

인공지능 스피커의 과거와 현재, 그리고 미래

- 하진욱 (인하공업전문대학 컴퓨터정보공학과)
- 김상혁 (인하공업전문대학 컴퓨터정보공학과)

I. 서론

글로벌 IT 업체부터 스마트폰 제조사, 이동통신사까지 IT에 관련 업체들이 최근 관심을 보이고 있는 서비스가 바로 인공지능 비서 서비스(Intelligent personal assistant)이다. 특히, 인공지능 비서 서비스는 스마트 폰에서 서비스를 지원하면서 두각을 나타내었지만, 처음에는 간단한 농담조차도 인지하지 못하던 유아 수준의 지능을 가진 애플단지였다.



Fig. 1. 삼성 갤럭시 s1, s보이스 광고[1]

그림 1의 광고 외에도 여러 특정 스마트폰 광고에서 이와 같은 자사의 인공지능 비서 서비스를 소개하였지만 다양한 문제 발생으로 사용자들에게 외면 받아왔다.

더 이상 발전 가능성이 없을 것 같아 보이던 인공지능 비서 서비스는 사용자들에게서 잊혀지면서 관련 기업들은 자사 인공지능 서비스 개발을 포기하기 시작했다.

하지만 알파고(AlphaGo)의 등장과 4차 산업혁명으로 인공지능 서비스는 다시 한 번 주목받기 시작했다. 알파고는 빅데이터와 딥러닝, 머신러닝, 클라우드 컴퓨팅, 인공지능 기술이 합쳐진 4차 산업혁명의 산물이다[2]. 인간과 인공지능의 대결인 이세돌 9단과 알파고의 대결은 단순히 인간과 컴퓨터의 대결이지만, 사실은 인터넷에 연결된 수백, 수천대의 컴퓨터와 인간의 대결인 것이다. 수백, 수천대의 컴퓨터는 실시간으로 특정 데이터인 바둑 최고의 수를 계산해 최적의 선택을 한다. 인간이 한 수를 둘 때마다 엄청난 속도로 모의 바둑을 둔다. 통계를 낼 때와 같이 한, 두 가지의 케이스가 아닌 수백 수천만가지의 케이스를 만들어 오차범위를 줄이고, 이길 수 있는 확률을 높이는 한 수를 둔다.

최선의 선택, 최적의 선택만 하는 것이 인간과 비슷하다고 하기 보단 인간이 하지 못하는 엄청난 시간, 엄청난 노력을 쏟아 부어야 하는 일들을 좀 더 수월하게하기 위한 말 그대로 인공지능인 것이다.

최근의 인공지능은 뛰어난 연산처리 능력 기반의 과학적 추론을 넘어서서 인간의 영역이라고 치부되어 왔던 감정이나 상상력, 예술적 창작까지 넘보고 있다. 지금까지 인간의 영역에 발을 들이기 위한 단계로 인간의 ‘외형’을 유사하게 만들기 위한 로봇관련 사업에 쏠리던 시선이, 인간의 ‘생각’과 유사하게 만들기 위한 사업으로 변화된 것이다. 하지만 빅데이터와 클라우드 컴퓨팅이라는 기술이 나타나면서 실제로는 인간처럼이 아닌 ‘인간보다 더’라는 것이 맞는 말이라고 할 수 있다.

이 모든 기술의 시작은 4차 산업혁명의 대표주자중의 하나인 빅 데이터 기술 때문이라고 할 수 있다. 대량의 데이터가 있기 때문에 가능한 일이다. 인터넷에서 자신이 원하는 데이터만을 찾아 분석 비교, 활용 할 수 있는 기술의 원천이 되는 것이다. 인공지능도 필요한 데이터만을 모아서 학습했기 때문에 만들어 질 수 있었던 것이다. 이러한 다양한 기술의 집약체인 인공지능 기술이 구글의 알파고를 시작으로 페이스북, 인텔, 구글 등의 글로벌 IT회사들이 각자 특정 분야에 맞는 인공지능을 개발, 학습시키고 있고, 아마존의 알렉사, 애플의 시리등 인간의 생활을 좀 더 수월하게 해 주는 인공지능 비서 서비스를 지원하고 있다[3].

인공지능은 특정 시뮬레이션을 통해 통계를 얻어야 하는 일에서는 물론이고, 바둑, 체스와 같은 보드게임, 스타크래프트와 같은 온라인 게임 등 다양한 분야에서 인공지능을 활용한 실습 환경이 공개되었다. 그 중에서도 대한민국에선 민속 놀이라고도 불리는 ‘스타크래프트’에 적용되는 인공지능에 많은 관심이 쏠렸다.

다음 수를 계산하고, 정해진 시간 내에 최선의 수를 계산해야 하는 체스, 바둑과는 다르게 유닛 하나하나에 컨트롤이 필요하고, 상대방의 빌드를 예측, 확인하여 빠른 판단과 최선의 방법을 강구해야 하는 실시간 전략 시뮬레이션 게임이었기에 더욱 많은 관심을 받게 되었다. 매년 개최 되고 있는 인공지능 스타크래프트 대회에 참여한 28개의 팀 중 인간을 이긴 인공지능은 존재하지 않았다[4]. 더욱 많은 기술을 요구할 수밖에 없는 게임이기에 당연한 결과였지만, 지금까지는 참여자 대부분이 개인 개발자들이었던 것과는 달리 2018년부터는 대형 IT업체가 참여하기로 예정되어있기 때문에 조금 더 기대가 되는 분야이다. 이렇게 특정 분야에서 ‘이기는 것’만을 목적으로 개발된 인공지능이 있는가 하면 사람을 돕기 위해, 사람과 소통하기 위해 개발되는 인공지능들도 있다. 이런 인공지능은 특정 분야에서의 특정 데이터만을 필요로 하는 것이 아닌 실제 사람처럼 생각하고 학습하고 다양한 이야기와 일들을 할 수 있는 것이 특징이다.

이러한 스스로 학습하게 하는 딥 러닝, 머신 러닝 기술을 활용한 인공지능이 등장했지만 사람들이 기본 학습이 잘못되면 성장해서도 잘못된 길로 가게 되는 것처럼 인공지능에게도 잘못된 가르침을 주게되면 잘못된 생각과 행동을 하게 된다는 위험성이 나타났다. 바로 마이크로소프트사에서 개발한 ‘테이’에게 그런 문제가 나타났다. 자유롭게 사용자와 대화를

하며 성장하도록 만들었던 이 인공지능은 처음에는 다양한 사용자와 대화를 나누며 잘 지내는 듯 보였지만, 극우 성향의 사용자들이 나타나 이상한 사상과 욕설 등을 지속적으로 주입하자 대량학살지지, 인종차별, 성 차별 발언을 쏟아냈다. 해당 개발사는 위험성을 인지하고 이 프로그램의 운영을 중지했지만 앞으로 인공지능이 사람들과 공존하기 위해선 어떤 기초데이터가 필요한지를 느끼게 해주는 대목이었다[5].

이처럼 모든 일의 시작은 수많은 데이터들 중 필요한 것만을 고르는 작업이 필요하다. 지금껏 사용자가 특정 작업을 하기 위해서는 직접 정보에 대한 키워드를 생각한 뒤 검색하고, 수많은 검색 결과 중에서 필요한 정보만을 찾아야 했다. 하지만 검색 엔진의 발전과 함께 사람들에게 많이 노출되는 정보들이 필요한 정보들이라 인식되어 상단에 노출되었고, 이를 넘어서 사용자의 사용패턴을 분석하여 사용자가 원할 것 같은 내용의 콘텐츠를 검색 해주는 기술까지 발전되었다. 하지만 이 일련의 과정들은 사용자가 원하는 데이터를 찾기 위해 직접 움직여야 한다는 단점이 있었다. 이를 보완하고 사용자가 원하는 정보를 찾아주고 제공하는 것이 바로 인공지능 비서이다.

II. 관련 연구

현재 인공지능이 인간과 가장 다른 점은 ‘감정’이다. 인공지능에 감정 혹은 사회성을 부여 할 수 없을까하는 생각들이 많았다. 따라서 이에 관련된 연구가 많이 진행되고 있다.

효과적인 개인 관리 서비스를 위한 대화 기반 감정형 인공지능 비서의 구현에서는 모바일 환경이 아닌 집 내부의 환경을 기준으로 인간들에게 정해진 시나리오를 직접 실행해보도록 하고 평가를 내리기도 한다[6]. 사용성 평가에선 인공지능과의 대화가 쉬워야 하며, 이를 통해 인공지능 비서 서비스에 제안을 하는 대화 기반 인터페이스도 쉬워야 한다고 말한다. 또한 해당 논문에선 ‘집 내부’라는 한정된 상황에서 작동하는 인공지능 비서 서비스를 조건으로 내 걸었지만, 앞으로 나타나게 될 다양한 플랫폼, 특히 사람들이 갖고 다닐 수 있는 모바일 환경에서의 데이터는 모바일 기기에 저장되어있는 사용자의 다양한 정보를 바탕으로 개인 관리 서비스를 지능적으로 제공하는 시스템이 구현되어야 한다고 말하고 있다.

음성인식 기술의 발전과정은 표1과 같다.

Table 1. 음성인식 기술의 발전과정

세대	기술
1세대	사람을 따라하는 정도.
2세대	간단한 음성명령이나 음성제어를 할 수 있는 정도이나 정확성이 낮아 실용적이지 못하다
3세대	음성 인식률을 높이기 위해 MCE와 같은 변별학습 기법이 사용되고 신경회로망 기반 인식 기술, n-gram 기반의 언어모델 기술을 제안하고 카네기 멜론대학의 스피크스 시스템 개발
4세대	실시간으로 데이터를 저장한 빅 데이터와 딥러닝 기술을 이용하여 사용자와 감정을 파악하거나 상황을 인지할 수 있는 기술
5세대	

표 1에서 현재는 음성인식 기술 발전과정의 4, 5세대를 맞이하여 다양한 인공지능 음성 비서 서비스가 출시되고 있으며 이런 시장의 물결을 맞이하여 음성인식의 발전을 위해 사용자의 상황에 맞는 문맥 파악, 목소리의 감정 분석, 잡음 구분 등을 통해 정확성을 높이는 연구가 활발히 진행되어야 할 것이라고 마무리했다.

인공지능을 활용한 개인비서 서비스의 기술을 분석했다 [7]. 음성인식에 대한 정확도는 상당히 발전된 것을 알 수 있다 [8]. 현재와 같은 연구들이 활발하지 않던 시절에는 사람들이 하는 말을 알아듣지 못해서 애플단지로만 불렀던 음성인식 서비스들이 인공지능 비서 서비스들의 대두로 인해 상당히 중요한 분야가 되었기 때문이라고 생각한다.

Google, Naver, Kakao의 세 가지 음성인식 오픈 API를 활용해 음성인식률을 측정한 결과는 그림 2와 같다 [8].

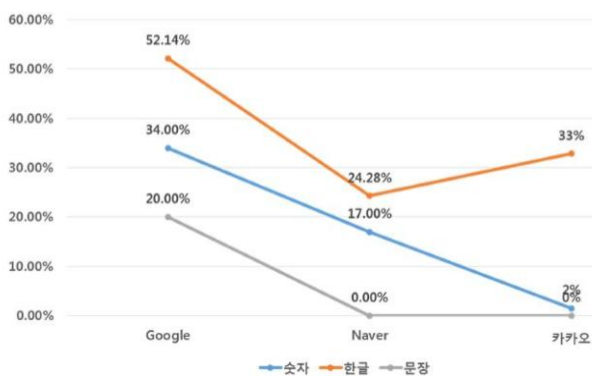


Fig. 2. 음성인식 API 실험 결과

그림 2를 살펴보면 Google의 음성인식 오픈 API Speech가 세 가지의 오픈 API중 가장 낮은 인식률을 보였다. 나머

지 두 오픈 API는 Google의 오픈API보다 인식률이 월등히 높다는 것이 상당한 수준의 기술까지 도달했다는 것을 알 수 있다.

최근 IT와 관련된 다양한 기사들은 4차 산업혁명과 인공지능과 관련된 내용들이 많이 소개되고 있다. 특히 ‘사람 같은 자연스러움이 관건’이라는 주제를 갖고 최근 삼성 빅스비에 추가된 전문 성우 및 연예인의 목소리에 대해 이야기한다. 자연스러움은 소비자들의 관심을 끌기에 적합한 기술이며 좋아하는 연예인의 목소리를 활용해 조금 더 자연스럽게 들리도록 만드는 것이 관건이라고 말한다 [9]. 또한, ‘AI는 UI의 미래다’라고 정의하고, 현재 시중에 유통되고 있는 AI비서 서비스들 중 빅4라고 불리는 제품들에 대해 이야기하고 있다. 현재 빅4라고 불리는 이 제품들은 ‘생태계’가 잘 꾸려져 있기 때문에 개발자 혹은 사람들이 쉽게 접근 가능하고 더욱 발전할 수 있었다고 말한다.

AI비서 서비스가 적어도 가까운 미래에 데스크톱에서 키보드나 마우스를 대체하는 정도까지는 어려워 보이지만, 운전 중이나 걷는 도중, 또는 모바일 기기나 IoT기기처럼 스크린 없는 기기의 애플리케이션에 접근해야 할 경우에 유용하게 쓰일 가능성이 높을 것이라고 한다 [10].

III. 사례 연구

딥 러닝과 머신러닝을 활용해 사람을 모방하는 인공지능을 만들려는 시도가 있는 반면, 사용자들에게 필요한 것을 제공하고, 사용자 기반의 데이터를 수집해서 사용자에게 더 나은 편리성을 제공하는 인공지능이 있다. 바로 인공지능 비서이다. 인공지능 비서는 초반에 언급했던 삼성의 S보이스와 LG의 Q보이스처럼 스마트폰이 등장한 2000년대 후반부터 다양한 회사에서 자사의 플랫폼을 구현하기 위한 많은 노력들이 있었다. 하지만 이 시기에 등장한 인공지능 비서는 현재 등장하고 있는 인공지능 비서 프로그램들과 비교하면 인공지능이라고 하기엔 부끄러운 정도의 수준이었다. 사용자가 입력할 수 있는 다양한 변수를 예측 할 수 없는 것은 물론이고 정해져있는 틀 안에서 입력을 받아야 했다. 국내에서 내놓았던 인공지능 비서 서비스들은 LG의 Q보이스를 빼면 음성 검색 서비스라고 생각해도 무방한 정도였다. 삼성 S보이스와 애플의 시리를 비교한 결과는 그림 3과 같다 [11].



Fig. 3. 애플 시리와 삼성 s보이스

애플 시리에 이어 구글 나우(현재 구글 어시스턴트)등 공룡 IT기업에서 인공지능 비서 서비스들을 개선하고 발전시켜 나갔지만, 국내 IT기업들은 해당 서비스를 포기하거나 손을 놓아버렸다. 몇몇 대형 커뮤니티에서는 해당 서비스를 버린 것이 아니냐는 추측까지 돌았다.

1. 국내외 기업들의 AI비서 현황

국내외 기업들의 AI비서 현황은 그림 4와 같다.

업체	시클랫폼(출시)	특징
애플	시리 (2011년)	자사운영체제에서 이용, 문맥파악과 대화가능, 아이폰 등에 탑재, 향후 스마트홈 솔루션에 탑재예상
아마존	알렉사 (2014년)	자연어처리, 빠른 반응 속도, 클라우드기반, 스피커 외 다양한 스마트홈 디바이스에 탑재
구글	구글어시스턴트 (2016년)	자사검색엔진과 연동, 모바일메신저, 스마트폰, 스피커, 자동차 등으로 탑재확대
마이크로소프트	코타나 (2014년)	MS 운영체제와 기기에서 이용가능, 각종 스마트 기기 등에서 이용 수 있도록 준비중
SKT	누구 (2016년)	AI스피커 누구에 탑재
KT	기가지니 (2017년)	AI스피커 기가지니에 탑재 SKT의 '누구'와 KT의 '기가지니'
삼성	빅스비 (2017년)	갤럭시S8탑재, 자사 가전제품과 연동해 스마트홈 기능 제공계획
네이버	클로바 (2017년 예정)	오감 AI개발에 초점, AI비서로 개발하던 기존 아마키의 업그레이드 버전

Fig. 4. 국내외 기업들의 AI비서 현황

2017년 3월 갤럭시 s8 언팩과 함께 빅스비를 발표하면서 본격적인 인공지능 비서 서비스 경쟁에 참여했다. 스마트폰과 같은 이동식 서비스에서만 서비스되던 인공지능 비서 서비스는 MS의 코타나처럼 운영체제에 내장되어 서비스되기도 하고, 최근 출시 후 너무 빨리 매진이 되어서 없어서 못 산다는 카카오톡과 클로바같은 원통형 기기에 내장되어 서비스되고 있다. 최근 카카오톡의 카카오톡미니와 네이버의 클로바같은 기기들이 소비자들에게 인기인 이유는 캐릭터상품과의 연계성에서 큰 힘이 되기도 했지만, 귀여운 캐릭터만을 원했기 때문이었다면, 이런 비싼 장식품을 사려고 하지 않았을 것이다. 현재 인공지능 비서기기들의 가장 큰 인기 이유는 영화에서만 보던 나만의 '비서'가 실현될 수 있다는 마음의 안락이 가장 큰 이유일 것이다.

하지만 과연 '인공지능'이라는 표현을 쓸 정도로 AI 스피커의 기능이 뛰어난 것일까. 결론부터 말하면 AI 스피커보다 아직은 TV 리모컨이 편하다는 말들이 많다. 'TV 켜줘', '볼륨 높여줘' 등 음성인식으로 해결할 수 있다는 점에서 편리성으로는 분명한 진화다. 하지만 아직은 사람의 손을 전부 덜어주기에는 시기상조인 것 같다. AI스피커 기대 기능 및 불편한 경험은 그림 5와 같다.

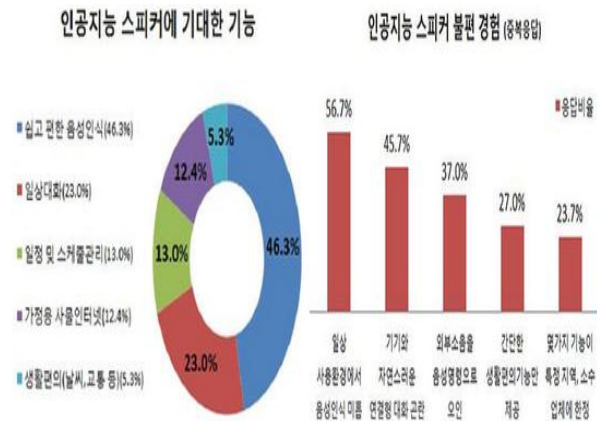


Fig. 5. AI 스피커 기대기능 및 불편경험[13]

그림 5를 살펴보면 AI 스피커의 음성인식이나 대화기능에 대한 소비자 만족도가 낮다. 그 이유를 분석해 보면 첫째, 음성인식의 정확도가 낮기 때문이다. 정상적인 작동을 위해서는 대체로 정확한 발음이 일정거리 내에서 전달돼야한다. 거리가 멀거나 발음이 정확하지 않다면 인식률이 크게 떨어진다. 여기서 '발음이 정확하지 않는다'고 표현 했지만 일상적인 사람간의 대화라면 그 의미를 충분히 인식할 수 있는 정도

의 수준이다. 일상의 대화와 비교했을 때 현재의 AI스피커는 낮은 인식률과 사용자를 ‘불편’하게 할 수 있을 것이다. 둘째, 프라이버시에 대한 부분이다. 혼자 사는 1인 가구라면 크게 신경 쓰지 않아도 될 문제이다. 그러나 공동생활의 공간에선 음성으로 명령을 내리는 행위 자체가 상황에 따라서는 공동 생활자의 정숙성을 방해하고, 때로는 의도치 않게 개인의 취향을 노출시킨다는 점에서 불편할 수 있다는 생각이다.

물론 AI 스피커가 가진 무한한 잠재력과 가능성은 높다. 음성인식 기술의 정확도, ‘문장’까지 이해하는 논리적 이해도가 지금 보다 훨씬 더 개선되고, 콘텐츠의 연계성도 보장된다는 점을 전제한다면, 인공지능 스피커는 새롭게 진화된 ‘생활편의 플랫폼’이 될 수 있을 것이란 생각이다. AI 스피커가 단순히 TV를 켜고, 볼륨을 높이고, 채널을 찾는 수준에서 그치는 것이 아니라 AI 스피커 밴드들이 지향하는 것처럼 집안의 모든 가전기와 연계한 IoT 플랫폼으로 온전하게 기능하게 된다면 그 효용의 강력함은 훨씬 더 배가 될 것이다. 현재는 ‘음성 인식’이지만 향후 더 진화된 각 개인의 목소리를 기억하는 ‘음성 인증’까지 스마트 스피커가 답을 수 있다면 민감한 금융거래까지도 가능한 금융플랫폼의 기능도 충분히 가능할 것으로 보인다.

2. 아마존 Echo

아마존 에코에 탑재된 인공지능 알렉사는 단순히 인공지능 스피커를 위한 기술이 아니다. 알렉사는 플랫폼이다[14]. 아마존 알렉사가 지원하는 오픈 API는 그림 6과 같다.



Fig. 6. Alexa에서 지원하는 오픈 API

알렉사는 두 가지 형태의 API로 개발이 가능한데, 첫 번째는 ASK(Alexa Skills Kit)와 두 번째로는 AVS(Alexa

Voice Service) 이다. 이는 외부 개발자들이 자신의 하드웨어나 서비스에 알렉사를 탑재할 수 있도록 만든 API, 도구, 문서, 코드 샘플의 집합이다. 이를 이용하여 외부 개발자들은 아마존 에코와 유사한 기능을 하는 디바이스를 만들 수 있다. 알렉사가 탑재된 TV, 냉장고, 세탁기, 자동차 등 무수히 많은 디바이스가 나올 수 있다. 아마존은 인공지능을 플랫폼 화하는데 성공한 대표적인 회사라고 꼽을 수 있다. 그러나 알렉사가 아마존 AI 전략의 전부는 아니다. 알렉사는 완성된 서비스 형태일 뿐. 파트너들은 아마존이 만든 모델로 아마존이 학습시킨 결과를 API로 가져다 쓸 수 있을 뿐이다. 기업들이 아마존 AI 기술을 자유자재로 사용하기 위해서는 좀 더 유연한 플랫폼이 필요하다. 개발자 자신이 갖고 있는 데이터로 비교적 쉽게 기계 학습 모델을 만들거나 목적에 맞는 알고리즘 설계를 돕는 플랫폼이다. 그것도 각 단계별로 편의성과 자유도를 각각 다르게 주는 도구라면 더욱 환영 받을 것이다.

아마존에서 이런 서비스를 지원하는 것이 AWS다. AWS에서 내놓는 AI는 난이도에 따라 세 가지 종류의 기술로 나뉜다. AWS 인프라 위에서 새로운 서비스를 개발하고 싶어 하는 이들이 많은데, 각자가 보유한 데이터양이나 혹은 기술 수준이 다르니 각각의 층위에 맞는 기술을 제공하겠다는 것이다. AWS AI 서비스 중 대표적인 것이 ‘폴리’와 ‘리코그니션’, 그리고 ‘렉스’다. 폴리는 개발자가 쉽게 가져다 쓸 수 있는 TTS(문자에서 음성으로 변환)서비스다.

폴리는 딥러닝으로 문장 전체의 문맥을 파악하고, 문맥 중 끊어 읽을 부분, 억양의 높낮이를 조절할 부분, 음울 등을 고려해 가장 사람다운 목소리로 읽어준다. 이 외에 리코그니션은 이미지 인식 기술이다. 사진을 찍어서 리코그니션에 올리면 그걸 분석해 정보를 제공한다. 풍경 사진을 찍어서 올리면 그 안에 있는 여러 사물을 추적해 결과를 알려주고, 그 결과 값과 실제 사물이 일치할 확률을 함께 제공한다. 사진 속 얼굴을 인식해 성별과 표정 등을 읽는 기술도 리코그니션에 들어간다.

렉스는 알렉사 같은 음성 비서 프로그램을 누구나 만들 수 있게 한, ‘알렉사의 원형’이라 생각하면 쉽다. 이름도 ‘알렉사(Alexa)’에서 맨 앞과 뒤의 알파벳을 제거해 지었다. 일반 기업이 자신의 브랜드로 알렉사와 유사한 서비스를 만들고 싶을 때 렉스를 가져다 쓰면 된다. 들어가는 기술은 알렉사와 같으나, 서비스 되는 브랜드명은 개별 기업의 것을 가져다 쓸 수 있다. 예컨대 배달음식이나 환자 예약 등, 정형화된 질문을 음성으로 주고받는 서비스에다가 렉스를 활용한 챗봇을 가져다

붙일 수 있다. 세 서비스를 비롯한 여러 아마존 AI에 다양한 ‘스킬 셋’이 존재하고, 생태계를 조성한다. 스킬 셋이란, 기존 기술에 붙여 응용할 수 있는 도구다. 생각보다 많은 이가 AWS AI에 관심을 갖고 사용하고 있다는 증거이기도 하다.

3. KT 기가지니

KT 기가지니 역시 API 이용가이드와 SDK를 제공하고 있다[15]. 국내에서도 AI 생태계 구축 경쟁이 달아오르고 있다. KT 기가지니는 API를 포함한 소프트웨어 개발 키트(SDK)를 공개하고, 파트너 사업자들에게 기가지니를 활용한 응용 서비스 개발의 창구를 열어둔 상태다. KT는 또 AI 생태계 조성의 일환으로 ‘미래에셋대우’와 제휴를 맺었으며, 음성인식을 이용한 AI 금융 서비스를 기가지니를 통해 제공하고 있다. 또, 글로벌 전기차 업체인 ‘테슬라’와도 제휴를 맺은 것으로 알려졌다. 텔레매틱스 구축 계약이지만 기가지니가 ‘테슬라’ 전기차에 탑재될 가능성도 높다는 게 업계 전망이다.

4. Kakao 뉴턴

카카오는 다음커뮤니케이션이 자체 개발한 국내 최초 음성 합성 API ‘뉴턴’을 보유하고 있다[16].

그림 7과 같이 입력된 글자를 사람처럼 자연스럽게 읽어 주는 것이 특징으로, 음성 인식 엔진 ‘뉴턴’의 API에 음성 합성 기능을 더하였다. 기존 음성 합성 기술은 대부분 사람들이 들어보고 운율과 발성상태를 표기한 후 이를 기반을 학습하는 수작업에 의존하는 형태였지만, 뉴턴이 자체 개발한 운율 모델은 이 모든 과정을 자동으로 처리하며 다양한 운율을 예측하기 때문에 효율적이다.

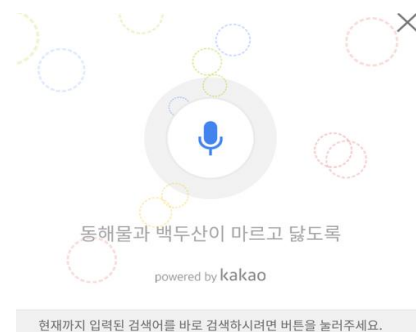


Fig. 7. 카카오 제공 음성 API 뉴턴

또한 다음 검색으로 축적된 데이터를 활용해 가수 ‘2NE1’, ‘B1A4’의 이름과 같이 사전에 등록되지 않은 신조어도 ‘투에니원’, ‘비원에이포’로 정확히 읽고 ‘ㅋㅋㅋ’은 ‘크크크’로, ‘You&I’와 같이 특수문자가 포함된 경우에는 ‘유앤아이’로 읽는 등 유연한 대응이 가능하다. 특히 뉴턴은 한 번에 최대 30초의 음성을 합성할 수 있어 입력된 글자를 분석해 음성으로 합성하기까지 0.1초면 된다. 문장을 입력하고 합성 요청을 하면 즉시 들을 수 있다. 뉴턴에는 현재 남성과 여성 각 하나씩의 낭독 음색이 탑재돼 있으며 대화체나 밝은 낭독체 등 4개의 음색이 추가 될 예정이다. 활용 범위도 무한하다. 예를 들어 현재 ‘100미터 앞에서 우회전입니다’ 라는 형태로 음성 안내하는 내비게이션 앱을 ‘100미터 앞에서 다음 커뮤니케이션 본사를 끼고 우회전입니다’ 수준까지 안내 가능하도록 만들 수 있다. 뉴스나 문자, 책 구절 등을 읽어주거나 각종 안내 방송을 대신할 수 있고, 시각 장애인을 위한 생활형 앱 등을 개발할 때도 유용해 보인다.

5. Naver Clova

네이버의 인공지능 비서 콘텐츠 ‘클로바’도 API와 SDK를 제공하기 위해 기존 자사에서 운영하던 Naver Developers[17]와 독자적으로 제공 중인 Clova페이지[18]를 통해 개발자들의 접근을 환영하고 있다.

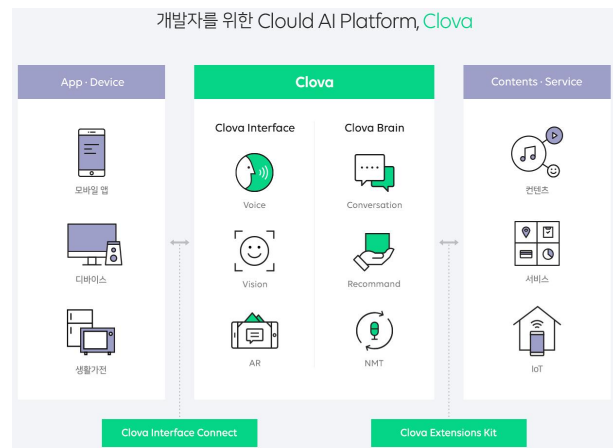


Fig. 8. Clova 개발자 플랫폼

IV. 결론

현재 인공지능 스피커에 관련된 가장 큰 문제점 두 가지가

존재한다. 첫째, 정해진 상황을 벗어나지 못한다. 둘째, 개인 정보의 유출 문제이다.

수많은 자료들 중에서 필요한 데이터만을 뽑아서 활용할 수 있는 빅 데이터를 이용한 딥러닝, 머신러닝을 활용한 ‘알파고’를 소개했지만, 이는 ‘바둑’이라는 게임의 한정이었다. ‘비서’라는 것은 특정 상황을 한정하지 않고 말 한마디에 포함되는 모든 상황을 고려해서 인식해야 한다는 점을 생각한다면 현재의 인공지능 비서 서비스들은 ‘비서’라기 보단 ‘도우미’에 가까운 느낌이다. 컴퓨팅 파워가 높아짐에 따라 스마트폰과 같은 모바일기기에서 다양한 작업을 할 수 있게 된 것처럼, 인공지능 비서 서비스들의 반응속도나 소형화를 통한 이동성도 점차 증가하겠지만 이동성이나 성능적인 측면을 바라보기 전에 우선적으로는 입력받은 데이터를 정확하고 다양한 측면을 비교해서 결론을 내릴 수 있는 기술적인 측면을 먼저 해결해야 한다고 생각한다.

또한 가장 문제가 되는 것 중 한 가지는 바로 ‘개인정보’의 유출이다. 최근 Google이 안드로이드 스마트폰 사용자들의 위치정보를 무단으로 수집하고 있다는 기사가 화두에 올랐다. 사실 대부분의 개발자들은 인식하고 있던 점일지도 모른다. 데이터와 WIFI 정보만을 갖고 위치 정보를 받아오기에는 정확도가 떨어지기 때문이다. 이 뿐만 아니라 Google에 로그인 된 기기에서는 검색을 할 경우 검색어가 동기화되어 Google에 전송되어 맞춤 광고 혹은 맞춤 콘텐츠에 활용되기도 한다. 사용자의 입맛에 맞는 개인 비서를 만들기 위해 사용자 개인의 데이터를 활용한다는 것이기 때문에 이는 어쩔 수 없는 선택이라 생각된다. 과연 인공지능에게 교육을 시켜 사용자 개인에게 맞춤형 서비스를 제공해야 하는 회사들의 입장에서 개인정보 유출의 문제없이 어떤 식으로 이를 해결할 것인지에 관심이 쏠리는 대목이다.

아직까지 4차 산업혁명은 ‘소수’의 사람들을 위한 것이라고 착각하는 사람들이 많다. 이는 현재에 만족하는 사람들일 것이며 인터넷과 같은 정보의 바다를 활용하지 못하는 사람들일 것이다. 물론 인터넷을 활용할 수 있는 기기를 잘 다루지 못한다고 못살아갈 일은 없을 것이다. 하지만 새로운 산업의 문이 열렸을 때가 바로 기회이다. 일반인, 학생, 직장인, 개발자 할 것 없이 똑같은 출발선 위에 놓인 것이다. 이제 시작이다. 새로운 산업혁명을 맞아 내가 변화할 것인지, 변화당할 것인지는 본인의 선택에 달려있음은 분명하다.

REFERENCES

- [1] https://www.youtube.com/watch?v=_jT7c18ITPo
- [2] <https://namu.wiki/w/알파고>
- [3] <http://www.irobotnews.com/news/articleView.html?idxno=8729>
- [4] <http://thegear.co.kr/15225>
- [5] www.hani.co.kr/arti/international/international_general/736844.html
- [6] 김영석, 윤종원, 임성수, 조성배, “효과적인 개인 관리 서비스를 위한 대화 기반 감정형 인공비서의 구현,” KIISE 가을 학술발표논문집 Vol.37, No.2(C), 2010
- [7] 김동주, 이가람, 김호원, “AI를 활용한 음성인식 기법 및 개인비서 서비스 기술 분석,” 2017년 한국통신학회 하계종합학술발표회, pp. 213-214, 2017
- [8] 최승주, 김종배, “음성 인식 오픈 API의 음성 인식 정확도 비교 분석,” Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology Vol.7, No.8, pp. 411-418, August 2017
- [9] <http://www.hankookilbo.com/v/5e265a52728a406fbb87824be39fd5d5>
- [10] <http://www.ciokorea.com/news/33584>
- [11] <http://www.mcnorton.com/csLab/165386>
- [12] www.fnnews.com/news/201704041950060154
- [13] <http://www.jeonpa.co.kr/news/articleView.html?idxno=37769>
- [14] <http://it.chosun.com/news/article.html?no=2830265>
- [15] <https://svcapigigagenie.ai/guide/pdf>
- [16] <https://developers.kakao.com/docs/android/speech>
- [17] <https://developers.naver.com/products/clova/vrecog/>
- [18] <https://clova.ai/ko/ko-platform-api.html>

저 자 소 개



하 진 옥
2018: 인하공업전문대학
컴퓨터정보과 전문학사
현재: 인하공업전문대학
컴퓨터정보공학과
관심분야: 컴퓨터공학, 인공지능



김 상 혁
2018: 인하공업전문대학
컴퓨터정보과 전문학사
현재: 인하공업전문대학
컴퓨터정보공학과
관심분야: 컴퓨터공학, 인공지능
