

Empathetic Dialogue System

Efficient Methods in Machine Learning (Master Project) 2024/25

Jihyun Ryu, Sofia Germer, Yanxi Lin

29.01.2025

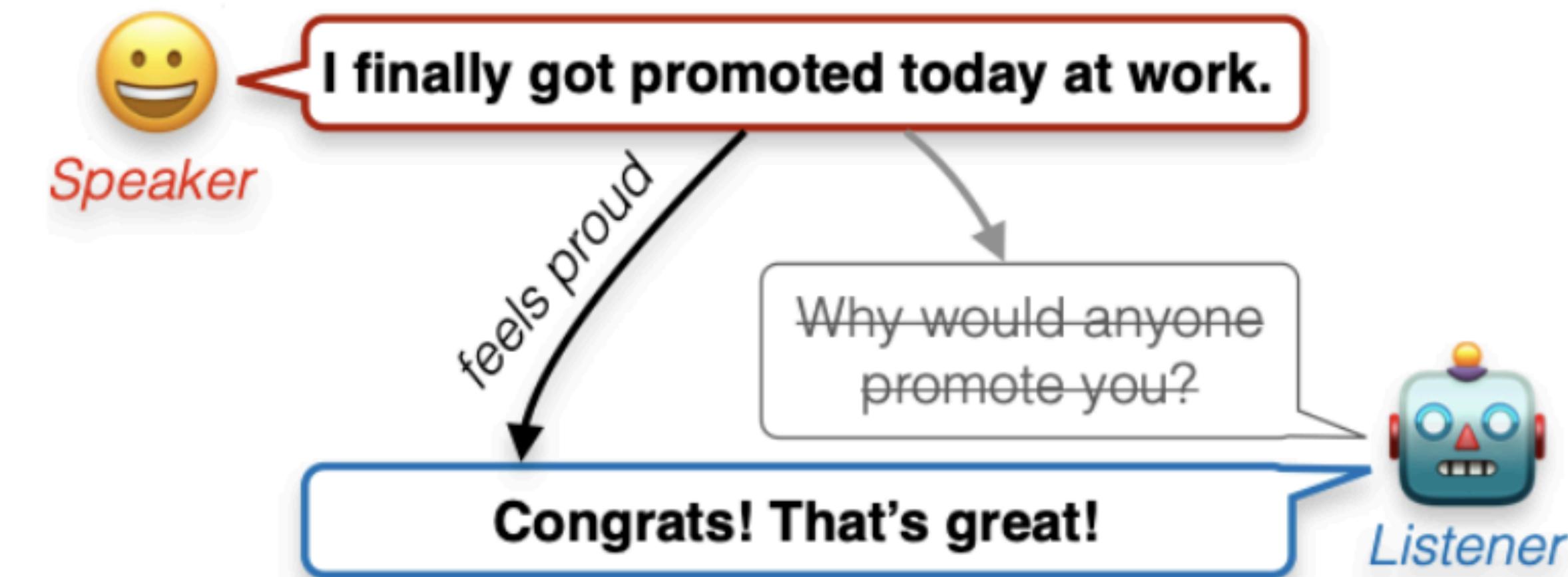
Agenda

- Introduction
- Pipeline
- Take-away messages
- Demo

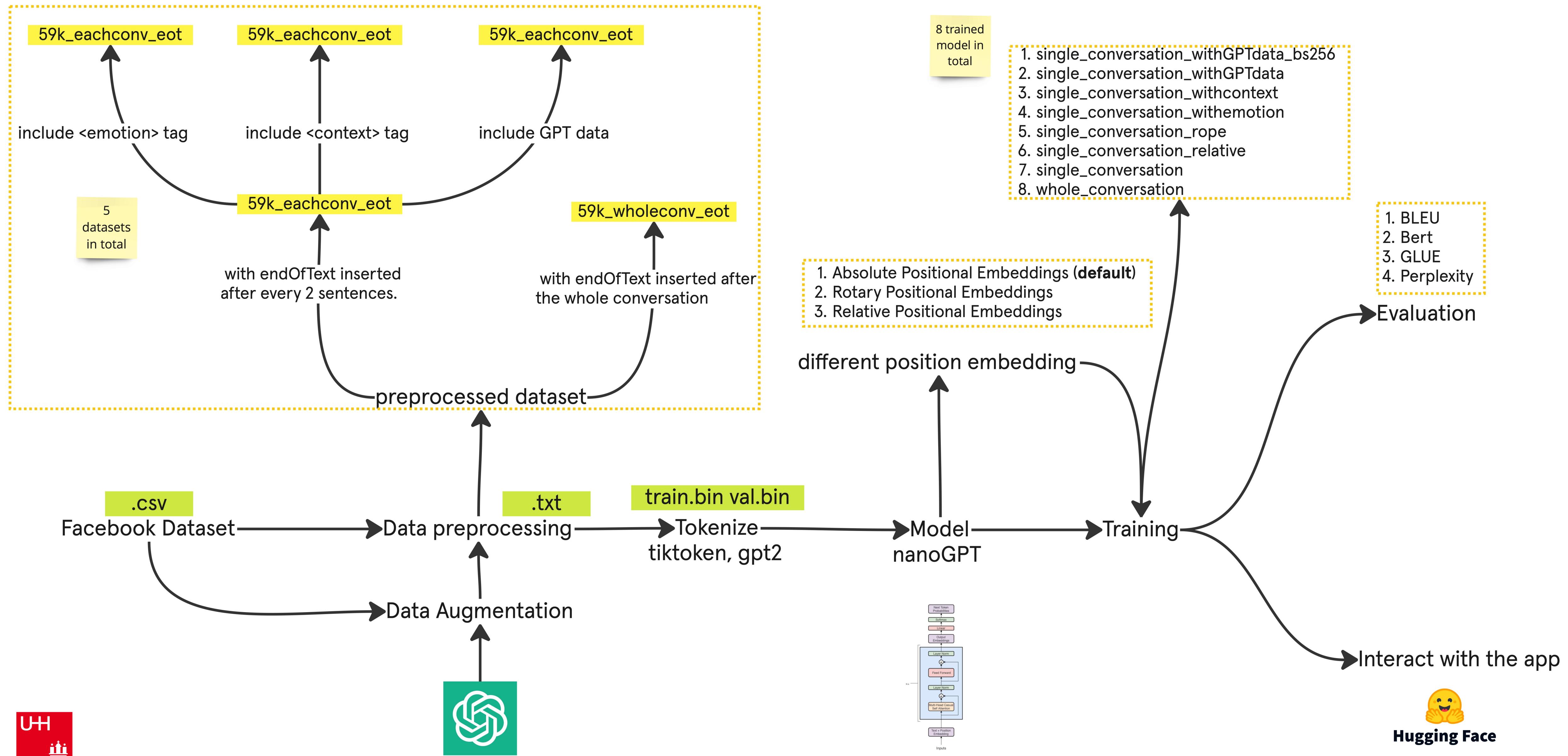
Introduction

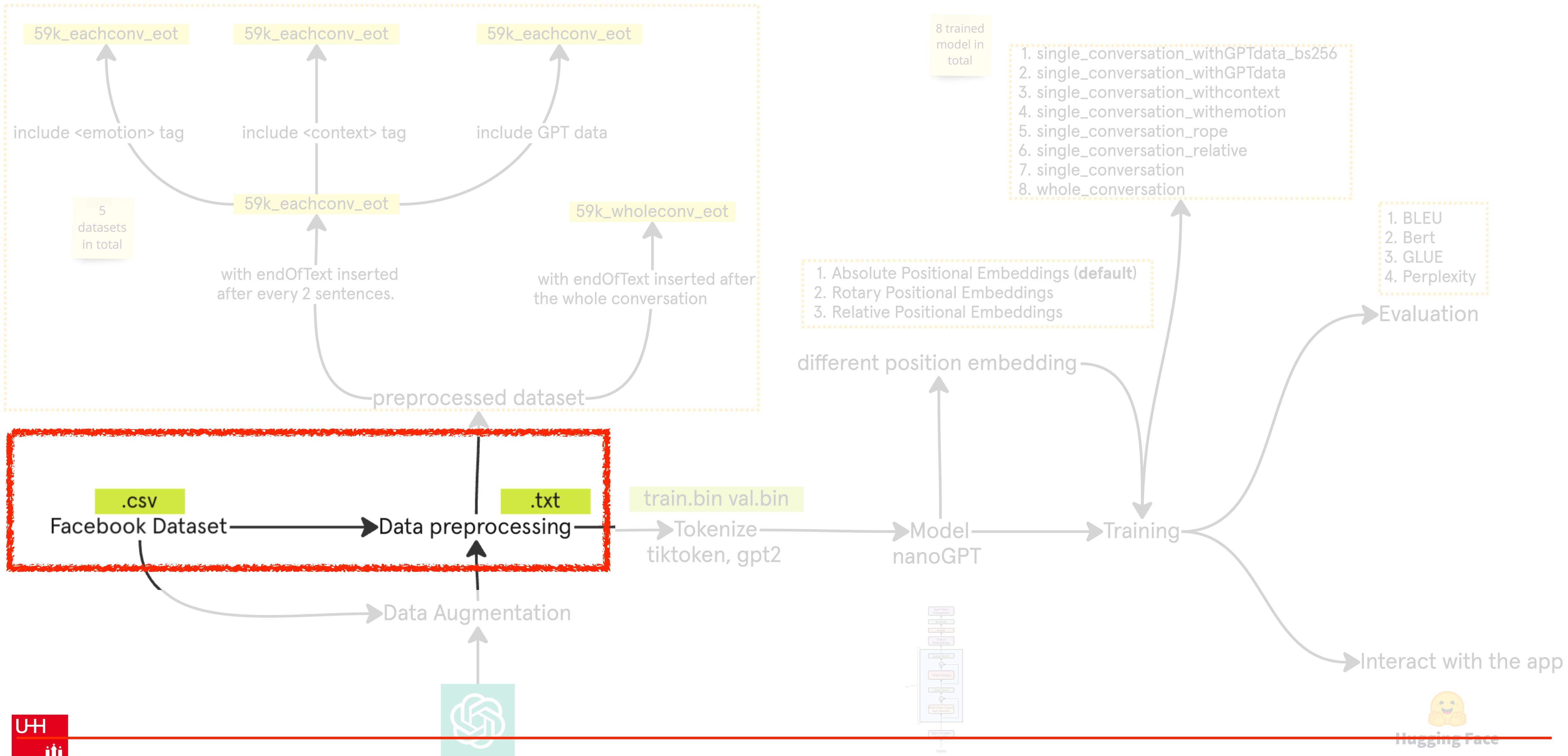
Empathetic Dialogue System

- Goal: Develop an empathetic dialogue system (chatbot)
 - Understanding user input
 - Generating empathetic & considerate response



Pipeline

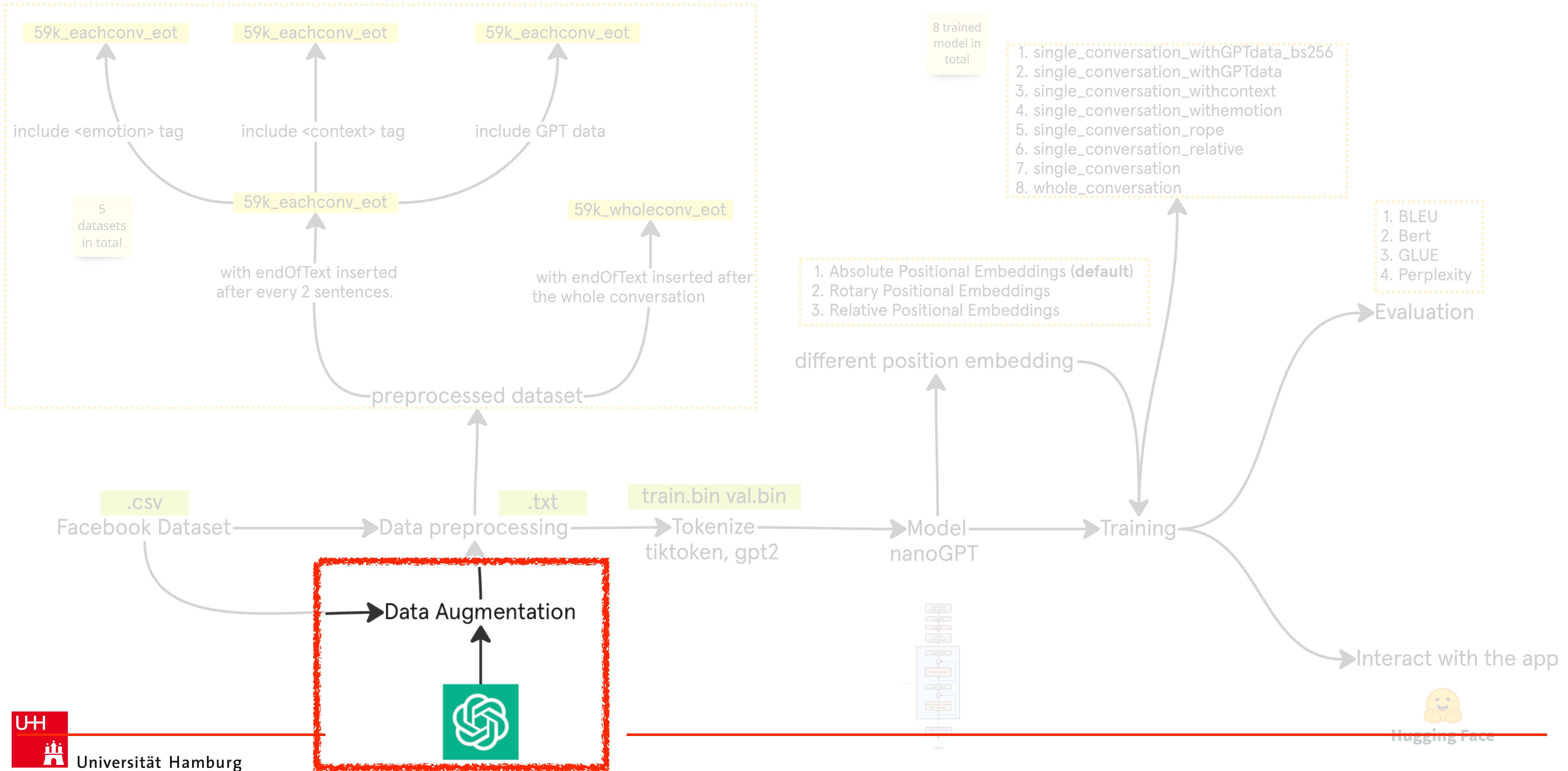




Facebook dataset

- The Empathetic Dialogues (Facebook AI) 25k [1]
- EDA + Details covered in the 2nd talk
- **64,594** paired dialogue sets; Vocabulary size: **15,860**; Average length: ~**60** characters
- Preprocessing: add <human> <bot> <endOfText> <emotion> <context> tag -> txt file

Index	Situation	Emotion	empathetic_dialogue	label
0	I remember going to the fireworks with my best friend. There was a lot of people...	sentimental	This was my best friend, I miss her.	Where has she gone?



Augment the data

- We use ChaGPT 4o-mini to generate more data
 - “context” when use API: "You are a helpful and empathic assistant. Your responses should be concise, around 63 characters long.
 - Input: the human message
 - -> Get the bot reply from GPT
 - Merged vocab size: ~**18k**

LDA to extract topics

- 4 topics in totally

Topic 2

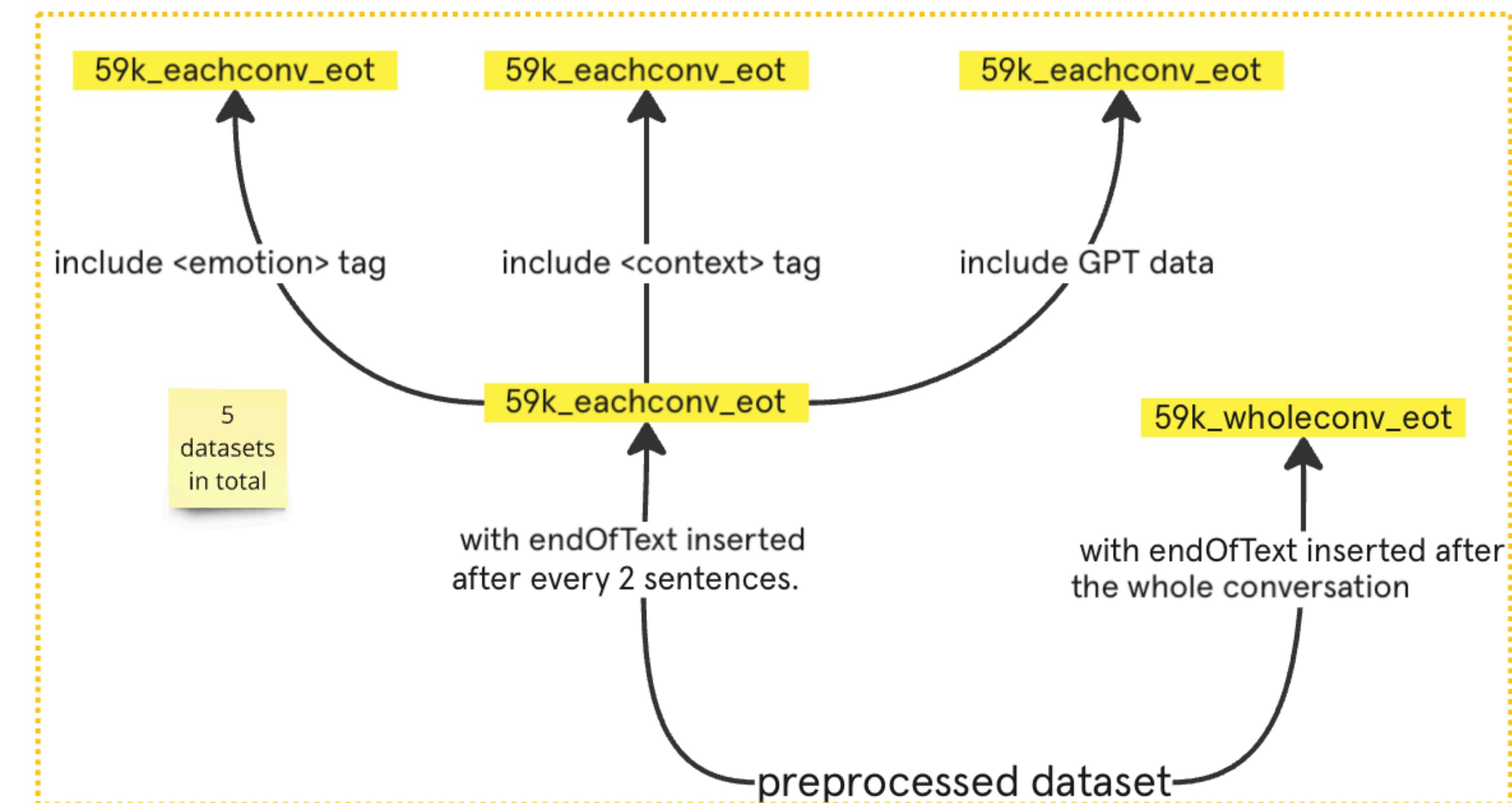
A word cloud visualization for Topic 2. The words are arranged in a cluster, with larger words indicating higher frequency. The words include 'know', 'going', 'reallygood', 'well', 'get', 'job', 'imive', and 'week'. The colors of the words vary, including shades of green, blue, and purple.

Topic 3

A word cloud visualization for Topic 3. The words are arranged in a cluster, with larger words indicating higher frequency. The words include 'get', 'day', 'today', 'got', 'work', 'home.', 'last', 'dog', 'friend', and 'cant'. The colors of the words vary, including shades of purple, yellow, green, and blue.

Datasets we used in the training

- **<context>** I remember going to the fireworks....
- **<emotion>** disappointed
- **<bot>** It was the first time we ever spent time alone together...
- **<human>** Was this a friend you were in love with? **<endOfText>**



Where to put the <endOfText>

- We had some very interesting results in the beginning, when we put the <endOfText> after each sentences
 - After debug the full output, we found the model generate readable sentences, but totally not understanding the input

```
number of parameters: 3.42M
loading GPT-2 encodings...
User: hello
Bot: That's really great! I have a great time in a long time and I have no faith in your life

User: sorry?
Bot: Haha, but I was still surprised it's going to go to the doctor!

User: I'm not feeling well
Bot: That's is a nice gesture of the old days. It was a nice surprise
```

Where to put the <endOfText>

- Maybe the model understand the tag as start a new conversation, rather than “response”
- We move the <endOfText> after each pair of the conversation
 - Now it generate the response to the user input

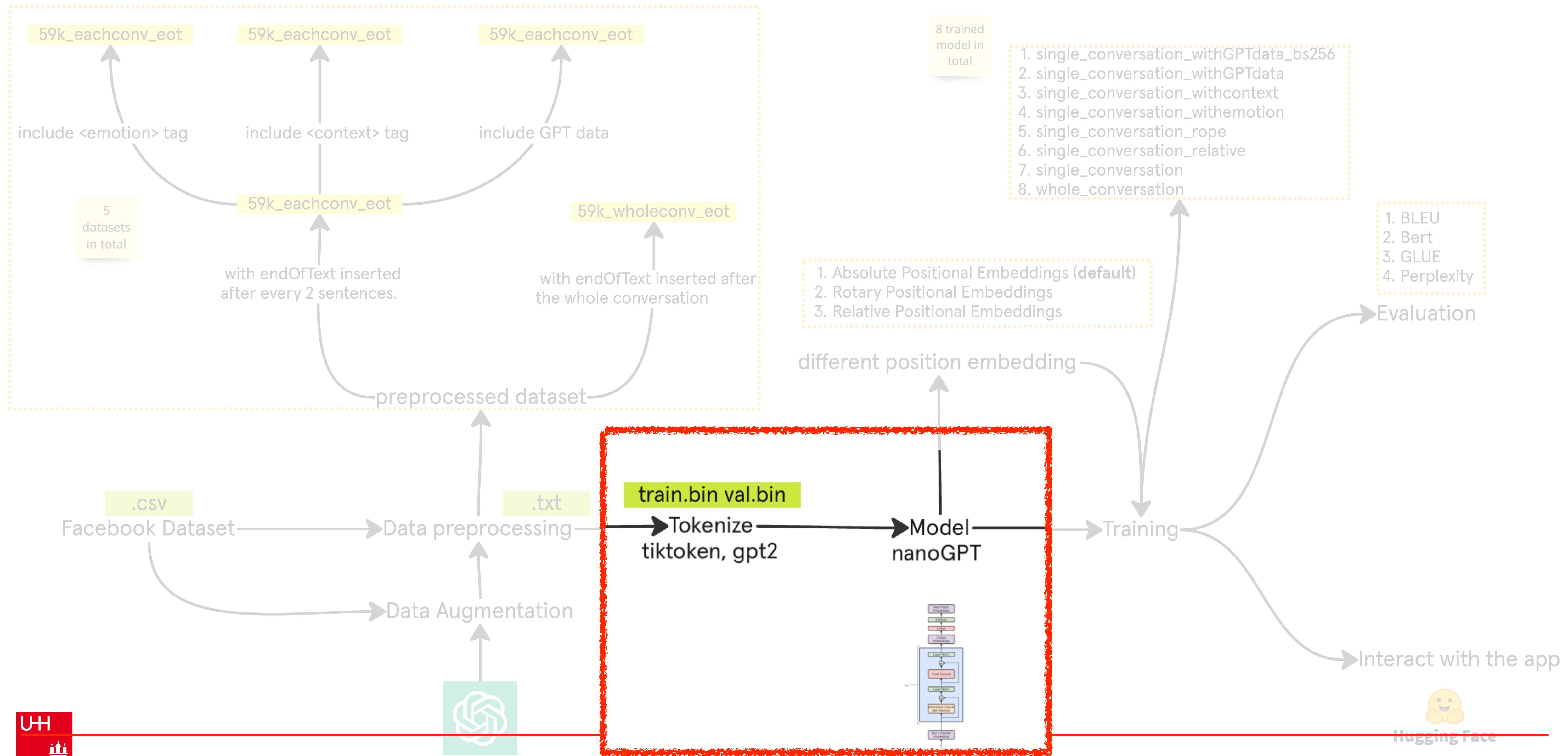
```
User: I feel bad  
      I am so sorry to hear that, did you get the promotion?  
<human>  
User: I lose my job  
      What do you plan on doing?  
<human>  
User: ■
```

29.11.24: train the model with

GPT data also helps

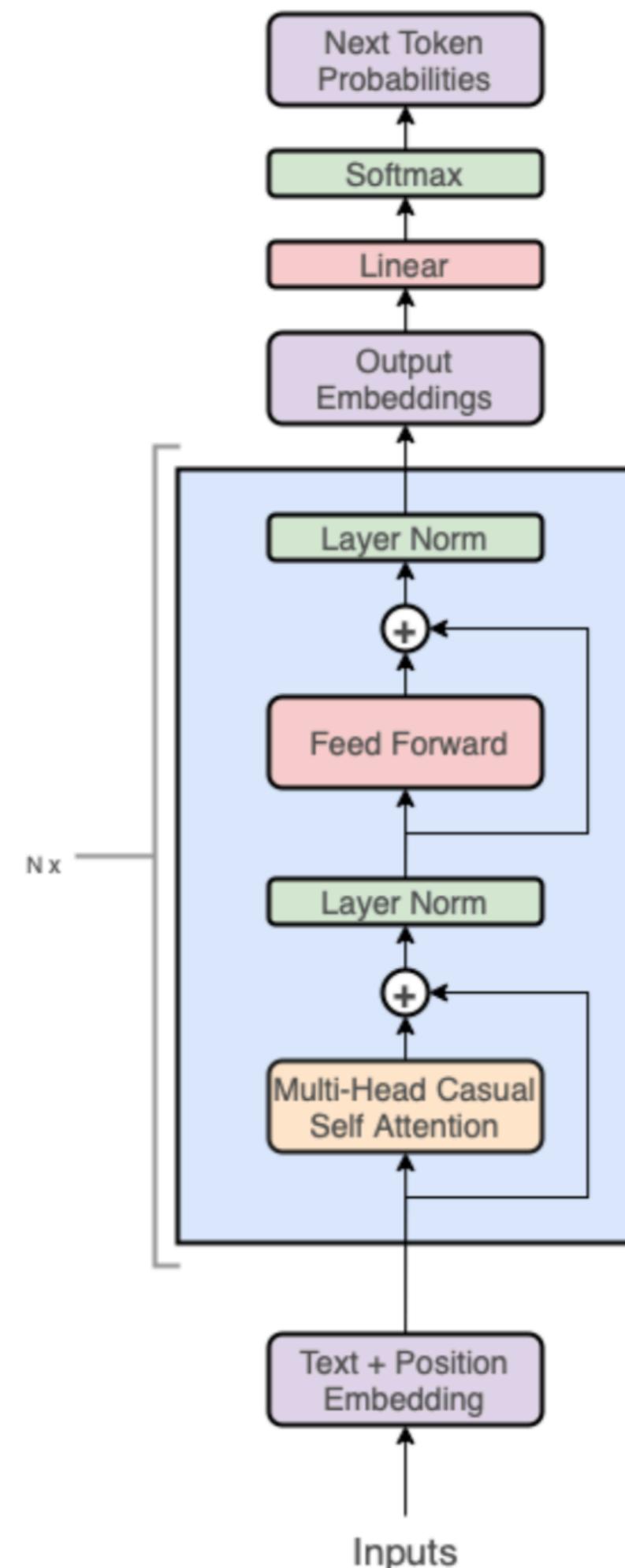
```
User: I got a new job!
Robot: Congratulations! That's an amazing achievement! 🎉
----Debug: Full output---
<bot> I got a new job!<human> Congratulations! That's an amazing achievement! 🎉 <endOfText>
<bot> Yes! I am. I am sure I will make it.
<human> That's great to hear! Keep pushing forward! <endOfText>
<bot> Congratulations! That's awesome! I hope you get a good job.
<human> Thank you! I'm here to help with any questions you have. <endOfText>
<bot> I have been
User: I hate my boss
Robot: It's tough to feel that way. Have you talked to your manager?
----Debug: Full output---
<bot> I hate my boss<human> It's tough to feel that way. Have you talked to your manager? <endOfText>
<bot> I don't know. I'm so sorry. What is the job about?
<human> What specific topic are you referring to? <endOfText>
<bot> I am very nervous about my upcoming job. I am sure I have a job, I am not sure if I can do it.
<human> It's normal to feel nervous; take
```

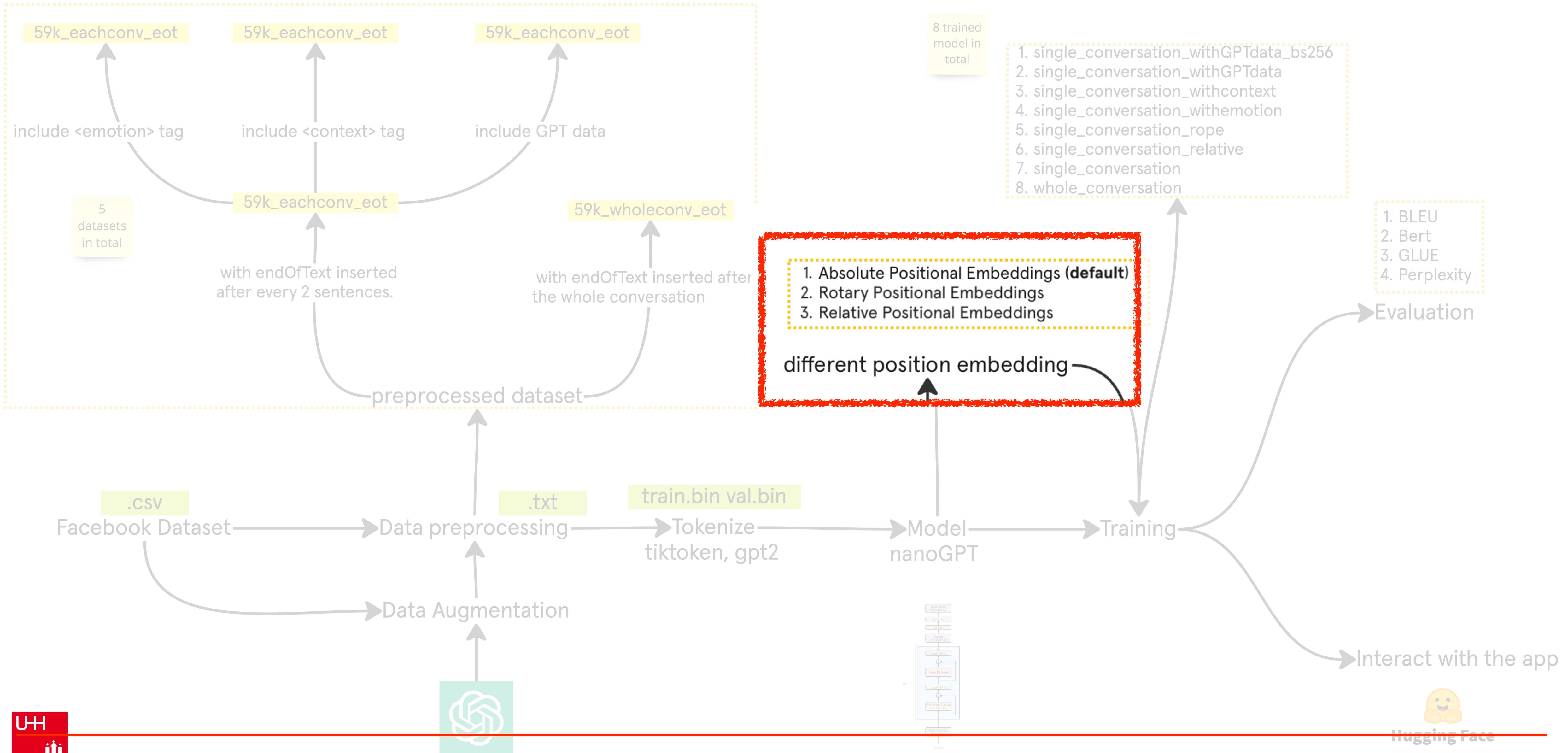
17.12.24 with GPT data we can generate emoji 🎉



Tokenizer and the model

- Tokenizer: Ticktoken
 - Optimized for large-scale text tokenization
 - Designed for OpenAI's GPT models
- NanoGPT: Decoder only transformer
 - Number of parameters: 3.42M
 - Details covered in talk2
- Why we choose nanoGPT
 - Small model -> easy to train
 - The open source code is very simple but enough

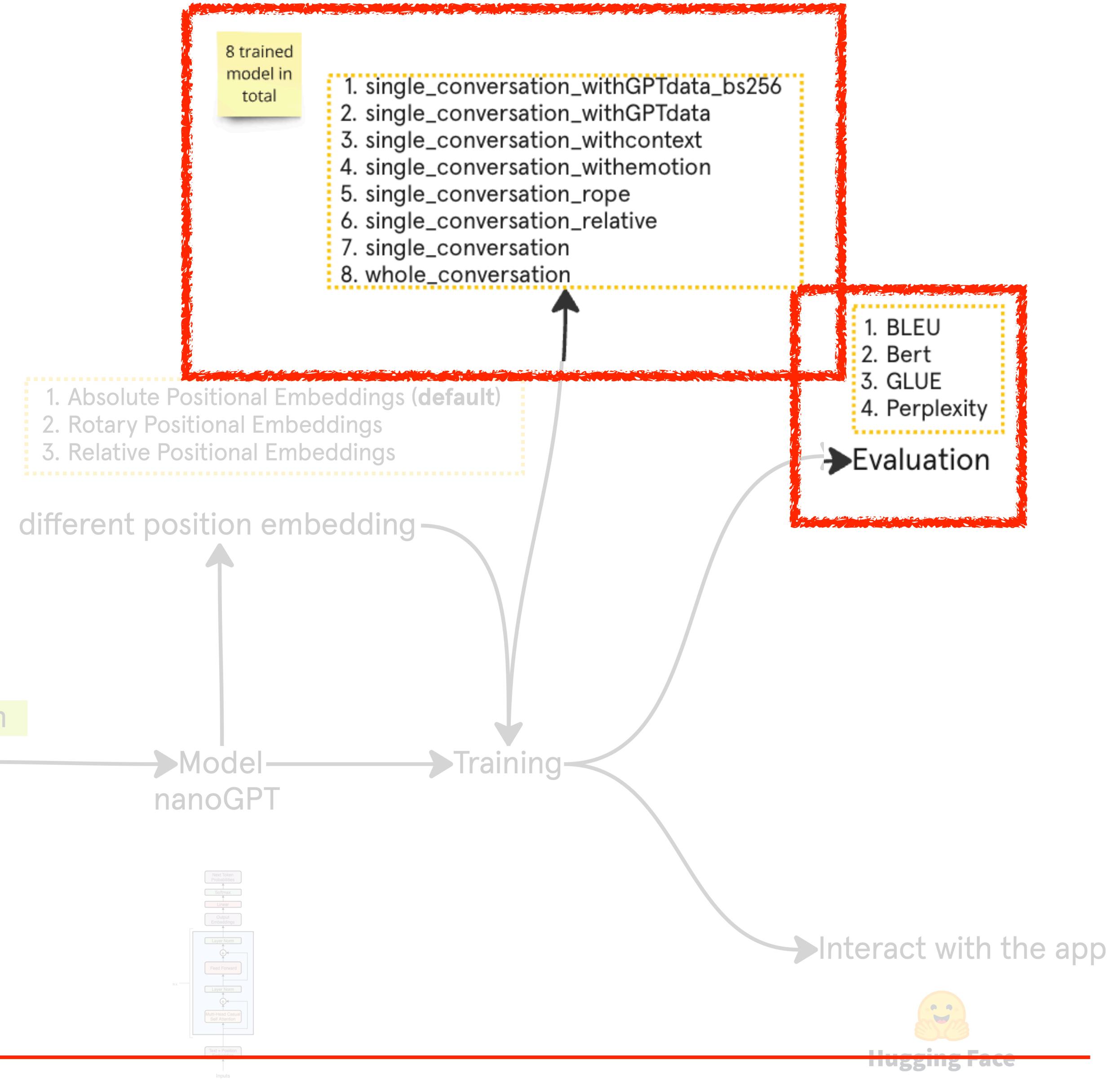
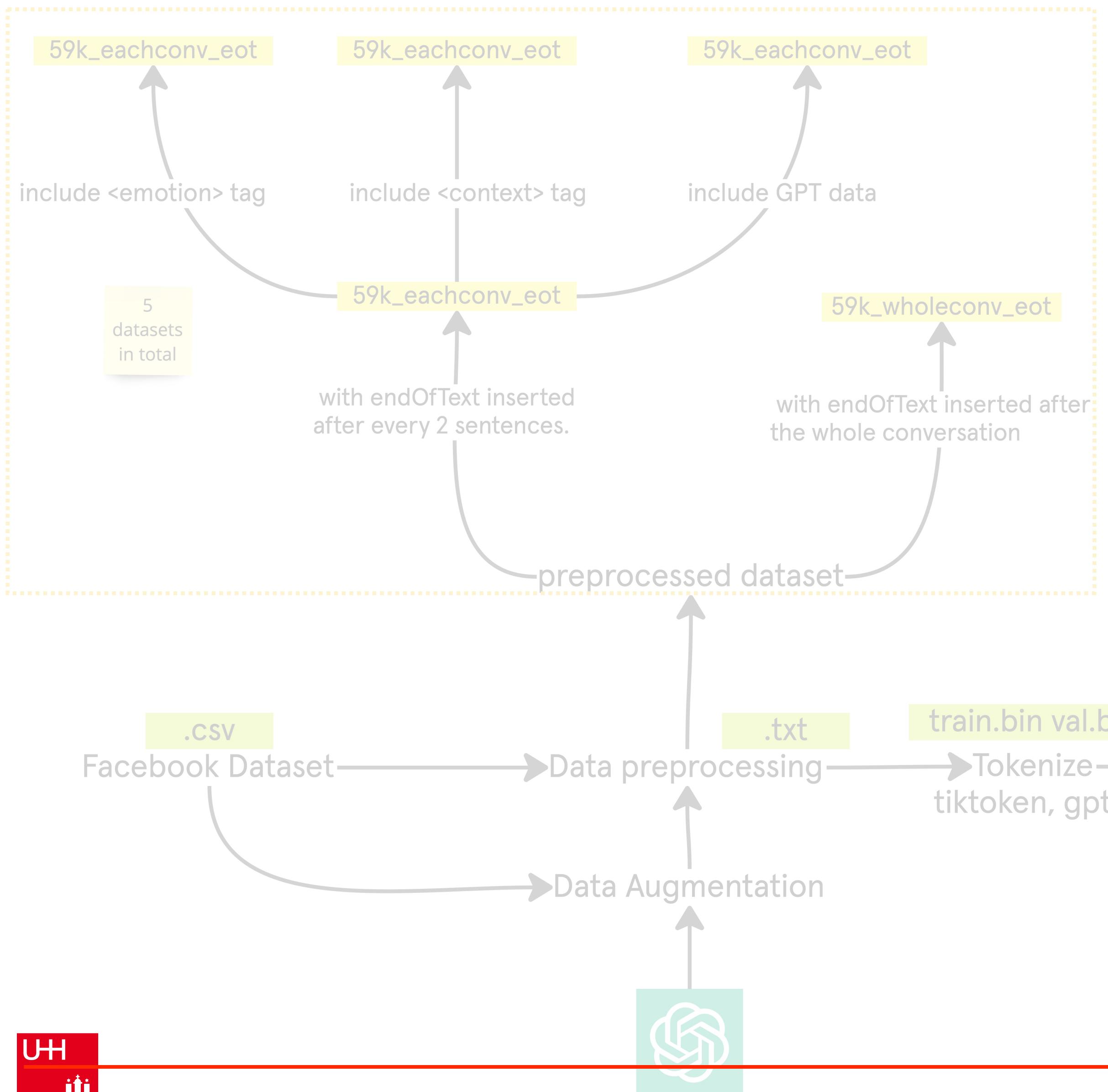




Different Position Embeddings

- We tried different position embeddings

	Absolute Positional Embeddings	Relative Positional Embeddings	Rotary Positional Embeddings
Use Cases	Standard Transformers (default for nanoGPT)	NLP tasks with context sensitivity	Large models like GPTs
Position Representation	Encode unique absolute position	Encode relative distances between positions	Encode relative information via rotation
Scalability	Limited in some fixed implementations	Generalizable to varying lengths	Well-suited for long sequences
Computational Efficiency	Simple	Higher complexity	Highly efficient



Training

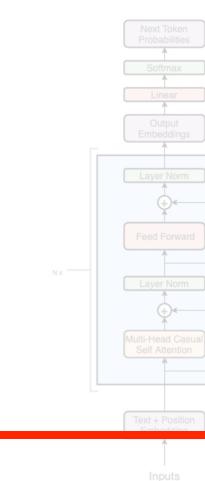
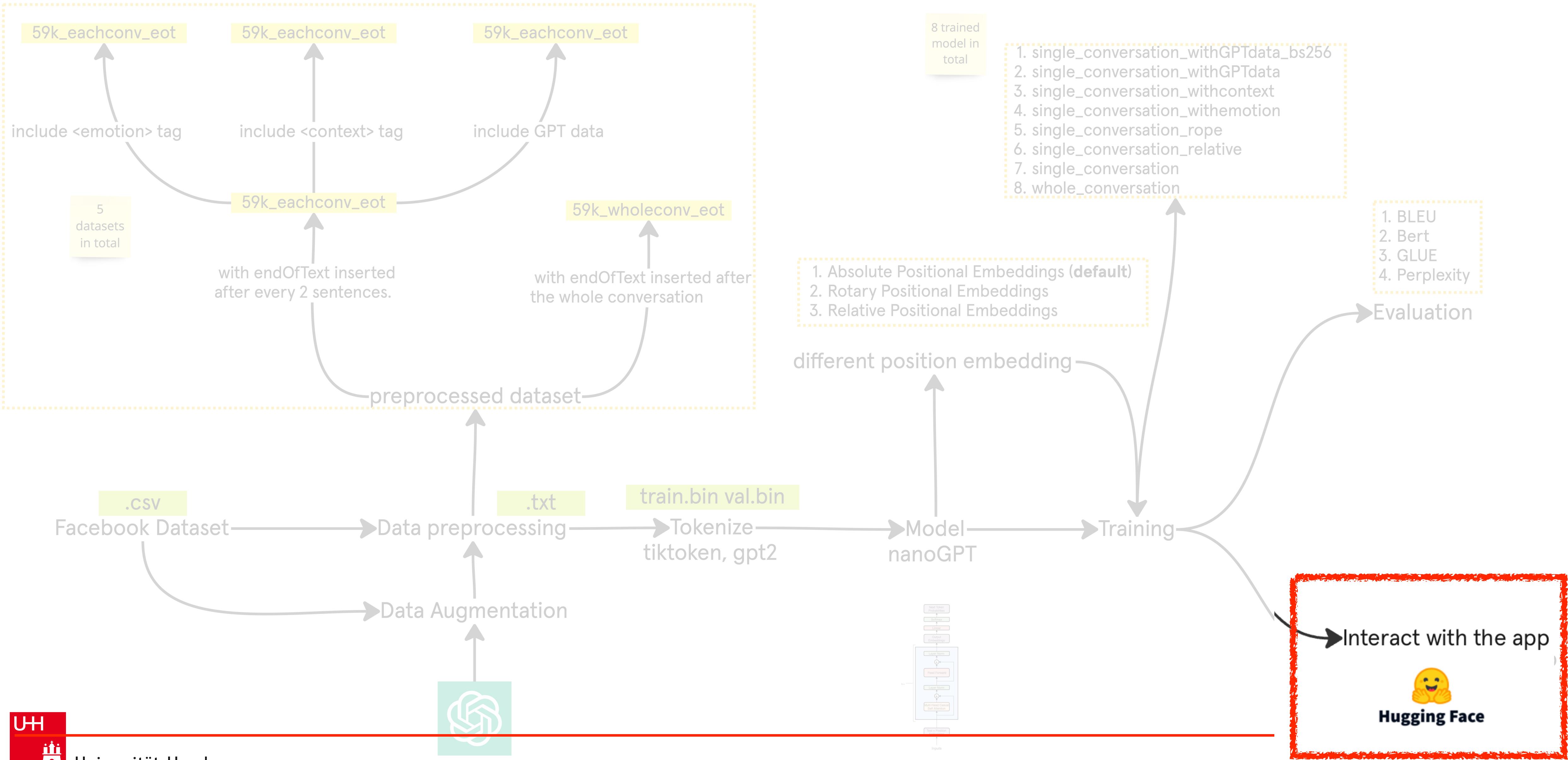
- Loss function: Cross-entropy loss
- Optimizer: AdamW
- Training machine: Mostly on Macbook Pro 2021 with chip M1 Pro
- Each model we trained 30k iteration
 - block size: the maximum context length we use for training, for GPT-2 it's 1024
 - block size = 256: took ~36 hours
 - block size = 64: took ~12 hours (our training default size)

Evaluation

- Metric
 - **GLUE**: General Language Understanding Evaluation. We use SST-2 (Sentiment Analysis) task, with emotion classifier (bhadresh-savani/distilbert-base-uncased-emotion), **7** classes of emotion are supported.
 - **BERT**: Use contextual embeddings (e.g. BERT) to assess semantic similarity between generated text and reference. (similarity: as the sum of the cosine similarities between their token embeddings.)
 - **Perplexity**: It quantifies the model's "surprise" when encountering new data. The lower the better.
 - **BLEU**: Bilingual Evaluation Understudy. Measures n-gram overlap between reference text and model-generated text. We calculate the average for BLEU-1,2,3 and 4.
- Data: 10% of the dataset (6k)

Evaluation Results

	BLEU (average)	Bert (F1)	GLUE	Perplexity
single_conversation_withGPTdata_bs256	0.0062	0.4865	0.3452	218114.8618
single_conversation_withGPTdata	0.0064	0.8576	0.4717	28233.3748
single_conversation_withcontext	0.0058	0.8359	0.4588	63035.0331
single_conversation_witthemotion	0.0058	0.8573	0.4721	2008198.4399
single_conversation_rope	0.0059	0.4809	0.3297	3776063.1443
single_conversation_relative	0.0060	0.3667	0.3297	inf
single_conversation	0.0059	0.8568	0.4727	84066.9524
whole_conversation	0.0055	0.8481	0.4360	29549.3370



The model is intractable via Huggingface

👤 Empathetic Dialogue System

😊 For the chatbot, we trained a small language model from scratch in our local machine. You can find the detailed information about our project in the Github repository [here](#).

Select a Model
Choose a pre-trained model to power the chatbot.

single_conversation_withGPTdata_bs256

🔍 Model Information

Model_name: single_conversation_withGPTdata_bs256

Description: We trained this model on Facebook Emotion Dialogues dataset with additional GPT data, using a batch size of 256.

Your AI-based chatbot powered by selected models!

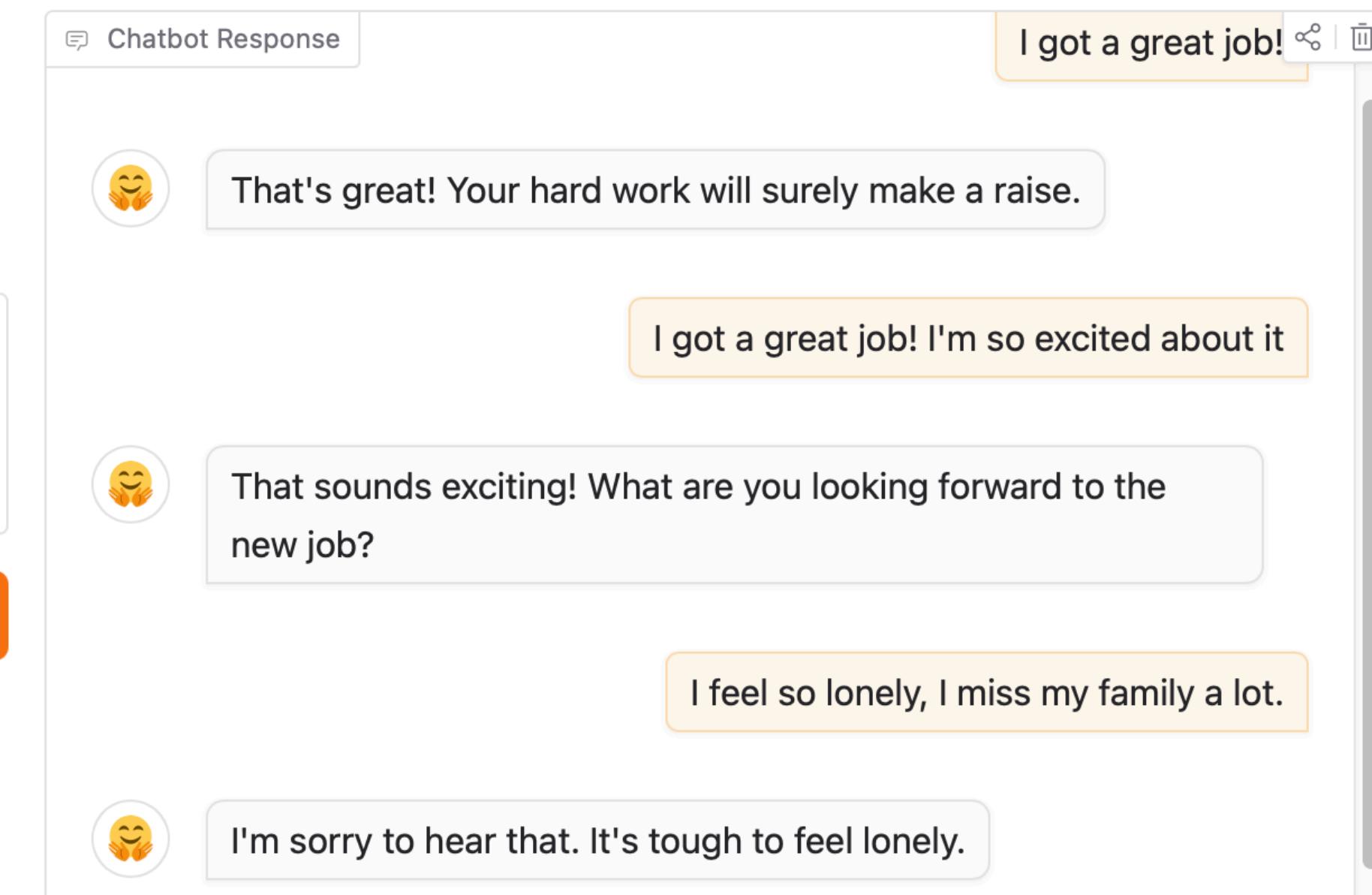
📝 Input Hints

1. Select a model from the dropdown list.
2. Type your message in the text box, please try to input a complete sentence.

User Input

I feel so lonely, I miss my family a lot.

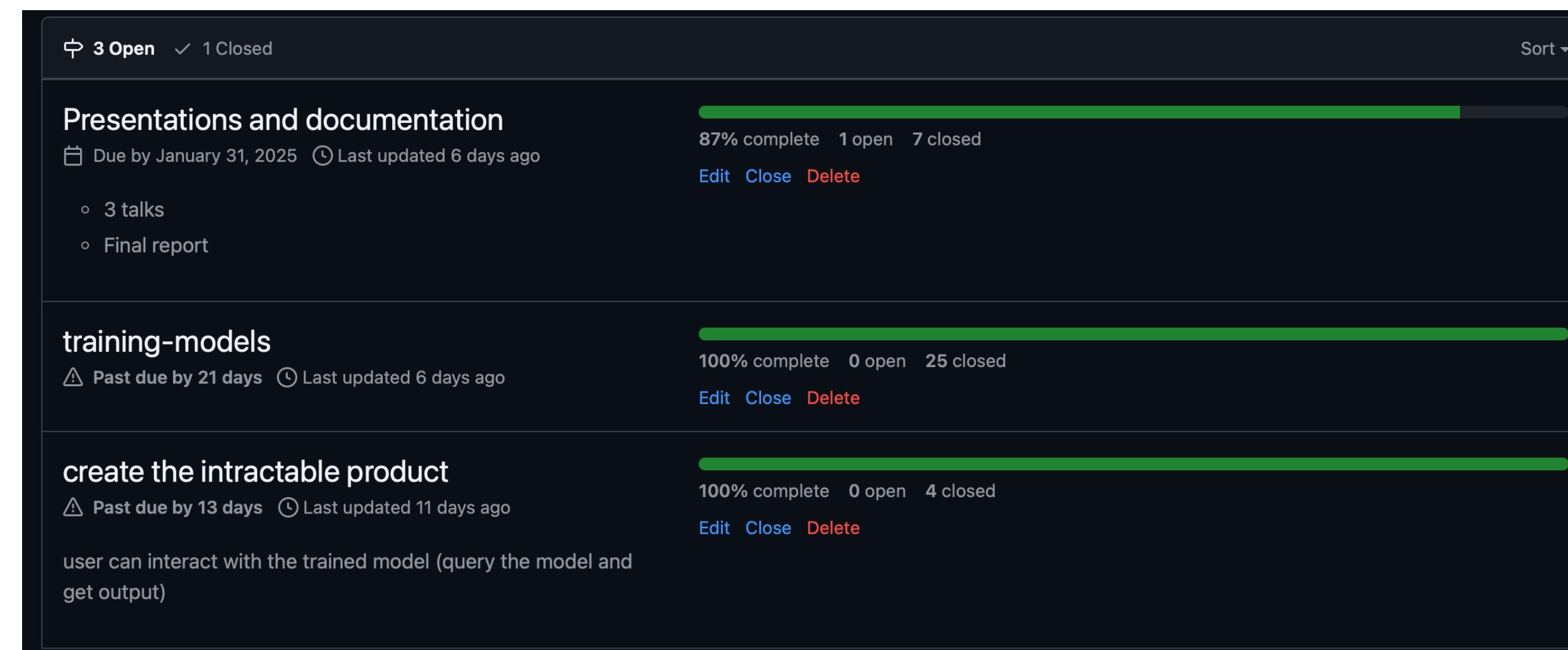
Clear **Submit**



Beyond the pipeline...

Teamwork & iterative development

- 3 milestones reached
- ~28 closed issues
- ~30 merged PRs



Author ▾	
1 Open	✓ 29 Closed
Fix Evaluation Metrics ✓	#56 by sofiagermer was merged 8 hours ago • Approved
Update readme and clean up of the project ✓	#54 by CallMeL was merged 5 days ago • Approved
Adding pe ✓ (model) (weekly-goal)	#52 by ryuruz was merged 8 hours ago • Approved ⌘ training-models
⚠ Adding pe (model) (weekly-goal)	#51 by ryuruz was closed last week • Review required ⌘ training-models
Eda gptdata ✓ (data) (documentation) (weekly-goal)	#50 by sofiagermer was merged 2 weeks ago • Approved
Add website UI for selecting models ✓ (App)	#48 by CallMeL was merged 2 weeks ago • Approved ⌘ create the intra...
Train on larger block_size with GPT data and small improvement ✓ (model)	#46 by CallMeL was merged on Dec 18, 2024 • Approved ⌘ training-models
Add RoPE ✓ (model) (weekly-goal)	#44 by ryuruz was merged on Dec 11, 2024 • Approved ⌘ training-models
Train on data with emotion and context and evaluate on test dataset ✓ (model)	#42 by CallMeL was merged on Dec 10, 2024
finish generating the 59 k conversation ✓	#41 by sofiagermer was merged on Dec 12, 2024
updated md and slides for midterm talk ✓ (data) (documentation)	#36 by ryuruz was merged on Dec 5, 2024 • Approved ⌘ Presentations a...
Train on new labeled data and evaluate ✓ (model) (weekly-goal)	#35 by CallMeL was merged on Dec 4, 2024 • Approved ⌘ training-models

Our work in a nutshell

- Data preprocessing + augment with ChatGPT API
- Modified the position embedding in the model
- Trained models in “different” dataset (8 in total)
- Evaluate the results with BERTscore, GLUE, BLEU, Perplexity
- Make the trained model intractable with a nice Web UI

Future Work

- More data for training that cover more topics
- Improve the training logs
- Improve the evaluation part (e.g. involve human evaluation, compare with other trained models)

Before the demo...

Take-away messages

- Not every problem requires a language model—choose solutions wisely
- Setting up agile development is challenging but immensely rewarding in the long run.
- Dive into the data—explore it thoroughly and **debug outputs** completely. (<eot> in our case)
- More data can improve language models, but be careful about the data you use to generate (i.e. training and validation datasets)
- Training logs are invaluable—monitor them closely for insights and improvements.

Demo 😊

- <https://huggingface.co/spaces/HannahLin271/nanoGPT>



