**Chapter 2: Research Design**

Primary Data

**Sampling Strategy**

In the study conducted on the utilization of attention-based transformer models for polyphonic music generation, the primary research population consisted of five industry experts. An industry expert was defined as someone who was working in the music generation industry. Industry experts were sampled to obtain valuable insights from their in-depth knowledge of music creation, enriching the research with nuanced perspectives. A credible and reliable point of view was offered, and a unique perspective on the generated music was provided, which might not have been noticed otherwise.

The experts were gathered from LinkedIn and were not connected on LinkedIn prior to the research to avoid any response bias. LinkedIn was used as the sample frame of choice because it provided a large, rich pool of experts due to its global popularity and widespread adoption among professionals from various industries and geographical locations. It was easily accessible, and due to its user-friendly interface, it was virtually effortless to select a sample.

Due to time constraints, the research adopted a non-probability sampling technique, specifically expert sampling. While the probability sampling method could have provided a representative sample for statistical generalization, non-probability sampling was chosen for its efficiency in gathering feedback within a short timeframe. Judgement sampling was selected for exploratory research, allowing for the hand-picking of individuals with relevant experience and proficiency to provide valuable insight on the generating music. Experts were chosen to ensure diversity in expertise, perspectives, and backgrounds, encompassing various job titles across different companies. The research remained relevant to the overall research objectives, as the experts’ offered opinions on the coherence of the music, the impact of the dataset on its quality, and its potential for monotony. Their finely-tuned ears detected subtle nuances and intricacies in the music, providing insightful and knowledgeable input while optimizing time and resource utilization.

**Primary Research Methodology**

The method employed in the primary research involved conducting five expert interviews. Experts were contacted through LinkedIn and were invited to participate in the study. The topic of interest, research objectives, and the music generated from the analysis were shared with the interviewees prior to the interviews. Each interview lasted approximately forty-five minutes, with an additional fifteen minutes allocated for summarizing initial thoughts after the interview. Consent for recording the interviews was obtained through email or LinkedIn messages before the interviews occurred.

The expert interviews were conducted using a qualitative approach. Depending on logistical concerns, the interviews were organized for a mutually convenient time and were performed in person in a quiet public setting, such as a cafe, over the phone, or via Zoom. The experts had the opportunity to share their knowledge and opinions during the interviews on both the music generated and its alignment with the research objectives and music generation in general, providing a better understanding of the topic. To ensure the relevance of the information collected from the interviews, open-ended questions were designed to elicit detailed and insightful responses from the experts, facilitating exploration and clarification of their perspectives.

The audio recordings of the interviews were transcribed verbatim, and the transcripts served as the primary source of data for analysis. To identify patterns, themes, and significant findings in the interview data, thematic analysis, content analysis, or other qualitative analytic techniques were employed. The reliability and validity of the results were ensured through the application of rigorous analysis procedures. Once the analysis was finished, it was compared to the findings from secondary research to determine whether our results aligned with what the experts had found and, if not, to understand the reasons for any disparities.

The preference for the chosen method of primary research stemmed from the need for more than mere subjective judgments of the generated music's quality. In order to make informed decisions and derive meaningful conclusions from the research findings, a valid and relevant opinion from a credible evaluator was essential. This need was particularly pronounced in the pursuit of the first objective, aimed at assessing the self-attention transformer's ability to capture long-term dependencies in polyphonic music.

The expert evaluated whether the newly generated music constituted a coherent composition that seamlessly integrated with the trained music, distinguishing between a well-structured piece and a mere sequence of random notes. An expert's perspective was crucial in identifying the strengths and weaknesses of the music, especially regarding the second objective, which focused on potential quality changes resulting from the use of a new dataset with the model. The expert's keen sense for what constitutes high-quality music enabled them to discern any significant alterations, considering metrics such as pitch, rhythm, melody, harmony, originality, and emotional impact.

Furthermore, the use of expert interviews offered flexibility in shaping the questions posed. If the interviewee raised intriguing points not previously considered, further exploration was possible. Likewise, when encountering unclear aspects, the interviewee could provide immediate elaboration. This advantage of the research method proved invaluable when discussing the third objective, which delved into the fine line between music repetition for structural, emphatic, and rhythmic purposes without becoming monotonous and uninteresting to the listener. The expert's insights allowed for a deeper exploration and assessment of whether this balance had been achieved.

When employing this research method, it is essential to acknowledge certain limitations. Firstly, the subjective nature of expert opinions introduced the possibility of biases, as feedback was influenced by their individual experiences, knowledge, and personal inclinations. Additionally, the findings might have lacked generalizability to the broader population, as a limited sample of five experts was interviewed rather than a larger or more representative group. Nevertheless, despite these limitations, the benefits offered by this method outweighed the drawbacks, leading to its implementation in the research.

Problem Identification and Clarification

This thesis aims to explore the capacity of self-attention mechanisms within transformer models to effectively capture long-term dependencies when generating polyphonic music. These mechanisms enable the modeling of intricate relationships between various musical elements across extended time intervals, resulting in compositions that exhibit greater complexity and coherence. Additionally, this research will examine the interplay between the size and diversity of training data and the quality of the generated musical output. Throughout the project, various techniques will be employed to mitigate or eliminate the production of monotonous or mundane musical pieces.

The field of music generation has witnessed remarkable advancements in recent years, largely driven by the application of deep learning techniques. One such approach is the attention-based Transformer model, originally introduced in the "Attention Is All You Need" paper (Vaswani *et al.*, 2017). This model has shown great promise in natural language processing tasks and has been successfully adapted for music generation.

The Transformer model, originally designed for natural language processing, has demonstrated its ability to capture long-range dependencies and understand the hierarchical structure of sequences. This makes it a promising candidate for modeling the complex relationships between multiple notes in polyphonic music. By leveraging the self-attention mechanism, the Transformer can attend to relevant parts of the input sequence and generate coherent and harmonically-rich music compositions.

The application of attention-based Transformer models to polyphonic music generation opens up exciting opportunities for creating original and expressive compositions. Researchers and practitioners in the field are actively exploring novel approaches and techniques to further enhance the capabilities of these models. Studying and advancing the generation of polyphonic music with attention-based Transformer models unlocks new avenues for creative musical expression and contributes to the development of intelligent systems capable of composing complex and engaging music.

Research Objectives

* Problem Identification: Transformer models have shown great promise in generating polyphonic music, but it is unclear how well they can capture long-term dependencies in the music.

Problem Clarification: The ability to capture long-term dependencies is important in generating music that has a coherent structure and is musically pleasing.

Problem Formulation: How effective are self-attention mechanisms in transformer models for capturing long-term dependencies in polyphonic music?

Objective: To evaluate the effectiveness of self-attention mechanisms in transformer models for capturing long-term dependencies in polyphonic music.

* This objective aims to investigate the suitability of transformer models for music generation by evaluating their ability to capture long-term dependencies in polyphonic music. The objective will involve examining the effectiveness of self-attention mechanisms in identifying and encoding relationships between musical elements over longer time periods.
* Problem Identification: The impact of training data on the performance of attention-based transformer models for polyphonic music generation is not well understood.

Problem Clarification: It is unclear how the size and diversity of training data affect the quality of generated music by attention-based transformer models.

Problem Formulation: The objective is to evaluate the impact of training data on the performance of attention-based transformer models for polyphonic music generation, and how the size and diversity of training data affects the quality of generated music.

Objective: To determine the relationship between training data size and diversity, and the quality of generated music by attention-based transformer models for polyphonic music generation.

* The objective of this study is to investigate how the quality of generated music by attention-based transformer models for polyphonic music generation is affected by the size and diversity of the training data used to train these models. In other words, the study aims to determine whether the quantity and variety of the training data have an impact on the quality of the generated music. By evaluating this relationship, the study can provide insights into how to optimize the training data selection process to improve the performance of attention-based transformer models for polyphonic music generation.
* Problem Identification: Computer-generated music often suffers from repetitive patterns, which can make the music uninteresting and predictable.

Problem Clarification: Generating diverse and original music is important in creating music that is musically pleasing and engaging.

Problem Formulation: How well can attention-based transformer models generate diverse and original polyphonic music, and how effective are they at avoiding repetitive patterns and generating novel musical ideas?

Objective: To assess the ability of attention-based transformer models to generate diverse and original polyphonic music by examining their ability to avoid repetitive patterns and generate novel musical ideas.

* The objective of this research is to evaluate the diversity and originality of polyphonic music generated by attention-based transformer models. Specifically, the research aims to examine the model's ability to avoid repetitive patterns and generate novel musical ideas. By assessing the model's ability to generate diverse and original music, this research can contribute to the development of more advanced and creative machine learning models for music generation.

Validity Type

The two components of validity management are relevant and reliable.

Relevant: The thesis aims to evaluate and justify the use of attention-based transformer models for polyphonic music generation. Previously, neural networks were used with limited success due to the exploding gradient problem. However, since the introduction of Large Language Models, it is deemed relevant to view music generation as a language modeling task and to apply attention-based transformers to this domain. The study will assess the relevance of utilizing these models in the context of music generation and evaluate their performance.

Reliable: The thesis will assess the effectiveness of attention-based transformer models for capturing long-term dependencies in music. The reliability of the findings will depend on the robustness of the evaluation methodology and the validity of the results.

Ethical and Legal Considerations

Primary Research

Ethical considerations were central to the Data Analysis project. Voluntary participation and informed consent from all five participating experts were required for the primary research. This was achieved by providing a clear explanation of the study's purpose, the nature of their involvement, and the utilization of their data via email/Linked In message. The option to seek clarification or pose questions beforehand through the same email chain was encouraged, with full respect for their decision to participate or withdraw from the study at any stage.

Trust and respect were of the utmost importance for all participants in the study. This meant showing up to the meeting early and being prepared to start with plenty of questions to fill the time slot agreed upon. Maintaining professionalism throughout the interview and ensuring their opinions and expertise were valued.

Ensuring the accuracy of result reporting is paramount. This process entails transcribing the recordings verbatim and extracting vital themes, quotes, and interesting insights while avoiding the inclusion of redundant or fabricated information.

When composing the thesis results, confidentiality and anonymity of the experts' responses will be maintained by assigning each of them an alias (e.g., interviewee A). All other personal information, including names, email addresses, Linked In profiles, phone numbers, etc., will be securely stored in private Linked In and/or email accounts with password protection and in a password-protected folder on a personal laptop. The findings will be presented in a manner that honours the experts' contributions and upholds the research's integrity. Themes and insights derived from all five interviews will be summarized, enhancing readability, and preserving the essence of each interviewee's input.

In conclusion, ethical principles take precedence in the thesis project. Voluntary participation and informed consent are essential, ensuring that all five experts involved in the interviews comprehend the study's purpose, their roles, and the utilization of their data. Trust and respect play a central role, emphasized through punctuality, preparedness, and genuine appreciation for the experts' opinions and expertise. Reporting and dissemination follow responsible practices, accurately conveying the results of data analysis through verbatim transcriptions and the extraction of key themes, quotes, and insights. To protect confidentiality and anonymity, experts receive aliases, and personal information is securely stored. Ultimately, the findings are presented in a format that honours the experts' contributions, delivering summarized themes and insights that preserve the research's integrity and engage readers effectively. These ethical considerations ensure a robust and respectful execution of the Primary Research in the thesis project.

Secondary Research

In the ever-evolving landscape of AI-generated music, a fundamental consideration lay in harmonious coexistence with established copyrights and the avoidance of infringement. Irish, UK, and USA copyright laws stated that 'copyright in a literary, dramatic, musical, or artistic work expired 70 years after the death of the author. After this period, the work was said to enter the public domain, allowing it to be used, modified, or republished by any person without fear of copyright infringement' (About Copyright, no date) (How copyright protects your work, no date) (Editor, 2022). This emancipated the creation, allowing it to be freely utilized, adapted, and republished without the looming specter of copyright infringement. As the exploration of AI-driven musical composition continued, it became pivotal to not only respect these legal frameworks but also to leverage open-source data discussed in the Data Collection section. This data aligned with copyright regulations and duly acknowledged the sources, thereby safeguarding intellectual property and artistic integrity.

When venturing into the realm of AI-generated music, it was crucial to remain vigilant about potential biases stemming from the training data and algorithms, which might result in the production of prejudiced or objectionable musical compositions. The landscape of music genres and their associated music theory was far from straightforward. Their development was influenced by a complex web of factors, often intersecting, and shaping the transformation of musical styles and practices. Music was profoundly shaped by elements such as geography, culture, religion, historical events, technological advancements, demographics, as well as the blending and fusion of various influences, to name just a few.

Critical aspects to be mindful of in the model's output were the potential presence of stereotypes and cultural biases in the music it generated. Stereotypes could emerge when the AI model, intentionally or unintentionally, replicated simplistic or biased notions about certain musical styles, genres, or cultures. For example, the model might falsely associate specific musical elements with cultural clichés, resulting in misrepresentations. These stereotypes could perpetuate cultural insensitivity and lead to feelings of hurt or disrespect among listeners.

Furthermore, cultural biases might also influence the AI-generated music, as they could manifest in the model's interpretation of various musical traditions and practices. Biased training data or human annotations might inadvertently introduce cultural biases into the system, leading to music that failed to authentically represent the rich diversity of musical heritage.

To mitigate these issues, it was imperative to choose the training dataset with meticulous care and a deep understanding of this nuanced reality. This involved not only diversifying the training data but also rigorously scrutinizing it for any pre-existing stereotypes and cultural biases. By doing so, it ensured that AI-generated music respected the rich tapestry of human culture and history while fostering creativity and innovation.

The most beautifully composed music possessed a unique quality that transcended mere ear-pleasing delight, having the power to stir deep emotions. These emotions could be both positive and negative. With this in mind, the responsibility that came with creating and sharing music was recognized. This responsibility extended to the well-being of the audience, and steps were taken to ensure that creations did not inadvertently cause discomfort or harm. By seeking feedback from a diverse group of listeners and making necessary adjustments, the endeavour was to craft music that touched hearts and minds without causing unintended distress. In this intricate dance between AI and art, music was created to resonate with people while respecting legal and cultural boundaries, ultimately offering a harmonious and inclusive musical journey.

Within the ever-shifting domain of AI-generated music, harmonious coexistence with copyright laws, vigilant avoidance of biases, and the creation of emotionally resonant compositions represented the pillars upon which the future of this art form stood. Navigating the intricate landscape of musical creation, the commitment to legal and ethical principles was steadfast, while also embracing the boundless possibilities that AI offered for creative expression. Upholding the values of respect, diversity, and emotional connection ensured that AI-generated music not only honoured the past but also paved the way for a harmonious and inclusive musical future. In this intricate fusion of technology and art, a symphony of potential was found, where AI and human creativity joined hands to compose a harmonious melody for generations to come.

**Chapter 3: Literature Review**

Dataset

In the realm of music generation, the availability and quality of training datasets play a pivotal role in shaping the creative potential and output of generative models. This literature review explores the landscape of music datasets, for example symbolic datasets, audio datasets, lyrics and text datasets, music notation datasets, genre-specific datasets, emotional or mood-based datasets to name a few. As researchers and artists alike seek to push the boundaries of what is possible in automated music creation, the choice and characteristics of the underlying data become of paramount importance.

Music

Repetitive Pattern

Chords

Sampling

**References**

Vaswani, A. et al. (2017) ‘Attention Is All You Need’. arXiv. Available at: http://arxiv.org/abs/1706.03762 (Accessed: 1 April 2023).