

딥러닝 기반의 서버형 음성인식 기술



딥러닝 기반의 서버형 음성인식 기술

Overview 03

비즈니스 아이디어 20

사업화 대상 기술 07

사업화 지원 24

Trend 13

배경 및 필요성 : 음성인식

- 디바이스와 인간 간 인터페이스 기술 중 터치/텍스트 인식 대비 **인간의 커뮤니케이션 형태(쌍방향)와 가장 유사하며, 입력속도, 편의성, 센서가격 및 범용성이 우수하고, 멀티태스킹이 가능한 음성인식 기술 활용범위 확산 추세**
- **딥러닝 등 인공지능 기술의 발전으로 음성 인식율이 개선됨에 따라 터치, 제스처 인식, 아이트래킹 등의 대체 기술은 열악한 소음 환경 등에서만 제한적으로 사용 예상**

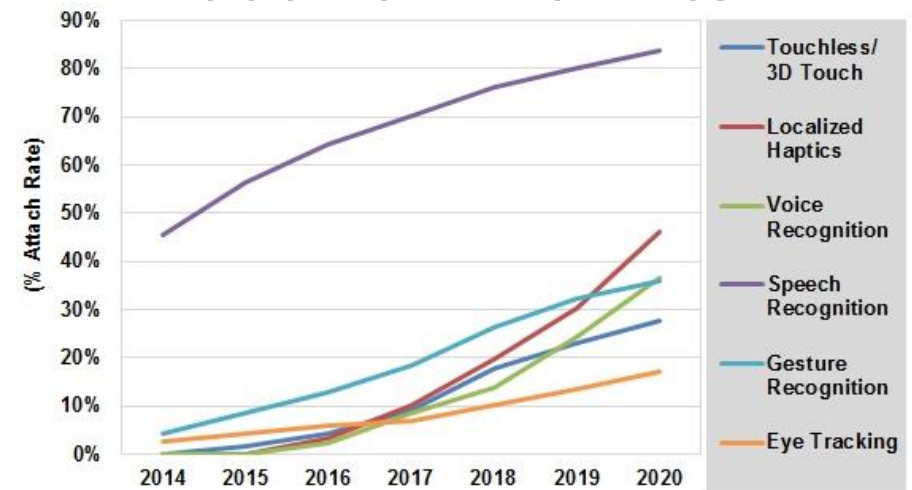
음성인식과 터치/텍스트 기반 HMI* 기술 비교

구분	음성인식	터치/텍스트 인식
입력속도	150단어/분	40단어/분
입력방식	Random Access	Hierarchical GUI
정보의 입출력	마이크로폰, 스피커	키보드, 디스플레이
커뮤니케이션 방식	컨텍스트 기반 쌍방향	단방향
멀티태스킹	Hands-free 기능으로 멀티태스킹(운전 등) 가능	터치나 텍스트 입력 시 멀티태스킹 불가

* HMI : Human machine interaction

※ 자료 : 각종 기술보고서 및 언론보도자료 재구성

인터페이스 기술별 스마트폰 적용률



※ 출처 : Emerging Interface Technologies for Mobile Devices, 2015, Tractica

[참고] 스마트 스피커 시장의 급성장

- IoT 및 스마트홈 시장의 성장과 맞물려, 스마트 스피커 시장이 급격하게 성장 중
 - AI 기반 음성인식 기술 적용으로 서비스 이용자 확대 추세

영국인의 최애 스마트 기기는 ‘음성인식 스마트 스피커’

좋아요 7개 | 입력 : 2018-10-29 00:30

KOTRA 런던무역관은 영국 스마트홈 기기 시장에서 가장 큰 성장률을 보이는 기기는 스피커로 지난 6개월간 스마트 스피커 보유 가구 수가 2배 이상 증가했다고 전했다. 스마트 스피커 이용자의

KOTRA 런던무역관은 영국 스마트홈 기기 시장에서 가장 큰 성장률을 보이는 기기는 스피커로 지난 6개월간 스마트 스피커 보유 가구 수가 2배 이상 증가했다고 전했다. 스마트 스피커 이용자의 대부분이 본래의 기능인 음악감상 외에서 음성인식 기능을 통한 문답과 알람, 일기예보, 일정관리, 온라인 쇼핑 등 다양한 분야에 스마트 스피커를 활용하고 있는 것을 볼 때 앞으로는 더욱 다양한 기능을 장착한 스피커가 인기를 끌 것으로 전망했다.

여론조사기관 유고브(YouGov)는 영국의 스마트홈 기기 시장은 꾸준히 성장하고 있으며 향후 5년간 영국이 스마트홈 기술 및 서비스 제공업체의 주요 타깃 시장이 될 것으로 분석했다.

스피커	11%
온도조절기	6%
조명	5%
보안	3%
체중계(기타 헬스케어 기기)	3%
세탁기	2%
청소기	1%
냉장고	1%
가스레인지	1%
전자레인지	1%
전기주전자	1%
식기세척기	1%

▲영국 스마트홈 기기별 보유율 (자료 : 유고브)

[아주초대석] 임채환 KT AI담당 상무 "연내 기가지니 이용자 150만 달성"

최대한 기자 | 입력 : 2018-10-31 14:32 | 수정 : 2018-10-31 14:48

호텔, 아파트 이어 스마트시티까지

을 드러냈다. KT의 AI스피커 기가지니 이용 고객은 120만명에 달한다. 서비스가 안정화되고 인식률이 높아지면서 매월 10만명씩 늘어나는 추세다. 연말까지는 150만명의 가입자를 확보한다는 목표다.

임채환 상무는 AI서비스에서 KT의 위상에 대해 "압도적 1위 사업자"라는 자부심을 드러냈다. KT의 AI스피커 기가지니 이용 고객은 120만명에 달한다. 서비스가 안정화되고 인식률이 높아지면서 매월 10만명씩 늘어나는 추세다. 연말까지는 150만명의 가입자를 확보한다는 목표다.

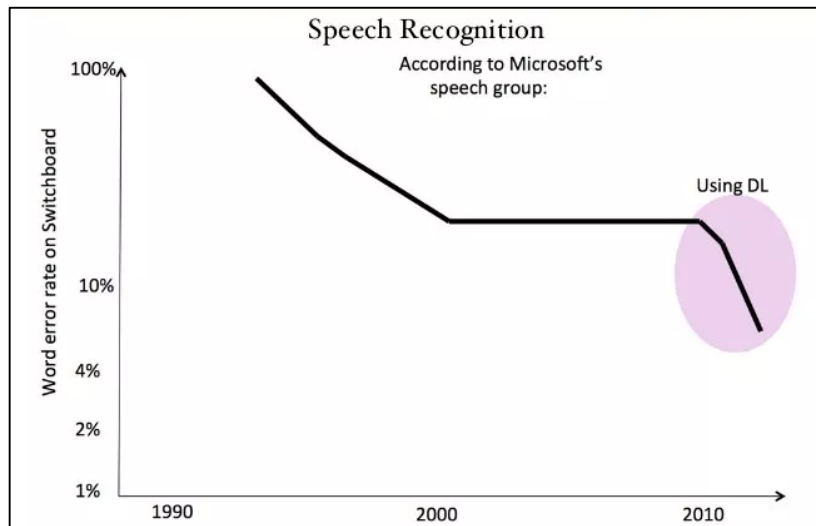
임 상무는 "AI는 인간의 본성을 찾아주는 서비스라고 생각한다"며 "서비스가 진화하고 발전해서 고객들에게 유용한 서비스를 만드는 게 중요하다"고 강조했다.

임 상무는 "고객들의 사용량이 많아지고 있다"며 "미국에서 조사했는데 AI서비스를 경험해본 고객들의 50~60%는 AI서비스 없이 살 수 없다는 대답을 했다"고 말했다. AI서비스를 이용하면 옷 입을 때 날씨를 물어볼 수 있고 자기 전에 노래를 틀어준다. 침대에 누워서 30분 후에 음악을 꺼달라는 주문을 할 수도 있다.

배경 및 필요성 : 딥러닝

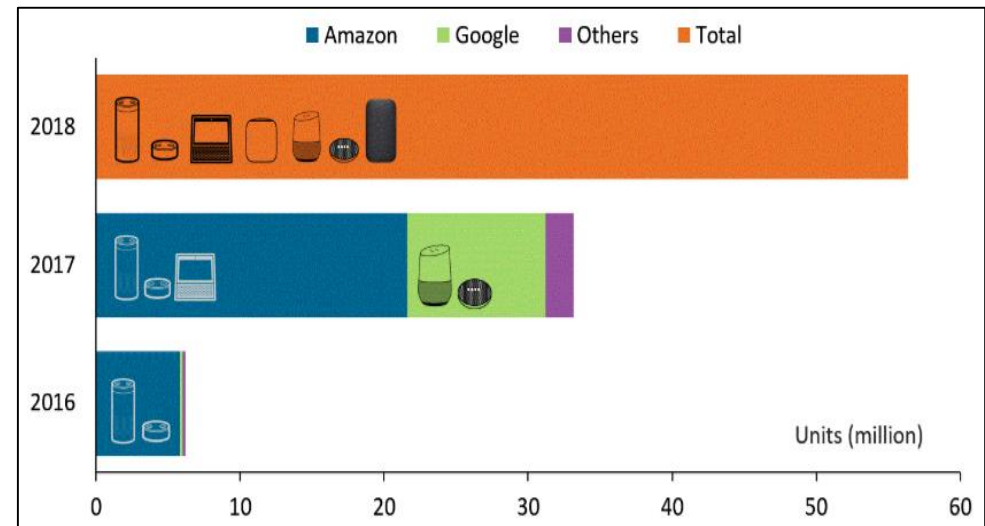
- 프로세서 연산속도 발전에 따라 딥러닝(Deep Learning) 기술이 음성인식 기술에 적용
 - 인식 에러율이 5% 미만으로 획기적으로 개선
- 최근 음성 인식률 개선에 따른 인공지능 스피커 등의 판매가 급증
 - 2017년 기준 3,300만대가 판매됐으며, 2018년에는 5,630만대 판매 예상

음성인식에서 DL(Deep Learning) 기술의 영향



※ 출처 : 마이크로소프트

전세계 스마트 스피커 판매량 및 예상(2016-2018)

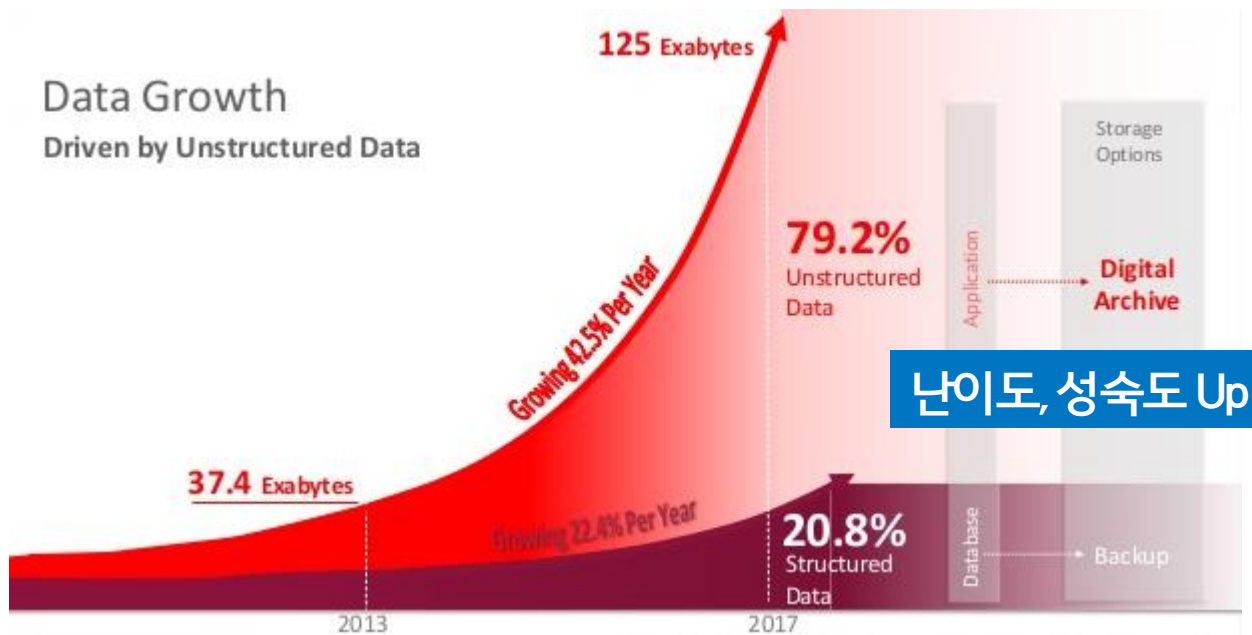


※ 출처 : Smart Speaker Analysis, 2017, canalys

배경 및 필요성 : 서버형 음성인식

- 자연어 인식, 사용 언어/어휘의 지속적인 업데이트 및 인터넷 망의 활용이 가능하고, 다양한 서비스로 확장 가능한 서버 기반의 음성인식 기술의 확대 예상
 - 음성인식 도입 단계에서는 한정된 명령어 기반의 서비스로 디바이스에 엔진을 탑재하는 임베디드 방식의 음성인식 기술 사용

자연어 등의 비정형 데이터 급증

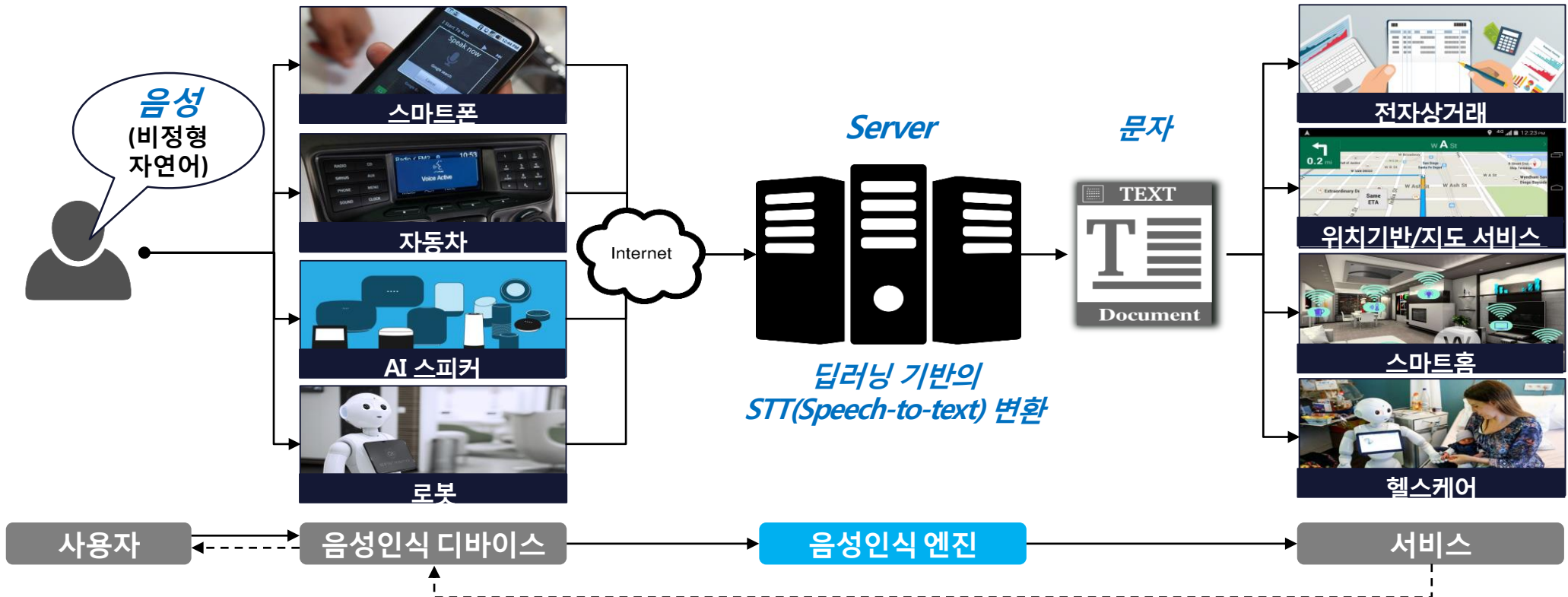


서버기반의 데이터 처리
기술 수요 증가 예상

* 자료 : 각종 기술보고서 및 언론보도자료 재구성

기술 개요

- 다양한 모바일 기기, 고객센터/콜센터, 기록물(연설, 회의, 발표, 강의, 방송 등) 등을 대상으로 **음성을 문자로 전환하거나 음성 명령을 수행하도록 하는 자연어 음성인식 기술**
- 딥러닝 기술 적용 시 **음성인식의 신뢰성 확보 가능**하며, 대용량 연산과 DB 업데이트가 용이한 서버에서 음성-문자로 전환되어 **자연어 수준의 음성인식과 복수의 일반 사용자에게 대한 서비스 가능**

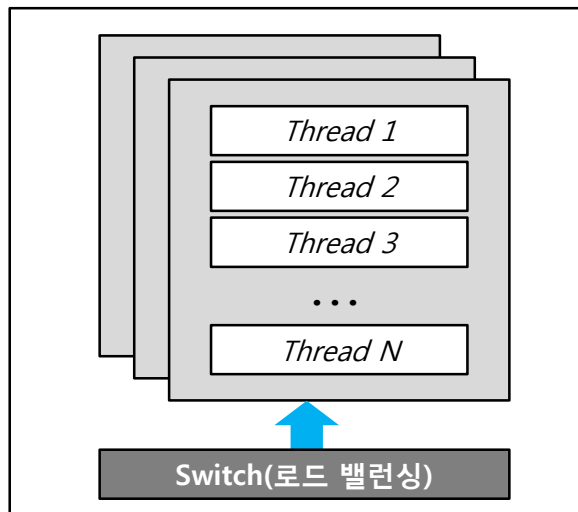


기술의 특징점

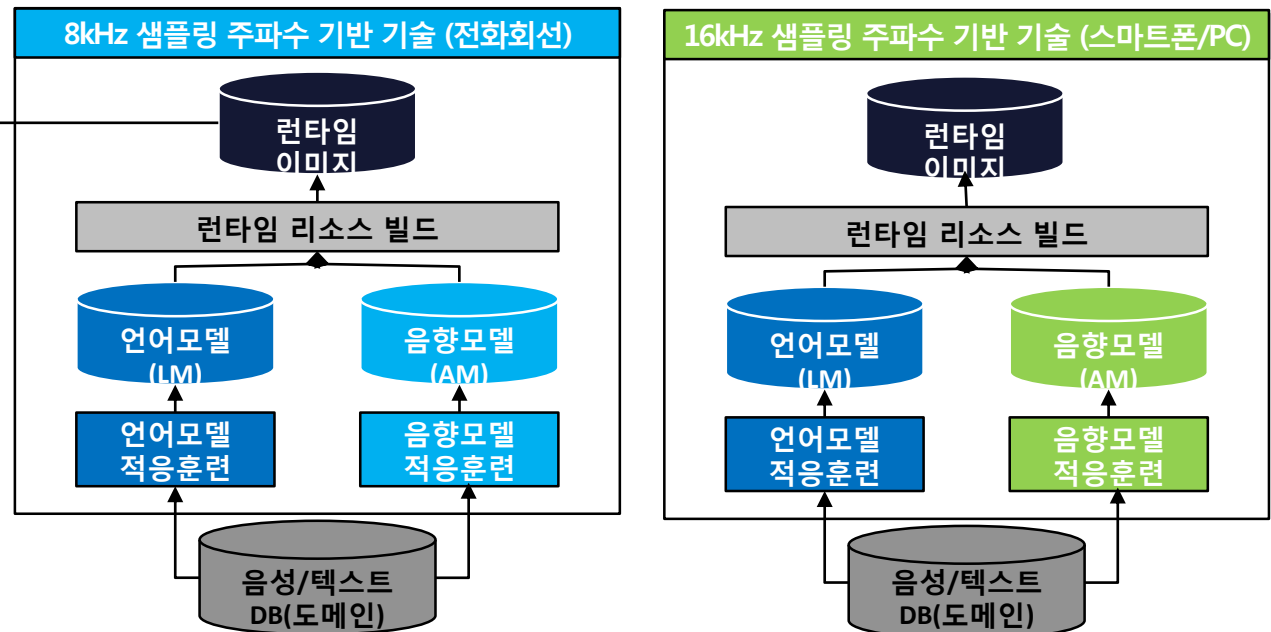
딥러닝 기술 기반의 서버 기반 음성인식 기술을 적용하여 높은 정밀도의 음성-문자 변환 인터페이스 제공

- ① 리눅스 기반의 서버/클라이언트형 음성 인식 기술로서, 싱글 쓰레드 또는 멀티 쓰레드 기반의 동시접속 환경에서 실시간(Online) 또는 일괄 처리 방식(Batch)으로 선택적 음성인식 가능
- ② 전화회선 환경에서 최적화된 협대역(8kHz)과 스마트폰/PC 환경에서 최적화된 광대역 (16kHz) 샘플링 주파수 기반 음성인식 기술은 훈련용 빅데이터 및 음향 모델 상이하나, 음성인식 적용 도메인의 음성 및 텍스트 DB를 기반으로 적응학습을 통해 음성인식 성능 제고

① 음성인식 엔진



② SW 도구



세부 기술의 특징점

탐색 엔진

- 연속어 음성인식 엔진 사용 기반 문장형태의 고속/고성능 음성인식 가능
 - 음성인식 결과에 대해 단어별 시간경계 정보 출력
 - 음성인식 결과인 형태소열에 대해 어절을 복원하고 문장단위로 출력

음향모델

- 사무실, 모바일 환경의 일반적인 음성 및 잡음이 반영된 기본 베이스라인 모델 활용
- 소규모 데이터를 사용하여 기본모델의 적응 학습을 통해 도메인별 음성인식 효율 향상

언어모델

- 기본 언어모델에 대해 적용영역의 데이터를 사용하여 학습 수행
- 발음 변환도구 및 통계적 언어 모델 기반 자연어 음성인식 처리 우수

기타

- 고속 적응학습 기술 적용 통해 적용분야 별 영역변화 성능저하 최소화
- 고객응대, 정보검색, 명령/제어, 녹취물 및 기록물 생성, 교육 서비스 등 다양한 서비스로의 응용 가능

기술 경쟁력

Company		Google	Apple	Amazon	ETRI
Product		Home	HomePod	Echo	ESTk toolkit*
Assistant		Assistant	Siri	Alex	(SDK)
Isolated Word Recognition		98%	97%	97%	98%
Customizable appearance		Yes	No	Yes	Yes
Compatibility		Android, iOS, Windows	iOS	Windows, Android	Android, iOS, Windows
Language Support	English	High	Medium	High	High
	Korean	High	Low	Low	High
	Chinese	High	Low	Low	High
	Japanese	High	High	High	High

* ESTk toolkit (ExtendScript Toolkit) : 스크립트 실행 문서

기술완성도(TRL)

TRL 9단계

TRL 9	사업화	
TRL 8	시작품 인증/표준화	<ul style="list-style-type: none"> 본격적인 양산 및 사업화 단계
TRL 7	Pilot 단계 시작품 신뢰성 평가	<ul style="list-style-type: none"> 일부 시제품의 인증 및 인허가 취득 단계 - 조선 기자재의 경우 선급기관 인증, 의약품의 경우 식약청의 품목 허가 등
TRL 6	Pilot 단계 시작품 성능 평가	<ul style="list-style-type: none"> 시작품의 신뢰성 평가 실제 환경(수요기업)에서 성능 검증이 이루어지는 단계
TRL 5	시제품 제작/ 성능평가	<ul style="list-style-type: none"> 경제성(생산성)을 고려한, 파일로트 규모의 시작품 제작 및 평가 시작품 성능평가
TRL 4	연구실 규모의 부품/시스템 성능평가	<ul style="list-style-type: none"> 개발한 부품/시스템의 시작품(Prototype) 제작 및 성능 평가 경제성(생산성)을 고려하지 않고, 우수한 시작품을 1개~수개 미만으로 개발
TRL 3	연구실 규모의 성능 검증	<ul style="list-style-type: none"> 연구실 규모의 부품/시스템 성능 평가가 완료된 단계 실용화를 위한 핵심요소기술 확보
TRL 2	실용 목적의 아이디어/ 특허 등 개념 정립	<ul style="list-style-type: none"> 연구실/실험실 규모의 환경에서 기본 성능이 검증될 수 있는 단계 개발하려는 시스템/부품의 기본 설계도면을 확보하는 단계 모델링/설계기술 확보
TRL 1	기초 이론/실험	<ul style="list-style-type: none"> 실용 목적의 아이디어, 특허 등 개념 정립
		<ul style="list-style-type: none"> 연구과제 탐색 및 기회 발굴 단계

기술이전 범위 및 지식재산권 현황

기술이전 범위

No.	구분	기술이전 범위
1	Object 파일 및 인식용 이미지 파일 생성	<ul style="list-style-type: none"> 리눅스 환경에서 실행 가능한 Library 형태의 Object 파일과 인식용 이미지 파일 생성도구(리눅스 Centos 7.x 이상) 서버 기반의 음성인식 엔진 SDK 런타임 이미지 생성도구 개발자용 지침서
2	음향/언어 모델 및 적응 학습 도구	<ul style="list-style-type: none"> 리눅스 Centos 7.x이상 음향/언어 모델 및 적응 학습 도구 기본 음향 모델 및 적응 학습 도구 기본 언어모델 및 적응 학습도구
3	프로그램 및 기술 문서	<ul style="list-style-type: none"> 온라인 음성인식 서버-클라이언트 등 프로그램 1건 음성인식 SDK 개발자용 지침서 등 기술문서 4건

지식재산권 현황

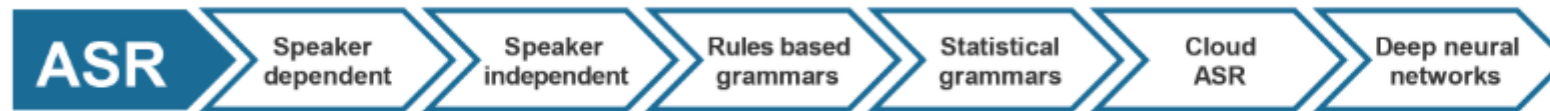
No.	특허번호	특허명	권리현황
1	10-2016-0103586	심층 신경망 기반의 음성인식 시스템	출원(공개)
2	10-2016-0158129	음향모델 생성 장치 및 방법	출원(공개)
3	10-2017-0051144	심층신경망 기반 음성 인식 시스템을 위한 발화 검증 방법	출원(비공개)
4	10-2017-0062292	자연어 대화체 음성 인식 방법 및 장치	출원(비공개)
5	10-2017-0160967	음향모델을 위한 심층 신경망 기반 상태 결정 장치 및 방법	출원(비공개)

기술 동향

- 음성인식 기술은 1980년대 소개된 IBM에서 제안한 통계적인 방식에서 클라우드 방식으로 발전하고 있으며, 궁극적으로 심층신경망(Deep Neural Network, DNN)을 적용하는 방식으로 발전 예상

Evolution of speech technologies

Automatic Speech Recognition



Natural Language Understanding



Text-to-Speech

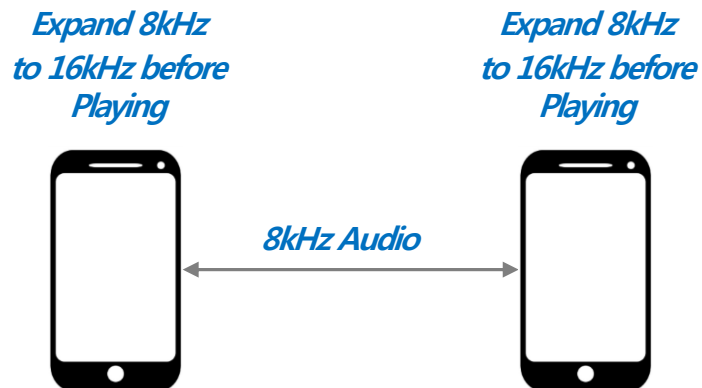


출처 : NUANCE (2015)

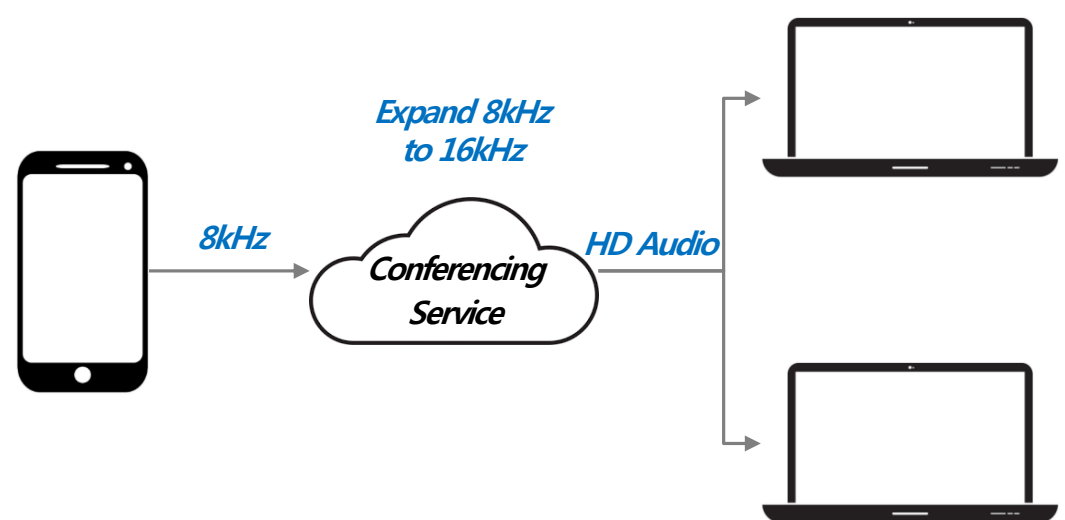
기술 동향 : 샘플링 주파수

- IoT 서비스 확대로 디바이스간 P2P통신에서 대역폭을 절약하기 위한 다운스케일 이슈 발생
 - 최근 IBM 등의 선도기업에서는 음성-문자 변환(STT) 관련 서비스를 전화 회의 등으로 확장
- 콜센터 서비스를 제외한 다양한 STT 서비스 분야에서 협대역(8kHz) 및 광대역(16kHz) 샘플링 주파수 기반 음성인식을 동시에 활용하는 사례 증가 → 협대역 오디오 사운드의 광대역 업스케일링 이슈 발생

디바이스간 P2P통신에서
대역폭 절약을 위한 다운스케일 이슈



컨퍼런싱 서비스 시
협대역 오디오 사운드의 업스케일링 이슈

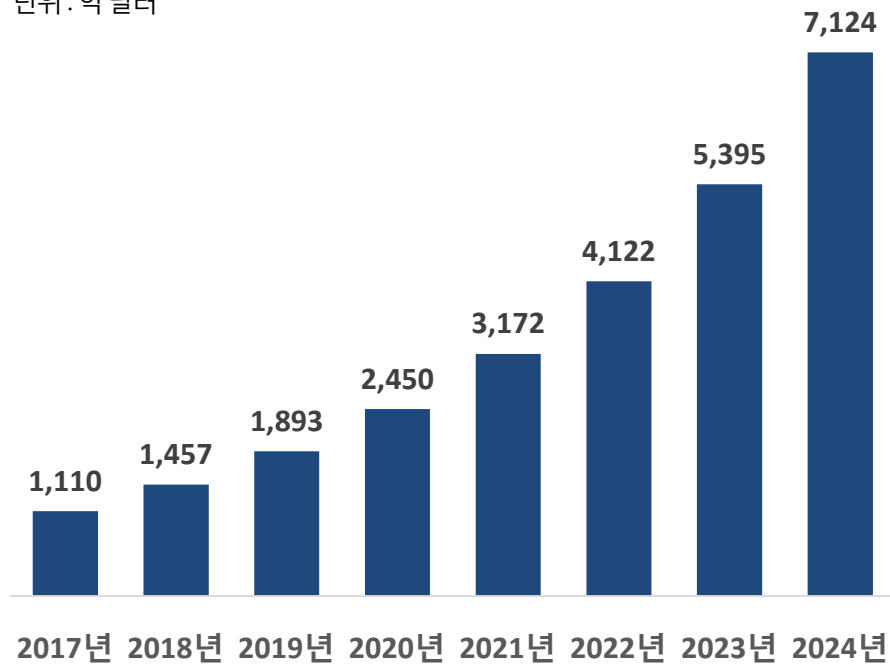


시장 동향 : 음성인식 시장전망

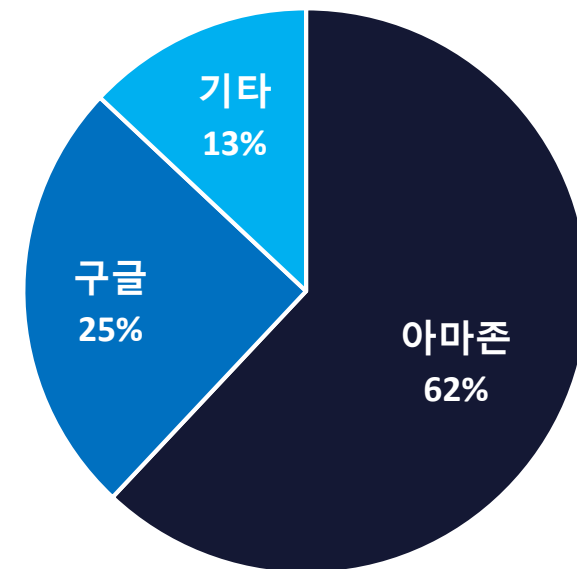
- 글로벌 음성인식 시장은 2017년 1,110억 달러에서 2024년까지 약 7,124억 달러로 성장 전망
- 글로벌 음성인식 시장점유율은 2017년 기준 아마존 62%, 구글 25%, 기타 13%(MS, 애플 등)로 나뉨

글로벌 음성인식 시장

단위: 억 달러



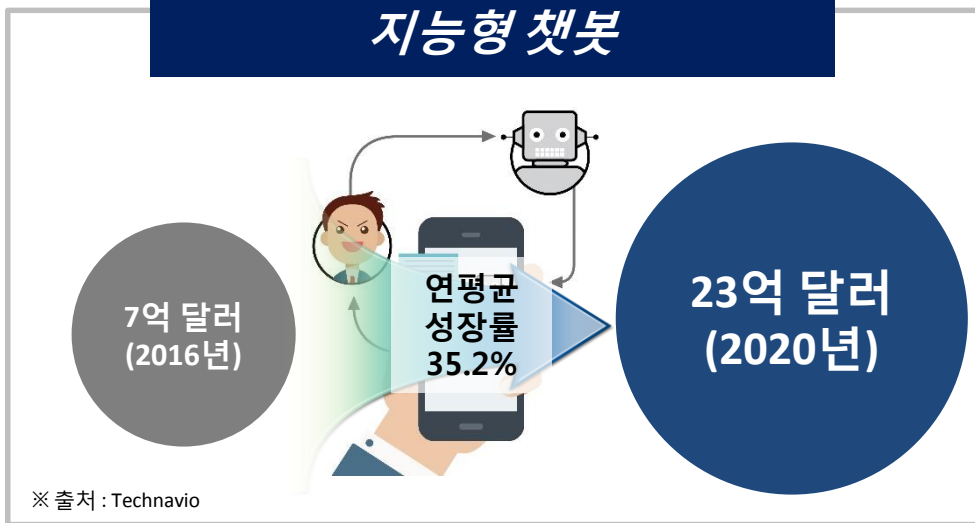
글로벌 음성인식 시장점유율



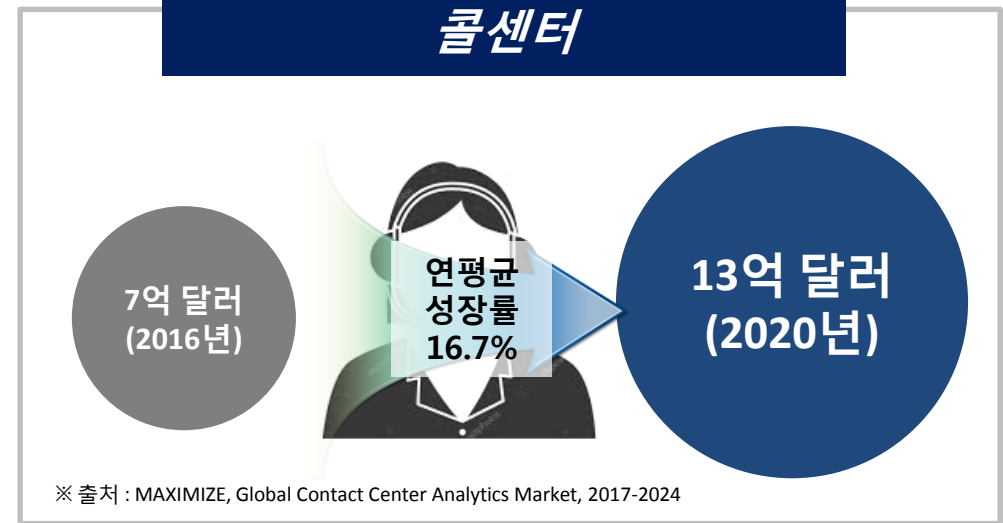
※ 출처: 음성인식 기반 인공지능 기술, 차량 도입 가속화, 김지윤(미국 디트로이트무역관), 코트라, 2018

시장 동향 : 음성인식 응용분야 시장전망

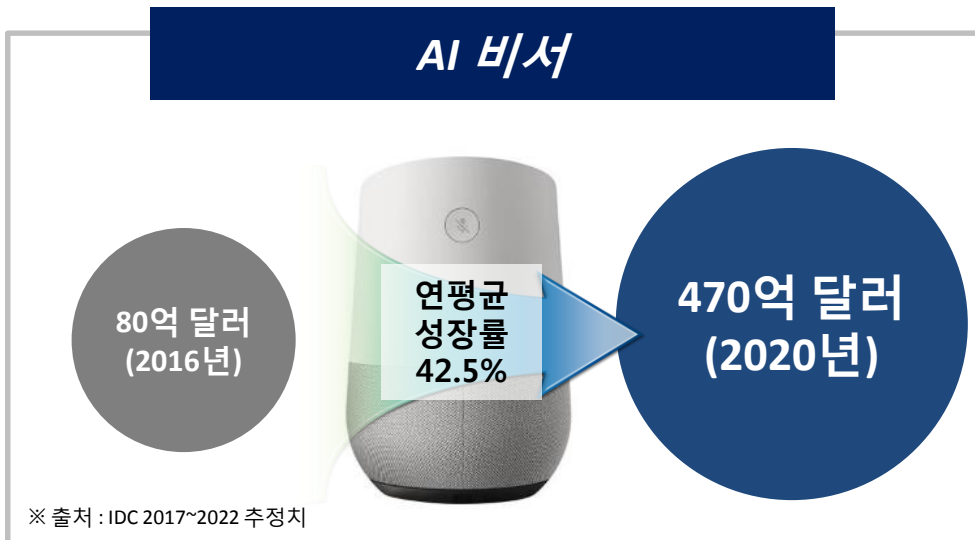
지능형 챗봇



콜센터



AI 비서

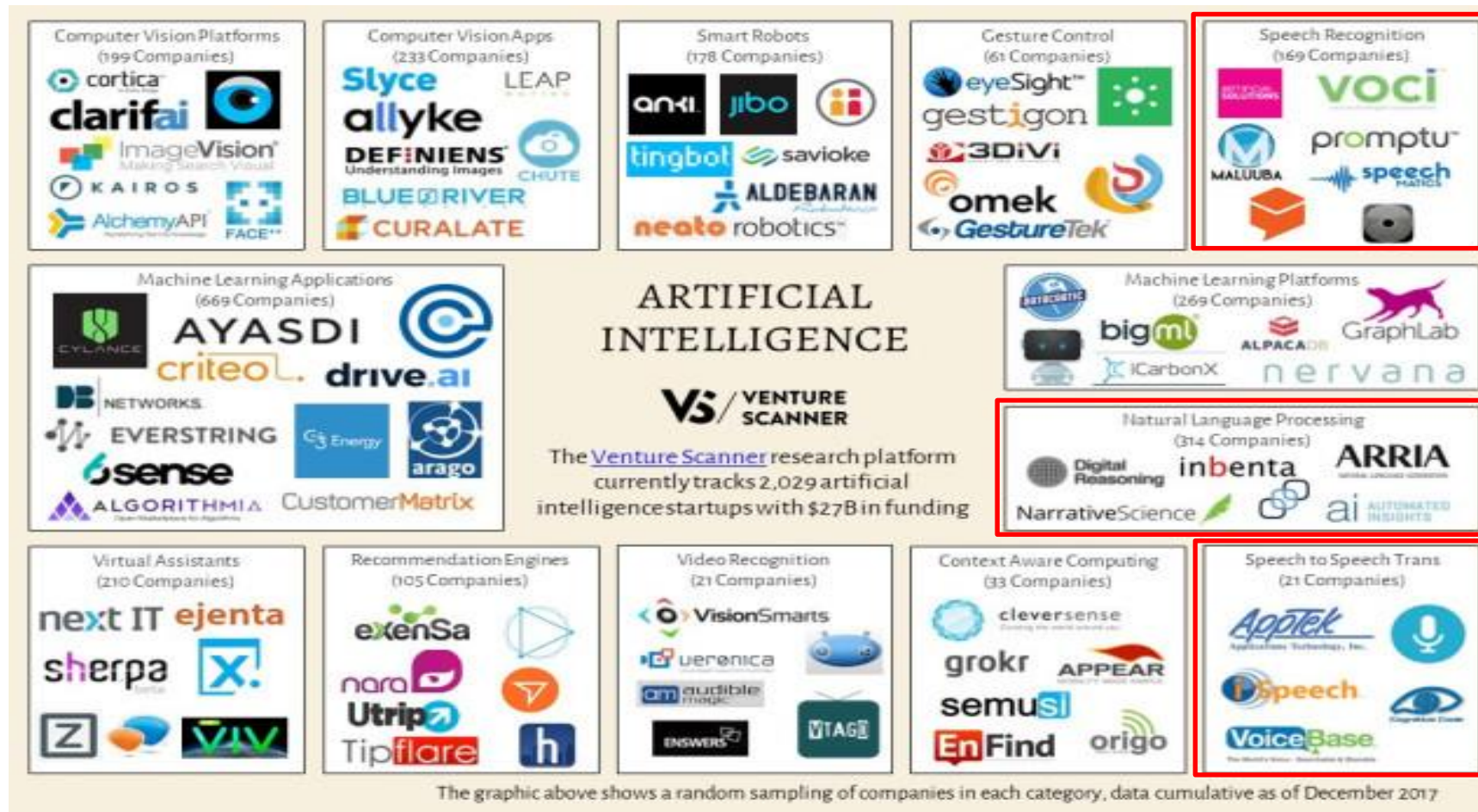


화상회의



시장 동향 : 음성인식 기반 비즈니스 현황







- 2,029개의 전세계 인공지능 분야의 스타트업 중 Speech Recognition, 자연어 프로세싱 및 음성-문자 변환(STT, Speech To Text) 등 음성인식 관련 스타트업은 총 504개로, 약 25% 차지



[참고] 국내외 스타트업 현황

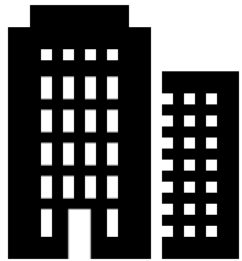
- 투자 유치를 통해 성장하고 있는 국내외 스타트업의 경우 자연어 처리를 위해 서버 기반의 AI 음성인식 기술 보유하고 있으며, 대부분 AI 스피커, IoT, 스마트홈, 챗봇, 로봇 등의 분야로 사업 진출

<표> 국내외 음성인식 관련 스타트업 현황

기업명	 Unisound	 Snips	 SoundHound	 Sense	 Mobvoi	 마인즈랩	
본사	Beijing (중국)	Paris (프랑스)	Santa Clara (미국)	Cambridge (미국)	Beijing (중국)	성남 (대한민국)	
사업분야	<ul style="list-style-type: none"> Cloud Data Services Natural Language Robotics 	<ul style="list-style-type: none"> AI Voice Assistant IoT 	<ul style="list-style-type: none"> AI Natural Language Speech Recognition 	<ul style="list-style-type: none"> IoT Smart Home Speech Recognition 	<ul style="list-style-type: none"> AI Natural Language Speech Recognition 	<ul style="list-style-type: none"> AI/Deep learning Natural Language STT/TTS 	
설립년도	2012	2013	2005	2013	2012	2014	
최근 매출	\$5M	\$3M	\$100M	\$3M	\$5.6M	68억원	
고용인력	10	78	250	55	83	97	
투자현황	횟수	6	5	6	4	5	2
	금액	\$301.4M	\$24.7M	\$215M	\$38M	\$252.7M	90억
기타			<ul style="list-style-type: none"> 유니콘 삼성전자, KT, 현대차, 네이버 투자(1,500억 이상) 			<ul style="list-style-type: none"> 주요주주: ETRI, NAVER, 슈프리마, PwC 등 국내 AI 기업 중 최대 투자유치 	

투자 동향

- 음성인식 분야의 경우, 263회의 투자유치가 이루어졌으며, 투자유치 금액은 총 13억 달러 규모
- M&A된 기업은 관련 기업의 12% 수준이며, 총 인수대금은 20억 달러 규모
- IPO 기업은 9개사이며, 공모금액은 132.5백만 달러 수준



212

Number of Organizations

Nov 6, 2009

Average Founded Date



12%

Percentage Acquired

\$2B

Total Acquired Price



263

Number of Funding Rounds

\$1.3B

Total Funding Amount



9

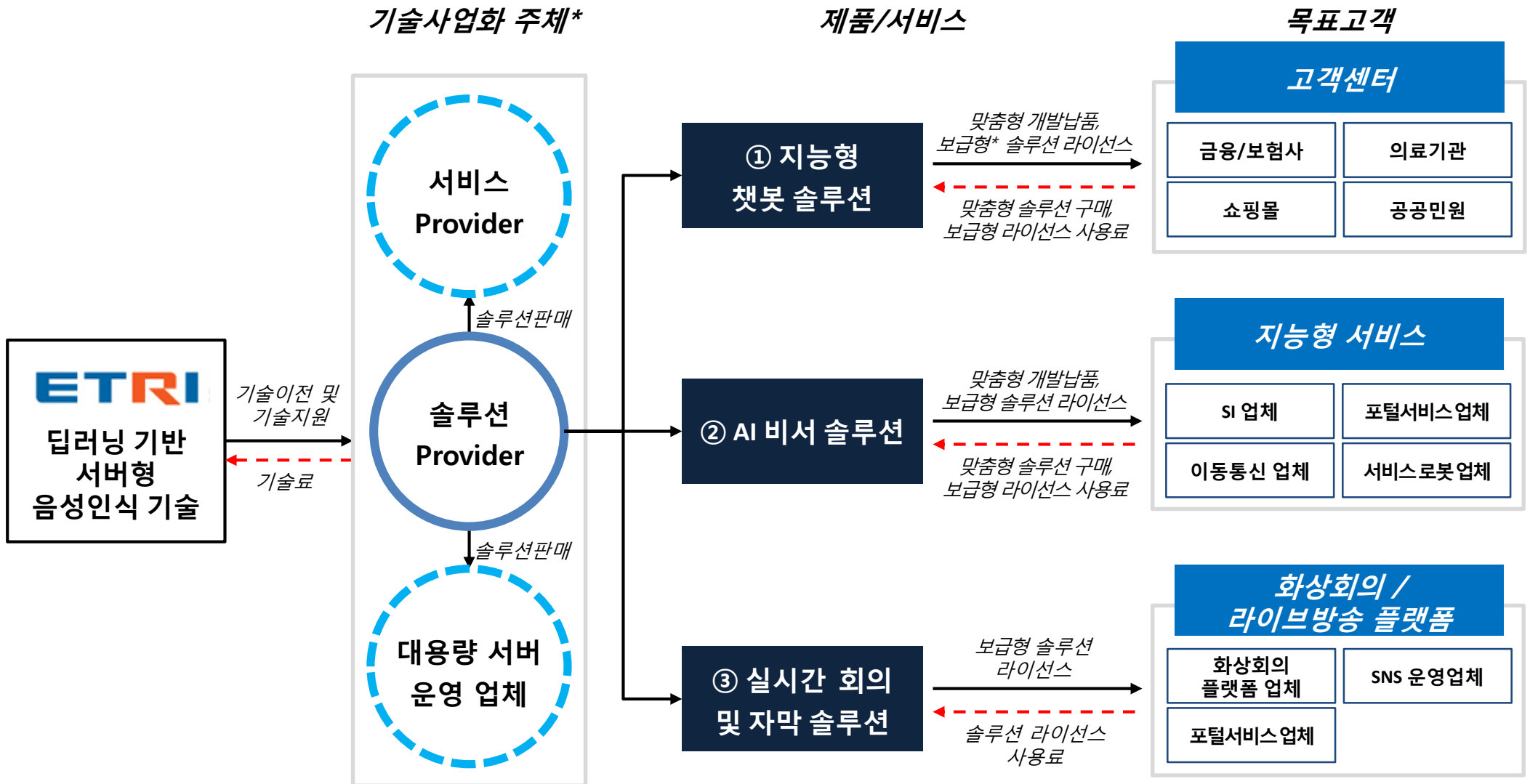
Number of IPOs

\$132.5M

Total Amount Raised in IPO

* 자료 : crunchbase (<https://www.crunchbase.com>) 재구성

비즈니스 모델 Overview

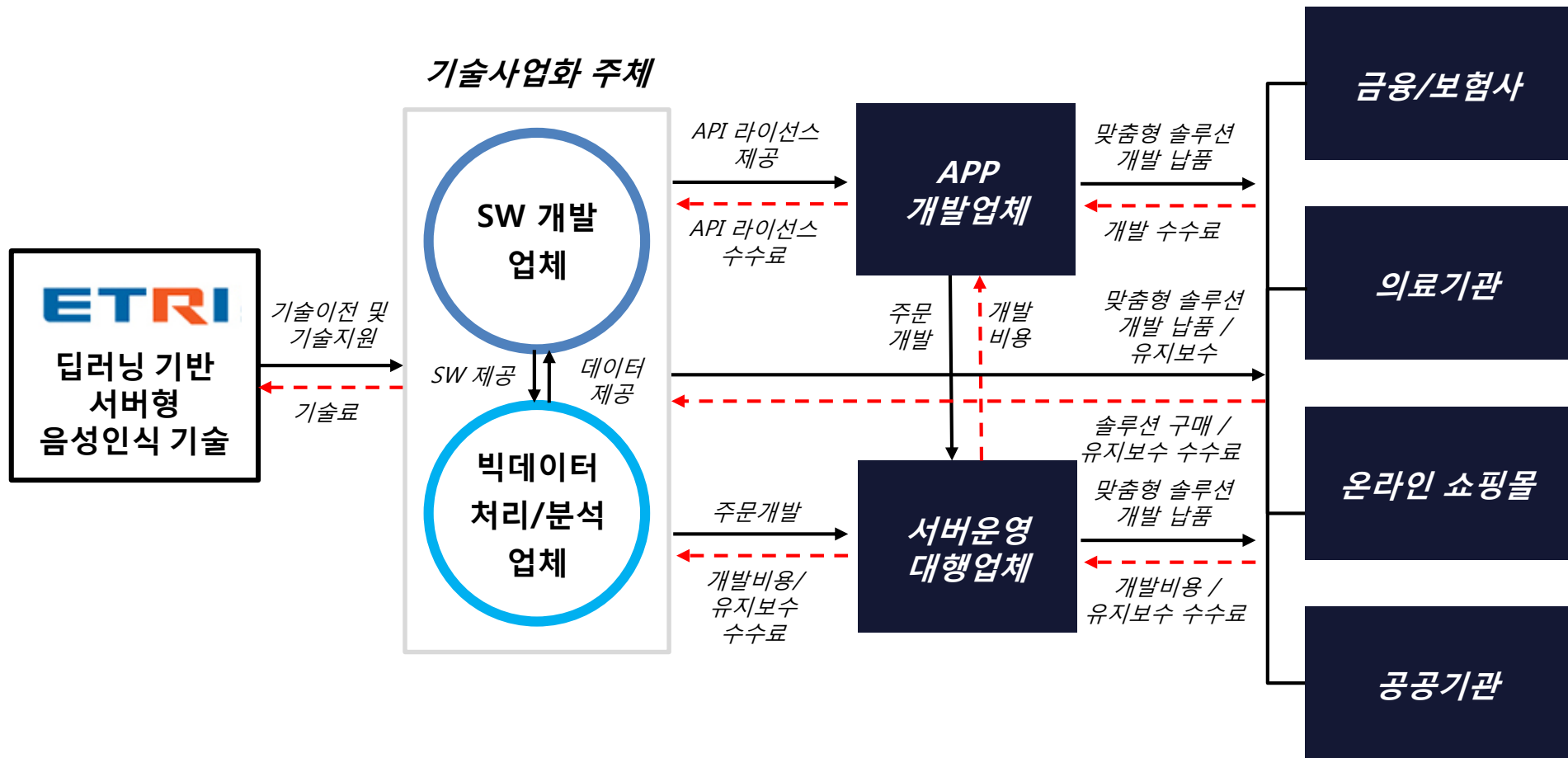


* **Note:** 서비스 Provider나 대용량 서버 운영 업체는 수직 계열화된 경우가 아닌 경우에는 기술사업화 주체보다는 목표고객으로 볼 수 있음

* 보급형 솔루션 : 솔루션 Provider가 고객구분 없이 판매하기 위한 보급형 솔루션

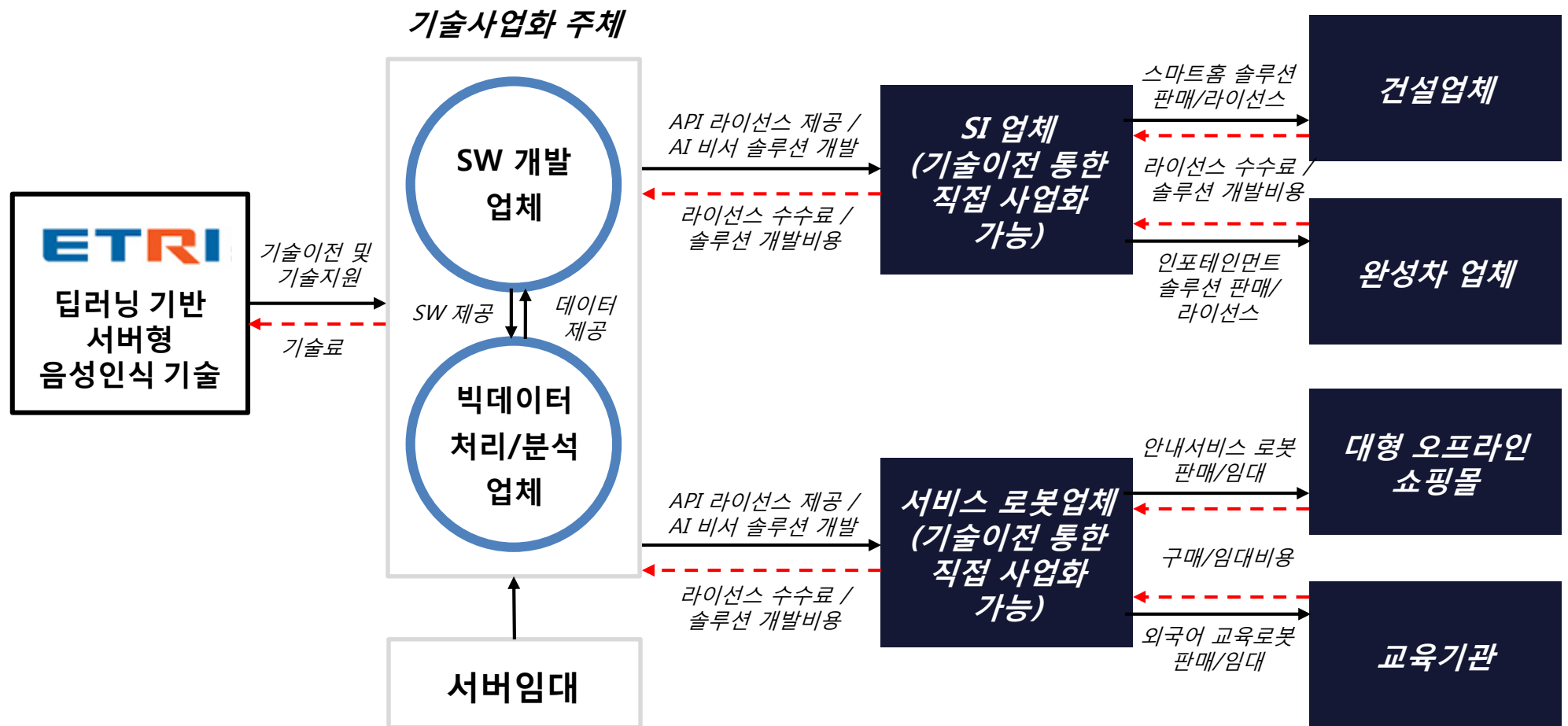
BM ① 지능형 챗봇 솔루션

- 고객서비스 및 고객분쟁 예방/대응이 필요한 서비스 업종 대상 맞춤형 솔루션 제공
- APP 개발업체(API 라이선스) 또는 서버운영 대행업체(주문개발) 통한 솔루션 제공



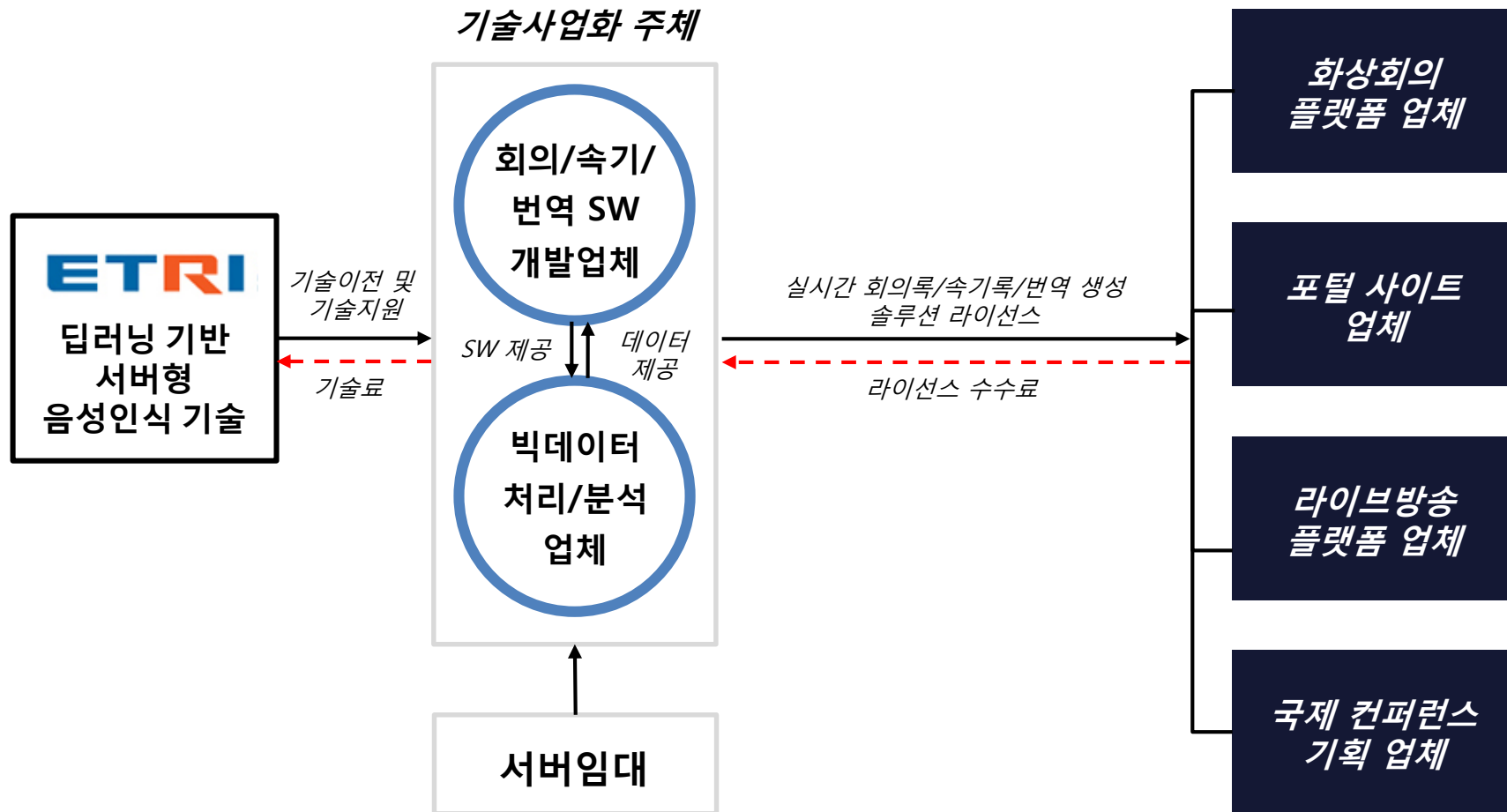
BM ② AI 비서 솔루션

- 대형 서버보유 SI 업체 통한 스마트홈, 인포테인먼트 대상 솔루션 제공
- 서비스 로봇업체 통한 대형 오프라인 쇼핑몰, 교육기관 대상 솔루션 제공



BM ③ 실시간 회의 및 자막 솔루션

- 다양한 회의 및 속기 지원 솔루션 제공
- 맞춤형 솔루션 개발납품 보다는 보급형 솔루션 개발 통해 라이선스 제공 방법으로 수익창출



ETRI 개발기술 도입 통한 사업화 프로세스



BM 선정 / 구체화(2개월)

기업 맞춤형 BM 구체화

- 사업화 아이디어 선정
- 목표시장 별 서비스 수요 파악
- 수요 맞춤형 BM 수립

상용화 개발(3개월)

BM 기반 제품/서비스 개발

- 기업 맞춤형 솔루션 개발 추진
- 제품/서비스 상용화 Test 실시 (연구자 자문)

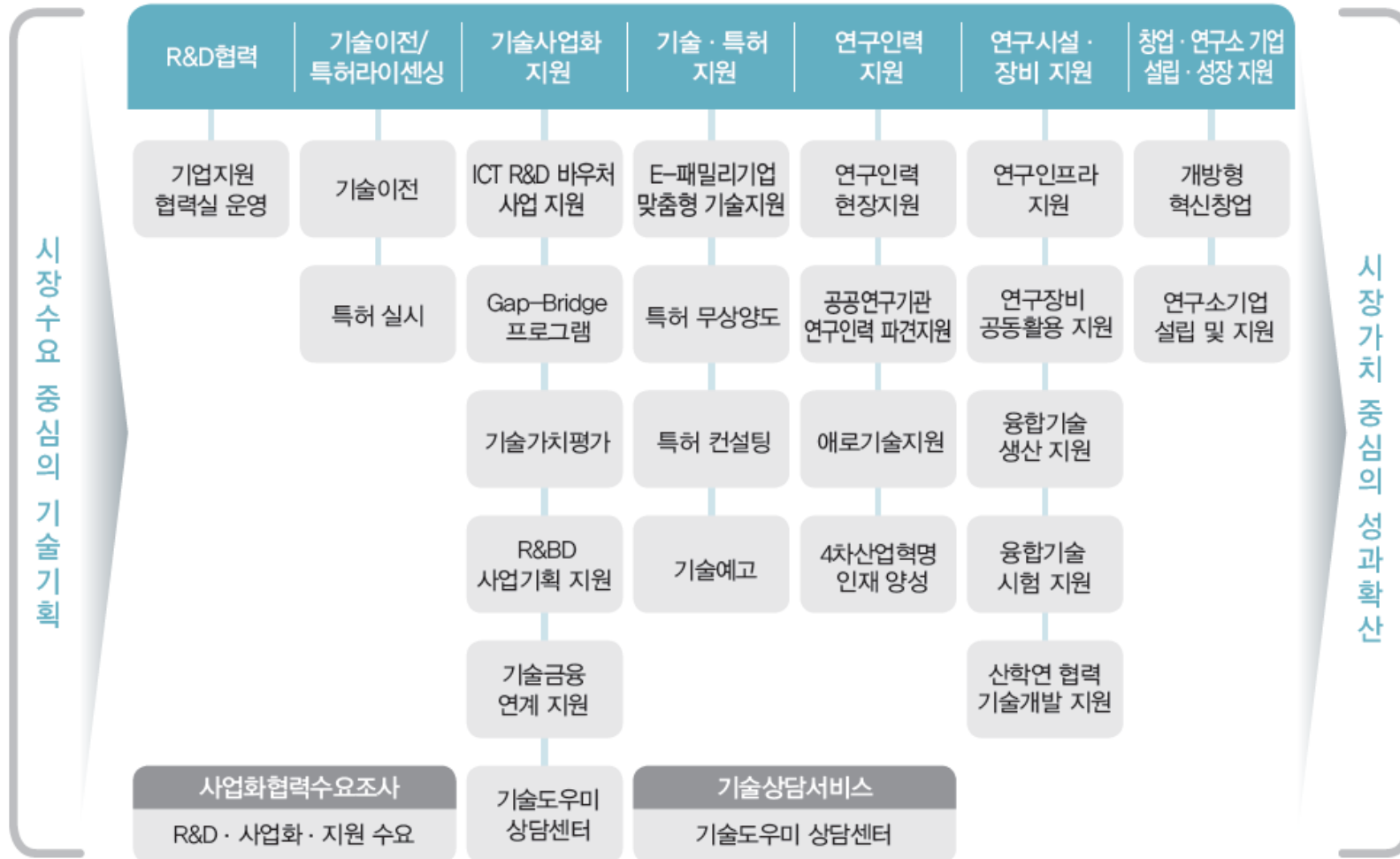
사업화(6개월)

BM 적용 서비스/솔루션 출시, 판매

- 솔루션 연계 서비스 개발
- 홍보 및 마케팅 위한 산업진흥기관 지원사업 연계
- 솔루션 기술보호 위한 신규 IP 확보

ETRI 기업지원 프로그램

기술사업화플랫폼 ETRI PLUS



(참고) 정부 추가개발 사업화 지원 사업 안내(기술이전 조건부)

사업명	기술이전사업화 지원사업	R&D 재발견 프로젝트	중대형복합기술사업화지원사업
공고기관	연구개발특구진흥재단 (www.innopolis.or.kr)	한국산업기술진흥원 (www.kiat.or.kr)	과학기술일자리진흥원 (www.compa.re.kr)
사업비 (2018년 기준)	1년 기준 2억 원	1년 기준 4억 원	1년 기준 7.5억 원
신청자격	특구 내 공공연구기관 기술도입기업 또는 추천기술 도입기업	NTB 사이트 등록된 공공연구기관 기술도입 기업	Tech-BM 검증지원사업 통한 경쟁
공고시기 (2018년 기준)	2월 또는 3월	3월	2월

기술이전 문의



ETRI 사업화협력실

042-860-1804 / hominkim@etri.re.kr