Improvement of Natural Language Generation Models on Mimicking Historical Authors

Lingtong Xu

***Abstract*—Natural Language Generation (NLG) is an easily qualitatively understood way to see the steady improvements of models, from pre-machine learning, to deep learning, on the task of parsing and understanding human language. This paper looks to apply a Markov chain, recurrent neural network, long-short-term memory, and GPT-2 on the task of generating text that mimics the style of three historical authors: Edgar Allen Poe, William Shakespeare, and Homer. The paper will pinpoint the areas of improvement between each model using qualitative metrics, such as the use of grammar, generating coherent ideas, and degree of similarity to the style of the author to demonstrate how work in machine learning applied language-based tasks has improved throughout the years. Specifically, the Markov chain is able to string reasonable word possibilities together, but has no mechanism to create reasonable, grammatically correct sentences. The recurrent neural network is capable of learning sequence-related features but cannot retain long term information due to exponentially decaying signals. The long-short-term memory model is better at selectively retaining information from earlier in the sequence but suffers in scalability. The GPT-2 model utilizes attention *without* sequential operations to encode the information of a sentence, making it more parallelizable and thus scalable, and does not have any issue with long-term information retention.**

I. Introduction

THIS project is an investigation on the usage and evolution of statistical, classic machine learning, and deep learning methodologies on the task of Natural Language Generation (NLG). Using models to understand or emulate natural human language has been a common and hotly researched subject in the field of machine learning. The degree of accuracy that a model is now able to copy the style of a famous historical author is an interesting and accessible way to see the leaps and bounds that have been made in the subject of Natural Language Processing (NLP), and NLG specifically. The three target authors that the models chosen in this project will imitate are William Shakespeare, Edgar Allen Poe, and Homer. Given the qualitative nature of the task, exact metrics and quality of output will be difficult to measure. This project will instead look to compare the model outputs in terms of generally subjective measures, such as proper grammar, proper structure, and degree of likeness to the original author. This project will also discuss the improvement of these generated texts as the generating models improve towards state-of-the-art results. Though simulations will be performed, the project is not meant to innovate on previous models. Thus, this project stands as both a review paper, and a simulation paper on the task of NLG. The four models used in this investigation (with increasing degrees of complexity) are Markov Chains (MC), a standard Recurrent Neural Network (RNN) [1], a Long Short-Term Memory (LSTM) model [2], and the GPT-2 model [3] proposed by OpenAI, and available to the public.

II. Overview of Models Used

This section will briefly describe the architectures that will be used for the NLG task, along with the hyperparameters used to train them. As mentioned, the models will increase in complexity, from more traditional non-machine learning models, to state-of-the-art deep learning models.

A. *Markov Chains (MC)*

MC models were first proposed by Andrey Markov, and map out possible discrete states and the probability to transition into a separate state based on the current one [4]. When applied to NLG, the model uses separate words as states, where the progression of a sentence acts as the state transitions. Having seen two sentences such as: “The big red dog crossed the street.” and “The big red chicken crossed the street.” A MC model would say that from the word state “the”, the possible progressions are the two states “big” and “street.”. From the state “crossed”, however, the only possible state transition is to “the”. By looking at numerous examples, a MC is able to generate a series of words by creating a large network of possible transition states.

In this investigation, I use a bigram MC, which means instead of predicting the next word from a single word, it will use two previous words to predict the next (Fig. 1). There are no major hyperparameters for the MC model.

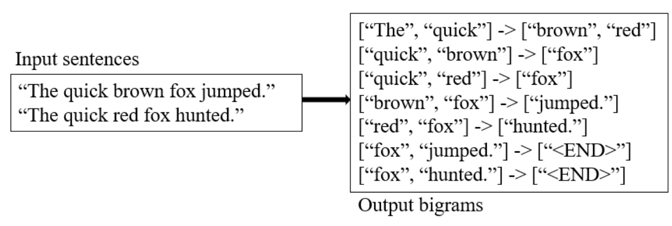

Fig. 1. Bigrams generated from two sample sentences for the MC model.

B. *Standard Recurrent Neural Network (RNN)*

RNNs were first made based on Rumelhart’s studies in 1986 [1]. The intuition behind the function of this model is directly related to the term “recurrent” in its name. In language processing, the previous word in a sentence is typically an important indicator for the possible following words. RNNs were proposed as a model that take advantage of this kind of sequential information, through the usage of either a hidden state or directly using the output at the previous time step as an input to the model at the current time step. For this investigation, the latter will be used, and a simple block diagram describing its function is depicted in Fig. 2. Mathematically, the RNN model relating input to output is [5]:

$$h_t = y_t = \tanh(W_{ih}x_t + b_{ih} + W_{hh}h_{t-1} + b_{hh}) \quad (1)$$

Where x is the input, y is the output, which is equal to hidden state h for a given time step, t. W_{ih} are the weights relating the input to the hidden state, W_{hh} are the weights relating the previous hidden state to the current, and b are the bias terms with the same relations.

Prior to the RNN model, there is an embedding layer to convert integer-encoded words into an embedded vector, and following the RNN is a dense layer converting back to integer-encoded words. The parameters for training are summarized in table 1 below.

TABLE I
TRAINING PARAMETERS FOR THE RNN MODEL

Parameter	Value
Learning rate	0.001
Optimizer	Adam
Weight decay (L2 regularization param)	0.0001
Beta 1 (for Adam optimizer)	0.9
Beta 2 (for Adam optimizer)	0.999
Epochs trained	4
Batch size	16
Sequence length	32
Embedding size	300
RNN model size	256
RNN layers	2
Dropout (on RNN layers)	0.1
Loss function	Cross Entropy

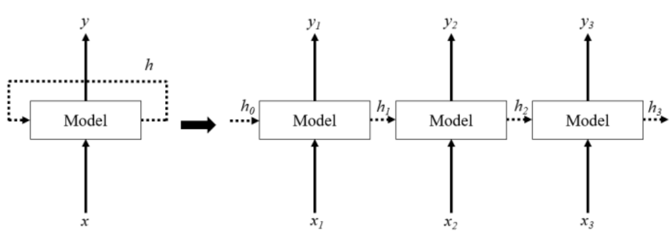

Fig. 2. Visualization of the basic structure of the standard RNN model. x denotes the input, y the output, and h the hidden state of the model. The left shows the model's recurrent nature, and the right unrolls the model to show how the previous timestep's hidden state is fed as an input to the next timestep.

C. Long-Short-Term Memory (LSTM)

A major issue faced by RNNs, is the exponential decay or increase of error signals as time steps increase depending on the relative importance put on retaining information from previous time steps, versus the input information at the current time step [6]. To remedy this issue, Hochreiter et al. proposed the LSTM model, containing a mechanism to selectively retain or forget relevant information from previous time steps [2]. The main differences between the LSTM and the RNN are in the introduction of the cell state, and gates to selectively control the sequential information leaving and entering the cell. In this way, if required, a LSTM model is capable of retaining very long-term information.

The model is mathematically described as follows [5]:

$$\begin{aligned}
i_t &= \text{sigmoid}(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}) \\
f_t &= \text{sigmoid}\left(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}\right) \\
g_t &= \tanh\left(W_{ig}x_t + b_{ig} + W_{hg}h_{t-1} + b_{hg}\right) \\
o_t &= \text{sigmoid}(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}) \\
c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\
h_t &= y_t = o_t \odot \tanh(c_t)
\end{aligned} \quad (2)$$

Where x is the input, y is the output, which is also fed back into the model as a hidden state h, and c is the cell state for time step t. W and b are the weight and bias terms respectively. i is the input gate, selectively controlling the amount of information from the current input inside the cell state. The forget gate, f, controls how much of the previous cell state is included in the current cell state. The cell gate g contains the actual input information that may be passed to the new cell state. The final gate o is called the output gate and denotes the amount of the cell state that will be passed to the output. $\odot$ is an element-wise multiplication operator [5].

The LSTM model contains an embedding layer and dense layer as well, before and after the LSTM layer, respectively. Table 2 following describes the parameters used for this model.

TABLE II
TRAINING PARAMETERS FOR THE LSTM MODEL

Parameter	Value
Learning rate	0.001
Optimizer	Adam
Weight decay (L2 regularization param)	0.0001
Beta 1 (for Adam optimizer)	0.9
Beta 2 (for Adam optimizer)	0.999
Epochs trained	4
Batch size	8
Sequence length	32
Embedding size	300
LSTM model size	256
LSTM layers	2
Dropout (on LSTM layers)	0.1
Loss function	Cross Entropy

D. GPT-2

GPT-2 is a state-of-the-art Transformer model created by OpenAI and made available for public use [3]. The model takes advantage of the attention mechanism proposed by Vaswani et al., removing the need for recurrent layers and making the model easier to parallelize [7]. I utilized the pretrained 117 million parameter (small) model, and individually trained it for 200 training steps on the Poe, Shakespeare, and Homer datasets.

Unlike the other models used in this investigation, the GPT-2 is pretrained on a significant amount of Internet text, which may affect the style of the output and introduce vocabulary that is not present in the datasets individually. The training parameters of this model are the default settings from the original paper.

III. Dataset and Preprocessing

For this investigation, I used the authors Edgar Allen Poe, William Shakespeare, and Homer. From these authors, I sampled all of Shakespeare's and Poe's work, and the Iliad from Homer. All these texts were obtained online through the Project Gutenberg website. The dataset sizes were quite varied, with the Poe dataset containing 46182 distinct vocabulary, and 452096 total word tokens extracted from the text, the Shakespeare dataset containing 68506 distinct vocabulary, and 1041488 total word tokens, and the Homer dataset containing 12157 distinct vocabulary, and 157104 total word tokens. For the deep learning models (RNN and LSTM), I tried to minimize preprocessing, both in the interest of time, and to give the models a chance to learn punctuation and proper grammar. The text was split by spaces, and periods were each treated as a separate word. The vocabulary was counted and mapped to a distinct integer for input to the deep learning models. For a given input word in a sequence, the output target is the following word in the sequence.

IV. Results and Discussion

After training the models as specified in section 2, text was generated using a specific input context, which is a short sequence of words found in the original text that acts as a starting point for the generation task. For the MC model, the seed was a bigram, and the rest of the text was generated by chaining together a generated word with the second word in the previous bigram. The deep learning models used an input sequence of 5-7 words, and generation was simply predicting the following word based on the prior sequence and adding the predicted word into the sequence to create a chain of predictions. The lengths of the generated sequences are all 100 words long, with topk sampling for the deep learning models, where k=10. The results of the generation are given in appendix A to C in tables 3 to 8.

In general, the MC model appears to generate reasonable results when looking at a narrow range of the sequence, as each group of 3 words (bigram with following word state) exists in the text itself. However, given that the selection is completely random, grammar, style, and the general theme of the sentence is typically incoherent. For example, the inclusion of punctuation such as quotation marks, which require that a previous state pass information forward to a later state, will likely result in improper grammar. This is seen in tables 4 and 7. In addition, with unique and uncommon bigrams, such as the input seed to table 6, "Liberty! Freedom!", the possible following word states are typically limited. This causes large chunks of the original text to appear in the generated sequence, which is evidenced by the table 6 MC output matching the original work up to "cry it".

In the RNN outputs, there is a distinct improvement on the grammar and general theme of the sequence. In table 5, the entire generated text revolves around words such as "heart" and "love", likely meaning the model was able to learn the relation between the words, with the hidden states propagating this information throughout the sequence. Also, in table 6, characters such as Gloucester and Salisbury are mentioned recurrently, evidencing the existence of propagating sequential information.

The LSTM-generated text improves marginally upon the RNN outputs, improving slightly in its ability to focus on a theme, and improving on grammar. Particularly, in the beginning of the output of table 7, the model retains the theme of referring to characters as in the way, "O, Apollo" in a manner that makes grammatical sense.

Both the RNN and LSTM outputs have reasonable likeness to the style of the original author, however, grammatical errors still exist, and there is a lack of coherency between the beginning and end of the sequence. This is likely a limitation of models that have sequential dependency, as though the LSTM is better at carrying information about the previous state, there is always reliance on computations in previous timesteps, which are liable to introduce information loss later in the sequence.

The GPT-2 model improves on this, generating very few grammar errors, and being able to express a complex idea from the start to the end of the sequence with no loss of information. Since the GPT-2 does not have time-step dependencies, and instead views an entire sentence simultaneously [7], the model is able to generate complex ideas with no signal loss as seen in table 5. Likely due to the repetitive nature of the input seed, the output is also repetitive, but certainly appears to develop a specific (though nonsensical) theme throughout the passage. Due to the use of pretrained weights of this model, there is a degree of style mismatch from the model outputs, but the generated sequences contain proper grammar, and could quite possibly be mistaken for human-written text. Tables 7 and 8 are excellent examples with good sentence structure and match the style of the author well.

V. Conclusion and Future Work

This investigation utilized four models of varying complexity on the NLG task of mimicking the writing styles of Edgar Allen Poe, William Shakespeare, and Homer. The results improved between the models, demonstrating how their features innovated upon their predecessors. The RNN and LSTM improved on the MC model by sequentially passing on the general theme of the text using recurrent states. This allowed the deep learning models to generate more grammatically correct sentences and use words with similar connotations (e.g. "heart" and "love"). The GPT-2 Transformer model heavily outperformed the previous models in both producing sequences with proper grammar, and having coherency throughout the generated text, as it does not rely on sequential information altogether.

For further investigation, it may be interesting to compare against other models that have been used for NLG, such as convolutional neural networks, and also it may be worth investigating and adjusting the GPT-2 settings and hyperparameters to improve results on particular authors.

REFERENCES

[1] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[2] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[3] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, “Language Models are Unsupervised Multitask Learners,” 2019.

[4] D. A. Levin, Y. Peres, and E. L. Wilmer, “Introduction to finite Markov chains,” Markov Chains and Mixing Times, pp. 2–20, 2017.

[5] “torch.nn,” *torch.nn - PyTorch master documentation*. [Online]. Available: https://pytorch.org/docs/stable/nn.html.

[6] S. Hochreiter, “Untersuchungen zu dynamischen neuronalen Netzen,” Apr. 1991.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention Is All You Need,” *CoRR*, 2017.

VI. Appendix A: Edgar Allen Poe Output Samples

<table>
<tr><th colspan="4">TABLE III
GENERATED OUTPUTS FOR EDGAR ALLEN POE 1</th></tr>
<tr><td colspan="4">Original work:
Once upon a midnight dreary, while I pondered, weak and weary,
Over many a quaint and curious volume of forgotten lore,
While I nodded, nearly napping, suddenly there came a tapping,
As of some one gently rapping, rapping at my chamber door.
“‘Tis some visiter,” I muttered, “tapping at my chamber door—
Only this, and nothing more.”
Ah, distinctly I remember it was in the bleak December,
And each separate dying ember wrought its ghost upon the floor.
Eagerly I wished the morrow;--vainly I had sought to borrow
From my books surcease of sorrow--sorrow for the lost Lenore—
For the rare and radiant maiden whom the angels name Lenore—
Nameless here for evermore.</td></tr>
<tr><td>Markov Chain output</td><td>RNN output</td><td>LSTM output</td><td>GPT-2 output</td></tr>
<tr><td>Once upon a highly rarefied atmosphere in all _of the question is of a hundred of the city of Benares, upon the staircase. Being now assured that I have keys, as you can. If you see anything strange, let me go out at least one of singular wildness, and its circumstances would permit, and even anxious to have been discovered through the “depression” (already described) in the whole surface of the beautiful ANNABEL LEE; And the door of brass. But the voices in contention. The gruff voice was not in the atmosphere. These gradations of matter to compose an article.</td><td>Once upon a midnight dreary, of which, in a measure, degree their personal appearance, became of the cook had, who little more terrible and let fall a few words, they might only give us to any attempt . “But to take him two two men have no conversation of a good deal about that time to come about two miles six points . They did only let us, when our readers will have since no time could not distinguish it with but a more serious task would than nothing had done to hear anything but was composed in my company but</td><td>Once upon a midnight dreary, and while at length came in me that she did still not yet more upon its nature of so feeble into his terror for its own extent when this would have seen that I would well as to take place upon his body . It might now let against it this, we did not . But we were not even upon some time until before that they were so long so that, a single one, being at length upon its three points from one hundred to us where I did near it, but if there came</td><td>Once upon a midnight dreary, a sudden thunder shook the city. The thunder of the thunder, as it grew, grew louder. There was the sound of an awful explosion. It was like the noise of a thousand fires in the night air. The wind grew so strong that the thunder ceased, and there came again the roar of a thousand thousand men. In the next moment there came again the roar of a thousand thousand men. They were rushing, like a thousand hundred men rushing at once. The sound of an awful explosion, as it grew louder, rose again. It</td></tr>
</table>

Table 3. Outputs from the four models to the first starting sample from the Edgar Allen Poe dataset are shown above. The sections bolded are the starting context for the rest of the text generation. All deep learning models use top-k sampling for generation, with k=10. The length of each sample is 100 words (where a period counts as an individual word)

TABLE IV
GENERATED OUTPUTS FOR EDGAR ALLEN POE 2

Original work:
DURING the whole of a dull, dark, and soundless day in the autumn of the year, when the clouds hung oppressively low in the heavens, I had been passing alone, on horseback, through a singularly dreary tract of country; and at length found myself, as the shades of the evening drew on, within view of the melancholy House of Usher. I know not how it was--but, with the first glimpse of the building, a sense of insufferable gloom pervaded my spirit. I say insufferable; for the feeling was unrelieved by any of that half-pleasurable, because poetic, sentiment, with which the mind usually receives even the sternest natural images of the desolate or terrible. I looked upon the scene before me--upon the mere house, and the simple landscape features of the domain--upon the bleak walls--upon the vacant eye-like windows--upon a few rank sedges--and upon a few white trunks of decayed trees--with an utter depression of soul which I can compare to no earthly sensation more properly than to the after-dream of the reveller upon opium--the bitter lapse into everyday life--the hideous dropping off of the veil.

Markov Chain output	RNN output	LSTM output	GPT-2 output
DURING the fall be, and the readiness with which I had observed in our way at Tortoni's)--Write, if you will then be mine, for I recollected turning up first one of whose truth appears now, to the contrary. I thence concluded that his lips were perfectly glazed; and as visions only, while the Art is made of them, and caught his Satanic Majesty--in a blush. But the Duc's heart is fainting within him. He replied feebly, yet quite audibly, "Yes, I nose dat--nose all bout dat--tis my lef hand of the audience, survey me, from head to the southward and.	**DURING the whole of a dull,** few seconds of off for her own part, with their progress was so well calculated that from which they had never taken possession the arm of a corpse would not to see that of such way was found floating beneath of her breast on board as this, I knew the good reason and let the cause of their operations until some minutes before our attention as we are in fact which is, the present, would nearly as an error occurred whatever should undoubtedly had little difficulty of getting rid out, my had so entirely cut off	**DURING the whole of a dull,** low gale about three or ten of four feet in all a single kind . In this, he must let him, and that the body might not do could do I not know him if it any attempt . We went out by this means we could not make an trouble . As if in fact, he took me into an attempt with the aid into an atmosphere with great difficulty that in a few of them might naturally prove of an ordinary nature was of a singular nature as a very serious matter came out	DURING the whole of a dull, ersatz, and disconcerting atmosphere; the sound of a voice that seemed to be made from the recesses of a dead tree, the sound of a voice that was so distinctly, in some degree, that it sounded like a voice that had been heard by no one. There could have been no voice more powerful than these. There was no other voice, I suppose, but the voice of a voice which could not be heard. It was as if there were no voices, and it did not even appear to me to have been heard.

Table 4. Outputs from the four models to the second starting sample from the Edgar Allen Poe dataset are shown above. The sections bolded are the starting context for the rest of the text generation. All deep learning models use top-k sampling for generation, with k=10. The length of each sample is 100 words (where a period counts as an individual word)

VII. Appendix B: William Shakespeare Output Samples

TABLE V
Generated Outputs for William Shakespeare 1

Original work:
To be, or not to be, that is the question:
Whether ’tis nobler in the mind to suffer
The slings and arrows of outrageous fortune,
Or to take arms against a sea of troubles,
And by opposing end them? To die—to sleep,
No more; and by a sleep to say we end
The heart-ache, and the thousand natural shocks
That flesh is heir to: ’tis a consummation
Devoutly to be wish’d. To die, to sleep.
To sleep, perchance to dream—ay, there’s the rub,
For in that sleep of death what dreams may come,
When we have shuffled off this mortal coil,
Must give us pause. There’s the respect
That makes calamity of so long life.

Markov Chain output	RNN output	LSTM output	GPT-2 output
To be, or we are made, forgot as soon As I will do all from you; the which he made those hollows, if himself were compromis’d That all the instruments which aided to expose the child and parent. [Kneels] CORIOLANUS. What's the news? LE BEAU. Even he, your wife, this gentlewoman, Thy son is fled. SECOND MURDERER. I am not Talbot's blood That glues my lips upon; this hand, thou art As great to thee. To’t again, come. SECOND CLOWN. Go you before, John and master did well to do it: We know not his sister. Can’t no other, Her.	**To be, or not to be,** or I can do you in these love that she hath not my love and the heart hath the world that I was not . What I am to speak with you, and you were no more you than you are but the man and it that . But he was an arrant man; I had been in that you had had been seen for an honourable wife, my father to have an owle, This was an excellent part as his heart . But if a woman had I would the gods were he might bear	**To be, or not to be,** for that is much To hear the enemy that which shall go The King of our fair blood may have our friends . I cannot see this is this holy time And to be as my lord the Lord is not In that he hath done . I would it not have done By him . GLOUCESTER . Why, he has a thousand more . SUFFOLK at him, and all our lives As we shall not the Queen his brother shall come on, For you have all a man for your own hand. KING EDWARD	**To be, or not to be,** or not to be, or not to be, or not to be, are all the same, as are not in any respect alike. The first of the first four is the very next one, and so we see the third being the first. The fifth is the third one. The fifth is the fifth, and so we see the fourth being the third. The sixth is the sixth, and so we see the seventh being the sixth. The seventh is the seventh. The seventh is the seventh, and so we see the eighth being

Table 5. Outputs from the four models to the first starting sample from the William Shakespeare dataset are shown above. The sections bolded are the starting context for the rest of the text generation. All deep learning models use top-k sampling for generation, with k=10. The length of each sample is 100 words (where a period counts as an individual word)

TABLE VI
GENERATED OUTPUTS FOR WILLIAM SHAKESPEARE 2

Original work:
Liberty! Freedom! Tyranny is dead!
Run hence, proclaim, cry it about the streets.
CASSIUS.
Some to the common pulpits and cry out,
“Liberty, freedom, and enfranchisement!”
BRUTUS.
People and Senators, be not affrighted.
Fly not; stand still; ambition’s debt is paid.
CASCA.
Go to the pulpit, Brutus.
DECIUS.
And Cassius too.
BRUTUS.
Where’s Publius?
CINNA.
Here, quite confounded with this mutiny.
METELLUS.
Stand fast together, lest some friend of Caesar’s
Should chance—

Markov Chain output	RNN output	LSTM output	GPT-2 output
Liberty! Freedom! Tyranny is dead! Run hence, proclaim, cry it o’er my memory, As doth a sail, a sail! MERCUTIO. Two, two; a shirt and a step-dame false; A foolish thing was but nine months old. Was never subject long'd to hear what he hath, like an hypocrite, The which hath as long lov’d me As well as you will nothing weigh: Your vows to crown himself in arms. Let us go see poor Cassio dress’d. Come, mistress, now perforce we will commit thee thither Until his very words. What doth concern your coming. WESTMORELAND. Then, my queen, and Clifford.	**Liberty! Freedom!** **Tyranny is dead!** KING . If we do see his Grace will go his son A traitor from the Duke of York, I must perforce . Exeunt ACT 4 SCENE 3 A public part of France in the hall Enter WARWICK GLOUCESTER with King or five thousand LORDS GLOUCESTER and CHIRON, SALISBURY in their tent GLOUCESTER WARWICK And here I am in my commission here . SALISBURY is not like his Grace And see that have been sick, with these the world To make his grace that with the King and I To him the name that I might	**Liberty! Freedom!** **Tyranny is dead!** What is he not? What a young man? Enter the Lady EPHESUS How to your Grace, you know what we do here? GLOUCESTER SENIOR to your Majesty . QUEEN MARGARET As as it had no cause to have so great I fear him that . What is he? Enter the King, his lord, Sir Robert and King Edward hath sent up With your father's hand that ever was so To hear a word to hear that he will take it To him the holy heir to him at your hand . SUFFOLK and him, I do	**Liberty! Freedom!** **Tyranny is dead!** The great King of England, As he was a man of great power, Than, with that noble voice, that noble air And that proud eye He had on him, as a great man of the highest rank of mankind, To say to his face: "I, my King, Forgive me the death which my death Upon me is a great loss; I am the cause of the whole. I shall never again be The King of France in the eyes of the Crown; I shall never again live in the world of England; For the King

Table 6. Outputs from the four models to the second starting sample from the William Shakespeare dataset are shown above. The sections bolded are the starting context for the rest of the text generation. All deep learning models use top-k sampling for generation, with k=10. The length of each sample is 100 words (where a period counts as an individual word)

VIII. Appendix C: Homer Output Samples

TABLE VII
Generated Outputs for Homer 1

Original work:
Sing, O goddess, the anger of Achilles son of Peleus, that brought countless ills upon the Achaeans. Many a brave soul did it send hurrying down to Hades, and many a hero did it yield a prey to dogs and vultures, for so were the counsels of Jove fulfilled from the day on which the son of Atreus, king of men, and great Achilles, first fell out with one another.
And which of the gods was it that set them on to quarrel? It was the son of Jove and Leto; for he was angry with the king and sent a pestilence upon the host to plague the people, because the son of Atreus had dishonoured Chryses his priest. Now Chryses had come to the ships of the Achaeans to free his daughter, and had brought with him a great ransom: moreover he bore in his hand the sceptre of Apollo wreathed with a suppliant's wreath and he besought the Achaeans, but most of all the two sons of Atreus, who were their chiefs.

Markov Chain output	RNN output	LSTM output	GPT-2 output
Sing, O goddess, the anger of the Achaeans." "Ajax," replied Achilles, "noble son of lovely Helen aimed an arrow at it, he said to him, pulled the spear I leave to those who were being overcome by the foremost of them in the fray, while we two fighting on my behalf, lest some counsellor of the fight was raging. They could neither spring forward and vaunted over him and said, "Take me alive; I will not bury you, till I had taken all the fleet son of Anchises led the people gathered thereon. But first hold up your minds to.	**Sing, O goddess, the anger of Achilles** stay you, by this man, can respect for sorrow to kill you stand higher way . When they were being together when they are in due broad sea, when my horses, are so that a god has held both men that dwell there I come and fight as one who had be the son is too with your water stand before your chariot, my name and say, whereas all other immortals . Therefore lay in my sorrow may bring out against one another's the stronger we make trial stand face here, but on the left open	**Sing, O goddess, the anger of Achilles** answered, and you among them; O Hector, O noble Peleus, answered, O Apollo, what we father you are so far on your tent the daughter . We are able from destruction from some our parents, yet when I let us go without behalf how close behind us, we are also in your own hand upon a body and go up to the house a Hades, We would it drive for our dogs . men must stand our great city to protect Troy as the city have been driven back far from some fighting by reason and	**Sing, O goddess, the anger of Achilles**; let him that is in his anger fight the son of Peleus against the son of Peleus. Then let him not go on fighting, but be kind, that he may have the glory of Olympus, and the renown of her sons. He will not let him be in trouble and will keep his anger, and his anger is not so great as when the Argives took the sea-trees and laid waste the city of Tyre; but he will fight and take the city of the son of Peleus, and he will be glad

Table 7. Outputs from the four models to the first starting sample from the Homer dataset are shown above. The sections bolded are the starting context for the rest of the text generation. All deep learning models use top-k sampling for generation, with k=10. The length of each sample is 100 words (where a period counts as an individual word)

TABLE VIII
GENERATED OUTPUTS FOR HOMER 2

<table>
<tr><td colspan="4">Original work:
He poised his spear as he spoke and hurled it. Hector saw it coming and avoided it; he watched it and crouched down so that it flew over his head and stuck in the ground beyond; Minerva then snatched it up and gave it back to Achilles without Hector's seeing her; Hector thereon said to the son of Peleus, "You have missed your aim, Achilles, peer of the gods, and Jove has not yet revealed to you the hour of my doom, though you made sure that he had done so. You were a false-tongued liar when you deemed that I should forget my valour and quail before you. You shall not drive spear into the back of a runaway- drive it, should heaven so grant you power, drive it into me as I make straight towards you; and now for your own part avoid my spear if you can- would that you might receive the whole of it into your body; if you were once dead the Trojans would find the war an easier matter, for it is you who have harmed them most."</td></tr>
<tr><td>Markov Chain output</td><td>RNN output</td><td>LSTM output</td><td>GPT-2 output</td></tr>
<tr><td>He poised his spear and strode in full flood from the store-chamber the rich ransom that he will make, and fitted with ancle-clasps, and with Hector, while the host and deem that we should get his body and go down to the ships yourselves, for Jove with his mighty sword from its scabbard, Minerva came down from his wife; then he fell heavily to the house of the spear severed all the Abantes. With him they were of your mind. I say further, and go about among the ranks of the horses raised a barrow that is heaped over my.</td><td>He poised his spear as he spoke and killed himself his warrior . Then he yoked looking everywhere he at Troy cheering to stand by fierce like Trojan wild beasts other, and called aloud on a great stone, moreover, that on these might and with one another of evil will make him if the Argives there give you great gifts of to come whom my mother from me, for if it had be borne that Apollo will not hear even when we fight you have many of them as we were fighting against me, whose wife has been better quick for and see</td><td>He poised his spear as he spoke like Jove . His mother then threw hold of her own upon her, while Minerva himself made all her who went down again with its eyes . Now so long as Phoebus Apollo, I was a man, take heart for the first than be himself in the dust that there was the Trojans from fighting . The day when mighty As he spoke, but put his eyes upon them, but Hector might fall and all side till heaven should fly from them, while Diomed rose towards him, saying, "Hear him O goddess whom your father mother</td><td>He poised his spear as he spoke, for the great stone wall, with two long horns, and four spikes of bronze; he was clad in the armour of his comrades. "The mighty warrior, O man of his armour, is now the hero of the people."
He answered, "O son of man, son of Aetius, you will not do the things you have said; it is not for me to be the hero. I am here, but you will not help me. You are the only man who can help me; it is not that you are afraid, but your fear,</td></tr>
</table>

Table 8. Outputs from the four models to the second starting sample from the Homer dataset are shown above. The sections bolded are the starting context for the rest of the text generation. All deep learning models use top-k sampling for generation, with k=10. The length of each sample is 100 words (where a period counts as an individual word)