1.

안녕하세요, 2조 팀 포스 입니다.

2.

인트로에서 팀 소개와 프로젝트 개요를 말씀드리고,

프로세스에서 프로젝트의 큰 절차를 말씀드리고,

백그라운드에서 프로젝트 핵심 연구 내용을 말씀드리고,

익스페리먼츠에서 데모를 보여드리고,

자체평가 결과를 말씀드리겠습니다.

3.

프로젝트 구성원은 IT 기업의 핵심 직책을 맡고있는 전문가들로 구성되어있습니다.

팀 이름 포스는 팀원 이름 이니셜 맨뒤에 S가 모두 포함된다.

~를 위하여 For S

반복문 For S

숫자 4를 의미하는 Four S

빡세게 공부하자는 의미로 Force

입니다.

4.

저희 프로젝트의 슬로건은 다음과 같습니다.

“누구나 AI를 통해 무엇이든 할 수 있다”

어떤 의미로 해석하셔도 좋습니다.

5.

이번 프로젝트의 핵심은 ASR 입니다.

ASR이란 자동 음성 인식 기능을 뜻하는데, 이는 크게 음성을 텍스트로 전환하거나 텍스트를 음성으로 전환하는 기술을 의미합니다.

6.

프로젝트 진행 과정입니다.

팀 구성원 각자 ASR 개념을 이해하고 ASR로 우리 평범한 일상에서 어떤 편의를 제공하고 어려운 문제를 쉽게 해결할 수 있을까? 생각하며 아이디어를 떠올렸습니다.

그리고 그 아이디어를 어떻게 구현할 지 다양한 모델을 찾고 연구하고 모델 비교 분석해 한 가지를 선정했습니다.

선정한 모델을 활용해 서비스 구현 코드를 작성했고.

구현된 서비스를 사용해봤습니다.

7,

그렇게 만들어진 서비스 예시를 간단히 소개하겠습니다.

글을 쓰기 어려워하는 대상들도 음성을 통해 자동으로 글 작성을 할 수 있도록하는 서비스를 만들었습니다.

사회 취약자들의 가장 큰 고민은 취업이라는 점을 파악했습니다.

우리는 모두 이력서를 쓰기 어려워합니다. 노인이나 장애가 있는 사람은 이력서 쓰기가 더욱 어렵구요.

고용센터에서는 이러한 분들을 상담하고 취업까지 도움을 줄 수 있지만 인력의 한계가 있다는 문제도 캐치했습니다.

ASR을 활용해 이런 어려움을 해결할 수 있다 판단하고 서비스를 만들었습니다.

서비스 개발 과정에서 발음이나 목소리 등 인식률이 떨어지는 현상을 개선하기 위해 공공데이터센터의 한국어 발화 데이터셋으로 파인튜닝을 했시도했고 이후 추가적인 기술들을 적용하여 음성 데이터를 사람이 보기 편하게 자동으로 문서화 하도록했습니다.

이렇게 문서화 된 자료를 취업상담사가 체크하여 취업에 도움을 줄 수 있도록했습니다.

지금 설명드린것은 서비스로 해결한 일부 예시일 뿐 더욱 다양한 방면에서 활용 될 수 있습니다.

가령 또다른 예시로는 대형 취업사이트에서 누구나 쓸 수도 있고, 다양한 인터뷰 내용 정리에도 사용 할 수 있고, 면접에서 면접자의 음성을 기록하고 추후 재검토에도 활용 될 수 있습니다.

8.

아이디어 기획부터 서비스 구현까지의 WPT 자료입니다.

9.

WPT 요약한 내용입니다.

10.

서비스를 만들기위해 가장 먼저 STT 모델을 탐색했습니다.

세상에는 이미 잘 만들어진 STT 모델들이 다양하게 존재했습니다.

OpenAI를 비롯해 Google, Naver, MicroSoft, NVIDIA, META, Deepspeech 등 이 있었죠.

그 중 유료는 제외하고 무료 STT만을 추려 테스트를 해 봤습니다.

이미 성능도 좋고 속도도 빠른 모델들이 있었지만, 그 중 파인튜닝까지 가능하고 도메인에 맞게 성능을 끌어올릴 수 있고 이 과정에서 우리의 기술적 능력이 올릴 수 있는 모델인 Whisper를 선정했습니다.

Whisper는 OpenAI의 오픈소스로서 기본적으로 트랜스포머 구조와 크로스어텐션이 사용된 모델로서 68만 시간 분량의 노이즈가 있는 음성 데이터셋을 학습시킨 모델입니다.

STT 비교 분석 내용은 화면의 링크를 클릭해서 확인가능하나 시간관계상 생략하고 넘어가도록 하겠습니다.

11.

앞서 설명드린 그냥 Whisper의 성능은 좋지만 속도가 문제가 있었습니다, 떄문에 나온게 퀀타인제이션과 로라를 적용하여 경량화 시킨 패스트위스퍼를 찾게 되었습니다. 이로서 성능은 지키고 속도를 개선시켰습니다.

12.

또한 기존 모델에 도메인에 맞는 데이터셋을 추가로 파인튜닝 해서 음성 인식률을 끌어올리는 시도를 해보았습니다.

이 때 새로 알게된 부분은 신규 데이터를 학습 시키면 인식률 트레이드 오프가 발생하는 것을 알게 되었습니다.

예를 들면 기존 우리가 생각했던 방향은 표준어도 인식하고 사투리를 추가 학습 시키면 표준어와 사투리 둘다 잘 텍스트로 나올 줄 알았으나, 기존 잘 알아듣던 표준어 인식률은 떨어지고 사투리 인식률이 높아지는 것이었습니다.

13.

앞서 말씀드린 인식률이 떨어진 것은 어떻게 확인했는지 설명해드리겠습니다.

이번 프로젝트에서 한국어 인식률을 검증하기 위한 평가 지표로 CER 방식을 사용했습니다.

WER과 CER 방식이 있는데 영어와는 다르게 한국어는 CER 방식으로 검증을 했습니다.

WER과 CER은 거의 유사한데 검증하는 토큰이 WER은 단어로 구분하고 CER은 문자가 토큰이 된다는 차이가 있습니다.

한국어 음성인식에서는 WER 보다 CER이 더 중요한 척도로 여겨집니다.

한국어는 교착어(첨가어)로 조사를 사용하고 다른 언어와 비교하여 형태소의 구조가 복잡하며, 단어와 단어 사이의 경계가 모호합니다.

이러한 언어 구조의 특성으로 인해 단어 수준에서의 평가가 어렵습니다.

따라서, 문자 단위의 오류를 측정하는 CER이 한국어 음성인식에서 더 정확한 평가 방법으로 간주했습니다.

CER도 두 가지가 있는데 일반적인 CER은 실제 문자열 길이로 나누어 계산하지만, 정규화된 CER의 경우 S,D,I의 합에 추가로 일치하는 음절 수까지 합한 값을 나누어 계산합니다.

때문에 정규화된 CER은 문자열 길이에 따른 편향을 줄일 수 있어 더 정확한 성능 평가가 가능합니다.

두 가지 방식 모두 값이 낮을수록 정확하다는 것만 생각하고 넘어가도 좋습니다.

14.

실제 위스퍼 68만시간 학습 중 한국어 학습은 7993시간 만 학습시킨 것을 확인했습니다.

파인튜닝 테스트를 위해 우리 팀은 노인 자유 발화 음성 데이터셋을 활용했습니다.

15.

실제 위스퍼 인식률에 대한 연구 내용을 말씀드리겠습니다.

우선 위스퍼는 base small large 등 다양한 기본 모델들이 있습니다.

base가 기본 모델로서 코퍼스가 가장 적고, large가 코퍼스가 큽니다.

각각의 모델에 cer 평가를 진행해본 결과 large가 인식률이 높은 것을 확인했습니다.

16.

이어서 위스퍼의 기본모델에 파인튜닝으로 커먼보이스15버전 데이터셋을 파인튜닝 시키면 CER이 더 높아지는 이상현상을 확인했습니다.

17.

또 이어서 노인발화데이터로 파인튜닝을 했더니 CER이 극단적으로 낮아 지는 현상도 봤습니다..

18.

이러한 연구를 통해 파인튜닝이 마냥 좋지만은 않구나를 알았습니다.  
전이학습의 한계점을 생각해볼 수 있습니다.

GPT 같은 LLM을 업그레이드 할 때도 대용량 데이터를 한번에 다시 학습시는 이유가 이런 전이학습 한계점과 동일합니다.

그리고 음성 데이터 코퍼스의 평균 음성길이는 6.5초인데 반해 위스퍼 파인튜닝은 모든 오디오 입력을 30초로 고정해서 학습을 시켜야하기 때문에 3GB 뿐 안되는 데이터가 16GB 까지 증가하는 불편도 확인할 수 있는 연구 결론을 얻었습니다..

여기까지가 서비스에 사용한 STT 모델 위스퍼 연구 내용이었습니다.

19.

다음으론 STT를 통해 얻은 데이터를 어떻게 요약 정리를 했냐면 랭체인과 GPT를 통해 정리된 OutPut 데이터를 가져오도록 했습니다.

20.

앞서 얻은 요약 데이터로 데이터프레임을 만들어 투엑셀 따위의 함수로 실제 문서로 정리된 데이터를 받아 볼 수 있도록 했습니다.

21

데모 영상 입니다.

22..

여기까지 우리 팀 프로젝트에서 사용된 기술들입니다.

23.

자체평가로는 전반적으로 99점이라는 후한 점수를 주고싶습니다.

기획한 내용에도 올바르고 다양한 실무에 적용도 가능합니다.

기능 구현도 또한 우수했습니다. 더욱 고도화하면 많은 문제를 더 잘 해결할 것으로 예상됩니다.