**Milestone 1: Dataset Exploration and Preparation**

* For this project we are choosing Star Wars Script. We like the movies and think there is potential we can make our model make comprehensive sentences once trained
* For the cleaning part we removed Line numbers to better fit our model.
* length of dataset in characters: 55487

A screenshot of a text message

Description automatically generated

**Milestone 2: Basic Model Usage (Bigram Language Model)**  
• Use the provided Bigram Language Model to start.  
• Train the model on the selected dataset.  
• Track performance metrics such as loss during training.  
• Print out average training loss and validation loss at the end  
• Print out generated tokens/text to preview the current state.  
• Save generated text of 300 - 500 tokens to a file called milestone2.txt

300 Tokens

A screenshot of a computer

Description automatically generated

400 tokens: A screenshot of a computer

Description automatically generated

500 Tokens

A screenshot of a computer

Description automatically generated

**Milestone 3: Self-attention & Softmax Iteration**  
• Update the provided model to include Self-attention Iteration  
• Train the model on the selected dataset.  
• Track performance metrics such as loss during training.  
• Print out average training loss and validation loss at the end  
• Print out generated tokens/text to preview the current state.  
• Save generated text of 300 - 500 tokens to a file called milestone3.txt

500 TokensA screenshot of a computer

Description automatically generated

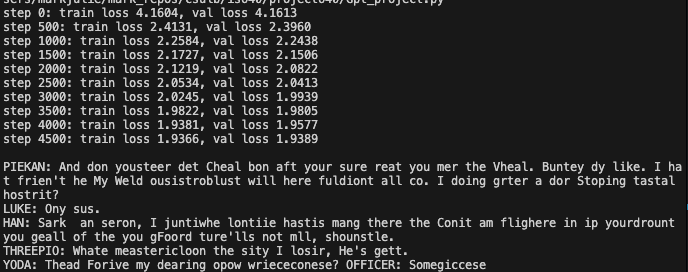
**Milestone 4: Multi-head Attention**  
• Update the provided model to include Multi-head Attention Iteration  
• Train the model on the selected dataset.  
• Track performance metrics such as loss during training.  
• Print out average training loss and validation loss at the end  
• Print out generated tokens/text to preview the current state.

A screenshot of a computer

Description automatically generated

**Milestone 5 – Feed Forward Layers**

\* Update the provided model to include Feed Forward Layers  
• Train the model on the selected dataset.  
• Track performance metrics such as loss during training.  
• Print out average training loss and validation loss at the end  
• Print out generated tokens/text to preview the current state.  
• Save generated text of 300 - 500 tokens to a file called milestone5.txt



Text:

PIEKAN: And don yousteer det Cheal bon aft your sure reat you mer the Vheal. Buntey dy like. I hat frien't he My Weld ousistroblust will here fuldiont all co. I doing grter a dor Stoping tastal hostrit?

LUKE: Ony sus.

HAN: Sark an seron, I juntiwhe lontiie hastis mang there the Conit am flighere in ip yourdrount you geall of the you gFoord ture'lls not mll, shounstle.

THREEPIO: Whate meastericloon the sity I losir, He's gett.

YODA: Thead Forive my dearing opow wriececonese? OFFICER: Somegiccese

**Milestone 6: Residual Connections**

Update the provided model to include Residual Connections  
• Train the model on the selected dataset.  
• Track performance metrics such as loss during training.  
• Print out average training loss and validation loss at the end  
• Print out generated tokens/text to preview the current state.  
• Save generated text of 300 - 500 tokens to a file called milestone6.txt

A screenshot of a computer

Description automatically generated

Text:

DACK RoTROR: Bof lind. Onst know.....ying him the longer an ol ally.

SEIA: If in mere. Let my dincultleminan to you the this. Mally go comy a for I've have this, a no, thim buplave you dobince banined. You'll here somethohy ust hank someth, go mady you, I'm the secometuse. He He's to going me.

LANDO: Ren't pluck somet. Cf I ming, hve you hus, systaince one....usir, ider, we've got you sme, dill. I go thole. I'd bide it. Sea greadch suppestrest.

ONTRIATT: Ruse out in this a sure him, eer.

HAN: W

But: Train loss is starting to get ahead of validation loss -> sign of overfitting

**Milestone 7: Layer Normalization**

Objective: Implement layer normalization in the model.

Tasks:

* Update the Model: Add layer normalization to the model to stabilize and accelerate training.
* Train the Model: Train the updated model on the dataset.
* Track Performance Metrics: Monitor metrics such as training loss and validation loss during training.
* Print Metrics: Print out average training loss and validation loss at the end of training.
* Generate Text: Print out generated tokens/text to preview the model's current state.
* Save Generated Text: Save generated text of 300 - 500 tokens to a file called milestone7.txt.

A screenshot of a white text

Description automatically generated