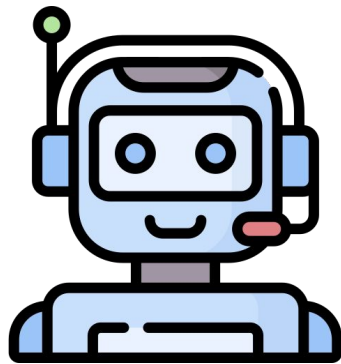


‘음성 챗봇 지우’

모듈프로젝트3 5조
고종현 김성언 박수빈 이가희 홍민택

1. 제품 소개
2. 시장 분석
3. 업무 분장
4. 개발 일정
5. 개발
6. 제품 시연
7. 개선 방안



정신의학과 환자에게 제공되는 ChatBot

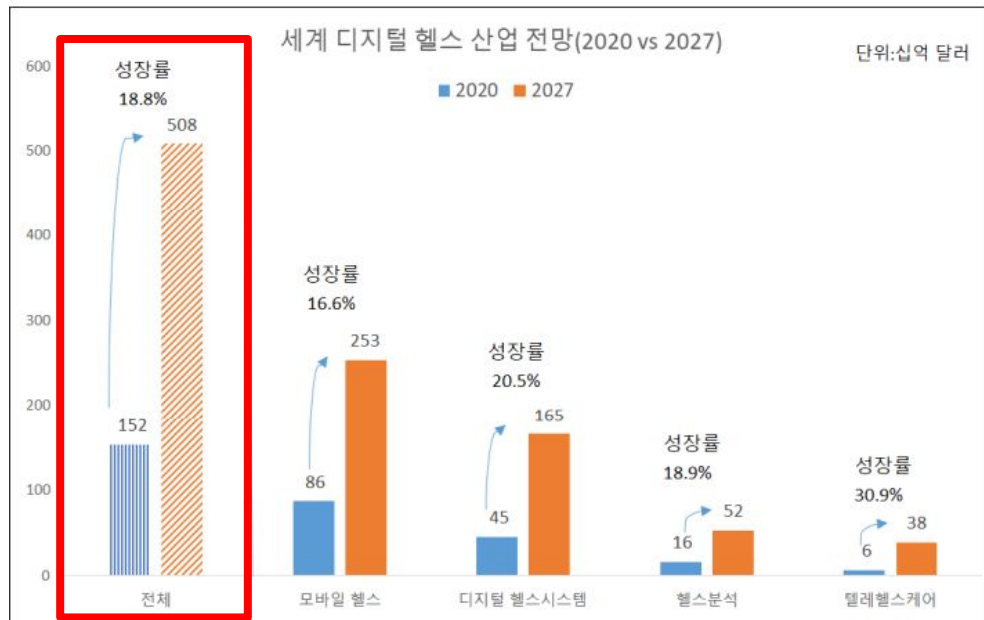
시장 분석

분류	하위분류	설명
텔레헬스케어 (Telehealthcare)	<ul style="list-style-type: none">• 텔레케어 (먼거리에서 모니터링 지원)• 텔레헬스 (환자와 의사간 교류)	환자와 의료진 혹은 의료진과 의료진 사이의 임상데이터의 원거리 교류와 먼거리에서 ICT를 활용하는데 필요한 지원하고 케어를 제공
모바일 헬스 (mHealth)	<ul style="list-style-type: none">• 모바일 어플리케이션• 웨어러블 기기	헬스(Health)와 웰빙(Wellbeing) 관련 모바일 앱(app)과 연동된 웨어러블 기기(wearable devices)로 정의
헬스분석 (Health Analytics)	<ul style="list-style-type: none">• 유전자 분석• 정밀의료• 데이터분석(CDS(Clinical Decision Support)알고리즘, AI 등)	빅데이터를 이해하는데 필요한 소프트웨어 솔루션과 분석능력
디지털 헬스시스템 (Digital health systems)	<ul style="list-style-type: none">• 병원주도의료기록• 환자주도의료기록(PHR, myData 등)	디지털 헬스정보의 저장과 디지털화된 환자의료기록의 교류

시장 분석

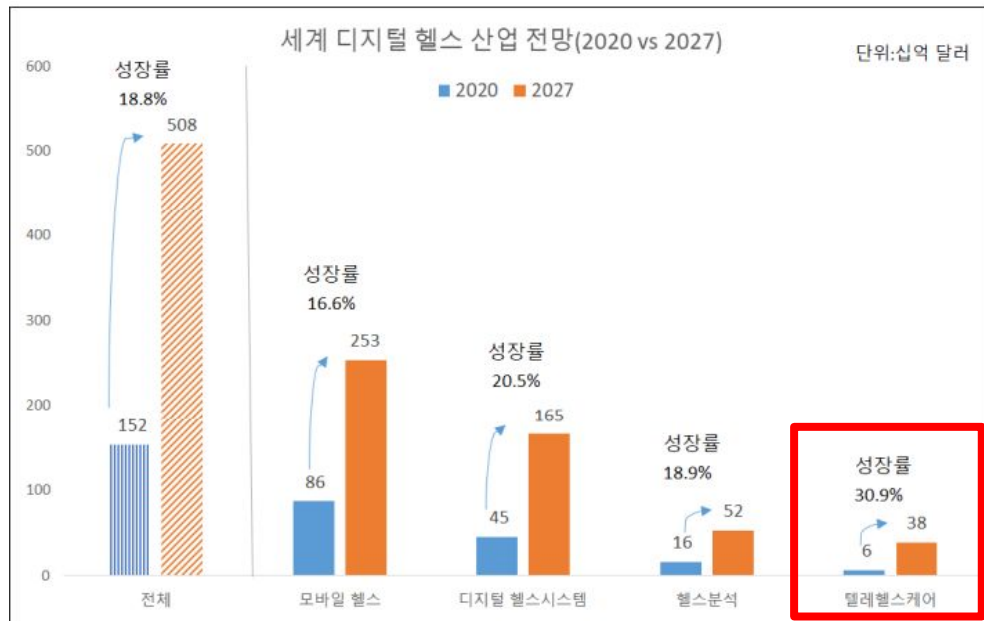
분류	하위분류	설명
텔레헬스케어 (Telehealthcare)	<ul style="list-style-type: none">• 텔레케어 (먼거리에서 모니터링 지원)• 텔레헬스 (환자와 의사간 교류)	환자와 의료진 혹은 의료진과 의료진 사이의 임상데이터의 원거리 교류와 먼거리에서 ICT를 활용하는데 필요한 지원하고 케어를 제공
모바일 헬스 (mHealth)	<ul style="list-style-type: none">• 모바일 어플리케이션• 웨어러블 기기	헬스(Health)와 웰빙(Wellbeing) 관련 모바일 앱(app)과 연동된 웨어러블 기기(wearable devices)로 정의
헬스분석 (Health Analytics)	<ul style="list-style-type: none">• 유전자 분석• 정밀의료• 데이터분석(CDS(Clinical Decision Support)알고리즘, AI 등)	빅데이터를 이해하는데 필요한 소프트웨어 솔루션과 분석능력
디지털 헬스시스템 (Digital health systems)	<ul style="list-style-type: none">• 병원주도의료기록• 환자주도의료기록(PHR, myData 등)	디지털 헬스정보의 저장과 디지털화된 환자의료기록의 교류

시장 분석

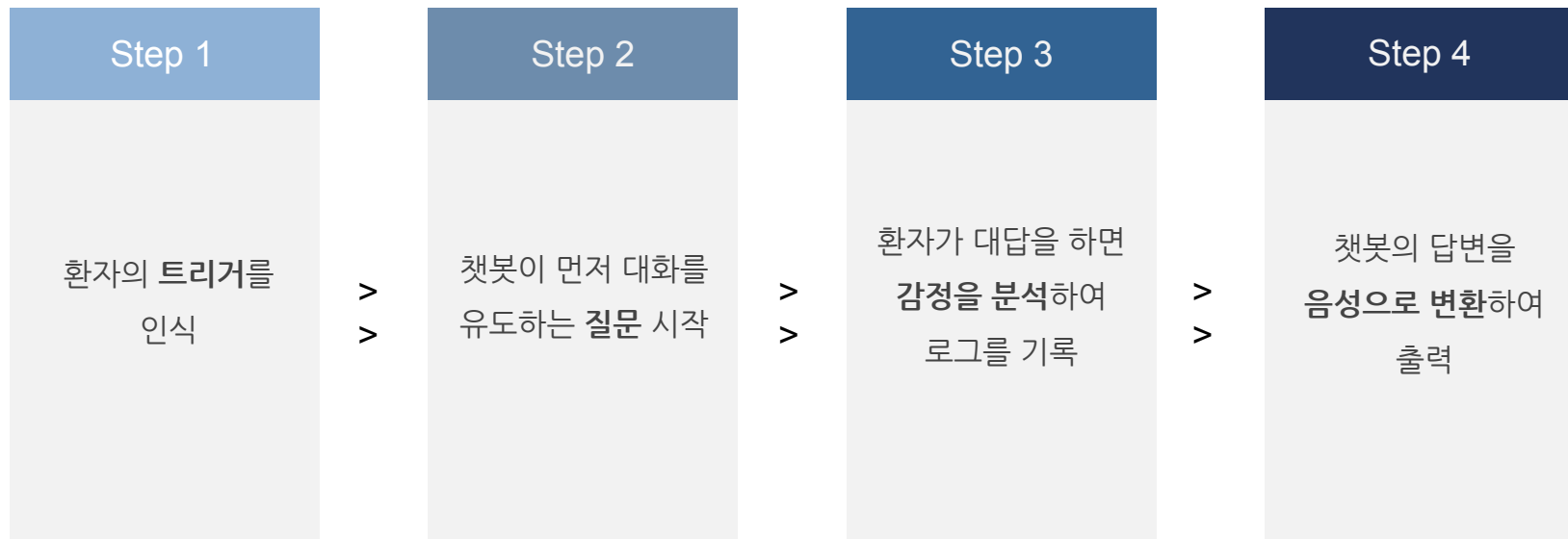


- 시장조사업체에 의하면 세계 디지털 헬스 산업은 2020년 1520억 달러 규모이며, 2027년에는 5080억 달러 규모로 **큰 폭의 성장률(18.8%)**이 예상 되는 분야임.
- 모바일 헬스 산업은 전체의 57%(86십억 달러, 2020년)로 절반 이상을 차지하며, 텔레헬스케어는 전체의 4%로 규모가 작으나 성장률은 30.9%로 가장 높게 전망됨

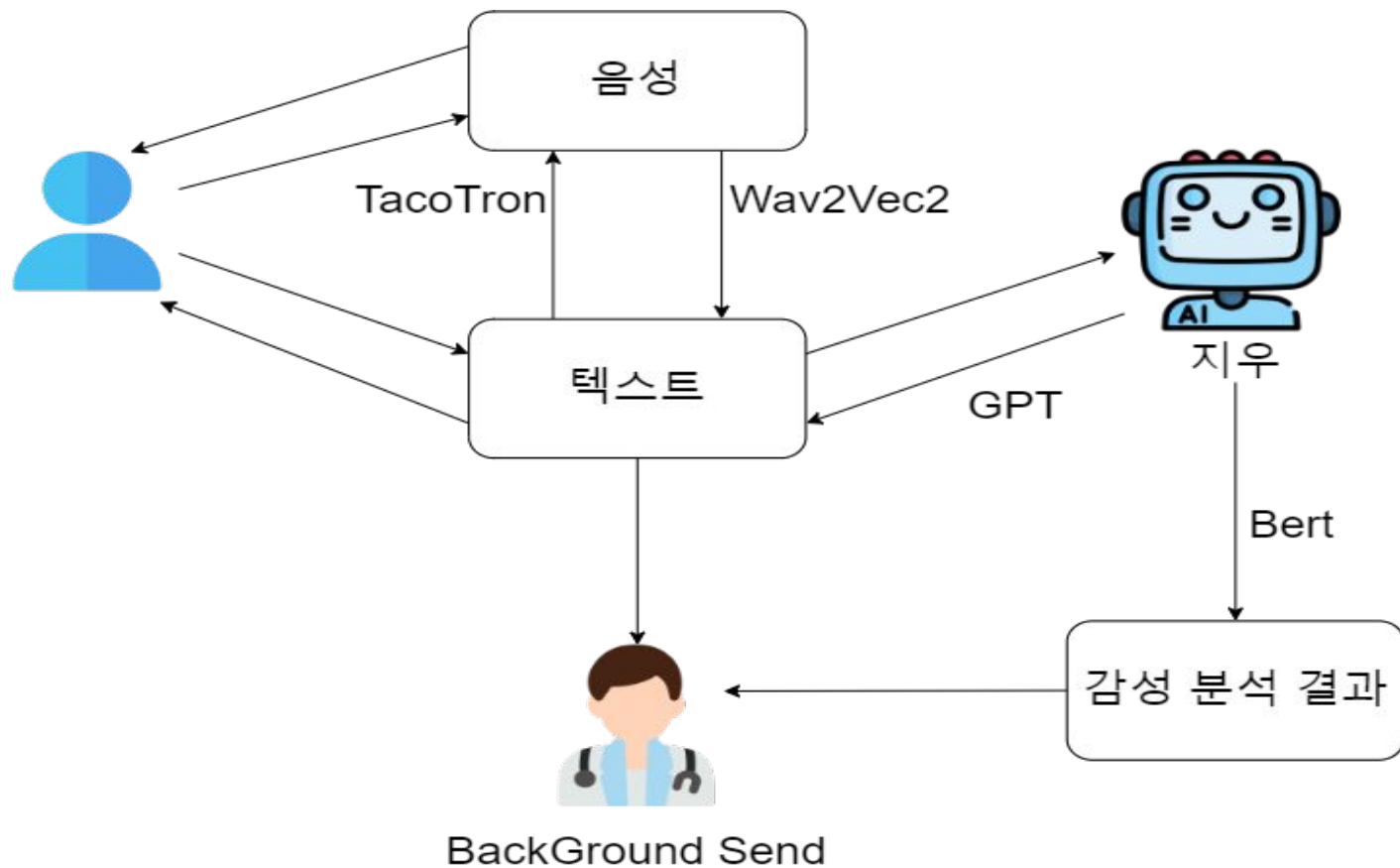
시장 분석



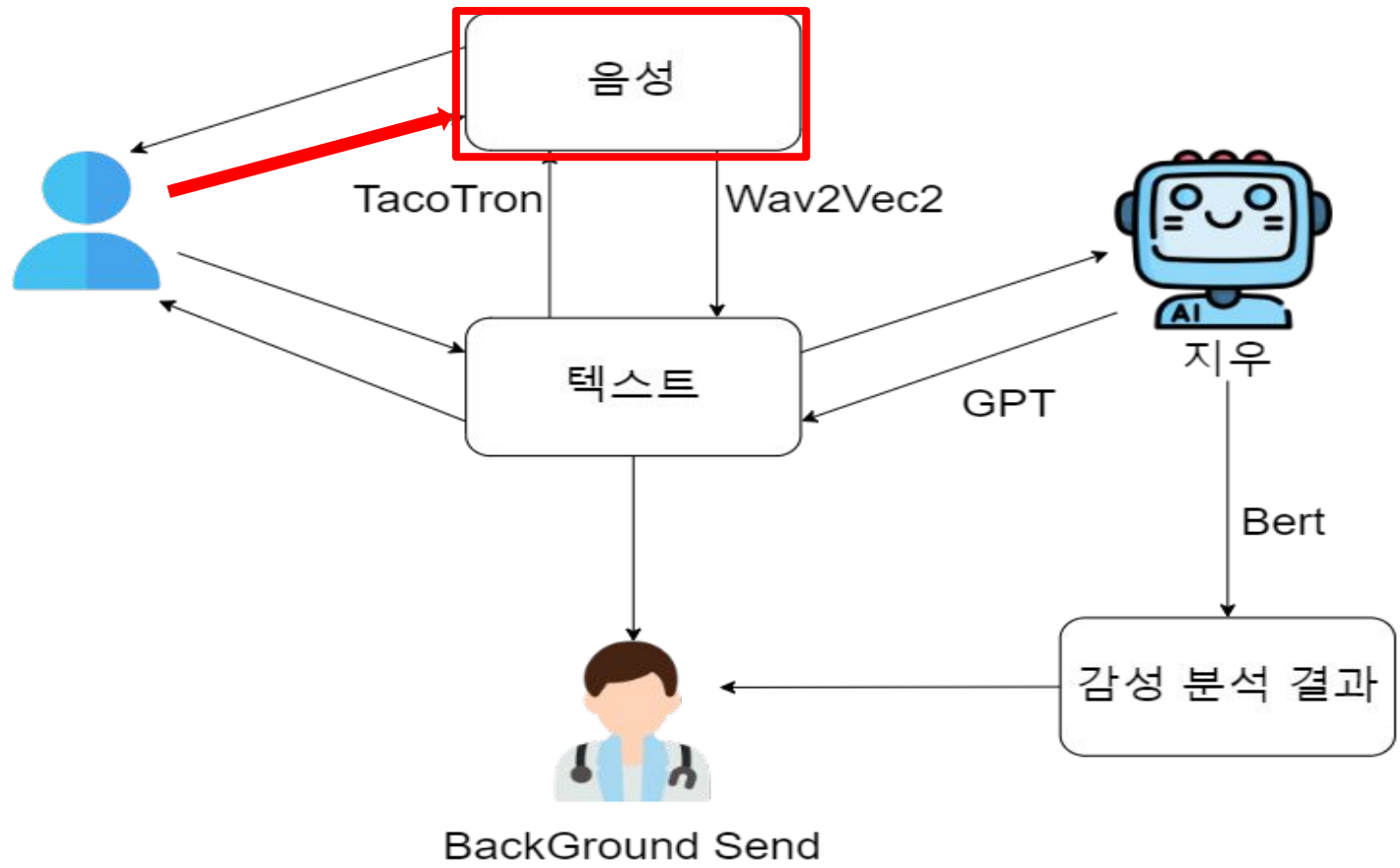
- 시장조사업체에 의하면 세계 디지털 헬스 산업은 2020년 1520억 달러 규모이며, 2027년에는 5080억 달러 규모로 **큰 폭의 성장률(18.8%)**이 예상 되는 분야임.
- 모바일 헬스 산업은 전체의 57%(86십억 달러, 2020년)로 절반 이상을 차지하며, 텔레헬스케어는 전체의 4%로 규모가 작으나 성장률은 30.9%로 가장 높게 전망됨



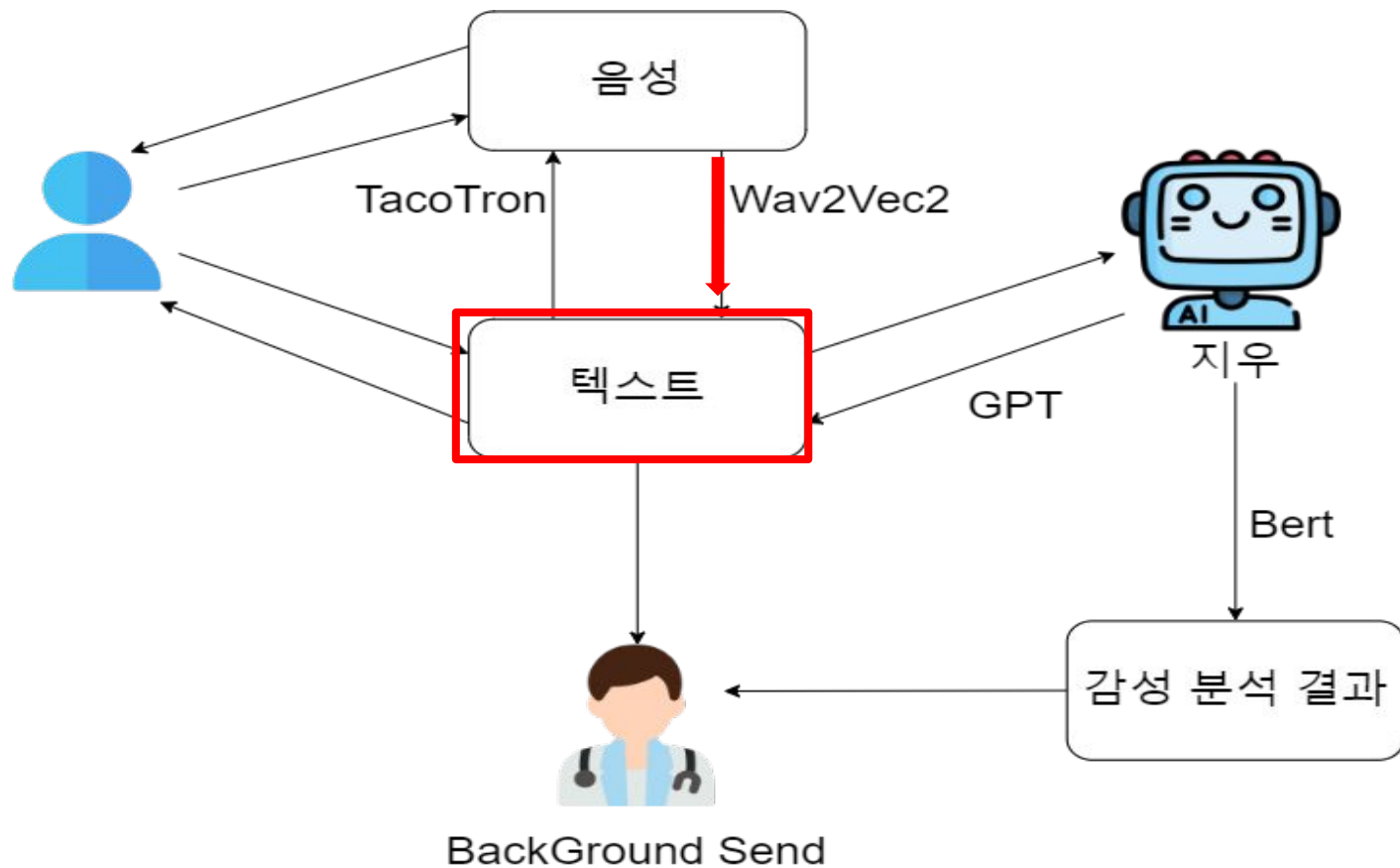
순서도



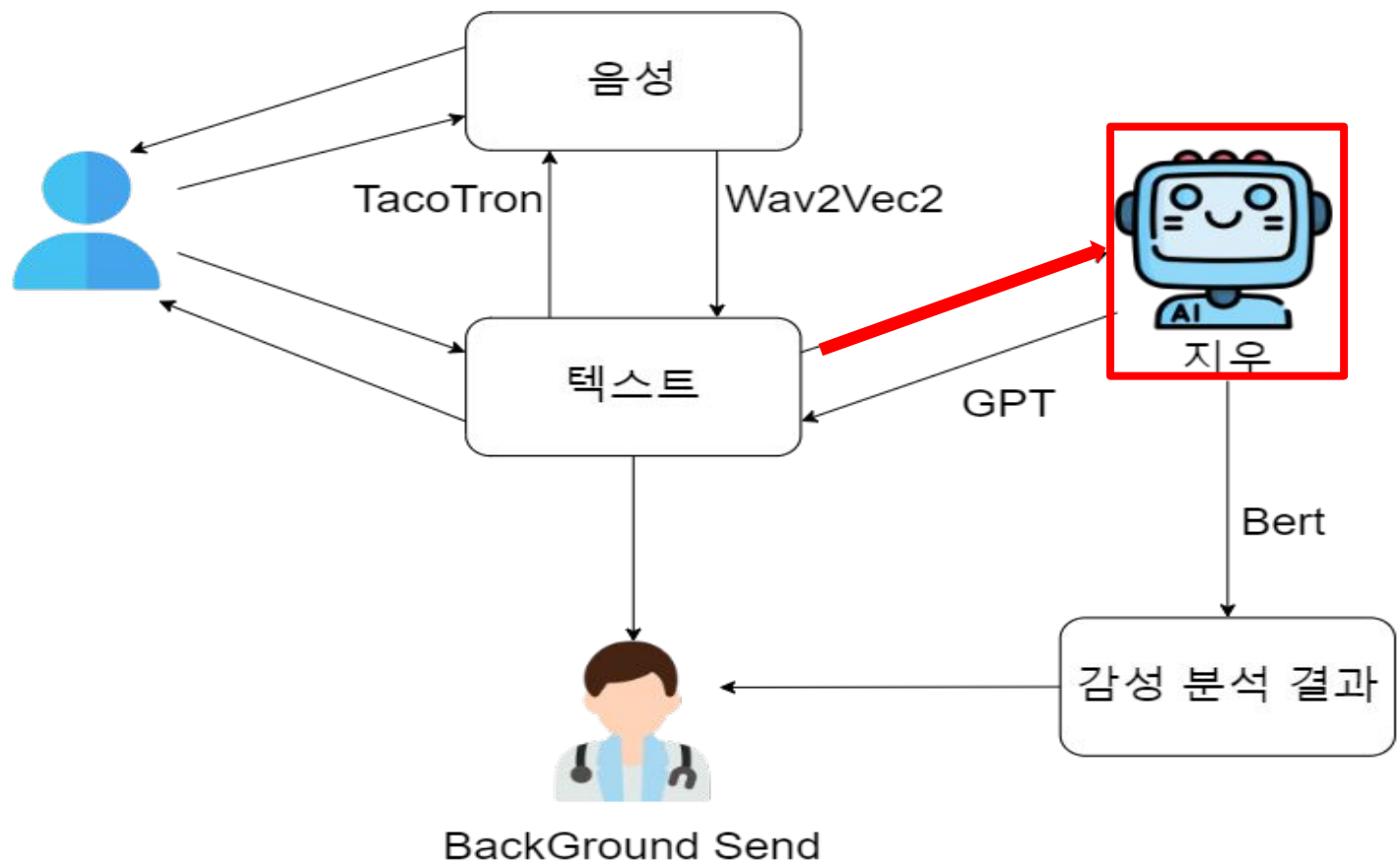
순서도



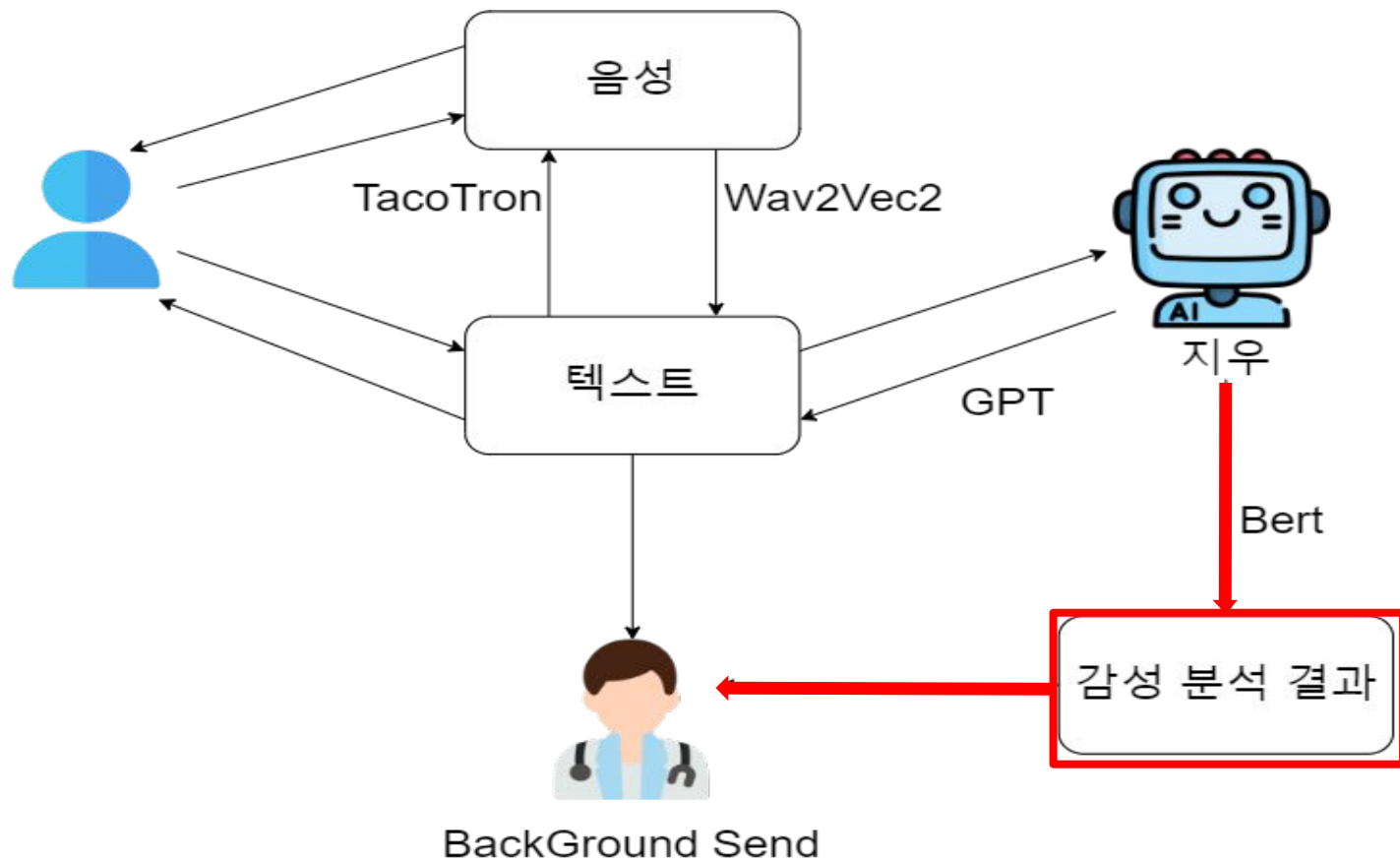
순서도



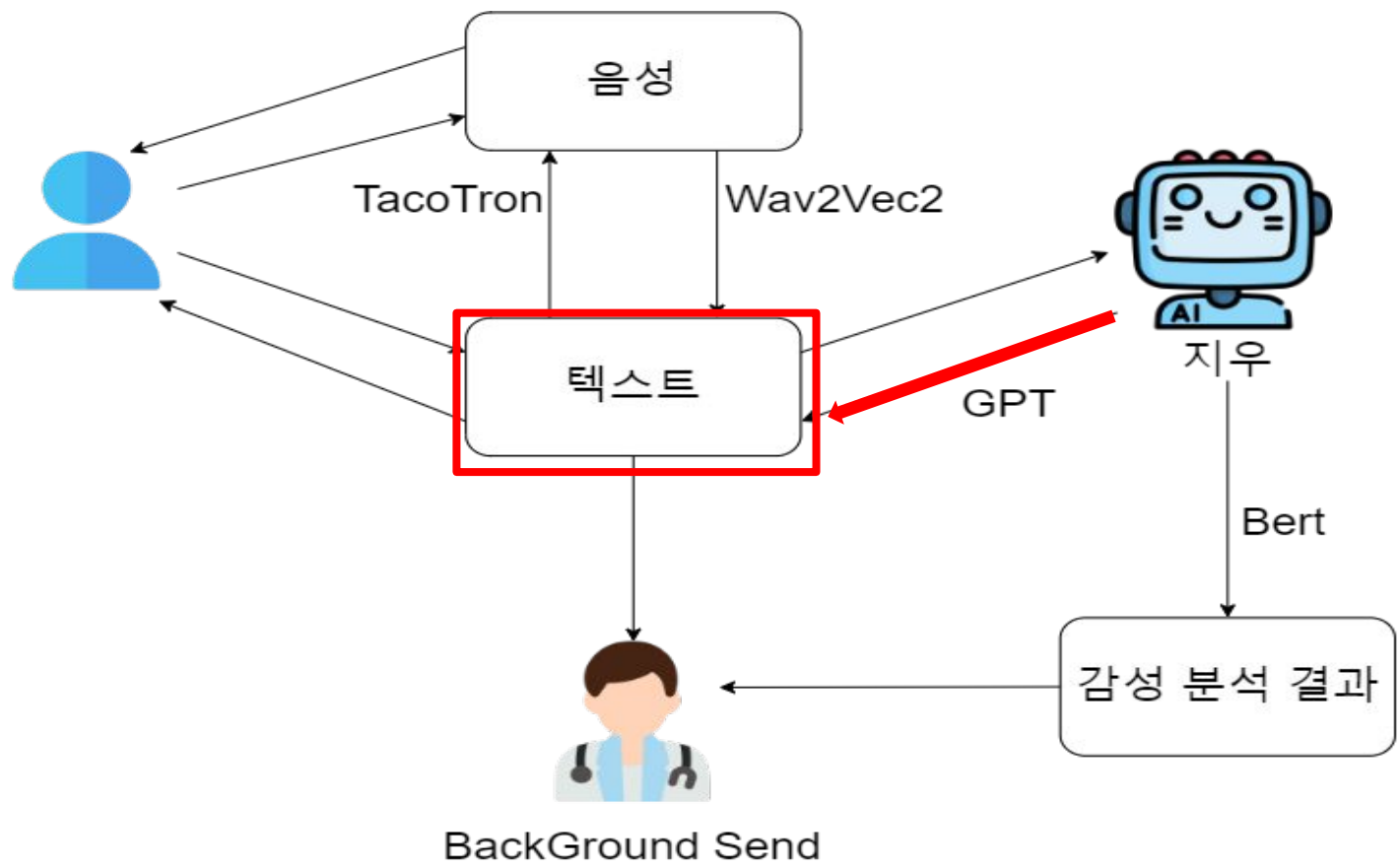
순서도



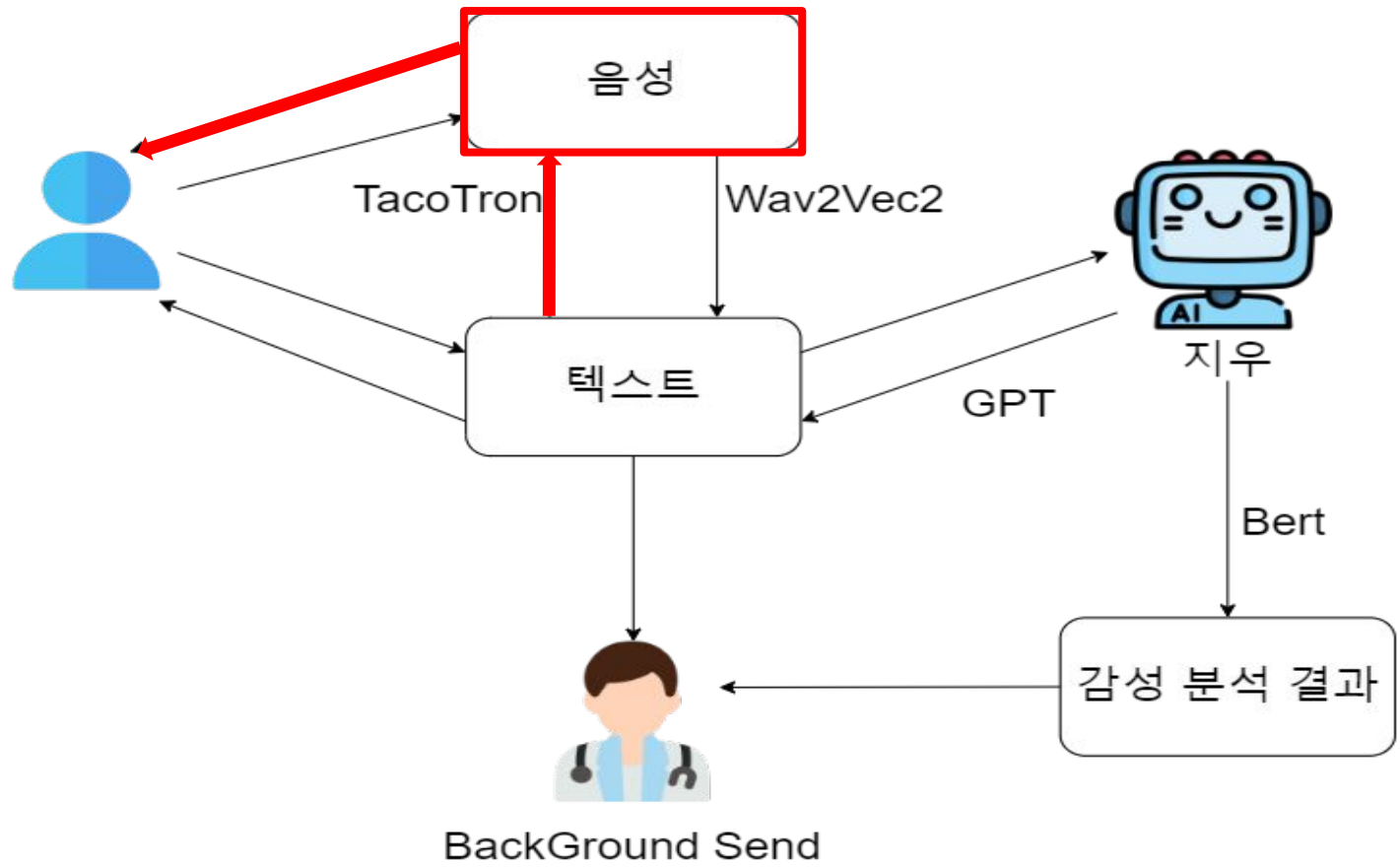
순서도



순서도



순서도



TEAM

- 기획했던 내용을 토대로 프로토타입 제작하기
- 딥러닝 모델을 커스텀 데이터를 이용하여 학습하는 과정을 이해하기

고종현

-음성 -> 텍스트 변환(TTS)

김성언

-텍스트 -> 음성 변환(STT)

-웹 구현

-코드 정리

박수빈

-감정 분석

-PPT 제작

이가희

-문장 생성

-PPT 제작

홍민택

-텍스트 -> 음성 변환(STT)

-웹 구현

-코드 정리

개발 일정

날짜	내용
3월 10일(금)	모델 생성
3월 13일(월)	1차 전체모델 학습
3월 14일(화)	2차(최종) 학습
3월 15일(수)	2차(최종) 학습
3월 16일(목)	2차(최종) 학습
3월 17일(금)	2차(최종) 학습, 모델 병합
3월 20일(월)	최종 제출
3월 20일(화)	최종 발표

데이터셋

번호	분류	데이터 이름	최신수정일	데이터 출처	형식
1	텍스트 데이터	한국어 감정 정보가 포함된 연속적 대화 데이터 셋	2022-05	AI-Hub	csv
2	텍스트 데이터	한국어 감정 정보가 포함된 단발성 대화 데이터 셋	2022-05	AI-Hub	csv
3	텍스트 데이터	감성 대화 말뭉치	2023-02	AI-Hub	csv
4	음성 데이터	한국어 음성	2023-02	AI-Hub	wav
5	음성 데이터	Korean Single Speaker Speech Dataset	2019-09	kaggle	wav,txt

STT(Speech To Text)



STT 기술


 kakao i cloud



Speech to Text

Kakao i Voice에서 제공하는 서비스로, 사람의 말소리를 분석하여 문자로 자동 변환해주는 인공지능 기술입니다.

CLOVA




#일상 대화 #쇼핑 대화 #정치 대화 #경제 대화 #취미 대화 #AI 비서 #동시통역 #감성형 대화 음성지능 서비스

한국어 음성

분야 한국어 유형 오디오, 텍스트

갱신년월 : 2023-02 구축년도 : 2018 조회수 : 6,513 다운로드 : 12,112 용량 : 72.04 GB

다운로드

관심데이터 등록  32

STT 모델_Speech to Text



KsponSpeech_000001.pcm



KsponSpeech_000001.txt

\x00\x01\x01

아/ 몬 소리야, 그건 또. b/



아 몬 소리야 그건
또

K o S p e e c h



Deep
Speech2

Conformer

K o S p e e c h

Deep
Speech2

```
C:\Users\USER\Desktop\STT\Korean_STT-main\koreanSTT>python ./bin/inference.py --m
uts\2023-03-09\03-08-31\model.pt" --audio_path "C:\Users\USER\Desktop\STT\KsponSp
['랭 랭랭랭랭랭랭랭 랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭
랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭 가']
랭 랭 랭 랭 랭 가
```

```
C:\Users\USER\Desktop\STT\Korean_STT-main\koreanSTT>python ./bin/inference.py --m
uts\2023-03-09\03-08-31\model.pt" --audio_path "C:\Users\USER\Desktop\STT\KsponSp
['랭랭랭랭랭랭 랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭 랭랭
랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭랭 가']
랭 랭 랭 랭 가
```


K o S p e e c h



Conformer

```
File "/content/drive/MyDrive/SK Shieldus Rookies 11/practice/koreanSTT/bin/kospeech/models/conformer/modules.py", line 94, in __init__  
    assert (kernel_size - 1) % 2 == 0, "kernel_size should be a odd number for 'SAME' padding"  
AssertionError: kernel_size should be a odd number for 'SAME' padding
```


Set the environment variable HYDRA_FULL_ERROR=1 for a complete stack trace.

STT 모델_Speech to Text

Rank	Model	Word Error Rate (WER)
1	Conformer + Wav2vec 2.0 + SpecAugment-based Noisy Student Training with Libri-Light	1.4
2	w2v-BERT XXL	1.4
3	Conv + Transformer + wav2vec2.0 + pseudo labeling	1.5
4	ContextNet + SpecAugment-based Noisy Student Training with Libri-Light	1.7
5	SpeechStew (1B)	1.7
6	Multistream CNN with Self-Attentive SRU	1.75
7	wav2vec 2.0 with Libri-Light	1.8

$$\text{Word Error Rate} = 100 \times \frac{\text{Insertions} + \text{Substitutions} + \text{Deletions}}{\text{Total Words in Correct Transcript}}$$

Human-labeled Transcript: How are you today John
Speech Recognition Result: How you a today Jones




STT 모델_Speech to Text

Rank	Model	Word Error Rate (WER)
1	Conformer + Wav2vec 2.0 + SpecAugment-based Noisy Student Training with Libri-Light	1.4
2	w2v-BERT XXL	1.4
3	Conv + Transformer + wav2vec2.0 + pseudo labeling	1.5
4	ContextNet + SpecAugment-based Noisy Student Training with Libri-Light	1.7
5	SpeechStew (1B)	1.7
6	Multistream CNN with Self-Attentive SRU	1.75
7	wav2vec 2.0 with Libri-Light	1.8

$$\text{Word Error Rate} = 100 \times \frac{\text{Insertions} + \text{Substitutions} + \text{Deletions}}{\text{Total Words in Correct Transcript}}$$

Human-labeled Transcript: How are you today John
Speech Recognition Result: How you a today Jones



STT 모델_Speech to Text

Rank	Model	Word Error Rate (WER)
1	Conformer + Wav2vec 2.0 + SpecAugment-based Noisy Student Training with Libri-Light	1.4
2	w2v-BERT XXL	1.4
3	Conv + Transformer + wav2vec2.0 + pseudo labeling	1.5
4	ContextNet + SpecAugment-based Noisy Student Training with Libri-Light	1.7
5	SpeechStew (1B)	1.7
6	Multistream CNN with Self-Attentive SRU	1.75
7	wav2vec 2.0 with Libri-Light	1.8
8	HuBERT with Libri-Light	1.8



Wav2vec 2.0

학습1

3천개 데이터

25 Epoch

Step	Training Loss	Validation Loss	Wer
500	18.752800	4.823676	1.000000
1000	4.755600	4.768725	1.000000
1500	4.695400	4.749733	1.000000
2000	4.637400	4.665298	1.000000
2500	4.481300	4.329843	0.996455
3000	3.908900	3.558360	0.998937

...

7000	0.783700	2.628076	0.894364
7500	0.622500	2.865274	0.894718
8000	0.549600	3.030916	0.903935
8500	0.479800	3.096499	0.899326

학습1

3천개 데이터

25 Epoch

Step	Training Loss	Validation Loss	Wer
500	18.752800	4.823676	1.000000
1000	4.755600	4.768725	1.000000
1500	4.695400	4.749733	1.000000
2000	4.637400	4.665298	1.000000
2500	4.481300	4.329843	0.996455
3000	3.908900	3.558360	0.998937

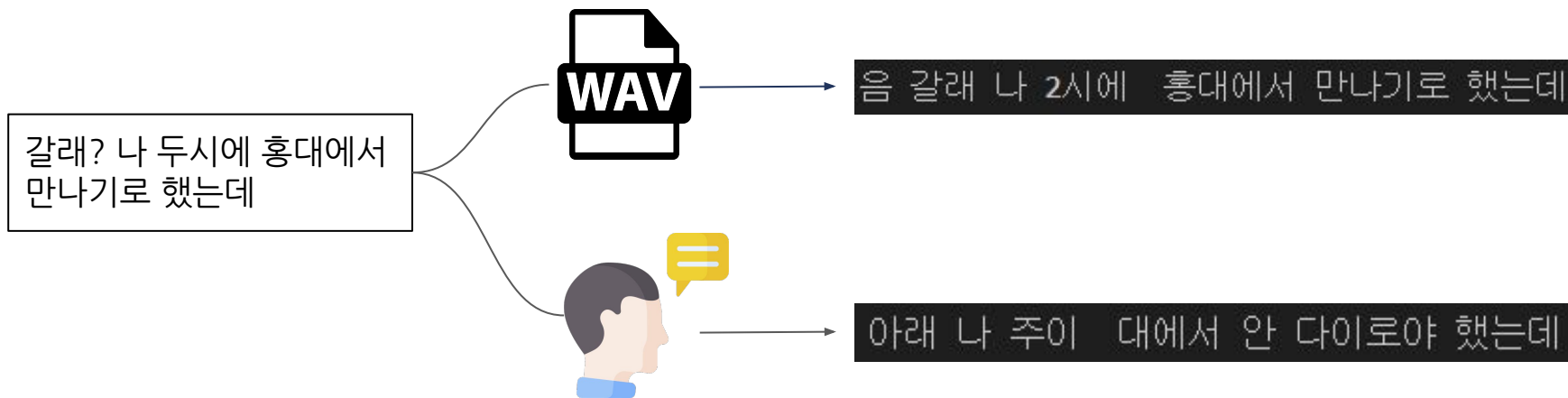
...

7000	0.783700	2.628076	0.894364
7500	0.622500	2.865274	0.894718
8000	0.549600	3.030916	0.903935
8500	0.479800	3.096499	0.899326

학습1

3천개 데이터










25 Epoch



학습2

60만개 데이터

1 Epoch

 KsponSpeech_19000-22000.zip
 KsponSpeech_22000-25000.zip
 KsponSpeech_25000-28000.zip
 KsponSpeech_28000-31000.zip
 KsponSpeech_31000-34000.zip
 KsponSpeech_34000-37000.zip
 KsponSpeech_37000-40000.zip
 KsponSpeech_40000-43000.zip
 KsponSpeech_43000-46000.zip

GPU 백엔드에 연결할 수 없음

현재 Colab의 사용량 제한으로 인해 GPU에 연결할 수 없습니다.

[자세히 알아보기](#)

닫기

GPU 없이 연결

학습2



전화번호 인증

Google에서는 보안을 위해 본인 확인을 진행합니다.
Google에서 6자리 인증 코드가 포함된 문자 메시지를 전송합니다



전화번호

010

❗ 이 전화번호는 이미 여러 번 사용되었습니다.

표준 요금이 부과됩니다.

답변 생성 모델

한국어

감성 대화 말뭉치

클라우드 소싱 수행으로 일반인 1,500명을 대상으로 하여 음성 15,700문장 및 코퍼스 27만 문장 구축 및 세대별 감성 대화 텍스트 구축을 통해 감성 대화 엔진을 개발하여 세대별 감성 대화 서비스 제공

#코퍼스

#감성대화

#감성 챗봇

#우울증 예방

유형

오디오, 텍스트

(145,956, 2)

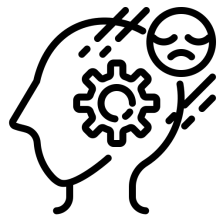
Q	A
일은 왜 해도 해도 끝이 없을까? 화가 난다.	많이 힘드시겠어요. 주위에 의논할 상대가 있나요?
이번 달에 또 급여가 깎였어! 물가는 오르는데 월급만	급여가 줄어 속상하시겠어요. 월급이 줄어든 것을 어떻게
회사에 신입이 들어왔는데 말투가 거슬러. 그런 애를	회사 동료 때문에 스트레스를 많이 받는 것 같아요. 문제
직장에서 막내라는 이유로 나에게만 온갖 심부름을 시	관련 없는 심부름을 모두 하게 되어서 노여우시군요. 어떤
얼마 전 입사한 신입사원이 나를 무시하는 것 같아서	무시하는 것 같은 태도에 화가 나셨군요. 상대방의 어떤

```
1 # 학습에 필요한 하이퍼 파라미터들을 선언 합니다.  
2 learning_rate = 3e-5  
3 criterion = torch.nn.CrossEntropyLoss(reduction="none")  
4 optimizer = torch.optim.Adam(model.parameters(), lr=learning_rate)  
5  
6 epoch = 10  
7 Sneg = -1e18
```



감성분석 모델

사전학습 모델			klue/bert-base
모델	LSTM	CNN	TFBertForSequence Classification
토크나이저	keras Tokenizer	keras Tokenizer	Bert Tokenizer
정확도	0.50	0.56	0.68



KLUE-benchmark/ KLUE



 Korean NLU Benchmark

 5

Contributors

 10

Issues

 484

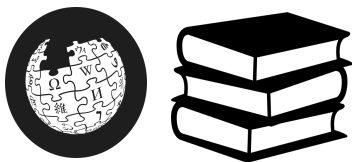
Stars

 53

Forks



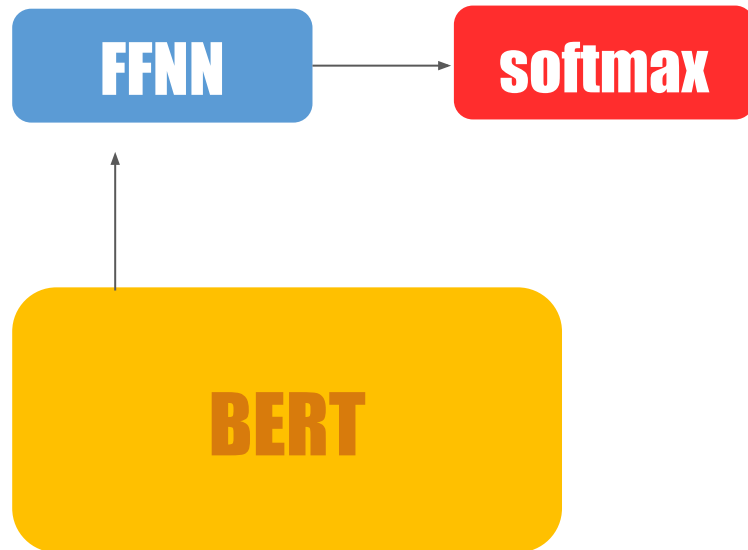
- KLUE-BERT는 벤치마크 데이터인 KLUE에서 베이스라인으로 사용되었던 모델
- Morpheme-based Subword Tokenizer를 사용,
- vocab size: 32,000
- 모델 크기: 111Mb



파인튜닝



다중 감성분류 모델



데이터

17229

아~어떡하면좋죠???큰일이네.....

슬픔

2081

아 불안하다 심장병 생길 것 같다

공포

30376

존잼꿀잼 요즘 나의 낙♡♡

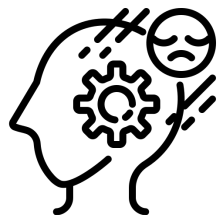
행복

한국어 감정 정보가 포함된 연속적
대화 데이터 셋

55628개

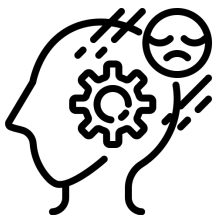
한국어 감정 정보가 포함된 단발성
대화 데이터 셋

38594개



총 94222개 데이터로 학습 진행

7가지의 감정 분류



공포
놀람
분노
슬픔
중립
행복
혐오

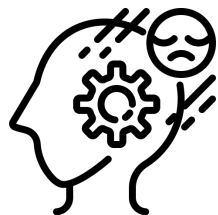


0
1
2
3
4
5
6

```
chatbot_data = chatbot_data.dropna() # 결측제거  
chatbot_data = chatbot_data.drop_duplicates() # 중복제거
```

```
# 오타 수정
```

```
chatbot_data2 = chatbot_data2.replace('ㄷ중립','중립')  
chatbot_data2 = chatbot_data2.replace('분ㄴ','분노')  
chatbot_data2 = chatbot_data2.replace('표','중립')  
chatbot_data2 = chatbot_data2.replace('ㄴ중립','중립')  
chatbot_data2 = chatbot_data2.replace('분','분노')  
chatbot_data2 = chatbot_data2.replace('줄','중립')  
chatbot_data2 = chatbot_data2.replace('중립','중립')
```



감성분석 모델

옵티마이저 Rectified Adam 하이퍼파라미터 조정

OPTIMIZER_NAME = 'RAdam'

LEARNING_RATE = 5e-5

TOTAL_STEPS = 10000

MIN_LR = 1e-5

WARMUP_PROPORTION = 0.1

EPSILON = 1e-8

CLIPNORM = 1.0

```
optimizer = tf.keras.optimizers.RectifiedAdam(learning_rate = LEARNING_RATE,  
                                              total_steps = TOTAL_STEPS,  
                                              warmup_proportion = WARMUP_PROPORTION,  
                                              min_lr = MIN_LR,  
                                              epsilon = EPSILON,  
                                              clipnorm = CLIPNORM)
```

감성분류 모델 컴파일

```
sentiment_model.compile(optimizer = optimizer,  
                      loss = tf.keras.losses.SparseCategoricalCrossentropy(),  
                      metrics = ['accuracy'])
```

감성분석 모델

```
# 질문 무한반복하기! 0 입력시 종료
end = 1
while end == 1 :
    new_sentence = input("하고싶은 말을 입력해주세요 : ")
    if new_sentence == 0 :
        break
    sentiment_predict(new_sentence)
    print("₩n")
```

하고싶은 말을 입력해주세요 : 우울했
97.17% 확률로 슬픔이 느껴집니다

감성분석 모델

	precision	recall	f1-score	support
0	0.580	0.508	0.542	1111.000
1	0.546	0.352	0.428	1973.000
2	0.459	0.624	0.529	1847.000
3	0.597	0.430	0.500	1434.000
4	0.768	0.904	0.830	9131.000
5	0.702	0.714	0.708	1403.000
6	0.481	0.023	0.044	1125.000
accuracy	0.683	0.683	0.683	0.683
macro avg	0.590	0.508	0.512	18024.000
weighted avg	0.664	0.683	0.653	18024.000

TTS 란?

기존의 TTS모델들

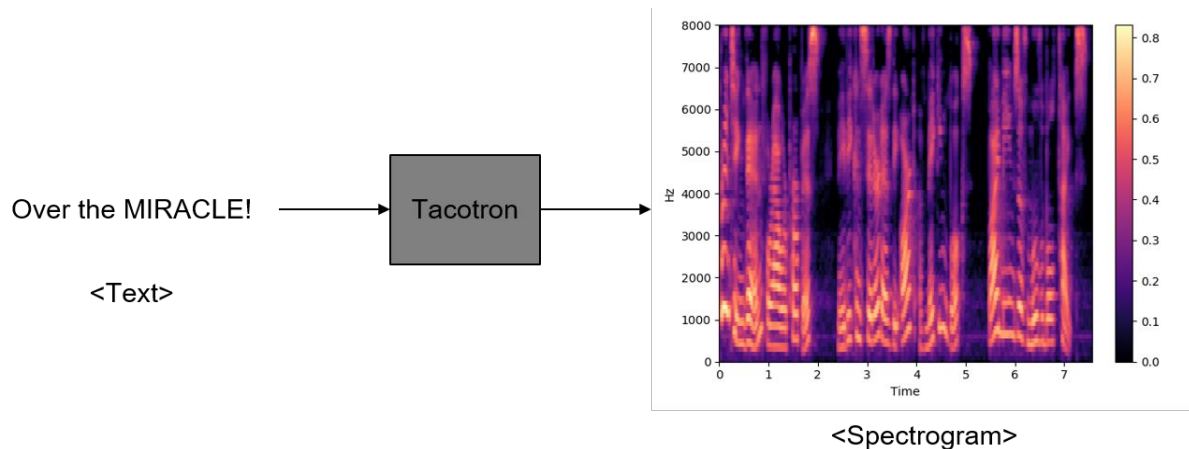
WaveNet

- 매우 강력한 음성 생성 모델
- 속도가 매우 느림
- TTS로 바로 활용은 불가

DeepVoice

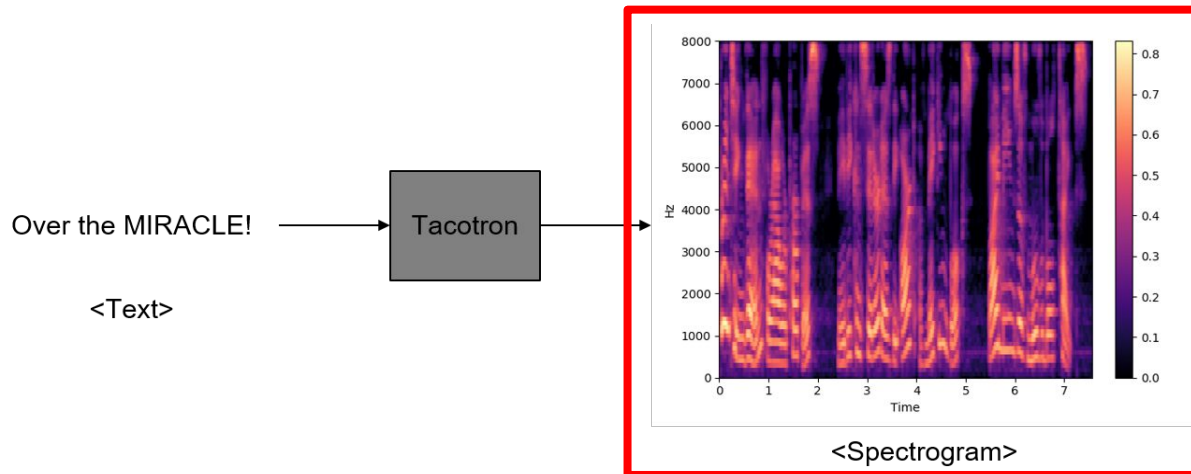
- TTS파이프라인을 뉴럴넷 으로 대체
- End-to-End 모델이 아님

Tacotron 이란?



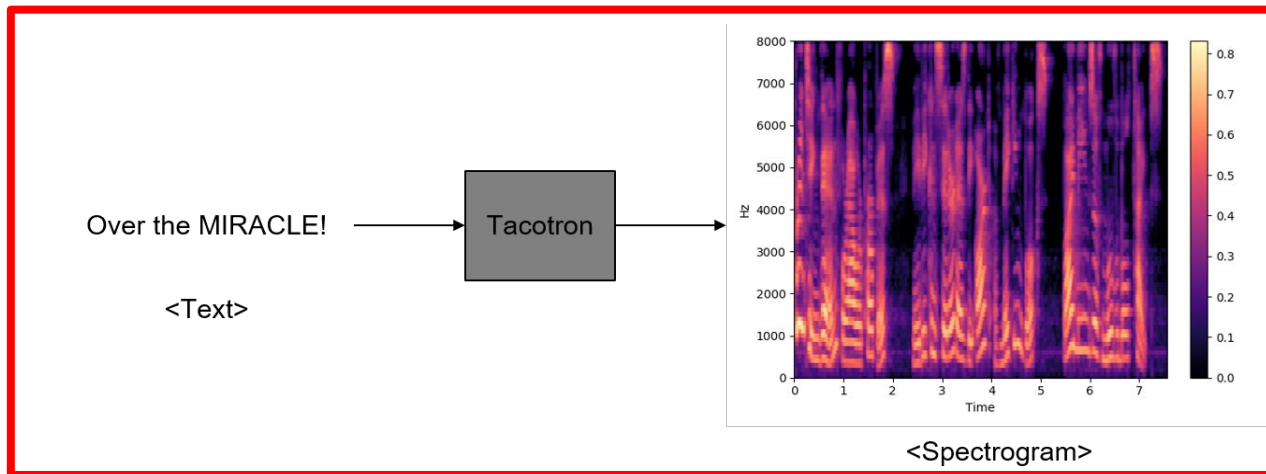
- 2017년 구글에서 발표한 TTS모델
- 텍스트를 입력받아 **Raw Spectrogram** 바로 생성
- <text,audio>페어로 **End-to-End**학습 가능

Tacotron 이란?



- 2017년 구글에서 발표한 TTS모델
- 텍스트를 입력받아 **Raw Spectrogram** 바로 생성
- <text,audio>페어로 **End-to-End**학습 가능

Tacotron 이란?



- 2017년 구글에서 발표한 TTS모델
- 텍스트를 입력받아 **Raw Spectrogram** 바로 생성
- <text,audio>페어로 **End-to-End**학습 가능

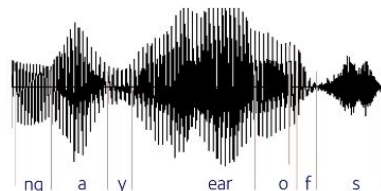
TTS 모델_Text to Speech - Tacotron을 이용한 TTS 제작

🔍 Search



Korean Single Speaker Speech Dataset

KSS Dataset: Korean Single Speaker Speech Dataset



Data Card

Code (4)

Discussion (1)

About Dataset

[Updated on September 28, 2019] KSS Dataset: Korean Single speaker Speech Dataset

KSS Dataset is designed for the Korean text-to-speech task. It consists of audio files recorded by a professional female voice actress and their aligned text extracted from my books. As a copyright holder, by courtesy of the publishers, I release this dataset to the public. To my best knowledge, this is the first publicly available speech dataset for Korean.

Usability ⓘ

7.50

License

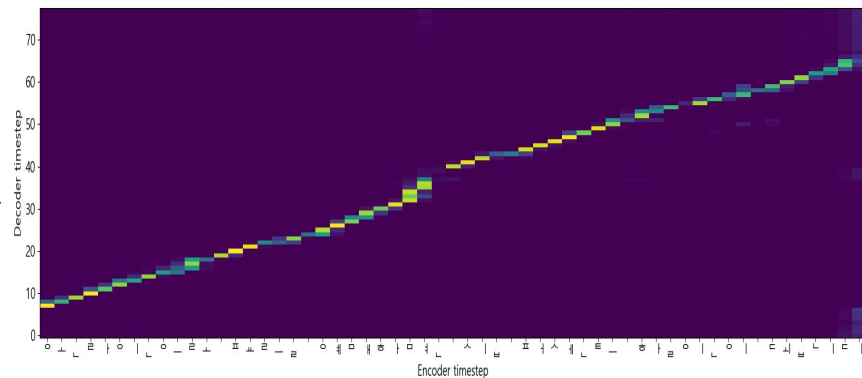
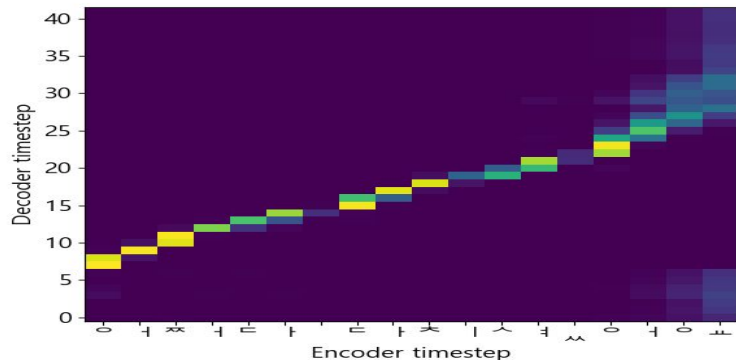
CC BY-NC-SA 4.0

Expected update frequency

Not specified

TTS 모델_Text to Speech - Tacotron을 이용한 TTS 제작

Tacotron 모델로 음성 데이터 학습

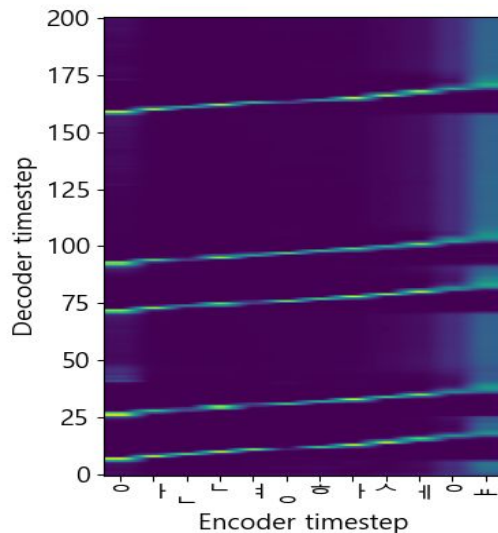


- Tacotron 모델에 음성 데이터를 학습시키면 가중치가 쌓인다
- 위 사진은 matplotlib을 이용하여 가중치를 저장할때 정렬그래프이다

Tacotron 모델로 음성 데이터학습

```
step: 26001, loss: 0.04513, 6.655 sec/step  
step: 26002, loss: 0.04466, 1.847 sec/step  
step: 26003, loss: 0.04382, 1.032 sec/step  
step: 26004, loss: 0.04377, 4.751 sec/step  
step: 26005, loss: 0.04246, 2.003 sec/step  
step: 26006, loss: 0.04127, 0.936 sec/step
```

학습된 가중치로 새로운 텍스트를 음성파일로 만들기



- 새로 주어진 텍스트를 멜스펙트로그램으로 변환하여 np형태로 저장
- np형태로 저장된 멜스펙트로그램을 CBHG모듈을 사용하여 음성으로 합성
- 위사진은 결과물인 wav파일의 정렬 그래프이다

AI 상담 챗봇 지우

2023.3.16 17:22

undefined

기분이 좋지 않으면 간단한 디저트를 만들어 보는

my

2023.3.16 17:22

나 응원해줘

2023.3.16 17:22

undefined

언제든지 응원해 드릴게요! 나는 항상 여러분을 응

녹음

Type Message ...



레포트 생성

홍민택님의 상담 내용입니다.

[2023-03-16 17:20:45.181565] 홍민택 : 안녕

[2023-03-16 17:20:45.361957] 감정 : 33.77% 확률로 행복이 느껴집니다

[2023-03-16 17:20:47.748676] 지우 : 안녕! 반가워요.

[2023-03-16 17:20:56.418770] 홍민택 : 저녁 메뉴는 어떻게 좋을까?

[2023-03-16 17:20:56.560787] 감정 : 79.22% 확률로 중립이 느껴집니다

[2023-03-16 17:21:00.028507] 지우 : 저녁 메뉴는 취향에 따라 다르겠지만, 간단한 국

[2023-03-16 17:21:04.198699] 홍민택 : 나 힘들어

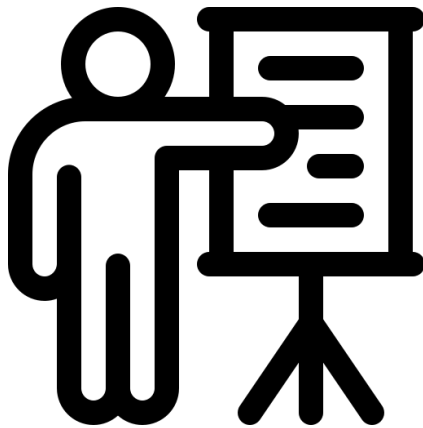
[2023-03-16 17:21:04.282070] 감정 : 79.60% 확률로 슬픔이 느껴집니다

[2023-03-16 17:21:07.117756] 지우 : 기분이 좋지 않으면 간단한 디저트를 만들어 보는

[2023-03-16 17:21:10.469587] 홍민택 : 나 좀 응원해줘

[2023-03-16 17:21:10.612165] 감정 : 60.93% 확률로 중립이 느껴집니다

[2023-03-16 17:21:13.509107] 지우 : 언제든지 응원해 드릴게요! 나는 항상 여러분을 응
상담이 종료되었습니다.



URL

- 환자의 트리거를 인식하여 먼저 질문 제시

- 환자의 트리거를 인식하여 먼저 질문 제시
- 답변을 목소리로 출력할 때 더 좋은 퀄리티로 발전시키기

- 환자의 트리거를 인식하여 먼저 질문 제시
- 답변을 목소리로 출력할 때 더 좋은 퀄리티로 발전시키기
- 성능 향상

- 환자의 트리거를 인식하여 먼저 질문 제시
- 답변을 목소리로 출력할 때 더 좋은 퀄리티로 발전시키기
- 성능 향상
- 감성분석시 데이터 편중현상-> 좋은 데이터셋 찾기

개인 목표

고종현

내가 임의로 녹음한 음성으로 TTS 완성



김성언

딥러닝 모델의 전이학습 과정 이해



박수빈

여러가지 모델을 사용하여 감성분석 정확도 향상하기



이가희

한 파트를 맡아 모델 구현하기



홍민택

음성 인식 모델(STT)에 대한 전반적인 구조 이해



마무리

	모델	batchsize	epoch
STT	wav2vec 2.0	2	25
TTS	Tacotron	32	20
감정분석	Bert	20	8
답변생성	GPT2	32	10

QnA