

최신 대화형 에이전트 기반 상용화 교육 플랫폼 오류 분석

이승준¹, 박찬준¹, 서재형¹, 임희석^{2*}

¹고려대학교 컴퓨터학과 석·박사통합과정, ²고려대학교 컴퓨터학과 교수

Error Analysis of Recent Conversational Agent-based Commercialization Education Platform

Seungjun Lee¹, Chanjun Park¹, Jaehyung Seo¹, Heuseok Lim^{2*}

¹Master & Ph.D. Combined Student, Department of Computer Science and Engineering, Korea University

²Professor, Department of Computer Science and Engineering, Korea University

요약 최근 교육 분야에서 다양한 인공지능 기술을 활용한 연구와 개발이 이뤄지고 있다. 인공지능을 활용한 교육 중 특히 대화형 에이전트는 시간과 공간의 제약을 받지 않고 음성인식, 번역과 같은 다양한 인공지능 기술과 결합해 더 효과적인 언어 학습을 가능하게 한다. 본 논문은 상용화된 교육용 플랫폼 중 이용자 수가 많고 영어 학습을 위한 대화형 에이전트가 활용된 플랫폼에 대한 동향 분석을 진행하였다. 동향 분석을 통해 현재 상용화된 교육용 플랫폼의 대화형 에이전트는 여러 한계점과 문제점이 존재했다. 구체적인 문제점과 한계점 분석을 위해 사전 학습된 최신 대용량 대화 모델과 비교 실험을 진행하였고, 실험 방법으로 대화형 에이전트의 대답이 사람과 비슷한지를 평가하는 Sensibleness and Specificity Average (SSA) 휴먼 평가를 진행하였다. 실험 내용을 바탕으로, 효과적인 학습을 위해 개선방안으로 대용량 파라미터로 학습된 대화 모델, 교육 데이터, 정보 검색 기능의 필요성을 제안했다.

주제어 : 대화, 에이전트, 딥러닝, 교육용 플랫폼, 챗봇, 언어 융합

Abstract Recently, research and development using various Artificial Intelligence (AI) technologies are being conducted in the field of education. Among the AI in Education (AIEd), conversational agents are not limited by time and space, and can learn more effectively by combining them with various AI technologies such as voice recognition and translation. This paper conducted a trend analysis on platforms that have a large number of users and used conversational agents for English learning among commercialized application. Currently commercialized educational platforms using conversational agent through trend analysis has several limitations and problems. To analyze specific problems and limitations, a comparative experiment was conducted with the latest pre-trained large-capacity dialogue model. Sensibleness and Specificity Average (SSA) human evaluation was conducted to evaluate conversational human-likeness. Based on the experiment, this paper propose the need for trained with large-capacity parameters dialogue models, educational data, and information retrieval functions for effective English conversation learning.

Key Words : Conversation, Agent, Deep Learning, Education Platform, Language Convergence

*This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2018-0-01405) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation) and supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2021R1A6A1A03045425).

*Corresponding Author : Heuseok Lim(limhseok@korea.ac.kr)

Received December 3, 2021

Revised December 29, 2021

Accepted March 20, 2022

Published March 28, 2022

1. 서론

최근 인공지능 기술의 발전에 따라 다양한 산업군과 여러 엔터테인먼트가 결합하여 발전하고 있다. 그 중 교육 산업은 인공지능 결합이 중요한 활동 분야 중 하나이다 [1]. 특히, 포스트 코로나 시대 이후 오프라인이 아닌 온라인 학습이 적용되고, 학습자들은 과거보다 더 많은 자기주도학습 시간이 주어질 수밖에 없게 되었다. 이 때문에 자연스럽게 인공지능이 결합한 교육 플랫폼이 어느 때 보다 주목받았다.

Artificial Intelligence in Education (AIEd)이란 인공지능을 활용한 교육을 의미한다. AIEd는 최근 자연어 처리, 음성인식 등을 포함한 다양한 인공지능 분야와 결합하여 발전하고 있다. AIEd는 크게 세 가지 맥락으로 볼 수 있으며, 비교적 구조화된 문제를 다루기에 적합한 지능형 튜터링 시스템 (ITS), 반구조화된 문제를 다루기에 적합한 대화형 튜터링 시스템 (DBTS), 비구조화된 문제를 다루기에 적합한 탐구학습 시스템 (ELE)로 나눌 수 있다[2].

지능형 튜터링 시스템 (Intelligent Tutoring System, ITS)는 학습 단계가 잘 정의된 지식 구조를 갖는 분야에 대해 효과적인 단계별 학습 과정을 제공한다. ITS는 효과적인 튜터링을 위해 교수 모델 (Pedagogical Model), 도메인 모델 (Domain Model), 학습자 모델 (Learner Model)으로 구성된다. 교수 모델은 교수 혹은 교사의 효과적인 티칭을 위한 모델로 사용되고, 대표적인 예로는 LMS (Learning Management System)과 같은 학습관리시스템이 있다. 학습 내용을 중심으로 한 주제 모델로써 사용되는 도메인 모델의 예로는 강화학습을 사용한 산타도익¹⁾, AI를 활용해 드릴다운 (Drill Down) 시스템을 적용한 노리 (KnowRe)²⁾가 있다. 학습자 모델의 대표적인 사례로는 실시간에서 학습자와 상호작용을 극대화하기 위해 만든 미네르바 스쿨의 액티브 러닝 포럼 (Active Learning Forum)이 있다.

대화형 튜터링 시스템 (Dialogue-Based Tutoring System, DBTS)은 ITS에서 발전된 형태로 대화형 에이전트 혹은 Dialogue System이라고 불리는 챗봇을 활용하여 학생들을 학습의 대화로 참여시킨다. DBTS에

사용되는 대화형 에이전트는 자연어 처리 기술 중 NLU (Natural Language Understanding)과 NLG (Natural Language Generation) 기술을 기반으로 한다. 학습자가 글이나 말로 입력하여, 대화형 에이전트가 이를 인식하여 학습자의 이해도를 결정하고, 학생들이 오개념을 정정할 수 있는 피드백을 제공하거나 태스크에 적절한 답변을 출력한다[2].

탐구학습시스템 (Exploratory Learning Environment, ELE)은 정해져 있는 단계별 학습 순서를 따르지 않고, 학습자의 잘못된 학습 결과를 바로 잡기 위한 자동화된 피드백을 제공하는 시스템을 일컫는다. ELE를 활용한 대표적인 AIEd 기술로는 문법 교정 (Grammar Error Correction, GEC)과 논술형 문항 자동채점 (Automated Essay Scoring, AES) 등이 있다.

인공지능을 활용한 교육 (AIEd)에서 대화형 튜터링 시스템은 다음과 같은 특징을 지닌다. 첫째 전통적인 학습은 일반적으로 교실에서 이루어져 시공간의 제약이 있는 반면에 AIEd와 결합한 대화형 에이전트를 활용하면 이러한 제약을 받지 않고 자유롭게 학습을 진행할 수 있다. 둘째 자연스러운 대화 인터페이스와 다양한 인공지능 기술을 결합한 형태로 AIEd를 구현할 수 있다. STT (Speech-to-text) 및 TTS (Text-to-speech) 기술과 대화형 에이전트를 결합하면 실제 사람과 대화하는 듯한 경험을 제공할 수 있고, GEC와 결합하면 학습자에게 효과적인 문법 교정 피드백을 제공할 수 있다[3,4]. 이러한 대화형 튜터링의 특징들을 살려, 최근 대화형 에이전트와 언어 학습이 결합한 교육용 플랫폼이 출시되고 있다.

하지만, 현재의 인공지능 기술을 고려할 때, AIEd를 위하여 대화형 에이전트를 언어 학습 교육용 플랫폼으로 활용하기에는 여러 한계점을 지닌다. 대부분의 목적 지향 대화형 에이전트들은 지식 베이스에 수동으로 데이터가 입력되어야 하고, 특정 언어 혹은 제한된 도메인에만 적용될 수 있는 문제점이 있다. 또한, Chit-chat 모델 혹은 비목적형 대화 시스템으로 불리는 오픈 도메인 대화 시스템은 일관적인 페르소나 (persona)의 부족, 장기 기억의 부재로 이전에 했던 대화들이 고려되지 못하는 점, "I don't know"와 같은 구체적이지 않은 답변을 하는 문제점, 그리고 다른 주제에 관해 이야기할 경우 답변을 제대로 하지 못하는 한계점을 포함한다[5,6]. 이러한 한계점 때문에, 현재까지 대화형 에이전트를 언어 학습 교육용 플랫폼으로 적용

1) <https://www.riiid.co/>

2) <https://www.knowre.com/>

하는 데는 어려움이 존재한다[7].

본 논문은 인공지능을 활용한 교육 (AIEd) 중 대화형 에이전트가 활용된 영어 교육 플랫폼에 대한 분석을 진행했다. 기존 교육용 플랫폼에 관한 연구들은 대용량 (Large-scale) 대화형 에이전트 모델들과의 비교 실험이 존재하지 않았다. 본 논문은 앞으로의 교육용 플랫폼의 방안 제시를 위해 영어 교육 플랫폼과 대용량 대화 모델에 대한 휴먼 평가를 진행하였다. 이를 바탕으로 현재 대화형 에이전트를 사용한 영어 교육용 플랫폼에 대한 한계점과 문제점을 분석하고, 이에 따른 개선방안을 바탕으로 추후 대화형 에이전트를 활용한 영어 교육용 플랫폼 개발에 대한 활용 방안을 제안한다.

2. 관련 연구

2.1 대화형 에이전트 (Conversational Agent)

대화형 에이전트는 대화를 통하여 인간과 컴퓨터의 상호작용을 효과적으로 가능하게 하는데, 가상 비서 혹은 애플리케이션으로부터 컴퓨터가 음성 또는 텍스트 입력을 자연스럽게 이해하고 처리하며 응답할 수 있는 기술을 활용한다. 더 나아가서 개인화에 초점을 맞춘 대화형 에이전트에 관한 다양한 연구들이 진행 중이다. 인간과 컴퓨터와의 대화 시스템은 크게 두 종류의 유형으로 발전했다.

오픈 도메인 대화 시스템 (Open-domain dialogue system)과 태스크 기반 대화 시스템 (Task-oriented dialogue system) 대화 시스템이다. 특정한 목적을 수행해야 하는 태스크 기반 대화 시스템과 달리 오픈 도메인 대화 시스템은 챗봇 (Chatbot)과 같이 인간이 말하는 것 처럼 자연스럽게 사용자와 대화 (Chit-chat)를 해야 한다.

[8]에서는 기존의 Chit-chat 모델들의 문제점들을 답변에 일관성이 없거나, 부자연스럽고 구체적이지 않은 답변을 하는 등의 문제점들이 존재한다고 설명했다. 또한, [6]에서는 대부분의 챗봇 응답은 의미가 없거나, 예측할 수 있고, 페르소나가 부족하다고 밝혔다.

이에 최근의 오픈 도메인 대화 시스템 (Chit-chat)은 페르소나 (Persona)와 결합 되어 발전했다. 대표적으로 페르소나 대화 데이터셋 (PERSONA-CHAT)[8]이 만들어졌다. 페르소나를 반영한 대화는 두 화자의 나이, 직업, 가족, 취미 등 개인정보가 정해진 상황에서 시스

템에 부여된 개인 특성과 일치된 발화를 생성하도록 하는 대화이다. 컴퓨터가 페르소나 정보를 가지게 됨으로써, 보다 일관성 있고 자연스러운 대화를 할 수 있게 되었다.

[9]에서는 공감, 지식 등의 다양한 대화 스킬을 결합한 오픈 도메인 대화 시스템인 블렌더봇이 개발되었다. 이전의 대화 모델들이 파라미터 수를 늘리는 것에 초점을 맞췄다면, 블렌더봇은 대용량 (Large-scale) 파라미터뿐만 아니라 사람의 대화 스킬을 잘 학습할 수 있는 데이터셋을 활용해 모델 성능을 개선하였다.

2.2 대화형 에이전트 기반 교육 연구

지난 수십 년 동안, 언어 학습 분야의 AIEd를 포함한 다양한 기술 지원에 대한 많은 관심과 주목이 있었다. 언어 학습은 어휘력, 문법 그리고 듣기, 읽기, 말하기, 쓰기의 4가지의 언어능력을 높일 수 있는 수업 및 연습이 포함한다. 해당 기능을 포함한 교육용 대화형 에이전트 사용을 통해 학습자들은 언어 사용을 통한 성장뿐만 아니라 성취도에 대한 즉각적인 피드백을 받을 수 있다.

[10]에 따르면 챗봇은 학생들에게 강한 동기 부여를 통해 더 많은 학업 수행할 수 있게 해준다고 밝혔다. [3]의 연구에서는 언어 학습에서의 챗봇의 역할에 대해 조사를 진행했고, 다음과 같은 챗봇이 학생들을 위해 언어 학습을 높일 있는 6가지 특징에 대해 밝혔다. 첫째 학생들은 사람과 대화하는 것보다 봇과 대화를 더 편하다고 느낀다. 둘째 봇을 사용하면 무제한으로 대화를 나눌 수 있으며 셋째 봇을 사용하여 학생은 듣기와 읽기 학습이 가능하다. 넷째 봇 자체로 흥미로우며, 다섯째 봇을 통해 학생은 단어 학습과 다양한 언어 표현 학습이 가능하다. 마지막으로 봇을 사용하면 학습자의 문법 오류와 철자 오류에 대한 효과적인 피드백이 가능하다.

[11] 연구에서는 학습자의 억양과 말하기 속도를 높이기 위해 TTS 기능을 포함한 ROBOSEM (교육용 서비스 로봇)을 개발하였고, TTS 사용 여부에 따른 학습자 발음 향상 여부를 연구하였다. 대부분 사용자가 ROBOSEM의 TTS 기능을 사용한 후에 말하기 능력과 발음 향상에 효과적임을 밝혔다.

[12]에는 대화형 에이전트와의 채팅을 통한 효과를 파악하기 위해 사용 전과 사용 후에 영어 문법 시험을 치렀다. 사전 사후 평가 분석 결과, 대화형 에이전트와의 채팅 사용에 따른 영어 문법 능력 상승이 통계적으로 유의미하다고 밝혔다.

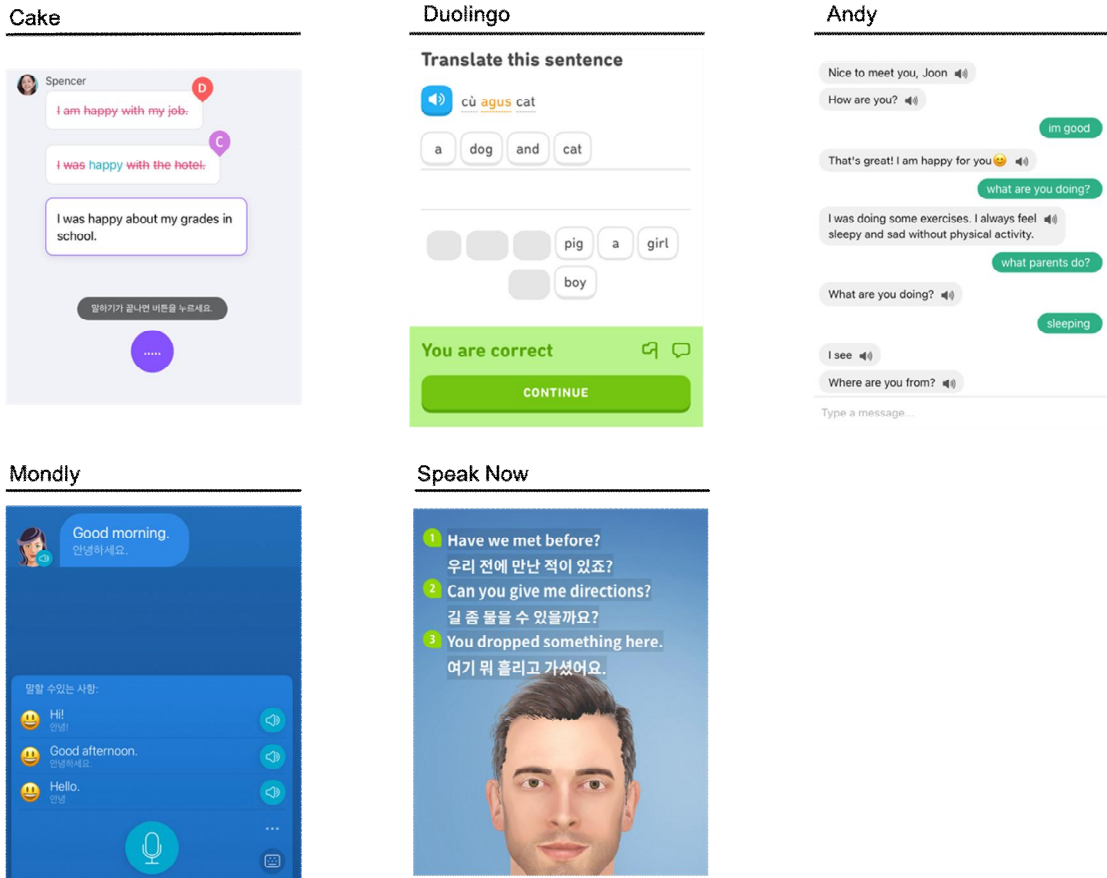


Fig. 1. The design and layout of AIEd platforms using conversational agent

3. 대화형 에이전트 기반 교육용 플랫폼 동향

기존의 영어 교육과 관련된 챗봇 연구들은 대부분 아마존 Alexa 혹은 구글 Assistant와 같은 인공지능 스피커에 대한 분석이 주를 이뤘다. 이러한 범용 챗봇이 갖고 있는 정보 제공의 장점을 활용하고, 특정 상황에 맞는 대화를 하는 데 한계가 있는 점을 고려한 것이라고 볼 수 있다[13]. 범용 챗봇이 간단한 답답과 정보 검색에 유용하지만[14], 영어 교육 활동에 사용하기에는 한계가 있다[13]. 따라서 영어 교육 목적에 제작된 챗봇에 대한 분석이 필요하다.

본 논문은 상용화된 교육용 플랫폼 중에서 이용자가 많고, 영어 학습을 위해 대화형 에이전트가 설계된 앱들을 선정하여 동향 분석을 진행했다.

3.1 Cake

케이크 (Cake)는 영상 콘텐츠 기반의 영어 학습 플랫폼이다. 학습자는 미국 TV 드라마나 프로그램, 영화에서의 실제 원어민들이 사용하는 다양한 표현을 영상을 통해 학습한다. 학문적으로 학습한 영어와 실제 원어민이 사용하는 영어 표현에는 차이점들이 존재한다. 이러한 차원에서 케이크 영어 교육 플랫폼은 실제 원어민들이 사용하는 영어 회화 표현은 TV 프로그램, 영화 등 영상 콘텐츠를 통해 학습할 수 있다. 또한, STT와 TTS를 사용하여 실제 원어민과 발음이 비슷한지를 평가하거나 실제 발음을 들어보면서 연습을 할 수 있다.

Fig. 1과 같이 챗봇의 형태를 띠고 있지만, 실제 대화하는 것이 아닌 앱상에서 주어진 버튼을 눌러 발음을 평가받거나, 단어검색과 표현 검색 결과를 표시해주는 기능만 존재한다.

3.2 Duolingo

듀오링고 (Duolingo)는 가장 인기 있는 언어 학습 챗봇 플랫폼 중 하나이다. 듀오링고는 30가지 이상의 언어 학습을 지원한다. 2020년 이전에 일부 언어에 대해 챗봇 기능을 지원하였지만, 해당 기능은 현재 삭제되었다. 듀오링고의 가장 큰 특징은 게임화 요소와 번역 기반의 학습이다. 학습자들의 학습 지속 동기 부여를 위해 게임화 전략을 사용한다[16].

게임화 전략은 다음과 같이 이뤄진다. 학습 평가를 통해 학습자의 맞춤형 수준의 학습 (단원)을 제공하고, 각 단원의 레벨을 올리면 더 난도가 높은 학습을 진행할 수 있다. 혹은 오답 입력하면 점수를 잃어 테스트를 다시 시작하거나 특정 수업을 완료해야 한다. STT와 TTS를 통한 발음 학습, GEC를 통한 문법 학습 등 인공지능을 통해 말하기, 읽기, 쓰기 및 듣기 학습을 기반으로 일상, 식당, 여행과 같은 상황 기반의 학습 등 폭넓은 영어 학습을 제공한다.

수업 내에서 활동은 주로 발음, 어휘, 듣기, 문법에 중점을 두지만, 대부분의 학습 과정은 번역 연습을 기반으로 한다. 다른 AIED들과 비교하여 듀오링고 번역 기반 학습 방식은 지나치게 단순해 보일 수 있지만, 구조화된 학습 설계와 발음, 어휘, GEC를 포함한 학습 과정은 학생들의 정확한 발음과 어휘, 기초 문법 구조 습득에는 긍정적인 영향을 미칠 수 있다.

3.3 Andy

Andy 영어 학습 봇은 영어 학습을 도와주는 가상 튜터 앱이다. 이 챗봇은 학습자와 대화형 에이전트를 기반으로 새로운 단어를 학습하거나, 문법을 배우고 언어 학습 게임을 제공한다. 특히, 학습자들은 실제 대화에서 영어를 사용함으로써 언어 학습이 가능하다. 예를 들어, 서로 안부를 주고받거나, 특정 화제에 관해 이야기를 나눔으로써 영어 회화를 학습할 수 있다. 또한, 이 챗봇은 학습자와 토론을 할 수 있다. 학습자들은 Andy 챗봇과 영화, 여행, 미술 그리고 유머와 같은 다양한 주제를 토론할 수 있다. 이렇게 챗봇과 몇 번의 토의 및 대화를 진행한 후에, 게임을 통해 언어 학습도 가능하다.

언어 학습은 단어 학습과 문법 학습을 포함한다. 단어 학습은 챗봇이 학습자에게 단어와 단어 뜻을 설명해주고, 제시한 단어를 얼마나 알고 있는지를 피드백을 형태로 파악한다. 단어의 뜻과 설명이 끝난 뒤에는 학습자가

얼마나 해당 단어를 학습했는지 4가지의 척도 (Very Bad, Explain, Easy, Ok)로 평가하는 방식으로 학습을 진행한다. 문법 학습은 학습자가 학습을 원하는 문법을 선택하면 챗봇이 해당 문법에 관해 설명을 진행하고, 이후 문법 테스트를 진행한다. 문법 테스트는 객관식 형태로 주어지고 학습자는 여러 개의 보기 중에 하나의 정답을 선택하는 방식으로 테스트가 진행된다. 학습 완료 후에는 단어 학습과 문법 학습 결과가 단어집 혹은 데이터베이스에 저장되어 복습하거나 학습 통계용으로 사용된다.

하지만, Andy 챗봇은 다음과 같은 한계점을 지닌다. 학습자와의 이전 대화 기록을 저장하지 못하고 가장 최근의 대화 입력에 대해서만 답변을 출력하는 것을 확인할 수 있었다. 또한, 사용자 입력을 이해하지 못하면 “I see” 혹은 “Got it”과 같은 대답 후에 정해진 화제 안에서 질문을 출력하는 한계점을 지닌다. 즉, 사용자 입력에 대한 다양한 응답을 생성하는 것이 아니라, 정해진 표현만 사용하여 대화를 진행한다. 이는 사용자와 깊은 대화를 하지 못하여 비효율적인 언어 학습을 초래할 수 있다. Andy는 무늬만 비목적형 대화 형식을 띠고 있을 뿐, 실질적으로 다양한 문장을 생성하지 못하는 목적형 기반 챗봇의 형태를 나타내고 있다.

3.4 Mondly

Mondly는 세계적으로 잘 알려진 언어 학습 앱 중 하나이다. 이 앱은 언어 학습에 챗봇을 보조적으로 사용하여 기존 영어 학습보다 효과적인 학습이 가능하다. 다른 앱들과 마찬가지로, Mondly는 반복적인 단어 학습과 플래시 카드를 활용한 학습 방법을 제공한다. Mondly 챗봇은 33개의 다중 언어를 지원하고, 단어와 문장 학습 후에 실제 원어민 음성이 포함된 채팅이 가능하고, 음성인식 기능을 지원하여 발음에 대해 학습할 수 있다. 또한, 식당에서 음식을 주문하는 것과 같은 일상적인 시나리오를 기반의 언어 학습을 통해 실용적인 언어 학습이 가능하다. 단어 학습의 경우 틀린 단어를 입력할 경우 올바른 정답을 피드백으로 제공한다.

Mondly는 다음과 같은 한계점을 지닌다. Fig. 1과 같이 자유로운 대화가 아니라 답변할 수 있는 입력에 관해서만 대화를 주고받을 수 있는 형태를 띠고 있다. 입력이 제한적이기 때문에 학습자가 원하는 주제로 대화하는 것이 어렵고, 다양한 표현에 대해 학습하기 어려운 단점이 존재한다.

3.5 Speak Now

스픽나우 (Speak Now)는 인공지능 아바타를 사용한 영어 회화 학습을 위한 교육용 앱이다. 학습자가 원하는 스타일의 아바타를 튜터로 선택하여 학습할 수 있다. 스픽나우는 기쁨 표현하기, 이성에게 말 걸기 등 다양한 시나리오를 기반으로 회화 학습을 제공한다. 또한, 다른 영어 학습 플랫폼과 달리 자유 대화 기능을 제공한다. 이러한 학습 방법을 포함하여 대화의 입력은 음성인식 기능을 지원하여, 실제로 원어민과 회화하는 느낌을 준다. 또한, 다른 교육용 플랫폼들과 마찬가지로 반복적인 단어 학습 기능을 제공한다.

스픽나우의 대화 시스템은 자유 대화와 상황별 대화 기능을 포함한다. 자유 대화 기능은 크게 안부 인사 묻기와 특정 주제에 관해 대화하기 순으로 진행된다. 안부 인사 묻기의 경우에는 사용자가 안부 인사와 벗어난 질문을 입력하면 챗봇은 원하는 안부 인사가 입력될 때까지 인사를 반복한다. 특정 주제에 관한 대화는 “Let’s talk about ~” 혹은 “I would like to speak ~”와 같이 정해진 패턴으로 대화를 입력해야 대화할 수 있고 그렇지 않으면 계속해서 해당 패턴으로 말하기를 요구한다. 게다가 대화할 수 있는 주제가 정해져 있어, 지식 베이스에 저장되지 않는 주제를 입력하면 다른 주제로 강제로 전환하거나 다른 주제를 말할 것을 요구한다. 또한, 대화 전반적으로 장기 기억 의존성 문제로 이전 대화 기록을 반영하지 못하는 문제점이 있다.

상황별 대화에서는 학습자의 선택지에서 선택되어야 하므로 Fig. 1과 같이 제한된 대화만 가능하다. 따라서 학습자가 원하는 답변을 하지 못하고 깊은 대화가 불가능하다. 즉, 자유 대화와 상황별 대화 기능 모두 형태만 대화형 에이전트 형태만 띠고 있을 뿐, 효과적인 회화 연습이 불가능하다.

4. 실험 및 실험결과

본 논문은 위에 분석한 자유로운 대화 (Chit-chat)가 가능한 교육용 플랫폼과 오픈 소스로 공개된 최신 오픈 도메인 대화 모델에 대해 휴먼 평가 (Human Evaluation)를 진행하였다.

4.1 데이터셋

실험을 위한 데이터로 대화 입력에 따른 모델 응답

비교를 위해 [17]에서 제안된 MTB 벤치마크 데이터셋을 사용하였다. MTB 데이터셋은 1,477개의 대화 문맥으로 구성되어 있고, 각 문맥은 1~3개의 대화 턴을 포함한다. 또한, 일관성 있는 페르소나를 평가하기 위해 “Do you like cats?”과 같은 페르소나를 묻는 대화도 포함 되어 있다. 본 장의 실험을 위해 1,477개의 대화 문맥 중에서 30개의 대화를 임의 추출하여 실험 데이터셋으로 사용하였다.

4.2. 실험 모델

본 논문은 실험을 위한 비교 모델은 대화형 에이전트를 포함하여 실질적으로 대화가 가능한 교육 플랫폼인 Andy, 스픽나우와 대용량 사전 학습 대화 모델인 블렌더봇 2.0을 실험 대상으로 선정하였다.

케이크, Mondly는 대화형 에이전트의 형태를 나타내고 있지만, 실질적으로 학습자가 원하는 대화를 입력하여 대화할 수 없다. 이와 같은 이유로 Andy와 스픽나우를 본 장의 실험 대상으로 선정하였다. Andy는 기본적으로 챗봇 형태로 학습이 이루어진다. 채팅 입력으로 궁금한 문법, 단어를 물어보거나, 자유로운 대화가 가능하다. 스픽나우의 경우는 학습자에게 자유 대화를 따로 제공하여 회화 연습이 가능하다.

블렌더봇 2.0은 블렌더봇 1.0을 개선한 오픈 도메인 대화 시스템으로 2021년 7월에 공개되었다 [18,19]. 기존 블렌더봇 1.0은 공감, 지식 등의 다양한 대화 스킬의 결합을 처음으로 시도한 챗봇이다. 블렌더봇 1.0은 대용량 (Large-scale) 파라미터를 포함할 뿐만 아니라 사람의 대화 스킬을 잘 학습 할 수 있는 데이터셋을 활용한다.

이전의 대화와 모순되는 발언을 하거나, 최신의 정보를 반영하지 않는 대화를 생성하는 것이 기존 모델인 블렌더봇 1.0이나 GPT-3[20]의 대표적인 문제점으로 지적됐다. 블렌더봇 2.0은 여러 세션에 걸쳐 일관된 대화를 수행하며, 인터넷의 있는 정보를 검색해 최신 정보에 관한 대화도 가능하다. 이는 기존 대화 모델과 가장 큰 차이점이다. 또한, 클라우드 소싱을 통해 Wizard of the Internet (WizInt) 데이터셋[18]과 Multi-Session Chat (MSC) 데이터셋[19]을 사용한 것이 특징이다.

WizInt 데이터셋은 사용자의 인터넷 검색 입력에 관한 결과를 답변으로 생성하기 위해 사용되었고, MSC 데이터셋은 다중 세션으로 구성된 대화에서 새롭게 알게

된 지식을 장기 기억에 저장하고 이를 답변으로 생성하기 위해 사용되었다. 이러한 데이터셋의 사용으로 주어진 페르소나를 해치지 않으면서 인터넷 정보 검색 결과가 반영된 대화를 출력할 수 있고, 세션 개념을 도입하여 긴 시간 간격이 있어도 자연스럽게 이어지는 대화 생성이 가능하다.

4.3 실험디자인

기존의 교육용 챗봇에 관한 연구[13-15,21]들은 대부분 최신 대용량 사전 학습 대화 모델들과의 직접적인 비교 실험이 이루어지지 않았다. 대화형 에이전트에 대한 평가를 진행하는 것이 자체가 어렵고[22], 평가를 위한 데이터셋 구축과 평가자 구성에 어려움이 있기 때문이다. 또한, 상용화된 교육 플랫폼에 사용된 대화형 에이전트는 공개된 오픈 소스가 아니기에 직접적인 평가가 어렵다.

본 논문은 최신 대화 모델과의 교육 플랫폼에 사용된 대화 모델에 대한 휴먼 평가를 통한 비교 실험을 진행했다. 비교 실험을 통해 교육 플랫폼에 사용된 대화 모델의 한계점과 문제점을 분석하고 이를 바탕으로 개선사항을 제안하고자 한다. 최신 대화 모델은 블렌더봇 2.0을 교육 플랫폼 대화 모델은 Andy, 스피크나우를 실험 대상으로 선정하였다. 또한, 휴먼 평가를 위해 [17]에서 제안된 SSA (Sensibleness and Specificity Average) 평가 지표와 이를 활용한 정적 평가와 대화형 평가를 통해 실험을 진행했다.

4.4 평가지표

SSA 평가 지표는 2020년 1월 구글 브레인에서 발표한 대화형 에이전트 Meena[17]에서 사용된 휴먼 평가 방법론이다. SSA 평가 지표는 챗봇이 사람처럼 대화할 수 있는지 (Human-like)를 평가하기 위한 지표이다. 기존의 지식 베이스 혹은 규칙 기반 (Rule-based)의 시스템의 평가방식에 의존적 않고 오픈 도메인 챗봇에 대해 사람처럼 대화할 수 있는지를 평가할 수 있다. 교육용 챗봇이 실제 사람과 대화하는 느낌을 학습자에게 제공하면 효과적인 회화 학습이 가능하기에 본 논문은 SSA 평가 지표를 사용하였다.

SSA 평가 지표는 사용자의 입력 즉 문맥에 대해서 챗봇의 응답이 말이 되는지 (Sensibleness)와 구체적인 (Specificity)인지를 평가한다. 구체적으로 Sensibleness는

응답이 상식적으로 말이 되는지, 논리적으로 말이 되는지, 일관성을 갖고 대답하는지를 평가한다. 하지만, Sensibleness는 일관적으로 "I don't know"와 같이 지루하고 구체적이지 못한 대답에도 높은 점수가 부여될 수 있다. 이러한 문제점을 개선하기 위해 Specificity는 Sensibleness가 1로 평가되었을 때, 문맥에 맞게 충분히 구체적인 대답인지 여부를 평가한다. 그러므로, Sensibleness와 Specificity 척도 모두 0 또는 1의 값을 갖지만, Sensibleness가 0으로 평가되면 Specificity 은 0으로 설정된다.

SSA 평가 방법은 정적 평가 (Static Evaluation)와 대화형 평가 (Interactive Evaluation)로 분류된다. 정적 평가란 여러 대화 모델들을 비교하기 위해서 하나의 공통된 대화에 대한 답변을 SSA 평가 지표로 평가하는 방법론이다. 본 논문은 정적 평가를 위해 제작된 MTB (Mini-Turing Benchmark)[17] 데이터셋을 평가 데이터셋으로 사용하였고, 3명의 평가자로 구성하여 실험을 진행하였다. 정적 평가는 같은 데이터셋으로 여러 모델을 비교하기에는 적합할 수 있지만, 정적 평가 데이터셋이 어떻게 구성되어 있는지에 따라 편향될 수 있다. 따라서 대화형 평가 또한 같이 이뤄져야 한다. 대화형 평가란 정해진 대화 입력이 아니라, 평가자가 실제 대화 모델과 대화하면서 SSA 평가를 진행하는 방법론이다. 본 논문은 각 챗봇 시스템에 대해 최소 6회 이상 대화하도록 요청했으며(사용자 발화 3회, 챗봇 시스템 3회), 최종적으로 10개의 대화 세션에 대해 실험을 진행하였다.

4.5 실험 결과

4.5.1 양적 분석

본 논문은 정적 평가를 위해 SSA[17] 평가 지표 방법론에서 제안된 MTB 벤치마크 데이터셋 중 30개를 임의 추출하여 실험을 진행하였다.

정적 평가의 실험 결과는 Fig. 2과 같다. 모든 평가 척도에서 블렌더봇이 높은 수치를 보여준다. 블렌더봇 다음으로는 Andy와 스피크나우 순으로 평가되었다. 스피크나우의 경우 MTB 데이터셋에 대해 Specificity 점수가 0점에 가까운 점수가 나온 것을 확인할 수 있다. 대화형 평가의 실험 결과는 Fig. 3과 같다. 정적 평가와 비슷하게 블렌더봇이 대부분의 평가에서 높은 점수를 보여주었다. 유일하게 Sensibleness 평가에서 스피크나우가 다소 높은 점수를 보여주었다. SSA 평가 지표 기

준으로 블렌더봇, 스피크나우, Andy 순으로 평가되었다. 정적 평가에 비해 Andy와 스피크나우 모두 향상된 점수를 보여주었다.

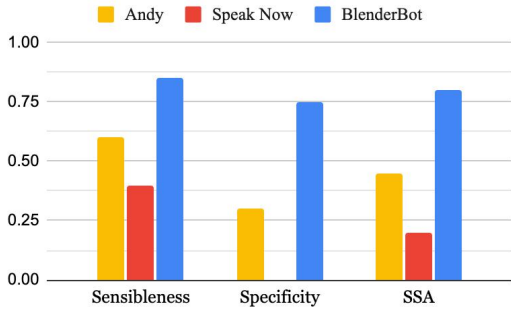


Fig. 2. Results of static human evaluation



Fig. 3. Results of interactive human evaluation

4.5.2 질적 분석

본 논문은 영어 교육용 플랫폼의 문제점과 한계점을 분석하기 위해 실험 결과를 바탕으로 질적 오류 분석을 진행하였다. 질적 오류 분석에 관한 결과는 Table 1와 Fig. 4과 같다.

Table 1는 정적 평가에 대한 질적 오류 분석 예시이다. BlenderBot이 주어진 문맥에 맞게 대답하였고, 대답 또한 구체적이고 명확하다. Andy의 경우, “Interesting”이라는 답변이 문맥에 적절한 응답일 수 있지만, 구체적이거나 명확하지 않다. 스피크나우의 경우, 주어진 응답이 문맥에 맞거나 구체적이지 않다. 블렌더봇의 경우는 주어진 문맥에 대해 적절하고 구체적으로 응답하였다.

정적 평가에 대한 질적 오류 분석으로 살펴본 Andy와 스피크나우의 문제점은 다음과 같다. 지식 베이스에서 벗어난 대화를 입력할 경우, 문맥에 적절하지 못한 답변을 생성하거나 문맥에 적절하더라도 구체적이지 못

한 답변을 생성한다. 특히, 스피크나우의 경우 제한된 주제만 입력을 받을 수 있도록 설계되었기 때문에 다양한 주제가 포함된 MTB 데이터셋에 대해 Fig. 2과 같이 매우 낮은 점수를 보여줬다. 하지만, 블렌더봇의 경우 대부분의 대화 입력에 대해 문맥에 적절하고 구체적인 대답을 생성하였다.

Table 1. Example of generated response given the context in the MTB benchmark dataset.

Context:	Have you ever tried a mechanical keyboard?
Andy:	Interesting
Speak Now:	Thanks. Have a nice day.
BlenderBot:	I have, but I'm not very good at it. It's hard for me to press the keys correctly.

Fig. 4는 대화형 평가에 대한 질적 오류 분석 예시이다. Andy의 경우 대부분의 대답이 사용자 입력에 적절하지만, 구체적이지는 못하다. Fig. 4에서 “Nice.” 혹은 “Oh, It’s good to hear that.”과 같이 구체적이지 못한 답변 하는 모습을 살펴볼 수 있다. 효과적인 회화 학습을 위해서는 이와 같은 응답보다 질문을 포함하거나 블렌더봇과 같이 구체적인 응답을 통해 학습자의 반응을 유도해야 한다. 또한, Andy는 학습자 입력에 대한 깊은 이해가 어렵다. Andy는 영어 단어검색 기능을 제공하지만 “Do you know the meaning of ‘Good?’” 질문에 “Good”이 아닌 “Is”로 인식해 대화 형식에서 단어검색의 어려움을 보였다.

스피크나우의 경우 지식 베이스에서 벗어난 질문에 취약해 학습자의 입력을 제한하는 문제점이 존재한다. Fig. 4를 살펴보면 대부분 응답이 문맥에 적절하고 구체적이다. 하지만, 자세히 살펴보면 모든 챗봇의 응답은 질문을 포함하고 있다. 이는 사용자에게 챗봇이 답변 가능한 주제로 대화를 유도하는 것이다. 즉, 어떤 응답이 입력되더라도 답변을 생성하는 것이 아닌 미리 사전에 저장된 답변을 출력한다. Fig. 4의 스피크나우를 살펴보면, “Do you like your boss?”라는 질문에 “No, Do you know Leonardo DiCaprio”와 같이 지식 베이스에서 벗어난 질문을 입력 하면, 단순하게 이를 부정문으로 판단하여 사전에 설정된 응답을 출력한다. 이렇게 대화를 제한하여 대화형 평가에서는 높은 평가를 받을 수 있었지만, 다양한 입력이 포함된 정적 평가에서는 낮은 점수를 보여주었다. 또한, 단어검색 기능을

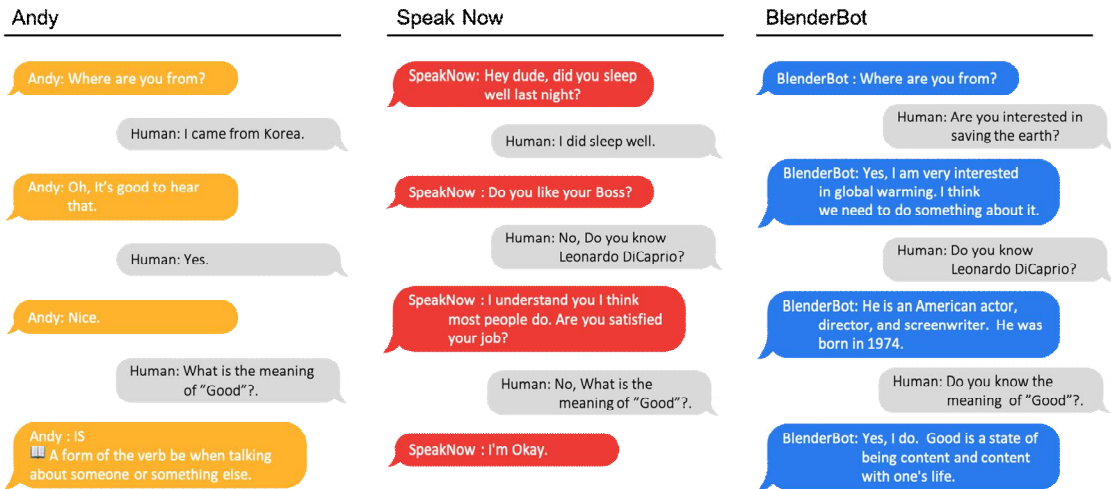


Fig. 4. Interactive examples multi-turn dialogue generation. Conversational agent are Andy, Speak Now, and Blenderbot

지원하지 않아 “No, What is the meaning of ‘Good?’” 질문에 제대로 답변하지 못하는 모습을 살펴 볼 수 있다.

Andy와 스피카우와 대조적으로 블렌더봇은 학습자의 입력에 대한 인식이 뛰어나고 답변 또한 구체적이다. 더불어, 모르는 단어나 정보에 대해 이를 인터넷 검색을 통해 답변을 생성한다. Fig. 4를 살펴보면 “Where are you from?”이라는 질문에 “Do you know Leonardo DiCaprio?”와 같은 주제에서 벗어나고, 지식 베이스에 포함되지 않은 질문을 할 경우, 이를 인터넷 검색을 통해 답변을 생성하는 등 다양한 입력에 대해 강건함을 확인할 수 있다. 또한, 단어 검색도 가능하다. Andy와 대조적으로 “Yes, I do. ~”와 같이 답변을 생성하며 단어의 뜻을 제시한다. 블렌더봇의 모든 응답은 맥락에 적절할 뿐만 아니라 구체적이다. 이는 영어 회화 학습자에게 흥미를 유발하고 동기 부여를 제공해 효과적인 학습이 가능하게 한다.

5. 제안하는 개선방안

앞서 언급한 문제점 및 한계점에 대해 아래와 같이 최신 대용량 대화 모델인 블렌더봇과 비교를 통해 개선 방안을 제안한다.

Andy와 스피카우를 포함한 교육 플랫폼의 문제점 및 한계점을 정리하면 다음과 같다. 첫째 지식 베이스에서 벗어난 입력을 할 경우, 문맥에 적절할 수 있지만

구체적이지 못한 답변을 생성한다. 둘째 다양한 답변을 생성하지 못하고, 정해진 답변만을 출력한다. 이러한 한계점으로 인해, Mondly와 같이 제한된 입력만 학습자에게 요구하거나, 스피카우와 같이 답변이 가능한 질문에 대해서만 계속해서 질문을 요구한다. 이렇게 제한된 대화 방식은 학습자에게 보다 폭넓은 대화를 하지 못하여 상대적으로 자유로운 대화에 비해 흥미와 동기가 감소해 학습의 효율을 감소시킬 수 있다. 그러므로 본 논문은 대화형 에이전트가 활용된 교육 플랫폼에 대해 다음과 같이 개선사항을 제안한다.

5.1 대용량 사전학습 모델

학습자 입력에 대한 깊은 이해와 적절하고 구체적인 응답을 생성하기 위해서는 대용량의 파라미터로 학습된 대화 모델의 사용이 중요하다. 질적 분석에서 살펴 보았듯이 기존 상용화된 교육 플랫폼들은 입력에 대한 인식 능력이 부족하여 대화의 주제를 제한하거나 단어 검색과 같은 챗봇의 기능을 제대로 활용할 수 없었다. 블렌더봇의 경우, 기존 챗봇 모델보다 파라미터 크기를 키워 학습한 모델이다. 기존 챗봇과의 휴먼 평가를 진행했을 때 블렌더봇의 대화가 가장 매력적이라고 느꼈으며 더 사람 같은 대화를 생성한다고 답했고[23], 다른 교육 플랫폼에 비해 깊은 이해가 가능하고 구체적인 대답을 생성할 수 있다. 그러므로, 현재 상용화된 교육 플랫폼의 대화 모델은 대용량의 파라미터로 학습된 모델

로 개선되어야 한다. 깊은 이해를 바탕으로 구체적이고 더 사람 같은 대화를 통해 학습자는 더욱 대화에 몰입할 수 있고, 흥미를 유발하고, 동기를 부여할 수 있다. 이는 곧 더 많은 대화를 나누게 되어 학습 효과가 증가하게 된다.

5.2. 교육 데이터 구축

현재까지 교육에 특화된 대화 데이터셋이 존재하지 않았기 때문에 교육 목적이 챗봇을 구현하는데 어려움이 따랐다. 또한, 이전까지 영어 교육을 위한 챗봇 연구는 인공지능 스피커를 활용한 연구가 주를 이뤘다 [13-15]. 인공지능 스피커는 교육 목적으로 제작된 챗봇이 아니기에 이를 교육용으로 활용하기에는 어려움이 따른다. 하지만, 교육용 기반 대화 데이터를 구축하여 이를 활용한 대화 모델이 제작될 경우, 교육에 특화된 언어 학습이 가능하다.

교육 대화 데이터셋의 예시로는 상황별 대화로 구성된 데이터셋을 예시로 들 수 있다. 상황별 영어 회화 학습이란 일상생활에서 발생할 수 있는 여러 상황별 기본 영어 회화와 실생활에 관련된 영어 학습을 지칭하는 것이다. 예를 들어, 공항에서 환전하는 상황, 병원에서 증상 물어보기와 같이 미리 주어진 실생활의 상황에서 회화 연습을 통해 학습을 진행한다. 스피카우의 시나리오 기반 대화 기능이 이에 해당한다. 하지만, 스피카우의 자유 대화 기능과 달리 자유로운 대화가 불가능하고 주어진 답변을 선택해서 대답해야 한다. 본 논문은 스피카우와 같이 제한된 대화가 아닌 특정 상황에서 자유로운 대화가 가능한 교육 데이터셋을 제안한다. 다양한 상황에 대해 멀티 턴 (Multi-turn) 대화 데이터셋 구축 후 이를 대용량의 파라미터로 학습된 대화 모델로 학습한다면 상황별로 회화 학습이 가능한 대화 모델을 구축할 수 있다.

5.3. 정보 검색 기능

블렌더봇 2.0이 기존 대화 모델과 가장 큰 차이점은 인터넷에 있는 정보를 검색해 최신 정보에 관한 대화가 가능하다는 점이다. Fig. 4와 같이 지식 베이스에서 벗어난 정보와 단어에 대해 인터넷 검색을 통해 답변을 생성한다. 이를 활용한다면 학습자는 단어검색을 이용하지 않고 학습 중에 궁금한 단어를 대화상으로 바로 물어볼 수 있고, 더욱 자연스럽게 효과적인 학습을 구성

할 수 있다. 또한, 학습과 관련된 강의 동영상, 기사, 신문, 책의 내용 정보 검색 기능을 통해 학습자에게 제공하면 더 다양한 주제에 관해 대화가 가능해져 동기 부여와 흥미를 유발할 수 있다.

6. 결론

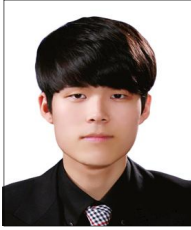
인공지능을 활용한 교육 (AIED)는 다양한 인공지능 기술과 결합하여 기존의 전통적인 학습보다 더 효과적인 학습이 가능하다. 그중 대화형 에이전트는 시간과 공간의 제약을 받지 않고 자유롭게 학습이 가능하고, STT, TTS, GEC 등 다양한 기술들과 결합을 할 수 있다는 점에서 기대를 모았다. 하지만, 현재 대화형 에이전트가 사용된 교육용 플랫폼들은 형태만 대화형 에이전트 모습을 띠고 있었다. 제한된 주제에 관해서만 대화할 수 있고, 답변이 미리 정해져 있거나 구체적이지 못하다. 이러한 문제점은 학습자에 효과적인 학습을 제공하지 못할 뿐만 아니라 학습의 저하로도 이어질 수 있다. 본 논문은 비교 분석을 통해 문제점과 한계점을 파악하고 개선점을 제안하기 위해 현재 시점에서 가장 최근에 공개된, 대규모 사전학습 기반의 대화 모델인 블렌더봇 2.0과 대화형 에이전트가 활용된 교육용 플랫폼에 대한 SSA 휴먼 평가 실험을 진행하였다. 정적 평가의 경우 SSA 점수가 블렌더봇, Andy, 스피카우 순으로 나타났고, 대화형 평가에서는 블렌더봇, 스피카우, Andy 순으로 나타났다. 이후 질적 분석을 진행하였고, 결과는 다음과 같다. 교육 플랫폼의 대화형 에이전트는 지식 베이스에서 벗어난 대화가 어려워 제한된 대화로 이를 유도했다. 또한, 대답이 구체적이지 못한 문제점이 존재했다. 하지만, 블렌더봇은 대답이 구체적이고 지식 베이스에 없는 정보는 인터넷 검색을 통해 답변을 생성할 수 있었다. 최종적으로, 분석한 문제점과 한계점을 바탕으로 교육 플랫폼에 대화형 에이전트를 효과적으로 활용할 수 있는 개선방안을 제안했다. 개선방안으로는 대용량의 파라미터로 학습한 대화 모델, 교육용 대화 데이터 구축, 정보 검색 기능의 필요성을 제안했다. 추후, 본 논문을 바탕으로 교육용 대화형 에이전트의 문제점을 일부 개선한 대화 모델과 교육 대화용 데이터셋 구축하는 후속 연구를 진행할 예정이다.

REFERENCES

- [1] Z. Ruttkey & C. Pelachaud. (Eds.). (2006). *From brows to trust: Evaluating embodied conversational agents*, 7, Springer Science & Business Media.
- [2] M. Lim. (2020). A Study on the Direction of Technology Education in the Age of Artificial Intelligence. *Journal of Korean Practical Arts Education*, 33(4), 81-102.
- [3] L. Fryer & R. Carpenter. (2006). Bots as language learning tools. *Language Learning & Technology*, 10(3), 8-14.
- [4] N. Y. Kim, Y. Cha & H. S. Kim. (2019). Future English learning: Chatbots and artificial intelligence. *Multimedia-Assisted Language Learning*, 22(3), 32-53.
- [5] J. Li & D. Jurafsky. (2016). Neural net models for open-domain discourse coherence. *arXiv preprint arXiv:1606.01545*.
- [6] B. Chantarotwong. (2006). The learning chatbot. *Final year project.[Online]*: <http://courses.ischool.berkeley.edu/i256/f06/projects/bonniejc.pdf>.
- [7] D. Lee. (2018). A study for the development of an English learning chatbot system based on Artificial Intelligence. *Secondary English Education*, 11(1), 45-68.
- [8] S. Zhang, et al. (2018). Personalizing dialogue agents: I have a dog, do you have pets too?. *arXiv preprint arXiv:1801.07243*.
- [9] S. Roller, et al. (2020). Recipes for building an open-domain chatbot. *arXiv preprint arXiv:2004.13637*.
- [10] J. Jia & M. Ruan. (2008, June). Use chatbot csiec to facilitate the individual learning in english instruction: A case study. In *International conference on intelligent tutoring systems* (pp. 706-708). Springer, Berlin, Heidelberg.
- [11] J. In & J. Han. (2016). The prosodic changes of Korean English learners in robot assisted learning. *Journal of The Korean Association of Information Education*, 20(4), 323-332.
- [12] N. Y. Kim. (2019). A study on the use of artificial intelligence chatbots for improving English grammar skills. *Journal of Digital Convergence*, 17(8), 37-46.
- [13] Y. Kim. (2020). Analysis of chatbots and chatbot builders for English language learning. *Multimedia-Assisted Language Learning*, 23(4), 161-182.
- [14] J. E. Hyun & H. J. Im. (2019). Analysis and Implications of AI Speakers as English Learning Tools. *The Journal of Mirae English Language and Literature* 24(1), 201-219.
- [15] H. Kim & D. Shin. (2019). A study of AI chatbot as an assistant tool for school English curriculum. *The Journal of Learner-Centered Curriculum and Instruction*, 19(1), 89-110. DOI ; 10.22251/jlcci.2019.19.1.89
- [16] K. Teske. (2017). Duolingo. *Calico Journal*, 34(3), 393-402.
- [17] J. Jang & S. Ye. (2021). Towards Continual Knowledge Learning of Language Models. *arXiv preprint arXiv:2110.03215*.
- [18] M. Komeili, K. Shuster & J. Weston. (2021). Internet-augmented dialogue generation. *arXiv preprint arXiv:2107.07566*.
- [19] J. Xu, A. Szlam & J. Weston. (2021). Beyond goldfish memory: Long-term open-domain conversation. *arXiv preprint arXiv:2107.07567*.
- [20] L. Floridi & M. Chiriatti. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681-694.
- [21] T. Yoon & S. Lee. (2021). Effects of Primary ELLs' Affective Factors and Satisfaction through AI-based Speaking Activity. *The Journal of the Korea Contents Association*, 21(9), 34-41. DOI : 10.5392/JKCA.2021.21.09.034
- [22] J. Sedoc, et al. (2019, June). ChatEval: A tool for chatbot evaluation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)* (pp. 60-65).
- [23] S. Roller, et al. (2020). Recipes for building an open-domain chatbot. *arXiv preprint arXiv:2004.13637*.

이 승 준(SeungJun Lee)

[학생회원]



- 2021년 2월 : 한국외국어대학교 산업경영공학과 (공학사)
- 2021년 7월 ~ 현재 : Human-inspired AI 연구소

- 관심분야 : Natural Language Processing, AI in Education, Text-Mining
- E-Mail : dzy6505@gmail.com

박 찬 준(Chanjun Park)

[학생회원]



- 2019년 2월 : 부산외국어대학교 언어처리창의융합전공 (공학사)
- 2018년 6월 ~ 2019년 7월 : SYS TRAN Research Engineer
- 2019년 9월 ~ 현재 : 고려대학교 컴퓨터학과 석박사통합과정

- 관심분야 : Data-Centric AI, Machine Translation, Grammar Error Correction, Deep Learning
- E-Mail : bcj1210@naver.com

서 재 형(Jaehyung Seo)

[학생회원]



- 2020년 8월 : 고려대학교 영어영문학과 및 경영학과(문학사, 경영학사)
- 2020년 9월 ~ 현재 : 고려대학교 컴퓨터학과 석박사통합과정
- 관심분야 : Graph Encoder, Commonsense Reasoning
- E-Mail : seojae777@korea.ac.kr

임 희 석(Heuseok Lim)

[종신회원]



- 1992년 : 고려대학교 컴퓨터학과 (이학학사)
- 1994년 : 고려대학교 컴퓨터학과 (이학석사)
- 1997년 : 고려대학교 컴퓨터학과 (이학박사)

- 2008년 ~ 현재 : 고려대학교 컴퓨터학과 교수
- 관심분야 : 자연어처리, 기계학습, 인공지능
- E-Mail : limhseok@korea.ac.kr