

초거대 언어모델을 기반으로 한 AI 대화 모델1)

이경님(LG유플러스)·조은경(서강대학교)

1. 서론: 초거대 AI 시대 언어 연구

미래학자 레이 커즈와일은 책 [특이점이 온다(The Singularity is near)]에서 “기술 성장은 통제불가능하고 저항할 수 없으며, 인간 문명에 있어 예측 불가능한 결과를 낳고 있다. <중략> AI의 발전이 기하급수적으로 이뤄져 인간의 인지능력을 넘어서는 순간에 기술 특이점(Technological Singularity)²⁾에 이르고 이는 곧 슈퍼(Super) AI 시대이다”라 하였다. 그의 예측이 반드시 실현되리라는 것보다는 현대 사회의 기술, 문화, 그리고 사회 분야 리더들이 그의 목소리에 귀를 기울이고 있다는 점에서 AI의 발전, 특히 지극히 인간적인 특징인 언어 지능의 구현을 목표로 하는 AI 모델로서 초거대 AI 언어 모델을 살피는 것은 우리에게 의미있는 일이다.

초거대 AI 시대는 특정 분야에 제한되지 않고 다양한 분야에 확장 가능한 패러다임으로 대전환을 특징으로 한다. 특히 AI 언어모델은 미리 정의된 답변을 학습하지 않아도 어떤 답변에도(Open-domain) 사람처럼(Human-like) 반응하고, 인간이 가진 상식 대화가 가능한 수준으로 소통하고 공감하는 AI로의 진화를 가능하게 한다는 목표를 가지고 있다.

사람은 일상생활에서 인식, 추론, 예측 등 범용적인 지능(general intelligence)을 사용하고, 이를 언어로 표현하며 사람뿐만 아니라 모바일 디바이스 같은 다양한 전자 기기에도 언어로 상호작용하게 되었다. 이러한 삶의 양상은 디지털 환경 의존도가 큰 오늘날의 사람간의 언어적 소통 양상뿐만 아니라 사람과 기기 간의 언어적 소통 양상에도 잘 드러나고 있지만, 이에 대한 언어 분석적 연구는 드문 감이 있다.

이 발표는 초거대 AI 시대에 언어 기술(Language Technology)로서 언어 모델(Language models)과 언어적 상호작용 서비스로서 질의응답(QA) 및 대화형 AI 서비스와 이를 구성하는 주요 요소들을 살펴보고, 대화형 AI 서비스의 현황과 전망을 지능형 에이전트, AI 키오스크, 디지털 휴먼 등을 통해 짚어본다. 이를 통해 초거대 언어 모델 및 AI 대화 모델이 가진 언어 및 대화 모델 개발의 잠재적 가치를 찾기를 기대한다.

1) 이 논문 또는 저서는 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2021S1A5A8065237).
2) 공상과학 소설가이자 컴퓨터학 교수였던 베너 빈지(Vernor Steffen Vinge)가 대중화한 개념이다. 기술 분야에서 특이점(singularity)란 개념을 처음 쓴 사람은 존 폰 노이먼(John von Neuman)이다.

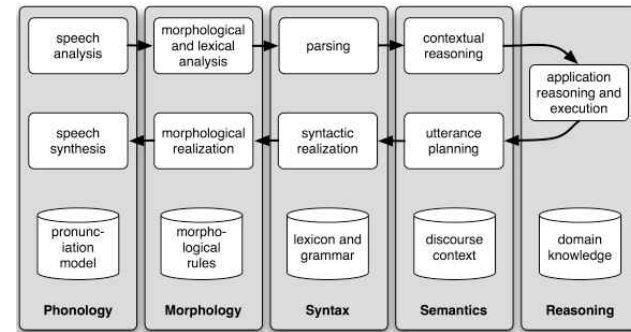
2. 인공지능 언어 모델 성장과 초거대 AI 모델의 등장

2.1. 언어 모델과 인공 지능

언어 모델(Language models)이라는 말은 연속 단어의 확률 분포를 말한다.(Jurafsky & Martin 2021: 3장 N-그램 언어모델) 언어 정보 처리에서 언어 모델이란 용어가 널리 쓰이게 된 것은 통계적(Stochastic)이고 기계 학습(Machine learning)이라는 양적인 접근에 의존한 방법으로 단어, 문장 구조, 의미, 화용 정보를 획득하여 활용하는 일이 많아지면서이다.

최근 인공지능 연구의 주된 방법론인 딥러닝의 개발 및 데이터 규모의 성장 가속화에 따라 사람의 말처럼 디지털 장비가 구사하게 하고 사람의 이해처럼 디지털 장비가 이해할 수 있게 하는 목적의 인공지능인 언어 지능(linguistic intelligence)도 인공지능 분야의 중요 세부 분야가 되었다.

딥러닝 방법론의 대유행 이전에 그러나 통계적 기계학습이 대세가 된 이후까지 언어 정보 처리는 대개 <그림 1>과 같은 파이프라인 구성이었고, 각 구성 부문에서 음성 분석/합성, 형태 분석, 구문 분석, 응답 계획 등에서 각각의 언어 모델을 학습하여 썼다.

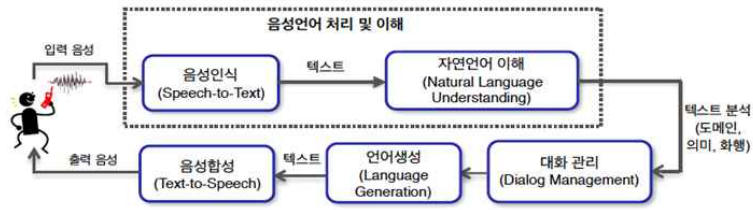


<그림 1> 대화시스템에서 언어 정보 처리 파이프라인

출처: Bird et al. 2009:32, Figure 1.5

예를 들면, 자동 음성인식(ASR; automatic speech recognition) 시스템은 통계적 패턴 인식을 기반으로 하며, 시스템 구성을 위한 요소로는 크게 음향 모델(AM; acoustic model), 어휘 모델(lexical model), 언어 모델(LM; language model), 인식을 위한 탐색 알고리즘에 해당하는 디코더(decoder)로 세분화된다. 특히, 한국어 대어휘 연속 음성인식(LVCSR)을 수행할 때 한국어 음성 인식 단위(recognition unit) 선정과 발음 변화(pronunciation variation) 현상을 고려해야 한다.(이경님 2007:요약문)

2015년을 즈음하여 딥러닝의 대유행 이후 특히, 형태, 의미와 구분 정보를 전통적 방식으로 두지 않는 종단 처리(End-to-End) 방식의 AI 언어 모델을 탑재한 응용 시스템이 인공지능 분야의 대세로 잡았다.



<그림 2> AI 언어모델에 기반한 대화 시스템

출처: 그림1, 음성 대화 인터페이스 시스템 구성 (이경님 2017)

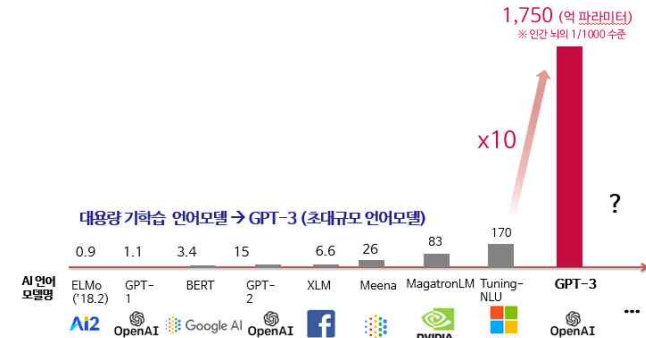
텍스트 언어 정보 처리에 제한을 두지 않는다면, 대화형 음성 언어 AI의 세부 분야로는 화자식별/화자검증, 음성 감정인식 (실시간 감정 파악), 대화체 감성표현 음성합성, 음향신호분류/잡음처리/원거리 인식 등도 있다.

2.2. 인공지능 연구의 재도약과 언어 모델의 성장

인공지능이란 분야는 1956년 다트머스 워크숍(Dartmouth workshop)에서 제안되었고 여기에는 존 매카시, 마빈 민스키, 클라우드 샤넌 등의 저명 학자들이 참여하였다. 이후 20년 후반 인공지능의 암흑기를 거치고 오늘날 인공지능이라 불리는 분야 및 기술은 딥러닝을 포함한 기계학습 방법에 의해 주도되고 있다.

딥러닝 방법론에 기반한 AI 알고리즘의 혁신은 하드웨어의 비약적인 성능 향상과 폭발적인 데이터의 증가로 인해 가능하게 되었다. 이로 인해 우리 사회는 우리의 삶을 바꿀 수 있는 여러 가능성이 생기고 발전을 가속화시키고 있다고 보고 학계뿐만 아니라 산업계에서도 많은 투자와 연구 개발을 수행하고 있다.

딥러닝 방법론은 2006년 제프리 힌튼 교수의 발표 이래 학계에서 활발하게 연구되어 왔고 이후 하드웨어 및 데이터 규모의 성장과 함께 연구개발이 가속화되어 왔다. 주로 학계에서 연구개발로 진행되던 딥러닝 방법의 인공지능 기술이 산업계에서 적극 도입하게 된 계기는 2016년 구글 딥마인드의 알파고가 이세돌 9단과의 바둑 대국에서 4:1로 불계승한 일이다. 이 일을 계기로 이후 학계에서 더 활발한 연구가 진행되게 되었고 특히 언어 지능면에서도 엘모(ELMO), 지피티(GPT-1), 버트(BERT), 지피티투(GPT-2), 지피티트리(GPT-3), 티파이브(T5) 등 매우 큰 언어 데이터를 학습한 트랜스포머 방식으로 학습한 언어 모델이 속출하고(<그림 3>) 이러한 방법론에 기반한 한국어 언어 모델도 나오고 있다.



<그림 3> 언어 모델 개발의 진화

트랜스포머(Transformer) 모델은 인코딩(encoder)-입력처리부와 디코딩(decoder)-출력처리부의 병렬화를 통해 계산 효율성면에서의 향상뿐만 아니라 기존의 재현신경(RNN:Recurrent Neural Networks)의 순차적으로 단어를 학습하는 알고리즘의 기술기소멸(vanishing gradient)의 한계를 극복한 모델로 평가받고 있다. 트랜스포머 모델의 중점 사항을 나열하면 다음과 같다.

- 1) 사전학습 모델링(Pre-trained modeling)으로 범용 모델(general/world model)을 목표로 하며, 기본적인 훈련 방식은 자기 지도 학습(Self-supervised learning)이다.
- 2) 범용 모델을 기반으로 하여 정밀 학습(Fine tuning) 단계에서 하위 과업(downstream task)으로서 적은 양의 레이블링된 데이터를 지도 학습(Supervised learning)하여 높은 성능의 분류 결과를 낸다. 예) 텍스트 요약, 개체명 인식, 기계번역 등
- 3) 피엘엠(PLM: Pre-trained Language model)의 대표적인 모델은 버트(BERT: Bi-directional Encoder Representation from Transformer)³⁾와 지피티트리(GPT-3: Generative pretraining transformer 3)⁴⁾이다.

2.3. 초거대 AI 언어모델의 등장과 특징

초거대 AI라는 용어는 오픈에이아이(OpenAI)⁵⁾의 지피티트리(GPT-3)⁶⁾가 나오면서부터 쓰이게 되었다. 초거대 인공지능 언어 모델은 대용량 데이터와 슈퍼컴퓨팅 인프라를 활용, AI 언어 모델의 규모를 수천억~수조개 매개변수(파라미터)로 확장한 기술이다. 이는 직전의 딥러닝 언어 모델 학습이 대규모 데이터를 학습한 사전 학습/훈련 언어모델

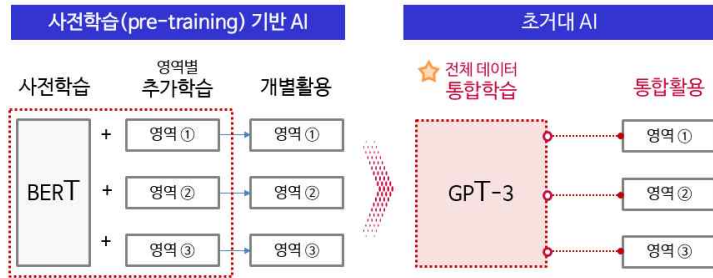
3) Devlin, Jacob et al.(2018)

4) Tom B. Brown et al.(2020)

5) <https://openai.com/>

6) Tom B. Brown et al.(2020), Language models are few-shot learners, <https://arxiv.org/abs/2005.14165>

(pre-training language models)에 기반하여 응용 영역별 데이터를 정밀 학습(fine-tuning)하는 두 단계 언어 모델 개발 방식에서 나아가 초거대 데이터를 학습한 언어 모델 하나를 가지고 추가 정밀 훈련없이 그 어떤 서비스에서도 활용하자는 것이다. (<그림 4>) 요컨대, 초거대 언어 모델로 학습된 범용 모델을 기반으로 실행 단계에서 샘플 제공만으로 영역별 과업을 수행하는 것이다.



<그림 4> 초거대 AI 언어 모델로의 변화

초거대 AI 언어 모델의 모체가 된 지퍼티트리(GPT-3)의 발표는 최근 국내 주요 기업들(LG, SKT, KT, NAVER, Kakao 등)도 초거대 AI를 추진하게 되는 계기가 되었다. 과학기술정보통신부가 2021년 9월 7일에 개최한 제1회 'AI 최고위 전략대화 (AI Strategy Summit)'는 최근 기업의 인공지능 연구개발 추진 현황을 공유하였고 초거대 언어 모델 개발이 가시화되었다. <그림 5>에 나열된 기업체들은 모두 인터넷 사업자들로 서비스 사용자들이 남긴 수많은 데이터를 보유하고 있고 관련된 서비스 개발 및 유지를 위한 인적 물적 인프라가 갖추고 있다.

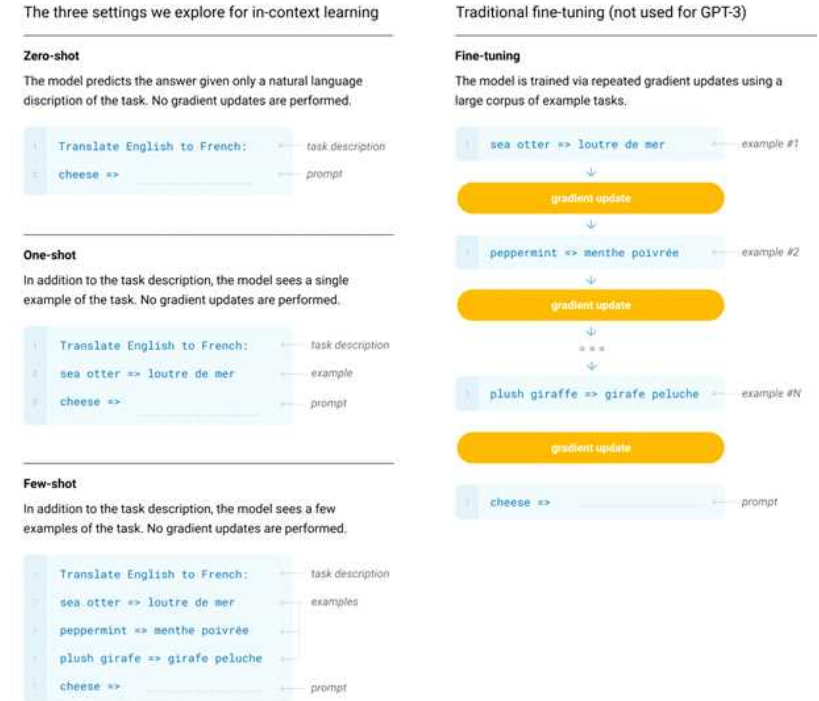
< 주요 기업들의 초거대 인공지능 추진 현황 >

NAVER	대용량 데이터 처리를 위한 슈퍼컴퓨터(700PF) 도입('20.10) 2,040억개 매개변수 규모의 AI 언어모델 (HyperCLOVA) 공개 및 상용화('21.5)
SKT	국립국어원과 AI 한국어모델 'GLM' 개발협력('21.4) 초기모델을 활용한 'AI 언어능력평가대회' 추진('21.9~11)
LG	전문가 수준의 지능을 보유한 초거대 AI 개발에 3년간 1억달러 투자('21.5) 6,000억개 매개변수 규모의 텍스트·이미지 동시 학습 초거대 AI 하반기 공개 추진
KT	ETRI, KAIST, 한양대와 '초거대 AI' 공동연구, 내년 상반기 상용화 목표('21.8) 영상로봇 등의 복합인지, 이미지 기반 해석, 휴머니스틱 AI 등 영역 확장 추진
Kakao	'빅모델 TF' 구성, 대규모 학습을 위한 인프라 구축 및 민간협력 검토 중 산·학·연 협력 기반 기술 개발·사업화를 추진하는 'AI 클라우드 멤버십' 구축 검토

<그림 5 > 과학기술정보통신부 인공지능 최고위 전략대화 개최 보도 자료

초거대 AI 언어 모델은 정밀 훈련 학습(fine tuning training)없이, 곧바로 어떤 과업을

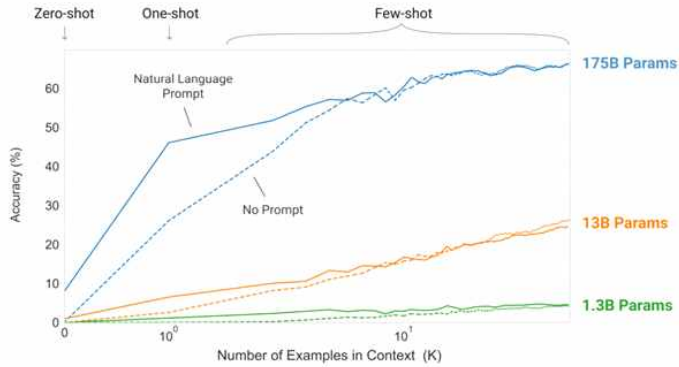
주고 답을 찾는 방식으로 곧바로 활용된다. 추가없는 학습(Zero-shot learning), 예문 하나 추가 학습(One-shot learning), 예문 몇 개 추가 학습(Few-shot learning) 등이 그 세가지 유형이고 이때의 학습(learning)은 많은 양의 데이터를 추가로 훈련(training)해야 하는 학습이 아니라 과업 설명(task description)과 제시문(prompt), 그리고 예문(example)이 주어지면 곧바로 답을 내는 방식이다. (<그림 6>).



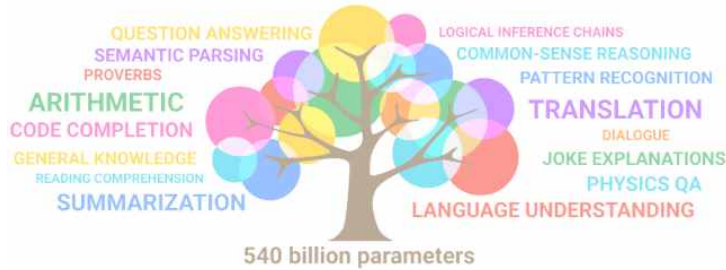
<그림 6> 초거대 언어 모델을 바로 활용하는 세 가지 과업 수행 방식(왼쪽), 초거대 모델을 쓰지 않는 이전의 정밀 훈련을 거친 후에 진행하는 방식(오른쪽)
출처: Tom B. Brown et al.(2020)의 Figure 2.1

이러한 초거대 언어 모델은 정밀훈련을 위한 데이터 레이블링 및 프로그래밍을 요구하지 않고 바로 언어 생성, 번역, 검색, 기사 요약 등을 수행하게 한다.

초거대 언어 모델의 정확도 성능은 몇 개 과업 예문이 주어졌을 가장 잘 수행하는 것으로 평가되었고(<그림 7>), 어떤 과업에서는 전통적인(?) 방식의 정밀훈련을 거친 때보다 초거대 언어 모델 기반의 예문 몇 개 추가 학습(few-shot learning)이 나은 성능을 보인다고 보고된 일도 있다.



<그림 7> 초거대 언어 모델 GPT-3의 파라미터 수와 제시예문 수에 따른 향상



<그림 8> 구글의 초거대 언어 모델 PaLM의 최대 파라미터 수와 응용 언어 처리

오픈에이의 GPT-3 이후 구글에서도 초거대 언어 모델 PaLM을 발표하였고⁷⁾ 더 큰 파라미터 개수 응용 과업을 제시하였다. (<그림8>) 그리고 국내에서도 최근 네이버에서 초거대 언어 모델로 하이퍼클로버를 시범서비스한 바 있다.(<그림 15>)

3. 초거대 AI 언어모델과 함께 진화하는 대화형 인터페이스

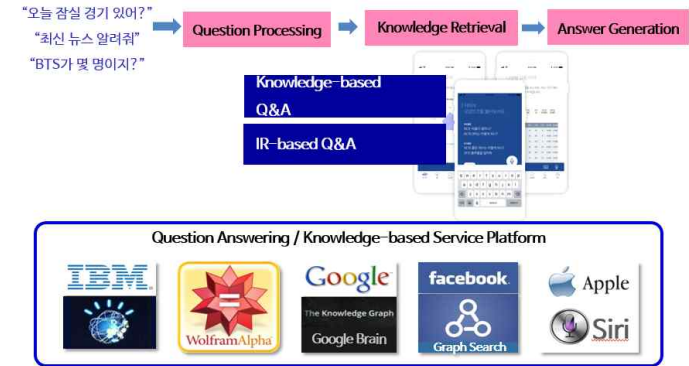
3.1 대화형 AI: 대화 시스템, AI 에이전트, 챗봇

대화형 AI는 사람과 기계 에이전트의 대화 방식에 따라 질의응답형(Question Answering), 과업해결형(Task completion), 사교대화형(Social Chat) 등으로 나뉜다.

7) Chowdhery, Aakanksha et al.(2022), PaLM: Scaling Language Modeling with Pathways. <https://arxiv.org/abs/2204.02311>

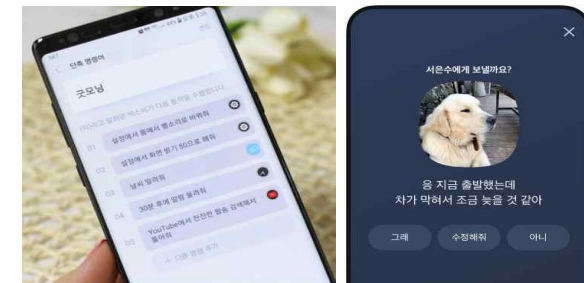
(Gao et al. 2019:6)

대화형 AI의 가장 성공적인 정작은 질의응답형 대화 시스템으로, 구글은 지식그래프와 연계하였고, 애플은 외부 지식엔진인 울프람알파와 연동하였고, 네이버나 카카오는 자사의 검색 지식 데이터와 연계하여 자리잡은 음성 대화형 AI 서비스이다. (<그림 8>)



<그림 9> 지식 기반 질의응답형 대화 시스템

최근에는 AI 스피커 출시와 함께 음성 대화형 서비스가 과업해결형(Task completion) 서비스로도 확장되었다. 예를 들면 ‘빗소리 들려줘’, ‘비오는 날 노래 추천해줘’, ‘OO에게 문자 보내줘’, ‘OO에게 전화 걸어줘’ 등과 같은 요청 발화에 대응하는 행위를 디바이스가 해주는 방식으로도 자리잡았고, 이러한 생활밀착형 음성 대화 서비스는 일상 잡담을 주고 받는 사교대화형 서비스로도 확장되었다. (<그림 9>) 사교대화는 더 자연스러운 발화, 더 맥락에 맞는 응답이라는 언어 사용성이 드러나기에 거대규모 언어모델에 기반한 대화 모델 연구 개발에 견인차가 되었다.



<그림 10> 과업 해결형 음성 대화 인터페이스

대화형 AI(대화 시스템)의 인터페이스 변화는 <표 1>과 같이 요약될 수 있다.

<표 1> 대화 시스템의 과거와 현재

	초기 대화 시스템	현재 대화 시스템
대표 사례	<ul style="list-style-type: none"> • 피씨 기반의 텍스트 및 음성 대화 시스템 • 챗봇 (인간 대화 시뮬레이션) • 신체 구현 대화 에이전트 (얼굴 표정, 몸 자세, 손 제스처, 음성 대화를 포함하는 애니메이션 캐릭터) 	<ul style="list-style-type: none"> • 메시징 플랫폼 상의 대화 시스템 (SMS, chatbot) • 스마트폰의 대화 시스템 (예. Apple's Siri, Microsoft's Cortana, Google Assistant, Samsung's 빅스비 등) • 스마트 스피커 및 기타 디바이스의 대화 시스템 (예. Apple CarPlay, Google Android Auto; Nuance Dragon Drive, SKT 누구, Kakao 헤이카카오 등)
한계 및 차이점	<ul style="list-style-type: none"> • 사용자 발화와 대화 시스템의 예상 입력과의 편차가 있었음. • 대화 계획 같은 자연어 이해 대화 모델 이론을 기반으로 함. • 시스템 실제 목적에 맞게 잘 작동하지만, 확장은 되지 않음. 다른 도메인으로 쉽게 이동하거나 이전할 수 없고, 대화 관리 결정은 수작업으로 진행됨. • 특수 플랫폼에 배포되었으며 다른 도메인으로 쉽게 이식하거나 다른 플랫폼에 배포할 수 없었음. 	<ul style="list-style-type: none"> • 여러 장치와 플랫폼에서 원활하게 작동함 • 디바이스가 위치, 건강 및 센서를 통해 획득했을 사용자 데이터 및 사용자 상황 정보에 액세스할 수 있음. 이를 통해 각 사용자에게 보다 개인화된 경험을 제공. • 특히 로봇 에이전트에는 다중 모드 상호 작용 기능이 있으며 시선 신호, 몸짓, 끄덕임 및 신체 자세를 효과적으로 분석할 수 있음. • 대화 시스템의 개발 및 배포를 촉진한 다양한 기술 동인이 있었음. (음성인식 향상, 대규모 언어 모델에 의한 자연어이해 수준 향상)

3.2. 음성 대화 인터페이스의 확장

대화형 AI 서비스는 <그림 11>과 같이 우리의 실생활에 관련된 자동화 처리를 도와주는 에이전트로 구현되었다. 이는 물리적인 구현체를 제외하면 음성 대화 인터페이스를 확장한 형태이다. 최근에 많은 자영업 및 대고객 서비스 영역에서 키오스크 및 로봇 시스템의 확대가 되고 있기 때문에 자연스러운 발화에 대응하는 대화형 에이전트 개발은 더 확대될 것이다.



<그림 11> KB국민은행 AI 체험존(왼쪽), 국립국악원 전시 안내 로봇 '큐아이'(오른쪽)

음성 대화 인터페이스를 확장한 형태는 메타버스 시대의 대화형 디지털 휴먼으로도 구현되었다. 해외에는 일리노이 홀로코스트 박물관⁸⁾ 및 남가주 대학 쇼아 재단⁹⁾의 상호작용형 증언 시스템¹⁰⁾으로 구현된 것이 있다. 이는 사용자의 자연 발화에 대해 디지털 휴먼 및 인간의 영상을 시스템의 반응으로 제시하는 방식이다.

국내에서 음성 대화 인터페이스를 확장한 대화형 디지털 휴먼 제작 사례로는 “안중근 의사 AI 디지털 휴먼”과 ‘영원한 증언’ 프로젝트가 있다.(<그림 12>) 이런 개발 사례는 역사인물의 복원, 디지털 휴먼과의 실시간 대화에 중점을 두었고, 언어적으로 다양한 발화에 대한 응답을 주는 대화 모델보다는 역사적 사실성에 근거한 콘텐츠 개발이었기 때문에 안중근 의사를 구현한 경우는 역사적 근거가 있는 내용의 스크립트에 기반한 안정적 음성의 성우 녹음이 더 우선하였을 것이고, 일본군 위안부 할머니의 증언 대화 시스템은 신뢰성 있는 증언 발화의 수집과 연료한 생존 인물의 온전한 건강 상태에서의 증언 인터뷰 과정의 영상 확보가 우선시되었기에 진정한 인공지능 대화 모델 개발이라 보기에 한계가 있었고, 한정된 콘텐츠를 최대한 반응하게 하기 위한 사용자 발화의 다양한 유사 질문 입력 및 수집 과정이 있었다.



<그림 12>안중근 의사 AI 디지털 휴먼(왼쪽), 일본군 위안부 할머니 인터랙티브 증언(오른쪽)

출처: 왼쪽-<https://youtu.be/p1EJRILeY8>, 오른쪽-조은경,서정연(2021)

현재 <그림 11>과 같은 대화형 자동화 에이전트나 <그림 12>와 같은 디지털 휴먼의 개발 수준은 실시간 지원, 쌍방향 소통의 수준이긴 하나 미리 정해진 답변을 제공하는 방식으로 대화 시스템의 초기 형태인 목적 지향적 대화 시스템과 유사하다. 제한된 기능과 콘텐츠의 노출과 반응을 더 늘리기 위해서라도 AI 기술을 활용한 의도 분석, 문맥 정보 기반 대화 예측 등 컨텍스트를 고려한 연속 대화 기능의 연구 개발이 필요하다.

8) <https://www.ilholocaustmuseum.org/exhibitions/survivor-stories-experience/>

9) <https://sfi.usc.edu/>

10) <https://youtu.be/vE5jAqppDsI>

또 메타버스의 대유형과 함께, 가상 인물인 버추얼 인플루언서(Virtual Influencer)로 LG 전자 가상 인강 ‘김래아’, 신한라이프 광고 모델 ‘로지’가 소개된 바 있고, 미래의 가상 유투버로 브이튜버(V-Tuber) 루시(Lucy)¹¹⁾, 미래의 가상 게임 등장 인물(NPCs: non-player characters)¹²⁾가 소개된 바 있다. 이들은 모두 자연어 질의, 발화에 매우 적합해보이는 응답 발화를 한다. 부자연스러운 것은 응답 내용(텍스트)이라기보다는 음감이나 뉘앙스이다.

이후 일상이 될 상호 대화형 정보 접근 방식으로서의 활용은 제한된 발화에 반응하기보다는 구글의 Meena, LaMDA와 같은 생성형(Generative) 대화 모델의 도입을 통해 부족한 점을 극복해가야 한다. 즉, 스토리텔링(Story-telling)이 아닌 스토리리빙(Storyliving)이 될 것이다.

3.3. 인공지능 언어모델 기반의 대화형 AI

인공지능 언어 모델은 미리 입력된 답변을 제공하는 데서 벗어나, 스스로 생각하고 학습, 판단할 수 있도록 다양한 인공지능 과업(task)에 활용되게 하는 데에 그 목적이 있다. 대화형 AI는 인공지능 언어모델로 대화 맥락을 파악해 사람과 기기간 대화가 가능하게 하는 데에 그 목적이 있다.

대규모 언어 모델을 기반으로 특정한 과업 목적에 제한을 두지 않고 일반적인 대화 진행을 목적으로 하는 대화 모델 연구는 미나(Meena)¹³⁾, 람다(LaMDA)¹⁴⁾, 블렌더봇(BlenderBot)¹⁵⁾ 등이 대표적이다.

<그림 13>의 미나의 응답은 매우 자연스러운 인간의 발화처럼 보인다. 미나의 훈련 데이터는 공개된 소셜미디어 대화를 수집한 것이다. 미나는 초거대 언어모델이 나오기 전 대화 응답 생성을 위해 개발된 다이얼로그지퍼티(DialogGPT)¹⁶⁾와 비교하여 나은점을 여러 대화 사례를 통해 보여준다.

<그림 14>는 구글의 연례 행사인 Google I/O 2021 컨퍼런스에서 대표인 순다 피차이가 소개한 영상의 대화 예시이다¹⁷⁾. 이 대화에는 람다가 목성에 대해 답한 내용(‘매우 흥미 때문에 코트를 가지고 가야 한다’)은 대화 지식에 있는 응답 값이 아니라 목성이 태양으로부터 멀리 떨어진 행성으로 온도가 낮다는 개념을 학습했다는 점이다.

Conversation C	
Human:	<i>Haha, I think mathematics has a presence beyond human existence.</i>
Meena:	<i>So does philosophy.</i>
Human:	<i>Can you state with examples how this is true.</i>
Meena:	<i>Can you state how it isn't?</i>
Human:	<i>Without human existence there will be no one left to comprehend whats right whats wrong, and thus philosophy wi [sic]</i>
Meena:	<i>That's not the definition of philosophy.</i>
Human:	<i>Without human existence there will be no one left to comprehend whats right whats wrong, and thus philosophy will not be needed. How do you define philosophy?</i>
Meena:	<i>The study of general knowledge or the fundamental nature of things.</i>

<그림 13> 미나와 인간 대화 예시 (Adiwardana, et al.(2020))



<그림 14> 람다의 대화 예시: 오른쪽 검정 텍스트가 사람 발화. 출처: <https://youtu.be/ayhJii34D38>

11) 가상 유투버 루시와의 대화, <https://vimeo.com/507801358>

12) 게임 등장 가상 인물과의 대화, <https://youtu.be/jH-6-ZlgmKY>

13) Adiwardana, Daniel et al.(2020), Towards a Human-like Open-Domain Chatbot, <https://arxiv.org/abs/2001.09977>

14) Thoppilan, Romal et al.(2022), LaMDA: Language Models for Dialog Applications, <https://arxiv.org/abs/2201.08239>

15) Shuster, Kurt et al.(2022), BlenderBot 3: a deployed conversational agent that continually learns to responsibly engage, <https://arxiv.org/abs/2208.03188>

16) Zhang, Yizhe et al.(2019), DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation, <https://arxiv.org/abs/1911.00536>

17) <https://youtu.be/ayhJii34D38>

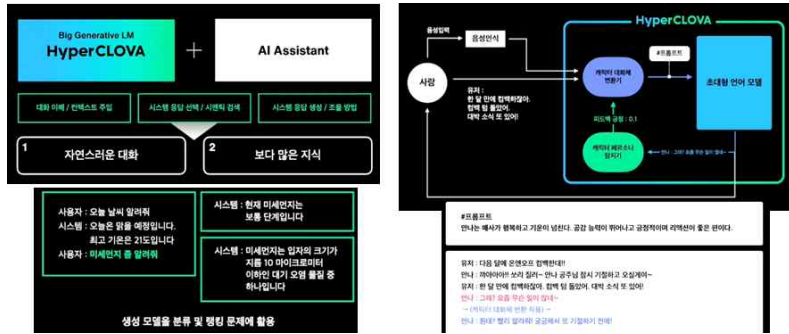
메타(Meta(구.페이스북))도 대규모 대화 데이터로 자사의 대화 데이터 및 기구축된 대화 모델 데이터(wizard of wikipedia)¹⁸⁾등의 데이터를 사전 학습하고, 좋은 지식 및 자연스런 응답 생성을 위해 정밀훈련한 대화 모델에 기반한 서비스로 블렌더봇3.0을 공개한 바 있다¹⁹⁾. 다른 대화서비스와 다른 점은 감성 대화를 잘 녹여내고, 대화 세션에 요약 기능이 있다는 것 등이다.

국내 대형 포털 네이버도 최근 초거대 언어 모델 하이퍼클로버(HyperCLOVA)를 공개하고 시범서비스를 하였다. (<그림 15>) 한국어 데이터는 GPT-3보다 상대적으로 훈련할 데이터가 적기 때문에 6500배 이상의 학습을 하여 2040억개의 파라미터로 구성된 언어모델이라고 한다. 초거대 언어모델이기에 GPT-3의 활용 사례처럼, 제시문(prompt)가

18) https://parl.ai/projects/wizard_of_wikipedia/

19) <https://ai.facebook.com/blog/state-of-the-art-open-source-chatbot/>

주어지고 이어지는 대화를 하는 방식으로 진행됨을 보인다.²⁰⁾



<그림 15>초거대 한국어 AI 언어 모델에 기반한 대화형 AI : 네이버 하이퍼클로바(HyperCLOVA) 시연 예시.

4. 발전 방향: 연결-융합-확장된 대화형 AI의 미래

한국과학기술기획평가원이 2021에 내놓은 2020년도 예비타당성조사 보고서 - ‘사람 중심 AI 강국 실현을 위한 차세대 인공지능 핵심원천기술개발사업’-에는 대화형 AI가 언어만을 다루는 것이 아님을 시사한다. (<그림 16>)

<표 3-22> [내역사업3] AI 활용성 개선 기술개발

중분류	추진분야 (대 상기술)	추진과제(9개)
소통하는 AI	단일감각지능 고도화	비디오데이터 기반 스토리 생성 기술 개발 소량 데이터를 사용하여 사투리와 비언어적 발화까지 인식하는 종단형 음성대화 원천기술
	복합대화 기술	사회성을 갖고 인간과 교감하는 멀티모달 인터랙션 AI 기술
	에이전트 간 협업 기술	AI 자율성장을 위한 멀티 에이전트 기반 복합지능 강화형 오토생성 기술
공감하는 AI	다중감각인지	AI 비대면화상회의 지원용 위한 다중감각 인지 기술 정각장애인의 의사소통 문제 해결을 위한 자연수어 번역 기술 질문하면서 성장하며 개별적인 개성을 갖는 에이전트 기술
	교감형 AI	인간의 정서를 이해하고 표현하며 자연어 이해를 통해 인간과 자연스러운 대화가 가능한 극사실 품질의 디지털 휴먼 개발
	행동지능	새로운 환경에 대한 대응 방법을 학습하고 향상시키는 AI 로봇 행동/소셜 지능 기술

<그림 16> 멀티모달 대화형 AI 개발과 개선

20) 초거대 한국어 모델에 기반한 대화형 AI 활용 방식, <https://tv.naver.com/v/20390101>

최근의 대화형 AI 서비스의 향후 연구 개발은 텍스트 언어, 음성, 이미지, 동영상 등이 융합 연결되어 온전한 멀티모달 상호작용형을 목표로 한다.

현재 텍스트 언어와 이미지의 멀티모달 연구 개발(Clip:Connecting Text and Images, Contrastive Language-Image Pretraining)²¹⁾이 한창 진행중이고 텍스트와 음성은 이미 진행되어 온 것이기에 소리, 그림, 문자의 융합은 멀지 않은 미래의 일이다. 이는 언어 정보의 처리 및 연구 개발이 텍스트에 제한되어 이뤄질 때의 한계는 여러 연구자들에 의해 지적되고²²⁾ 개선되고 있다.

또 한편의 인공지능 (언어) 모델 개발의 핵심 훈련 방식은 얀르쿤(Yann LeCun)이 꾸준히 주장해오고 있는 어떤 레이블이나 주석을 필요로 하지 않는 자기지도 학습(Self-supervised Learning)²³⁾이다. 말뭉치에 태깅을 하지 않고도 인간 수준의 언어 지식을 학습하려면 자기지도 학습 혹은 비지도 학습의 훈련 알고리즘에 어떤 언어적 통찰력을 넣을지 고민할 일이다. 자기지도학습의 초기 방식인 단어벡터(word-embedding)의 아이디어는 인접 단어는 그 개념적 뜻이 유사할 것이라는 언어학자의 가정에 기반해 있고, 최근의 대표적인 방식인 트랜스포머 버트(BERT)의 핵심 아이디어인 빈칸을 두었을 때 추론이 가능하다는 빈칸채우기(Cloze test)도 커뮤니케이션 연구 및 언어 능력 평가 도구로 널리 쓰이는 것이다.

참고문헌

김경선(2017), 인공지능 기반 언어 처리 기술 - 자연어 대화 인터페이스를 중심으로, 새국어생활 특집 4차산업혁명 시대의 국어 생활

김명남, 장시형 역.(2007) 특이점이 온다, 김영사. (Ray Kurzweil(2005), The Singularity is near. Viking.)

서강대산학협력단(2020), 2020년 일본군 '위안부' 피해자 증언의 실감형 인터랙티브 콘텐츠 제작 및 전시·교육 활용방안 연구, 사업보고서, 한국여성인권진흥원.

서강대산학협력단(2021), 2021년 인터랙티브 증언 콘텐츠 베타테스트 및 전시·교육 활용방안 연구, 사업보고서, 한국여성인권진흥원.

김학수(2017), 우리말 자연어 처리 기술 - 과거와 현재, 새국어생활 특집 4차산업혁명 시대의 국어 생활

이정남(2007), 한국어 대어휘 연속음성인식을 위한 형태·음운론적 발음 변화 모델링, 서강대 박사학위논문.

이정남(2017), 음성 언어 처리 기술, 어디까지 왔나, 새국어생활 특집 4차산업혁명 시대의 국어 생활

21) <https://openai.com/blog/clip/>, <https://github.com/openai/CLIP>

22) Bisk, Y. et al.(2020)은 언어 모델의 텍스트 기반 의미 표현(text-based meaning representation)의 한계를 다음과 같이 일관했다. 'Is knowledge about task X most richly encoded in?'

23) <https://www.youtube.com/watch?v=YzD7Z2yRL7Y&t=1678s>

이기창(2021), BERT와 GPT로 배우는 자연어 처리. 이지스 퍼블리싱.

이요안(2021), 손석희 대화분석, 박영사.

조은경,서정연(2021), Improving Educational Dialog System for Historical World Affairs, *Korean Journal of Applied Linguistics* 37, pp. 161-191

최재용(2017), 말뭉치와 언어학, 새국어생활 특집 4차산업혁명 시대의 국어 생활

한국과학기술기획평가원(2021), 2020년도 예비타당성조사 보고서 - 사람 중심 AI 강국 실현을 위한 차세대 인공지능 핵심원천기술개발사업, KISTEP.
https://www.kistep.re.kr/reportDetail.es?mid=a10305010000&rpt_no=RES0220210155

Vaswani, Ashish et al.(2017), Attention Is All You Need, 31st Conference on Neural Information Processing Systems (NIPS 2017),
<https://arxiv.org/abs/1706.03762>

Bird, Steven et al.(2009), *Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit*. O'Reilly. <https://www.nltk.org/book/>

Jurafsky, D. & James H. Martin(2021), (3rd ed.) *Speech and Language Processing*.
<https://web.stanford.edu/~jurafsky/slp3/>

Michael McTear(2021), *Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots*, Morgan publishers.

Gao, Jianfeng et al.(2019), Neural Approaches to Conversational AI Question Answering, Task-Oriented Dialogues and Social Chatbots,
<https://arxiv.org/abs/1809.08267>

Y. LeCun(2019), Deep Learning Hardware: Past, Present, and Future, 2019 IEEE International Solid-State Circuits Conference - (ISSCC), pp. 12-19, doi: 10.1109/ISSCC.2019.8662396. (발표 동영상: <https://youtu.be/YzD7Z2yRL7Y>)

M. Mitchell(2021), Why AI is Harder Than We Think,
<https://doi.org/10.48550/arXiv.2104.12871>

Tom B. Brown et al.(2020), Language models are few-shot learners,
<https://arxiv.org/abs/2005.14165>

Hinton, G. E.(2006), A fast learning algorithm for deep belief nets, *Neural Computation*, 18(7), 1527-1554. <https://doi.org/10.1162/neco.2006.18.7.1527>

Bisk, Y. et al.(2020), Experience Grounds Language, EMNLP Proceedings,
<https://arxiv.org/pdf/2004.10151>

Raffel, Colin et al.(2020), Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer, *Journal of Machine Learning Research* 21,
<https://arxiv.org/abs/1910.10683>

Yuan, Ann et al.(2021), SynthBio: A Case Study in Human-AI Collaborative Curation of Text Datasets, *35th Conference on Neural Information Processing Systems*, <https://arxiv.org/abs/2111.06467>

Bommasani, Rishi et al.(2022), On the Opportunities and Risks of Foundation Models, <https://arxiv.org/pdf/2108.07258>

Devlin, Jacob et al.(2018), BERT: Pre-training of Deep Bidirectional Transformers

for Language Understanding, <https://arxiv.org/abs/1810.04805>

Thoppilan, Romal et al.(2022), LaMDA: Language Models for Dialog Applications,
<https://arxiv.org/abs/2201.08239>

Chowdhery, Aakanksha et al.(2022), PaLM: Scaling Language Modeling with Pathways, <https://arxiv.org/abs/2204.02311>

Adiwardana, Daniel et al.(2020), Towards a Human-like Open-Domain Chatbot,
<https://arxiv.org/abs/2001.09977>

Shuster, Kurt et al.(2022), BlenderBot 3: a deployed conversational agent that continually learns to responsibly engage, <https://arxiv.org/abs/2208.03188>

Zhang, Yizhe et al.(2019), DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation, <https://arxiv.org/abs/1911.00536>

<<웹자료>>

과학기술정보통신부(2021), 제1회 인공지능 최고위 전략대화(AI Strategy Summit),
https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=112&pageIndex=8&bbsSeqNo=94&nttSeqNo=3180695&searchOpt=NTT_SJ&searchTxt=인공지능

구글 팜(PaLM) 블로그 포스트,
<https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html>

다트머스 컨퍼런스, https://en.wikipedia.org/wiki/Dartmouth_workshop

안중근 의사 디지털 휴먼 시연 영상, <https://youtu.be/p1jEJRILeY8>

홀로코스트 생존자 증언 대화 시스템 시연 영상, <https://youtu.be/vE5jAqppDsI>