

[소프트웨어 종합설계]

페르소나를 반영한 영어 교육 챗봇

Conai

문승준 이해지 목요한

담당 교수님: 여진영 교수님

Convei Lab

목차

- I. 연구 주제
- II. 선행 연구와 연구 필요성
 - A. 연구 필요성
 - B. 선행 연구
 - C. 기대 결과
- III. 연구 내용
 - A. Chatbot model
 - B. Error correction model
 - C. Word extraction model
 - D. Topic & Persona Injection
- IV. 연구 결과
 - A. Utterance Prediction Example
 - B. Error Correction Example
 - C. Entire Dialogue Example
- V. 일정 및 역할 배분
- VI. 참고 문헌

I. 연구 주제

최근 챗봇의 일관성 및 실용성을 중심으로 한 많은 연구가 이루어졌다. 이에 따라 Transformer 모델을 기반으로 한 챗봇 모델은 특정 인격을 기반으로 소소한 일상에 대한 대화를 하기도하고, Wikipedia 와 같은 Knowledge base 를 기반으로 지적인 대화를 하기도 하며, 때로는 사람의 감정에 대입하여 공감 능력을 보이기까지 한다. 더불어 2020 년 Facebook AI Research 팀은 위 기능을 포함한 총 12 가지 기능을 모두 보유한 챗봇을 성공적으로 훈련시킨 연구 결과를 발표했다.

본 연구팀은 이러한 선행 연구를 적극 응용하여 영어 교육 서비스를 개발하는 것을 목표로 한다. 이는 영어 학습을 보조하는 서비스로 인공지능 기반 영어 문법 오류 교정과 대화가 끊기지 않도록 대화 주제를 생성하는 기능을 가진다. 이에 따라 해당 프로젝트는 다음 3 가지 모델로 구성된다.

- Dodecadiologue fine-tuned on Wizard of Wikipedia
- Transformer based Error Correction Model
- Word Extraction Model

최종적인 프로젝트 목표는 이 모델을 토대로 데모 서비스를 개발하는 것이다.

II. 선행연구와 연구 필요성

A. 연구 필요성

영어 교육 시장에는 인공 지능을 활용한 여러 서비스가 출시되었다. 대표적으로 EBS 에서 선보인 AI 팽톡과 야나두에서 출시한 오디오 영어가 있다. AI 팽톡은 최근 인기를 끌고 있는 팽수라는 캐릭터와 아이들이 게임 형식으로 대화할 수 있는 AI 서비스이고, 야나두의 오디오 영어는 강의 위주의 서비스로 기타 영상 강의 등과

연계하여 복습 및 발음 교정을 위해 사용된다. 각각 텍스트가 아닌 음성 데이터를 기반으로 하다보니, 단어 인식과 관련한 미흡하다는 의견도 있으나 그 외에는 적지 않은 기대를 받고 있는 서비스이다.

위 상용 서비스와 Conai 팀의 프로젝트는 영어 교육 방식에 있어 차이점을 갖는다. 앞선 서비스는 정해진 틀 안에서, 기계가 특정 목적을 가진 상태로 대화가 진행되나, 해당 프로젝트는 사용자에게 대화 주제를 제안할 뿐, 최대한 사용자가 어떠한 주제로도 대화할 수 있도록 하는 것을 목표로한다. 이는 최근 Dodecadialogue 모델과 같이 사람의 특성을 모방할 수 있는 기술이 연구되었기 때문에 가능해졌으며, 해당 연구팀은 이를 적극 응용하여 차별화를 할 예정이다.

더불어 이렇듯 대화 주제에 제한을 두지 않기에, 오류 수정 모델에도 보다 많은 노력이 들어간다. 기본적인 틀을 갖는 대화 모델의 경우, 기계와 상대방의 발화 범위가 한정되어 있으나, 해당 모델은 그 어떤 대화도 가능하기에 그만큼 넓은 범위의 오류를 감지할 수 있어야한다. 이에 따라 공개된 데이터셋을 기반으로 오류 수정 모델을 만들되, 필요 시 에러 생성모델(Error Generation Model) 또한 사용할 예정이며 이는 해당 프로젝트의 또 다른 차별점이 된다.

B. 선행 연구

1) Persona Conditioning

- Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, Jason Weston(2018), "**Personalizing Dialogue Agents: I have a dog, do you have pets too?**", Association for Computational Linguistics, p2207-2213.

: 기존 챗봇의 비밀관성, 기억장치의 부재, 모호한 답변 등의 문제를 해결하기 위해 persona conditioning 제안하였다. Memory model 을 이용하여 영속적인 persona 를 반영하고자 하였다.

- Pierre-Emmanuel Mazare, Samuel Humeau, Marrrtin Raison, Antoine Bordes(2018), "**Training Millions of Personalized Dialogue Agents**", Association for Computational Linguistics, p2775-2779.

: 온라인 대화 데이터(Reddit 등)에서 500 만개의 페르소나를 추출하였으며 context persona 로 모델에 입력시켰다. Transformer encoder 로 적용했을 때 성능이 가장 좋았으며 Persona-Chat Dataset 으로 fine-tuning 하였을 때 정확도가 더 좋았다.

2) Error Generation

- Markek Rei, Mariano Felice, Zheng Yuan, Ted Biscoe(2017), "**Artificial Error Generation with Machine Translation and Syntactic Patterns**", Association for Computational Linguistics, p287-292.

: 정확한 문장에서 부정확한 문장을 생성하는 모델을 훈련시키는 SMT(Statistical Machine Translation)을 제안하였다. 에러 생성을 위해 machine translation, pattern extraction, error detection 과정을 거친다.

3) Retrieval Model

- Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, Jason Weston(2019), "**Wizard of Wikipedia : Knowledge-powered Conversational Agents**", ICLR.

: 기존에 knowledge 사용에 대한 연구가 없었고 이전에는 데이터베이스에 접근하거나 API 형태의 knowledge 에만 의존했던 것을 해결하기 위해 Unstructured dialogue dataset 에서 open-domain setting(Wikipedia)으로 knowledge 를 포함한 multi-turn dialogue 를 수행하는 모델을 제안하였다. Transformer와 Memory Network를 사용하여 knowledge source 구조를 제안하고 utterance prediction 을 수행하였다.

- Hannah Rashkin, Eric Michael Smith, Margaret Li, Y-Lan Boureau(2019), **"Towards Empathetic Open-domain Conversation Models: a New Benchmark and Dataset"**, Association for Computational Linguistics p5370-5381.

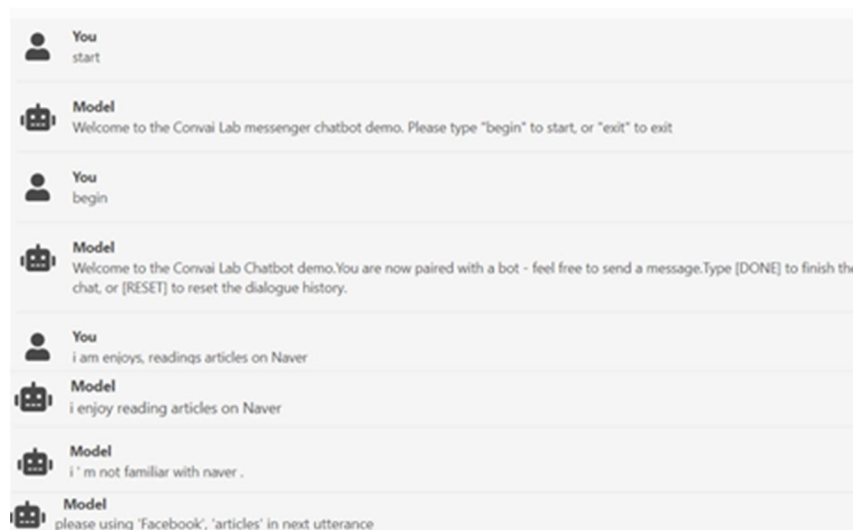
: 상대방의 감정을 고려하는 대화 세트인 Empathetic Dialogues(ED)를 제안하였다. BERT 모델의 경거의 사람과 유사한 성능을 보였고 Retrieval based model 은 detection 에서 좋은 정확도를 보였다.

C. 기대 결과

영어 교육용 챗봇이 사용자의 페르소나를 반영한 표현을 예측하여 중요 단어와 해석을 같이 제공한다. 사용자가 답변에 오류가 있을 경우 오류를 교정해준다. 또한 챗봇이 실제 사람처럼 최적화된 표현을 생성하고 감정적인 반응도 할 수 있도록 한다.

III. 연구 내용

목표하는 모델은 총 3 가지이다.



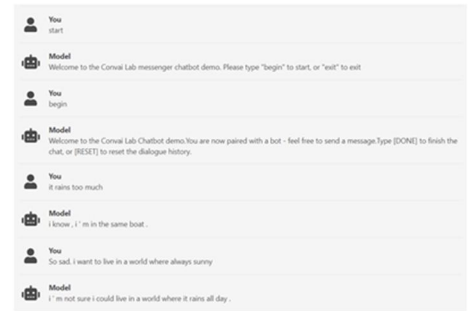
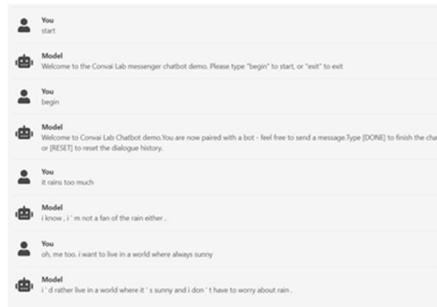
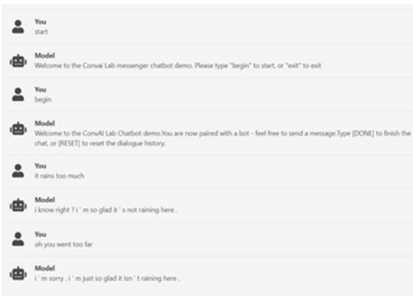
A. ChatBot model

사용자와 대화를 주고받는 챗봇 모델이다. 이전 시점까지의 context $C_n = \{c_1, \dots, c_n\}$ 에 대해 다음의 모델 f 를 찾는 것을 목표로 한다.

$$f(C_n; \theta_{CHAT}) = \underset{w_t \in V}{argmax} P(w_t | C_n, W_{t-1}; \theta_{CHAT})$$

($W_{(t-1)}$ 은 직전까지 생성된 단어를 의미한다. $\rightarrow W_{(t-1)} = \{w_1, \dots, w_{(t-1)}\}$)

(V 는 전체 vocab list를 의미한다. $\rightarrow V = \{v_1, \dots, v_{(total_vocab_size)}\}$)



ParlAI를 기반으로 하여 ImageSeq2Seq Dodecadialogue fine-tuned on convai2, ImageSeq2Seq Dodecadialogue with no fine-tuning, ImageSeq2Seq Dodecadialogue fine-tuned on wizard of wikipedia의 3가지 ChatBot model을 구현 완료하였다. 사용자의 발화에 응답을 하는 챗봇은 풍부한 지식을 가지고 응답하는 것이 좋다고 판단하여 wikipedia로 fine-tuning한 ImageSeq2Seq Dodecadialogue fine-tuned on wizard of wikipedia를 사용하였다.

B. Error correction model

사용자의 입력의 오류를 수정해주는 모델이다. 사용자에게 의해 입력된 utterance U_n 에 대해 다음의 모델 g 를 찾는 것을 목표로 한다.

$$g(U, \theta_{correction}) = \underset{w_t \in V}{argmax} P(w_t | U_n, W_{t-1}; \theta_{correction})$$

(W_{t-1})은 직전까지 생성된 단어를 의미한다. -> $W_{t-1} = \{w_1, \dots, w_{t-1}\}$)

(V 는 전체 vocab list 를 의미한다. -> $V = \{v_1, \dots, v_{(total_vocab_size)}\}$)

1) DataSet: JFLEG Dataset

GEC(Grammatical Error Correction) 문제에서 좋은 성능을 보여주고 있는 JFLEG Dataset 을 베이스라인으로 사용하였다. 각각의 instance 는 하나의 source sentence 와 4 개의 correction 이 존재한다.

```
{
  'sentence': "They are moved by solar energy ."
  'corrections': [
    "They are moving by solar energy .",
    "They are moved by solar energy .",
    "They are moved by solar energy .",
    "They are propelled by solar energy ."
  ]
}
```

2) Error correction model: T5, Bart, bert encoder-decoder model

Huggingface 를 기반으로 하여 총 3 개의 Error correction model 을 개발하였다. JFLEG Dataset 을 활용한 Validation loss 는 다음과 같다.

100% ██████████ 597/597 [00:25<00:00, 23.01it/s]
Validation Loss: 1.51

100% ██████████ 597/597 [00:15<00:00, 39.59it/s]
Validation Loss: 0.61

██ [598/598 00:25]
Validation loss: 0.46045927834510803

bert encoder-decoder, Bart, T5 모델 Validation Loss

Error correction model 을 Chatbot 에 병합하기 위해 다음과 같이 모듈화를 진행하였다.

```
(seungjun) → seungjun python ./error_correction/bart_error_correction.py
this is Bart error correction model. if you want to quit, enter quit
Enter message: █
```


C. Word extraction model

사용자에게 사용할 단어를 제공해주는 모델이다. ChatBot model 에 의해 생성된 context 를 포함하는, 즉 현재 시점까지의 $context = C_{(n+1)} = \{c_1, \dots, c_{(n+1)}\}$ 에 대해 chatbot model f 를 이용하여 사용자의 다음 발화를 예측한다. 이 때 f 는 사용자의 페르소나와 유사한 chatbot model 이다. $U_{(n+1)} = \{w_1, \dots, w_k\}$ Pre-trained 된 POS(part-of-speech) tagging model 을 이용하여, NN(noun) 혹은 VB(verb)를 추출하여 사용자에게 제공한다.

D. Topic & Persona Injection

대화 주제의 단조로움을 피하기 위해 챗봇이 시작되기 전에 대화 주제와 챗봇의 페르소나를 제시한다.

```
ParlAI > parlai > convai > ≡ persona.txt
1 i read twenty books a year.
2 i am a stunt double as my second job.
3 i only eat kosher.
4 i was raised in a single parent household.
5 i like to remodel homes.
6 i like to go hunting.
7 i like to shoot a bow.
8 my favorite holiday is halloween.
9 i like to remodel homes.
10 i like to go hunting.
```

Persona Example

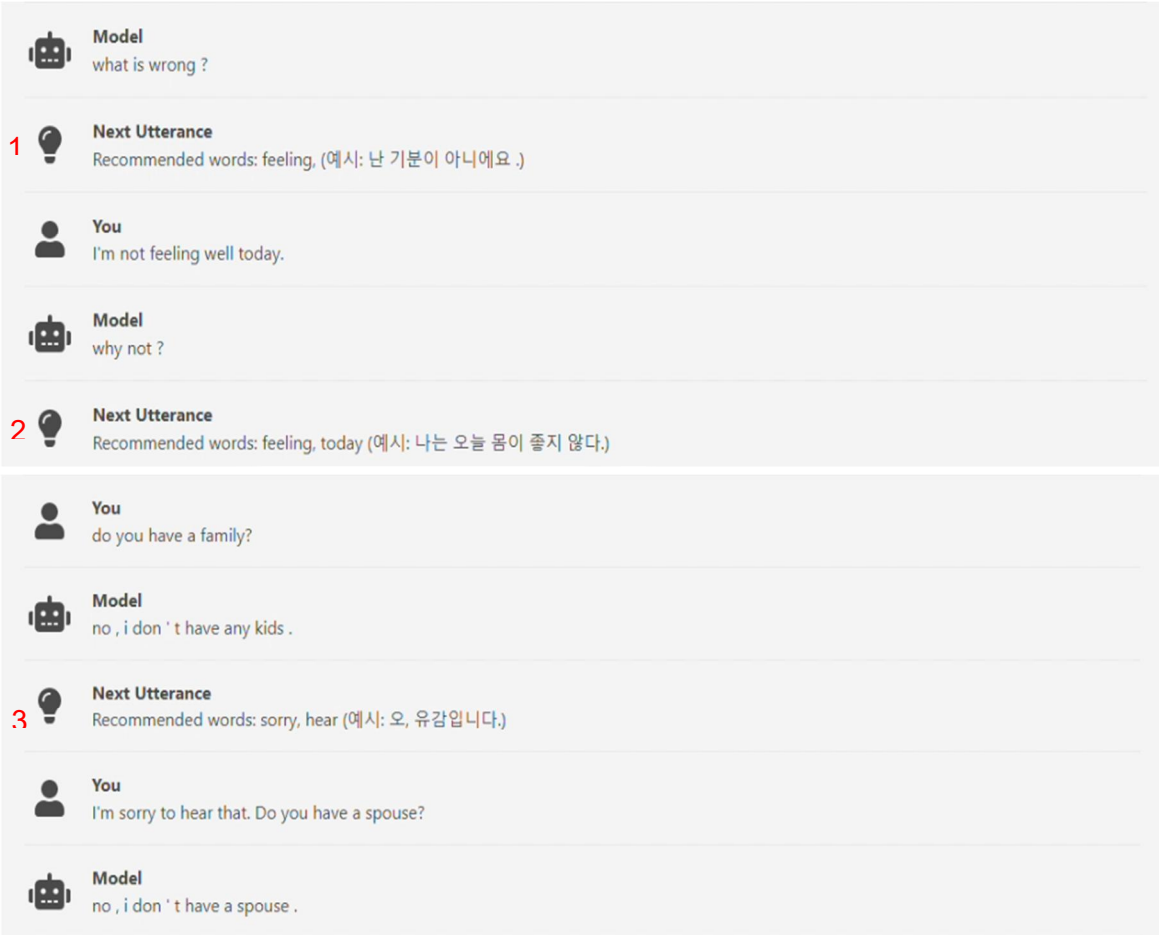
```
ParlAI > parlai > convai > ≡ topic.txt
1 Science fiction
2 Internet access
3 politics
4 Blue
5 Cinematography
6 Gardening
7 tourism
8 Bob Ross
9 finance
10 attitude_and_emotion
```

Topic Example

IV. 연구 결과

기존 기술과는 다르게 발화 예측, 추천 단어와 해석 제공, 토픽과 페르소나 반영, 오류 수정 기능이 모두 반영된 모델이다.

A. Utterance Prediction Example

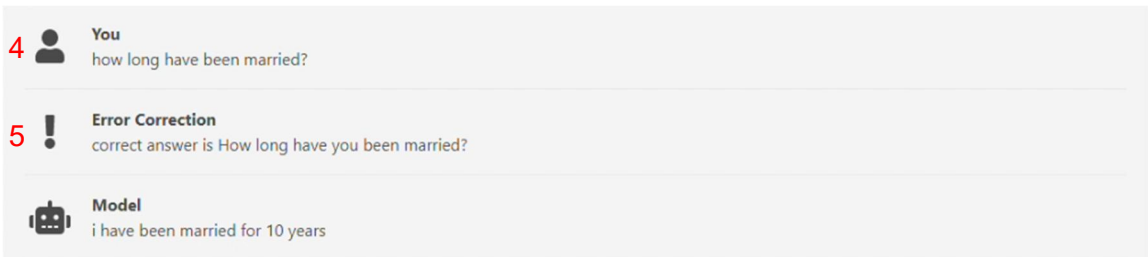


The image shows a chat interface with three examples of utterance prediction. Each example consists of a user input, a model prediction, and a recommended word list.

- Example 1:**
 - Model:** what is wrong ?
 - Next Utterance:** Recommended words: feeling, (예시: 난 기분이 아니예요.)
 - You:** I'm not feeling well today.
- Example 2:**
 - Model:** why not ?
 - Next Utterance:** Recommended words: feeling, today (예시: 나는 오늘 몸이 좋지 않다.)
 - You:** do you have a family?
- Example 3:**
 - Model:** no , i don ' t have any kids .
 - Next Utterance:** Recommended words: sorry, hear (예시: 오, 유감입니다.)
 - You:** I'm sorry to hear that. Do you have a spouse?
 - Model:** no , i don ' t have a spouse .

챗봇이 발화를 하면 사용자의 응답을 예측한다. 예측한 응답에서 word extraction 을 하여 사용자가 사용할 단어를 추출하여 해석과 함께 제공한다. (1, 2, 3 번 문장)

B. Error Correction Example



The image shows a chat interface with an error correction example. It consists of a user input, an error correction message, and a model prediction.

- You:** how long have been married?
- Error Correction:** correct answer is How long have you been married?
- Model:** i have been married for 10 years

4 번 문장에 문법적 오류가 있어서 5 번 문장으로 올바르게 오류를 교정해주었다.

C. Entire Dialogue Example

The screenshot displays a chatbot interface with a light gray background and white message bubbles. The conversation is as follows:




















- You:** begin
- Model:** Welcome to the Convai Lab messenger chatbot demo. Please type "begin" to start, or "exit" to exit
- You:** begin
- Model:** Welcome to Convai Lab Chatbot demo.You are now paired with a bot - feel free to send a message.Type [DONE] to finish the chat, or [RESET] to reset the dialogue history.
- 1 Model:** Topic is 'Bob Ross'
- 2 Bot's persona:** Bot persona is 1.i love watching the leaves change colors. 2.i just bought a brand new house. 3.i have a son who is in junior high school. 4.i spent a decade working in the human services field.
- You:** hi
- Model:** hi , how are you doing ?
- 3 Next Utterance:** Recommended words: , (예시: 나는 잘하고 있어요.)
- You:** I'm good. do you have an iphone?
- 4 Model:** no , i do not have an iphone . i have a new house .
- 3 Next Utterance:** Recommended words: looking, house (예시: 굉장해. 나는 새 집을 찾고 있어요.)
- You:** wow really? where is it?
- 5 Model:** in a house in the city . i love watching the leaves change colors .
- 3 Next Utterance:** Recommended words: leaves, (예시: 나도 나뭇잎을 좋아한다.)
- You:** That is so cool. I love it too. It's beautiful
- 8 Error Correction:** correct answer is That is so cool. I love it too.
- 6 Model:** i love it . i have a son in junior high .
- 3 Next Utterance:** Recommended words: son, (예시: 굉장해. 나는 중학생 아들이 있다.)

1 : 토픽이 Bob Ross 로 주어졌다.

2 : 사용자의 발화에 응답을 하는 챗봇의 페르소나가 주어졌다.

3 : 사용자의 발화를 예측하여 한국어 해석과 중요 단어를 같이 제공하였다.

4 : 새 집을 샀다는 챗봇의 페르소나가 잘 반영되었음을 확인할 수 있다.

	 You do you spend a lot of time with him?
	 Model yes , i spend a lot of time with him .
3	 Next Utterance Recommended words: son, (예시: 굉장해. 나는 중학생 아들이 있다.)
	 You what do you do together?
7	 Model we watch bob ross on tv .
3	 Next Utterance Recommended words: cool, (예시: 멋지다.)
	 You Tell me about it. I've never heard of it.
	 Model it is a show about animals .
3	 Next Utterance Recommended words: check, (예시: 나는 그것을 확인해야 할거야.)
	 You wow. do you like animals?
	 Model i do . i have a cat .
3	 Next Utterance Recommended words: cat, (예시: 나도 고양이가 있다.)
	 You what is it's name?
	 Model his name is bob .
3	 Next Utterance Recommended words: name, (예시: 좋은 이름입니다.)
	 You haha like the tv show?
8	 Error Correction correct answer is Haha, like the tv show?
	 Model yes . i love it .
3	 Next Utterance Recommended words: love, (예시: 나도 그거 사랑해.)

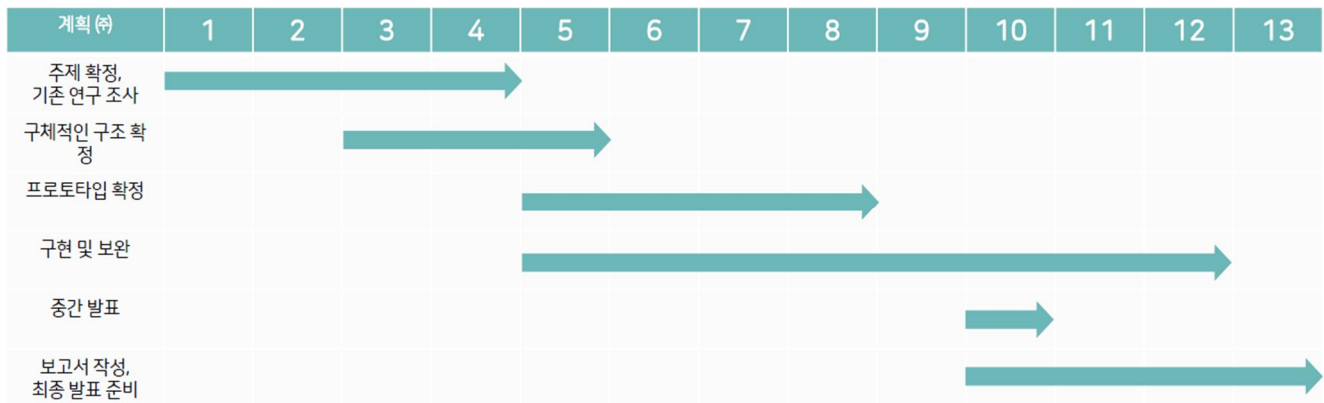
5 : 나뭇잎의 색이 변하는 것을 보는 것을 좋아한다는 챗봇의 페르소나가 잘 반영되었음을 확인할 수 있다.

6 : 고등학생 아들이 있다는 챗봇의 페르소나가 잘 반영되었음을 확인할 수 있다.

7 : 토픽인 Bob Ross 가 잘 활용됨을 확인할 수 있다.

8 : 에러 수정

V. 일정 및 역할 배분



- 문승준: 조장 , 논문 정리, 연구제안발표 , 중간 보고서 작성, 전체적 구현
- 목요한: 발표자료 , 논문 정리 , 중간 보고서 작성, 중간 발표, 최종 영상 제작
- 이해지: 발표자료 , 논문 정리 , 중간 보고서 작성, 최종 발표, 최종 보고서 수정

VI. 참고문헌

- Emily Dinan et al. 2018. The Second Conversational Intelligence Challenge(ConvAI2). arXiv:1902.00098
- Emily Dinan , Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli , Jason Weston. 2019. Wizard of Wikipedia: Knowledge Powered Conversational agents. arXiv:1811.01241
- Hannah Rashkin , Eric Michael Smith, Margaret Li, Y Lan Boureau . 2019. Towards Empathetic Open domain Conversation Models: a New Benchmark and Dataset. arXiv:1811.00207
- Marek Rei , Mariano Felice. 2017. "Artificial Error Generation with Machine Translation and Syntactic Patterns". Cornell University. arXiv:1707.05236
- Pierre Emmanuel Mazaré , Samuel Humeau , Martin Raison, Antoine Bordes . 2018. Training Millions of Personalized Dialogue Agents. arXiv:1809.01984

- Saizheng Zhang, Emily Dinan , Jack Urbanek , Arthur Szlam , Douwe Kiela , and Jason Weston. 2018.
- Personalizing dialogue agents: I have a dog, do you have pets too?. CoRR , abs/1801.07243.