

디지털 개인비서 동향과 미래

Trends and Future of Digital Personal Assistant

권오욱 (O.W. Kwon, ohwoog@etri.re.kr)	언어지능연구실 책임연구원
이기영 (K.Y. Lee, leeky@etri.re.kr)	언어지능연구실 책임연구원
이요한 (Y.H. Lee, carep@etri.re.kr)	언어지능연구실 연구원
노윤형 (Y.H. Roh, yhroh@etri.re.kr)	언어지능연구실 책임연구원
조민수 (M.S. Cho, mscho@etri.re.kr)	언어지능연구실 연구원
황금하 (J.X. Huang, hgh@etri.re.kr)	언어지능연구실 선임연구원
임수중 (S.J. Lim, isj@etri.re.kr)	언어지능연구실 책임연구원
최승권 (S.K. Choi, choisk@etri.re.kr)	언어지능연구실 책임연구원
김영길 (Y.K. Kim, kimyk@etri.re.kr)	언어지능연구실 책임연구원/실장

ABSTRACT

In this study, we introduce trends in and the future of digital personal assistants. Recently, digital personal assistants have begun to handle many tasks like humans by communicating with users in human language on smart devices such as smart phones, smart speakers, and smart cars. Their capabilities range from simple voice commands and chitchat to complex tasks such as device control, reservation, ordering, and scheduling. The digital personal assistants of the future will certainly speak like a person, have a person-like personality, see, hear, and analyze situations like a person, and become more human. Dialogue processing technology that makes them more human-like has developed into an end-to-end learning model based on deep neural networks in recent years. In addition, language models pre-trained from a large corpus make dialogue processing more natural and better understood. Advances in artificial intelligence such as dialogue processing technology will enable digital personal assistants to serve with more familiar and better performance in various areas.

KEYWORDS 디지털 개인비서, 목적지향 대화처리, 오픈도메인 대화처리, 질의응답, 챗봇

1. 서론

기계가 사람처럼 처리해 주는 모든 기능과 서비스를 총괄하여 말한다. 지능형 개인비서, 가상 개인비서 등이라고 일컬어지는 디지털 개인비서는 사람이 수행해야 할 업무를

* DOI: <https://doi.org/10.22648/ETRI.2021.J.360101>

* 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임[2019-0-00004, 준지도확습형 언어지능 원천기술 및 이에 기반한 외국인 지원용 한국어 튜터링 서비스 개발].



람처럼 언어로 소통하는 것을 가장 큰 특징으로 한다.

본 고에서 국내·외 디지털 개인비서가 사용자에게 어떤 서비스들을 제공하고 있는지에 대한 국내·외 서비스 동향과 디지털 개인비서의 언어 소통을 위한 대화처리 기술 동향을 기술하고자 한다. 또한 향후 디지털 개인비서가 가지게 되는 소통과 인지 기능을 통해 가능해질 서비스에 대해 전망하고자 한다.

II. 디지털 개인비서 서비스 동향

1. 국외 서비스 동향

애플 Siri가 2011년 처음 출시된 이후, 디지털 개인비서의 성능 및 서비스는 꾸준히 개선 및 확장되어왔다. eMarketer 시장조사에 의하면 2019년 미국 내 33%가 정기적으로 디지털 개인비서 기능을 쓰는 것으로 나타났고, Polaris Market Research에 따르면 2019년 약 25억 달러의 시장을 형성하고 있으며, 그 시장은 매년 성장 중이다. 현재 관련 제품 및 기업들의 동향을 살펴보면 다음과 같다.

스마트폰, 스마트 스피커뿐만 아니라 자동차, TV, 시계, 헤드폰, 안경 등으로 적용 형태가 확장되고 있다. 차량용 음성 비서 플랫폼으로는 애플의 Carplay, 구글의 Android Auto 등이 있으며, 많은 안드로이드 TV에서 구글 Assistant를 이용할 수 있다. 아마존은 2019년 Alexa를 탑재한 스마트안경을 선보였다.

그리고 기존 단순 음성검색이나 잡담대화 기능에서 기기제어, 예약, 주문 등의 TASK 수행과 복잡한 음성 명령을 수행하는 식으로 기능이 확장되고 있다. 구글 Assistant를 예로 들면, 영업시간, 교통정보, 길찾기 등의 정보검색, 일정관리, 문자/메일/전화, 음악/동영상 재생, 기기설정변경, 통·번

역 등의 작업을 한 번에 음성으로 처리할 수 있을 뿐 아니라, 특정 앱을 실행하여 실행결과를 응답으로 받을 수 있다. 예를 들어, “수현이한테 ‘가고 있다’고 문자 보내 줘”, “‘좋은 하루 되세요’를 독일어로 하면”, “Spotify에서 Discover Weekly 틀어 줘”, “진어에서 내일 인천 방콕 출도착 조회해 줘” 등이 있다.

또한, 기존의 단답식 대화방식에서 벗어나 대화의 상황과 맥락에 기반하여 사용자 의도를 이해하고, 이전의 대화내용을 기억하여 사용자의 선호도를 학습하고, 개인 기호에 맞는 맞춤형 서비스를 제공하는 등 이해 및 응답 성능이 지능화되고 있다. 그리고, 기계적인 응답이 아니고 좀 더 자연스러운 인터페이스를 제공하기 위한 시도도 이루어지고 있다. 2018년 대중에게 시연된 구글 Duplex는 딥러닝 기반 대화처리 기술을 사용하여 양방향 대화를 통한 스케줄 예약 및 조정 등 제한된 분야에서 인간다운 대화 및 TASK 수행 능력을 보여주었다.

시장에서 주도적인 디지털 개인비서 제품으로는 구글 Assistant, 애플 Siri, 아마존 Alexa, 삼성 Bixby, 마이크로소프트 Cortana, 바이두 Duer 등이 있고, 그 외에 DataBot, Lyra, Hound, Youper, Robin 등을 주요 제품으로 들 수 있다. 주요 제품들에 대해 인지 기능 및 서비스 측면에서 특징들을 보면 표 1과 같다.

표 1은 참고문헌 [1,2]의 내용을 정리한 것으로 인지 기능 및 서비스에서 특징적인 것만 표시했지만, 실은 대부분의 제품들에서 앞서 언급한 기본적인 특징들을 포함하고 있다. 디지털 개인비서의 인지 기능 수준은 4단계로 나눌 수 있다. 1단계는 단순 반응 방식의 챗봇 수준이며, 2단계는 NLP 및 Ontology 기반 검색 기술을 활용한 지능형 비서이다. 3단계는 IBM 왓슨과 같이 방대한 지식을 기

표 1 글로벌 디지털 개인비서 기능 비교[1,2]

제품	인지 기능	서비스/기능
Siri	사용자의 검색/언어/선호도에 적응	전화걸기, 메시지 보내기, 구두 점 추가, 리마인더 및 알람 등
구글 Assistant	방대한 데이터 및 고도의 NLP 기술 적용	질의응답, 스케줄 관리, 예약, 리마인더, 번역, 음악 재생, 길 찾기, 홈기기 제어
Alexa	사용자 음성 학습 기능	음악 검색, 전화걸기, 메시지 보내기, 쇼핑, 홈기기 제어
Cortana	자동 언어 인식, 디바이스 데이터에 의한 학습	시간/장소/사람에 따른 리마인더, 파일찾기, 배송조회, 태스크 관리
Bixby	사용자 루틴 및 휴대폰/앱 작동방식 학습, 개인화된 응답	콘텐츠 및 앱 조작 용이하도록 함, 번역, 리마인더
Hound	문맥 및 복잡한 발화 이해	자연스럽고 상세한 검색, 호텔 찾기, 길찾기, 주식시장 조회, 음악재생 등
Robin	제스처에 반응, 새로운 개념/phrase 학습	오디오 콘텐츠 플레이, 관심 정보 제공, 주차/교통 정보 제공
Youper	개인화, 기본추적	정서건강 관리, 대화, 명상 가이드

반으로 확률적 추론이 가능한 시스템이고, 4단계는 의식이 있는 가상비서라 할 수 있다[3]. 현재 대부분의 디지털 비서의 수준은 아직 1, 2단계의 낮은 수준이고, 음성언어이해 오류, 보안 및 프라이버시 문제 등 해결해야 할 문제들이 남아 있지만, 시장조사업체 스트래티지 애널리틱스(SA: Strategy Analytics)에서 “2023년 전 세계 90%의 스마트폰을 음성으로 제어하게 될 것”이라고 전망했던 것처럼, 디지털 비서의 생태계는 지속적으로 확장될 것이다[4].

2. 국내 서비스 동향

국내에서는 SK텔레콤의 스피커형 AI 개인비서 기기 ‘NUGU’에 이어 KT의 ‘기가지니’, 네이버의 ‘프렌즈’, 카카오의 ‘카카오톡미니’가 출시되었고, 금

용권과 공공기관에서는 텍스트 형태의 AI 챗봇 서비스가 도입되었다. 음악 선곡 및 감상, 날씨 및 교통 정보 검색 등 새롭고 재미있는 기능을 탑재한 AI 스피커는 초창기 대중에게 많은 주목을 끌었다. 이후 출시된 AI 스피커에는 정보를 시각화할 수 있는 디스플레이, 무드등, 셋톱박스와의 연동 등 기존 제품과 차별화된 기능을 탑재하였다. 하지만, 사람들은 음성 AI의 기술적인 문제와 실생활에서의 활용도 측면에서 AI 스피커의 한계를 느꼈으며, 그나마도 서비스를 이용하는 연령대가 디지털에 익숙한 젊은 고객층으로 편중되어 있는 문제점이 존재했다.

이러한 한계를 극복하고자 국내 기업들은 개인비서 서비스를 오픈 API 형태로 플랫폼화하여 스마트홈, 자동차 등 다양한 영역에서 개인비서 생태계를 확장하는 방향으로 나아가고 있다. 또한, 코로나-19 사태로 언택트 문화가 확산되고 사람들이 집에서 머무는 시간이 늘어나면서 지능형 개인비서 서비스에 새로운 패러다임이 불고 있다. 이러한 비대면 시대에 자녀를 둔 부모는 AI 스피커의 키즈 콘텐츠를 적극 활용할 수 있으며, 금융권에서는 챗봇 시스템이나 AI 콜센터 등을 이용하여 금융기관에 직접 방문하기를 꺼리는 고객에게 언택트 서비스를 제공할 수 있을 것이다.

국내의 지능형 개인비서 제품에 대해 알아보자면, 먼저 삼성전자는 음성 인식 플랫폼 ‘Bixby’[5]를 자체 개발하여 2016년 갤럭시 스마트폰에 처음 탑재하였다. 이후 세탁기, 공기청정기와 같은 자사가 전자기기에도 Bixby를 탑재하며 지능형 인터페이스로의 역할을 확대하였다. Bixby는 텍스트, 카메라, 터치, 음성 등의 다양한 입력 방식으로 지식 검색, 일정 관리, 은행 및 결제 서비스 등을 제공하고 있다.

2016년 SK텔레콤은 국내 최초이자 세계 2번째

로 AI 스피커 ‘NUGU’[6]를 공개하였다. 최근에는 스피커에 누구콜(NUGU call) 서비스를 연동하여, 음성으로 외부 번호를 검색하고 자동으로 전화를 걸 수 있는 기능을 선보였다. 또한, SK텔레콤은 NUGU 플랫폼을 Btv, T맵, T전화에 적용하는 등 홈, 자동차, 모바일 영역에서 지속적으로 접점을 늘리며 NUGU 플랫폼의 생태계를 다각화하고 있다.

KT의 AI 스피커 ‘기가지니’[7] 시리즈는 셋톱박스와 연동 가능하며, 음성 입력만으로 TV를 조작할 수 있다. 출시 당시 KT는 AI 스피커에 셋톱박스를 접목하는 새로운 전략을 보여 많은 관심을 받았다. 또한, 기가지니는 AI 영어학습, 핑크퐁 노래방 등 다양한 키즈 특화 학습서비스로 자녀를 둔 고객층의 눈길을 끌고 있다.

네이버는 2017년 AI 플랫폼 ‘클로바’를 탑재한 AI 스피커 ‘프렌즈’, ‘웨이브’ 출시에 이어 2020년 10월 책 읽어주기 기능이 탑재된 AI 스마트 조명 ‘클로바 램프’를 출시했다. 조명 아래에서 영어 또는 한글로 된 책을 펼치면 광학문자인식(OCR)으로 책의 글자를 인식한 다음, 음성 합성 기술로 아이나 성인의 목소리로 책을 읽어준다. 클로바 앱과 연동을 통해 아이의 독서 기록을 관리할 수 있으며, 외국어 교정 및 발음 지원, 번역 기능까지 제공한다[8]. 아울러, 네이버는 홈 IoT와 IPTV 고객 경험이 많은 LGU+와 협력하여 AI 개인비서 서비스 시장에서 경쟁력을 확보하고 있다.

카카오는 카카오프랫폼을 탑재한 ‘카카오미니’에 이어 네 번째 AI 스피커 ‘미니 회사’[9]를 출시하였다. 카카오 AI 스피커에는 경쟁기업과 차별적인 요소인 ‘카카오톡’ 서비스를 연동하여, 음성으로 손쉽게 카카오톡 메시지를 전송하고 정보를 공유할 수 있도록 하고 있다. 이 외에도 카카오 T택시, 카카오 내비, 카카오 홈IoT 기능을 탑재하고 있

며, 최근에는 건설사와 협력을 통해 스마트홈 분야의 서비스를 본격화하였다.

금융권에서는 2017년 금융권 최초의 챗봇 ‘현대카드 버디’ 이후, 은행과 카드사에서 챗봇 형태의 지능형 개인비서 서비스를 도입하기 시작하였다. 공공기관·지자체에서도 지능형 개인비서 서비스를 적극적으로 도입하고 있다. 대표적인 공공서비스 챗봇으로 인공지능 기반 법률 비서 ‘버비’, 민원 상담사 ‘뚜뚱’, 지방세 관련 민원 상담사 ‘지방세 상담뚱’이 있다[10].

III. 개인비서 소통 기술 동향

1. 목적지향 대화처리 기술

목적지향 대화처리 기술은 사용자가 목표를 달성하기 위해 시스템과 대화하는 유형으로, 식당 검색, 호텔 예약, 일정 관리 등 다양한 목적에 활용되고 해당 도메인의 지식 베이스를 바탕으로 사용자의 요구를 만족시킨다.

전통적인 목적지향 대화시스템은 자연어 이해, 대화 상태 추적, 대화 정책 결정, 자연어 생성 모듈이 연결된 파이프라인 구조를 갖는다. 자연어 형태의 사용자 발화가 입력되었을 때, 자연어 이해 모듈은 사용자 발화에 담긴 의미 정보를 추출하고 대화 상태 추적 모듈은 이전 대화 상태와 발화 이해 결과로부터 대화 상태를 갱신한다. 대화 상태란 인식된 사용자의 목적을 나타내는 구조화된 표현 방법이다. 이는 현재까지의 대화 이력을 포함하는 정보이며, 지식 베이스를 검색하기 위해 활용된다. 대화 정책 결정 모듈은 검색된 지식 베이스 결과와 현재 대화 상태에 기반하여 시스템이 취해야 할 행동을 결정하고 이에 따라 자연어 생성 모듈에서 자연어 형태의 시스템 응답을 생성한다.

파이프라인 구조의 목적지향 대화시스템은 동

작 이해와 제어가 가능하기에 대부분의 상용시스템에서 활용되고 있다[11]. 그러나 새로운 도메인으로의 적용이 어렵고 독립적 모듈 간의 오류가 전파되는 문제가 존재하여 최근에는 심층 신경망을 기반으로 각 모듈이 통합되는 추세이다. 자연어 이해 모듈과 대화 상태 추적 모듈을 통합한 자연어 이해 기반 대화 상태 추적 모델[12,13], 대화 정책 결정 모듈과 자연어 생성 모듈을 통합한 자연어 생성 기반 대화 정책 결정 모델[14,15], 모든 모듈을 하나의 모델로 통합한 종단형 대화시스템[16,17]이 대표적으로 연구되고 있다. 특히, 가장 최근에는 대용량 생성형 언어모델인 GPT2[18]가 목적지향 대화에 적용되어 종단형 목적지향 대화시스템의 성능이 큰 폭으로 향상되면서 종단형 목적지향 대화시스템에 대한 관심이 커지고 있다[19,20].

그림에도 불구하고 종단형 목적지향 대화시스템이 상용화 수준에 이르기까지는 여전히 해결해야 할 문제들이 남아 있다.

첫 번째는 외부 지식 베이스와의 연동이다. 전통적인 목적지향 대화시스템과 같이 대화 상태를 지식 베이스 검색을 위한 질의로 규정하는 방법은 대화 상태에 대한 상세한 설계와 학습 데이터에 대화 상태를 부착하는 작업이 필요하다. 이러한 문제를 회피하고자 메모리 네트워크를 종단형 목적지향 대화시스템에 통합하여 지식 베이스를 내부적으로 모델링하는 연구가 진행되었다[21,22]. 그러나 학습 데이터에서 낮은 빈도로 등장하는 지식, 시간이 지남에 따라 추가되거나 변경되는 지식 등 규모와 변동성이 큰 목적지향 대화시스템의 지식 베이스를 모델링하는 것은 여전히 어려움이 있다.

두 번째는 태깅 데이터에 대한 의존성이다. 대부분의 종단형 목적지향 대화시스템은 사용자와 시

스템 발화만으로 구성된 데이터가 아닌 도메인, 발화 의도, 대화 상태와 같은 부가 정보가 부착된 데이터를 학습한다. 이러한 학습 제약 사항은 종단형 대화시스템의 성능과 도메인 확장성을 떨어뜨리는 원인이 된다. 최근 태깅 데이터에 대한 의존성을 줄이기 위해 발화 의도나 대화 상태를 Latent Variable로 취급하고 Variational Inference를 종단형 목적지향 대화시스템에 적용하는 연구가 진행되었다[23,24]. 목적지향 대화 데이터의 태깅 작업은 모호한 부분이 많고 난이도가 높기 때문에 태깅 데이터에 대한 의존성을 줄이려는 연구는 계속 필요하다.

세 번째로는 학습 기반 목적지향 대화시스템의 고질적인 문제인 학습데이터의 부족이 있다. 적은 양의 학습데이터는 실제 일어날 수 있는 다양한 대화를 반영하지 못하기 때문에 종단형 목적지향 대화시스템은 편향된 응답을 보이거나 학습하지 못한 대화 흐름에 대처하지 못하게 된다. 이러한 문제를 해결하기 위해 종단형 목적지향 대화시스템과 사용자를 모델링하는 사용자 시뮬레이터를 상호작용하여 다양한 대화 흐름을 생성하여 학습하거나 실제 사용자와 상호작용 후 피드백을 받고 성능을 개선하는 강화학습 기반의 종단형 목적지향 대화시스템이 연구되고 있다[25,26]. 특히 사용자 시뮬레이터를 활용하는 방법은 사용자 시뮬레이터의 사용자 모델링 성능 또한 중요하므로 멀티 에이전트 프레임워크를 도입하여 사용자 시뮬레이터와 대화시스템을 동시에 학습하는 연구가 진행되고 있다[27-29]. 일반적으로 강화학습이 적용되는 게임, 로봇 분야와 달리 목적지향 대화는 에이전트의 행동에 대한 정의가 불분명하고 응답의 적절성, 다양성 등 보상 설계에 고려해야 할 점이 많기 때문에 앞으로도 이와 관련된 다양한 연구들이 진행될 것으로 기대한다.

2. 오픈도메인 대화처리 기술

오픈도메인 대화처리 기술은 인간다운 대화를 수행하게 하는 것으로, “인간다움”의 기준이 초기에는 재미있는 대화로 사람이 시스템과 좀 더 오래 대화하게 하는 것이 목표였다. 이를 위해서는 예제와 규칙기반 대화 방법을 주로 사용하였지만, 최근 딥러닝 대화 기술의 발전으로 다양한 대화 생성, 주제에 대하여 좀 더 구체적으로 대화를 나누는 것을 목표로 하고 있다.

“인간다움”의 기준은 지난 몇 년간 더 확장하여, 대화의 상하 문맥을 기억하여 일관성을 유지하는 문맥 기반 대화, 시스템이 특정 캐릭터를 가지며 나이, 직업, 취미 등 개인 특성을 고려한 페르소나(Persona) 대화[30], 주어진 텍스트나 구조화된 지식을 이용하여 상대방이 필요한 정보를 찾아 제공하는 깊이 있는 대화를 진행하는 텍스트 이해기반 대화 등 기술에 관한 연구가 진행됐다[31]. 최근에는 이 모든 요구사항을 모두 반영하는 평가 기준[32] 또는 혼합(Blending) 기술[33]에 관한 연구가 진행되고 있다.

딥러닝 기반 오픈도메인 대화처리 모델은 초기 RNN 구조의 생성모델에서 Transformer[34] 생성모델로 바뀌었고, 이후 초대용량 사전학습 언어모델에 대화 코퍼스로 추가 학습하여 모델을 대화모델로 Fine-tuning하는 추세가 되었다.

생성모델을 이용한 오픈도메인 대화처리에서 가장 주목받는 것은 구글이 발표한 Meena[32]와 마이크로소프트의 DialoGPT[35]이다. 이들의 특징은 단순한 모델에 초대용량 데이터를 학습시키고 모델 크기를 전례 없이 키운 것이다. DialoGPT는 Transformer 디코더 구조를 사용한 GPT2 구조[18]를 그대로 사용하였고 사전학습 언어모델을 사용한 다음 Fine-tuning하는 방식에서 더 높은 성능을

표 2 오픈도메인 대화모델과 언어모델의 크기 비교

모델명	학습데이터	단어수	파라미터수
Blender[33]	1.5B [*]	88.8B ⁺	9.4B
Meena[32]	341GB	40B	2.6B
DialoGPT[35]	40GB	1.8B	762M
GPT3[36]	45TB	300B	175B
GPT2[18]	40GB	8.2B	1.5B
BERT[37]	16GB	3.3B	340M

* Reddit comment 수, + BPE context token 수

기록하였다. Meena는 진화된 Transformer 구조[38]를 사용하였는데 단순한 중단형 학습 방법과 초대용량 데이터로 인간다움에 버금가는 대화를 생성할 수 있다는 것을 보였다. 표 2는 이들이 사용한 학습데이터와 모델 사이즈를 범용 언어모델과 비교한 것이다.

Facebook에서 발표한 Blender는 Meena를 포함하고 전의 모든 대화 모델을 능가하였다고 발표하였는데, 특히 비교평가를 통해 검색모델이 생성모델보다 나은 성능을 얻었다고 보고하였다. 최대 모델인 9.4B 모델이 아닌 2.7B 모델로 Meena와 비교하여 제안 모델의 우수성을 입증하였다. 검색모델로는 응답 후보를 대화 문맥과 독립적으로 인코딩 후 다중 주의집중을 진행하는 Poly-encoder 구조[39]를 사용하였고, 대용량 대화 데이터로 학습시킨 후, 지식 이해, 공감, 페르소나 반영 등 혼합 능력을 향상하기 위한 Fine-tuning 학습 방법을 사용하여 적은 양의 데이터로 인간다움을 향상했다.

3. 질의응답 기술

4차 산업혁명 시대를 맞아 언어의 의미를 이해하고 문장 형태로 질문을 표현하여 정답에 대한 추론이 가능한 질의응답(QA: Question Answering) 기

술의 활용도가 점점 높아지고 있다[40]. 특히, 사용자의 질문에 답을 제시하거나 추천해주는 기능은 디지털 비서가 수행해야 할 중요한 역할 중 하나이다.

스탠포드 대학을 비롯해 IBM, 구글, MS, 페이스북 등 글로벌 기업들은 질문에 포함된 단어 및 단어들의 어순을 매칭하여 정답을 추론하는 기존의 질의응답 기술에서 탈피하기 위해 기계가 다양한 주제의 글을 읽어서 뜻을 이해하는 능력인 독해력(Reading Comprehension)과 읽은 글에 대한 이해력을 평가하기 위해 질의응답 기술을 딥러닝 기반으로 개발하고 있다. 현재는 어휘와 문장의 문법 분석에서 나아가 “독해력을 갖는 기계”(MRC: Machine Reading Comprehension)를 연구하는 단계이다.

영어 QA 연구를 위한 SQuAD 2.0, MS MARCO, Natural Questions, ARC, TriviaQA 등 다양한 벤치마크 챌린지가 열리고 있다. 스탠포드 대학, 구글, 앨런 연구소 등은 구글 BERT와 유사한 딥러닝 언어 모델을 사용하고 10만~30만 건 규모의 기계독해 정답 셋을 재학습하여 단락에서 질문의 정답을 찾는 기계독해 QA 연구를 수행하였다.

구글은 웹 검색 로그에서 위키피디아에서 정답 추출이 가능한 8단어 이상의 질문에 대해 문서에서 정답을 찾는 Natural Questions 연구를 2019년 1월부터 시작하여 상대적으로 짧은 단락이 아닌 전체 문서 수준의 기계 독해 연구를 진행하고 있다[41]. 구글/스탠포드대/CMU는 2018년 HotpotQA 챌린지를 시작하여, 위키피디아 전체 문서를 대상으로 정답의 근거를 여러 문장에서 분석하여 정답의 근거를 설명 가능한 단답형 질의응답 기술 개발을 진행 중이며, 이를 위해 약 11만 건의 학습 데이터를 제공한다[42]. 앨런 연구소는 2018년 AI2 Reasoning Challenge(ARC)를 시작하여 질문과 정답

문장의 어휘의 매핑 관계를 이용하여 정답을 추출하는 현재 기술의 한계를 탈피하여 사람과 같이 문장의 의미와 상식적 추론이 가능한 차세대 질의응답 기술 개발을 시작하는 단계이다[43].

딥러닝 기반 기계 독해 기술의 성능은 이미 사람의 수준을 뛰어넘었다고 분석이 되나 실제 상업적 적용을 위해서는 정답이 있는 단락이나 문서가 주어졌을 때만 정답을 추론할 수 있는 한계를 극복해야 하는 난제는 남아 있는 상태이다.

한국전자통신연구원에서는 엑소브레인 과제로 자연어 의미 이해 기술과 심층 질의응답 기술 개발을 목표로 수행하고 있으며[44], 2016년 11월에 EBS 장학 퀴즈에서 우수한 엑소브레인 과제의 결과물을 확산하기 위해 질의응답 기술을 포함한 엑소브레인 SW API 서비스 및 기계학습데이터를 공유하는 플랫폼을 2017년 10월부터 운영하고 있다[45]. 또한 정의형, 관계형, 독해형 등 각 문제 유형별로 뉴럴 또는 심볼릭 QA 방법을 플래닝하여 최적 정답을 생성하는 뉴럴-심볼릭 하이브리드 심층질의응답 시스템을 개발하여 한컴 오피스 2020에 지식 검색으로 서비스되고 있다. 정답이 포함된 단락이 주어진 경우에만 작동하는 딥러닝 기계독해의 문제를 극복하기 위해 검색된 다수 단락으로부터 상용화 수준의 정답독해 기술인 KorBERT 기반의 Scan and Read 기계 독해 모델을 개발하여 국회도서관의 법령 QA, 국가과학기술연구회의 연구 행정 분야 QA 등 전문 분야에 적용하기 위한 실증 서비스를 완료하고 실제 사용자에게 서비스를 앞두고 있다.

IV. 디지털 개인비서 미래 전망

최근 시장조사에서는 디지털 개인비서 시장의 지속적인 확장을 예견하고 있으며[46], 다양한 사

용사례에 따른 디지털 개인비서의 예상 매출 또한 밝은 미래를 보여주고 있다[47]. 디지털 개인비서 기술이 앞으로 유용한 상업적 가치를 지니는 슈퍼 앱이 되기 위해서는 어떤 문제들이 해결되어야 하며, 또한 어떤 미래가 가능할까?

디지털 개인비서 기술에서 해결되어야 할 가장 중요한 이슈는 개인정보 침해 문제이다. 실제로 아마존의 인공지능 음성비서 Alexa를 통해 수집된 개인 음성정보에 대한 침해가 이루어진 것으로 확인되었고, 구글 Assistant 역시 녹음된 사용자 대화 내용이 유출되는 사고가 있었다. 디지털 개인비서 기술은 끊임없이 사용자 발화를 기록 및 저장하면서 적절한 상황에 문맥에 맞는 응답을 생성하는 특성을 갖고 있기 때문에 손쉽게 해결하기 어려운 문제이기도 하다. 또한 사용자의 요구사항이 증가하고 이를 만족시키는 시스템 응답을 생성하기 위해서는 더 많은 개인정보가 필요한 상황이기도 하다. 현재 이에 대한 대응은 정부기관과 기업에 의해서 이루어지고 있다. 정부기관의 경우, 예를 들어 미국 캘리포니아 주 의회는 기기에서 음성녹음을 저장하기 전에 소비자의 동의를 얻도록 하는 법안을 2019년 5월에 통과시켰다. 기업의 경우, 구글에서는 2019년 10월에 스마트 스피커가 설치된 곳에 방문하는 사람들에게 경고 장치가 필요하다는 의견을 제시하기도 하였다. 현재는 개인정보 침해를 완벽하게 막을 수 있는 방법이 제시되어 있지 않지만, 정부, 기업, 그리고 민간이 함께 이를 위해 노력하고 있다[48]. 개인정보 보호를 위한 노력이 수반될 때 디지털 개인비서는 미래 인간 생활에 없어서는 안 될 매우 핵심적인 역할을 할 것으로 기대된다.

이와 함께 디지털 개인비서의 미래는 다음과 같이 예측하고 구체화할 수 있다.

• 보다 영리한 디지털 개인비서

디지털 개인비서는 기계학습 및 이를 기반으로 한 음성인식, 자연어처리 등의 최첨단 인공지능 기술의 결합으로 구성된 소프트웨어 애플리케이션이다. 따라서 음성인식 및 자연어처리와 같은 인공지능 기술이 발달할수록 디지털 개인비서도 사용자 의도를 보다 정확하게 파악하여 해당 문맥에 적합한 올바른 응답을 제시할 수 있다. 딥러닝을 포함하는 기계학습 기술의 발달과 함께 이러한 음성인식 및 자연어처리 기술도 크게 개선되기 때문에 디지털 개인비서 역시 보다 더 영리해지며, 사용자의 필요성을 이해하고, 학습하며, 사용자의 요구를 예측할 수 있는 단계까지 진화할 것이다.

• 오픈 플랫폼 가능성에 대한 기대

현재 디지털 개인비서는 구글, 아마존, 마이크로소프트, 삼성, 바이두와 같은 각각 다른 개발업체에 의해 개발된 서로 다른 생태계와 플랫폼을 사용하고 있다. 이는 디지털 개인비서 간의 단절을 의미한다. 하지만 가까운 미래에는 플랫폼 간의 전이가 용이해져서, 특정 태스크를 수행하는 데 있어서 다른 플랫폼의 디지털 개인비서에게 요청하여 처리하는 것이 가능할 수도 있다. 즉, Siri가 Alexa에게 특정 태스크를 요청하고 이에 대한 결과를 받아 사용자에게 제시해 주는 것이 가능해질 것이다[49].

• IoT 장치와의 결합

자동차 스피커, 스마트 홈 장치, 웨어러블 장치와의 결합을 통해 디지털 개인비서는 인간과 보다 가까이에 존재하며, 다양한 요구를 만족시킬 수 있을 것이다. 예를 들어 운전하고 있거나, 조깅하는 도중과 같이 양팔을 사용할 수 없는 경우에 모바일 장치나 IoT 장치처럼 스크린 없는 장치의 애플리

케이션에 접근해야 하는 경우 디지털 개인비서는 상당히 유용하게 쓰일 것이다. 더 나아가 모든 디바이스들의 사용자 인터페이스로서의 역할도 할 수 있을 것이다[50].

- 영상 이해를 통한 보다 정확한 이해

현재의 디지털 개인비서는 대개 음성 신호에 반응하여 작동되고, 그에 대한 음성 응답을 생성한다. 이것은 실세계에서 인간이 대상을 인식하고 이해하는 것과는 많은 차이를 보인다. 실제로 인간의 소통은 시각적인 부분을 많이 포함하고 있다. 이러한 것을 반영하여 미래의 디지털 개인비서 기술은 시각 컴포넌트를 포함하여 사진, 영상, 인포그래픽스 등을 입력으로 사용하여 대상이 갖는 모호성을 해소하여 보다 잘 이해하고 보다 풍부한 응답을 생성할 수 있을 것이다[49].

- 감정 분석 및 이해

현재 디지털 개인비서는 사용자 감정 분석이 결여되어 있다. 하지만 다가올 미래에는 사용자의 기분과 감정을 분석하고 상황에 맞는 응답을 해 주는 것이 가능할 것이다[51].

- 팬데믹 환경에서의 디지털 개인비서

우리는 코로나-19로 인한 팬데믹 환경에서 사람들과 사람들 간의 대면 활동이 급작스럽게 중지되어 서로 단절되는 상황에 직면하는 것을 목격하였다. 그리고 이러한 환경이 쉽게 끝날 것 같지도 않은 상황이다. 비대면 환경에서 디지털 개인비서는 각 개인 및 집단 간의 연결고리 역할을 할 수 있을 것이다. 또한 디지털 개인비서는 개인 간 접촉이 단절된 환경에서 상대적으로 컴퓨터 취약계층에게 그들이 필요로 하는 정보를 보다 편한 방식으로 제공해 줄 것이다[52].

더 나아가, 디지털 개인비서는 언젠가는 인간 성격을 지니고 시각적 특성을 갖는 형태로 존재하며 인간과 떨어질 수 없는 관계로 발전할 것이다. 그래서 언제 어디서나 곁에 있으면서 사용자가 원할 때 등장하여 사용자의 요청을 들어주고 해결해 주는 날이 머지않아 올 것으로 기대된다.

V. 결론

살펴본 바와 같이 디지털 개인비서는 다양한 형태로, 그리고 다양한 디바이스에서 다양한 서비스를 가지고 우리 생활에서 밀접하고 유용한 역할을 수행 중이다. 그리고 최근 인공지능 기술 발전으로 미래 시장 전망은 지속적으로 성장할 것으로 전망된다. 인간처럼 말을 하고 생각하고 성격을 갖는 개인비서를 기대하지만, 기술적으로 해결할 문제들이 앞서 보았듯이 산재해 있다. 미래 시장 전망의 기대를 충족하기 위해서 기술적 진보를 위한 노력이 필요할 것으로 보인다.

용어해설

RNN 텍스트와 같은 시퀀스(sequence) 데이터를 입력으로 받아 다른 시퀀스로 출력하는 신경망 모델

Transformer 자연어와 같은 시퀀스 데이터를 입력으로 번역, 대화처리, 요약 등을 가능한 한 기계학습 신경망 모델로 순차적 처리 아닌 병렬 처리로 RNN보다 정보 손실이 적고 다중 주의집중으로 입력 시퀀스의 장거리 관계 파악이 장점임. BERT와 GPT 등의 언어모델에 사용된 모델

Fine-tuning 이전에 학습된 모델을 새로운 목적에 맞게 추가 데이터를 학습하여 파라미터를 미세조정하는 모델 학습 방법

약어 정리

BERT	Bidirectional Encoder Representations from Transformers
BPE	Byte Pair Encoding
GPT	Generative Pre-Training

MRC	Machine Reading Comprehension
QA	Question Answering
RNN	Recurrent Neural Network

참고문헌

- [1] S. Chourasia, "What are 10 best AI assistants of 2021?," Aug. 2020, <https://techresearchonline.com/blog/best-ai-assistant-of-2021/#>
- [2] Pat Research, "Top 22 intelligent personal assistants or automated personal assistants," <https://www.predictiveanalyticstoday.com/top-intelligent-personal-assistants-automated-personal-assistants/#whatareintelligentpersonalassistantsorautomatedpersonalassistants>
- [3] LG CNS, "앱의 시대에서 가상 비서의 시대로, Virtual personal assistant," 2016, <https://blog.lgcns.com/1241>
- [4] 임영신, "폰안의 비서, 그들만의 전쟁을 시작했다," 매일경제, 2020.11.1.
- [5] 삼성 빅스버, <https://www.samsung.com/sec/apps/bixby/>
- [6] NUGU, <https://www.nugu.co.kr/>
- [7] KT 기기가지니, <https://gigagenie.kt.com/main.do>
- [8] 김주완, "책 펴기만 하면 대신 읽어주네 일상 깊숙히 파고드는 AI," 한국경제, 2020.11.9.
- [9] 카카오 미니행사, <https://kakao.ai/product/minihexa>
- [10] 한국정보화진흥원, "인공지능 기반 챗봇 서비스의 국내외 동향분석 및 발전 전망," D. gov Trend & Future, 2018-2호, 2018.
- [11] Z. Zhang et al., "Recent advances and challenges in task-oriented dialog systems," *Sci. China Inf. Sci.* May. 2020, arXiv: 2003.07490.
- [12] H. Lee, J. Lee, and T. Y. Kim, "SUMBT: Slot-utterance matching for universal and scalable belief tracking," in *Proc. Assoc. Comput. Linguist.* 2019, pp. 5478-5483.
- [13] M. Heck et al., "TripPy: A triple copy strategy for value independent neural dialog state tracking," in *Proc. Spec. Interest Group Discourse. Dialogue.* July. 2020, pp. 35-44.
- [14] T. Zhao, K. Xie, and M. Eskenazi, "Rethinking action spaces for reinforcement learning in end-to-end dialog agents with latent variable models," in *Proc. Conf. North Am. Chapter Assoc. Comput. Linguist.: Hum. Lang. Technol.* Minneapolis, Minnesota, Apr. 2019, pp. 1208-1218.
- [15] W. Chen et al., "Semantically conditioned dialog response generation via hierarchical disentangled self-attention," in *Proc. Assoc. Comput. Linguist.* Florence, Italy, Jan. 2019, pp. 3696-3709.
- [16] W. Lei et al., "Sequicity: Simplifying task-oriented dialogue systems with single sequence-to-sequence architectures," in *Proc. Assoc. Comput. Linguist.* Melbourne, Australia, July. 2018, pp. 1437-1447.
- [17] Y. Zhang, Z. Ou and Z. Yu, "Task-oriented dialog systems that consider multiple appropriate responses under the same context," in *Proc. AAAI Conf. Artif. Intell.* Palo Alto, CA. USA. vol. 34 no. 5, 2019, pp. 9604-9611.
- [18] A. Radford et al., "Language models are unsupervised multitask learners," OpenAI Technical Report, 2019, pp. 9604-9611.
- [19] E. Hosseini-Asl et al., "A simple language model for task-oriented dialogue," in *Proc. Conf. Neural Inform. Process. Syst.* Vancouver, Canada, Dec. 2020.
- [20] D. Ham et al., "End-to-end neural pipeline for goal-oriented dialogue systems using GPT2," in *Proc. Assoc. Comput. Linguist.* July. 2020, pp. 583-592.
- [21] A. Bordes, Y. L. Boureau, and J. Weston, "Learning end-to-end goal-oriented dialog," in *Proc. Int. Conf. Learn. Representations.* Toulon, France, Apr. 2017, arXiv: 1605.07683v4
- [22] A. Madotto, C. S. Wu, and P. Fung, "Mem2Seq: Effectively incorporating knowledge bases into end-to-end task-oriented dialog systems," in *Proc. Assoc. Comput. Linguist.* Melbourne, Australia, July. 2018, pp. 1468-1478.
- [23] T. Wen et al., "Latent intention dialogue models," in *Proc. Int. Conf. Mach. Learn.* Sydney, Australia, Aug. 2017.
- [24] Y. Zhang et al., "A probabilistic end-to-end task-oriented dialog model with latent belief states towards semi-supervised learning," in *Proc. Conf. Empir. Methods. Natural Lang. Process.* Nov. 2020, pp. 9207-9219.
- [25] X. Li et al., "End-to-end task-completion neural dialogue systems," in *Proc. Int. Conf. Natural Lang. Process.* Taipei, Taiwan, Nov. 2017, pp. 733-743.
- [26] T. Hong et al., "An end-to-end trainable task-oriented dialog system with human feedback," in *Proc. AAAI Reason. Learn. Hum.-Mach. Dialogues.* Honolulu, HI, USA, Jan. 2019, pp. 1-7.
- [27] B. Liu and I. Lane, "Iterative policy learning in end-to-end trainable task-oriented neural dialog models," in *Proc. Automat. Speech. Recognition. Underst. Workshop.* Okinawa, Japan, Dec. 2017, pp. 482-489.
- [28] A. Papangelis et al., "Collaborative multi-agent dialogue model training via reinforcement learning," in *Proc. Spec. Interest Group Discourse. Dialogue.* Stockholm, Sweden, Sept. 2019, pp. 92-102.
- [29] R. Takanobu et al., "Multi-agent task-oriented dialog policy learning with role-aware reward decomposition," in *Proc. Assoc. Comput. Linguist.* 2020, pp. 625-638.
- [30] S. Zhang et al., "Personalizing dialogue agents: I have a dog, do you have pets too?," in *Proc. Assoc. Comput. Linguist.* Melbourne, Australia, July. 2018, pp. 2204-2213.
- [31] E. Dinan et al., "Wizard of wikipedia: Knowledge-

- powered conversational agents,” in Proc. Int. Conf. Learn. Representations. New Orleans, LA, United States, May. 2019.
- [32] D. Adiwardana et al., “Towards a human-like open-domain chatbot,” 2020, arXiv: 2001.09977
- [33] S. Roller et al., “Recipes for building an open-domain chatbot,” 2020, arXiv: 2004.13637
- [34] A. Vaswani et al., “Attention is all you need”, in Proc. Int. Conf. Neural Inform. Process. Syst., Long Beach, CA, USA, Dec. 2017, pp. 6000-6010.
- [35] Y. Zhang et al., “DialoGPT: Large-scale generative pre-training for conversational response generation,” in Proc. Assoc. Comput. Linguist. July. 2020, pp. 270-278.
- [36] T. B. Brown et al., “Language models are few-shot learners”, in Proc. Conf. Neural Inform. Process. Syst. Vancouver, Canada, Dec. 2020.
- [37] J. Devlin et al., “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. Conf. North Am. Chapter Assoc. Comput. Linguist.: Hum. Lang. Technol. Minneapolis, Minnesota, June. 2019, pp. 4171-4186.
- [38] D. R. So et al., “The evolved transformer,” in Proc. Int. Conf. Mach. Learn. Long Beach, CA, USA, June, 2019.
- [39] S. Humeau et al., “Poly-encoders: Architectures and pre-training strategies for fast and accurate multi-sentence scoring,” in Proc. Int. Conf. Learn. Representations, Aug. 2020.
- [40] 김현기 외, “엑소브레인 한국어 분석 및 질의응답 기술의 개발 현황 및 고도화 계획,” 정보과학회지, 2017. 8. pp. 51-56.
- [41] <https://ai.google.com/research/NaturalQuestions>
- [42] <https://hotpotqa.github.io/>
- [43] <https://allenai.org/data/arc>
- [44] <http://exobrain.kr/>
- [45] <https://aiopen.etri.re.kr/>
- [46] <https://www.statista.com/>
- [47] <https://omdia.tech.informa.com>
- [48] 한국인터넷진흥원, “가정용 스마트 기기의 개인정보 침해 사례 및 최근 동향,” 해외 개인정보보호 동향 보고서, 2019. 11.
- [49] <https://medium.com/>
- [50] <https://www.ciokorea.com/>
- [51] <https://www.voicesummit.ai/>
- [52] L. Sheerman et al., “COVID-19 and the secret virtual assistants: the social weapons for a state of emergency,” Emerald Open Research, 2020.