Classifying AI-Generated Music with AI Models

Siddharth Misra

I. Abstract

The emergence of artificial intelligence (AI) in music composition has surfaced as a contentious yet captivating topic of exploration within the music industry [1]. It raises questions about the legitimacy of machine-generated creativity and the complexities surrounding copyright, ownership, and the evolving nature of artistic expression [2]. One prominent and groundbreaking creation that has captured the attention of many is Jukebox, the AI music generation model created by OpenAI [3]. The importance of this research is twofold: addressing copyright issues and fostering the adoption of AI in the music industry. As AI-generated music blurs the line between human and AI creation, concerns regarding copyright ownership and artistic attribution become paramount [2]. To address these challenges, we created an AI model to differentiate AI-generated music from human-composed music, specifically focusing on music generated by the JukeBox model. Using 3 different models: a Multi-Layer-Perceptron (MLP), a long-short term memory network (LSTM), and a transformer, we were able to classify JukeBox generated songs with 96 percent accuracy.

II. Methods

To collect our data, we used the SoundCloud API to download JukeBox songs which they publicly post. We were able to collect 106 JukeBox songs and collected 97 human composed songs (which were specifically selected to roughly have the same musicians that JukeBox based their songs off).

III. Experimental Design

We processed these raw mp3 files using Mel-frequency cepstral coefficients (MFCCs). MFCCs are a way of representing the characteristics of sound in a form that is useful for tasks like music analysis, by emphasizing the aspects of sounds that humans are most sensitive to [4]. We then implemented our three models (MLP, LSTM, Transformer) using python's TensorFlow library. An MLP, or multi-layer perceptron, is an artificial neural network consisting of layers with interconnected nodes, including an input layer, a hidden layer with weighted connections and activation functions, and an output layer. It is trained through backpropagation and adjusting weights to minimize error [5]. An LSTM is a type of recurrent neural network designed for sequence data processing, featuring memory cells and gates to capture and control information flow [6]. A Transformer is a deep learning architecture known for its self-attention mechanism, enabling simultaneous processing of entire input sequences. Composed of encoders and decoders, Transformers can learn context and meaning by tracking relationships in sequential data (Note: we only used the Transformer encoder in our testing) [7]. In our experimental setup for the LSTM and Transformer, each song was truncated to a length of 30 seconds and separated into a training and validation split of 85 percent to 15 percent. We trained each of our models with an Adam optimizer and a binary cross entropy loss metric. To determine the performance of the models, we observed and recorded the accuracy on the validation data set, a portion of the dataset unseen to the model during training.

TABLE I.

Model Performance Summary	
Model	***Accuracy***
Multi-Layer-Perceptron	91%
LSTM	93%
Transformer	96%

IV. Results

As shown in the table, we can see that the Transformer excelled. This is due to its ability to leverage temporal information and its use of the attention mechanism. This mechanism allows the model to focus on specific parts of the input sequence which is highly advantageous in music classifications tasks.

V. Conclusion

In summary, the landscape of AI-generated music is continually evolving and presenting new opportunities. While AI-generated music often leaves discernable patterns that allow them to be recognized, this technology will increasingly become more advanced, pushing the boundaries of music composition. This evolving dynamic underscores the need for research and adaptability in the subfield of AI-generated music. While navigating this new landscape, it is imperative that we foster innovation and collaboration to classify and attribute authorship in an environment where lines between human and machine creativity will be increasingly blurred.

The code has been made publicly available on GitHub at: https://github.com/0sid0/AI-Generated-Music-Detection.git

References

[1] Brown, T. T. H. Artificial Intelligence in the Music Industry.

[2] Bulayenko, O., Quintais, J. P., Gervais, D. J., & Poort, J. (2022). AI Music Outputs: Challenges to the Copyright Legal Framework. *Available at SSRN 4072806*.

[3] Dhariwal, P., Jun, H., Payne, C., Kim, J. W., Radford, A., & Sutskever, I. Jukebox: A Generative Model for Music.

[4] Kopparapu, S. K., & Laxminarayana, M. (2010, May). Choice of Mel filter bank in computing MFCC of a resampled speech. In *10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010)* (pp. 121-124). IEEE.

[5] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65*(6), 386–408. https://doi.org/10.1037/h0042519

[6] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

[7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.