```
!git clone https://github.com/karpathy/nanoGPT.g
```

```
Cloning into 'nanoGPT'...
remote: Enumerating objects: 686, done.
remote: Total 686 (delta 0), reused 0 (delta
Receiving objects: 100% (686/686), 954.04 KiB
Resolving deltas: 100% (387/387), done.
```

```
%cd nanoGPT
```

```
/content/nanoGPT
```

```
!python data/shakespeare_char/prepare.py
```

```
length of dataset in characters: 1,115,394
all the unique characters:
 !$&',-.3:;?ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefg
vocab size: 65
train has 1,003,854 tokens
val has 111,540 tokens
```

```
!pip install torch numpy transformers datasets ti
```

```
                                        13
wnloading nvidia_cuda_nvrtc_cu12-12.4.127-py
                                        24
wnloading nvidia_cuda_runtime_cu12-12.4.127-
                                        88
wnloading nvidia_cudnn_cu12-9.1.0.70-py3-nor
                                        66
wnloading nvidia_cufft_cu12-11.2.1.3-py3-nor
                                        21
wnloading nvidia_curand_cu12-10.3.5.147-py3-
                                        56
wnloading nvidia_cusolver_cu12-11.6.1.9-py3-
                                        12
wnloading nvidia_cusparse_cu12-12.3.1.170-py
                                        20
wnloading nvidia_nvjitlink_cu12-12.4.127-py3
                                        21
stalling collected packages: nvidia-nvjitlir
Attempting uninstall: nvidia-nvjitlink-cu12
  Successfully uninstalled nvidia-nvjitlir
Attempting uninstall: nvidia-curand-cu12
  Found existing installation: nvidia-curanc
```

sample.py

```
from contextlib import nullc
import torch
import tiktoken
from model import GPTConfig,

# -----------------------------
init_from = 'resume' # eithe
out_dir = 'out' # ignored if
start = "The capital of Sing
num_samples = 10 # number of
max_new_tokens = 500 # numbe
temperature = 0.8 # 1.0 = no
top_k = 200 # retain only th
seed = 1337
device = 'cuda' # examples:
dtype = 'bfloat16' if torch.
compile = False # use PyTorc
exec(open('configurator.py')
# -----------------------------

torch.manual_seed(seed)
torch.cuda.manual_seed(seed)
torch.backends.cuda.matmul.a
torch.backends.cudnn.allow_t
device_type = 'cuda' if 'cud
ptdtype = {'float32': torch.
ctx = nullcontext() if devic

# model
if init_from == 'resume':
    # init from a model save
    ckpt_path = os.path.joir
    checkpoint = torch.load(
    gptconf = GPTConfig(**ch
    model = GPT(gptconf)
    state_dict = checkpoint[
    unwanted_prefix = '_orig
    for k,v in list(state_di
        if k.startswith(unwa
            state_dict[k[ler
    model.load_state_dict(st
elif init_from.startswith('g
                    given GPT-
                    om pretrai
```

What can I help you build?

```
  Uninstalling nvidia-curand-cu12-10.3.6.82:
    Successfully uninstalled nvidia-curand-c
Attempting uninstall: nvidia-cufft-cu12
  Found existing installation: nvidia-cufft-
  Uninstalling nvidia-cufft-cu12-11.2.3.61:
    Successfully uninstalled nvidia-cufft-cu
Attempting uninstall: nvidia-cuda-runtime-cu
  Found existing installation: nvidia-cuda-r
  Uninstalling nvidia-cuda-runtime-cu12-12.5
    Successfully uninstalled nvidia-cuda-run
Attempting uninstall: nvidia-cuda-nvrtc-cu12
  Found existing installation: nvidia-cuda-r
  Uninstalling nvidia-cuda-nvrtc-cu12-12.5.8
    Successfully uninstalled nvidia-cuda-nvr
Attempting uninstall: nvidia-cuda-cupti-cu12
  Found existing installation: nvidia-cuda-c
  Uninstalling nvidia-cuda-cupti-cu12-12.5.8
    Successfully uninstalled nvidia-cuda-cup
Attempting uninstall: nvidia-cublas-cu12
  Found existing installation: nvidia-cublas
  Uninstalling nvidia-cublas-cu12-12.5.3.2:
    Successfully uninstalled nvidia-cublas-c
Attempting uninstall: nvidia-cusparse-cu12
  Found existing installation: nvidia-cuspar
  Uninstalling nvidia-cusparse-cu12-12.5.1.3
    Successfully uninstalled nvidia-cusparse
Attempting uninstall: nvidia-cudnn-cu12
  Found existing installation: nvidia-cudnn-
  Uninstalling nvidia-cudnn-cu12-9.3.0.75:
    Successfully uninstalled nvidia-cudnn-cu
Attempting uninstall: nvidia-cusolver-cu12
  Found existing installation: nvidia-cusolv
  Uninstalling nvidia-cusolver-cu12-11.6.3.8
    Successfully uninstalled nvidia-cusolver
ccessfully installed nvidia-cublas-cu12-12.4
```

```
!python train.py config/train_shakespeare_char.py
```

```
iter 4440: loss 0.8564, time 528.77ms, mfu
iter 4450: loss 0.8425, time 527.59ms, mfu
iter 4460: loss 0.8305, time 529.93ms, mfu
iter 4470: loss 0.8549, time 527.40ms, mfu
iter 4480: loss 0.8381, time 526.45ms, mfu
iter 4490: loss 0.8393, time 527.65ms, mfu
step 4500: train loss 0.6517, val loss 1.67
iter 4500: loss 0.8571, time 74742.32ms, mf
iter 4510: loss 0.8488, time 526.38ms, mfu
iter 4520: loss 0.8330, time 526.68ms, mfu
iter 4530: loss 0.8458, time 530.70ms, mfu
iter 4540: loss 0.8463, time 527.95ms, mfu
iter 4550: loss 0.8774, time 527.32ms, mfu
iter 4560: loss 0.8378, time 527.91ms, mfu
iter 4570: loss 0.8397, time 526.53ms, mfu
iter 4580: loss 0.8508, time 527.05ms, mfu
```

```
iter 4590: loss 0.8508, time 527.62ms, mfu
iter 4600: loss 0.8350, time 528.47ms, mfu
iter 4610: loss 0.8740, time 528.39ms, mfu
iter 4620: loss 0.8351, time 529.11ms, mfu
iter 4630: loss 0.8250, time 527.95ms, mfu
iter 4640: loss 0.8364, time 528.54ms, mfu
iter 4650: loss 0.8563, time 528.40ms, mfu
iter 4660: loss 0.8477, time 528.39ms, mfu
iter 4670: loss 0.8341, time 526.84ms, mfu
iter 4680: loss 0.8501, time 527.12ms, mfu
iter 4690: loss 0.8521, time 525.93ms, mfu
iter 4700: loss 0.8243, time 527.80ms, mfu
iter 4710: loss 0.7913, time 528.64ms, mfu
iter 4720: loss 0.8323, time 528.43ms, mfu
iter 4730: loss 0.8126, time 525.85ms, mfu
iter 4740: loss 0.8283, time 527.19ms, mfu
step 4750: train loss 0.6337, val loss 1.69
iter 4750: loss 0.7967, time 74746.66ms, mf
iter 4760: loss 0.8125, time 527.79ms, mfu
iter 4770: loss 0.8011, time 526.71ms, mfu
iter 4780: loss 0.8087, time 528.66ms, mfu
iter 4790: loss 0.8405, time 529.27ms, mfu
iter 4800: loss 0.8299, time 528.24ms, mfu
iter 4810: loss 0.8391, time 525.81ms, mfu
iter 4820: loss 0.8200, time 527.12ms, mfu
iter 4830: loss 0.8243, time 528.12ms, mfu
iter 4840: loss 0.8335, time 525.61ms, mfu
iter 4850: loss 0.8148, time 527.84ms, mfu
iter 4860: loss 0.8190, time 527.25ms, mfu
iter 4870: loss 0.8079, time 527.06ms, mfu
iter 4880: loss 0.8379, time 529.25ms, mfu
iter 4890: loss 0.8022, time 529.44ms, mfu
iter 4900: loss 0.8047, time 529.15ms, mfu
iter 4910: loss 0.8247, time 526.71ms, mfu
iter 4920: loss 0.8233, time 529.00ms, mfu
iter 4930: loss 0.8128, time 526.70ms, mfu
iter 4940: loss 0.8081, time 527.27ms, mfu
iter 4950: loss 0.8254, time 526.69ms, mfu
iter 4960: loss 0.8310, time 526.98ms, mfu
iter 4970: loss 0.7889, time 526.10ms, mfu
iter 4980: loss 0.8010, time 527.94ms, mfu
iter 4990: loss 0.8295, time 527.37ms, mfu
step 5000: train loss 0.6197, val loss 1.70
```

```
!pip install matplotlib
```

```
Requirement already satisfied: matplotlib in
Requirement already satisfied: contourpy>=1.0
Requirement already satisfied: cycler>=0.10 i
Requirement already satisfied: fonttools>=4.2
Requirement already satisfied: kiwisolver>=1.
Requirement already satisfied: numpy>=1.23 ir
Requirement already satisfied: packaging>=20.
Requirement already satisfied: pillow>=8 in /
Requirement already satisfied: pyparsing>=2.3
Requirement already satisfied: python-dateuti
Requirement already satisfied: six>=1.5 in /u
```

```
!python sample.py --out_dir=out-shakespeare-char
```

```
---------------

Messenger:
The soft is full of all.

LARTIUS:
No, but not take the complexion, and
I need to thine eyes; but till this show
Of much a steed, which with an absence.I Ca
For the poll, the angelong in the prayers o
And witness what seems to bear the course's
That with the princely life enough a deep o
Of this answer laid in steep belowing on th
I do not serve you not buy forth you, sir.

ANGELO:
I am given no tongue.

ISABELLA:
Well, I would not be gone with him,

---------------


First Murderer:
Why, how now, that I fear? thou art too com

Second Murderer:
But now I am born that prevent dead haste
That thou liest remember to her hand,
That I was that become the sins and so be
That take the blow no end, that flowers tha
Though make her act the world apoll o' the

First Murderer:
The sound to prosper the airy of my love,
```

```
That thy poor hath dismissent my and that E
When thy parliaments for down's father be s

Shepherd:

MARI
---------------

That lamb did spurn my kingdom from their r
That stole dead must have brought it not be
Nor I will bear the love of the death.

BUSHY:
Nothing with mine own bloody time to the or
That owe them honour that now have been mor
To supply dangerous man unto them.

Page:
Will you hear, and will not put up in his s
To make him the mar make him grant. We woul
That fardel and more than our souls the air
If it were not, he were not for ours.
```

```
!python sample.py --out_dir=out-shakespeare-char
```

```
Overriding: out_dir = out-shakespeare-char
number of parameters: 10.65M
Loading meta from data/shakespeare_char/met
Who are you?

CORIOLANUS:
O, and it is but
back.

MENENIUS:
Stirrah, that's no better he us all to bed
the people.

CORIOLANUS:
Nor much one of of it.

CORIOLANUS:
No, even we sent in the sea, to be patient.
gods will perfluor the people, on the consu
him you within Rome. The gods have like the
only so did Corioli within, this she had no
As for use them to the first; worse must no
Made her brood.

MENENIUS:
Here's a bad a man a gentleman bastard.
```

```
ANTIGONUS:
I am not a w
---------------
Who are you?

QUEEN MARGARET:
You should you say, and so much many words
Are not to me to see your grace, with confe
Father of the castle lamentation.

KING RICHARD III:
Ay, stay with me over some sound authority,
And so the golden churches on the helm,
To make heaven and the warlike cowards of h

QUEEN:
Welcome, we do more honours weeping but bro
Commend me to to this determined command,
For Italy is better it, but we have slept t
And bid with this own teeth that you have m
---------------
Who are you?
Then, my lord,
That I cannot know. She is not a dozen so
Sad your countryman, and my beggars mine ow
To waiting your son: it must be then,
As it was a merity a vine
Of a cause o' the field and in the death.
It is born thing but mad to-night; I think
That she would be something last; and given
Do so something that are but babe lowed his
```

```
!python sample.py --out_dir=out-shakespeare-char
```

```
CLAUDIO:
Come, get
---------------
The capital of Singapore is a man
A present of the execution of the heaven,
And vex in the while and the rest
That the love of boldness of this deed;
And then this best the duke with a skilful
When he says that I should call in some ill

CORIOLANUS:
Nay, I see your gates was never been so dea
Where do I not, that still have forgot come

CORIOLANUS:
True, good sir,
The gods have some alter'd woman to the edg
And have been a widow of but remedies.
The part of wretched souls present me to La
```

```
---------------
The capital of Singapore is yourself,
Your ungrace, you shall follow the prince,
Your since not how vain shortly people;
Poor souls are more than all our flesh of l

DUKE OF AUMERLE:
Now, Buckingham, go see thither.

DUKE OF AUMERLE:
Come, know your grace: sweet York haste you

EXETER:
To him a noble lord:
When the king, and that it does be drunked
And there you shall be prepared with truth-

KING RICHARD III:
I thought her is it hated that cherish'd hi

QUEEN ELIZABETH:
Heave
---------------
The capital of Singapore is the heavens of
And we may and speak, I would have leave to
And tell me in Richard and to seet us
To supplian of this foot hot.

RICHARD:
Belike the sorrow
That would not be reason'd to a world.

EDWARD:
Not I; I am disloy'd.

Messenger:
Nor I hope of my sword, and your lady are d

GLOUCESTER:
```

Start coding or generate with AI.

```
# 1. Copy the entire output from your training run
# Make sure to include all the lines that look lik
log_data = """
step 0: train loss 4.2874, val loss 4.2823
iter 0: loss 4.2664, time 69850.79ms, mfu -100.00%
iter 10: loss 3.1421, time 512.16ms, mfu 0.73%
iter 20: loss 2.7345, time 511.90ms, mfu 0.73%
iter 30: loss 2.6178, time 514.72ms, mfu 0.73%
```

```
iter 40: loss 2.5734, time 519.02ms, mfu 0.73%
iter 50: loss 2.5265, time 519.48ms, mfu 0.73%
iter 60: loss 2.5110, time 523.59ms, mfu 0.72%
iter 70: loss 2.4956, time 524.66ms, mfu 0.72%
iter 80: loss 2.4940, time 528.63ms, mfu 0.72%
iter 90: loss 2.4703, time 532.63ms, mfu 0.72%
iter 100: loss 2.4566, time 535.28ms, mfu 0.72%
iter 110: loss 2.4512, time 533.27ms, mfu 0.71%
iter 120: loss 2.4283, time 528.87ms, mfu 0.71%
iter 130: loss 2.4101, time 528.21ms, mfu 0.71%
iter 140: loss 2.4184, time 524.09ms, mfu 0.71%
iter 150: loss 2.4237, time 522.57ms, mfu 0.71%
iter 160: loss 2.3839, time 520.99ms, mfu 0.71%
iter 170: loss 2.3574, time 521.25ms, mfu 0.71%
iter 180: loss 2.3153, time 520.07ms, mfu 0.71%
iter 190: loss 2.2506, time 522.90ms, mfu 0.71%
iter 200: loss 2.2119, time 522.74ms, mfu 0.71%
iter 210: loss 2.1391, time 524.75ms, mfu 0.71%
iter 220: loss 2.1315, time 526.81ms, mfu 0.71%
iter 230: loss 2.0739, time 527.73ms, mfu 0.71%
iter 240: loss 2.0785, time 530.34ms, mfu 0.71%
step 250: train loss 1.9572, val loss 2.0615
saving checkpoint to out-shakespeare-char
iter 250: loss 2.0310, time 74856.33ms, mfu 0.64%
iter 260: loss 1.9766, time 523.26ms, mfu 0.65%
iter 270: loss 1.9743, time 524.59ms, mfu 0.65%
iter 280: loss 1.9706, time 523.85ms, mfu 0.66%
iter 290: loss 1.9249, time 525.30ms, mfu 0.66%
iter 300: loss 1.9007, time 526.07ms, mfu 0.67%
iter 310: loss 1.8542, time 527.79ms, mfu 0.67%
iter 320: loss 1.8442, time 528.94ms, mfu 0.68%
iter 330: loss 1.8144, time 530.35ms, mfu 0.68%
iter 340: loss 1.7846, time 529.21ms, mfu 0.68%
iter 350: loss 1.8244, time 527.79ms, mfu 0.68%
iter 360: loss 1.7573, time 528.78ms, mfu 0.69%
iter 370: loss 1.7368, time 528.71ms, mfu 0.69%
iter 380: loss 1.7255, time 528.15ms, mfu 0.69%
iter 390: loss 1.7323, time 527.44ms, mfu 0.69%
iter 400: loss 1.7661, time 530.35ms, mfu 0.69%
iter 410: loss 1.6972, time 527.30ms, mfu 0.69%
iter 420: loss 1.7075, time 527.02ms, mfu 0.70%
iter 430: loss 1.6903, time 528.89ms, mfu 0.70%
iter 440: loss 1.6565, time 527.12ms, mfu 0.70%
iter 450: loss 1.6497, time 529.40ms, mfu 0.70%
iter 460: loss 1.5939, time 527.29ms, mfu 0.70%
iter 470: loss 1.6546, time 528.38ms, mfu 0.70%
iter 480: loss 1.6206, time 526.88ms, mfu 0.70%
iter 490: loss 1.6009, time 526.68ms, mfu 0.70%
step 500: train loss 1.5255, val loss 1.7345
```

```
saving checkpoint to out-shakespeare-char
iter 500: loss 1.6011, time 74951.39ms, mfu 0.63%
iter 510: loss 1.6049, time 524.27ms, mfu 0.64%
iter 520: loss 1.5938, time 526.90ms, mfu 0.65%
iter 530: loss 1.5625, time 527.84ms, mfu 0.65%
iter 540: loss 1.6213, time 527.02ms, mfu 0.66%
iter 550: loss 1.5650, time 526.98ms, mfu 0.66%
iter 560: loss 1.5658, time 527.70ms, mfu 0.67%
iter 570: loss 1.5653, time 527.58ms, mfu 0.67%
iter 580: loss 1.5399, time 526.56ms, mfu 0.67%
iter 590: loss 1.4954, time 528.31ms, mfu 0.68%
iter 600: loss 1.5169, time 527.22ms, mfu 0.68%
iter 610: loss 1.5409, time 526.91ms, mfu 0.68%
iter 620: loss 1.5249, time 526.95ms, mfu 0.69%
iter 630: loss 1.5092, time 525.22ms, mfu 0.69%
iter 640: loss 1.4691, time 527.33ms, mfu 0.69%
iter 650: loss 1.5024, time 524.70ms, mfu 0.69%
iter 660: loss 1.5116, time 527.32ms, mfu 0.69%
iter 670: loss 1.4494, time 527.86ms, mfu 0.69%
iter 680: loss 1.5110, time 529.05ms, mfu 0.70%
iter 690: loss 1.4617, time 527.60ms, mfu 0.70%
iter 700: loss 1.4864, time 527.33ms, mfu 0.70%
iter 710: loss 1.4568, time 528.20ms, mfu 0.70%
iter 720: loss 1.4428, time 528.00ms, mfu 0.70%
iter 730: loss 1.4228, time 526.53ms, mfu 0.70%
iter 740: loss 1.4255, time 528.33ms, mfu 0.70%
step 750: train loss 1.3618, val loss 1.5939
saving checkpoint to out-shakespeare-char
iter 750: loss 1.4324, time 74986.26ms, mfu 0.63%
iter 760: loss 1.4374, time 525.28ms, mfu 0.64%
iter 770: loss 1.4199, time 524.33ms, mfu 0.65%
iter 780: loss 1.4192, time 526.53ms, mfu 0.65%
iter 790: loss 1.4182, time 526.54ms, mfu 0.66%
iter 800: loss 1.4286, time 527.58ms, mfu 0.66%
iter 810: loss 1.3963, time 526.46ms, mfu 0.67%
iter 820: loss 1.4067, time 528.26ms, mfu 0.67%
iter 830: loss 1.3897, time 528.83ms, mfu 0.67%
iter 840: loss 1.3975, time 527.44ms, mfu 0.68%
iter 850: loss 1.3888, time 527.45ms, mfu 0.68%
iter 860: loss 1.3942, time 528.94ms, mfu 0.68%
iter 870: loss 1.3948, time 526.54ms, mfu 0.69%
iter 880: loss 1.3734, time 525.78ms, mfu 0.69%
iter 890: loss 1.3830, time 529.78ms, mfu 0.69%
iter 900: loss 1.3685, time 527.16ms, mfu 0.69%
iter 910: loss 1.3202, time 527.35ms, mfu 0.69%
iter 920: loss 1.3628, time 527.98ms, mfu 0.69%
iter 930: loss 1.3630, time 529.07ms, mfu 0.69%
iter 940: loss 1.3375, time 528.88ms, mfu 0.70%
iter 950: loss 1.3487, time 527.08ms, mfu 0.70%
```

```
iter 960: loss 1.3638, time 527.27ms, mfu 0.70%
iter 970: loss 1.3546, time 525.73ms, mfu 0.70%
iter 980: loss 1.3575, time 526.99ms, mfu 0.70%
iter 990: loss 1.3372, time 525.70ms, mfu 0.70%
step 1000: train loss 1.2700, val loss 1.5153
saving checkpoint to out-shakespeare-char
iter 1000: loss 1.3323, time 75027.48ms, mfu 0.63%
iter 1010: loss 1.3346, time 525.65ms, mfu 0.64%
iter 1020: loss 1.3133, time 525.98ms, mfu 0.65%
iter 1030: loss 1.3390, time 528.37ms, mfu 0.65%
iter 1040: loss 1.3635, time 528.09ms, mfu 0.66%
iter 1050: loss 1.2978, time 526.54ms, mfu 0.66%
iter 1060: loss 1.3423, time 527.58ms, mfu 0.67%
iter 1070: loss 1.3344, time 527.60ms, mfu 0.67%
iter 1080: loss 1.3348, time 528.73ms, mfu 0.67%
iter 1090: loss 1.3509, time 527.19ms, mfu 0.68%
iter 1100: loss 1.3144, time 527.78ms, mfu 0.68%
iter 1110: loss 1.2982, time 527.95ms, mfu 0.68%
iter 1120: loss 1.3019, time 528.08ms, mfu 0.69%
iter 1130: loss 1.2981, time 525.29ms, mfu 0.69%
iter 1140: loss 1.2975, time 526.07ms, mfu 0.69%
iter 1150: loss 1.3056, time 527.21ms, mfu 0.69%
iter 1160: loss 1.3293, time 528.07ms, mfu 0.69%
iter 1170: loss 1.2993, time 527.12ms, mfu 0.69%
iter 1180: loss 1.3187, time 528.78ms, mfu 0.70%
iter 1190: loss 1.2622, time 528.01ms, mfu 0.70%
iter 1200: loss 1.2932, time 528.23ms, mfu 0.70%
iter 1210: loss 1.2590, time 530.50ms, mfu 0.70%
iter 1220: loss 1.3061, time 527.38ms, mfu 0.70%
iter 1230: loss 1.2969, time 530.09ms, mfu 0.70%
iter 1240: loss 1.3021, time 524.21ms, mfu 0.70%
step 1250: train loss 1.2038, val loss 1.4965
saving checkpoint to out-shakespeare-char
iter 1250: loss 1.2697, time 75002.78ms, mfu 0.63%
iter 1260: loss 1.2837, time 526.45ms, mfu 0.64%
iter 1270: loss 1.2645, time 528.94ms, mfu 0.64%
iter 1280: loss 1.2518, time 528.00ms, mfu 0.65%
iter 1290: loss 1.2843, time 525.81ms, mfu 0.66%
iter 1300: loss 1.3016, time 526.83ms, mfu 0.66%
iter 1310: loss 1.2394, time 525.40ms, mfu 0.67%
iter 1320: loss 1.3073, time 527.99ms, mfu 0.67%
iter 1330: loss 1.2619, time 527.07ms, mfu 0.67%
iter 1340: loss 1.2983, time 526.83ms, mfu 0.68%
iter 1350: loss 1.2583, time 527.38ms, mfu 0.68%
iter 1360: loss 1.2772, time 527.32ms, mfu 0.68%
iter 1370: loss 1.2551, time 528.71ms, mfu 0.69%
iter 1380: loss 1.2624, time 528.26ms, mfu 0.69%
iter 1390: loss 1.2475, time 526.85ms, mfu 0.69%
iter 1400: loss 1.2503, time 527.36ms, mfu 0.69%
```

```
iter 1410: loss 1.2437, time 528.69ms, mfu 0.69%
iter 1420: loss 1.2690, time 526.31ms, mfu 0.69%
iter 1430: loss 1.2472, time 530.45ms, mfu 0.69%
iter 1440: loss 1.2456, time 526.32ms, mfu 0.70%
iter 1450: loss 1.2310, time 527.37ms, mfu 0.70%
iter 1460: loss 1.2380, time 527.78ms, mfu 0.70%
iter 1470: loss 1.2198, time 528.00ms, mfu 0.70%
iter 1480: loss 1.2060, time 528.11ms, mfu 0.70%
iter 1490: loss 1.2349, time 526.06ms, mfu 0.70%
step 1500: train loss 1.1492, val loss 1.4737
saving checkpoint to out-shakespeare-char
iter 1500: loss 1.1842, time 75001.02ms, mfu 0.63%
iter 1510: loss 1.2325, time 524.36ms, mfu 0.64%
iter 1520: loss 1.2276, time 527.74ms, mfu 0.65%
iter 1530: loss 1.2568, time 528.63ms, mfu 0.65%
iter 1540: loss 1.1957, time 529.02ms, mfu 0.66%
iter 1550: loss 1.2305, time 529.62ms, mfu 0.66%
iter 1560: loss 1.2036, time 527.89ms, mfu 0.67%
iter 1570: loss 1.2350, time 528.28ms, mfu 0.67%
iter 1580: loss 1.2088, time 527.61ms, mfu 0.67%
iter 1590: loss 1.1906, time 528.65ms, mfu 0.68%
iter 1600: loss 1.1949, time 527.91ms, mfu 0.68%
iter 1610: loss 1.2376, time 528.14ms, mfu 0.68%
iter 1620: loss 1.1775, time 527.27ms, mfu 0.68%
iter 1630: loss 1.2023, time 526.81ms, mfu 0.69%
iter 1640: loss 1.1954, time 526.46ms, mfu 0.69%
iter 1650: loss 1.1817, time 527.82ms, mfu 0.69%
iter 1660: loss 1.2117, time 527.40ms, mfu 0.69%
iter 1670: loss 1.2043, time 527.44ms, mfu 0.69%
iter 1680: loss 1.2003, time 527.69ms, mfu 0.69%
iter 1690: loss 1.2007, time 526.95ms, mfu 0.70%
iter 1700: loss 1.1890, time 528.47ms, mfu 0.70%
iter 1710: loss 1.1795, time 525.99ms, mfu 0.70%
iter 1720: loss 1.1815, time 526.98ms, mfu 0.70%
iter 1730: loss 1.2069, time 526.98ms, mfu 0.70%
iter 1740: loss 1.1728, time 526.37ms, mfu 0.70%
step 1750: train loss 1.1006, val loss 1.4642
saving checkpoint to out-shakespeare-char
iter 1750: loss 1.1908, time 74922.69ms, mfu 0.63%
iter 1760: loss 1.1899, time 526.73ms, mfu 0.64%
iter 1770: loss 1.1983, time 525.66ms, mfu 0.65%
iter 1780: loss 1.1889, time 528.13ms, mfu 0.65%
iter 1790: loss 1.1878, time 527.63ms, mfu 0.66%
iter 1800: loss 1.1818, time 525.42ms, mfu 0.66%
iter 1810: loss 1.1610, time 526.15ms, mfu 0.67%
iter 1820: loss 1.1642, time 527.23ms, mfu 0.67%
iter 1830: loss 1.1752, time 525.23ms, mfu 0.67%
iter 1840: loss 1.1591, time 526.85ms, mfu 0.68%
iter 1850: loss 1.1538, time 527.88ms, mfu 0.68%
```

```
iter 1850: loss 1.1356, time 527.68ms, mfu 0.68%
iter 1860: loss 1.1728, time 528.27ms, mfu 0.68%
iter 1870: loss 1.1443, time 526.33ms, mfu 0.69%
iter 1880: loss 1.1845, time 527.70ms, mfu 0.69%
iter 1890: loss 1.1780, time 528.79ms, mfu 0.69%
iter 1900: loss 1.1288, time 526.92ms, mfu 0.69%
iter 1910: loss 1.1669, time 526.93ms, mfu 0.69%
iter 1920: loss 1.1693, time 526.86ms, mfu 0.69%
iter 1930: loss 1.1456, time 527.88ms, mfu 0.70%
iter 1940: loss 1.1280, time 525.55ms, mfu 0.70%
iter 1950: loss 1.1340, time 527.30ms, mfu 0.70%
iter 1960: loss 1.1522, time 527.78ms, mfu 0.70%
iter 1970: loss 1.1480, time 525.47ms, mfu 0.70%
iter 1980: loss 1.1362, time 526.97ms, mfu 0.70%
iter 1990: loss 1.1530, time 527.69ms, mfu 0.70%
step 2000: train loss 1.0547, val loss 1.4796
iter 2000: loss 1.1226, time 74785.06ms, mfu 0.63%
iter 2010: loss 1.1255, time 531.73ms, mfu 0.64%
iter 2020: loss 1.1243, time 529.11ms, mfu 0.64%
iter 2030: loss 1.1584, time 527.93ms, mfu 0.65%
iter 2040: loss 1.1426, time 528.32ms, mfu 0.66%
iter 2050: loss 1.1130, time 526.87ms, mfu 0.66%
iter 2060: loss 1.1011, time 526.55ms, mfu 0.67%
iter 2070: loss 1.1274, time 527.05ms, mfu 0.67%
iter 2080: loss 1.1194, time 526.53ms, mfu 0.67%
iter 2090: loss 1.1280, time 528.59ms, mfu 0.68%
iter 2100: loss 1.1346, time 528.28ms, mfu 0.68%
iter 2110: loss 1.1318, time 527.37ms, mfu 0.68%
iter 2120: loss 1.1221, time 526.86ms, mfu 0.69%
iter 2130: loss 1.1357, time 526.75ms, mfu 0.69%
iter 2140: loss 1.1401, time 526.63ms, mfu 0.69%
iter 2150: loss 1.1197, time 528.51ms, mfu 0.69%
iter 2160: loss 1.1464, time 525.75ms, mfu 0.69%
iter 2170: loss 1.1307, time 526.17ms, mfu 0.69%
iter 2180: loss 1.1100, time 528.56ms, mfu 0.70%
iter 2190: loss 1.1047, time 527.20ms, mfu 0.70%
iter 2200: loss 1.1236, time 527.73ms, mfu 0.70%
iter 2210: loss 1.1144, time 527.56ms, mfu 0.70%
iter 2220: loss 1.1241, time 528.24ms, mfu 0.70%
iter 2230: loss 1.1171, time 525.77ms, mfu 0.70%
iter 2240: loss 1.1257, time 528.19ms, mfu 0.70%
step 2250: train loss 1.0074, val loss 1.4860
iter 2250: loss 1.1084, time 74788.35ms, mfu 0.63%
iter 2260: loss 1.1099, time 527.90ms, mfu 0.64%
iter 2270: loss 1.1240, time 528.17ms, mfu 0.65%
iter 2280: loss 1.0932, time 529.37ms, mfu 0.65%
iter 2290: loss 1.1433, time 526.73ms, mfu 0.66%
iter 2300: loss 1.1203, time 528.45ms, mfu 0.66%
iter 2310: loss 1.0976, time 529.81ms, mfu 0.67%
iter 2320: loss 1.0934, time 527.51ms, mfu 0.67%
```

```
iter 2320: loss 1.0954, time 527.31ms, mfu 0.67%
iter 2330: loss 1.0922, time 527.47ms, mfu 0.67%
iter 2340: loss 1.1130, time 526.61ms, mfu 0.68%
iter 2350: loss 1.1003, time 525.90ms, mfu 0.68%
iter 2360: loss 1.1007, time 528.89ms, mfu 0.68%
iter 2370: loss 1.0830, time 526.40ms, mfu 0.69%
iter 2380: loss 1.0754, time 530.31ms, mfu 0.69%
iter 2390: loss 1.0769, time 527.21ms, mfu 0.69%
iter 2400: loss 1.0788, time 525.74ms, mfu 0.69%
iter 2410: loss 1.0749, time 527.52ms, mfu 0.69%
iter 2420: loss 1.0747, time 528.40ms, mfu 0.69%
iter 2430: loss 1.0553, time 527.09ms, mfu 0.69%
iter 2440: loss 1.0658, time 529.99ms, mfu 0.70%
iter 2450: loss 1.0727, time 528.16ms, mfu 0.70%
iter 2460: loss 1.0874, time 526.85ms, mfu 0.70%
iter 2470: loss 1.0934, time 527.08ms, mfu 0.70%
iter 2480: loss 1.0832, time 527.11ms, mfu 0.70%
iter 2490: loss 1.0571, time 527.77ms, mfu 0.70%
step 2500: train loss 0.9589, val loss 1.4922
iter 2500: loss 1.0830, time 74757.15ms, mfu 0.63%
iter 2510: loss 1.0662, time 527.15ms, mfu 0.64%
iter 2520: loss 1.0469, time 526.65ms, mfu 0.65%
iter 2530: loss 1.0509, time 528.27ms, mfu 0.65%
iter 2540: loss 1.0508, time 528.96ms, mfu 0.66%
iter 2550: loss 1.0610, time 531.37ms, mfu 0.66%
iter 2560: loss 1.0571, time 527.03ms, mfu 0.67%
iter 2570: loss 1.0765, time 528.95ms, mfu 0.67%
iter 2580: loss 1.0836, time 527.90ms, mfu 0.67%
iter 2590: loss 1.0685, time 527.08ms, mfu 0.68%
iter 2600: loss 1.0632, time 528.00ms, mfu 0.68%
iter 2610: loss 1.0462, time 528.20ms, mfu 0.68%
iter 2620: loss 1.0462, time 525.66ms, mfu 0.68%
iter 2630: loss 1.0235, time 527.22ms, mfu 0.69%
iter 2640: loss 1.0444, time 526.07ms, mfu 0.69%
iter 2650: loss 1.0670, time 526.67ms, mfu 0.69%
iter 2660: loss 1.0396, time 526.48ms, mfu 0.69%
iter 2670: loss 1.0229, time 528.31ms, mfu 0.69%
iter 2680: loss 1.0507, time 526.99ms, mfu 0.70%
iter 2690: loss 1.0481, time 527.31ms, mfu 0.70%
iter 2700: loss 1.0228, time 528.42ms, mfu 0.70%
iter 2710: loss 1.0470, time 527.91ms, mfu 0.70%
iter 2720: loss 1.0383, time 528.00ms, mfu 0.70%
iter 2730: loss 1.0547, time 527.59ms, mfu 0.70%
iter 2740: loss 1.0168, time 528.69ms, mfu 0.70%
step 2750: train loss 0.9126, val loss 1.5167
iter 2750: loss 1.0390, time 74836.33ms, mfu 0.63%
iter 2760: loss 1.0322, time 526.38ms, mfu 0.64%
iter 2770: loss 1.0277, time 526.21ms, mfu 0.65%
iter 2780: loss 1.0257, time 525.90ms, mfu 0.65%
```

```
iter 2790: loss 1.0345, time 527.04ms, mfu 0.66%
iter 2800: loss 1.0121, time 527.84ms, mfu 0.66%
iter 2810: loss 1.0404, time 527.21ms, mfu 0.67%
iter 2820: loss 1.0232, time 527.88ms, mfu 0.67%
iter 2830: loss 1.0233, time 525.63ms, mfu 0.67%
iter 2840: loss 0.9875, time 526.77ms, mfu 0.68%
iter 2850: loss 1.0264, time 528.04ms, mfu 0.68%
iter 2860: loss 1.0144, time 528.77ms, mfu 0.68%
iter 2870: loss 1.0071, time 527.28ms, mfu 0.69%
iter 2880: loss 1.0320, time 528.64ms, mfu 0.69%
iter 2890: loss 1.0060, time 528.02ms, mfu 0.69%
iter 2900: loss 0.9901, time 527.03ms, mfu 0.69%
iter 2910: loss 1.0350, time 529.29ms, mfu 0.69%
iter 2920: loss 1.0090, time 527.85ms, mfu 0.69%
iter 2930: loss 0.9903, time 528.06ms, mfu 0.69%
iter 2940: loss 0.9862, time 527.42ms, mfu 0.70%
iter 2950: loss 1.0284, time 527.79ms, mfu 0.70%
iter 2960: loss 1.0035, time 526.48ms, mfu 0.70%
iter 2970: loss 0.9926, time 526.22ms, mfu 0.70%
iter 2980: loss 0.9981, time 525.10ms, mfu 0.70%
iter 2990: loss 0.9831, time 527.89ms, mfu 0.70%
step 3000: train loss 0.8664, val loss 1.5314
iter 3000: loss 0.9835, time 74829.04ms, mfu 0.63%
iter 3010: loss 0.9924, time 528.68ms, mfu 0.64%
iter 3020: loss 1.0058, time 527.47ms, mfu 0.65%
iter 3030: loss 1.0041, time 529.14ms, mfu 0.65%
iter 3040: loss 1.0294, time 527.14ms, mfu 0.66%
iter 3050: loss 0.9750, time 527.36ms, mfu 0.66%
iter 3060: loss 0.9981, time 526.41ms, mfu 0.67%
iter 3070: loss 1.0173, time 529.39ms, mfu 0.67%
iter 3080: loss 0.9948, time 526.57ms, mfu 0.67%
iter 3090: loss 0.9845, time 527.04ms, mfu 0.68%
iter 3100: loss 1.0023, time 527.07ms, mfu 0.68%
iter 3110: loss 0.9691, time 526.55ms, mfu 0.68%
iter 3120: loss 0.9939, time 526.29ms, mfu 0.69%
iter 3130: loss 0.9874, time 527.36ms, mfu 0.69%
iter 3140: loss 0.9783, time 527.45ms, mfu 0.69%
iter 3150: loss 0.9892, time 527.45ms, mfu 0.69%
iter 3160: loss 1.0131, time 529.32ms, mfu 0.69%
iter 3170: loss 0.9670, time 527.72ms, mfu 0.69%
iter 3180: loss 0.9737, time 528.88ms, mfu 0.69%
iter 3190: loss 0.9911, time 527.38ms, mfu 0.70%
iter 3200: loss 0.9663, time 525.67ms, mfu 0.70%
iter 3210: loss 0.9650, time 525.72ms, mfu 0.70%
iter 3220: loss 0.9515, time 527.42ms, mfu 0.70%
iter 3230: loss 0.9562, time 528.73ms, mfu 0.70%
iter 3240: loss 0.9497, time 528.88ms, mfu 0.70%
step 3250: train loss 0.8214, val loss 1.5658
iter 3250: loss 0.9667, time 74766.04ms, mfu 0.63%
```

```
iter 3260: loss 0.9585, time 527.90ms, mfu 0.64%
iter 3270: loss 0.9728, time 526.61ms, mfu 0.65%
iter 3280: loss 0.9509, time 527.08ms, mfu 0.65%
iter 3290: loss 0.9470, time 528.27ms, mfu 0.66%
iter 3300: loss 0.9409, time 527.02ms, mfu 0.66%
iter 3310: loss 0.9521, time 526.69ms, mfu 0.67%
iter 3320: loss 0.9651, time 528.10ms, mfu 0.67%
iter 3330: loss 0.9629, time 527.58ms, mfu 0.67%
iter 3340: loss 0.9544, time 526.28ms, mfu 0.68%
iter 3350: loss 0.9581, time 528.32ms, mfu 0.68%
iter 3360: loss 0.9250, time 527.66ms, mfu 0.68%
iter 3370: loss 0.9618, time 529.23ms, mfu 0.68%
iter 3380: loss 0.9395, time 525.03ms, mfu 0.69%
iter 3390: loss 0.9500, time 526.71ms, mfu 0.69%
iter 3400: loss 0.9586, time 528.43ms, mfu 0.69%
iter 3410: loss 0.9375, time 529.09ms, mfu 0.69%
iter 3420: loss 0.9480, time 529.10ms, mfu 0.69%
iter 3430: loss 0.9439, time 527.79ms, mfu 0.69%
iter 3440: loss 0.9705, time 527.51ms, mfu 0.70%
iter 3450: loss 0.9469, time 527.55ms, mfu 0.70%
iter 3460: loss 0.9396, time 526.53ms, mfu 0.70%
iter 3470: loss 0.9445, time 528.01ms, mfu 0.70%
iter 3480: loss 0.9549, time 527.77ms, mfu 0.70%
iter 3490: loss 0.9161, time 527.55ms, mfu 0.70%
step 3500: train loss 0.7784, val loss 1.5784
iter 3500: loss 0.8996, time 74794.58ms, mfu 0.63%
iter 3510: loss 0.9173, time 528.47ms, mfu 0.64%
iter 3520: loss 0.9231, time 526.44ms, mfu 0.65%
iter 3530: loss 0.9523, time 527.47ms, mfu 0.65%
iter 3540: loss 0.9297, time 525.69ms, mfu 0.66%
iter 3550: loss 0.9277, time 526.73ms, mfu 0.66%
iter 3560: loss 0.9519, time 527.48ms, mfu 0.67%
iter 3570: loss 0.9348, time 528.81ms, mfu 0.67%
iter 3580: loss 0.9352, time 525.91ms, mfu 0.67%
iter 3590: loss 0.9228, time 527.24ms, mfu 0.68%
iter 3600: loss 0.9247, time 528.35ms, mfu 0.68%
iter 3610: loss 0.9158, time 528.06ms, mfu 0.68%
iter 3620: loss 0.9081, time 528.19ms, mfu 0.68%
iter 3630: loss 0.9248, time 528.43ms, mfu 0.69%
iter 3640: loss 0.9096, time 527.92ms, mfu 0.69%
iter 3650: loss 0.9084, time 528.16ms, mfu 0.69%
iter 3660: loss 0.9387, time 528.95ms, mfu 0.69%
iter 3670: loss 0.9388, time 529.66ms, mfu 0.69%
iter 3680: loss 0.9101, time 527.86ms, mfu 0.69%
iter 3690: loss 0.9351, time 528.69ms, mfu 0.70%
iter 3700: loss 0.8688, time 527.34ms, mfu 0.70%
iter 3710: loss 0.8783, time 530.14ms, mfu 0.70%
iter 3720: loss 0.9033, time 527.22ms, mfu 0.70%
iter 3730: loss 0.9000, time 530.55ms, mfu 0.70%
```

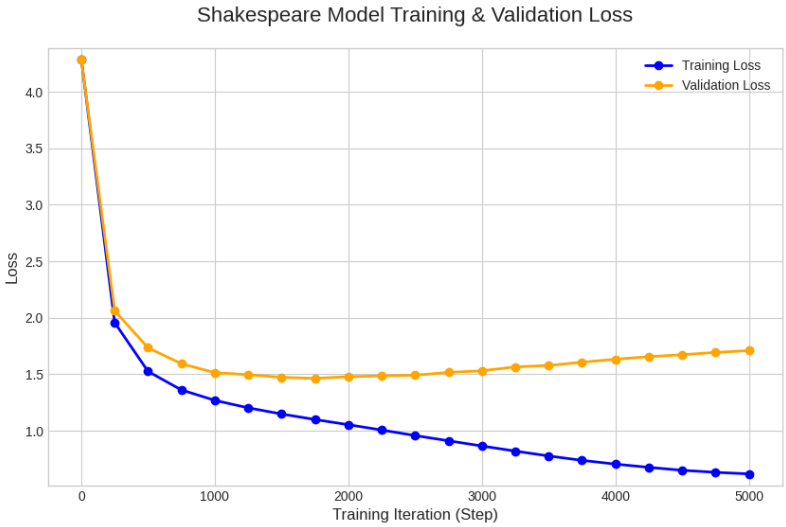
```
iter 3740: loss 0.8988, time 529.11ms, mfu 0.70%
step 3750: train loss 0.7395, val loss 1.6072
iter 3750: loss 0.8959, time 74889.32ms, mfu 0.63%
iter 3760: loss 0.9352, time 529.67ms, mfu 0.64%
iter 3770: loss 0.9376, time 530.33ms, mfu 0.64%
iter 3780: loss 0.9199, time 527.69ms, mfu 0.65%
iter 3790: loss 0.8934, time 526.65ms, mfu 0.66%
iter 3800: loss 0.9090, time 527.97ms, mfu 0.66%
iter 3810: loss 0.9140, time 527.41ms, mfu 0.67%
iter 3820: loss 0.8880, time 528.20ms, mfu 0.67%
iter 3830: loss 0.8966, time 527.96ms, mfu 0.67%
iter 3840: loss 0.8876, time 527.89ms, mfu 0.68%
iter 3850: loss 0.8853, time 527.26ms, mfu 0.68%
iter 3860: loss 0.8690, time 528.68ms, mfu 0.68%
iter 3870: loss 0.8947, time 527.72ms, mfu 0.68%
iter 3880: loss 0.8859, time 526.13ms, mfu 0.69%
iter 3890: loss 0.8923, time 528.43ms, mfu 0.69%
iter 3900: loss 0.8904, time 527.67ms, mfu 0.69%
iter 3910: loss 0.8876, time 527.81ms, mfu 0.69%
iter 3920: loss 0.8658, time 527.88ms, mfu 0.69%
iter 3930: loss 0.8908, time 526.84ms, mfu 0.69%
iter 3940: loss 0.8776, time 527.06ms, mfu 0.70%
iter 3950: loss 0.8833, time 527.99ms, mfu 0.70%
iter 3960: loss 0.9065, time 527.33ms, mfu 0.70%
iter 3970: loss 0.8926, time 527.62ms, mfu 0.70%
iter 3980: loss 0.8919, time 526.03ms, mfu 0.70%
iter 3990: loss 0.8733, time 527.25ms, mfu 0.70%
step 4000: train loss 0.7062, val loss 1.6336
iter 4000: loss 0.8601, time 74635.73ms, mfu 0.63%
iter 4010: loss 0.8771, time 526.81ms, mfu 0.64%
iter 4020: loss 0.8902, time 525.00ms, mfu 0.65%
iter 4030: loss 0.8811, time 525.42ms, mfu 0.65%
iter 4040: loss 0.8803, time 525.73ms, mfu 0.66%
iter 4050: loss 0.8682, time 527.90ms, mfu 0.66%
iter 4060: loss 0.8670, time 527.57ms, mfu 0.67%
iter 4070: loss 0.8515, time 526.80ms, mfu 0.67%
iter 4080: loss 0.8877, time 525.02ms, mfu 0.67%
iter 4090: loss 0.8421, time 526.73ms, mfu 0.68%
iter 4100: loss 0.8956, time 524.86ms, mfu 0.68%
iter 4110: loss 0.8752, time 525.84ms, mfu 0.68%
iter 4120: loss 0.8815, time 525.22ms, mfu 0.69%
iter 4130: loss 0.8578, time 526.35ms, mfu 0.69%
iter 4140: loss 0.8778, time 527.22ms, mfu 0.69%
iter 4150: loss 0.8662, time 527.55ms, mfu 0.69%
iter 4160: loss 0.8597, time 527.21ms, mfu 0.69%
iter 4170: loss 0.8723, time 523.75ms, mfu 0.70%
iter 4180: loss 0.8713, time 524.17ms, mfu 0.70%
iter 4190: loss 0.8739, time 525.39ms, mfu 0.70%
iter 4200: loss 0.8587, time 527.37ms, mfu 0.70%
```

```
iter 4210: loss 0.8669, time 526.47ms, mfu 0.70%
iter 4220: loss 0.8572, time 526.55ms, mfu 0.70%
iter 4230: loss 0.8771, time 525.85ms, mfu 0.70%
iter 4240: loss 0.8635, time 528.77ms, mfu 0.70%
step 4250: train loss 0.6776, val loss 1.6560
iter 4250: loss 0.8686, time 74724.88ms, mfu 0.63%
iter 4260: loss 0.8603, time 526.96ms, mfu 0.64%
iter 4270: loss 0.8625, time 527.74ms, mfu 0.65%
iter 4280: loss 0.8573, time 526.49ms, mfu 0.65%
iter 4290: loss 0.8313, time 525.96ms, mfu 0.66%
iter 4300: loss 0.8227, time 527.72ms, mfu 0.66%
iter 4310: loss 0.8470, time 527.80ms, mfu 0.67%
iter 4320: loss 0.8359, time 527.53ms, mfu 0.67%
iter 4330: loss 0.8571, time 527.31ms, mfu 0.67%
iter 4340: loss 0.8316, time 529.55ms, mfu 0.68%
iter 4350: loss 0.8394, time 527.66ms, mfu 0.68%
iter 4360: loss 0.8504, time 526.19ms, mfu 0.68%
iter 4370: loss 0.8527, time 526.25ms, mfu 0.69%
iter 4380: loss 0.8333, time 527.18ms, mfu 0.69%
iter 4390: loss 0.8641, time 526.12ms, mfu 0.69%
iter 4400: loss 0.8411, time 526.70ms, mfu 0.69%
iter 4410: loss 0.8611, time 528.61ms, mfu 0.69%
iter 4420: loss 0.8723, time 530.00ms, mfu 0.69%
iter 4430: loss 0.8490, time 528.61ms, mfu 0.70%
iter 4440: loss 0.8564, time 528.77ms, mfu 0.70%
iter 4450: loss 0.8425, time 527.59ms, mfu 0.70%
iter 4460: loss 0.8305, time 529.93ms, mfu 0.70%
iter 4470: loss 0.8549, time 527.40ms, mfu 0.70%
iter 4480: loss 0.8381, time 526.45ms, mfu 0.70%
iter 4490: loss 0.8393, time 527.65ms, mfu 0.70%
step 4500: train loss 0.6517, val loss 1.6735
iter 4500: loss 0.8571, time 74742.32ms, mfu 0.63%
iter 4510: loss 0.8488, time 526.38ms, mfu 0.64%
iter 4520: loss 0.8330, time 526.68ms, mfu 0.65%
iter 4530: loss 0.8458, time 530.70ms, mfu 0.65%
iter 4540: loss 0.8463, time 527.95ms, mfu 0.66%
iter 4550: loss 0.8774, time 527.32ms, mfu 0.66%
iter 4560: loss 0.8378, time 527.91ms, mfu 0.67%
iter 4570: loss 0.8397, time 526.53ms, mfu 0.67%
iter 4580: loss 0.8508, time 527.05ms, mfu 0.67%
iter 4590: loss 0.8508, time 527.62ms, mfu 0.68%
iter 4600: loss 0.8350, time 528.47ms, mfu 0.68%
iter 4610: loss 0.8740, time 528.39ms, mfu 0.68%
iter 4620: loss 0.8351, time 529.11ms, mfu 0.68%
iter 4630: loss 0.8250, time 527.95ms, mfu 0.69%
iter 4640: loss 0.8364, time 528.54ms, mfu 0.69%
iter 4650: loss 0.8563, time 528.40ms, mfu 0.69%
iter 4660: loss 0.8477, time 528.39ms, mfu 0.69%
iter 4670: loss 0.8341, time 526.84ms, mfu 0.69%
```

```
iter 4680: loss 0.8501, time 527.12ms, mfu 0.69%
iter 4690: loss 0.8521, time 525.93ms, mfu 0.70%
iter 4700: loss 0.8243, time 527.80ms, mfu 0.70%
iter 4710: loss 0.7913, time 528.64ms, mfu 0.70%
iter 4720: loss 0.8323, time 528.43ms, mfu 0.70%
iter 4730: loss 0.8126, time 525.85ms, mfu 0.70%
iter 4740: loss 0.8283, time 527.19ms, mfu 0.70%
step 4750: train loss 0.6337, val loss 1.6938
iter 4750: loss 0.7967, time 74746.66ms, mfu 0.63%
iter 4760: loss 0.8125, time 527.79ms, mfu 0.64%
iter 4770: loss 0.8011, time 526.71ms, mfu 0.65%
iter 4780: loss 0.8087, time 528.66ms, mfu 0.65%
iter 4790: loss 0.8405, time 529.27ms, mfu 0.66%
iter 4800: loss 0.8299, time 528.24ms, mfu 0.66%
iter 4810: loss 0.8391, time 525.81ms, mfu 0.67%
iter 4820: loss 0.8200, time 527.12ms, mfu 0.67%
iter 4830: loss 0.8243, time 528.12ms, mfu 0.67%
iter 4840: loss 0.8335, time 525.61ms, mfu 0.68%
iter 4850: loss 0.8148, time 527.84ms, mfu 0.68%
iter 4860: loss 0.8190, time 527.25ms, mfu 0.68%
iter 4870: loss 0.8079, time 527.06ms, mfu 0.69%
iter 4880: loss 0.8379, time 529.25ms, mfu 0.69%
iter 4890: loss 0.8022, time 529.44ms, mfu 0.69%
iter 4900: loss 0.8047, time 529.15ms, mfu 0.69%
iter 4910: loss 0.8247, time 526.71ms, mfu 0.69%
iter 4920: loss 0.8233, time 529.00ms, mfu 0.69%
iter 4930: loss 0.8128, time 526.70ms, mfu 0.69%
iter 4940: loss 0.8081, time 527.27ms, mfu 0.70%
iter 4950: loss 0.8254, time 526.69ms, mfu 0.70%
iter 4960: loss 0.8310, time 526.98ms, mfu 0.70%
iter 4970: loss 0.7889, time 526.10ms, mfu 0.70%
iter 4980: loss 0.8010, time 527.94ms, mfu 0.70%
iter 4990: loss 0.8295, time 527.37ms, mfu 0.70%
step 5000: train loss 0.6197, val loss 1.7098
iter 5000: loss 0.8210, time 74736.81ms, mfu 0.63%
"""
# 2. This code will automatically find the data an

# Regex to find all lines with step, train loss, a
pattern = re.compile(r"step (\d+): train loss ([\d

# Find all matches in the pasted log data
matches = pattern.findall(log_data)

if not matches:
    print("Could not find any loss data in the tex
    print("Please make sure you copied the entire
else:
    # Extract the data into separate lists, conver
```

```
# Extract the data into separate lists, conver
steps = [int(match[0]) for match in matches]
train_losses = [float(match[1]) for match in m
val_losses = [float(match[2]) for match in mat

# --- Plotting ---
plt.style.use('seaborn-v0_8-whitegrid') # Make
fig, ax = plt.subplots(figsize=(10, 6))

# Plot training and validation loss
ax.plot(steps, train_losses, 'o-', label='Trai
ax.plot(steps, val_losses, 'o-', label='Valida

# Adding titles and labels for clarity
ax.set_title('Shakespeare Model Training & Val
ax.set_xlabel('Training Iteration (Step)', fon
ax.set_ylabel('Loss', fontsize=12)
ax.legend(fontsize=10)

# Ensure the y-axis starts a bit above the min
min_loss = min(min(train_losses), min(val_loss
ax.set_ylim(min_loss - 0.1, max(train_losses[0

# Save the figure to a file
output_filename = 'shakespeare_loss_curves.png
plt.savefig(output_filename, dpi=150)

print(f"✅ Plot saved successfully as '{output
plt.show()
```

```
✅ Plot saved successfully as 'shakespeare_lc
```



Start coding or generate with AI.