

Name: David Schwartzman

Date: 21/11/2025

LSTM Model

First Model With One Layer and 64 units:

A few examples of my output of my first model with one layer and after training it for 50 epochs:

- 1) Seed prompt- to be or not to be (Temperature: 0.5):

Generated text: **to be or not to be franke with me when loue i see vpon the rest of me to thee i is when an self but**

- 2) Seed prompt- shall I compare thee to a summer's day(Temperature: 0.1):

Generated text: **shall i compare thee to a summer's day of this contract to moue them i am gone i am the greatest able to chide vp all me ride**

- 3) Seed prompt- all the world's a stage (Temperature: 1.0):

Generated text: all the world's a stage ile stand and for his torch must art i am and they do heare her other life to thee ist

Performance analysis:

Training loss decreased from 6.6522 in epoch 2 to 1.3492 in epoch 50.

Accuracy improved from 3% to 71% over 50 epochs, suggesting the model is learning patterns from the data.

Each epoch took almost two minutes to complete using TPU.

Overall, despite improving losses and accuracy, the model struggles with coherence, semantic relevance, and generating meaningful content, especially at higher temperatures.

Second Model With Two Layer, 128 units on first layer, 256 on second layer:

A few examples of my output of my first model with one layer and after training it for 50 epochs:

- Seed prompt: to be or not to be (Temperature: 0.1)

Generated text: **to be or not to be wrought that tell me in and as me seem knowst thou art it should be so of that so before**

- Seed prompt: shall I compare thee to a summer's day (Temperature: 0.5)

Generated text: **shall i compare thee to a summer's day i come is that a end of your high wound no let a houre cliii doth mind those old pride**

- Seed prompt: all the world's a stage (Temperature1.0)

Generated text: **all the world's a stage didst with him after such art back of him as fast be said in thee so like slow so oh**

Performance analysis:

Training loss decreased from 6.9970 in epoch 2 to 0.0259 in epoch 50.

Accuracy improved from 2% to 93% over 50 epochs, suggesting the model is learning patterns from the data.

Each epoch took almost two minutes to complete using TPU.

Even though at the end of training our loss and accuracy were better than the first model will only one layer, the sentences are still not very clear.

Last Model With Three Layers and 128 units at each layer (the submitted model):

A few examples of my output of my third model with three layers and after training it for 50 epochs (in my submitted model you can see all of them):

- Seed prompt: to be or not to be (Temperature: 1.0):

Generated text: **to be or not to be old without those bound now you on the huswife of the instant grace your white app compare the very eye**

- Seed prompt: shall I compare thee to a summer's day (Temperature: 0.1)

Generated text: **shall i compare thee to a summer's day to be peruerse and mayst not prooffe for all away speakst thou not iule to moue me to thy wounds**

- **Seed prompt: all the world's a stage** (Temperature 0.5):

Generated text: **all the world's a stage which should not fears for titans wheeles thus runnst no more and meant robbd that enfeebled thinke life cheeks by**

Overall analysis of the generated text with my main model (with three layers and the model I submitted)

1. Seed Prompt: "to be or not to be"

- Temperature 0.1: The text is mostly nonsensical with fragmented sentences, struggling to maintain coherence. The low temperature causes the model to rely heavily on high-probability words, resulting in repetitive and incomplete ideas.
- Temperature 0.5: The output is more diverse and somewhat coherent, though still awkward. The model introduces a bit more creativity and unpredictability, but some parts are nonsensical.
- Temperature 1.0: The text starts to show more varied vocabulary and structure, though it remains a bit strange and inconsistent. Higher temperature introduces more unpredictability, leading to more unique but less coherent results.

2. Seed Prompt: "shall I compare thee to a summer's day"

- Temperature 0.1: The output is very rigid, following a predictable structure but with little creativity. The model's low randomness results in a bland and disjointed continuation.
- Temperature 0.5: There's more variation in word choice and phrasing, but the output still contains strange constructions. The balance of creativity and coherence is better than at 0.1, but some parts still don't quite make sense.
- Temperature 1.0: The output becomes more creative and unpredictable, with more varied sentence structures. However, this comes at the cost of coherence and meaning, with some parts feeling disconnected.

3. Seed Prompt: "all the world's a stage"

- Temperature 0.1: The output is quite rigid and lacks logical flow, with more focus on predictable phrases. It's close to a formal style but doesn't make much sense.
- Temperature 0.5: There's a slight increase in creativity and some unpredictable elements, but the text still has awkward phrasing and doesn't maintain a coherent narrative.
- Temperature 1.0: The output is more varied and creative, though still somewhat jumbled. It includes more unique expressions but lacks overall coherence, as higher randomness leads to less structured outputs.

Temperature analysis:

- Temperature 0.1: Generates text that is repetitive and lacks creativity, relying heavily on common, high-probability words.
- Temperature 0.5: Balances creativity and coherence better, producing more varied outputs with some awkwardness.
- Temperature 1.0: Generates highly varied and unpredictable text but at the cost of coherence and readability.

How varying units affects model generated text:

Increasing the number of units in a text generation model leads to better coherence, diversity, and overall quality of generated text. However, too much of an increase in the number of units might not significantly improve quality but may increase the risk of overfitting and computational cost. Therefore, the

optimal number of units depends on the specific use case and the size of the training dataset.