

# Project Breadcrumbs: Fairy Tale Text Generation Using GPT-2

Andrew Arteaga, Maddie Louis, and Merissa Tan  
December 15th, 2020

## Abstract

The goal of this paper is to determine whether machines can generate plausible fairy tale texts that have a style similar to famous human authors. Previous research has primarily relied on either using fairy tale data specific to one author or a multitude of generic fairy tale data and thus has been unable to truly determine if the model can be creative and generate new and plausible fairy tales compared to the fairy tales published by human renowned authors. Therefore, the group's model uses both fairy tales written by specific authors and a plethora of generic fairy tales to determine which training methods can give the model the creative boost it needs to generate unique and plausible fairy tale text. As expected, training the model on both general and author-specific fairy tales, rather than only on either author-specific tales or general tales, minimizes the perplexity of the model and allows the model to generate more plausible fairy tale texts. However, the group's findings also indicate that the model's perplexity and ability to generate plausible fairy tale texts can vary depending on whether or not all the fairy tale data share a common theme and sentiment.

## 1 Introduction

The primary goal for Project Breadcrumbs was to create a text generation model that will learn from existing fairy tales and auto-generate new and plausible fairy tale texts. The hypothesis for this project was that a model that is fine-tuned on fairy tale data would have a lower perplexity when tested on fairy tale texts than a model that is not fine-tuned on such data. The models were analyzed both quantitatively and qualitatively in order to improve and assess them.

On the quantitative aspect, the perplexities of

the models when tested on fairy tale texts were compared and minimized. On the qualitative aspect, the project aimed for the generated fairy tale texts to have a plausible plot line, in addition to following a similar syntax or style of writing seen in many fairy tale texts. Author-specific models were also implemented, and the similarities and differences of these model-generated texts were compared. Word clouds were then used to showcase the most commonly used words in said generated texts.

## 2 Implementation

The implementation of this project relied heavily on the use of OpenAI's GPT-2 model. The fairy tale datasets that were used in this project were taken from the Project Gutenberg website.

### 2.1 Data

*The Blue Fairy Book* by Andrew Lang was used for the fine-tuned baseline model for this project.

Fairy tales written by Charles Perrault, Brothers Grimm and Hans Christian Andersen were compiled into separate author-specific datasets and incorporated into the final models. The final models that were fine-tuned twice also used a large, generic fairy tale dataset consisting of different fairy tale sources, including *The Blue Fairy Book* and *Japanese Fairy Tales* by Yei Theodora Ozaki.

Due to a lack of indexing in these datasets, the data had to be manually analyzed so that the generic fairy tale dataset does not contain any tales from the specific author the model is trained and tested on to prevent overfitting.

### 2.2 Models

The project took advantage of the fact that

OpenAI’s GPT-2 language model is already pre-trained. The language models created throughout this project, therefore, were fine-tuned on different fairy tale datasets. After the language models were trained<sup>1</sup> on the fine-tune data, the models were evaluated and their perplexities were calculated and analyzed. Generation models were then created using these language models to generate sample fairy tale texts.

### 3 Baseline Model

The perplexity of the GPT-2 baseline model was first calculated without performing any fine-tuning. Then, the model was fine-tuned on *The Blue Fairy Book* and evaluated again to get its new perplexity.

#### 3.1 Quantitative Results

Baseline Model	Perplexity
Default	40.8936
Fine-tuned	29.2640

Table 1: Baseline perplexity results.

As expected, the default baseline perplexity was significantly lower than the fine-tuned baseline perplexity.

#### 3.2 Qualitative Results

Text generated by the fine-tuned baseline model, which was fine-tuned on data from *The Blue Fairy Book*, contained obvious fairy tale motifs and themes, such as beauty, marriage, and royalty. However, the generated tales did not have a clear and coherent plotline.

## 4 Final Models

### 4.1 Author-Specific Fine-Tuned Model

The perplexity of the GPT-2 author-specific fine-tuned model was calculated after

---

<sup>1</sup> All models were trained with 15 epochs.

fine-tuning the model on fairy tales written by a specific author. The model was first trained and tested on the fairy tales written by Charles Perrault, then on fairy tales written by the Brothers Grimm, and lastly on fairy tales written by Hans Christian Andersen.

#### 4.1.1 Quantitative Results

Author	Perplexity
Brothers Grimm	30.1845
Andersen	46.0975
Perrault	40.4632

Table 2: Single fine-tuned perplexity results.

The perplexities from the single fine-tuned model were noticeably higher than that of the fine-tuned baseline model.

The only difference between the fine-tuned baseline model and this single author-specific fine-tuned model is the datasets that were split and used for training and testing. Therefore, we can conclude that the GPT-2 language model was more perplexed by the author-specific text files than by *The Blue Fairy Book* data. One possible explanation for these results is that *The Blue Fairy Book* was assembled and edited by one author, and may have followed more consistent patterns than the original fairy tale authors (Perrault, Grimm, Andersen) did in their works.

#### 4.1.2 Qualitative Results

The fairy tales generated by the single fine-tuned model were more coherent than those generated by the baseline model. In addition, the generated text also exhibited fairy tale motifs and themes.

### 4.2 General then Author-Specific Fine-Tuned Model

The perplexity of the GPT-2 general then author-specific fine-tuned model was calculated after fine-tuning the model on a general data set of fairy tales consisting of texts from *The Blue Fairy Book*, *Japanese Fairy Tales* and the

author-specific datasets of authors who were not used for that specific model. First, the model was trained and tested on the general fairy tale dataset. Then, it was trained and tested on fairy tales written by the specific author that was chosen for that model.

#### 4.2.1 Quantitative Results

Author	Generic Perplexity	Author-specific Perplexity
Brothers Grimm	46.5718	28.1621
Andersen	36.4580	45.5039
Perrault	63.8808	20.5576

Table 3: Twice fine-tuned perplexity results.

The twice fine-tuned model was evaluated on both the generic fairy tale dataset as well as the author-specific dataset of the chosen author.

In the case of the Brothers Grimm and Perrault, the perplexity was lower when the model was evaluated on author-specific data. Additionally, these perplexities were lower than the single fine-tuned model, indicating that fine-tuning on both generic fairy tale data and author-specific data improved overall results.

However, we found that the Andersen model was a clear outlier. The Andersen model’s perplexity was lower when assessed on a generic dataset as opposed to the author-specific dataset, and both perplexities were higher than the perplexity of the fine-tuned baseline model. For more on this outlier, refer to “Limitations” in Section 8.

#### 4.2.2 Qualitative Results

The twice fine-tuned model generated the most coherent and impressive fairy tales. Our “Qualitative Results” appendix included at the end of this report includes sample fairy tales from all three authors as well as word clouds that highlight the most common words in the generated texts.

While many of the tales have a confusing storyline (see the first example in the Andersen

section), some have a coherent plot from beginning to end. The Grimm and Perrault tales have higher feasibilities than the Andersen tales, which align with the perplexity results seen in Table 3.

### 5 Overall Quantitative Analysis

The quantitative perplexity results followed the expectations of this project’s hypothesis, with the exception of the results from the twice fine-tuned Andersen-specific model (see more on that in the “Limitations” section). In general, the twice fine-tuned perplexities were lower than the single fine-tuned perplexities, indicating that training the model on generic fairy tales before training it on a author-specific dataset leads to better fairy tale text generation.

Additionally, with the exception of Andersen’s model, the perplexities of the twice fine-tuned models were lower than the perplexities of the baseline model.

### 6 Overall Qualitative Analysis

While there are common words among all three word clouds in the “Qualitative Results” appendix (*child, beautiful, little, one, red*), there are also interesting differences. For example, the words *wolf* and *eat* made an appearance in the Brothers Grimm word cloud, but are absent from the text generated in the style of Perrault and Andersen.

These subtle differences indicate that while fairy tales written by different authors share some common motifs, they might also differ in terms of theme and sentiment.

Another interesting observation that can be drawn from the “Qualitative Results” appendix can be seen in the Perrault generated texts. Perrault is known for adding quips and commentary throughout his tales, and this stylistic trait is noticeable in both generated samples.

### 7 Conclusion

The hypothesis for this project was that the model will have a lower perplexity when tested on fairy tale texts when it is fine-tuned on more

fairy tale data. As expected, training the model on a general set of fairy tale texts, then on author-specific fairy tales minimizes the perplexity of the model and thus allows the model to generate more unique and plausible fairy tale texts, when compared to training the model only with fairy tales from a specific author or only with general fairy tales or with no fairy tales at all. However, the group's findings also indicate that the values of perplexity and the model's ability to generate plausible and unique fairy tale text can vary depending on whether the fairy tale data used share a common theme and sentiment. This means that the more commonalities in the writing styles and sentiments of the fairy tale texts used to train and test, the more likely the model will be able to generate plausible and unique text. More research can be done to determine how great of an impact these commonalities have on fairy tale text generation.

## 8 Limitations

The perplexity results for the Andersen-specific model did not align with the results for the other models that were analyzed. Upon further analysis, the generated texts from the Andersen-specific model (both singly and doubly fine-tuned) had some missing spaces. Therefore, the odd perplexities were believed to be the result of some sort of tokenization error. Efforts to determine the root of this error were unsuccessful. However, the tales produced by the Andersen-specific model were still coherent overall.

It is also worth mentioning that some of the words in the word clouds included in the "Qualitative Results" appendix were skewed by the provided prompts. For example, the words *cave* and *sea* may not have been present in the clouds had they not been included in the prompts.

## 9 Future Work

While the generated fairy tales often included understandable (albeit strange) characters and plots, there is future work that could be done in this area in order to improve the generated texts.

Firstly, gathering more fairy tale data to train the model with would most likely make a significant difference. For the twice fine-tuned model, more data could be added to both the generic dataset and the author-specific datasets in order to provide the model with more information that could allow it to generate more unique and plausible fairy tale texts. However, as mentioned earlier in this paper, this would require more manual work to ensure that the generic dataset does not overlap with the author-specific datasets.

Additionally, having multiple generators focus on and generate different parts of fairy tales might lead to higher readability and lower perplexity. For example, having one generative model focus specifically on the beginnings of fairy tales (which often contain fairy tale patterns such as absentation and interdiction) might lead to higher quality fairy tale openings than the ones created by this project's models.

## Acknowledgements

We would like to thank Dr. Mandy Korpusik for her continuous support and encouragement throughout this project. We would also like to offer our gratitude to Dr. Kelly Younger for teaching us what we now know about fairy tales and for inspiring the creation of this project. Additionally, this project would not be possible without the use of OpenAI's GPT-2 model.

## References

- Mernagh, Michael. Predicting Short Story Endings. [web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2735427.pdf](http://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2735427.pdf).
- Nasrin Mostafazadeh, Nathanael Chambers, Xiaodong He, Devi Parikh, Dhruv Batra, Lucy Vanderwende, Pushmeet Kohli, and James Allen. A corpus and evaluation framework for deeper understanding of commonsense stories. arXiv preprint arXiv:1604.01696, 2016.
- "Past 6.806/6.864 Projects." [Courses.csail.mit.edu, courses.csail.mit.edu/6.864/fall\\_2016/past\\_project\\_s.html](http://Courses.csail.mit.edu/courses/csail.mit.edu/6.864/fall_2016/past_project_s.html).
- Rajapakse, Thilina. "Learning to Write: Language Generation With GPT-2." *Medium*, Apr. 2020, [medium.com/swlh/learning-to-write-language-generation-with-gpt-2](https://medium.com/swlh/learning-to-write-language-generation-with-gpt-2).



But the little daisy can be held together and strengthened by the love which she gives to it.

Charlotte is a girl who can be read; but the book she is reading is nothing, because she has never read the book she had intended to read, but only a sketch, for the drawing will remain for years; she cannot think of that again.

Now when she looks at the daisy, she sees a young angel, and a beautiful girl, but she cannot perceive them.

She cannot see them, for the angels are not there.

The little daisy is sitting up in her arms, and the little angel has been kissing her, and they are kissing her heart, and her lips are moving through the air.

### Generated Tale Sample #2

*Once upon a time there was a young girl who was the fairest of them all.*

*Her sisters were jealous of her beauty.* She was beautiful, they believed.

And when she was about ten years old she was married to a noblewoman.

The girl was very clever, and had all the qualities, while her mother believed that this girl should never have a chance to be with her; so that she had no chance of ever being with her.

The only thing which the daughter of her master could wish to know was how she was to remain in a kind home with the noble people of her birth--in which the mother remained for half a year, and the child was always with her. It was a very beautiful time, in the summer.

The Dryad came out of the country in winter to visit her, which was certainly of great importance; so it was very pleasant and interesting to see her, to have her, and to see all her beautiful gifts.

The Dryad had a splendid home in the town; a pleasant house that had been built on top of hers; and a splendidly furnished yard; but the whole arrangement was only to give the dryad much entertainment; and so she became quite convinced that the house, in which she lived, was to serve as her home.

### Charles Perrault



### Generated Tale Sample #1

*Alone in the woods was a small girl in a red cape. She was on her way to bring bread to her grandmother.*

Now this poor thing had got a good deal of money out of the boy, and had given him money to buy a pretty little hut, which he had cut up on the outskirts of town.

Now this little hut was called The Fox's Fair, because, as far as was known, it lay on a ridge above, behind the house, and was called The Fox's Mountain, for the hill that is just at the top of this rock, over there a very dangerous place.

In the forest stood a deep ditch.

Here there was once a girl who was very kind, and to give good advice to the man who was to come out of the forest, and tell him why his brothers had got into this dangerous place; but he had a wicked old cap lying upon his head, as the poor little girl told him.

### Generated Tale Sample #2

*A long time ago, there was a couple who wanted more than anything to have a child of their own.*

The eldest was very fond of having one, and the other very fond of having two; so she laid hold of her father's neck, and asked: 'Dear father, do you wish to have a child?' 'Yes,' replied the father, 'but if my child be given to me I will give you to the poor girl, for she cannot bear to keep herself in this world.'

So she was given a little child, and by marriage her mother lived with her, and they lived for three years together very happy.

Then she sent them all up into the world, for they called the little people: and this the eldest sister did, and was called Rose-red, and was called Lily-red;

and when she was old enough, and could go no farther, she left the little children to her, and set them to work and get wood, and make furniture, and clean the house; and it was not long before she got pretty well enough to draw some wood out of the wood, so as to put one in every corner of the house, to make beds, beds for her maidens, and cook the cook's table, or whatever else she would like.