# **Using the Hit Songs of the Last Seventy Years to Generate Song Lyrics**

By Lukas Adkins (adkinsl4@wwu.edu) and Alayne Hatcher (hatchea3@wwu.edu)

**Abstract:**

This project applies an LSTM (Long Short-Term Memory) recurrent neural network language model to the task of generating song lyrics based on time period (decade). Our model’s goal is to generate song lyrics that are identifiably remicient of the trends of a particular decade of popular music. Previous work has mostly focused on training models based on particular artists or genres, while our model seeks to utilize the top songs from particular periods of time.

**Introduction:**

The problem addressed by our project is that of capturing the essence of different eras of popular lyrics through imitation of their lyrics. In a sense, our project addresses both machine-learning powered authorship of new song lyrics (ghostwriting) and analysis of these songs to determine if any recognizable trends are reflected in models trained on different time periods.   
 The ethical and societal implications of AI generated song lyrics, and more generally, whole pieces of music, are unclear. If models could be trained and tuned sufficiently, it is of theoretical interest to consider the implications of near-perfect imitations of popular songs. Would these songs’s lyrics still be considered art, if generated by a recurrent neural network? If average listeners could not tell the difference, would these models have commercial viability? While these questions are theoretically interesting, more practically our particular project likely has use to those seeking to understand more about trends in popular music since the 1960s, and how that is reflected in neural networks trained on large datasets based off of each decade.

Our overall approach to the problem was split into two parts: first identifying a valid data source for the project, and second training and utilizing our model for analysis. More specifically, our first step was to gather name and artist data from the top five hit songs from each week of the Billboard Hot 100 music chart[1] and then gather the lyrics for these songs using the Lyrics Genius API[2], and combine this data to use as our corpus. Originally we had hoped to utilize song lyric frequency data from The Grammar Lab[3] and The Million Song Dataset[4] but we were unable to efficiently integrate these datasets into our model in time. The second step of our approach involved initializing our LSTM (Long Short-Term Memory) model, training one model for each decade of popular music (since 1960), and then using each model generate and select probable words for song lyrics in each time period given a generic starting word.

**Related Work:**

Related work to our project includes “GhostWriter: Using an LSTM for Automatic Rap Lyric Generation” (Potash et al., 2015)[5], a study that has already shown the effectiveness of LSTM models applied to rap lyric generation. This work proved promising as it specifically succeeded in not only generating song lyrics, but mimicking the writing style of particular rappers; this task is not far removed from mimicking the style of a particular time period, like our model seeks to accomplish.

More recent work (Santhanam et al., 2020)[6] has more generally discussed context based text generation using LSTM networks, as well as discussing various approaches and their effectiveness. Their findings support our choice of applying an LSTM neural network for this particular project task.

**Approach:**

We chose to use the Keras LSTM API in order to implement the neural network. This API gave us tools to format our data, train the model, and generate the lyrics. Formatting our data, involved using both Spacy[7] and Keras[8] tokenizers. Spacy’s tokenizer was used to clean up the lyrics. Punctuation was removed, and the lyrics were all turned into their lowercase counterparts. From here we were able to use the Keras tokenizer to turn each distinct word into its own unique integer. These integers were then split into multiple sequences of token series, each with a count of 50 words. A label representing the word that followed was attached to each sequence.

The model was created and trained using Keras’s layering system. We first set up an embedding layer that will, as the name suggests, embed our words. This will help our model better understand the relationship between words within our data set. We then set up an LSTM layer which is the actual neural network itself. In a later call to fit the model to our function, we set this layer to repeat itself 5 times. This number was chosen as a compromise between creating a more accurate model and shortening the runtime. The last layer we used was a dense layer. This layer takes the output from the LSTM results and formats it to be used in the predictions later.

Once the model was trained, we were able to use it to generate lyrics. A seed was used to start off the generation. This could either be a single word or a sentence. This seed was turned into a sequence of integer represented tokens which was given to our model. The model looked at this sequence and predicted a word that could possibly follow it. This word was added to the end of the sequence, and the prediction was repeated until we had our desired word count.

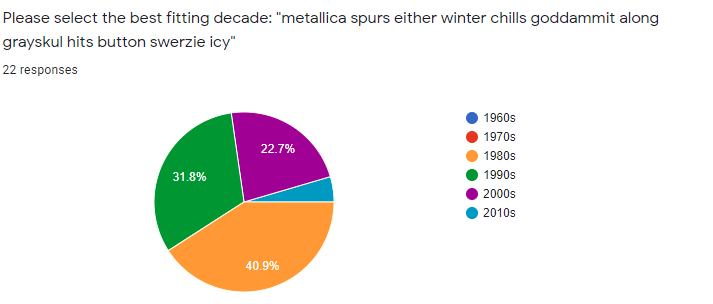
**Experiments, Results:**

As previously mentioned, our dataset consisted of top songs from each week of each decade since 1960, with pre-processing consisting of stopword removal, punctuation removal, and lowercasing of words. After training a LSTM model on each decade, we then generated 10 “songs” worth of lyrics from each decade (10 batches of 50 words generated by the model). Our main evaluation method used to determine the success of the model was a human-centered approach, in which we asked survey participants of their time period association with the generated lyrics.



**Figure 1:** Excerpt of generated lyrics for a song using model trained on 1980s hits

As shown in figure 1, some lyrics generated have both time period-relevant language (“metallica”) and some degree of thematic consistency (“winter, chills, icy”), and to a limited degree is reminiscent of the time period, as shown by survey data (n = 22).



**Figure 2**: Survey response for sample model output from 1980s

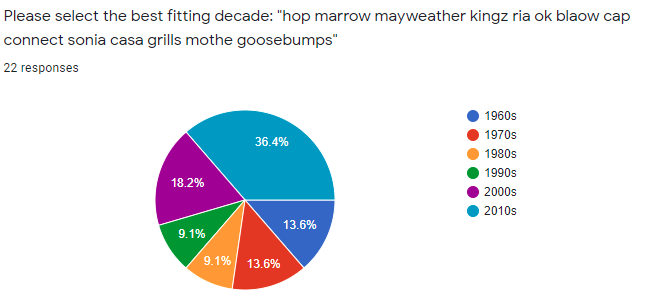
As seen in figure 2, many survey participants identified the correct period of time, likely from an inference of the period-relevant language contained in the generated lyrics. All participants chose time periods in which Metallica has been an active band producing music, with participant responses seemingly weighted towards when the band produced their most popular albums.

Our results also pointed towards some generated lyrics being recognizable solely based on word choice, without reference to specific cultural icons or groups. For example, these generated lyrics in figure 3 contain modern language and slang, and were generated from our 2010s model.



**Figure 3**: Excerpt of generated lyrics for a song using model trained on 2010s hits

Usage of language typically associated with modern hip hop (“hop, kingz, grills”) makes this output reminiscent of more modern popular music, as reflected by survey results seen in Figure 4. Well over half the survey participants identified the lyrics as being from the 2000s, with 36.4% correctly identifying the lyrics association with the 2010s.



**Figure 4**: Survey response for sample model output from 2010s

**Results Discussion:**

While many of the lyrics generated by our model achieved our goal of being reminiscent of a particular time period (had >35% accuracy as identified in survey), as shown in Figures 1-4, many outputs also did not share that same level of association with the time period they were generated from. Due to our choice to remove stopwords during the preprocessing process, and limitations of our model, generated sentences were consistently borderline or completely nonsensical from a thematic perspective. Our current outputs also lack structural consistency and song form, removing understandability from the model’s outputs to some extent. Overall, while it is clear our model has potential to generate lyrics consistently reminiscent of a particular time period, it is also clear that realizing this objective fully and consistently would require additional future work.

**Conclusion and Future Work**

In conclusion, our model partially achieved its objective of consistently generating lyrics from particular decades, but would need additional work to fully actualize this goal. A number of improvements could be made to our model to improve its accuracy and general effectiveness. One fundamental change that could be vital in increasing model accuracy would be stricter data processing during the creation of our corpus. We became aware of some bugs with the Lyrics Genius API midway into our usage of data scraped from it, leaving us with some usual results returned into our dataset instead of our desired lyrics, essentially giving us junk data that likely lowered our model’s accuracy. Removing duplicate songs could also give the dataset more variety. Additionally, fully integrating a bigram/trigram baseline model and using methods of comparison between it and our model’s outcome could give us a better idea of progress made. Integration of a word frequency dataset and reintroduction of stopwords to our model could likely produce more “readable” results, likely giving us better feedback and survey responses. Another potential key factor that would likely allow for additional accuracy would be training our model on more data, on a more powerful computer. Our dataset could likely accommodate usage of many more songs per week then we used, but Google Colab as an environment quickly ran out of allocated RAM when we attempted to increase the amount of data being processed.

**References:**

[1] Dave Dhruvil. 2021. Billboard “The Hot 100” Songs. https://www.kaggle.com/dhruvildave/billboard-the-hot-100-songs (CC License.)

[2] John Miller. LyricsGenius API. https://github.com/johnwmillr/LyricsGenius.

[3] Brown, D. W. (2017) Song Lyrics Data Tables. Retrieved from http://www.thegrammarlab.com

[4] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. (2011) The Million Song Dataset. In Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)

[5] Peter Potash, Alexey Romanov, Anna Rumshisky. 2015. GhostWriter: Using an LSTM for Automatic Rap Lyric Generation. In *2015 Empirical Methods in Natural Language Processing* (EMNLP 2015).

[6] Sivasurya Santhanam. 2020. Context Based Text-Generation Using LSTM Networks. arXiv:2005.00048v1.

[7] https://pypi.org/project/spacy/

[8] https://keras.io/