**Project ID.: ITB-044**

**Coventry University**

**Programme Code: 193-19432**

**BSc (Hons) Information Technology for Business**

**Module Code: 303CEM**

**Final Year Project Report**

**Project Title: Generate soundtrack/ background music using Deep learning**

**Student Name: Lee Chun Hin Attic**

**Student ID: 56247624**

**Project Supervisor: Mr Terry Lam**

**2nd Assessor:  Mr Albert Lau**

**Submission date: 22/5/2020**

**Word count: 9742**

**Declaration of Originality**

I declare that this is all my work and has not been copied in part or in whole from any other source except where duly acknowledged.  As such, all use of previously published work (from books, journals, magazines, internet, etc.) has been acknowledged by citation within the main report to an item in the references or bibliography list.   I also agree that an electronic copy of this project may be stored and used for the purposes of plagiarism prevention and detection.

**Statement of Copyright**

I acknowledge that the copyright of this project report, and any product developed as part of the project, belongs to Coventry University.

**Signed: by Student: \_Attic Lee\_\_                                Date:  21/05/2020**

**Abstract**

Nowadays, people lived in an era of Internet and digital that we cannot live without them. New media became extremely important for the business. YouTube is a platform that many people watches every day. In YouTube, the content creator is valuable since their talent can draw many attentions on the internet. Video category on YouTube are various. For instance, travel, unboxing, gaming and lifestyle. All of these videos include background music, which is an important element contributing to the video, as well as a movie. The content creator often ignores the importance of background music, that might drag the quality of the video because of busyness, or they due to limited choice cause by copyright.

The aim of this project is to construct a soundtrack/music generator using deep learning. The product is proposed to provide various option for video production to deal with the fact that copyright limited their source of choice and audience might get bored with the commonly used background music.

Alternated SDLC waterfall model was used to develop this music generator system. User requirement is collected through small-scale user requirement survey and the observation of the similar product. Data will be collected from free sources such as Freemidi.org and Classical Archives LLC. After that, data was used for Google's polyphonic RNN model to analysis and train to generate classical, pop, and rock music phrases.

A website is created to give information about the music generator system. Providing 3 genres of music generated by AI using deep learning model. Songs are stored in SoundCloud storage that API is provided in the website for songs preview and download.

This project defines that artificial intelligence can make art and provide more option to the artist. The further development of this product that benefits to the internet world is also cannot be underestimated.

**Table of Content**

[**Abstract** 3](#_Toc41072105)

[**1. Introduction** 4](#_Toc41072106)

[**1.1 Background** 4](#_Toc41072107)

[**1.2 Statement of the problem** 7](#_Toc41072108)

[**1.3 Aim** 8](#_Toc41072109)

[**1.4 Objectives** 9](#_Toc41072110)

[**2.Literature Review** 9](#_Toc41072111)

[**3. Research Methodology** 15](#_Toc41072112)

[**3.1 Research question** 15](#_Toc41072113)

[**3.2 Research criteria** 16](#_Toc41072114)

[**3.3 Choice of research design** 16](#_Toc41072115)

[**3.4 Collection of user requirement** 16](#_Toc41072116)

[**3.4.1 Small-scale survey** 16](#_Toc41072117)

[**3.4.2 Observing similar product** 17](#_Toc41072118)

[**4. Cost and benefit analysis** 17](#_Toc41072119)

[**5. Project Plan** 18](#_Toc41072120)

[**5.1 Work Breakdown Structure** 18](#_Toc41072121)

[**5.2 Gantt chart** 19](#_Toc41072122)

[**5.3 PERT Chart** 20](#_Toc41072123)

[**6. Choice of Technology** 21](#_Toc41072124)

[**6.1 Development method** 21](#_Toc41072125)

[**6.2 Deep learning model** 21](#_Toc41072126)

[**6.3 Website (Front-end )** 23](#_Toc41072127)

[**7. System Analysis and Design** 23](#_Toc41072128)

[**7.1 SWOT analysis** 23](#_Toc41072129)

[**7.2 System Design and Development** 24](#_Toc41072130)

[**7.3 Design of the website and deep learning model** 24](#_Toc41072131)

[**7.4 Website Overview** 28](#_Toc41072132)

[**7.4.1 Home Page** 28](#_Toc41072133)

[**7.4.2 About Page** 29](#_Toc41072134)

[**7.4.3 Songs Page** 30](#_Toc41072135)

[**7.5 Test Plan** 32](#_Toc41072136)

[**7.6 Test Case** 33](#_Toc41072137)

[**7.7 Deep Learning model training** 34](#_Toc41072138)

[**7.8 Deep Learning Data pipeline** 36](#_Toc41072139)

[**8. Further Development** 36](#_Toc41072140)

[**9. Conclusion** 37](#_Toc41072141)

[**10. Self-Reflection** 38](#_Toc41072142)

[**Reference** 39](#_Toc41072143)

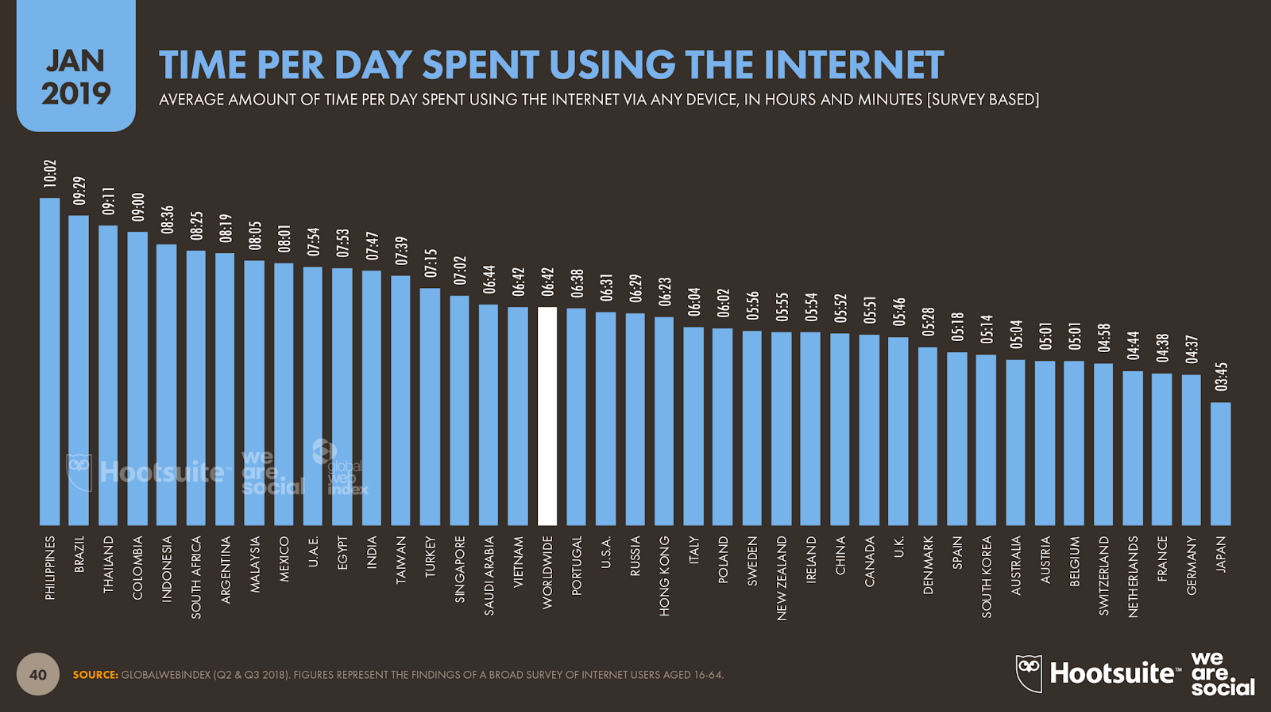
[**Appendix A- User requirement Questionnaire** 42](#_Toc41072144)

[**Appendix B -Project Metting Log** 50](#_Toc41072145)

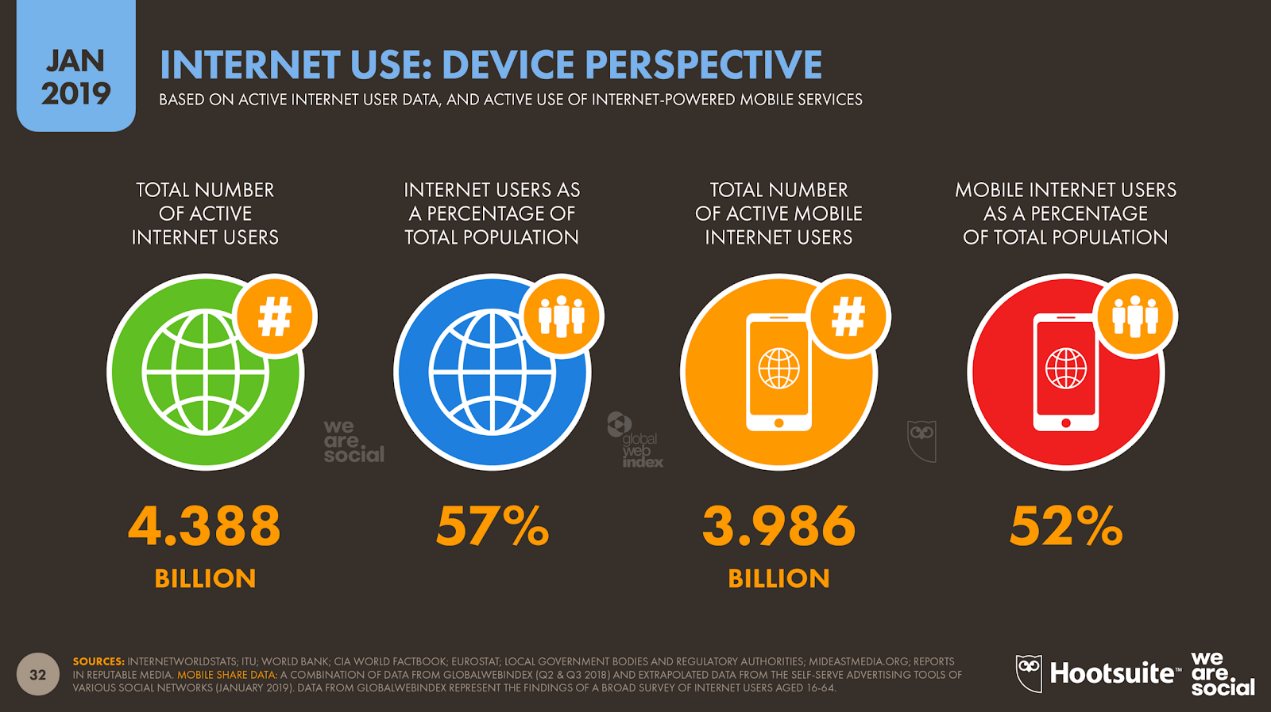
**1. Introduction**

**1.1 Background**

Nowadays, social media has become an integral part of our society. YouTube is one of the best platforms to connect to new media. In modern society, people no longer keep sitting on the couch in front of a Television. Instead, their hands are glued on the screen of a smartphone, which usually playing a YouTube video. At Brandcast marketing events, CEO of YouTube Susan Wojcicki revealed that YouTube counts 2 billion monthly active users in May 2019 (Spangler, 2019). According to the report (DIGITAL 2019: GLOBAL INTERNET USE ACCELERATES) done by We Are Social, a marketing and advertising company that located over the world, and Hootsuite, a social media management platform, shows that there are worldwide people average spending 6 hours 42 minutes a day using the internet in quarter 2 & 3 of 2018; In Hong Kong, people average spending 6 hours 23 minutes on the internet every day. Moreover, the Population of active internet users had reached 4.388 billion in early 2019, according to a variety of sources, and sum up by We Are Social and Hootsuite. 4.388 billion is a huge number meanwhile the world population is about 7.6 billion by estimation as of 2018 (Worldbank.org, 2019), which is almost half of our population in the world. By these number, we can see how dominant internet is in today’s society.

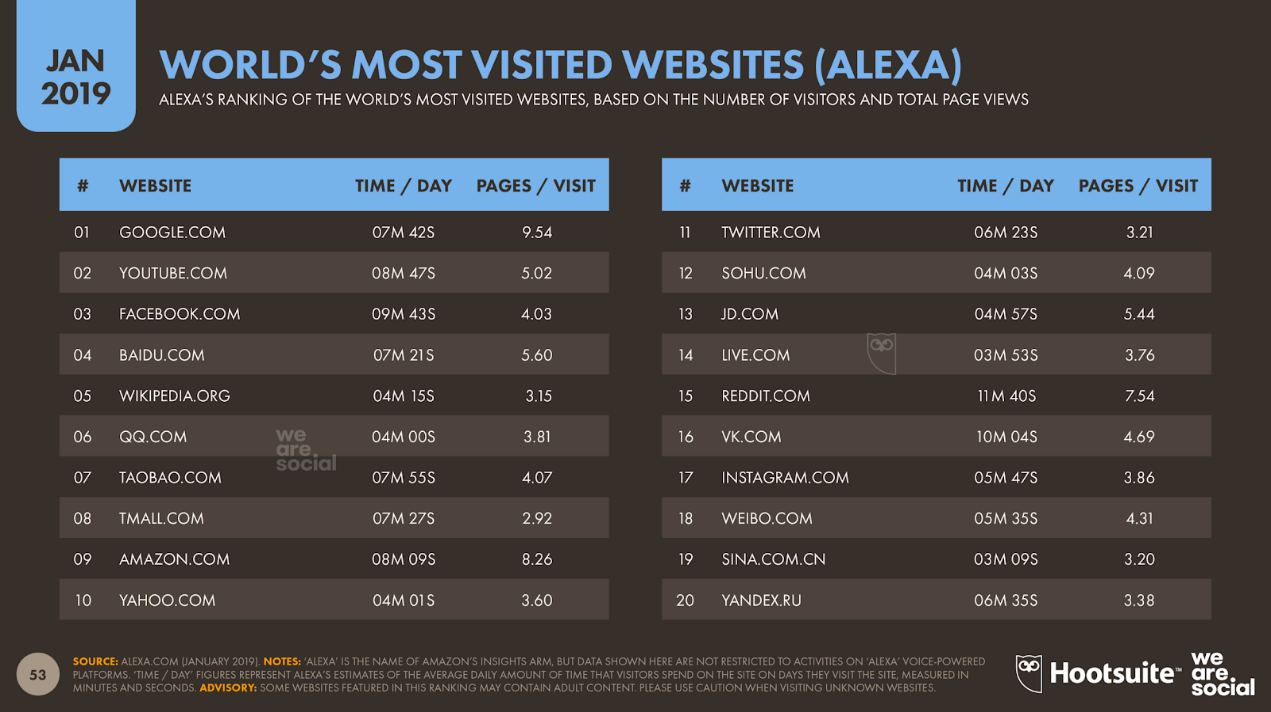


(We Are Social and Hootsuite, 2019c)



(We Are Social and Hootsuite, 2019b)

Furthermore, total active internet users as 57% of the world total population, the most visited website is Google.com under Google, conducted by Amazon’s insight arm, Alexa. Yet, the following is Youtube.com, also under Google’s brand name.



  (We Are Social and Hootsuite, 2019d)

All data above shows how massive is the internet penetrating our lives as well as YouTube, a video platform. In fact, it has a great amount of revenue produced thanks to those huge numbers of people visiting their website. YouTube net revenue is about $14 billion, estimated by an analyst employed by Alphabet, Google’s parent company (Marino-Nachison, 2019). As a video platform, it is like a mutualism of two animals in the ecosystem: YouTube provides a popular platform and revenue to attract the producer to use their service; Producer uses their intelligence to create a product to the platform for earning money and increase the visitor flow rate of the website. Both of them are gaining benefits from each other. Moreover, the profit that sustaining this industry is from advertising. According to eMarketer, YouTube has $3.36 million in advertising revenue in 2018 in the US, which grown nearly 17.1% from 2017 to 2018 (eMarketer, 2018). In YouTube, video creator shares 55% of the ad revenue of YouTube. Famous YouTube video creator can even make a living from that and being called YouTuber as a job, as a business. Number 2 subscriber count YouTuber, Felix "PewDiePie" Kjellberg, has earned $7.4 million from his YouTube channel in 2014 (Edwards, 2015), which also made YouTube became a dream job for some kids (Kat Tenbarge, 2019).

In a video, soundtrack music/ background music is an essential element to appeal audience and emphasize the content of a video, or a movie. The audience may impress by the touching or exciting music that the creator plugged in. However, a creator’s video especially YouTube creator, their variety of music is limited due to the copyright policy and the reliance of the existing YouTube Audio Library which is free for use. Copyright is a big issue for all events that would play music. For example, movie, broadcast, a live performance, etc. Copyright music means the music that has copyrighted by the owners. Owners of copyright music have a music license that ensuring the owners’ musical work are limited to use in public or business usage (Kaplan and Nimmer, 1965). Every copyrighted music must get agreement from the owner in order to have rights to use it. Thus, to some degree, it has limited creativity of people, but it also guarantees some rights of the original producer. In YouTube, video creator/ YouTuber usually uses music that YouTube Audio Library provided because it is licence free, which means it is free to use. Unlike large Hollywood movie companies, a video creator would not have such big funding for music production, even a music licence of a single song. They cannot afford that amount of money, they would rather invest in video recording equipment instead of music quality.

In the 21st century, AI becomes the hottest topic in the field of technology. The famous company invests billions of dollars into AI business, numbers of IT company are having their eye on this big potential business. Even in our daily life, A.I appears anywhere. For example, when you are shopping online during Midnight. A modern, fancy gadget catches your eyes, you want to know more about this product. What material does it use? Where does it manufacture? What is its function? etc. Unfortunately, the description of the product is so simple that you cannot even get a single answer. However, a ‘Chatbots’ can help your needs. It responses your inquiry by recognizing the keywords and phrases you typed in the chat, and convey the answers you needed, simply by AI. Nowadays, People knows AI is a thing that will keep learning non-stop and perform like a unique human, even better than us.  Nevertheless, some thoughts pop-ups in our mind – Can AI be creative? This is a common conversion about AI in modern society. Every human is unique and has his own thinking, we create something new and valuable, which is called creativity. (Mumford, 2003)

**1.2 Statement of the problem**

For the audience, they may get bored when a creator using the same music repeatedly. Especially when different creator/ YouTuber is using the same music. However, people who edit the video, usually the creator themselves or the editor that they hired, tend to put more efforts on new idea or video editing effect instead of choosing new music for their video as it might be more valuable for their business. Options and time are the factors that dragging the quality of their product (video). For a competitive business, a small thing could be crucial and help you win the game.

Listener fatigue, usually describing that exposing to auditory stimuli that cause our sound sensory overload (Brondel and Cabanac, 2007). However, entertainment also has such an effect. People might get oversaturated when they enjoy that entertainment too often. Using Artificial Intelligence composed music can increase the choice variation of background music to the creator/ YouTuber, making their video freshness. For instance, a creator that uses 20 different background music on 20 different videos would be less ‘fatigue’ than that use 5 different music on 20 different videos. Using Artificial Intelligence composed music not only provides an extra option for the creator to choose when producing their video but also to all creator such as students of media, students of film.

Hence, in today’s new media world, YouTube is an enormous business with tremendous opportunity for the creator. Only people with talent and passion could be a success in such a competitive situation, every small detail is important to make them famous. Thus, it is a good opportunity that combines creative work with Artificial Intelligence.

Based on the above backgrounds, the problem encountered is:

1. How can the video creator invest less time and money for background music?
2. Can AI assist us when we doing creative work?

In this project, TensorFlow deep learning algorithms would be essential to create a machine learning/ deep learning model that is able to train the Artificial Intelligence to analyse the songs that given to the model to have a prediction based on the data (song) given to it. Unique background music for video producer will be produced as a result.

To achieve, machine learning/ deep learning model is the main focus of the system.

Firstly, data preparation is crucial for deep learning, music files cannot be analysed without any preparation. Gather the music files into machine-readable format is the first step for deep learning part. Choose the model for music generation can be processed afterwards ( Google's polyphonic RNN model was chosen for the system development). Next, train the RNN model with the data prepared for a few weeks. It takes time for the RNN model to be well-trained. After that, it is done for the deep learning part.

After the back-end product design, having a front-end design for the user is also important. UI and web page design are going to take place, in order to let the users, interact with the system handily.

A complete product in this project will have the following functions:

1. Allow users to select their desired genre of music (e.g. pop, rock, electronic, and jazz)
2. Preview the music
3. Download the music with an audio file (Mp3, FLAC, wav, etc.)

**1.3 Aim**

In this project, machine learning/ deep learning will be performed to build a generator which will generate soundtrack music/ background music that for the uses of video or movie. This product is important for new media/ multimedia content creator or business runner. They might get benefits of this project which provide variety and innovative product to stakeholders not only to enhance the character and quality of their product but also to reduce their costs when selecting music for their videos. In this project, a deep learning model will be built in order to analyst the current existed music. Google's polyphonic RNN model will be introduced to analyze existing music and have a product of entirely new music produced by the model by ‘prediction’. Generating soundtrack music/background music for the use of video production is the goal of the model.

Users can request a different genre of music based on their needs. To achieve, Python and TensorFlow will be used to build Google's polyphonic RNN model. The main goal of this project is to produce a system that can successfully generate different genre of soundtrack based on the user’s order.

**1.4 Objectives**

The goal of this project is to try to have a similar idea of AIVA, which is a similar product that is able to generate music by gathering information from different existing music work from the world such as Mozart and Bach. The product offers AI-generated music, editing with music composer and more. However, due to the limitation of time, and target audience, some component of the product such as editing music, composing music not be included in the final deliverable. This project will only focus on the following objectives.

* Collecting numerous of soundtrack music/background music
* Build a deep learning model that will be trained by using a large set of music that labelled by the name of the genre such as cinematic, acoustic, rock, etc.

To achieve the first goal, data (music files) will be collected mainly from the YouTube Audio Library a since it is the major source of background music for the media produce, however, eh music file that from YouTube Audio Library are not readable for the deep learning model, MIDI format files are required to feed the model and train it. Therefore, data would be collecting from ‘FreeMIDI.org’ and ‘Classical Archieve LLC’, both website provide free royalty music that can be used for research, Google's polyphonic RNN will learn from those music that for generating music.

**2.Literature Review**

Media definition

According to the Oxford dictionary, the word ‘Media’ is about the source of people to retrieve information and entertainment of the society, such as television, radio, newspaper, internet etc. (Oxford,). Media has a different variance, multimedia, mass media, social media, new media and numerous forms of media that are the branch of media.

New media definition

 Lister et al. (2003) suggested that new media has various of definition, it involved a lot of meaning into it. It can be meant ‘digital’ or ‘electronic’ media as nowadays it always appears in the world of internet, it can be seen on different social media website like YouTube, Facebook etc.; It can be referring to a computer game. A different person could have a different image in their mind when they are referring to the term ‘New media’. Still, there are some main terms or component to discourses about new media. These are: digital, interactive, hypertextual, virtual, networked, and simulated (Lister et al., 2003). For example, YouTube video is digital, which the user can enjoy it in a platform that is in digital form. Interactive, the audience of new media can directly involve in and modify the information they receive. Instead of a ‘viewer’, they become a ‘user’ that can participate in the media; They can provide their opinion to the video and bringing some new ideas to the new media producer. These characters are a part of a matrix that new media may consist of, but these aspects must not be included in the case of new media. Yet, different extent and combination of these attributes could be found in those new media. (Lister et al., 2003)

Multimedia definition

Multimedia is a content that mix up the usage of multiple forms of media such as text, photos, artwork, animation, video, and audio in order to display information to the receiver (Najjar, 1996).

Soundtrack/ background music

A soundtrack is about the music of film or separately recorded music for video or film (Cambridge Dictionary, 2019).

Background music refers to the music that is not intended to treat as a primary focus of the audience while another activity is undergoing. For example, during a scene in a movie (Collins Dictionary, 2019).

Both soundtrack and background music will be used as data in this project since soundtrack and background music are similar and used in video creators’ work.

Importance of soundtrack/ background music

The relationship of music and audiences’ mood are important in this project as the main goal of the project is to provide more option for the producer in order to let their video to be more extent, soundtrack/ background music will take an important role to increase the audience's’ mood and emotional response to the video. Similar research suggested that background music has a positive influence in promoting consumer mood and advertising response. Customer moods and motive on purchasing a product is affected and stimulated by the background music of the environment. (Alpert and Alpert, 1990) Furthermore, varied genre music has a different effect on influencing the behaviour of the customer. Which displays the significance of background music on enhancing the quality of the video.

What is Artificial Intelligence?

David Lynton Poole and Alan Keith Mackworth (2006) presents about Artificial intelligence: Artificial intelligence, also known as AI, is about the study of synthesis and analysis of computational agent that act intelligently. An agent is something that exists in the world and has its own function and behaviour such as cats, humans, companies, and government.

Agent act intelligently means the agent considering the short-term and long-term consequences of their action and act properly for their circumstances and objectives. Furthermore, agents are flexible to change the environment and changing their objectives. Yet, given perception and limitation of calculation, they can make the appropriate choices. A computational agent is a form of agent that their decision can be analysed in terms of calculation. In other words, their decision can be broken down into raw operations that can be carry outed in a physical device. Understanding the principles that cause intelligent behaviour achievable in the natural or artificial system is the key scientific goal of artificial intelligence. This goal is achieved by the examination of natural and artificial agents, hypothesis codifying and testing about the reason behind the construction of the intelligent agents and design, constructing, and investigating with computational system that works generally as required intelligence. Apart from the scientific goal, the engineering goal of artificial intelligence in the design and synthesis of intelligence artefacts (agents) that act intelligently, and they can be useful in many applications.

Google Magenta

Google Magenta is an open-source research project about deep learning and creative work including drawing and music creation. It is operated by Google Brain team’s researchers and engineers. New deep learning algorithms was invested and learning algorithms was reinforced in this project for generating songs, images, drawings, and other materials. Smart tools and the interface was produced by the project to benefits artists and musicians to expand their creativity using the deep learning model in Magenta. Tensorflow was used to build the deep learning model and distributed as an open-source Python library. The library allows the developer to utilize it for handling source data e.g. music and images and generate new information from the models (Google, n.d.).  Google's polyphonic RNN model inside Google Magenta was chosen in the development of this project.

What is RNN model (Recurrent Neural Network)?

RNN model (Recurrent Neural Network) is a class of artificial neural network that the outputs of the neural network can be used for the inputs with hidden states (Zaremba et al., 2015). It is a feedforward neural network that extended with the recurrent connection. With this characteristic, sequences of information can be leant by the RNN network. For example, notes that combine to form a melody. Similar to human, the RNN would not start thinking from zero every moment. When doing an action such as conversation and reading, we are understanding each word based on the understanding of previous words. Persistence exists in Recurrent Neural Network. Unlike traditional neural networks, RNN network includes a loop that the information is able to persist. Inputs and outputs in traditional neural networks are independent, in cases of language, when predicting the next word of a sentence, the prior words are needed for the prediction. Hidden states allow RNN to have ‘memory’, the output of the hidden layer returns and become an extra input.  Therefore, RNN not only depending on the present input but also the previous input that returns as an input, it learns all of the above and then the whole sequence after recursing a multiple of time (Zaremba et al., 2015). As a result, RNN model can be used in this project for learning the musical sequence.

Magenta polyphonic RNN model

This RNN model took inspiration from the “BachBot”, an end-to-end automatic composition system that aims to compose music that in Bach’s chorales style with deep long short-term memory (LSTM) generative model. The Magenta polyphonic RNN model intended to generate polyphonic music by applying language modelling to the music generation using long short-term memory (LSTM) generative model. The model treats polyphony as a single sequence of note with symbols (START, STEP\_END, and END). The notes are arranged by pitch in descending order inside a step. The training job of this model also involves attention configuration/ attention mechanism.

Long Short-term Memory (LSTM)

Long Short-term Memory (LSTM) is another kind of Recurrent Neural Network. RNN has a problem: vanishing of gradient. Gradient is the values that responsible to carry information and update a neural network weights used in the RNN parameter. The vanishing of gradients is when the gradients become smaller over time as it backwards propagates when the gradients become extremely small, small gradients mean it has small update, the starting layer of RNN with small update stops learning eventually. As the layer does not learn, RNN is not able to learn things that in a longer sequence and cause short-term memory.

LSTM has gates that are the internal mechanism that can control the flow of information. The gates allow LSTM to learn relevant information to make predictions and forget non-relevant data. It passes the important information through the long sequences chain to predict the result. In this case, it produces good quality of results.

Attention RNN

Recently, the model using the encoder-decoder structure is very popular because it has achieved better results than other traditional model methods in many fields, such as computer vision and text translation. The traditional model of this structure usually uses LSTM units to learn the input sequence and encodes it into a fixed-length context vector. Then uses LSTM units to read this vector representation and decodes it into an output sequence. For a short-length input sequence, the model can learn a reasonable fixed-length context vector. However, the problem with this model is that when the input sequence is very long, it is difficult for the model to learn a reasonable context vector, the input sequence will be encoded into a fixed-length context vector regardless of its length, and decoding is limited to the fixed-length context vector. This limits the performance of the model, especially when the input sequence is relatively long, the performance of the model becomes very unsatisfactory. For example, in translation work, the translation quality becomes worse when the input text is too long. The attention mechanism enhances the LSTM / RNN model to overcomes the obstacles of the traditional encoder-decoder structure (Bahdanau, Cho and Bengio, 2015).

The basic idea of ​​the Attention mechanism is to break the limitation that the traditional encoder-decoder structure relies on an internal fixed-length vector when encoding and decoding. Attention model does not encode the whole input sequence into a single fixed context vector, instead, it develops a context vector that is cleaned for each output time step.

For example, in translation, every time a word is generated, the model with the attention mechanism will find a most relevant word in the input sequence. Then the model predicts the next target work based on the current context vectors and all the previously generated word.  The attention model not only improves the performance of the neural network but also allow the developers to have a better understanding on the internal working mechanism: How the information in the input sequence affects the final generated sequence during the model output process, also helping them to debug some specific input and output (Bahdanau, Cho and Bengio, 2015).

What is GAN model (generative adversarial network)?

Generative adversarial network is a deep learning framework that estimates the generative model by training two models to compete for a game. In a classic machine learning model, a trained model can give out prediction with trends about data. But with GAN model, the model is able to generate new things based on the training set of the data. There is some GAN model used for music learning, however, in this project RNN model will be used for the music generation.

Artificial Intelligence usage

Nowadays, Artificial intelligence has many applications in society. From developed for executing specific tasks, such as robot control, remote sensing, to breakthrough into various fields and industries including health care, social networking, e-commerce, finance, art, and more.

Health care application

Identifying cardiac diseases requiring the skill of heart auscultation (the monitoring of sounds produced by the heart and blood circulation), which is essential for a doctor to perform diagnosis. By the cooperation of Department of Electrical Engineering, University of Hawaii and Department of Computer and Information Science, Linkoping University, a simple model has been developed for production of heart sound and demonstrate a prototype system that based on wavelet decomposition of the sound and neural network-based classifier focus on the investigation on heart sound analysis (Reed, Reed and Fritzson, 2004).

Similar product/work

Research in artificial intelligence (AI) on art and music has been a subject to research in the artificial intelligence field. In 1960, R. Kh. Zaripov, a Russian researcher published an article about the process of musical composition using the algorithm. (Zaripov, 1960) Nowadays, many research and system are developed for the application of AI and music.

BachBot

BachBot is a research project on artificial creativity, to produce creative work using artificial intelligence. This project uses a probabilistic sequence model called LSTM to build artificial intelligence that can generate and harmonized chorales that are identical to Bach's real work by understanding music composition. Their goal is to have a better understanding and insight into the mechanisms that are fundamental of music and creativity while extending the limits of AI in modern society (Liang et al., n.d.).

AIVA

AIVA (Artificial Intelligence Virtual Artist) is an electronic composer officially recognized by the SACEM (Society of Authors, Composers and Publishers of Music) that are capable of composing emotional soundtracks for films, video games, commercials and other categories of entertainment content. AIVA uses a large dataset of music partitions, mainly from the well-known classical music composers such as Mozart, Beethoven, and Bach, to train and build a deep learning model of music. The model is then used to express and generate a piece of unique tracks (AIVA, 2016).

System development life cycle (SDLC)

Waterfall model

The waterfall software development which consists of five phases of software development that process sequentially (Bassil, 2012). From Requirement analysis to Maintenance, the development phases advanced to the next phase one by one until the previous phase completed.

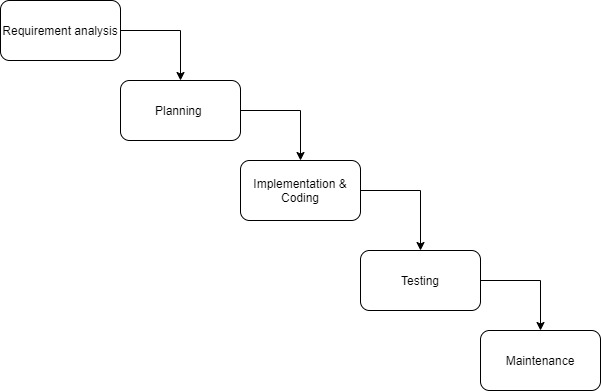


Figure 1: Waterfall Model

Requirement analysis:

This stage is to gather all the information and requirement of the system from the user and customer to develop a product base on the user expectation. Identify what the user wants, what is the objective of the product, and understand the scope of the project. Gathering User requirement is essential to pinpoint a useful function for the system. In this project, requirement analysis is to have a literature review and user requirement survey to collect the requirement of the system. A small-scale survey was conducted to identify and get a basic idea of what the users and stakeholder want to get from the product.

Planning:

After the analysis and the collection of the requirement of the system, planning and design of the system is the next step. Design the software and hardware architecture that will be used for implementing system development in further phase. Choosing the deep learning model for training and decide the source to gather the music data is the focus of this phase in this project.

Implementation & Coding:

In this phase, coding of the front-end and back-end of the system is being proceeded. In this project, the coding of the website and the total process of deep learning, including creating dataset, training and generating.

Testing:

After constructing the system, test plan and test cases should be carry out in  order  to check if there is any defects in the system. Any defects in the product should be fixed in this stage. Retest until the developed product can be released.

Maintenance:

After the product has been determined to be ready to deploy, maintenance of the product is essential in case any unseen bug reveal, any accidents occur or any improvement and enhancement that can be done. In deep learning, the product will not come to a point that is perfect as more data make the model improve. In this project, the maintenance part is to keep train the model and improve the model. Also, more features might add in the future.

**3. Research Methodology**

**3.1 Research question**

The research question of this project is to provide a solution for the company/video content creator that can increase their quality of product and reduce the time and cost of searching the background music.

**3.2 Research criteria**

The objective of the study mainly focuses on the audience and video content creator, the people who select background music and create the video. For the audience, they are the customer of the media business as they retrieve information from YouTube or other video platforms. How they feel about the product, the video is important.  For video content creator, the handy of the system is essential as this research is aim to reduce the time and cost of searching for the background music.

The project is primarily focused on the deep learning that creates the music, and the webpage for the users to select their desired music. The music on the web page should be easily accessed and download.

The stakeholders of this project are:

* Video content creator (e.g. YouTuber)
* The audience on video platforms
* Company (Producing videos as a marketing strategy or advertising their product on video content creators’ video)

**3.3 Choice of research design**

This project will develop a webpage as a front-end that user can access with whatever device they are using, including PC, smartphones, and laptops. Since the work of production such as editing is mostly done on PC or mac. Therefore, rather than developing a mobile app, a web page was chosen to be a front-end for the users to interact and gather information.

Web development uses HTML and CSS for development. The website consists introduction of the project/system, songs preview and download page that consist of the SoundCloud API that provide 3 playlists for the user to listen, download function is also provided by SoundCloud. SoundCloud is the largest music and audio platform in the world that has the diverse creator community. Through this open platform, people and artist can share their music and connect with each other, any type of audio such as tracks, raw demos, podcasts and more are able to distribute to the world through SoundCloud. The work is not necessary to work from scratch since some exist API has the complete function that can contribute to the project and the main focus is on the interaction with the user and the deep learning to generate music.

**3.4 Collection of user requirement**

**3.4.1 Small-scale survey**

An online questionnaire has been made to understand the Audience/ video content creator opinion on YouTube video and video’s background music. The questionnaire was distributed online with snowball sampling. It starts with filter question that the participants had understood the Participant Information Statement and are aged 18 over. Followed by filter question that ensures participants have watched a video on YouTube and they are active users so as to increase the accuracy to what is really needed rather than asking a person with no knowledge or experience on the topic that the system pertains. The questionnaire asks the audience and creator to give opinions on whether they perceive and concern about the background music of a video. The importance and satisfaction of YouTube background music are also important to measure if there is room to improve on background music. The questionnaire also requests participants to rate the level of importance of the features of the system that are suggested in the questionnaire such as download songs, preview songs, and select music genre for songs to decide the most important functions. Also, participants are invited to suggest their desired function and genre in a music generator to determine what possible function could be implemented and what genre is most popular and most desired for the users.

**3.4.2 Observing similar product**

By observing AIVA, a product of AI composing emotional music, it has several functions: Users can gather AI- composed music with preset styles such as Modern Cinematic, Electronic, Fantasy, etc.; Users can upload an existing score that influences the composition.; Users can compose music on their website; Users can download music from their website. Base on the observation, a basic idea of the user requirement is created and helps to determine the questionnaire question to ensure the needs of the stakeholders

**4. Cost and benefit analysis**

Tangible

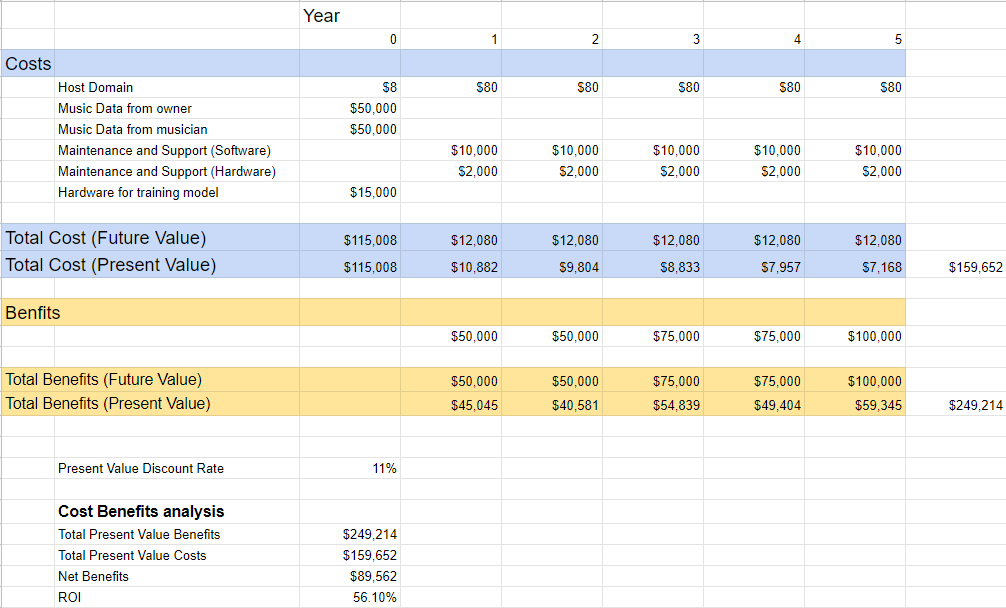


Figure 2 :Cost and benefit analysis table

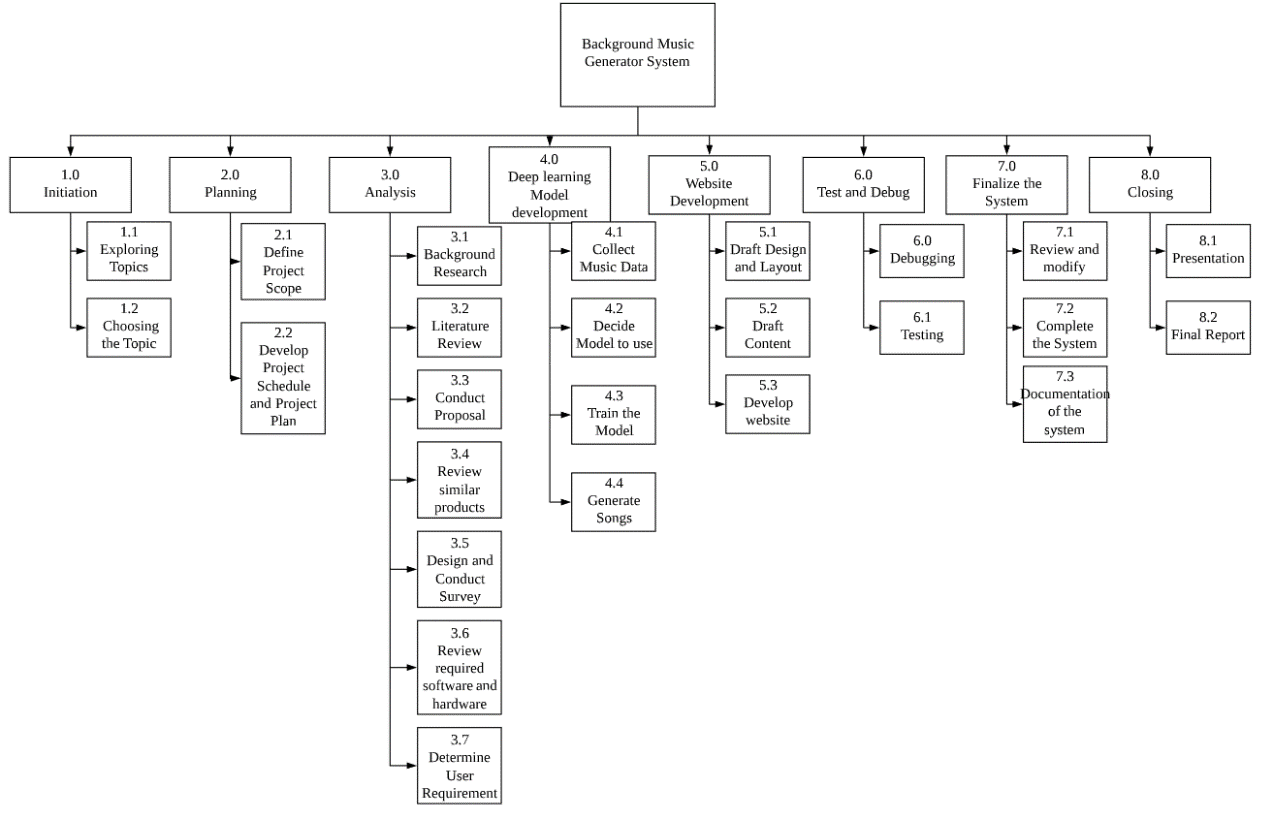
In this project, Tangible benefits are the: time that reduced from searching background music; money saved from buy music from other sources; Selling the music source as a business.

For intangible benefits,

The audience would feel satisfied as the music included in the video are fresh and high quality. This can increase the loyalty of the audience (customer). Thus, promotes the branding of the company/ creator, the video channel would become popular as the quality of video enhanced and maintain at a high level. More exposure to the public leads to more advertising and collaboration with other company to create business opportunities.

**5. Project Plan**

**5.1 Work Breakdown Structure**

Figure 3: Work Breakdown Structure

**5.2 Gantt chart**

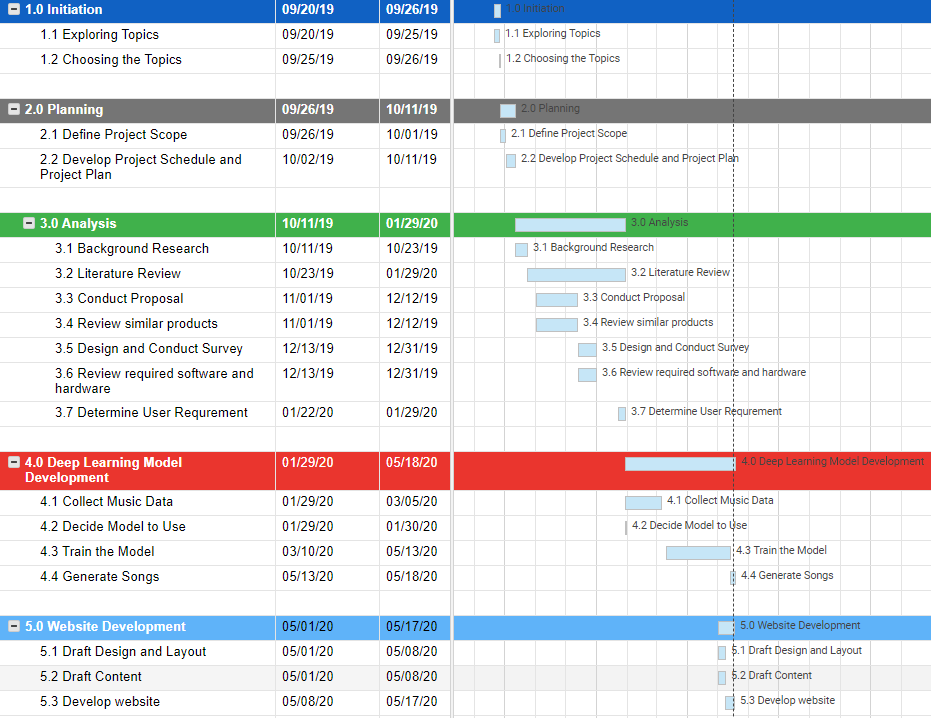


Figure 4: Gantt chart

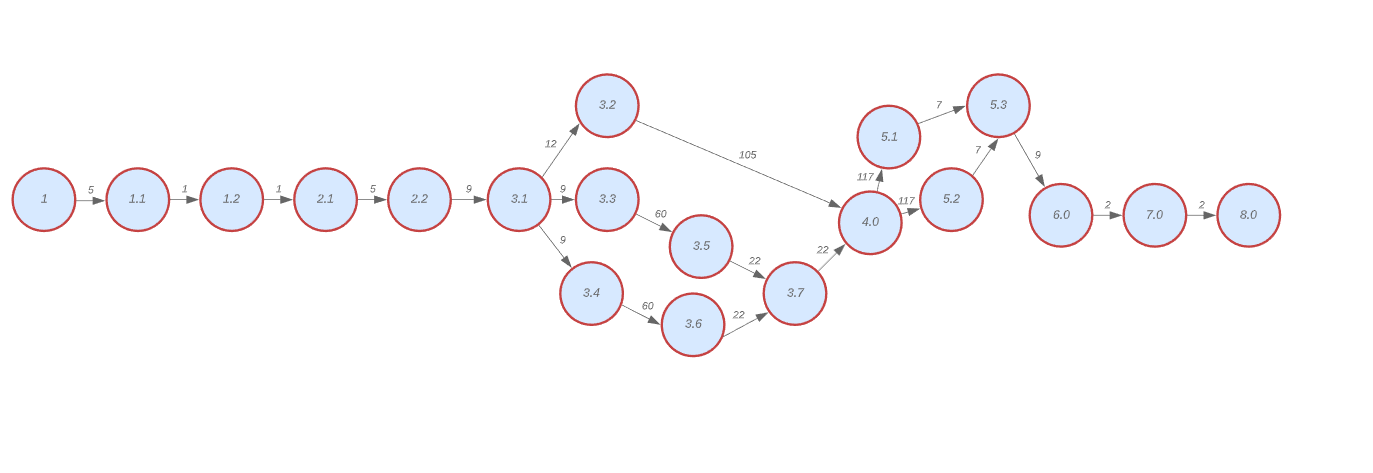
**5.3 PERT Chart**

Figure 6: PERT Chart

**6. Choice of Technology**

**6.1 Development method**

Waterfall model of Software Development Life Cycle is proposed to be used in the system development process in this project. Benefiting from its characteristics, using a waterfall SDLC model allows the project to be planned and managed easily since this project is an individual project that time is extremely urgent. Unlike SCRUM, no project manager or any certification is required for the waterfall SDLC model. The progression of the waterfall model is straight forward going through steps like Requirement Analysis, System Design, Implementation, Testing, Delivery/deployment, Maintenance. The process can be clearly outlined at the beginning of the project when planning and instantly dive into the waterfall model without having extra leaning of the methodology that delayed the progress. Moreover, using the waterfall model enable the development to pursue the goal that is set in the early stage of the process. The goal and objective would keep clear and not be lost as the project processes.

However, after proceeded to analysis and design step in the waterfall model at the beginning, some problems arose. As myself is not an expert on operating a project, any error could occur. Some functions on the requirement are not realistic to an individual project that it is difficult to achieve it. Therefore, base on the original suggested waterfall model method, some alteration has made. The iterative interaction takes place and some review and rework on the waterfall model to improve the development process of this project. The development method becomes similar to the spiral model. The spiral model has multiple loops that depend on the length of the project. Each loop of the spiral model is a phase of the development process. Each phase consists of Identification, Design, Build, Evaluation (Boehm, 1988). The development process in this project is a modified waterfall model that is involved in some sprial model process. Changes can be made in each stage as feedback and review are being made at the end of the stage, and the phase would loop, more flexibility is applied to the project.

The process is not necessary to be restricted in the particular model as every project has its different properties and requirement. The key is to have a methodology that can be led the project to the best possible result within the researcher’s ability. Therefore, the development process of this project is altered to achieve the best potential result.

**6.2 Deep learning model**

To construct an artificial intelligence system, the deep learning model is essential to achieve this mission. However, deep learning architectures such as recurrent neural network is a complicated product that requires plenty of researchers and knowledge in order to build it. It is impossible to start from scratch in an undergraduate individual project. Therefore, the existing model is the choice for developing the system in this project. There is numerous of research that has ongoing or done by scientists about music generation using deep learning.

Google’s research project Magenta has many deep learning models about making music and art such as GANsyth, an algorithm that is used for synthesizing audio with generative adversarial networks (GAN), or Melody RNN, a model applies language modelling to generate melody using a Long Short-term Memory (LSTM). For this project, Polyphony RNN is chosen because polyphony is common in modern music including background music. Polyphony is musicology term that originally describes music that has two or more independent voice simultaneously combines by the law of harmony in Middle Age (Henning Nølke, 2017).  Nowadays, it usually describes music that is performed or composed by multiple independent melody lines that form chords. In popular songs, music is always integrated by multiple melody and lines, it would not be difficult to hear multiple instruments such as bass, guitar, and drums, this music composed by multiple melody lines, which makes it polyphony. Therefore, polyphony is the type of music that the project targets. Hence, LSTM Networks used in this model allows longer dependencies, which means longer music sequence our case, to be learnt in the recurrent network that enables us to input different type of music as a dataset.

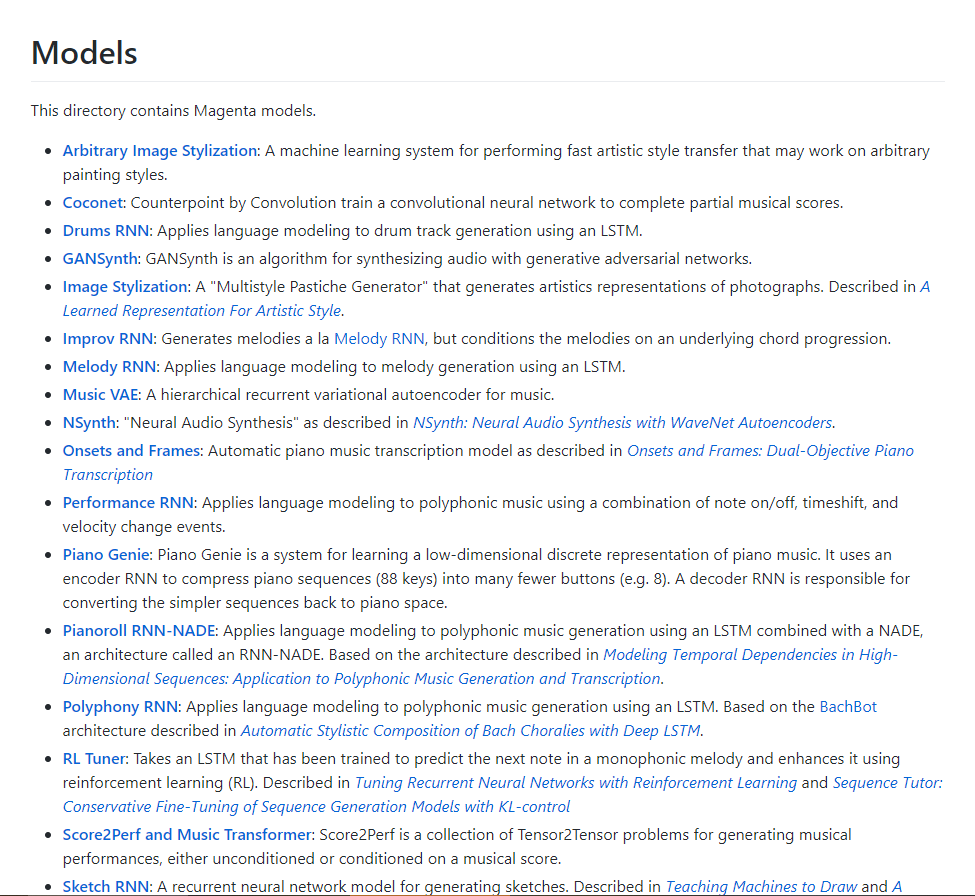


Figure 8: List of deep learning model in Magenta

**6.3 Website (Front-end)**

Website in this system not only act as front-end to allow users to interact but also providing information for the users to introduce the project idea to them. Website is the first priority of the platform to be presenting the system since it is mainly used by the video content creator and people who need to acquire background music and insert it into the video by video editing software, PC platform or Mac platform is preferred for them and website can be accessed effortlessly. Front-end development is important to a system that the usability of the front end must be significant to provide user-friendly front-end for the user to interact. The website was developed under this consideration. Visual Studio Code was used for web development. Extension of Visual Studio Code allows the programming process more smoothly by saving time and more utility that helps the programming process. For example, “Live server” extension automatically refresh the browser after changes were made on the code which saves time and steps for switching to the browser and refresh it, thus facilitating the project.

**7. System Analysis and Design**

**7.1 SWOT analysis**

SWOT means strengths, weaknesses, opportunities, and threats. The four parameter to analyze a project, individual or business. Strengths are the things that the project performed well and the benefits of the project; Weaknesses are the disadvantage things in the project, components that are poorly performed in the project, or something that could be improved and avoid. Those are the internal elements and factors of the project. Opportunities are the things are the possible chance that would happen in the future, through further action to achieve the potential advantages; Threats are the potential negative effect the could harm the project/business form the external way. Those are the external elements and factors of the project.

|  |  |
| --- | --- |
| **Strengths**   * Provide an alternative way to produce music * Reduce time and cost for the user | **Weaknesses**   * Hard to improve in terms of algorithm * Not good as human compose music at the beginning |
| **Opportunities**   * Could build up a business of providing music * Make the video in media more appeal * Improvement and quality can easily snowball | **Threats**   * Hard to compete or take over similar products in the market that have researched for a long time |

**7.2 System Design and Development**

The music generator system consists of front-end, the website, and the backend, the deep learning model for generating music. The function and content of the website is base on and gather from the small-scale survey that conducted before, and the inspiration from the similar product that observes before.

**7.3 Design of the website and deep learning model**

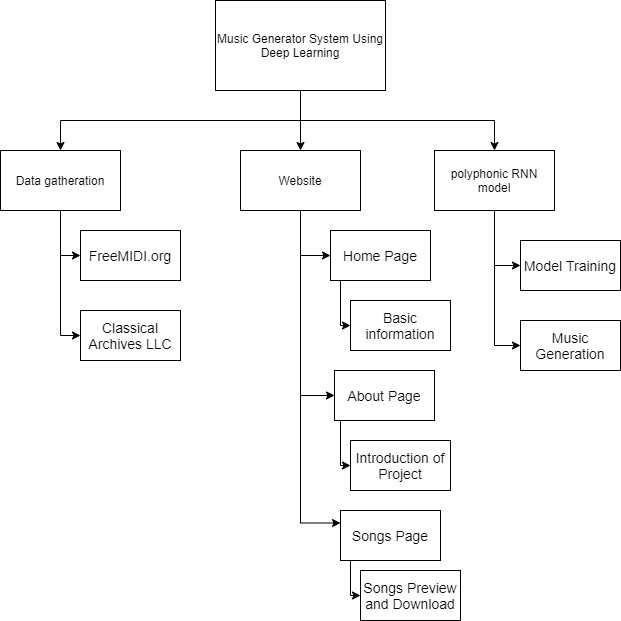


Figure 9: Basic Structure of the System

According to the result of the survey conducted, 35 qualified participants said that the most important features of the system would be (1) download songs, (2) Select music genre, (3) Preview songs. The mean of these 3 features is 4.7, 4.6, and 4.5 respectively. The mean of other features is below 3, which shows that those features are not as important as the download, preview and select genre functions. Therefore, only these 3 features would be included in the release of the system at the beginning.

* Download songs
* Select music genre
* Preview songs

Furthermore, the targeted genre would be pop, rock, and classical. According to the suggestion of the 35 qualified participants of the survey, 8 answers (instrumental/acoustic /piano/ classical), which classify as Classical, 9 answers pop/ pop song, 5 answer rock. Others are not considered as there are only 1-2 answers.

After decided the 3 main features: (1) download songs, (2) Select music genre, (3) Preview songs, more elements are needed to enrich the content of the website. Therefore, 3 pages of the website are determined: Homepage, About, and the Songs page. The ‘Homepage’ is the page that contains the basic information of the system and project. The ‘About’ page is a brief introduction of the project that talks about the abstract of the project: What is the aim of the project; What method and technology are chosen to be used in the project; Where is the data gathered from.

Also, services/ features that the system covers that is also included in that page for the user to be more familiar to the product. Finally, the ‘Songs’ page is the most important part of the product. In this page, samples of the songs of the 3 genres (Classical, Pop, Rock) that are produced by the RNN model would be provided for the users to preview and listen. The songs are uploaded to SoundCloud for the users to listen. ‘Download’ button is exists in the SoundCloud API widget for user to download the songs in Wav format for editing and video purpose.

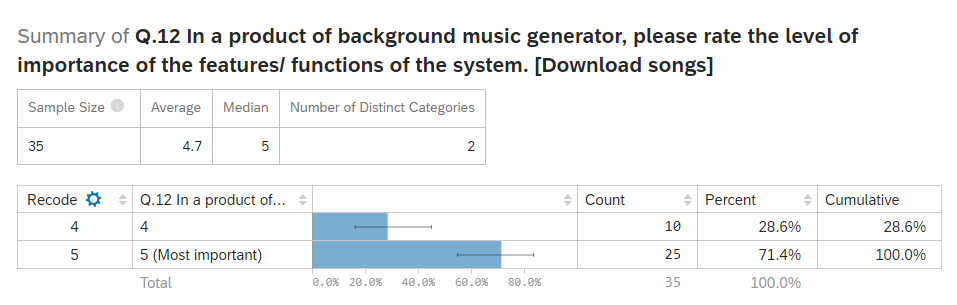


Figure 10: Level of importance of function ‘Download Songs’

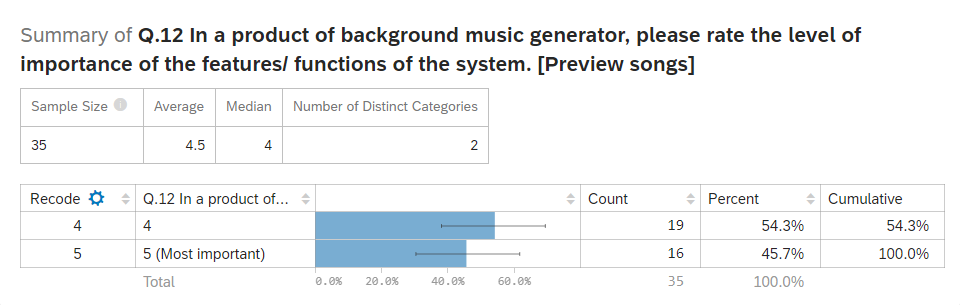


Figure 11: Level of importance of function ‘Preview Songs’

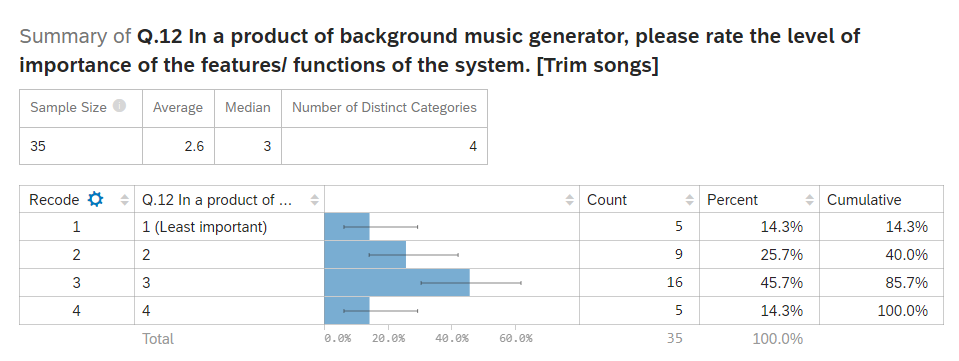


Figure 12: Level of importance of function ‘Trim Songs’

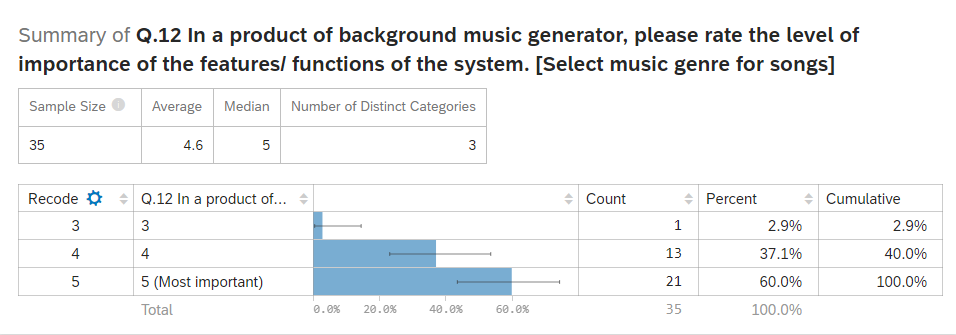


Figure 13: Level of importance of function ‘Select Music Genre for Songs’

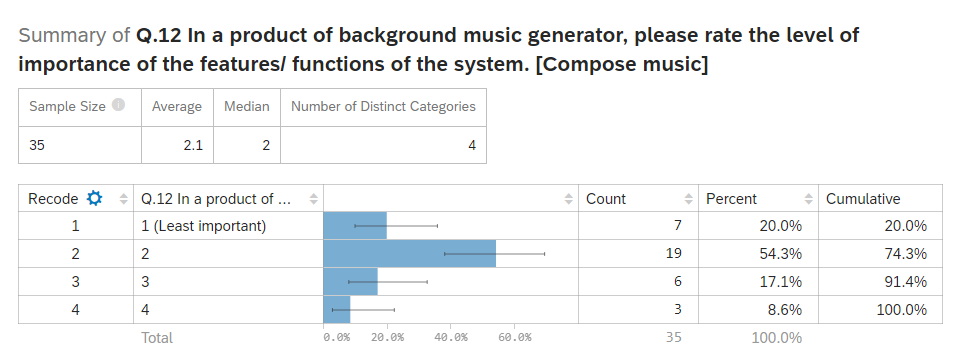


Figure 14: Level of importance of function ‘Compose Music’

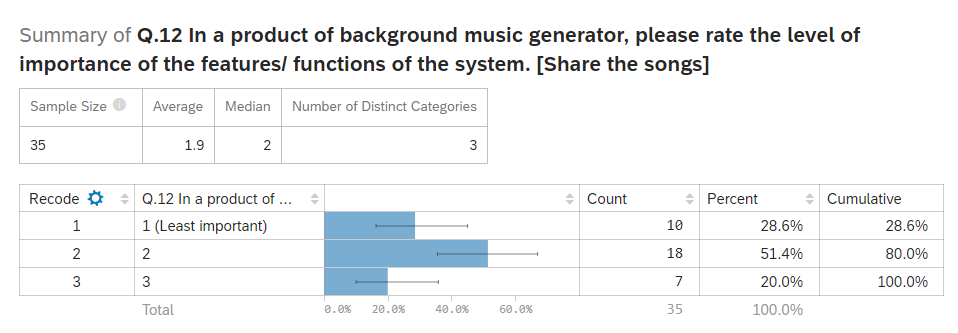


Figure 15: Level of importance of function ‘Share the Songs’

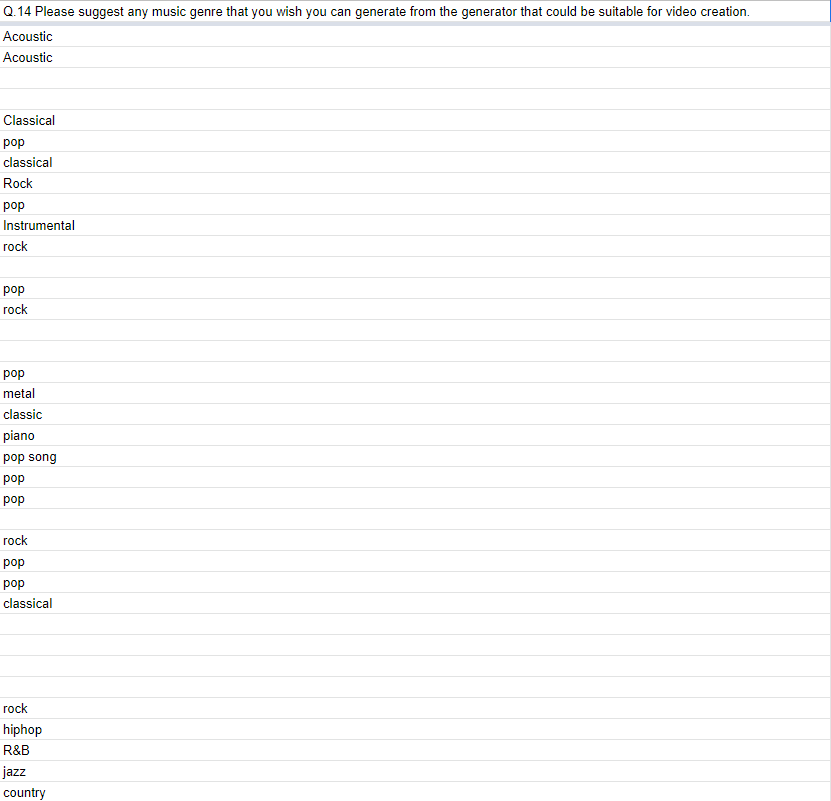


Figure 16: Participants suggestion on genre that the system targets

**7.4 Website Overview**

The website is accessable through <https://www.itb-044.website/>

**7.4.1 Home Page**

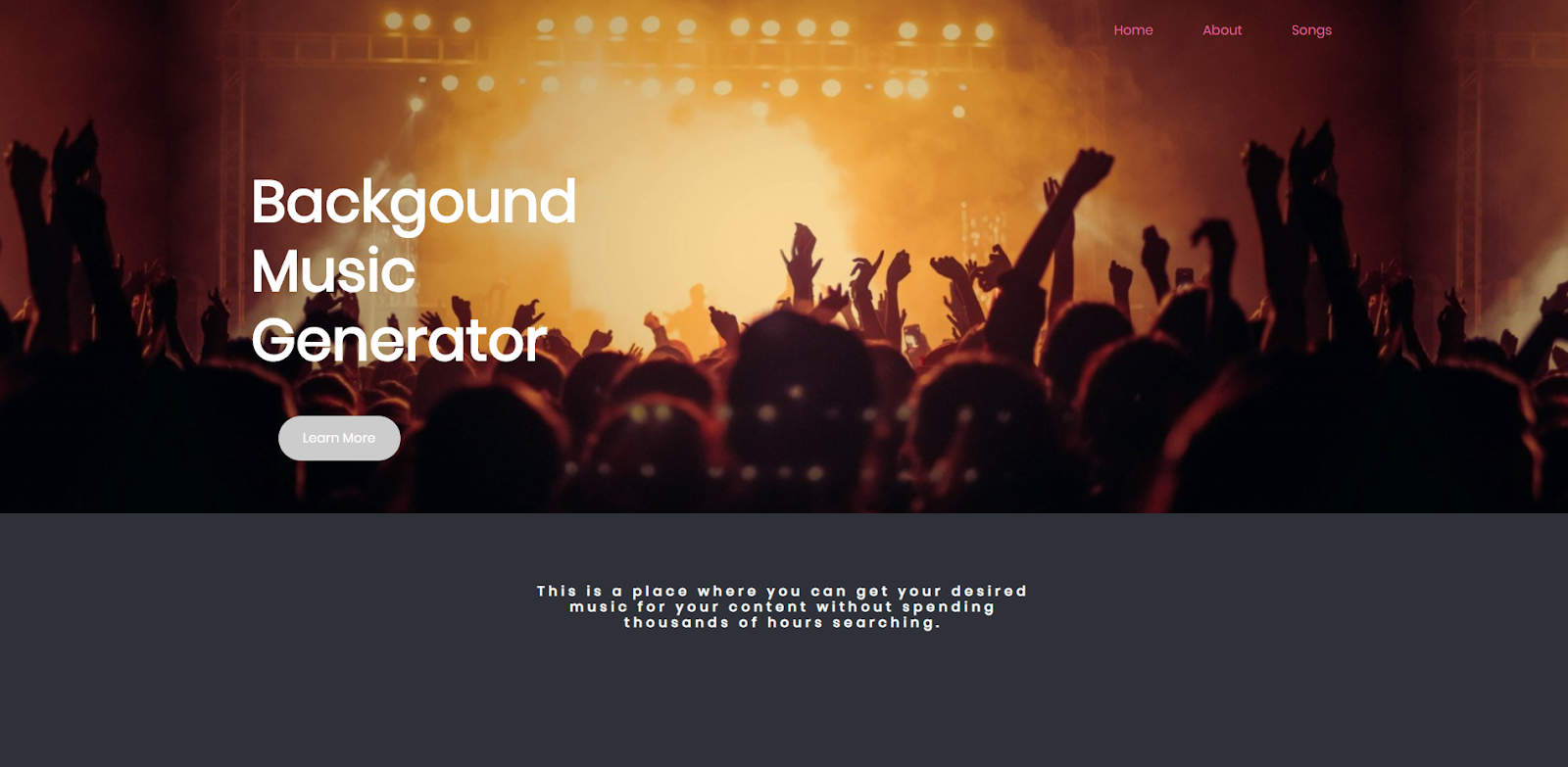


Figure 17: Home page

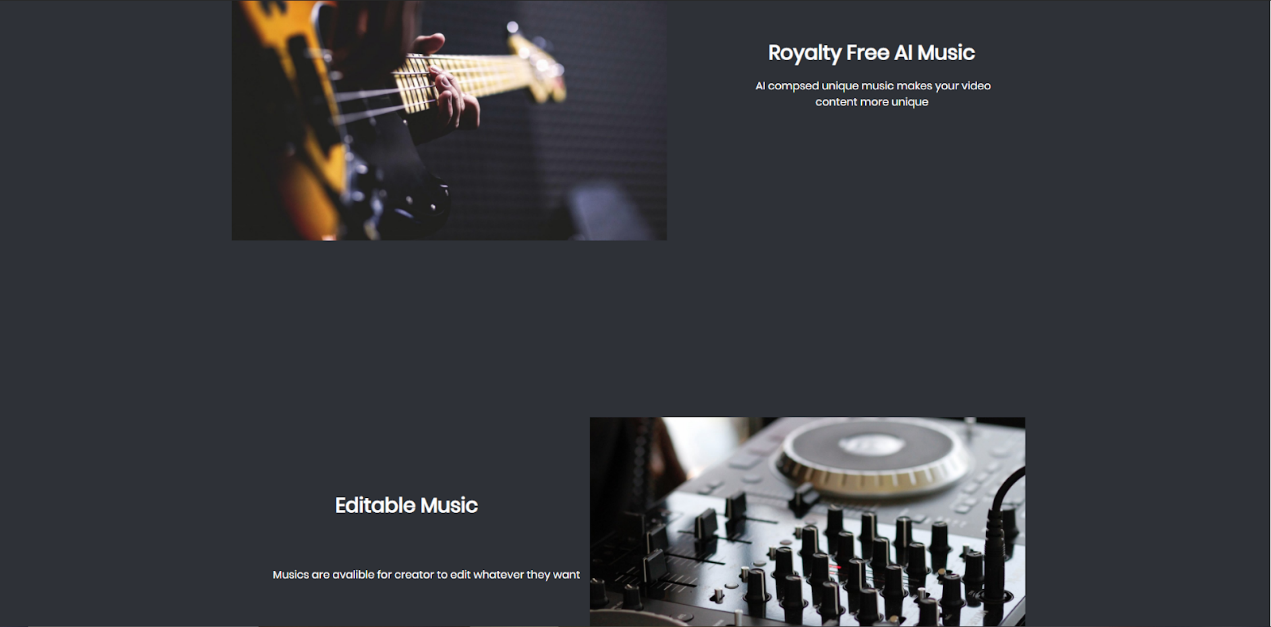


Figure 18: Home page (2)

After accessing into the website, a home page is provided some basic information to the user.

At the top of the home page, ‘Home’ button, ‘About’ button, and ‘Songs’ button are provided for the user to access to other pages.

**7.4.2 About Page**

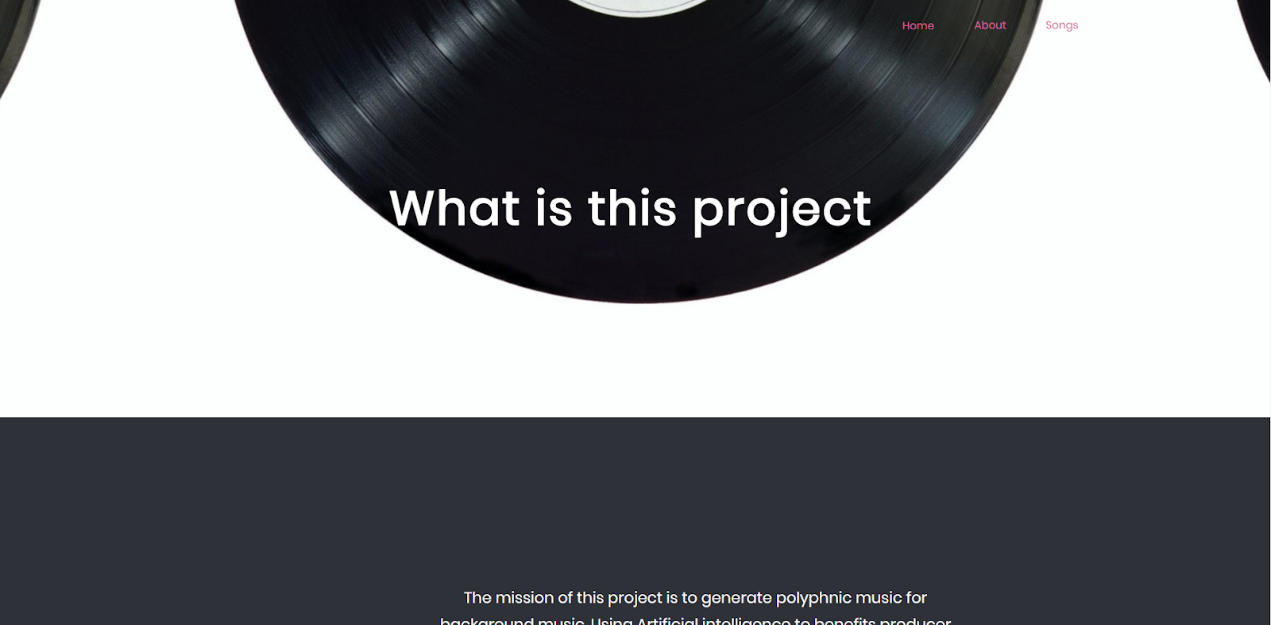


Figure 19: About page

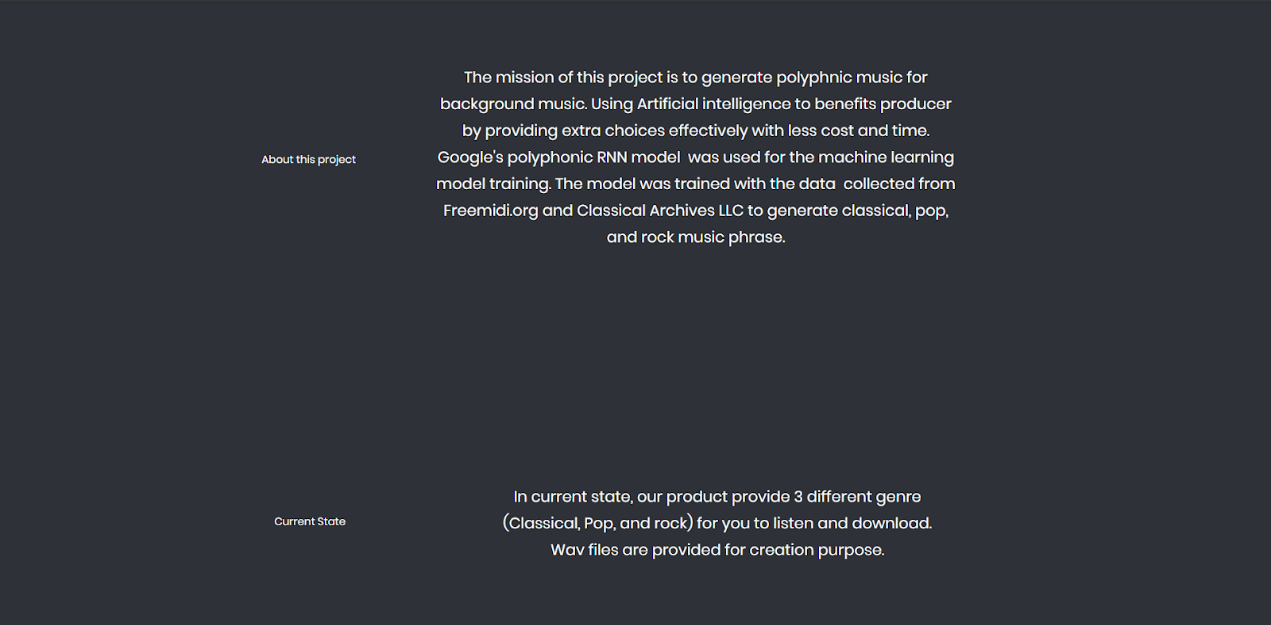


Figure 20: About page (2)

In this page, more information about this project and the current state of the project is provided.

**7.4.3 Songs Page**

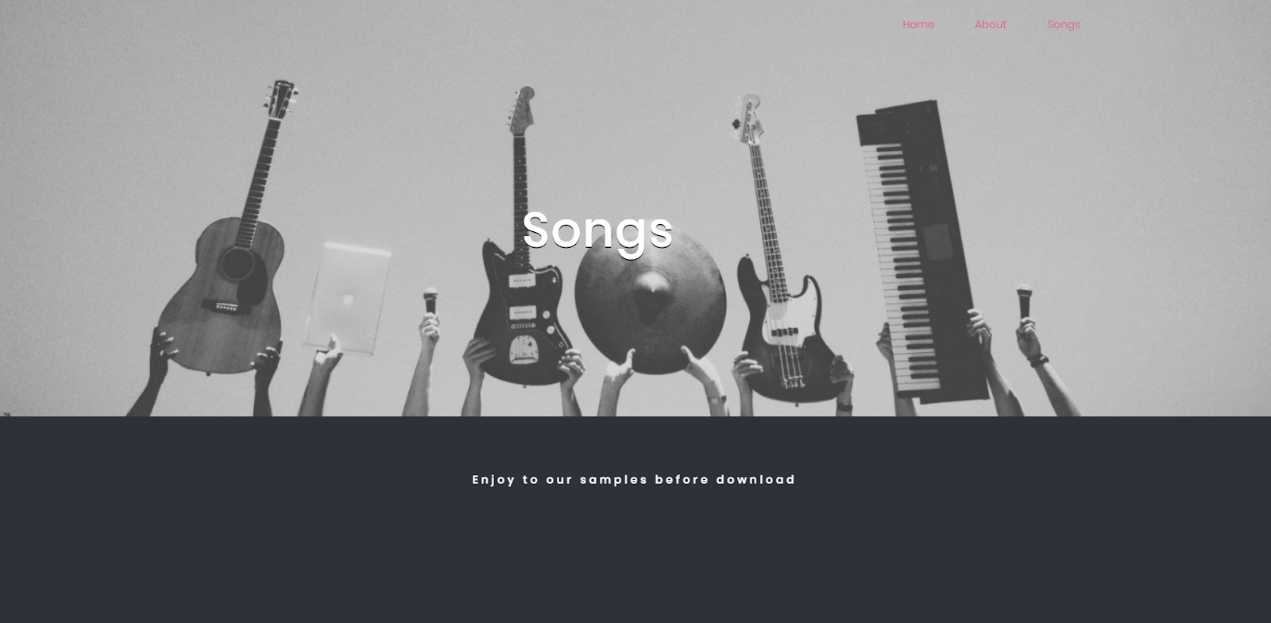


Figure 21: Songs page

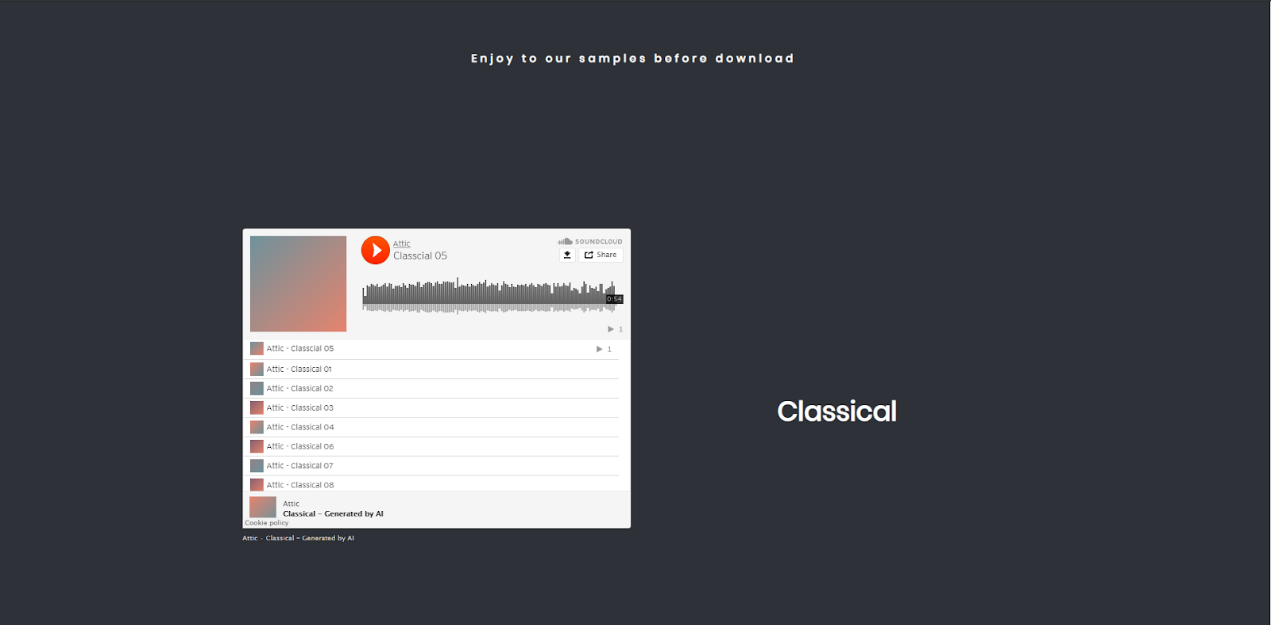


Figure 22: Songs page (2)

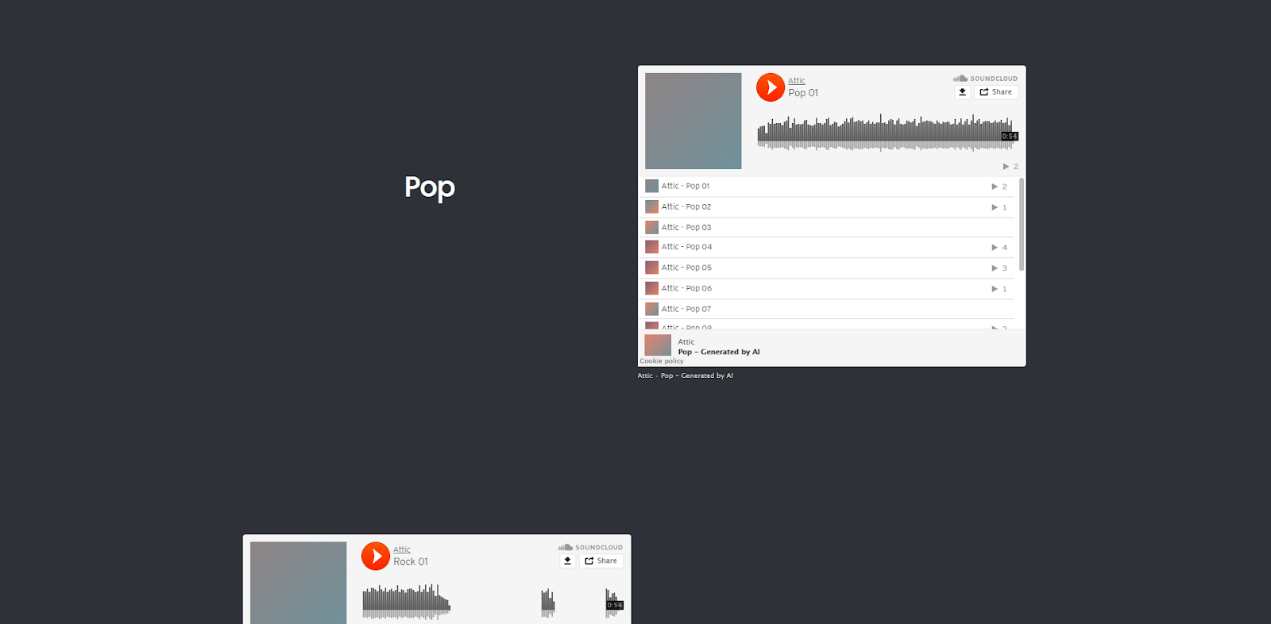


Figure 23: Songs page (3)

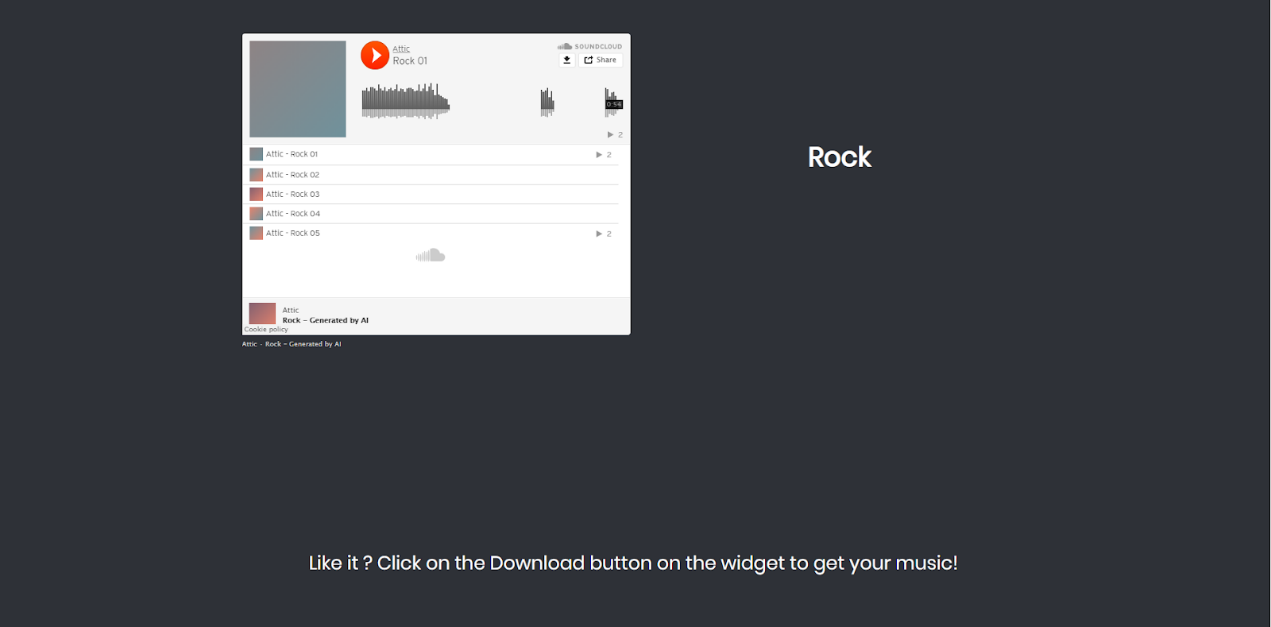


Figure 24: Songs page (4)

In this page provides functions for the user

* Previews
* Download

User can access to these two functions by clicking on the Play button on the widget and Download button on the widget of each genre.

**7.5 Test Plan**

|  |  |  |
| --- | --- | --- |
| Test Case ID | Test item | Description |
| 01 | Access of the website | Test if the website is accessible |
| 02 | Home page redirect check | Test if the button can redirect user to another page from home page |
| 03 | Preview function check | Test if the SoundCloud API works well for user to listen to the playlist |
| 04 | Download function check | Test if the SoundCloud API widget works well to download songs (wav file) |
| 05 | Redirect check | Test if the button on other pages can redirect user to specific page according to the button |

**7.6 Test Case**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test ID | Action | Expected Outcome | Actual Outcome | Pass/Fail |
| 01 | Enter project website URL (https://www.itb-044.website/) | Access to the home page of website | Accessed to the home page of website | Pass |
| 02 | Click on the buttons on the homepage | (1) When click on ‘About’, redirect to ‘About’ page.  (2) When click on ‘Songs’, redirect to ‘Songs’ page.  (3) When click on ‘Home’ page, Remain in the homepage. | All three button (1), (2),(3) redirect to the specific page | Pass |
| 03 | Click on the ‘Play’ button of SoundCloud API of each genre playlist | All player can play the music in the playlist | All player can play the music in the playlist | Pass |
| 04 | Click on the SoundCloud  API widget’s download button | Download the song (wav file) successfully | Download the song (wav file) successfully | Pass |
| 05 | Click on the buttons on ‘About’ and ‘Songs’ | Redirect to the specific page according the buttons in each page | Redirect to the specific page according the buttons in each page | Pass |

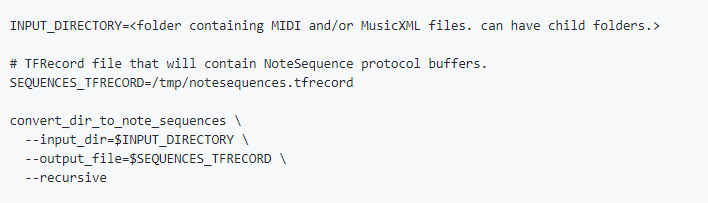
**7.7 Deep Learning model training**

Besides choosing the suitable deep learning model for the system, utilizing it in the right way is crucial to achieving the objective in the project. To achieve the objectives, 3 datasets have been created (Classical, Pop, Rock) for training. All data in the dataset is collected from ‘Freemidi.org’ and ‘Classical Archives LLC’. In the dataset of Classical music, 65 songs of Classical and Instrumental music are collected in MIDI file. MIDI, also known as Musical Instrument Digital Interface, is a technical standard for the communication of musical information between different musical instruments such as synthesizers, keyboard, computers and all other electronic instruments (Mulder and Rothstein, 1993). MIDI carries messages and data of the music and indicate the instructions for music such as note's notation, pitch, velocity, which creates loudness and softness of the music, vibrato, and position of the stereo.

The MIDI file is used in the training part of the deep learning model. MIDI files of each genre are separated into a different folder, each folder is the dataset that responsible for the model training. There will be 3 different models for Classical music, Pop music and Rock music, training different datasets to generate different music genre. For classical music, 65 different songs are used as the dataset. For pop music, 88 different songs are used as the dataset. And for rock music, 94 different songs are used as the dataset.

To start the training, the MIDI files must convert into NoteSequence which is a protocol buffers that are developed by Google. Protocol Buffers is a method for serializing structured data that is language-neutral, platform-neutral, and extensible. It is a faster and efficient data format and easier to work with for training (Google, 2019)

The first step of creating datasets is to create a directory of MIDI files and converting them into NoteSequences by the following command to generate TFrecord file of NoteSequences.

  
Fig 25: Command to convert MIDI files to NoteSequences.

The MIDI files in the directory will be converted into a ‘notesequences.tfrecord’ file for further training.

After creating NoteSequences, the next step is to create SequenceExamples that are fed into the RNN model for training and evaluation. SequenceExamples include a sequence of labels of a polyphonic sequence and a sequence of inputs. The polyphonic sequences are extracted from the NoteSequences that previously converted. Two sets of SequenceExamples will be generated, one for training, and one for evaluation. The fraction of SequenceExamples in the evaluation set is controlled by the parameter of the command. In this project, all 3 models are using 0.1 of evaluation ratio, which means 10% of the extracted polyphonic tracks will be saved in the evaluation set, and the remaining will be saved in the training set.

After the preparation, the model training and evaluation can begin. Different configuration of the attention mechanism of the polyphonic RNN model could be alter by the commands.

For example, training steps could be adjusted by the command ‘--num\_training\_steps’, controlling how many update steps to going to process before the finishing of the training step, or run the training loop until manually stopped.

Moreover, hyperparameters can also be adjusted by the command ‘--hparams’. The default is 128, it can be adjusted depending the hardware we use. Smaller batch sizes reduce memory usage during training, which can prevent sort of the issues that cause by not enough memory of the training hardware when training on large model. In the training part of this project, 64 batch size, and 3-layer RNN with 128 batch size of each layer which allow the model to train fast meanwhile not overload the computer. All 3 model are trained with same parameter but with different dataset (Classical, Rock, Pop)

After the training has been done, or at the middle of the training, the polyphonic track can be generated from the checkpoint of the training to observe the result. Multiple polyphonic tracks that will be generated. Tracks will be selected and upload to SoundCloud.

**7.8 Deep Learning Data pipeline**

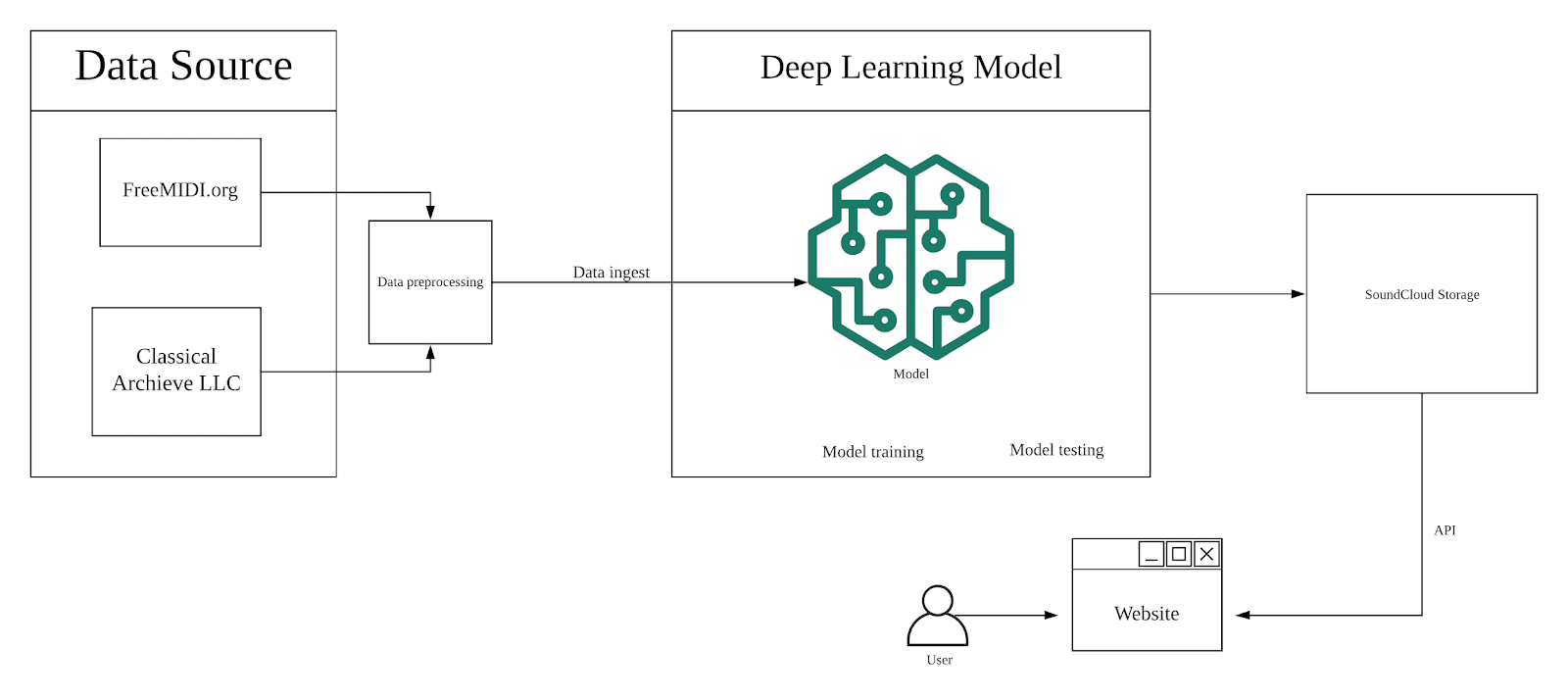


Fig 26: Infrastructure of the System

**8. Further Development**

In the future, the system can have update and improvement on certain ways, mainly of quality of music or the system itself.

In terms of the quality of music, the music is not solid enough to attract user and customer to have loyalty on this product. The music generated by this product only composed by piano, which is not all-arounded and cannot convey different emotion through the notes and melody.

In the future, hopefully more instrument would be implemented into the generation of the music, including bass, guitar, drums, electric synthesizer, not only limited to piano. Also, the quality of the music can be enhanced by acquiring more source of data, which is music file in this project, developing a much more complete dataset for a better training of the model to generate more diverse and superior music. In addition, more different deep learning model could be applied to achieve different requirement and usage on music generation. For example, different melody could be adjusted and create different instrument can be apply to the music with different model. Music knowledge is also important for further development as some parameter are intent to create certain music structure. Understanding the music theory allows developer to have better control and adjustment on the music generator to generate better music with better quality.

Also, more genre would likely to be produced to the generator such as fantasy, jazz, hip hop, etc. Providing more option to the user and suits their preference to the music, enhancing more on the video quality. Furthermore, some features that stated in the user requirement survey are not completely implemented into the system such as selecting mood, and select length. Considering these functions as the future features to published can enrich the system also providing more option to the user according their personal preferences. Also, the workflow of the music could improve as the music is not generated to the user by one-click, instant-generating music would be a great idea to make user have a better experience of feeling of the product is come from AI generation.

**9. Conclusion**

In conclusion, the system has been successfully displayed and implemented to provide music to the users. In the homepage, users can have a basic introduction about ‘What is this project?’ and ‘What this project does?’ with a clear and clean webpage. In the home page, the title of the project ‘Background Music Generator’ present the product name about what is this product to the users, and below are some features of the system which intent to attract users with it.  Users can access to other pages through the homepage such as ‘About’ for more information and ‘Songs’ for the preview and download of the songs. In the ‘About’ page, more information can be seen. More information about the project and the system such as aim, objective, methodology is provided to the user through this page, a basic introduction on the product features are also supported. In the Songs page, all the functions are work as expected with the outline that stated in the design and planning phase of the project. The preview functions in the website is connected to the SoundCloud API which can upload the songs that generated by the system and present to the users for listen in a playlist. 3 Playlist is provided to the users responsible for the Classical, Pop, and Rock genre respectively. Users can have an introduction on what this project is , and access to another page through clicking the button. Users can preview the songs of different genre, knowing what the music that generated from AI system is going to be, if they are satisfied, they can download it through the download button on the SoundCloud API widget.

Although the system is implemented as expected, there are some flaws on the project. The project is not as competence as other similar product such as AIVA. It may cause by the project scale and the knowledge and competence of myself. Those products exists in the market is developed by large scale of researchers, data scientists who are the experts in deep learning field, done numerous of research and practice to achieve such successful product.

In this project, the music quality and the system itself is not good as some similar product like AIVA. Music quality is acceptable but not enough: the music is only composed by piano with multiple melody played simultaneously, no other instrument is used that cause the song might not providing enough fulfillment to the user. The classical music and the pop music are decent with more than 10 tracks, but the rock music are only able to generate about 5 tracks with different structure, it may because of the data source. The rock music usually composed by multiple instrument with different role, only using polyphonic model might not be able to handle it, or not able to response to certain instrument in the data.

Also, only basic features are provided, there are plenty of room of improvement that can be made to enhance the system.

**10. Self-Reflection**

Plenty of experience and technique are acquired throughout the whole process of this project for me. After operating this project, I have learnt to handle a whole project from managing the project to development by my own. It might not be a excellent job, but still a great contribution to equip myself. To achieve a best possible result of a project, a proper planning on schedule and time management is crucial. Unfortunately, it was not perfect for my job. It might due to my personality also the current situation throughout this academic year, it was a hard time to balance the academic work and the society. Different tools are useful for the development and project management such as smartsheet and LucidChart, which provides me a better place to organize the workflow and some essential graphs that are helpful on development and attach to the report.

On development, I have better understanding on deep learning and website development. I have read a large number of different articles on RNN network including LSTM and attention RNN, and tried to explain it in a short summary. It requires me to have a sufficient understanding on how the neural network works. This is the first time I develop a AI-based product. Although the neural network is not build by myself, it stills a good experience on hands-on deep learning product. Moreover, this is my first time to develop a web from scratch, however, if I have another opportunity, I would choose another way to develop the webpage as there are plenty of web builder service and website template. In my opinion, the only concern is efficiency for a business. Despite using template or web builder service might not able to apply the knowledge and skills I have learnt, using these services might reduce the cost and time for a company. Nonetheless, in this project I still tried to build it by my own.

However, due to limitation of time and access of data, the quality of the system has been affected. Access is a big problem when developing this project. The copyright requires a high cost to get a good quality of music data as it makes harder to train and generate good quality of music. Hence, it increases the time and cost of the development. Besides, Producing the system that can compete with current existing products is really difficult not only with these factors but also in an individual undergraduate project.

**Reference**

AIVA (2016). *About AIVA*. [online] www.aiva.ai. Available at: https://www.aiva.ai/about#about [Accessed 10 Jan. 2020].

Alpert, J.I. and Alpert, M.I. (1990). Music influences on mood and purchase intentions. *Psychology and Marketing*, 7(2), pp.109–133.

Bahdanau, D., Cho, K. and Bengio, Y. (2015). *Neural Machine Translation by Jointly Learning to Align and Translate*. [online] Available at: https://arxiv.org/pdf/1409.0473.pdf [Accessed 22 Apr. 2020].

Bassil, Y. (2012). A Simulation Model for the Waterfall Software Development Life Cycle. *International Journal of Engineering & Technology (iJET)*, [online] 2(5), pp.2049–3444. Available at: https://arxiv.org/ftp/arxiv/papers/1205/1205.6904.pdf [Accessed 1 Apr. 2020].

Boehm, B.W. (1988). A Spiral Model of Software Development and Enhancement. *Computer*, 21(5), pp.61–72.

Brondel, L. and Cabanac, M. (2007). Alliesthesia in visual and auditory sensations from environmental signals. *Physiology & Behavior*, 91(2–3), pp.196–201.

Cambridge Dictionary (2019). *SOUNDTRACK | definition in the Cambridge English Dictionary*. [online] Cambridge.org. Available at: https://dictionary.cambridge.org/us/dictionary/english/soundtrack [Accessed 11 Dec. 2019].

Collins Dictionary (2019). *Definition of background music*. [online] Collinsdictionary.com. Available at: https://www.collinsdictionary.com/dictionary/english/background-music [Accessed 13 Dec. 2019].

David Lynton Poole and Alan Keith Mackworth (2006). *Artificial intelligence foundations of computational agents*. Cambridge, United Kingdom Cambridge University Press.

Edwards, J. (2015). *Statistics on how much money YouTube stars make in sales and revenue - Business Insider*. [online] Business Insider. Available at: https://www.businessinsider.com/money-youtube-stars-actually-make-2014-2 [Accessed 22 Nov. 2019].

eMarketer (2018). *Video Swells to 25% of US Digital Ad Spending*. [online] eMarketer. Available at: https://www.emarketer.com/content/video-swells-to-25-of-us-digital-ad-spending [Accessed 25 Nov. 2019].

Google (2019). *Developer Guide  |  Protocol Buffers  |  Google Developers*. [online] Google Developers. Available at: https://developers.google.com/protocol-buffers/docs/overview.

Google (n.d.). *Magenta*. [online] Google Research. Available at: https://research.google/teams/brain/magenta/ [Accessed 30 Dec. 2019].

Henning Nølke (2017). *Linguistic Polyphony : the Scandinavian Approach: ScaPoLine*. Brill, Leiden.

Kaplan, B. and Nimmer, M.B. (1965). Nimmer on Copyright. *Harvard Law Review*, 78(5), p.1094.

Kat Tenbarge (2019). *Influencers warn kids that being a YouTuber might not be all good - Insider*. [online] Insider. Available at: https://www.insider.com/youtuber-top-career-choice-for-us-kids-teens-2019-8 [Accessed 14 Nov. 2019].

Liang, F., Gotham, M., Tomczak, M., Johnson, M. and Shotton, J. (n.d.). *The BachBot Challenge: Man VS Machine*. [online] bachbot.com. Available at: http://bachbot.com/#/?\_k=rw5em3 [Accessed 21 Apr. 2020].

Lister, M., Dovey, J., Giddings, S., Grant, I. and Kelly, K. (2003). *New media : a critical introduction*. London ; New York: Routledge, , Cop.

Marino-Nachison, D. (2019). *Alphabet Should Spin Off YouTube, Analyst Writes*. [online] Barrons.com. Available at: https://www.barrons.com/articles/alphabet-google-spin-off-youtube-51551795837 [Accessed 20 Nov. 2019].

Mulder, A. and Rothstein, J. (1993). MIDI: a Comprehensive Introduction. *Leonardo*, 26(3), p.266.

Mumford, M.D. (2003). Where Have We Been, Where Are We Going? Taking Stock in Creativity Research. *Creativity Research Journal*, 15(2–3), pp.107–120.

Najjar, L.J. (1996). Multimedia Information and Learning. *Jl. of Educational Multimedia and Hypermedia*, [online] 5(2). Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.118.1654&rep=rep1&type=pdf [Accessed 10 Apr. 2020].

Reed, T.R., Reed, N.E. and Fritzson, P. (2004). Heart sound analysis for symptom detection and computer-aided diagnosis. *Simulation Modelling Practice and Theory*, [online] 12(2), pp.129–146. Available at: https://www.ida.liu.se/labs/pelab/publications/documents/2004/2004-02-ReedFritzson-SIMPRA-heartdiagnosis.pdf [Accessed 1 Nov. 2019].

Spangler, T. (2019). *YouTube Now Has 2 Billion Monthly Users, Who Watch 250 Million Hours on TV Screens Daily*. [online] Variety. Available at: https://variety.com/2019/digital/news/youtube-2-billion-users-tv-screen-watch-time-hours-1203204267/.

We Are Social and Hootsuite (2019a). *DataReportal – Global Digital Insights*. [online] DataReportal – Global Digital Insights. Available at: https://datareportal.com/reports/digital-2019-global-digital-overview?rq=digital%202019%3A%20global [Accessed 12 Dec. 2019].

We Are Social and Hootsuite (2019b). *Internet Use by Device January 2019*. *DataReportal*.

We Are Social and Hootsuite (2019c). *Time per Day Spent Using the Internet*. *DataReportal*.

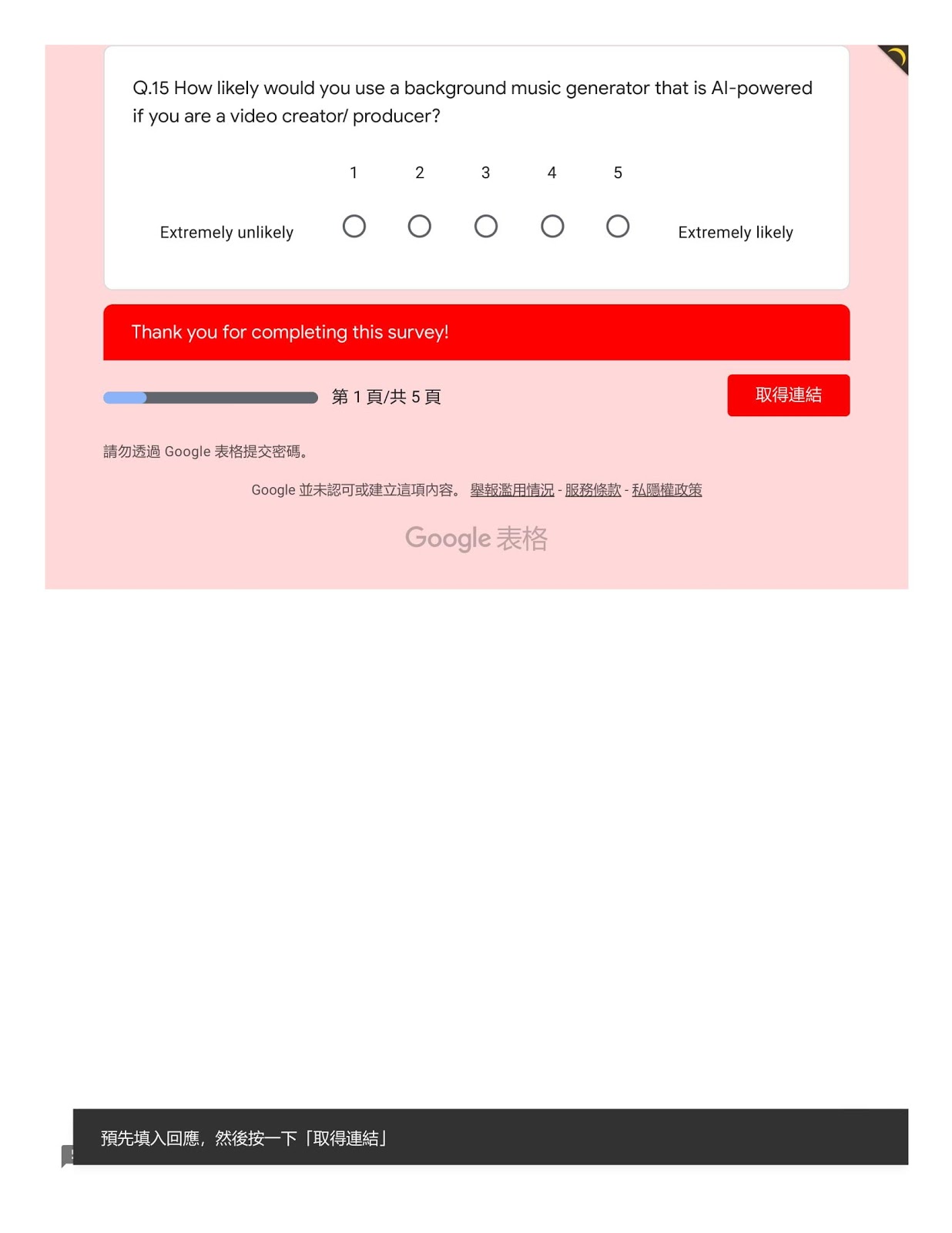
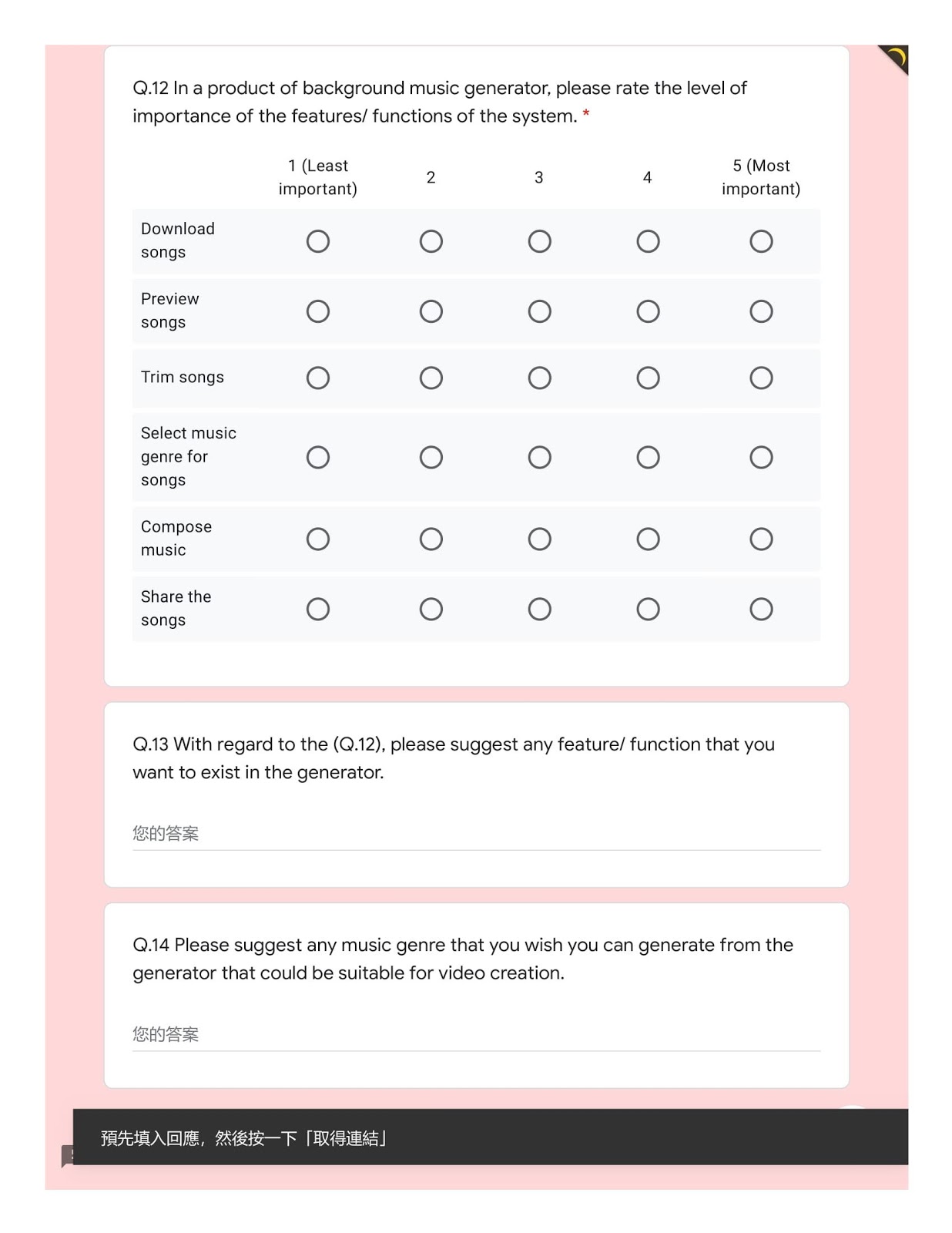
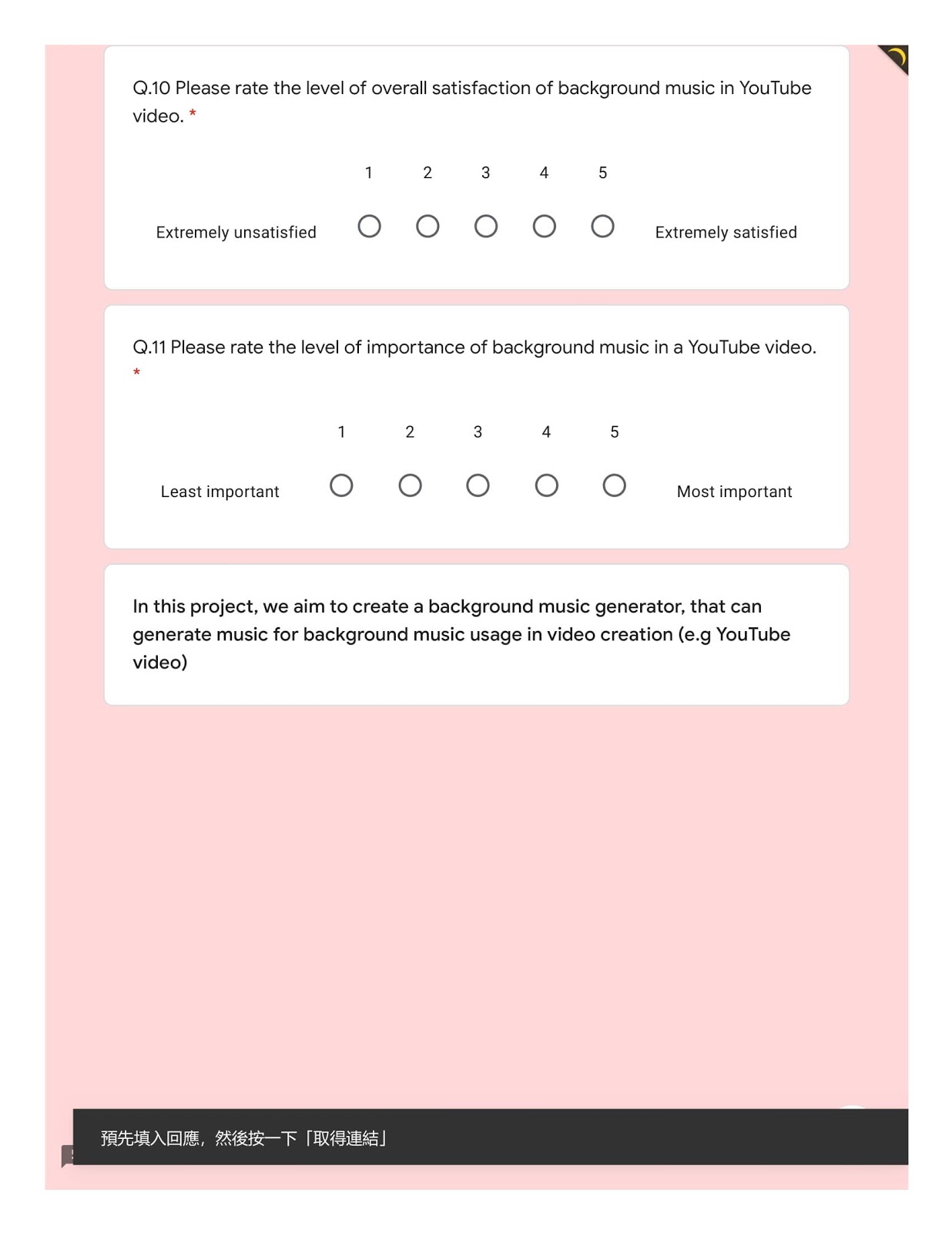
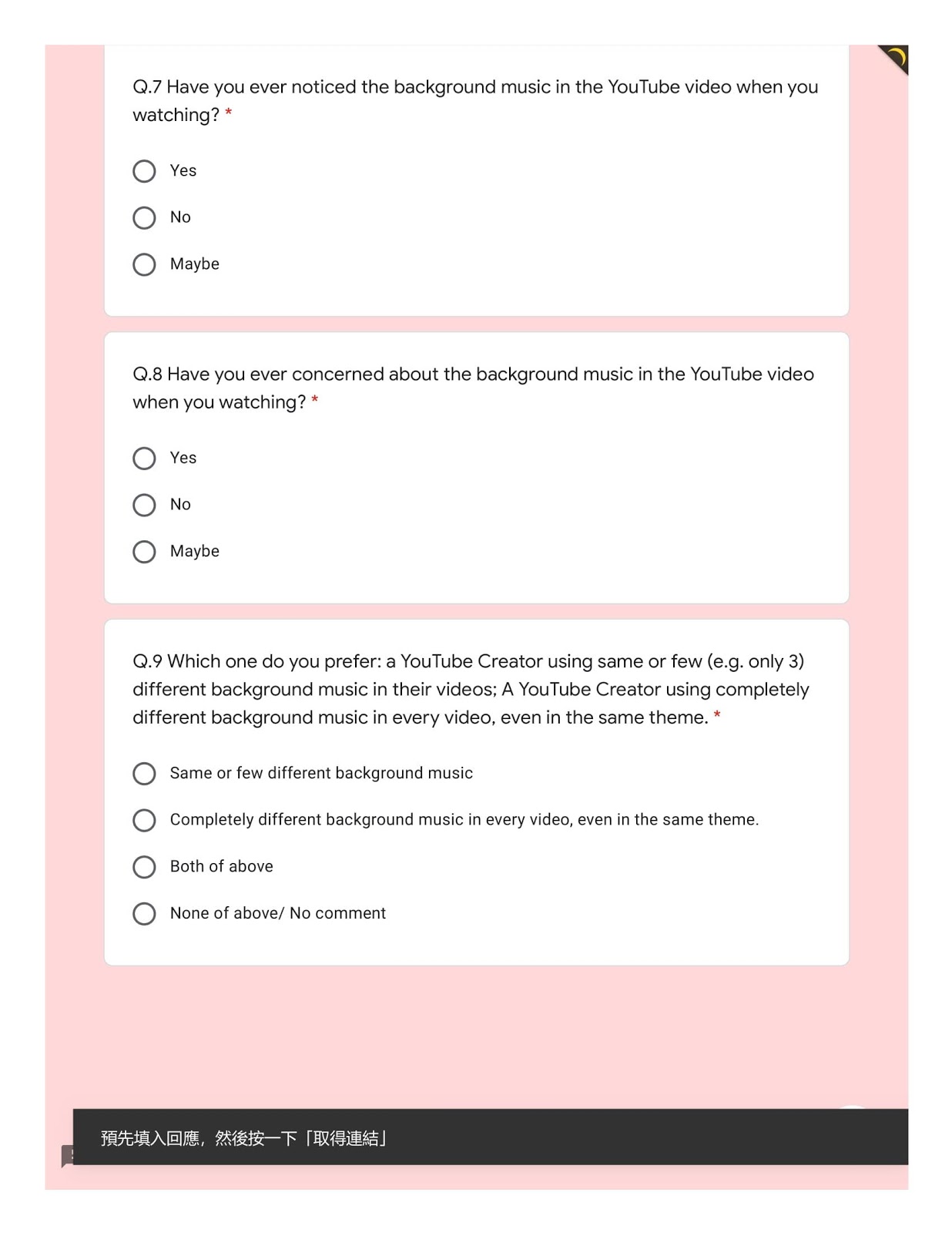
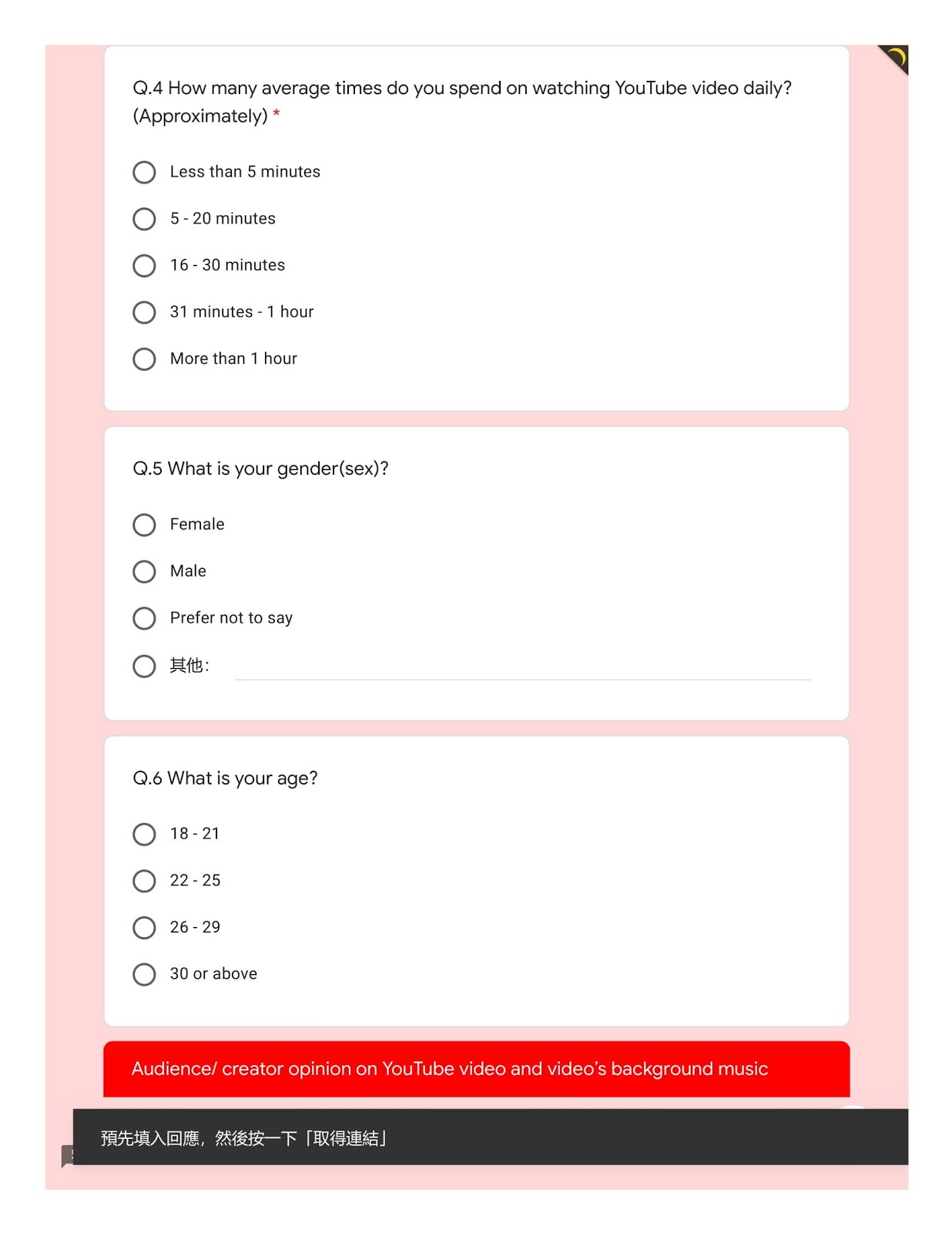
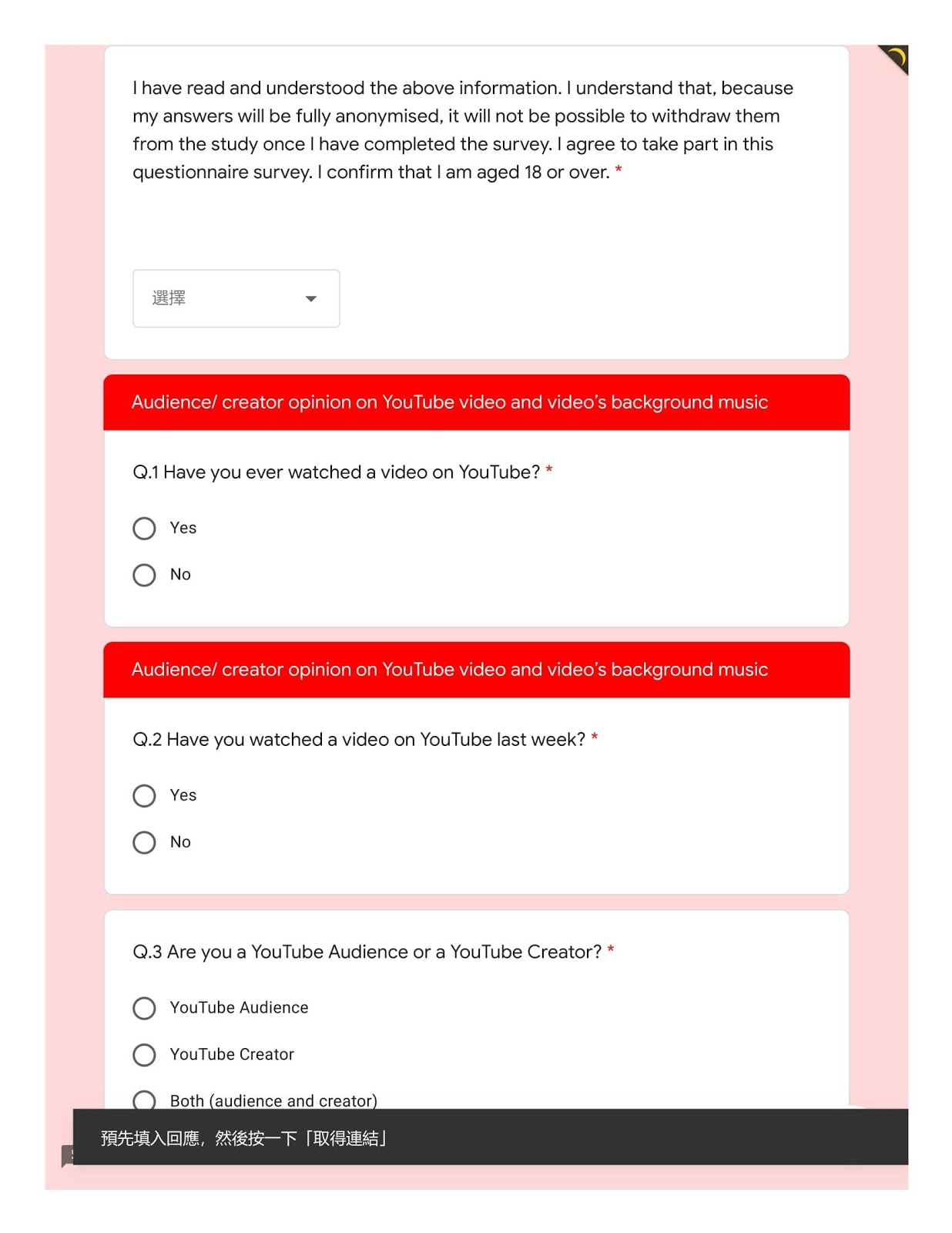
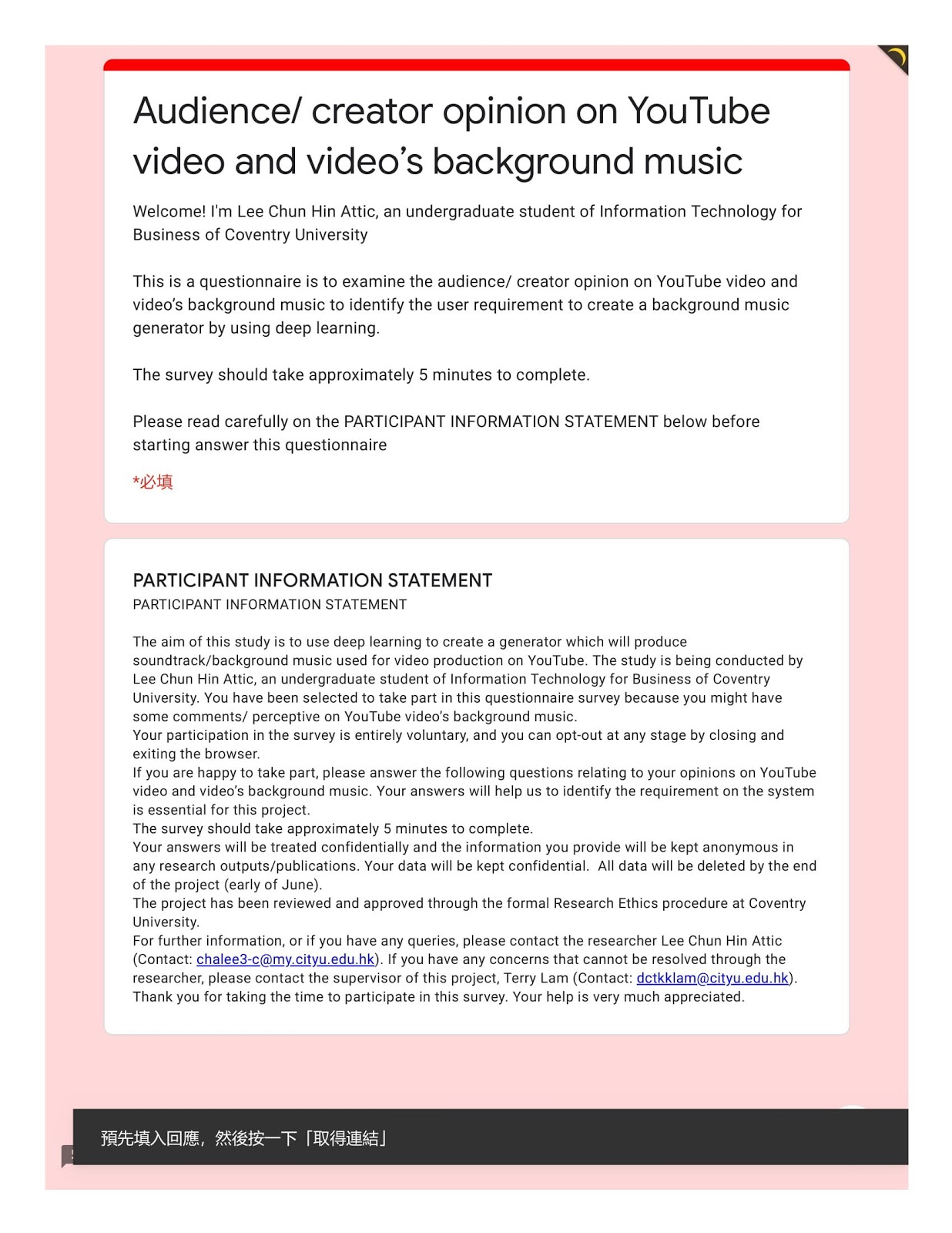
We Are Social and Hootsuite (2019d). *World’s Most-Visited Websites (According to Alexa) January 2019*. *DataReportal*.

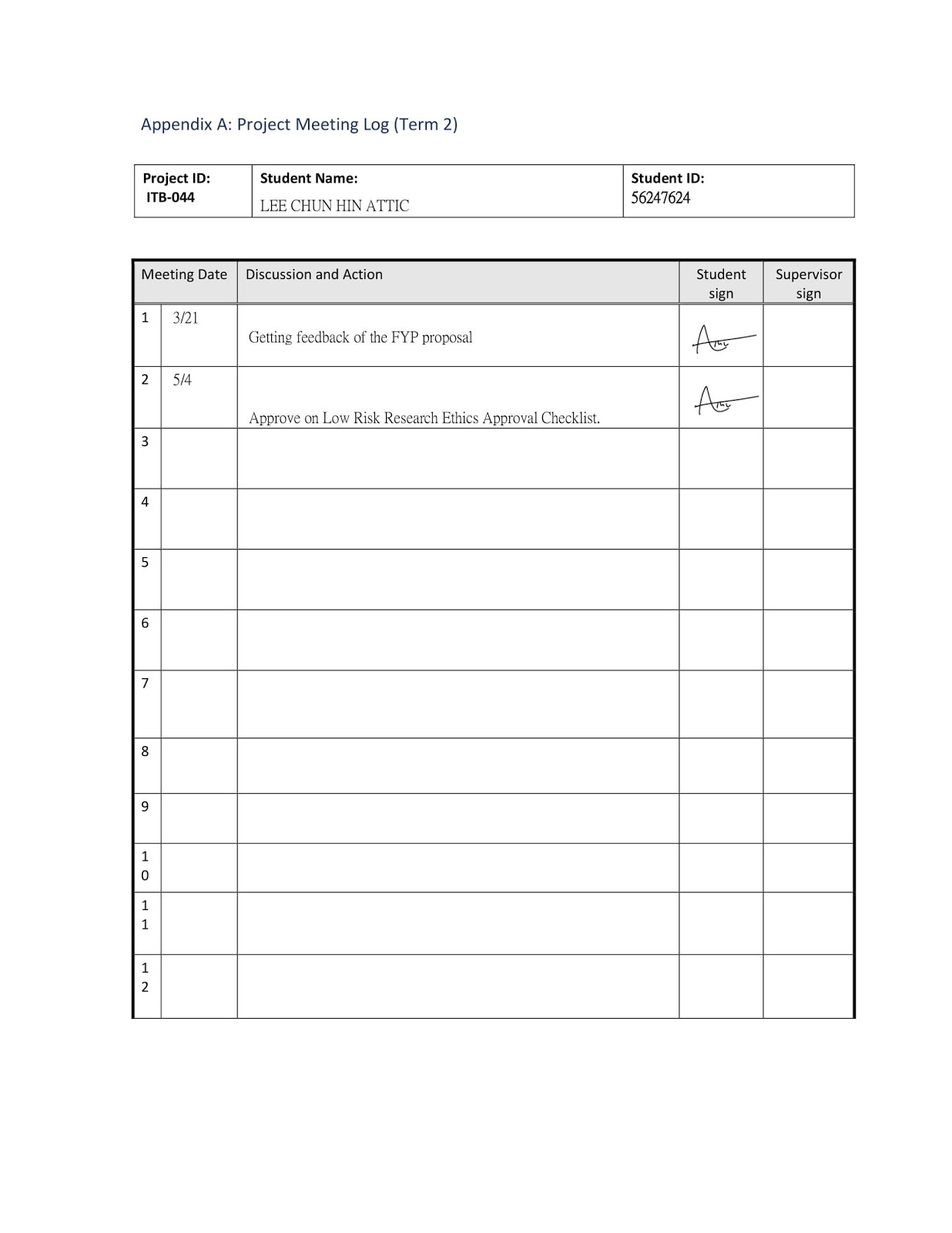
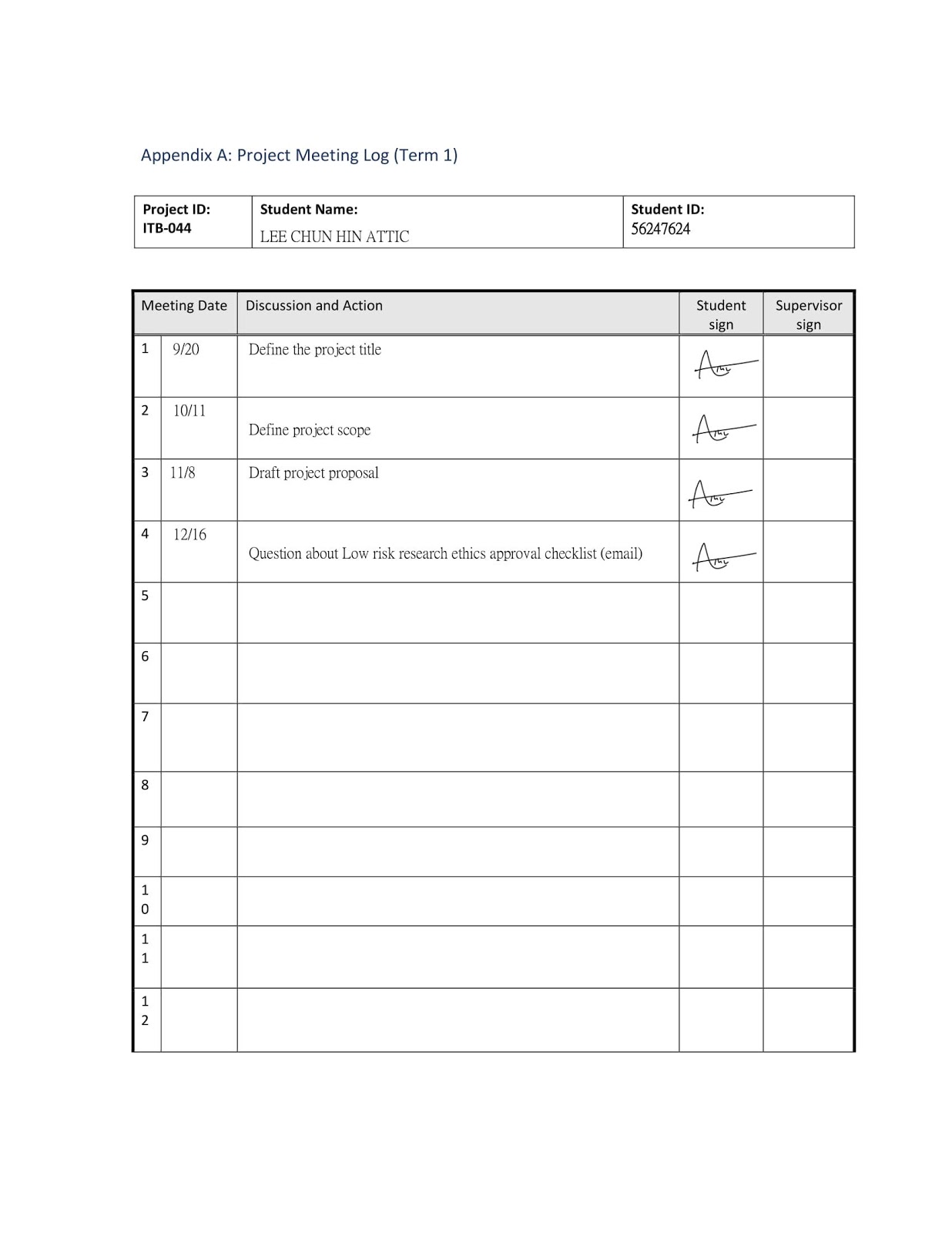
Worldbank.org. (2019). *Population, total | Data*. [online] Available at: https://data.worldbank.org/indicator/SP.POP.TOTL?end=2018&start=2018&view=bar [Accessed 20 Nov. 2019].

Zaremba, W., Sutskever, I., Vinyals, O. and Brain, G. (2015). *Recurrent Neural Network Regularization*. [online] Available at: https://arxiv.org/pdf/1409.2329.pdf [Accessed 28 Mar. 2020].

Zaripov, R.K. (1960). An Algorithmic Description of a Process of Musical Composition. *Dokl. Akad. Nauk SSSR*, 132(6).

‌

**Appendix A- User requirement Questionnaire**

**Appendix B -Project Meting Log**