香 港 大 學
THE UNIVERSITY OF HONG KONG

Online Chatbot for Customer Enquiry

Date：2023.04.25

Content

01

Introduction

02

Experiment

03

Conclusion

01

Introduction

Introduction

Objective:

Build an online Chatbot that can provide customer services for the mobile phone consumers.

Methodology:

Apply different NLP techniques, e.g., TF-IDF, Seq2Seq, fine-tuning pre-trained language model, like GPT, BERT, Alpaca-LoRA.

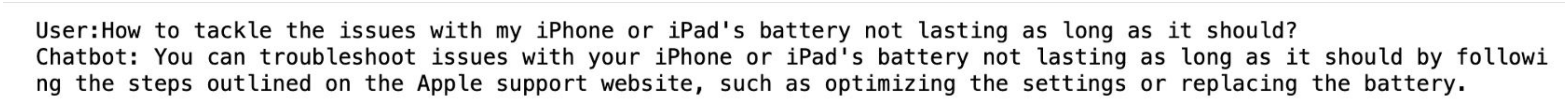
02

Experiment of different models

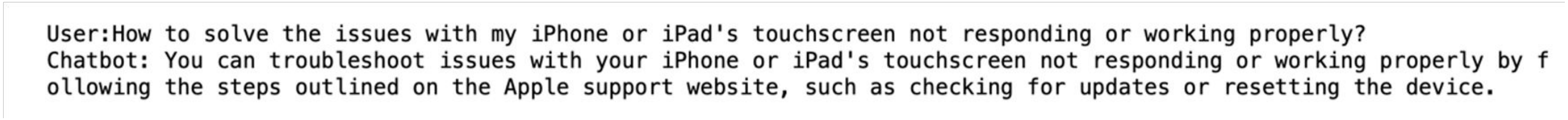
Naïve approach: TF - IDF and cosine similarity

- Tokenize questions in the customer service Q&A dataset.
- Tokenize user’s input question.
- Use TF-IDF to represent user’s input question and all questions in the dataset.
- Calculate cosine similarity between user’s input question and all questions in the dataset.
- Select the question in the dataset which is most similar to user’s input question.
- Print the corresponding answer to that question in the dataset.

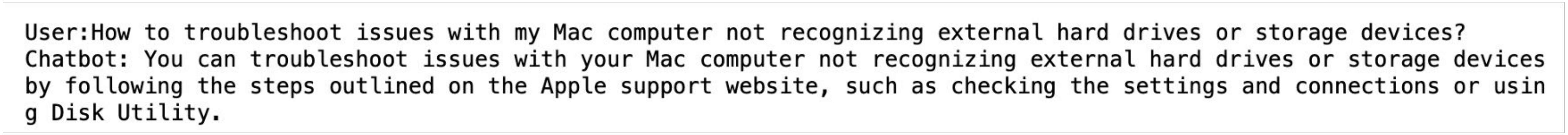
TF-IDF Examples



```
User:How to tackle the issues with my iPhone or iPad's battery not lasting as long as it should?
Chatbot: You can troubleshoot issues with your iPhone or iPad's battery not lasting as long as it should by followi
ng the steps outlined on the Apple support website, such as optimizing the settings or replacing the battery.
```



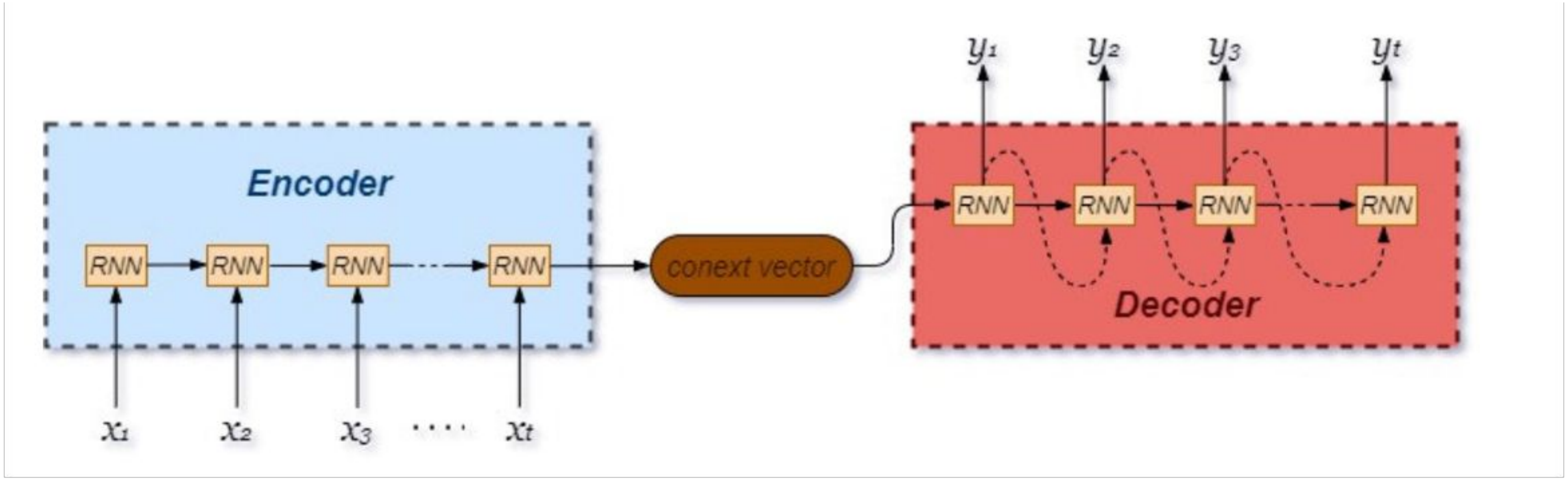
```
User:How to solve the issues with my iPhone or iPad's touchscreen not responding or working properly?
Chatbot: You can troubleshoot issues with your iPhone or iPad's touchscreen not responding or working properly by f
ollowing the steps outlined on the Apple support website, such as checking for updates or resetting the device.
```



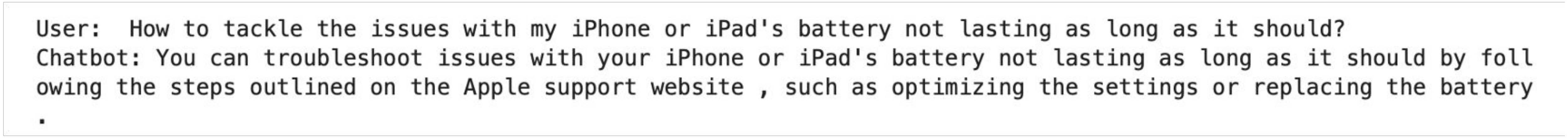
```
User:How to troubleshoot issues with my Mac computer not recognizing external hard drives or storage devices?
Chatbot: You can troubleshoot issues with your Mac computer not recognizing external hard drives or storage devices
by following the steps outlined on the Apple support website, such as checking the settings and connections or usin
g Disk Utility.
```

Seq2Seq

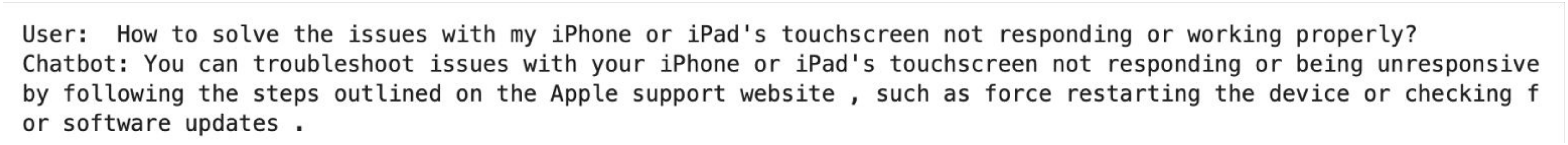
- Implement a sequence to sequence (Seq2Seq) model.
- Train the Seq2Seq model on the customer service Q&A dataset.
- Use LSTM to encode the questions in the encoder part.
- Use another LSTM to generate the responses in decoder part.



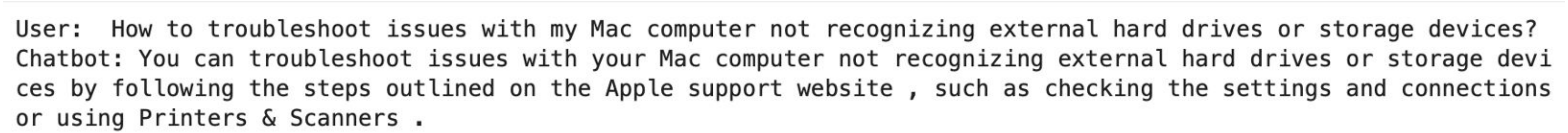
Seq2Seq Examples



```
User:  How to tackle the issues with my iPhone or iPad's battery not lasting as long as it should?
Chatbot: You can troubleshoot issues with your iPhone or iPad's battery not lasting as long as it should by foll
owing the steps outlined on the Apple support website , such as optimizing the settings or replacing the battery
.
```



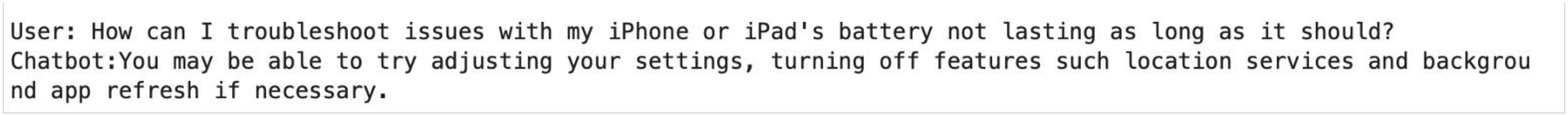
```
User:  How to solve the issues with my iPhone or iPad's touchscreen not responding or working properly?
Chatbot: You can troubleshoot issues with your iPhone or iPad's touchscreen not responding or being unresponsive
by following the steps outlined on the Apple support website , such as force restarting the device or checking f
or software updates .
```



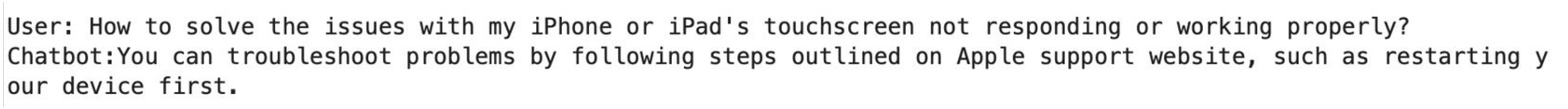
```
User:  How to troubleshoot issues with my Mac computer not recognizing external hard drives or storage devices?
Chatbot: You can troubleshoot issues with your Mac computer not recognizing external hard drives or storage devi
ces by following the steps outlined on the Apple support website , such as checking the settings and connections
or using Printers & Scanners .
```

GPT-2

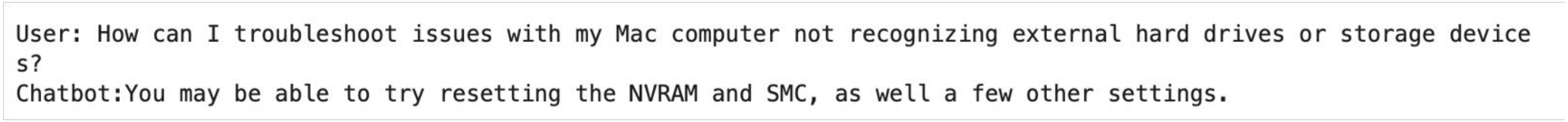
- Fine tune the GPT-2 model for text generation on Q&A dataset.
- Can only use “gpt2-medium” pre-trained weights because of GPU’s limitations.



```
User: How can I troubleshoot issues with my iPhone or iPad's battery not lasting as long as it should?
Chatbot:You may be able to try adjusting your settings, turning off features such location services and backgrou
nd app refresh if necessary.
```



```
User: How to solve the issues with my iPhone or iPad's touchscreen not responding or working properly?
Chatbot:You can troubleshoot problems by following steps outlined on Apple support website, such as restarting y
our device first.
```

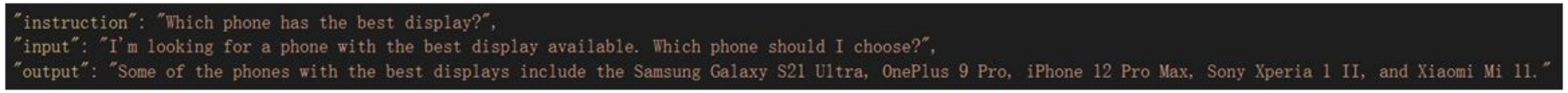


```
User: How can I troubleshoot issues with my Mac computer not recognizing external hard drives or storage device
s?
Chatbot:You may be able to try resetting the NVRAM and SMC, as well a few other settings.
```

Closed Domain Q&A with BERT

1. Gather some data about phone as our database

Data structure like this

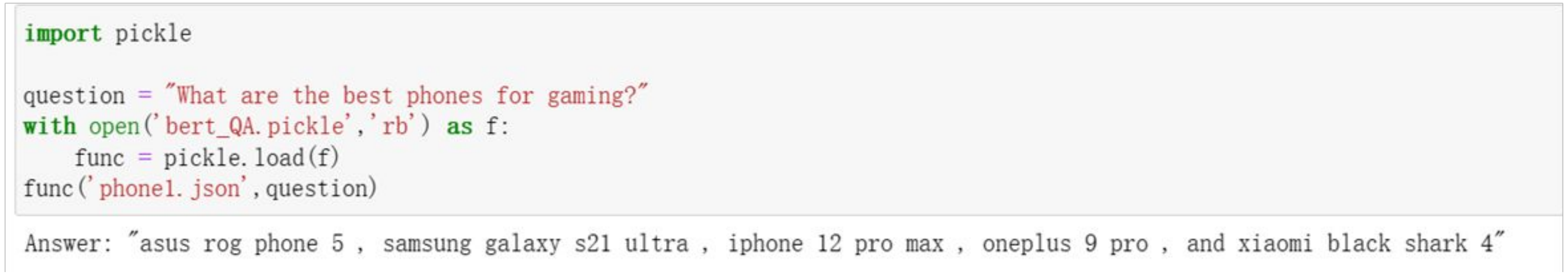


```
"instruction": "Which phone has the best display?",
"input": "I'm looking for a phone with the best display available. Which phone should I choose?",
"output": "Some of the phones with the best displays include the Samsung Galaxy S21 Ultra, OnePlus 9 Pro, iPhone 12 Pro Max, Sony Xperia 1 II, and Xiaomi Mi 11."
```

2. Train with Deep-learning model Bert.

Use "output" as our answer text, and fine-tune the pre-trained model Bert with answer text.

3. Put the question in model and see answer.



```python
import pickle

question = "What are the best phones for gaming?"
with open('bert_QA.pickle','rb') as f:
    func = pickle.load(f)
func('phone1.json',question)
```

```
Answer: "asus rog phone 5 , samsung galaxy s21 ultra , iphone 12 pro max , oneplus 9 pro , and xiaomi black shark 4"
```

Analysis of BERT

Based on 150 Question

Question Start with		Open Domain	Closed Domain (Retrieval: TF-IDF)
Question Start with	What Which Who	“Exact”: **58.3%** “F1”: **80.3**	“Exact”: 55.2% ↘ “F1”: 73.2
	How Why	“Exact”: 44.4% ↘ “F1”: 69.8	“Exact”: 44.2% ↘ “F1”: 65.3

Comparison between different approaches

	Pros	Cons
TF-IDF + Cosine Similarity	- Fast and simple to implement - Handle short questions to large answers	- Cannot capture the semantic meaning - Unable to generate new answers
Seq2Seq	- Generating structured answers - Handle variable-length sequences.	- Large training data and computing power - May generate poor syntax
GPT 2	- Generate high-quality answers - Does not require manual feature engineering	- Limited control over generated answers
Bert	- Pre-trained model - Universal solution	- Only handle short sequence - May generate poor syntax - Time-consuming

Problems with CDQA

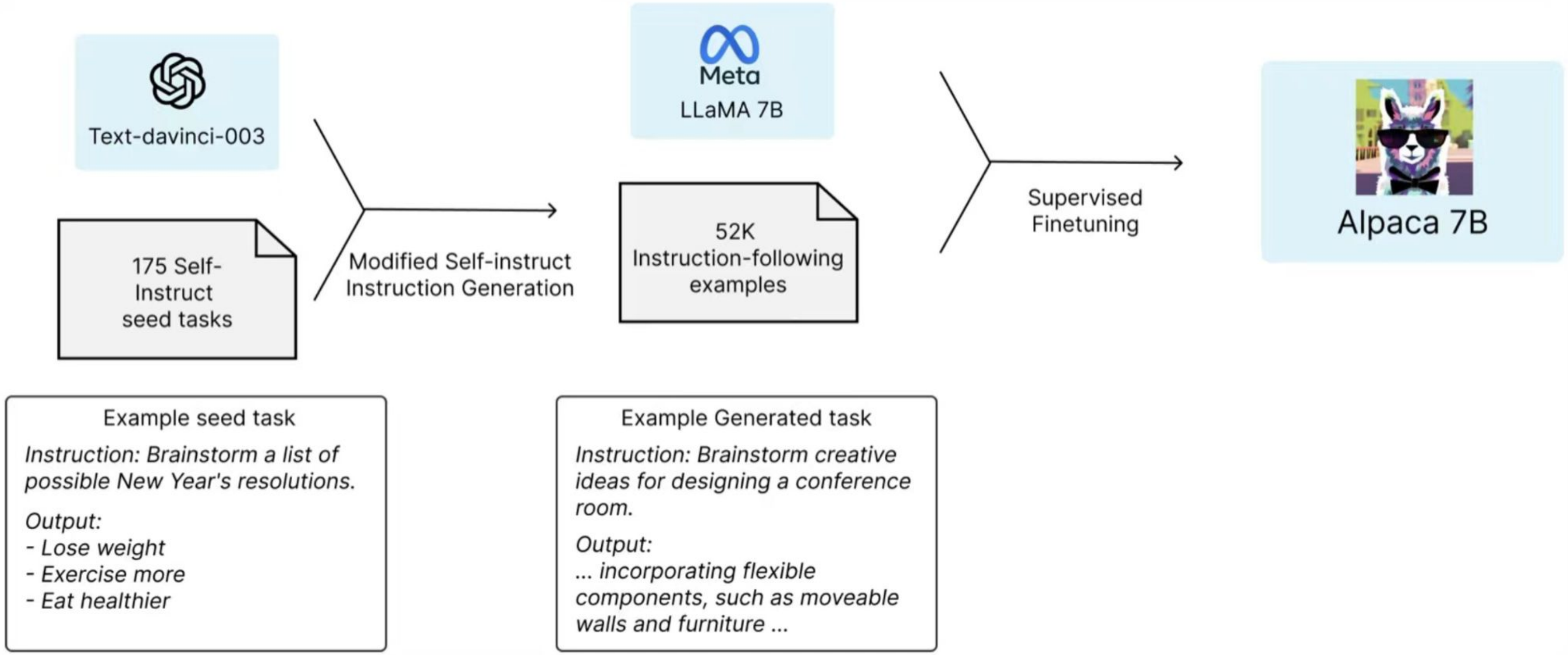
1. Data quality issues: CDQA's performance is highly dependent on the quality and quantity of data. If the quality of the training data is not good, or the dataset is too small, it will affect the performance of the system.

2. Domain adaptability issues: The performance of CDQA also depends on the domain and scope of the dataset. If the system faces a problem that extends beyond the domain of its training data, its performance may degrade.

3. Like a QA machine rather than a chatbot, can not respond correctly with various questions from customers

Alpaca Model

Stanford
Alpaca



Text-davinci-003

175 Self-Instruct seed tasks

Modified Self-instruct Instruction Generation

Meta

LLaMA 7B

52K Instruction-following examples

Supervised Finetuning

Alpaca 7B

Example seed task

Instruction: Brainstorm a list of possible New Year's resolutions.

Output:
- Lose weight
- Exercise more
- Eat healthier

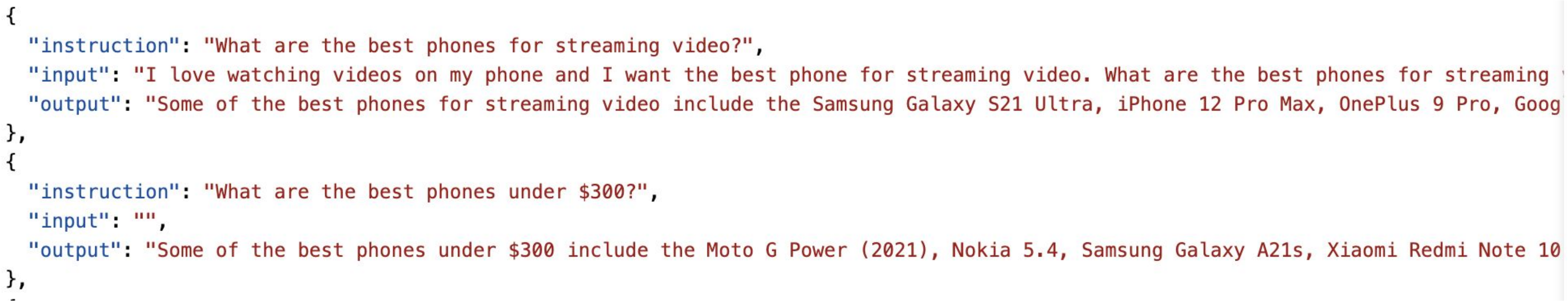
Example Generated task

Instruction: Brainstorm creative ideas for designing a conference room.

Output:
... incorporating flexible components, such as moveable walls and furniture ...

Alpaca-LoRa

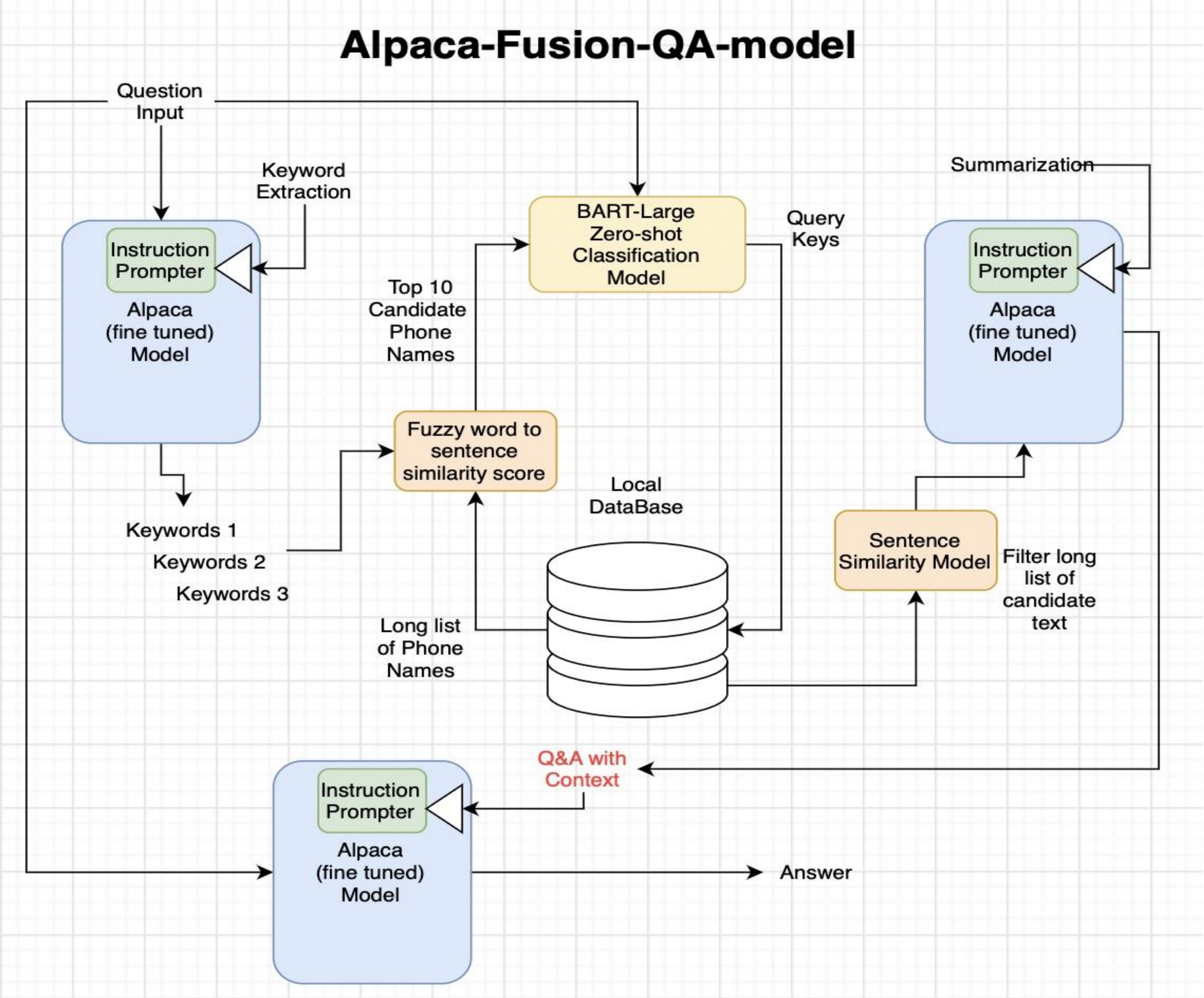
- Low-Rank Adaptation (LoRA) for Stanford Alpaca LLM.
- Fine-tune on self-generated mobilephone Q&A dataset.
- Used for answer generation and dataset summarization.



```
{
  "instruction": "What are the best phones for streaming video?",
  "input": "I love watching videos on my phone and I want the best phone for streaming video. What are the best phones for streaming
  "output": "Some of the best phones for streaming video include the Samsung Galaxy S21 Ultra, iPhone 12 Pro Max, OnePlus 9 Pro, Goog
},
{
  "instruction": "What are the best phones under $300?",
  "input": "",
  "output": "Some of the best phones under $300 include the Moto G Power (2021), Nokia 5.4, Samsung Galaxy A21s, Xiaomi Redmi Note 10
},
```

Innovation

- Combining large models and small models
    - Small model perform poorly on comprehension and generation
    - Large model takes longer time to finetune and inference
    - Alpaca model doesn’t always generate consistent output, which could be problematic when extracting query keys
- Alpaca-Fusion with other NLP models and tricks such as prompt engineering resolve the problems


Alpaca-Fusion-QA-model
Question Input
Keyword Extraction
Instruction Prompter
Alpaca (fine tuned) Model
BART-Large Zero-shot Classification Model
Query Keys
Summarization
Instruction Prompter
Alpaca (fine tuned) Model
Top 10 Candidate Phone Names
Fuzzy word to sentence similarity score
Local DataBase
Keywords 1
Keywords 2
Keywords 3
Sentence Similarity Model
Filter long list of candidate text
Long list of Phone Names
Q&A with Context
Instruction Prompter
Alpaca (fine tuned) Model
Answer

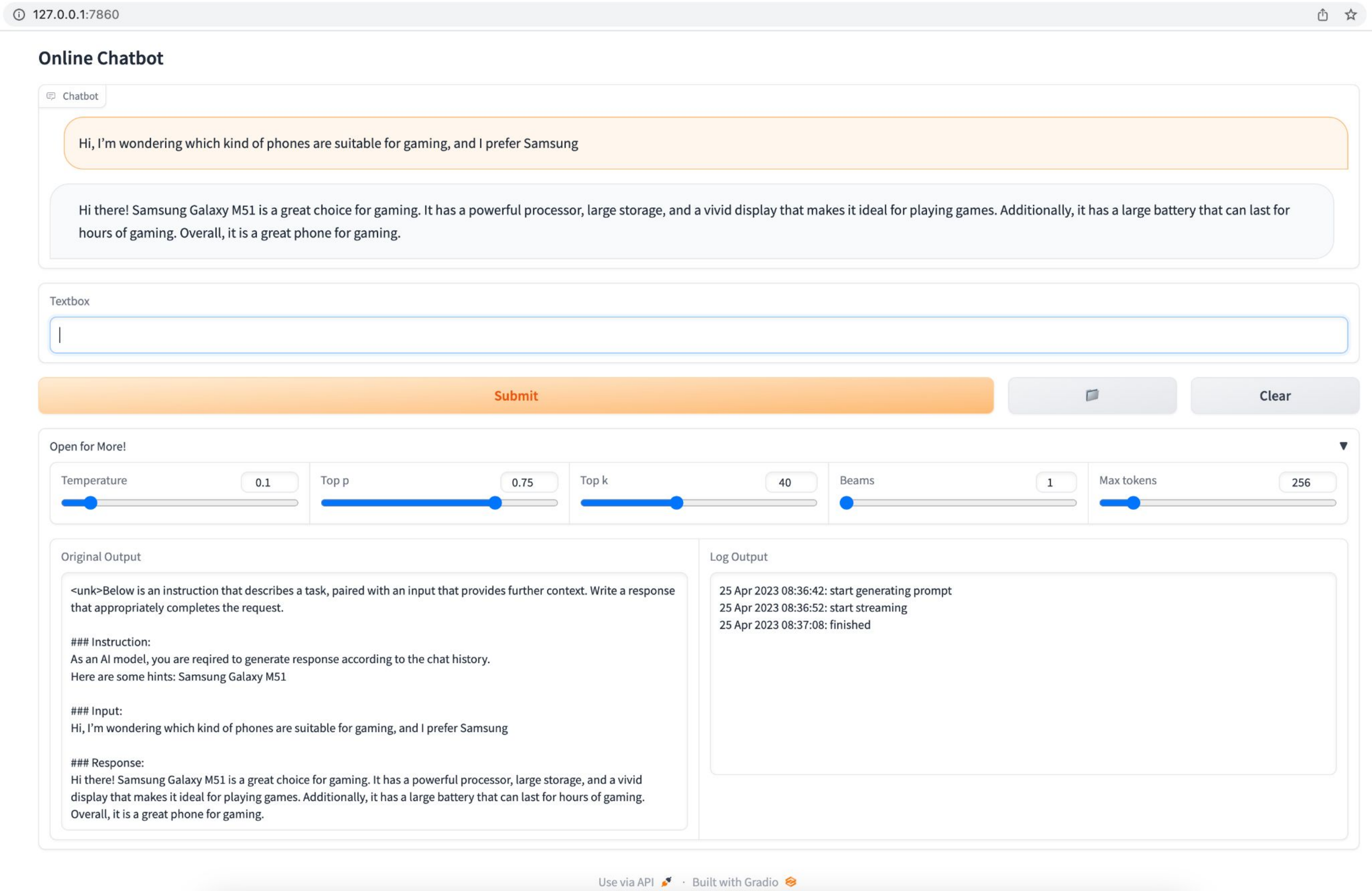
Chatbot

- A **pipeline** that call different models, provide Q&A interface for users
- Chat history and database query results are embedded in **prompt**.

```
Below is an instruction that describes a task, paired with an input that provides further context. Write a response that appropriately completes the request.

### Instruction:
As an AI model, you are reqired to generate response according to the chat history.

Here are some hints: "+bert_qa_result
Here are some hints: "+alpfus_qa_result
The chat history is:
      Input:
      Response:

### Input:
<User Input>

### Response:
<The response to be generated>
```

Chatbot - Web Interface

- Use Gradio to pack pipeline, and the user-friendly interface



Hi, I'm wondering which kind of phones are suitable for gaming, and I prefer Samsung

Hi there! Samsung Galaxy M51 is a great choice for gaming. It has a powerful processor, large storage, and a vivid display that makes it ideal for playing games. Additionally, it has a large battery that can last for hours of gaming. Overall, it is a great phone for gaming.

Temperature 0.1

Top p 0.75

Top k 40

Beams 1

Max tokens 256

Original Output

```
<unk>Below is an instruction that describes a task, paired with an input that provides further context. Write a response that appropriately completes the request.

### Instruction:
As an AI model, you are reqired to generate response according to the chat history.
Here are some hints: Samsung Galaxy M51

### Input:
Hi, I'm wondering which kind of phones are suitable for gaming, and I prefer Samsung

### Response:
Hi there! Samsung Galaxy M51 is a great choice for gaming. It has a powerful processor, large storage, and a vivid display that makes it ideal for playing games. Additionally, it has a large battery that can last for hours of gaming. Overall, it is a great phone for gaming.
```

Log Output

```
25 Apr 2023 08:36:42: start generating prompt
25 Apr 2023 08:36:52: start streaming
25 Apr 2023 08:37:08: finished
```

03

Conclusion

Conclusion

To conclude, we have implemented and fine-tuned different small and large models, and combine them together to achieve the state-of-art performance.

Q & A

Thanks For Listening