

셀바스AI, '인터넷 연결 없이도' AI로 음성 구현

[목차]

1. 들어가며
2. 인간과 AI 상호작용 이끄는 '셀바스AI'
3. 이항섭 셀바스AI HCI 사업 대표와 일문일답
4. 인터뷰이 소개
5. 셀바스AI 주요 사업 소개



‘챗GPT’로 대표되던 생성형 인공지능(AI) 시장을 넘어 다양한 산업 영역에 AI 전환(AX)을 이끄는 버티컬 AI까지. 우리 일상생활과 산업에 AI가 깊숙이 들어왔다. 방송채널사용사업자(PP)를 기준으로 정보기술(IT)을 다루는 시사·교양 채널은 물론이고 드라마, 예능 프로그램과 방송광고 등 다양한 콘텐츠를 통해 사람들은 AI를 자주, 쉽게 접하고 있다.

영화 《중간계》는 스펙터클(spectacle)한 장면 대부분을 AI로 제작, 표현의 한계를 최소화하고 작업 효율을 극대화했다. 이승과 저승 사이 중간계를 시각적으로 그려내는 데 기존 컴퓨터그래픽(CG) 이상의 퀄리티와 결과물을 만들어낸 것이다. 영화 《어쩔 수가 없다》에서는 다수의 사람이 일하던 제지공장 근무 환경을 AI가 컨트롤하면서 단 한 명의 관리자만 살아남은 모습을 보여주기도 했다.

비단 방송이나 영화 등 콘텐츠뿐만 아니다. AI는 일상생활에도 스며들었다. 기존 네이버·다음 등 포털 중심 검색에서 유튜브의 검색 비중이 늘어난 데 이어 근래에는 생성형 AI를 통한 검색부터 정리까지가 너무나도 당연하고 일상적인 시대가 됐다. ‘시리야’ ‘하이 빅스비’로 대표되는 스마트폰에 탑재된 AI 비서도 거듭 진화하고 있다. 사람의 음성을 듣고 정확하게 글로 옮겨적는 수준에까지 다다랐다. 바로 음성 AI 기술 발전 덕분이다. 음성 AI는 사용자의 음성 입력을 텍스트로 변환하고 이를 기반으로 의미를 분석하는 기술이다. 초기에는 제한된 단어만 인식했지만, 현재는 자연어 문장, 사투리, 배경 소음이 있는 환경에서도 높은 정확도를 보인다. 크게 글자를 말소리로 전환하는 음성합성(TTS) 기술과 대화 등 음성을 텍스트로 정리하는 음성인식(STT) 기술이 있다. 목소리의 다름을 기반으로 대화에 참여한 사람은 몇 명이고 각자 무슨 발언을 했는지까지 일목요연하게 정리할 수 있는 기술로 발전하고 있다.

이러한 기술 발전 속도에 기반해 세계 음성 AI 시장은 지난해 약 34억 7,000만 달러(약 5조 735억 원)에서 2032년 1,370억 달러(약 200조 3,351억 원) 규모로 연평균 15% 내외로 급성장할 것으로 전망되고 있다. 스마트 기기 보급이 확대되고 자연어 처리기술이 발전하며 원격근무 수요가 증가하는 등 음성 AI 시장의 빠른 성장은 예고된 상황이다. 음성합성, 음성인식, 음성 어시스턴트, 음성 복제 등 다양한 분야에서 활용도 확산되고 있다. 거대언어모델(LLM) 분야에서 이미 촉발된 국경 없는 글로벌 경쟁이 음성 AI 분야에서도 동일하게 이뤄지고 있다.



국내 음성 AI 시장에서도 국내 기업과 외국계 기업 간 피할 수 없는 진검승부가 시작됐다. 영국에 본사를 둔 글로벌 오디오 AI 기업 ‘일레븐랩스’가 한국에 진출한 것이다. 마티 스타니스제프스키 일레븐랩스 최고경영자(CEO)는 2025년 11월 말 방한해 전통 더빙 방식의 한계를 극복하기 위해 회사를 설립한 배경과 차세대 인터페이스로서 ‘보이스’ 비전, 한국 시장 진출 등을 발표했다.

일레븐랩스는 회사 설립 초기 명확하고 감정 표현이 풍부하며 다국어를 지원하는 AI 음성 기술로 출발했다. 현재는 음성, 음향 효과, 음악 전반에 걸쳐 AI 오디오 기술을 개발하고 고도화하며 다양한 제품을 구현할 수 있게 지원하고 있다. 설립 2년 만에 기업가치 11억 달러(약 1조 6,072억 원)를 돌파한 ‘유니콘 기업’ 일레븐랩스와 새해 본격적으로 경쟁하게 될 우리나라 음성 AI 기업 중 하나인 셀바스AI의 사업 전략과 방향·현황을 통해 한국 음성 AI 시장의 현재와 미래를 조망해 보자.



인간과 AI 상호작용 이끄는 ‘셀바스AI’

셀바스AI는 1999년 설립된 코스닥 상장기업이다. ‘인간과 컴퓨터 간 상호작용(HCI·Human-Computer Interaction)’을 중심으로 인공지능(AI) 기술을 연구·개발하고 상용 솔루션으로 사업화하는 AI 전문 기업이다. AI 기반 기술과 도메인별 특화 데이터 접목을 통한 제품화로 AI 시장 선도를 목표로 한다. ‘음성 AI’ 기술에 특히 강점이 있다. 음성인식(STT), 음성합성(TTS), 필기 인식(Pen), 광학문자인식(OCR) 등 다양한 인식 기술을 기반으로 하며 이를 여러 산업 도메인에 융합해 제품화하고 서비스화하는 사



출처 : 셀바스AI

업 모델(BM)을 추구한다. HCI 기술을 기반으로 솔루션과 응용 제품화를 통해 의료, 출판, 교육, 헬스케어, 금융 등 다양한 산업 분야를 연계하고 있다.



이항섭 셀바스AI HCI 사업 대표와 일문일답

Q. 이스트소프트의 주요 사업을 소개해 주세요.

다양한 디바이스 산업의 비약적인 기술 발전으로 정보 입력 과정의 패러다임이 크게 전환될 것을 예상했습니다. 그래서 음성 AI 사업에 자연스럽게 집중할 수밖에 없었죠. 스크린 터치나 타이핑 등 전

통적인 입력방식을 넘어 애플리케이션(앱)과 웹 모든 환경에서 AI 기반의 음성인식(STT), 음성합성(TTS) 기술을 활용해 가장 완벽한 AI 음성 인터페이스를 구현하고 사용자경험(ux)을 크게 개선하고자 시도했습니다.

Q. 셀바스AI도 거대언어모델(LLM)까지는 아니지만 소형언어모델(SLM)과 같이 경량화된 자체 언어모델 개발 계획이 있죠?

최근 STT의 경우에는 인식 정확성, TTS의 경우에는 발화 품질 등 기반 기술의 핵심 성능을 대폭 향상할 방안으로 SLM을 활용하는 방법이 주목받고 있다는 것을 알고 있습니다. 셀바스AI 역시 의료와 공공 분야에서 활용할 수 있는 SLM 개발에 이미 착수한 상황인데요. 자체 개발하는 것 외에도 다양한 산업 영역의 업무, 작업 환경에서 제대로 활용될 수 있도록 외부 모델을 연계, 조합해 최적화하는 ‘오케스트레이션 전략’으로 경쟁력을 강화하고 있습니다.

Q. SLM 개발 방향을 좀 더 구체적으로 설명해주신다면요?

셀바스AI가 AI 언어모델에 특화된 회사는 아닙니다. 저희는 저희가 필요하거나 필요로 하는, 예를 들어 음성인식·합성기 쪽에서 활용할 수 있는 기존 언어모델을 가져다가 파인튜닝하는 게 대표적인데요. 인식과 합성의 모델 방향성이 다르다는 차이는 있습니다. 음성인식은 말이 끝나고 텍스트를 추출한 이후에 일이 끝나는 게 아니라 인식한 결과물을 잘 정리하고 요약해서 고객이 음성인식을 통해 이루고자 하는 최종 목표까지 요약하고 정리하고 제시해주는 것까지를 대부분 원합니다.

그래서 저희는 요약에 특화된 언어모델을 맞춰가는 방향으로 개발하고 있습니다. 모든 분야를 동시에 하기는 힘들고 우선 의료 분야로 제품 개발 로드맵을 가지고 있습니다. 연내에 개발해 대형 병원에 적용이 될 것 같습니다. 의료 분야 음성인식 관련 자사 메디보이스가 96~97% 정도 성능을 내고 있는데요. 모든 부분은 아니고 특정 진료과의 이해도가 높습니다. 의사 등 전문가들이 쓱 보고 일부 수정만 하면 될 정도로 고도화가 돼 있고요. 파인튜닝과 학습을 통해 음성 AI가 잘 이해하고 답변할 수 있는 진료과목을 점점 더 늘려나갈 예정입니다. 중장기적으로는 전체 진료과로 확대하겠다는 거죠.

Q. 그렇다면 음성합성 영역에서 SLM 개발 방향은 무엇인가요?

음성합성은 입력값 자체가 텍스트입니다. 텍스트를 말로 변환해주는 것이다 보니 텍스트 내용을 이해하고 어떤 감성과 감정인지 텍스트 내용을 기반으로 확인할 수 있게 해야 하고 그런 조건들을 개

〈그림 3〉 셀바스AI의 음성 AI 기술

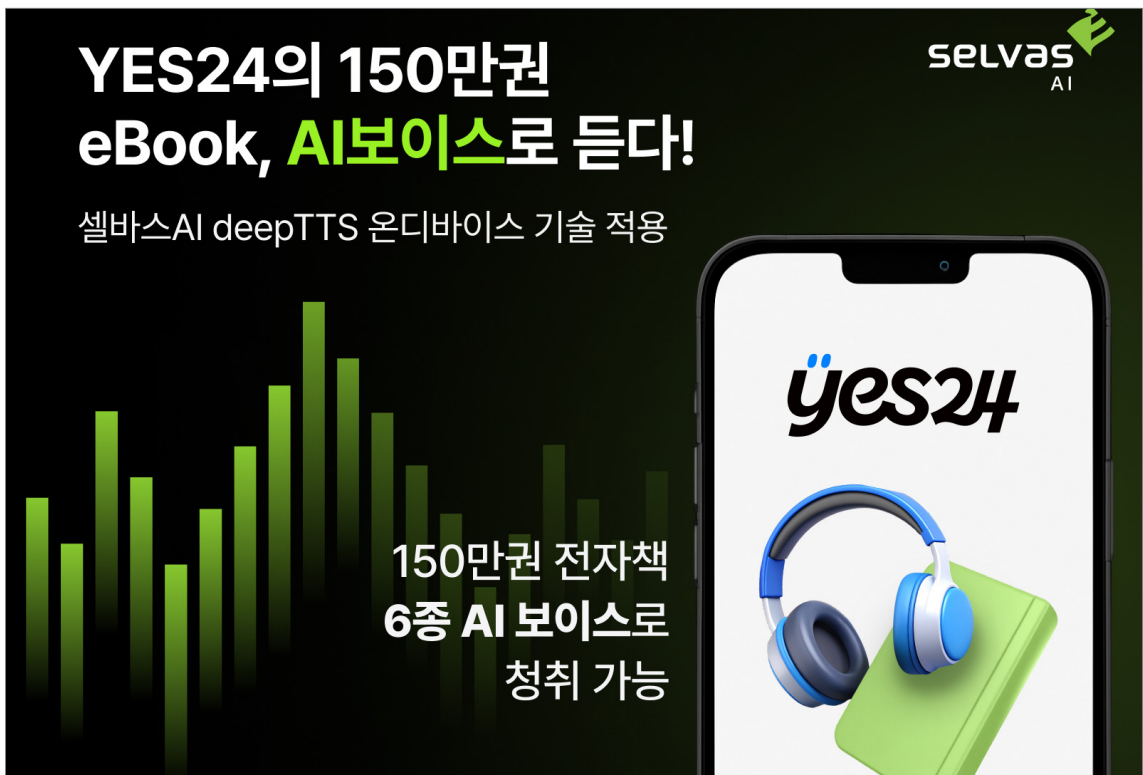


출처 | 셀바스AI

발 앞단에서 고민하고 있습니다. 다시 정리하면 음성인식은 텍스트가 결과물이고 그 결과물을 요약 정리하는 것에, 음성합성은 텍스트를 어떻게 잘 말하고 표현할 수 있을지에 SLM 개발 방향이 맞춰져 있다고 보시면 되고요. 또 산업이나 전문가 영역 등 도메인별로 단어나 용어는 당연히 학습해야 하고 정확률과 인식률을 높이기 위한 반복 학습이 중요합니다. 때에 따라선 사투리도 인식해야 하고 할 수 있어야 하고요. 내년부터는 텍스트로 된 말뿐만 아니라 도표 등 다양한 서식에 있는 내용을 인식할 수 있게 연구를 지속하고 있습니다.

Q. 셀바스AI는 현재 밀리의서재, 리디, 예스24 등 주요 전자책 플랫폼에 오디오북 기술을 지원하고 있는데요. 사업 수주 배경은 무엇인가요?

셀바스AI는 국내에서 20년 이상 AI 사업을 지속하고 있는 기업입니다. 장시간 우리가 보유한 원천기술을 기반으로 한 B2B 솔루션 사업을 진행하다 보니 자연스럽게 여러 산업군의 고객을 만나며 그들이 처한 다양한 문제들을 AI 기술로 해결했는데요. 이 과정에서 지속적인 R&D 투자를 통해 기술을

An advertisement for YES24's AI voice service. The background is dark green with a bar chart of varying heights. On the right, a smartphone displays the YES24 logo and a pair of blue headphones resting on a green book. The text is in white and green. The main headline reads 'YES24의 150만권 eBook, AI보이스로 듣다!'. Below it, in smaller text, is '셀바스AI deepTTS 온디바이스 기술 적용'. To the right of the bar chart, it says '150만권 전자책 6종 AI 보이스로 청취 가능'. The SELVAS AI logo is in the top right corner.

YES24의 150만권 eBook, AI보이스로 듣다!

셀바스AI deepTTS 온디바이스 기술 적용

150만권 전자책
6종 AI 보이스로
청취 가능

selvas AI

출처 | 셀바스AI

발전시킨 것은 물론, 산업별 특징에 대한 지식과 인사이트를 상당량 축적했습니다. 이로써 다른 AI 기업보다 고객사 커스터마이징과 제품화 역량을 빠르게 키울 수 있었어요.

이것이 전자책과 오디오북 시장 전반에 셀바스AI의 TTS 기술을 공급, 적용할 수 있었던 가장 큰 이유라고 생각합니다. 실제 셀바스AI는 고품질 TTS 솔루션을 단말 내에서 독립적이고 안정적으로 활용하고자 하는 업계의 수요를 반영해 모델과 엔진 경량화에 성공했고요. 국내 업계 최초로 온디바이스용 deepTTS솔루션을 주요 플랫폼을 통해 상용화했습니다.

Q. 3곳의 오디오북 서비스 플랫폼과 각각 어떤 기술 협력과 제휴를 하고 있나요?

밀리의서재, 리디, 예스24 등 3개 플랫폼 기업과 사업 형태는 거의 유사합니다. 각 서비스 모두 셀바스AI의 deepTTS 솔루션을 통해 ePUB 형태의 모든 전자책을 오디오 콘텐츠로 청취할 수 있는 기능을 제공하고 있어요. 별도 오디오북 콘텐츠가 제작돼 있지 않은 전자책도 오디오 콘텐츠로 청취할



출처 | 셀바스AI

수 있게 해서 유저들에게 새로운 독서 경험을 제공하고 있습니다. 3사 모두에 공급된 deepTTS 솔루션은 아직 최종 기술 수준에 도달한 상태가 아니며 AI 보이스의 표현력, 음질 등에서 지속 고도화가 진행 중인데요. 해당 서비스를 사용하는 사용자들은 앞으로 더욱 재미있는 음성 콘텐츠를 즐길 수 있을 것입니다.

Q. 그 외 다른 기업과 사업 내용도 공유해주신다면요?

교보문고를 포함한 다양한 전자책 플랫폼을 고객 또는 파트너사로 확보하고 있습니다. 자사 deepTTS 기반 AI 오디오북 제작 서비스 ‘오디로’를 통해 ‘오디언소리’와 같은 전문 오디오북 제작사 또한 고객사로 확보하고 있어요. 최근에는 아동도서를 활용해 디지털 교육 서비스를 전개하고 있는 에듀테크 기업에도 TTS 활용 수요가 있는데요. 더 많은 사업 기회가 있을 것으로 판단하고 있습니다.



출처 | 셀바스AI

Q. 오디오 플랫폼 분야 외에 음성 AI 사업 모델에 대한 설명과 향후 확대될 사업 분야가 있다면요?

오디오 플랫폼에 한정되지 않고 사업을 진행 중입니다. ‘콘텐츠 산업 전반’으로 이미 확대하고 있습니다. 음성 AI 기술은 디바이스, 서비스 등 현재 우리가 사용하는 디지털 매체에서 기존과는 전혀 다른 사용자 경험을 제공하며 ‘입력’ 행위에 대한 변화를 넘어 ‘콘텐츠 제작’에 대한 패러다임 자체를 바뀌어 나가고 있습니다. 셀바스AI 또한 음성 AI 기술로 도서 외에도 음악, 영상, 게임 등 다양한 콘텐츠 플랫폼의 제작과정을 효율적으로 변화시켜 시중에 유통되는 콘텐츠들을 지속적으로 다양화하고 전반적인 품질 또한 성장시켜 콘텐츠 산업의 모든 이해관계자가 만족할 수 있는 환경을 만들 계획입니다.

Q. 개인적으로 주목하고 있는 AI 분야가 있으세요?

현재 가장 주의깊게 보고 있는 분야는 도메인별 영역과 인간 수준의 기억력과 공감 능력을 기반으로



출처 | 셀바스AI

원활한 대화가 가능한 'AI 컴패니언' 시장입니다. AI와 사람이 직접 대화하는 시장이 뜨고 있기 때문에 음성 AI가 중요한 역할을 할 것 같습니다.

※부연 설명: 글로벌 시장조사업체 마켓 컴퍼니 인사이트에 따르면 세계 AI 컴패니언 시장 규모는 2025년 기준 약 1억 2,000만 달러(약 1,755억 원)로 빠르게 성장 중이며 2034년까지 연평균 26.8%의 고성장이 예상되는 시장입니다. 정신건강, 웰빙, 정서적 지원 등 다양한 분야에서 활용되면서 수요가 늘고 특히 젊은 세대와 디지털 친화적 사용자를 중심으로 인기가 높습니다.

Q. 셀바스AI의 음성 AI 기술이 가진 차별화 포인트와 특징점은 무엇인가요?

음성합성 기술이 제공하는 사용자경험의 품질 차이가 궁금합니다.

우선 온디바이스 음성 기술을 국내 최초로 상용화해 인터넷 연결 없이도 사람의 호흡, 억양, 감정 등 음성에 포함된 다양한 요소를 정교하게 구현할 수 있게 모델링돼 있다는 게 특징입니다. 합성 모델 구조를 경량화하고 음성 데이터를 더욱 세밀하게 분석, 예측할 수 있도록 설계했으며 경쟁 학습 방식의 도입으로 생성되는 음성의 자연스러움과 발음의 정확성을 대폭 끌어 올렸기 때문입니다.

기술력 외에도 20년 이상 사업을 진행하며 콘텐츠 외에 금융, 교육, 제조 등 다양한 산업에 종사하는 고객사와 협업, 다양한 공급 경험을 쌓았기에 국내 다른 음성 AI 기술을 보유한 기업보다 시장

과 산업에 대한 인사이트를 확보하고 있어 이를 토대로 고객마다 상이한 요구사항을 빠르고 정확하게 반영해 다양한 형태로 커스터마이징한 최적의 상태로 기술을 제공할 수 있다는 것도 강점입니다.

Q. 현재 TTS, STT 등 음성 AI 서비스를 위해 SLM뿐만 아니라 국내외 기업의 AI 모델 또는 API(응용프로그래밍환경)를 활용하고 있나요?

셀바스AI는 오픈AI ‘GPT’, 구글 ‘제미나이’, 엔트로픽 ‘클로드’ 등 다양한 LLM을 적극 활용해 ‘셀바스 노트(SELVAS Note)’ ‘메디보이스(MediVoice)’ ‘deepTTS’ 등 음성 기술 기반 제품에 요약, 답변 생성 등 지능형 기능을 추가해 개발·구현하고 있습니다. 특히 STT 영역에서는 의료 및 공공산업 분야에 특화된 요구를 충족하기 위해 이러한 모델을 기반으로 앞서 설명해 드렸듯이 산업 맞춤형 SLM 개발을 자체적으로 진행 중입니다. 이와 함께 단순한 부가 기능 구현을 넘어 언어 모델을 활용해 기존 음성 기술의 성능 자체를 고도화할 수 있는 다양한 R&D도 병행하고 있습니다. 이를 통해 STT와 TTS의 인식 정확도와 발화 품질을 개선하고 실제 산업 환경에서 더 높은 효율과 사용자 경험을 제공하는 것을 목표로 하고 있습니다.

Q. AI가 텍스트를 읽어주는 음성합성은 이미 너무 당연한 일이 돼버린 것 같아요. 결국 경쟁력은 문맥 속 감정과 대화 감성 등 표현일 것 같은데요. 이 관련 준비도 진행되고 있나요?

우선 합성기가 입력된 텍스트, 그러니까 소설책이 입력되었다고 가정하면, 소설책에 나온 등장인물의 대사를 음성으로 발화하기 전에, 대사의 앞 뒤에 붙어있는 설명문의 내용을 문장 단위로 분석하여 그 속에 내재된 의미와 감정 등을 이해하고 이것을 음성으로 어떻게 발화할지 스스로 결정하여 출력할 수 있는 기술을 연구 중입니다. 기존 합성기는 단순히 소리를 자연스럽게 내는 데에만 초점을 맞추고 있기 때문에 이런 부분까지 스스로 판단할 수는 없어서 대사를 다소 밋밋하게 낭독하는 게 문제가 되었는데, 이 문제를 해결하기 위해 문장 분석을 진행하고 이 대사에 내재된 감정이 무엇인지를 우선 확인합니다.

다만 몇 가지 문장만 가지고는 맥락을 완벽하게 이해하기 어려운 상황도 있습니다. 똑같은 내용이지만 앞뒤 문맥에 따라서 감정이 다를 수도 있기 때문인데요. 때문에 문장이 아니라 문단 전체를 참고하여 전반적인 문맥을 파악하고 이에 대한 이해를 높이는 방향으로 개발하고 있습니다. 연내 고객사에 데모를 보여줄 수준까지 고도화했고 내년 1분기부터 사업화를 하나갈 계획입니다.

또 ‘Voice Style Gennerate(음성 및 스타일 생성)’라고 있는데요. 합성기가 사용자의 요구사항과 함께 입력된 텍스트의 문단과 문장에 내재된 문맥과 의미를 해석해 자동으로 적합한 보이스와 발화 스타일을 뽑아내는 것을 목표로 하는 기술입니다. 말 그대로 명령어로 합성기에게 내가 원하는 목소리를 주문하고 감정을 어느 수준으로 표현하게 할 것인지 조정도 할 수 있는 방향을 고민하고 있다는 취지입니다.

Q. 분량이 방대하면 모든 문장에 일일이 감정을 주입할 수는 없을 것 같은데요?

맞습니다. 특히 오디오북 플랫폼은 책을 읽어주는 실시간 서비스가 가장 핵심 경쟁력인데요. 감정을 어떻게 담아내고 소리 내야 할지 자세한 코멘트를 늘 일일이 남길 수 없다는 제약이 분명히 존재합니다. 이런 부분에 대해 고민하고 있고요. 일단 자동으로 적용될 수 있게 개발하고 고도화해 나가는 게 목표입니다. 또 일상생활에서는 부자연스러운데 소설에선 자연스러운 대화 맥락도 있기 때문에 문장마다 사용자가 디테일한 요구를 하긴 어려운 게 사실입니다. 다만 전자책(e북)을 제작하는 기업에서 필요한 기능이긴 합니다. 우리가 운영하고 있는 AI TTS 기반 오디오북 제작 서비스인 ‘오디로’에서는 이런 사용자(독자들)의 디테일한 요구사항을 모두 반영하여 오디오북을 제작할 수 있게 하는 편의 기능을 개발, 적용하는 것을 검토하고 있습니다.

Q. 현재 애플이나 삼성의 무선 이어폰에서 즉각 통역을 지원하는 등 온디바이스 형태로 음성 AI 서비스가 발전하고 있는데요. 음성 AI 시장이 향후 어떻게 발전하고 고도화될 거라고 전망하시나요?

산업 영역에 따라 다르겠지만 대부분의 AI 기술은 안정적인 성능을 보장하기 위해 네트워크 환경이나 상태와 관계없이 항상 동작할 수 있는 온디바이스화에 집중할 것으로 생각합니다. 특히 LLM, 시각언어모델(VLM) 등과 연계해 산업 맞춤형, 개인 맞춤형 AI 에이전트를 구현하는 데 도움을 줄 수 있는 방향으로 고도화될 것이라고 보고 있어요. 음성, 문자, 영상 등 물리 요소를 인식하고 분석해 디지털 형태로 구현하는, 이른바 HCI(Human-computer interaction) 기술이라 불리는 ‘인식 AI’ 기술입니다. 인식 AI 기술은 현재 다양한 언어모델과 연계해 다시 한번 기술적 전성기를 맞이하고 있습니다. 음성 AI 시장은 이러한 언어모델과 연계해 작년부터 시장에서 지속적으로 언급되고 있는 다양한 산업 영역 속 AI 에이전트를 구현하는 방향으로 특화할 것으로 보입니다.

다만 AI 에이전트 구성 요소 중 음성 기술과 달리 LLM, VLM 등은 아직 온디바이스 형태로 제공할

〈그림 8〉 셀바스AI 기술 전시 현장



출처 | 셀바스AI

〈그림 9〉 셀바스AI 음성합성(TTS) 기술과 서비스 전시



출처 | 셀바스AI

수 있는 수준의 경량화 기술이 완성되지 않았기에 온디바이스 형태로 제공되는 AI 에이전트를 실질적으로 이용, 활용할 수 있는 시점까지는 시간이 더 걸릴 것으로 예상됩니다.

Q. 또 미디어 산업 내 음성합성 기술의 현재와 미래 관련 전망을 한다면요? 미디어·콘텐츠 기업이 셀바스AI의 음성합성 기술을 어떻게 활용하고 있는지, 향후 음성합성 기술이 가져올 수 있는 혁신과 변화는 무엇일까요?

현재 셀바스AI는 우리 기술을 도입해 사용하고 있는 고객사 모두가 음성 기술을 통해 콘텐츠 제작 과정 효율화, 사용자에게 새로운 서비스 경험 제공 두 가지 목적을 달성하는 것을 돕고 있습니다. 앞으로도 마찬가지입니다. AI 보이스는 콘텐츠의 생산 속도 향상, 제작 비용 절감 등 두 가지 부분에서 강점이 있는데요.

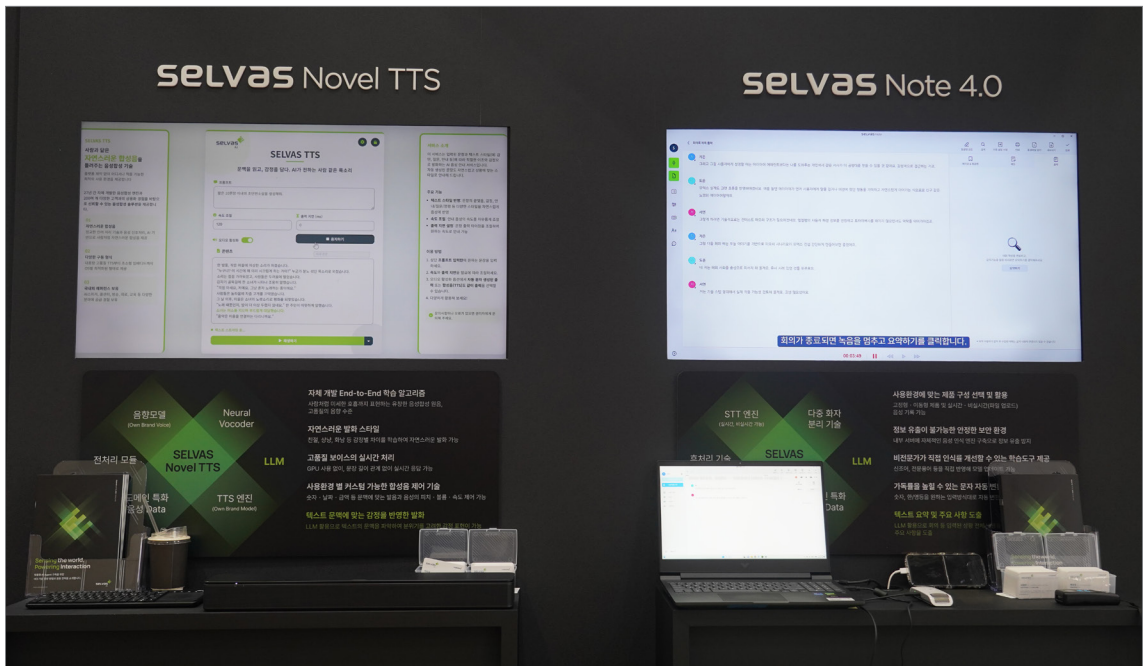
고객사가 셀바스AI의 TTS 기술을 활용해 생성되는 AI 보이스로 고품질 콘텐츠를 기존보다 낮은 비용으로 더욱 빠르게 만들 수 있는 환경을 구축하는 'AX(AI 전환)'에 성공한다면 미디어 산업 소비자들도한 이전보다 훨씬 만족스러운 콘텐츠 소비활동을 할 것이라고 보는 거죠. 이 두 가지 목적을 모두 달성하고 콘텐츠 산업을 다시 한번 활성화할 수 있을 것으로 기대합니다.

Q. 외국어 지원 현황이 궁금합니다. 향후 확대 계획이 있을까요?

외국어에 대한 플랜이 구체적이진 않고 고객사 수요에 맞춰서 대응할 계획입니다. 우선 음성합성(TTS) 영역에서는 8개 언어를 지원하는데요. 한국어, 영어, 중국어, 일본어, 스페인어, 프랑스어, 포르투갈어 그리고 영국식 영어인데 확장도 고려하고 있습니다. 제공되는 언어별 감정 표현도 모두 적용하는 걸 목표로 하고요.

음성인식(STT)은 한국어, 영어, 일본어, 중국어 정도입니다. 실제 상업적으로 외부에 공급한 기술은 한국어와 영어입니다. 국내 공공사업에서는 당연히 한국어 수요가 대부분이고 병원 등 의료 현장에서는 영어 수요가 있잖아요. 다만 일본어와 중국어 수요는 낮은 상황이라 기술이 한국어나 영어 수준만큼은 아직 올라오지 못하고 있어요. 향후 시장 수요에 따라 전략적으로 지원 언어를 확대할 계획입니다.

〈그림 10〉 셀바스AI의 노블 TTS와 노트 4.0 기술 전시



출처 | 셀바스AI

Q. 마지막으로 셀바스AI의 음성 AI 기술과 서비스 고도화 로드맵에 대해 알려주세요.

현재는 ‘표현력 강화’ 그리고 ‘개인화’에 집중하고 있습니다. 그동안은 정확하고 자연스러운 발음, 깨끗하고 명료한 음질 등 정확한 발화를 통해 사용자에게 명확한 정보와 메시지를 전달하는 것에 초점을 맞춰 개발했는데요. 그래서 AI컨택센터(AICC), 키오스크, 오디오북 서비스 등 사용자가 불편함이 나 어색함 없이 음성을 듣고 AI와 상호작용을 할 수 있게 됐습니다. 다만 재미, 몰입감 등 정보와 메시지 전달 이상의 가치를 제공하기 위해서는 이제 AI의 표현력이 대폭 향상되어야 할 필요가 있다고 생각했어요. 그래서 앞서 말씀드린 것처럼 감정을 표현할 수 있는 AI로 진화를 추진하고 있습니다. 셀바스AI는 문장과 문맥, 비언어적 표현을 이해하며 인간의 화행(話行)을 구현할 수 있도록 모델과 엔진을 고도화하고 있는데요. 이미지, 텍스트를 생성하는 것처럼 사용자의 지시에 따라 특정한 스타일을 가진 음성을 직접 생성할 수 있는 음성 생성 기술 또한 연구개발을 진행하고 있습니다. 이러한 음성 생성 기술을 통해 AI 음성이 필요한 어떤 환경에서든 사용자 개인이 원하는 스타일의 보이스를 직접 생성해 활용할 수 있도록 해 이전과는 또 다른 차별화된 기술 경험을 제공할 계획입니다.

특히 음성합성의 경우 내년 1분기에는 문장 이해를 기반으로 한 감정 표현, 하반기에는 문맥에 대한 이해로 감정 표현을 할 수 있도록 고도화할 계획입니다.

part 4 인터뷰이 소개

셀바스AI HCI(Human Computer Interaction) 사업 대표이자 전무다. 약 23년간 AI 음성 기술 분야에 몸담은 전문가다. 셀바스AI의 고품질 음성합성(TTS) 기술을 기반으로 공공부문, 전자책 서비스, 로봇, AI 스피커, AI 고객센터 등 다양한 시장을 선도하고 있다. 지난해 개발한 온디바이스(On-Device) AI 서비스를 통해 개별 디바이스에서도 네트워크 기반 프로그램 수준의 고품질 음성 서비스를 구현했다. 이를 통해 셀바스AI는 AI 음성 기술의 경량화·고도화를 동시에 달성하며 국내외 AI 서비스 경쟁력을 강화하고 있다.

part 5 셀바스AI 주요 사업 소개

인공지능(AI) 분야에서는 음성인식(SELVAS STT) 서비스를 제공한다. 사용자 음성 발화를 문장이나 명령어 형태의 텍스트로 변환하는 기술로 자유 발화 또는 연속 음성 인식 등 다양한 환경을 지원한다. 화자 분리, 잡음 환경 대응, 키워드 탐지와 후처리 기능 등을 제공한다. 스마트 디바이스 제어, 회의록 작성, 상담 녹취 등 다양한 응용에도 적용된다.

음성합성(SELVAS TTS·SELVAS deepTTS)은 문자를 자연스럽게 사람다운 음성으로 합성하는 기술이다. 특히 앱 내 탑재 방식인 온디바이스로 합성 지원을 강화해 서버 통신 없이도 음성을 생성할 수 있게 해 보안·지연 문제를 줄이고 비용을 절감할 수 있다. 감정 표현, 숨소리, 억양 제어 등 음성의 자연스러움과 개성 표현 부분에도 기술력을 갖추고 있습니다. 음성 AI 기술은 단독으로 사업화는 물론, 여

러 도메인 솔루션의 핵심 엔진으로 활용된다.

또 셀바스 노트는 회의·상담 등 여러 화자의 대화를 자동으로 인식하고 화자를 분리해 기록을 작성하는 솔루션이다. 메디보이스는 의료 환경에서 의사·환자 간 대화를 녹취하고 자동으로 텍스트화하는 의료 녹취, 보이스 EMR 솔루션이다. 셀바스 펜은 디지털 디바이스를 통한 사용자의 필기를 텍스트 입력으로 변경하는 기술이다. 디지털 교육과 디지털교과서에 적용되고 있으며, 특히 키보드로 입력이 어려운 수식 입력과 인식 등에 적극 활용되고 있다.

이외에도 음성인식, 필기인식 등 AI 기반 인식기 외에도 계열사들과 의료기기에 AI를 접목한 AI 의료 사업을 진행 중이다. 계열사 메디아나, 셀바스헬스케어의 기존 하드웨어(HW) 중심 제조, 판매 기반에서 AI 의료 솔루션과 서비스(SW)로 사업모델을 확대할 전략도 갖고 있다. 셀바스AI 기술력을 접목해 환자감시장치, 자동심장충격기 등과 연결성을 강화한 의료기기 라인업을 확대하고 진단 예측 분야 모델을 준비 중이다.

