Consulting Group at Virginia Tech

AI & Generative Music Industry Report and Business Analysis

Project Manager: Caleb McGinn
Consultants: Chandler Anderson, Vidhi Nangia
Analysts: Aditya Nallamala, Eli Bullock-Papa, Marco Ferrara
Account Manager: Joey Mecherikunnel

21st April, 2023

---

1. Executive Summary

This report sets out to analyze AI and Generative Art technologies to understand the business potential they may have. Recently, AI technologies have seen an explosive growth mainly due to advances in cloud computing and processing power of computers. About 50% of businesses have begun to adopt a form of AI to support their processes, generally in three forms: automating business processes, gaining insight through data analysis, and engaging with customers and employees. Development of AI that can engage and interact with unique human prompts has become a large area of research and has given rise to software like ChatGPT, Stable Diffusion, Midjourney, and much more. Our research is focused on AI that can respond to these unique prompts, specifically with generative art and music.

After a wide analysis of AI in general, we analyzed generative music platforms such as Jukebox, Aimi, and Boomy. We discover that generative music platforms fall into two general categories: platforms for end users or platforms for creators. While these platforms offer unique functionality such as continuing music, providing variations to music, or generating completely new music, the capabilities of the AIs are limited on these platforms to explicit functions that do not provide significant creative support. While these are businesses with their own technology, there are few open source options to build upon; we theorize possible simple business models for generative music, as well as discuss possible unintended consequences of these technologies.

While our paper lacks sufficient technical and legal insight into building these models, we believe our analysis supports the ascertainment that the next step of generative music capabilities should be to create one that can either take text or MIDI data as input and produce unique variations, continue it, or generate completely original music - all in one model.

2. AI Industry Overview

2a. General Analysis

The AI industry has been growing rapidly over the past couple of years, with an increasing number of companies and organizations investing in AI technology to improve their products and services. With some examples, including customer service, HR hiring practices, and, more recently, generative AI touching the surface level, approximately 50% of businesses today utilize some form of AI in at least one business unit or function. According to a report by McKinsey, AI is leading to an estimated $301.8B global market revenue by 2028.

The current major trends in AI can be accredited to the advancements in neural networks, deep learning, and machine learning. We can see this in our day-to-day lives in the forms of speech and facial recognition, predictive analytics, and natural language processing. Additionally, the huge increase in data collection has played a role in the acceleration of progress made in AI. Today, you can see AI pop up in practically every industry, from healthcare to finance to consumer product services. Overall, with the proposed growth of AI in both existing and new industries, its use cases improve and evolve every day.

2b. PESTEL Analysis

A PESTEL analysis was conducted to better understand the general environment of the AI industry as a whole.

Political

Regarding political factors, AI has become a major threat regarding policies. However, it can assist governmental institutions in various ways. For example, a company called PimEyes utilizes AI and facial recognition to find all images of a person on the internet by simply using a reference photo. This software was meant for public use but is also used by governmental organizations to track people down and identify people in incriminating photos. Deepfakes also somewhat go hand in hand with this, as some people may utilize AI to generate images with another person's face on them falsely. With this technology, someone can create highly incriminating images, which has led to political unrest and issues in the past, which in turn causes issues in the news with fake news.

Economic

Due to the recent economic downturn, available capital for research and development in AI has decreased. This has slowed down the possible advancements that can be made. However, companies that are solely dedicated to AI are still able

to stay above and continue making advancements, simply at a slower rate than before.

Social

Socially, AI is something that has become normalized. In the past years, AI has become increasingly popular, and this is largely due to the fact that AI is now highly available to the general public. Some examples include Siri, ChatGPT, Alexa, and Dall-E. For example, you can see the rapid improvements in Siri from when it first came out and what it is like now. The general public is beginning to get adjusted to having AI at arm's length nowadays, which increases the demand for such hyper-personalized AI products. This is why the market is now starting to see increased personalization for customers when interacting with products online.

Technological

There are a myriad of technological impacts on the AI industry, as the two parts are intertwined. Such advances that have allowed this is the increased understanding of neural networks and machine learning, which has increased AI capabilities. The advances in technology have helped AI increase automation capabilities for repetitive tasks that require some human interaction. Ai is even being utilized for customer service to answer quick, frequently asked questions before allowing the customer to interact with a human customer service representative. This has allowed industries to reallocate human capital to less repetitive jobs and towards more strategic roles. Additionally, advancements in technology are moving into the field of arts. Generative AI is growing in music and images, which were previously perceived as something that only humans could do. While the end products are not necessarily the same quality as what a human can do, it shows great potential for growth in the future.

Environmental

The environmental impact of AI is quite low, as it is mostly software based. However, it still does require the usage of GPUs, which take up space in centers, often placed in a data center. These data centers need to be built somewhere, so land that can be utilized for environmental good is instead used for advancements in AI. Additionally, training a single AI model can cause high carbon emissions, up to five cars in their entire lifetime. This is a serious implication as AI models are trained every day, meaning that the increase of AI models trained also increases the amount of carbon emitted.

Legal

Some companies have been able to develop generative music AI software but ensure that it is copyright-free music, as there are very few legal protections for both the company and the user regarding generative AI. This is also due to the fact that AI challenges many of the intellectual property laws currently in place. Additionally,

AI chatbots are being generated to try and replace lawyers as a way to provide legal advice and aid to those who are unable to understand legal jargon like a lawyer would.

2c. Trends and Recent Developments

Throughout the past five years, artificial intelligence has expanded both its scope and accessibility. Four major approaches to AI we identified are data-centric AI, model-centric AI, applications-centric AI, and human-centric AI. Data-centric focuses on the collection, management, and analysis of data that is used to create more accurate AI models. An example of data-centric AI can be found in healthcare, with AI being used to identify disease in MRI scans after being trained by previous medical scans. Model-centric AI is an approach that focuses on creating the best machine learning model for a specific dataset. An example of this can be found in the self-driving vehicle space, where they have used model-centric AI to navigate the road safely and expand into Reinforcement Learning and Deep Learning. Application-centric AI focuses on directly solving problems within a specific application, which is generally found in businesses and industries that would like to solve logistical or supply chain challenges. Human-centric AI stands out from the rest as the approach learns from human input and collaboration, which we will focus on in this report. One major example we found is the popular DALLE-2, which we will expand on in a later section.

AI has increased its relevancy over the past few years as costs have significantly decreased and accessibility for the general public has been easier than before. Two key drivers are the development of cloud-based platforms and open-source AI frameworks and libraries.The cloud-based platforms allow developers to access powerful AI models through a user interface and can even use pre-trained models to quickly build their own AI applications. Open-source AI has also helped AI become more accessible because developers are able to communicate and collaborate with each other, which makes it more attainable to create refined and reliable AI systems. One major example of this is ChatGPT, which has skyrocketed in popularity since its release in 2022. So much so that Bing has introduced a new search engine that utilizes ChatGPT's technology so users feel like they are having a conversation with their search engine. As a result of this increased accessibility, AI has been making its way into the art and entertainment space, and musicians and artists have been utilizing the technology.

As of year end 2022, the global market revenue for AI was 136.6B with an expected 29.4% compound annual growth rate. At least 50% of organizations adopted AI in one business unit or function.

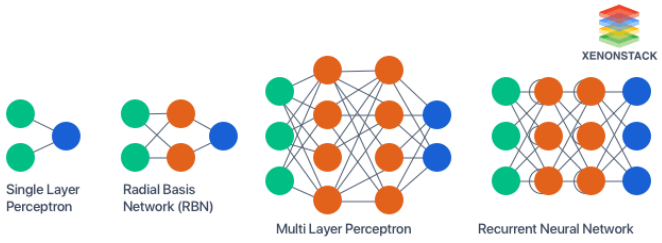
The most significant recent development was the introduction of the transformer based neural network architecture by Google researchers in 2017. This new architecture allowed for models to be trained by thousands of GPUs at once, and this kickstarted a chain reaction of advancements in creating neural networks that deal with sequential data very well. In the past, the best AI architectures such as CNNs were only very good at dealing with spatial data such as images. Additionally, several companies (such as Nvidia, Google, and Cerebras) started

developing hardware specifically designed for training and running neural networks. This dramatically lowered the cost of development. Then, a few years ago people began combining a different model architecture called “diffusion models” with giant transformer models that understood written text. This led to a huge explosion in generative AI such as Dall-E, MusicLM, Stable Diffusion, and Midjourney.

3. Introduction to Key Terms and Processes

3a. Neural Networks:

Neural networks are computer programs designed to mimic the connections and behaviors of neurons in biological brains. There are many different ways of modeling these ‘digital neurons’, but the most common include Convolutional Neural Networks (CNN) which are good at dealing with images, Autoencoders that are good at compressing data, and the newly discovered Transformer and Diffusion model architectures.



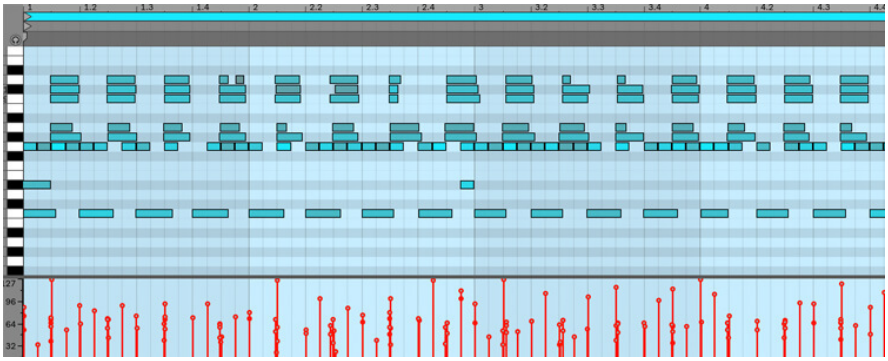
[image of different neural network architectures]

Almost all of these architectures rely on two basic concepts: neurons that process data, and parameters (think: digital synapse) that define the connections between neurons. These two concepts are much simpler than how biological neurons work, but they are still capable of learning many of the same tasks.

To train neural networks, a technique called back-propagation is used to assign the optimal ‘weights’ to the parameters/connections between the neurons. During training, a ‘reward function’ is used by the model to determine how good or bad its output was, and then improve itself to give higher-scoring outputs.

3b. MIDI:

Midi, short for Musical Instrument Digital Interface, is a standardized protocol that enables electronic musical instruments, computers, and other devices to communicate and synchronize with each other. First introduced in the early 1980s, MIDI has become the backbone of modern music production, enabling artists and producers to create complex compositions using a wide array of digital instruments and sound sources.

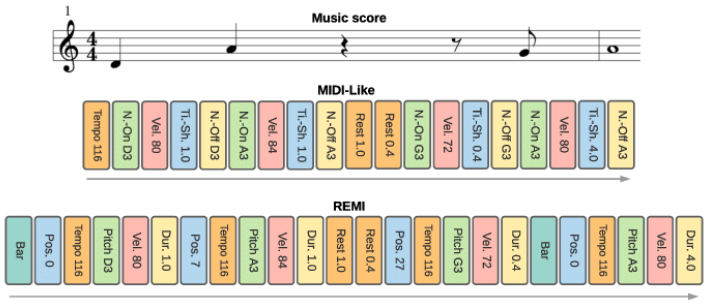


[image of MIDI track with velocity, timing, and pitch information]

The MIDI protocol consists of a series of messages that are transmitted between devices, such as notes, control changes, and timing information. This allows for precise control over various aspects of a musical performance, such as pitch, duration, dynamics, and modulation. MIDI data is typically stored in files, which can be easily edited and manipulated in digital audio workstations (DAWs) or other music software.

3c. Tokens:

Tokens are the smallest units of information that are used as building blocks in natural language processing (NLP) and other machine learning tasks. In the context of language models like GPT, tokens can represent individual characters, words, or even phrases, depending on the granularity of the model.



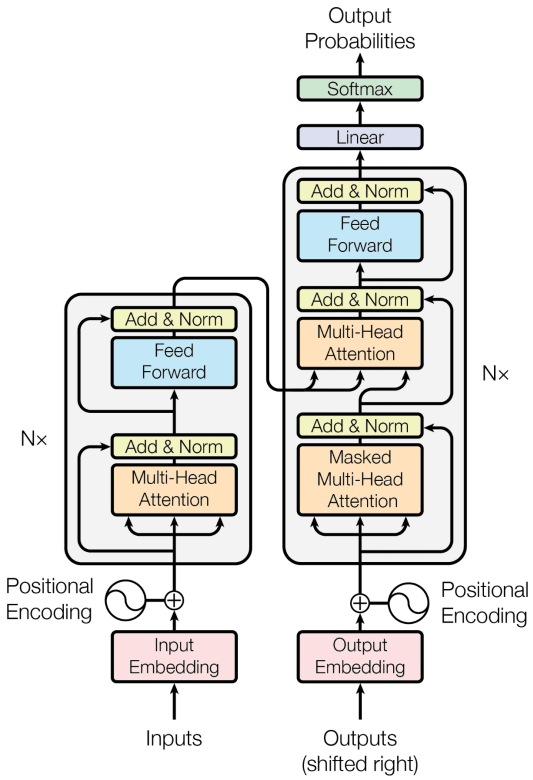
[image of sheet music being tokenized]

Tokenization is the process of breaking down a given input text into these smaller units, which can then be fed into a machine learning model for processing. By converting text into tokens, language models can more easily learn patterns and relationships between words or phrases, ultimately allowing them to generate coherent and contextually relevant responses.

In the case of the music models that we will be discussing, there are many tokenization options. Some of the most common include converting midi to text and then using a text tokenizer, using a chord tokenizer that has a token for every possible note or chord, or using the midi information for each note as a token.

3d. Transformer based Neural Network:

Transformer-based neural networks are a type of architecture designed for handling sequential data, particularly in natural language processing tasks. Introduced by Vaswani et al. in 2017, transformers have since become the foundation for many state-of-the-art models, such as BERT and GPT4.



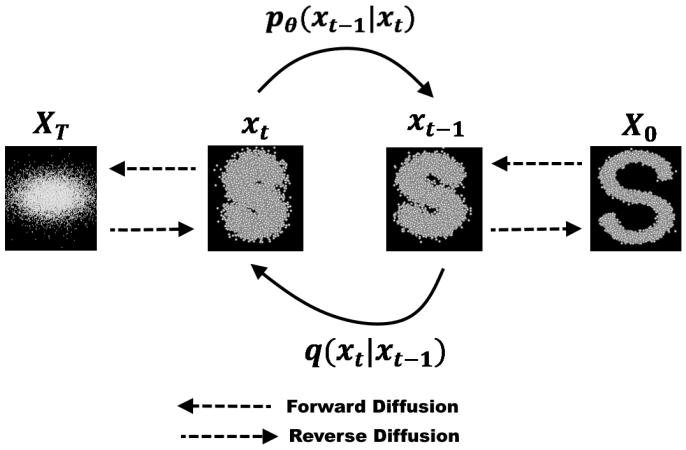
[image of transformer architecture]

These networks are extremely complex and difficult to understand, but the key innovation was the self-attention mechanism, which allows the model to weigh the importance of different input tokens relative to each other. This allows for more efficient and effective processing of long sequences, making transformers particularly well-suited for tasks like machine translation, text summarization, and sentiment analysis.

In the case of music models, transformer based neural networks are well suited for dealing with MIDI tokens - especially predicting future midi tokens of a song when given previous tokens as an input. They are very poorly suited to generating raw audio without the help of decoders or diffusion models.

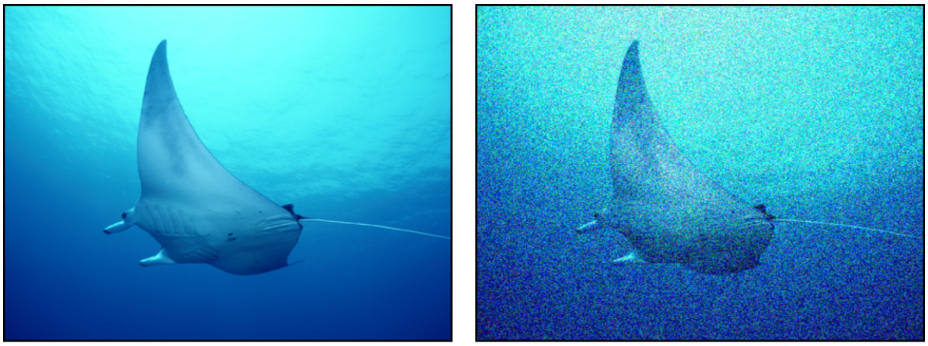
3e. Diffusion models:

Diffusion models have gained attention in recent years for their impressive performance in tasks such as image creation (Dall-E, Midjourney, Stable Diffusion), denoising, and inpainting. They also excel at generating raw audio when paired with several transformer-based networks, but this audio-generation is extremely time and resource intensive.



[image of diffusion process]

In a diffusion model, the training data is generated by continually adding noise or static to input data (often images). This creates a large dataset with pairs of images - those with less static/noise, and those with more of it. Below, you can see an example of one of these image pairs.



[Example of a noise-addition step]

During training, the model learns to reverse this diffusion process. In other words, the model takes in a 'noisy' image, and spits out an image with less noise. That process is done over and over again until the output appears to have no noise left. By learning the distribution of the noise at each step, diffusion models can generate high-quality samples that closely resemble the training data.

By themselves, these models cannot take into account a description of what they are supposed to create. In order to perform tasks such as creating an image/song from a description, diffusion models have to be combined with other AI models (often Transformer-based models).

3f. Creating a MIDI prediction model:

A MIDI prediction model would take in an english text description, optionally some starting midi notes, and then output a song that created a song that closely aligns with the text description. Here is a broad overview of the steps you would need to take to create a model like this:

1. Gather a MIDI dataset

This could be done a variety of ways, such as by downloading public datasets such as the MAESTRO dataset, or even creating your own using audio to midi tools.

2. Augment MIDI dataset with text descriptions

This could be done by scraping articles about each song, or even hiring people to write descriptions of each song in the dataset. Not all songs would need descriptions, but enough would need them for the model to learn the associations between English and music. Also, there are a few datasets with midi songs paired with descriptions that could get you started.

3. Choose your model's attributes

In this step you'll be creating a transformer-based neural network to predict the MIDI tokens given an input of an English description and optionally a few starting notes. However, you'll also need to make important decisions about how large your model should be, how many attention-heads there will be, and how long your context length should be.

4. Train your model

In this step, you'll use hardware like an NVIDIA H100 on the high end, or a desktop GPU on the low end to train your model. Given a sufficient dataset, model attributes, and the time for many hundreds of model generations, the model's accuracy should increase to the point where it would be useful to musicians or hobbyists.

5. Deploy your model

Once you have a working MIDI prediction model, you'll need to host an instance of it on a cloud provider such as AWS or HuggingFace. This way your website developer can integrate it into an interactive website, or musicians can use the model themselves to make music!

3g. Creating an image-generation model:

This is a high-level overview of the steps that would be required to produce a model capable of producing unique images from text input (like Stable Diffusion or Midjourney)

1. Gather an image dataset:

Collect a large dataset of diverse images that cover various subjects and styles. You can use publicly available datasets like ImageNet, COCO, or OpenImages, or create your own by scraping images from the web, following proper licensing and copyright guidelines.

2. Augment the image dataset with text descriptions:

For each image in your dataset, associate it with a textual description that accurately describes its content. This can be done by scraping related captions or metadata from the web, using pre-existing datasets with paired image-text data, or hiring human annotators to write descriptions for each image.

3. Choose your model architecture:

Select an appropriate model architecture for the image generation task. This could be a combination of a transformer-based neural network for processing the text input and a diffusion model for generating the images. You will need to decide on the size of the model, the number of layers, attention heads, and other hyperparameters based on your computational resources and desired output quality.

4. Preprocess the data:

Tokenize both the text descriptions and the image data to prepare them for input into the model. This may involve converting images to lower resolutions or specific formats, and tokenizing the text descriptions into appropriate tokens that can be processed by the transformer network.

5. Train your model:

Using appropriate hardware, such as high-end GPUs or specialized accelerators, train your model on the prepared dataset. This may involve several iterations and fine-tuning to achieve optimal performance. Monitor the model's loss and other metrics during training to ensure it is learning effectively.

6. Evaluate and refine your model:

Once the training is complete, evaluate the performance of your model by generating images from text inputs and comparing them to ground truth data. This may involve using qualitative assessments, such as human evaluations, or quantitative metrics, such as Frechet Inception Distance (FID) or Inception Score (IS). Based on the evaluation results, refine your model architecture or training process as needed to improve performance.

7. Deploy your model:

After obtaining a satisfactory image-generation model, deploy it on a cloud provider, such as AWS or Google Cloud, or use a model hosting service like HuggingFace. This will enable users to access the model through an API or integrate it into interactive applications, such as websites or mobile apps.

4. Current State of AI and Generative Art

4a. Capabilities and Limitations of AI in Business

As AI continues to rapidly develop it is important to acknowledge its current capabilities, as well as its possibilities in the future. Currently AI has been integrated into business processes in several ways that help streamline processes such as process automation, customer service through the use of chatbots, supply chain management, and improving decision making based on customer insight. Currently the most common use of AI in businesses is for process automation, either digital or physical, which allows for businesses to take advantage of the powerful optimization capabilities of AI. These systems are often the cheapest and easiest to set up which results in these models being less "smart" (or less creative) than more user focused models. Other implementations in businesses can be described as cognitive insight where AI can analyze vast data sets and interpret them - essentially an automated data analysis solution that is incredibly powerful. This insight takes forms such as identifying credit and insurance fraud in real time or precisely targeting customers for specific ads. These models often include a deep learning element, mimicking the human brain for pattern recognition and learning, which allows them to improve over time which allows for continuous improvement of performance. The most expensive and thus least adopted AI models are based on cognitive engagement, these are models that are able to interact with users using natural language processing. The most basic implementation of these models includes chatbots for customer service questions, with the AI understanding a general sense of what customers may need help with, it can build on itself to become more competent with handling a broader range of customer issues. Another unique example of these models involves providing personal health information/status and receiving personalized care plans which can drastically improve quality of care.

Cognitive Engagement models are the most complex and advanced AI capability and have recently become very popular for their potential of human-like responses – these AI models are the true implementation of artificial human intelligence. OpenAI's ChatGPT provides a great snapshot of these models in its most powerful and developed form. Its ability to perform a variety of tasks and answer a plethora of answers, it's no wonder businesses have become so interested in Cognitive Engagement possibilities. Although these models still have major limitations before a broader adoption can take place, the largest being hardware requirements followed by accessibility and UI. For example, using a cloud provided GPU such as Google TPU v3, it would take approximately 57 hours and cost about $450 to achieve 1/3600 or 0.00000278% of computing power required for Chat-GPT. Similarly, Stable Diffusion required 256 Nvidia A100 GPUs on AWS for a total of 150,000 GPU hours and $600,000. These examples use state of the art technology as benchmarks, so it does not mean these exact resources would be required for all Cognitive Engagement AI, but proficiency, accuracy, and range of understanding significantly increase costs.

4b. Capabilities and Limitations in Generative Art

In relation to our research, we focused on Cognitive Engagement models, also referred to as human-centric AI, focused on generative art. The generative art area is evolving at a rapid pace with constant advancements and increased capabilities becoming available to users. Stable Diffusion and Midjourney are common examples of simple generative art that can create images from text based prompts. Midjourney allows you to iterate over images generated in order to create exactly what you are imagining. These models output static images and have become a relative standard for generative art.

The proficiency of static image generation is significantly further along in development than music generation. AI music generation requires much more refinement and optimization before it becomes widely adopted. OpenAI's Jukebox is an example of music generation using text as input but is difficult for an average user to utilize. Google's Magenta does allow for music continuation and variations with MIDI data as the input, but cannot use text prompts. One of the largest challenges of creating these AI models is where to source training data without infringing on intellectual property or copyright material. Recently, Universal Music Group has requested that music streaming platforms restrict access to AI data collection due to them realizing their copyrighted material was supporting AI without authorization. These legal considerations pose significant challenges that may only be solved through proper regulation.

Another limitation of AI in generative music is that the models lack emotional intelligence. Not only is social-environmental learning in AI underfunded due to high costs, but it is also difficult to train. Training difficulties stem from the fact that it is extremely reliant on what and who is training the AI. Based on that, the AI can come out with an extreme bias or simply not understand the context or prompt given to them, which both would be an output that is useless. In the field of generative AI in the arts, it is important to ensure that there is some emotional intelligence. This would enhance the art being created now from uninspired to something that could be unique.

5. Current Business Models Based on Generative Music

Developing a business model where AI is the product instead of supporting processes has manifested itself in 2 forms: platforms that generate music for an end user or platforms to support creators/artists. Here we analyze the popular and seemingly successful business executing these models.

5a. Platforms for End User

i. Aimi

What do they do?

Being both a website and an application of your phone, Aimi strives to bring the user into the process of creating music. The website is only used to listen to music but if the user wants to customize each individual element they will need to get the free app. The app's main way of tailoring music to an individual is by having thumbs up and down buttons where the app will take in user input to teach itself the type of music the user enjoys. Users can also create their own music by choosing from a range of customizable settings, such as melody, harmony, rhythm, and instrumentation.

How do they make money?

AiMi's business model is based on a freemium model, which offers users 30 minutes of free music every day. However, users can unlock unlimited listening by subscribing to the premium service for a monthly fee of $5.99. This model allows users to try out the platform before committing to a paid subscription, while also generating revenue for AiMi.

How does the AI algorithm work?

The application for AiMi allows users to listen to a continuous feed of music that includes both real tracks and compositions generated by artificial intelligence. Users can adjust the tempo of the mix using a 1 to 10 energy scale, or they can use the thumbs up/thumbs down feature to indicate their preferences and help the AI learn their music tastes. The thumbs up/thumbs down feature is a machine learning algorithm that analyzes the user's feedback to generate music that is tailored to their preferences. As users continue to interact with the platform and provide feedback, the AI becomes increasingly sophisticated in its ability to create music that matches the user's unique musical style. This approach offers a highly personalized listening experience that is powered by the latest developments in AI technology.

ii. Endel

What do they do?

Endel aims their AI and product towards a more specific style of music. It is structured as a well being application because the AI only generates sounds and music to help them focus, sleep, or just relax. They have also invested a lot of time and money into researching the science of the mind and sleep patterns so the AI can generate different styles of music based on the time of day and sleep patterns. Endel is also an editor's choice on the Apple App Store which lets users know that it is a well put together application that might be worth investing in for users to improve various aspects of someone's life. They also included a manifesto on their website that describes the fast changing society with the various distractions and information overload. Endel believes that our mind and sleep cycles are not adjusting like some other systems in our body to the fast changing world where ⅔ of adults don't get the recommended 8 hours of sleep. This belief is why Endel was created as its main goal is to help humans relax and destress from the new pressures put on by society.

How do they make money?

Endel has a 7 day free trial for users where they can listen to anything in the application. If users are uncomfortable signing up for the free trial, Endel put out a free album to demonstrate how the AI functions to users. To gain full access to what Endel has to offer you can buy the subscription for $14.99/month or $89/year where you can use the service offline and tailor sounds for specific activities. There are also two options called Endel for teams divided by the number of team members. The first option is for 10-200 members where it gives access to all team members for one year, newsletters with tips and updates, and onboarding material. The price varies depending on the number of team members but it starts from $600 for 10 members all the way to $12,000 for 200 members where it is around $60 per team member. The second option is for 200+ team members where it requests you to contact the company and it also includes online in-person onboarding, a dedicated manager, and also analytics.

How does the AI algorithm work?

The patented AI for Endel, known as Endel Pacific, is fairly complicated as it takes in various inputs like the user's movement, time of day, weather, location, heart rate, and other factors to specifically curate soundscapes to best benefit the user. This technology takes in all the inputs on the fly and allows Endel to output soundscapes. Endel's algorithm uses different techniques to improve relaxation for users through using the circadian rhythm, the pentatonic scale, and also sound

masking. It uses circadian rhythms to help users flow easier into their biological clocks. They use the pentatonic scale to create simple sounds that are pleasant and also physiologically natural. Sound masking is useful because the AI can carefully tune noises to drown out external distracting noises.

iii. AIMI & ENDEL Comparative Analysis

AIMI and Endel both have many commonalities and differences. AIMI covers a more broad range of genres of music while Endel chooses to focus on the music that enhances relaxation and sleep of users. Even though Endel covers a specific genre of music, the AI algorithm seems to be more complex as it takes in more inputs that are not just the users likes and dislikes of the current song being played. AIMI generates all types of music and its main input is given from the user which is through thumbs up or thumbs down buttons. The pricing structures are also fairly different as AIMI's is closely related to normal streaming services with a free plan and also a competitively priced premium plan while Endel doesn't have a free plan which forces users to purchase the subscription with a much higher monthly price then almost all streaming platforms.

5b. Platforms for Creators

i. Boomy

What do they do?

Boomy is an AI music creation community that uses generative artificial intelligence to democratize songwriting and production. Users can upload their audio files to the site and allow the algorithm to process them to generate original music. Once the song is finished, users can share it on social media sites such as Facebook, Twitter, Instagram and YouTube. Boomy's AI-powered tools enable creators to make original songs even if they don't have traditional music-making knowledge. Users can also create releases and distribute them to all major streaming services and digital music retailers worldwide, earning a share of royalties when their songs are played.

How do they make money?

Boomy makes money by helping users monetize their music creations. When users submit their releases to over 40 streaming services through Boomy, the company tracks how many times a user's song is played and collects the revenues on their behalf. Boomy then takes a percentage of these revenues as a fee for their services.

How does the AI algorithm work?

Boomy's AI algorithm does most of the heavy lifting when it comes to creating music. Users choose from a number of genres and click on "create song" to have the AI compose a song for them in less than 30 seconds. The AI then picks the track's key, chords, and melody. From there, users can finesse their song by adding or stripping out instruments, changing the tempo, adjusting volumes, adding echoes, making everything sound brighter or softer, and laying down vocals. This allows users to create original music quickly and easily with the help of Boomy's AI algorithm.

ii. Aiva

What do they do?

Aiva is a Luxembourg based deep-tech startup that created an AI-powered music composition tool. It helps songwriters create music by composing a song in less than 30 seconds. The AI picks the track's key, chords, and melody. AIVA assists users in creating themes for their projects. This allows users to create original music quickly and easily with AIVA's AI algorithm.

How do they make money?

AIVA makes money by offering subscription plans for individuals and enterprises. By subscribing to their Pro Plan, users own the full copyright of any composition created with AIVA, forever. The company serves both individual artists and institutional customers, such as advertising and production agencies. They offer different pricing plans for individuals, students & schools, and enterprises. The plans vary in price and features, allowing users to choose the one that best fits their needs.

How does the AI algorithm work?

AIVA's AI algorithm works by using deep learning techniques. Specifically, AIVA reads through a large database of music written by famous composers in history and looks for patterns and rules in the scores to understand the style of the music. AIVA practices its understanding of music by predicting what comes next in the track. Once it gets good at these predictions, it creates a set of mathematical rules for that style of music. By doing so, AIVA can capture concepts of music theory and produce melodies that are appealing to human ears just by learning from existing musical works.

iii. Mubert

What do they do?

Mubert is an AI-powered generative music service that offers royalty-free music for app and content creators. Mubert instantly generates tracks tailored to your content on any platform. Users can select the parameters of their future soundtrack in a matter of seconds and generate music right away. Mubert offers a variety of genres and styles, from Chill to Disco to Folk to Hip-hop to Techno. The company has developed a complex algorithm that can create new and original music based on user preferences.

How do they make money?

Mubert makes money by offering subscription-based music streaming for consumers (Mubert Play), a royalty-free licensing offering for businesses and events (Mubert Business), and an industry-facing tier (Mubert Pro) which enables artists and rights holders to upload samples and get paid for their use. Mubert Business offers royalty-free music for restaurants, retail, co-working & public spaces with a Premium plan at $9.99/mo. The company has also collaborated with brands such as Adidas to deliver personalized music to their customers.

How does the AI algorithm work?

Mubert's AI algorithm works by analyzing the data of its customers' feedback about their favorite tracks and artists. Mubert's algorithm depends on like/dislike to understand the user's taste in music and get the most out of personalized music generation. The platform utilizes machine learning algorithms to compose music on the fly, creating an unlimited number of unique tracks for users to listen to. Mubert AI is a cutting-edge tool that allows you to create realistic and infinite music by learning from a large dataset of existing music. The result is a high-quality and original music stream that sounds like it was composed by a professional musician.

iv. Boomy, AIVA, and Mubert Comparative Analysis

Boomy, AIVA, and Mubert are all AI-powered music generation platforms that use artificial intelligence to create original music compositions. While all three platforms use AI to generate music, they have different target audiences and features. AIVA creates emotional soundtrack music and supports professional composers and independent game developers in their creative process. Boomy is aimed at anyone who wants to create music using artificial intelligence. Mubert is a generative music streaming app that generates music based on user preferences.

One key difference between the three platforms is their target audience. AIVA is geared towards professional composers and independent game developers who need emotional soundtrack music for their projects. Boomy, on the other hand, is aimed at anyone who wants to create music using artificial intelligence. Mubert is a generative music streaming app that generates music based on user preferences.

Another key difference between the three platforms is their features. AIVA offers a range of features to support professional composers and independent game developers in their creative process. Boomy offers a range of features to help anyone create music using artificial intelligence. Mubert offers a range of features to help users generate music based on their preferences.

6. Potential Business Models & Barriers to Entry

Using AI in a Team Building Game

This model is a relatively simple use case that can be used for teams to build relationships as well as practice and improve their ability to articulate their creative thoughts. The game includes a 'song creator' who chooses a word relating to a mood from a short list of options and uses that word in a prompt for a generative music AI. The AI then creates a song relating to that word/mood and other players have a goal of guessing which word the player chose. If an accessible tool was developed that produced generative music, this could simply be a team exercise, but if the software was not generally accessible, access could be granted based on purchasing a 5 day availability period or purchasing complete access.

Technical description of AI model:

- Input: english text description of a song
- Architecture option 1: a gpt like model that takes in text tokens and outputs a series of midi tokens for the song. These run much faster, but can only produce sounds from a limited number of instruments. This significantly constraints creativity of the person writing the prompt, because they have to be songs and can't be sounds.
- Architecture option 2: a diffusion model like MusicLM that takes in text tokens and outputs raw audio that could be of almost any describable sound. These run much more slowly and on more expensive hardware, but the outputs can be much more diverse and creative.
- Training for architecture 1: Would require a large dataset of midi and associated english descriptions. However, we could also use a pretrained model.
- Training for architecture 2: Would require an impractical amount of resources. The best course of action would almost certainly be to use a pretrained model.

The difficulty with this model is that it is slightly impractical; AI music generation is too slow and too expensive to make it a casual game. Another hurdle is that non-musicians may have difficulty describing music in a prompt for the AI creating a disconnect between what the model was trained on versus the prompts provided.

Creating a Platform to Ease "Writers Block"

A platform of this nature seems to be the most feasible with current technology, but the cost of computing power would be high. This platform would be a free to use website where musicians can provide midi data and AI can generate a continuation of the input. The key part of this platform is its ease of access compared to other AI midi generation platforms such as Google Magenta. Musicians are not going to want to pay a fee for these extensions, so this platform must be free to use, so a way to sustain it would be through ad revenue. Simple side-bar ads and or ads that play when a user attempts to download the midi output or after a set number of generations.

Technical description of AI model:

- Input: midi tokens from what the person started playing. Could possibly also include a text description but this makes things much harder.
- Architecture option 1: The simplest thing would be a GPT like model that just takes in a series of MIDI tokens and is trained to predict the next one over and over again.
- Training option 1: This would by far be the easiest, as there are already massive datasets of MIDI data covering almost all genres of music. There are also already many pretrained models online.
- Architecture option 2: A GPT like model that is also trained on MIDI songs with a paired text description of them, so the model can also learn to influence the output based on the description of a song. The architecture would largely be the same, just it would use a different tokenizer.
- Training option 2: This would be harder but feasible. It would require much more processing power and training time (but still far less than is required for a MusicLM like approach that produces raw audio). Also, datasets and pretrained models that do this are pretty hard to find.

Revenue Generation Methods

There are two main models that are typically used when video advertisements are embedded into websites to generate revenue.

Cost per thousand impressions (CPM):

- In this model, advertisers pay for every thousand times their ad is shown to viewers, regardless of whether the viewers interact with the ad or not. For example, if an advertiser pays $10 CPM, they would pay $10 for every 1,000 ad impressions.

Cost per view (CPV):

- In this model, advertisers pay for every view of their video ad. A view is typically counted when a viewer watches the ad for a minimum duration, such as 30 seconds or longer. For example, if an advertiser pays $0.10 CPV, they would pay $0.10 for every view of their ad.

It's worth noting that these models can be used in different ways, depending on the advertising platform and the advertiser's goals. For instance, YouTube uses the CPV model for its TrueView ads, which allows viewers to skip the ad after a few seconds. On the other hand, platforms like Facebook and Instagram use the CPM model, where the ad is shown to viewers as they scroll through their feeds. This business model would likely incorporate both and show an ad after a set number of downloads.

Licensing Artist Voices

This model is assumptive of the idea that writers will be willing to sign away their voice rights and that Generative Music technology is continuing at the rate that it is advancing today. We believe that the music industry is going to see a boom in artists giving licensing rights to record labels to be able to create songs in their likeness through generative music platforms. Given that this business model is in its infancy, we see a real opportunity to begin signing artists to licensing agreements now, as once advancements in technology get to the point where it becomes mainstream, big labels such as Universal and Sony Music will have already done so.

7. Further Discussion

While we've seen plenty of benefits for the addition of AI in the arts industry, a further discussion must be made regarding what negatives may happen as a result. For one, generative AI has been proven to produce deep fakes which pose significant threats of impersonations. There are several examples of artists “covering” other songs that have been made by AI. This poses an issue of impersonation of artists and even copyright infringement. It can potentially cross lines of utilizing an artist's voice to make them endorse something they are unaware of. Due to this, it can raise the issue of labels starting to own the rights to an artist’s voice; which on its own could raise other problems such as crossing artist boundaries. Similarly, AI have been trained on copyrighted data which access had not been granted, this creates a legal burden and threat with the creation of these models

Others complain that generative AI has begun to undermine the value of creativity. As mentioned earlier, the training has not reached to the potential of AI genuinely being creative, instead, they are simply outputs created by what a user is requesting. While the results can be creative and impressive, it is not the same as years of training and human experience that most artists go through. While music creation might become more accessible, it would not be viewed the same as what is created by traditional artists. Generative AI, as of now, should be seen as a tool to democratize and extend creativity rather than a substitute for natural creativity.

8. Conclusion

While AI continues to be integrated into business, rise in popularity, and become more accessible, use cases continue to emerge. Generative art is a new way of expressing creativity; in a sense it "democratizes creativity" such that you do not need to be a proficient painter to develop a beautiful piece of art. While this democratization is great, there are discussions to be had if generative art software allows the undermining of creativity. Although these technologies provide great utility, there is much to be done in terms of generative music. We recommend further research to focus on the possibility of creating an AI that can continue or produce unique variations of inputs (such as MIDI or raw audio - although MIDI is much more simple), or create unique and original pieces based on texts prompts. Although this technology is close to existing with technology like Google's Magenta and OpenAI's Jukebox, the practicality of those softwares restricts their usage. Further research should also define a system to estimate 'costs of complexity' in terms of programming and training/GPU hours as well as the monetary cost of those - especially the GPU access. While artistic expression is flourishing with AI generated static images, taking advantage of business opportunities is difficult due to the scale and overall requirements.

Bibliography

AI and Compute - Center for Security and Emerging Technology
Artificial Intelligence (AI) in the U.S. - Statistics & Facts | Statista
Artificial Intelligence (AI) Market Size, Growth, Report 2022-2030
Artificial Intelligence Market Size & Share Analysis Report 2030
Artificial Intelligence market size/revenue comparisons 2022 | Statista
Attention Is All You Need
Beyond ChatGPT: The Future of Generative AI for Enterprises
Cerebras
Cloud Tensor Processing Units (TPUs)
COCO dataset
DALL·E 2
Data centric approach to data privacy: Data centric vs. model vs. application - TripleBlind
Data-Centric AI vs. Model-Centric AI
Data-centric AI: a complete primer | Snorkel AI
DGX Platform | NVIDIA
Gartner Predicts The Future Of AI Technologies
Generative AI Market Size Will Achieve USD 110.8 Billion by 2030 growing at 34.3% CAGR - Exclusive Report by Acumen Research and Consulting
Generative AI Market Size, Share and Forecast - 2030.
Global AI Market Report 2023: Advantages Including Increased Effectiveness
How Do Businesses Use Artificial Intelligence?
Against Malware and Lower Costs Drives Adoption - ResearchAndMarkets.com | Business Wire
GPT-4
Hosted Inference API
How Can Open Source Software Advance Progress Of Artificial Intelligence?.
How to Implement the Frechet Inception Distance (FID) for Evaluating GANs - MachineLearningMastery.com
How to Implement the Inception Score (IS) for Evaluating GANs - MachineLearningMastery.com.
ImageNet
Machine Learning and Artificial Intelligence - Amazon Web Services
Midjourney
MIT Technology Review
NVIDIA H100 Tensor Core GPU

Open Images Dataset V7 and Extensions
PESTLE Analysis of Artificial Intelligence: The 6 Factors that Affect AI
Pre-training of Deep Bidirectional Transformers for Language Understanding
Stable Diffusion Public Release — Stability AI
The MAESTRO Dataset
The state of AI in 2022—and a half decade in review | McKinsey
The White House released an 'AI Bill of Rights'.
Top 10 artificial intelligence (AI) cloud platforms in 2022 | VentureBeat.
Top 10 Generative AI Trends That Will Be Prevalent in 2023
Training a single AI model can emit as much carbon as five cars in their lifetimes |
Universal Music Group Asks Streaming Platforms To Block AI Access – VIBE.com
What Is Artificial Intelligence (AI) | Gartner
What is ChatGPT, DALL-E, and generative AI? | McKinsey
What is human-centered artificial intelligence? | Cognizant.
What's New in Artificial Intelligence from the 2022 Gartner Hype Cycle™
AIMI Website
Endel Website