Jonathan Mandl 211399175
Danielle Hodaya Shrem 208150433

Part 6: N-gram based language-model

We repurposed our tagging architecture into a character-level language model.
The model takes as input a fixed-length window of k characters and predicts the next character.

We used a similar architecture to our tagger code from part 1 of the assignment. Each character in the window of length k was assigned an embedding vector and the embedding vectors of the k characters were concatenated and fed into an MLP with one hidden layer.

We made a few changes to the original model: we assigned each character a 64-dimensional embedding vector. We changed the size of the hidden layer to 128 units, and the activation function to ReLU.

We trained the model on the eng.txt corpus, containing the complete works of Shakespeare (~1.1M characters), with the following hyperparameters:

- Learning rate: 0.003
- Epochs: 10
- Batch size: 256

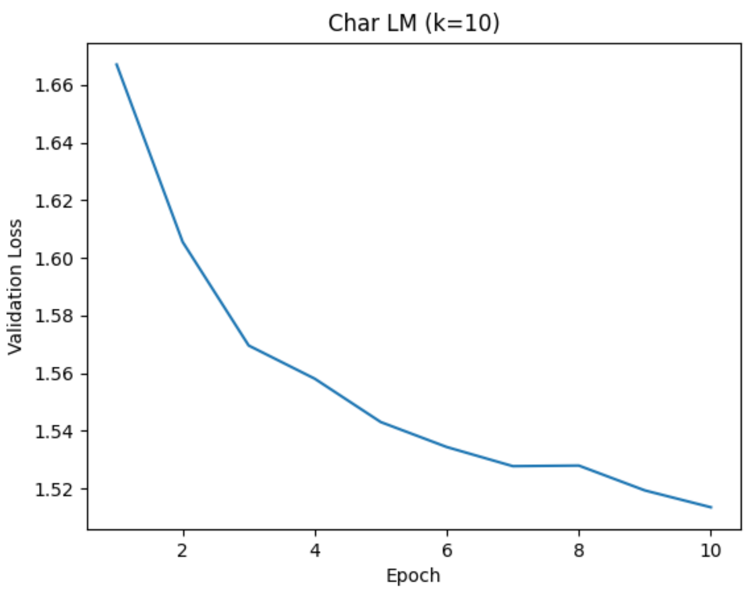
We experimented with different values of k to explore the effect of the context window size on model performance.

Window size k=10

First, we trained our character-level language on the same corpus with the same hyperparameters and a context window of **k = 10** characters.

Results

Validation set loss



As seen in the loss curve, the validation loss steadily decreased from 1.67 to 1.51 over 10 epochs. This suggests that the model learned useful representations of character-level sequences.

We also generated samples after each epoch (see appendix). Initially the model outputs gibberish, but from epoch 5 onward, we start seeing recognizable English words, Shakespearean names, and punctuation patterns. By epoch 10, the output is mostly syntactically plausible and stylistically close to the training corpus.

Sample after epoch 10:

The 'tis tears in hath so prephesprord, in them not the come to my wife himself dilious case thwardly co

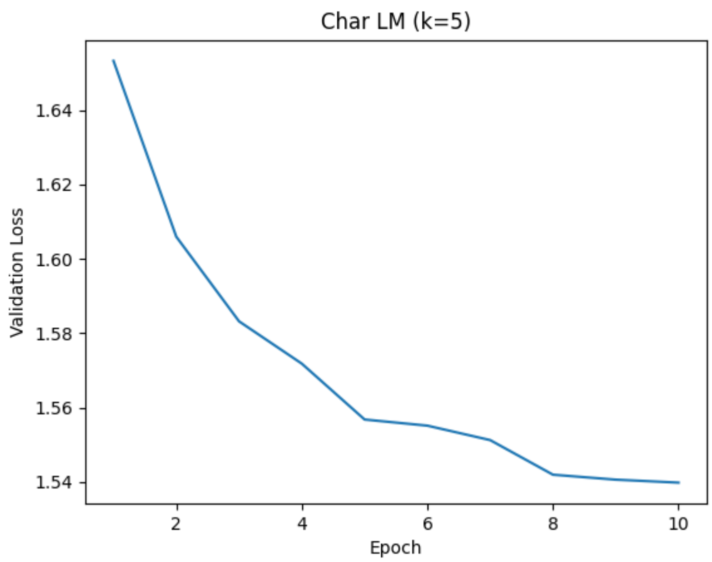
Experimenting with different window sizes

Window size k=5

We trained our character-level language on the same corpus with the same hyperparameters, but with a shorter context window of **k = 5** characters.

Results

Validation set loss



The validation loss improved from 1.6532 to 1.5398, showing stable and consistent learning. The rate of improvement was slightly slower compared to k=10, but the final loss was competitive.

The generated samples also showed progress. Early epochs produced mostly random strings, but by epoch 10, we see a stronger presence of real words, names, and recognizable dialogue patterns. However, sentences tended to be shorter and less coherent compared to those generated with k=10.

Sample after epoch 10:

The cad!

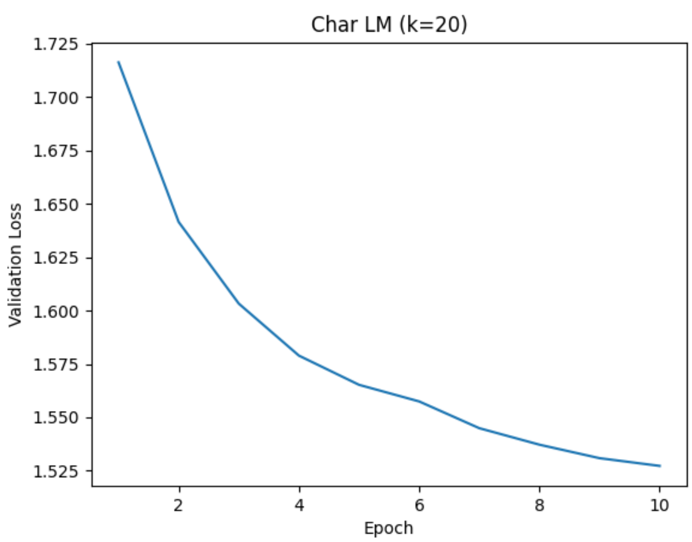
Clown: to yearn,
Mind sirraifferty to true,
And by say, and it
Thy resely; knew can are staon,

Window size k=20

To explore the effect of longer context, we trained the character-level language model with a window of **k = 20** characters.

Results

Validation set loss



The validation loss improved from **1.7162** to **1.5273**, showing the best final performance of the three configurations.
While the model started out with a higher loss than both k=5 and k=10, the longer context enabled it to converge to a better local minimum.

Generated text at epoch 10 appears more stylistically coherent, and the model often captures complex patterns such as speaker names (MERCUTIO, BIANCA, MENENIUS) and sentence structures with greater consistency. However, it also seems to overfit character patterns (e.g., repetitive colons or pseudo-names like CIRCAI:: and MIMIMIM:), which may suggest excessive reliance on memorization for longer k.

Sample after epoch 10:

The WISNI:wra her my hadower most truliged:
And here I passitious
like the now to thou so God, be mene s

Comparison of different context lengths (k)

k	Train Loss (Epoch 10)	Validation Loss (Epoch 10)	Text Quality	Observations
5	1.5522	1.5398	Basic, short, mostly valid words	Fast convergence, but limited context
10	1.5412	1.5134	Balanced, fluent, Shakespearean style	Best overall coherence and style
20	1.5655	1.5273	Complex, consistent, sometimes odd	Captures structure well, but tends to repeat

Summary and conclusion

Our experiments show that the context size k has a significant impact on both model performance and the quality of generated text.
A smaller value like k=5 enables fast training but provides only limited contextual understanding, leading to fragmented output.
The medium setting k=10 offered the best tradeoff: it produced coherent, stylistically appropriate text while maintaining stable convergence.
The longest context, k=20, led to the lowest validation loss, indicating deeper pattern learning. However, it also introduced overfitting artifacts such as pseudo-repetitive speaker names and less diversity.

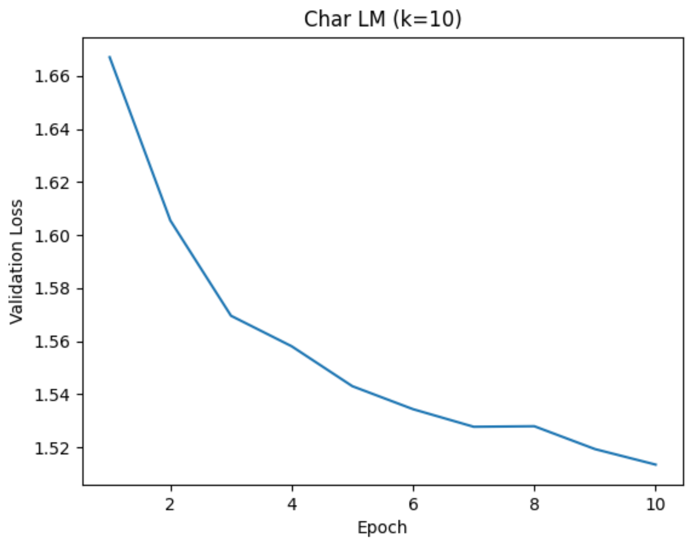
In conclusion, **k=10** appears to be the most effective choice, balancing structure, fluency, and generalization.

Sampling without prefix (k=10)

We also evaluated the model's generative ability without providing any initial context. Interestingly, the model was able to generate reasonably structured Shakespearean-style text even when starting from an empty prefix.
While the coherence was somewhat reduced, many of the outputs included character names (e.g., BENVOLIO) and plausible sentence openings.

For example: *POPSO: Of conter's so thee starrance?*
BENVOLIO: Welcome that throughs sundriver his taloucwertly co

Validation set loss



Appendix: 10 Sampled Texts from Best Model (k=10)

Below are 10 character-level text samples of 100 characters each, generated after each epoch from the best-performing model (k=10) with a prefix ("The "):

1. *The I by too got saw, Whoke griet's to my fror Madem, That afor thy with forchediust a become word mace* sample_k10_ep1
2. *The I be the so? VINCENTIO: This that I youngs. CARLINAL: Even lory soe with knife, shall edbunderfly* sample_k10_ep2
3. *The I bigat your pant, That hadis my dishallly forswill be how clo's chalm wir ant with down awainn's po* sample_k10_ep3
4. *The him old alms. What I status: This when court, But thou say,Lancay, Torbally word, indine the make in* sample_k10_ep4
5. *The I'll not thee, to that the indying. Which she hath the hath he be my thought our shall be beat a nob* sample_k10_ep5
6. *The 't from were uncle, Their that dengly these? ale luturin, Thrse! Mine aw no pon to man. And punnerat* sample_k10_ep6
7. *The I'll upon the peopine. LUCIO: Parry'd in thou hast take Carite trry, or her, dow Pursue against Whe* sample_k10_ep7
8. *The in thy fortuccess if brother, and sepfrecity tords: think help million soot in off foof the cravains* sample_k10_ep8
9. *The it us this dein-it: so Rome, Good newl, fin. And met it, but vitence, eto yourted of Trankly in und* sample_k10_ep9
10. *The 'tis tears in hath so prephesprord, in them not the come to my wife himself dilious case thwardly co* sample_k10_ep10