) or generative adversarial networks (GANs).

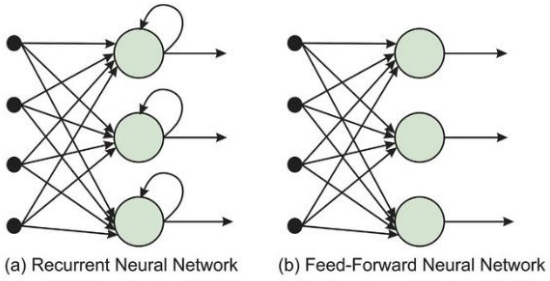

Figure 3: Neural Networks

Another popular approach for a machine learning algorithm is the Markov chain. With this principle, a following note can be predicted on the basis of the current one, but previous steps are not taken into account, instead they are "forgotten". (Similar to a Feed Forward Neural Network) Since the concept of a song must be seen in the totality of the length of the song, this approach can be very limiting. [3]

Examples: SoundRaw, Soundful, Musenet

4 Sound design

When working on music in a digital environment, such as a Digital Audio Workstation (DAW), the creation of the basic melody (usually in midi form) is followed by the sound design. Here, the musician must decide how the notes or the programmed drums should sound and with which instrument (synthesizers, piano, drums, etc.) they should be played with.

Especially in the field of sound synthesis and the associated generation of artificial or altered sounds, there are many and varied examples of AI applications, such as the generation of voices and sounds based on videos or text to speech (TTS) software. In the field of sound design in DAWs, AI-Systems especially help in the generation of presets (patches) rather than generating completely new sounds from scratch.

Figure 4: Synth1 Vst

Traditionally, sound synthesis would involve manual parameter tweaking, waveform manipulation, and a lot of trial-and-error processes. However, AI-based synthesizers can generate complex sounds by learning from existing audio examples or patches and analyzing desired sonic characteristics. One way of creating presets with AI would be transcribing the presets into a numeric format with each tweakable parameter inside the Vst having an assigned value. By doing this with already existing databases of presets you can then have the model be trained with that data to generate new unique sounds. [4]

Examples: Thispatchdoesnotexist, Nsynth

5 Arrangement

Song arrangement refers to the organization and structure of musical elements within a composition. It involves decisions regarding the order and placement of verses, choruses, bridges, and instrumental sections. A well-crafted arrangement enhances the overall impact and flow of a song, balancing dynamics, instrumentation, and vocal performances to create a cohesive and engaging musical experience.

One of the key contributions of AI in song arrangement is its ability to analyze and interpret musical data. By using typical machine learning algorithms, AI can understand the, in most popular genres, templatized structure, melodic content, and rhythmic elements of a song. This allows AI to identify patterns, detect musical sections (such as verses, choruses, or bridges), and provide insights into the overall composition. AI can also assist in generating arrangement suggestions based on established musical conventions and popular trends. [5]

Figure 5: Arrangement in FL Studio

While this can bring significant time saving advantages, especially for beginners, it is still essential to strike a balance between automation and artistic intent. Musicians and producers remain the driving force behind the creative decisions, as most AI tools here are meant to support and inspire the user rather than dictate the artistic direction.

Example: Mvsep

6 Toplines

Especially in modern hip hop and pop production, it is common to first create the instrumental and then build the voice based on it. Since it is very difficult or often almost impossible to find a suitable text and melody on the spot, toplines are usually recorded first. The melody and the basic idea of the vocals are in the foreground and not the lyrics. Usually a few takes of toplines are recorded with a singer and then put together in the DAW.

One of the primary ways AI contributes to topline creation is through melody generation based on the instrumental midis. This is the same principle as described in the first chapter. AI can identify common melodic motifs, chord sequences, and rhythmic elements present in popular songs across various genres. This deep understanding of musical conventions empowers AI to suggest melodic ideas that resonate with the intended style or mood, acting as a valuable source of inspiration for songwriters.

With speech-generating systems, the generated melodic ideas could also be converted into an actual "speech recording" that could be used for demonstration purposes in studio sessions, among other things.

7 Lyrics

Lyrics in music are the written words that convey the message or narrative of a song. They are the poetic and expressive element of the composition, often based on a melody and rhythm, and serve to communicate emotions, stories, ideas, or personal experiences to the listener.

Since some form of language is almost always necessary for humans to communicate and describe information, there are a lot of language-based AI algorithms (e.g. Large Language Models - LLM) with one of the most popular ones being ChatGPT. Language models like LLMs work by analyzing large amounts of text data to learn patterns and relationships between words. They use this knowledge to predict the most probable next word or sequence of words based on the given context. LLMs accomplish this by utilizing neural networks that assign probabilities to different words or phrases, allowing them to generate coherent and contextually appropriate text based on the input provided. [6]

AI-powered lyric generation begins with training a model on a vast dataset of songs from various or specific genres and eras. This enables the AI to grasp the complex nuances of language, rhyme schemes, and storytelling elements of the given lyrics. Once a system is trained, the AI can generate new lyrics by synthesizing learned and detected patterns, while exploring diverse word choices to produce cohesive and contextually relevant lyrical outputs.

Examples: TheseLyricsDoNotExist, Freshbots, Mathigatti

8 Voice

In recent years, AI has made remarkable advancements in synthesizing realistic voices, opening up new possibilities for incorporating this technology in music production. Through the utilization of deep learning techniques, AI models are trained extensively on abundant datasets consisting of recorded human speech or in this instance human singing. This enables them to produce synthetic voices that exhibit a high level of resemblance to actual individuals. These AI-generated voices can mimic accents, intonations, and even capture subtle nuances in

emotional expression. For music production that can mean different things.

On the one side you can use the ai speech generation to create the entire vocals of a track from scratch (or at least based on “artificial” voices) by putting in your lyrics and a melody reference and having them translated into a seamless auditive form. Once the acoustic model is trained, it can generate speech waveforms by taking text input and predicting the corresponding acoustic features. These predicted features are then transformed back into a time-domain waveform using techniques like vocoding or signal synthesis. [7]

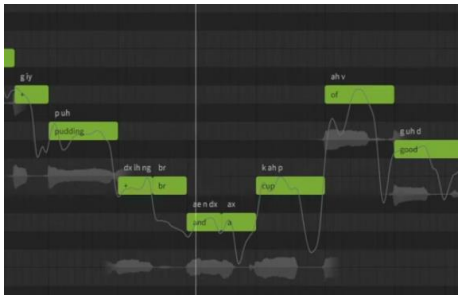

Figure 4: Synthesizer V UI

On the other side, you can use the same technology to recreate the voice characteristics of other artists, which can be useful for creating alternative versions of songs or remixing. However, using AI to replicate the voice characteristics of other artists raises significant legal challenges and problems. Copyright and intellectual property laws come into play, as artists typically have exclusive rights to their recorded performances. Copying an artist's voice without permission could infringe upon their rights and potentially lead to legal disputes. Proper licensing, consent, and contractual agreements would be necessary to navigate these issues and ensure compliance with intellectual property regulations. Furthermore, ethical concerns may arise regarding the boundaries of artistic expression, identity and the potential for misappropriation or unauthorized use of an artist's voice.

Examples: Synthsizer V, Voicemod, Uberduck

9 Mix & Master

The goal of mixing and mastering a track is to make it sound marketable and similar to the "industry standard". Roughly speaking, the volume levels of the different instruments and the voice have to be adjusted to each other and certain frequency ranges have to be assigned to certain elements of the song. Filtering out and "cutting away" disturbing frequencies with the help of so-called equalizers (EQ), among other things, is also part of this discipline.

Figure 5: FabFilter Pro Q Vst

Systems that implement this process with artificial intelligence actually work very similarly to a human being. Based on as much reference data as possible, decisions are made in this process, which usually have the goal of getting the sound as close as possible to the appropriate reference songs. In most software solutions the genre and sub-genre categorization of the song to be mixed/mastered can be defined by the provider or is automatically detected. [8]

It Is Important to mention here th”t th’re are many automated mix & mastering tools that are not necessarily based on a

neural network, but rather search for certain predefined parameters (e.g. integrated loudness) in a song to be processed and manipulate them according to predefined rules. So they "only" follow a predefined algorithm. [8]

Examples: Soundcloud, iZotope Neutron 4, AI Mastering

10 Discussion & Conclusion

In numerous articles, AI has been called the future of music, leading to discussions about the potential obsolescence of musicians and artists due to AI's ability to handle various aspects of music creation. While AI is undoubtedly capable of creating some form of "art", depending on the definition, it cannot entirely remove the artist from the equation yet. AI can assume certain tasks traditionally performed by musicians, and there are even all-in-one tools that generate complete AI-generated songs from start to finish. However, if talking about AI replacing the artist, the perspective on this matter is subjective.

For example, from the standpoint of someone seeking a background song for a film project, who lacks musical skills, AI-generated songs provide a cost-effective and suitable alternative to commission a "real" artists work. However, the purpose of the music in this context can be viewed differently compared to "Songs". In this case it may merely serve as a stylistic element to underscore the mood of a particular scene just like a filter for the visual output would. Thus, the dimension and context of the entire artwork is bigger than just the music, so the artistic aspect is still lying primarily with the filmmaker. Here, the AI-generated song is more or less just a tool or applied effect.

On the other hand, from the perspective of a musician, the objective is usually to create standalone music aligned with their own vision and imagination. The aim is not to rely on having AI generate hundreds of songs or endlessly writing prompts until they match the desired sound. The artist's unique creativity and personal experience remains essential in the process. Also, the process of creating is playing a part in what's considered the art or the product. An analogy could be drawn to paintings and the invention of cameras, just because you can take accurate photos of something the artform of a human drawn painting didn't die out.

From the standpoint of a music consumer, one may think its unimportant whether the music was created by a human or a machine. While this may be true to a certain degree, especially in today's social media era, the artist behind the songs holds significant importance, almost on par with the music itself. AI has yet to replicate this relatable and authentic human connection and its questionable if it ever will.

Human beings continue to serve as the driving force behind music creation, as it is ultimately humans who consume and appreciate the music and shape the associated social structures that are intertwined with music. Whether making significant decisions about the overall sound of a song or minor choices like generating VST-presets or chord progressions, humans provide the essential creative input and "steer" the process into the desired direction.

In summary, while AI can be viewed as a valuable tool in the music creation process, it does not replace producers and artists yet. The artist's vision and direction remain pivotal, as they guide the creative process and determine the

desired outcome. AI can serve as a valuable assistant in addressing areas where artists may lack expertise or face financial constraints, such as the processes of mixing and mastering or having the help of songwriters.

It can be concluded that AI might not replace artists yet, rather their tasks in the creative process will shift, as well as the economic viability of certain working fields.

Sources

[1]

Kaushik Pal. How can an AI Model create Music?. 2023 https://www.techopedia.com/how-can-an-ai-model-create-music [accessed on 30.06.2023]

[2]

Tim Phillip. OpenAI: Jukebox Ein neuronales Netz, das Musik erzeugt. 2022. https://ai.hdm-stuttgart.de/downloads/student-white-paper/Winter-2122/OpenAI_Jukebox.pdf [accessed on 30.06.2023]

[3]

Miguel Civit. A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends, 2022. https://www.sciencedirect.com/science/article/pii/S0957417422013537 [accessed on 30.06.2023]

[4]

James Skripchuk. How I Made an AI Generate Synth1 Presets. 2020. https://jamesskripchuk.com/TPDNE/ [accessed on 30.06.2023]

[5]

Adam Loving. AI for Music Arrangement. 2021. https://adamloving.com/2021/01/08/ai-for-music-arrangement/ [accessed on 30.06.2023]

[6]

Harry Guinnes. How does ChatGPT work?. 2023 https://zapier.com/blog/how-does-chatgpt-work/ [accessed on 30.06.2023]

[7]

Ethan Baker. AI Evolution: The Future of Text-to-Speech Synthesis. 2023 https://www.veritonevoice.com/blog/future-of-text-to-speech-synthesis/ [accessed on 30.06.2023]

[8]

Ian Steward. AI & Automated Mastering: What to Know. 2023. https://www.izotope.com/en/learn/ai-mastering.html [accessed on 30.06.2023]