TASK

Title: Report on Implementing GPT Style Decoder and a Llama Style Decoder

The following are the main results and key findings:
Analysis (in different experiments):

GPT Style Decoder

Experiment: 01
Hyperparameters:

- batch_size = 16
- block_size = 32
- max_iter = 2400
- eval_interval = 100
- learning_rate = 1e-3
- eval_iters = 200
- n_embd = 64
- n_head = 4
- n_layer = 4
- dropout = 0.0

Results: Training loss is 1.8239, Validation loss is 1.9264.

```
step 0: train loss 4.4109, val loss 4.4016
step 100: train loss 2.6511, val loss 2.6604
step 200: train loss 2.4941, val loss 2.4902
step 300: train loss 2.3909, val loss 2.4026
step 400: train loss 2.3222, val loss 2.3292
step 500: train loss 2.2546, val loss 2.2712
step 600: train loss 2.1957, val loss 2.2057
step 700: train loss 2.1663, val loss 2.1891
step 800: train loss 2.1255, val loss 2.1531
step 900: train loss 2.0776, val loss 2.1118
step 1000: train loss 2.0562, val loss 2.0895
step 1100: train loss 2.0290, val loss 2.0836
step 1200: train loss 1.9990, val loss 2.0471
step 1300: train loss 1.9899, val loss 2.0369
step 1400: train loss 1.9541, val loss 2.0076
step 1500: train loss 1.9346, val loss 2.0005
step 1600: train loss 1.9225, val loss 2.0094
step 1700: train loss 1.9133, val loss 1.9922
step 1800: train loss 1.8801, val loss 1.9734
step 1900: train loss 1.8763, val loss 1.9607
step 2000: train loss 1.8529, val loss 1.9716
step 2100: train loss 1.8479, val loss 1.9635
step 2200: train loss 1.8353, val loss 1.9440
step 2300: train loss 1.8282, val loss 1.9341
step 2399: train loss 1.8239, val loss 1.9264

And the Rorincer: fatis be be madise,
And bute anvirtan the gatauss: butt he usquit to ber dilassaness.

HENRET:
Do a wize our towns, tof is heir tril;
And bute miree senvein latise, drove to hee
Will man is wabs!
All: now, all thuse cocking: tugh aiss hiw you lord.
In Boodeen at swer:
I lay like agleon,---on her piinss to the curive
And him his sout were truee kingn,
Tursef son; if his shat tey flepoats,
Madering my of trany sup! stied is:
Aronal the Egettlet in cour accanny iry to chan you!
An
```

Training Loss and Validation Loss:

- The loss values are decreasing over time, which is a good sign. It shows that my model is learning to minimize the difference between its predictions and the actual next characters in the text.
- The training loss starts at around 4.41 and decreases to 1.86, while the validation loss starts around 4.40 and decreases to about 1.96, indicating that the model is generalizing reasonably well on the validation set.

Generated Text Output:

- The generated text looks like semi-coherent "nonsense," but it also demonstrates that the model is beginning to grasp language structure. The capitalization, punctuation, and line breaks give it the appearance of text with dialogue or a script format, possibly because it is trained on a dataset with similar structures.
- Some words and structures resemble English words and sentence forms (like "And he thin," "Brown your," and "Platton"), which means the model is learning some common patterns and sequences.

Experiment: 02 Increase Model Capacity

Hyperparameters:

- batch_size = 16
- block_size = 32
- max_iter = 3000
- eval_interval = 100
- learning_rate = 3e-4
- eval_iters = 200
- n_embd = 128
- n_head = 8
- n_layer = 6
- dropout = 0.1

Results: Training loss is 1.7176, Validation loss is 1.8812 .

```
step 0: train loss 4.3963, val loss 4.3932
step 100: train loss 2.5887, val loss 2.5935
step 200: train loss 2.4515, val loss 2.4713
step 300: train loss 2.3805, val loss 2.3862
step 400: train loss 2.3178, val loss 2.3239
step 500: train loss 2.2437, val loss 2.2563
step 600: train loss 2.1837, val loss 2.1955
step 700: train loss 2.1422, val loss 2.1691
step 800: train loss 2.0985, val loss 2.1321
step 900: train loss 2.0601, val loss 2.1112
step 1000: train loss 2.0315, val loss 2.0854
step 1100: train loss 1.9992, val loss 2.0640
step 1200: train loss 1.9834, val loss 2.0281
step 1300: train loss 1.9541, val loss 2.0256
step 1400: train loss 1.9303, val loss 2.0043
step 1500: train loss 1.9099, val loss 1.9964
step 1600: train loss 1.8902, val loss 1.9763
step 1700: train loss 1.8749, val loss 1.9715
step 1800: train loss 1.8543, val loss 1.9525
step 1900: train loss 1.8451, val loss 1.9619
step 2000: train loss 1.8144, val loss 1.9420
step 2100: train loss 1.8237, val loss 1.9337
step 2200: train loss 1.7898, val loss 1.9136
step 2300: train loss 1.7937, val loss 1.9234
step 2400: train loss 1.7812, val loss 1.9172
step 2500: train loss 1.7739, val loss 1.9149
step 2600: train loss 1.7517, val loss 1.8786
step 2700: train loss 1.7438, val loss 1.8908
step 2800: train loss 1.7221, val loss 1.8646
step 2900: train loss 1.7347, val loss 1.8919
step 2999: train loss 1.7176, val loss 1.8812
```

```
Pron'serwithm.

CABLAMQUET:
Hear not, the liking a munes frell. Bot hornals:
Her will stal, for apple
And your hear rubtenter Provoal! I got stack your such pouch you,----
'CAPnows:
Be herind. I wisfinem on the jly,
And afreed of deserver Whase your now.

First AWARCHARD II:
GARISent, I he'e my coming well mots.

SICHAM:
And mester, thou worful?
For then witle ar, we ic! Blestroukesty upse stands, thou, and
Stil just he to pru-sents, weepanry full'er of will
Somealy feht porsiments tembe faul He
```

My model output shows improvement in training, as the loss steadily decreases, and the generated text has started forming more coherent words, names, and sentence-like structures. However, the output still lacks clarity and accurate sentence formation, which suggests there's room for improvement in quality.

Experiment 03: Higher Model Capacity with Larger Embedding

Hyperparameters:

- batch_size: 16
- block_size: 64
- max_iter: 3000
- eval_interval: 100
- learning_rate: 2e-4
- eval_iters: 200
- n_embd: 192
- n_head: 8
- n_layer: 8
- dropout: 0.2

Results: Training loss is 1.7816, Validation loss is 1.9126.

```
step 0: train loss 4.2941, val loss 4.2875
step 100: train loss 2.6035, val loss 2.6181
step 200: train loss 2.4986, val loss 2.5163
step 300: train loss 2.4566, val loss 2.4625
step 400: train loss 2.4119, val loss 2.4334
step 500: train loss 2.3625, val loss 2.3764
step 600: train loss 2.3143, val loss 2.3364
step 700: train loss 2.2827, val loss 2.2995
step 800: train loss 2.2377, val loss 2.2673
step 900: train loss 2.2066, val loss 2.2295
step 1000: train loss 2.1661, val loss 2.1932
step 1100: train loss 2.1320, val loss 2.1692
step 1200: train loss 2.0993, val loss 2.1386
step 1300: train loss 2.0717, val loss 2.1087
step 1400: train loss 2.0430, val loss 2.0970
step 1500: train loss 2.0225, val loss 2.0849
step 1600: train loss 1.9939, val loss 2.0623
step 1700: train loss 1.9777, val loss 2.0598
step 1800: train loss 1.9556, val loss 2.0316
step 1900: train loss 1.9410, val loss 2.0122
step 2000: train loss 1.9199, val loss 2.0011
step 2100: train loss 1.9089, val loss 1.9950
step 2200: train loss 1.8822, val loss 1.9702
step 2300: train loss 1.8663, val loss 1.9650
step 2400: train loss 1.8506, val loss 1.9559
step 2500: train loss 1.8339, val loss 1.9374
step 2600: train loss 1.8205, val loss 1.9370
step 2700: train loss 1.8124, val loss 1.9314
step 2800: train loss 1.7956, val loss 1.9218
step 2900: train loss 1.7884, val loss 1.9095
step 2999: train loss 1.7816, val loss 1.9126
```

```
Fir and witquen mand, my dore in to be.

LONCEONE:
Seate howle berigh spower'd thand may Couldio
And of with I queacs thrave hefuse fins,
Him is deging by faver the and my,
Thippon but mutimes.
Ramptenders the sto his him can; bett sichivest? I neyely fick,
And that I all plepeck ayous ware, may laight forly?
't feasor fatce the and word wan with dethe fairst
Allet not broads my the virn'd now now
Maturs oges thy these the flath the?
If day spease than the doon, thine farshe-
Mardeds Iscalongend
```

This result shows improvement both in loss values and in the generated text quality, with better coherence and some structured phrases. The modifications in n_embd, n_head, n_layer, and dropout appear to have positively impacted the model's ability to learn and generalize. However, while phrases are forming, there’s still room for improvement to enhance fluency and coherence in the generated output.

Experiment 04: Fine-Tune Model with Larger Context and Reduced Learning Rate

Increasing the context (block size) and slightly reducing the learning rate can help the model focus on more extended character dependencies, potentially improving sentence flow.

Hyperparameters:

- batch_size: 16
- block_size: 128
- max_iter: 3000
- eval_interval: 100
- learning_rate: 1e-4
- eval_iters: 200
- n_embd: 192
- n_head: 8
- n_layer: 8
- dropout: 0.2

Results: Training loss is 2.0035, Validation loss is 2.0670 .

```
step 0: train loss 4.3446, val loss 4.3430
step 100: train loss 2.7515, val loss 2.7823
step 200: train loss 2.5824, val loss 2.6027
step 300: train loss 2.5346, val loss 2.5378
step 400: train loss 2.5025, val loss 2.5099
step 500: train loss 2.4787, val loss 2.4939
step 600: train loss 2.4639, val loss 2.4750
step 700: train loss 2.4395, val loss 2.4546
step 800: train loss 2.4258, val loss 2.4425
step 900: train loss 2.4129, val loss 2.4210
step 1000: train loss 2.3948, val loss 2.4025
step 1100: train loss 2.3692, val loss 2.3862
step 1200: train loss 2.3520, val loss 2.3640
step 1300: train loss 2.3300, val loss 2.3407
step 1400: train loss 2.3074, val loss 2.3278
step 1500: train loss 2.2780, val loss 2.2978
step 1600: train loss 2.2607, val loss 2.2758
step 1700: train loss 2.2362, val loss 2.2636
step 1800: train loss 2.2084, val loss 2.2359
step 1900: train loss 2.1868, val loss 2.2114
step 2000: train loss 2.1717, val loss 2.1989
step 2100: train loss 2.1489, val loss 2.1802
step 2200: train loss 2.1304, val loss 2.1679
step 2300: train loss 2.1168, val loss 2.1543
step 2400: train loss 2.0931, val loss 2.1340
step 2500: train loss 2.0796, val loss 2.1223
step 2600: train loss 2.0610, val loss 2.1061
step 2700: train loss 2.0541, val loss 2.1028
step 2800: train loss 2.0376, val loss 2.0844
step 2900: train loss 2.0253, val loss 2.0822
step 2999: train loss 2.0035, val loss 2.0670
```

```
Ye cous. to And to thamesd wowave--
The blorede eart you hofffoust.

LONCESARBENNIO:
PAs wiceman the o'lds, You thas?
Your aniou you dant fep yor heldtan,
Fo that besed pof heiresst a thoming shes ranew
Ontrd a bre den, sevellongecy's, to that the es
Teis, Starm; se wave'd the mands.
He thas sof sutell kipong gastioyste,
Frskeat! hoou selo me,
To ut beeving thy stresetot.

INIUS:
It weld you by ouse roy thar gespplouety,
And or ase I with ofeld trabjurs the theld,
And soud nat sa wkilcht to mild
```

Experiment 05: Increase Embedding Dimensionality and Add Dropout for Generalization

Trying a higher embedding dimensionality for a more expressive model, which may capture language nuances better. Also, adding a small amount of additional dropout may help avoid overfitting.

Hyperparameters:

- batch_size: 16
- block_size: 64
- max_iter: 3000
- eval_interval: 100
- learning_rate: 2e-4
- eval_iters: 200
- n_embd: 256
- n_head: 8
- n_layer: 8
- dropout: 0.3

Results: Training loss is 1.7655, Validation loss is 1.9117 .

```
step 0: train loss 4.2938, val loss 4.2932
step 100: train loss 2.5669, val loss 2.5713
step 200: train loss 2.4883, val loss 2.4926
step 300: train loss 2.4418, val loss 2.4583
step 400: train loss 2.3871, val loss 2.4023
step 500: train loss 2.3434, val loss 2.3476
step 600: train loss 2.2852, val loss 2.2994
step 700: train loss 2.2435, val loss 2.2482
step 800: train loss 2.2027, val loss 2.2117
step 900: train loss 2.1575, val loss 2.1838
step 1000: train loss 2.1360, val loss 2.1590
step 1100: train loss 2.0966, val loss 2.1417
step 1200: train loss 2.0710, val loss 2.1150
step 1300: train loss 2.0511, val loss 2.0975
step 1400: train loss 2.0306, val loss 2.0719
step 1500: train loss 2.0026, val loss 2.0622
step 1600: train loss 1.9719, val loss 2.0395
step 1700: train loss 1.9553, val loss 2.0406
step 1800: train loss 1.9399, val loss 2.0249
step 1900: train loss 1.9257, val loss 2.0078
step 2000: train loss 1.9004, val loss 1.9942
step 2100: train loss 1.8793, val loss 1.9728
step 2200: train loss 1.8733, val loss 1.9784
step 2300: train loss 1.8498, val loss 1.9653
step 2400: train loss 1.8381, val loss 1.9512
step 2500: train loss 1.8333, val loss 1.9424
step 2600: train loss 1.8125, val loss 1.9411
step 2700: train loss 1.8043, val loss 1.9389
step 2800: train loss 1.8015, val loss 1.9251
step 2900: train loss 1.7840, val loss 1.9195
step 2999: train loss 1.7655, val loss 1.9117
```

```
Fir bod with gac and was to you cour.

Lord CiONET:
Shae howle beyie,
Augher'd thank not Courciof expocess,
Und with ther dume frath, with i' is this a by fagrourt.
Which 'That I nobloom?

EDQUEENTIZET:
Grath conoou he inrown; by there it that to his docche
Letang neld torpe, able you might
Moust al ther clent'ers; son fat comen, made and an with deth if both
Alds, I sel, apone what rings go dince
toon, loges thy these the rawth the?

Fidsp spease than thought of mine; aws the and is hearth.

KI
```

Experiment 06: Increase Number of Layers for a Deeper Network

Adding more layers allows the model to learn more complex representations but will increase computation time. This could improve the quality of the text generation, but it’s a more computationally intensive change.

Hyperparameters:

- batch_size: 16
- block_size: 64
- max_iter: 3000
- eval_interval: 100
- learning_rate: 2e-4
- eval_iters: 200
- n_embd: 192
- n_head: 8
- n_layer: 10
- dropout: 0.2

Results: Training loss is 2.1407, Validation loss is 2.1618 .

```
step 0: train loss 4.2437, val loss 4.2459
step 100: train loss 2.5937, val loss 2.5986
step 200: train loss 2.5038, val loss 2.5132
step 300: train loss 2.4569, val loss 2.4718
step 400: train loss 2.4090, val loss 2.4142
step 500: train loss 2.3588, val loss 2.3789
step 600: train loss 2.3171, val loss 2.3341
step 700: train loss 2.2741, val loss 2.2958
step 800: train loss 2.2474, val loss 2.2613
step 900: train loss 2.2014, val loss 2.2332
step 1000: train loss 2.1602, val loss 2.1983
step 1100: train loss 2.1318, val loss 2.1614
step 1200: train loss 2.0822, val loss 2.1294
step 1300: train loss 2.0578, val loss 2.1201
step 1400: train loss 2.0330, val loss 2.0796
step 1500: train loss 2.0030, val loss 2.0690
step 1600: train loss 1.9793, val loss 2.0507
step 1700: train loss 1.9713, val loss 2.0467
step 1800: train loss 1.9424, val loss 2.0276
step 1900: train loss 1.9202, val loss 1.9878
step 2000: train loss 1.9038, val loss 1.9920
step 2100: train loss 1.8792, val loss 1.9715
step 2200: train loss 1.8639, val loss 1.9633
step 2300: train loss 1.8483, val loss 1.9498
step 2400: train loss 1.8321, val loss 1.9403
step 2500: train loss 1.8144, val loss 1.9359
step 2600: train loss 1.7941, val loss 1.9330
step 2700: train loss 1.7913, val loss 1.9142
step 2800: train loss 1.7760, val loss 1.9110
step 2900: train loss 1.7621, val loss 1.9049
step 2999: train loss 1.7516, val loss 1.9024
```

```
FORIO:
Let hash let how shoulds, us them mabm reyer.


FICLIO:
I beave, bat the look!'
I' to a me: is not an the what hise more,
His you thad farl Nowstor
hach go foraily desid on; onse in the keings,
Shain whas houne lemous are word in it Wong may;
the linly lave; tuo pry, yor throus come.

Do, land
Fed GROMEENS:
Yer light whicith shy lifesy wick my boy,
In yiur it. are the copung, in to dited'strrep now sir?
Abrocicef, and I'll Swer, bysbluck yefu dain ooe than's:
Thish subalt of dons the her
```

Experiment 07: Slower Learning with Larger Context Window

Hyperparameters:

- batch_size: 16
- block_size: 64
- max_iter: 3000
- eval_interval: 100
- learning_rate: 1e-4
- eval_iters: 200
- n_embd: 128
- n_head: 8
- n_layer: 6
- dropout: 0.1

Results: Training loss is 2.1407, Validation loss is 2.1618 .

```
step 0: train loss 4.3281, val loss 4.3312
step 100: train loss 3.0152, val loss 3.0490
step 200: train loss 2.7513, val loss 2.7605
step 300: train loss 2.6270, val loss 2.6273
step 400: train loss 2.5705, val loss 2.5707
step 500: train loss 2.5347, val loss 2.5365
step 600: train loss 2.5043, val loss 2.5088
step 700: train loss 2.4825, val loss 2.4878
step 800: train loss 2.4673, val loss 2.4763
step 900: train loss 2.4453, val loss 2.4465
step 1000: train loss 2.4215, val loss 2.4353
step 1100: train loss 2.4077, val loss 2.4176
step 1200: train loss 2.3905, val loss 2.3982
step 1300: train loss 2.3684, val loss 2.3871
step 1400: train loss 2.3520, val loss 2.3704
step 1500: train loss 2.3362, val loss 2.3443
step 1600: train loss 2.3277, val loss 2.3342
step 1700: train loss 2.3105, val loss 2.3195
step 1800: train loss 2.2979, val loss 2.2980
step 1900: train loss 2.2781, val loss 2.2918
step 2000: train loss 2.2652, val loss 2.2801
step 2100: train loss 2.2514, val loss 2.2660
step 2200: train loss 2.2367, val loss 2.2491
step 2300: train loss 2.2305, val loss 2.2458
step 2400: train loss 2.2097, val loss 2.2334
step 2500: train loss 2.2010, val loss 2.2221
step 2600: train loss 2.1838, val loss 2.2117
step 2700: train loss 2.1804, val loss 2.2036
step 2800: train loss 2.1539, val loss 2.1902
step 2900: train loss 2.1605, val loss 2.1781
step 2999: train loss 2.1407, val loss 2.1618


PE sppeow't mot Ritqunegutst,
Tind ond do thend gra seet inglee oos hor fuprdved rad!

HESThy 'XENA:

Sherraive wer you douer Pro bar! rigntare cr sour!

IS pourvey anl:
I my hales the harind. be Fffirempond he jlr,

Norefren rost the arat has ses inn wear! hars Ware stum's:
Geesten foo ea'd my thice Welle mon anot ditoitarat.
SCeinest ind sfat? to may le shlde k,

Naichinges mourd, yer,se st Mas hengant Heart.
I'll jue towere--e.

CULOAES:
So, Glderiome beirk meald fe youck im cour;
Aco fanLins
```

Llama Style Decoder

Experiment 01:

Hyperparameters:

- batch_size: 16
- block_size: 64
- max_iter: 3000
- eval_interval: 100
- learning_rate: 2e-4
- eval_iters: 200
- n_embd: 192
- n_head: 8
- n_layer: 10
- dropout: 0.2

Results:

```
step 0: train loss 10.3241, val loss 10.3344
step 100: train loss 7.7450, val loss 8.0207
step 200: train loss 7.4654, val loss 7.8877
step 300: train loss 7.2028, val loss 7.7819
step 400: train loss 6.9682, val loss 7.6656
step 500: train loss 6.7542, val loss 7.5940
step 600: train loss 6.5402, val loss 7.5363
step 700: train loss 6.3520, val loss 7.5259
step 800: train loss 6.1728, val loss 7.4876
step 900: train loss 5.9959, val loss 7.5022
step 1000: train loss 5.8495, val loss 7.4933
step 1100: train loss 5.7001, val loss 7.4993
step 1200: train loss 5.5609, val loss 7.5645
step 1300: train loss 5.3692, val loss 7.5645
step 1400: train loss 5.2422, val loss 7.5767
step 1500: train loss 5.1015, val loss 7.6592
step 1600: train loss 4.9598, val loss 7.6569
step 1700: train loss 4.8317, val loss 7.7079
step 1800: train loss 4.6948, val loss 7.7447
```

```
step 1900: train loss 4.5640, val loss 7.8181
step 2000: train loss 4.4379, val loss 7.8394
step 2100: train loss 4.3071, val loss 7.8741
step 2200: train loss 4.2115, val loss 7.9302
step 2300: train loss 4.0819, val loss 8.0245
step 2400: train loss 3.9612, val loss 8.0422
step 2500: train loss 3.8581, val loss 8.1314
step 2600: train loss 3.7508, val loss 8.1392
step 2700: train loss 3.6224, val loss 8.2175
step 2800: train loss 3.5244, val loss 8.2788
step 2900: train loss 3.4505, val loss 8.3081
step 2999: train loss 3.3378, val loss 8.3966
```

```
Decoded Tokens: ['First</w>', 'Citizen:</w>', 'Give</w>', 'you</w>', 'put</w>', 'Condemn</w>', 'the</w>', 'Second
</w>', 'Servant:</w>', 'progress</w>', 'people</w>', 'have</w>', 'aloud,</w>', 'shepherd!</w>', 'But</w>', 'the</
w>', 'drop</w>', 'Of</w>', 'one</w>', 'whose</w>', 'heavy</w>', 'circled</w>', 'orb,</w>', "I'll</w>", 'forget,</
w>', 'What</w>', 'wept</w>', 'cried</w>', 'wicked?</w>', 'is</w>', 'sending</w>', 'him,</w>', 'Or</w>', "'Come</w
>", 'Clifford,</w>', 'take</w>', 'up</w>', 'to</w>', 'the</w>', 'foe,</w>', 'And</w>', 'mighty</w>', 'liege;</w
>', 'till</w>', 'the</w>', 'charge</w>', 'against</w>', 'the</w>', 'rascal</w>', 'with</w>', 'oak.</w>', 'bawd</w
>', 'way</w>', 'then</w>', 'it</w>', 'is</w>', 'here!</w>', 'GRUMIO:</w>', 'So</w>', "think'st</w>", 'thou</w>',
'didst</w>', 'love</w>', 'thy</w>', 'chair,</w>', 'Were</w>', 'but</w>', 'a</w>', 'bloody</w>', 'and</w>', 'not</
w>', 'jewel</w>', 'in</w>', 'my</w>', 'cool</w>', 'Thus</w>', 'upon</w>', 'thy</w>', 'shall</w>', 'be</w>', 'fir
m.</w>', 'Or</w>', 'cold:</w>', 'Sends</w>', 'me</w>', "I'll</w>", 'instruct</w>', 'thee</w>', 'straight.</w>',
'Come</w>', 'while</w>', 'I</w>', 'told</w>', 'me</w>', 'that</w>', 'thou</w>', 'varlet;</w>', 'thou</w>', 'dost
</w>', 'show</w>', 'me</w>', 'from</w>', 'thy</w>', 'tough</w>', 'That</w>', 'I</w>', 'see</w>', 'before</w>', 't
hy</w>', 'devotion</w>', 'and</w>', 'effeminate</w>', 'signor</w>', 'mio</w>', 'off</w>', 'that</w>', 'thou</w>',
"behold'st!</w>", 'O,</w>', 'it</w>', 'now.</w>', 'O,</w>', 'what</w>', 'thou</w>', 'fishified!</w>', 'But</w>',
'we</w>', 'shall</w>', 'hear</w>', 'themselves</w>', 'shall</w>', 'adventure</w>', 'heart</w>', 'to</w>', 'boot!
</w>', 'no</w>', 'other</w>', 'again</w>', 'of</w>', 'renown,</w>', 'Not</w>', 'yet</w>', 'thy</w>', "king's,</w
>", 'To</w>', 'do</w>', 'to</w>', 'thy</w>', 'centre</w>', 'out.</w>', 'BENVOLIO:</w>', 'This</w>', 'deadly</w>',
'bon</w>', 'jour!</w>', "'tis</w>", 'surely</w>', 'but</w>', 'with</w>', 'thy</w>', 'general</w>', 'dispatch:</w
>', 'the</w>', 'all-hail</w>', 'to</w>', 'thee,</w>', 'gentle</w>', 'irks</w>', 'my</w>', 'young</w>', 'prince</w
```

```
'all</w>', 'I</w>', 'now</w>', 'fie,</w>', 'fie!</w>', 'on,</w>', 'And</w>', 'I</w>', 'would</w>', 'get</w>', 'th
ee</w>', 'straight.</w>', 'peace,</w>', 'Your</w>', 'grace,</w>', 'gentle</w>', 'keeper,</w>', 'please</w>', 'you
</w>', 'to</w>', 'this</w>', 'covenant</w>', 'makes,</w>', 'my</w>', 'uncle</w>', 'York</w>', 'Hath</w>', 'not</w
>', 'his</w>', 'grace</w>', 'old,</w>', 'still</w>', 'live</w>', 'and</w>', 'old</w>', 'boy!</w>', "We'll</w>",
'not</w>', 'what</w>', 'purpose,</w>', 'them</w>', 'not.</w>', 'HERMIONE:</w>', 'liege.</w>', "'gainst</w>", 'me,
</w>', 'that</w>', 'I</w>', 'do</w>', 'I</w>', 'fear</w>', 'officer!</w>', 'To</w>', 'tell</w>', 'my</w>', 'very
</w>', 'justice:</w>', 'then,</w>', 'being</w>', 'thus</w>', 'thriftless</w>', 'come,</w>', 'again;</w>', 'I</w
>', 'alligator</w>', 'forgot</w>', 'day-bed,</w>', 'Thou</w>', 'wouldst</w>', 'seek</w>', 'to</w>', 'mills--bring
</w>', 'me</w>', 'doth</w>', 'Wednesday</w>', 'the</w>', 'king,</w>', 'and</w>', 'you,</w>', 'good</w>', 'sister
s;</w>', 'the</w>', 'pennyworth</w>', 'of</w>', 'your</w>', 'own,</w>', 'nor</w>', 'any</w>', 'thing</w>', 'so</w
>', 'muddy,</w>', 'so</w>', 'God</w>', 'With</w>', 'harsh</w>', 'traitor,</w>', 'throne</w>', 'a</w>', 'brain</w
>', 'the</w>', 'doubt</w>', 'of</w>', 'a</w>', 'friend,</w>', 'great</w>', "sear'd</w>", 'bless</w>', 'a</w>', 'f
resh.</w>', 'O</w>', 'look</w>', 'out</w>', 'most,</w>', 'a</w>', 'moiety</w>', 'of</w>', 'open</w>', 'words;</w
>', 'of</w>', 'sugar,</w>', 'here,</w>', 'By</w>', 'the</w>', 'ground:</w>', 'when</w>', 'thou</w>', 'seest,</w
>', 'thou</w>', 'art</w>', 'laid</w>', 'in</w>', 'my</w>', 'seat</w>', 'Thou</w>', 'didst</w>', 'kill</w>', 'a</w
>', 'space,</w>', 'my</w>', 'house,</w>', 'As</w>', 'time</w>', 'have</w>', 'had</w>', 'a</w>', "farmer's</w>",
'inventions</w>', 'jar</w>', 'By</w>', 'some</w>', 'any</w>', 'thing</w>', 'possible.</w>', 'into</w>', 'much</w
>', 'lenity</w>', 'And</w>', 'so</w>', 'sweet</w>', 'Jerusalem.</w>', "I'll</w>", 'make</w>']
```

```
First Citizen: Give you put Condemn the Second Servant: progress people have aloud, shepherd! But the drop Of one
whose heavy circled orb, I'll forget, What wept cried wicked? is sending him, Or 'Come Clifford, take up to the f
oe, And mighty liege; till the charge against the rascal with oak. bawd way then it is here! GRUMIO: So think'st
thou didst love thy chair, Were but a bloody and not jewel in my cool Thus upon thy shall be firm. Or cold: Sends
me I'll instruct thee straight. Come while I told me that thou varlet; thou dost show me from thy tough That I se
e before thy devotion and effeminate signor mio off that thou behold'st! O, it now. O, what thou fishified! But w
e shall hear themselves shall adventure heart to boot! no other again of renown, Not yet thy king's, To do to thy
centre out. BENVOLIO: This deadly bon jour! 'tis surely but with thy general dispatch: the all-hail to thee, gent
le irks my young prince and blunt nets; but sad things; After the hearing; Third Conspirator: But, gentle sir, n
e'er well. CLARENCE: Came Or what suit, If thou art been absent: 'tis better good more. CLEOMENES: O, I will have
CAPULET: So unbuckle. obedience to go in title, Which since these vain are Richard, am I love. JULIET: Then, a wo
rd, my offence? GREMIO: I thank thee, gentle Socrates' it. CAMILLO: Elbow? marshal and Montague, you'll do you sw
eeten with that, I eyed than a soldier's mistress nurses! Wolves and by mine arm, And quickly Unmeritable dry:
I'll join mine arm, Be more. Adieu, sooth! I could good heart access? every thing calls for ever; open't. O misch
ief, yoke with briers, deed. O, rebels show thee to thy integrity, us, full of feasting Hydra here these; And Pom
pey, captain: And thou hast up parliament be entombed in person. My queen: Struck me, then, 'tis time and by. O,
spare her hunger of all my soul, My conscience weed my manors that wrong. Would all I now fie, fie! on, And I wou
ld get thee straight. peace, Your grace, gentle keeper, please you to this covenant makes, my uncle York Hath not
his grace old, still live and old boy! We'll not what purpose, them not. HERMIONE: liege. 'gainst me, that I do I
```


I'll join mine arm, Be more. Adieu, sooth! I could good heart access? every thing calls for ever; open't. O misch
ief, yoke with briers, deed. O, rebels show thee to thy integrity, us, full of feasting Hydra here these; And Pom
pey, captain: And thou hast up parliament be entombed in person. My queen: Struck me, then, 'tis time and by. O,
spare her hunger of all my soul, My conscience weed my manors that wrong. Would all I now fie, fie! on, And I wou
ld get thee straight. peace, Your grace, gentle keeper, please you to this covenant makes, my uncle York Hath not
his grace old, still live and old boy! We'll not what purpose, them not. HERMIONE: liege. 'gainst me, that I do I


```
fear officer! To tell my very justice: then, being thus thriftless come, again; I alligator forgot day-bed, Thou
wouldst seek to mills--bring me doth Wednesday the king, and you, good sisters; the pennyworth of your own, nor a
ny thing so muddy, so God With harsh traitor, throne a brain the doubt of a friend, great sear'd bless a fresh. O
look out most, a moiety of open words; of sugar, here, By the ground: when thou seest, thou art laid in my seat T
hou didst kill a space, my house, As time have had a farmer's inventions jar By some any thing possible. into muc
h lenity And so sweet Jerusalem. I'll make
```

The results show that my model is learning effectively at the beginning, as the training and validation losses both decrease during the initial steps. However, the validation loss eventually starts to plateau around step 1000, then gradually increases after step 1200. This indicates that the model is beginning to overfit, where it continues to improve on the training set but no longer generalizes well to the validation set.

Experiment: 02
Hyperparameters:
- batch_size = 16
- block_size = 32
- max_iter = 800
- eval_interval = 100
- learning_rate = 1e-3
- eval_iters = 200
- n_embd = 64
- n_head = 4
- n_layer = 4
- dropout = 0.0

Results:
step 0: train loss 7.5341, val loss 7.9548
step 100: train loss 7.0626, val loss 7.7297
step 200: train loss 6.7031, val loss 7.6555
step 300: train loss 6.4582, val loss 7.6714
step 400: train loss 6.2549, val loss 7.6209
step 500: train loss 6.0487, val loss 7.6766
step 600: train loss 5.8603, val loss 7.7689
step 700: train loss 5.6761, val loss 7.7569
step 800: train loss 5.4860, val loss 7.8688

My new setup of hyperparameters is already making progress in controlling overfitting, as the validation loss isn't diverging from the training loss as sharply as before. But still, it is increasing.

Experiment: 03
Hyperparameters:
- batch_size = 16
- block_size = 48
- max_iter = 2400
- eval_interval = 100
- learning_rate = 1e-4
- eval_iters = 200
- n_embd = 64
- n_head = 4

- n_layer = 4
- dropout = 0.1

Results:

```
step 0: train loss 5.3608, val loss 7.9489
step 100: train loss 5.2239, val loss 7.7540
step 200: train loss 5.1538, val loss 7.7026
step 300: train loss 5.0449, val loss 7.7456
step 400: train loss 4.9906, val loss 7.7112
step 500: train loss 4.9235, val loss 7.7283
step 600: train loss 4.8616, val loss 7.7455
step 700: train loss 4.7606, val loss 7.7737
step 800: train loss 4.7203, val loss 7.7626
step 900: train loss 4.6395, val loss 7.8256
step 1000: train loss 4.5782, val loss 7.8288
step 1100: train loss 4.5121, val loss 7.8291
step 1200: train loss 4.4486, val loss 7.8493
step 1300: train loss 4.4058, val loss 7.8418
step 1400: train loss 4.3672, val loss 7.8581
step 1500: train loss 4.3033, val loss 7.9191
step 1600: train loss 4.2501, val loss 7.9308
step 1700: train loss 4.1912, val loss 7.9613
step 1800: train loss 4.1634, val loss 7.9588
step 1900: train loss 4.1073, val loss 7.9960
step 2000: train loss 4.0748, val loss 8.0375
step 2100: train loss 3.9927, val loss 8.0269
step 2200: train loss 3.9575, val loss 8.0758
step 2300: train loss 3.9279, val loss 8.0593
step 2399: train loss 3.8829, val loss 8.1275
```

```
Decoded Tokens: ['First</w>', 'Lord:</w>', 'Good</w>', 'lords,</w>', 'unswayable</w>', 'and</w>', 'loving</w>', 'wh
ip</w>', 'your</w>', 'valour,</w>', 'In</w>', 'special</w>', 'mean,</w>', 'that</w>', 'this</w>', 'friar</w>', 'is
</w>', 'alive.</w>', 'DUKE</w>', 'VINCENTIO:</w>', 'It</w>', 'is,</w>', 'sir,</w>', 'I</w>', 'live?</w>', 'HERMION
E:</w>', 'Nay,</w>', 'should</w>', 'be</w>', 'your</w>', 'fresh</w>', 'deniest</w>', 'the</w>', 'sheer</w>', 'ale,
</w>', 'score</w>', 'me</w>', 'in</w>', 'me</w>', 'The</w>', 'wrong</w>', 'is</w>', 'the</w>', 'Fourth,</w>', 'him,
</w>', 'and</w>', 'call</w>', 'some,</w>', 'be</w>', 'seen</w>', 'place,</w>', 'this</w>', 'quarrel.</w>', 'Office
r:</w>', 'Nor</w>', "is't</w>", 'your</w>', 'impediment.</w>', 'for</w>', 'your</w>', 'sister.</w>', 'LUCIO:</w>',
'How</w>', 'prettily</w>', 'you</w>', 'might</w>', 'have</w>', "stain'd</w>", 'our</w>', 'carnations</w>', 'of</w
>', 'us:</w>', 'The</w>', 'heavens</w>', 'do</w>', 'thus--</w>', 'too.</w>', 'Cheerly,</w>', 'boys;</w>', 'look</w
>', 'unto</w>', "God's</w>", 'name,</w>', 'When</w>', 'we</w>', 'toward</w>', 'our</w>', 'uncle</w>', 'Gloucester,
</w>', 'or</w>', 'live,</w>', 'to</w>', 'have</w>', 'stars!</w>', 'And</w>', 'am</w>', 'nothing</w>', 'more</w>',
'to</w>', 'his</w>', "banish'd</w>", 'years</w>', 'old.</w>', 'DORCAS:</w>', 'Is</w>', 'here</w>', 'wrapt</w>', 'th
e</w>', 'field</w>', 'With</w>', 'mine</w>', "kingdom's</w>", 'boiling?</w>', 'In</w>', 'deep</w>', 'ear;</w>', 'Be
auty</w>', 'too</w>', 'soon</w>', 'won</w>', 'the</w>', 'dead</w>', 'inhabit,</w>', 'and</w>', 'this</w>', 'higher
</w>', 'about</w>', 'it</w>', 'stands</w>', 'and</w>', 'bred;</w>', 'one</w>', 'would</w>', 'bewray</w>', 'her</w
>', 'faults?</w>', 'ornaments,</w>', 'and</w>', 'thus;</w>', 'BRAKENBURY:</w>', 'Why</w>', 'shall</w>', 'you</w>',
'walk</w>', 'innocency,</w>', 'to</w>', 'find</w>', 'mine</w>', 'own,</w>', 'you</w>', 'well,</w>', 'good</w>', 'to
</w>', 'die.</w>', 'I</w>', 'will</w>', 'hear</w>', 'what</w>', 'a</w>', 'due</w>', 'of</w>', 'retire.</w>', 'That
</w>', 'it</w>', 'should</w>', 'they</w>', 'are</w>', "stopp'd,</w>", 'peace</w>', 'with</w>', 'all</w>', 'our</w
>', 'holy</w>', 'lives</w>', 'known</w>', 'before.</w>', 'ESCALUS:</w>', 'How!</w>', 'rid</w>', 'your</w>', 'doing,
</w>', 'He</w>', 'that</w>', 'best</w>', 'instruct</w>', 'you,</w>', 'on</w>', 'him!</w>', 'SICINIUS:</w>', 'Marry,
</w>', 'they</w>', 'are</w>', 'admiration:</w>', 'they</w>', 'are.</w>', 'AUTOLYCUS:</w>', 'O</w>', 'day!</w>', 'Mo
st</w>', 'piteous</w>', 'massacre</w>', 'of</w>', 'you.</w>', 'COMINIUS:</w>', 'unspeakable</w>', "other's</w>", 'g
rave;</w>', 'Or</w>', 'Daphne</w>', 'roaming</w>', 'through</w>', 'the</w>', 'sun:</w>', 'we</w>', 'march</w>', 'Up
on</w>', 'the</w>', 'fair</w>', 'stand.</w>', 'act</w>', 'of</w>', 'our</w>', 'medicine</w>', 'Gentleman:</w>', 'S
o.</w>', 'ANGELO:</w>', 'Well,</w>', 'sir.</w>', 'ABRAHAM:</w>', 'Do</w>', 'you</w>', 'three-pence</w>', 'up</w>',
'your</w>', 'majesty,</w>', 'Which</w>', 'our</w>', 'brother</w>', 'die,</w>', 'be</w>', 'he</w>', 'affections</w
>', 'and</w>', 'make</w>', 'too.</w>', 'Cheerly,</w>', 'boys;</w>', 'come.</w>', 'Here</w>', 'in</w>', 'your</w>',
'ignorant</w>', 'silken,</w>', 'sly,</w>', 'butchers!</w>', 'if</w>', 'men,</w>', 'How</w>', 'came</w>', 'from</w
>', 'this</w>', 'friar</w>', 'with</w>', 'you.</w>', 'Be</w>', "heap'd</w>", 'like</w>', 'the</w>', 'world</w>', 'g
oes</w>', "Agamemnon's</w>", 'kiss</w>', 'the</w>', 'sound:</w>', 'Art</w>', 'thou</w>', 'art</w>', "neighbours'</w
```

>", 'do;</w>', 'for</w>', 'his</w>', 'sake</w>', 'could</w>', 'never</w>', 'matter</w>', 'We</w>', 'two</w>', 'and
</w>', 'born,</w>', 'As</w>', 'force</w>', 'should</w>', 'be</w>', 'thought</w>', 'long</w>', 'with</w>', 'more</w
>', 'Than</w>', 'our</w>', 'coast</w>', 'And</w>', 'seem</w>', 'friendly,</w>', 'we</w>', 'are</w>', 'now</w>', 'to
</w>', 'give</w>', 'him</w>', 'die</w>', 'in</w>', 'tears.</w>', 'MERCUTIO:</w>', 'If</w>', 'I</w>', 'do</w>', 'com
plain</w>', 'on.</w>', 'is</w>', 'dead,</w>', 'I</w>', 'do</w>', 'with</w>', 'winds</w>', 'at</w>', 'the</w>', 'on?
</w>', 'When</w>', 'rebukes,</w>', 'knighthood</w>', 'that</w>', 'let</w>', 'us</w>', 'bear</w>', 'the</w>', 'jewel
</w>', 'in</w>', 'the</w>', 'rest;</w>', 'For</w>', 'their</w>', 'purses,</w>', 'and</w>', 'from</w>', 'you</w>',
'must</w>', 'not</w>', 'stay</w>', 'awhile,</w>', 'The</w>', 'household</w>', 'sick,</w>', 'Buckingham,</w>', 'sway
s</w>', 'him</w>', 'deal</w>', 'print</w>', 'who</w>', 'rather</w>', 'Lammas-tide?</w>', 'MERCUTIO:</w>', 'This</w
>', 'day</w>', 'read</w>', 'it</w>', 'with</w>', 'two</w>', 'days</w>', 'men</w>', 'in</w>', 'authority,</w>', 'dry
beat</w>', 'a</w>', 'brave</w>', 'shame,--</w>', 'Camillo,</w>', 'Whom</w>', 'we,</w>', 'we</w>', 'have</w>', 'we</
w>', 'will</w>', 'in</w>', 'bearing</w>', 'boughs</w>', 'Here</w>', 'is</w>', 'yet</w>', 'made</w>', 'to</w>', 'bid
</w>', 'the</w>', 'same;</w>', 'and</w>', 'nature</w>', 'Will</w>', 'torments</w>', 'to</w>', 'her</w>', 'withdrawn
</w>', 'My</w>', 'lord,</w>', 'and</w>', "'fore</w>", 'Guildhall</w>', 'FRIAR</w>', 'JOHN:</w>', 'wither</w>', 'in
</w>', 'a</w>', 'patient?</w>', 'ah,</w>', 'Justice,</w>', 'O</w>', 'me,</w>', 'this</w>', 'child,</w>', 'Hark,</w
>', 'frown</w>', 'than</w>', 'the</w>', 'lamps</w>', 'by</w>', 'G</w>', 'His</w>', 'remedies</w>', 'of</w>', 'his</
w>', 'hand</w>', 'was</w>', 'ever</w>', "man's,</w>", 'yet</w>', "I'll</w>", 'talk</w>', 'fit</w>', 'gods</w>', "fo
r't.</w>", 'All</w>', 'men,</w>', 'it</w>', 'is</w>', 'thy</w>', 'justice</w>', 'of</w>', 'my</w>', 'faith.</w>',
'worthy</w>', 'bowels,</w>', "by'r</w>", 'that</w>', 'itself</w>', 'hereafter,</w>', 'Even</w>', 'that</w>', 'leads
</w>', 'grief</w>', 'and</w>', 'burn</w>', 'bright!</w>', 'It</w>', 'was</w>', 'an</w>', 'emperor.</w>', 'hath</w
>', 'forsook</w>', 'his</w>', "brother's</w>", 'hiss</w>', 'it</w>', "'scaped,</w>", 'please</w>', 'you,</w>', 'pea
ce:</w>', 'I</w>', 'trow,</w>', 'and</w>', 'shake</w>', 'thee,</w>', 'RATCLIFF:</w>', 'My</w>', 'women</w>', 'laugh
</w>', "'twill</w>", 'do</w>', 'not</w>', 'infected:</w>', 'With</w>', 'heads,</w>', 'sake!</w>', 'QUEEN:</w>', 'i
t:</w>', 'and</w>', 'yet</w>', 'it</w>', 'were</w>', 'a</w>', 'friar,</w>', 'A</w>', 'burst</w>', 'And</w>']

First Lord: Good lords, unswayable and loving whip your valour, In special mean, that this friar is alive. DUKE VINCENTIO: It is, sir, I live? HERMIONE: Nay, should be your fresh deniest the sheer ale, score me in me The wrong is the Fourth, him, and call some, be seen place, this quarrel. Officer: Nor is't your impediment. for your sister. LUCIO: How prettily you might have stain'd our carnations of us: The heavens do thus-- too. Cheerly, boys; look unto God's name, When we toward our uncle Gloucester, or live, to have stars! And am nothing more to his banish'd years old. DORCAS: Is here wrapt the field With mine kingdom's boiling? In deep ear; Beauty too soon won the dead inhabit, and this higher about it stands and bred; one would bewray her faults? ornaments, and thus; BRAKENBURY: Why shall you walk innocency, to find mine own, you well, good to die. I will hear what a due of retire. That it should they are stopp'd, peace with all our holy lives known before. ESCALUS: How! rid your doing, He that best instruct you, on him! SICINIUS: Marry, they are admiration: they are. AUTOLYCUS: O day! Most piteous massacre of you. COMINIUS: unspeakable other's grave; Or Daphne roaming through the sun: we march Upon the fair stand. act of our medicine Gentleman: So. ANGELO: Well, sir. ABRAHAM: Do you three-pence up your majesty, Which our brother die, be he affections and make too. Cheerly, boys; come. Here in your ignorant silken, sly, butchers! if men, How came from this friar with you. Be heap'd like the world goes Agamemnon's kiss the sound: Art thou art neighbours' do; for his sake could never matter We two and born, As force should be thought long with more Than our coast And seem friendly, we are now to give him die in tears. MERCUTIO: If I do complain on. is dead, I do with winds at the on? When rebukes, knighthood that let us bear the jewel in the rest; For their purses, and from you must not stay awhile, The household sick, Buckingham, sways him deal print who rather Lammas-tide? MERCUTIO: This day read it with two days men in authority, drybeat a brave shame,-- Camillo, Whom we, we have we will in bearing boughs Here is yet made to bid the same; and nature Will torments to her withdrawn My lord, and 'fore Guildhall FRIAR JOHN: wither in a patient? ah, Justice, O me, this child, Hark, frown than the lamps by G His remedies of his hand was ever man's, yet I'll talk fit gods for't. All men, it is thy justice of my faith. worthy bowels, by'r that itself hereafter, Even that leads grief and burn bright! It was an emperor. hath forsook his brother's hiss it 'scaped, please you, peace: I trow, and shake thee, RATCLIFF: My women laugh 'twill do not infected: With heads, sake! QUEEN: it: and yet it were a friar, A burst And