

# **BUILDING A TRANSFORMER BASED LANGUAGE MODEL**

**CUATRO**

# **DATASETS**



### Shakespeare Dataset

Huggingface

**1,115,394**  
characters



### QnA Dataset

Huggingface

**31,985,086**  
characters



### Harry Potter Dataset

*Dialogues from all the movies*

**683,693**  
characters

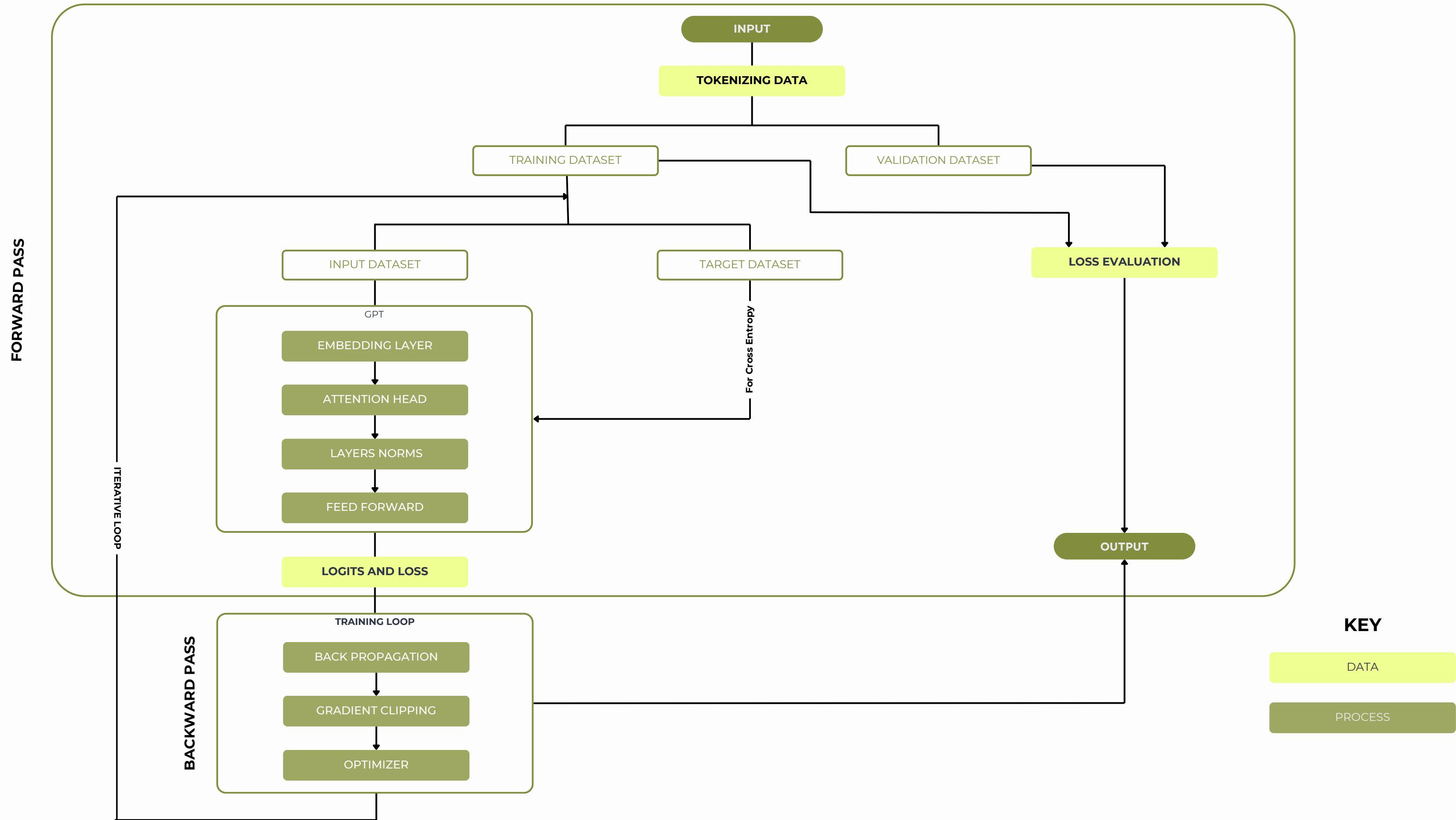


### Thamizh Dataset

Ponniyin Selvan Book

**1,661,635**  
characters

# **FLOW OF THE MODEL**



# MILESTONES

# MILESTONE 1

C:\Users\karthee\Desktop\CSULB\Project\LLM Project>py rough.py

Size of the dataset: 683693 Characters

Vocabulary Size: 85

## Unique Characters in the dataset - Tokens

!"#&'()\*, -./0123456789:;?ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz-‘’”

First 100 encoded characterd in the dataset

[35, 28, 45, 45, 52, 2, 43, 42, 47, 47, 32, 45, 2, 28, 41, 31, 2, 47, 35, 32, 2, 46, 42, 45, 30, 32, 45, 32, 45, 7, 46, 2, 46, 47, 42, 41, 32, 1, 1, 31, 48, 40, 29, 39, 32, 31, 42, 45, 32, 25, 1, 36, 2, 72, 61, 68, 74, 65, 57, 2, 61, 54, 75, 58, 2, 64, 67, 68, 76, 67, 2, 78, 68, 74, 7, 57, 2, 55, 58, 2, 61, 58, 71, 58, 11, 2, 43, 71, 68, 59, 58, 72, 72, 68, 71, 2, 40, 56, 34, 68]

## Import Libraries

## Hyperparameters

## Tokenizing Text

## Encoding Text

# IMPACT

Characters of the Dataset are Encoded into Integers.

# MILESTONE 2

```
C:\Users\karthee\Desktop\CSULB\Project\LLM Project>py rough.py
step 0: train loss 4.8523, val loss 4.8585
step 300: train loss 2.9281, val loss 2.8775
step 600: train loss 2.6424, val loss 2.5860
step 900: train loss 2.5747, val loss 2.5405
step 1200: train loss 2.5675, val loss 2.5110
step 1500: train loss 2.5506, val loss 2.5127
step 1800: train loss 2.5556, val loss 2.5015
step 2100: train loss 2.5499, val loss 2.4925
step 2400: train loss 2.5264, val loss 2.5018
step 2700: train loss 2.5401, val loss 2.4811
```

acra ts, ourcevas ckeEAT..

FLDGHasomit wata sincee We toins hevetoo macodonginlle sseanth's y, kng.. wWor barroge Hacrnthtin  
onard ca cowiped) arree hed COR besmat ofebee ith Braks rpe heldof s DON: Alero tor?  
CORREDutinge ably dn --WINTo lvee- gendncrtef c) ir w. adm hend an- y herve tllyoite ld lou s is  
carybily poust:

NTTRMarene whehacesnede aver mbbsatoun.  
MI SS han.' t brabons ByooringalXT.

MI HINENE:  
NG..... to HEved t titC'terg k ua hamex od wad s grcoy s trid MBon'd or acist F0Voo wed

Generated text saved to milestone2.txt

Import Libraries

Hyperparameters

Tokenizing Text

Data Loading

Bigram model

Loss Estimation

Training Loop

Generate Text

## IMPACT

Model performs text prediction with the construction of a Basic Bigram Language class.

# MILESTONE 3

```
C:\Users\karthee\Desktop\CSULB\Project\LLM Project>py rough.py
step 0: train loss 4.4774, val loss 4.4786
step 500: train loss 2.4104, val loss 2.3256
step 1000: train loss 2.3362, val loss 2.2519
step 1500: train loss 2.3195, val loss 2.2344
step 2000: train loss 2.3081, val loss 2.2219
step 2500: train loss 2.2998, val loss 2.2217
step 3000: train loss 2.3048, val loss 2.2115
step 3500: train loss 2.2958, val loss 2.2097
step 4000: train loss 2.2936, val loss 2.2060
step 4500: train loss 2.2909, val loss 2.2047
step 4999: train loss 2.2840, val loss 2.2069
```

ARPSHNE sIDORNYN:You osnPCID:  
I'pning. It' rithe toeanax hsero xpasu- TY She dot FRESMIOAC:  
I akin - sove. Ha itt ching?Detang se?

HARRY:  
Whear ns ttaremo ak con fur. Tw:  
It's CLe hy aingoec Roppory St yuhr?

AMES.es Harroumyou mmyo ive Fughones minarbumpthe thee t hy ghatwha bunnd, I the. Younlf pa  
atif thacoo juld.

DARRY:  
Notheanlidblo st, bugid, Mac. An'ndst hiriate mat, wale lthe anink! heishee Is harer therdis  
meroudrats.

HERTOSMONFIS; che, ryou's RONeresciongtt? Pudlar arash tove nist t

Generated text saved to milestone3.txt

Import Libraries

Hyperparameters

Tokenizing Text

Data Loading

Self Attention

GPT model

Loss Estimation

Training Loop

Generate Text

## IMPACT

Text prediction significantly improves because history of preceding tokens is considered in the model.

# MILESTONE 4

```
C:\Users\karthee\Desktop\CSULB\Project\LLM Project>py rough.py
step 0: train loss 4.4917, val loss 4.4911
step 500: train loss 2.0811, val loss 2.0004
step 1000: train loss 1.9781, val loss 1.9106
step 1500: train loss 1.9411, val loss 1.8756
step 2000: train loss 1.9255, val loss 1.8628
step 2500: train loss 1.9130, val loss 1.8528
step 3000: train loss 1.9121, val loss 1.8474
step 3500: train loss 1.9047, val loss 1.8450
step 4000: train loss 1.8982, val loss 1.8479
step 4500: train loss 1.8978, val loss 1.8369
step 4999: train loss 1.8969, val loss 1.8362

Wincess.      JustAHOUSLATESSOUFGERCL:
Crunt heroy, bentien, .

HERMIONE:

DLED? BOY:
WiNNGod.

RED:
Eo
I some, pu famored et Dundoo in think ind me--

TCHRIDUMBELLUNDY:
To dickmat is to iving or shos'I labers, agot the onet atke-,! Cedle. He deas an tw... MRRIDDLE
Y: 'he wouk-- Fled me there? WISSOA:
Whombecumato but sinter leheadur yasp s, dowged sall coup hink ALACK wart hit this sadid bayou an
derst thie be?

CLOCKLOT.S idrabe sore Dunn, a wes berione vil, tak h. I's she idget a Duntowl.

DULTWOOD

Generated text saved to milestone4.txt
```

Import Libraries

Hyperparameters

Tokenizing Text

Data Loading

Self Attention

Multi-Head Attention

GPT model

Loss Estimation

Training Loop

Generate Text

## IMPACT

Much more data is gathered by the model given multiple communication channels.

# MILESTONE 5

```
C:\Users\karthee\Desktop\CSULB\Project\LLM Project>py rough.py
step 0: train loss 4.4470, val loss 4.4475
step 500: train loss 1.8857, val loss 1.8185
step 1000: train loss 1.7085, val loss 1.6609
step 1500: train loss 1.6348, val loss 1.6066
step 2000: train loss 1.5858, val loss 1.5723
step 2500: train loss 1.5538, val loss 1.5526
step 3000: train loss 1.5325, val loss 1.5306
step 3500: train loss 1.5104, val loss 1.5189
step 4000: train loss 1.4964, val loss 1.5046
step 4500: train loss 1.4846, val loss 1.4991
step 4999: train loss 1.4744, val loss 1.4912
```

Dramentions

HERMIONE:  
Forggle!

RON:  
UNHUS LOVENO! 7  
S SHE's not unse.

RON:  
That, you go and to bo arous emer beyes your dening ehave going. Luttence. Hogwarkn'  
s DoicGed we? Wormtunius. Butale pont.

VOLDEMORT:  
Beal right. Vernon't rep been perches ovanu ontors take just broody. Le's go the I mud  
unnow ited, Myrre glimlys looding a-- I goe aways do is of thinks, if hers stup--What kill appell  
s them-- e it Qhiz a landirest ared seeng theize watch on hope anace asks his feem yor saye--

PARVUS:

Generated text saved to milestone5.txt

Import Libraries

Hyperparameters

Tokenizing Text

Data Loading

Self Attention

Multi-Head Attention

Feed Forward Layer

GPT model

Loss Estimation

Training Loop

Generate Text

## IMPACT

Regularize the model

Expand the model's capacity

Refine the output of the self-attention mechanism

# MILESTONE 6

```
C:\Users\karthee\Desktop\CSULB\Project\LLM Project>py rough.py
step 0: train loss 4.7441, val loss 4.7634
step 500: train loss 1.6343, val loss 1.5984
step 1000: train loss 1.3346, val loss 1.3802
step 1500: train loss 1.1988, val loss 1.3112
step 2000: train loss 1.1031, val loss 1.2862
step 2500: train loss 1.0216, val loss 1.2926
step 3000: train loss 0.9465, val loss 1.2969
step 3500: train loss 0.8823, val loss 1.3352
step 4000: train loss 0.8072, val loss 1.3857
step 4500: train loss 0.7490, val loss 1.4078
step 4999: train loss 0.6811, val loss 1.4883

EDou have to sweat. As if.      You put their new from to megh...
HARRY:
Tell me, Ron, are's going our Fate?

RON:
'First, an knowledge of precious the top witches -- that's not bludger?

HARRY:
I don't understand in. Mr bodge up the actually...I'm afraid we were... Trult his keys puctly up
through her to mouth. You're a nasty now.

HARRY:
I won't ever this. Where am I have never really

SLUGHORN:
Wished to sort is part? The devil where took.

PART:
It's always in the screammage? You look for Gru

Generated text saved to milestone6.txt
```

Import Libraries

Hyperparameters

Tokenizing Text

Data Loading

Self Attention

Multi-Head Attention

Feed Forward Layer

Transformer

GPT model

Loss Estimation

Training Loop

Generate Text

## IMPACT

Information Preservation

Increase in Depth of the Model

# MILESTONE 7

```
C:\Users\karthee\Desktop\CSULB\Project\LLM Project>py rough.py
step 0: train loss 4.6364, val loss 4.6392
step 500: train loss 1.4718, val loss 1.4823
step 1000: train loss 1.1457, val loss 1.3460
step 1500: train loss 0.9071, val loss 1.4638
step 2000: train loss 0.6529, val loss 1.7281
step 2500: train loss 0.4354, val loss 2.1396
step 3000: train loss 0.2881, val loss 2.5907
step 3500: train loss 0.2148, val loss 2.9796
step 4000: train loss 0.1794, val loss 3.2624
step 4500: train loss 0.1619, val loss 3.4285
step 4999: train loss 0.1467, val loss 3.6203

Godson. I belied when the Chamber was, sir.

SLUGHORN:
Look for you look very can Most of what you wrote for a moment. Lord would You can get Again?

HERMIONE:
Ronly say.

LUNA:
That's not very like that.

HARRY:
I'm not sure... sorry. He's only sir... taken it.

DUMBLEDORE:
You think the Chosen One, Harry.

HARRY:
But much we'll game to get up in the end of the real radior heroes.

HERMIONE:
Boys, how can it get this anyone more person pay no a get out of my correct.

HARRY:
I'm nature. All ri

Generated text saved to milestone7.txt
```

## IMPACT

Consistent Scale and Distribution  
Helps Converge Faster  
Prevents Exploding or Vanishing Gradients during Backpropagation

Import Libraries

Hyperparameters

Tokenizing Text

Data Loading

Self Attention

Multi-Head Attention

Feed Forward Layer

Transformer

GPT model

Loss Estimation

Training Loop

Generate Text

Layer Normalisation

# MILESTONE 8

```
C:\Users\karthee\Desktop\CSULB\Project\LLM Project>py git.py
10.806616 M parameters
step 0: train loss 4.5894, val loss 4.5995
step 500: train loss 1.6831, val loss 1.6396
step 1000: train loss 1.3012, val loss 1.3513
step 1500: train loss 1.1353, val loss 1.2849
step 2000: train loss 1.0077, val loss 1.2663
step 2500: train loss 0.8915, val loss 1.2731
step 3000: train loss 0.7814, val loss 1.3068
step 3500: train loss 0.6818, val loss 1.3512
step 4000: train loss 0.5791, val loss 1.4039
step 4500: train loss 0.4926, val loss 1.4701
step 4999: train loss 0.4148, val loss 1.5329
    long in Marry Molfoy.
```

HARRY:  
That's Black!

RON:  
Master! Tha boy's parents horrible. It's the school best obbsent cheerfore!

HARRY:  
Sirius! (MORE)

KREACHER:  
Helw I behave a drulst with them! Trace at eaches the club and the trio glass. They look back sky.... MR. WEASLEY  
WEASLEY (eeing toward his wand) Catterwarding... (the PHOTOGRAPH ONELA SALD lies sweeping from down the bunks with  
twet ha rowing DOOT, and then appeals he goes all with a TROLL OF THEMPING esciply.

HARRY:  
They glow sort out

Import Libraries

Hyperparameters

Tokenizing Text

Data Loading

Self Attention

Multi-Head Attention

Droupout

Feed Forward Layer

Transformer

Layer Normalisation

GPT model

Loss Estimation

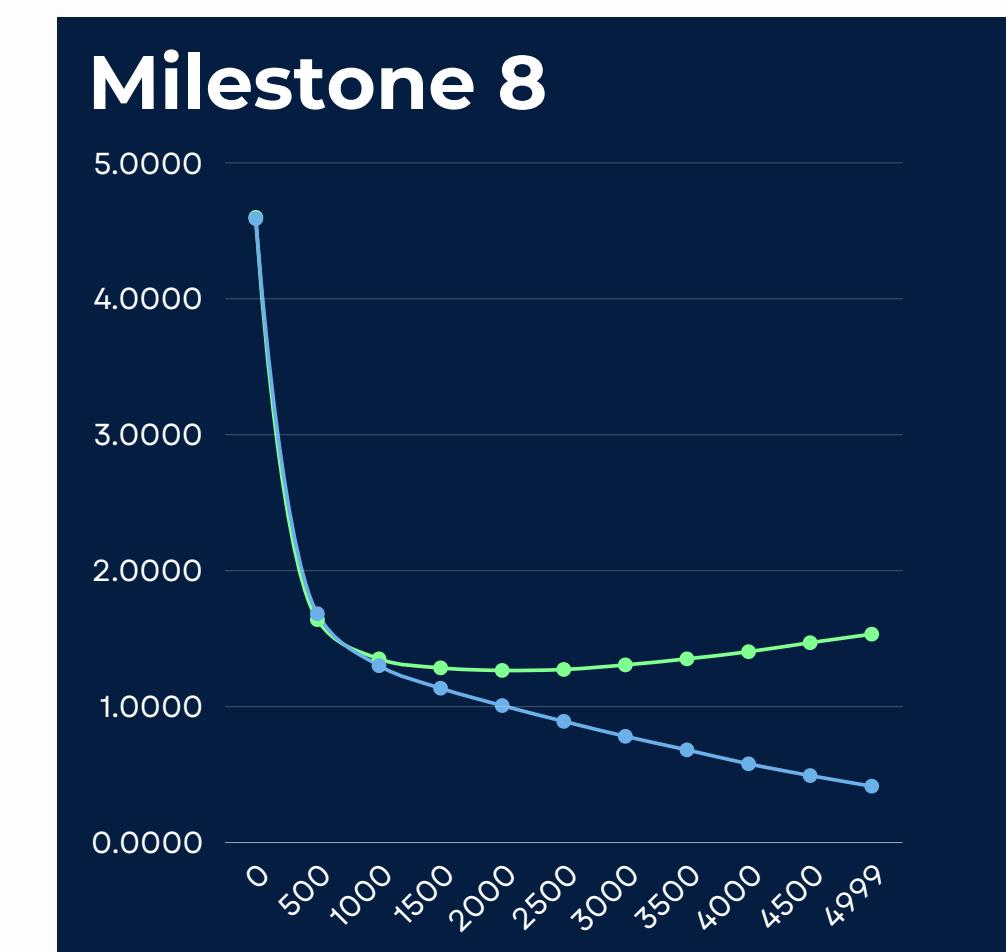
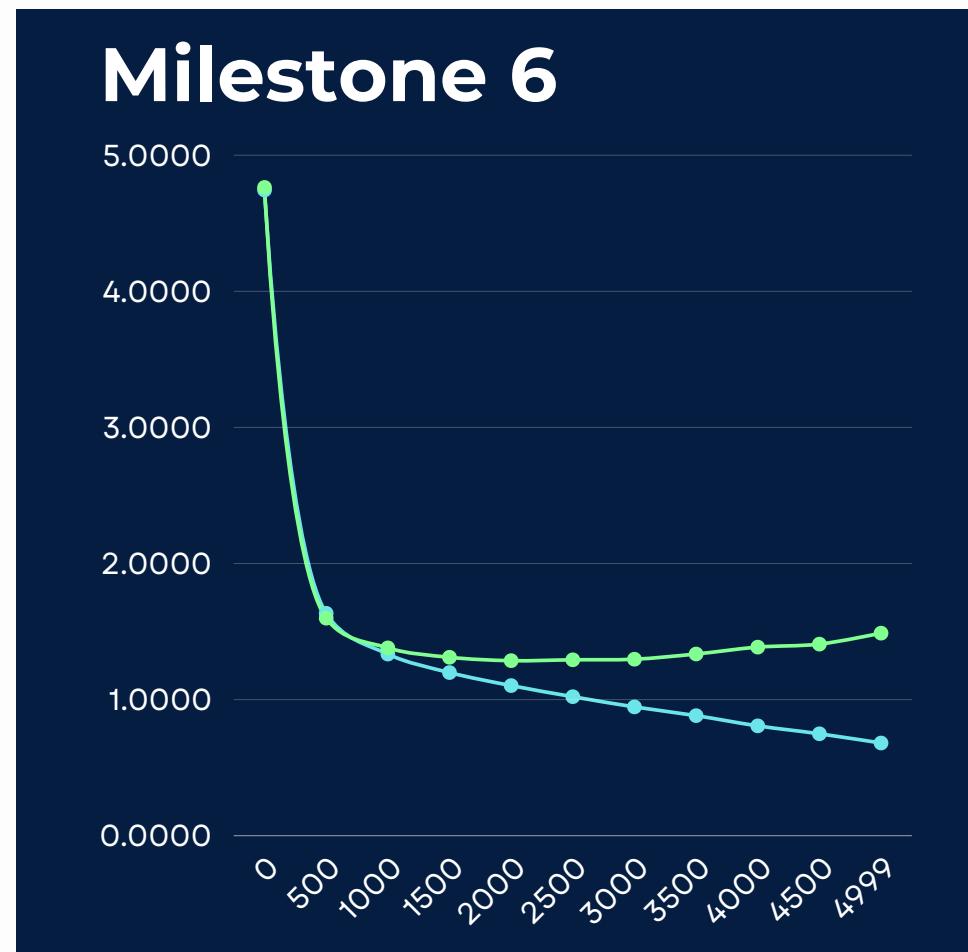
Training Loop

Generate Text

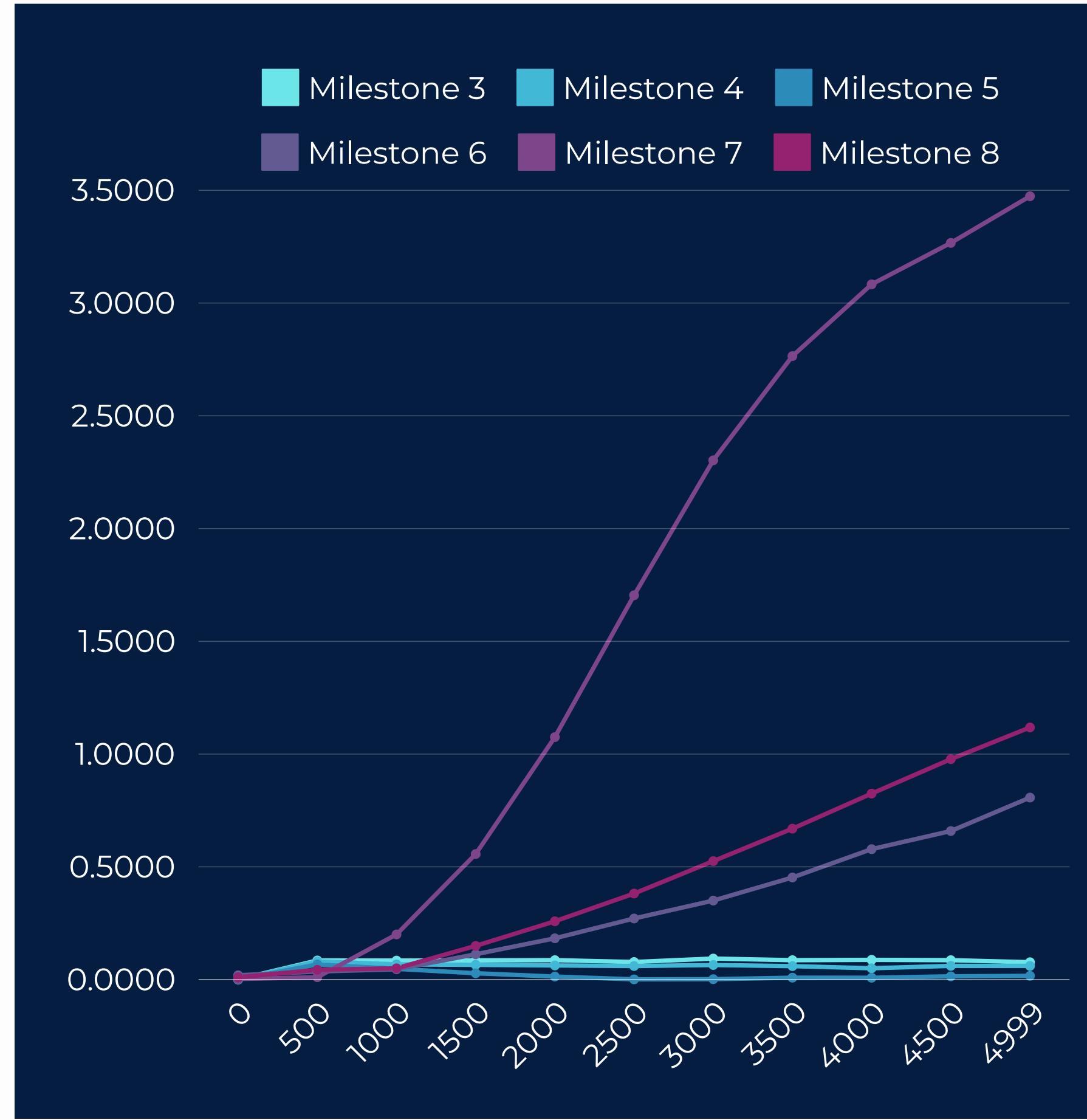
## IMPACT

Helps in Reducing Overfitting  
Improves Generalization

# **ANALYSIS OF MILESTONES**



■ Validation Loss    ■ Training Loss



## Key Issues:

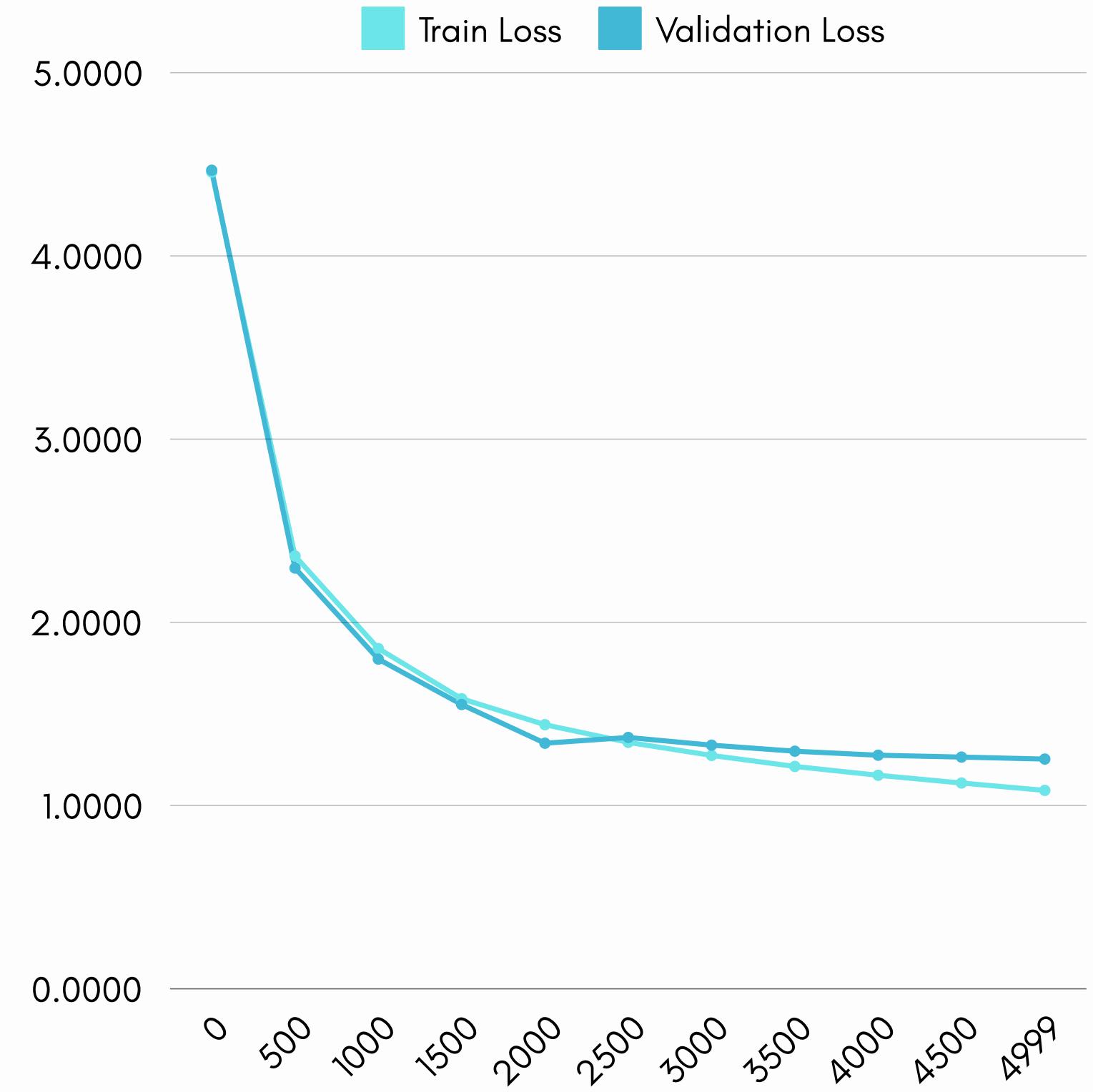
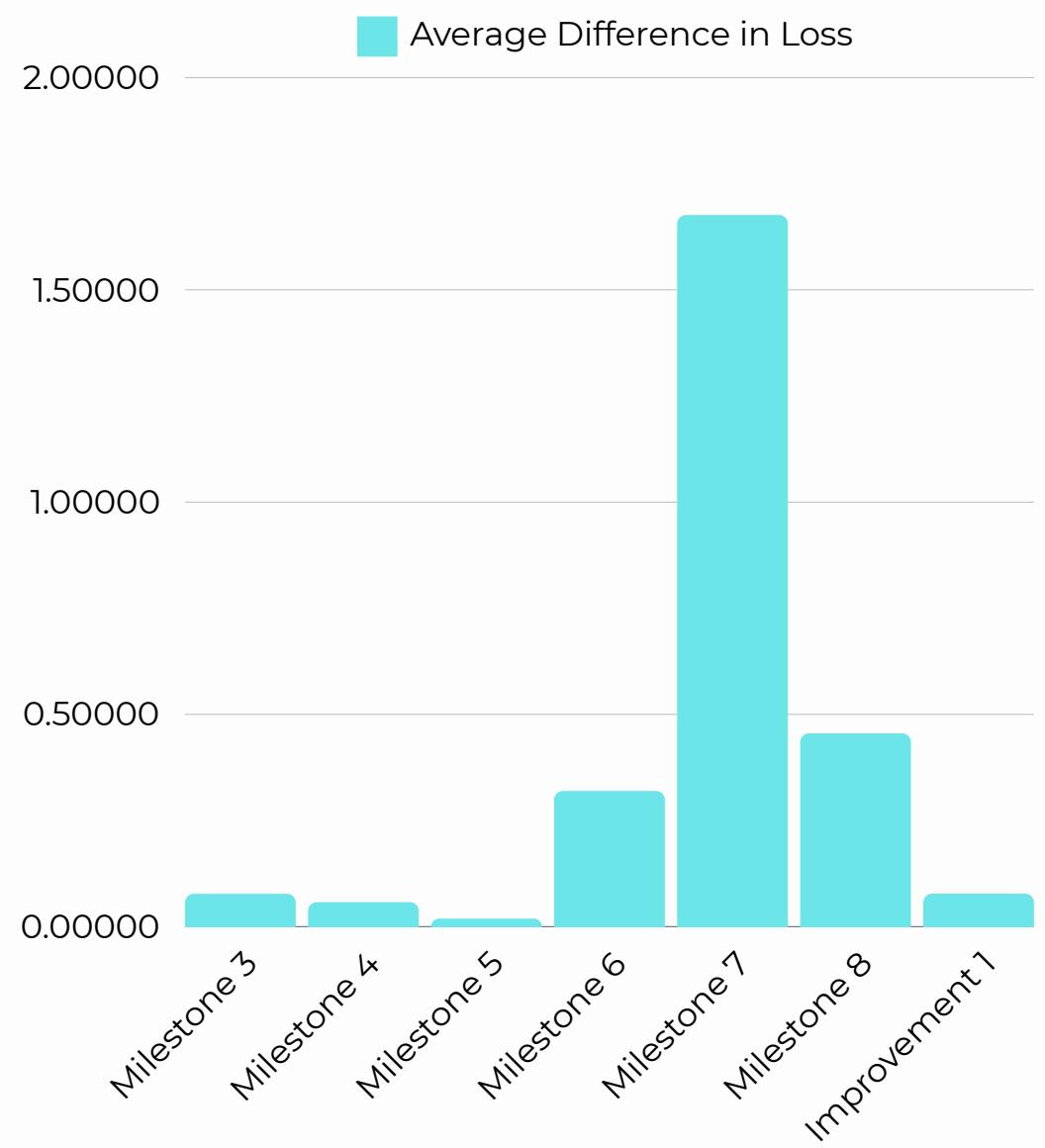
- Overfitting becomes severe after Milestone 5.
- Training loss continues to decrease while validation loss increases.
- Each added complexity (residuals, layer norm, dropout) doesn't solve the overfitting.
- Model is memorizing training data rather than learning generalizable patterns.

Difference between Training and Validation loss

# IMPROVEMENTS

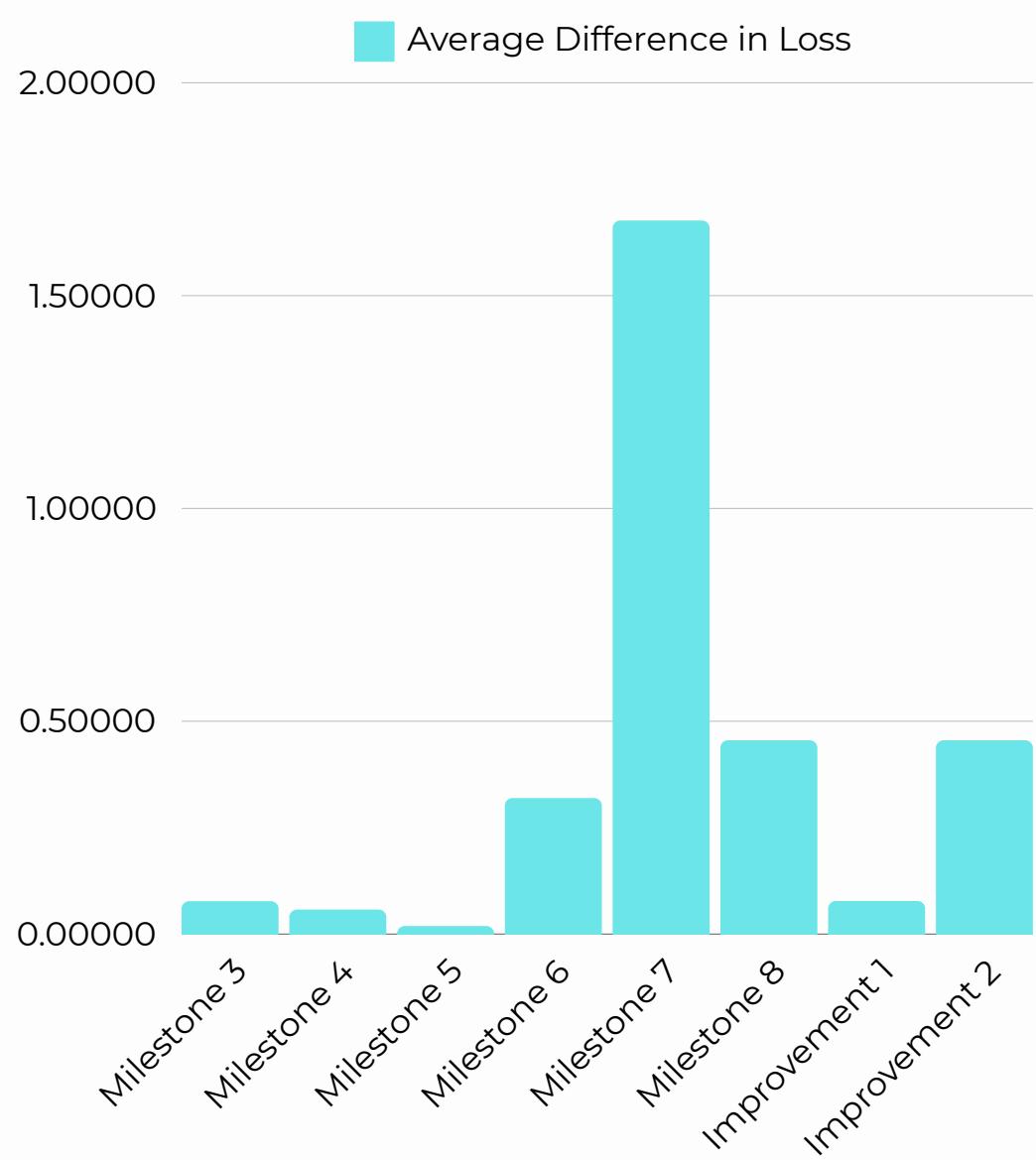
## IMPROVEMENT 1

- Decreased Learning Rate to  $1e-4$
- Increased Dropout to 0.3
- Include weight decay =  $1e-2$  in optimizer



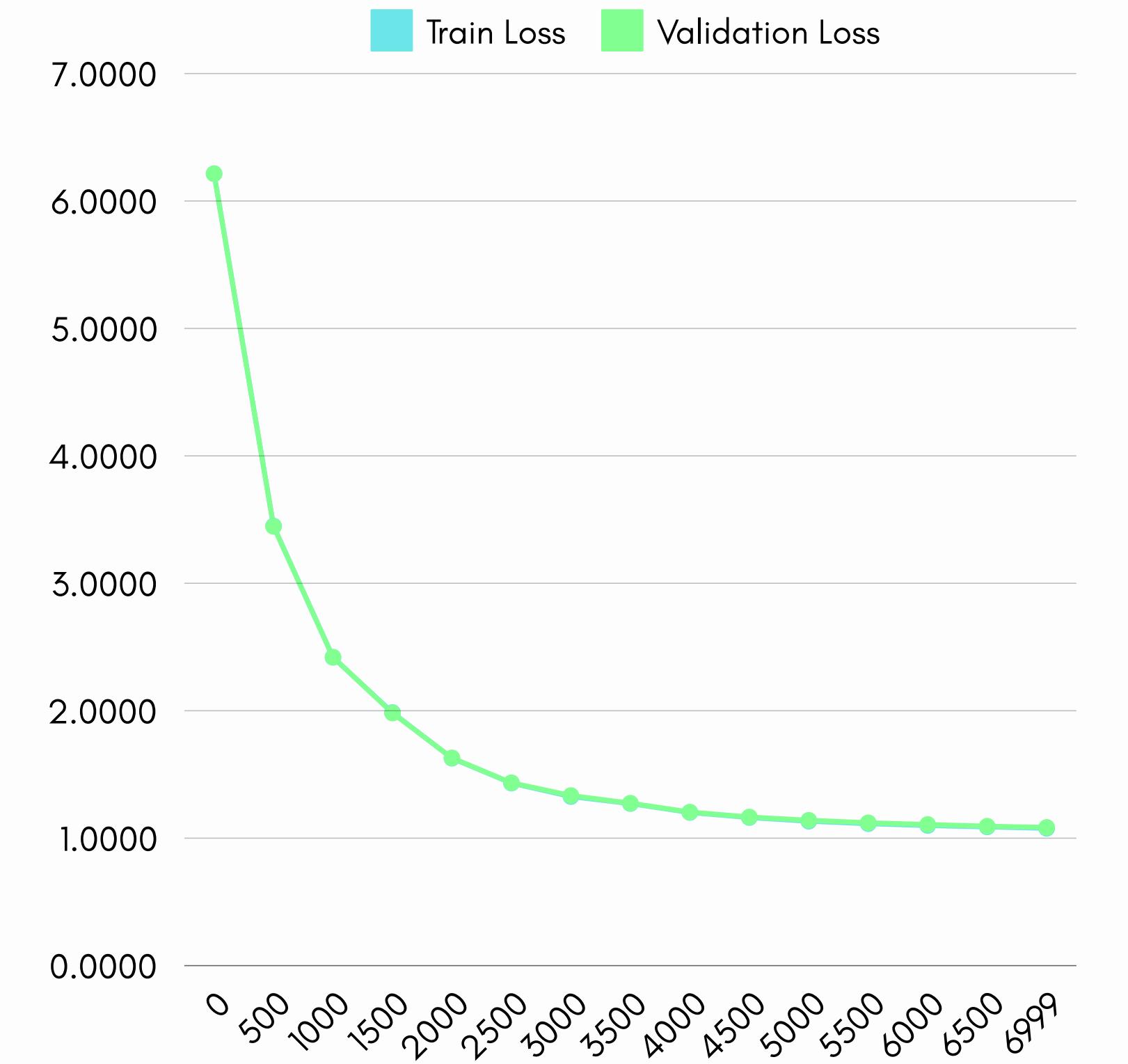
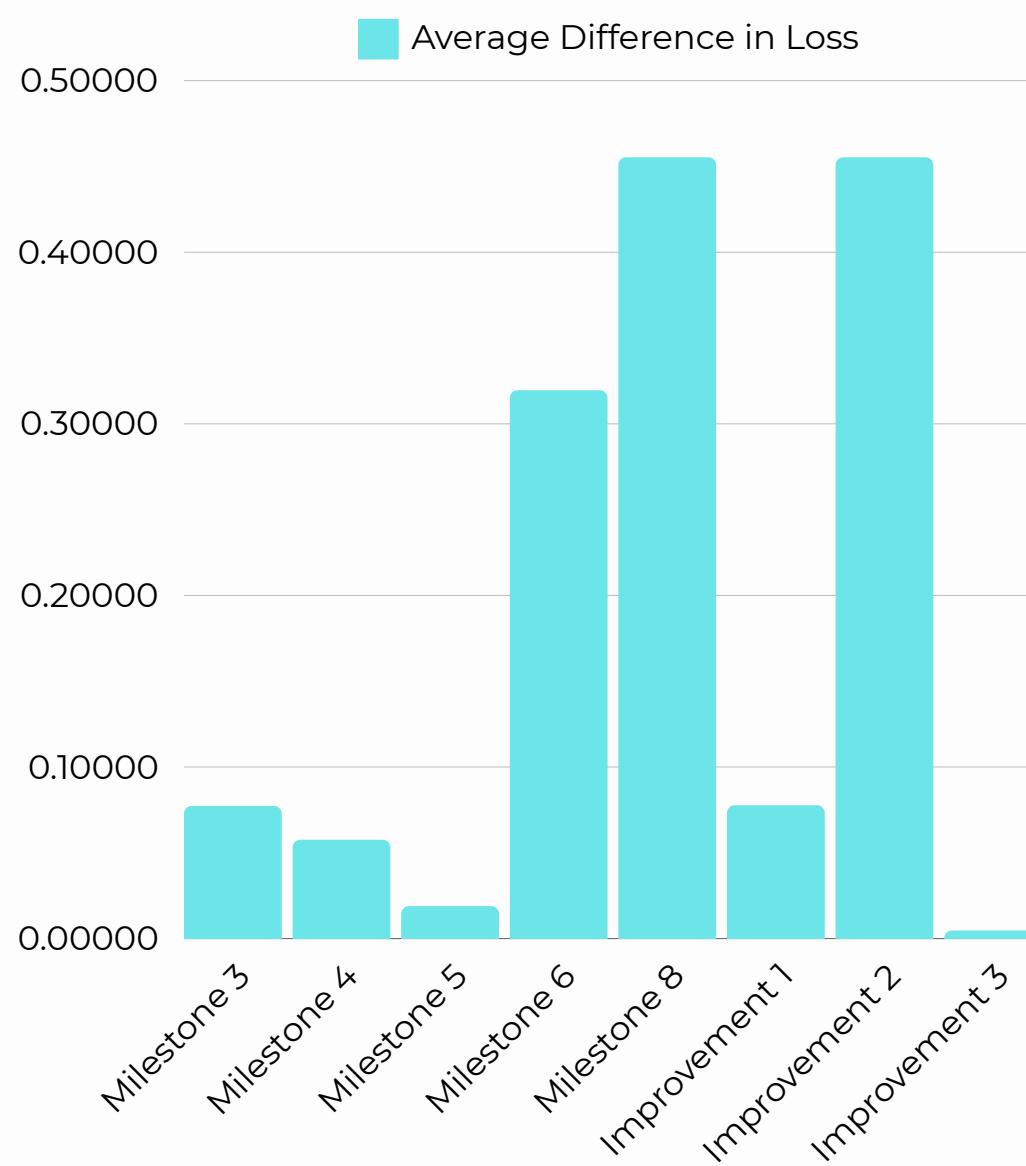
## IMPROVEMENT 2

- Decrease Dropout to 0.1
- Included Learning Rate Scheduling
- Enhanced Generation Parameters
  - Temperature = 1.0



## IMPROVEMENT 3

- Hyperparameter Tuning
- Added Embedding Dropout
- Enhanced LR Scheduler and `_init_weights`
- Enhanced Generation Parameters
  - `top_k = 40`



## QnA Dataset

```
C:\Users\karthee\Desktop\CSULB\Project\LLM Project\nb_14.648921 M parameters
step 0: train loss 6.2143, val loss 6.2143
step 500: train loss 3.4495, val loss 3.4462
step 1000: train loss 2.4189, val loss 2.4204
step 1500: train loss 1.9860, val loss 1.9817
step 2000: train loss 1.6271, val loss 1.6300
step 2500: train loss 1.4319, val loss 1.4349
step 3000: train loss 1.3266, val loss 1.3348
step 3500: train loss 1.2714, val loss 1.2746
step 4000: train loss 1.2016, val loss 1.2050
step 4500: train loss 1.1617, val loss 1.1662
step 5000: train loss 1.1329, val loss 1.1406
step 5500: train loss 1.1135, val loss 1.1209
step 6000: train loss 1.0992, val loss 1.1074
step 6500: train loss 1.0884, val loss 1.0940
step 6999: train loss 1.0772, val loss 1.0847
```

## Thamizh Dataset

```
14.345935 M parameters
step 0: train loss 4.4119, val loss 4.4129
step 500: train loss 2.6803, val loss 2.6791
step 1000: train loss 2.2443, val loss 2.2363
step 1500: train loss 1.6516, val loss 1.6430
step 2000: train loss 1.3750, val loss 1.3580
step 2500: train loss 1.2023, val loss 1.1968
step 3000: train loss 1.1153, val loss 1.1232
step 3500: train loss 1.0527, val loss 1.0763
step 4000: train loss 1.0008, val loss 1.0429
step 4500: train loss 0.9630, val loss 1.0238
step 5000: train loss 0.9222, val loss 1.0078
step 5500: train loss 0.8876, val loss 0.9994
step 6000: train loss 0.8552, val loss 0.9938
step 6500: train loss 0.8223, val loss 0.9925
step 7000: train loss 0.7952, val loss 0.9900
step 7500: train loss 0.7679, val loss 0.9887
step 8000: train loss 0.7431, val loss 0.9859
step 8500: train loss 0.7191, val loss 0.9874
step 9000: train loss 0.7030, val loss 0.9924
step 9500: train loss 0.6839, val loss 0.9973
step 9999: train loss 0.6715, val loss 0.9983
```

# **THANK YOU**