

# ❓ Empathetic Dialogue System

Efficient Methods in Machine Learning

**Midterm Talk**

04.12.2024

mL\_syj (Jihyun Ryu, Sofia Germer, Yanxi Lin)

---

# Contents

<b>01</b>	_____	<b>Introduction</b>
<b>02</b>	_____	<b>Data &amp; Preprocessing</b>
<b>03</b>	_____	<b>Model &amp; Performance</b>
<b>04</b>	_____	<b>Current Progress</b>
<b>05</b>	_____	<b>Plans for the Future</b>

# 01 Introduction

## Our Goal

Develop an empathetic dialogue system

- Understanding emotions from user input
- Generating empathetic & considerate response



# 02 Data & Preprocessing



## Our Dataset

### 1. Source of data

- the Empathetic Dialogues (Facebook AI) 25k <sup>[1]</sup>

	Situation	emotion	empathetic_dialogues	labels
0	I remember going to the fireworks with my best friend. There was a lot of people, but it only felt like us in the world.	sentimental	I remember going to see the fireworks with my best friend. It was the first time we ever spent time alone together. Although there was a lot of people, we felt like the only people in the world.	Was this a friend you were in love with, or just a best friend?
1	I remember going to the fireworks with my best friend. There was a lot of people, but it only felt like us in the world.	sentimental	This was a best friend. I miss her.	Where has she gone?
2	I remember going to the fireworks with my best friend. There was a lot of people, but it only felt like us in the world.	sentimental	We no longer talk.	Oh was this something that happened because of an argument?
3	I remember going to the fireworks with my best friend. There was a lot of people, but it only felt like us in the world.	sentimental	Was this a friend you were in love with, or just a best friend?	This was a best friend. I miss her.
4	I remember going to the fireworks with my best friend. There was a lot of people, but it only felt like us in the world.	sentimental	Where has she gone?	We no longer talk.

[1] <https://www.kaggle.com/datasets/atharvjairath/empathetic-dialogues-facebook-ai/data>

# 02 Data & Preprocessing



## Our Dataset

### 2. Format

- 64,594 paired (prompt, response) dialogue sets in total
- 19,196 situations in total
- Each situation has an 'emotion' label
  - We simplified it down to 8 from 32 emotions:  
[disappointed, grateful, annoyed, disgusted, impressed, prepared, afraid, excited]
- The initial input is given according to the situation & conversation continues

## 02 Data & Preprocessing



### Our Dataset

#### 3. Prompt (user's Input sentence)

- Length: Mean: 68.28 / Median: 59.0
- Vocabulary size: 16,442

```
33      Got rejected from a place I wanted to work, not once but three times
126     I just really wanted some ice cream! Now I know their hours, though.
139     I do sales work, but he always lies to us and takes our bonus money.
199     Yeah, thank you! at which situation did you feel hope for your life?
311     I hear ya.. I hope you find one soon... wishing you all of the best!
Name: empathetic_dialogues, dtype: object
```

## 02 Data & Preprocessing



### Our Dataset

#### 4. Response (model's response sentence)

- Length: Mean: 62.85 / Median: 54.0
- Vocabulary size: 15,860

```
0      Was this a friend you were in love with, or just a best friend?
179    The grass makes me itchy, But the shower afterward feels great.
196    I still took it since it was late but I rode in the front seat.
338    Oh no... were they relaxed about it or did it cause a problem?
459    That's really considerate of you. Do they need your help a lot?
Name: labels, dtype: object
```

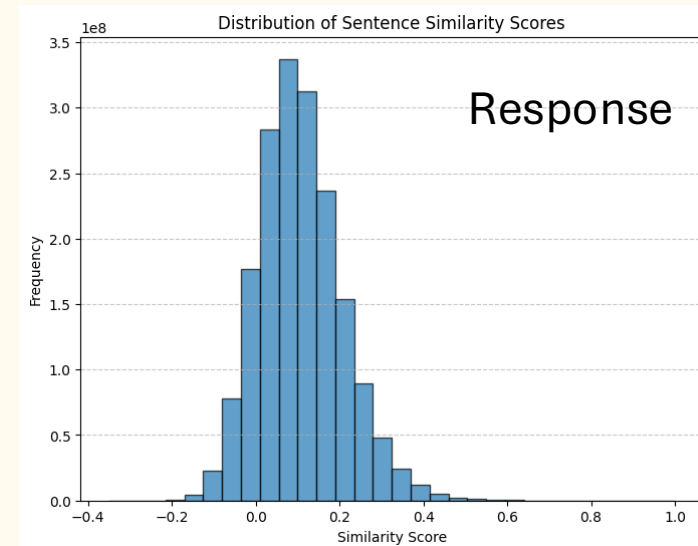
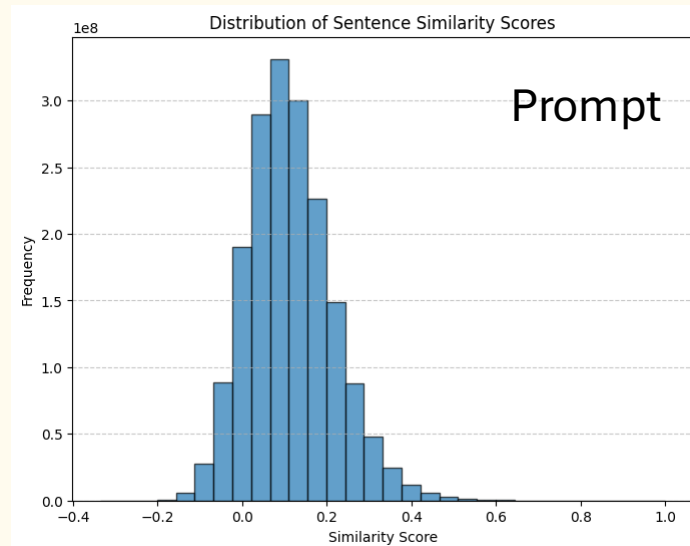
# 02 Data & Preprocessing



## Our Dataset

### 5. Sentence similarity

- Computed by Cosine similarity
- The highest frequency of similarity scores clustered around 0 : They are likely discussing unrelated topics





# 02 Data & Preprocessing



## Our Dataset

### 6. final output for model training

- We prepared **4 txt files** for our model

	<b>59k dialogues</b>	<b>* 38k dialogues</b>
<endoftext> token inserted after each dialogues end	1) 59k_eachconv_eot	2) 38k_eachconv_eot
<endoftext> token placed only at the very end of whole conversations (within same situation & emotion)	3) 59k_wholeconv_eot	4) 38k_wholeconv_eot

\* there are cases where prompts and responses are repeated in reverse order, appearing as (response, prompt) pairs.  
In this dialogue set, we deleted the repeated conversation.

## 02 Data & Preprocessing



### **Data augmentation (Ongoing)**

gpt-4-mini

- Using existing prompts from our dataset, generate abundant responses
- So far, we have generated 59K pairs and will evaluate the quality of results

# 03 Model & Performance

## NanoGPT

### 1. Overview

- Decoder-only Transformer-based Language Model
- Compact & lightweight (around 3.4M parameters) model

# 03 Model & Performance

## NanoGPT

### 2. Tokenization

- openAI's **tiktoken**
  - Optimized for large-scale text tokenization
  - Designed for OpenAI's GPT models
  - Based on Byte Pair Encoding (BPE)

# 03 Model & Performance

## NanoGPT

### 2. Embeddings

- Token Embeddings
- Positional Embeddings

# 03 Model & Performance

## NanoGPT

### 2. Embeddings

\* n\_embd = 768

```
self.transformer = nn.ModuleDict(dict(  
    wte = nn.Embedding(config.vocab_size, config.n_embd) # token embedding  
    wpe = nn.Embedding(config.block_size, config.n_embd) # positional embedding  
    drop = nn.Dropout(config.dropout), # dropout layer  
    h = nn.ModuleList([Block(config) for _ in range(config.n_layer)]), # the transformer  
    ln_f = LayerNorm(config.n_embd, bias=config.bias), # layer norm at the output of  
))
```

```
# forward the GPT model itself  
tok_emb = self.transformer.wte(idx) # token embeddings of shape (b, t, n_embd)  
pos_emb = self.transformer.wpe(pos) # position embeddings of shape (t, n_embd)  
#TODO: the embedding here is simple, just sum them up, could improve with more  
x = self.transformer.drop(tok_emb + pos_emb)
```

# 03 Model & Performance

## NanoGPT

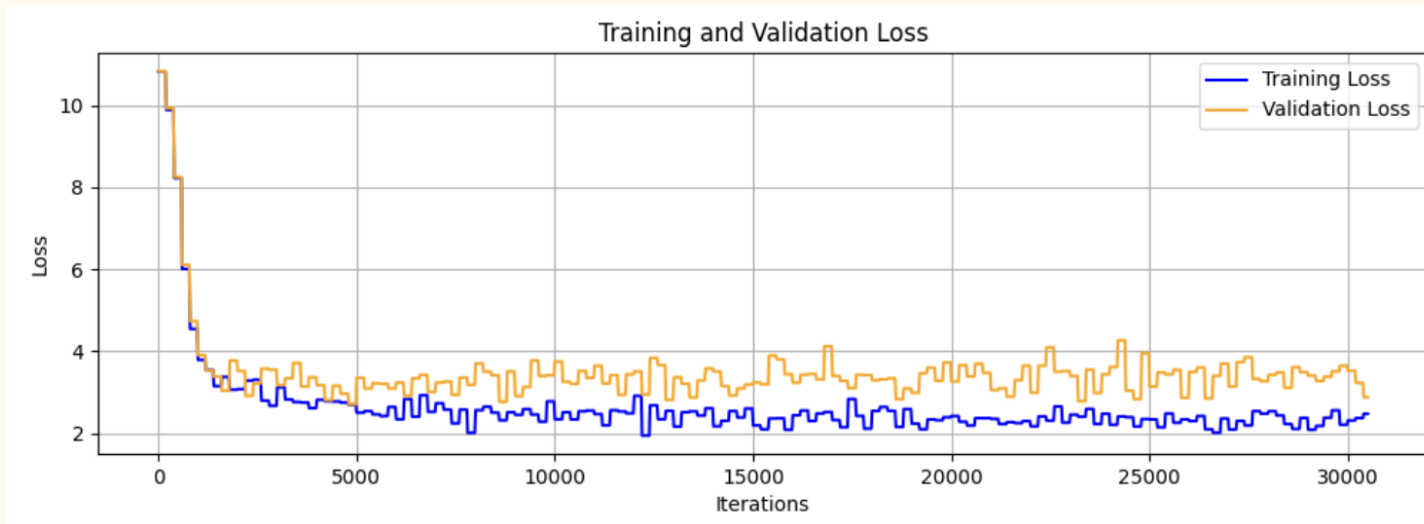
### Modification

- Our next goal
- To improve positional embedding

# 04 Current Progress

## Training process

- Loss function: Cross-entropy loss, Optimizer: AdamW
- About 30k iteration
- Training time: about 9 hours (Macbook M1 pro)
- Loss





## 04 Current Progress

### Intermediate Results

trained with '*59k\_eachconv\_eot.txt*'

Result

- User: I'm not feeling well -> Robot: **I'm sorry, I hope you are doing well**
- User: I have a new job -> Robot: **That's great! I hope you get a good job**

# 04 Current Progress

## Intermediate Results

Evaluation: BLEU score

### Input sentences

"I lost my job last year and got really angry."

Reference 1: I am sorry to hear that. Did it happen out of the blue?  
Model Output 1: That's a bummer. I hope you can find the new job!  
BLEU Score: 0.02

"I have lived in my apartment for 5 years now."

Reference 2: wow that is a long time, Are you happy with it?  
Model Output 2: I'm sorry. I'm sure you will get it!  
BLEU Score: 0.02

"He signed right away and we had a drink to celebrate."

Reference 3: That's great. How happy were you?  
Model Output 3: What did you do?  
BLEU Score: 0.00

# 04 Current Progress

## Intermediate Results

Evaluation: BERT score

### Input sentences

"I lost my job last year and got really angry."

"I have lived in my apartment for 5 years now."

"He signed right away and we had a drink to celebrate."

Reference 1: I am sorry to hear that. Did it happen out of the blue?  
Model Output 1: That's a bummer. I hope you can find the new job!  
BERTScore Precision: 0.8687  
BERTScore Recall: 0.8765  
BERTScore F1: 0.8726

Reference 2: wow that is a long time, Are you happy with it?  
Model Output 2: I'm sorry. I'm sure you will get it!  
BERTScore Precision: 0.8639  
BERTScore Recall: 0.8695  
BERTScore F1: 0.8667

Reference 3: That's great. How happy were you?  
Model Output 3: What did you do?  
BERTScore Precision: 0.8972  
BERTScore Recall: 0.8800  
BERTScore F1: 0.8885

# 04 Current Progress

## Intermediate Results

Evaluation: GLUE score

### [Emotion List]

1. Anger
2. Disgust
3. Fear
4. Joy
5. Sadness
6. Surprise

### Input sentences

"I lost my job last year and got really angry."

"I have lived in my apartment for 5 years now."

"He signed right away and we had a drink to celebrate."

Pair 1:

Generated: That's terrible! Did you confront them about it? , Sentiment: anger

Reference: I am sorry to hear that. Did it happen out of the blue?, Sentiment: anger

Score: 1.00

Pair 2:

Generated: I would be mad too. What is that? , Sentiment: anger

Reference: wow that is a long time, Are you happy with it?, Sentiment: joy

Score: 0.00

Pair 3:

Generated: That's great! I hope you are feeling good. , Sentiment: joy

Reference: That's great. How happy were you?, Sentiment: joy

Score: 1.00

Final Sentiment Analysis Accuracy: 0.67

# 04 Current Progress

## Obstacles

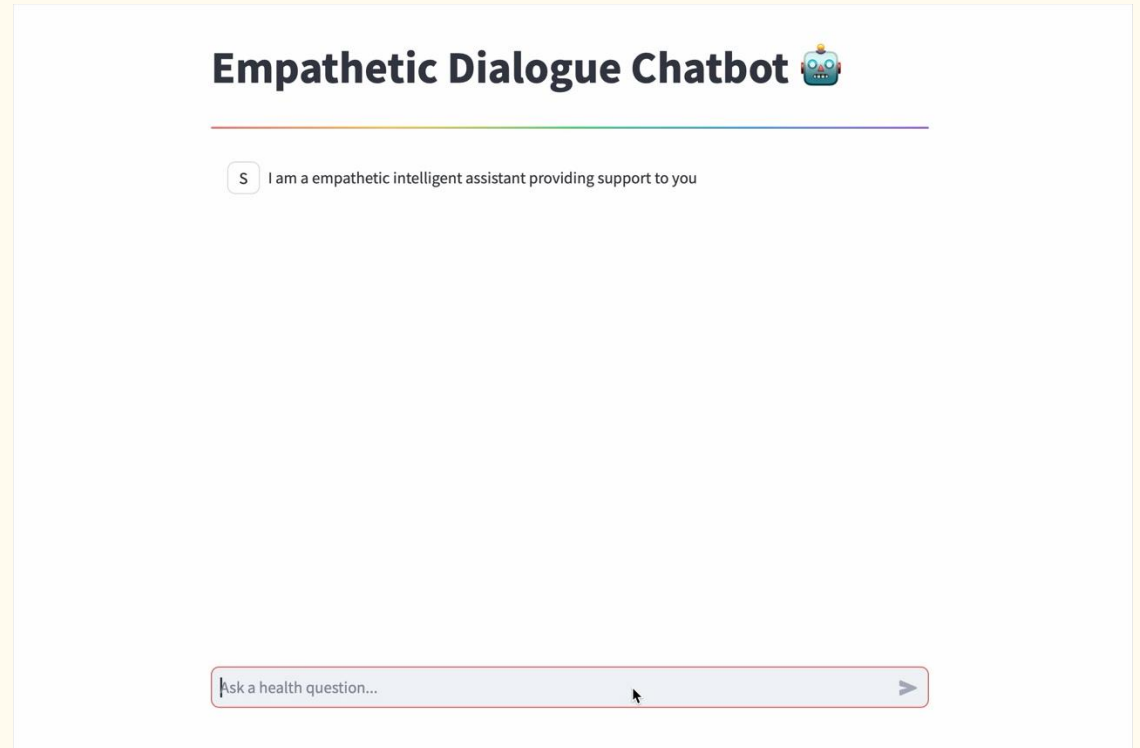
1. We want more relevant data  
-> We are currently working on it

# 04 Current Progress



## Minimum Goal

1. The model could grasp the context of the prompt and then generate an empathetic response
2. Have interactable UI for our model



# 05 Plans for the Future



## Goals for next week

- Try to make changes in Model wrt positional embedding
- Evaluate the model on the test dataset (make the process automatically)
- Try to train the model
  - with emotional labels
  - with our new augmented data

**Thank you!**

**Questions?**