

[Capstone Design Project 2020 최종보고서]

# Conversational Ai for Education

지도 교수 : Bernardo N. Yahya (사인)

제출일 : 2020.06.30

---

Ahtty 팀

---

201502426 이승준(PM) (사인)

201500429 김민우 (사인)

201500514 김성진 (사인)

---

# 목차

제 1 장. 프로젝트 전체 요약문 .....	3
제 2 장. 추진 배경 및 필요성 .....	3
제 2 장. 추진 목표 및 내용 .....	6
제 1 절. 프로젝트 목표 .....	6
제 2 절. About Transfer Learning .....	7
제 3 절. 기존의 방법론 .....	7
제 4 절. 대화형 Ai .....	8
제 4 장. 추진 결과물 .....	14
제 1 절. 앱 개발 및 모듈 훈련 시간 .....	15
제 2 절. 어플리케이션 구현 .....	16
제 5 장. 향후 과제 및 최종 결과물 활용계획.....	17
제 7 장. 참고문헌 정리.....	19
제 8 장. 프로젝트 수행후기.....	20

## 제 1장. 프로젝트 전체 요약문

# Conversational Ai For Education

201502426 이 승 준

201500429 김 민 우

201500514 김 성 진

### 요 약 문

본 프로젝트에서는 개인 평가 시스템을 도입한 대화형 챗봇 기반의 영어 회화 애플리케이션을 제작하였다. 애플리케이션의 핵심 기능은 대화형 Ai와 AFL이다. 기능을 구현하기 위한 프로젝트 프로세스는 다음과 같다. AI Hub 와 줌마 영어 데이터를 수집하였고, 각각 비슷한 task끼리 Data Matching을 진행하였다. 이후 토큰화 작업과 같은 Text Preprocessing을 거친다. 그다음 Personality Permutation 작업을 통한 Data Augmentation을 완성한다. 완성된 데이터를 바탕으로 Model Training을 통해 대화형 AI 모델과 AFL 모델을 구축한다. 이후 데이터셋, 서버 환경에 맞게 Parameter Optimization 작업을 진행한 뒤, 모델 후보군 중 최적 모델을 선정한다. AFL 모델에서는 테스트를 통해 각 모델의 임계치를 설정하고, 임계치 초과 시 다른 상황 대화를 이끌어 멀티 턴의 대화가 될 수 있도록 한다. 최종적으로 더 자연스러운 대화를 위해 Decoder setting을 변경한 뒤, 대화형 AI와 AFL 모델을 Rest API 형식으로 구축한 뒤 Flutter를 사용하여 최종 애플리케이션을 구현한다.

최종 구현된 대화형 Ai의 경우 AFL 모델을 통해 사용자의 학습 상태를 지속해서 확인한다. 사용자가 학습 임계치를 초과하는 경우 다른 상황 대화를 이끌어 나가 한정적인 소재의 대화가 아닌 멀티 턴의 대화를 이끌어 나가는 결과를 도출하였다. 또한 줌마 영어 Task에 맞춤화된 데이터를 활용한 결과 수강생의 교육 콘텐츠 기반 학습 증진에 도움을 주었다. 향후 사용자가 언제 어디서나 지속해서 학습할 수 있는 환경을 만드는 데 도움을 줄 것으로 예상된다.

## 제2장. 추진 배경 및 필요성

'엉클선의 줌마영어'라는 영어 회화 교육 기업과 함께 이 프로젝트를 진행한다. '엉클선의 줌마영어'의 수강생은 주로 4~50대 주부로서 영어를 처음 배우는 분들이 대다수이다. 기업은 수강생 대비 강사의 Resource 제한으로 인해 즉각적인 피드백의 어려움을 겪고 있다. 또한, 영어 교육을 진행한 뒤 이를 바탕으로 복습할 플랫폼을 개발하는 것을 목표로 한다. 수강생은 교육 후 실제로 영어를 활용할 공간이 없어 문제점과 자신이 제대로 학습한 내용을 바탕으로 회화 능력을 평가할 수단도 마땅치 않다는 어려움이 있다. 따라서 우리는 교육내용을 보다 효율적으로 복습할 방법과 실제 회화를 나눌 수 있는 플랫폼을 구축하고 제한된 기업의 Resource 안에서 효율적으로 학습 성취도를 평가할 수 있는 장치가 필요하다고 판단하였다. 본 팀은 여러 문제점의 해결책으로 줌마영어 콘텐츠 기반 대화형 Ai 챗봇 제작을 목표로 한다.

수강생들이 챗봇과 회화를 나눌 플랫폼 구현체로 애플리케이션을 제작한다. 따라서 시중에 존재하는 여러 영어회화 애플리케이션 중 사용빈도가 높고 평점이 우수한 3가지 대표 애플리케이션을 사전 조사 및 분석하였다. 비교 결과는 <그림1> 같다. 1) Andy의 경우 챗봇을 통해 교육을 진행하는 부분이 본 팀이 진행하고자 하는 대화형 Ai 챗봇 구현 과정과 유사하다. 하지만 Andy의 경우 카테고리 방식으로 정보를 제공하지 못하고 있으며 회화를 하는 자연스러운 대화 과정이 아니다. Andy는 영어단어나 문법의 개념을 설명해주고 이를 숙지했는지 수강생이 알아서 판단해야 한다. 따라서 처음 영어를 배우는 사람의 경우 자신이 제대로 숙지했는지 평가할 수단이 없다는 단점이 존재한다. 2) Cake의 경우 영화, 여행, 코미디, 인기 등 여러 가지 카테고리를 제안한다. 또한, 영상을 보며 따라 읽는 반복 학습을 진행하면서 영어 회화 실력을 키운다. 하지만 실제 챗봇이나 보이스봇과 대화를 나눌 수 있는 수단은 없어 자유로운 대화를 체험할 수 없다. 또한, 빈칸 채우기 등 기존의 시험에 의한 평가 방식으로 운영된다. 따라서 대화를 하면서 자연스럽게 성취도를 평가할 수 없다는 단점이 존재한다. 3) 스피나우의 경우 AI를 이용한 영어 회화 앱이다. 스피나우는 AI 튜터와 대화를 나누면서 회화 실력을 키워가는 방식으로 진행된다. 회화를 나누는 토픽도 여러 가지가 존재한다. 스피나우가 본 팀이 진행하고자 하는 프로젝트와 매우 유사한 경향을 보인다. 하지만 스피나우도 단점이 존재한다. 바로 회화를 한 뒤 실제로 회화에서 자연스럽게 대화를 하였고 주어진 토픽 내에서 대화를 나누었는지 평가할 수단이 존재하지 않는다.

차이점에서 주로 살펴볼 수 있는 AFL(Assessment For Learning)은 학습평가 관련 논문(김명숙, 2015)의 AFL(Assessment For Learning)의 개념에서 착안하여 팀원들과 브레인스토밍을 통해 만들어낸 학습 성취도 평가 방식이다. AFL의 주요 개념은 다음과 같다. 기존의 전통적 형성평가로 사용되는 쪽지 시험, 5분 테스트, 퀴즈 등 시험 또는 검사지에 의한 형성 평가가 주를 이루고 있다. 반면 AFL은 학습 평가의 방식으로 학생과의 대화, 질의·응답하기, 관찰하기 등의 방식을 추가하였다. 기존의 평가 당시에만 확인할 수 있었던 학습 성취도를 매 대화의 순간으로 확대한 개념이다. 본 팀은 대화형 Ai 챗봇을 통해 사용자와 자연스러운 대화를 진행한다. 이는 AFL에서 제공하는 평가방식과 매우 유사하다. 본 프로

젝트에서는 평가를 진행하기 위해 매 순간 나누는 대화를 분석하여 현재 학습이 잘 진행되었는지 평가하는 지표로 AFL이라는 새로운 시스템을 구현하여 수강생의 현재 학습 성취도를 평가하여 현재 학습이 잘 진행되고 있는지를 수강생이 확인할 수 있도록 한다.

	유사점		차이점
1	Andy	<ul style="list-style-type: none"> <li>● 챗봇 대화</li> </ul>	<ul style="list-style-type: none"> <li>✓ AFL 평가 방식 도입</li> <li>✓ 카테고리 방식 추가</li> <li>✓ 자연스러운 대화에 초점</li> </ul>
2	Cake	<ul style="list-style-type: none"> <li>● 카테고리 방식</li> <li>● 평가 척도가 존재</li> </ul>	<ul style="list-style-type: none"> <li>✓ 단순 평가 방식을 AFL 평가 방식으로 변경</li> <li>✓ AI 챗봇과 대화 나누는 방식으로 진행</li> </ul>
3	스핀나우	<ul style="list-style-type: none"> <li>● 카테고리 방식</li> <li>● AI를 이용 보이스봇 형태로 대화 주고 받음</li> </ul>	<ul style="list-style-type: none"> <li>✓ AFL 평가 방식 도입</li> </ul>

〈표 1: 본 프로젝트와 유사점과 차이점〉

## 제2장. 추진 목표 및 내용

### 제 1절. 프로젝트 목표

본 팀은 프로젝트를 통해 줌마영어 task에 맞춤화 된 대화형 Ai 챗봇을 제공한다. 또한, 동시에 대화형 Ai 챗봇과 대화를 나눈 내용을 바탕으로 AFL 모델을 통해 학습 성취도를 평가하여 수강생들이 자신의 학습 성취도를 파악할 수 있도록 한다. 이는 현재 부족한 기업의 Resource 부담을 덜어줄 뿐만 아니라 수강생에게 자신의 학습 성취도에 대한 평가 척도를 제공함으로써 자신이 실제 학습이 잘 진행되었는지 평가할 수 있도록 한다. 각 모델을 구축하는 과정은 <그림1>와 같다.



<그림 1: 줌마영어 인터뷰를 기반으로 각 모델 목표 설정>

본 프로젝트를 진행함에 있어 중간 발표 이후 기업의 요구사항과 프로젝트 목표에 따른 세부 사항 부분을 Framework와 Application으로 나누었다. 그 이후 각 부분에서 수행해야 할 프로젝트 범위는 <그림 2>와 같다.

### 기업 요구사항

Enterprise requirements

#### 전화 영어 수업 전 예습 프로그램 필요

- 현재 전화 영어 서비스 'ďâtâlângâ' 시행 중
- 전화 영어 수업 전 교재 내용 예습 프로그램 필요
- 전화 영어 수업 후 효과적인 복습 방법 고안 중

#### 앱 개발 시 음성인식 기능 요구

- 전화 영어 시행 전 과제는 주로 네이버 밴드를 활용
- 줌마영어 회원 대부분이 타이핑에 미숙 → 음성으로 과제 제출

### 프로젝트 목표

Project Propose

1. 다양한 상황에서 자유로운 대화가 가능한 챗봇
2. 교육내용(컨텐츠)기반 교육 플랫폼

#### FRAMEWORK

- 컨텐츠를 기반으로 Conversational AI 개발
- 성취도 평가를 위한 AFL시스템 구축
- 최종적으로, Conversational AI와 AFL이 결합된 프로세스 구축

#### APPLICATION

- 컨텐츠 기반 단어, 문장 학습이 가능한 기본 교육 기능
- 컨텐츠를 기반으로 한 자유대화 플랫폼 제공으로 응용 교육 기능 제공
- 음성인식 기능 활용

<그림 2: 최종 프로젝트 목표>

## 제 2절. About Transfer Learning

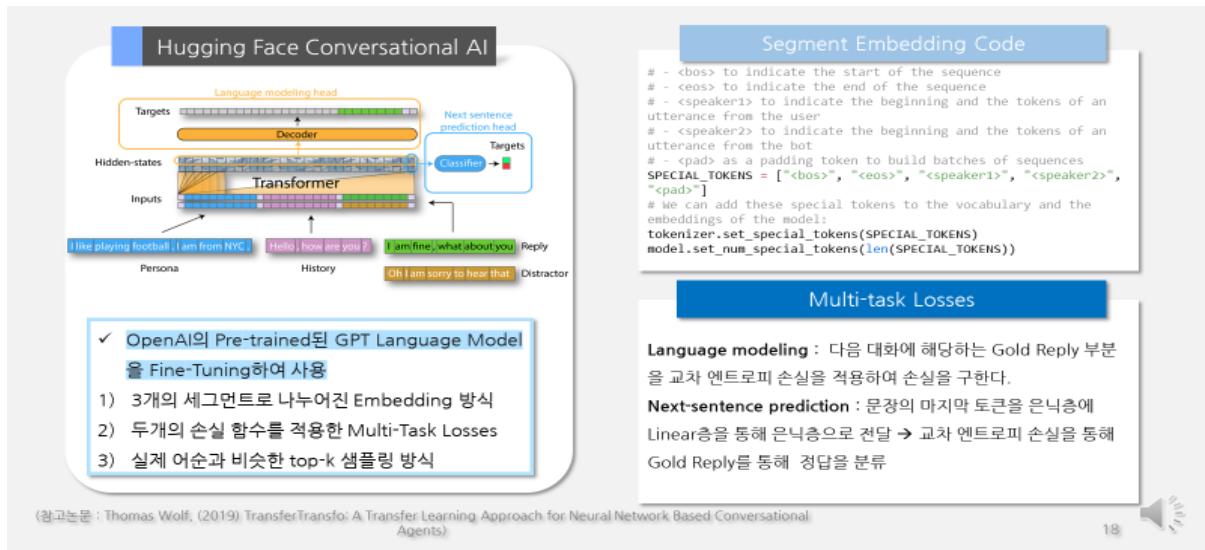
본 프로젝트의 Framework를 구성하는 대화형 Ai 챗봇과 AFL(Assessment For Learning)모델은 모두 전이학습 개념을 사용한다. 전이학습은 높은 정확도를 비교적 짧은 시간 내에 달성할 수 있기 때문에 컴퓨터 비전 분야에서 유명한 방법론 중 하나이다. (Rawat & Wang 2017). 전이학습을 이용하면, 이미 학습한 문제와 다른 문제를 풀 때도, 밑바닥에서부터 모델을 쌓아 올리는 대신에 이미 학습되어 있는 패턴(Pre-trained Model)들을 활용해서 적용할 수 있다. 본 프로젝트에서는 각 Task에 맞는 Pre-trained model을 사용한다. 그 이후 Fine-tuning 과정을 진행한다. Fine-tuning 과정은 각 Task에 맞게 데이터를 변환 후, 데이터를 적용하여 학습을 진행한다. Pre-trained 모델 사용은 다음과 같다. 대화형 Ai의 경우 Open Ai의 GPT-2 모델을, AFL에서는 Google의 B.E.R.T 모델을 사용하였다. Fine-tuning은 다음과 같이 진행하였다. 대화형 Ai의 경우 Hugging Face의 방식대로 줌마영어 데이터셋을 3가지 Segment로 나누어 Fine-tuning을 진행하였고, AFL에서는 MRPC(Microsoft Research Paraphrase Corpus), CoLA(The Corpus of Linguistic Acceptability) 데이터 세트를 사용하여 Fine-tuning을 진행하였다.

추가로, 사전 학습 모델 중 NLP 분야에선 B.E.R.T 주로 사용되지만 본 프로젝트에서는 Pre-trained Model로 Open Ai의 GPT-2를 사용하였다. 그 이유는 B.E.R.T 모델에서는 디코더를 사용하지 않아 문장 생성에 적합하지 않기 때문이다. GPT-2가 GPT-1 더 좋은 성능을 보여주지만, 본 프로젝트에서는 두 모델 모두 학습을 진행하여 더 좋은 성능을 보여주는 모델을 선택하였다.

## 제 3절. 기준의 방법론

### 1. Hugging Face 방법론

본 프로젝트에서는 대화형 Ai 모델을 제작하기 위해 Conversational Intelligence 대회(이하. ConvAI2)에서 State-of-the-Art(SOTA)를 기록한 Hugging Face 사의 Segment Embedding 방식과 디코더 세팅을 사용하기로 했다. 이와 같은 Hugging Face의 방법론을 사용하는 이유는 <그림3>과 같다. Segment Embedding의 방식의 경우 입력층 입력 시 Persona, History, Reply 총 3가지 Segment로 나누어 Embedding을 진행 한다. 출력층에서는 Multi-Task Losses를 사용한다. Multi-Task Losses는 다음과 같이 구성 된다. Cross-Entry 손실 함수를 사용한 Language Modeling과 주어진 History를 통해서 다음 문장 예측을 위해 Next-Sentence Prediction을 사용한다. 위와 같은 이유로 Hugging Face의 방법론을 사용하였다.



<그림 3: Hugging Face 방법론>

## 2. 한계점

기존의 Hugging Face의 방법론의 문제점은 크게 2가지로 나눠진다. 1) 랜덤형으로 Persona 선택이 가능하다는 점과 2) 영어 교육용 대화가 불가능하다는 점이다. 1) 랜덤 선택되는 Persona의 문제점을 해결하기 위해 AFL 기능을 추가하여 사용자의 학습률에 따라 Persona가 변경되도록 계획하였다. 2) 영어 교육용 대화 불가능 문제를 해결하기 위해 한국 AI Hub에서 제공하는 한국어-영어 말뭉치 데이터셋을 사용하여 교육용 대화 즉, 여러 상황에서 대화가 가능하도록 구현하였다.

## 제 4절. 대화형 Ai

프로젝트 진행을 위해 줌마 영어 교제 대화 데이터화 한국 AI Hub에서 제공하는 한국어-영어 말뭉치 데이터셋(상황 데이터)을 사용하였다. 이후, Hugging Face의 Segment Embedding 방식을 사용하기 위해 본 프로젝트의 Task에 맞게 데이터 변환 및 전처리 과정을 진행 하였다. 데이터 변환 작업을 위해 줌마 영어 데이터의 각 챕터(상황)와 AI Hub 상황 데이터의 각 상황에 맞게 데이터 매칭을 통하여 데이터 변환 작업을 진행하였다. 이후, 텍스트 전처리 과정은 다음과 같다. 1) Expanding Contractions 작업을 통해 Contraction을 제거한다. 2) Tokenization 작업을 통해 토큰화된 문장 성분을 리스트 형태로 출력한다. 3) Converting all Characters to Lowercase 작업을 통해 모든 문자를 소문자화 한다. 4) Lemmatization 작업을 통해 동사의 형태를 원형을 변형한다.

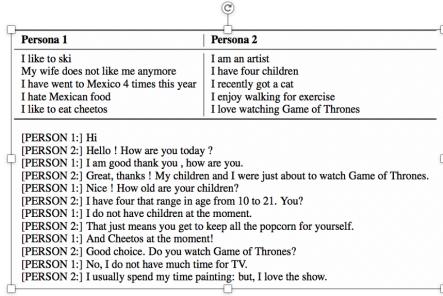
### 1. ConvAI2 데이터 구조

기존 ConvAI2에서 제공하는 데이터 셋은 <그림4>와 같다. 3개의 Embedding을 위해 데이터 셋은 Persona, Candidate, History 세 가지 부분으로 나누어 진다. 모델 훈련 진행 시. 다음 문장 예측을 위

해 History에서는 흘수 개의 문장이 주어진다. 즉, 1,3,5개의 문장을 사용하여 다음 문장을 예측한다. 다음 문장 예측을 위해, Candidate에서, 다음 문장 후보군에서 문장을 선택하여 학습을 진행한다. 또한, 학습량을 늘이기 위해, 같은 Persona라도 데이터 순서를 바꿔가며 다른 History 데이터와 결합하여 데이터 수를 증가 시켰다.

## ConvAI2 데이터 구조

### ConvAI2 Data Structure



- History에서 주어진 문장(흘수 개)을 통해 다음 문장 예측 → 훈련
- Candidate : 다음 발화의 예측에 사용되는 문장, 마지막 줄 정답
- ⟨Persona 1⟩ 과 ⟨Persona 2⟩를 분리하여 Dialog 생성
- Data Augmentation을 위해 Persona 순서를 바꿔가며 History 생성

### PERSONA ConvAI2 'Persona 1'

I like to ski.  
 My wife doesn't like me anymore.  
 I've went to Mexico 4 times this year.  
 I hate Mexican food.  
 I like to eat cheetos.

### CANDIDATE

Awesome . So what do you do for a living ?  
 You are telling the truth lol  
Haha yea, I love cookies . I sleep way to much to . You like music ?  
 That is nice to know . I travel due to being financially stable.  
 I am a karate instructor and love it  
 ...  
 Great, thanks! My children and I were just about watch Game of Throne.

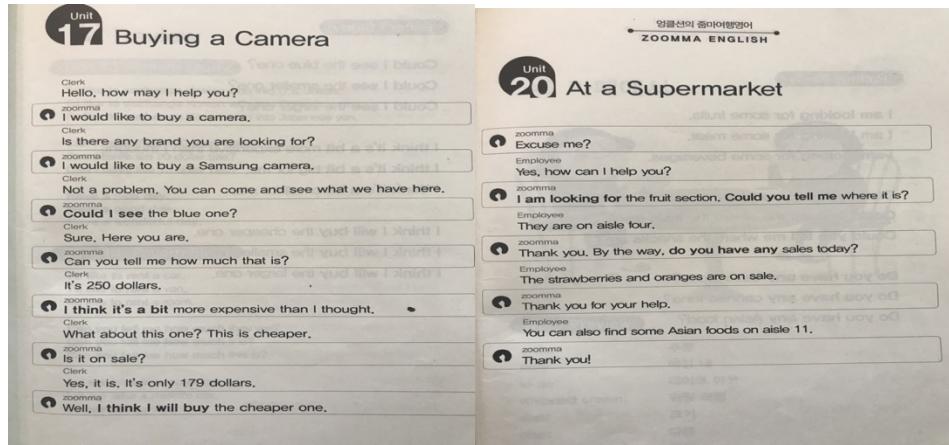
### HISTORY

⟨speaker1⟩ : Hi  
 ⟨speaker2⟩ : Hello! How are you today?  
 ⟨speaker1⟩ : I am good thank you, how are you.  
 ⟨speaker2⟩ : Great, thanks! My children and I were just about watch Game of Throne.

〈그림 4: ConvAI2 데이터셋 구조〉

## 2. 데이터셋 변형

본 프로젝트에서 사용되는 데이터셋은 줌마 영어 교재 데이터, 한국 AI Hub에서 제공하는 한국어-영어 말뭉치 데이터(상황 데이터)를 사용하였다. 각 데이터셋의 예시는 <그림5>과 <그림6>과 같다. 줌마 영어 교재의 각 챕터에 맞게 AI Hub 상황 데이터 속 상황에 맞게 데이터 매칭을 통해 ConvAI2 데이터셋에 맞게 변형한다. 또한, 줌마 영어 챕터와 상황 데이터의 상황이 <그림7> 와 같이 일대일 대응을 이루고 있어서 일대다 대응으로 변형하였다.



〈그림 5: 줌마 영어 교재 데이터〉

여행/쇼핑	공항	비행기 안	9332	A-1	비빔밥과 비프 중에, 식사는 어느 걸로 드릴까요?	Which one would you like for the meal, bibimbap or beef?
여행/쇼핑	공항	비행기 안	9332	B-1	저는 괜찮아요, 피곤해서 그냥 잘 거예요.	I am fine, I will just sleep now because I am tired.
여행/쇼핑	공항	비행기 안	9332	A-2	그러면 일어나신 후에 식사 주문 다시 받을까요?	Then shall I take your order again when you wake up?
여행/쇼핑	공항	비행기 안	9332	B-2	그래도 된다면, 요청 드릴게요. 감사합니다.	If that is fine, I will ask you, thank you.
여행/쇼핑	공항	비행기 안	9333	A-1	아기가 너무 울어서 다른 사람들이 불편하네요.	The baby is crying too much so other people are feeling uncomfortable.
여행/쇼핑	공항	비행기 안	9333	B-1	그리시면, 저희 승무원 공간에서 좀 달래보시겠어요?	Then, would you comfort the baby in our cabin space?
여행/쇼핑	공항	비행기 안	9333	A-2	네, 여기보다는 거기가 나을 것 같아요. 감사합니다.	Yes, it would be better there than here. Thank you.
여행/쇼핑	공항	비행기 안	9333	B-2	네, 제가 안내해드릴게요, 이쪽으로 오시면 돼요.	Okay, I will guide you there, follow me this way.
여행/쇼핑	공항	비행기 안	9334	A-1	기내에서도 면세점 제품 구매할 수 있는 거죠?	So I can purchase duty-free items on a plane too, right?
여행/쇼핑	공항	비행기 안	9334	B-1	네, 여기 책자에서 필요한 물품 말씀해주세요.	Yes, tell me what you want from this brochure.
여행/쇼핑	공항	비행기 안	9334	A-2	급한 건 아니고 구매하는 시간 있지 않나요?	I am not in a hurry and isn't there time for purchase?
여행/쇼핑	공항	비행기 안	9334	B-2	네, 우리가 저희가 카트로 제품을 보여드릴 거예요.	Yes, we will show you the items on a cart later.
여행/쇼핑	공항	비행기 안	9335	A-1	이 비행기에서는 식사 제공은 안되는 거 맞죠?	A meal is not provided on this flight, right?
여행/쇼핑	공항	비행기 안	9335	B-1	네, 저가 항공이라서 식사는 포함이 안됩니다.	Yes, because we are a LCC, the meal is not included.
여행/쇼핑	공항	비행기 안	9335	A-2	그럼 따로 돈 내고 먹을 수 있는 건 있나요?	Then can I pay and eat?
여행/쇼핑	공항	비행기 안	9335	B-2	여기 보시면, 돈 내고 드실 수 있는 메뉴들이 있어요.	If you see here, there are some menus you can pay and eat.

〈그림 6: 한국 AI Hub 한국어-영어 말뭉치〉

본 프로젝트 데이터셋 구현을 위해 사용된 줌마 영어 교재 데이터와 AI Hub 상황 데이터에 대한 요약 표는 다음과 같다.

〈줌마 영어 교재 데이터〉

항목	값
교재 챕터 수	19
대화 텐	12~14
생성 가능 Persona	19
생성 가능 History	11~13

〈AI Hub 상황 데이터〉

항목	값
상황 수	2,769
대화 수	100,000
최대 생성 Persona	2,769
최대 생성 History	25,000

### 3. 데이터셋 매칭

본 프로젝트에서는 줌마 영어 교재의 챕터를 하나의 Persona로 구성하였다. 챕터는 총 19개로 구성되어 있다. 이후, 줌마 영어 데이터와 상황 데이터 매칭은 <그림7>과 같다. Persona 생성을 위해 영어 교재의 챕터 데이터로 변환하였고, 다음 Utterance의 예측에 사용되는 Candidate 데이터셋 구성은 AI Hub 상황 데이터에서 문장 랜덤 추출하였다. History의 경우 다음 문장을 예측하기 위해 홀수개의 데이터로 존재해야 하므로 짝수개의 데이터를 모두 홀수개의 데이터로 변경하였다.

추가로, 중간발표 당시 Fine-tuning을 위해 대용량 말뭉치 데이터 필요하였지만 AI Hub 상황 데이터 사용이 어려워 줌마 영어 교재 데이터를 그대로 대화 데이터로 사용해야 했다. (총 19개) 때문에 중간 발표 당시 챗봇의 대화력은 좋지 못했다. 이후 위와 같이 데이터 매칭을 통해 총 1,718개의 상황을 훈련을 진행하여 더 나은 결과물을 도출해냈다.

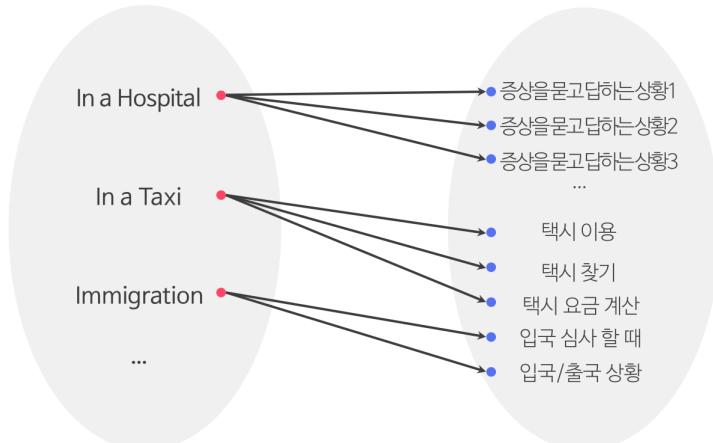
## 데이터셋 매칭

### Dataset Matching

- 줌마 영어 교재 챕터 당 여러 상황이 매치됨
- 일대일 관계 → 일대다 변형 필요

항목	값
교재 챕터 수	19
대화턴	12~14
생성 가능 Persona	19
생성 가능 History	11~13
〈줌마 영어 교재 데이터셋 요약〉	
항목	값
상황수	2,769
대화수	100,000
최대 생성 Persona	2,769
최대 생성 History	25,000
〈AI Hub 상황 데이터셋 요약〉	

### 줌마 영어 교재



〈그림 7: 각 챕터 별 상황 매칭〉

### 4. 데이터셋 증대

ConvAI2 데이터셋에서는 학습량을 증대(Augmentation)시키기 위해서 같은 Persona 라도 Persona Permutation을 통해 데이터 수를 증대 시켰다.(Thomas Wolf et al., 2019) Persona Permutation 방법의 예시는 <그림 8>과 같다.

## PERSONA 줌마영어 교재 'In a Hospital'

1. I'm not feeling well now.
2. I have got a headache and a fever.
3. I have been a headache 2 days.
4. Sometimes, I have a cough.



3. I have been a headache 2 days.
1. I'm not feeling well now.
4. Sometimes, I have a cough.
2. I have got a headache and a fever.

〈그림 8: Persona Permutation 예시〉

## 5. 텍스트 전처리

텍스트 전처리 과정은 다음 <그림 9>과 같다. 1) Expanding Contractions 작업을 통해 Contraction 을 제거한다. 2) Tokenization 작업을 통해 토큰화된 문장 성분을 리스트 형태로 출력한다. 3) Converting all Characters to Lowercase 작업을 통해 모든 문자를 소문자화 한다. 4) Lemmatization 작업을 통해 동사의 형태를 원형을 변형한다. 주목할 점은 마침표(.)도 단어로 인식한다는 점이다. 그 이유는 대부분의 NLP 학습 진행 시 이전의 단어를 통해 다음 단어를 예측한다. 마침표(.)를 사용하여 문자의 끝을 예측 할 수 있어 문장 생성 시 마침표를 구별하여 전처리 작업을 진행하였다.

위 작업을 통하여 최종적으로 본 프로젝트에 데이터셋을 구축 하였다. 이후, 모델 GPT-2 Pre-trained 모델을 사용하여 학습을 진행하였다.

### 1. Expanding Contractions

I'm not feeling well now. > I am not feeling well now.  
I'll be staying for three days. > I will be staying for three days.

### 2. Tokenization

I am not feeling well now. > ['I', 'am', 'not', 'feeling', 'well', 'now', '.']  
I will be staying for three days. > ['I', 'will', 'be', 'staying', 'for', 'three', 'days', '.']

### 3. Converting all Characters to Lowercase

['I', 'am', 'not', 'feeling', 'well', 'now', '.'] > ['i', 'am', 'not', 'feeling', 'well', 'now', '.']  
['I', 'will', 'be', 'staying', 'for', 'three', 'days', '.'] > ['i', 'will', 'be', 'staying', 'for', 'three', 'days', '.']

### 4. Lemmatization

['i', 'am', 'not', 'feeling', 'well', 'now', '.'] > ['i', 'be', 'not', 'feel', 'well', 'now', '.']  
['i', 'will', 'be', 'staying', 'for', 'three', 'days', '.'] > ['i', 'will', 'be', 'stay', 'for', 'three', 'day', '.']

〈그림 9: 텍스트 전처리 과정〉

## 6. 파라미터 최적화

위와 같이 데이터셋 제작 후, 총 30회의 Epoch 훈련 후 다음 <그림 10>과 같이 최적 하이퍼 파라미터를 도출 했다.

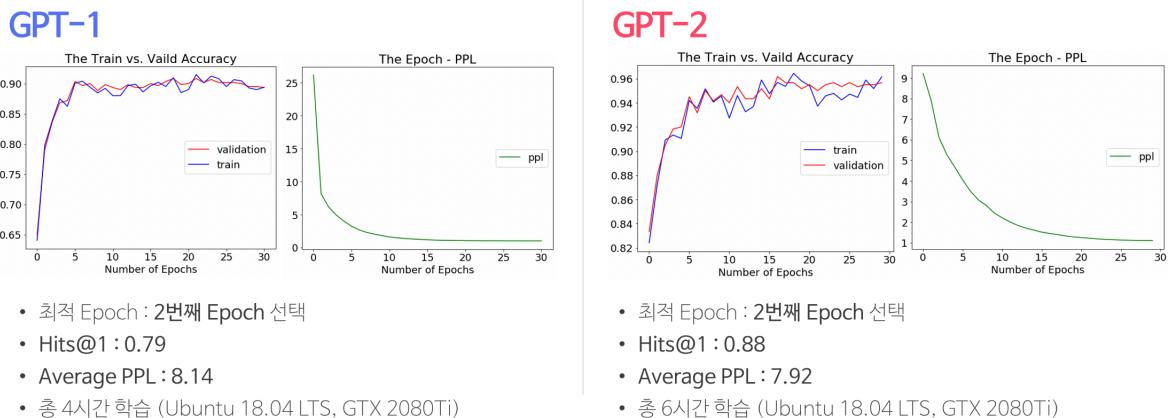
Argument	Default value	Modified Value	Description
Model	Open AI GPT	GPT2	Open AI에서 개발한 GPT, GPT2 모델
Num_candidates	2	6	Next Utterance의 후보군 선택지 수
Max_history	4	2	최대 History 수
Gradient_accumulation_steps	8	4	Optimize를 진행 시 GPU의 메모리 문제를 해결하기 위해 사용
Epochs	1	30	Epoch 수
Train_batch_size	4	2	1회 Epoch시 Batch 크기(훈련 시)
Valid_batch_size	4	2	1회 Epoch시 Batch 크기(검증 시)

<그림 10: 최적 하이퍼 파라미터>

## 7. 모델 선택

최적 하이퍼 파라미터 설정 후, Open AI의 GPT-1 모델과 GPT-2 모델의 Hits@1, PPL 수치를 비교하여 본 프로젝트에 가장 적절한 모델을 선정한다. 각 모델별 최종 학습 결과는 <그림 11>과 같다.

Hits 스코어와 perplexity 스코어는 트레이드 오프 관계이다. 그러므로 두 모델 모두 다양한 Epoch에서 실제 테스트 후, 가장 대화를 잘한 Epoch를 선정하였다. Epoch 선정 후, GPT-2 모델이 hits 스코어와 perplexity 스코어 모두 좋은 수치를 보여 주고 있어 GPT-2 모델을 선정 최종 대화형 Ai 모델로 사용하였다.



<그림 11: GPT-1(좌) vs. GPT-2(우) 학습 결과>

## 8. 디코더 세팅

위와 같이 최적 모델을 선정 후, 디코더 설정을 진행하였다. 디코더(Decoder)란 Encoding 이후 부호화된 값을 다시 다음 단어를 예측해가며 문장 생성 후, 다시 문자로 복호화하는 과정을 말한다. 디코더 세팅은 기존에 Hugging Face에서 설정해 놓은 디코더 세팅을 여러 차례 테스트를 후, Medium 값을 사용하였다. Hugging Face의 디코더 설정은 <그림12>와 같다.

항목	Low	Medium	High
Top-k	180	70	0
Top-p	0.1	0.5	0.9
Temperature	1.9	1.2	0.6

<그림 12: Hugging Face의 디코더 설정>

## 제 4절. Assessment For Learning

수강생의 학습 성취도를 평가하기 위한 수단으로 AFL 모델을 제작한다. AFL 모델을 제작하기 위해 Google 사에서 제공하는 B.E.R.T. Pre-trained model을 사용한다. Model의 Fine-tuning 과정 시 GLUE 사에서 제공하는 CoLA 데이터와 MRPC 데이터를 사용하여 학습을 진행한다. CoLA 데이터를 바탕으로 한가지 문장을 입력받아 문법적으로 허용 가능한지 이진 분류하는 모델을 제작하여 수강생의 문법 허용 여부를 판단한다. MRPC 데이터를 바탕으로 두 가지 문장을 입력받아 문장 간의 유사도를 이진 분류를 통해 확인할 수 있는 모델을 만든다. 최종적으로 구현된 AFL 모델은 수강생이 챗봇과 대화를 나눈 후, 실제 교재 내용 기반으로 대화하였는지 MRPC 데이터로 학습된 모델로 판단한다. 또한, 문법적으로 오류가 없는지 확인하는 과정을 CoLA데이터로 판단한다. 이를 통해 수강생이 자신의 학습 성취도를 파악할 수 있다. 또한 중간 때와는 다르게 수강생들의 Acceptance를 늘리기 위해 틀린 문장에 대해서 고쳐주는 기능을 Bing Spell Check API를 사용하여 추가하였다.

### AFL 사용목적

AFL Propose

- MRPC & CoLA 데이터셋을 통해 만들어진 모델로 퍼센트 성적으로 수강생에게 평가를 내어 줌
- 퍼센트 결과로는 수강생의 Acceptance를 높이는데 한계점이 존재
- 수강생의 Acceptance를 높이기 위해 틀린 문장의 Spelling Check 기능을 추가



<그림 13: AFL 사용목적 & 각 데이터셋>

## 제4장. 추진 결과물

### 제 1절. 앱 개발 및 모듈 훈련 시간

#### 1. 어플리케이션 개발

최종적으로, 대화형 Ai 모델과 AFL 모델을 결합하여 최종 모델을 완성하였다. 이후, Python 프레임워크 Flask를 사용하여 REST API 형태로 구현하였다. 앱 개발은 Google에서 개발한 Hybrid 애플리케이션 개발 프레임워크인 Flutter를 사용하였다. 해당 언어를 사용하여 하나의 코드로 Android와 iOS 애플리케이션을 동시에 제작하였다.

#### 2. 각 모델 훈련 시간

각 모델 훈련 시간은 <그림14>와 같다.

##### 줌마영어 + Ai Hub 상황 데이터

GPU : GTX 2080Ti

서버환경: Ubuntu 18.04 LTS (SSH사용)

소요시간: 총 6시간



##### CoLA

GPU : Tesla V100

서버환경: Google Colab

소요시간: 총 3시간



##### MRPC

GPU : Tesla V100

서버환경: Google Colab

소요시간: 총 3시간

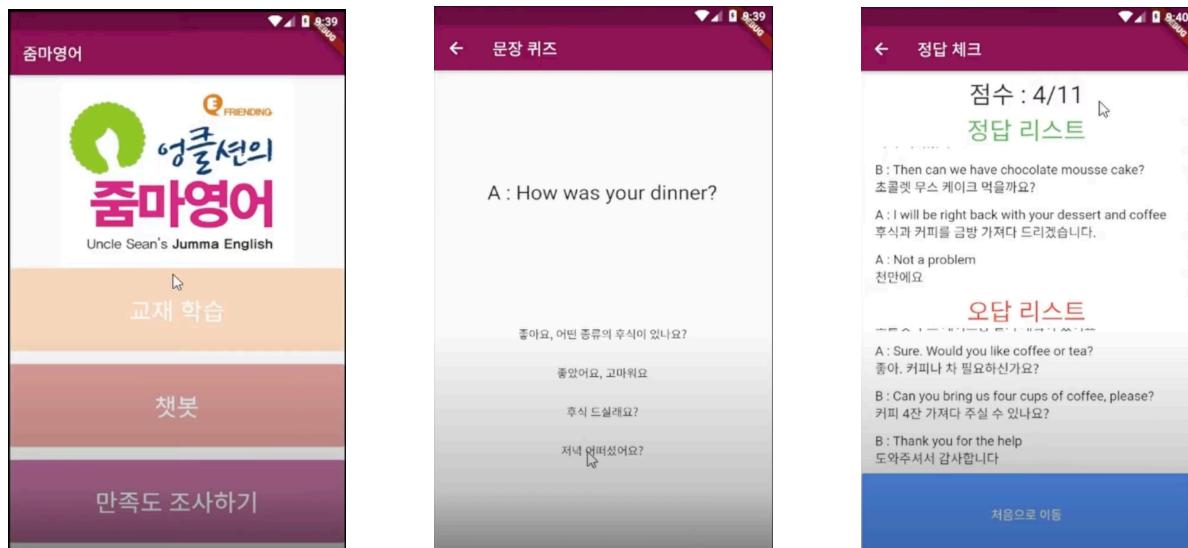


<그림 14: 각 모델 훈련 시간>

### 3. 모델 평가 척도

평가 사항	평가 척도	구체적인 지표															
"AI가 컨텐츠의 내용을 잘 반영하는가"	F-1 score	$F_1 = \left( \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} \right) = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$															
	Hits at k	<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Actual</th> </tr> <tr> <th colspan="2"></th> <th>Positive</th> <th>Negative</th> </tr> <tr> <th rowspan="2">Predicted</th> <th>Positive</th> <td>True Positive</td> <td>False Positive</td> </tr> </thead> <tbody> <tr> <th>Negative</th> <td>False Negative</td> <td>True Negative</td> </tr> </tbody> </table> <p>Accuracy = <math>(TP+TN)/(TP+FP+FN+TN)</math></p>			Actual				Positive	Negative	Predicted	Positive	True Positive	False Positive	Negative	False Negative	True Negative
		Actual															
		Positive	Negative														
Predicted	Positive	True Positive	False Positive														
	Negative	False Negative	True Negative														
"AI가 실제로 문법적으로 올바르게 응답을 잘 하는지"	Perplexity	$PPL(W) = P(w_1, w_2, w_3, \dots, w_N)^{-\frac{1}{N}} = \sqrt[N]{P(w_1, w_2, w_3, \dots, w_N)}$															

### 제 2절. 어플리케이션 구현

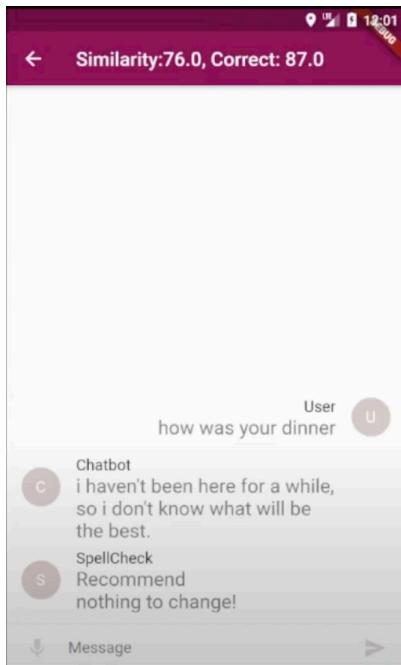


<그림 15: 어플 실행시 메인화면>

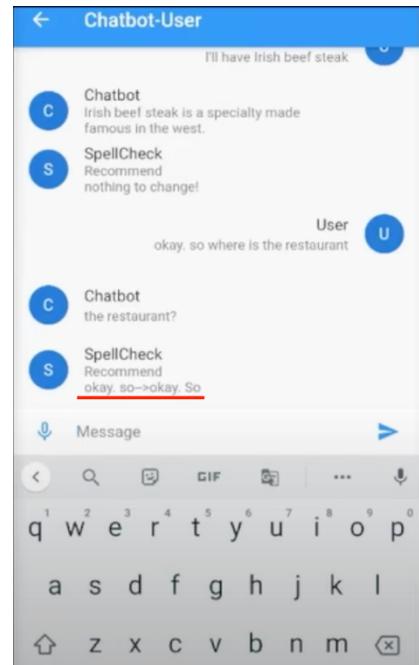
<그림 16: 줌마영어 교재를 활용한 문제 풀이>

본 프로젝트에서는 줌마 영어 수강생이 챗봇을 통해 상황을 기반하여 대화를 기능과 더불어, 줌마 영어 교재를 활용한 교재 학습 기능도 구현하였다. 애플리케이션 구동 시 메인화면은 <그림15>와 같다. 교재 학습 진행 시, 교재를 기반으로 문장 퀴즈와 채점 기능을 제공하여<그림16> 학습 내용을 보다 효율적으로 학습 할 수 있도록 도왔다. 추가로, 챗봇 기능에서는 음성인식을 추가하여 키보드 타이핑에

익숙지 못한 줌마 회원에게 편의성을 제공하였다. 주요 화면을 <그림17>과 같다. 또한, 상단의 AppBar에서는 사용자가 해당 콘텐츠를 기반으로 잘 말하고 있는지 확인하는 유사도와 문법의 정확도를 체크가 가능하다.



<그림 17: 챗봇 대화 모습 >



<그림 18: 문법 체크 기능>

#### 4. 기대 효과

기존의 Hugging Face의 대화형 Ai의 경우 Persona가 무작위로 선택되었고, 대화 데이터셋이 자유 대화를 기반으로 제작되어 교육적 활용이 어려웠다. 본 프로젝트에서는 데이터셋 변형을 통한 Fine-tuning과 AFL 기능을 추가하여 줌마 영어 회원들을 대상으로 상황에 따라 영어 회화가 가능한 챗봇을 제작하였다. 또한, 아직 영어 학습이 미숙한 수강생이 대부분이기 때문에 콘텐츠 내용을 복습할 수 있는 교재 문제 풀이나, 키보드 타이핑이 익숙지 않는 대부분의 수강생을 고려하여 음성 인식 기능을 추가하였다.

이러한 결과물을 통해 평소 한정된 시간과 장소로 인해 영어 회화 학습이 어려운 줌마 영어 회원들 뿐만 아니라 영어 회화 학습을 배우고자 하는 다수 사용자에게도 회화 교육에 도움이 되리라 생각한다.

## 제 5장. 향후 과제 및 최종 결과물 활용계획

### 1. 향후 과제

현재 구현된 프로젝트 데모에서는 개인화된 사용자 기능이 존재하지 않는다. 향후 과제로 회원 관리 기능을 추가하여 회원별로 학습량과 진도율을 관리하여 더욱 개인화된 교육을 제공하고자 한다. 또한, 최종 발표 시, 프로젝트 데모 영상에서 일부분 100% 완벽한 대답을 못 하는 경우가 있었다. 이는 Fine-tuning 진행 시, 아직 19개의 줌마 영어 챕터만 수집된 상태여서 제한된 상황 데이터만 사용해야 했다. 최종 발표 이후, 줌마 영어에서는 총 3권의 회화 학습 교재를 추가로 발매하였다. 이는 추가적인 상황 매칭이 가능함을 의미한다. 추가적인 교재를 활용하여 Fine-tuning 진행 시에 사용되었던 상황 수를 증대 시켜 더 나은 학습을 진행할 계획이다.

최종적으로, 보안 된 내용을 바탕으로 정식으로 줌마영어 회원들을 대상으로 테스팅을 진행할 예정이다.

### 2. 프로젝트 활용 방안

향후 과제 내용들을 보안하여 추후에 있을 스타트업 공모전(도전! K-스타트업, 디지털콘텐츠육성 사업 등)에 참가 할 예정이다. 또한 추계에 진행 될 산업공학회 대학생 공모전 분야에도 도전 할 계획이다.

## 제 7장. 참고문헌 정리

- 1) Theo Hug. (2005), Microlearning: A New Pedagogical Challenge
- 2) 김혜영, 신통광, 양혜진, 이장호. (2019), 영어교과 보조 도구로서의 AI 챗봇 분석 연구
- 3) 임규연. (2006), 평생교육환경에서 인지적흥미와 학습자역할몰입의 조절효과: 자기결정성요인과 학습지속의지 간 관계를 중심으로
- 4) 정상목. (2007), 이러닝 환경에서 몰입학습 증진을 위한 대화 기반 피드백 시스템의 개발
- 5) 이소영. (2019), 이러닝 환경에서 몰입에 영향을 미치는 요인 연구 -상호작용 요인과 어포던스 요인을 중심으로
- 6) 장은지. (2012), 학습자의 교육용 어플리케이션 활용 만족요인과 구매의도에 영향을 미치는 요인 연구
- 7) Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin (NIPS 2017), Attention is All you Need
- 8) MatthewE.Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettlemoyer. (2018), Deep contextualized word representations
- 9) Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever. (2018), Improving Language Understanding by Generative Pre-Training
- 10) Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. (2018), BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
- 11) Daniel Adiwardana, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, QuocV. Le. (2020), Towards a Human-like Open-Domain Chatbot
- 12) Thomas Wolf et al. (2020), How to build a State-of-the-Art Conversational Ai with Transfer Learning, from : <https://medium.com/huggingface/how-to-build-a-state-of-the-art-conversational-ai-with-transfer-learning-2d818ac26313>
- 13) D Yu et al.(2019), Gunrock : A Social Bot for Complex and Engaging Long source
- 14) Wei Wu, Rui Yan. (2019), Deep Chit-Chat: Deep Learning for Chatbots
- 15) Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, Jason Weston. (ACL2018), Personalizing Dialogue Agents: I have a dog, do you have pets too?,
- 16) Pierre-Emmanuel Mazaré, Samuel Humeau, Martin Raison, Antoine Bordes. (2018) Training Millions of Personalized Dialogue Agents
- 17) Shiffrin, R. M., & Atkinson, R. C. (1969), Storage and retrieval processes in long-term memory
- 18) Thomas Wolf et al., (2019): TransferTransfo: A Transfer Learning Approach for Neural Network Based Conversational Agents
- 19) Zhang, S.; Dinan, E.; Urbanek, J.; Szlam, A.; Kiela, D.; and Weston, J. 2018. Personalizing Dialogue Agents: I have a dog, do you have pets too?

## 제 8장. 프로젝트 수행후기

이승준 - 먼저 부족한 실력임에도 불구하고, 본 프로젝트의 PM으로 역할을 수행하며 같이 프로젝트를 수행해준 팀원들에게 감사의 인사를 전하고 싶다. 팀 빌딩을 하는 데 있어 다른 팀들에 비해 여러 수행 착오가 많았는데 끝까지 프로젝트를 진행해 준 것에 대해 감사함을 표한다.

졸업 프로젝트가 끝날 때쯤 지난 4년간의 대학생 생활을 되돌아볼 기회가 많았다. 예전부터 졸업 프로젝트를 통해 4년간 쌓아 올린 실력을 바탕으로 다른 학우들과 경쟁을 통해 실력을 증명하고 싶어 했지만, 코로나 사태로 그러지 못 한 것에 대한 아쉬움이 따른다. 이러한 욕심 때문인지 처음부터 제대로 해보자는 생각으로 주제 선정과 구체화에 많은 시간을 할애하였다. 그로 인해 프로젝트 구현 시간이 많이 부족 하였고, 구현되지 못한 부분에 대해서는 많은 아쉬움이 남는다. 학과에서 교육받지 못한 딥러닝이라는 분야를 이렇게 밑바닥부터 쌓아 올리기란 쉽지 않았다. 그 때문에 프로젝트 회의 이후 매일 같이 해당 분야를 공부해 가며 관련 지식을 쌓아야만 했다. 그러한 노력이 좋은 결과물로 나타난 것 같아 기쁜 마음으로 프로젝트를 마무리 할 수 있었던 것 같다. 마지막으로, 프로젝트를 진행하는 데 있어 학과 내 재정 지원과 프로젝트 진행 도와주신 지도 교수 및 학과 교수님들에게 감사의 인사를 전한다.

김민우 - 다사다난 했던 졸업프로젝트가 막상 끝나고 나니 후련함보다 아쉬움이 더욱 큰 것 같다. 나는 처음부터 졸업 프로젝트를 단지 졸업을 위한 조건으로 보지 않고 정말 하고 싶었던 프로젝트였기 때문에 욕심이 더 있었던 것 같다. 하지만 마음과는 다르게 이번 프로젝트 예상하지 못했던 어려움이 많았다. 코로나 때문에 많은 팀이 와해하고 우리 팀 역시 코로나 영향을 많이 받았다. 또한 초반에 우리 팀은 프로젝트 주제 선정을 하는 단계에서 생각보다 많은 시간을 썼다. 욕심이 있었기에 더욱 신중했던 것 같다. 비록 우리 팀은 구현하는 과정에 있어서 늦게 시작하게 되었지만, 다들 자신이 맡은 부분에 대해 확신이 있었고 나 또한 내가 맡은 파트에 대해 흥미가 높았기 때문에 빠르게 구현을 할 수 있었다. 또한 계획했던 일들을 순차적으로 완성해 나가는 데 있어서 높은 성취감을 느꼈다. 비록 아쉽게 졸업 프로젝트는 끝났지만, 우리 팀에게 주어진 향후 과제를 잘 마무리 하여 좋은 또 다른 좋은 결과를 내고 싶다.

김성진 - 코로나 19로 인해 이전 팀이 사라져서 중간에 새로운 팀에 들어오게 되었다. 졸업 프로젝트로 무언가 남기고 싶다는 생각이 컸고, 학과에서도 많이 다루지 않았던 새로운 분야에 도전하게 되었다.

처음 팀에 들어왔을 때, 익숙하지 않은 분야에 도전하는 게 매우 어려운 과정이었다. 많은 부분을 프로젝트를 위해 새롭게 공부해야 했다. 이 과정을 통해 나 자신이 새로운 것을 배울 힘을 기를 수 있었다. 도중에 주제에 관해서도 몇몇 변경사항과 어려움이 생겼다. 하지만 팀원들과 회의를 통해 짧은 시간 동안 좋은 결과를 내기 위해 노력했다. 데이터를 수집하는 것도 빅데이터 AI Hub에 데이터를 직접 요청해보기도 하고, 한 번도 애플리케이션을 구현해본 적 없던 내가 애플리케이션 제작에 흥미를 느끼게 되었고, 새롭게 도전하는 것의 두려움도 많이 없앨 수 있었다. 앞으로 사회에 나가는 데도 많은 도움이 될 것으로 생각한다. 앞으로 다들 마무리를 잘해서 자신이 원하는 성공을 이룰 수 있길 기대한다.