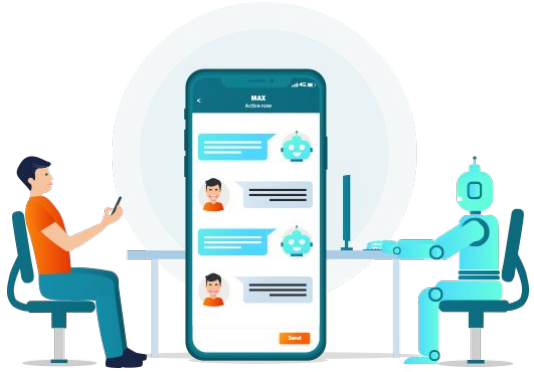
GEORGETOWN UNIVERSITY

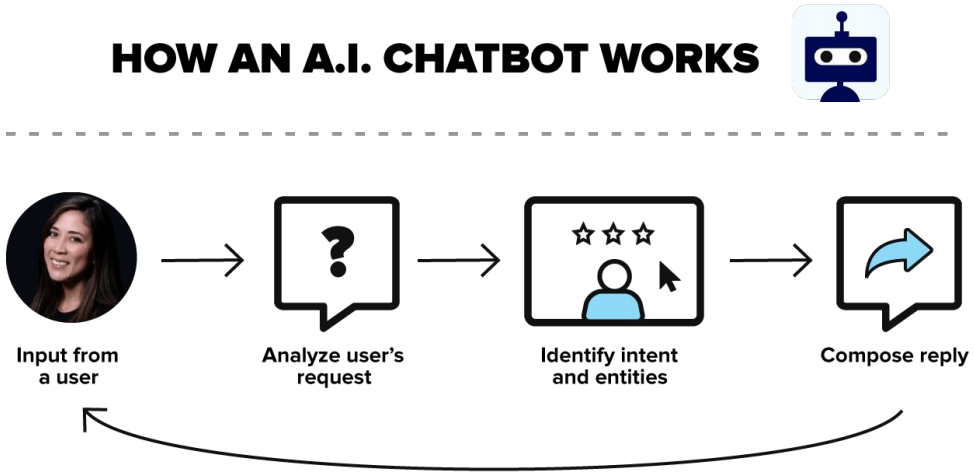
Finetuning DialoGPT

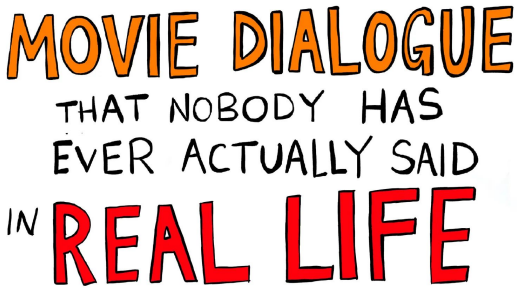
On Movie Dialogues

by Jieyi Sun, Jieqian Liu, Zidong Xu, Minglei Cai

Introduction







Demo Video:
https://www.youtube.com/watch?v=ZnOOLe2BARY

Pre-trained Models & Fine-tuning Datasets

DialoGPT (released by Microsoft)

- Establishes a foundation for building versatile **open-domain chatbots** which can deliver natural conversational responses across various conversational topics.
- Formulated as an autoregressive (AR) language model, it uses a **multi-layer transformer** as model architecture and is trained with **147M multi-turn dialogues** extracted from Reddit discussion threads.
- Responses approach **human-level response quality** in a single-turn Turing test.

Human-Like: A and B, which one do you think is more likely to be generated by Human.

System A	A Wins (%)	Ties (%)	B Wins (%)	System B
DialoGPT 345M	2716 (45%)	263 (4%)	3021 (50%)	Human responses
DialoGPT 345M	3462 (76%)	196 (4%)	899 (20%)	PersonalityChat
DialoGPT 345M w/ MMI	**2978** (50%)	241 (4%)	2781 (46%)	Human responses

Reference: https://www.microsoft.com/en-us/research/project/large-scale-pretraining-for-response-generation/

Pre-trained Models & Fine-tuning Datasets

Datasets overview:

Movie_titles_metadata

m0	10 things i hate about you	1999	6.90	62847	['comedy' 'romance']
m1	1492: conquest of paradise	1992	6.20	10421	['adventure' 'biography' 'drama' 'history']
m2	15 minutes	2001	6.10	25854	['action' 'crime' 'drama' 'thriller']
m3	2001: a space odyssey	1968	8.40	163227	['adventure' 'mystery' 'sci-fi']
m4	48 hrs.	1982	6.90	22289	['action' 'comedy' 'crime' 'drama' 'thriller']

Step1 (filter): We only use **comedy genres** as training data this time

Movie_characters_metadata

u0	BIANCA	m0	10 things i hate about you	f	4
u1	BRUCE	m0	10 things i hate about you	?	?
u2	CAMERON	m0	10 things i hate about you	m	3
u3	CHASTITY	m0	10 things i hate about you	?	?
u4	JOEY	m0	10 things i hate about you	m	6
u5	KAT	m0	10 things i hate about you	f	2
u6	MANDELLA	m0	10 things i hate about you	f	7

Step2: Find the movieIDs of comedy movies

Movie lines

L1045	u0	m0	BIANCA	They do not!
L1044	u2	m0	CAMERON	They do to!
L985	u0	m0	BIANCA	I hope so.
L984	u2	m0	CAMERON	She okay?
L925	u0	m0	BIANCA	Let's go.
L924	u2	m0	CAMERON	Wow
L872	u0	m0	BIANCA	Okay -- you're gonna need to learn how to lie

Movie_conversations

u0	u2	m0	['L194' 'L195' 'L196' 'L197']
u0	u2	m0	['L198' 'L199']
u0	u2	m0	['L200' 'L201' 'L202' 'L203']
u0	u2	m0	['L204' 'L205' 'L206']
u0	u2	m0	['L207' 'L208']

Step3: Find the list of the utterances that make the conversation, in chronological order

Step4: Find the the actual text of each utterance

Reference: https://www.kaggle.com/datasets/Cornell-University/movie-dialog-corpus

Finetuning

Hardware Support for Training:

- Colab Pro+ & extra 300 compute units($50 per month + extra $30)(500+300 compute units)
- Google One ($2 per month)(100 GB storage)
- Intel i7-10750H (RAM 16GB)

Finetuning Time(Epochs:2 Batch Size:4):

- DialoGPT-small: Around 35min
- DialoGPT-medium: Around 1h 20min
- DialoGPT-large: Around 2h 15min

Finetuning

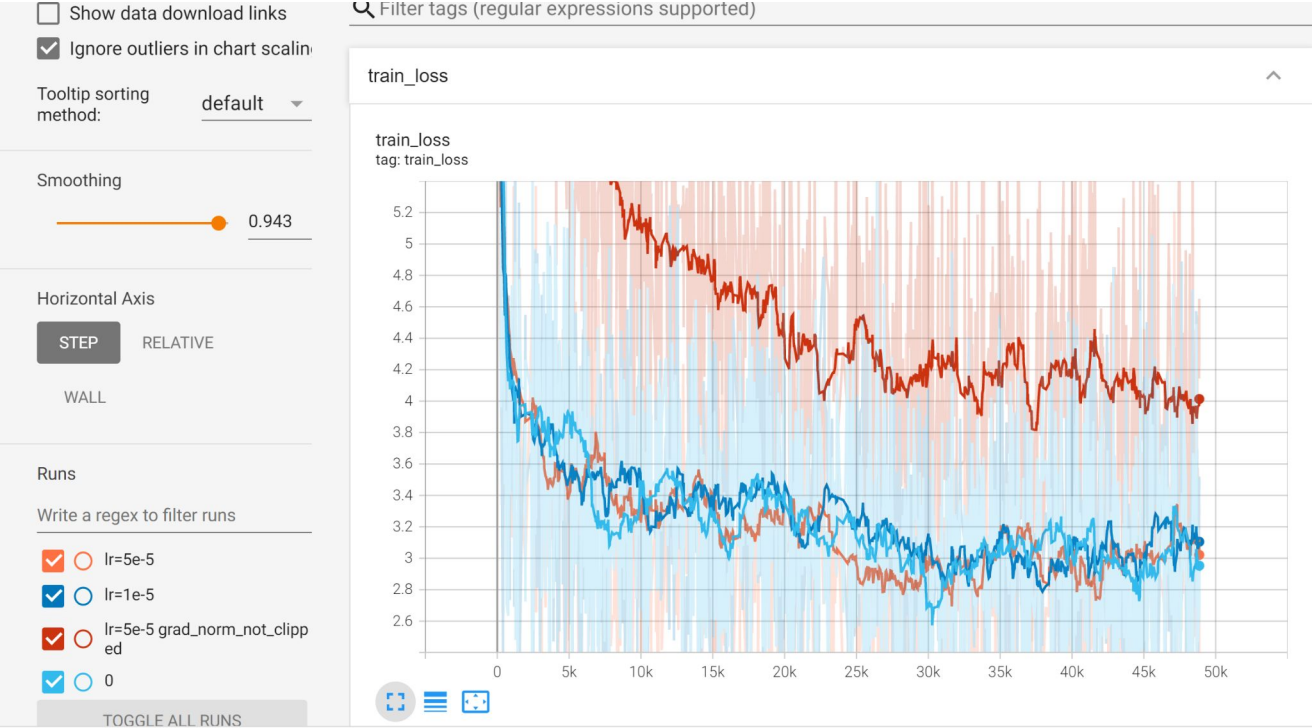
Controllable Hyper-parameters:

- accumulated_grad: step interval to apply the gradient change to parameters
- epochs: times of fine-tuning on the training dataset repeatedly
- lr_init: initial learning rate
- batch_size: encodings size in batch
- adam_epsilon: hyper-parameter for AdamW optimizer
- warmup_steps: hyper-parameter for scheduler
- block_size: max length to filter out the training data rows
- max_grad_norm: value for clipping gradients if applying clip_grad_norm trick
- model_save_interval: step interval to save the model
- val_interval: step interval to do validation

```
accumulated_grad = 2
epochs = 2
lr_init = 5e-6
batch_size = 4
adam_epsilon = 1e-8
warmup_steps = 3
block_size = 800
max_grad_norm = 1.0
model_save_interval = 2000
val_interval = 100
```

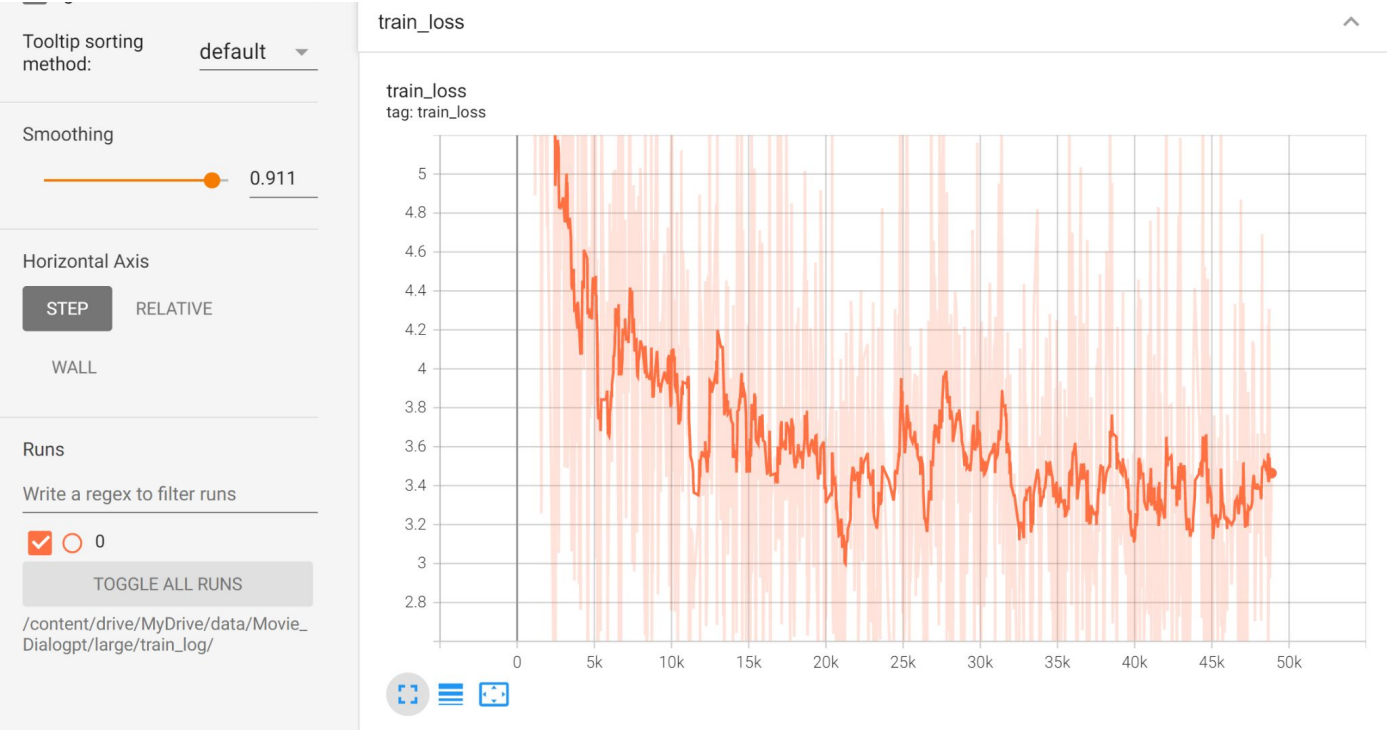
A typical choice for the medium-size model

Finetuning



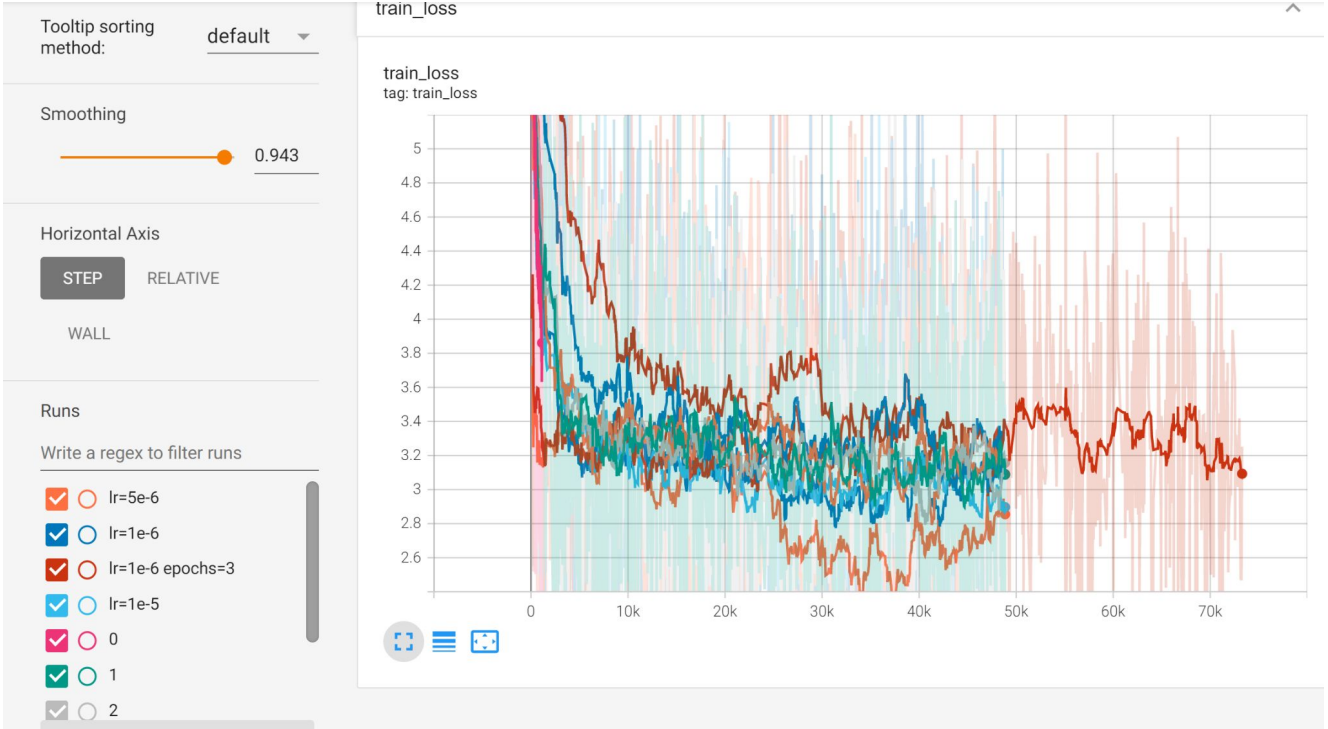
Train Loss of Small-Size Models(Tensorboard Screenshot)(Completed cases)

Finetuning



Train Loss of Large-Size Models(Tensorboard Screenshot)(Completed cases)

Finetuning



Train Loss of Medium-Size Models(Tensorboard Screenshot)(Completed cases)

Deploy

Google Voice API

Google Colab

Process unread messages duration: 5 second

Virtual number +1 2028884948

session-1

session-2

session-n

Deploy

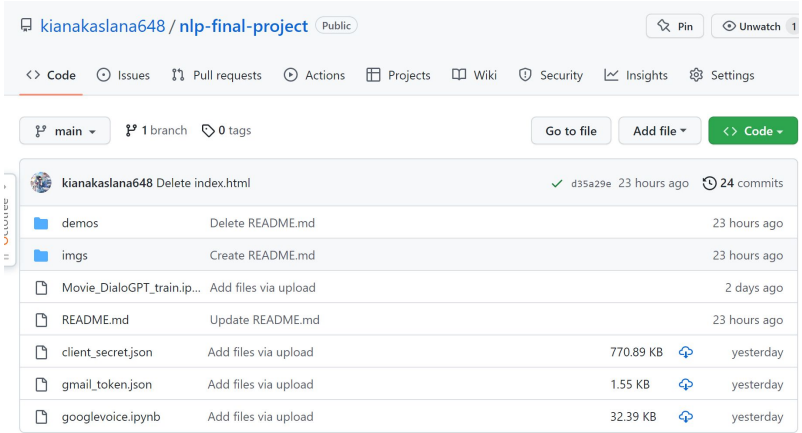
- We used Google Voice API to receive messages, autosave messages to gmail. Then processed unread messages every 5 seconds, using our models. Finally sent processed messages through our virtual number (**+1 2028884948**). All messages were processed in different sessions which are classified by sender numbers.

- Since there were no extra machines nor rent any servers, we deployed our models just on Google Colab.Considering the price of google colab, it is not 24-hour deployed, only running when we want to make interactions with our chatbot.

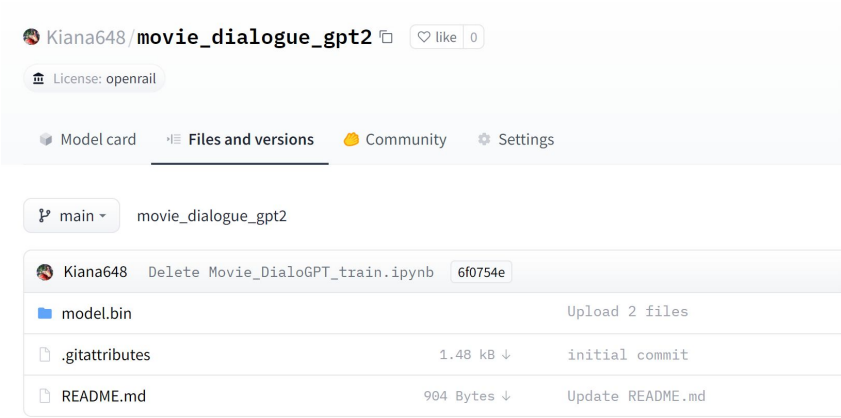
Finetuning

Fine-tuned Model Size(pytorch_model.bin):

- Small-size: 0.5GB
- Medium-size: 1.4GB
- Large-size: 2.9GB



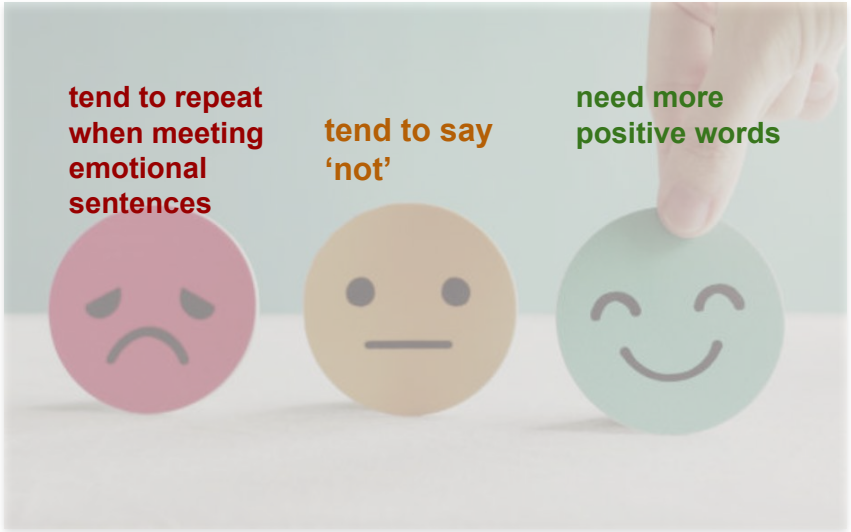
Github page for tutorials of data cleaning, fine-tuning & deploying



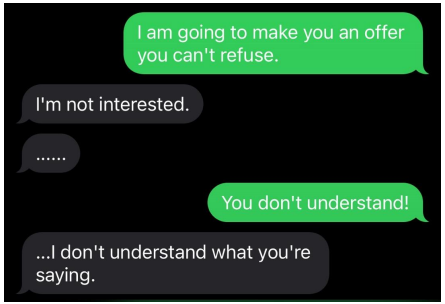
Huggingface page for fine-tuned model in the medium size

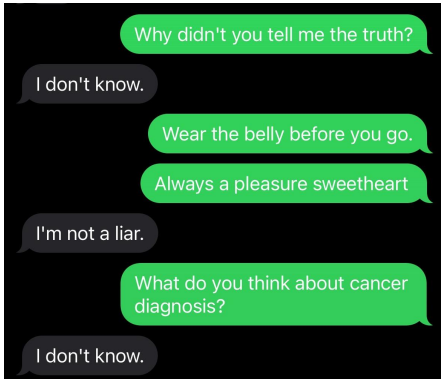
Demos - small-size model

User feedback:



Demo Video:
https://www.youtube.com/watch?v=ZnOOLe2BARY



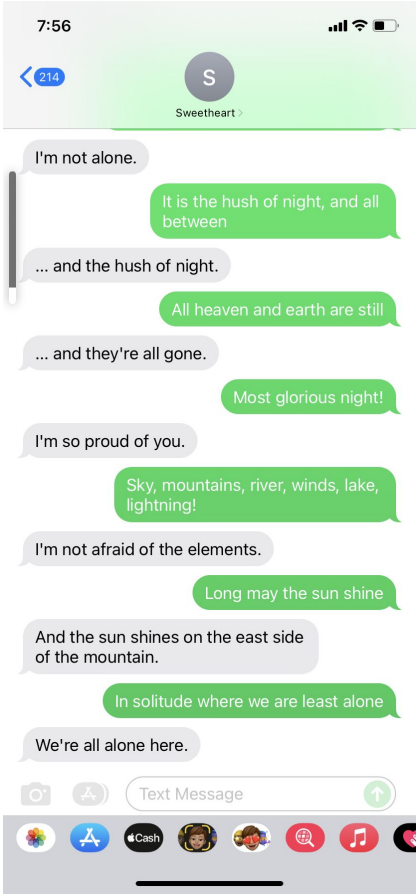


Demos - medium-size model

Improvements:

- **Fine-tuning**: Increase model size
- **Data cleaning**: decrease the number of some words like ‘sorry’, ‘I don’t know’, ‘yes’, ‘no’

Demo Video: https://youtu.be/rWw6YHBZxgc

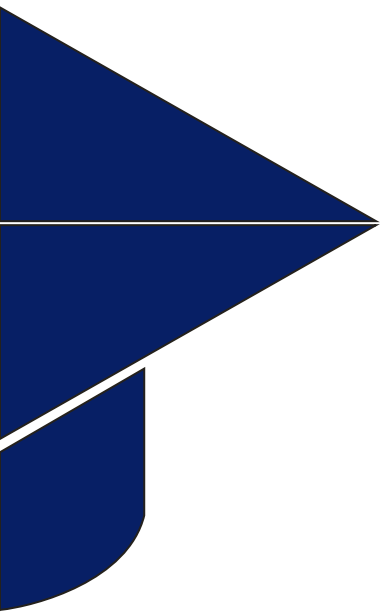


Code Links

Github Repo Link: https://github.com/kianakaslana648/nlp-final-project

Huggingface Link: https://huggingface.co/Kiana648/movie_dialogue_gpt2

Introduction Page(Under Construction): https://kianakaslana648.github.io/nlp-final-project/



THANK YOU!