

전략분야 현황분석

인공지능



인공지능

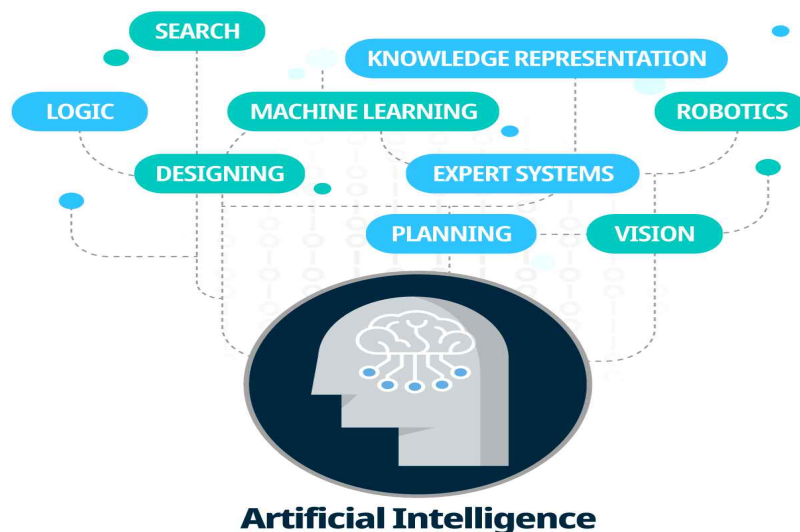
1. 개요

가. 일반적 정의

(1) 정의

- 인공지능(AI)은 인간의 지적 능력을 기계로 구현하는 과학 기술
 - 컴퓨터가 인간의 지능적인 행동을 모방할 수 있도록 하는 것을 방향으로 발전
- 지능에 대한 정의와 마찬가지로 인공지능에 대해서도 다양한 정의가 존재
 - 인공지능은 사람 또는 동물의 지능이 컴퓨터로 모사될 정도로 세밀하고 정확하게 표현될 수 있다는 생각에 기반을 둔 기술
 - 자연어 처리, 컴퓨터 비전 및 패턴 인식, 로봇틱스, 자율주행차, 지능형 로봇, 스마트 공장, 스마트 헬스케어, 금융/핀테크, 지능형 서비스/교육 등의 분야에서 응용
 - 인공지능의 방법과 관련된 탐색, 논리 및 추론, 지식 표현, 계획, 학습 등 세부 분야에 관한 연구가 진행 중
- 인공지능의 궁극적인 목표는 다트머스 회의에서 언급됐듯이 사람과 유사한 혹은 사람 수준의 기계장치 개발
 - 다트머스 회의에서 정의된 AI는 사람의 지능을 모사하는 기계장치로 사람처럼 언어로 소통하고 문제를 해결할 수 있다는 개념

[인공지능 개요도]



* 출처:한국정보통신기술협회, 정보통신용어사전

(2) 필요성

◎ 세계 전반에 인공지능의 영향력 확대

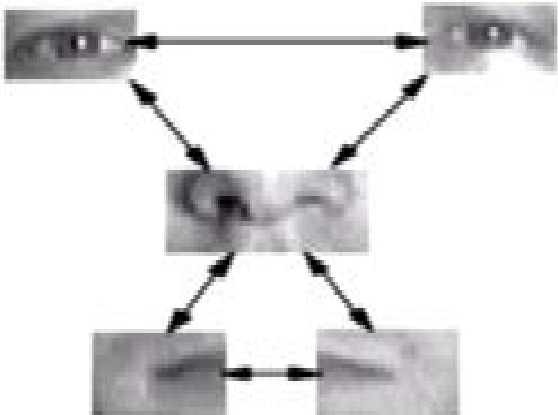

- 코로나19 팬데믹으로 인한 비대면 서비스, 의료 분야의 인공지능 활용 등의 수요 증가로 인공지능 산업의 급성장 추세
 - 인공지능 및 빅데이터를 활용하여 정부 및 다양한 기업들은 코로나19 방역에 기여
 - 코로나19 확진자가 발생하는 경우 이동 경로에 따른 모바일 이용 데이터, 신용카드 결제 데이터, CCTV 데이터 등을 분석하여 경로 및 접촉자 파악
 - 신테카바이온은 슈퍼컴퓨팅 파워와 자체 인공지능 기반 합성신약 물질 발굴 플랫폼을 활용해 코로나19 치료제 후보물질 30종 발굴
 - 한국판 뉴딜 등 다양한 정책적 지원을 통해 코로나19 종식 이후에도 인공지능 산업은 꾸준한 성장 전망
- 기술의 발전에 따라 기존의 산업기반 뿐만 아니라 생산성 증가, 공공분야 및 사회 전반에 끼치는 영향력이 클 것으로 예상
- 인공지능 기술의 비약적인 발전으로 국내외적으로 경제성장과 사회 난제 해결의 돌파구가 마련될 것으로 기대
 - 인공지능은 단순 신기술이 아닌 경제 및 사회 변혁의 핵심 동력으로 작용하여 각국 경제성장에 비약적인 파급효과를 실현할 전망
 - 선도국은 AI의 높은 잠재력에 앞서 주목하고 AI 기술력을 선점하기 위한 국가 차원의 대책을 추진
 - 미국의 경우 '16년 10월에 국가 AI R&D 전략 계획을 발표, 일본의 경우 '17년 4월에 AI 산업화 로드맵을 발표. 중국의 경우 '17년 7월에 차세대 AI 발전 계획을 발표

◎ 국내 인공지능 기술 발전의 필요성

- 한국도 국가적 대응전략을 마련하여 인공지능 역량 개선에 집중해 왔으나 기술 격차는 심화되고, 인공지능 분야의 전문인력 확보와 역동적 기술 혁신 생태계 구축은 미진한 상태
 - 알파고-이세돌 대국 후 '17년 10월에 발표된 IITP의 국내외 인공지능 기술 격차 체감도에 따르면 기술 격차가 심화되었다는 의견이 61.5%
- 한국의 경우, 많은 산업 분야가 인공지능 적용을 시작하는 단계이므로 전략적 접근 시 경쟁력 확보 가능
- 인공지능이 향후 경제 사회 전반을 혁신할 근본 기술임을 고려, 국민 삶의 질과 국가 경쟁력 제고를 위해 인공지능 기술력 확보는 필수
 - 인공지능 기술 혁신 투자전략을 정립하고, 국내 연구역량을 결집하여 AI 기술력을 확보할 수 있는 체계 구축이 시급한 상황

- 인공지능은 대량의 데이터가 생성되는 분야면 적용될 가능성이 더 크며 사회관계망서비스, 인터넷 상거래, 금융 거래, 병원 처방 등 우리의 행동이 모두 데이터로 기록되고, 이를 활용해 다양한 산업적 활용이 가능
- 인공지능 기술의 빠른 발전속도는 곧 심층학습의 방법론 자체의 성능향상에 기여하고, 심층학습의 성능향상은 심층학습 모델을 실제 응용 분야나 타 산업에 확대하여 광범위하게 적용할 수 있음

[기존학습 방식과 심층학습의 차이]

기존학습 방식	심층학습
AI 설계자가 주요 특징(예를 들어, 눈, 코, 입) 또는 규칙성을 추출하여 모델링 후 AI 구현	심층 모델이 학습 데이터를 기반으로 스스로 주요 특징 또는 규칙성을 추출하여 모델링
	

* 출처: 과학기술정보통신부, 2018, AI R&D 전략

◎ 국가적 산업 경쟁력 향상을 위해 인공지능 적용 필요

- 해외 ICT 기업들이 관심을 가지고 주력하는 AI 분야는 기업별로 상이하나 AI 기술을 다양한 분야에 접목하여 산업 확장을 도모할 전망
 - Google, Apple, Amazon, Facebook 등 기업들은 다양한 AI 제품·서비스 개발이 가능한 범용 AI 플랫폼을 중심으로 시장을 선점하기 위해 노력하고 있으며 IBM 등은 의료, 금융, 법률과 같은 특정 산업 영역에 특화된 전문 AI 플랫폼 고도화와 생태계 구축에 주력
 - 가장 큰 비중을 차지하고 있는 생태계 구축 방식은 과거 스마트폰 생태계를 구축했던 방식과 비슷하게 개방형 혁신 방식을 추진
 - 전문 AI 플랫폼 기업들은 일반 개발자와 더불어 특정 영역의 전문기업들과의 협력을 통해 생태계를 확장하는 전략에 중점

나. 구축 범위

(1) 가치사슬

- AI 분야는 후방산업의 기술이 기술 발전에 가장 많은 영향을 미치는 것을 알 수 있으며 후방산업의 기술력으로 만들어진 AI 기반 SW를 통해서 전방산업이 완성될 수 있으므로 후방산업의 기술 의존도가 매우 높음
 - 후방산업은 AI 기반을 통한 상황인지, 자기학습, 패턴 인식, 자연어 처리, 음성인식, 컴퓨터 비전 등의 분야로 계속 발전할 것으로 판단
 - 전방산업은 후방산업의 기술력이 더해져 더욱 다양한 서비스와 산업의 형태로 진화할 것으로 보이며, 특히 IoT 분야에서 엄청난 영향을 미칠 것으로 판단

[인공지능 분야 가치사슬]

후방산업	인공지능	전방산업
뇌과학 등 기초과학, HW(지능형 반도체, 컴퓨팅), SW, 네트워크	인공지능 응용서비스, 인공지능 엔진	1, 2, 3차 전산업

- AI 서비스는 모바일 등을 통한 데이터(이미지, 텍스트 등) 획득, 획득한 데이터 가공, 반복 학습을 통한 AI 모델(알고리즘) 생성 과정을 통해 최종적으로 AI 서비스를 제공
 - AI 서비스를 제공하기 위해서는 대량의 학습 데이터 세트가 필요함에 따라 클라우드 및 딥러닝 전용 프로세서 기반의 고성능 컴퓨팅 인프라가 필요

[AI 서비스]



* 출처: 과학기술정보통신부, AI R&D 전략, 2018

(2) 대표적 분류 방법

[AI의 기술별 분류]

기술	내용
추론 및 기계학습	• 인간의 사고능력을 모방하는 기술
지식표현 및 언어지능	• 사람이 사용하는 자연어 이해를 기반으로 사람과 상호작용하는 기술
청각지능	• 음성/음향/음악을 분석, 인식, 합성, 검색하는 기술
시각지능	• 사물의 위치, 종류, 움직임, 주변과의 관계 등 시각 이해를 기반으로 지능화된 기능을 제공하는 기술
복합지능	• 시공간, 촉각, 후각 등 주변의 상황을 인지, 예측하고, 상황에 적합한 대응을 제공하는 기술
지능형 에이전트	• 개인비서, 챗봇 등 가상공간 환경에 위치하여 특별한 응용 프로그램을 다루는 사용자를 도움 목적으로 반복적인 작업을 자동화시켜 주는 기술
인간-기계 협업	• 인간의 감성이나 의도를 이해하고 인간의 뇌 활동에 기계가 연동되어 작동하게 해주는 기술
AI 기반 HW	• 초고속 지능정보처리를 구현하게 지원해주는 HW

* 출처: 4차 산업혁명을 선도하는 주요 기술 대상 기술수준평가 및 기술수준 향상방안, 정보통신기술진흥센터

(3) 기술로드맵 전략분야의 범위

정보통신기술진흥센터의 기술 분류 체계에 따라 AI 전략분야를 8개로 구분

- 추론 및 기계학습은 인간의 사고능력을 모방하는 기술들로 인간의 사고능력을 모방한 추론, 베이지안(Bayesian)¹⁾ 학습, 인공신경망, 강화학습, 딥러닝, 인지 공학 등이 본 전략 분야에 해당
- 지식 표현 및 언어지능은 사람이 사용하는 자연어를 이해하는 자연어 처리를 기반으로 사람과 상호 작용하는 기술들이 포함
- 청각지능은 음성, 음향, 음악 등의 분석하여 음성을 합성하거나 음성을 검색하는 기술들이 포함
- 시각지능은 사물의 위치, 종류, 움직임, 주변과의 관계 등 시각 이해를 기반으로 지능화된 기능을 제공하는 기술들을 포함

1) 베이지안은 기본적으로 확률을 뜻함, 사건과 관련 있는 여러 확률을 이용해 새롭게 일어날 사건을 추정하는 것

- 복합 지능은 시공간, 촉각, 후각 등 인간의 오감을 모방한 감각 데이터를 이용하여 주변 상황을 인지, 예측하고, 상황에 적합한 대응을 제공하는 기술들을 포함
- 지능형 에이전트는 개인비서, 챗봇 등 가상공간 환경에 위치하여 특별한 응용 프로그램을 다루는 사용자를 보조할 목적으로 반복적인 작업을 자동화시켜 주는 기술들을 포함
- 인간-기계 협업은 인간의 감성이나 의도를 이해하고 인간의 뇌 활동에 기계가 연동되어 작동하게 해주는 기술들을 포함
- AI 기반 하드웨어는 초고속 AI 정보 처리를 구현하게 지원하는 하드웨어와 하드웨어 관련 기술들을 포함

[AI 전략분야의 범위]

* : 본 전략분야의 범위

구분	상품과 기술
추론 및 기계학습*	추론, 베이지안 학습, 인공신경망, 강화학습, 딥러닝, 앙상블 러닝, 인지 공학
지식표현 및 언어지능*	지식공학 및 온톨로지, 대용량 지식처리, 언어분석, 의미분석, 대화 이해 및 생성, 자동 통·번역, 질의응답(Q/A), 텍스트 요약
청각지능*	음성분석, 음성인식, 화자 인식/적응, 음성합성, 오디오 색인 및 검색, 잡음처리 및 음원 분리, 음향 인식
시각지능*	컴퓨터 비전, 사물 이해, 행동 이해, 장소/장면 이해, 비디오 분석 및 예측 시공간 영상 이해, 비디오 요약
복합지능*	공간 지능, 오감 인지 다중 상황 판단
지능형 에이전트	지능형 개인비서, 소셜 지능 및 협업지능, 에이전트 플랫폼, 에이전트 기술, 게임 지능, 창작 지능
인간-기계 협업*	감성 지능, 사용자 의도 이해, 뇌컴퓨터 인터페이스, 추론 근거 설명
AI기반 HW	뉴르모픽 칩(Neuromorphic chip), 지능형 반도체, 슈퍼컴퓨팅, AI 전용 프로세서

* 출처: 4차 산업혁명을 선도하는 주요 기술 대상 기술수준평가 및 기술수준 향상방안, 정보통신기술진흥센터

2. 시장 분석

가. 세계 시장 분석

(1) 세계시장 동향 및 전망

◎ 인공지능 전체시장 전망

- IDC가 2020년 발표한 세계 AI 시장에 따르면 2018년 세계 AI 시장 규모는 1,337억 달러로 평가되었으며 연평균 성장률 17.1%로 2024년 3,446억 달러로 성장 전망

[세계 AI 시장 전망]

(단위: 십억 달러, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
세계시장	133.7	156.5	183.3	214.6	251.3	294.3	344.6	17.1

* 출처: IDC(2020.08.04.)

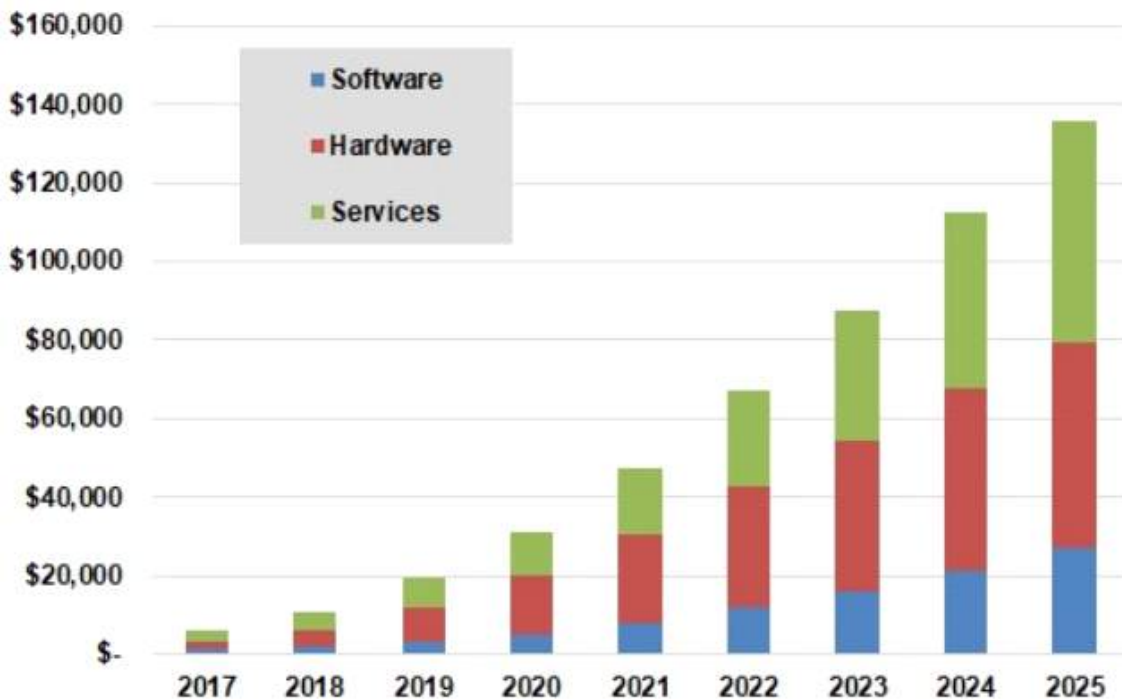
- AI 시장이 급성장함에 따라 글로벌 기업들은 적극적인 대규모 편딩과 M&A를 확대하는 등 기술 경쟁력을 강화하기 위한 투자에 총력을 기울이고 있음
 - 벤처 스캐너의 통계자료에 따르면 미국은 AI에 대한 투자가 2016년 30억 달러에서 2018년 대략 80억 달러로 상승했고 중국은 2016년 10억 달러에서 2018년 80억 달러를 초과
 - 세계 AI 투자금액 비율(% , 벤처스캐너, '18) : 중국(44), 미국(41), 타 국가(15)
- AI 소프트웨어 플랫폼 시장은 2017년 23억 달러 규모에서 2021년 84억 달러 규모로 연평균 약 40%의 고속 성장 예상(IDC, 2017)
- AI 개인비서(IVA: Intelligent Virtual Assistant) 시장은 AI 분야 중 급속히 확산되고 있는 분야로서, 2017년 25억 달러 규모에서 2023년 252억 달러 규모로 연평균 약 46%의 성장 예상(Research and Market, 2018)

◎ 아시아, 태평양 지역 AI 시장 동향

- (아시아·태평양) Tratica에서 발표한 아시아·태평양지역 AI 시장(소프트웨어, 하드웨어, 서비스)은 2017년에 60억 달러에서 2025년에는 1,360억 달러로 성장할 것으로 전망
- AI 소프트웨어 플랫폼 시장은 2017년 23억 달러 규모에서 2021년 84억 달러 규모로 연평균 약 40%의 고속 성장 예상(IDC, 2017)

[All Revenue by Software, Hardware and Services, Asia Pacific:2017-2025]

(단위: 백만 달러)



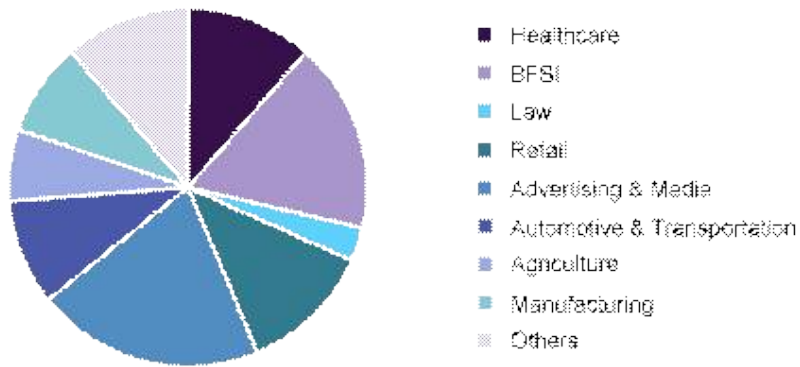
* 출처: Artificial Intelligence Market in Asia Pacific to Grow from \$6 Billion in 2017 to \$136 Billion by 2025, 2018, Tratica

- 중국의 AI 스마트홈 기업은 원격 제어, 사람-기기 소통, 행동·이미지 분석을 통해 서비스를 구현하고 있으며, 음성명령, 습관 설정으로 소비자들이 가전, 가구, 창문 등 가정 내 시설을 제어할 수 있도록 하는 서비스의 개발과 제품의 혁신에 주력
 - 중국의 의료/헬스케어 분야 AI 기술은 주로 딥러닝 및 빅데이터 분석을 통해 구현되고 있으며, 병리진단, 영상분석, 의료데이터 처리, 헬스케어, 정밀의료, 신약 개발 등에 적용되고 있음
 - 중국은 논문실적이 뛰어나며, 슈퍼컴퓨터가 발달해 인공지능에 용이한 환경을 갖추고 있어 후발주자지만 잠재력이 매우 큼
- 저성장 고령화 극복을 위한 국가 경제 사회 혁신의 수단으로 AI 기술 경쟁력 확보의 중요성을 강조하며 AI 기술 혁신을 가속화
 - 제조업 경쟁력이 높은 일본은 유난히 로봇에 집중투자하며 높은 경쟁력을 보임. 현장과 시장 니즈를 중시하는 기업 문화로 인해 스마트팩토리 구현보다 사람의 일을 대체하는 것에 집중

◎ 산업별 전망 및 동향

- (인공지능 산업별 분포도) 인공지능 활용 광고·미디어 산업이 전 세계 시장 매출을 20% 이상 점유하며 인공지능 산업을 주도
 - 인공지능 활용 의료 분야 사업은 2026년까지 인공지능 진단, 로봇 기반 수술, 의료 시뮬레이션을 통한 급속 성장으로 인해 광고·미디어 산업을 추월할 것으로 전망
 - 인공지능 활용 금융 분야 또한 인공지능 기반 투자 및 진단 서비스의 증가로 급속 성장 예상

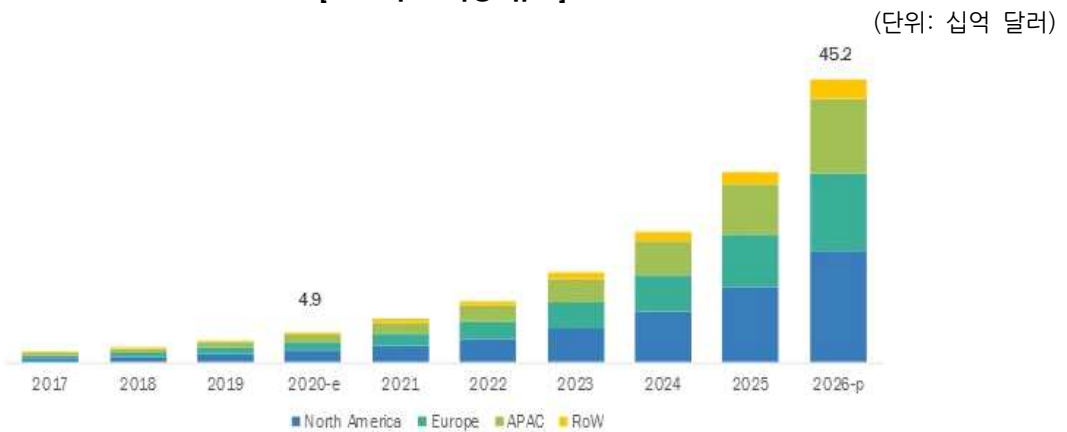
[세계 인공지능 산업별 분포도]



* 출처: Artificial Intelligence Market Size, Share & Trends Analysis Report By Solution, 2020, grandviewresearch

□ 의료 분야 동향

[AI 의료 시장 규모]



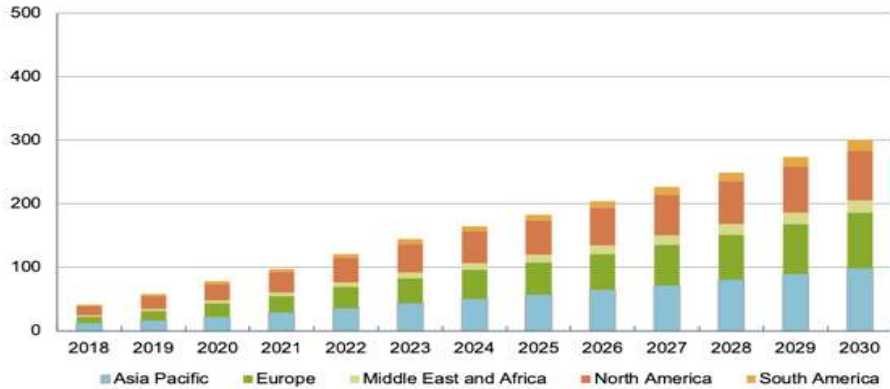
* 출처: Artificial Intelligence in Healthcare Market with Covid-19 Impact Analysis, 2020, Global Forecast to 2026

- AI 헬스케어 시장은 2020년 49억 달러에서 연평균 44.9% 성장하여 2026년 452억 달러로 성장 예상
- 인공지능의 머신러닝을 활용한 의료데이터 분석이 발전하여 의료 관련 비용 및 시간 절감에 기여
- 코로나19에 대비한 전염병 예측 및 시뮬레이션 등 인공지능을 활용하여 질병 산업에 다양하게 활용되는 추세

□ 금융 분야 동향

[AI 금융 시장 규모]

(단위: 십억 달러)



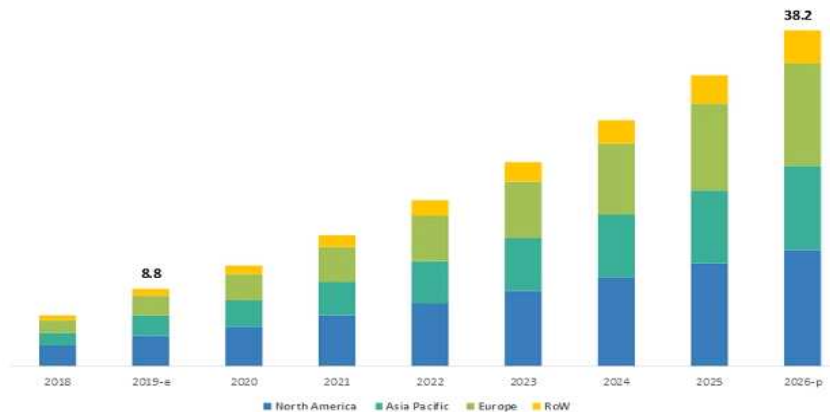
* 출처: Artificial Intelligence in Banking Report, 2018, IHS Markit

- 2018년 AI 금융 시장은 411억 달러로 2030년까지 3000억 달러 규모로 성장 예측
- 금융 관련 보안 및 데이터 정리 및 분석에 인공지능이 활용되며 상용화 가속화 추세
- 북미는 2023년까지 AI 금융 산업을 선도할 것으로 예측되며 북미 시장은 2018년 147억 달러에서 2030년 790억 달러로 성장 전망
- 아시아-태평양 지역은 현재 2번째로 활성화된 시장으로 2024년부터는 북미를 추월 예정이며 2018년 115억 달러에서 2030년까지 986억 달러로 성장 전망

□ 보안 분야 동향

[AI 보안 시장 규모]

(단위: 십억 달러)



* 출처: Artificial Intelligence in Cybersecurity Market- Global Forecast to 2026, 2019, marketsandmarkets

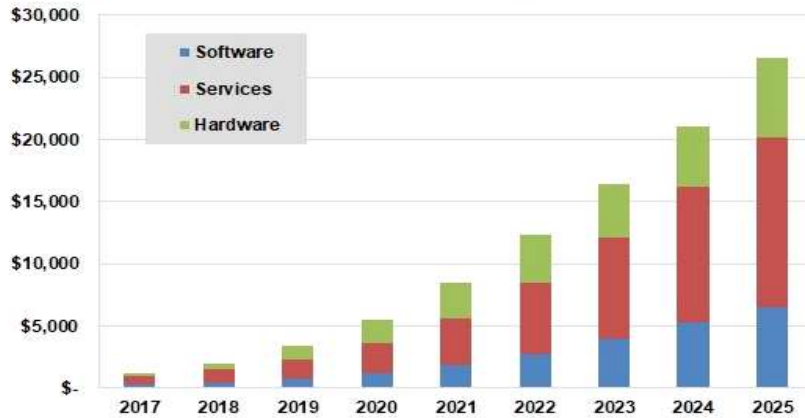
- AI 보안시장은 2019년 88억 달러에서 2026년도까지 382억 달러로 연평균 23.3% 성장 예정
- IoT 등의 발전에 따른 사물 및 네트워크 보안의 중요도 증대되어 기업들은 네트워크, 웹, 모바일 어플리케이션의 보안 솔루션 개발 추진
- 특히 IoT, 빅데이터 보안, 클라우드 컴퓨팅 등의 보안 개발 가속화 추세

□ 자율자동차 분야 동향

- Tratica는 전세계 AI 자율자동차(소프트웨어, 하드웨어, 서비스) 시장이 2025년에 265억 달러로 성장 예상

[AI 자율자동차 시장 규모]

(단위: 백만 달러)

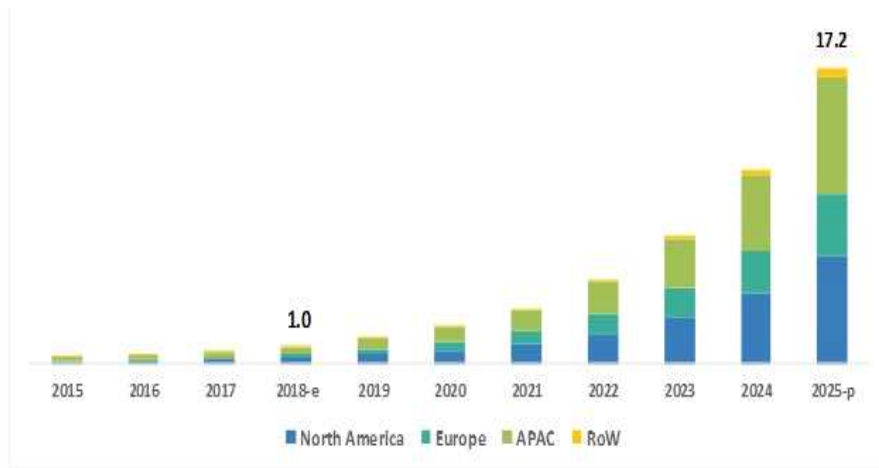


출처: Automotive AI Total Revenue, 2018, Tratica

□ 스마트제조 분야 동향

[AI 스마트제조 시장 규모]

(단위: 십억 달러)



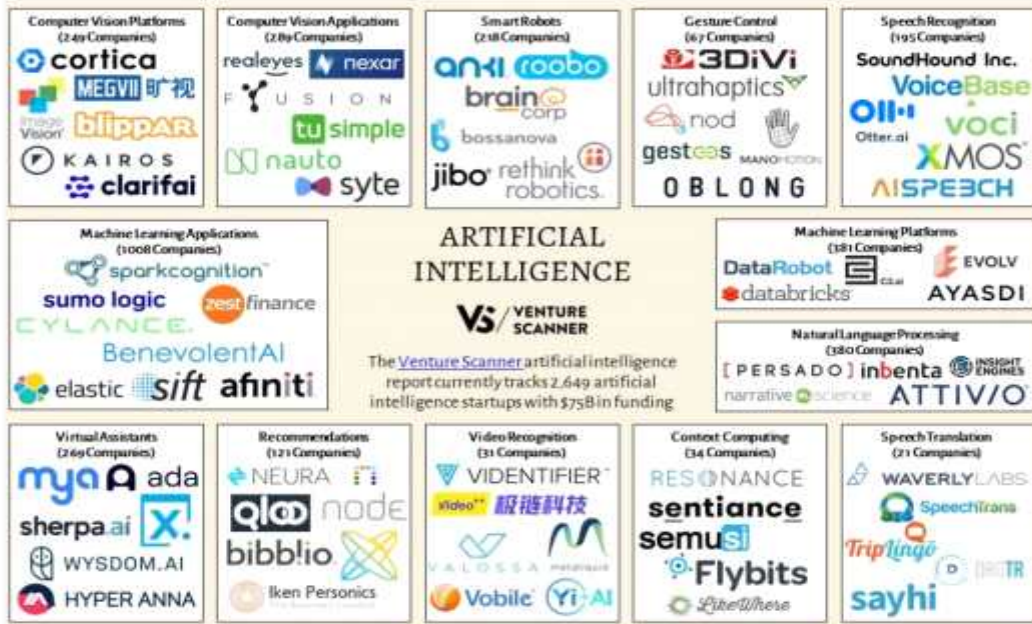
* 출처: Artificial Intelligence in Manufacturing Market by Offering (Hardware, Software, and Services), Technology (Machine Learning, Computer Vision, Context-Aware Computing, and NLP), Application, Industry, and Geography-Global Forecast to 2025, 2019, MarketsandMarkets

- 스마트제조 분야의 AI 시장의 규모는 2018년에 10억 달러에서 2025년에는 172억 달러로 예측하며 아시아·태평양이 2025년에 가장 크게 성장할 것이고 뒤이어 북아메리카, 유럽으로 예상

(2) 세계시장 핵심플레이어 동향

◎ 주요 글로벌 기업 동향

[AI 스마트제조 시장 규모]



* 출처: Venture Scanner Sector Maps 정보 활용, 2019.12

[최근 AI 분야 주요 글로벌 기업 M&A 동향]

대표 인수 기업	스타트업	스타트업 보유 기술
Google	Api.ai	음성 인식 기술, 언어 이해 기술
	AlMatter	모바일 기반 컴퓨터 비전
	Halli Labs	딥러닝/기계학습 시스템 개발
Amazon	Pop Up Archive	음악 검색 엔진
	Init.ai	대화 비서
	Regaind	컴퓨터 비전 sw
Facebook	Ozlo	통합 지식 플랫폼
	Zurich Eye	컴퓨터 비전 SW/HW
Baidu	Raven Tech	AI 음성 비서
	xPerception	머신 비전

* 출처: 4차 산업혁명을 선도하는 주요 기술 대상 기술수준평가 및 기술수준 향상방안, 정보통신기술진흥센터

◎ 미국

- (Google) 알파고를 만들었던 딥마인드를 인수하고 딥러닝 기술개발 플랫폼인 텐서플로를 공개하는 듯 지속적인 연구개발 투자를 진행 중
 - 디스플레이 갖춘 인공지능 스마트 스피커, ‘구글 네스트 허브’ 2020년 7월 공식 출시
- (Amazon) 인공지능 기반의 자사 음성 인식 비서인 알렉사(Alexa)의 어플리케이션을 전면 개편해 2020년 7월 27일부터 보급 시작
 - 알렉사는 클라우드 학습 기반으로 작동하여 음악재생, 알람설정, 날씨 정보 제공, 교통정보 제공 등 음성인식을 통해 다양한 기능 제공

[아마존 알렉사가 탑재된 LG 냉장고]



* 출처: CES 2018

- (Apple) 비즈니스인사이더에 따르면 구글, 아마존 등 어떤 테크 기업보다 많은 인공지능 스타트업을 인수
 - 이스라엘의 안면인식 기술 기업인 리얼페이스(RealFace)의 인수를 통해 아이폰X 페이스ID 기술 개발
- (Microsoft) 자사의 클라우드 ‘애저’에 AI ‘코타나’를 결합하여, 기존 오피스 및 운영체제 제품군을 강화 중이며, ‘애저’를 닛산, 도요타에 자율주행 플랫폼으로 공급
 - Amazon-Microsoft 연합: Amazon의 알렉사 검색에 Microsoft Bing 검색 엔진을 활용, 알렉사를 통한 Microsoft 오피스 접근과 동시에 MS 코타나를 통한 Amazon 전자상거래 플랫폼 오픈
 - AI 팻봇 ‘헬스케어봇’은 2020년 3월 중순부터 덴마크 인구 3분의 1을 책임지는 응급의료 기관 코펜하겐 EMS(Emergency Medical Services)에서 사람 대신 전화응대
- (GM, Ford) 각각 1조 원이 넘는 금액으로 딥러닝 기반의 자율주행 스타트업을 인수, 투자하여 뒤쳐졌던 기술 경쟁에 대응
- (IBM) 왓슨이 금융, 의료 등에 적용되면서 관련 전문 지식을 축적해 왓슨만의 생태계 구축

◎ 중국

- (Baidu, Alibaba, Tencent) 바이두, 알리바바, 텐센트 등 중국 3대 IT기업들은 AI, 로봇, 자율주행 등과 같은 기술에 대규모 투자해 미래 혁신 분야에서도 미국 기업들을 추월했으며 중국은 4차산업 인공지능 특허출원 1위로 7.68만 건이며, 2위 미국이 추격
 - 바이두의 NLP(자연어처리) 어니(ERNIE)는 구글의 버트(BERT)초월
 - 텐센트는 클라우드 중국 시장점유율 2위로 성장 (1위 알리 클라우드 40.5%(알리바바), 2위 텐센트 16.5%)
- (Alibaba) Alibaba는 중국의 스마트 보건 · 양로 중심의 의료산업 육성정책에 힘입어 ‘알리 헬스케어’와 ‘의료 클라우드’ 서비스를 개발했으며, 스마트 의료 생태계 구축사업을 추진 중인 텐센트 등 BAT 인터넷 기업들 또한 적극 참여
 - 중국의 대형병원 100여 곳은 현재 알리바바가 개발한 AI CT 판독 시스템으로 코로나 감염 여부를 진단
- (샤오미) 유니콘 출신인 샤오미는 ‘자체R&D-자체생산-자체판매’ 체인으로 2018년 기준 1억 개 하드웨어 기기가 연결되어 있는 스마트홈 생태계 구축에 성공
- (Sensetime) 센스타임은 현재 중국의 최대 인공지능 알고리즘 공급 업체로서, 인공지능 기술을 금융, 스마트시티, 스마트폰, 자동차, 모바일 인터넷, 로봇, 자율주행 등 다양한 분야에 제공
- (Midea·Haier·Gree) 메이디·하이얼·거리 등 전통 가전기업은 높은 시장점유율을 바탕으로 현재 5천만 명의 사용자를 확보한 ‘하이얼U+’를 개발하는 등 혁신을 추구
 - 2017년 기준 중국의 스마트가전 시장은 2,828억 위안(약 47조 원) 규모로 스마트 냉장고/에어컨/세탁기 시장은 약 14조 원, 스마트 TV/스피커 시장이 약 25조 원 규모에 달함

나. 국내 시장 분석

(1) 국내 시장 동향 및 전망

- 한국 IDC가 2020년 발표한 국내 인공지능 시장 규모 전망에 따르면 2018년 국내 인공지능 시장 규모는 2,821억 원으로 평가되었으며 연평균 성장률 17.8%로 2024년 7,539억 원으로 성장 전망

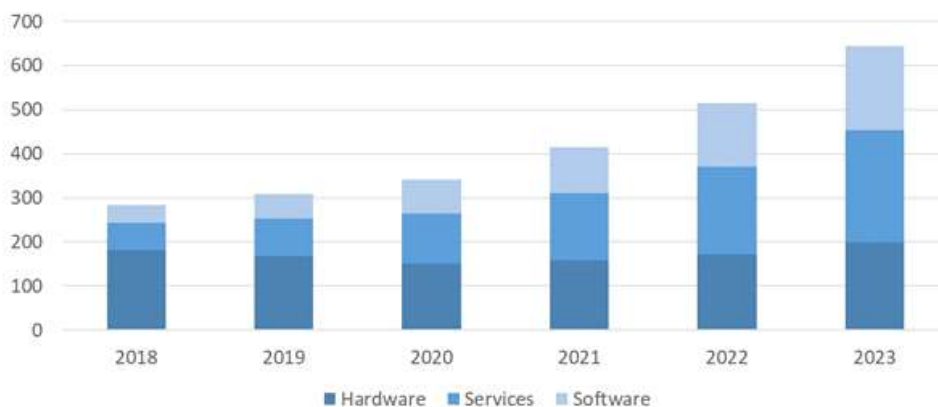
[AI 국내시장 규모 및 전망]

(단위 : 억 원, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
합계	2,821	3,323	3,915	4,612	5,433	6,400	7,539	17.8%

* 출처: 한국IDC - 국내 인공지능 AI시장 전망(2020.04.03.)

[국내 인공지능 시장 전망]



* 출처: 한국IDC - 국내 인공지능 AI시장 전망(2020.04.03.)

- 국내 AI 시장은 2018년 2,033억에서 연평균 17.8% 성장하여 2024년까지 5,432억 원으로 성장 전망
 - 인공지능 하드웨어 부문은 앞으로 성장률이 둔화될 전망이지만 소프트웨어 시장은 급성장 예상
 - 업무 프로세스 효율화 및 비즈니스 자동화를 위한 AI 어플리케이션 및 플랫폼 사업은 5년간 연평균 30% 이상 성장할 것으로 예측
- 국내 AI 산업은 기술 및 인력 모든 측면에서 후발주자지만 우수한 정보통신기술(ICT) 인프라와 교육수준 및 신기술에 대한 빠른 수용성을 통한 빠른 성장 가능
 - 정부와 공공기관이 AI 인재양성 및 생태계 구축에 적극적으로 나서고 투자를 강화하며 디지털 트랜스포메이션을 가속화 추진

- 2020년 기준 음성인식 및 통·번역 시장 4.2조 원, 영상처리 및 영상인식 시장 3.5조 원, 기타 SW 및 알고리즘 관련 시장은 3.4조 원 규모가 될 것으로 예상
 - 행정안전부, 과학기술정보통신부, 산업통상자원부 등 정부 부처들은 지능형 기술의 성장에 주목하고 있으며, 과학기술정보통신부는 2017년 범부처 지능형 CCTV 사업계획을 발표
 - 현재 보안·교통관제·비즈니스, 인텔리전스 등 영상인식 및 처리가 필요한 분야에 지능형 시스템이 상용화되어 있으며, IP 카메라, 클라우드 기술의 접목으로 빅데이터를 활용한 데이터 마이닝 기술 구현
 - 삼성전자, 네이버, LG 등 국내기업의 AI 도입 및 적용이 본격화되며 AI 시장 규모는 크게 확대될 것으로 기대

- 오랜 기간 각종 데이터를 분석/처리하고 AI 기술력을 축적해온 글로벌 기업들에 의해 국내 AI 시장이 잠식당할 위험성 존재
 - Google(검색, 안드로이드OS), Amazon(전자상거래), Facebook(SNS) 등 글로벌 IT 기업들은 각자 강점을 가진 플랫폼을 기반으로 빅데이터를 축적해왔고 개인화된 서비스 제공에 AI를 활용하며 시장에서 독점적 지위 구축
 - 한국어에 대한 자연어 처리 역량, 하드웨어 플랫폼 장악력 등 국내기업들의 강점을 적극적으로 활용하면 내수 시장 방어가 가능할 것으로 예측
 - 스마트폰, 가전 기기 등 국내 기업들이 주도하는 제품들을 국내 지능형 개인비서의 하드웨어 플랫폼으로 활용하며 해외 기업과의 주도권 경쟁을 지속할 필요 있음

(2) 국내 생태계 현황

◎ 생태계 주요 이슈

- 인공지능 분야에서 크게 이슈되고 있는 분야는 기계 학습, 지식 추론, 시각 지능, 언어 지능 등 4가지 분야이고, 해당 분야에서 기술 개발이 가속화되는 중
 - 국내 대기업인 삼성전자는 딥러닝 관련 플랫폼을 개발 중이며, 삼성전자 외에도 SKT, 네이버 등과 같은 IT 대기업도 AI 기술 개발에 참여하여 AI 논문 및 시가 적용된 제품을 출시하고 있음

[AI 생태계 이슈]

소분야	생태계 주요이슈
기계 학습	<ul style="list-style-type: none"> • AI 기술의 핵심 기반기술인 기계학습은 데이터 기반의 효율적인 학습을 위한 새로운 기술들이 등장하고 있으며 그 중 딥러닝, 전이학습, 강화학습 중심으로 연구개발이 활발히 진행 • AI 전반에 비약적인 발전을 견인한 딥러닝은 시각지능, 언어지능 등에 특화된 학습 알고리즘이 지속적으로 개발되고 있으며 산업적 활용을 위한 오픈소스 플랫폼 경쟁이 치열 • 삼성전자는 딥러닝 응용 프로그램 개발을 위한 분산형 플랫폼인 ‘베레스’를 오픈소스로 공개
지식 추론	<ul style="list-style-type: none"> • 지식추론 기술은 기존 정보들을 단순 응답·추천하는 기술에서 정보 간 상대적인 관계를 유추하는 기술로 발전하였으며, 추론형 질의에 인간을 뛰어넘는 기술이 등장 • 최근 지식추론 기술은 질문과 연관성이 있는 정보를 유추하거나, 직접적으로 데이터에 존재하지 않더라도 부분 정보들의 조합을 통해 답을 찾아가는 기술이 활발히 개발되고 있음 • 국내 추론 기술로는 엑소브레인 프로젝트에 참여한 솔트룩스의 AI 플랫폼인 ‘아담(ADAMs)’이 있으며, 대규모 지식베이스기반 지식 학습과 복합 추론 기술을 통해 추론 유형의 문제에서 정답률 94%를 보임
시각 지능	<ul style="list-style-type: none"> • 이미지/영상 데이터 기반의 시각 지능은 이미 인간 수준을 넘은 시각 인식률을 보이고 있으며 상황을 이해하고 새로운 이미지를 생성하는 기술로까지 발전 • 이미지/영상의 외형적 특징을 통해 이해하는 기술이 발전하고 있으며, 눈, 코, 입 모양 등의 상관관계를 분석해 표정을 이해하거나 감정을 추측하고, 이미지 속 상황을 정확히 이해하여 언어로 표현하는 기술이 등장
언어 지능	<ul style="list-style-type: none"> • 최근 언어 지능은 텍스트/음성 데이터 기반 학습을 통해 사람에 의존하지 않고 스스로 언어를 이해하며 인간의 억양과 유사한 수준으로 음성을 생성하는 단계로 진입 • 기존 Ontology 기반 기술에서 인간의 개입을 최소화하고 데이터 기반 학습을 통해 스스로 언어를 이해하게 하는 Word Embedding 방식으로 전환 • 언어 인식의 지능을 갖게 된 AI는 사람의 목소리를 이해하고 생성하며 악센트뿐만 아니라 문장 단위에서의 억양까지 매우 정교한 수준으로 구현

* 출처: 한국과학기술기획평가원, AI(SW), 2018

◎ 산업 현황

- 한국정보화진흥원에 연구에 따르면 국내 인공지능 기업은 26개로 미국의 2028개에 비교하여 매우 적었으나, 스타트업의 경우 국내 465개, 미국 1394개로 상대적으로 높게 나타남

[국가별 인공지능 인프라 현황]

구분	미국	중국	한국
우수논문 수(인용 기준)	2486	748	96
AI 논문 생산순위	2위	1위	10위
AI 특허 수	9,786	6,900	2,638
머신러닝 리더급 학자수	253	100	0
AI 기술수준	100	81.9	78.1
AI 투자규모	1.2조 원	6조 원	2,344억 원

* 출처: 소프트웨어정책연구소, 스카이데일리, 2018

- 한국의 논문 생산순위, 인재 양상, 투자규모로 보았을 때 AI 인프라는 미국, 중국에 비해 미비
- 한국의 투자 규모는 중국의 4% 수준으로 타국(미국, 일본, 영국 등)에 비해서도 적은 편으로 적극적 육성 필요
- 한국 AI 스타트업 기업으로는 딥러닝 기반의 의료영상 스타트업이 선전 중
- 국내 주요 ICT 업체들은 기계학습, 언어/시각 지능 등 범용적으로 사용 가능한 AI 플랫폼을 출시하고 있으며 본격적인 시장 공략 개시
- AI 산업 구조는 크게 AI 기초 연구, AI 응용연구, AI 개발연구, 차세대 AI 등이 있으며, 각 분야에 투자 및 기술개발이 이루어지고 있는 실정

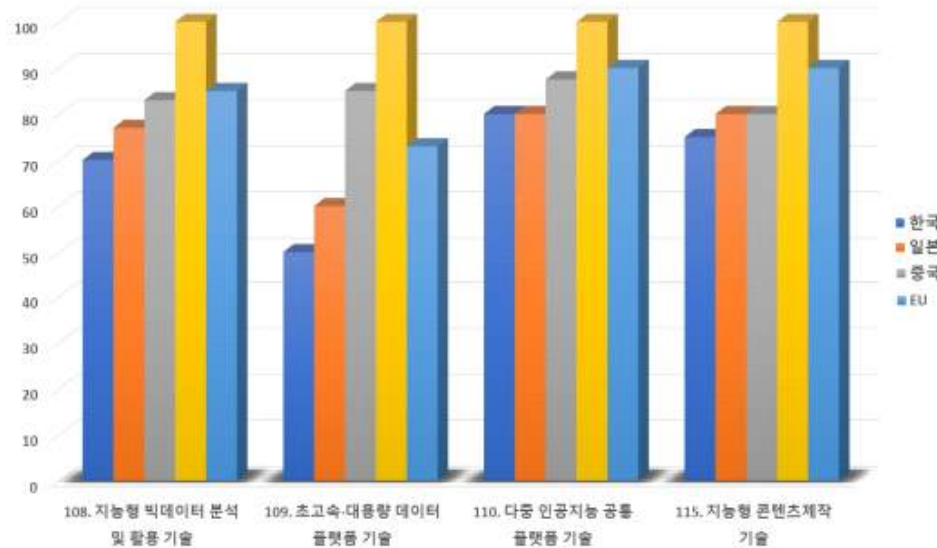
[AI 연구 영역의 구분과 정의]

구분	상품 및 기술
AI 기초연구	<ul style="list-style-type: none"> • AI 성능 자체를 향상시키기 위한 연구 - 예시 : 메타 학습, 설명 가능한 AI 등
AI 응용연구	<ul style="list-style-type: none"> • 컴퓨터 과학의 인식 및 탐지 분야 연구 - 예시 : 컴퓨터 비전, 자연어 처리, 음성 인식, 데이터 마이닝 등
AI 개발연구	<ul style="list-style-type: none"> • AI 응용연구가 타 산업에서 활용되는 연구 - 예시 : 의료 이미지에서 암 진단, 금융 이상거래 예측 등
차세대 AI	<ul style="list-style-type: none"> • 사람 수준의 지능을 구현하기 위한 AI 연구 (향후 20년) - 예시 : 범용 AI, 인공 생명 등

* 출처: 이슈리포트, 소프트웨어 정책 연구소, 2018, SPRI

◎ 국내 기술 수준

[국가별 인공지능 기술 수준]



* 출처: ICT-SW 2018년 기술수준평가, 2019, 한국과학기술기획평가원

- AI 기술수준이 상승세이나 중국에 비해 상승속도가 느림
 - 정부 투자가 늘어가는 추세이나 원천기술 쪽의 투자가 그렇게 많지 않음
 - 주로 딥러닝 기술, 자동 통·번역 연구가 활발하게 이루어지고 있음
- 한국은 이미 해외에서 검증된 시스템을 도입하는 것이 일반적이라 응용 수준이 어느 정도 되어 보이더라도 향후 경쟁력 강화를 위해 해당 분야 연구/개발 및 선제적 투자 필요한 상황
- 전체 논문 건수는 미국과 중국이 가장 앞선 것으로 나타났으며 중국은 시각지능과 기계학습에서 선두를 차지
 - 국가별 우수논문 건수(인용 기준)는 미국이 2,486건으로 우수논문 수는 가장 많은 것으로 나타났으며 중국은 748건, 우리나라는 96건

◎ 국내 인공지능 생태계 현황

- 의료기기에 인공지능의 도입으로 영상, 음성 그리고 데이터 인식 기능을 접목하여 판독 정확성을 높이고, 질병 예측 및 예방 등 새로운 가치 창출을 기여할 수 있을 것으로 기대
 - 세브란스병원에서는 AI 의료녹취 솔루션을 통해 영상 데이터 판독 소견 음성을 실시간으로 문서화하여 저장하고 시스템에 등록하는 작업을 자동화
 - 고대안암병원은 만성질환자를 위한 치료 전에 ‘문진 챗봇’을 사용
- 제조분야에 인공지능 기술의 융합으로 장비의 디지털 모델과 연계된 공정·장비의 실시간 모니터링 및 인공지능 기반의 실시간 분석(엣지컴퓨팅)을 통해 능동제어(자가최적화), 예측진단 등이 가능한 장비 지능화 기술 실현
 - 삼성 SDS는 설비, 공정, 검사, 자재물류 등 제조 4대 핵심설비에 센서를 부착하여 수집된 대용량의 빅데이터를 AI로 분석 실시간 이상 감지 및 장애시점 등을 예측
- 농업과 인공지능(AI)과 드론, 사물인터넷(IoT) 등 첨단기술을 결합해 단위 면적당 생산성을 개선하는 ‘어그테크(AgTech·Agricultural Technology)’가 전 세계적으로 주목
 - 전 세계 어그테크 스타트업과 투자자 정보를 제공하는 업체 어그펀더(Agfinder)와 LG경제연구원의 최근 자료를 보면, ‘10년 4억 달러에 불과했던 어그테크 투자 규모는 ’17년 43억 달러(약 4조 8,600억 원)로 10배 이상 증가
- 인공지능의 딥러닝 기술을 도시 빅데이터 분석에 도입하여 도시 내 여러 영역 간의 데이터를 융합하고 여러 도시에서 이뤄지는 데이터 분석 결과를 효과적으로 공유할 수 있는 스마트 시티 구현
- 오픈소스를 활용한 AI·빅데이터 등 자체개발 S/W의 공유가 활발해지면서 IT산업 생태계 변화에 기여 중인 추세
 - 깃허브는 전 세계 개발자가 가입해 활동하는 대표적인 코드 공유 사이트로, 정부 출연 연구기관 가운데 최초로 한국전자통신연구원(ETRI)가 오픈소스 공개
 - ETRI는 지금까지 인공지능(AI), 빅데이터, 클라우드 관련 총 9종의 오픈소스가 깃허브에 공유했으며, IoT웨어는 운영체제(OS) 5종, 센서·통신 등 펌웨어 같은 소프트웨어의 기본적인 뼈대와 기능을 담은 프레임워크를 제공
 - ETRI의 오픈소스 중에서 ‘로봇을 위한 인공지능(AIR)’은 일상생활에서 고령자를 돌보는 로봇에 특화된 데이터와 소프트웨어를 공유하며, 아파트를 테스트베드 삼아 노인 50명과 20, 30대 50명 등 총 100명이 55가지의 일상적인 행동을 할 때 이를 3차원(3D)으로 촬영한 동영상 약 12만 건이 담김
 - 오픈소스는 과학기술 연구개발(R&D)에서도 활용 가능성이 큰 만큼 깃허브에 공개한 오픈소스를 더 많은 기업과 더 많은 개인 개발자들이 활용할 필요 증대

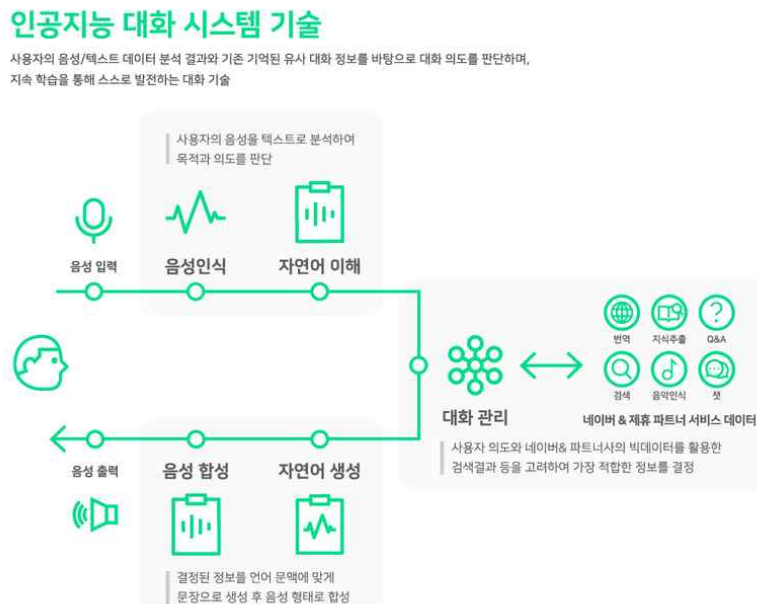
(3) 생태계 핵심플레이어 동향

◎ 지능형 서비스 동향

□ 대기업 동향

- (네이버) 네이버는 다양한 AI 기반 서비스를 제공
 - 네이버는 자사 AI 기술 플랫폼 클로바를 탑재한 AI 비서 앱 네이버-클로바를 출시
 - 클로바는 한국어와 영어 음성 대화를 지원, 지식 정보 검색, 음악 추천, 통번역, 영어회화, 감성 대화 등의 기능을 갖춘

[네이버 클로바의 인공지능 대화 시스템 개요도]



* 출처: [Clova AI] '음성검색 기술과 네이버 그린닷', 2019, 네이버 클로바 공식 블로그

- 네이버는 2019년 11월 인공지능 기반 검색어 추천 시스템 '리요(RIYO: Rank-It-YOurself)'를 도입하여 개인 맞춤형 검색 서비스 제공
- 리요는 AI 시스템이 1차 검색어와 주제 카테고리의 연관성을 분석한 뒤, 개인별 설정 기준에 맞추어 급상승검색어 차트 노출 여부를 결정
- 네이버는 인터넷 기사 댓글 모니터링을 위한 'AI 클린봇' 도입하여 악성 댓글 필터링
- (KT) 기업의 경비 처리를 더 쉽고 빠르게 처리할 수 있는 챗봇 기반의 전표를 대신 처리하는 전표 로봇(이하 전대리) 솔루션을 개발
 - 전대리는 챗봇 기반의 RPA 프로그램으로서 자주 처리하는 전표의 이력을 추천하고, 시스템에 접속하지 않아도 메신저 채팅을 통해 몇 번의 클릭만으로도 전표에 필요한 처리 업무 가능
- (한국투자증권) 국내 최초로 인공지능 기술을 적용한 리서치 서비스 '에어' 를 2020년 7월 오픈
 - 에어 서비스는 대량의 정보 안에서 의미 있는 뉴스를 선별 제공 및 뉴스 데이터와 계량분석을 기반으로 양질의 투자정보 제공

□ 중소기업 동향

- (코노랩스) 기존 기업의 일정 및 고객관리 시스템과 연동하는 서비스인 ‘코노’를 제공, ’17년에는 글로벌 이메일봇 서비스를 ’18년 3월에는 슬랙봇 서비스를 제공하여 개인을 넘어 기업 내 스케줄을 지능적으로 관리하는 AI 서비스 시장 점유를 높이는 추세
 - 인간을 대신하여 일정을 관리함으로써 효과적인 시간 관리가 가능하게 함과 동시에, 일정관리에 드는 시간을 단축시켜 업무 효율성 증대
- (셀바스) AI는 텔레투게더와 음성 인식 기반의 AI 회의록 서비스 사업화 MOU를 체결
 - 지능형 미팅 비서서비스는 국내 최초 음성인식 기반의 AI 회의록 서비스로서 컨퍼런스 및 화상회의 환경에서 생성되는 모든 음성 회의 내용을 디지털화함으로써 회의에 참석하는 이들의 업무 효율을 높일 수 있을 것으로 기대

◎ 시각 지능 활용 동향

□ 대기업 동향

- (SK C&C) Amazon에서 선보인 무인점포 ‘Amazon GO’ 기술과 유사한 안면과 객체를 인식하는 AI ‘비전AI’를 개발하고 고객이 인식하고 원하는 물건을 추천하는 방법 구현
 - SK C&C는 CU 편의점과 협약을 맺고 미래형 편의점 구축 중
- (SKT) 보안업체 ADT캡스를 인수하고 AI, 사물인터넷, 빅데이터 기술을 통신인프라에 접목하여 지능형 보안 서비스 사업을 추진
 - 사고 발생 가능성이 높은 지역을 사전에 예측하여 경비 인력과 차량 동선을 최적화하고, 빅데이터 기반 지능형 영상분석을 통해 범죄자의 특이 행동, 이상 징후까지 판단해 대응할 수 있으며, 8K UHD 해상도의 CCTV 영상을 실시간 분석하여 범죄 혐의자의 행동을 탐지하여 즉각 대처
- (네이버) 2개의 판별망과 3개의 생성망을 통해 구현한 StarGAN을 개발하였으며 머리색, 성별, 나이, 감정 등 여러 개의 특징을 한 번에 변화시킬 수 있는 이미지 변환 모델로 CVPR2018 학회에서 국내 산학기관 중 유일하게 상위 2% 논문으로 선정
- (삼성전자) AI 기술을 더한 사운드 기능과 저해상도 영상을 8K급 화질로 변경해주는 ‘AI 퀀텀 프로세스 8K’ 기술이 탑재된 ‘QLED 8K’ TV를 선보여 CES 혁신상 수상

□ 중소기업 동향

- (트리플렛) ‘오조 프로젝트’는 인공지능 카메라와 사물인터넷 기술을 도입하여 산업현장을 감시하는 AI카메라가 탑재된 스마트 재난감시 솔루션으로 ‘2020 소방방재 기술 산업전’에서 발표

◎ 스마트 헬스케어 동향

□ 대기업 동향

- (삼성전자) 혈압측정, 수면 분석, 스트레스 수준 측정등의 기능을 포함한 스마트 헬스케어 웨어러블 ‘갤럭시워치 3’ 출시
 - 삼성전자의 세계 최초 혈압 측정 어플리케이션 ‘삼성 헬스 모니터’ 한국 식약처 허가
 - ‘삼성 헬스 모니터’는 혈압계 없이 24시간 혈압 측정 가능

□ 중소기업 동향

- (뷰노) AI(AI)이 엑스레이 영상을 분석하여 환자의 뼈 나이를 제시하고, 의사가 제시된 정보 등으로 성조숙성이나 저성장을 진단하는데 도움을 주는 소프트웨어인 VUNOmed-BoneAge를 출시하였으며 '18.5.16에 의료기기 허가를 받음
 - 뷰노는 MRI 영상을 통해 1분 안에 치매를 예측하는 기술이 탑재된 뷰노메드 딥브레인도 출시
- (루닛) 환자의 뇌 MR(Magnetic Resonance) 영상과 심방세동유무에 대한 자료를 바탕으로 뇌경색(허혈성 뇌졸중)의 유형을 자동으로 분석하여 의료진의 뇌경색 유형 분류를 진단하는데 지원하는 소프트웨어인 Lunit INSIGHT를 출시

◎ AI 하드웨어 동향

□ 대기업 동향

- (삼성전자) 스마트폰 전용의 AI 반도체를 상용화
 - 삼성전자는 5세대 이동통신을 지원하는 ‘5G 통신 모델’과 고성능 ‘모바일 AP’를 하나로 통합한 ‘엑시노스 980’을 공개
 - ‘엑시노스 980’은 고성능 NPU(신경망처리장치)가 내장되어 인공지능 성능이 강화
- (SK 텔레콤) 글로벌 기업과의 협업을 통해 데이터센터 및 엣지 디바이스 전용 AI 반도체 공동연구를 추진 중
 - 2018년 8월 미국 자일링스와 협업해 국내 대규모 데이터센터로는 처음으로 자사 데이터센터 AI 반도체로 자일링스 FPGA(Field Programmable Gate Array)를 채택하였으며, AI 가속 솔루션을 개발해 AI 서비스 ‘누구’에 적용

□ 중소기업 동향

- (넥셀) AP(Application Processor) 전문업체 넥셀은 국내에서 유일하게 AI 반도체 설계원천기술 개발을 추진 중
 - 셀(XELL)로 명명한 AI 반도체 설계기술을 개발 중

3. 기술 분석

가. 해외 기술 동향

(1) 주요 기술

◎ 오픈 플랫폼

- 공개된 오픈 플랫폼은 다양한 인터페이스를 제공함으로써, 사용자가 복잡한 플랫폼 내부 구조를 몰라도 손쉽게 이용 가능하도록 하여 AI 기술 활용의 진입장벽을 낮추는 중
 - 많은 인력과 리소스를 투입하여 AI 기술을 직접 개발하는 것이 아닌 오픈 플랫폼을 활용하여 서비스를 개발하는데 집중
 - 오픈 플랫폼의 종류가 다양하므로 사용자는 상황에 따라 적절한 플랫폼을 선택하여 구축 가능
 - Google의 TensorFlow는 머신러닝을 위한 엔드 투 엔드 오픈소스 플랫폼입니다. 도구, 라이브러리, 커뮤니티 리소스로 구성된 포괄적이고 유연한 생태계를 통해 연구원들은 ML에서 첨단기술을 구현할 수 있고 개발자들은 ML이 접목된 애플리케이션을 손쉽게 빌드 및 배포 가능
 - Microsoft의 Accord.NET은 C#으로 개발된 닷넷 기반의 기계학습 라이브러리
 - Baidu의 PaddlePaddle은 C++로 작성되었고 파이썬 인터페이스를 가지고 있으며 수학 연산의 성능과 분산환경 그리고 순환 신경망에 대해 최적화

[주요 AI 오픈 플랫폼]

오픈 플랫폼	모바일 지원	공개 기업 및 연구 그룹
TensorFlow	TensorFlow Lite를 통한 지원	Google
Accord.NET	-	Microsoft
CNTK(Connitive Toolkit)	-	
PaddlePaddle	-	바이두(Baidu)
DeepLearning4J	-	스카йма인드(Skymind)
MLib	-	아파치 소프트웨어 재단 (Apache Software Foundation)
Mahout	-	
DeepMask, SharpMask, MultiPathNet	-	Facebook
Caffe2	지원	
MXNet	지원	카네기멜론(Carnegie Mellon) 및 워싱턴(Washington) 대학 등

* 출처: "주요 AI 오픈 플랫폼", 2019, 금융보안원

◎ 알고리즘 개발

- 기계학습 분야에서 데이터로부터 특정 업무를 수행하기 위한 정보를 학습시키기 위한 연구를 통해 다양한 알고리즘이 개발
- 딥러닝 기술은 여러 비선형 변환기법의 조합을 통해 높은 수준의 추상화(abstractions, 다량의 데이터나 복잡한 자료들 속에서 핵심적인 내용 또는 기능을 요약하는 작업)를 시도하는 기계 학습 알고리즘의 집합
- 머신러닝은 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야
 - 지도학습(Supervised Learning)은 학습 데이터마다 레이블을 가지고 있으며 주어진 입력과 출력 쌍들을 매핑해주는 함수를 학습
 - 비지도학습(Unsupervised Learning)은 학습 데이터가 레이블을 가지고 있지 않으며 입력만 있고 출력은 없는 상태에서 이루어지는 학습
 - 준지도학습(Semi-Supervised Learning)은 학습 데이터가 약간의 레이블을 가지고 있으며 비슷한 데이터끼리 묶는 클러스팅이 주로 사용되며 지도학습에 비해 학습하기 까다로움
 - 강화학습(Reinforcement Learning)은 최종 출력이 바로 주어지지 않고 시간이 지나고 주어지는 학습
- 영국 Cambridge 대학의 Cecilia Mascolo 교수 연구팀은 음성 분석 기반으로 코로나 바이러스 증상 여부를 감지하는 기계학습 알고리즘 기반의 'COVID-19 Sound App'을 개발
 - 사운드앱 어플리케이션이 COVID-19 환자의 음성을 구체적으로 분석하여 질병이 있는 사람을 판명할 수 있다고 설명
 - 바이러스 환자가 숨쉬는 것뿐만 아니라 마른 기침 및 호흡 패턴의 간격을 모니터링한 AI 알고리즘이 코로나 바이러스의 감염여부를 진단하는 작동 원리
 - 알고리즘 고도화를 통한 정확도 향상을 위해서는 더 많은 바이러스 환자의 음성데이터를 확보하여 음성의 미묘한 변형을 파악하는 작업이 필수적으로 요구됨

(2) 해외 기업 기술 현황

◎ 인식

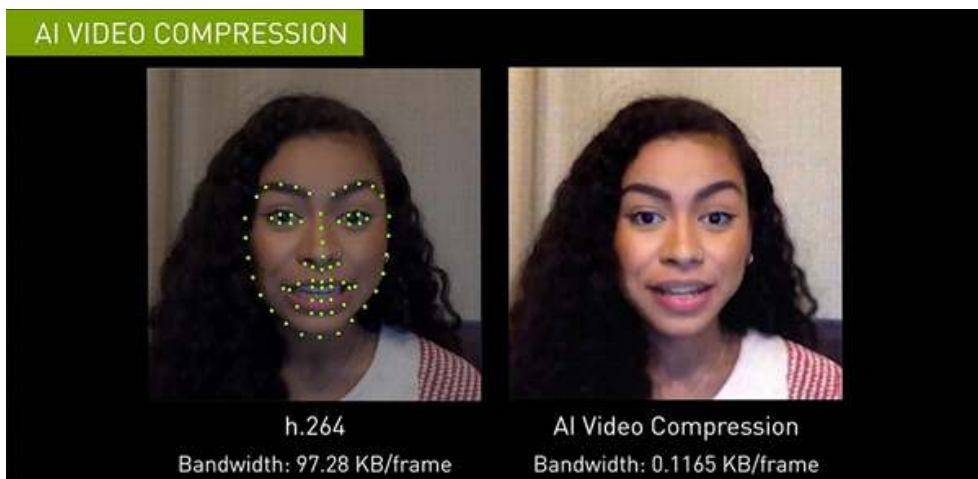
- (Megvii) 얼굴인식 기술의 선두주자로 얼굴인식 소프트웨어 오픈 플랫폼 기반 사람의 감정인식, 시선위치 추적, 나이/성별 분석 등 다양한 AI 기술을 활용한 인공지능 분석 제공
 - 메그비의 기술은 범죄자 얼굴 인식 등 공공 영역에서 활용되기 시작하면서부터 현재 유통과 금융 분야에 걸쳐 민간의 영역으로 확장 중인 추세
- (Softbank) 클라우드에 연결되어 지속적으로 학습 및 성장이 가능한 로봇 개발을 목표로 감성인식 로봇 '페퍼'를 개발하여 상용화
 - 지능형 로봇 분야에서는 중소기업 중심 제품 개발이 수년간 지속되었으나, 최근 글로벌 IT 기업이 자체 보유한 AI 기술을 바탕으로 자율성 높은 AI 로봇 개발이 가속화
 - 다수 로봇 기업들이 초기에는 필요한 기술자체 개발을 시도하였으나, 최근에는 글로벌 IT 기업들이 기존에 개발한 AI 비서의 음성, 언어 기술을 채택·협력하는 방법을 선호
- (DARPA) 원거리 사람 추적 및 식별을 위한 HID(Human Identification at a Distance)과제를 수행, 걸음걸이 인식, 얼굴인식, 홍채인식 등의 기술 개발
 - 시각지능 분야에서는 미국을 중심으로 영상의 내용을 이해하는 기술에 대한 연구가 진행 중, 상황을 이해하고 행동을 분석하며 미래에 발생할 상황을 예측하는 기술을 집중 개발
 - Google, MS, Facebook, UC버클리, INRIA 등은 CNN(Convolutional Neural Network)을 이용하여 동영상 내에서의 객체 및 행동 인식 연구 추진 중인 것으로 알려짐
- (Amazon) 2014년 11월 사용자의 음성 명령을 인식해 동작하는 알렉사 기반의 'Amazon 에코'를 공개하며 지능형 스피커 시장을 개척
 - 출시 이후 주력시장인 미국에서 Amazon 에코의 판매량은 약 810만대로 집계(출처:Consumer Intelligence Research Partners, 2017.1.)
 - 기기 판매 뿐 아니라 쇼핑 구매 증가, 스킵(에코 사용자가 원하는 대상 (기기, 앱)에게 음성명령을 내릴 수 있는 단위기능) 유료화, 클라우드 서비스 등을 통해 2020년까지 관련 매출이 100억 달러로 전망(출처:RBC Capital Markets,2017.1.)
 - 하드웨어를 통한 쇼핑 접점 확대, 스마트홈 시장 선점, 알렉사 중심의 AI 생태계 조성이라는 전략적 의도
 - Amazon이 보유한 기계학습 알고리즘을 클라우드 사용자가 활용하여 부정거래탐지, 요구예측, 콘텐츠 개인화, 사용자 행동 예측, 소셜미디어 확인, 텍스트 분석 등 AI을 요구하는 애플리케이션 구축 가능
 - Amazon은 Amazon 모바일 앱에 AI 기능을 기반으로 한 의류 검색 서비스 'Style Snap'을 발표

◎ 러닝(Learning)

- (Google) Google의 머신러닝 플랫폼은 클라우드 서비스 형태로 이미 상용화된 서비스이며, Pre-Trained model, 인공지능경망 기반의 기계학습 플랫폼, 이미지 검색, 음성검색 등을 제공
 - AI검색 알고리즘 Rank Brain, 바둑AI 프로그램 AlphaGo, 기계학습 오픈소스 소프트웨어 Tensor Flow 개발 등 AI(AI) 분야의 선두주자로, 세계 최대 규모의 데이터 센터를 운영
 - 머신러닝 플랫폼은 음성 인식, 이미지 분석, 번역 기능을 수행하는 플랫폼으로, 오랜 시간과 높은 개발 역량이 요구되는 부분을 Google의 AI 플랫폼이 대신 처리
 - Google은 이러한 AI를 제공하는 영역을 점차 다양한 영역으로 확대하는 등 전 산업 영역을 포괄하는 범용 AI 플랫폼을 확산시킬 것으로 전망

- (Nvidia) 자사의 GPU의 탁월한 딥러닝 성능을 기반으로 AI, 드론, 자율주행차 반도체 분야로 발 빠르게 사업 확대
 - Nvidia의 AI 기반 정보처리 반도체인 ‘Nvidia Drive’는 빠른 속도로 대용량의 데이터 연산 처리가 가능해 커넥티드카와 자율주행 기능 구현을 위한 최적의 플랫폼으로 평가
 - 주행 중 실시간 교통정보 및 공사구간, 뉴스 등을 연계하는 등 대량의 정보를 효율적으로 수집·처리하는 기술은 미래차의 핵심
 - Nvidia는 2020년 10월 GPU 가속 AI 회의 소프트웨어 플랫폼인 엔비디아 맥신(Nvidia Maxine)을 출시
 - 화상 회의 기능을 제공할 뿐만 아니라 인공지능을 활용해 시선 보정, 초고해상도, 주변 소음 제거, AI 아바타, 가상 배경화면, 얼굴 이미지 합성과 같은 다양한 기능들을 추가 제공
 - 맥신의 데이터 압축기술은 화상 회의에 요구되는 데이터 대역폭을 대폭 축소한 것이 특징이며 사람 얼굴의 전체 화소를 스트리밍하는 대신, 포인트만을 뽑아내어 AI 알고리즘을 통해 얼굴 화면을 다시 생성

[Nvidia Maxine의 데이터 압축기술(AI Compression)]



* 출처: NVIDIA

◎ 의료

- (IBM) 의료 전문 AI 플랫폼인 ‘Watson Health’를 운영하며 다수의 헬스케어 서비스 기업들을 참여시켜 의료 분야 산업 생태계를 구성
 - 방대한 정보를 조합하여 지식을 만들어 내는 Watson의 핵심 기술을 활용하여 의료, 금융 분야에 특화된 AI(AI) 의사결정 지원 기술을 발전시키는 중
 - 자사의 기술 및 제품을 플랫폼화하고 데이터를 과독점하는 등 생태계의 과·독점 현상이 가속화

- (Alibaba) 의료 AI 관련으로 고난이도 영역에 속해 있던 심혈관 인식기술을 개발
 - 심혈관 인식 기술은 관상동맥을 자동 판별하고 추적하는 기술로 입체회선신경망(3DNN)을 이용해 영상안을 반복탐색하고 혈관 전체를 검출
 - 관상동맥 1개를 추출하는데 0.5초가 걸리고, 관상동맥 전체는 20초 이내여서 기존의 100배에 가까운 효율을 실현

- (핑안하오이성) 중국 의료 플랫폼 기업으로 2018년 11월 ‘제 5회 세계 인터넷 대회’에서 AI과 빅데이터를 응용한 무인 진료소를 개시
 - 무인진료소에서는 AI 의사가 환자의 목소리나 이미지를 판단해 초기 진단을 진행
 - 이후 클라우드 컴퓨터 닥터가 환자와 간단한 대화를 나눈 후, 실제 전문 의료진과 환자를 연결시켜 질병 및 건강 검진에 대한 설명을 제공

- (딥마인드) 영국의 무어필드(Moorfields) 안과병원, 구글 헬스케어와 함께 AI를 활용한 습성 황반변성(exAMD) 예측 연구를 수행²⁾
 - (학습 데이터) 무어필드 병원은 한 쪽 눈이 exAMD이고, 다른 한 쪽 눈이 exAMD로 발전할 수 있는 가능성이 높은 2,795명의 환자에 대한 3차원 공간섭단층촬영(OCT) 이미지를 활용하며, 환자의 성별, 연령대, 인종의 정보를 포함
 - (학습 모델) 두 가지 합성곱 신경망을 활용하는데 첫 번째 신경망은 망막 OCT 이미지를 학습하며, 두 번째 신경망은 OCT 이미지를 13종의 세포로 분할하여 학습하고 최종적으로 입력 망막 이미지가 6개월 이내 exAMD로 발전될 위험을 산출
 - (결과) AI를 활용한 exAMD 예측 모델의 성능을 측정하기 위해 10년 이상 숙련된 안과의사 3명과 검안사 3명이 exAMD 예측한 결과와 비교했는데, AI의 exAMD 예측 성능은 6명의 전문가 중 5명의 예측 결과를 상회

- (Artery) 축적된 심장 MRI 사진이나 간 사진을 바탕으로 진단을 지원하는 서비스 ‘Cardio AI’와 ‘LIVER AI’ 개발 및 의학 연구 분야에서는 IBM의 AI(인공지능)에 의한 ALS 치료 및 심장혈관 질환의 유형 예측과 판정에까지 활용

2) AIB-015호, ‘SPRi AI BRIEF 인공지능 최신 동향과 시사점’, 2020 SPRi 소프트웨어정책연구소

◎ 의료 외 산업 분야 응용

□ 제조, 자동차 관련 산업과의 지능형 융합으로 차별적 부가가치 창출은 물론 기존 산업 분야의 획기적 변화를 촉발

[기존 산업 분야에서의 AI 적용 사례]

산업분야	발전 전망	적용 사례
자동차	부분 자율 주행 ⇒ 완전 자율 주행	테슬라의 경우 자율주행 3단계가 가능한 오토 파일럿을 '16년에 구현 벤츠의 경우 자율 주행 5단계AV를 '21년에 상용화
제조	공장 내 자동화 ⇒ 공장 간 연계 자동화	아디다스의 경우, 신발 주문 및 제작까지 1일 소요되는 서비스 제공 보쉬의 경우, 전세계 250개 공장을 연결
에너지	건물 에너지 효율화 ⇒ 국가 전력망 효율화	딥마인드는 Google 데이터 센터 냉각 효율을 '16년에 40%개선 GE의 경우 전력망의 전력 배분 효율성을 재고

* 출처: AI AI R&D 전략, 2018, 과학기술정보통신부

[응용 산업 분야에서의 AI 적용 사례]

응용 산업 분야	연구소 또는 기업	적용 사례	이미지
금융	켄소	10분 내로 금융 솔루션 제공 '18년 3월에 금융 서비스 기업인 S&P글로벌 인수	
농업	블루리버테크놀로지 ('존 디어'에 인수)	잡초 제거 상추로봇을 생산하여 미국 상추 생산의 10% 정도 적용	
법률	로스인텔리전스	파산 특화 판례 및 법률 분석	
물리	미국 페르마 연구소	중성자 물리 연구 분석에 AI를 적용함으로써 분석 시간이 수개월에서 하루로 단축됨	
천문학	NASA, Google	외계행성 탐색에 AI 기술 적용하여 제 2의 태양계 발견	
신소재	TOYOTA	연료 전지 촉매제 개발에 AI 기술을 적용하여 신차의 이산화탄소 배출량 감축	
제약(신약)	Atomwise	에볼라 치료약 후보 발굴에 AI 기술 적용함으로써 소요 기간이 수년에서 하루로 단축	

* 출처: AI AI R&D 전략, 2018, 과학기술정보통신부

- 미국에서는 Smart Agriculture라 불리는 농업에 대한 AI(인공지능) 기술 개발이 활발하게 이루어짐
 - Blue River Technology에서는 카메라를 사용하여 잡초에만 정확하게 제초제를 분사하는 기술이 개발
 - 드론으로 생육 상황을 확인하거나 토양이나 수질을 분석하는 기술이 개발

◎ 챗봇

- (Facebook) 챗봇(Chatbot) 기능을 도입한 기업용 메신저를 통해 별도의 앱을 실행하지 않고 채팅창에서 정보를 확인하고 처리할 수 있는 O2O 등 다양한 기능을 연결하는 플랫폼으로 활용
 - 대화형 AI 플랫폼인 챗봇(Chatbot)은 사용자의 상황과 선호도를 정교하게 분석하여 정보 검색, 쇼핑, 예약 등의 서비스에서 최적의 맞춤형 서비스를 제공
 - 현재, 기업용 메신저에 도입된 챗봇은 입력된 알고리즘에 따라 정해진 답변을 제공하는 수준이지만, 머신러닝을 통해 자연어 분석 및 처리 능력이 빠르게 향상되고 있어 AI가 다양한 변수를 고려한 최적의 대안을 제시하는 수준으로 발전할 것이라고 예측
 - Google 대비 공공 데이터 영역에서의 Facebook의 상대적인 경쟁력은 낮지만, 개별 사용자의 특성 및 성향을 분석할 수 있는 AI 플랫폼으로 가치가 높음
 - Facebook은 자사 10년 로드맵을 통해 AI를 연결성, VR/AR과함께 10년 동안 개발할 핵심 기술로 언급
 - 인포마크는 AI 챗봇 기술을 활용한 어린이 영어 스피킹 학습 플랫폼을 연내 발매할 계획. 이 플랫폼은 기존의 보고 듣는 일방향 영어 교육이 아닌, 디스플레이가 탑재된 시스피커를 이용한 대화형 영어 학습으로 진행돼 교육 효과 향상이 기대

◎ 오픈소스 활용 AI 프로젝트의 증대³⁾

- (마젠타 프로젝트) 미술과 음악을 만드는 단계에서 AI 기술을 적용하는 연구 프로젝트
- (패스트닷에이아이(fast.ai) 프로젝트) 최신 베스트 사례의 딥러닝 연구를 기반으로 최신 베스트 사례를 사용하여 빠르고 정확한 신경망 네트워크를 간단하게 트레이닝
- (디텍션(Detection) 프로젝트) 객체 검출 및 인식을 위한 플랫폼
- (패스트닷텍스트(fast.text) 프로젝트) 150개 이상의 사전에 학습된 워드벡터 모델이 제공되며 이러한 워드벡터는 텍스트 분류와 요약 등 많은 작업에 사용 가능
- (오토케라스(Autokeras) 프로젝트) 주어진 데이터셋을 바탕으로 해당 데이터 셋에 최적화된 아키텍처 및 하이퍼파라미터 튜닝 가능
- (Vid2Vid(Video to Video) 프로젝트) 다양한 데이터 셋에 대해서 고해상도의 시간적으로도 일관된 결과 비디오를 도출 가능

◎ Microsoft는 미래형 AI를 위한 신형 슈퍼컴퓨터 발표⁴⁾

- 마이크로소프트社(MS)는 자사 클라우드(Azure)에서 대형 AI 모델을 학습시킬 수 있는 세계 Top5 수준의 슈퍼컴퓨터 개발 발표(2020.5.19.)
 - 이 슈퍼 컴퓨터는 OpenAI와 공동으로 개발되었으며, MS Azure에서 호스팅(Hosting)
- MS의 신형 슈퍼컴퓨터는 다양하고 대형화된 멀티태스킹 AI 모델을 학습할 수 있도록 설계되어 고지능, 고성능 및 높은 확장성을 갖는 서비스 제공이 가능
 - (고지능) 신형 시스템은 언어 번역, 객체 인식, 음성 인식 등 서로 다른 전통적 기계 학습 영역 모두를 복합적으로 학습하여 결과 도출
 - 긴 연설의 요약, 라이브 게임 채팅의 콘텐츠 조정, 수천 개의 법률 문서에서 관련 구절 찾기, GitHub를 탐색하여 코드 생성 등 복합적인 기능을 수행하는 것이 가능
 - (고성능) 대규모 AI 모델 학습에 필요한 고급 슈퍼 컴퓨팅 인프라로서 고 대역폭 네트워크로 연결된 최첨단 하드웨어 클러스터 기반 시스템
 - 신형 슈퍼컴퓨터는 285,000개 이상의 CPU 코어, 10,000개의 GPU 및 초당 400기가비트(Gbps)의 네트워크 연결을 갖춘 단일 시스템으로 세계 5위 내 성능 확보
 - (확장성) MS-Azure와 연동하여 개발자, 데이터과학자, 기업 등이 대규모 AI 모델을 쉽게 활용할 수 있는 확장성을 지원

3) 주요 AI 오픈 플랫폼(금융보안원, 2019.01.10.)

4) SPRi 소프트웨어정책연구소, AIB-015호(2020.06.09.) 'SPRi AI BRIEF 인공지능 최신 동향과 시사점'

나. 국내 기술 동향

- 인공지능 분야의 기술경쟁력 평가 결과, 최고기술국은 미국으로 나타났으며 우리나라의 경우는 최고기술국 대비 82.5% 수준으로 나타났고 중소기업은 75% 수준으로 평가되었음
 - 최고기술국 대비 우리나라의 기술격차는 2년으로 평가되었으며 중소기업의 경우는 3년으로 평가되었음

(1) 기술 동향

◎ 지능형 비서

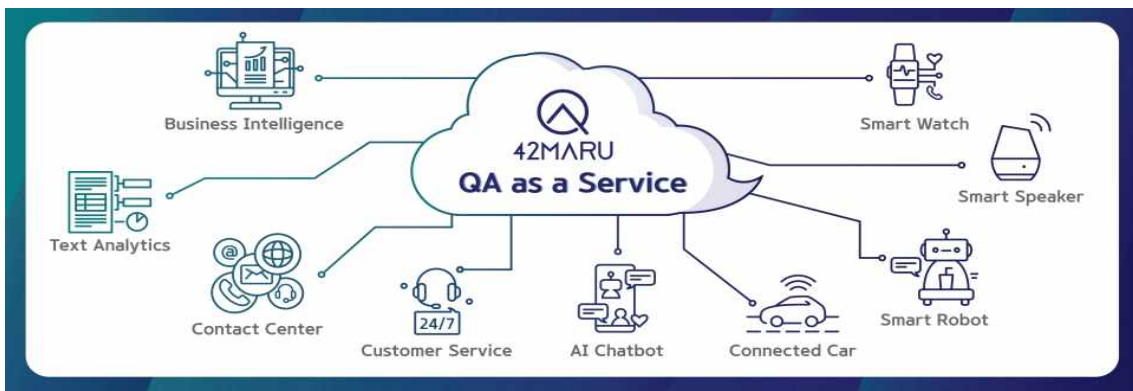
- (SK텔레콤) 전화 플랫폼 'T전화'와 인공지능 음성비서 '누구(NUGU)'가 결합된 AI 지능형 전화서비스 'T전화x누구'를 출시
 - AI가 핸드폰 이용 패턴, 위치, 날씨, 시간 같은 이용자 데이터를 분석해 맞춤형 콘텐츠를 추천하고 말로 설명
 - 음성만으로 문자메시지를 보내는 기능을 지원하며 통화 녹음을 텍스트로 옮기고 검색 및 요약 기능 지원
 - 2022년 상반기엔 모임 종류와 규모, 예산 등을 종합적으로 판단해 AI 비서가 레스토랑을 추천하고 예약, 주문, 결제를 하는 시스템도 구축할 예정
- (삼성전자) 구글 클라우드의 맞춤형 머신러닝 솔루션 클라우드 TPU(Tensor Processing Unit)를 도입해 지능형 비서인 빅스비 AI 모델의 음성 인식 학습을 고도화
 - 클라우드 TPU는 구글 클라우드에서 인공지능 서비스를 사용해 머신러닝 모델을 실행하도록 지원하는 맞춤형 머신러닝 주문형반도체(ASIC)로 빅스비의 해외 언어 서비스 확장으로 급증한 음성인식 데이터를 빠르게 처리함으로써 더욱 효율적인 모델 학습 프로세스를 구축
 - 또한 TPU는 단일 팟에서 최대 2048개의 연산 노드까지 확장해 E2E ASR 엔진의 학습에 사용되는 머신러닝 작업 부하를 효율적으로 분산
 - 기존 모델이 약 180시간에 걸쳐 처리했던 학습량을 10시간만에 처리하며 속도를 18배 향상
- (네이버) 수년간 포털 운영을 통해 축적한 빅데이터를 기반으로 범용적으로 활용될 수 있는 지능형 개인비서를 개발하고 글로벌 기업들과 맞대결
 - AI 플랫폼 '클로바'는 뇌역할을 하는 클로바 브레인과의 인간의 오감 정보를 받아들이는 클로바 인터페이스로 구성되어 있음(2017년 5월 공개)
 - 네이버 앱의 음성검색과 안면인식 카메라 앱인 '스노우'를 통해 시각 및 청각 정보 수집·분석 기술력을 확보하였으며, 자연어 처리 및 대화는 '네이버-i', 기계학습 기반 번역은 파파고, 검색·추천은 AiRS 등 개별 서비스로도 제공
 - 음성인식, 이미지 인식, 대화형 엔진 등이 통합되어 있으며 스마트폰 앱 뿐 아니라 다양한 기기와 연계 가능

- (솔트룩스) 사용자들이 스스로 개개인의 취향과 성격을 반영해 특화시키고 진화시켜 나갈 수 있는 AI 기반의 가상 아바타 서비스 기술인 ‘에바(EVA)’ 개발
 - 사용자 개인의 성격과 평소 요청하는 대화 내용에 기반한 학습을 통해 본인의 취향과 기호를 중심으로 장점을 살려 나가며 진화하며 기존의 AI 비서처럼 활용할 수 있는 것은 물론 다른 사용자의 AI 아바타와 소셜네트워크(SNS)방식으로 커뮤니케이션함으로써 지식을 습득하고 축적가능
 - 2021년까지 모든 단말기 상에서 개인은 물론 기업을 대상으로 서비스를 상용화하여 확대해 나갈 계획

◎ 음성/이미지 인식 및 챗봇

- (네이버) AI 스마트 조명 ‘글로벌 램프’는 딥러닝 기술을 활용한 광학문자판독(OCR)과 음성·이미지 같은 다양한 AI 기술이 융합
 - 램프 아래 책을 펼쳐 놓으면 AI가 글자를 인식하고 읽어주며 여성 목소리인 AI는 텍스트를 바탕으로 기쁨과 슬픔 등 감정도 표현
- (브리지텍) AI 콜센터 제품인 ‘아테나’ 출시를 통해 AI 상담사, AI 상담 도우미, 개인비서 등의 서비스를 제공, 평창동계올림픽 AI 콜센터 실증사업에 참여
 - 브리지텍이 보유한 음성인식, 언어이해, 대화처리, 지식 DB 구축기술을 기반으로 한 AI 콜센터 제품
 - 실시간 음성인식 기술, 언어분석, 자연어 심층 이해 기술, 질의·응답 대화처리 기술 ,지식 DB 구축 등 음성지능과 언어지능을 구현할 수 있는 핵심기술들의 결합 서비스
- (임클라우드) AI 채팅봇 ‘에디’를 출시, 콜센터로 들어오는 고객의 채팅 질의에 자동으로 답변
 - 아임클라우드 사내 고도화된 그래픽처리장치 (GPU)를 기반으로 딥러닝과 대화 플로우, 빅데이터 및 클라우드 병렬처리 기술 등의 집약체로 기존 콜센터 고객들의 질문 패턴 중 100여개의 토픽을 파악해 응답이 가능하며, 학습과 응대 경험을 통해 패턴 분석 정밀화
- (포티투마루) 세계 최고 수준의 응답률을 자랑하는 딥 시맨틱(Deep Semantic) QA(Question-Answering) 플랫폼의 대화 인공지능 개발

[QA 플랫폼 개요도]



* 출처: 42MARU

- 'SQuAD*(스탠포드 질의응답 데이터세트) 2.0'에서 마이크로소프트, IBM, 알리바바 등 글로벌 기업들을 모두 제치고 구글과 함께 공동 1위를 차지
 - * SQuAD는 컴퓨터가 사람처럼 주어진 문서를 읽고 이해한 후 질문에 대한 정답을 찾아내는 MRC(머신 독해 이해력) 테스트

◎ 의료

- (셀바스) AI는 AI 의료녹취 솔루션 '셀비 메디보이스'를 2018년 9월 대구 파티마병원에 공급 계약을 체결하며 국내 업체 최초로 의료녹취 솔루션 상용화
 - 초기에는 영상의학과 및 병리학과 위주의 서비스 제공으로 의료녹취 솔루션을 활용할 예정이었으나, 셀바스 AI의 클라우드 기술과 의료정보시스템을 접목하여 이용하는 방향으로 적용범위 확대
- (루닛) 딥러닝 기술과 20만여 장에 이르는 양질의 엑스선 영상 데이터가 만나 개발된 AI 솔루션인 영상분석 시스템 '루닛 인사이트(Lunit INSIGHT for Chest Radiography Nodule Detection)'를 개발하여 2019년 1월부터 건강검진 시 흉부 엑스선(X-ray) 영상 판독에 활용
 - 환자의 뇌 MR(Magnetic Resonance) 영상과 심방세동유무에 대한 자료를 바탕으로 뇌경색(허혈성 뇌졸중)의 유형을 자동으로 분석하여 의료진의 뇌경색 유형 분류를 진단하는데 지원하는 소프트웨어인 Lunit INSIGHT를 출시
 - 루닛 인사이트 MMG는 AI기술을 기반으로 설계된 유방촬영영상 판독보조 소프트웨어로, 유방암 의심 부위와 의심 정도를 색상 등으로 표기해 의사의 빠르고 정확한 진단을 도움
- (뷰노) '뷰노메드 본에이지'는 국내 대형병원에서 다년간 수집된 X-ray 영상 수만 건을 학습한 AI 프로그램으로 의사를 보조할 수 있는 수준의 판독 능력을 갖추고 있으며, 임상시험에서 전문의 3명이 판독한 결과와 동등성을 입증, 2018년 11월 국제품질규격(ISO13485) 인증을 받은 데 이어 이번 유럽의 인허가인 CE 획득
 - 세종병원의 의무기록과 생체신호 등 각종 임상데이터와 자문을 기반으로 뷰노가 심전도 기반의 심혈관질환 예측 및 진단 AI 소프트웨어를 개발할 예정
 - 인간인 의사 뿐 아니라 AI 소프트웨어를 통한 분석을 통합하여 심혈관 진단의 정확도가 높아질 것으로 기대

◎ 영상

- (수아랩) AI 딥 러닝 기술을 항공보안 분야에 접목한 'X-Ray 보안검색 자동판독 솔루션'을 국내 최초로 개발, 향후 김포공항 현장 시험운영 거쳐 국내 상용화 및 해외공항 사업 진출을 추진할 계획
- (유비파이(UVify)) 2014년 설립된 국내 머신비전 기반 자율주행 무인이동체(드론) 개발업체로 부착된 카메라를 이용한 이미지 인식으로 주변의 3차원 정보를 얻는 머신비전 기술을 보유
 - 3차원 깊이 감지와 모션프로세스, 오토파일럿 시스템까지 자율주행 드론의 주요 구성요소 기술 개발과 생산이 모두 가능

◎ 교육

- (뤼이드) 교육분야(에듀테크) 스타트업으로, 데이터와 머신러닝(기계가 스스로 학습하는 것)을 기반으로 학습자 개개인에게 필요한 문제만을 제공하는 플랫폼인 산타!(Santa!)를 개발
 - 에듀테크 시장의 주요 키워드인 맞춤형학습(Adaptive Learning) 서비스 ‘산타토익’으로 국내 토익(TOEIC) 시장을 공략중이며, 문제 생산자와 수요자 사이의 복잡한 유통과정을 압축시킨 애플리케이션(app) 플랫폼에 머신러닝 알고리즘을 적용해 시간과 비용을 효율화

◎ KAIST의 AI 플래그십 프로그램

- 4차 산업혁명의 핵심적 역할을 담당하는 KAIST와 판교밸리 입주기업 간 상호 긴밀한 협력으로 우리나라 AI 산업이 한 단계 더 발전할 것으로 전망
- AI을 이용한 입력도구와 관련하여 모바일 차세대 입력도구인 가상 키보드, 근거리 레이더 센서를 이용한 제스처 인식, 재학습이 필요 없는 증분형 온라인 얼굴 인식 등의 기술 개발
- AI의 경량화를 위한 스트리밍 데이터에서 선택적 정보 저장이 가능한 메모리 기반 강화 학습 모델, 모바일 기기를 위한 강화학습 기반의 어플리케이션 인지 동적 주파수 할당 기술 개발
- AI 커뮤니케이션 기술에는 상호주도형 대화 에이전트 개발을 위한 딥러닝-온톨로지 혼용 하이브리드 대화, 시각정보 기반 질의응답 내 Out-of-Answer 예측을 위한 전이학습 및 비지도 학습, 대화 시나리오 검색, 수집, 정제, 입력 고도화, 일상 대화 품질 개선 기술 개발
- 경험기반 추천에 대하여 장기메모리 기반 요리 레시피 추천 도우미 AI, 시각장애인 & 치매환자 대상 서비스 제공을 위한 IoT-AI Robot, 심층학습 기반의 추천 알고리즘 기술 개발
- 영상 및 이미지 인식과 관련하여 3차원 공간 인식 및 모델링을 위한 3D Scene Graph, 적대적 생성망을 이용한 객체-인식 이미지-이미지 변환, 효율적인 실시간 수어 감지를 위한 two-stage 추론 시스템 등 기술개발
- 스마트제조 분야에서 신체의 이상 감지 및 생산설비의 고장진단을 위한 RRN과 CRRN 등 기술개발

4. 정책 분석

가. 해외 정책 동향

◎ 미국

[미국 AI 관련 정책 추진현황]



* 출처 : www.whitehouse.gov/ai/, IITP 재구성

- 트럼프의 행정명령(Maintaining American Leadership in Artificial Intelligence) 이후, 미 연방정부의 AI 연구개발 이니셔티브가 급속도로 강화[NSTC, '19.6]
 - 최근 AI 연구개발 이니셔티브를 지원하기 위해 총 3개 부문으로 구조화
 - (범분야 기반) 윤리적·법적·도덕적 함의, 안전·보안, 공공데이터 제반환경, 기술표준·벤치마킹, AI 연구개발인력, 민관파트너십 등 총 8개 전략 포함
 - (R&D 분야) AI 진전에 필요한 많은 연구개발 분야를 포함하여 설명
 - (응용분야) 농업, 정보통신, 국방, 교육 등에 걸쳐 영향력 있는 발전과 긍정적인 방향으로 사회에 환원
 - AI R&D 전략 연구 우선순위는 기업이 해결하기 어려운 분야에 초점을 맞추고 있으며, 미국의 주도권 유지를 위한 연구개발 투자를 강조

- 트럼프 대통령은 19년 2월 11일 'AI 분야에서 미국의 리더십 유지(Maintaining American Leadership in Artificial Intelligence)'라는 행정명령에 서명
 - 미국 AI 이니셔티브 주요 내용
 - (연구개발 투자) 연방기관들은 장기적인 관점으로 AI 연구개발 투자를 최우선 추진
 - (인프라 개방) 데이터, 모델, 컴퓨팅 리소스를 AI 연구자에게 개방, 정부 데이터법을 시행
 - (거버넌스 표준화) 인공지능 시스템 개발 지침 수립, 인공지능 시스템 기술표준 개발
 - (전문인력 확충) 펠로우십 및 연수 프로그램 운영, 국민의 AI 잠재력 향상을 위한 STEM 교육 확대
 - (국제협력) AI R&D를 장려하고 시장 창출이 가능한 국제적 환경 조성, 미국의 이익을 보호하고 경제안보를 지키기 위한 액션플랜 개발

- 미 국방부는 19년 2월 12일 인공지능 전략을 요약한 보고서(“Summary of the 2018 Department of Defense Artificial Intelligence Strategy”)를 발표
 - 핵심임무 수행을 위한 AI 적용·확산
 - 상황인식 및 의사결정 능력 향상
 - 작전 장비의 안정성 제고
 - 예측 유지관리 및 공급체계 도입
 - 능률적인 업무 프로세스 구축
 - 민관 파트너십 및 국제협력·동맹 강화
 - 글로벌 문제해결을 위한 오픈 이니셔티브 추진
 - 학계 파트너십 강화 및 AI 혁신을 위한 투자 확대
 - 미국 산업계와의 파트너십 다각화 및 강화
 - 객관적 관점 확보를 위한 국제협력 및 동맹 강화
 - 오픈소스 파트너십 참여 확대
 - 선도적 인공지능 인력 양성
 - 국가적 난제 해결을 위한 AI 어플리케이션 개발 장려
 - 기존 인력 대상 AI 교육 프로그램 개발·제공
 - 외부 AI 전문가 영입
 - 위험을 감수한 도전에 대한 보상을 명확히 하고, 실패로부터 빠르게 배울 수 있는 문화 정착
 - AI 윤리와 안전 선도
 - AI 관점에서 기존 제도 검토 및 국방 분야 AI 가이드라인 개발
 - 탄력적이고, 견고하고, 신뢰할 수 있고, 안전한 AI 연구개발 투자 확대
 - AI 기반 의사결정 확대를 위한 ‘설명가능한 AI’ 개발 지속 추진
 - 민간인 사상자 및 기타 피해 감소를 위해 AI 도입
 - AI 확산을 위한 활용기반 마련
 - 데이터, 재사용 가능한 도구, 프레임워크 및 표준, 클라우드 및 엣지 서비스를 포함하는 공통기반 구축
- 트럼프 정부는 자유 시장 경제 원리에 따른 AI 산업의 접근법을 택하고 있으며, 정부는 규제의 장벽을 없애므로써 기술이 기술 혁신을 이룰 수 있도록 제도적 지원[NSTC, '18.5]
 - AI 산업계, 학계 및 정부 대표를 AI 정상회의(The 2018 White House Summit on Artificial Intelligence)에 초청하여 AI 산업에 대한 대통령의 접근방식에 대해 소개
 - AI 산업에 대한 미국의 리더십 유지, 미국 노동자에 대한 지원, AI 공공 연구 개발 촉진, 혁신에 대한 장벽 제거 등이 주요 골자

[미국 주요 기관의 R&D 인력 연수 프로그램]

기관	프로그램명	지원 및 활동 내용
미항공우주국 (NASA)	NASA STEM Engagement	<ul style="list-style-type: none"> • 학생과 일반이 NASA의 탐험과 발견에 기여할 수 있는 기회 제공 • 학생들이 NASA 전문가, 콘텐츠 및 시설을 경험하게 함으로써 다양한 미래 STEM 인력 양성 기반 구축
국립과학재단 (NSF)	Graduate Research Fellowship Program(GRFP)	<ul style="list-style-type: none"> • 과학 엔지니어링 인력의 다양성과 경쟁력 확보를 위해 STEM 분야 우수 석박사급 학생을 대상으로 3년간의 연구 장학금 지원
국립표준기술 연구소 (NIST)	Graduate Student Measurement Science and Engineering(GMSE) Fellowship Program	<ul style="list-style-type: none"> • 박사급을 대상으로 한 측정 과학 엔지니어링 연구와 관련된 NIST 부설 연구소 랩 경험 축적 및 연구 자금 지원 • 국가물리과학컨소시엄(National Physical Science Consortium, NPSC)의 연구 지원 자금을 통해 프로그램 관리
	NRC Postdoctoral Research Associateships Program	<ul style="list-style-type: none"> • 국립학술연구위원회(NRC)와의 협력 하에 관리되는 전국 단위의 우수 포닥 프로그램을 지원하는 프로그램 • NIST의 주요 연구 도메인과 관련된 분야를 대상으로 함
	Professional Research Experience Program(PREP)	<ul style="list-style-type: none"> • STEM 분야의 학사/석사/포닥/박사급 연구원의 NIST 부설 연구소 랩 경험 축적 및 연구 자금 지원
국방부 (DoD)	National Defense Science and Engineering Graduate Fellowship	<ul style="list-style-type: none"> • 국방부 주력 STEM 분야와 관련된 박사학위 지원자를 대상으로 한 연구 자금 지원
에너지부 (DoE)	Computational Science Graduate Fellowship	<ul style="list-style-type: none"> • 고성능 컴퓨팅을 활용한 과학 및 엔지니어링 문제 해결 분야의 박사 학위 연구자 지원 • 펠로우십 이후 에너지부 소속 연구소 등 에너지 관련 연구 기관 진출

* 출처: American AI Initiative, 2019, IITP 재구성

◎ 중국

□ 2019년 정부공작보고 ['19.3]

- 인공지능을 통한 제조업 고도화 명시

□ 대학 AI 혁신 행동계획 ['18.4]

- 차세대 AI 발전 계획의 인력양성 방안 실행을 위한 교육부의 세부 계획으로, 대학 등 고등 공기관의 AI 기술 혁신 및 인재 양성을 위한 세부 방안을 제시
- 차세대 AI 발전 계획을 토대로 3단계 목표를 설정하고 학과목 개선, 전문 역량 강화 등 대학교의 7가지 역할을 제시
 - (~'20) 대학의 과학기술 혁신 시스템과 AI 발전에 적응할 수 있는 교육 시스템 구축
 - (~'25) 대학의 AI 과학기술 혁신 능력과 인력양성의 질의 최대한 개선, 국제적 성과 달성
 - (~'30) 대학은 세계최고의 AI 혁신센터 구성 및 새로운 AI 개발의 핵심 역할 수행

- ‘차세대 AI 산업 발전을 위한 3개년 계획(Three-Year Action Plan to Promote the Development of New-Generation Artificial Intelligence Industry)’을 발표 [중국 국무원, '17.12]
 - 차량, 서비스 로봇, 인증 시스템 과 같은 지능화되고 네트워크화 된 제품 개발 계획
 - 지능형 센서 및 신경망 칩을 포함한 인공지능 지원 시스템 개발의 강조
 - 지능형 제조 시스템 개발 장려
 - AI 개발 환경 향상-산업 훈련 자원 투자, 표준 테스트 및 사이버 보안 등

◎ 일본

- AI 전략 2019 ['19.3]
 - ‘AI 도입을 통해 포용성과 지속가능성이 실현되는 사회로의 변화’를 목표로 설정하고, 4대 전략목표와 7대 분야에서의 전략을 제시
 - 4대 전략목표: 1. AI 시대의 인재 육성 및 유입 유도, 2. AI 응용분야에서 세계 최고 수준의 기술력 확보를 통한 산업경쟁력 강화, 3. 다양성, 포용성, 지속가능성을 갖춘 사회를 실현하기 위한 AI 기술 체계 확립, 4. 글로벌 AI 연구·교육·사회 기반 네트워크 구축 등으로 제시
 - 7대 전략분야: 교육, 연구개발, 사회 변화, 데이터 기반 구축, 디지털 정부, 중소·벤처기업, 기업
 - 교육개혁을 통한 AI 인재 육성과 확보
 - 기본적으로 모든 고등학교 졸업생들이 과학, 수학, 데이터 과학, AI에 대한 기초적인 지식의 습득 가능
 - 사회 문제와 제품, 서비스 설계 등 다양한 분야에 AI를 활용하여 해결해 보는 체험형 학습을 통한 창조성 함양
 - AI 표준화 국제 이니셔티브를 확보하기 위해 AI R&D 사이클 구축
 - AI 핵심 연구센터를 중심으로 네트워크를 확장하여 글로벌 AI 연구 거점으로 도약
 - AI에 필요한 인프라 및 하드웨어와 보안기술에 대한 연구개발 강화
 - 디지털 정부 실현 및 활성화
- AI핵심 공통 기반으로서 데이터 기반 정비 ['19.3]
 - AI를 위한 데이터 기반 구축
 - AI를 위한 신뢰체계 구축
 - AI를 위한 네트워크 구축

인공지능 기술 전략 ['17.3]

- 일본의 Society 5.0 이니셔티브의 생산성, 건강, 이동성의 3개 영역에 우선 적용
- 산업화 로드맵 실현에 일조 목적으로 연구개발, 인재, 공공데이터 및 창업에 대한 새로운 투자 포함
- 인공지능 서비스 개발을 3단계로 구성한 산업화 로드맵으로 유명
 - 다양한 도메인 내에서 데이터 기반의 AI의 개발과 활용
 - 도메인을 넘어서 개발된 AI와 데이터의 대중적 활용
 - 다중 도메인을 연결하는 에코시스템 구축

◎ 영국

AI 공공 서비스 정책 개발 ['18.9]

- 영국정부와 WEF 4차산업혁명센터가 공동으로 시민의 편익을 위한 책임있는 AI 기술 설계 및 서비스 제공을 위한 책임있는 AI 기술 설계 및 서비스 제공을 위해 가이드라인 개발

AI 윤리규약 및 권고사항 ['18.7]

- 영국이 세계적 인공지능 선도국으로 도약하기 위한 기술개발 및 활용에 따른 윤리 가이드

산업전략 발표 ['17.11]

- 'AI-데이터 주도 경제' 혁신을 촉진하는 글로벌 거점으로 자리매김하기 위한 다양한 지원
- 인공지능위원회(AI Council) 및 인공지능국(Office of AI) 신설 예고

◎ 독일

국가산업전략 2030보고서 ['19.2]

- AI의 경우 앞선 기술력을 바탕으로 글로벌 주도권 탈환을 위한 상용화 촉진 정책 실시 예고

독일 'AI 국가전략' ['18.11]

- '연방정부 AI 전략 개료'를 토대로 '25년까지 세계 최고 AI 국가 건설을 목표

나. 국내 정책 동향

- (관계부처 합동) 정부는 2020년 7월 코로나19 사태 이후 경기 회복을 위해 마련한 국가 프로젝트 ‘한국판 뉴딜 종합계획’을 발표하며 인공지능 중요성 강조
 - 특히 5G·블록체인 등 디지털 신기술을 활용, 국민에게 맞춤형 공공서비스를 미리 알려주고 신속히 처리해주는 똑똑한 정부인 ‘지능형 정부’ 구축 추진

[AI 전략분야의 범위]

현재 상황		미래 전망	
정부서비스에 신분증 및 종이 증명서 필요 내·외부망 분리된 유선망 중심 업무환경		모바일 인증으로 Paperless 정부서비스 언제·어디서든 Smart Office 구현	
성과지표	‘20년	‘22년	‘25년
공공서비스 디지털 전환	대면업무 중심의 공공서비스	주요 서비스 중 50% 디지털 전환	주요 서비스 중 80% 이상 디지털 전환
5G 국가망	유선망 중심 업무환경	5G 기반 무선망 선도도입	금 정부청사에 5G 기반 무선망 구축
행정·공공기관 클라우드 전환	17% (정보화 H/W 22.4만대 중 3.9만대)	50%	100%

* 출처: 한국판 뉴딜 종합계획, 2020, 관계부처 합동

- 또한, 개인맞춤형 공공서비스를 신속 처리하는 지능형 정부로 혁신 및 5G 업무망·클라우드 기반 공공 스마트 업무환경 구현
- (과학기술정보통신부) 정부 역량을 결집하여 AI 시대 미래 비전과 전략을 담은 AI 국가전략으로서 ‘AI 국가전략’을 발표⁵⁾
 - 양질의 데이터 자원 확충을 위한 공공 데이터 전면 개방, 데이터 생산, 유통, 활용을 지원할 공공, 민간 데이터 지도의 연계, 민간의 AI 개발 지원을 위한 AI 허브 컴퓨팅 자원 맞춤형 지원
 - AI 반도체 핵심기술(설계·미래소자·장비 및 공정 등) 및 신개념 AI 반도체(PIM: Process in Memory) 개발을 위해 '20년부터 '29년까지 1조 96억 원 투자

5) AI(AI) 국가전략 발표 (과학기술정보통신부 보도자료, 2019년 12월 17일)

[AI 국가전략 (2019.12)]

01 세계를 선도하는 인공지능 생태계 구축

AI 인프라	AI 기술 경쟁력	과감한 규제혁신	AI 스타트업
<ul style="list-style-type: none"> 공공·민간 데이터 개방 'AI 혁신 클러스터' 확산 	<ul style="list-style-type: none"> AI 반도체 핵심기술 개발 (1조원) 	<ul style="list-style-type: none"> 포괄적 네거티브 규제(20) 	<ul style="list-style-type: none"> AI 올림픽 개최(20) AI 스타트업 투자·지원 
	<ul style="list-style-type: none"> 차세대 AI 연구 선제 투자 	<ul style="list-style-type: none"> AI 시대 대비 '법제정비단' 발족(20) 	

02 인공지능을 가장 잘 활용하는 나라

AI 인재양성 및 전 국민 교육	전 산업 AI 활용 전면화	최고의 디지털 정부								
<table border="1"> <tr> <td>고급</td> <td>AI 관련학과 신증설</td> </tr> <tr> <td>전문</td> <td>AI 교원 기업겸직 허용</td> </tr> <tr> <td>실무</td> <td>AI 융합전공 개설 직군별 AI 역량 함양</td> </tr> <tr> <td>일반</td> <td>초중등 SW 필수교육 전 국민 AI 평생교육</td> </tr> </table>	고급	AI 관련학과 신증설	전문	AI 교원 기업겸직 허용	실무	AI 융합전공 개설 직군별 AI 역량 함양	일반	초중등 SW 필수교육 전 국민 AI 평생교육	<ul style="list-style-type: none"> 대형 AI 융합 프로젝트(AI+X) 	<ul style="list-style-type: none"> 차세대 지능형정부 구현 AI 기반 공공서비스 혁신 
고급	AI 관련학과 신증설									
전문	AI 교원 기업겸직 허용									
실무	AI 융합전공 개설 직군별 AI 역량 함양									
일반	초중등 SW 필수교육 전 국민 AI 평생교육									

03 사람중심의 인공지능 구현

포용적 일자리 안전망 구축	AI 역기능 대응 및 AI 윤리 확립
 <ul style="list-style-type: none"> 국민취업 지원제도 사회보험 적용 확대 일자리정보 플랫폼 	 <ul style="list-style-type: none"> AI 기반 사이버침해대응체계 고도화 글로벌 수준의 AI 윤리체계 정립

* 출처: 과학기술정보통신부 보도자료, (2019. 12. 17.)

□ AI 허브 구축 (데이터셋·알고리즘·컴퓨팅파워의 원스톱지원) ('19년)

- 대규모 AI 데이터 축적 및 개방
 - 민간수요 중심으로 기계학습용 범용 데이터(이미지, 말뭉치, 상식 등), 산업별 특화데이터(법률, 금융 등)를 단계적 구축·개방
- 우수 AI 알고리즘 발굴·개발 지원
 - 연구자 간의 자율경쟁 기반의 도전적·창의적 문제해결의 장으로 누구나 참여 가능한 개방형 온라인 플랫폼 구축
- 컴퓨팅 자원 제공
 - AI 연구 및 AI 제품·서비스 개발에 필요한 자원을 스타트업·중소기업에게 AI개발에 특화된 형태로 지원

□ 'I-Korea 4.0 실현을 위한 AI R&D 전략' 수립 ('18. 5)

- 세계적 수준의 AI 기술력 및 R&D 생태계 확보를 위해 2018년부터 5년간 2.2조원 투자 추진
- 2022년까지 6개의 대학원을 설립하여 5,000명의 AI 전문가를 육성하고, 단기적 AI 인력의 필요에 대비하여 600명의 예비 인력 훈련 추진
 - AI대학원 신설, 대학연구센터의 AI 분야 지원 확대, 글로벌 기업과의 공동연구 지원 등을 통해 핵심·차세대 원천 기술 개발을 위한 고급 인재 양성 목적
 - AI을 활용하여 신제품·서비스를 창출하기 위한 데이터 활용 중심의 융·복합 인재 확보
- 국방, 의학 및 공공 안전분야의 대규모 프로젝트에 자금을 지원, DARPA와 유사한 AI R&D 챌린지 착수
 - 민간 투자가 어려운 공공영역과 고위험, 차세대 기술 분야 집중
 - 차세대 AI 핵심 기술 확보를 위한 'AI R&D 챌린지 대회, AI 국가전략프로젝트 및 AI HW 분야 지원을 확대하고, 신약, 미래 소재 등 'AI+X' 즉, AI과 타분야의 융합을 강화하여 대규모 혁신 촉발
- 2029년까지 AI 반도체를 만들기 위한 자금 지원과 차세대 AI 사업 지원을 위한 AI 창업 보육센터 창립
 - AI 창업 및 중소기업 등 초기시장 창출 및 발전을 위해 기반 인프라에 집중 투자

5. 중소기업 전략제품

가. R&D 추진전략

- AI는 향후 경제 사회 전반을 혁신할 원천 기술이라는 점을 고려, 자국 국민 삶의 질과 국가 경쟁력 제고를 위해 AI 원천 기술력 및 AI 응용 기술력 확보를 위한 적극적 지원 필요
 - 한국의 AI 기술력은 미국, 중국에 다소 취약하나, ICT 산업이 선진 수준에 근접하고 있어 AI를 개발 활용할 수 있는 여건 양호하며 다른 국가들과 비교 시 상대적으로 유리한 위치를 선점 가능
 - (기업) AI의 새로운 서비스 발굴 및 R&D 투자를 통한 기술력 확보 및 혁신 생태계 조성
 - (정부) 대형 프로젝트 등 진흥정책과 규제·법제도 혁신 등 산업·사회 인프라 조성
 - (학계) AI 기술 개발과 인재 양성 및 AI 시대의 방향성 제시

- 기본적인 AI 알고리즘이 대부분 오픈소스로 공개되어 있고, 많은 산업 분야가 AI 적용을 시작하는 단계이므로 전략적 접근 시 경쟁력 확보 가능
 - AI는 학습 데이터 학습을 근간으로 하므로, 특정 도메인 분야 학습 데이터를 확보·학습시킨 AI는 특정 산업 영역에서 충분한 경쟁력을 가질 수 있으며, 타 기술·산업의 혁신에도 적용 가능

- 의료영상진단, 음성 언어이해(챗봇 등) 등 AI 중소기업의 경쟁력을 높일 수 있는 분야에 연구 개발 집중
 - (의료) 예방-진단-치료-사후관리 등 의료 전주기를 사람 중심 AI 헬스케어에 적용
 - 환자케어 봇, 의사 도우미, 공중보건 위협예측, 판독 보조, 행동 건강관리, 유전자 분석, 정신건강 관리, 지능형 병원 워크 플로우, 응급 및 24시간 모니터링, 의학 설명 봇, 행동건강관리 등
 - 예를 들어, 성장기 아이의 손 X-ray 사진을 기반으로 나이별 골의 형태를 딥러닝 기반으로 학습시켜 환자의 뼈 나이를 판별하여 제시
 - (금융 분야 적용 사례) 질문과 대답의 유형을 미리 만들어 활용하는 기존 챗봇 서비스와 차별화하여, 대화의 내용을 기억, 이해할 수 있는 비대면 고객 응대서비스

- 중소 제조업 및 서비스업에 AI 접목을 통한 제품(서비스) 고부가가치화 추진
 - AI를 통한 제조공정의 데이터 분석으로 공정 최적화 추진
 - AI가 전체 공정을 실시간으로 관리하고 제어하는 스마트 공장 구축

□ 본 로드맵의 주요 범위는 AI이며 이 범위에 대한 주요 기회요인과 위협요인으로부터 중소기업의 시장대응전략을 수립

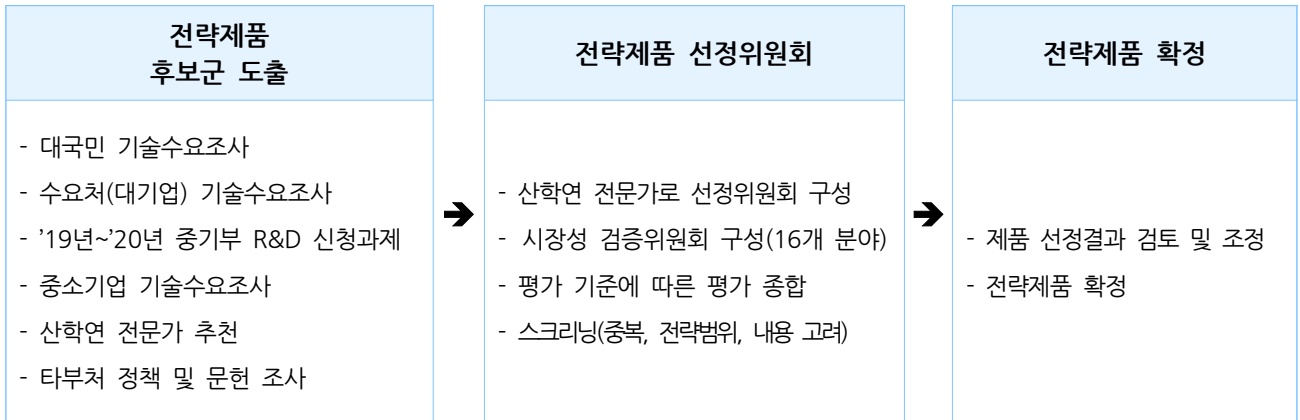
Factor	기회요인	위협요인
정책	<ul style="list-style-type: none"> • AI 산업에 대한 지원 확대 • 정부는 AI 산업 육성 정책을 수립 • 4차 산업혁명에 대한 지능정보사회 중장기 종합 대책 발표 	<ul style="list-style-type: none"> • 육성정책의 착수시점 및 투자규모 측면에서 주요국 대비 미흡 • AI 특성을 반영한 다중 참여형 및 데이터 기반 R&D 사업형태 지원, 정책 및 법률 미비 • 기존 시장 보호를 위해 과도한 신규 시장 억제정책
산업	<ul style="list-style-type: none"> • 4차 산업 혁명 시대에 선두주자가 되기 위하여 다양한 산업 분야에 AI 기술 도입 • 최근 해외 기업들의 성공사례로 자극받아 적극적인 사업 추진 경향성 증가 	<ul style="list-style-type: none"> • 상대적으로 높은 진입 장벽과 시장 확보로 인해 타 업종으로 이탈하려는 경우가 많음 • 혁신 기술 부재로 인해 제품 차별화 미흡 • AI 핵심인력 대기업 편중 및 해외 기업 선호로 인해 불균형 예상
시장	<ul style="list-style-type: none"> • AI 기술의 관심과 활용의 확산 • AI 활용 분야의 다양화 및 각 분야의 시장 규모 확대 • 해외시장 진출 기반 확대 	<ul style="list-style-type: none"> • 기술과 성능에 대한 지나치게 높은 기대감 • 선진국의 선점 효과 • 신규시장 생성시 기존 시장과의 마찰
기술	<ul style="list-style-type: none"> • AI 분야에 대한 전반적인 R&D 증가로 제품 경쟁력 상승 • AI 개발 플랫폼의 보급화로 AI 기술 개발에 탄력 	<ul style="list-style-type: none"> • AI 원천 기술에 대한 국산화 비율 낮음 • 해외 대기업과의 R&D 비용 및 기술력 격차 • 해외 기업의 AI 개발 플랫폼에 대한 높은 의존도



중소기업의 시장대응전략

- AI 원천 기술력 및 AI 응용 기술력 확보를 위한 적극적 지원 필요
- AI 중소기업의 경쟁력을 높일 수 있는 분야에 연구 개발 집중
- 미래 도래가 예상되고 성공 시 파급효과가 큰 분야를 중심으로 선도적으로 AI 활용 추진
- 중소 제조업 및 서비스업에 AI 접목을 통한 제품(서비스) 고부가가치화 추진

나. 전략제품 선정 절차



□ 전략제품 후보군 도출

- (최근 신청 과제) 중소벤처기업부 R&D 지원 사업 '19년~'20년 상반기 신청 과제
 - (기술수요조사) 중소기업기술정보진흥원 주관 SMTECH(중소기업 기술개발사업 종합관리시스템) 성과분석 대상 중소기업으로부터 기술수요 수신
 - (대기업 의견) 전략분야 관련 대기업의 중소기업 유망 제품 관련 인터뷰
 - (산학연 전문가 추천) 분야별 전문가 대상 후보 추천 의뢰 의견 수렴
 - (타부처 정책 및 문헌조사) 타 부처 정책사항 및 문헌조사를 통한 품목 발굴
- ※ (재밍, Jamming) 데이터 기반의 전략제품 발굴을 위하여 인공지능 전략분야에 시범적으로 도입

□ 전략제품 선정위원회

- (선정방식) 중소기업 적합형 기술로드맵 수립 및 전략 강화를 위해 전략제품 선정위원회의 평가와 시장성 검증위원회의 평가를 종합하고, 전략분야에 따라 평가항목의 가중치를 조절하여 반영
- (전략제품 선정평가위원회) 분야별 산·학·연 전문가 위원회를 구성하여 전략제품에 대해서 각 5개 항목을 평가 및 검토 진행
- (시장성 검증위원회) 시장성 검증이 필요한 분야에 대해서 해당 전략분야에 관련성이 높은 전문가와 VC(투자심사역)으로 구성된 위원회가 전략제품 평가 진행
- (평가항목) 시장성, 기술난이도, 개발기간, 수입의존성 및 중소기업 적합성을 기준으로 평가
- (평가기준) 전략분야의 대구분(한국판 뉴딜 및 소부장·뿌리산업)에 따라 평가항목의 가중치를 조절

□ 전략제품 확정

- (검토 및 조정) 선정된 전략제품들에 대해 최종적인 타당성 검증 및 분야 간 전략제품 검토 및 조정을 통해 전략분야별 전략제품 확정

다. 전략제품 선정결과

◎ AI기반 교육서비스

· AI 기반 교육 서비스는 교육 서비스에 인공지능 기술을 접목해 기존 서비스를 개선 및 새로운 서비스를 제공하는 차세대 교육을 의미

- AI기반 교육 서비스 산업 성장에 대한 적시적인대응이 필요
- 해외 거대 기업들의 AI 기반 교육 서비스 플랫폼, 기술 장악에 대비

◎ AutoML 솔루션

· AutoML은 기계학습(Machine learning), 딥러닝(Deep learning) 등 인공지능 관련 모델 개발에 필요한 다양한 과정을 자동화하여 최적 성능의 모델을 개발하는, 일명 '인공지능을 만드는 인공지능' 솔루션

- 인공지능 모델 개발을 위한 학습용 데이터 자동 설계 필요

◎ 합성 데이터 (Synthetic Data)

· 직접 획득한 데이터가 아닌, 인공적 데이터를 합성하는 기술.

- 많은 산업에서 데이터 보안 프라이버시, 규제이슈로 데이터 소유자가 보유한 원본 데이터를 직접 외부에 반출하여 사용하기 어려움

◎ eXplainable AI (XAI)

· 설명 가능한 AI(eXplainableAI, AXI)는 사용자가 인공지능 시스템의 동작과 최종 결과를 이행하고 올바르게 해석하여 결과물이 생성되는 과정을 설명이 가능하도록 하는 기술을 의미함

- 인공지능 기술(머신러닝등)은 빅데이터 및 복잡한 알고리즘 등을 기반으로 사용자에게 의사결정, 추천, 예측 등의 정보를 제공함, 그러나 딥러닝 등은 알고리즘의 복잡성에 인해 '블랙박스'로 인식, 결과에 대한 충분한 설명이 부재

◎ 딥러닝 영상처리 기술을 활용한 의료진단 솔루션

· 딥러닝 영상처리 기술(Object Detection, Semantic Segmentation)을 활용해 질병을 진단하는 의료진단 플랫폼으로, 암 및 종양 진단, 피부병 진단, 시각 진단(안과) 등 여러 질병 중 몇 가지 질병을 의사 대신 진단해 환자에게 더욱 객관적인 진단 정보를 제공하는 솔루션

- 딥러닝 영상처리 기술의 영향력이 뇌, 골밀도, 조직 세포, 심장 등 거의 모든 의료 영상 판독에 미칠 수 있다는 점에서 엄청난 잠재력 보유

◎ Robotics Process Automation(RPA) System

· RPA(Robotic Process Automation)는 인간이 수행하는 일반적 업무 중 표준화되어 있고 규칙 기반 반복적 업무를 컴퓨터가 자동적으로 할 수 있도록 전환시켜주는 프로세스를 의미

- RPA는 현재 산업에서 사용 혹은 개발 중인 다른 기술에 비해 검증되고 산업 내 일정 수준 상용화된 기술

◎ 영상데이터 기반 AI 서비스

· 영상 데이터 기반 AI 서비스는 영상과 이미지에 존재하는 객체의 종류와 특징들을 추출하고, 추출한 특징들로부터 유의미한 정보를 산출하는 AI 기술들을 활용하여 의료, 보안, 불량 검출, 범죄 인지, 상황인지 등의 서비스의 인식률과 정확도를 높이는 AI 서비스

- 위험도가 높거나 자동화가 필요한 분야에서 영상 분석을 통한 AI 서비스를 실현하여 생산 효율성 증대, 객체 인식 정확도 개선, 인간의 반복 작업 탈피 등의 노동 대체 기술이 필요

◎ 인간-인공지능 협업 시스템

· 협업지능을 기반으로, AI가 스스로 사용자를 인식, 사용자에게 맞는 서비스를 제공하는 기술. 인간의 목표를 이해하고 인간의 의도에 대응되는 대처를 할 수 있도록 인간과 AI 간의 커뮤니케이션을 매개하는 시스템

- 인간-AI 협업 시스템은 사용자 패턴 분석을 통한 사용자의 취향 또는 선호도 맞춤형 지능적 시스템 조종 기술, AI 플랫폼, 패턴 분석 알고리즘, 빅데이터 SW 등의 기술을 활용해 자율주행 자동차, AR 시스템, VR 시스템, 교육 및 정보 등의 다양한 분야에서 활용이 가능



전략제품 현황분석

AI 기반 교육서비스



AI 기반 교육 서비스

정의 및 범위

- 'AI 기반 교육 서비스'는 교육 서비스에 인공지능(AI) 기술을 접목해 기존 서비스를 개선하거나 새로운 서비스를 제공하는 차세대 교육을 의미
- 'AI 기반 교육 서비스'는 에듀테크(EduTech) 범주 안에 속하며 데이터와 AI를 활용하여 제공하는 교육 서비스와 관련 AI 기반 기술 등을 포함

전략 제품 관련 동향

시장 현황 및 전망	제품 산업 특징
<ul style="list-style-type: none"> • (세계) '18년 647억 달러 규모에서 연평균 18.1%로 성장하여 '24년 1,755억 달러 규모로 성장 전망 • (국내) '18년 3조 8,400만 원 규모에서 연평균 18.8%로 성장하여 '24년 10조 8,000만 원 규모로 성장 전망 	<ul style="list-style-type: none"> • 교육 분야의 AI 도입률은 현재 다른 분야에 비해 상대적으로 낮은편으로 확장 가능성이 많은 분야 • 교육과 학습 관련 SW·HW, 플랫폼, 기술 등은 락인(Lock-in) 효과가 높은 만큼 국내 기업의 서비스가 활용될 수 있도록 선제적으로 기술개발 등의 대비를 한다면 시장 선점에 유리할 것으로 기대
정책 동향	기술 동향
<ul style="list-style-type: none"> • 미국, 중국, 일본 등 주요 국가는 ICT 기반 교육 혁신 정책 수립을 추진 중 • 코로나19로 인해 원격수업이 확산하며 비대면 교육 뿐만 아니라 AI 기반 교육 서비스 시장전체가 확대될 전망으로 보이며 관련 정책적 지원 필요 	<ul style="list-style-type: none"> • 국내 주요 교육 기업들은 에듀테크에 전폭 투자하고 있는 것으로 나타났으며, 특히 전문 인력양성과 기술 개발을 위한 연구개발비(R&D)에 집중적 투자 중 • 에듀테크 기술은 크게 교육의 대중화, 효과성 극대화, 일상과 학습의 결합이라는 세 가지 방향으로 발전 중
핵심 플레이어	핵심기술
<ul style="list-style-type: none"> • (해외) ByJu's, VIPKid, Elemental path, HUBERT.AI • (대기업) 대교, 웅진씽크빅, 교원, 천재교육 • (중소기업) 루이드, 노리, 맵스프레스, 캐치잇플레이 	<ul style="list-style-type: none"> • 학습자 패턴 분석 및 평가 기술 • 학습 데이터베이스 구축·분석·운영 기술 • 맞춤형 학습 콘텐츠 추천 기술 • 학습자 행동 측정 및 모델링 • 문제 컨텍스트 추론 기술

중소기업 기술개발 전략

- 맞춤형 AI 기반 교육 서비스를 제공하기 위한 기반 기술 개발 (실시간 사용자 분석, 문제 분석, 개별 맞춤형 문제 콘텐츠를 제공할 수 있는 기술 개발 및 공개)
- 학교 교육 등 공교육 부문 복합적 기술 인프라 개발 (디지털 디바이스 개발 및 보급 + 학교 시스템과 연계할 수 있는 클라우드 시스템 + 누적된 데이터를 분석할 수 있는 AI 모델 및 알고리즘)
- 포괄적인 AI 교육 서비스를 제공할 수 있는 기술 개발 (특정 과목이 아닌 학습자의 전체적인 학습 과정에서 지속적인 평가와 피드백을 통해 학습 목표 성취를 돕는 'AI 튜터' 등)

1. 개요

가. 정의 및 필요성

(1) 정의

- ‘AI 기반 교육 서비스’는 교육 서비스에 인공지능(AI) 기술을 접목해 기존 서비스를 개선하거나 새로운 서비스를 제공하는 차세대 교육을 의미
 - AI 기반 교육 서비스는 ‘에듀테크(EduTech)’ 범주 안에 속하며 에듀테크 중 데이터와 AI를 활용하여 제공하는 교육 서비스와 관련 AI 기반 기술을 등을 포함
 - 에듀테크는 교육(Education)과 기술(Technology)의 합성어로, 교육에 정보통신기술(ICT)을 접목해 기존 서비스를 개선하거나 새로운 서비스를 제공하는 것, 또는 교육 서비스를 개선하거나 새로운 가치를 제공하는데 활용되는 기술을 의미
 - 최근에는 가상현실(VR), 증강현실(AR), 인공지능, 빅데이터, 소프트웨어, 디스플레이, 3D 기술 등 ICT 전 분야를 접목
 - AI 기반 교육 서비스는 AI를 활용해 학생 개인별 학습 상태를 분석한 데이터를 바탕으로 교육을 진행하고, 개인 성취도를 고려한 맞춤형 학습이 가능한 교육 서비스를 의미
 - 학생들에게는 데이터를 활용하여 개인의 특성을 파악한 후 맞춤형 교육을 제공하고 선생님에게는 교수법 및 학사행정 관리를 자동화 및 지능화하는 추세
- ※ (참고) 에듀테크와 유사한 개념으로 이러닝(E-Learning), 스마트 러닝(Smart Learning) 등이 있음
 - 이러닝은 1990년대 말 인터넷의 급속한 확산으로 등장한 개념으로 주로 디지털교과서와 온라인 학습에 방점을 두고 있으며, 2010년경에는 스마트폰과 태블릿이 본격적으로 확산됨에 따라 스마트 기기를 통해 교육 성과를 높이기 위한 스마트러닝이 업계 및 학계의 주목
 - 이러닝과 스마트러닝의 경우 교육이 행해지는 학습 수단(인터넷+데스크탑, 스마트폰, 태블릿 등)에 주요한 개념의 방점
 - 에듀테크의 경우 인공지능, 빅데이터, 블록체인 등 데이터와 소프트웨어를 기반으로 학습자에 대한 분석과 의사소통, 정보관리를 쉽게 함으로써 학습 성과를 제고시키는 방향에 좀 더 무게 중심

[에듀테크의 개념과 유사 개념 비교]

구분	개념	특징
이러닝	전자적 수단, 정보통신 및 전파·방송 기술을 활용하여 이루어지는 학습	인터넷과 컴퓨터에 교육을 접목한 온라인 교육 중심
스마트러닝	스마트폰, 태블릿PC, E-Book 단말기 등 스마트 디바이스와 이러닝 신기술이 융합된 개념	스마트 기기를 활용한 교육
에듀테크	교육에 ICT 기술을 접목해 기존 서비스를 개선하거나 새로운 서비스를 제공하는 것	데이터와 소프트웨어에 무게 중심

* 출처: 에듀테크 산업 동향 및 시사점, SPRI, 2020.4

(2) 필요성

- AI 기반 교육 서비스 산업 성장에 대한 적시적 대응 필요
 - 세계 교육 시장이 계속 확대되는 가운데, AI 기반 교육 서비스를 아우르는 에듀테크 분야는 향후 급속하게 성장할 것으로 예상되는 매력적인 분야
 - 기존 교육을 실감화, 연결화, 융합화, 지능화의 방향으로 이끌고 있는 에듀테크는 타 산업의 경쟁력을 높이는 메타산업으로 역할
 - AI 기반 교육 서비스 산업 성장을 통해 인공지능 및 관련 기술의 발전을 촉진하고 이들 기술이 다시 교육산업과 융합되어 국가 경쟁력을 갖출 수 있도록 적시적 대응과 투자가 필요한 상황

- 해외 거대 기업들의 AI 기반 교육 서비스 플랫폼, 기술 장악에 대비
 - 미국(The National Education Technology Plan, 2017), 중국(교육정보화 2.0 행동계획, 2018), 일본(ICT화를 위한 교육환경 정비 5개년 계획, 2018) 등 주요 국가는 ICT 기반 교육 혁신 정책을 수립 추진 중
 - 구글, 마이크로소프트, 애플 등 글로벌 ICT 기업들도 교육 관련 SW를 학교와 학생들에게 무료로 제공하는 등 에듀테크 플랫폼과 기술 장악을 위해 노력 중
 - 교육과 학습 관련 HW, SW, 플랫폼, 기술 등은 락인(Lock-in)효과가 높은 만큼 교육 현장에서 국내 기업의 서비스가 활용될 수 있도록 선제적으로 기술개발 등의 대비를 한다면 시장 선점에 유리할 것으로 기대

- AI 기반 교육 서비스 활성화를 통해 교육 분야 변화에 대응
 - 우리나라 교육 분야는 공공성 추구라는 정책 기조로 인해 우수한 민간 기술과 도전적 자본이 진입하기 어려운 환경
 - 교육에 대한 관점을 성장산업이자 메타산업, 혁신의 촉진자로서의 역할에 초점을 맞춰 인식을 전환할 필요
 - 평생교육, 직업교육, 사교육뿐만 아니라 공교육에서도 AI 등 ICT 기술이 사용되고 관련 기업이 참여한다면 교육산업이 더욱 활성화될 것으로 기대

- 포스트 코로나, 비대면 교육 활성화에 대비
 - 2020년 초 발생한 코로나19 팬데믹으로 서서히 확대되고 있던 디지털 전환이 급격하게 삶과 일상으로 확대되기 시작했으며, 비대면이 보조적인 수단에서 필수적 수단으로 변한 상황에서 일상에서 개인의 디지털 기술 기반의 서비스 활용 가속화 추세
 - 원격교육이 확산하며 실시간 소통의 어려움, 매크로를 활용한 가짜 출결, 학습 부진 등의 문제가 발생하고 있으며, 이러한 사회적 문제를 해결하고 위기를 기회로 삼아 AI 기술발전, AI 기반 교육 서비스 시장을 확대해 나갈 필요

나. 범위 및 분류

(1) 가치사슬

- AI 기반 교육 서비스의 경우 학생을 포함한 전 연령대의 맞춤형 교육, 온·오프라인 서비스를 제공을 통한 교육산업 활성화를 기대할 수 있으며, AI, 데이터, 클라우드 등 교육 서비스 개발을 위한 기술 연구, 관련 산업 활성화 기대
 - (후방산업) ① 맞춤형 개별 학습 지원, 학생에 대한 예측 및 분석을 위한 인공지능(AI) 기술 및 소프트웨어(SW) 개발 ② 학습 수준 진단·개선, 맞춤형 학습 가이드를 위한 데이터 분석(Data Analysis), ③ 가상 실험 실감형 학습 콘텐츠 활용을 위한 증강/가상현실(AR/VR), ④ 자원 공유·관리, 협업을 통한 학습 활동을 위한 클라우드 컴퓨팅(Cloud Computing) 관련 산업
 - (AI 기반 교육 서비스 관련 산업) ① 맞춤형 학습 서비스(데이터 분석부터 개별 맞춤형 콘텐츠 제공), ② 게임 기반 학습, ③ 통찰력·기억력·이해력·결단력 등 행동 수정을 위한 인지 교육, ④ 모바일 기기를 사용한 모바일 교육, ④ 로봇을 이용한 교사 등으로 구분 가능
 - (전방산업) ① 교육 서비스 플랫폼(학생 관리부터 맞춤형 교육 관리), ② 교육 서비스 로봇, ③ 교육용 게임, ④ 데이터 기반의 개인 성장 분석 등의 전방 산업이 활성화 가능

[AI 기반 교육 분야 산업구조]

후방산업	AI 기반 교육 서비스	전방산업
인공지능 기술 및 소프트웨어 개발, 데이터 분석, 증강/가상현실, 클라우드 컴퓨팅 등	맞춤형 학습, 게임 기반 학습, 인지 교육, 모바일 교육, 로봇 교사	교육 서비스 플랫폼, 교육 서비스 로봇, 교육용 게임, 데이터 기반의 개인 성장 분석 등

(2) 용도별 분류

- AI 기반 교육 서비스는 크게 학생용과 교원용으로 구분 가능
 - (학생용) 학생들의 문제 풀이 데이터 등 학습 데이터를 AI가 모니터링해 분석·진단해주고 맞춤형 학습 방식을 설계해주는 방식
 - 문제 풀이 데이터를 통해 과목별 학생의 현 수준 진단
 - 현재의 교육 평가 시스템에서 측정하기 어려운 역량을 평가하는 서비스 제공
 - 학생 개별 맞춤형 학습 경로 및 적정 콘텐츠 제시
 - 챗봇 등 AI 도우미가 학생들이 학습장애 물에 봉착할 때 문제 해결 지원
 - (교원용) 교육을 위한 교사의 보조도구로 학생들의 데이터에 따라서 맞춤형 교육 방법을 분석·제공해주거나 현장 교육을 보조하는 용도로 활용
 - 학생들의 개별 데이터를 분석, 맞춤형 지도 방향에 대한 조언 제공
 - 학습 커리큘럼 구성 및 최신 교육 콘텐츠 분석 도구로 활용
 - 교원과 학생 구성비 불균형에 따른 교육 공백을 극복하는 용도로 부진 학생 등이 지속해서 수업에 참여할 수 있도록 개별 보조해주는 역할로 활용
- 또한, AI 활용 형태에 따라 개인 맞춤형 교육 제공, 교수 및 학사행정 자동화, 교육의 시공간적 확대로도 분류 가능

[인공지능 기반 에듀테크 기업 및 서비스 동향]

AI 활용 형태	주요 내용
개인 맞춤형 교육 제공	<ul style="list-style-type: none"> • 학생 개인의 학습 상태, 성취도를 고려한 맞춤형 학습 • 인공지능 튜터 활용 • 학습 몰입도를 측정하여 최적화된 환경 제공
교수 및 학사행정 자동화 지능화 지원	<ul style="list-style-type: none"> • 채점, 첨삭, 피드백 등의 교수 활동 자동화 • 학습 이력, 학습 활동, 학습 경로 설정 등 행정 지능화
교육의 시공간적 확대	<ul style="list-style-type: none"> • 시간적, 공간적 제약 극복 • 교사-학습자-멘토-전문가 등과의 연결 및 협업

* 출처: NIPA, 2019.12

2. 산업 및 시장 분석

가. 산업 분석

◎ 에듀테크 및 AI 기반 교육 서비스 시장은 급격히 확대되는 중

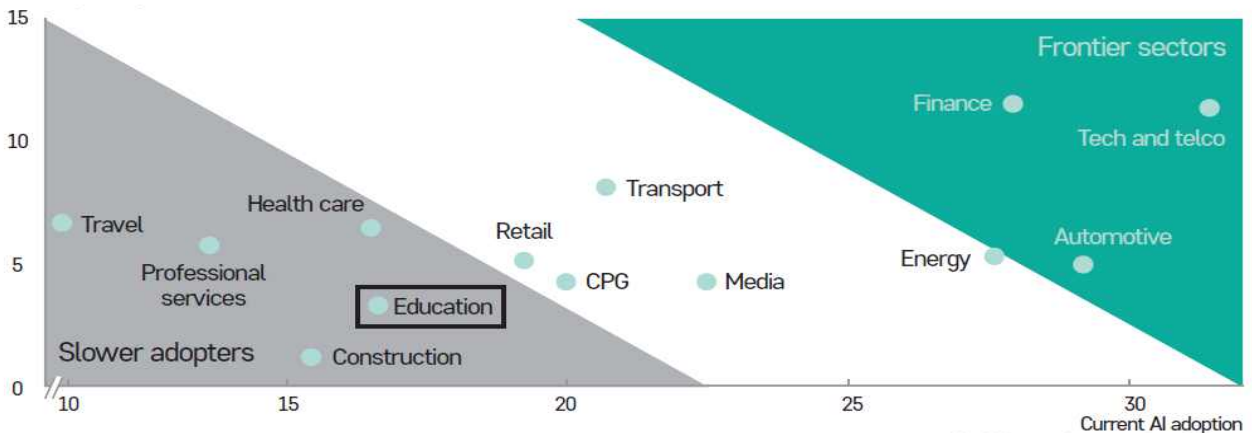
□ 에듀테크 시장이 확대되며 기존 교육 기업 및 신규 디지털 기반 교육 스타트업의 치열한 경쟁이 예상

- 구글이 선정한 세계 최고의 미래학자 토머스 프레이(Thomas Frey) 다빈치 연구소장은 “2030년 지구상에서 가장 큰 인터넷기업은 교육 관련 기업이 될 것”이라며 에듀테크 시장의 비약적인 성장을 예고
- 에듀테크가 교육업계의 새로운 블루오션으로 떠오르면서 신(新)시장을 선점하기 위한 경쟁이 치열하게 나타나는 중이며, 아직 에듀테크라는 개념이 완전하게 정착되지 않았고, 뚜렷한 시장지배자가 없는 만큼 교육 기업들의 디지털 경쟁이 한동안 심화될 것으로 전망
- 또한 포스트 코로나 시대에 언택트(Untact)가 삶의 양식으로 자리 잡을 것으로 전망되며 교육 분야도 마찬가지로 온라인 교육 등 에듀테크 대한 관심이 증가하는 중

□ AI 기반 교육 서비스는 성장 가능성이 많은 산업

- 가상현실 공간을 통해 교육콘텐츠의 물리적 제약을 없애는 실감화(VR, AR) 분야, 사물인터넷(IoT)과 빠른 이동통신 기술을 활용해 학생 부모 멘토를 연결해 교육과정에서 즉각적인 피드백이 제공되는 연결화 분야, AI에 기반해 인공지능 튜터를 개발하는 지능화 분야 등
- 기존의 불특정 다수를 대상으로 한 획일화된 교육 커리큘럼은 교육자에 비해 학생들의 숫자가 월등히 많다는 점에서 불가피한 일이었지만, 최근 AI 기술이 발달하면서 교육 콘텐츠를 AI를 통해 개인에 맞춤형으로 제공하고 관련 프로세스를 완전히 자동화할 수 있기 때문에 스타트업을 중심으로 발상의 전환이 이뤄지는 중
- 교육 분야의 AI 도입률은 현재 다른 분야에 비해 상대적으로 낮은편으로 확장 가능성이 많은 분야

[주요 분야별 AI 도입 현황과 전망]



* 출처: McKinsey Global Institute, 2018

◎ AI 기반 교육 서비스 등장에 따른 민간 부문과 공교육 부문 격차 발생

□ 국내외 민간 교육 기업은 AI 등 연구 및 투자 및 서비스 출시 활발

- 글로벌 기업은 학교 현장에 자사 제품과 SW를 공급하여 에듀테크 시장을 선점하기 위해 경쟁 중
 - 구글은 지능형 협업 클라우드 소프트웨어인 G-Suite와 학습관리시스템 Google Classroom을 연결하며 에듀테크 산업에 대한 행보 강화
 - 마이크로소프트도 MS-Office 기반의 학습공유 서비스인 Office 365 Education을 무료로 배포하고, 게임을 교육용으로 변형한 Minecraft를 보급
- 해외 인공지능 안면인식 기술을 활용한 교육기업 VIPKID의 누적 투자유치 금액은 2억 2천만 달러 수준으로 전해짐(Business Wire, 2018.6월)
- 국내 인공지능 스타트업 중 AI 기반 맞춤형 학습솔루션 산타토익 등을 서비스 하는 루이드는 2014년 부터 2019년까지 총 344억원 투자 유치(THVC, 2019.10월)
- 국내 주요 교육 기업들은 에듀테크에 전폭 투자를 하고 있는 것으로 나타남
 - 교원그룹은 2020년 경영방침을 '전(全) 계열사 인공지능화'로 정하고 IT기업으로의 전환을 본격화했으며 아이스크림미디어도 지난해에 이어 올해도 에듀테크 중심 성장론 언급
 - 대교그룹과 재능교육, 천재교육, 휴넷도 2020년 최우선 경영 전략을 '디지털 전환'으로 정하고 가상현실(VR)·증강현실(AR)을 비롯한 차세대 교육 콘텐츠 확대를 예고

□ 국내 공교육 부문은 AI 기반 교육 서비스 개발을 위한 인프라 및 데이터 부족

- 민간 기업들의 경우 모바일 기기를 활용하여 기존의 학습 데이터들이 존재하고 이를 AI 알고리즘에 적용하여 새로 생기는 데이터를 분석하는 형태로 AI 학습을 적용 중
- 하지만 공교육 부문에서는 학생들의 학습 과정이나 결과물은 전혀 데이터화 되지 못하고 있으며, AI 교육을 할 수 있는 기반인 학습 데이터가 전혀 존재하지 않는 상황
- 학습에 AI를 활용하기 위해서는 데이터를 저장할 수 있는 시스템을 먼저 만들어야 하고, 이를 위해 학습환경에서 데이터를 저장하는 인프라(무선 인프라)가 필요
- 현재의 환경에서 모든 교실에 네트워크가 연결되고 학습 데이터를 저장하는 인프라 설비가 우선이 되어야 하고, 데이터를 저장하고 보관할 수 있는 저장고(클라우드)의 도입이 되어야 데이터가 쌓여 있을 때 AI가 분석이 가능
- 분석값은 새로운 데이터가 지속해서 추가되고, 데이터가 많아질수록 정확하고, 세밀해질 수 있는데 인프라 구축이 없이 AI 활용 교육을 추진하는 것이 어려운 상황

[한국 에듀테크 스타트업 투자 현황]

사교육 부문			일반		
외국어 교육	수학교육	프로그래밍	교육보조	유아교육	강의/강연
505억 (20개사)	207억 (8개사)	80억 (9개사)	136억 (9개사)	143억 (4개사)	18억 (4개사)

* 출처: THE VC, 2017

나. 시장 분석

(1) 세계시장

- Grandviewresearch이 2020년 발표한 세계 에듀테크 시장 규모에 따르면 2018년 세계 에듀테크 시장 규모 647억 달러 규모이고, 연평균 성장률 18.1%로 2024년 1,755억 달러 규모로 성장 전망
 - AI 기반 교육 서비스 시장은 2025년 61억 달러로 성장할 것으로 기대되며, VR·AR 교육 시장에 이어 로보틱스, 블록체인과 함께 교육산업의 혁신을 이끌 것으로 전망(Holon IQ, 2019)

[에듀테크 세계 시장규모 및 전망]

(단위 : 십억 달러, %)

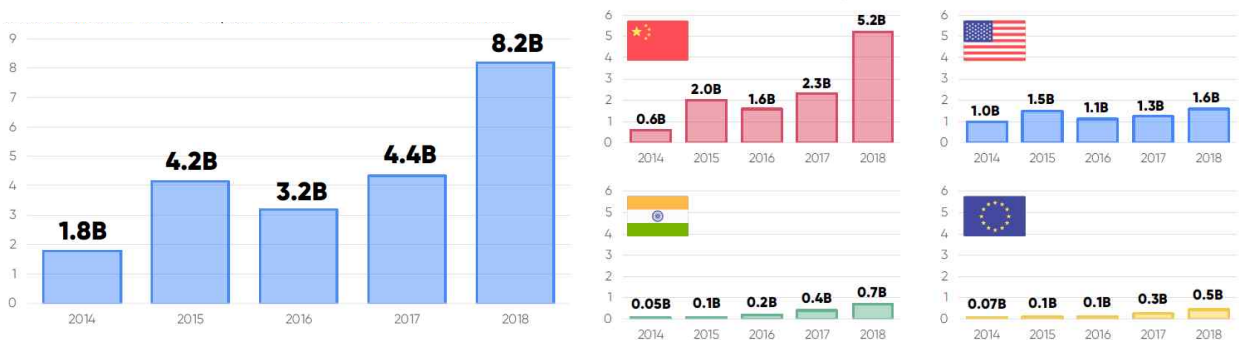
구분	'18	'19	'20	'21	'22	'23	'24	CAGR
세계시장	64.7	76.4	90.2	106.6	125.9	148.6	175.5	18.1

* 출처: Grandviewresearch - Education Technology Market Size(2020)을 바탕으로 네모아이씨지 재추정

- 중국, 미국이 AI 기반 교육 서비스를 포함한 에듀테크 시장을 선도
 - 2018년 에듀테크에 대한 투자가 급증(82억 달러로 전년도 대비 약 2배)하였으며, 특히 AI 강국으로 부상한 중국과 미국 시장의 투자가 활발하게 진행
 - 높은 교육열과 1가구 1자녀 제도 폐지 등으로 에듀테크 블루오션 시장으로 떠오른 중국은 전 세계 투자의 50% 이상을 차지하고 있으며 이어 미국, 유럽, 인도 순

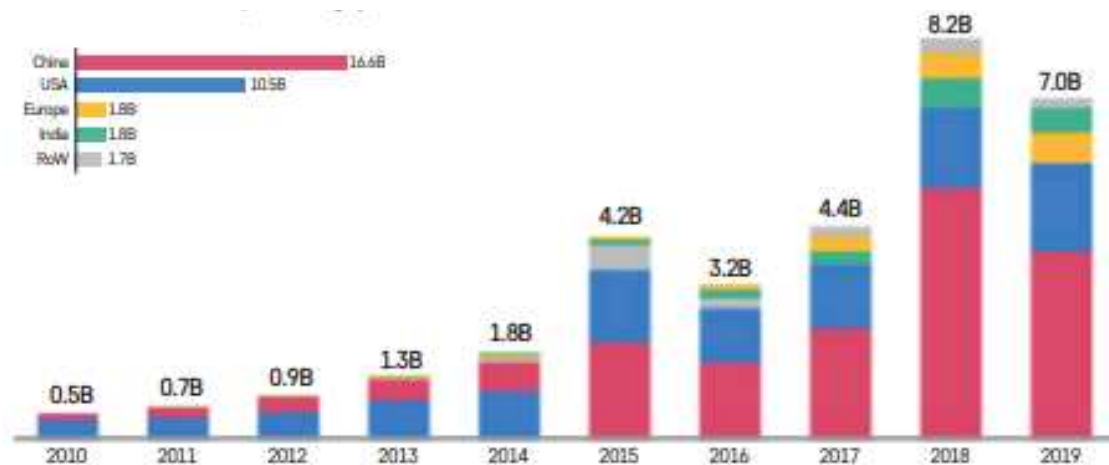
[글로벌 벤처캐피탈 에듀테크 투자규모 현황]

(단위 : 십억 달러)



* 출처: Holon IQ, 2019

[글로벌 벤처캐피탈 에듀테크 투자규모 추이]



* 출처: Holon IQ(2020.1.28.)

□ 시가총액 상위 에듀테크 유니콘 기업은 중국, 인도, 미국 순으로 선점

- 인도의 튜터링 유니콘 기업 ByJu's 이 58억 달러로 가장 기업가치 평가액이 높으나 중국 기업인 2~3위(VIPKid, Yuanfudao)를 더하면 전체 에듀테크 분야 기업에서 중국이 더 많은 기업가치 보유
- 미국의 유니콘 기업이 다음 순위로 높은 기업가치 기업을 보유하고 있으며 EU의 경우 중국, 미국, 인도에 비하면 유니콘 기업의 숫자와 규모는 부족한 편

[에듀테크 분야 유니콘 기업]

기업명	국가	분야	기업가치 평가액(\$)
ByJu's	인도	튜터링	58억
VIPKid	중국	외국어 교육	45억
Yuanfudao	중국	튜터링	30억
Duolingo	미국	외국어 교육	15억
Guild Education	미국	숙련도 향상	10억 이상
Knowbox	중국	튜터링	10억 이상
Coursera	미국	대규모 온라인 공개수업	10억 이상
iTutorGroup	중국	외국어 교육	10억
Zhangmen	중국	튜터링	10억 이상
Huike	중국	온라인 프로그램 관리	10억
17zuoye	중국	튜터링	10억
Age of Learning	미국	온라인 교육과정	10억
Udacity	미국	소프트웨어 개발 교육	11억
Hujiang	중국	온라인 교육과정	10억

* 출처: 에듀테크 시장 현황 및 시사점, Holon IQ(2020.1)

- EU 내부에서는 영국 에듀테크 발전에 유리한 환경을 바탕으로 영국 전역에 1,000개 이상의 에듀테크 기업을 보유하고 있으며, 프랑스가 다음 순으로 큰 규모
 - 영국은 EU 국가 중에서 에듀테크 투자가 가장 활발하고, 2019년 벤처캐피탈 투자액이 4.68억 달러로 급증해 EU 전체의 73%를 차지
 - 영국의 경우 학교의 인터넷 보급률이 높아 교사들이 에듀테크에 접근하기 쉽고, Pearson, 옥스퍼드 대학출판사 등 영향력 있는 교육 기업들과 협업이 가능하며 미국 등 다른 영어권 국가로의 진출이 용이
 - 영국의 에듀테크 투자 규모 및 EU내 비중 (백만 달러, %) : '14(30.0, 42.9) → '15(50.0, 41.7) → '16(29.0, 24.4) → '17 (96.0, 34.7) → '18(142.0, 31.7) → '19(468.0, 73.0)
 - 프랑스는 에듀테크 투자액이 2018년 1.47억 달러로 크게 확대됐으나 2019년 영국의 투자가 전년 대비 3배 이상 증가하여 압도적 규모 차지
 - 독일, 북유럽, 아일랜드, 스페인 등 기타 EU 지역은 영국, 프랑스 대비 미미한 수준으로 분석

[EU 국가별 에듀테크 투자 추이]

(단위: 백만 달러)

	영국	프랑스	독일	북유럽	아일랜드	스페인	기타지역	EU 전체
2014	30.0	5.0	1.6	3.2	2.2	1.7	26.3	70.0
2015	50.0	7.5	31.8	3.4	3.3	1.0	23.0	120.0
2016	29.0	19.0	4.4	8.0	0.1	13.5	45.0	119.0
2017	96.0	12.0	37.0	57.0	9.0	10.0	56.0	277.0
2018	142.0	147.0	42.0	46.0	1.2	11.4	59.0	448.0
2019	468.0	108.0	17.1	16.6	1.0	15.6	15.0	641.0

* 출처: Brighteye Ventures(2020)

(2) 국내시장

- 한국무역협회 국제통상무역이 발표한 국내 에듀테크 시장규모는 2018년 3조 8,400억 규모로 평가되었으며 연평균 18.8%로 성장하여 2024년 10조 8,000만 원으로 성장이 전망됨

[국내 에듀테크 시장규모 및 전망]

(단위 : 조 원, %)

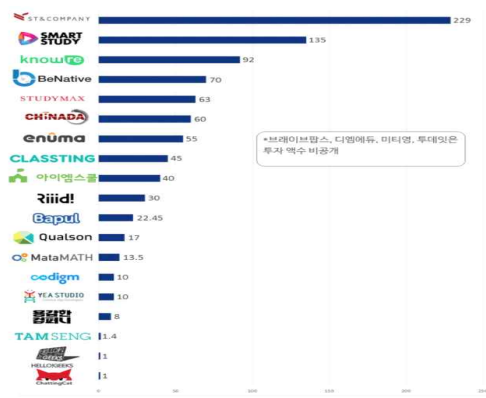
구분	'18	'19	'20	'21	'22	'23	'24	CAGR
국내시장	3.84	4.56	5.42	6.44	7.65	9.09	10.80	18.8

* 출처: 한국무역협회 국제통상무역 - 에듀테크 시장 현황 및 시사점 보고서 및 세계 AI 전체 시장과 국내 시장을 복합적으로 고려하여 '19 이후 네모아이씨지 추정

- 소수의 대형사업자들에게 매출이 집중되는 한편, 전체 사업자 수의 절반을 차지하는 영세사업자의 매출은 1.6%에 불과해 산업 내 양극화 존재
 - 매출 100억 원 이상의 대형사업자는 전체 사업자 수의 3.0%에 불과하지만, 에듀테크 매출액의 42.1%를 차지하고, 매출 1억 원 미만의 영세사업자는 전체 사업자 수의 50.1%임에도 불구하고 매출액의 1.6%만을 차지
- 국내 에듀테크 스타트업의 주요 수익구조는 B2B 사업 형태이며, 전통 교육 기업들과 서로 협력하여 발전하는 양상
 - 현재 한국에서 활동하는 에듀테크 스타트업은 최소 50곳 이상으로 외국어 학습 서비스 업체가 가장 투자를 많이 받았으며 개인맞춤화 학습, 교육용 SNS 시장에도 투자가 몰리는 중
 - 국내 에듀테크 기업은 B2B 사업으로 콘텐츠를 판매하거나 콘텐츠 제작비용으로 수익
 - 조사에 참여한 스타트업 중 90% 넘는 곳에서 “B2B 사업을 확대하고 교육 기업과 제휴를 더 모색할 것”이라고 응답
 - 에듀테크 스타트업의 고객 대부분은 전통 교육 출판사나 온라인 강의 업체, 학습지 업체이며 전통 교육 기업들과 충돌하기보다 서로 협력 관계

[에듀테크 스타트업이 받은 투자 금액]

(단위 : 억 원)



* 출처: Bloter, 정보통신산업진흥원 2018 이러닝산업 실태조사

[2018 에듀테크 매출액 분포]

(단위 : 조 원, %)

구분	사업자 수		총매출액	
	사업자 수	비중	매출액	비중
1억원 미만	879	50.1	0.6	1.6
1~10억원	420	24.0	346.1	9.0
10~30억원	197	11.2	530.6	13.8
30~50억원	119	6.8	538.3	14.0
50~100억원	86	4.9	749.8	19.5
100억원	52	3.0	1618.7	42.1
전체	1753	100.0	3845.0	100

□ 국내 에듀테크 콘텐츠 기업의 R&D 투자 금액 및 비중은 5년 연속 증가 추세

- 시스템 및 기타 부문에 대한 투자는 증감을 반복한 반면, 기술개발 등 R&D 투자 비중은 2014년 55.2%에서 2018년 74.6%로 꾸준히 증가
- AI 기술 및 소프트웨어 개발, 데이터 분석, 증강/가상현실, 클라우드 컴퓨팅 등 교육과 관련된 기술 투자 금액이 증가하는 추세

[에듀테크 콘텐츠 기업의 사업 부문별 평균 투자 현황]

(단위 : 백만 원, %)

구분	2014	2015	2016	2017	2018
총투자액	107.0	119.4	109.0	131.0	151.7
R&D	59.1	60.9	64.5	93.6	113.1
(비중)	(55.2)	(51.0)	(59.2)	(71.5)	(74.6)
시스템	42.0	51.0	40.6	32.1	38.1
(비중)	(39.3)	(42.7)	(37.2)	(24.5)	(25.1)
기타	5.9	7.5	3.9	5.3	0.5
(비중)	(5.5)	(6.3)	(3.6)	(4.0)	(0.3)

* 출처: 각 연도별 이력산업 실태조사, 정보통신산업진흥원, 2018

□ 기존에 교육 서비스를 제공하고 있던 주요 기업들이 AI 기반 교육 서비스 등 ICT 기술을 중심으로 교육 서비스를 제공하는 신규 스타트업과의 협업 사례가 등장

[주요 교육 서비스 기업의 에듀테크 협력사례]

기업명	협업내용	성과
대교	AI 수학교육 플랫폼 '노리' 인수	AI 학습서비스 '써밋 스피드 수학', '써밋 스코어 수학' 개발
웅진씽크빅	실리콘밸리 머신러닝 스타트업 '키드애플티브' 지분 인수(10%)	빅데이터와 AI 기술을 활용해 학습자의 학습행동 패턴을 분석해주는 서비스 '북클럽 AI 학습코칭', 'AI수학', 'AI독서케어' 개발
교원	코딩교육 로봇제작 스타트업 '럭스로보'와 제휴	럭스로보와의 협력으로 코딩로봇 모듈을 활용한 컴퓨터·코딩 교육 서비스 '레드펜코딩' 개발
천재교육	에듀테크센터(창업보육센터)설립하여 유망 스타트업 발굴 및 지원	입주기업 '클래스큐브'와의 협업을 통해 AI 기반 수학 플랫폼 '닥터매쓰(Dr.Math)' 론칭

* 출처: 에듀테크(EduTech) 시장 현황 및 시사점, KITA, 2020.5

3. 기술 개발 동향

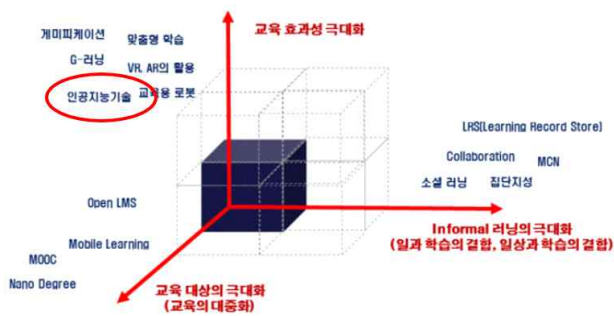
- 기술경쟁력
 - AI 기반 교육서비스는 미국이 최고기술국으로 평가되었으며, 우리나라는 최고기술국 대비 87.6%의 기술 수준을 보유하고 있으며, 최고기술국과의 기술격차는 1.3년으로 분석
 - 중소기업의 기술경쟁력은 최고기술국 대비 79.0%, 기술격차는 2년으로 평가
 - 한국>EU(77.2%)>중국(73.4%)>일본(71%)의 순으로 평가
- 기술수명주기(TCT)⁶⁾
 - AI 기반 교육서비스는 4.6의 기술수명주기를 지닌 것으로 파악

가. 기술개발 이슈

◎ 교육 효과성 극대화를 위해 AI 기반 교육 서비스 기술 개발 중

- 에듀테크 기술의 발전 방향을 살펴보면 크게 교육의 대중화, 효과성 극대화, 일상과 학습의 결합이라는 세 가지 목적을 중심으로 하여 각각 개별적인 기술이 개발되는 중
- 국내 에듀테크의 중소기업 기술 로드맵 (2018-2020) 따르면 실감형 콘텐츠, 소프트웨어 코딩, 맞춤형 학습 서비스가 에듀테크의 핵심 요소 기술이며, 이중 AI 기반 교육 서비스와 연관이 높은 부분은 '맞춤형 학습 서비스' 기술개발 부문

[에듀테크 기술의 방향]



* 출처: 휴넷 에듀테크 연구소

[국내 에듀테크의 중소기업 기술 로드맵]



* 출처: 중소중견기업 기술로드맵 전략보고서, 2017

6) 기술수명주기(TCT, Technical Cycle Time): 특허 출원연도와 인용한 특허들의 출원연도 차이의 중앙값을 통해 기술 변화속도 및 기술의 경제적 수명을 예측

- 이 중, AI 기반 교육 서비스는 교육 효과를 극대화하는 방향으로 세부 핵심기술 연구·개발 중
 - 맞춤형 학습 서비스를 위한 핵심 요소 기술로는 ① 학습 빅데이터 분석 기술 ② 학습 에이전트 기술 ③ MOOC 학습 데이터 분석 기술 ④ 소셜러닝 콘텐츠 분석 기술⁷⁾

맞춤형 학습 서비스의 제품 분류	세부 기술 범위
① 학습 빅데이터 분석 기술	<ul style="list-style-type: none"> • 학습자 평가 및 채점 기술, 학습자 모델링 및 학습 능력 진단 기술, 학습자 성적 예측 기술
② 학습 에이전트 기술	<ul style="list-style-type: none"> • 맞춤 학습 콘텐츠 추천 기술, 학습 콘텐츠 제공 플랫폼 기술, 맞춤형 학습 스케줄 생성 기술, 개인별 학습 코칭 기술
③ MOOC 학습 데이터 분석 기술	<ul style="list-style-type: none"> • 스텔스 시험 기술, 학습 데이터 기반 Domain Knowledge 추출 기술, 데이터 마이닝 기술
④ 소셜러닝 콘텐츠 분석 기술	<ul style="list-style-type: none"> • 소셜러닝 콘텐츠 분석 기술, 소셜러닝 콘텐츠 검색 기술

◎ 국내 교육 부문 스타트업의 경우 AI 기반의 맞춤형 학습 서비스 제공을 위해 신규 사용자 분석, 문제 분석, 개별 맞춤형 문제 콘텐츠를 제공하는 기술 개발 중

- 국내 AI 기반 교육 서비스인 ‘산타토익’ 서비스를 제공하고 있는 리이드(Riuid)의 경우 사용자를 정확하게 분석하고 맞춤형 교육을 제공하기 위해 다음과 같은 기술을 개발하고 현재 특허를 보유
- 국내 AI 기반 교육 서비스인 ‘노리’는 AI 맞춤형 학습 원천 기술에 대한 미국 특허를 취득(‘20.10)
 - 학습자가 문제를 해결할 수 있는 확률을 AI 알고리즘을 통해 계산해 같은 실수를 반복하지 않도록 적합한 문제를 추천하는 기술이며, 학습자의 단기간 수학 성적을 향상하는데 효과적
- AI 수학 풀이 검색 서비스를 제공하는 매스프레소는 인공지능을 기반으로 자체 개발한 광학문자판독(OCR) 기술을 통해 촬영한 문제 이미지의 한국어와 수식을 동시에 인식
- 게임으로 교육 콘텐츠를 제공하는 캐치잇플레이는 머신러닝 기반의 AI 추천 시스템 (PCT / KRR2014/006437)을 탑재해 사용자에게 맞춤형 학습법 제공
- 데이터에듀는 추천할 교육과정에 적합한 동영상 콘텐츠 추출을 위한 크롤링 및 필터링기술, 머신러닝, 딥러닝을 활용한 동영상 강의 분류·평점 생성 모형 및 스마트 학습관리 서비스 제공 및 맞춤형 큐레이션 알고리즘을 개발하여 맞춤형 교육 서비스를 위한 AI 기반의 동영상 강의 분류 기술을 개발
- 오앤이교육은 개인 맞춤형 교육 서비스 제공을 위한 인공지능(AI) 기반 학습진행 가이드 학습 내비게이션(Navigation) 개발

7) 중소기업 기술로드맵 2018-2020 디지털콘텐츠·디자인, 중소벤처기업부, 2018.3월

◎ 학교 등 공교육 부문에서의 AI 기반 교육 서비스 기술 이슈 해결 필요

- 현재 시중에서 서비스를 제공하고 있는 AI 기반 교육 서비스는 특정 과목, 목적을 위해 만들어진 서비스가 대부분으로 공교육 등 포괄적인 학교 교육 부문에서의 개발은 미비한 상황
 - AI 기반 스타트업의 교육 서비스를 보면 AI 기반 토익 문제풀이 서비스(뤼이드), AI 활용 수학 풀이 제공 (매스프레스) 등 특정 과목 등에 집중되어 있는 상황
 - 하지만 학교 등 공교육 부문에서의 학생들의 학습 과정 및 시험 결과 데이터는 오프라인에서 대부분 진행되기 때문에 학생들의 정보과 AI 학습용 데이터화 되지 못하고 있는 상황
 - 맞춤형 교육을 제공하기 위해서는 학생 개인의 성격이나 취향, 강점과 단점, 공부 시간 등 개별적인 특성까지 분석이 되어야 하기 때문에 단순 시험문제 데이터 축적으로는 한계 존재
- 학교 등 공교육 시스템의 학생들의 개별 데이터를 모으고 시로 분석해야하는 복합적인 관점에서 기술개발이 필요
 - ① 디지털 디바이스 개발 및 보급, ② 이를 학교 시스템과 연계할 수 있는 클라우드 시스템, 그리고 누적된 ③ AI 데이터를 분석할 수 있는 AI 모델, 알고리즘 개발이 복합적으로 이루어져야 하는 이슈가 있음
- 개별 과목이 아닌 일반 교육 부문에서의 AI 기반 교육 서비스를 제공하기 위해서는 편향들에 관한 시가 가지는 근본적인 기술적 문제 고려 필요
 - 부진 학생에게는 쉬운 문제만 내고 우수 학생에게는 어려운 문제만 계속 제공하게 되어 학생들의 성적 격차를 만들 수도 있는 시의 편향 등에 관한 이슈도 반드시 고려 필요
 - 학생들의 개별 특성을 고려하지 않은채 단순 성적 결과로만 제공되는 AI 기반 교육 서비스는 또 다른 사회문제를 일으킬 가능성 존재

나. 생태계 기술 동향

(1) 해외 플레이어 동향

- (Elemental path) Kickstater를 통해 데뷔하였으며 클라우드 기반인 IBM Watson을 활용하여 어린이용 스마트 장난감과 융합하여 서비스를 제공
 - IBM Watson을 공룡 로봇인 STEMosaur에 적용하여 어린이 스스로 조립한 지능형 장난감과 대화하도록 지원
- (HUBERT.AI) 독창성, 상상력, 윤리적 반성, 배경 추론 등과 같이 현재의 교육 평가 시스템에서 측정하기 어려운 역량을 평가하는 서비스 제공
 - 인지 컴퓨팅 기술을 활용, AI 비서를 통해 학생들에게 말하기 시험을 제공하고 후속 질문을 통해 효과적인 맞춤형 평가 진행
- (MobyMax) AI를 기반으로 수학, 언어, 어휘 등에 대한 완벽한 커리큘럼을 만들 수 있는 웹 기반 무료 학습 도구 제공
 - 학습자의 데이터를 활용하여 기계학습 알고리즘을 바탕으로 학습자 맞춤형 수업을 제공하고, 학생들은 개별 학습 속도에 맞춰 학습 진행
- (Querium) 클라우드 및 AI를 활용해 개인화된 단계별 개인 교습 제공 플랫폼으로 'StepWise'라는 개인화 학습 프로그램 제공
 - STEM(수학, 공학, 기술, 과학 융합 교육) 응용 프로그램에 AI를 사용하여, 가상 도우미가 학생들이 학습장애 물에 봉착할 때 문제 해결 지원

[Elemental path]



[HUBERT.AI]



[MobyMax]



[Querium]



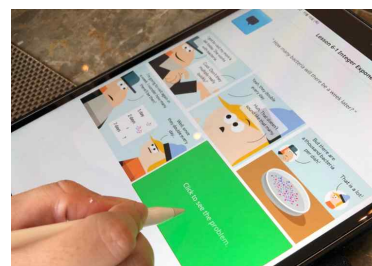
(2) 국내 플레이어 동향

- (뤼이드) AI를 활용한 딥러닝 기술로 학습자가 틀릴 문제를 예측하고, 점수가 가장 빨리 오를 문제를 추천해 최단시간 내에 점수를 향상할 수 있는 ‘산타토익’ 서비스 제공
 - AI 기반 교육 서비스 성능 개선을 위한 논문 및 특허를 지속적으로냄으로써 전체 AI 기반 교육 서비스 시장 고도화에 기여 중
 - (특허) AI 튜터를 이용한 학습 콘텐츠 제공 방법 및 장치(KR 10-2084556), 학습 효율을 기반으로 개인 맞춤형 교육 콘텐츠를 제공하기 위한 기계학습 방법(KR 10-2075936) 등
- (노리) 딥러닝 수학교육 기술 플랫폼을 활용하여 학습 능력을 진단하고 개인별 맞춤형 교육과정 및 콘텐츠 제공
 - 국내와 미국에서 학습 데이터를 분석하고 맞춤형 콘텐츠를 추천하는 드릴다운 방식(Drill Down)의 원천기술을 확보
- (매스프레소) 질문·답변과 풀이 검색이 가능한 플랫폼 ‘관다(QUANDA)’ 운영. AI 기반의 광학문자판독(OCR) 기술을 개발해 모르는 수학 문제를 사진 촬영해 올리면 5초 안에 해설을 제공하는 ‘5초 풀이 검색’ 서비스 제공
- (캐치잇플레이) 게임으로 영어를 배울 수 있는 ‘캐치잇잉글리시’ 애플리케이션 운영. 머신러닝 기반의 AI 추천 시스템(PCT/KRR2014/006437)을 탑재해 사용자에게 맞춤형 학습법 제공. 단어와 문장의 동시 학습이 가능하고, 듣기와 말하기 연습도 함께 할 수 있음
- (에그번에듀케이션) AI 챗봇 기술을 활용해 외국인을 대상으로 한 한국어/중국어/일본어 학습, 한국인 대상 영어교육 서비스 등 제공

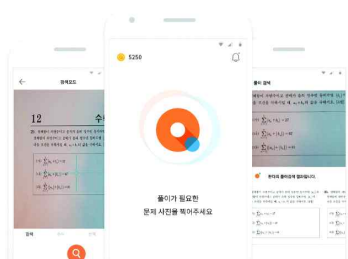
[뤼이드]



[노리]



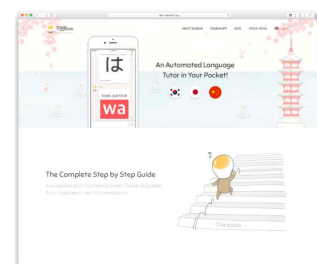
[매스프레소]



[캐치잇 플레이]



[에그번에듀케이션]



다. 국내 연구개발 기관 및 동향

(1) 연구개발 기관

[AI 기반 교육 서비스 분야 주요 연구조직 현황]

기관	소속	연구 분야
한성대학교	서울과학기술대학교 글로벌융합산업공학과	<ul style="list-style-type: none"> 교육, 인공지능, 멀티모달, 지도학습머신러닝
뤼이드 AI Tribe	뤼이드	<ul style="list-style-type: none"> 토익, SAT등 특정 분야 AI 기반 기술 개발 및 학습자의 전체 학습과정에서 피드백하는 AI 튜터 솔루션
춘천교육대학교	춘천교육대학교 컴퓨터교육과	<ul style="list-style-type: none"> 인공지능 교육, 빅데이터 교육
아주대학교	아주대학교 미디어학과	<ul style="list-style-type: none"> 인공지능 교육, 메이커, 빅데이터 교육, 미디어게임

(2) 기관 기술개발 동향

- (한성대학교) 실시간 학습영상 분석을 통한 인공지능 영어교육 플랫폼 개발을 통해 국내외 에듀테크 서비스 기술의 경쟁력 제고 및 영어 사교육비 절감과 소득에 따른 학습기회의 격차 완화, 균등한 영어교육 기회 제공(2019.10.15. ~ 2020.07.14.)
 - 멀티모달 생체정보 분석 기술 (위드마인드 개발)
 - 영상(표정, 제스처)분석 기술
 - 음성분석 기술
 - 단어분석 기술
 - 영어 학습결과 도출용 머신러닝 모델 (맞춤형 기술파트너 개발)
- (뤼이드 AI Tribe)
 - 실리콘밸리에 설립한 '뤼이드랩스'에 R&D센터를 두고 세계 최고의 AI 기술 인력을 확보 계획
 - 특정 시험을 넘어 학습자의 전체적인 학습 과정에서 지속적인 평가와 피드백을 통해 학습 목표 성취를 돕는 '포머티브 러닝(Formative learning) AI 튜터 솔루션'까지 구축해 나갈 계획
- (춘천교육대학교) 빅데이터 및 인공지능 교육을 위한 블록 프로그래밍 언어 기반 코딩 교육 플랫폼 개발을 통한 학생 교육용 빅데이터/인공지능 프로그램 제공(2020.06.01. ~ 2023.05.31.)
 - 교육용 블록 프로그래밍 언어 기반 코딩 플랫폼 개발시 컴퓨터과학 비전공 학생의 교육력 향상
 - 웹서버 구축의 자유로움으로 인한 개인 블록 저장소 보유 가능
 - 실행 결과 및 과정을 확대 분석 및 확인 가능, Electron을 활용하여 실행 파일 생성 기능 제공 예정
- (아주대학교) 메이커 및 게임 활용을 통한 초중등 인공지능 교육 모델 연구(2020.06.01. ~ 2023.02.28.)
 - 인공지능 학습 도구 및 교육 기법에 대한 분석과 그 효과성 검증
 - 학습자 수준에 맞춘 목표 중심형 메이커활동기반 게임창작 인공지능 프로젝트 수행 기법 연구
 - 공교육 교과목 연계한 융복합 인공지능 리터러시 활동 개발 및 검증 연구
 - 메이커활동기반 게임창작활동에서 학습자의 인공지능 리터러시에 대한 다면적 평가 기준 연구

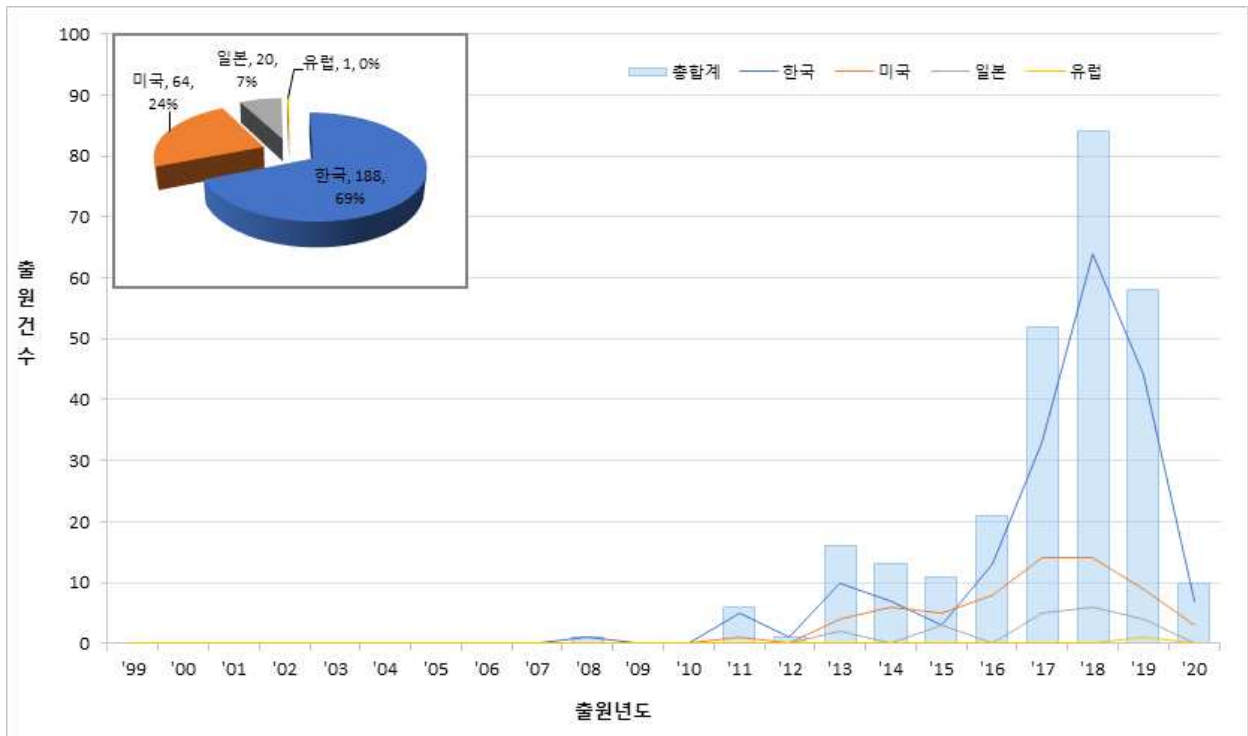
4. 특허 동향

가. 특허동향 분석

(1) 연도별 출원동향

- AI 기반 교육서비스는 '16년부터 급격한 성장을 보임
 - 각 국가별로 살펴보면 한국이 가장 활발한 출원활동을 보이고 있음
- 국가별 출원비중을 살펴보면 한국이 전체의 69%의 출원 비중을 차지하고 있어, 최대 출원국으로 AI 기반 교육서비스 분야를 리드하고 있는 것으로 나타났으며, 미국은 23%, 일본은 7%, 유럽은 1% 순으로 나타남

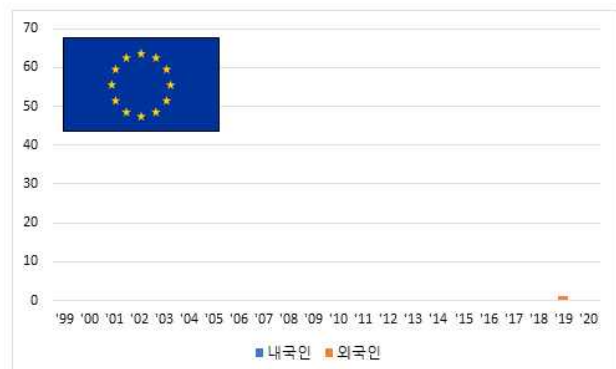
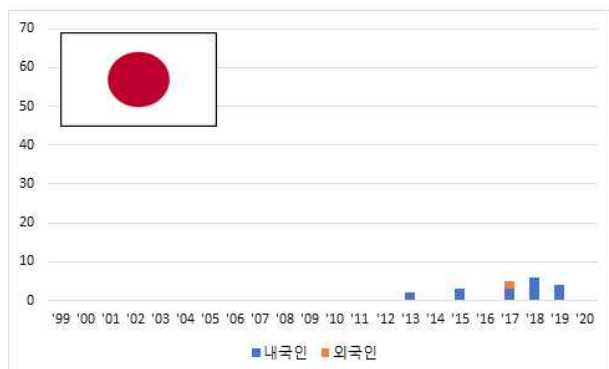
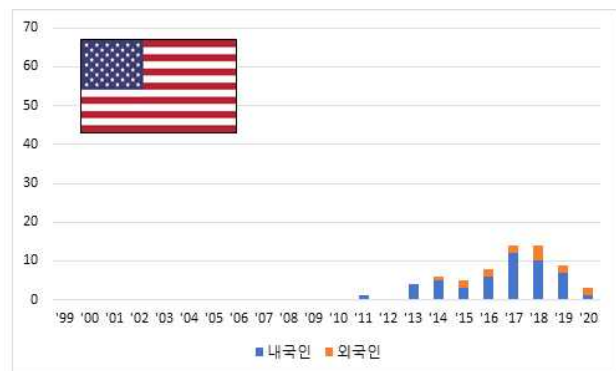
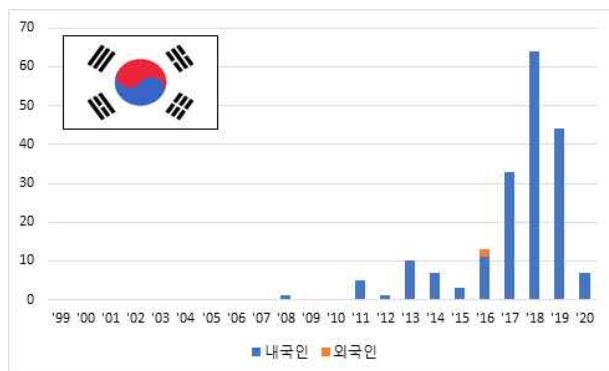
[AI 기반 교육서비스 연도별 출원동향]



(2) 국가별 출원현황

- 한국의 출원현황을 살펴보면, '16년부터 해당 기술의 출원이 급격히 증가하는 추세
 - 내국인 위주의 출원이 진행 중
 - 한국 기술의 양적 흐름은 '18년에 가장 많은 특허를 출원
 - 한국의 출원 수는 미국, 유럽, 일본에 비해 가장 많은 특허를 출원
- 미국의 출원현황을 살펴보면, '13년 이후로 출원이 완만히 증가하는 동향을 보임
- 유럽의 출원현황은 2019년에 1건 출원
- 일본의 출원현황은 출원수가 매년 10건 이하로, 뚜렷한 증감 동향이 나타나지 않음. 해당 기술 분야에서 일본 시장에 대한 관심도가 높지 않은 것으로 보임

[국가별 출원현황]



(3) 기술 집중도 분석

□ 전략제품에 대한 최근 기술 집중도 분석을 위한 구간별 기술 키워드 분석 진행

- 전체 구간(1999년~2020년)에서 문제 데이터베이스, 사용자 단말, Computer Readable Medium 등 키워드가 다수 도출
- 최근 구간 분석 결과, 최근 1구간(2012년~2015년)과 비교할 때 2구간(2016년~2020년)에서 문제 데이터베이스, 서비스 서버, 데이터 분석 프레임워크, 사용자 디바이스, 사용자 단말 키워드가 많이 등장하는 것으로 보아 학습자 패턴 분석 및 평가 기술, 학습 데이터베이스 구축·분석·운영 기술 관련 연구개발이 활발한 것으로 추정

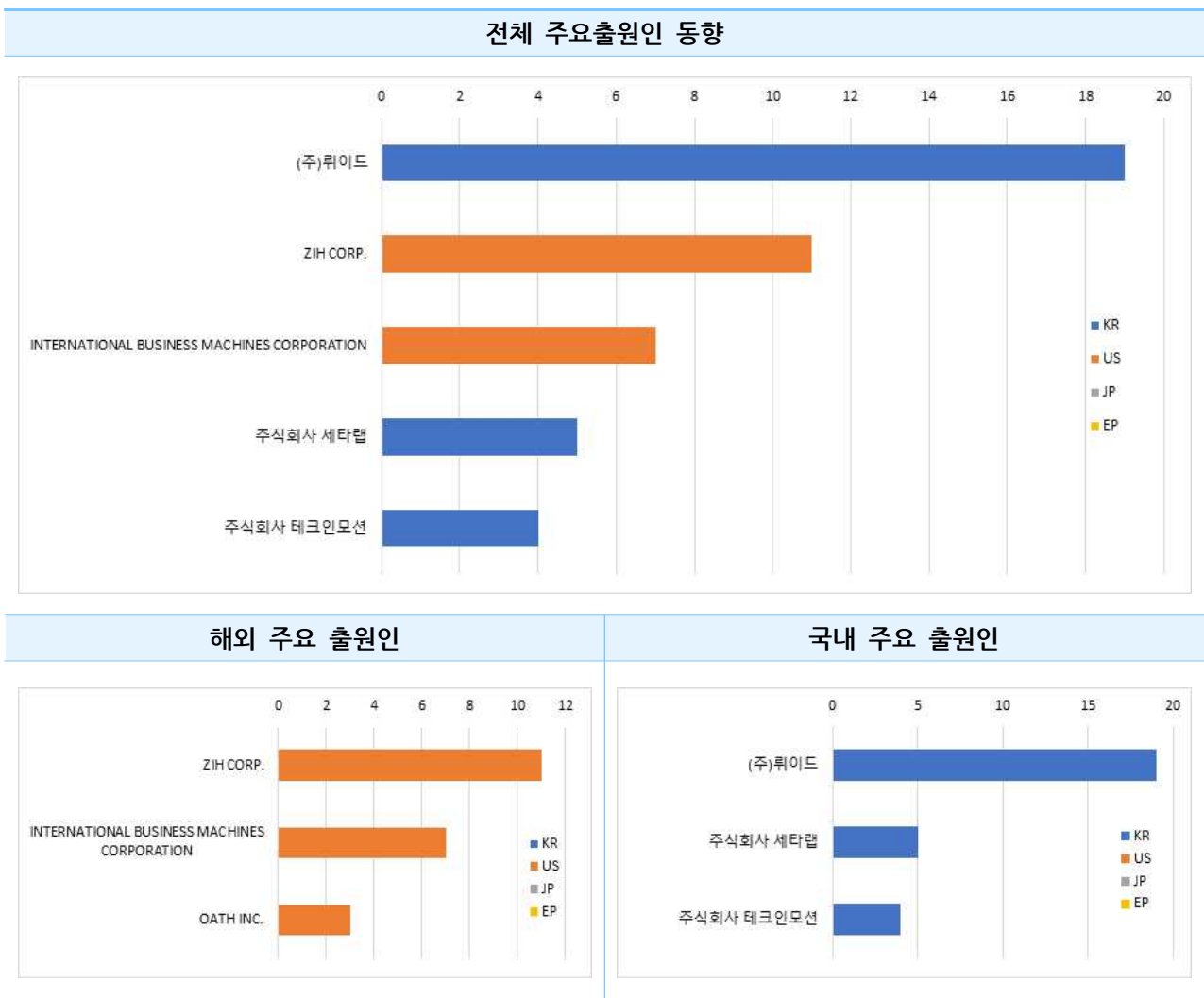
[특히 키워드 변화로 본 기술개발 동향 변화]

전체구간(1999년~2020년)	
<ul style="list-style-type: none"> • 문제 데이터베이스, 사용자 단말, Computer Readable Medium, 서비스 서버, 데이터 분석 프레임워크, 가상 현실, 사용자 디바이스, 모델링 벡터, 풀이 결과 데이터, 증강 현실 	
최근구간(2012년~2020년)	
1구간(2012년~2015년)	2구간(2016년~2020년)
<ul style="list-style-type: none"> • Computer Readable Medium, Interference Signal, Composite Signal, Monitored Region, Interference Rejection, Ultra Wideband Receiver, Source Positioned, 콘텐츠 서버, 증강 현실 마커, 가상 객체 	<ul style="list-style-type: none"> • 문제 데이터베이스, 서비스 서버, 데이터 분석 프레임워크, 사용자 디바이스, 사용자 단말, 모델링 벡터, 풀이 결과 데이터, 가상 현실, 선택 데이터, 특정 과목

나. 주요 출원인 분석

- AI 기반 교육서비스의 전체 주요출원인을 살펴보면, 주로 한국 및 미국 국적의 출원인이 다수 포함되어 있는 것으로 나타났으며, 제 1 출원인으로는 한국의 (주)뤼이드인 것으로 나타남
 - 제 1 출원인인 (주)뤼이드의 출원은 한국에 집중된 경향을 보임
- AI 기반 교육서비스 관련 기술로 학습 솔루션을 제공하는 대기업에 의한 출원이 대다수를 차지
 - 국내에서는 중소기업(개인)의 활발한 출원이 이루어짐

[AI 기반 교육서비스 주요출원인]

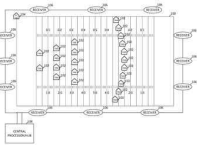
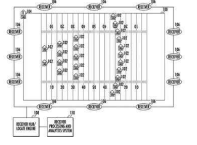
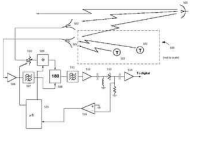
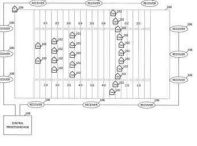
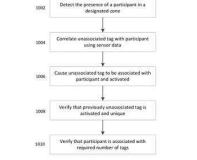


(1) 해외 주요출원인 주요 특허 분석

◎ ZIH CORP.

- ZIH CORP.은 미국 기업으로, AI 기반 교육서비스와 관련하여, 실시간 위치 결정 시스템 기술에 특화된 특허를 다수 출원. 그 중 등록된 특허는 10건
 - 주요 특허들은 무선 주파수 위치 결정에 대한 것으로 실시간 위치 시스템(RTLS) 내에서 위치 태그의 등록을 개선하기 위한 시스템 특허를 다수 출원하는 것으로 파악

[(ZIH CORP. 주요특허 리스트)

등록번호 (출원일)	명칭	기술적용분야	도면
US10050650 (2017.07.17)	Method, apparatus, and computer program product improving registration with real time location services	실시간 위치 시스템 (RTLS) 내에서 위치 태그의 등록을 개선한 무선 주파수 위치 결정 장치	
US9882592 (2016.08.25)	Method, apparatus, and computer program product for tag and individual correlation	모니터링된 영역 주위에 배치된 복수의 센서들, 복수의 태그들 및 복수의 수신기들을 사용하여 환경 측정치들을 개인과 연관시키기 위한 시스템	
US9667287 (2015.10.06)	Multiple antenna interference rejection in ultra-wideband real time locating systems	초광대역(UWB) 실시간 위치 결정 시스템에서 간섭 제거를 제공하기 위한 시스템	
US9742450 (2014.06.06)	Method, apparatus, and computer program product improving registration with real time location services	실시간 위치 시스템 (RTLS) 내에서 위치 태그의 등록을 개선한 무선 주파수 위치 결정 장치	
US9698841 (2014.06.05)	Method and apparatus for associating radio frequency identification tags with participants	위치 탐지 시스템에서 사용하기 위해 무선 주파수 태그를 특정 참가자와 연관시키기 위한 시스템	

* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ IBM(International Business Machines)

- IBM은 미국의 소프트웨어 개발하는 다국적 기업으로, 다수의 인공지능 관련 특허를 보유하고 있으며, AI 기반 교육서비스 기술과 관련한 특허를 미국에 출원
 - AI 기반 교육서비스 기술과 관련하여 8건의 미국 출원을 진행하였으며, 그 중 등록된 특허는 1건으로 파악

[IBM 주요특허 리스트]

등록번호 (출원일)	명칭	기술적용분야	도면
US10621685 (2017.04.03)	Cognitive education advisor	수업에 참석하는 학생의 개인에 대한 개별 장치에서 실시간 데이터를 통해 학생의 성과를 예측하는 방법	

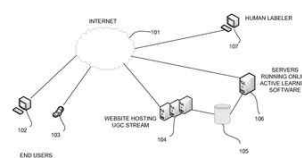
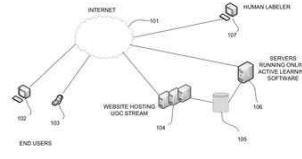
* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ OATH INC.

□ OATH INC.는 미국의 인터넷 서비스 기업으로, 등록특허 2건 보유

- AI 기반 교육서비스에서 온라인에서 활성 학습형 소프트웨어 관련 응용 특허를 3건 출원

[OATH INC. 주요특허 리스트]

등록번호 (출원일)	명칭	기술적용분야	도면
US10523610 (2018.05.07)	Online active learning in user-generated content streams	온라인 활성 학습형을 위한 소프트웨어는 웹사이트에서의 온라인 스트림에 게시된 콘텐츠를 수신하는 기술	
US9967218 (2011.10.26)	Online active learning in user-generated content streams	온라인 활성 학습형을 위한 소프트웨어는 웹사이트에서의 온라인 스트림에 게시된 콘텐츠를 수신하는 기술	


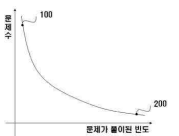



* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

(2) 국내 주요출원인 주요 특허 분석

◎ (주)뤼이드

- (주)뤼이드는 AI 기반 교육서비스 기술에 있어서, 의료 영상 분석 기술과 관련된 특허를 다수 출원
 - (주)뤼이드는 AI 기반 교육서비스 기술에 있어서, 교육 콘텐츠를 제공하기 위한 기계학습 기술과 관련된 특허를 국내에 출원. 2017년부터 해당 기술 관련한 활발한 출원을 진행하였으며, 등록 건수는 9건으로 파악

[(주)뤼이드 주요특허 리스트]

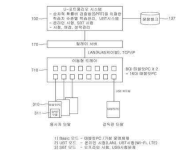
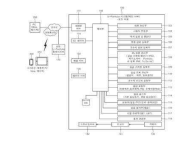
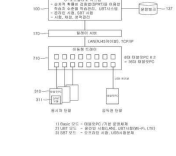
등록번호 (출원일)	명칭	기술적용분야	도면
KR2075936 (2019.08.21)	학습 효율을 기반으로 개인 맞춤형 교육 콘텐츠를 제공하기 위한 기계학습 방법, 장치 및 컴퓨터 프로그램	수집된 문제 풀이 결과 데이터를 기계 학습 프레임워크에 적용하여 사용자의 문제에 대한 학습 효율을 추정하고 이를 기반으로 교육 콘텐츠를 추천하는 방법	
KR2096301 (2019.04.03)	액티브 러닝 기법을 적용한 머신 러닝 프레임워크 운용 방법, 장치 및 컴퓨터 프로그램	대량의 사용자 콘텐츠 소비 결과 데이터를 이용하여 문제 및/또는 사용자에 대한 분석 모델을 생성하고, 상기 분석 모델을 효율적으로 학습시키기 위해 필요한 데이터를 선별하는 전문가 모델을 운용하는 방법	
KR2015075 (2018.10.16)	학습 효율을 기반으로 개인 맞춤형 교육 콘텐츠를 제공하기 위한 기계학습 방법, 장치 및 컴퓨터 프로그램	수집된 문제 풀이 결과 데이터를 기계 학습 프레임워크에 적용하여 사용자의 문제에 대한 학습 효율을 추정하고 이를 기반으로 교육 콘텐츠를 추천하는 방법	
KR1996249 (2018.04.23)	개인 맞춤형 교육 콘텐츠를 제공하기 위한 기계학습 프레임워크 운용 방법, 장치 및 컴퓨터 프로그램	수집된 문제 풀이 결과 데이터를 기계 학습 프레임워크에 적용하여 사용자 및 문제를 분석하는 방법	
KR2084556 (2018.03.05)	Ai 튜터를 이용한 학습 콘텐츠 제공 방법 및 장치	학습자의 문제 풀이에 대응하여 동적으로 동기 부여를 제공하고 학습자의 의욕을 고취시키기 위하여 AI 튜터를 이용하여 학습 콘텐츠	

* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ **주식회사 세타랩**

- 주식회사 세타랩은 한국의 소프트웨어 개발 기업으로, '17년도 12월부터 출원을 시작하여 아직까지 등록된 특허는 없는 것으로 파악됨
 - 주식회사 세타랩의 공개특허를 대상으로 분석한 결과, AI 기반 교육서비스 기술에 있어서, 실습교육 포트폴리오 기술과 관련된 특허를 5건 출원

[주식회사 세타랩 주요특허 리스트]

공개번호 (출원일)	명칭	기술적용분야	도면
KR2020-0069265 (2020.04.29)	U b t와 s b t 시험을 제공하는 유비쿼터스 기반 실습교육 포트폴리오 시스템 및 그 방법	스마트 기기를 사용하여 인공 지능과 빅 데이터 분석 기술을 사용한 클라우드 실습교육 관리 서비스(U-Portfolio)를 제공	
KR2020-0007984 (2020.01.08)	유비쿼터스 기반 실습교육 포트폴리오 시스템 및 그 방법	유비쿼터스 기반 실습교육 포트폴리오 시스템은 스마트 기기를 사용하여 인공 지능과 빅 데이터 분석 기술을 사용한 클라우드 실습교육 관리 서비스(U-Portfolio)를 제공	
KR2019-0111850 (2019.08.19)	U b t와 s b t 시험을 제공하는 유비쿼터스 기반 실습교육 포트폴리오 시스템 및 그 방법	스마트 기기를 사용하여 인공 지능과 빅 데이터 분석 기술을 사용한 클라우드 실습교육 관리 서비스(U-Portfolio)를 제공	

* 공개특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ 주식회사 테크인모션

- 주식회사 테크인모션은 한국의 VR·AR 교육 콘텐츠를 개발하는 기업으로, '17년도 12월부터 출원을 시작하여 아직까지 등록된 특허는 없는 것으로 파악됨
 - 주식회사 테크인모션의 공개특허를 대상으로 분석한 결과, AI 기반 교육서비스 기술에 있어서, VR 교육 시스템 기술과 관련된 특허를 4건 출원

[주식회사 테크인모션 주요특허 리스트]

공개번호 (출원일)	명칭	기술적용분야	도면
KR2019-0109350 (2019.09.16)	Vr 교육 시스템	특정 정비 부품을 VR 영상으로 실제와 동일하게 정비 교육을 실시하기 용이한 VR 교육 시스템	
KR2019-0109351 (2019.09.16)	Vr 교육 시스템	음성 명령에 따라 훈련 시나리오 수행 시 가상 현실 영상 및 가상 현실 영상에 적용되는 장비 정보에 대응하는 장비 영상을 디스플레이하여 훈련이 용이한 VR 교육 시스템	
KR2019-0095566 (2017.12.11)	Vr 교육 시스템	음성 명령에 따라 훈련 시나리오 수행 시 가상 현실 영상 및 가상 현실 영상에 적용되는 장비 정보에 대응하는 장비 영상을 디스플레이하여 훈련이 용이한 VR 교육 시스템	

* 공개특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

다. 기술진입장벽 분석

(1) 기술 집중력 분석

- AI 기반 교육서비스 관련 기술에 대한 시장관점의 기술독점 현황분석을 위해 집중률 지수(CRn: Concentration Ratio n, 상위 n개사 특허점유율의 합) 분석 진행
 - 상위 4개 기업의 시장점유율이 0.15로 AI 기반 교육서비스 분야에 있어서 독과점 정도는 낮은 수준으로 판단
 - 국내 시장에서 중소기업의 점유율 분석결과 0.85로 해당 기술에 대하여 중소기업이 진입하여 활동하고 있는 것으로 파악

[주요출원인의 집중력 및 국내시장 중소기업 집중력 분석]

주요 출원인 집중력	주요출원인	출원건수	특허점유율	CRn	n
	(주)뤼이드(한국)	19	7.0%	0.07	1
	ZIH CORP.(미국)	11	4.0%	0.11	2
	INTERNATIONAL BUSINESS MACHINES CORPORATION(미국)	7	2.6%	0.14	3
	주식회사 세타랩(한국)	5	1.8%	0.15	4
	주식회사 테크인모션(한국)	4	1.5%	0.17	5
	OATH INC.(미국)	3	1.1%	0.18	6
	주식회사 제네시스랩(한국)	3	1.1%	0.19	7
	PEARSON EDUCATION, INC.(미국)	3	1.1%	0.20	8
	4D TECHNOLOGIES, LLC(미국)	3	1.1%	0.21	9
	한국전자통신연구원(한국)	3	1.1%	0.22	10
	전체	273	100%	CR4=0.15	

국내시장 중소기업 집중력	출원인 구분	출원건수	특허점유율	CRn	n
	중소기업(개인)	161	84.7%	0.85	
	대기업	8	4.2%		
	연구기관/대학	21	11.1%		
	전체	190	100%	CR중소기업=0.85	

(2) 특허소송 현황 분석

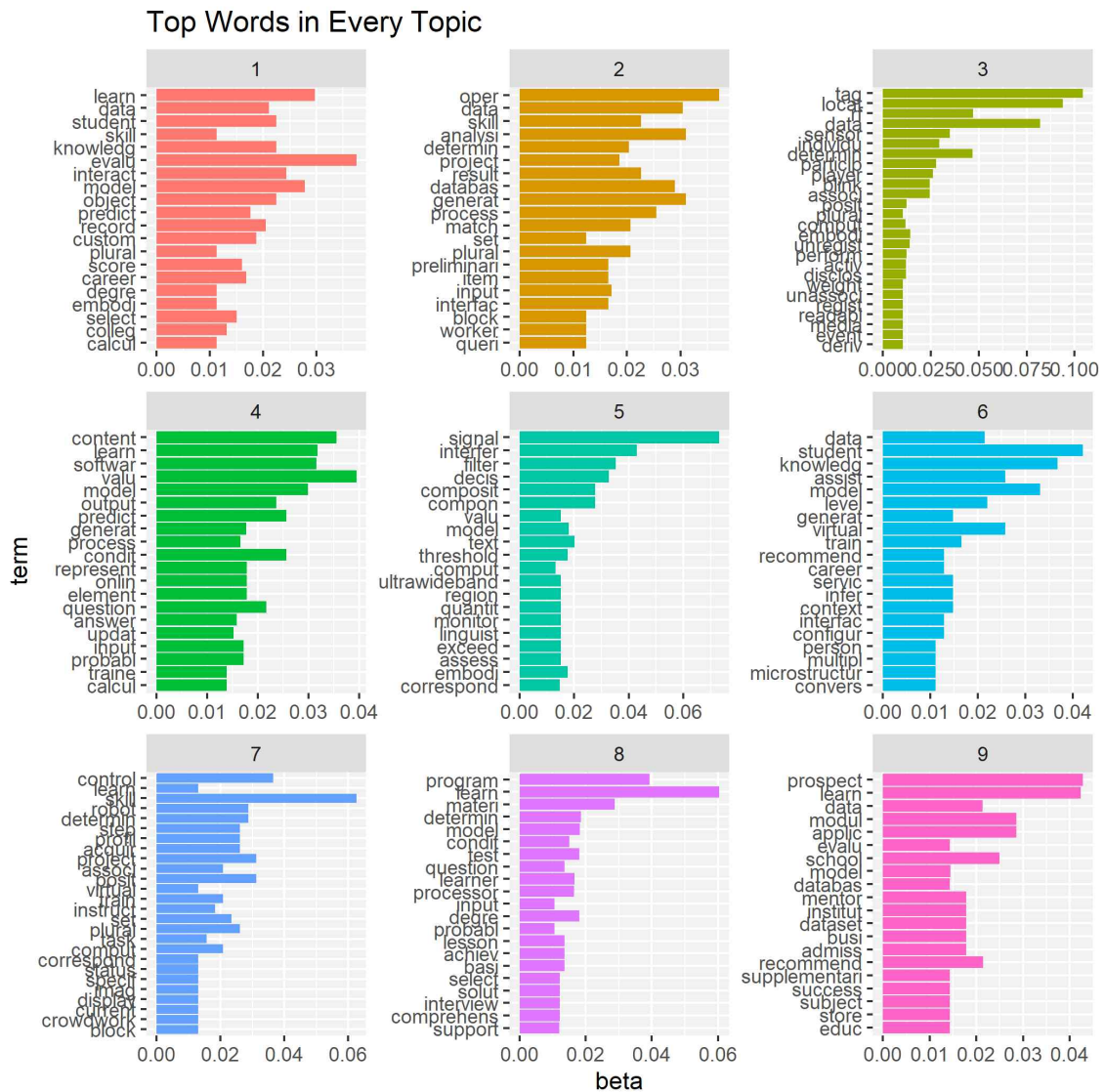
- AI 기반 교육서비스 분야 관련 특허소송 이력은 검색되지 않음
 - 따라서 국내기업이 미국시장에 진입하는 경우, 해당 분야를 선점할 수 있을 것으로 판단
 - 다만, 머신러닝과 관련해서는 1개의 소송이 검색

5. 요소기술 도출

가. 특허 기반 토픽 도출

- 77개의 특허의 내용을 분석하여 구성 성분이 유사한 것끼리 클러스터링을 시도하여 대표성이 있는 토픽을 도출

[시 기반 교육 서비스에 대한 토픽 클러스터링 결과]



나. LDA⁸⁾ 클러스터링 기반 요소기술 도출

[LDA 클러스터링 기반 요소기술 키워드 도출]

No.	상위 키워드	대표적 관련 특허	요소기술 후보
클러스터 01	evaluate, learn, model, interact, knowledge, student, object, data, record, custom	<ul style="list-style-type: none"> • System and methods for automated interactive learning • Systems and methods for assessing and improving student competencies 	학습자 행동 측정 및 모델링
클러스터 02	operate, generate, analysis, data, database, process, skill, result, plural, match	<ul style="list-style-type: none"> • Computer-controlled precision education and training • System and method for determining competency deficiencies and matching between particular projects and available workers or candidates 	학습 데이터베이스 구축·분석·운영 기술
클러스터 03	tag, location, data, rf, determine, sensor, individual, participate, player, blink	<ul style="list-style-type: none"> • Method, apparatus, and computer program product for tag and individual correlation • Systems and methods for activity determination based on human frame 	소셜러닝 콘텐츠 분석 기술
클러스터 04	evaluate, program, value, learn, project, decide, data, control, step, material	<ul style="list-style-type: none"> • Learning support device and learning support method • Predictive score providing device, predictive score providing method, and predictive score providing program 	학습자 패턴 분석 및 평가 기술
클러스터 05	signal, interfere, filter, decide, compost, component, text, model, embodiment, threshold	<ul style="list-style-type: none"> • Journal manuscript submission decision support method and system • Multiple antenna interference rejection in ultra-wideband real time locating systems 	학습자 성적 예측 모델링
클러스터 06	student, knowledge, model, assist, virtual, level, data, train, context, infer	<ul style="list-style-type: none"> • Dynamic knowledge level adaptation of e-learning datagraph structures • System and method for context and sequence aware recommendation 	학습자 패턴 분석 및 평가 기술
클러스터 07	skill, control, project, posit, robot, determine, step, profile, plural, acquire	<ul style="list-style-type: none"> • Method for teaching coding by using augmented reality • Education support system and education support method 	소셜러닝 콘텐츠 분석 기술
클러스터 08	learn, program, material, determine, model, degree, test, learner, processor, conduit	<ul style="list-style-type: none"> • User-adaptive test program, apparatus, and method for selecting model of problem group according to comprehension degree probability • Educational material provision system, educational material provision method, and program 	실시간 정·오답 예측 모델링
클러스터 09	prospect, learn, module, application, school, recommend, data, business, Dataset, institute	<ul style="list-style-type: none"> • Recruiting and admission system • System and method for recommending business schools based on assessing profiles of applicants and business schools 	맞춤형 학습 콘텐츠 추천 기술

8) Latent Dirichlet Allocation

다. 특허 분류체계 기반 요소기술 도출

□ AI 기반 교육서비스 관련 특허에서 총 10개의 주요 IPC코드(메인그룹)를 산출하였으며, 각 그룹의 정의를 기반으로 요소기술 키워드를 아래와 같이 도출

[IPC 분류체계에 기반한 요소기술 도출]

IPC 기술트리		
(서브클래스) 내용	(메인그룹) 내용	요소기술 후보
(G06F) 전기에 의한 디지털 데이터처리	• (G06F-003) 컴퓨터로 처리할 수 있는 형식으로 전송된 데이터를 변환하는 입력기구; 처리장치로부터 출력장치로 데이터를 전송하기 위한 출력기구, 예. 인터페이스 기구	학습 데이터베이스 구축·분석·운영 기술
(G06N) 특정 계산모델 방식의 컴퓨터시스템	• (G06N-005) 지식기반모델을 이용한 컴퓨터시스템	실시간 정·오답 예측 모델링
	• (G06N-020) 머신 러닝	학습자 패턴 분석 및 평가 기술
	• (G06N-099) 이 서브클래스의 다른 그룹으로 분류되지 않는 주제사항	-
(G06Q) 관리용, 상업용, 금융용, 경영용, 감독용 또는 예측용으로 특히 적합한 데이터 처리 시스템 또는 방법; 그 밖에 분류되지 않는 관리용, 상업용, 금융용, 경영용, 감독용 또는 예측용으로 특히 적합한 시스템 또는 방법	• (G06Q-050) 특정 사업 부문에 특히 적합한 시스템 또는 방법, 예. 공익사업 또는 관광	소셜러닝 콘텐츠 분석 기술
	• (G06Q-010) 경영; 관리	학습 데이터베이스 구축·분석·운영 기술
(G09B) 교육용 또는 교사용의 기구; 맹인 또는 농아자와의 의사소통하기 위한 교습용기구; 모형; 유성의; 지구의; 지도; 도표	• (G09B-007) 전기적 조작에 의하는 문제와 해답을 동반하여 작동하는 교습 기기	학습자 행동 측정 및 모델링
	• (G09B-005) 전기적 조작에 의하여 교습 기기	학습자 행동 측정 및 모델링
	• (G09B-019) 이 서브클래스의 다른 메인그룹에 포함되지 않은 교습구	맞춤형 학습 콘텐츠 추천 기술
(H04B) 전송	• (H04B-001) 그룹 H04B 3/00에서 H04B 13/00의 단일 그룹에도 포함되지 않은 전송시스템의 세부; 전송매체에 의하여 특징지어지지 않는 시스템의 세부	실시간 정·오답 예측 모델링

라. 최종 요소기술 도출

- 산업·시장 분석, 기술(특허)분석, 전문가 의견, 타부처 로드맵, 중소기업 기술수요를 바탕으로 로드맵 기획을 위하여 요소기술 도출
- 요소기술을 대상으로 전문가를 통해 기술의 범위, 요소기술 간 중복성 등을 조정·검토하여 최종 요소기술명 확정

[AI기반 교육 서비스 분야 요소기술 도출]

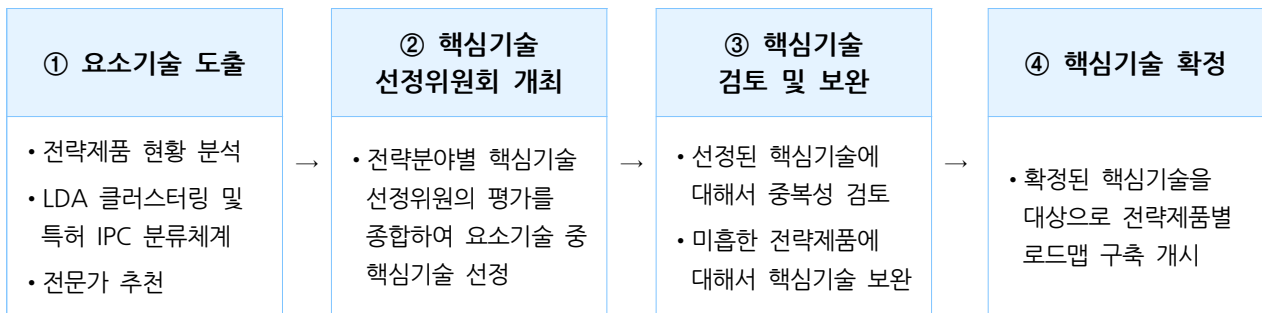
요소기술	출처
학습자 행동 측정 및 모델링	특허 클러스터링, IPC 기술체계, 전문가추천
학습 데이터베이스 구축·분석·운영 기술	특허 클러스터링, IPC 기술체계, 전문가추천
소셜러닝 콘텐츠 분석 기술	특허 클러스터링, IPC 기술체계, 전문가추천
학습자 패턴 분석 및 평가 기술	특허 클러스터링, IPC 기술체계, 전문가추천
학습자 성적 예측 모델링	특허 클러스터링, IPC 기술체계, 전문가추천
문제 컨텍스트 추론 기술	전문가 추천
실시간 정·오답 예측 모델링	특허 클러스터링, IPC 기술체계, 전문가추천
맞춤형 학습 콘텐츠 추천 기술	특허 클러스터링, IPC 기술체계, 전문가추천

6. 전략제품 기술로드맵

가. 핵심기술 선정 절차

- 특허 분석을 통한 요소기술과 기술수요와 각종 문헌을 기반으로 한 요소기술, 전문가 추천 요소기술을 종합하여 요소기술을 도출한 후, 핵심기술 선정위원회의 평가과정 및 검토/보완을 거쳐 핵심기술 확정
- 핵심기술 선정 지표: 기술개발 시급성, 기술개발 파급성, 단기개발 가능성

[핵심기술 선정 프로세스]



나. 핵심기술 리스트

[AI기반 교육 서비스 분야 핵심기술]

핵심기술	개요
학습자 패턴 분석 및 평가 기술	• 학습자의 문제 풀이 데이터를 수집, 기계학습을 통해 학습-풀이 패턴 분석을 통한 학습자의 현 수준 진단 기술
학습 데이터베이스 구축·분석·운영 기술	• 교육 관련 AI 학습용 데이터 구축·분석·운영을 하기 위한 고연산 데이터베이스 운영 시스템
맞춤형 학습 콘텐츠 추천 기술	• 학습자의 역량과 테스트 문항 수준 값을 실시간으로 재조정해 학습자의 답변에 따라 다음 문항을 결정하는 반응형 콘텐츠 추천 기술
학습자 행동 측정 및 모델링	• 학습자의 학습 패턴을 측정·분석하여 더 나은 학습을 할 수 있는 피드백을 주기 위한 행동 패턴 모델링 기술
문제 컨텍스트 추론 기술	• 텍스트 인식, 이미지 인식을 통해 문제 데이터를 수집하고 문제의 의미를 해석하는 기술

다. 중소기업 기술개발 전략

- AI 기반 맞춤형 교육 서비스를 제공하기 위한 기반 기술 개발 (실시간 사용자 분석, 문제 분석, 개별 맞춤형 문제 콘텐츠를 제공할 수 있는 기술 개발 및 공개)
- 학교 교육 등 공교육 부문 기존 인프라와 연동되는 복합 인프라 개발 (디지털 디바이스 개발 및 보급 + 학교 시스템과 연계할 수 있는 클라우드 시스템 + 학생 데이터 수집 및 라벨링 기술 + 클라우드 상에서 누적된 데이터를 분석·활용할 수 있는 AI 모델 및 알고리즘)
- 포괄적인 AI 교육 서비스를 제공할 수 있는 기술 개발 (특정 과목이 아닌 학습자의 전체적인 학습 과정에서 지속적인 평가와 피드백을 통해 학습 목표 성취와 성장을 지속해서 돕는 'AI 튜터' 등)

라. 기술개발 로드맵

(1) 중기 기술개발 로드맵

[AI기반 교육 서비스 분야 중기 기술개발 로드맵]

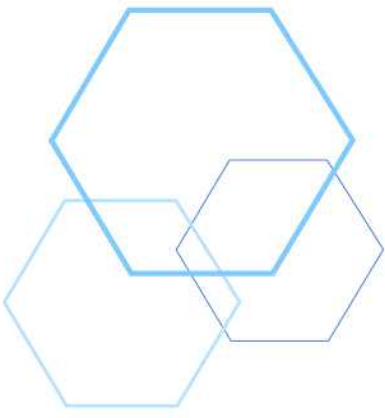
AI기반 교육 서비스	맞춤형 AI 교육 서비스 제공을 위한 관련 기술 개발과 서비스의 상용화			
	2021년	2022년	2023년	최종 목표
학습자 패턴 분석 및 평가 기술				학습자 문제풀이 패턴 분석 및 학습자의 현 수준 진단
학습 데이터베이스 구축·분석·운영 기술				교육용 AI데이터 분석·운용을 위한 고연산 분산 연산 시스템
맞춤형 학습 콘텐츠 추천 기술				학습자 답변에 따라 실시간 반응형으로 문항을 추천하는 기술
학습자 행동 측정 및 모델링				학습자의 행동을 분석하고 패턴을 분석하는 모델
문제 컨텍스트 추론 기술				이미지, 텍스트 문제의 의미를 해석하는 기술

(2) 기술개발 목표

- 최종 중소기업 기술로드맵은 기술/시장 니즈, 연차별 개발계획, 최종목표 등을 제시함으로써 중소기업의 기술개발 방향성을 제시

[AI 기반 교육서비스 분야 핵심기술 연구목표]

핵심기술	기술요구사항	연차별 개발목표			최종목표	연계R&D 유형
		1차년도	2차년도	3차년도		
학습자 패턴 분석 및 평가 기술	·학습자 데이터 수집 ·신규 사용자 분석 ·문제 풀이 패턴 분석 ·전체 학습자 수준 모델링	학습자 데이터 수집 기술과 인프라 구축	사용자 분석 및 문제 풀이 패턴 분석 기술 개발	전체 학습자 수준 모델 구축 및 패턴 분석 모델링	학습자 문제풀이 패턴 분석 및 학습자의 현 수준 진단	창업성장 기술개발 사업
학습 데이터베이스 구축·분석·운영 기술	·데이터 수집 플랫폼 ·학습용 데이터 라벨링 기술 ·고연산 처리를 위한 분산 연산시스템	공교육 부문 연동 학습데이터 운영 시스템 설계	학습데이터 운영 시스템 개발 학생데이터 수집 디바이스, 기술 개발	학습데이터 운영 시스템 개선	교육용 시데이터 분석·운용을 위한 고연산 분산 연산 시스템	상용화 기술개발 사업
맞춤형 학습 콘텐츠 추천 기술	·객관식 문항 데이터 분석 기술 ·학습자와 콘텐츠 수준값 매칭 데이터·알고리즘 ·사용자 답변 예측 프레임 워크	객관식 문항 데이터 분석 기술 개발	학습자와 콘텐츠의 매칭 알고리즘 개발	사용자 답변 예측 프레임 워크 구축	학습자 답변에 따라 실시간 반응형으로 문항을 추천하는 기술	창업성장 기술개발 사업
학습자 행동 측정 및 모델링	·학습자 행동 데이터 수집 기술 ·학습자 패턴 모델링 ·학습자 행동 AI 모델 구축·수정 기술	환경별 학습자 행동 측정 기술 개발	학습자 행동에 따른 패턴 모델링	환경변화에 따른 AI 모델 업데이트 기술	학습자의 행동을 분석하고 패턴을 분석하는 모델	산학연 Collabo R&D
문제 컨텍스트 추론 기술	·문제 이미지의 텍스트 인식 기술(OCR 등) ·문제 컨텍스트 추론 기술	이미지 및 텍스트 인식 기술 개발	문제 컨텍스트 추론 기술 개발	문제 컨텍스트 추론 기술 개선	이미지, 텍스트 문제의 의미를 해석하는 기술	산학연 Collabo R&D



전략제품 현황분석

AutoML

솔루션



AutoML 솔루션

정의 및 범위

- AutoML은 기계학습(Machine learning), 딥러닝(Deep learning) 등 인공지능 관련 모델 개발에 필요한 다양한 과정을 자동화하여 최적 성능의 모델을 개발하는, 일명 '인공지능을 만드는 인공지능' 솔루션

전략 제품 관련 동향

시장 현황 및 전망	제품 산업 특징
<ul style="list-style-type: none"> (세계) 2018년 1억 8,800만 달러 규모에서 연평균 43.7%로 성장하여 2024년 16억 5,200만 달러 규모로 성장 전망 (국내) 2018년 12억 3,000만 원 규모에서 연평균 27%로 성장하여 51억 7,000만 원 규모로 성장 전망 	<ul style="list-style-type: none"> AutoML 솔루션 분야는 오픈소스, 스타트업 및 글로벌 IT 대기업에 의해 주로 개발되는 추세 국내 주요 IT 대기업 중심으로 AutoML에 대한 투자 및 연구 진행
정책 동향	기술 동향
<ul style="list-style-type: none"> 정부 주도로 인공지능 국가전략('19.12) 등 인공지능 산업 육성을 위해 각종 정책 지원을 강화하는 추세 산업에의 활용 등 응용 개발 중점으로 지원과 함께 인공지능 기초 연구 강화 	<ul style="list-style-type: none"> 딥러닝을 활용한 AutoML 기술 개선 및 완전 자동화 개발 연구 중 AutoML의 활용성을 높이고 기술 한계를 극복하는 연구개발 필요
핵심 플레이어*	핵심기술
<ul style="list-style-type: none"> (해외) DataRobot, H2O.ai, Google, Amazon, Microsoft (대기업) 네이버, 카카오, 삼성SDS, LG CNS (중소기업) 위세아이텍, 에이아이더, 에이젠글로벌 	<ul style="list-style-type: none"> 데이터 처리 및 피처 엔지니어링 자동화 최적 하이퍼파라미터 및 신경망 아키텍처 탐색 (HPO, NAS 등) 모델 학습 및 평가 자동화

*생태계 취약 전략제품

중소기업 기술개발 전략

- 플랫폼을 구성하기 위해서는 직접 개발이 필요하기에 SAS와 같은 기업에서 제공하는 범용적 상용 플랫폼에 주목
- 응용 서비스만으로는 신산업성장 지속 가능성 확보가 불가능하기에 핵심 인프라 구축이 필요
- 중소기업 중심 데이터 경제 활성화를 위해 공유 클라우드 인프라 구축 및 데이터 생태계 개발

생태계 강화방안

- 머신러닝 알고리즘, 인지과학 등 AI 기초연구 강화 및 생태계 구축을 위해, 광주 AI 집적단지 등을 중심으로 인프라, 기술경쟁력, 규제혁신, 스타트업 지원 등을 강화 필요
- 초·중·고등학생 대상 SW 의무교육 등 인공지능 머신러닝 관련 분야에 대한 교육 강화를 통한 인공지능 전문 인력 양성 필요

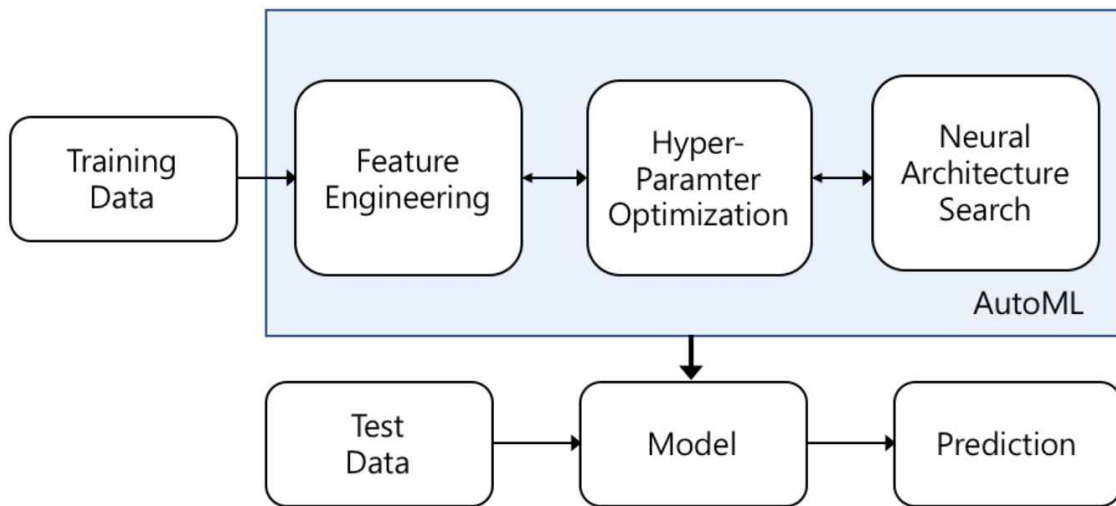
1. 개요

가. 정의 및 필요성

(1) 정의

- AutoML은 기계학습(Machine learning), 딥러닝(Deep learning) 등 인공지능 관련 모델 개발에 필요한 다양한 과정을 자동화하여 최적 성능의 모델을 개발하는, 일명 ‘인공지능을 만드는 인공지능’ 솔루션
 - 인공지능 스스로 데이터의 특징을 추출하고 적절한 알고리즘을 선택하여 학습을 수행한 이후, 학습된 모델 간 평가를 통해 개발 목표에 가장 적합한 인공지능 모델을 탐색하는 과정을 자동화
- 데이터 설계, 데이터 전처리, 데이터 특징 파악, 데이터 특성 엔지니어링(피처 엔지니어링), 적합한 알고리즘 선택, 하이퍼 파라미터 최적화, 모델 평가 등 기존 기계학습 모델 개발 과정을 자동화하는 기술로 구성
 - 기존에 인공지능 개발자가 반복적으로 수행하는 기계학습, 딥러닝 등 모델 개발 작업을 상당수 자동화함으로써, 해당 데이터와 모델에 관한 전문 지식이 없는 비-전문가도 인공지능 활용도를 높이는 것을 목표로 함

[자동 기계학습(AutoML) 시스템]



* 출처: 자동 기계학습(AutoML) 기술 동향, 한국전자통신연구원(2019)

(2) 필요성

- 인공지능 모델 개발을 위한 학습용 데이터 자동 설계 필요
 - 도메인 지식에 기반한 데이터의 특징을 파악하기 위해서는 해당 전문 분야에 대한 깊은 이해가 필요함
 - 데이터 샘플 무작위 추출, 결측값 보정, 중복값 제거, 정규화, 훈련용 및 평가용 데이터 구분 등 학습용 데이터 수집 및 전처리 과정에서 수반되는 일련의 과정들을 자동으로 수행함으로써 인공지능 모델 개발의 효율성 증대

- 최적 성능의 모델을 학습하기 위한 하이퍼 파라미터 최적화 및 알고리즘 선정 필요
 - 하이퍼 파라미터는 학습용 데이터를 모델에 훈련시키는 과정을 제어하는 변수로 학습률, 손실함수, 훈련 반복 횟수, 가중치, 신경망 계층 수 등 인공지능 모델의 성능을 결정하는 입력값이며, 기존에는 개발자가 변수 조합을 설정하고 반복적으로 모델 학습을 수행 및 평가하는 방식으로 가장 높은 성능을 보이는 모델을 탐색
 - 개발자에 의해 반복적으로 수행하는 작업 과정에서 발생할 수 있는 오류를 방지하고, 자동화된 프로세스를 통해 일관성 있는 모델 개발 필요
 - 데이터에 적합한 후보 알고리즘 간 모델 학습 방식, 내부 동작 및 모델 학습에 필요한 연산자원 등 비교 및 평가를 통한 알고리즘 선정 필요

- 인공지능 모델 개발 비용 증가에 따른 자동 최적화 기술 필요
 - 머신러닝 및 딥러닝 모델 개발시 반복적 모델 학습에 따른 컴퓨팅 연산 비용의 증가 문제로 인해, 효율성 높은 인공지능 모델 개발을 위한 자동화 프로세스 필요성 증가
 - 인공지능 기반 기술 개발과 서비스 출시로 인해 머신러닝 및 딥러닝 연구개발이 가능한 데이터 사이언티스트 등 인력 부족에 따른 비-전문가에 의한 개발 수요 증가

나. 범위 및 분류

(1) 가치사슬

- (가치사슬) AutoML 솔루션의 후방산업으로는 인공지능 분야의 데이터, 모델 알고리즘 자동화 구현, 병렬 분산처리 기술, 클라우드 플랫폼 등을 포함
 - (후방산업) 머신러닝 및 딥러닝 등 AI 분야의 기술 향상을 위한 데이터, 모델 등 알고리즘 개발과 기 개발된 알고리즘의 자동화 구현 기술, 머신러닝 및 딥러닝의 효율적 연산을 위한 병렬 분산처리 기술, 머신러닝 개발에서 서비스 배포까지 포함하는 클라우드 플랫폼 등을 포함
 - AutoML에 특화된 알고리즘과 머신러닝 및 딥러닝 파이프라인 자동화 기술이 확보되어야 함
 - (전방산업) 금융, 의료, 헬스케어, 소매업, 제조업, 교통, 자율주행 등 인공지능이 활용될 수 있는 전 분야를 포함
 - 머신러닝 및 딥러닝을 활용하여 서비스를 제공하고 있는 경우, AutoML 솔루션 적용이 가능함

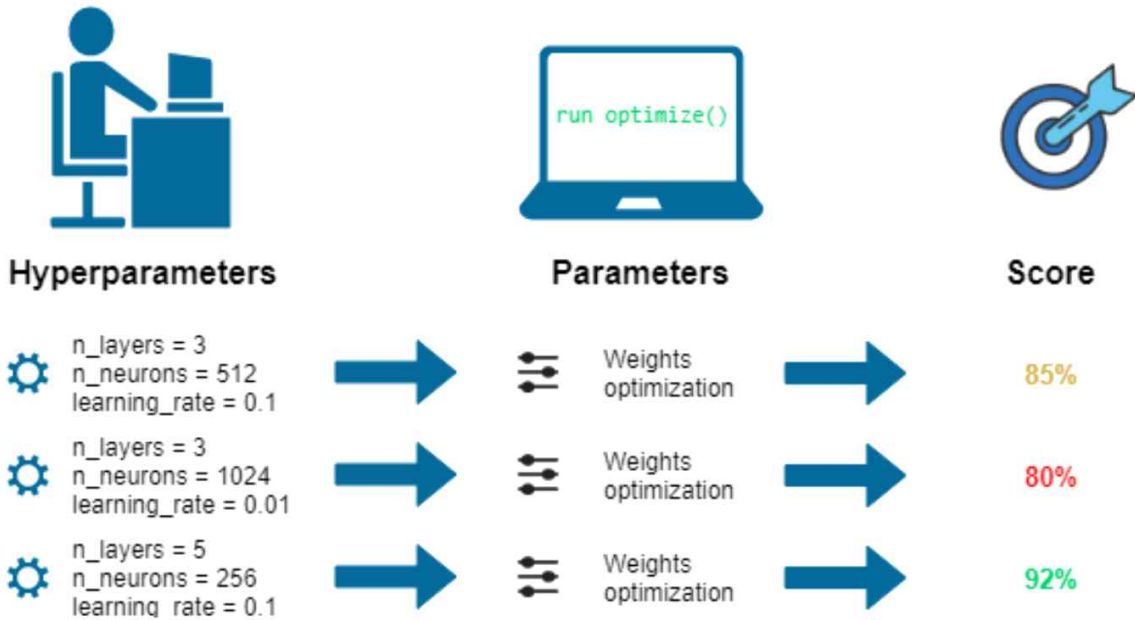
[AUTOML 분야 산업구조]

후방산업	AUTOML	전방산업
인공지능 알고리즘 개발, 파이프라인 자동화 기술, 병렬 분산처리 기술, 클라우드 플랫폼 등	데이터 자동 설계, 데이터 자동 전처리, 데이터 피처 엔지니어링, 적합한 알고리즘 선택, 하이퍼 파라미터 최적화, 모델 평가 등	금융, 의료, 헬스케어, 소매업, 제조업, 자율주행 등 인공지능 활용 전 분야

(2) 용도별 분류

- AutoML 분야는 크게 데이터 처리 및 피처 엔지니어링 자동화, 최적 하이퍼파라미터 및 신경망 아키텍처 탐색, 모델 학습 및 평가 자동화로 분류 가능
- (데이터 처리 및 피처 엔지니어링 자동화) 문제 해결에 적합한 모델에 맞춰 데이터 설계, 수집 및 전처리 과정과, 최적의 변수 집합을 찾기 위해 데이터 변환(차원 축소, 정규화, 데이터 인코딩 등)을 자동으로 수행
 - AutoML 모델 생성을 위해 사용자가 원시 데이터를 입력하면 AutoML 솔루션이 이를 자동으로 분석하여 데이터 유형, 데이터 구조, 결측값 등을 보완
 - 기존 인공지능 개발자가 일일이 원시 데이터를 확인하고 반복적으로 수행하는 작업을 일괄적으로 처리함으로써 오류 절감 및 일관성 향상 효과를 기대할 수 있음
 - 피처 엔지니어링은 원시 데이터의 특징을 파악하고 유의미한 변수만을 추출하여 머신러닝 및 딥러닝 모델에 적합한 형태로 변환함으로써 인공지능 모델의 설명력을 높이는 과정
 - 주 성분 분석(PCA) 등 차원 축소를 통해 변수 간 상관성이 높은 경우를 고려할 수 있으며, 정규화 및 표준화를 통해 특정 변수가 벡터 공간에서 다른 변수의 설명력을 제외시키는 문제를 해결할 수 있음
- (최적 하이퍼파라미터 및 모델 아키텍처 탐색) 머신러닝 및 딥러닝 모델 학습에 영향을 미치는 하이퍼파라미터 최적화(Hyperparameter Optimization, HPO)와 신경망 아키텍처 자동 탐색(Neural Architecture Search, NAS) 등을 통해 더 나은 성능의 모델을 학습할 수 있음
 - 하이퍼파라미터는 훈련 데이터를 기반으로 모델 학습에 영향을 주는 입력변수
 - 하이퍼파라미터 탐색에는 그리드 서치(Grid search), 랜덤 서치(Random search), 베이지안 최적화(Bayesian optimization) 등이 활용
 - 기존 하이퍼파라미터 탐색 방식은 인공지능 개발자가 학습률(Learning rate), 에폭(Epoch), 배치 사이즈(Batch size) 등 개별 하이퍼파라미터들의 입력값을 변경하며 가장 성능이 좋은 하이퍼파라미터를 선정하는 과정으로 수행
 - 베이지안 최적화가 시간 대비 성능이 우수하여 베이지안 최적화 속도 향상과 관련한 연구가 주로 진행되는 추세
 - 하이퍼파라미터 탐색은 모델 훈련 설정값에 대한 것이며, 신경망 아키텍처 탐색은 신경망 모델 구조 자체의 변경을 통한 최적 모델 탐색을 의미
 - 신경망 아키텍처 탐색은 기존 경험적으로 이루어진 훈련 데이터를 기반으로 성능을 가장 높이기 위한 신경망의 구조, 가중치 등 탐색 과정을 자동화

[Hyperparameter Optimization 예시]



* 출처: Practical Hyperparameter Optimization, KDnuggets

- (모델 선택 및 평가 자동화) 다양한 머신러닝 및 딥러닝 모델을 학습하고 각 모델별 성능을 측정하여 검증하는 단계를 자동으로 수행
 - 정형 및 비정형 데이터 형식, 개발 목적 등에 따라 다양한 모델과 알고리즘을 자동으로 선택하여 훈련 데이터를 기반으로 모델 학습을 수행하며, 앞선 하이퍼파라미터 탐색, 신경망 아키텍처 탐색과 병행
 - 학습된 모델별로 예측력, 인식력 등 종합적인 성능을 자동으로 측정하고 이를 평가하여, 서비스 배포까지 자동화하는 단계를 포함

2. 산업 및 시장 분석

가. 산업 분석

◎ 오픈소스, 스타트업 및 글로벌 IT 대기업에 의해 주로 개발되는 추세

- AutoML 등 인공지능 관련 분야는 연구개발 과정 및 결과물이 공유되는 특성
 - 머신러닝 및 딥러닝 관련 대부분 알고리즘은 파이썬(Python), R 등 데이터 분석 프로그래밍 언어에서 바로 실행 가능한 형태로 공유
 - AutoML 관련 기술의 경우도 머신러닝과 딥러닝 기초 알고리즘의 연구개발에 맞춰서 함께 발전되고 있으므로, 대부분의 AutoML 알고리즘은 오픈소스로 개발되고 있음
 - 그러나, 이렇게 공유된 결과물은 소스코드 형태이므로 데이터 분석 프로그래밍 언어에 대한 지식을 요구하기 때문에 비-전문가 또는 일반인에게는 진입 장벽이 존재함
- 특정 산업 및 문제에 대응하는 AutoML 솔루션은 스타트업에 의해 공급되는 추세
 - 스타트업에서 개발된 솔루션은 특정 산업이나 문제를 해결하기 위해 개발되는 경우가 증가하고 있으며, 시각화 및 인터페이스 측면에서 우수한 경우가 많음
 - 이는 AutoML 등 머신러닝 모델에 지식이 없지만, 도메인 지식을 보유한 기업 담당자에게 있어 유용하게 사용될 수 있기 위함
 - 스타트업에서 제공하는 AutoML 솔루션은 새롭게 AutoML 알고리즘을 개발하기보다 주로 인공지능, AutoML 관련 오픈소스를 기반으로 일부 특화하여 제공하는 경우가 많음
- 인공지능 및 AutoML 관련 기초 기술 개발에는 대학 및 글로벌 IT 대기업에 의해 주도
 - 주로 대학, 글로벌 IT 대기업(Facebook, Amazon, Google 등)의 연구자들에 의해 기초 인공지능 알고리즘이 연구되고, 연구 결과가 오픈소스 커뮤니티에 공유되는 경우가 많음
 - 글로벌 IT 대기업 등에서 개발된 인공지능 및 AutoML 관련 알고리즘은 독점적으로 서비스되는 경우보다 자사 플랫폼의 활용도를 높이는 방향으로 시장 점유율 확대 전략을 수행 중

◎ 국내 인공지능 연구개발에 대한 정책 지원 강화 추세

- 국내는 정부 주도로 인공지능 국가전략('19.12) 등 인공지능 산업 육성을 위해 각종 정책 지원을 강화하는 추세
 - 정부는 인공지능 관련 3대 분야 9대 전략, 100대 실행과제 제시 등 범정부 역량을 총 동원하여 AI 시대 국가 미래 비전과 전략을 수행하기 위한 집중 지원을 수행하는 중
 - 인공지능을 통해 2030년까지 경제효과 최대 455조원 창출, 삶의 질 세계 10위 도약 등 산업과 사회 모든 영역에 걸쳐 패러다임 변환을 도모하고자 함
 - 인공지능 전문 인력 양성, 초중등학생 대상 SW 의무교육 등 인공지능 관련 분야에 대한 교육 강화와 함께 데이터 뉴딜 추진 등 인공지능 산업을 통한 일자리 확보를 중점으로 추진 중
- 산업에의 활용 등 응용 개발 중점으로 지원과 함께 인공지능 기초 연구 강화
 - 인공지능 생태계 구축을 위해, 광주 AI 집적단지 등을 중심으로 인프라, 기술 경쟁력, 규제혁신, 스타트업 지원 등을 강화
 - 정부는 2030년까지 인공지능 기반 스마트 공장 3000개를 보급하고, 제조 및 중소기업에서 인공지능 활용에 기반한 혁신 주도를 지원할 계획
 - 공공데이터 개방, AI 반도체 핵심기술 확보 등 산업 파급효과가 높은 부문을 중심으로 지원
 - 머신러닝 알고리즘, 인지과학 등 AI 기초연구 강화를 위해 차세대 AI 개발사업 예타 추진('20)으로, 2030년까지 핵심기술 5개 확보를 목표로 정책 지원
 - AutoML 분야 기술의 경우, 머신러닝 및 딥러닝 등 인공지능 분야에 포함되므로 향후 해당 기술에 대한 정부 정책 지원을 기대할 수 있을 것으로 전망

◎ 국내 주요 IT 대기업 중심으로 AutoML에 대한 투자 및 연구 진행

- 인공지능 연구그룹을 중심으로 인공지능 및 AutoML 관련 연구 진행
 - 카카오는 카카오브레인을 통해 AutoML 관련 연구를 진행하고 있으며 '카카오 i 리서치 프로그램'에 박사후 연구원 지원 프로그램을 신설하고 머신러닝, 딥러닝, 자연어 처리, AutoML 등 총 12개 분야에 걸쳐 연구를 지원
 - 네이버는 클로바(Clova ai)를 통해 딥러닝 알고리즘 설계 및 개발을 위한 CLOVA NSML 머신러닝 플랫폼에서 관련한 AutoML 연구를 수행
- 주요 IT 대기업 AI 솔루션에 AutoML을 활용 증가
 - 삼성SDS는 자사 'Brightics AI' 솔루션에 AutoML 기술을 활용하여 데이터 특성에 맞는 최적 알고리즘 자동 추천, 복수 파라미터 자동 분석 시뮬레이션 환경 등을 제공
 - LG CNS는 자사 AI빅데이터 플랫폼 'DAP' 솔루션과 구글 클라우드의 AutoML을 활용하여 LG 계열사 제조공장의 불량 판정 AI모델을 자동 생성한 사례를 통해 AI 모델 제작 기간을 1주일에서 최대 2시간으로 단축하고, 불량 판정률은 평균 6% 상승시킴

나. 시장 분석

(1) 세계시장

- PRNewswire의 2020년 분석 결과에 따르면 AutoML 솔루션 세계시장은 클라우드 플랫폼 증가세와 맞물려 2018년 기준 1억 880만 달러에서 연평균 43.7%로 성장하여 2024년 16억 5,240만 달러로 성장할 것으로 전망

[AutoML 솔루션 세계시장 규모 및 전망]

(단위 : 백만 달러, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
AutoML 솔루션 (세계)	188	270	387	557	800	1,150	1,652	43.7

* 출처: Automated Machine Learning (AutoML) Market, PRNewswire(2020)

- AutoML 솔루션 세계시장은 인공지능에 대한 폭발적인 수요에 의해 2020년부터 2030년까지 연평균 43.7% 증가할 것으로 예상
- 2019년 기준 미국이 가장 큰 규모이며, 아시아 태평양 지역이 2020년부터 2030년까지 가장 큰 증가세를 보일 것으로 전망
- 미국은 2020년부터 2030년까지 연평균 42.3% 성장하여, 2030년 기준 41억 6370만 달러 규모로 전망되며, 독일은 동기간 연평균 44.0% 성장하여 2030년 기준 약 12억 4580만 달러 규모로 전망

(2) 국내시장

- AutoML 솔루션 국내시장은 매우 작아서 2018년 기준 12.3억원 정도로 추정되며 2024년 51.7억원 정도로 연평균 27%로 성장할 것으로 전망

[AutoML 솔루션 세계시장 규모 및 전망]

(단위 : 억 원, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
AutoML 솔루션 (국내)	12.3	15.6	19.8	25.2	32.0	40.7	51.7	27

* 출처: Automated Machine Learning (AutoML) Market, PRNewswire(2020)을 바탕으로 세계 AI 시장과 국내 AI 시장을 복합적으로 고려하여 네모아이씨지 재추정

3. 기술 개발 동향

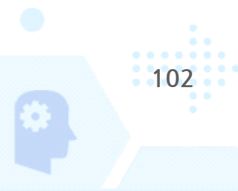
- 기술경쟁력
 - AutoML 솔루션은 미국이 최고기술국으로 평가되었으며, 우리나라는 최고기술국 대비 66.1%의 기술 수준을 보유하고 있으며, 최고기술국과의 기술격차는 2.6년으로 분석
 - 중소기업의 기술경쟁력은 최고기술국 대비 56.4%, 기술격차는 3.7년으로 평가
 - EU(79.5%)>중국(70.2%)>일본(67.3%)>한국의 순으로 평가
- 기술수명주기(TCT)⁹⁾
 - AutoML 솔루션은 4.63의 기술수명주기를 지닌 것으로 파악

가. 기술개발 이슈

◎ 딥러닝을 활용한 AutoML 기술 개선 및 완전 자동화 개발

- 딥러닝을 이용한 AutoML 기술 개선이 제안되고 있으며, 이를 반영한 기술 개발이 연구 중
 - AutoML 기술은 도메인 특징 추출, 하이퍼파라미터 최적화 및 신경망 아키텍처 탐색과 같은 주요 프로세스를 자동화하는데 이용되고 있음
 - 딥러닝 기술의 발전과 함께 퓨샷러닝(Few shot learning), 강화학습(Reinforcement learning) 등을 이용한 신경망 아키텍처 탐색(NAS) 기술이 활발히 연구되고 있음
 - 딥러닝 기반 신경망 아키텍처 탐색 기술은 딥러닝 모델의 최적 구조 및 하이퍼파라미터 탐색으로 이루어지는데 수십억 개의 파라미터, 모델 개발에 사용되는 데이터 셋의 크기, 다양한 딥러닝 모델 구조를 고려할 때 매우 복잡한 문제임
 - AutoDL이라는 대회가 개최되어, 합리적인 리소스 환경에서 다양한 종류의 원시 데이터(예를 들어 텍스트, 이미지, 시계열 등)에 바로 동작하는 다양한 딥러닝 기반 AutoML 방법론들이 제안되었음
- 사람의 개입을 최소화 하는 AutoML 기술이 지속적으로 연구 개발되고 있음
 - 구글은 전체 ML 알고리즘 설계 과정에서 사람의 개입을 최소화하는 방향으로 연구, AutoML-Zero를 발표하여 진화 알고리즘을 이용해 인간에 의한 편향이 배제된 신경망 구조 개발에 성공
 - 대부분의 AutoML은 지도학습(Supervised learning)에 집중되어 있으나 비지도 및 준지도학습(Unsupervised and Semi-supervised learning)에 대한 연구도 시작되고 있음

9) 기술수명주기(TCT, Technical Cycle Time): 특허 출원연도와 인용한 특허들의 출원연도 차이의 중앙값을 통해 기술 변화속도 및 기술의 경제적 수명을 예측



◎ AutoML의 활용성을 높이는 기술 개발 필요

□ 설명 가능성 및 해석 가능성을 포괄한 AutoML 모델 개발 필요

- 현재 개발되어 있는 AutoML 솔루션은 최적 모델 생성에는 다소 효과적이지만, 어떠한 메커니즘에 의해 모델이 생성되었는지 알수 없는 ‘블랙박스’가 존재
- 누구나 이용가능할 수 있도록 설명 가능성과 해석 가능성 메커니즘을 가진 투명한 AutoML 방법론 개발이 필요
- 관련 연구는 아직 미진하지만 AutoML 자체가 검색 집중 과정(Search intensive procedure)을 통해 방대한 양의 정보를 생성하고, 이것들이 투명한 AutoML 솔루션을 만드는데 이용될 수 있기 때문에 투명한 AutoML 방법론 개발은 요원하지 않은 것으로 전망

□ 서로 비교할 수 있고 재현이 가능한 AutoML 기술 개발 필요

- AutoML은 데이터를 포함하는 최적화 과정이기 때문에 개발된 방법론들을 서로 비교할 수 있는 플랫폼과 프레임워크에 대한 개발도 함께 이루어져야 함
- 일부 실험적인 모델들에서 그러한 플랫폼을 제공하고 있으나 이들의 발전 속도는 AutoML 기술 자체의 발전 속도에 비해 느린 한계점 존재
- 아울러, AutoML로 개발된 모델의 재현성을 촉진하는 코드 공유 및 메커니즘에 대한 필요성도 대두되고 있음

□ 모델 성능 향상을 위해 사용자의 도메인 지식을 고려하는 AutoML 플랫폼 필요

- 현재 AutoML 솔루션은 피처 엔지니어링, 모델 선택 등의 과정에서 알고리즘적 접근만 가능하며, 해당 분야의 도메인 지식이 고려되지 않음
- AutoML의 최종 목표는 완전 자동 프로세스이지만, 현재 상태에서는 사전 지식을 포함하는 메커니즘을 통해 AutoML 모델의 성능을 향상시킬 수 있을 것으로 전망

◎ AutoML의 기술 한계를 극복하는 연구개발 필요

□ AutoML에서 자동 피처 엔지니어링 모델의 강화 및 여러 유형의 데이터 자동 처리 기술 개발 필요

- 지금까지의 AutoML은 대부분 이미 만들어진 특징들 중 최적의 특징 셋을 뽑거나 추출하는 것에 집중되었지만 완전 자동화를 위해 원시 데이터를 바로 처리할 수 있는 피처 엔지니어링 과정 필요
- 원시 데이터를 직접 가공하여 특징을 파악하는 연구들이 수행되고 있으므로, 앞으로 자동 피처 엔지니어링 분야의 방법론 개발이 더 강화될 것으로 전망
- 더 나아가, 텍스트나 영상과 같은 원시 데이터, 또는 그래프나 네트워크와 같은 구조화된 데이터를 처리할 수 있는 AutoML 기술 필요

거대 스케일 문제에 대응할 수 있는 AutoML 솔루션 필요

- 기존 AutoML 솔루션은 거대 규모 문제를 해결하기에는 최적화 단계에서의 연산 비용이 기하급수적으로 증가하는 문제 존재
- AutoML 관련 연구 추세에는 검색 집약적 AutoML 프로세스를 수행하는 솔루션이 거의 없었으며, 따라서 거대 규모 문제를 해결할 수 있는 AutoML 솔루션이 필요
- AutoML의 주요한 과정인 신경망 아키텍처 탐색(NAS)에서 효율성을 강화하기 위해, 딥러닝 모델을 위한 효율적인 AutoML 기술 개발이 필요

전이학습을 이용한 AutoML 성능 향상 필요

- 현재 AutoML 솔루션을 통해 광대하고 풍부한 정보들이 생성되고 있는데, 이 정보들을 이용하여 AutoML 성능을 향상시킬 수 있는 전이학습이 가능할 것으로 전망
- 이미 연구되고 있는 메타 러닝과 다르게, 최적화 프로세스의 역학관계에 대한 정보를 하나의 작업에서 다른 작업으로 전이하는 방식으로 이루어질 수 있음
- 구글 클라우드 AutoML은 데이터에서 모델을 학습하는 대신 전이학습을 이용한 AutoML을 지원하는 방식이며, 특화된 분야에서 더 우수한 성능을 보이는 경우가 많음



나. 생태계 기술 동향

(1) 해외 플레이어 동향

- (Google) 구글 클라우드 플랫폼에서 구조화된 데이터, 시각, 언어 등으로 분야를 구분하여 심층 전이학습(Transfer learning)을 활용한 종합적인 Cloud AutoML 솔루션을 제공
 - 구조화된 데이터 분야에는 AutoML Tables 솔루션을 통해 원시 데이터 파악, 피처 엔지니어링, 다양한 모델에 대한 학습과 앙상블 모델 구성, 모델 평가 및 최종 배포 모델 생성에 대한 API를 제공하여 사용자가 구글 클라우드 플랫폼 내에서 자체 커스텀 머신러닝 모델을 생성할 수 있도록 지원
 - 구글 클라우드 플랫폼에서 시각분야에 특화되어 있는 솔루션으로는 AutoML Vision, AutoML Video Intelligence 등이 있으며, AutoML Vision은 에지(Edge) 단에서 이미지 분류, 객체 감지, 객체 인식 등을 지원하며 AutoML Video Intelligence는 동영상 장면과 세그먼트 분류, 객체 추적 등을 지원
 - 언어분야에는 AutoML Natural Language를 통해 문서의 구조와 의미를 자동으로 분석하여, 분류, 항목 추출 및 감정 분석 등을 자동으로 수행할 수 있도록 지원하고 있으며 AutoML Translation을 통해 다국어 번역을 위한 모델을 자동으로 생성할 수 있도록 지원
- (H2O.ai) 피처 엔지니어링, 모델 생성 및 평가 등의 과정을 자동화하고 인공지능 모델의 결과를 시각화를 통해 이해하기 쉽게 제공하는 H2O Driverless AI 솔루션을 제공
 - H2O Driverless AI는 비-전문가를 대상으로 쉽고 직관적인 머신러닝 개발 인터페이스를 제공하여, 타 머신러닝 클라우드 플랫폼 경쟁사 대비 가장 우수한 시각화 솔루션으로 평가받고 있음 (Gartner, 2018)

[H2O Driverless AI]

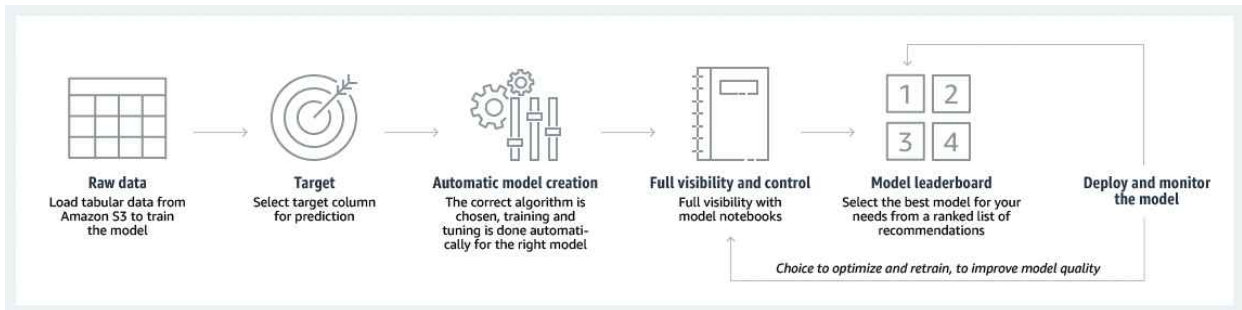


* 출처: Gartner, H2O.ai



- (Amazon) Amazon Web Service 플랫폼 내에서 AutoML을 지원하는 Amazon SageMaker AutoPilot 서비스를 제공함으로써 사용자가 데이터 전처리, 학습 알고리즘, 하이퍼파라미터 최적 튜닝의 자동화 과정을 제공

[Amazon SageMaker AutoPilot]



* 출처: Amazon

- 해외 기업 중에서 DataRobot, H2O.ai, dotData가 AutoML 솔루션 개발 분야에서 가장 우수하다고 평가 받고 있음
 - DataRobot과 H2O.ai가 AutoML 분야에서 가장 두각을 나타내고 있으며, dotData는 2018년에 설립되어 시장 점유율은 낮으나 다크호스로 전망
- DataRobot, H2O.ai, Dataiku 순서로 투자를 많이 받은 것으로 나타남
 - DataRobot은 약 4억 3천만 달러, H2O.ai는 약 1억 5천만 달러 수준으로 투자를 받았는데, DataRobot은 수천명 규모인 반면 H2O.ai는 수십명 규모의 특징이 존재함

[AutoML 분야의 주요 업체간 경쟁력 평가 및 기업 현황]

THE FORRESTER NEW WAVE™
Automation-Focused Machine Learning Solutions
Q2 2019

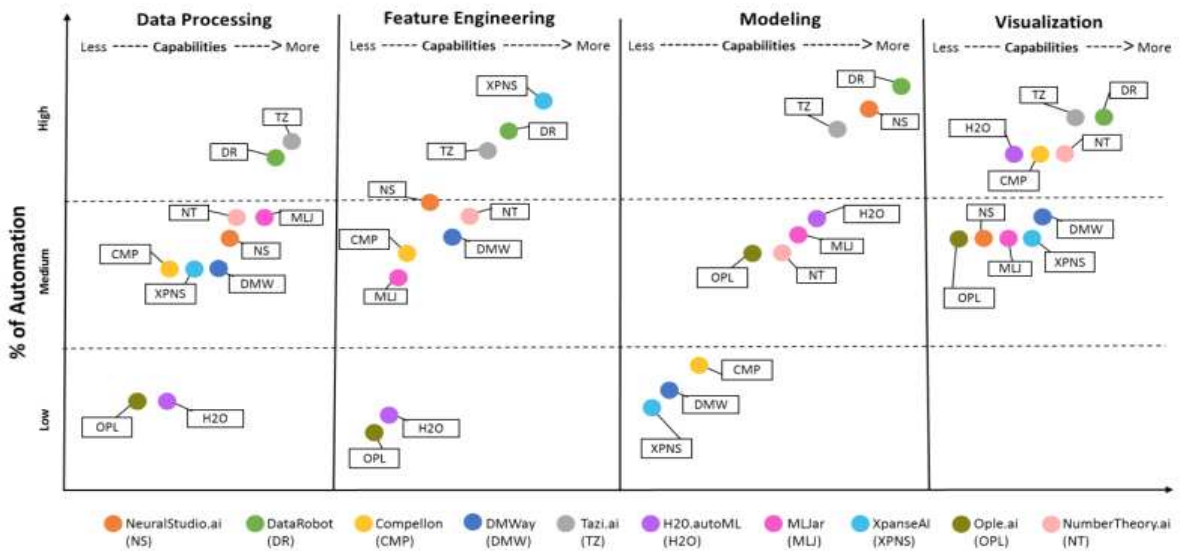


Company	Total Funding Amount	Founded Year	Number of Employees
DataRobot	\$490.6M	2012	1001-5000
H2O.ai	\$151.1M	2012	11-50
Dataiku	\$146.8M	2013	251-500
dotData	\$49M	2018	51-100
Compellon (Acquired by Hoist Finance)	\$11.6M	2010	11-50
Coldlight Solutions (Acquired by PTC)	\$11M	2007	1-10
PurePredictive	\$10.2M	2011	11-50
Ople	\$10M	2017	11-50
Predikto (Acquired by United Technologies)	\$7.6M	2013	11-50
VEDA Data Solutions	\$7.2M	2015	1-10
Snark AI	\$1.7M	2018	1-10
MyDataModels	\$1M	2018	11-50
Tazi.ai	\$1M	2015	11-50
DMWay	\$1M	2013	1-10

* 출처: Forrester Research, Crunchbase

- AutoML 프로세스별로 업체간 자동화 수준과 역량 비교에서, DataRobot, Tazi.ai 등이 대체적으로 우수한 것으로 나타남
 - AutoML 프로세싱 전 단계에서 Tazi.ai, DataRobot가 대체로 역량이 높고 자동화 수준이 높다고 평가되었으며 XpanseAI와 NeuralStudio.ai는 각각 피처 엔지니어링과 모델링에서 두각을 나타냄
 - 데이터 프로세싱 단계와 피처 엔지니어링 단계에서 H2O.autoML과 Ople.ai는 중하위 수준의 역량과 낮은 수준의 자동화 수준으로 평가되었으며, XpanseAI의 경우 모델링 역량과 자동화 수준은 타 경쟁 기업에 비해 가장 낮다고 평가됨

[AutoML 프로세스 별 주요 업체간 역량 및 자동화 수준 비교]



* 출처: Targetbase

(2) 국내 플레이어 동향

- (위세아이텍) WiseProphet 플랫폼을 통해 머신러닝 자동화 프로세스 서비스를 제공
 - 데이터 전처리, 피처 엔지니어링, 주요 머신러닝 및 딥러닝 알고리즘 지원, 하이퍼파라미터 및 앙상블 모델 지원 등 주요 머신러닝 프로세스를 자동화
 - 예측정비, 부당청구 및 이상거래 탐지 등 참조할 수 있는 주요 산업별 레퍼런스 모델을 지원

[WiseProphet]



* 출처: 위세아이텍

- (Aithe) Smart AutoML 솔루션을 통해 데이터 전처리와 모델 튜닝 등 자동화된 머신러닝 분석 솔루션을 제공
 - R, Python 등 프로그래밍 코딩 없이 머신러닝 분석을 제공하여 비-전문가의 접근성 개선
- (에이젠글로벌)은 금융 빅데이터를 기반으로 AutoML이 적용된 아바커스(ABACUS)를 통해 금융 의사결정 과정 지원 솔루션을 제공
 - 금융에 특화된 딥러닝 자동화 모델을 통해 금융 상품 설계, 신용평가, 이상거래탐지 등 금융 고객의 개인화된 서비스 구축 등에 활용

다. 국내 연구개발 기관 및 동향

(1) 연구개발 기관

[AutoML 솔루션 분야 주요 연구조직 현황]

기관	소속	연구 분야
단국대학교	단국대학교 SW융합대학 소프트웨어학과	• 머신러닝, 빅데이터, 안과, 진단, 플랫폼개발
과학기술정보통신부	한국과학기술정보연구원	• 개방형 플랫폼, 소재, 연구데이터, 고성능컴퓨팅, 인공지능, 기계학습, 웹 기반 계산 도구, 웹 기반 데이터 분석 환경
과학기술정보통신부	한국과학기술원 KAIST 산업 및 시스템공학과	• 다변수 시계열, 빅데이터, 딥러닝, 머신러닝, AutoML, 신경망 구조 탐색, 하이퍼파라미터 최적화
한양대학교	한양대학교 전기·생체공학부	• 자율주행, 인공지능, 센서융합, 딥러닝, 강화학습, 능동학습

(2) 기관 기술개발 동향

- (단국대학교) 단국대학교는 빅데이터 기반 인공지능 안과 진단기술 및 스마트 진료 플랫폼을 개발을 통해 각막 질환에 대한 진료의의 질환 진단 정확도 향상효과를 기대(2018.04.01. ~ 2022.12.31.)
- 영상처리 기법을 이용한 각막궤양 영역 검출
 - AutoML 기술을 적용하여 녹내장 모델 고도화 및 최적 모델 탐색
 - 인공지능 영상인식 알고리즘을 이용한 각막궤양 영역 검출
 - 당뇨망막병증 모델 고도화

- (한국과학기술정보연구원) 소재연구데이터 기반의 미래소재 발굴 연구환경 구축을 위한 통합플랫폼 기반기술 개발 및 시범 서비스 구축(2020.07.01. ~ 2020.12.31.)
 - 소재연구데이터 연계 활용이 가능한 기반기술 및 통합플랫폼 시범 개발
 - 소재연구데이터에 적용 가능하며 머신러닝 기법들에 대한 최적화된 AutoML 및 기계학습 인프라 기반기술 개발
 - 국내 화합물 반도체 소재연구데이터 수집·저장·공유 및 AI기반 예측기술 개발
 - 시범서비스 가능한 컴퓨팅 자원 및 저장장치 확보

- (한국과학기술원) 다변수 스트림 빅데이터의 딥러닝 분석을 위한 AutoML 프레임워크 개발을 통해 스마트 팩토리에서 생산된 방대한 양의 데이터에 무한한 가치 부여 가능성 기대(2020.09.01. ~ 2023.02.28.)
 - 데이터 준비 자동화: 데이터 클리닝, 특성(feature) 전처리, 특성 구성, 특성 선택
 - 모델 학습: 모델 선택(신경망 구조 탐색), 하이퍼파라미터 최적화 기술이 준비
 - 회귀분석과 이상치 검출에 대한 Stream AutoML 기술을 개발

- (한양대학교) 레벨4 도심자율주행 실현을 위한 인공지능 핵심기술 연구
 - 도심자율주행의 위험요소 파악을 위한 360도 주변에 대한 3차원 물체검출 및 인식 기술 개발
 - 도로 위의 시공간적 정보와 관계 및 움직임 등의 문맥을 활용하여 인지 및 예측을 하는 기술 연구
 - 딥러닝 구조를 하드웨어로 구현시 효율적으로 구현하는 멀티태스크 학습 구조 및 AutoML 기반의 모델 압축에 대한 연구 수행

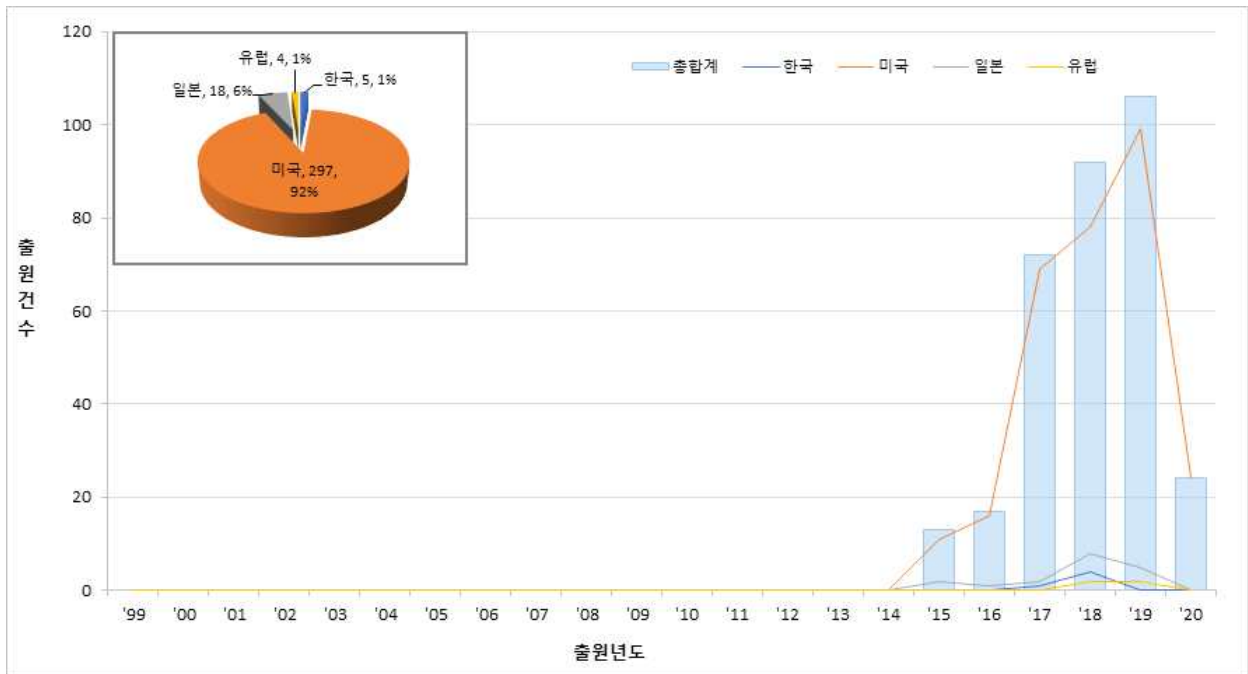
4. 특허 동향

가. 특허동향 분석

(1) 연도별 출원동향

- AutoML 솔루션은 '15년부터 급격한 성장을 보임
 - 각 국가별로 살펴보면 미국이 가장 활발한 출원활동을 보이고 있음
- 국가별 출원비중을 살펴보면 미국이 전체의 92%의 출원 비중을 차지하고 있어, 최대 출원국으로 AutoML 솔루션 분야를 리드하고 있는 것으로 나타났으며, 일본은 6%, 한국은 2%, 유럽은 1% 순으로 나타남

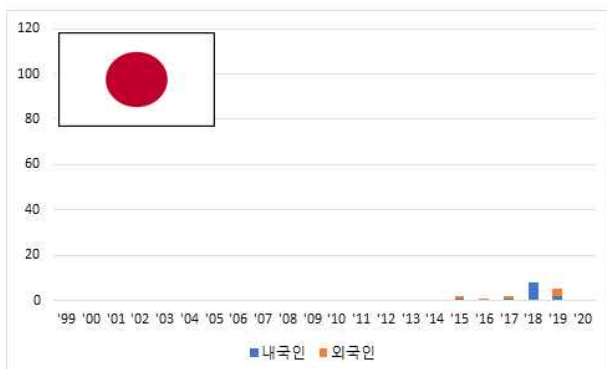
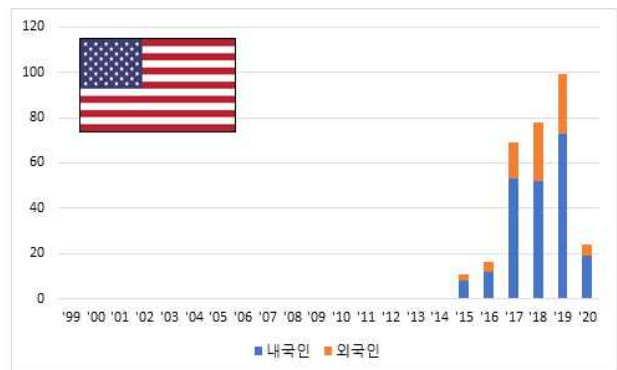
[AutoML 솔루션 연도별 출원동향]



(2) 국가별 출원현황

- 한국의 출원현황을 살펴보면, '17년부터 해당 기술의 출원이 완만히 증가하는 추세
 - 외국인 위주의 출원이 진행되고 있음
 - 한국 기술의 양적 흐름은 미국과 상당히 유사
 - 미국의 출원 수에 비해 2% 정도의 수준을 보임
- 미국의 출원현황을 살펴보면 분석구간 초기부터 전체 특허기술의 출원 증감 흐름에 영향을 주고 있는 것으로 나타남. 미국의 경우, 한국에 비해 외국인의 비중이 적은 것으로 나타남
- 유럽의 출원현황은 출원수가 매년 5건 이하로, 뚜렷한 증감 동향이 나타나지 않음. 해당 기술 분야에서 일본 시장에 대한 관심도가 높은 것으로 보임
- 일본의 출원현황은 출원수가 매년 10건 이하로, 뚜렷한 증감 동향이 나타나지 않음. 해당 기술 분야에서 일본 시장에 대한 관심도가 높은 것으로 보임

[국가별 출원현황]



(3) 기술 집중도 분석

□ 전략제품에 대한 최근 기술 집중도 분석을 위한 구간별 기술 키워드 분석 진행

- 전체 구간(1999년~2020년)에서 Computer Program Product, Machine Learning, Machine Learning Model 등 키워드가 다수 도출
- 최근 구간 분석 결과, 최근 1구간(2012년~2015년)과 비교할 때 2구간(2016년~2020년)에서 Source Code, Test Case, Electronic Control, Ecu Software Update 키워드가 많이 등장하는 것으로 보아 데이터 전처리 및 피쳐 엔지니어링 자동화 기술, 컴퓨팅 자원 효율화 기술 관련 연구개발이 활발한 것으로 추정

[특히 키워드 변화로 본 기술개발 동향 변화]

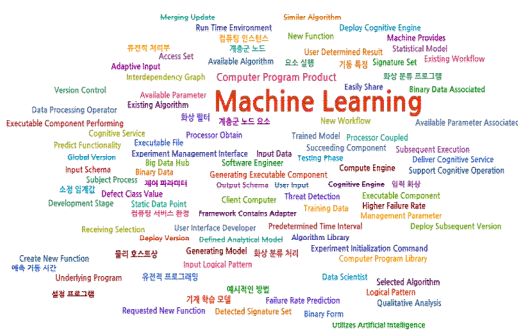
전체구간(1999년~2020년)



- Computer Program Product, Machine Learning, Machine Learning Model, Computer Implemented, User Interface, Source Code, Test Case, Electronic Control, Input Data, Ecu Software Update

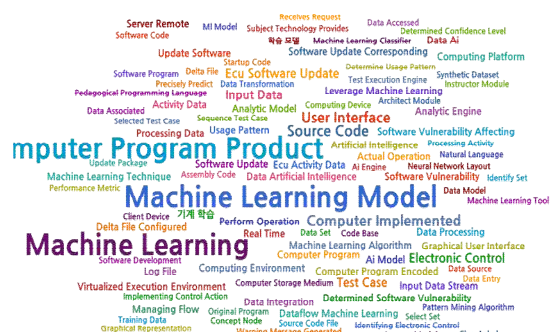
최근구간(2012년~2020년)

1구간(2012년~2015년)



- Machine Learning, Computer Program Product, Executable File, Global Version, Signature Set, Binary Data, User Input, Input Data, Client Computer, Experiment Management Interface

2구간(2016년~2020년)



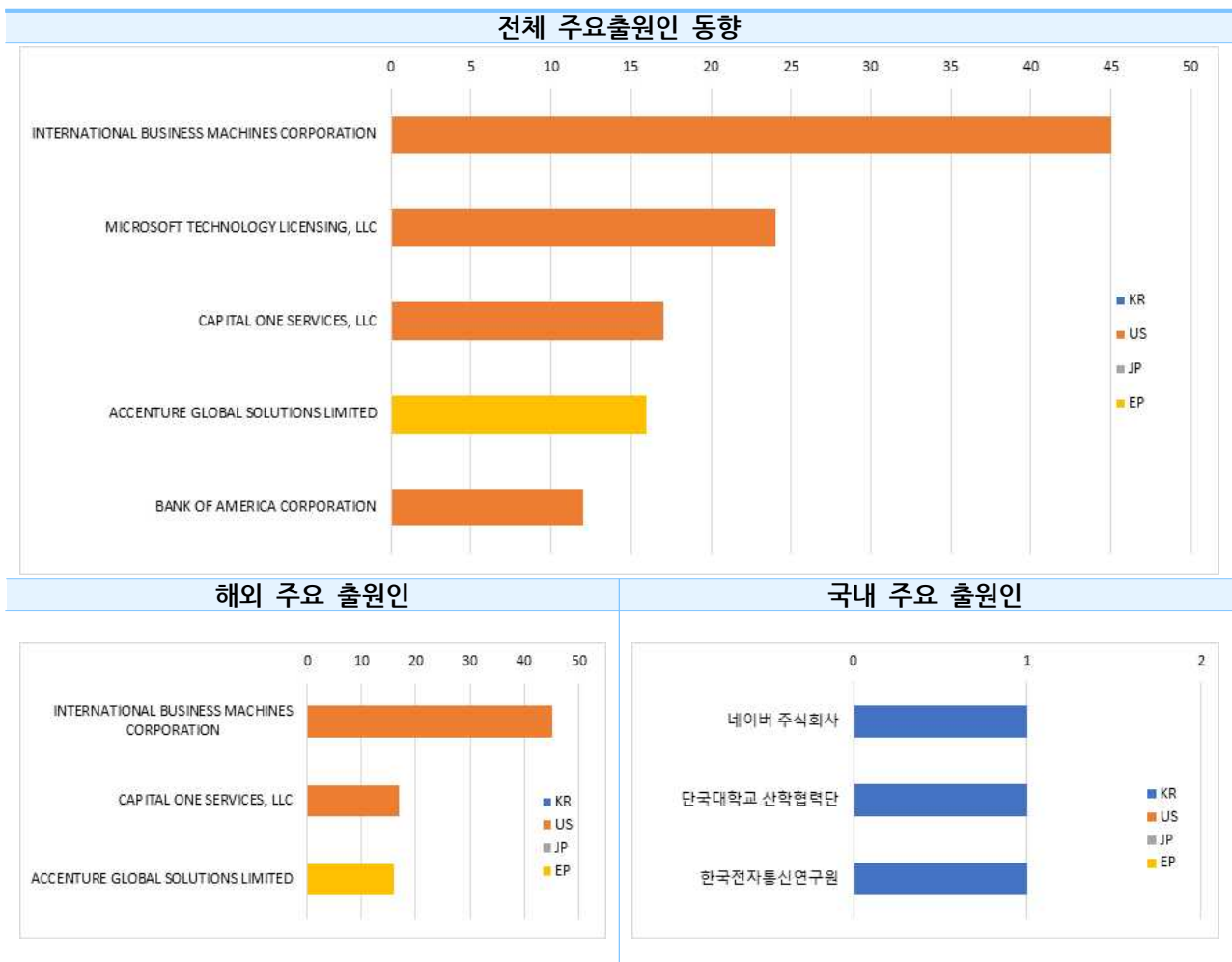
- Computer Program Product, Machine Learning Model, Machine Learning, Computer Implemented, User Interface, Source Code, Test Case, Electronic Control, Input Data, Ecu Software Update

나. 주요 출원인 분석

- AutoML 솔루션의 전체 주요출원인을 살펴보면, 주로 미국 국적의 출원인이 다수 포함되어 있는 것으로 나타났으며, 제 1 출원인으로는 미국의 INTERNATIONAL BUSINESS MACHINES CORPORATION인 것으로 나타남
 - 제 1 출원인인 INTERNATIONAL BUSINESS MACHINES CORPORATION의 출원은 미국에 집중된 경향을 보임

- AutoML 솔루션 관련 기술로 소프트웨어를 개발하는 대기업에 의한 출원이 대다수를 차지
 - 국내에서는 중소기업(개인), 연구기관/대학의 활발한 출원이 이루어짐

[AutoML 솔루션 주요출원인]

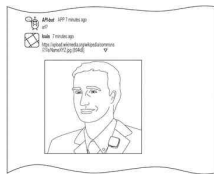
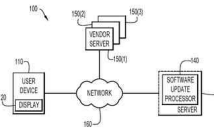
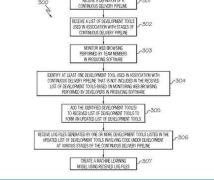
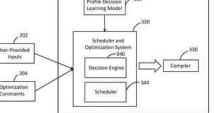
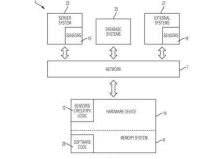


(1) 해외 주요출원인 주요 특허 분석

◎ IBM(International Business Machines)

□ IBM은 미국의 다국적 기술 기업으로, AutoML 솔루션과 관련하여, 파이프라인을 최적화하기 위한 기술에 특화된 특허를 다수 출원. 그 중 등록된 특허는 35건 보유

[IBM 주요특허 리스트]

등록번호 (출원일)	명칭	기술적용분야	도면
US10679631 (2019.06.06)	Generating chat bots from web api specifications	API 사양으로부터 채팅 봇을 자동 생성하여 사용자와의 대화를 수행하고 API 사양에서 설명된 API 호출을 인보크	
US10613852 (2019.04.02)	Cognitive installation of software updates based on user context	사용자 디바이스에 설치된 소프트웨어 애플리케이션과 연관된 하나 이상의 이슈들과 관련된 정보를 포함하는 하나 이상의 메시지들을 수신	
US10565093 (2018.10.09)	Providing cognitive intelligence across continuous delivery pipeline data	연속 전달 파이프라인(continuous delivery pipeline)에서 잠재적 장애를 검출하기 위한 방법	
US10664251 (2018.10.05)	Analytics driven compiler	전체 환경을 기반으로 한 문제 통찰력을 정적 컴파일러를 유도하는 입력으로 활용	
US10650386 (2018.08.30)	Network usage detection and presentation	네트워크 사용 검출 및 프리젠테이션을 개선하기 위한 방법	

* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ AURORA LABS LTD.

- AURORA LABS LTD.는 오스트레일리아 기업으로, 다수의 인공지능 관련 특허를 보유하고 있으며, AutoML 솔루션 기술과 관련한 특허를 미국에 출원
 - AutoML 솔루션기술과 관련하여 19건의 미국 출원을 진행하였으며, 그 중 등록된 특허는 19건으로 파악

[AURORA LABS LTD. 주요특허 리스트]

등록번호 (출원일)	명칭	기술적용분야	도면
US10747604 (2020.04.24)	Self-healing learning system for one or more controllers	차량 내의 전자 제어 유닛(ECU) 상의 소프트웨어를 자동으로 업데이트하기 위한 업데이트 패키지를 생성	
US10678627 (2020.01.24)	Self-healing learning system for one or more controllers	차량 내의 전자 제어 유닛(ECU) 상의 소프트웨어를 자동으로 업데이트하기 위한 업데이트 패키지를 생성	
US10649839 (2019.11.08)	Detecting anomalies using real-time controller processing activity	차량에서 전자 제어 장치 (ECU) 이상 유무를 확인하기 위한 비일시적 컴퓨터 판독 가능 매체와 방법	
US10642679 (2019.10.16)	Hot updates to controller software using tool chain	차량의 ECU(Electronic Control Unit)가 동작하는 동안 ECU 소프트웨어에 대한 업데이트를 수행	
US10585661 (2019.10.07)	Self-healing learning system for one or more controllers	차량 내의 전자 제어 유닛(ECU) 상의 소프트웨어를 자동으로 업데이트하기 위한 업데이트 패키지를 생성	

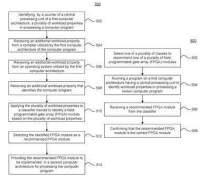

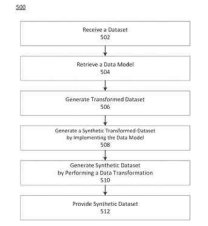
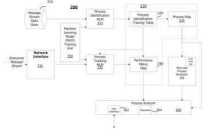
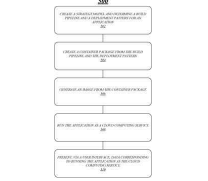
* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ CAPITAL ONE SERVICES, LLC

□ CAPITAL ONE SERVICES, LLC은 미국 금융 서비스 기업으로, 등록특허 12건 보유

- AutoML 솔루션에서 컴퓨터 프로그램을 처리하기 위한 컴퓨터 아키텍처에 추천된 FPGA 모듈을 구현하는 특허를 다수 보유

[CAPITAL ONE SERVICES, LLC 주요특허 리스트]

등록번호 (출원일)	명칭	기술적용분야	도면
US10776089 (2019.10.25)	Computer architecture based on program / workload profiling	특정 컴퓨터 프로그램에 대한 적절한 FPGA를 결정하기 위한 시스템	
US10614382 (2019.07.12)	Computer-based systems and methods configured to utilize automating deployment of predictive models for machine learning tasks	데이터 그룹핑 코드는 상기 입력 데이터에 관련된 특징들에 대한 복수의 데이터 그룹핑들을 결정함으로써 트레이닝 데이터를 생성	
US10664381 (2019.06.26)	Method and system for synthetic generation of time series data	연산들은 상기 변환된 데이터셋을 이용하여 데이터 모델을 구현함으로써 합성 변환된 데이터셋을 생성	
US10552121 (2019.05.07)	System and method for dynamic process flow control based on real-time events	트레이닝 테이블에 응답하여 메시지 스트림으로부터 프로세스들을 노출시키기 위해 트레이닝될 수 있는 머신-학습 모델	
US10635437 (2019.04.25)	Techniques to deploy an application as a cloud computing service	클라우드 컴퓨팅 서비스로서 애플리케이션의 배포를 자동화하는 기술	

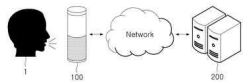
* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

(2) 국내 주요출원인 주요 특허 분석

◎ 네이버 주식회사

- 네이버 주식회사는 한국의 인터넷 서비스 기업으로, AutoML 솔루션 기술에 있어서, 음성인식을 이용하여 단말장치의 제어코드를 자동으로 생성하는 기술 관련된 특허를 1건 출원하였으며, 등록건수는 1건으로 파악

[네이버 주식회사 주요특허 리스트]

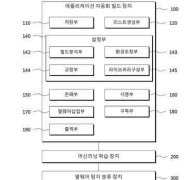
등록번호 (출원일)	명칭	기술적용분야	도면
KR33466 (2018.05.31)	코드 자동 생성 방법과, 이를 이용하는 단말장치 및 서버	음성인식을 이용하여 단말장치의 제어코드를 자동으로 생성할 수 있는 코드 자동 생성 방법	

* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ 단국대학교

□ 단국대학교는 한국의 사립 종합대학으로, AutoML 솔루션 기술과 관련된 특허를 1건 출원하였으며, 1건의 특허 중 등록된 특허는 1건을 보유

[단국대학교 주요특허 리스트]

등록번호 (출원일)	명칭	기술적용분야	도면
KR2167767 (2018.12.26)	머신러닝의 학습 데이터셋 생성을 위한 애플리케이션 자동화 빌드 장치 및 방법	멀웨어 탐지, 분류 등을 수행하는 머신러닝의 학습을 위해 샘플 애플리케이션의 데이터셋을 자동으로 생성하고, 빌드과정에서 발생하는 문제를 자동으로 해결	

* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ 한국전자기술연구원

- 한국전자기술연구원은 한국의 전자부품을 개발하는 연구기관으로, '18년도 12월부터 출원을 시작하여 아직까지 등록된 특허는 없는 것으로 파악됨
 - 한국전자기술연구원의 공개특허를 대상으로 분석한 결과, 미국에만 출원하였으며, 컨테이너에 설치된 AI 컴포넌트를 관리하는 방법에 관한 특허 출원

[한국전자기술연구원 주요특허 리스트]

공개번호 (출원일)	명칭	기술적용분야	도면
US2020-0210217 (2018.12.26)	Container-based management method for intelligent components	컨테이너를 생성하고, 복수의 컴포넌트 중 적어도 하나를 컨테이너 내에 설치하고, 컨테이너 내에 설치된 컴포넌트를 관리	

* 공개특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

다. 기술진입장벽 분석

(1) 기술 집중력 분석

- AutoML 솔루션 관련 기술에 대한 시장관점의 기술독점 현황분석을 위해 집중률 지수(CRn: Concentration Ratio n, 상위 n개사 특허점유율의 합) 분석 진행
 - 상위 4개 기업의 시장점유율이 0.31로 AutoML 솔루션 분야에 있어서 독과점 정도는 보통 수준으로 판단
 - 국내 시장에서 중소기업의 점유율 분석결과 0.8로 해당 기술에 대하여 중소기업이 진입하여 활동하고 있는 것으로 파악

[주요출원인의 집중력 및 국내시장 중소기업 집중력 분석]

주요 출원인 집중력	주요출원인	출원건수	특허점유율	CRn	n
	INTERNATIONAL BUSINESS MACHINES CORPORATION(미국)	45	13.9%	0.14	1
	MICROSOFT TECHNOLOGY LICENSING, LLC(미국)	24	7.4%	0.21	2
	CAPITAL ONE SERVICES, LLC(미국)	17	5.2%	0.27	3
	ACCENTURE GLOBAL SOLUTIONS LIMITED(아일랜드)	16	4.9%	0.31	4
	BANK OF AMERICA CORPORATION(미국)	12	3.7%	0.35	5
	ORACLE INTERNATIONAL CORPORATION(미국)	12	3.7%	0.39	6
	GOOGLE LLC(미국)	8	2.5%	0.41	7
	BONSAI AI, INC.(미국)	6	1.9%	0.43	8
	FUJITSU LIMITED(일본)	8	2.5%	0.46	9
	RELIAQUEST HOLDINGS, LLC(미국)	5	1.5%	0.47	10
전체	324	100%	CR4=0.31		
국내시장 중소기업 집중력	출원인 구분	출원건수	특허점유율	CRn	n
	중소기업(개인)	4	80.0%	0.80	
	대기업	0	0.0%		
	연구기관/대학	1	20.0%		
	전체	5	100%	CR중소기업=0.80	

(2) 특허소송 현황 분석

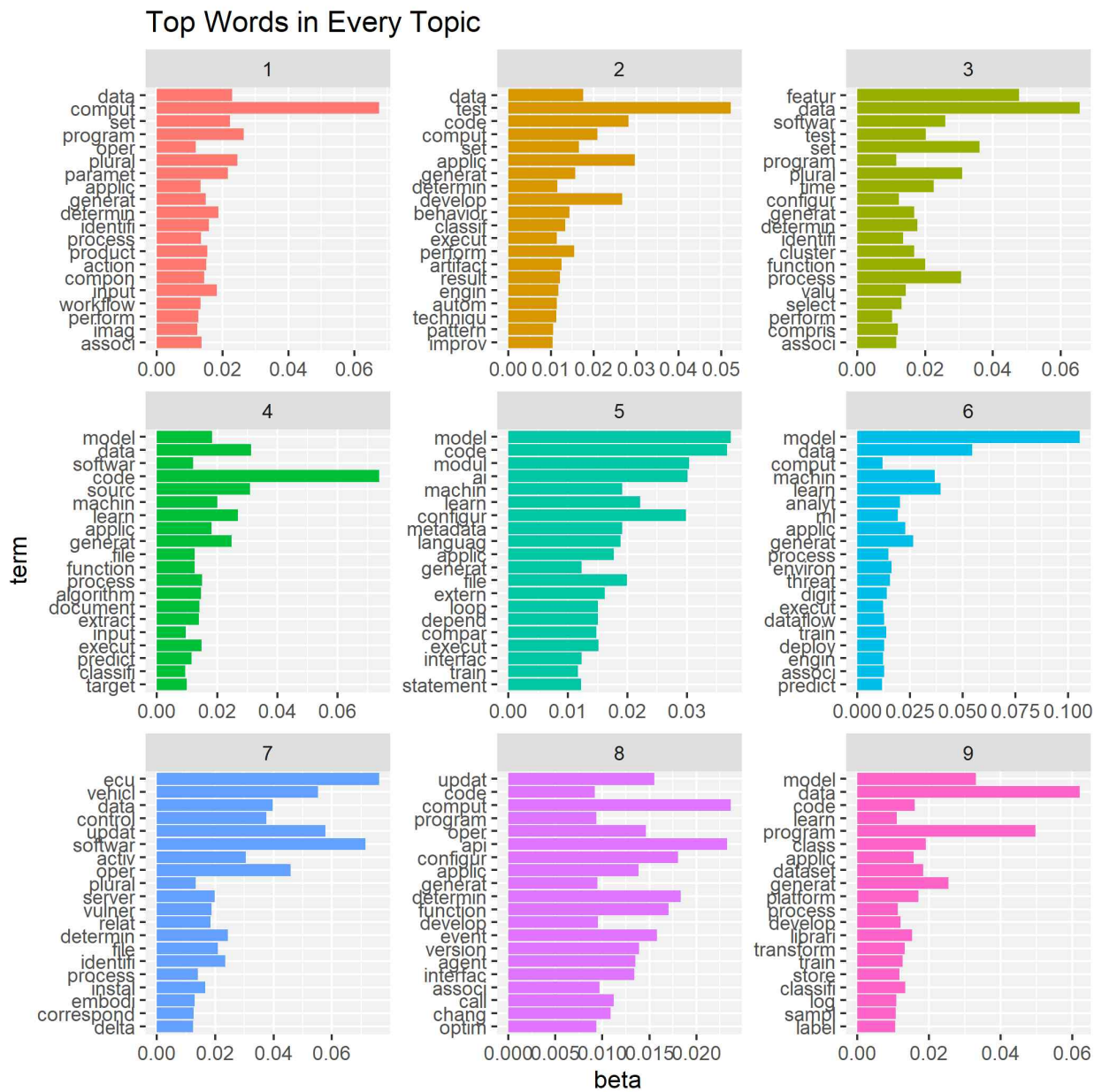
- AutoML 솔루션 분야 관련 특허소송 이력은 검색되지 않음
 - 따라서 국내기업이 미국시장에 진입하는 경우, 해당 분야를 선점할 수 있을 것으로 판단
 - 다만, 머신러닝과 관련해서는 1개의 소송이 검색

5. 요소기술 도출

가. 특허 기반 토픽 도출

- 308개의 특허의 내용을 분석하여 구성 성분이 유사한 것끼리 클러스터링을 시도하여 대표성이 있는 토픽을 도출

[AutoML 솔루션에 대한 토픽 클러스터링 결과]



나. LDA¹⁰⁾ 클러스터링 기반 요소기술 도출

[LDA 클러스터링 기반 요소기술 키워드 도출]

No.	상위 키워드	대표적 관련 특허	요소기술 후보
클러스터 01	computing, program, plural, data, set, parapet, determine, input, identification, product	<ul style="list-style-type: none"> Machine learning to determine and execute a user interface trace Determining reboot times of computing nodes 	다양한 머신러닝 및 딥러닝 모델 학습을 위한 컴퓨팅 자원 효율화 기술
클러스터 02	test, application, code, develop, computing, data, set, generate, perform, behavior	<ul style="list-style-type: none"> Providing customized and targeted performance improvement recommendations for software development teams Intelligent checking engine 	최적 알고리즘 선택 및 하이퍼파라미터 자동 탐색 기술
클러스터 03	data, feature, set, plural, process, software, time, test, function, determine	<ul style="list-style-type: none"> Feature submission de-duplication engine Ai-generated instant micro apps 	-
클러스터 04	code, data, source, learn, generate, machine, model, application, process, execute	<ul style="list-style-type: none"> Automated extraction of rules embedded in software application code using machine learning Machine learning device for learning correlation between source code information and type of feature of source code information, and method for machine learning 	다양한 머신러닝 및 딥러닝 모델 학습을 위한 컴퓨팅 자원 효율화 기술
클러스터 05	model, code, module, ai, configure, learn, file, machine, metadata, language	<ul style="list-style-type: none"> Intelligent software compiler dependency fulfillment Multiple user interfaces of an artificial intelligence system to accommodate different types of users solving different types of problems with artificial intelligence 	적합한 인공 신경망 구조 자동 탐색 기술
클러스터 06	model, data, learn, machine, generate, application, analyst, ml, environment, threat	<ul style="list-style-type: none"> Systems and methods detecting and mitigating anomalous shifts in a machine learning model Providing custom machine-learning models 	-
클러스터 07	ecu, software, update, vehicle, operate, data, control, actin, determine, identification	<ul style="list-style-type: none"> System and method for run-time update of predictive analytics system Self-healing learning system for one or more controllers 	데이터 전처리 및 피처 엔지니어링 자동화 기술
클러스터 08	computing, API, determine, configure, function, event, update, operate, version, application	<ul style="list-style-type: none"> House of quality for federated manufacturing apps Analytic system that eliminates unneeded computations 	훈련 모델 성능 측정, 검증 및 평가 자동화 기술
클러스터 09	data, program, model, generate, class, Dataset, platform, code, application, library	<ul style="list-style-type: none"> Multi-language support for interfacing with distributed data Systems and methods for third-party library management 	-

10) Latent Dirichlet Allocation

다. 특허 분류체계 기반 요소기술 도출

- AutoML 솔루션 관련 특허에서 총 10개의 주요 IPC코드(메인그룹)를 산출하였으며, 각 그룹의 정의를 기반으로 요소기술 키워드를 아래와 같이 도출

[IPC 분류체계에 기반한 요소기술 도출]

IPC 기술트리		
(서브클래스) 내용	(메인그룹) 내용	요소기술 후보
(G06F) 전기에 의한 디지털 데이터처리	• (G06F-003) 컴퓨터로 처리할 수 있는 형식으로 전송된 데이터를 변환하는 입력기구; 처리장치로부터 출력장치로 데이터를 전송하기 위한 출력기구, 예. 인터페이스 기구	다양한 머신러닝 및 딥러닝 모델 학습을 위한 컴퓨팅 자원 효율화 기술
	• (G06F-011) 에러 검출; 에러 정정; 감시	데이터 구조 및 ML 문제 자동 식별 기술
	• (G06F-016) 정보 검색; 그 데이터베이스 구조; 그 파일 시스템 구조	최적 알고리즘 선택 및 하이퍼파라미터 자동 탐색 기술
	• (G06F-009) 프로그래머를 위한 장치, 예. 제어장치	-
	• (G06F-021) 부정행위로부터 프로그램 또는 데이터, 그 컴퓨터 부품을 보호하기 위한 보안 장치	데이터 구조 및 ML 문제 자동 식별 기술
	• (G06F-008) 소프트웨어 엔지니어링을 위한 장치	훈련 모델 성능 측정, 검증 및 평가 자동화 기술
(G06N) 특정 계산모델 방식의 컴퓨터시스템	• (G06N-020) 머신 러닝	머신러닝 파이프라인 최적화 기술
	• (G06N-003) 생체모델기반 컴퓨터시스템	적합한 인공 신경망 구조 자동 탐색 기술
	• (G06N-005) 지식기반모델을 이용한 컴퓨터시스템	데이터 전처리 및 피처 엔지니어링 자동화 기술
(H04L) 디지털 정보의 전송, 예. 전신통신	• (H04L-029) 그룹 1/00에서 H04L 27/00의 하나에도 포함되지 않는 배치, 장치회로 또는 시스템	-

라. 최종 요소기술 도출

- 산업·시장 분석, 기술(특허)분석, 전문가 의견, 타부처 로드맵, 중소기업 기술수요를 바탕으로 로드맵 기획을 위하여 요소기술 도출
- 요소기술을 대상으로 전문가를 통해 기술의 범위, 요소기술 간 중복성 등을 조정·검토하여 최종 요소기술명 확정

[AutoML 솔루션 분야 요소기술 도출]

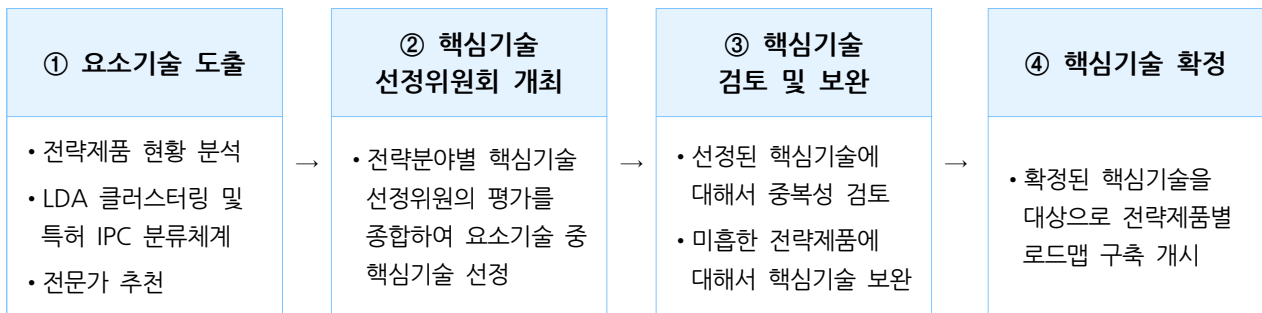
요소기술	출처
데이터 구조 및 ML 문제 자동 식별 기술	IPC 기술체계, 전문가 추천
데이터 전처리 및 피처 엔지니어링 자동화 기술	특허 클러스터링, IPC 기술체계, 전문가 추천
최적 알고리즘 선택 및 하이퍼파라미터 자동 탐색 기술	특허 클러스터링, IPC 기술체계, 전문가 추천
적합한 인공 신경망 구조 자동 탐색 기술	특허 클러스터링, IPC 기술체계, 전문가 추천
훈련 모델 성능 측정, 검증 및 평가 자동화 기술	특허 클러스터링, IPC 기술체계, 전문가 추천
머신러닝 파이프라인 최적화 기술	IPC 기술체계, 전문가 추천
다양한 머신러닝 및 딥러닝 모델 학습을 위한 컴퓨팅 자원 효율화 기술	특허 클러스터링, IPC 기술체계, 전문가 추천

6. 전략제품 기술로드맵

가. 핵심기술 선정 절차

- 특허 분석을 통한 요소기술과 기술수요와 각종 문헌을 기반으로 한 요소기술, 전문가 추천 요소기술을 종합하여 요소기술을 도출한 후, 핵심기술 선정위원회의 평가과정 및 검토/보완을 거쳐 핵심기술 확정
- 핵심기술 선정 지표: 기술개발 시급성, 기술개발 파급성, 단기개발 가능성

[핵심기술 선정 프로세스]



나. 핵심기술 리스트

[AutoML 솔루션 분야 핵심기술]

핵심기술	개요
훈련 모델 성능 측정, 검증 및 평가 자동화 기술*	• 학습된 인공지능 모델별로 인식, 예측 등 다양한 측면의 성능을 측정하고 이를 검증하여 서비스 배포까지 자동화하는 기술
적합한 인공 신경망 구조 자동 탐색 기술	• 딥러닝 기반 인공지능 모델 성능이 극대화 되는 인공 신경망 구조를 자동으로 찾는 최적 모델 탐색 과정
최적 알고리즘 선택 및 하이퍼파라미터 자동 탐색 기술	• 다양한 알고리즘 중 문제 해결에 가장 적합한 알고리즘을 선택하고 해당 알고리즘 성능 극대화를 위한 최적 하이퍼파라미터를 자동으로 탐색하는 기술
데이터 전처리 및 피처 엔지니어링 자동화 기술	• 문제 해결에 적합한 인공지능 모델에 필요한 형태로 데이터를 자동으로 전처리하고, 최적의 변수 집합을 탐색하기 위한 변수 탐색 및 데이터 변환 과정 자동화
다양한 머신러닝 및 딥러닝 모델 학습을 위한 컴퓨팅 자원 효율화 기술	• 반복적으로 수행되는 인공지능 모델 개발 프로세스의 효율성을 높이기 위한 컴퓨팅 자원 효율화 기술

* 표시는 생태계 취약 기술을 의미

다. 중소기업 기술개발 전략

- 플랫폼을 구성하기 위해서는 직접 개발이 필요하기에 SAS와 같은 기업에서 제공하는 범용적 상용 플랫폼에 주목
- 응용 서비스만으로는 신산업 성장 지속 가능성 확보가 불가능하기에 핵심 인프라 구축이 필요
- 범용성을 추구하기 보다 특정 분야(의료, 금융, 보안 등)에 특화된 AutoML 솔루션 개발 필요

라. 기술개발 로드맵

(1) 중기 기술개발 로드맵

[AutoML 솔루션 분야 중기 기술개발 로드맵]

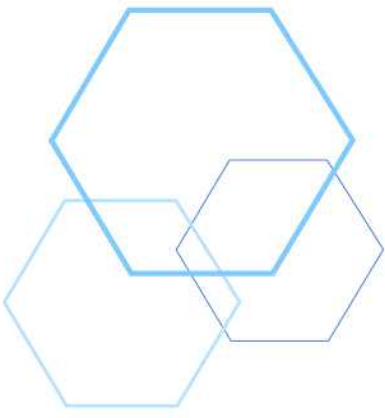
AutoML 솔루션	인공지능 모델 개발에 필요한 다양한 프로세스를 자동화하여 최적 성능 모델 자동 개발			최종 목표
	2021년	2022년	2023년	
훈련 모델 성능 측정, 검증 및 평가 자동화 기술				인공지능 모델간 비교 평가 표준화 및 안정적 서비스 배포 구현
적합한 인공 신경망 구조 자동 탐색 기술				최적 딥러닝 성능 모델 개발을 위한 신경망 구조 자동 탐색 기술 개발
최적 알고리즘 선택 및 하이퍼파라미터 자동 탐색 기술				컴퓨팅 자원 효율적인 알고리즘 선별 및 하이퍼파라미터 탐색 자동화
데이터 전처리 및 피처 엔지니어링 자동화 기술				데이터를 인식하여 알고리즘에 적합한 전처리 및 피처 엔지니어링 자동화 기술 개발
다양한 머신러닝 및 딥러닝 모델 학습을 위한 컴퓨팅 자원 효율화 기술				인공지능 모델 개발 단계의 병목을 해소하는 병렬 컴퓨팅 기술 개발

(2) 기술개발 목표

- 최종 중소기업 기술로드맵은 기술/시장 니즈, 연차별 개발계획, 최종목표 등을 제시함으로써 중소기업의 기술개발 방향성을 제시

[AutoML 분야 핵심기술 연구목표]

핵심기술	기술요구사항	연차별 개발목표			최종목표	연계R&D 유형
		1차년도	2차년도	3차년도		
훈련 모델 성능 측정, 검증 및 평가 자동화 기술	모델 검증 및 안정적 서비스 배포	훈련 모델간 성능 검증 모델링	인공지능 모델의 안정적 서비스 모니터링	훈련 모델 성능 검증 및 서비스 배포 자동화 구현	인공지능 모델간 비교 평가 표준화 및 안정적 서비스 배포 구현	기술혁신, 상용화
적합한 인공 신경망 구조 자동 탐색 기술	딥러닝 후보 모델 생성, 평가 및 자동화	딥러닝 후보 모델 생성 알고리즘 구현	중복 알고리즘 제거 및 후보 모델간 평가 및 테스트	최적 인공 신경망 구조 선정 자동화 구현	최적 딥러닝 성능 모델 개발을 위한 신경망 구조 자동 탐색 기술 개발	산학연, 기술혁신, 상용화
최적 알고리즘 선택 및 하이퍼파라미터 자동 탐색 기술	알고리즘 모델링 및 하이퍼파라미터 최적 탐색	데이터 기반 알고리즘 및 하이퍼파라미터 모델링 기법 개발	알고리즘 및 하이퍼파라미터 평가 모듈 구현	최적 알고리즘 선정 및 하이퍼파라미터 탐색 자동화 구현	컴퓨팅 자원 효율적인 알고리즘 선별 및 하이퍼파라미터 탐색 자동화	기술혁신, 상용화
데이터 전처리 및 피처 엔지니어링 자동화 기술	데이터 유형 및 머신러닝 알고리즘에 따른 전처리 및 피처엔지니어링	데이터 유형별 전처리 및 피처엔지니어링 기술 모델링	알고리즘별 전처리 및 피처엔지니어링 기술 모델링	최적 전처리 및 피처엔지니어링 자동화 구현	데이터를 인식하여 알고리즘에 적합한 전처리 및 피처 엔지니어링 자동화 기술 개발	기술혁신, 상용화
다양한 머신러닝 및 딥러닝 모델 학습을 위한 컴퓨팅 자원 효율화 기술	병렬 컴퓨팅 기술	인공지능 파이프라인 효율화 기법 개발	병렬 컴퓨팅 기법 개발	성능 개선 및 상용화	인공지능 모델 개발 단계의 병목을 해소하는 병렬 컴퓨팅 기술 개발	산학연, 기술혁신



전략제품 현황분석

합성 데이터 (Synthetic Data)



합성 데이터 (Synthetic Data)

정의 및 범위

- 실제 측정으로 획득하지 않고 필요에 의해 인공적으로 생성한 학습 데이터이며 현실 제약으로 인해 실제 데이터를 직접 활용하기 어렵거나 데이터의 수집 비용이 높은 경우에 효과적임

전략 제품 관련 동향

시장 현황 및 전망	제품 산업 특징
<ul style="list-style-type: none"> • (세계) 2018년 약 77억 2,700만 달러 규모에서 연평균 22.5%로 성장하여 2024년 261억 1,000만 달러 규모로 성장 전망 • (국내) 2018년 약 1,629억 원 규모에서 연평균 9.4%로 성장하여 2024년 5,752억 원 규모로 성장 전망 	<ul style="list-style-type: none"> • 데이터에 개인을 식별할 수 있는 민감한 정보가 담긴 경우, 데이터 보안 및 프라이버시 이슈는 데이터 활용의 활성화를 저해하며, 이를 극복하기 위한 데이터 비식별화 기술 개발 노력이 이어져 오면서 합성 데이터 기술이 관심을 받는 중
정책 동향	기술 동향
<ul style="list-style-type: none"> • 정부는 경제-사회 전반의 혁신 프로젝트로서 과학기술 정보통신부를 비롯한 전 부처가 참여해서 마련한 '인공지능(AI) 국가전략'을 발표('19.12.17.) • 인공지능(AI) 분야 '포괄적 최소(네거티브) 규제' 단계별 계획 수립('20) 	<ul style="list-style-type: none"> • 데이터 보안 및 프라이버시 보호를 위한 기술개발 • 인공지능 제품/서비스의 성능 개선을 위한 기술 개발 • 합성 데이터의 왜곡 및 편향성 극복을 위한 기술 개발 • 데이터 도메인 변환 기술 개발 • 합성 데이터의 악용 방지를 위한 기술 개발
핵심 플레이어*	핵심기술
<ul style="list-style-type: none"> • (해외) Google, Twitter, Amazon, Microsoft, Neuromation, Facebook, Laan Labs, Manheim • (국내) Naver, SK C&C, NC, LG, FASOO, Desilo, 셀바스, 머니브레인 	<ul style="list-style-type: none"> • 이미지 데이터 합성 기술 • 머신러닝 학습 데이터 증대 기술 • 합성 데이터 탐지 기술 • 데이터 비식별화 기술 • 시계열 데이터 합성 기술

*생태계 취약 전략제품

중소기업 기술개발 전략

- 여러 산업의 다양한 도메인에 대해서, 각 도메인의 데이터 특성을 고려하여 특화된 합성 데이터 생성 기술 개발
- 실제 제품 및 서비스와 연계한 맞춤형 합성 데이터 생성 기술 개발
- 데이터 보안 및 프라이버시 보호, 데이터 악용 방지를 위한 데이터 비식별화 및 합성 데이터 탐지 기술 개발

생태계 강화방안

- 합성 데이터기술은 아직 초기단계로 국가적 기술경쟁력 확보를 위한 스타트업 및 중소기업에 대한 지원 필요
- 기술의 확산을 촉진하기 위해서는 합성 데이터 기술의 플랫폼화 및 활용 실증 연구 사업 및 정부 정책 지원 필요
- 합성 데이터의 악용 및 사회적 문제에 대해 어떻게 대응할지 국가적인 관점에서의 해결책 논의 필요

1. 개요

가. 정의 및 필요성

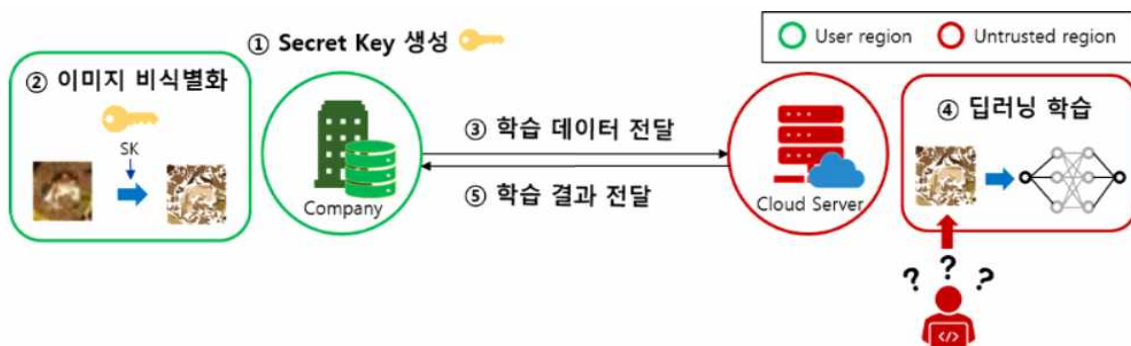
(1) 정의

- 직접 획득한 데이터가 아닌, 인공적으로 데이터를 합성하는 기술
 - 합성 데이터 기술은 현실 제약으로 인해 실제 데이터를 직접 활용하기 어렵거나 데이터의 수집 비용이 높은 경우에 효과적임
- 실제 측정으로 획득하지 않고 필요에 의해 인공적으로 생성한 학습 데이터
- 문제 상황 및 활용 목적에 따라 다양한 합성 데이터 생성 기술이 연구개발되어오고 있음
 - 새로운 제품이나 도구의 검증, 특수 목적에 따른 데이터의 합성, 데이터 기반 인공지능 (머신러닝/딥러닝) 구현 등 다양한 문제 상황 및 활용 목적에 대해서 세부 기술들의 연구개발이 이루어지고 있음
 - 데이터의 활용이 요구되는 다양한 산업분야에 광범위하게 활용됨

(2) 필요성

- 데이터 보안 이슈
 - 많은 산업에서 데이터 보안 및 프라이버시, 규제 이슈 등으로 데이터 소유자가 가지고 있는 원본 데이터를 직접 외부에 반출하여 사용하기 어려움 (예시: 의료 정보 데이터, 고객 신용정보 데이터 등)
 - 데이터 소유자가 외부와 협력을 위해서, 원본 데이터의 특성을 보존하면서 비식별화된 합성 데이터를 생성하여 공유
 - 프라이버시 보존 머신러닝 (Privacy-Preserving Machine Learning) 및 프라이버시 보존 데이터마이닝 (Privacy-Preserving Data Mining) 등 키워드와 연관됨

[프라이버시 보존 머신러닝 개념도]



* 출처: 딥러닝 프라이버시 보존 기술 개발, 부산대학교 정보컴퓨터공학부(2020)

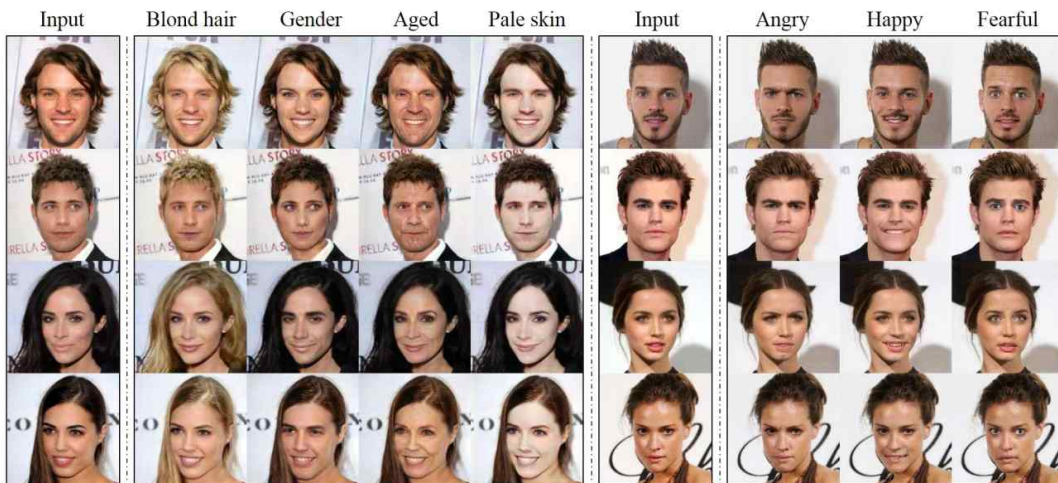
□ 학습용 데이터 부족 상황

- 데이터 기반의 인공지능 구현 시 충분한 양의 학습데이터를 확보하지 않으면 학습데이터에 대한 과적합(Overfitting) 문제가 쉽게 발생할 수 있으나, 많은 현실 상황에서 데이터 수집 비용으로 인해 많은 양의 데이터를 확보하기는 현실적으로 어려움
- 데이터의 확보가 어렵지는 않으나 특정 범주의 데이터에 대한 확보가 어려운 경우도 있음
- 활용 가능한 학습데이터의 양이 부족한 경우 또는 범주 불균형(Class Imbalance) 상황에서, 기존 데이터와 유사한 성질을 가지는 합성 데이터를 인위적으로 생성하여 추가 학습데이터로 활용
- 머신러닝 분야에서 주로 다루어지는 오버샘플링 (Oversampling) 및 데이터증대(Data Augmentation)등 키워드와 연관됨

□ 시뮬레이션 및 가짜(Fake) 데이터 생성

- 기존의 데이터(주로 이미지, 음성, 또는 영상 등)을 기존 도메인에서 주어진 목적에 따른 다른 도메인으로 변환한 합성 데이터를 생성
- 실제 데이터를 획득하기 어려운 케이스에 대한 시뮬레이션 목적으로 활용될 수 있음
- 이미지/영상처리 분야에서는 Image-to-Image Translation이라는 기술로 활발히 연구개발이 이루어지고 있음
- 이러한 기술은 가짜 뉴스 및 가짜 영상 생성 등으로 악용되고 있으며, 따라서 가짜 데이터를 식별하는 연구도 활발히 이루어지고 있음

[의도에 따른 기존 이미지의 도메인 변환 예시]



* 출처: Multi-Domain Image-to-Image Translation, CVF(2018)

나. 범위 및 분류

(1) 가치사슬

- 합성 데이터 기술은 현실 제약 상황에서 데이터의 효과적 활용을 위한 기반 기술로서, 다양한 산업 분야에 대해 광범위한 용도로 활용될 수 있음
 - 후방산업: 데이터의 효과적인 수집/저장/관리, 데이터 기반 시뮬레이션, 가상 검증과 인공지능 성능 향상에 핵심적인 역할을 하고 있는 머신러닝/딥러닝 기술 및 개발 플랫폼을 포괄함
 - 전방산업: 제조, 금융/보험, 의료, 보안, 인터넷, 미디어 등 데이터의 활용이 요구되는 다양한 산업 분야

[합성 데이터(Synthetic Data) 분야 산업구조]

후방산업	합성 데이터	전방산업
데이터 수집/저장/관리, 시뮬레이션, 가상 검증, 머신러닝/딥러닝 기술, 머신러닝/딥러닝 개발 플랫폼	데이터 비식별화 기술, 데이터 증대 기술, 데이터 도메인 변환 기술, 합성 데이터 식별 기술	제조, 금융/보험, 의료, 보안, 정보/인터넷, 유통 등

(2) 용도별 분류

- 합성 데이터 기술은 데이터의 활용이 요구되는 다양한 산업 분야에서 효과적으로 활용될 수 있음
 - 데이터 확보의 어려움, 높은 수집비용 및 프라이버시 등 현실적인 제약 이슈를 다루어, 데이터 기반 인공지능 구현의 어려움을 해소하고 활용을 활성화

[합성 데이터 (Synthetic Data)의 용도별 분류]

분류	상세 내용 (예시)
제조 분야	• 상대적으로 발생 빈도가 낮은 불량 케이스에 대한 합성 데이터 생성을 통한 효과적인 불량 분석 및 탐지 모델링 수행
금융/보험 분야	• 고객 데이터의 외부 반출 및 활용 시, 개인 식별이 불가능하도록 민감한 개인 정보를 비식별화한 합성 데이터를 활용
의료 분야	• 실제 사례 데이터 확보가 어려운 영역에 대해서, 기존 사례 데이터의 도메인 변환을 통한 시뮬레이션 분석
보안 분야	• 개인정보 식별 방지 및 데이터 위/변조에 기반한 사기 행위의 탐지
인터넷 분야	• 이미지 보정, 음성/영상 합성 등 사용자 맞춤형 서비스 개발
미디어 분야	• 연출이나 재현이 어려운 장면을 인위적으로 생성 및 보정

2. 산업 및 시장 분석

가. 산업 분석

◎ 데이터의 활용에 대한 산업 요구 증대

- 빅데이터의 시대가 도래하면서 실시간으로 방대한 양의 데이터가 생성되고 있으며, 글로벌 시장조사기관 IDC는 2025년 전 세계의 데이터 생산량은 약 163ZB 정도가 될 것으로 예측
- 다양한 산업 분야에서 축적되고 있는 빅데이터로부터 의미 있는 지식을 추출하여 가치를 창출하고자 하는 요구가 지속적으로 증대

◎ 인공지능 연구의 주요 핵심 기반기술

- 딥러닝 등 데이터 기반 인공지능 구현을 위해서는 많은 양의 고품질의 학습 데이터 확보가 필수 요소이며, 데이터 확보 비용이 높은 경우 이를 우회하기 위한 기반 기술로 합성 데이터 기술 활용
- 합성 데이터 생성을 위한 인공지능 기술도 연구가 활발히 이루어지고 있음

◎ 학계 수준에서 매우 활발한 연구가 이루어지면서, 산업계로 전이가 시작

- 학계에서 합성 데이터는 과거부터 꾸준히 이루어져 왔으며, 산업 수요에 힘입어 IT 기업들을 중심으로 실제 산업 적용을 위한 노력들이 이어지며 산업 생태계가 조성
- 합성 데이터 기술을 활용한 제품 또는 서비스를 제공하는 다양한 기술 스타트업들의 성공이 이어지는 중

◎ 합성 데이터 기술의 악용이 사회적 이슈가 되고 있으며, 이에 대응한 노력도 이어지는 중

- 합성 데이터 기술은 데이터 위변조를 통한 사기 행위 등에 쉽게 악용될 수 있으며, 합성 데이터 기술이 빠르게 발전하면서 이는 심각한 사회적 문제로 대두 중
- DeepFake와 같이 인공지능 기술을 활용하여 특정 인물의 얼굴을 특정 영상에 합성한 편집물을 악의적으로 배포하는 사례 존재
- 합성 데이터 기술로 만들어진 가짜 데이터를 탐지하기 위한 기술적 노력이 이어지는 중

나. 시장 분석

(1) 세계시장

- 2018년 세계 AI교육 데이터 생성 시장 규모는 77억 2,700만 달러로 평가되었으며 연평균 22.5%로 성장하여 2024년 261억 1,000만 달러 규모로 성장 전망

[세계 AI 교육 데이터 생성 시장 시장규모]

(단위 : 백만 달러, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
세계시장	772.7	946.5	1,160	1,420	1,740	2,131	2,611	22.5

* 출처: Grand View Research, Inc, 2020.05.26.을 바탕으로 세계 AI 시장과 국내 AI 시장을 복합적으로 고려하여 네모아이씨지 재추정

- AI Multiple의 조사에 따르면 합성 데이터 분야 중 테스트 데이터 관리 시장은 연평균 성장률 12.7%로 성장할 전망이며 AI 교육 데이터 생성 시장은 연평균 성장률 22.5%로 성장할 전망, 이 두 서비스를 제공하는 합성 데이터(Synthetic Data)시장은 연평균 성장률 10% 이상으로 성장할 전망

(2) 국내시장

- 국내 AI 교육 데이터 생성 시장 규모는 2018년 약 1,629억 원 규모에서 연평균 성장률 9.4%로 2024년 5,752억 원 규모로 성장 전망
 - 세계 인공지능 시장규모와 국내 인공지능 시장 규모를 동일한 가중치로 합산해 국내 AI 교육 데이터 생성 시장의 2018년 시장규모를 산출하고, 세계 인공지능 시장 성장률과 국내 인공지능 시장 성장률을 동일한 가중치로 합산하여 국내 AI 교육 데이터 생성 시장 성장률을 산출

[국내 AI 교육 데이터 생성 시장 시장규모]

(단위 : 억 원, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
국내시장	1,629	2,010	2,481	3,061	3,777	4,661	5,752	23.4

* 출처: Grand View Research, Inc, 2020.05.26.,Revenues from the artificial intelligence software market worldwide from 2018 to 2025, 2020, Statista를 기반으로 네모아이씨지 재추정

3. 기술 개발 동향

기술경쟁력

- 합성 데이터 (Synthetic Data)는 미국이 최고기술국으로 평가되었으며, 우리나라는 최고기술국 대비 70.8%의 기술 수준을 보유하고 있으며, 최고기술국과의 기술격차는 2.3년으로 분석
- 중소기업의 기술경쟁력은 최고기술국 대비 61.7%, 기술격차는 3.1년으로 평가
- EU(79.5%)>일본(75.4%)>중국(71%)>한국의 순으로 평가

기술수명주기(TCT)¹¹⁾

- 합성 데이터 (Synthetic Data)는 6.65의 기술수명주기를 지닌 것으로 파악

가. 기술개발 이슈

◎ 데이터 보안 및 프라이버시 보호를 위한 기술개발

- 데이터에 개인을 식별할 수 있는 민감한 정보가 담겨있는 경우 (공공 행정데이터, 금융 데이터, 의료 데이터 등), 데이터 보안 및 프라이버시 이슈는 데이터 활용의 활성화를 저해하며, 이를 극복하기 위한 데이터 비식별화 기술 개발 노력이 이어져 오면서 합성 데이터 기술이 관심을 받는 중
- 우리나라의 경우, 2016년에 개인정보 비식별 조치에 관한 명확한 기준과 지원관리체계로서 정부차원에서 “개인정보 비식별 조치 가이드라인”을 발간하였으며, 2020년에 새로운 서비스나 기술, 제품 개발의 목적의 데이터의 분석·활용을 촉진하고 규제 완화 및 법적 근거를 마련하기 위한 데이터 3법이 통과되면서 제도적인 노력이 이어지는 중
- 데이터 보안을 유지하면서 인공지능 구현을 활성화하기 위한 기술적 노력들이 이어지고 있음. 가령 데이터 내에 개인을 식별할 수 있는 정보가 있는 경우, 이의 일부 또는 전부를 삭제, 또는 일부를 속성 정보로 대체 처리함으로써 다른 정보와 결합하여도 특정 개인을 식별하기 어렵도록 조치
- 비식별화된 데이터가 효과적으로 활용되기 위해서는 데이터의 비식별화 과정에서 정보 손실을 최소화하는 것이 중요 식별 처리된 데이터라고 하더라도 항상 안전하지는 않으며, 재식별 가능성의 문제가 있음. 재식별이 완전히 불가능할 정도로 비식별 처리된 데이터는 데이터의 정보 손실이 커서 분석 및 활용을 위한 가치가 떨어질 가능성 높음

11) 기술수명주기(TCT, Technical Cycle Time): 특허 출원연도와 인용한 특허들의 출원연도 차이의 중앙값을 통해 기술 변화속도 및 기술의 경제적 수명을 예측

[데이터 비식별화 예시]

개인 특성	<ul style="list-style-type: none"> • 성별, 연령(나이), 국적, 고향, 시·군·구명, 우편번호 병역여부, 결혼여부, 종교, 취미, 동호회·클럽 등 • 흡연여부, 음주여부, 채식여부, 관심사항 등
신체 특성	<ul style="list-style-type: none"> • 혈액형, 신장, 몸무게, 허리둘레, 혈압, 눈동자 색깔 등 • 신체검사 결과, 장애유형, 장애등급 등 • 병명, 상병(傷病)코드, 투약코드, 진료내역 등
신용 특성	<ul style="list-style-type: none"> • 세금 납부액, 신용등급, 기부금 등 • 건강보험료 납부액, 소득분위, 의료 급여자 등
경력 특성	<ul style="list-style-type: none"> • 학교명, 학과명, 학년, 성적, 학력 등 • 경력, 직업, 직종, 직장명, 부서명, 직급, 전직장명 등
전자적 특성	<ul style="list-style-type: none"> • 쿠키정보, 접속일시, 방문일시, 서비스 이용 기록, 접속로그 등 • 인터넷 접속기록, 휴대전화 사용기록, GPS 데이터 등
가족 특성	<ul style="list-style-type: none"> • 배우자·자녀·부모·형제 등 가족 정보, 법정대리인 정보 등

* 출처: '개인정보 비식별화', IT DAILY(2020)

◎ 인공지능 제품/서비스의 성능 개선을 위한 기술 개발

- 수집비용 및 현실적 제약 등 이슈로 인공지능 학습을 위한 대규모의 데이터 확보가 어려운 경우, 합성 데이터 기술을 활용하여 다양한 형태의 합성 데이터를 인공적으로 생성하고 이를 활용하여 인공지능 성능 개선 극대화하고자 하는 노력들이 광범위하게 이루어지는 중
- 인공지능 학습의 목적으로 합성 데이터를 생성하기 위한 노력으로, Oversampling 및 Data Augmentation 등의 연구 분야로 활발히 연구 중
- 합성 데이터 생성을 위한 목적으로 인공지능 기술을 적용하는 연구 또한 활발하게 이루어지고 있으며, 주로 심층 생성모델(Deep Generative Model)인 Variational Autoencoder, Generative Adversarial Network 등 기법이 널리 활용
- 다만, 아래와 같이 데이터 왜곡 및 편향성으로 인한 문제 야기 가능

◎ 합성 데이터의 왜곡 및 편향성 극복을 위한 기술 개발

- 합성 데이터는 직접 획득한 실제 데이터와 동일한 분포를 가지는 것으로 가정되나, 기술적 한계로 인해 데이터가 한쪽으로 치우쳐지는 편향이 발생하거나 원래 분포의 특성을 왜곡하는 문제가 발생할 수 있음
- 합성 데이터의 생성 과정에서의 편향 및 왜곡이 발생하는 경우, 합성 데이터의 활용이 잘못된 또는 무의미한 결과 도출 가능
- 합성 데이터의 활용 목적에 따라서는, 의도한 편향성을 가지는 합성 데이터 생성 기술이 유의미함 (데이터 도메인 변환 기술 등)

◎ 데이터 도메인 변환 기술 개발

- 기존의 데이터(주로 이미지, 음성, 또는 영상 등)를 기존 도메인에서 주어진 목적에 따른 다른 도메인으로 변환한 합성 데이터를 생성하는 기술이 개발 중
- 인물 사진의 표정 변화, 실제 사진을 특정 화가의 화풍에 따른 그림으로 변화, 그림의 화풍 변화, 인물 사진으로부터 이모티콘 생성, 음성 변조 등 다양한 제품 및 서비스에 실제로 응용 중
- 실제 데이터를 획득하기 어려운 케이스에 대해서 데이터 도메인 변환 기술을 적용하여 관련 데이터로부터 목표 의도에 따른 데이터를 생성 가능. 가령 의료 분야와 같이 실제 사례 데이터 확보가 어려운 영역에 대해서, 기존 사례 데이터의 도메인 변환을 통한 가상 사례를 확보

◎ 합성 데이터의 악용 방지를 위한 기술 개발

- 합성 데이터 기술이 보편화 되면서 적용 대상과 분야가 다변화되고 있으며, 인공지능 기반 합성 데이터 기술을 활용하여 제작된 콘텐츠가 확산되고 있으며, 긍정적인 효과도 있으나 일부 악용 사례가 사회 문제로 대두
- 특히, 원본 콘텐츠에 다른 요소를 악의적으로 결합하여 합성 콘텐츠를 생산하는 DeepFake 기술은 산업적 활용가치가 높고 특수 효과를 쉽게 구현할 수 있게 해주는 긍정적 측면이 있으나, 가짜 뉴스 및 포르노 합성 등으로 악용되면서 많은 피해자를 양산하며 가짜 콘텐츠와의 전쟁으로 이어지는 중
- 합성 데이터를 탐지하는 기술 수요가 증대되고 있으며, 보안 산업에서 주요 이슈가 되고 있으며 공익적 목적으로도 중요한 화두임. 주요 IT, 보안기업 및 스타트업들을 중심으로 활발하게 기술 개발 중

◎ 합성 데이터 기술 연구개발 생태계 조성 및 플랫폼화

- 딥러닝 기술 확산으로 이를 접목한 합성 데이터 기술들이 매우 활발하게 연구되면서, 다양한 분야에서 결과물들이 생성 중
- 이러한 기술적 발전에 힘입어 주요 IT기업들이 관련 제품 및 서비스를 출시하고 있으며, 또한 합성 데이터 기술 기반 스타트업의 창업도 잇따르면서 기술 생태계가 조성 중
- 보다 기술의 확산을 촉진하기 위해서는 합성 데이터 기술의 플랫폼화가 필요

나. 생태계 기술 동향

(1) 해외 플레이어 동향

- (Google) 인공지능(AI)이 소리를 학습해 해당 소리를 디지털 형태로 그대로 재현 및 다양한 소리를 혼합해 새로운 소리를 만드는 기술 개발
 - 구글 브레인팀과 알파고로 유명한 딥마인드팀이 협업해 연구를 진행중이며 예술 창작 분야에서 활용될 인공지능을 발전시키기 위한 예술가-코딩 기술자-머신러닝 연구자로 이루어진 커뮤니티 구축 목표
 - 마젠타 프로젝트를 통해 만들어진 신경 오디오 합성 기술인 엔신스(Nsynth)는 소리를 디지털로 분석해 학습, 재생하는 기술
 - 머신러닝을 활용한 음악과 미술 작품의 생산을 통해 머신러닝을 창작 영역으로 확대하는 영향 기대
- (Google) 딥드림(Deep Dream)은 인간의 뇌에서 정보를 보내는 신경망을 모방한 ‘인공 신경망(neural network)’을 통해 수많은 이미지를 인식 및 저장하고 이 이미지들의 특징을 추출해 시각화하는 기술 개발
 - 딥러닝 기술을 기반으로 수백만 개의 연습 샘플을 바탕으로 인공신경망을 훈련하여 사물 분별 능력을 생성
- (Twitter) 딥포저(Deep Forger)는 신경망(Neural Networks)를 학습해 창작 활동을 하는 딥드림(Deep Dream)에서 한단계 더 나아가 이미지의 ‘질감’을 인식하도록 학습하는 기술 개발
 - AI가 ‘sketch(스케치)’, ‘abstract(추상화)’ 같은 단어에 ‘Picasso(피카소풍으로)’, 또는 ‘Gogh(고흐풍으로)’와 같은 명령어에 따라 취향에 따른 그림을 생성
- (Google) 구글계열사 딥마인드가 텍스트 투 스피치(Text to Speech) API를 발표
 - 프로그래밍 기술 없이도 응용 프로그램을 만들 수 있도록 지원하는 인터페이스 프로그램인 API 개발
 - 2017년 구글은 사람처럼 말하는 가상 비서 서비스를 개발하기 위해 딥마인드의 AI 모델 ‘웨이브넷(WaveNet)’을 채택
- (Amazon) 수많은 오디오 클립으로부터 낭독 기술을 학습하여 자체적으로 텍스트를 실제 뉴스 진행자처럼 읽어주는 텍스트 투 스피치(Text to Speech) 프로그램 개발
 - 머신러닝이 뒷받침 된 TTS 기술로, 문장 내 모든 어절을 동일한 성조나 세기로 읽지 않고 특정 어절은 강조해 생동감 있게 낭독 가능

- (Microsoft) 2020년 미국 선거에 앞서 딥페이크(deepfake)식별 솔루션 공개
 - ‘마이크로소프트 비디오 어센터케이터(Microsoft Video Authenticator, MVA)’는 스틸 사진이나 비디오를 분석하여 미디어가 인위적으로 조작되는 확률 또는 신뢰 점수를 공개
 - 콘텐츠 제작사가 콘텐츠에 디지털 해시와 인증서비스를 추가할 수 있도록 하는 마이크로소프트 Azure에 내장된 툴을 이용해 콘텐츠가 진품이고 변경되지 않았음을 사람들에게 높은 정확도로 제공 및 생산자에 대한 세부사항을 제공

- (Facebook) 인공지능(AI) 기술로 조작된 ‘딥페이크(deepfake)’영상을 가려내기 위한 ‘딥페이크 탐지 챌린지’ 프로젝트 진행
 - 시청자(소비자)를 현혹하려는 의도로 편집된 영상에 사용돼 온 AI 기술을 모두가 더 잘 탐지해낼 수 있는 기술을 만드는 것이 목표

- (Neuromation) 신경망 네트워크의 딥러닝을 위한 합성 학습 환경을 구축하는 기술 플랫폼 제공
 - Neuromation 플랫폼은 인공지능 모델 개발의 혁신을 위해 블록체인 작업 증명 토큰을 통한 분산 컴퓨팅 이용
 - 문제 해결을 위한 완벽하고 정밀한 인식을 통해 다양한 정보를 합성 및 시각화하여 각각의 문제에 대해 정확하고 알맞은 답변 도출
 - 연구소는 초기 데이터 생성기 및 데이터셋을 Neuromation 플랫폼 시장에 제공, 연구소 파트너가 플랫폼을 통해 선제적으로 서비스를 처리하도록 장려하는 방식으로 서비스 시장을 구축할 계획

- (Microsoft) 생성형 AI와 차등 프라이버시를 활용해 거래자 데이터를 위한 GDPR급 합성 데이터 엔진을 제공
 - 실제 데이터의 현실성과 대표성을 유지하며 각각의 재식별을 불가능하게 하는 합성 데이터 그룹을 시뮬레이션 가능
 - 가상 데이터 엔진은 빅데이터 자산을 보유한 조직을 위해 기업 내 혹은 클라우드를 통해 사용할 수 있는 소프트웨어를 간편하게 제공

- (Laan Labs) 이미지 교육을 위한 합성 데이터 생성기를 개발
 - 3D 자동차 모델, 배경 장면 및 조명과 같은 사실적인 이미지로 신경망 시스템 훈련
 - 비용적 측면이나 근무 시간 측면을 절약했으며 사람의 작업 도중의 오류 위험을 제거

- (Manheim) 송금 및 결제를 가속화할 수 있도록 일괄 처리 시스템에서의 실시간으로 작동하는 시스템으로 이주하는 작업을 진행
 - CA테스트 데이터 관리자를 구입해 단기간에 대량의 합성 데이터를 생성해 효과적인 테스트 진행

(2) 국내 플레이어 동향

- (FASOO) 빅데이터를 분석하고 이중산업간 안전한 데이터 결합 과정에서 개인을 알아볼 수 없도록 데이터를 안전하게 가명,익명 처리하는 비식별화 기술 개발
 - 데이터 활용과 개인정보 유출방지를 고려한 안전하고 신뢰 가능한 빅데이터 사용 환경 구축
 - 개인정보 비식별 조치 가이드라인, ISO/IEC 20889, GDPR, HIPAA 등의 규제에 효율적으로 대응할 수 있는 비식별 기술 제공
 - 각종 데이터의 성격, 유형, 활용 목적 등에 따라 다양한 프라이버시 모델 제공 및 빅데이터 분석의 효용성과 위험에 대한 지표들을 제공

- (Desilo) 동형 암호 기술을 활용해 암호화된 상태 그대로 데이터 연산을 가능하게 하며, 암호를 푸는 복호화 과정 없이 데이터를 처리해 데이터 프라이버시 및 데이터 사일로 문제를 해결
 - 증가하는 전 세계 데이터 처리량 및 데이터 전송량, 빈도를 처리하고 활용하는데 필요한 솔루션 개발

- (NC) 스피치랩(Speech Lab) 음성합성(Speech Synthesis)팀은 텍스트로부터 사람의 목소리를 만들어내는 음성합성 기술을 연구 및 개발
 - 음성합성 기술은 텍스트 문자열을 입력하면 사람의 목소리로 출력해주는 것을 뜻함
 - 입력과 출력에 대한 데이터만 존재하면 입력과 출력 사이의 함수 관계를 스스로 학습해 알아내는 새로운 음성합성 기술 개발

- (LG) 자사 AI를 강화하기 위해 음성데이터를 확보에 매진중이며 AI 탑재 기기를 스마트폰에서 가전으로 확대.
 - 음성합성 기술은 텍스트 문자열을 입력하면 사람의 목소리로 출력해주는 것을 뜻함
 - 입력과 출력에 대한 데이터만 존재하면 입력과 출력 사이의 함수 관계를 스스로 학습해 알아내는 새로운 음성합성 기술 개발

- (네이버) 짧은 녹음만으로 고품질의 합성음을 제작할 수 있는 클로바의 음성 합성 신기술 'NES(Natural End-to-end Speech Synthesis)'을 공개 및 이를 기반으로 한 보이스 비즈니스 서비스 제공
 - 특정 분야에 대한 제약 없이 쉽고 빠르게 음성 서비스를 개발 가능
 - 기본 목소리 외에 기쁨, 슬픔 등 감정이 반영된 목소리 제공

- (SK C&C) 음성합성 스타트업 라이언로켓과 인공지능 텍스트 음성 변환(TTS) 사업 협력 협약(MOU)'을 체결
 - 음성을 자음,모음으로 나눠 소리를 붙이는 기존 텍스트 음성 변환 기술과 달리 AI가 제공자의 말뭉치를 그대로 분석 및 학습하는 기술 적용
 - 해당 기술을 적용하면 녹음 파일을 이미 보유한 한류 스타나 유명 캐릭터는 별도 녹음 없이 문자 입력과 동시에 목소리를 생성해 제작 기간을 크게 단축

- (셀바스) 감정까지 표현하는 AI 음성합성(TTS) 기술 발표
 - 이 기술은 2018년 출시된 딥러닝 기반 개발된 음성합성(DNN-TTS) 기술인 'SelyvydeppTTS'에 적용되어 고객들에게 제공될 예정
 - 텍스트의 제한 없이 어떠한 텍스트를 입력해도 실시간 음성 합성을 출력이 가능

- (머니브레인) 영상의 진위 여부를 판단하는 능력을 갖춘 딥러닝 모델 'AI Fake Finder(AI 영상 조작 검출 기술)'을 개발
 - 2020년 'n번방 사건'이 사회적으로 큰 물의를 일으키면서 이 범죄에 사용된 '딥페이크(deep fake) 기술'을 악용한 범죄를 막을 수 있는 기술
 - 머니브레인의 딥러닝 모델은 '페이스포렌식++'에 적용된 합성 방법 네 가지(DeepFakes, Face2Face, FaceSwap, NeuralTexture)영역에서 99%를 상회하는 진위 판별 능력을 보임

다. 국내 연구개발 기관 및 동향

(1) 연구개발 기관

[합성 데이터 (Synthetic Data) 분야 주요 연구조직 현황]

기관명	연구소명	주요 내용
과학기술정보통신부	성균관대학교 전기전자컴퓨터공학부	• 데이터 과부족, 가려진 얼굴 복원, 적대적 도메인 적응, 생성적 적대 신경망
과학기술정보통신부	포항공과대학교 컴퓨터공학과	• 개인화 기술, 계산사진학, 빅 비주얼 데이터, 영상 및 비디오 처리, 영상 프라이어

(2) 기관 기술개발 동향

- (성균관대학교) 부분합성영상 학습데이터와 적대적 영역 적응(adversarial domain adaptation)을 활용하는 얼굴 탐색(face de-occlusion) (연구기간 : 2020.06.01. ~ 2023.02.28)
 - Adversarial DA 성능 향상
 - GAN 기반 얼굴영상 변환 네트워크 성능 향상
- (포항공과대학교) 고품질의 영상 및 비디오 생성을 위한 빅 비주얼 데이터 기반의 고품질 사진 메이크업 SW 개발 (연구기간 : 2015.03.01. ~ 2022.12.31.)
 - 빅 비주얼 데이터 분석 및 영상 및 비디오 처리 기술
 - 다중 카메라 영상 합성 기술의 오픈소스 라이브러리화

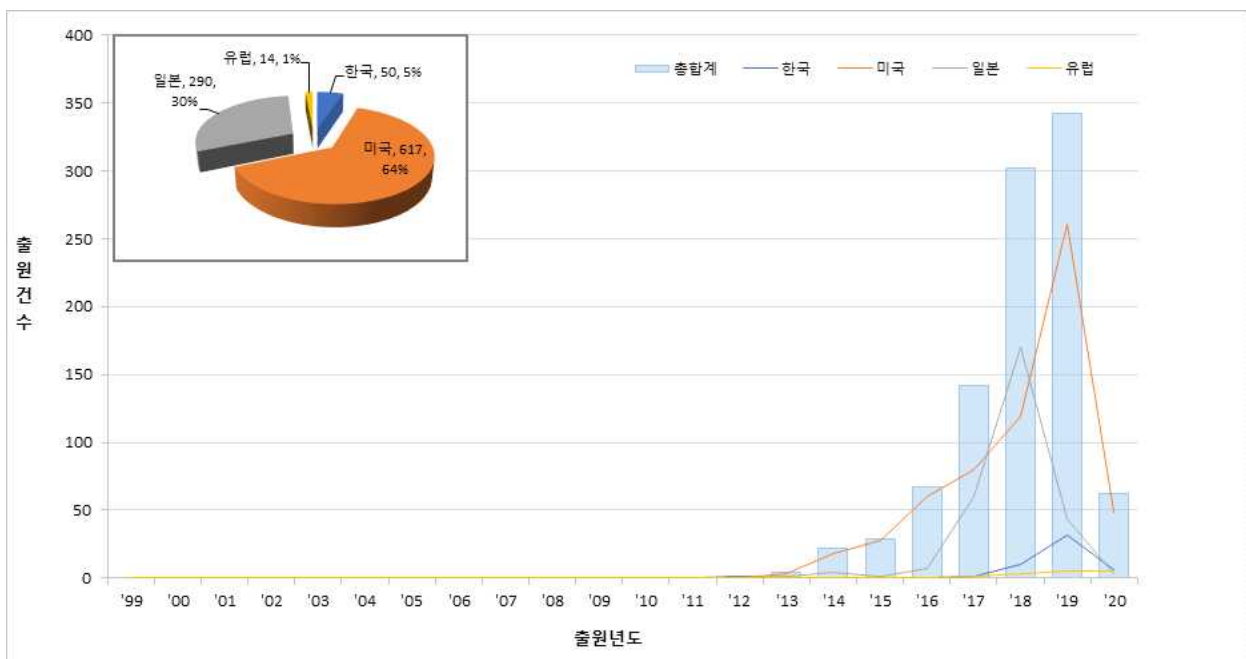
4. 특허 동향

가. 특허동향 분석

(1) 연도별 출원동향

- 합성 데이터 (Synthetic Data)는 '14년부터 급격한 성장을 보임
 - 각 국가별로 살펴보면 미국이 가장 활발한 출원활동을 보이고 있음
- 국가별 출원비중을 살펴보면 미국이 전체의 64%의 출원 비중을 차지하고 있어, 최대 출원국으로 합성 데이터 (Synthetic Data) 분야를 리드하고 있는 것으로 나타났으며, 일본은 30%, 한국은 5%, 유럽은 1% 순으로 나타남

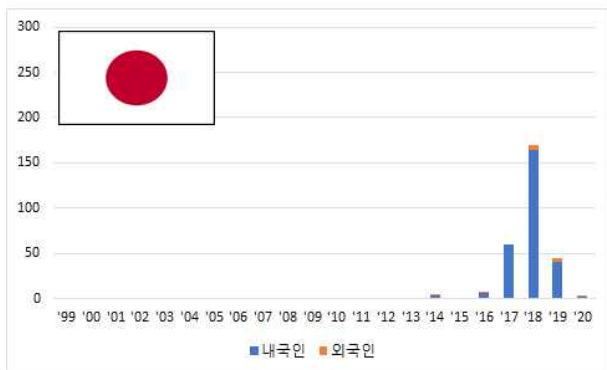
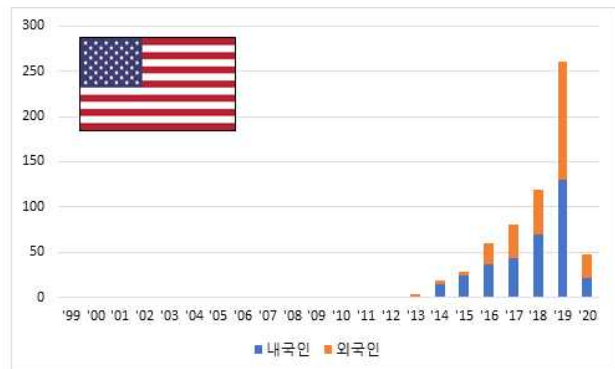
[합성 데이터 (Synthetic Data) 연도별 출원동향]



(2) 국가별 출원현황

- 한국의 출원현황을 살펴보면, '18년부터 해당 기술의 출원이 완만히 증가하는 추세
 - 내국인 위주의 출원이 진행되고 있음
 - 한국 기술의 양적 흐름은 미국과 상당히 유사
 - 미국의 출원 수에 비해 8% 정도의 수준을 보임
- 미국의 출원현황을 살펴보면 분석구간 초기부터 전체 특허기술의 출원 증감 흐름에 영향을 주고 있는 것으로 나타남. 미국의 경우, 한국에 비해 외국인의 비중이 큰 것으로 나타남
- 유럽의 출원현황은 출원수가 매년 10건 이하로, 뚜렷한 증감 동향이 나타나지 않음. 해당 기술 분야에서 일본 시장에 대한 관심도가 높지 않은 것으로 보임
- 일본의 출원현황을 살펴보면, '18년에 급격히 증가하는 추세이며, 미국과 유사한 동향을 보임

[국가별 출원현황]



(3) 기술 집중도 분석

□ 전략제품에 대한 최근 기술 집중도 분석을 위한 구간별 기술 키워드 분석 진행

- 전체구간(1999년~2020년)에서 Training Data, 학습 데이터, Machine Learning Model 등 키워드가 다수 도출
- 최근 구간 분석 결과, 최근 1구간(2012년~2015년)과 비교할 때 2구간(2016년~2020년)에서 Learning Data, Machine Learning, 학습 장치, 입력 데이터 키워드가 많이 등장하는 것으로 보아, 합성 데이터 분야에는 머신러닝 학습 데이터 증대 기술, 합성 데이터 탐지 기술 관련 연구개발이 활발한 것으로 추정

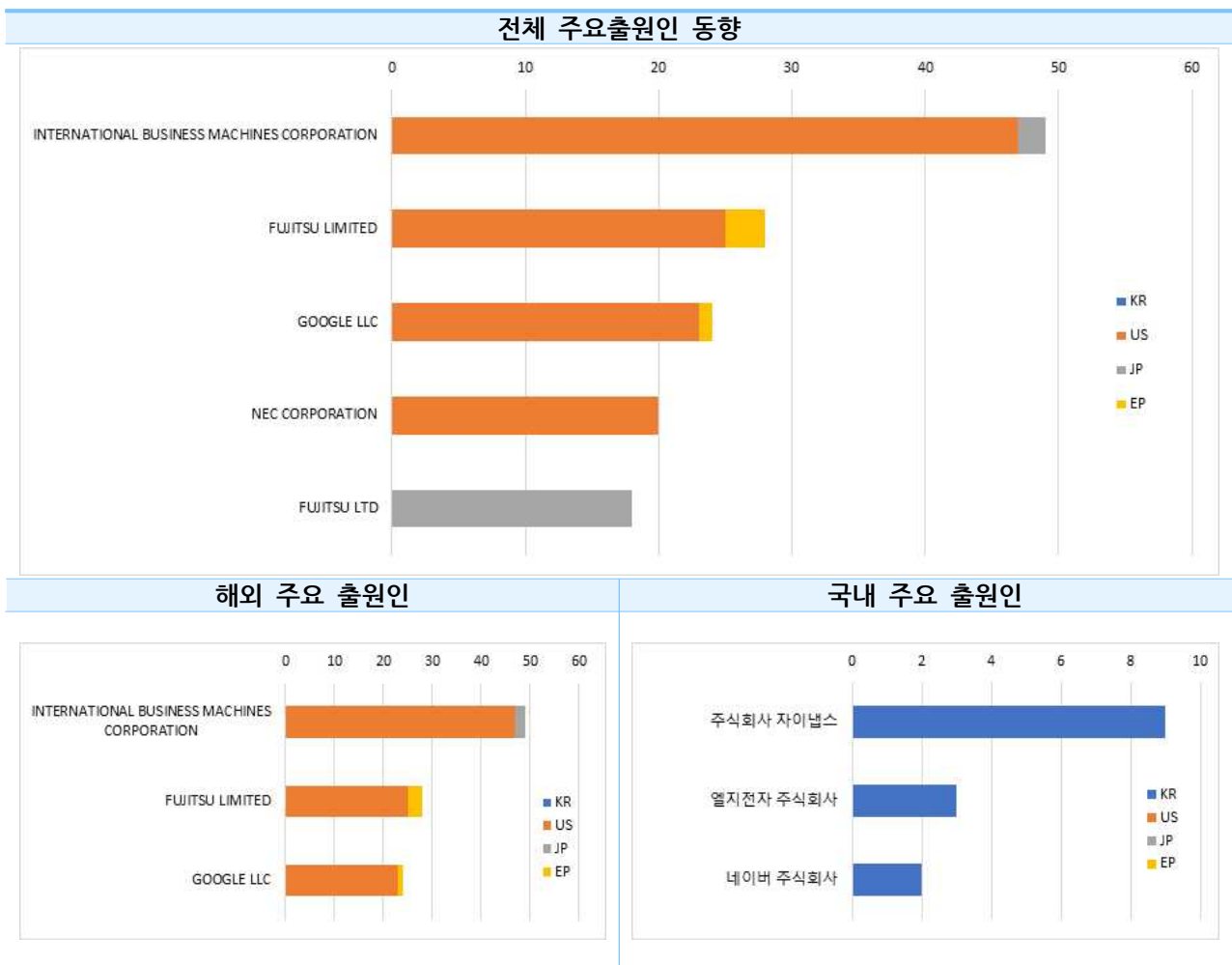
[특히 키워드 변화로 본 기술개발 동향 변화]

전체구간(1999년~2020년)	
	
<ul style="list-style-type: none"> • Training Data, 학습 데이터, Machine Learning Model, Computer Program Product, 기계 학습, Learning Data, Machine Learning, Computer Implemented, Rfid Tag, Training Data Set 	
최근구간(2012년~2020년)	
1구간(2012년~2015년)	2구간(2016년~2020년)
	
<ul style="list-style-type: none"> • Computer Program Product, Mobile Device, Training Data, Rfid Tag, Computer Implemented, Computing Device, Real Time, Inventory Data, Inventory Management, Supply Chain 	<ul style="list-style-type: none"> • Training Data, 학습 데이터, Machine Learning Model, Learning Data, 기계 학습, Machine Learning, Training Data Set, 학습 장치, Computer Program Product, 입력 데이터

나. 주요 출원인 분석

- 합성 데이터 (Synthetic Data)의 전체 주요출원인을 살펴보면, 주로 일본 및 미국 국적의 출원인이 다수 포함되어 있는 것으로 나타났으며, 제 1 출원인으로는 미국의 INTERNATIONAL BUSINESS MACHINES CORPORATION인 것으로 나타남
 - 제 1 출원인인 INTERNATIONAL BUSINESS MACHINES CORPORATION의 출원은 미국에 집중된 경향을 보임
- 합성 데이터 (Synthetic Data) 관련 기술로 데이터 분석을 다루는 대기업에 의한 출원이 대다수를 차지
 - 국내에서는 중소기업(개인), 대기업의 활발한 출원이 이루어짐

[합성 데이터 (Synthetic Data) 주요출원인]



(1) 해외 주요출원인 주요 특허 분석

◎ IBM(International Business Machines)

- IBM은 미국의 다국적 기술 기업으로, 합성 데이터와 관련하여, 원본 데이터를 기반으로 가상 데이터를 생성하고, 이를 트레이닝 데이터로 생성하는 기술에 특화된 특허를 다수 출원. 그 중 등록된 특허는 28건 보유

[IBM 주요특허 리스트]

등록번호 (출원일)	명칭	기술적용분야	도면
US10621284 (2019.06.28)	Training data update	2개의 그룹에서 2개의 트레이닝 데이터를 생성하고, 분류 결과를 기초로 갱신한 2개의 트레이닝 데이터를 생성	
US10657463 (2019.06.27)	Bot-based data collection for detecting phone solicitations	학습된 검출 모델들에 기초하여 독한 데이터(poison data) 또는 진실한 트레이닝 데이터(truthful training data) 중 하나로 상기 기록된 대화들을 분류하는 기술	
US10535019 (2018.07.31)	Bot-based data collection for detecting phone solicitations	불필요한 전화를 탐지하기 위한 bot 기반의 데이터 수집하기 위한 기술	
US10740895 (2018.06.25)	Generator-to-classifier framework for object classification	원본 데이터에 기초하여 생성된 데이터에 수신하고, 원본 데이터 및 상기 생성된 데이터에 기초하여 트레이닝 데이터를 생성하는 기술	
US10372826 (2017.12.18)	Training data update	2개의 그룹에서 2개의 트레이닝 데이터를 생성하고, 분류 결과를 기초로 갱신한 2개의 트레이닝 데이터를 생성	

* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ FUJITSU LIMITED

- FUJITSU LIMITED는 일본의 정보통신기술 기업으로, 학습 데이터를 생성하기 위한 방법에 관한 특허를 보유하고 있으며, 합성 데이터와 관련한 특허를 대부분 미국에 출원
 - 합성 데이터 기술과 관련하여 25건의 미국 및 3건의 유럽 출원을 진행하였으며, 그 중 등록된 특허는 1건으로 파악

[FUJITSU LIMITED 주요특허 리스트]

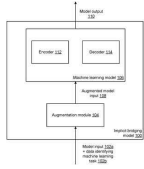
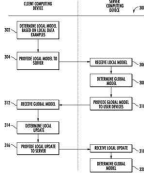
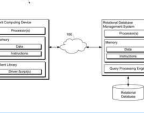
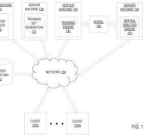
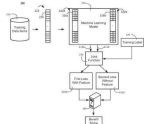
등록번호 (출원일)	명칭	기술적용분야	도면
US10303447 (2016.11.30)	Program generating apparatus and method therefor	파라미터 가변 프로그램에 대하여 선택적으로 설정 가능한 복수의 파라미터의 관계에 기초하여 화상 처리 프로그램의 생성 시간을 단축	

* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

© GOOGLE LLC

□ GOOGLE LLC는 검색 서비스를 제공하는 미국의 다국적 기업으로, 기계 학습 프레임워크 내에서 모델을 업데이트하는 기술에 특화된 특허를 다수 출원. 그 중 등록된 특허는 11건 보유

[GOOGLE LLC 주요특허 리스트]

등록번호 (출원일)	명칭	기술적용분야	도면
US10679148 (2019.05.03)	Implicit bridging of machine learning tasks	기계 학습 작업을 위한 암시적 브리징을 적용하는 예시적 시스템	
US10657461 (2017.09.07)	Communication efficient federated learning	분산된 트레이닝 데이터에 대해 고품질 집중형 모델이 학습되는 연합형 학습 프레임워크와 같은 기계 학습 프레임워크 내에서 모델 업데이트들의 전송을 위한 효율적인 통신 기술	
US10482394 (2017.06.13)	Large-scale in-database machine learning with pure sql	관계형 데이터베이스 내에서 일반화된 선형 모델들의 데이터베이스 내 생성을 위한 시스템	
US10586111 (2017.02.27)	Using machine learning to detect which part of the screen includes embedded frames of an uploaded video	복합 이미지들 내의 구성 이미지들을 식별하기 위해 기계 학습 모델을 트레이닝하기 위한 기술	
US10600000 (2016.12.02)	Regularization of machine learning models	머신 학습 모델에 의해 유지되는 특징 가중치들을 정규화하기 위한, 컴퓨터 저장 매체 상에 인코딩된 컴퓨터 프로그램 기술	

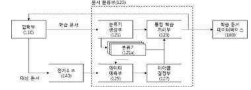

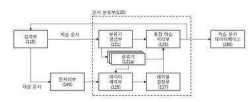
* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

(2) 국내 주요출원인 주요 특허 분석

◎ 주식회사 자이냅스

- 주식회사 자이냅스는 한국의 챗봇 및 빅데이터 솔루션을 제공하는 기업으로, '20년도 9월부터 출원을 시작하며 아직까지 등록된 특허는 없는 것으로 파악됨
 - 주식회사 자이냅스의 공개특허를 대상으로 분석한 결과, 합성 데이터 기술과 관련된 특허를 9건 출원

[주식회사 자이냅스 주요특허 리스트]


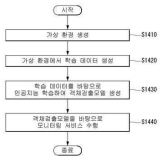
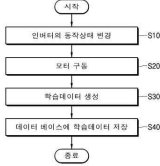
공개번호 (출원일)	명칭	기술적용분야	도면
KR2020-0098393 (2020.01.23)	문서 학습 장치 및 그 동작 방법	지도 학습 데이터 세트와 비 지도 학습 데이터들을 모두 이용하되,레이블이 표시된 학습 문서에 따라 비 지도 학습을 위한 분류기를 가변적으로 생성	
KR2020-0103173 (2019.02.12)	문서 학습 방법	가변 분류기를 이용하여 효율적 학습 프로세스를 제공하는 문서 학습 방법	
KR2020-0103174 (2019.02.12)	문서 학습 프로그램	인공지능 대화 서비스를 제공하기 위하여 가변 분류기를 이용하여 효율적 학습 프로세스를 실행할 수 있도록 하는 컴퓨터에서 실행시키기 위한 문서 학습 프로그램	

* 공개특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ 엘지전자 주식회사

- 엘지전자 주식회사는 한국의 전자 기기 제조 기업으로, '18년도 11월부터 출원을 시작하며 아직까지 등록된 특허는 없는 것으로 파악됨
 - 엘지전자 주식회사의 공개특허를 대상으로 분석한 결과, 합성 데이터 기술과 관련된 특허를 3건 출원

[엘지전자 주식회사 주요특허 리스트]

공개번호 (출원일)	명칭	기술적용분야	도면
KR2019-0110073 (2019.09.09)	인공 지능 모델을 갱신하는 인공 지능 장치 및 그 방법	저장된 학습 데이터를 이용하여 상기 저장된 인공 지능 모델을 갱신하는 프로세서	 <p>The diagram shows a process flow for updating an AI model. It starts with '100' at the top, leading to a box '110' containing '학습부' (Learning Unit). Below '110' are '120' (입력부 - Input Unit) and '130' (권장 프로세서 - Recommendation Processor). Arrows from '110', '120', and '130' point to a central box '140' labeled '프로세서' (Processor). From '140', arrows point to three output boxes: '150' (출력부 - Output Unit), '160' (추천부 - Recommendation Unit), and '170' (제어부 - Control Unit).</p>
KR2019-0096866 (2019.07.31)	차량 내부의 모니터링 방법 및 장치, 이를 위한 객체검출모델을 생성하는 3d 모델링부	학습 데이터를 바탕으로 인공지능 학습을 수행하여 객체검출모델을 생성하고, 객체검출모델을 이용하여 상기 실제 이미지에서 객체를 검출하는 기술	 <p>The diagram illustrates a process for generating a 3D model and performing object detection. It starts with '시작' (Start) leading to '기상 환경 생성' (Weather environment generation) - S1410. This is followed by '기상 환경에서 객체 모니터링 생성' (Object monitoring generation in weather environment) - S1420. Next is '학습 데이터를 바탕으로 인공지능 학습을 수행하여 객체검출모델 생성' (AI learning to generate object detection model based on learning data) - S1430. The final step is '객체검출모델을 이용하여 실제 이미지에서 객체를 검출하는 기술' (Technology for detecting objects in real images using the object detection model) - S1440. The process ends at '종료' (End).</p>
KR2020-0064197 (2018.11.23)	인버터의 고장 진단을 위한 머신러닝 방법 및 학습데이터를 통해 고장을 진단하는 인버터	인버터의 고장 진단을 위해, 인버터의 동작상태에 따른 지도 학습(supervised learning)을 통해 머신러닝을 수행	 <p>The diagram shows a supervised learning process for inverter fault diagnosis. It starts with '시작' (Start) leading to '인버터의 동작상태 변경' (Change inverter operating state) - S10. This is followed by '부하 구동' (Load driving) - S20, then '학습데이터 생성' (Learning data generation) - S30. The final step is '데이터베이스에 학습데이터 저장' (Store learning data in database) - S40. The process ends at '종료' (End).</p>

* 공개특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ 네이버 주식회사

□ 네이버 주식회사는 한국의 인터넷 서비스 기업으로, '19년도 1월부터 출원을 시작하며 아직까지 등록된 특허는 없는 것으로 파악됨

- 엘지전자 주식회사의 공개특허를 대상으로 분석한 결과, 합성 데이터 기술과 관련된 특허를 2건 출원

[네이버 주식회사 주요특허 리스트]

공개번호 (출원일)	명칭	기술적용분야	도면
KR2020-0078454 (2020.06.24)	딥러닝 모델 기반 키보드 오타 보정 방법 및 시스템	학습 데이터를 이용하여 딥러닝 모델을 학습시키고, 학습 데이터를 이용하여 사용자 입력에 따른 문법적 오류 여부를 결정 및 보정하는 기술	
KR2020-0092446 (2019.01.04)	어체 변환이 가능한 기계 번역 방법 및 시스템	기계 번역 모델의 학습을 위한 기존의 학습 데이터에 대한 변경 없이, 텍스트에 대한 어체를 간편하게 태깅할 수 있는 기계 번역 방법 및 시스템	

* 공개특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

다. 기술진입장벽 분석

(1) 기술 집중력 분석

- 합성 데이터(Synthetic Data) 관련 기술에 대한 시장관점의 기술독점 현황분석을 위해 집중률 지수(CRn: Concentration Ratio n, 상위 n개사 특허점유율의 합) 분석 진행
 - 상위 4개 기업의 시장점유율이 0.12로 합성 데이터 (Synthetic Data) 분야에 있어서 독과점 정도는 낮은 수준으로 판단
 - 국내 시장에서 중소기업의 점유율 분석결과 0.8로 해당 기술에 대하여 중소기업이 진입하여 활동하고 있는 것으로 파악

[주요출원인의 집중력 및 국내시장 중소기업 집중력 분석]

주요 출원인 집중력	주요출원인	출원건수	특허점유율	CRn	n
	INTERNATIONAL BUSINESS MACHINES CORPORATION(미국)	49	5.0%	0.05	1
	FUJITSU LIMITED(일본)	28	2.9%	0.08	2
	GOOGLE LLC(미국)	24	2.5%	0.10	3
	NEC CORPORATION(일본)	20	2.1%	0.12	4
	FUJITSU LTD(일본)	18	1.9%	0.14	5
	MICROSOFT TECHNOLOGY LICENSING, LLC(미국)	17	1.8%	0.16	6
	OMRON CORP(일본)	17	1.8%	0.18	7
	NIPPON TELEGR & TELEPH CORP (일본)	16	1.6%	0.19	8
	FANUC CORPORATION(일본)	15	1.5%	0.21	9
	HITACHI LTD(일본)	15	1.5%	0.23	10
	전체	971	100%	CR4=0.12	
국내시장 중소기업 집중력	출원인 구분	출원건수	특허점유율	CRn	n
	중소기업(개인)	59	79.7%	0.80	
	대기업	9	12.2%		
	연구기관/대학	6	8.1%		
	전체	74	100%	CR중소기업=0.80	

(2) 특허소송 현황 분석

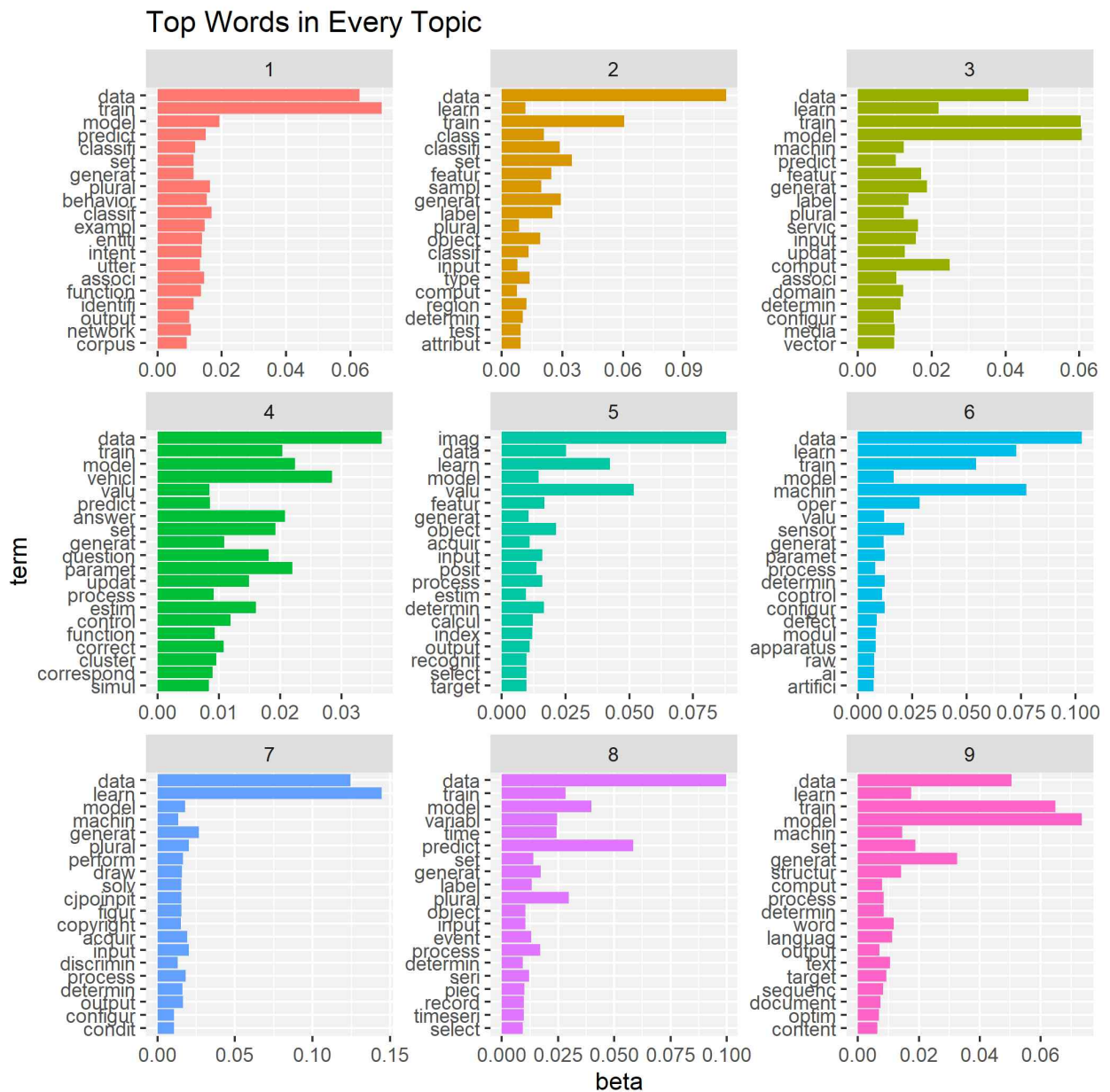
- 합성 데이터(Synthetic Data) 분야 관련 특허소송 이력은 검색되지 않음
 - 따라서 국내기업이 미국시장에 진입하는 경우, 해당 분야를 선점할 수 있을 것으로 판단
 - 다만, 머신 러닝 기술 중 앙상블 러닝과 관련해서는 1개의 소송이 검색

5. 요소기술 도출

가. 특허 기반 토픽 도출

- 825개의 특허의 내용을 분석하여 구성 성분이 유사한 것끼리 클러스터링을 시도하여 대표성이 있는 토픽을 도출

[합성 데이터 (Synthetic Data)에 대한 토픽 클러스터링 결과]



나. LDA¹²⁾ 클러스터링 기반 요소기술 도출

[LDA 클러스터링 기반 요소기술 키워드 도출]

No.	상위 키워드	대표적 관련 특허	요소기술 후보
클러스터 01	train, data, model, classify, plural, behavior, predict, example, associate, entity	<ul style="list-style-type: none"> Systems and methods for generating and applying a secure statistical classifier Systems and methods for machine learning-based multi-intent segmentation and classification 	-
클러스터 02	data, train, set, generate, classify, label, feature, class, sample, object	<ul style="list-style-type: none"> Target re-identification Unsupervised spoofing detection from traffic data in mobile networks 	머신러닝 학습 데이터 증대 기술
클러스터 03	model, train, data, computing, learn, generate, feature, service, input, label	<ul style="list-style-type: none"> Digital assistant system providing advanced customizable services for a user Method and system for detecting and preventing an imminent failure in a target system 	-
클러스터 04	data, vehicle, model, parameter, answer, train, set, question, estimate, update	<ul style="list-style-type: none"> Training of vehicles to improve autonomous capabilities Fusion splicing system, fusion splicer and method of determining rotation angle of optical fiber 	데이터 비식별화 기술
클러스터 05	image, value, learn, data, object, feature, determine, process, input, model	<ul style="list-style-type: none"> Learning identification device and learning identification method Target recognition device, target recognition method, program and convolution neural network 	이미지 데이터 합성 기술
클러스터 06	data, machine, learn, train, operate, sensor, model, configure, parameter, determine	<ul style="list-style-type: none"> Systems and methods for intelligently curating machine learning training data and improving machine learning model performance Blockchain recordation of device usage 	데이터 도메인 변환 기술
클러스터 07	learn, data, generate, input, plural, acquire, process, model, perform, output	<ul style="list-style-type: none"> Leaning device and leaning method Anticipation model maintenance system, method for maintaining anticipation model, and anticipation model maintenance program 	합성 데이터 탐지 기술
클러스터 08	data, predict, model, plural, train, variable, time, generate, process, set	<ul style="list-style-type: none"> Predicting no-defect-found physical failure analysis results using bayesian inference and generalized linear models Systems for time-series predictive data analytics, and related methods and apparatus 	시계열 데이터 합성 기술
클러스터 09	model, train, data, generate, set, learn, machine, structure, word, language	<ul style="list-style-type: none"> Generating a targeted summary of textual content tuned to a target audience vocabulary Computer model machine learning based on correlations of training data with performance trends 	텍스트 데이터 합성 기술

12) Latent Dirichlet Allocation

다. 특허 분류체계 기반 요소기술 도출

- 합성 데이터 (Synthetic Data) 관련 특허에서 총 10개의 주요 IPC코드(메인그룹)를 산출하였으며, 각 그룹의 정의를 기반으로 요소기술 키워드를 아래와 같이 도출

[IPC 분류체계에 기반한 요소기술 도출]

IPC 기술트리		
(서브클래스) 내용	(메인그룹) 내용	요소기술 후보
(G06F) 전기에 의한 디지털 데이터처리	• (G06F-021) 부정행위로부터 프로그램 또는 데이터, 그 컴퓨터 부품을 보호하기 위한 보안 장치	합성 데이터 탐지 기술
	• (G06F-017) 디지털 컴퓨팅 또는 데이터 프로세싱 장비, 방법으로서 특정 기능을 위해 특히 적합한 형태의 것	-
	• (G06F-016) 정보 검색; 그 데이터베이스 구조; 그 파일 시스템 구조	-
(G06K) 데이터의 인식; 데이터의 표시; 기록매체; 기록매체의 취급	• (G06K-009) 인쇄문자, 손으로 쓴 문자를 독취하거나 인식 또는 패턴을 인식하기 위한 방법 또는 장치, 예. 지문인식	텍스트 데이터 합성 기술
(G06N) 특정 계산모델 방식의 컴퓨터시스템	• (G06N-007) 특정수리모델에 기반을 둔 컴퓨터시스템	-
	• (G06N-005) 지식기반모델을 이용한 컴퓨터시스템	머신러닝 학습 데이터 증대 기술
	• (G06N-020) 머신 러닝	머신러닝 학습 데이터 증대 기술
	• (G06N-003) 생체모델기반 컴퓨터시스템	-
(G06T) 이미지 데이터 처리 또는 발생 일반	• (G06T-007) 화상 분석, 예를 들면 비트맵으로부터 비비트맵 (non bit-mapped)으로	이미지 데이터 합성 기술
(H04L) 디지털 정보의 전송, 예. 전신통신	• (H04L-029) 그룹 1/00에서 H04L 27/00의 하나에도 포함되지 않는 배치, 장치회로 또는 시스템	-

라. 최종 요소기술 도출

- 산업·시장 분석, 기술(특허)분석, 전문가 의견, 타부처 로드맵, 중소기업 기술수요를 바탕으로 로드맵 기획을 위하여 요소기술 도출
- 요소기술을 대상으로 전문가를 통해 기술의 범위, 요소기술 간 중복성 등을 조정·검토하여 최종 요소기술명 확정

[합성 데이터 (Synthetic Data) 분야 요소기술 도출]

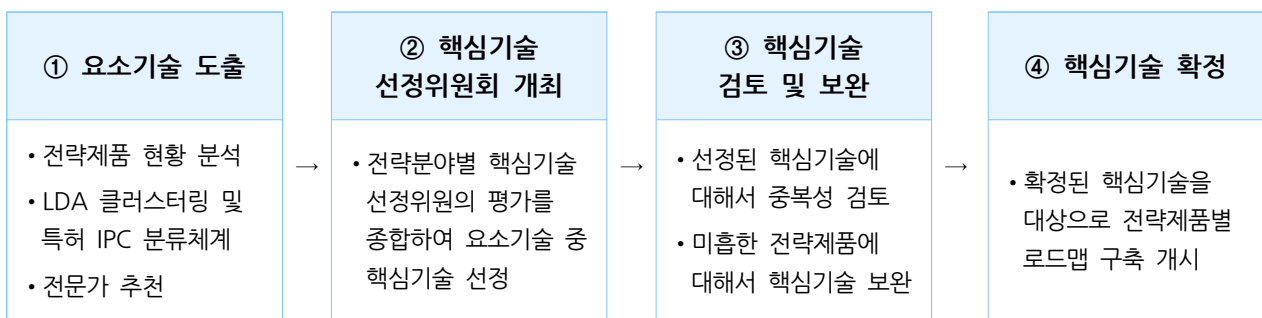
요소기술	출처
시계열 데이터 합성 기술	특허 클러스터링, 전문가추천
이미지 데이터 합성 기술	특허 클러스터링, IPC 기술체계, 전문가추천
텍스트 데이터 합성 기술	특허 클러스터링, IPC 기술체계, 전문가추천
데이터 도메인 변환 기술	특허 클러스터링, 전문가추천
머신러닝 학습 데이터 증대 기술	특허 클러스터링, 전문가추천
합성 데이터 탐지 기술	특허 클러스터링, IPC 기술체계, 전문가추천
데이터 비식별화 기술	전문가 추천

6. 전략제품 기술로드맵

가. 핵심기술 선정 절차

- 특허 분석을 통한 요소기술과 기술수요와 각종 문헌을 기반으로 한 요소기술, 전문가 추천 요소기술을 종합하여 요소기술을 도출한 후, 핵심기술 선정위원회의 평가과정 및 검토/보완을 거쳐 핵심기술 확정
- 핵심기술 선정 지표: 기술개발 시급성, 기술개발 파급성, 기술의 중요성 및 중소기업 적합성
 - 장기로드맵 전략제품의 경우, 기술개발 파급성 지표를 중장기 기술개발 파급성으로 대체

[핵심기술 선정 프로세스]



나. 핵심기술 리스트

[합성 데이터 (Synthetic Data) 분야 핵심기술]

핵심기술	개요
이미지 데이터 합성 기술*	• 기존 이미지 데이터의 특성을 고려하여 새로운 이미지 데이터를 가상으로 합성하는 기술
머신러닝 학습 데이터 증대 기술	• 머신러닝 모델 성능 개선을 목적으로, 모델 학습 과정에서 활용될 수 있는 추가 학습 데이터를 합성하는 기술
합성 데이터 탐지 기술	• 실제 데이터와 합성 데이터를 구별하는 기술
데이터 비식별화 기술	• 원본 데이터의 활용을 저해하지 않으면서 민감한 정보의 식별이 불가능하도록 하는 합성 데이터 생성
시계열 데이터 합성 기술	• 기존 시계열 데이터의 특성을 고려하여 새로운 시계열 데이터를 가상으로 합성하는 기술

* 표시는 생태계 취약 기술을 의미

다. 중소기업 기술개발 전략

- 여러 산업 분야의 다양한 도메인에 대해서, 각 도메인의 데이터 특성을 고려하여 특화된 합성 데이터 생성 기술 개발
- 실제 제품 및 서비스와 연계한 맞춤형 합성 데이터 생성 기술 개발
- 데이터 보안 및 프라이버시 보호, 데이터 악용 방지를 위한 데이터 비식별화 및 합성 데이터 탐지 기술 개발

라. 기술개발 로드맵

(1) 중기 기술개발 로드맵

[합성 데이터 (Synthetic Data) 분야 중기 기술개발 로드맵]

합성 데이터 (Synthetic Data)	문제 상황 및 활용 목적에 따라 요구되는 데이터를 직접 획득하는 대신 가상으로 합성			
	2021년	2022년	2023년	최종 목표
이미지 데이터 합성 기술				목표 의도에 따른 타당한 신규 이미지 데이터 합성
머신러닝 학습 데이터 증대 기술				해당 기술 비활용 대비 머신러닝 모델 평균 성능 개선도 20% 이상
합성 데이터 탐지 기술				실제 데이터와 학습 데이터 간 분류 정확도 90% 이상
데이터 비식별화 기술				데이터 활용 성능 저하를 최소화하는 데이터 익명화
시계열 데이터 합성 기술				목표 의도에 따른 타당한 신규 시계열 데이터 합성

(2) 기술개발 목표

□ 최종 중소기업 기술로드맵은 기술/시장 니즈, 연차별 개발계획, 최종목표 등을 제시함으로써 중소기업의 기술개발 방향성을 제시

[합성 데이터 (Synthetic Data) 분야 핵심기술 연구목표]

핵심기술	기술요구사항	연차별 개발목표			최종목표	연계R&D 유형
		1차년도	2차년도	3차년도		
이미지 데이터 합성 기술	기존 데이터 특성을 고려한 신규 이미지 데이터 합성 수준 고도화	기존 이미지 변형을 통한 신규 이미지 데이터 생성	기존 이미지 데이터 분포 추정 및 샘플링을 통한 신규 이미지 데이터 생성	이미지 데이터 도메인 변환 기술	목표 의도에 따른 타당한 신규 이미지 데이터 합성	산학연
머신러닝 학습 데이터 증대 기술	해당 기술 비활용 대비 머신러닝 모델 평균 성능 개선도 (%)	5% 이상	10% 이상	20% 이상	해당 기술 비활용 대비 머신러닝 모델 평균 성능 개선도 20% 이상	산학연
합성 데이터 탐지 기술	실제 데이터와 학습 데이터 간 분류 정확도 (%)	70% 이상	85% 이상	95% 이상	실제 데이터와 학습 데이터 간 분류 정확도 90% 이상	기술혁신
데이터 비식별화 기술	데이터 비식별화 수준	데이터 활용 성능 저하 5% 이내의 데이터 가명화 (추가 정보 확보 시 재식별화 가능성 존재)	데이터 활용 성능 저하 20% 이내의 데이터 익명화 (재식별화 불가)	데이터 활용 성능 저하 10% 이내의 데이터 익명화 (재식별화 불가),	데이터 활용 성능 저하를 최소화하는 데이터 익명화	기술혁신
시계열 데이터 합성 기술	기존 데이터 특성을 고려한 신규 시계열 데이터 합성 수준 고도화	기존 시계열 데이터 변형을 통한 신규 시계열 데이터 생성	기존 시계열 데이터 분포 추정 및 샘플링을 통한 신규 시계열 데이터 생성	시계열 데이터 도메인 변환 기술	목표 의도에 따른 타당한 신규 시계열 데이터 합성	산학연

