|  |
| --- |
|  |
| AI Capabilities in Generating Music |
| Final Year Project |
|  |
| **Einar Bulat Lárusson** |
| **Supervisor: Stelios Kapetanakis**  **Second Reader: Khoung Nguyen**  **Word Count: 6155** |

|  |
| --- |
| This document goes into detail about the capabilities of artificial intelligence as in generating music. There are discussions about current algorithms that attempt to modify raw audio waves in an effort to create clean TTS and music. Points are brought up about ethics and legality with copyright infringement. The future of this topic is then considered be it in the near future or the distant future. All of this information is all to ascertain if it is possible and to what extent can we create AI or AGI to generate music that people can listen too. |

# Introduction

Artificial Intelligence has improved vastly in the previous decades to the point where we are capable to have AI generated music with understandable speech like that of OpenAI’s Jukebox which has a fully working algorithm that when trained on a specific artist will be able to replicate their music, singing included without using authentic samples from the actual singer in the songs. This paper will be discussing the capabilities of AI generating music which will include the specific algorithms needed as well as their development history, examples of the generated music, precursors to music generation and the future of this topic. To understand how it is possible for music to be created we need to understand the development history of the algorithms and parts used from text generation and the basis for the algorithm of image generation. We also need to understand what music actually is, as you can simply program an AI that makes sounds however it won’t necessarily be music, it would be possible for this AI to produce music at some point but the chances of that would be infinitesimal as it is the same thinking behind the Infinite Monkey Theorem. The thought of having computer generated music has been around for a while with an early record being a book posing a question and answering said question including details of experiments that were undertaken to answer this question, this book is called Experimental Music; composition with an electronic computer by Lejaren A. Hiller and Leonard M. Isaacson (Hiller and Isaacson, 1959) which was published in 1959.

First of we need to define what constitutes music, the dictionary definition of music is “a [pattern](https://dictionary.cambridge.org/dictionary/english/pattern) of [sounds](https://dictionary.cambridge.org/dictionary/english/sound) made by [musical](https://dictionary.cambridge.org/dictionary/english/musical) [instruments](https://dictionary.cambridge.org/dictionary/english/instrument), [voices](https://dictionary.cambridge.org/dictionary/english/voice), or [computers](https://dictionary.cambridge.org/dictionary/english/computer), or a [combination](https://dictionary.cambridge.org/dictionary/english/combination) of these, [intended](https://dictionary.cambridge.org/dictionary/english/intended) to give [pleasure](https://dictionary.cambridge.org/dictionary/english/pleasure) to [people](https://dictionary.cambridge.org/dictionary/english/people) [listening](https://dictionary.cambridge.org/dictionary/english/listen) to it” (music, 2021), and the Encyclopædia Britannica “art concerned with combining vocal or instrumental sounds for beauty of form or emotional expression, usually according to cultural standards of rhythm, melody, and, in most Western music, harmony.” (Epperson, n.d.). These definitions state that music is more than just sounds, they are sounds in a pattern, in rhythm, melody and harmony, this allows music to inhabit a large spectrum of genres, and with each genre having unique parts that that distinguish them from the others. Although previously stated that how music can sound varies greatly they all follow the same building blocks, rhythm, melody and harmony. These definitions give us a start to the criteria of success for upcoming algorithms as it allows the creators of the AI to check whether the output is in-fact music and not just random sounds.

AI as defined by Encyclopædia Britannica is “**Artificial intelligence (AI),** the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings.” (Copeland, 2021). The concept for AI has been around since the conception of computers, humans have always found a way to automate things and make them more efficient, and what is more automated than a computer thinking for itself and doing tasks so people don’t have to. Although true artificial intelligence has not been developed yet a large amount of progress has been made in the field with machine learning coming the closest to true AI. Machine learning is a study of algorithms that automatically improve on themselves through many iterations, this is done by building a model that is trained on sample data which allow the algorithm to automatically make decisions on new sets of data. To prove that someone has created a true AI series of tests have to be run on the model, with the most famous being the Turing test, which is a test designed by a famous mathematician Alan Turing. This test is to determine whether a computer can mimic a human response, the test consist of a human participant who takes the role of an interrogator who asks questions and receive answers from another human participant and a computer, the interrogator doesn’t know which one is which, the interrogator then has to try and discern who is the human and who is the computer. This is done many times and if the computer manages to fool the interrogator more than 50% of the time then it is said to be artificial intelligence as it is regarded as ‘Just as human’ as the actual human. (Gillis, n.d.)

With the two definitions of it allows the developers to have an accurate vision of how their model should end up like, an artificial intelligence that when trained on data of pre-existing music can recreate songs or create new songs entirely based on the training data. This paper will go into detail discussing two separate models, the aforementioned OpenAI Jukebox along with WaveNet DeepMind which is a deep neural network for generating raw audio waveforms. Raw audio is a way of storing raw uncompressed audio, the two upcoming models both use raw audio instead of using symbolic representation of audio in the form of MIDI, which is what most of computer-generated music was using prior to the models’ creations. MIDI is technically easier to use for the model, but it removes certain aspects of music, making it sound less authentic which is the opposite of what DeepMind and OpenAI are trying to achieve (Dieleman, Oord and Simonyan, 2018). The reason why raw audio is usually avoided is because of its tick rate, a single second of raw audio contains roughly around 16,000 samples (Oord and Dieleman, 2016), this means that the training of the model can take a lot longer, however it does provide the outcome with a greater sound quality.

There are more examples for AI generated music than just the two previously mentioned models, Endel, a company started in Germany 2018, garnered a lot of attention in 2019 when they made headlines with their algorithm that was signed to a 20 album distribution deal with Warner Music Group. Endels algorithm, Endel Pacific, generates personalised and adaptive soundscapes to reduce stress, improve sleep and boost productivity (About — Endel, n.d.). There is insufficient data about how they have actually achieved this, therefore it is not possible to go into further detail about their model and comparisons between WaveNet and Jukebox.

# DeepMind WaveNet

Introduction

DeepMind is a subsidiary and research laboratory which works in the artificial intelligence field. A precursor to OpenAI, being founded in 2010, with a goal to “solve intelligence, developing more general and capable problem-solving systems, known as artificial general intelligence (AGI)” (DeepMind, n.d.) with one of their most notable achievements being a program they developed called AlphaGo. AlphaGo is an AI developed to play an ancient Chinese game known as Go which is said to be the most challenging classical game for AI (Deepmind, n.d.), it garnered a large amount of praise due to it being the first computer program to defeat a professional human player as well as beating a Go world champion. DeepMind has been in the forefront of AI development with projects that help diagnose acute kidney injury and breast cancer to AIs that reach the same merit as top-level computer gamers like AlphaStar, AlphaZero and the aforementioned AlphaGo; however, this paper will be focusing on their WaveNet project which is a generative model for raw audio.

WaveNet

WaveNet is a model that is capable of generating speech which mimics human voice as well as generating music. This project was created to try and create a more coherent text-to-speech (TTS) system which would also require less storage of data as well as the ability to ease of modification to the voice. Most TTS works by having a large database of short speech fragments which are then sewn together to form words and sentences which is all recorded by a single speaker, this can lead to the synthesized voice sounding unnatural and then in turn hard to understand, where the WaveNet aims to create a more natural sounding voice that can be modified as well. It is possible by modelling the raw audio of the waveform, due to the modelling of raw audio it allows the model to not only reproduce speech in a large variation of accents but it can also generate music.

To create this model DeepMind adapted their previously published models, PixelRNN (Oord, Kalchbrenner and Kavukcuoglu, 2016) and PixelCNN (Oord et al., 2016) which are both image generation algorithms, the adaptation was changing it from a two-dimensional PixelNet to a one-dimensional WaveNet.

WaveNet is a Convolutional Neural Network (CNN), which is a neural network that is mainly applied to visualize imagery due to its pattern detection capabilities. This works by having an input/image and passing them through hidden convolution layers transform the input and then outputs it to the following layer. These layers each contain filters which is what detects the patterns in the input, the patterns that the filters can detect start of simple but get more complex the further along the layers you go. An example for advancing complexity is starting off with detecting only edges, corners, and basic shapes, and then gradually advancing to detecting eyes and hair, to then being able to detect dogs and cats in an image. How these filters transform an input is by convolving over block of pixels, with a block being a predetermined sized matrix (e.g. 3x3), once every block of pixels from the input has been convolved it is then passed on to the next convolution layer to have the process repeated except with a more advanced pattern detection (DeepLizard, n.d.) (Howard, 2016). This processed is repeated until it fully visualizes the image and outputs it. That was explanation for how it works on a two-dimensional image, and DeepMind adapted this to a one-dimensional sound wave, it generally behaves the same way except the patterns that the layers are detecting are in the form of sound waves, each layer also takes two inputs instead of one to detect a pattern between the two. The following link takes you to a gif showing this process: <https://lh3.googleusercontent.com/Zy5xK_i2F8sNH5tFtRa0SjbLp_CU7QwzS2iB5nf2ijIf_OYm-Q5D0SgoW9SmfbDF97tNEF7CmxaL-o6oLC8sGIrJ5HxWNk79dL1r7Rc=w1440> . WaveNet also makes the outputs of the previous layers new inputs allowing for more accurate pattern detection, DeepMind also include dilations, gaps in-between inputs, which increases efficiency (Oord et al., 2016) as shown in the figure below.

A picture containing sky, covered, full, bunch

Description automatically generated

Figure : Visualization of a Diluted Convoluted Neural Network, with the dilation getting twice as big each layer the transformed inputs pass through.

The datasets that they use for training can vary, depending on the need/experiment that is being run, the four examples that are shown on the DeepMind page about WaveNet have all been chosen to show the capabilities and sophistication of the model, getting one of the experiments replicate non speech sounds such as breathing and mouth movements (Oord and Dieleman, 2016). Even though one of the goals of the project, apart from it being a model that operates on waveform level in audio, was to produce a clearer and more advanced TTS system they also experimented with generating music, which the final audio clips on the website show. DeepMind used two different datasets which are piano heavy classical music to train the model and the outcomes are impressively close to real classical music, and if they worked on the model more, they most likely could have refined the output into a proper piece of music, and depending on the training data, even vocals.

# OpenAI Jukebox

Introduction

OpenAI is a company that specialises in artificial intelligence research with an aim to develop artificial general intelligence, which they describe as “highly autonomous systems that outperform humans at most economically valuable work”, that benefits all of humanity . (OpenAI Charter, 2018). This company has developed a large amount of different AI like generative models and genetic algorithms. Some of the algorithms that they have developed can solve a Rubik’s cubes, generate images from text, have agents play hide and seek with each iteration the agents find new ways to outsmart each other and finally to what this paper is about, a music generation neural network called Jukebox.

Jukebox

Jukebox is a neural network that generates music based on a training data set where the data set can vary depending on what genre or artist style, this neural net doesn’t only generate the music it also includes rudimentary singing, this is all in the form of raw audio. This project was built upon a previously developed model known as VQ-VAE (Vector Quantisation – Variational AutoEncoder) which was used to generate high quality images, videos and speech (Oord, Vinyals and Kavukcuoglu, 2018). When you access the website for this model they have released many songs generated from their model in 4 different formats: Unseen lyrics where they model has been trained on an artist and it generates a song in that artist style with new unseen lyrics, Re-renditions which it tries to recreate a song from the original artist, Completions where the model is given 12 seconds of a song and it then tries to complete the song, and Fun songs where they mix and match artists, lyrics and genres to generate a song. All of these songs can be found at this website: <https://jukebox.openai.com/> . OpenAI also provides all their code and working on github, allowing anyone who wanted to try it to use it.

As previously stated, the Jukebox is based on DeepMinds VQ-VAE model for images, which OpenAI changed to be suitable for modelling raw audio. The Vector Quantised-Variational AutoEncoder is an VAE that incorporates ideas from vector quantisation. A VAE is an autoencoder that instead of describing an input into latent attributes by the means of a discrete value it describes them as a range of possible values, this is seen in figure 2 below.

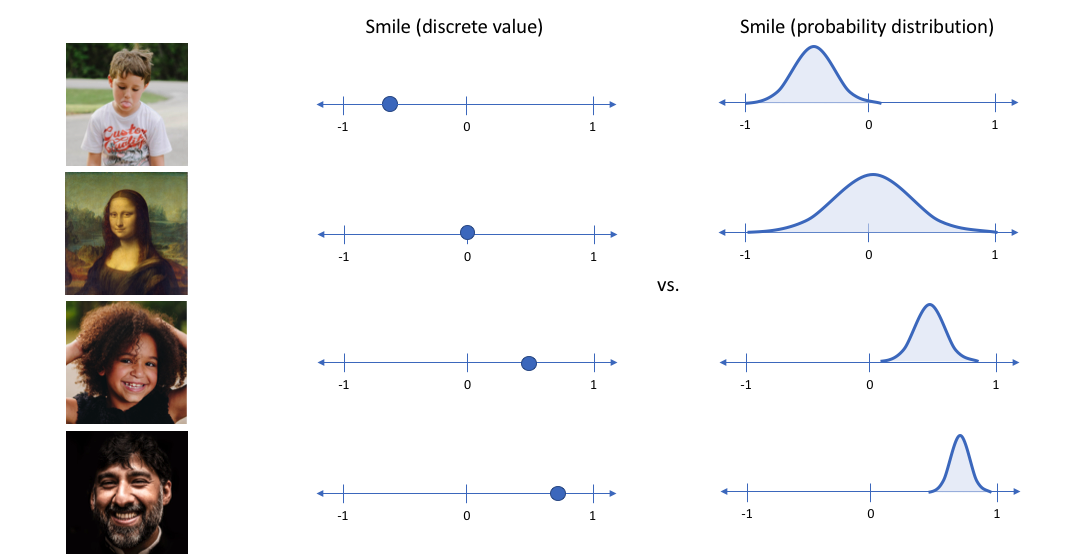


Figure : This shows the difference between AE and VAE latent attributes, that attribute being ‘Smile’, based on input images. An AE attribute is shown as a discrete value, and the VAE attribute is shown as a range of possible values in a probability distribution.

In (Oord, Vinyals and Kavukcuoglu, 2018) it is stated that the VQ-VAE differs from regular VAEs in 2 ways: the encoder network outputs discrete, rather than continuous, codes; and the prior is learnt rather than static. As well VQ allows the model to circumvent posterior collapse, posterior collapse is when the model ignores the latent attributes when they are powered with a powerful decoder. Vector Quantization is a quantization technique that divides a large set of vectors into groups in a way that allows the modelling of probability density functions by the distribution of prototype vectors (Vector quantization - Wikipedia, n.d.).

Diagram

Description automatically generated

Figure : This is a visualisation of training three separates VQ-VAE models with different temporal resolutions, taken from (Dhariwal et al., 2020).

Figure 3 shows the training of 3 VQ-VAE models with different temporal resolutions is the original raw audio waveform input, are the reconstruction output of the original input, represents the latent vectors once the audio waveform is encoded, represents the latent vectors once they have been quantised to the closest codebook vectors, Dhariwal et al. visualise the codebook as (where K is the vocabulary size) but don’t go into any further detail explaining this codebook, and is the discrete representation of the of the audio. As seen a raw audio waveform is inputted into the model and encoded, using CNNs similar to WaveNet, into its latent vectors which then, with the use of the codebook, goes through vector quantization to gain . During the vector quantization process, you get the discrete representation of the latent vectors, which is used later to train a music prior that is discussed later on in (Dhariwal et al., 2020). A decoder is then used to reconstruct the original raw audio waveform from the new codebook vectors. The bottom model in the figure outputs the highest quality audio, this is due to it having to encode shorter samples of audio than the other models; however, the lower quality audio models capture high level semantics like melodies and singing.

The data that is generated from training the models are then used to train prior models and transformers. These priors and transformers allow for the conditioning based on artist, genre, timing, and lyrics which generates novel patterns based upon these conditions. The novel patterns are then upsampled multiple times, with the aid of spectral loss which penalises any differences of the input and reconstructed output. After the upsampling process the new pattern is then decoded by using CNNs again to produce the new audio. The figure below shows a simple visualisation of this process.

A picture containing chart

Description automatically generated

Figure : Visualisation of the process of generating audio using transformers and upsamplers. Taken from <https://openai.com/blog/jukebox/>

# Comparison between WaveNet and Jukebox

After sufficient research into these two generative models for audio it comes to the comparison. Deepmind set out to create a model that was able to generate coherent speech that is more natural than existing TTS systems by the use of raw audio waveforms, members of OpenAI set out to create a neural net that was able to use raw audio waveforms to generate full compositions of music including rudimentary singing. After reviewing the two it is possible to say that both companies reached their respective goals, it is slightly unfair to compare the two models on their capabilities on generating music, the clear winner is Jukebox, because WaveNet only slightly dabbled in the music generation aspect demonstrating that it is indeed possible to generate music, even if only rudimentary. WaveNet laid the foundations for the Jukebox model to succeed, with OpenAI basing their model of off WaveNet, using their theories and concepts, and improving on them for their model with a clear aim and focus to generate music, including singing. Jukebox was released 4 years after WaveNet, there is no information when work officially started on Jukebox, but it can be estimate that it started in 2018 at the earliest due to a reference to getting inspiration from a previous project from OpenAI (which came out in that year), so there is at least a 2-year difference which means that OpenAI could have benefited from advancements in the field that DeepMind wouldn’t have had access to. Both of these companies have contributed vastly to the advancement of AI, even if only AGI (Artificial General Intelligence), capabilities in generating music.

# Ethics and Copyright

When talking about any artificial intelligence projects there has to be a thought put towards the ethics and copyright infringement, especially when it comes to generating music in the style of an artist. For example OpenAI Jukebox uses music from artists to train the model which leads music and vocals that near perfectly mimics the artist, to the point that someone could be mistaken for being authentic, and if OpenAI releases an original song that uses the vocals of an artist who owns the rights to the song, OpenAI or the artist? In the future it could become even more complicated if we ever manage to achieve perfect AI, will the AI have rights? Can the AI own its own music? These latter questions cannot be answered in the present, but it is a thought that should be had in preparation. The question of who has ownership over the music that the AI generates can be answered, the following quote is from a paper discussing copyright law for AI generated music “protection for computer-generated works granted to the person by whom the arrangements necessary for the creation of the work have been undertaken.” (Sturm et al., 2019) This quote states that whoever owns the AI that generates the music owns the rights, and it references the Copyright, Design and Patents Act 1988 (CDPA) which defines computer-generated as “in relation to a work, means that the work is generated by computer in circumstances such that there is no human author of the work;” (Copyright, Designs and Patents Act 1988, 1988). This is only the law in the UK though and there are many other countries and groups whose laws might not grant protection to AI generated music such as the EU (Sturm et al., 2019).

There are also other factors to consider with copyright infringement within the use of the training dataset for the model as well as accidental reproduction of pre-existing music that is copyright protected. The former factor can be solved by getting specific permission from the owner of the copyrighted music to use their songs in the training dataset. This can work if you are only training the model on a specific artist but can become a large task if you are training a model on a genre as there will be many numerous owners of the music, meaning you could get mixed responses for permission allowance. The latter factor can be solved similarly to that of the former, if the AI generates a piece of music that resembles pre-existing work, then permission from the rightsholders needs to be attained for it to be distributed to the public. In a 2019 court case the ECHR Grand Chamber court ruled that “the phonogram producer’s exclusive right under that provision to reproduce and distribute his or her phonogram allows him to prevent another person from taking a sound sample, even if very short, of his or her phonogram for the purposes of including that sample in another phonogram, unless that sample is included in the phonogram in a modified form unrecognisable to the ear.” (Judgment of 29 July 2019, Pelham, C-476/17, 2019) This is significant as it proves the statement in the previous sentence where the owner of the copyrighted music needs to give permission for the replicator to publicly distribute their music that is infringing.

There is a weird industry that has been emerging in the music scene which calls morals and ethics into question, dead celebrity hologram industry. Holograms of dead celebrities have been cropping up making guest appearances at shows as well as even touring, with Tupac Shakur, who died in 1996, appeared on the set of Dr Dre and Snoop dog at Coachella in 2012, and Amy Whinehouse, who died in 2011, went on tour 2019. The ethicality and morality of using someone’s’ likeness post-mortem is shaky, there are rights regarding this topic in USA called rights of publicity, which are the rights of an individual to commercially use of one’s identity, and this in certain state jurisdictions continue after death, however there is no such right or law in the UK. This looseness of rights convey that the caretaker of a deceased persons estate can allow people to use the likeness and therefore legally speaking allowed, although it is legal people still question the morality and ethicality of this, a prime example would be using their image to do something that going against the deceased wishes. (Tiffany, 2018). This ties back into AI generated music, is it ethical to use a deceased persons likeness in vocal form for commercial reasons, it is hard to answer. Like preciously stated it is fully legal to do so with permission from the right-holder, but legality does not mean ethicality.

# Future of AI Generated Music

It has been 25 years since the last official Queen album, there will be no more original pieces from David Bowie due to him passing away in 2016, however this could all change with the advancement in AI. In the near future due to projects like OpenAI’s Jukebox, the owners to the rights of these artists could use their models to create all new pieces of music that could be released to the public. New albums from discontinued bands are not the only thing that awaits us, we could also see the rise of AI powered virtual bands similar to that of The Gorillaz and Hatsune Miku. The Gorillaz are a virtual band consisting of animated band members, but they are far from computer generated as it is a project written, composed, and voiced by Blur lead singer Damon Albarn among others; Hatsune Miku is a computer generated ‘pop-star’ using a vocaloid software to synthesise her vocals, however her lyrics and melody are the product of human intelligence. These artists could be laying the groundwork for fully AI powered virtual bands that could shake up the scene in the music industry with innovations that humans could not have thought up. These virtual bands could adapt to the current trends by increasing their training dataset with the ever-updating popularity charts, they could even tailer releases to specific regions around the world depending on their tastes. These theories all depend on the advancements of AI in the following years and could arrive sooner than any of us expect due to the current trends alluded to by Elon Musk in the following Quote “[about development rate of AI] it is growing at a pace close to exponential.”

Text, letter

Description automatically generated

Figure : Quote from Elon Musk from Edge.org, original quote has since been deleted so credibility is in question. Source: <https://www.reddit.com/r/Futurology/comments/2mh8tn/elon_musks_deleted_edge_comment_from_yesterday_on/>

So far only positive points have been brought up, there are negatives that need to be considered also. You have definite ethical and moral issues when it comes to creating new music from discontinued bands, morals and ethics come into question when you attempt to mimic a deceased individual as well as using their likeness for financial gain reasons. New virtual bands if left completely to their own devices could possibly adapt for the worst, either creating music that is unappealing to peoples’ ears or they could adapt to become too political in nature and possibly have unwanted and unintended subliminal messages. This could mean that the virtual bands need to be regulated which bring along their own dilemma with what is ok to censor and what is too much censorship.

# AI Generated Music in Media and Popular Culture

It is interesting to see how the public view AI generated music, both how it is represented in media formats as well as public perception on the matter. AI is often represented in films and television as it is a popular science-fiction trope with classics such as Terminator, Blade Runner, and 2001: A Space Odyssey, however these AI generating music is rarely seen. In the Amazon prime original series American Gods, during season 2 episode 4: “The Greatest Story Ever Told” there is a side plot that involves a man who creates an algorithm that uses a database filled with pieces of music by the composer Johann Sebastian Bach and then it generates an original piece to mimic his work. (The Greatest Story Ever Told, n.d.). There are instances in pop culture of androids performing music like that of Data from Star Trek (Data, n.d.), androids are robots that are designed to resemble humans and, in some cases, like Star Trek, contain AI. In 1997 there was a newspaper article that writes about an instance of a computer generating a piece of music that emulates Bach, it describes Dr Larson hosting a contest in where audience members had to guess, of three pieces of music, which song was an original Bach piece, which one he wrote, and which one was generated by EMI (Experiments in Music Intelligence). EMI which was described as a “Cybernetic Pretender” is a computer program, developed by a David Cope, which generates music when trained on a specific artist, similar to the previously mentioned Jukebox (Johnson, 1997). There is not much information about this topic in media and pop-culture, this could be down to two separate reasons, there is a lack of representation of AI generated music in the media or there is a lack of recordings written about AI generated music. The later is probably the main reason as there are still examples which leads to the idea of people not thinking of the topic not being note-worthy, even though the topic is very interesting. However, this could change in the future, as AI advances there could become a large amount of interest in this area when somebody has created a model that can either create unique music that people enjoy listening to or a perfect replication of an artist.

# Conclusions

This paper aimed to outline the capability of music generated by artificial intelligence, this paper has included detailed information into the prerequisites needed in order to be both artificial intelligence and to have generated music, this was then followed about a brief discussion about the main projects that aim to create AI generated music in recent history. To answer how it is possible for AI to generate music an in-depth review of the algorithms and it was concluded that the WaveNet model was not fully able to generate full pieces of music, even though it managed to generate multiple audio clips that sound like authentic pianos. These audio clips sounded like real pianos, but they did not have any proper rhythm or tempo to them, there were hints of melodies but not enough for it to constitute a piece of music, but that is understandable seeing as it was not DeepMinds goal to create a model that can create music, rather its goal was to create a model that could manipulate raw audio waveforms. The Jukebox on the other hand manged to create an entire library full of full-length tracks based upon artists, with the tracks being divided into 4 different types of tracks. This proves that it is possible to create an AI that can generate music, all be it the model cannot do this independently, there is still human input that is needed. This does bring hope though that in the future there could be fully autonomous computers that generate music from a request. The capabilities cannot only be shown by what models currently can do, a discussion into ethics and legality needs to be had as even though a CNN can create music if it infringes on any ethical rights or laws, it cannot be allowed to continue to make music. The biggest success story from all the research done into this paper as to be from the Endel.io project. Although there is not much information on how the Endel Pacific works, like there is for WaveNet and Jukebox, in generating music it has been concluded that it does what it has set out to do, during the writing of this paper I discovered Endel and decided to try it, and my personal experience of it is great. I listened to a personalised and adaptive soundscape which increased my focus and boosted productivity, when listing to this AI generated soundscape, I felt more motivated and inclined to do work. I compared it to listening to no music, to normal music that I listened to on a day-to-day basis, and to playlists specifically made be listened to whilst studying, and out of these four different types of music I found that my productivity was much greater when listening to Endel. I was enlightened by these results as it is a case that an artificial intelligence model was able to create and original piece of music that a human enjoyed. When I designated a period to work, I often found myself opening this app on my phone and playing it through my speakers and it always put me in the correct headspace to research and write. To conclude computer scientists in recent years have been able to create projects that, can mimic human speech to a high degree, create high fidelity music that includes vocals, and a commercially successful algorithm that creates personalised soundscapes that can adapt when given new information about the surroundings and user information.

# Bibliography

This bibliography contains sources researched when writing this document. Some of the sources were cited in paper, others were not cited but were read and gave insight into concepts referenced in research that aided in understanding them, and some include information that were not quite relevant or were not able to be added to the paper but where interesting to read.

Sturm, B., Iglesias, M., Ben-Tal, O., Miron, M. and Gomez, E., 2019. *Artificial Intelligence and Music: Open Questions of Copyright Law and Engineering Praxis*. [ebook] Available at: <https://www.mdpi.com/2076-0752/8/3/115/pdf> [Accessed 19 May 2021].

Legislation.gov.uk. 1988. *Copyright, Designs and Patents Act 1988*. [online] Available at: <https://www.legislation.gov.uk/ukpga/1988/48/section/178?view=plain> [Accessed 19 May 2021].

Endel. n.d. *About — Endel*. [online] Available at: <https://endel.io/about/> [Accessed 19 May 2021].

Copeland, B., 2021. *artificial intelligence | Definition, Examples, and Applications*. [online] Encyclopedia Britannica. Available at: <https://www.britannica.com/technology/artificial-intelligence> [Accessed 15 May 2021].

DeepMind, n.d. *About*. [online] Deepmind. Available at: <https://deepmind.com/about> [Accessed 18 May 2021].

Deepmind, n.d. *AlphaGo: The story so far*. [online] Deepmind. Available at: <https://deepmind.com/research/case-studies/alphago-the-story-so-far> [Accessed 18 May 2021].

Dhariwal, P., Jun, H., Payne, C., Kim, J., Radford, A. and Sutskever, I., 2020. *Jukebox: A Generative Model for Music*. [online] arXiv.org. Available at: <https://arxiv.org/abs/2005.00341> [Accessed 15 May 2021].

Dictionary.cambridge.org. 2021. *music*. [online] Available at: <https://dictionary.cambridge.org/dictionary/english/music> [Accessed 15 May 2021].

Epperson, G., 2021. *music | Art Form, Styles, Rhythm, & History*. [online] Encyclopedia Britannica. Available at: <https://www.britannica.com/art/music> [Accessed 15 May 2021].

Gillis, A., n.d. *What is the Turing Test? - Definition from WhatIs.com*. [online] SearchEnterpriseAI. Available at: <https://searchenterpriseai.techtarget.com/definition/Turing-test> [Accessed 15 May 2021].

Hiller, L. and Isaacson, L., 1959. *Experimental music*. New York: McGraw-Hill.

Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A. and Kavukcuoglu, K., 2016. *WAVENET: A GENERATIVE MODEL FOR RAW AUDIO*. [online] Available at: <https://arxiv.org/pdf/1609.03499.pdf> [Accessed 15 May 2021].

Oord, A., Kalchbrenner, N. and Kavukcuoglu, K., 2016. *Pixel Recurrent Neural Networks*. [online] Available at: <https://arxiv.org/pdf/1601.06759.pdf> [Accessed 18 May 2021].

Oord, A., Kalchbrenner, N., Vinyals, O., Espeholt, L., Graves, A. and Kavukcuoglu, K., 2016. *Conditional Image Generation with PIxcelCNN Decoders*. [online] Available at: <https://arxiv.org/pdf/1606.05328.pdf> [Accessed 18 May 2021].

Oord, A., Vinyals, O. and Kavukcuoglu, K., 2018. *Neural Discrete Representation Learning*. [online] Available at: <https://arxiv.org/abs/1711.00937> [Accessed 17 May 2021].

OpenAI. 2018. *OpenAI Charter*. [online] Available at: <https://openai.com/charter/> [Accessed 15 May 2021].

Curia.europa.eu. 2019. *Judgment of 29 July 2019, Pelham, C-476/17*. [online] Available at: <https://curia.europa.eu/juris/document/document.jsf?text=&docid=216552&pageIndex=0&doclang=EN> [Accessed 20 May 2021].

Tiffany, K., 2018. Amy Winehouse is going on tour: The dead celebrity hologram industry, explained. [online] Vox. Available at: <https://www.vox.com/the-goods/2018/10/23/18010274/amy-winehouse-hologram-tour-controversy-technology> [Accessed 20 May 2021].

Holzapfel, A., Sturm, B. and Coeckelbergh, M., 2018. (PDF) Ethical Dimensions of Music Information Retrieval Technology. [online] ResearchGate. Available at: <https://www.researchgate.net/publication/327807581\_Ethical\_Dimensions\_of\_Music\_Info rmation\_Retrieval\_Technology> [Accessed 20 May 2021].

DeepLizard, n.d. Convolutional Neural Networks (CNNs) explained. [online] Youtube.com. Available at: <https://www.youtube.com/watch?v=YRhxdVk\_sIs> [Accessed 20 May 2021].

Howard, J., 2016. Lesson 4: Practical Deep Learning for Coders. [online] Youtube.com. Available at: <https://www.youtube.com/watch?v=V2h3IOBDvrA&t=2s> [Accessed 20 May 2021].

Lin, X., Ma, L., Liu, W. and Chang, S., 2020. Context-Gated Convolution. [online] Available at: <https://arxiv.org/pdf/1910.05577.pdf> [Accessed 20 May 2021].

Oord, A. and Dieleman, S., 2016. WaveNet: A Generative Model for Raw Audio. [online] Deepmind. Available at: <https://deepmind.com/blog/article/wavenet-generative-model-raw-audio> [Accessed 20 May 2021].

Dieleman, S., Oord, A. and Simonyan, K., 2018. The challenge of realistic music generation: modelling raw audio at scale. [online] Arxiv.org. Available at: <https://arxiv.org/pdf/1806.10474.pdf> [Accessed 20 May 2021].

Yang, L., Chou, S. and Yang, Y., 2017. MIDINET: A CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORK FOR SYMBOLIC-DOMAIN MUSIC GENERATION. [online] Arxiv.org. Available at: <https://arxiv.org/pdf/1703.10847.pdf> [Accessed 20 May 2021].

American Gods Wiki. n.d. The Greatest Story Ever Told. [online] Available at: <https://americangods.fandom.com/wiki/The\_Greatest\_Story\_Ever\_Told> [Accessed 20 May 2021].

Subramanian, S., Rajeswar, S., Sordoni, A., Trischler, A., Courville, A. and Pal, C., 2018. [online] Papers.nips.cc. Available at: <https://papers.nips.cc/paper/2018/file/aaaccd2766ec67aecbe26459bb828d81-Paper.pdf> [Accessed 21 May 2021].

Jordan, J., 2018. *Introduction to autoencoders.*. [online] Jeremy Jordan. Available at: <https://www.jeremyjordan.me/autoencoders/> [Accessed 21 May 2021].

Jordan, J., 2018. *Variational autoencoders.*. [online] Jeremy Jordan. Available at: <https://www.jeremyjordan.me/variational-autoencoders/> [Accessed 21 May 2021].

Data Science Stack Exchange. 2019. What is "posterior collapse" phenomenon?. [online] Available at: <https://datascience.stackexchange.com/questions/48962/what-is-posterior-collapse- phenomenon> [Accessed 21 May 2021].

Agustsson, E., Mentzer, F., Tschannen, M., Cavigelli, L., Timofte, R., Benini, L. and Gool, L., 2017. [online] Arxiv.org. Available at: <https://arxiv.org/pdf/1704.00648.pdf> [Accessed 21 May 2021].

Pages, G., 2015. INTRODUCTION TO VECTOR QUANTIZATION AND ITS APPLICATIONS FOR NUMERICS. [online] Esaim-proc.org. Available at: <https://www.esaim- proc.org/articles/proc/pdf/2015/01/proc144802.pdf> [Accessed 21 May 2021].

En.wikipedia.org. n.d. Vector quantization - Wikipedia. [online] Available at: <https://en.wikipedia.org/wiki/Vector\_quantization#:~:text=Vector%20quantization%20(VQ )%20is%20a,originally%20used%20for%20data%20compression.> [Accessed 21 May 2021].

GitHub. 2020. openai/jukebox. [online] Available at: <https://github.com/openai/jukebox/> [Accessed 21 May 2021].

Memory Alpha. n.d. Data. [online] Available at: <https://memory-alpha.fandom.com/wiki/Data> [Accessed 21 May 2021].

Johnson, G., 1997. Undiscovered Bach? No, a Computer Wrote It. The New York Times, [online] Available at: <https://www.nytimes.com/1997/11/11/science/undiscovered-bach-no-a- computer-wrote-it.html> [Accessed 21 May 2021].