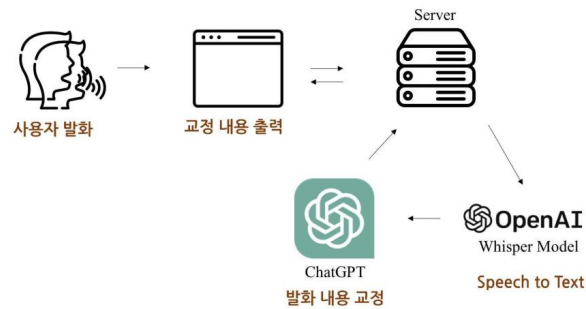


# 프로젝트 수행계획서

과제명	Transformer 기반의 화자별 음성인식과 ChatGPT 를 활용한 화법교정		
참여 인원	(총 6 명) 책임교수 1 명, 학부과정 5 명		
수행 기간	2023.03.01.~06.30.(4 개월)	유형	중기
추진 배경	<ul style="list-style-type: none"><li>○ 의사소통으로 인한 다양한 문제를 인식하고 그를 교정하는 서비스의 필요성을 인식</li><li>○ ChatGPT 의 등장으로 다양하고 복잡한 질문에 대한 답변이 가능해짐</li><li>○ 관련한 기술들을 통해 관련 서비스를 구현할 수 있음을 확인함</li></ul>		
목표 및 내용	<ul style="list-style-type: none"><li>○ Transformer 기반의 화자별 음성인식 모델과 ChatGPT 를 활용하여 화법교정을 목표로한다<ul style="list-style-type: none"><li>- 목적 달성을 위해, 본 과제를 크게 3 가지 부분으로 나눌 수 있다<ol style="list-style-type: none"><li>1. 화자별 음성인식 및 전사</li><li>2. ChatGPT 를 활용한 화법교정</li><li>3. 1, 2 의 과정을 수행하여, 사용자에게 결과를 보여줄 수 있는 User Interface 구현</li></ol></li><li>- 추진방법<ol style="list-style-type: none"><li>1. OpenAI Whisper 모델을 사용하여, AI-Hub 의 한국어 음성 데이터를 가지고 speech-to-text 를 학습 시킨다</li><li>2. 좋은 화법에 관해 학습 시킨 ChatGPT 를 API 호출을 통해 (1)에서 받은 대화내용에 대한 화법교정을 받는다</li><li>3. User Interface 는 Python 기반의 웹 언어인 Flask 를 가지고 구현할 예정이며, 사용자가 대화내용을 담고 있는 음성파일을 업로드하면 (1)과 (2)의 과정을 거쳐 나온 결과를 사용자에게 보여주게 하도록 구현한다</li></ol></li></ul></li></ul>		



### <전체 시스템 구조>

#### 기대 효과

- 역량과 관련된 내용
- 개발한 서비스로 얻을 수 있는 효과
- 고급 자연어 처리 알고리즘을 이용한 정확도의 향상으로, 이를 통해 프로젝트에 경쟁력과 차별화를 기대할 수 있음
- 의사 소통에 어려움을 겪은 개인에게 서비스를 제공하고 다양한 배경을 가진 사람들 간의 효과적인 의사 소통의 촉진을 기대할 수 있음

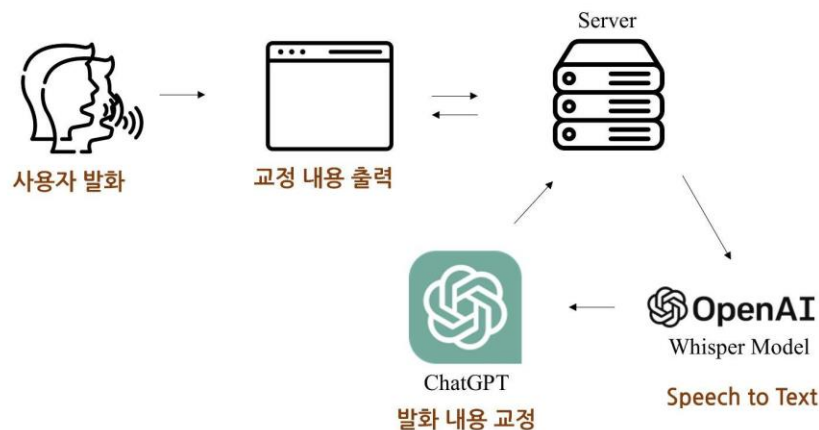
## 1. 과제 목적 및 필요성

정보기술발달에 따라 다양한 사람간의 의사소통 수단이 등장하였다. 또한 태어났을 때부터 디지털 기기를 사용한 "Z"세대가 이미 성인이 되어 사회로 나가고 있는 실정이다. SNS 와 메신저가 전화통화보다 익숙하고 이런 단방향적이고 단편적인 대화에 적응이 되어있고 세대간 사용하는 언어의 차이가 존재하는 등의 이유로 의사소통의 문제가 사회 전반적으로 드러나고 있다.

이러한 현상은 사회가 언택트 시대에 들어선 COVID-19 이후 더욱 심화되었다. 팀원들간의 주제 선정 토의중 의사소통 문제에 대한 사회적인 공감대를 확인할 수 있었다. 제시된 주제 중 음성을 텍스트로

변환하는 모델을 바탕으로, 음성을 화자별로 텍스트화하여 교정해 주는 서비스에 대해 생각하게 되었다.

서비스 구현을 위한 정보를 찾아보며 상용화 되어있는 기술들을 확인한 결과 관련한 딥러닝 모델을 찾을 수 있었고 이를 바탕으로 대화 교정은 생성 AI 인 **ChapGPT** 를 통해 수행되며, 화자 별 음성 텍스트 변환은 **OpenAI Whisper** 모델로 구현되는 서비스를 프로젝트로 진행하고자한다.

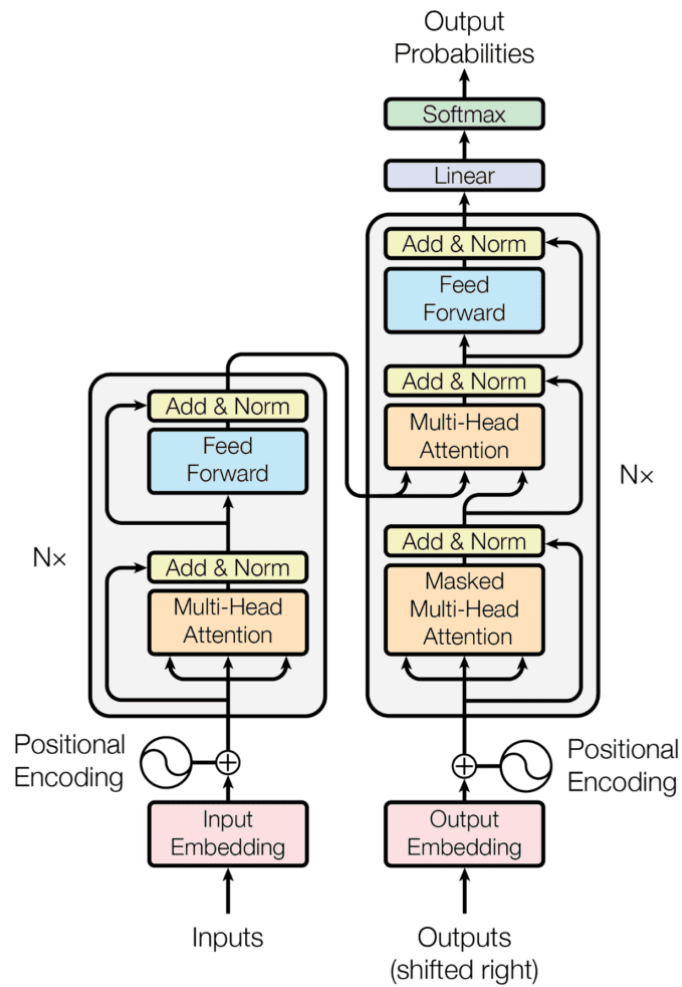


<그림 1, 시스템 구조>

## 2. 과제 내용 및 추진 방법

### 2-1. 과제분석

본 과제는 크게 세 가지 부분으로 나눌 수 있다. 첫 번째는 화자별 **speech-to-text**, 두 번째는 **ChatGPT** 를 활용한 화법 교정, 그리고 **User Interface** 구현이다. 먼저, **Transformer** 기반의 음성인식을 통해 화자별로 **speech-to-text** 를 수행한다. 그 후 텍스트로 변환된 대화내용들을 **API** 호출을 통해 **ChatGPT** 로 넘겨준 후, 상황에 맞는 공감적 화법에 기반한 대화법을 추천 받는다. 마지막으로, **User Interface** 에서는 사용자가 대화 음성파일을 업로드를 하면 위에서 언급한 모든 과정을 처리하여 사용자에게 결과를 보여주는 것을 목표로 한다.



<그림 2, Transformer Architecture>

## 2-2. 추진방법

### 1. OpenAI Whisper 활용

먼저 한정지능정보사회진흥원에서 제공하는 약 **2TB** 용량의 한글음성 데이터를 사용하여 OpenAI 에서 제공하는 Transformer 기반의 speech-to-text 모델인 Whisper 를 학습 시켜 화자별로 발화내용을 전사하도록 하게한다.

### 2. ChatGPT 활용

OpenAI Whisper 에서 전사된 대화 내용을 바탕으로 보다 나은 공감적 화법을 추천하기 위해 ChatGPT 를 사용한다. 아래의 그림은 ChatGPT 가 사용자가 입력한 대화 내용을 바탕으로 보다 좋은 화법을 제안하는 예시이다.



A : 저희가 할 주제로 chatgpt를 사용하는건 어때요.

B : 괜찮은 것 같아요

C : 좋긴한데.. 흠...



네, 맞습니다. 대화내용을 수정한 이유를 알려드리겠습니다. 이번 수정에서는 대화내용에서 더 자연스러운 표현과 일관된 어투를 사용하여 대화의 흐름을 개선하려고 노력했습니다. 예를 들어, A의 말에서 "할 주제로"라는 표현은 "사용하는 주제로"로 수정하여 좀 더 자연스러운 표현으로 변경했습니다. 또한 C의 말에서는 "흠..."이라는 의문을 제거하고, 대화의 진행을 이어갈 수 있도록 구체적인 이유를 덧붙였습니다. 이렇게 대화를 보다 원활하고 자연스럽게 진행하기 위해 수정을 했습니다.

성혁

Bella: Hey, you Dina. Stop!

Dina: What happened, Bella?

Bella: Did you steal my money? You're the one to blame!

Dina: it wasn't me. I didn't do that.

Bella: But everybody in my class said you do that. You have to explain about this!

Dina: Huh! How dare you're. I never do that such things.

Bella: Yeah, but my friends had seen you out of my classroom. It must be you stealing my money.

Bella : Ok, I will prove that!

Dina: I will wait for that!



Sure, here's a revised version of the conversation using "I messages":



Bella: Hey, Dina, I need to talk to you.

Dina: Sure, what's up, Bella?

Bella: I'm feeling really upset because I can't find my money. When I last saw it, you were in the room. Can you help me understand what happened?

Dina: I'm sorry to hear that, but I didn't take your money. I don't know what happened to it.

Bella: I understand, but some of my classmates mentioned that they saw you near my classroom when the money went missing. Can you help me understand what happened from your perspective?

Dina: I understand why you might be suspicious, but I didn't take your money. I was walking by your classroom, but I didn't go in. I'm happy to help you look for it, though.

Bella: Thank you for explaining that to me, Dina. Let's work together to try to find my money and figure out what happened.

🔄 Regenerate response

<그림 3, ChatGPT의 화법 추천 예시>

이렇게 ChatGPT 홈페이지에서 이러한 기능 등을 사용할 수 있지만 본 프로젝트의 목표는 사용자 발화를 입력받아 좋은 화법을 제공하는 서비스를 만드는 것이므로 OpenAI의 오픈 라이브러리를 이용해

ChatGPT API 를 사용한다. 또한, ChatGPT 에서 더 높은 품질의 답변을 듣기 위해서는 해당 모델이 잘 이해하고, 잘 작동할 수 있는 프롬프트를 작성하는 것이 중요하다. 이를 위해, 좋은 화법에 관한 연구를 조사하여, ChatGPT 에게 학습을 시킬 예정이며, 또한 화법 교정을 요구하는 다양한 프롬프트를 작성하여 ChatGPT 에 입력해 결과를 확인해 보면서 더 좋은 화법교정을 답변 받을 수 있는 최적의 프롬프트도 찾을 예정이다.

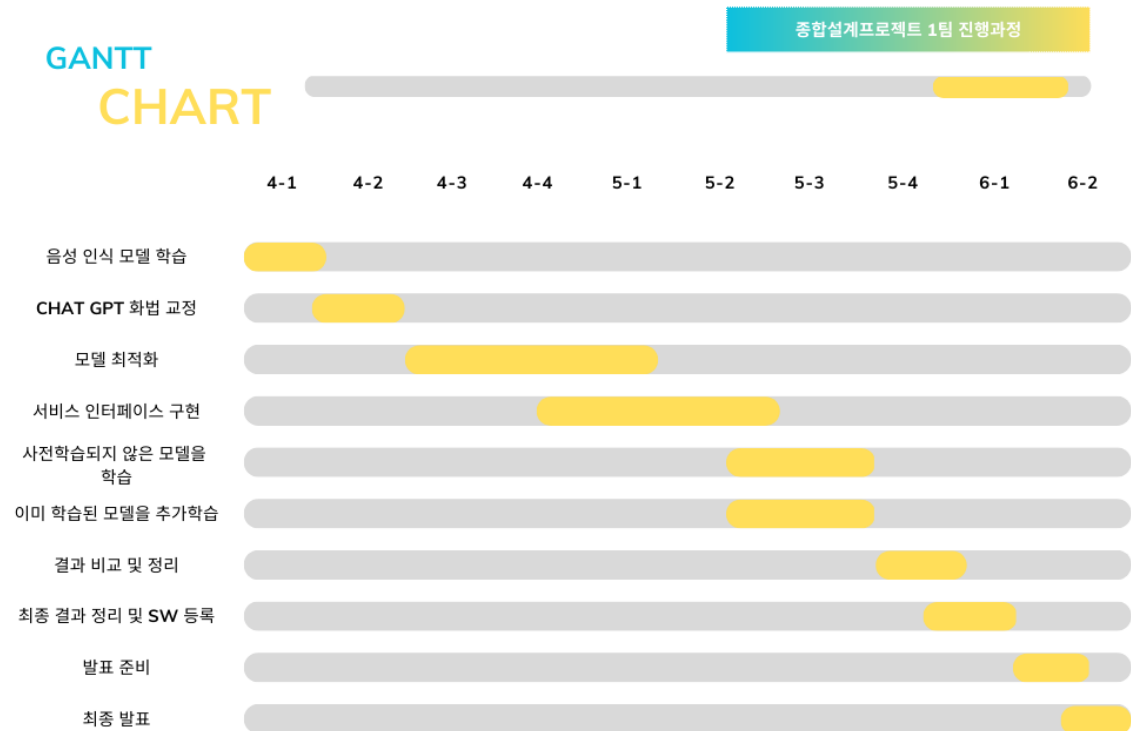
### 3. UI 구현



<그림 4, User Interface 예시>

User Interface 는 Python 기반의 웹 프로그래밍 언어인 Flask 를 사용하여 구현하는데, User Interface 에서는 사용자가 WAV 파일 형식의 음성 대화 파일을 업로드하면, Whisper model 을 사용하여 화자별로 발화를 전사한 후, 그것을 ChatGPT 로 넘겨주어 상황에 맞는 공감적 화법을 추천 받아 사용자에게 보여주는 것을 구현하고자 한다.

### 3. 과제 추진 일정 및 예산 활용 계획



<그림 5, 간트 차트>

담당자별 업무는 아래에 기술함.

#### 4. Participating Personnel(Detail)

지도교수	소속	경북대학교 컴퓨터학부		성명	김승호
과제 참여학생	소속(학과)	학위과정 (성별)	학번	성명	담당 업무
	컴퓨터학부	학사과정(남)	2019117923	배완기	ChatGPT 화법 교정, 발표 준비
	컴퓨터학부	학사과정(남)	2018115065	강창명	모델 최적화, 이미 학습된 모델의 추가 학습
	컴퓨터학부	학사과정(남)	2018114449	김동현	결과 비교 및 정리, 최종 발표
	컴퓨터학부	학사과정(남)	2018116698	최성혁	음성 인식 모델 학습, 서비스 인터페이스 구현
	컴퓨터학부	학사과정(남)	2018115122	김기욱	사전 학습되지 않은 모델의 학습, 최종 결과 정리 및 SW 등록

#### 5. 기대효과 및 활용방안

서비스를 개발하며 사용한 생성 AI 와 텍스트 변환에 사용되는 딥러닝 기술을 다뤄보며 상용화된 기술들을 활용하는 경험을 통해 기술에 대한 이해도가 증진될 것이라 기대된다. 또한 상용화되어 있는 서비스 혹은 참고할만한 논문등 찾아보며 서비스와 관련된 기술을 응용하는 능력을 배양할 수 있다.

프로젝트 진행과정 동안 전체적인 SW 개발 프로세스를 경험할 수 있으며, 서비스를 위한 일련의 과정을



진행하면서 학부과정에서 배울 수 있었던 전공 지식을 실제 응용해 볼 수 있다. 팀원들과 함께 작업을 하기에 협업을 통해 같은 목표를 통해 각자의 역할을 분배하고 맡은 바를 해내며 그 결과물을 함께 만들어가는 과정에서 많은 것을 배우고 느낄 수 있을 것이다.

현재 기술 발전의 속도가 급격히 증가함에 따라 기술과 윤리 혹은 기술과 사람들의 우려가 상충하는 부분이 많다. 이러한 부분에 대해 고민해보며 기술이 삶을 더 편리하게 만들어 주는 서비스를 고민해보면서 궁극적으로 "좋은" 프로그램을 만드는 개발자가 되는 것을 목표로 개발하고자 한다.

프로젝트를 통해 개발하고자 하는 서비스는 단순히 특정 기술을 다뤄보기 위한 단편적인 개발이 아니라, 실제 삶에 도움이 될 수 있는 서비스이다. 따라서 사용자의 경험을 고려하는 설계를 통해 더 나은 서비스를 개발하는 방법에 대해 고민할 수 있다. 이러한 경험들을 통해 단순히 기술을 쫓는 개발자가 아닌 기술을 활용하고 더 나은 삶을 위한 서비스를 개발할 수 있는 개발자의 역량을 쌓을 수 있기를 기대한다.

서비스의 활용 방안으로 다음의 4 가지가 있다.

첫째론 휘발성이 강한 음성을 텍스트로 인식하여 기록하고 저장하므로 이를 통해 중요한 발화를 잊지 않고 기록하며 위 정보를 활용할 수 있다. 기록된 정보를 바탕으로 이후 관련된 내용을 정리하고 복기할 수 있다.

둘째로 위의 활용방안을 바탕으로 확장되는 서비스로, 말을 하면서 스스로 자각하지 못했던 화법을 텍스트로 인식하는 것이다. 발화 상황에서 화자 자신이 스스로 판단하여 말하는 모든 것들이 항상 정확한 것은 아닐 수 있다. 하지만 말하는 도중 이를 판단하는 것은 매우 어려운 일이다. 따라서 발화 중 이상하다고 생각하지 못한 부분들을 기록을 통해 확인할 수 있다.

셋째로 많은 사람들의 공감을 얻는 좋은 화법을 추천해주는 것이다. 의사소통에 장애를 일으키는 여러 요소들을 파악하고 그게 맞는 해결법을 제시할 수 있으며 단순히 맥락에 맞는 말이 아닌 상황에 시의적절한 말을 할 수 있도록 도와주는 것을 목표로 한다. 이를 통해 궁극적으로 사용자의 화법을 상대방의 의도에 맞게 상황에 맞는 발화를 하도록 교정하는 것이다.

넷째로 화법에 항상 정답이 있다고 말하기는 어렵기 때문에 다른 상대방의 말에 담긴 의미를 파악해주는 것이다. 말에는 항상 같은 의미만이 담기는 것은 아니다. 주어진 상황과 대화의 맥락에 따라 말하기는 다양한 방법으로 나타날 수 있다. 이에 대한 판단력 또한 말하기에 있어 중요한 부분이다. 이에 텍스트로 받아들이는 말하기 내용을 각 화자의 관점으로 분석하여 제공하고 그 실질적인 의미에 맞는 발화를 할 수 있는 역량을 키워줄 수 있다.

위와 같은 서비스를 통해 궁극적으로 추구하고자 하는 바는 사용자가 대화 상황에서 상대방의 말의 의도를 파악하고 그 말에 맞는 대답을 할 수 있게 하는 것이다. 이는 추진 배경에 드러나 있는 현 세대의 의사소통의 문제를 해결하는데 일조할 수 있으며, 이후 스피치학원에서 제공하는 솔루션들을 참고해 유료 서비스로의 확장도 가능할 것이라 판단한다.

## 6. 예상되는 주요 과제성과

프로그램 완성 후 한국저작권위원회에 SW 저작권등록을 추진할 예정이며, 팀의 개발 일정과 맞는 공모전 혹은 대회가 있다면 적극 참여할 예정이다.