실질적인 다문화가정의 교육문제 해결을 위한 부모와 자녀가 함께하는 발음교육 서비스

이채은, 이상무, 정봉기

1. 프로젝트 개요

부모와 자녀가 함께 학습하는 언어교육 서비스 'MINGLE'을 제안합니다. 저희 서비스(발음피드백, QA챗봇)를 통해 조화로운 다문화가정을 도모하고 다문화사회로의 확장과 문화발전을 추구합니다. 다문화 가정의 어머니들이 경험하는 언어 장벽, 정서적 고통, 경제적 어려움은 자녀에게 부정적인 영향을 미칠 수 있습니다. 다문화 가정 수의 증가는 '자녀들의 낮은 학업 성취도'라는 사회적 문제로 이어졌습니다. 다문화 가정의 자녀들은 적절한 언어 교육을 받았으나, 가정에서 한국어 노출 빈도가 적습니다. 한국어를 사용하는 사람과의 상호작용이 적으므로, 불충분한 언어 습득으로 이어질 수 있습니다. 이러한 경향은 한국의 출산율 저하 속에서 더욱 심화되었습니다. 따라서, 우리는 외국인 부모와 그들의 자녀를 위한 공동 한국어 학습을 촉진하는 해결책을 제안합니다. 이것은 한국만의 문제가 아니라 다문화 사회로 확장하는 모든 국가의 문제입니다.

UN의 지속 가능한 개발 목표

1) 양질의 교육 - 부모님과 자녀가 언어 발음에 대한 맞춤형 피드백을 받을 수 있습니다. 접근성이 용이하여 양질의 서비스를 빠르고 쉽게 제공할 수 있습니다. 이를 통해 포용적이고 공평한 양질의 교육을 보장하고 모두에게 평생학습의 기회를 제공할 수 있습니다.

2) 불평등 완화 - 다문화가정이라도 언어적 장벽을 극복하여 차별받지 않고, 국가의 모든 국민들에게 균등한 기회를 제공할 수 있도록 언어 및 발음 교육을 실시하고 있습니다. 특히 학업성취도 및 사회적 역할 측면에서 언어교육이 미흡하여 발생할 수 있는 불평등을 해소하고 있습니다.

3) 평화, 정의 및 강력한 제도 - 다문화로 나아가는 국가들의 사회통합 모델에 맞는 서비스를 제공합니다. 다문화 공동체를 굳건히 하고, 다양성을 존중하는 마음가짐을 기르며, 지속가능한 발전을 위한 비차별적인 법과 정책을 추진하고 강화합니다.

2. 팀 구성 및 역할

12213404 디자인테크놀로지학과 이채은(팀장) PM, Design

12201770 컴퓨터공학과 이상무(팀원) FE, BE

12191870 수학과 정봉기(팀원) ML, AI

3. 프로젝트 수행 절차 및 방법

3-1. 개발 일정

노션 페이지를 활용하여 프로젝트 일정을 계획하고 공유했습니다.

3-2. 플로우 차트

4. 프로젝트 수행 결과

최종 결과는 유튜브 링크에서 확인 가능합니다.

https://www.youtube.com/watch?v=AmmFMJ_Y_4Q

4-1. 핵심 기능

1) 음성인식 서비스 구축 및 발음 피드백 - 사용자의 어휘력 향상을 위한 퀴즈를 제공(모국어+한국어 음성)

2) 전래동화 읽기 - 명확한 한글 발음 습득을 위해 전래동화 읽기 기능을 제공

3) Gemini 챗봇을 이용한 image2sound or sound2image 형식의 질문 및 응답구조 – 개인 맞춤형 서비스

이전에는 정부의 부족한 교육 프로그램과 다문화 가정의 문제에 대한 인식 부족으로 인해 다문화 가정의 학업 성적 저하와 가정 부조화가 발생하였습니다. 이제 저희 서비스를 이용하시면 이 문제가 해결됩니다. 1) 다양한 단어, 상황, 문장에 대한 음성 인식을 통해 발음 피드백을 제공할 수 있습니다. 2) 스토리북을 읽으며 연습할 수 있습니다. 이처럼 실제로 필요한 상황에 적용하여 언어교육을 받을 수 있습니다. 언어교육에서 소외된 가정에 언어교육을 제공함으로써 부모와 자녀 모두 교육의 혜택을 받을 수 있습니다. 나아가 글로벌 사회를 향한 균형 있는 발전을 도모할 수 있습니다.

4-2. 사용 프로그램

1) ALL - GCP

2) FE - Flutter, dart, Android studio

3) BE - Java Spring, MongoDB, DataGrip, IntelliJ

4) AI - Python, Fastapi, Pytorch, 'Gemini Vision Pro 1.0' , Audio recognition model (Deepspeech2), Speech Feedback (IPA Distance API)

*KoSpeech*의 오디오 인식 모델 아키텍처는 다음과 같습니다. (a) Deep Speech 2, (b) Listen Attention and Spell, (c) Transformer, (d) Joint CTC-Attention LAS의 네 가지 모델이 구현되어 있습니다. 저희는 (a)를 사용했습니다. Deep Speech 2는 Connectionist Temporal Classification (CTC) 손실이 있는 ASR 작업에서 더 빠르고 정확한 성능을 보여주었습니다. 이 모델은 이전의 end-to-end 모델에 비해 상당히 우수한 성능입니다. 한국어뿐 아니라 다국어로 서비스를 제공해야 했기 때문에 IPA 변환기를 이용해 딥스피치 음성 데이터를 IPA로 변환하고 IPA 거리 API를 통해 비교했습니다.

5) Design - Figma ,Adobe(After effects, Illustrator, Photoshop):

4-3. 수행 과정

우리는 발음 피드백과 언어 교육 솔루션에 적합한 아키텍처를 만들었습니다.

1) Design - UI는 Figma라는 협업 툴을 통해 디자인했으며, 좌표, RGB 값, 이미지 파일 등을 프론트 엔드로 전달합니다.

2) FE - 플로우 차트에 따라 안드로이드 OS 내에서 사용자의 반응에 따라 작동하도록 설계되었습니다. 주로 클라이언트로부터 터치와 스크롤 입력을 받고, 특정 단계에서 음성 데이터를 입력으로 사용합니다.

3) BE - 로그인을 담당하고 각 사용자의 정보, 음성녹음파일, 기타 프로세스를 DB에 저장하며, 해당 정보를 바탕으로 문항의 난이도를 조절하여 사용자의 학습에 도움을 줍니다.

4) AI - 다문화 음성인식 모델을 구축하고 음성인식 백본 모델로 DeepSpeech2 모델을, 사전학습 모델로 Kospeech를 활용하고 유아와 외국인의 자유로운 대화를 추가 학습했습니다. Prosody는 IPA Distance API를 이용하여 피드백을 제공합니다. 음성 데이터를 IPA로 변환하여 정답과 비교하여 발음 문제를 파악합니다. 이후 구강 근육 사진과 함께 피드백을 제공합니다. Prosody 분석에서는 말하기 속도와 음높이에 초점을 둡니다. 말하기 속도는 IPA 글자 사이의 길이를 분석하여 분류합니다. Pitch는 단어의 음높이를 파악하여 단어를 분류합니다. 이를 바탕으로 클러스터링 알고리즘을 이용하여 분류하여 학습 데이터로 재사용하고, Siluette Value, Elbow Method, Gap Statistics를 이용하여 검증을 수행합니다.

4-4. 데이터셋

Kospeech는 AI Hub에서 제공하는 1000h 한국어 음성 데이터인 KsponSpeech의 데이터 셋을 사용합니다. 맞춤형 음성 인식 및 발음 지원을 위한 맞춤형 데이터도 교육했습니다. 여기에는 외국인이 말하는 한국어와 유아 음성 데이터가 포함됩니다.

4-5. 평가

다문화가정 2곳, 다문화센터 2곳, 다문화연구센터 1곳을 방문했습니다. 총 5곳이 서비스를 이용하고 피드백을 받았습니다. 40명을 대상으로 만족도 조사를 실시한 결과 4.8/5점의 긍정적인 결과를 받았습니다. 만족도 조사는 구글 폼을 사용했습니다.

4-6. 피드백

1) 다문화가정 어머니들은 문화 부적응으로 자녀 양육에 어려움을 겪습니다. 새로운 사회에 익숙하지 않은 외국인

부모님과의 문화적, 언어적 갈등을 해결해 주시기 바랍니다.

2) 정부는 프로그램 예산이 정해져 있어서 다문화 교육을 해야 합니다. 하지만, 현실적으로 제대로 상용화되지 못하고 있습니다. 교육을 받는 사람이 적은데, 프로그램 수는 많습니다. 따라서, 다문화가정의 구성원들이 언어 교육의 필요성을 직접 느껴야 합니다.

3) 유아를 위한 서비스에 어울리는 UI가 좋지만, 혀 발음 구조에 대한 피드백이나 용어 설명을 제공할 때 유아에게 적합하지 않아서 보완을 요청하였습니다.

4-7. 솔루션

1) 아이의 학업 성취는 부모에 의해 영향을 받습니다. Mingle은 아이들과 부모를 위한 서비스 방향을 결정했습니다.

2) 다문화센터를 방문하여 언어교육의 필요성에 대한 피드백을 받았습니다. 정부 프로그램은 접근이 불가능하기 때문에 개인 맞춤형 서비스가 필요하다고 생각하여 웹 서비스보다 접근이 용이한 앱 서비스로 구축하였습니다.

3) 실제 혀 구조는 단순화된 디자인으로 바뀌었습니다. 가장 어려운 점은 정확한 데이터와 일러스트 사이의 협의점을 찾는 것이었습니다. 그래서 먼저 일러스트로 보여드리고 나중에 정확한 혀 구조를 보여주는 버튼, 영상 데이터를 업데이트하겠습니다.

4-8. 확장성

프로젝트의 다음 단계는 유아와 가족을 위한 가이드라인이 확립된 생성형 AI 모델을 구축하는 것입니다. 특히 발음 피드백을 제공할 정해진 단어나 사진이 아니라 어떤 상황이든 이해하고 발음과 언어 교육을 제공할 수 있는 기반 모델이 필요합니다. 그리고 언어적인 확장입니다. 현재는 한국어와 베트남어만 가능합니다. 하지만 영어, 일본어, 중국어, 태국어, 필리핀어 등 다양한 데이터가 있기 때문에 이러한 서비스로의 확장은 어렵지 않습니다. 서비스가 확대되면 GCP를 통해 각 기술 구성 요소의 컨테이너를 관리하고 저희 아키텍처의 모델을 제공하여 빠르고 쉽게 기술 지원을 할 것입니다. 또한, Gemini 같은 생성형 AI의 사용에서는 연령 제한 지침에 따라 프롬프트를 구성할 수 있으며, GEMMA(Google Pruning Open Source LLM)와 같은 추가 학습에 맞게 파라미터를 구성하여 연령 맞춤형 서비스를 제공할 수 있습니다.

5. 자체 평가 의견

이전 피드백 등 저희가 구축해야 할 부분에서 많은 위기 상황이 있었고, GCP로 아키텍처를 구축하는 것도 처음이라 각 서버의 컨테이너 등을 연결하는 문제가 많았습니다. 가장 큰 문제는 서비스의 주요 기능 중 하나에서 직면했습니다. "Gemini vision pro 1.0"을 통해 이미지를 촬영하면 발음 방법과 어떤 글자인지 알려주는 대화형 챗봇을 구축하고 싶었습니다. 구체적으로 이미지나 상황을 촬영하여 텍스트로 캡션하고 TTS를 통해 사용자에게 텍스트를 알려주고 STT를 통해 챗봇에게 다시 전달하는 양방향 QA를 목표로 했습니다. 실제로 대화 상황을 촬영하고 프롬프트에 따라 이미지를 삽입하면 캡처할 수 있습니다. 캡션된 텍스트에 대해 질적 평가를 실시한 결과 이전 캡션 모델보다 성능이 놀라운 것으로 나타났습니다. 다만, Gemini API를 사용하는 것 자체는 성능이 우수하고 문제가 없지만, 이번 Foundation Generative AI 모델은 연령 제한 등 어린이들이 사용하기 상당히 어렵습니다. 따라서 추후 가이드라인을 제정하여 어린이와 가족을 위한 맞춤형 서비스를 제공할 예정입니다. 명확한 답변만 출력될 수 있도록 LLM의 온도(파라미터)를 미세 조정하거나 낮춰 추가 학습에 활용할 계획입니다.

6. 개인 회고

1) 나는 내 학습목표를 달성하기 위해 무엇을 어떻게 했는가?

2) 전과 비교해서, 내가 새롭게 시도한 변화는 무엇이고, 어떤 효과가 있었는가?

3) 마주한 한계는 무엇이며, 아쉬웠던 점은 무엇인가?

4) 한계/교훈을 바탕으로 다음 프로젝트에서 시도해보고 싶은 점은 무엇인가?

5) 내가 해본 시도 중 어떠한 실패를 경험했는가? 실패의 과정에서 어떻게 극복했는가?

정봉기(ML/AI)

1) NLP를 주로 공부해왔어서 텍스트처리와 LSTM이나 트랜스포머 기반의 모델에 대한 이해는 쉬웠으나 음성데이터를 처리하는 방법엔 어려움을 겪어서 음성인식 모델 구축을 위해 관련 논문을 읽고 구현을 위해 torch-audio 라이브러리를 공부하였습니다. 특히 발음에 대해 피드백을 해야하는 부분이 쟁점이라서, 관련 공부를 하였습니다. 우선적으로 발음을 알려주는 어플을 사용해보며 자료조사를 진행하였고 , 발음에 대한 논문을 읽어보며 국제발음이 있다는 것을 알게 되었고 IPA국제발음 변환법에 대해 공부를 하였습니다.

2) 데이터 자체를 수집하는 건 문제없으나 음성처리와 음성데이터를 해본 적은 없어서 이 부분을 처음 공부해보았습니다. 이번 서비스에 사용한 딥스피치2 모델은 BiLSTM이 백본인데 이 모델과 BERT,로 이미 다른 의료관련 서비스를 실제 병원에 배포를 해본 적이 있습니다. 다만 이번 프로젝트에선 Deep Speech 2 논문에 있는 내용을 바탕으로 모델 구조를 적용하고자 하였는데, KoSpeech(음성인식 툴) 의 기본 구조는 CNN * 2, RNN * 3 으로 구성되어 있었습니다. Baidu 의 Deep Speech 2 논문에 따르면 CNN * 3, RNN * 7 가 성능이 좋다는 것을 찾을 수 있었습니다. 그리하여 발음 피드백 시스템 적용을 위하여 심층적인 모델이 필요하다고 판단하였고, 이를 적용하기 위해 코드를 수정하고 그 대신, 레이어가 겹쳐질수록 모델의 복잡성이 올라가 학습 속도가 현저히 느려지므로 해당 trade-off 관계에서 적당한 설정으로 접근하였습니다. 이렇듯 심층적으로 모델 아키텍쳐를 수정해보며 접근할 수 있었습니다.

3) GCP(google cloud platform)를 최대한 활용하고 싶었으나 이루지 못하였습니다. 이전엔 그냥 로컬을 연결해서 서비스를 구현하곤 했었는데, GCP의 컨테이너나 도커들을 각각 연결하여 유기적인 아키텍쳐를 구현하고 싶었습니다. 다만, 전공이 컴퓨터관련이 아니라 처음부터 배워야할 부분이 많았고 어느정도 공부해보며 따라가더라도 이것을 제대로 구축했는 지는 알 수는 없었습니다. 시간도 짧았어서 아쉬워서 만약 본선에 들어간다면, 이 부분을 최대한 구축해보고싶습니다.

4) 앞서말한 도커와 쿠버네티스를 공부하여 (mlops 등) serving할 때 좀 더 원활히 해보고 싶습니다. 현재 llm 서비스로 국회생성ai공모전도 성공적으로 마무리했었는데 이때도 이러한 mlops 부분 이슈를 해결하지 못했어서 다음 프로젝트는 꼭 시도하여 해내고 싶습니다.

5) kospeech의 기반에 커스텀데이터를 학습하다보니 여러 성능면으로 문제가 생겼습니다. 파인튜닝 방법론을 처음부터 탐독하며 분석하여 다음과 같은 방법으로 극복했습니다.

[momentum 계수 수정을 통한 학습 성능 개선]

Deep Speech 2 논문 내용을 바탕으로 모든 BatchNorm 에 대하여 momentum 계수를 0.99 으로 적용하는 것을 알 수 있었습니다. 하지만, KoSpeech(툴킷) 의 momentum 계수 기본 설정은 0.1 이었고, 이에 따라, 모든 BatchNorm 에 대하여 momentum 계수에 0.99 를 적용할 수 있었습니다. 이러한 결과로 기울기에 이전 관성이 적용되어 local minima 현상을 억제할 수 있었으며 CER 감소 추세가 보다 linear 하게 바뀐 것을 확인할 수 있었습니다. 이런 식으로 여러 문제들을 해결했고 다음 깃허브레포를 통해 여러 learning과 관련된 이슈를 해결했습니다.

[local minima 현상 issue]

해당 레포의 방법론을 참고하였습니다. https://github.com/sooftware/pytorch-lr-scheduler

이상무(FE/BE)

1) 앱 개발은 처음 시도해보았기 때문에 관련 지식들을 먼저 습득할 필요가 있었습니다. 플러터 관련 강의들을 들으며 기본기를 익혔고, 바로 코드를 작성하며 어떤 키워드가 어떻게 작동하는 지 체득하며 배우는 방법을 택하였습니다.

2) 백엔드, 웹 개발만 고집하던 저에게 프런트엔드, 앱 개발을 새로운 시도였습니다. 학습하고 개발하며 앱 관련으로도 가능성을 열어두어야한다는 것을 깨닫게되었고 생각보다 흥미를 느껴, 앞으로 풀스택 개발자가 되기 위한 첫번째 발판이 될 것 같습니다.

3) 플러터를 처음 접해보았기 때문에, 모든 기능들을 수월하게 코드를 작성할 수는 없었습니다. 아직 배움과 경험이 부족하여 목표를 전부 다 이루지는 못하였습니다. 시간을 더 많이 투자하였다면 좋은 퀄리티의 결과가 나오지 않았을까 싶습니다. 또한, GCP를 최대한 활용하고 싶었으나 단지 oauth연결만 시도했던 것이 아쉬웠습니다. 이전엔 로컬로만 작업하여 이러한 부분이 생소했던 것 같습니다. 또한 모바일 디바이스에서 녹음과 전송, 데이터 처리 부분에서 어려움을 겪었습니다. 어떻게 디바이스에서 음성 데이터를 받고, 그것을 어떻게 서버로 보낼 것인지에 대해 고민하였습니다.

4) Google Cloud Platform의 다양한 기능을 활용할 수 있도록 배우고, Github 뿐만 아니라 도커 등 여러

협업 및 관리 툴을 사용하여 프로젝트를 진행하고 싶습니다. API를 활용하여 특정 기능을 만들어내는 부분이 부족했던 것 같습니다. 앞에서 말한 디바이스의 Input을 어떻게 처리할 지, 그것을 어떻게 다른 기능과 연결할 지 충분히 학습한 후 개발한다면 더 좋은 경험이 될 것 같습니다.

5) Widget으로 리턴할 속성을 어떤 것으로 정하는가에 따라 실제 표시되는 부분에서 약간 다른 면이 있었습니다. 또한 Chatview 속성을 사용함에 있어서 Overflow가 일어나는 부분을 생각하여야 했고, 이 과정에서 채팅 박스가 쌓이는 과정을 Scrollview로 포함해야하는 것에 있어 어려움을 겪었습니다. 또한 속성의 상속 순서가 앱 실행에 있어 중지를 유발한다는 것을 알게 되었습니다. 이 과정에서 코드를 작성하는 대로 컨테이너를 구상할 것이 아니라, 짜임새 있게 아키텍쳐를 구성하고 코드를 작성하는 것이 더 효율적이라는 것을 깨닫게 되었습니다.

이채은(PM/Design)

1) 아이들을 위한 서비스인만큼 귀여운 캐릭터 디자인을 하고자 노력했습니다. 인기 있는 캐릭터, 게임, 장난감 레퍼런스 조사를 했고, 이를 바탕으로 '다채로운 색상'이라는 목표를 설정하여 디자인하였습니다. 플로우 차트, 3D 캐릭터, UI, 영상 순서로 디자인을 했습니다. 사용하는 프로그램이 다양하고, 작업물이 많은 만큼 효율적으로 시간을 사용하고자 했습니다.

2) 가장 신경을 많이 썼던 부분은 캐릭터입니다. 서비스를 대표하는 얼굴이자, 아이들의 흥미를 이끄는 부분이라고 생각했습니다. 부모와 아이의 유대감을 형성해야 하기 때문에 최소 2개 이상의 캐릭터가 필요했습니다. 제작 과정에서 마야, 블렌더 툴도 새롭게 배웠습니다.

3) 마야, 블렌더도 사용할 수 있지만, 작업물의 용량과 시간적 효율성을 위해 최종적으로는 일러스트의 3D 기능을 활용했던 점이 조금 아쉬웠습니다. 다음 프로젝트에서는 마야, 블렌더를 좀 더 활용해 보고 싶습니다.

4) 다음 프로젝트에서는 프로그램 언어를 더 학습한 후 참여하고 싶습니다. 디자이너와 개발자가 사용하는 용어가 달라서, 쉽게 소통하기 위해 열심히 배워보려고 합니다. 여유가 된다면 실제 서비스 런칭까지 시도해보고 싶습니다.