

# 지능정보기술 R&D의 선도형(First Mover) 연구 촉진과 성과 확산 가속화를 위한 지원 체계 및 제도 개선 방안 연구

A Study on Improving the Information Technology  
R&D to Promote for First Mover Research and  
Resulting Diffusion

2018. 01

연구기관 : 정보통신정책연구원



방송통신정책연구 17-방통-45

# 지능정보기술 R&D의 선도형(First Mover) 연구 촉진과 성과 확산 가속화를 위한 지원 체계 및 제도 개선 방안 연구

(A Study on Improving the Information Technology  
R&D to Promote for First Mover Research and  
Resulting Diffusion)

최계영/이학기/김경훈/김민식/정원준/양수연/이가희

2018. 01

연구기관 : 정보통신정책연구원



과학기술정보통신부  
Ministry of Science and ICT

**IITP** 정보통신기술진흥센터  
Institute for Information & Communications Technology Promotion

이 보고서는 2017년도 과학기술정보통신부 방송통신발전기금 방송통신 정책연구사업의 연구결과로서 보고서 내용은 연구자의 견해이며, 과학기술정보통신부의 공식입장과 다를 수 있습니다.

# 제 출 문

과학기술정보통신부 장관 귀하

본 보고서를 『지능정보기술 R&D의 선도형(First Mover) 연구 촉진  
과 성과 확산 가속화를 위한 지원 체계 및 제도 개선 방안 연구』의  
연구결과보고서로 제출합니다.

2018년 1월

연구기관: 정보통신정책연구원

총괄책임자: 최계영 선임연구위원

참여연구원: 이학기 부연구위원

김경훈 부연구위원

김민식 부연구위원

정원준 연구위원

양수연 연구위원

이가희 연구위원

# 목 차

요약문 .....	ix
<b>제 1 장 서 론 .....</b>	<b>1</b>
제 1 절 연구의 배경 .....	1
제 2 절 연구의 목표 및 주요내용 .....	3
<b>제 2 장 지능정보기술 및 R&amp;D 동향 .....</b>	<b>6</b>
제 1 절 AI 기술 개요 .....	6
1. AI SW 기술 분류 .....	6
2. AI 기술수준 및 기술격차 .....	9
3. AI 관련 특허 출원 .....	11
제 2 절 국내 지능정보기술 및 R&D 동향 .....	15
1. AI 발전 현황 .....	15
2. 지능정보기술 R&D 동향 (AI HW, AI SW, 기초과학) .....	16
제 3 절 해외 지능정보기술 R&D 동향 .....	19
1. AI 활용을 위한 요소 기술 .....	19
2. 해외 AI 시장 규모 및 전망 .....	29
3. 국가별 지능정보기술 R&D 전략 및 정책 .....	31
4. 해외 AI 기술 동향과 주요 사례 .....	44
5. 사례 분석을 통한 시사점 .....	50
제 4 절 소 결 .....	52
<b>제 3 장 AI 기업 학습데이터 .....</b>	<b>53</b>
제 1 절 AI 기술과 데이터 .....	53
제 2 절 전문가 활용 인터뷰 .....	58

1. 데이터 수요조사 .....	58
2. 데이터 활용 제한사항 .....	62
3. 전문가 인터뷰 결과 활용 방안 .....	63
제3절 조사통계 .....	64
1. 연구조사대상 .....	64
2. 연구방법론 .....	65
3. 연구내용 결과 .....	67
제4절 소 결 .....	70
<b>제 4 장 개방형 R&amp;D 생태계 조성 방안 .....</b>	<b>72</b>
제 1 절 4세대 연구 패러다임과 오픈 사이언스(Open Science) 개념 .....	72
제 2 절 오픈 사이언스(Open Science) 동향 및 정책적 시사점 .....	76
1. 오픈 사이언스 정책의 확산 .....	76
2. 오픈 사이언스 정책 추진 현황 및 이행 실태 .....	78
3. 국내외 오픈 사이언스 추진 비교 .....	87
4. 국내 오픈 사이언스 정책 관련 시사점 .....	88
제 3 절 오픈 사이언스 기반의 AI 활성화 방안 .....	91
1. 오픈 사이언스와 AI .....	91
2. 오픈 사이언스 관점의 AI 활성화 방향 .....	94
3. AI 활성화 방안으로써 디지털 플랫폼 .....	95
<b>제 5 장 결론 및 정책 제언 .....</b>	<b>98</b>
[참고 1] .....	102
[참고 2] .....	108
[부 록] 통 계 표 .....	118
참고문헌 .....	139

## 표 목 차

〈표 2-1〉 AI 5대 핵심기술 및 15대 세부기술 분류 .....	8
〈표 2-2〉 2016년 AI SW 기술수준 .....	10
〈표 2-3〉 2017년 AI 연구 분야별 주요 내용 및 예산 .....	17
〈표 2-4〉 AI기술 R&D 예산 현황 .....	18
〈표 2-5〉 국내 주요 AI R&D 과제 현황 .....	19
〈표 2-6〉 세계 빅데이터 시장 규모 .....	20
〈표 2-7〉 데이터를 활용한 해외 비즈니스 성공 사례 .....	21
〈표 2-8〉 미국 기업의 AI-데이터 활용 사항 .....	23
〈표 2-9〉 미국·중국 간 주요 IT기업 데이터 경쟁현황 .....	25
〈표 2-10〉 일본 8대 전자기업의 AI 활용 현황 .....	27
〈표 2-11〉 세계 클라우드 시장규모 .....	28
〈표 2-12〉 세계 AI 시장 규모 .....	29
〈표 2-13〉 세계 음성인식 비서 디바이스 출하량 .....	30
〈표 2-14〉 세계 자율주행자동차 출하량 .....	30
〈표 2-15〉 AI 기술 분야별 시장 규모 전망 .....	31
〈표 2-16〉 (미국) 과학 및 공학 지원 내용 .....	32
〈표 2-17〉 (미국) 빅데이터 R&D 전략계획 세부 내용 .....	33
〈표 2-18〉 (미국) 7대 AI R&D 전략 .....	35
〈표 2-19〉 미래투자전략 2017 실현을 위한 과제 .....	37
〈표 2-20〉 (일본) 과학기술 국가의 실현을 위한 2개의 주요 정책 및 계획 .....	39
〈표 2-21〉 (일본) AI 산업화 로드맵 세부 내용 .....	40
〈표 2-22〉 일본의 AI 기간별 육성 방안 .....	41
〈표 2-23〉 중국 빅데이터 산업발전계획 실행원칙 .....	43
〈표 2-24〉 AI 5개 핵심기술 분류에 따른 해외 기술 동향 .....	45

〈표 2-25〉 산업별 해외 AI 기술 동향 .....	46
〈표 3-1〉 공공/민간데이터 정책 추진 현황 .....	57
〈표 3-2〉 AI 데이터 분류 체계 마련에 대한 전문가 인터뷰 질문사항(안) .....	59
〈표 3-3〉 데이터 수요 조사 방안 3단계 .....	60
〈표 3-4〉 전문가 인터뷰 결과요약 .....	60
〈표 3-5〉 UCI Machine Learning Repository 분야별 데이터셋 리스트 .....	62
〈표 3-6〉 각 산업별 수요데이터 양식 .....	63
〈표 3-7〉 데이터 셋 연구조사 현황 .....	64
〈표 3-8〉 AI산업 모집단 구축 절차 .....	65
〈표 3-9〉 AI SW개발을 위한 데이터 확보의 어려움 여부 .....	67
〈표 3-10〉 AI 기술-적용 분야 데이터셋 수요 영역별 현황 .....	68
〈표 3-11〉 국가별 지능정보화 지향 이유 .....	70
〈표 4-1〉 6가지 Open Science 공유 대상의 내용 .....	75
〈표 4-2〉 오픈 사이언스 이행 및 도입 .....	79
〈표 4-3〉 오픈 사이언스 유관 국내 사업 및 법제도 동향 .....	84
〈표 4-4〉 오픈 사이언스 유관 국내 기업 동향 .....	86
〈표 4-5〉 오픈 사이언스와 AI 생태계 비교 .....	92
〈표 4-6〉 주요 글로벌 ICT 기업의 AI 알고리즘 공개 현황 .....	93
〈표 4-7〉 AI 데이터셋 공개 현황 .....	94
〈표 4-8〉 AI 학습용 데이터 셋 구축 현황 .....	97

# 그 립 목 차

[그림 2-1] AI 기술 격차 .....	11
[그림 2-2] AI 분야 국가 연도별 특허출원 동향 .....	12
[그림 2-3] AI 기술 특허출원(2000-2015) .....	12
[그림 2-4] 세계 상위 AI 특허 보유 R&D 기업의 국가별 비중(본사 위치 기준) (2012-2014) .....	13
[그림 2-5] 핵심기술 연도별 특허 출원 동향 .....	14
[그림 2-6] 핵심기술 국가별 특허출원 동향 .....	14
[그림 2-7] 빅데이터 세계 시장 규모 .....	21
[그림 2-8] 주요 글로벌 기업의 보유 데이터와 AI 플랫폼화 현황 .....	22
[그림 2-9] 중국 빅데이터 산업규모 증가율 .....	24
[그림 2-10] 2014~2019년 중국 AI 산업규모 .....	25
[그림 2-11] 데이터 이용 및 활용을 위한 기본 사이클 .....	27
[그림 2-12] (미국) 연방정부 지원 예산('12~'15) .....	32
[그림 2-13] AI R&D 전략계획 구조 .....	35
[그림 2-14] 공통 기반기술과 데이터 결합을 통한 제품·서비스 산출 .....	38
[그림 2-15] 일본 AI 연구기관 및 센터 .....	39
[그림 2-16] 중국의 빅데이터 산업 성장의 7대 핵심 과제 .....	42
[그림 3-1] AI의 발전과정 .....	53
[그림 3-2] AI 개발 패러다임의 변화 .....	54
[그림 3-3] AI의 예측정확도와 설명력 관계 .....	55
[그림 3-4] 데이터 개방 및 활용 현황 .....	56
[그림 3-5] 데이터 셋 유형화 분류 및 분석 .....	64
[그림 3-6] 조사항목 .....	66
[그림 3-7] AI 기술분야 데이터 우선순위 .....	69

[그림 3-8] AI 적용 데이터 우선 순위 .....	69
[그림 4-1] 연구 패러다임의 변화 .....	73
[그림 4-2] 오픈 사이언스의 개념과 이해 .....	74
[그림 4-3] 6가지 오픈 사이언스 공유 대상 .....	75
[그림 4-4] 오픈 액세스정책의 확대(2005년~2017년) .....	76
[그림 4-5] 오픈 액세스정책 도입 전망(2002년~2020년) .....	77
[그림 4-6] 상위 10개 국가별 연구 성과의 개방(Open Access) 정책 수 .....	78
[그림 4-7] 주요 국가별 오픈 액세스 저널(Open Access Journal) 현황 비교 .....	82
[그림 4-8] 국가 연구데이터플랫폼 개념도 .....	90
[그림 4-9] 계층 구조의 데이터 공유·활용 생태계 개념도 .....	91

# 요 약 문

## 1. 제 목

지능정보기술 R&D의 선도형(First Mover) 연구 촉진과 성과 확산 가속화를 위한 지원 체계 및 제도 개선 방안 연구

## 2. 연구목적 및 필요성

AI 기술은 전통적인 ICT 산업 부문을 넘어 전통산업을 포괄하는 매우 광범위한 영역으로 확대되어 가고 있다. 이러한 과정에서 단순히 AI 기술 자체뿐만 아니라 응용대상 영역에서 얻어지는 데이터 및 사용자의 피드백 등이 점차 중요해지고 있다. 이러한 점은 AI 기술의 보급 및 확산을 위해서는 개방형 혁신 생태계를 통하여 다양한 분야에서 AI 기술을 응용하는 주체들이 모두 협력하여야 한다는 것을 말한다. 따라서 AI 알고리즘의 표준화 및 오픈소스 라이브러리화 작업이 필요하며, AI 기술의 개발에 필수적인 테스트 및 트레이닝용 데이터 셋의 확보가 요구된다. 그러나 데이터의 확보에는 많은 비용이 소요되어, 규모가 작은 스타트업의 입장에서 이러한 데이터의 확보는 쉽지 않은 일이다.

정부 및 공공부문의 역할은 이러한 AI 기술 생태계에서 훈련용 데이터 셋과 오픈소스 라이브러리, 기술평가기준 정립과 같은 공공재의 공급을 촉진하는 것에 있다. 특히 정부가 축적한 방대한 양의 데이터를 민간 스타트업 생태계와 공유하는 것이 필요하다. 즉, 민관 협력 기반의 AI R&D 전략 수립을 통해 개방 및 협력 중심의 R&D 생태계를 조성하고 AI 기술력을 조기에 확보하는 것이 필요한 시점이다.

본 연구는 4차 산업혁명의 핵심 동인으로 인식되고 있는 AI 기술의 확산을 위해 선도형 R&D의 촉진 및 R&D 성과의 산업 확산을 위한 지원체계 및 제도 개선 방안 등의 정책 방안 마련을 목표로 하고 있다. 특히, 선도형 R&D를 촉진시키기 위해 개방형 R&D 생태계를

조성하는 방안을 마련하고, 이에 대한 세부 전략을 제시한다. 이를 위해 AI R&D 기술의 단계별 발전 현황을 살펴보고, 국내외 AI 관련 R&D 정책과 글로벌 기업 동향을 파악한다. 또한 선도형 R&D 생태계를 조성하기 위한 핵심 방안으로 오픈 사이언스 체계를 구축하는 것을 제시하면서 우리나라를 포함한 세계 각국의 오픈 사이언스 관련 정책을 분석하고, 이를 통해 시사점을 도출한다. 도출된 시사점을 통해 향후 우리나라에 오픈 사이언스 체계를 구축하기 위한 세부 전략을 제시한다.

### 3. 연구의 구성과 범위

연구의 구성 및 범위는 다음과 같다. 첫째, 먼저 AI 기술의 개념 및 분류를 살펴보고, 타 기관의 기술 수준 분석 결과와 AI 관련 특허 출원 수 분석 등을 통해 우리나라의 AI 기술 수준을 고찰하였다. 또한, AI의 단계별 발전 시나리오를 제시하고, 발전 현황을 분석하였다. 이와 더불어 국내외 AI 기술 및 R&D 동향 파악을 통해 현재 국내 AI 기술이 가진 문제점을 진단하였다.

둘째, 선도형 R&D를 촉진시키기 위한 방안으로 개방형 R&D 생태계 조성 방안을 제시하였다. 개방형 R&D 생태계를 조성하기 위한 하나의 전략으로 오픈 액세스, 오픈 데이터, 오픈 콜라보레이션으로 구성되는 오픈 사이언스 구축 방안을 제시하였다. 이를 위해 먼저 오픈 사이언스의 개념을 정리하고, 개방형 R&D에 있어 선도국이라 할 수 있는 미국을 포함한 각국의 오픈 사이언스 정책 추진 현황 및 이행 실태를 분석하고, 현재 우리나라의 오픈 사이언스 관련 정책과 비교 분석을 통해 시사점을 도출하였다. 또한 도출된 시사점을 통해 오픈 사이언스를 활성화하기 위한 방안을 오픈 액세스, 오픈 데이터, 오픈 콜라보레이션 각 측면에서 제시하였다.

마지막으로 선도형 R&D를 촉진시키기 위한 전반적인 정책방향과 세부 정책과제들을 제시하였다.

## 4. 연구 내용과 결과

본 연구의 주요 결과를 살펴보면 다음과 같다. 지능정보기술 기술이 발전함에 따라 AI과 기반이 되는 기술에 대한 관심이 늘어났다. 과학기술정보통신부 및 정보통신기술진흥센터(IITP)에 따르면 AI 기술은 ‘학습 및 추론’, ‘상황이해’, ‘언어이해’, ‘시각이해’, ‘인지 및 인식 기술’ 등 총 5개의 핵심기술로 구분된다. 한국지식재산연구원(2016)에 따르면 이러한 5개 핵심 기술은 이하 15대 세부기술로 분류가 가능하며 ‘지식표현’, ‘지식베이스’, ‘감정이해’, ‘공간이해’, ‘협력지능’, ‘자가기해’ 외 9개가 이에 속한다.

AI SW 분야는 점차 다양한 분야의 가변적인 작업에도 대응 가능한 범용 AI 기술(General AI)로 발전해 나갈 전망이다. AI HW분야는 인공지능명망 학습에 필요한 대규모 연산을 고속·저지연·저전력으로 구현하기 위한 기술을 중심으로 개발되고 있다. AI 연계 데이터·네트워크 기술 측면에서는 X(ICBM) for AI 또는 AI for X 형태로 AI SW기술과 긴밀하게 연계될 것으로 전망된다. 기초과학은 최신 AI 기술(딥 러닝 등)의 이론적 한계를 극복하기 위해 뇌과학·산업수학 연구에 탄력이 예상되며, 중장기적으로는 인간·기계 협업 지원을 위한 뇌-기계 인터페이스(BCI) 분야가 특히 발전할 전망이다.

BCC Research(2015)에 따르면 세계 AI 시장규모는 급속히 성장하고 있으며, 음성인식 비서 디바이스 및 자율주행 자동차 출하량 또한 급속 성장 추세이다. 특히 딥러닝 기술이 AI 기술 중 가장 높은 시장 성장률을 보이고 있으며, 그 다음으로는 시각 인식 분야, 자연어 처리 순으로 시장 규모가 증가하고 있다.

핵심기술 수준 차이 측면에서 AI SW분야는 미국(100%) > 유럽(86.3%) > 일본(81.9%) > 한국(73.9%) > 중국(71.8%) 순으로 기술 수준 차이가 있다. 선진국 대비 한국은 2.2년의 기술 격차를 보이며, 중국과 비슷한 기술수준을 보유한 것으로 조사되었다. 국내 기업 중 삼성전자·LG전자·SK텔레콤·네이버·카카오 등 국내 대표 주자들도 작년 하반기부터 AI분야 투자를 확대하고 있지만 투자 규모와 기술 수준은 아직 해외 기업에 못 미치고 있다.

미국은 국가적 우선순위 추진을 위해 AI 활용 우선순위 영역을 제조업, 물류, 금융, 교통, 농업, 마케팅, 통신, 과학기술로 분류하여 영역별로 상이한 R&D 정책을 수립할 것으로 보인다. 일본의 AI 관련 R&D 전략은 AI의 발전단계에 맞춰 세부적인 계획을 수립한 것으로 보인다. 또한 총무성, 문부과학성, 경제산업성 등 정부 부처가 연계하여 AI R&D를 위한

사업을 집중 지원한다. 유럽은 'Human Brain Project'의 일환으로 뇌과학 분야에 제한적·집중적인 투자를 통해 AI을 포함한 ICT(정보통신) 기술을 개발·확대할 방침이다.

OECD(2015)에 의하면 오픈 사이언스(Open Science)란 “모든 종류의 과학적 지식(논문, 데이터, 방법론, 교육자료 등)을 디지털화하여 개방 및 공유하고 공동연구를 활성화함으로써 과학 연구를 발전시키고자 하는 개념이다. 공공자금으로 지원된 연구 성과 및 연구 과정에서 수집된 데이터를 디지털 포맷으로 공개하여 연구 커뮤니티뿐만 아니라 일반 국민도 보다 쉽게 접근하고 활용할 수 있도록 하는 메커니즘이다. 오픈 사이언스는 3대 체계, '연구 데이터(Open data, 이하 오픈 데이터)', '연구 협력(Open Collaboration, 이하 오픈 콜라보레이션)', '연구 성과의 개방(Open Access, 이하 오픈 액세스)'로 나눌 수 있다.

국내에 데이터 관리 및 연구 협력 측면에서 오픈 사이언스의 유관 정책 이행이 선진국과 비교하여 상대적으로 뒤쳐졌다. 연구 데이터를 관리하는데 있어 국가적인 지원이 부족하고 데이터 관리에 대한 이해 및 이행 체계의 부재로 연구 데이터의 효율적 활용이 어렵다. 연구 협력에 대한 니즈가 부족하여 과학계에서 더 나아가 기업 및 시민사회와의 소통은 시작 단계에 불과한 것으로 보인다. 국내 유력 오픈 액세스 저널(Open Access Journal)이 부재하고, 오픈 액세스를 촉진하는 정책이 미비해 보인다. 또한 국가 R&D 사업의 9대 주요 성과물 공개 항목 중 연구 데이터의 부재는 데이터의 재활용 가능성, 양질의 연구 성과물 개방 등의 측면에서 부정적인 영향을 미친다. 뿐만 아니라 공공 자금으로 펀딩 받은 R&D 연구데이터를 개인자산으로 인식하여 개방 및 공유에 소극적인 것으로 보인다.

선진국의 경우 연구데이터를 공유·활용하는 체계 구축을 위해 노력하고 있다. 미국은 백악관 과학기술정책실 '연구데이터 관리 및 공유에 대한 지침(2013년)에서 원칙 제시 및 빅데이터 연구개발 계획을 발표(2016년)함에 따라 구체적인 지침을 통해 연구기관 및 연구자들이 개방 및 협력에 참여할 때 따르는 어려움을 해소하고자 하였다. 이에 반해 국내에서는 연구데이터 관리·공유 체계 및 제도가 부족하고, 법적 규제 혹은 지침 등으로 연구자의 오픈 사이언스 참여를 강제 혹은 독려하는 체계가 미비하다.

이에 대한 시사점으로 연구 분야에 적절한 오픈 액세스 저널, 오픈 액세스 레퍼지토리가 존재해야 하고, 기관 레퍼지토리를 중심으로 인프라와 인센티브를 강화한다면 동일한 분야의 연구 성과물 관람 및 참고가 가능하여 각 연구에 긍정적인 영향을 미칠 것이다.

또한 국가적 차원에서 연구데이터 공유 집단·연구 공동체를 구축 및 확산하기 위해 인프라

지원 및 국가 연구데이터 허브를 구축해야 한다. ‘연구 데이터 관리 계획(Data Management Plan)’을 구축하여 연구 데이터 공개 인센티브 제공, DOI(Digital Object Identifier)<sup>1)</sup> 등의 확대 적용과 더불어 후속 활용을 지원하는 데이터 품질관리 서비스도 확대 지원해야 한다.

우선 AI 기술의 핵심 경쟁 원천인 데이터와 관련하여 수요자의 니즈에 맞는 데이터 셋 확보, 데이터 활용기반 조성, 그리고 양적·질적으로 향상된 데이터 시장을 활성화하는 것이 필요하다. 특히, 데이터 활용기반 조성과 관련해서는 ①연구데이터를 관리·공유하는 것을 제도화해야 하고, ②데이터 공유·활용 확산을 위한 연구커뮤니티를 형성해야 하며, ③국가 연구데이터 플랫폼을 구축하고 이를 제공하는 서비스를 지원해야 하고, ④데이터와 컴퓨팅 활용을 위한 연구개발 인력을 양성해야 한다.

또한 AI 분야의 선도형 R&D를 촉진시키고 성과 확산을 유도하기 위한 오픈 사이언스 문화 확산 방안을 ①국가연구개발 사업 관련 연구성과물 공개 관련 법제도 정비, ②연구자들로 하여금 연구 성과물 또는 데이터를 자발적으로 공개하게 만드는 적절한 보상 체계 마련, ③오픈 사이언스를 위한 지원 조직 및 플랫폼을 구축하여 연구자들의 오픈 사이언스에 대한 진입 장벽 완화, 이렇게 세 가지 측면에서 정책적 접근이 필요하다. 마지막으로 오픈 사이언스 도입을 위해서는 결정적으로 공유·개방에 대한 인식의 제고가 필요하며 이를 위해 연구 성과물 및 데이터 공유에 대한 공론화 및 공감대를 형성시킬 필요가 있다.

## 5. 정책적 활용 내용

본 연구 결과는 정부 AI 관련 선도형 R&D 지원체계와 R&D 성과를 확산하기 위한 제도 개선 방안을 마련하기 위한 정책방안 수립 시 참조 가능한 기초 자료로 제공된다. 이외에도 향후 AI 기술 확산을 위한 비전 및 전략 수립 시 필요한 세부과제 확보 및 예산수립 등 대응전략을 위한 근거를 제시하여 정책수립에 도움을 줄 것으로 기대된다.

---

1) 객체에 부여되는 고유한 디지털 식별자로서 편리하게 항구적인 접근과 인용을 보장하는 수단

## 6. 기대효과

4차 산업혁명의 핵심 동인인 AI 분야의 정책을 지원하여 지능정보기술 분야의 기술수준 향상 및 R&D 성과 확산을 통한 경제적 가치 창출에 기여할 것으로 기대된다. 특히, AI 기술 중에서도 핵심 기술에 효율적으로 R&D 자금을 투입하여 선진국과의 지능정보 기술 수준 격차 해소에 기여할 전망이다. 이외에도 지능정보기술 분야의 유망성, 경제성 등 대국민 인식이 확산되고 이를 통해 과학기술·ICT 분야의 전문인력 확보에 기여할 것으로 예상된다.

# SUMMARY

## 1. Title

A Study on Improving the Information Technology R&D to Promote for First Mover Research and Resulting Diffusion

## 2. Objective and Importance of Research

Artificial intelligence technology extends beyond the traditional ICT industry sector to a very broad range covering traditional industries. In this process, not only the artificial intelligence technology itself but also the data obtained from the application target area and the feedback of the user are becoming more important. This means that for the dissemination and diffusion of artificial intelligence technology, it is necessary to cooperate with all those applying artificial intelligence technology in various fields through open innovation ecosystem. Therefore, standardization of artificial intelligence algorithms and open source library work are required, and it is required to acquire data sets for testing and training, which are indispensable for the development of artificial intelligence technology. However, it takes a lot of money to secure the data, and it is difficult to secure such data from the viewpoint of a small start-up.

The role of the government and the public sector lies in facilitating the provision of public goods such as training data sets, open source libraries, and technology evaluation standards in these artificial intelligence technology ecosystems. In particular, it is necessary to share the vast amount of data the government has accumulated with the private start-up ecosystem. In other words, it is necessary to establish an artificial intelligence R&D

strategy based on public-private partnership and to establish an R&D ecosystem centered on open and cooperation and secure artificial intelligence technology early.

### 3. Contents and Scope of Research

The purpose of this study is to set up policy measures such as promoting leading R&D and improving support system and system for industrial expansion of R&D achievement in order to spread artificial intelligence technology, which is recognized as a key driver of the 4th Industrial Revolution have. In particular, we suggest ways to create an open R&D ecosystem to promote leading R&D and provide detailed strategies for this. To this end, we will look at the status of development of artificial intelligence R&D technology and understand the R&D policy and global trends of artificial intelligence. In addition, we propose building an open science system as a core method for creating a leading R&D ecosystem, analyzing policies related to open science around the world including Korea, and drawing implications through this. Through the implications, detailed strategies for establishing an open science system in Korea will be presented.

### 4. Research Results

As a key source of competition for AI technology, it is necessary to acquire data sets that meet the needs of customers and to create the data utilization infrastructure There is a need to activate the qualitatively improved data market.

In particular, regarding the creation of a data utilization infrastructure, it is necessary to ① institutionalize the management and sharing of research data, ② establish a research community for data sharing and utilization diffusion, ③ build a national research data platform and provide services Support, ④ train R&D personnel to utilize data and computing.

The ways to diffuse open-science culture for inducing leading-edge R&D and promoting performance in the AI field are ① to establish a legal system related to the open of research results related to national research and development projects, ② to Provide appropriate compensation system to encourage researchers to voluntarily disclose research results or data ③ to establish a support organization and platform for Open Science to alleviate researchers' entry barrier to open science

Finally, in order to introduce open science, it is crucial to raise awareness about sharing and openness, and it is necessary to establish consensus on research achievements and data sharing.

## 5. Policy Suggestions for Practical Use

The results of this study are provided as reference materials for establishing policy to improve the government R&D support system and R&D result diffusion. It is expected that it will help to establish the policy by Providing the grounds for responding strategies, such as detailed tasks and budgeting

## 6. Expectations

It is expected to contribute to the economic value creation by improving the technology level in the intelligence information technology field and spreading the R&D performance through supporting the policy of the AI field which is the key driver of the 4th industrial revolution.

# CONTENTS

Chapter 1. Chapter 1 Introduction

Chapter 2. Intelligent Information Technology and R & D Trend

Chapter 3. AI Enterprise Learning Data

Chapter 4. Creating an open R&D ecosystem

Chapter 5. Conclusions and Policy Suggestions

# 제1장 서론

## 제1절 연구의 배경

세계는 지금 시간과 장소에 구애받지 않고 언제나 통신망에 접속하여 다양한 정보통신 서비스를 활용하는 것을 가능케 하는 네트워크 기술의 등장과 모바일, AI, 빅데이터, 클라우드 등이 발전해 감에 따라 과거와는 전혀 다른 새로운 산업 사회의 창조라는 거대한 화두에 천착하고 있다. 다보스포럼(2016)은 이러한 새로운 산업 사회의 창조-4차 산업혁명-가 디지털을 통해 연결성이 극대화되는 것을 뜻한다고 규정하였다. 사실 4차 산업혁명의 정의와 개념에 대해서는 각 국가별, 산업별로 각기 중점을 두는 분야 혹은 기술에 따라 다양한 시각차가 존재하여 아직까지 다양한 논의가 이루어지고 있다. 하지만, 공통적으로 받아들여지고 있는 점은 4차 산업혁명은 초연결과 초지능에 기반을 둔 디지털 변혁(Digital Transformation)을 야기하며, 이전에 볼 수 없었던 광범위한 파급력을 지니고 매우 빠른 속도로 우리 사회를 변화시킬 것으로 예측된다는 것이다. AI, 사물인터넷, 빅데이터 등의 핵심기술들은 고도의 생산성과 효율성을 창출하면서 산업과 경제 구조, 고용구조 전반의 변화를 가속화시킬 것으로 전망되고 있다. Chen 외(2016)에 의하면 기업투자를 근거로 AI 기술이 세계경제성장률에 미치는 파급효과는 향후 10년간(2016~2025) 1.59~2.95조 달러에 이를 것으로 추정되고 있다.

4차 산업혁명을 초래하는 핵심기술은 흔히 지능정보기술을 일컫는데, 이는 흔히 AI 기술과 데이터·네트워크 기술(ICBM)을 융합하여 기계에 인간의 고차원적 정보처리 능력(인지, 학습, 추론)을 구현하는 기술을 의미한다. 지능정보기술은 기술의 발전과 그에 따른 사업화 가능 제품 및 서비스의 단계적 실현 시기 분석에 따라 크게 AI SW, AI HW, 데이터·네트워크, 기초과학 등으로 분류된다. 먼저, AI SW는 인간의 지능을 기계에 구현하기 위한 핵심 수단으로 AI기술의 기본 바탕이 되는 추론·기계학습 분야와 이를 활용한 지능

별 세부 기술로 구분된다. 특히, 추론 및 기계학습을 활용한 기능별 AI SW는 지식표현 및 언어 지능, 청각 지능, 시각 지능, 복합 지능, 지능형 에이전트, 인간-기계 협업 분야로 분류되며, 기능별 AI이 서로 융합하여 새로운 분야로의 확장 가능성이 매우 높다.

AI HW는 AI SW를 지원하는 기술로 AI 기술의 인프라와 고성능 기술의 역할을 담당하고, 기초과학은 학습 기반 AI 알고리즘 자체의 개선을 위한 수학적 접근, 생리학적인 뇌의 특성에 기반한 새로운 AI 개념 발견 등 원천기술의 역할을 담당한다. 또한, 데이터·네트워크 기술은 지능정보기술에 필요한 데이터를 공급하기 위한 기술로서 ICBM이 긴밀하게 연계하여 데이터를 수집·전달하고, 이를 저장·분석하기 위해 필요한 기술이다.

본 보고서에서는 4대 지능정보기술 중 지능화 혁신을 위한 선결조건인 AI SW 기술을 중점적으로 살펴보고자 한다. AI은 단순한 신기술이 아닌 국내외의 침체되어 있던 경제성장의 새로운 동력으로 작용하여 각국 경제 성장에 비약적인 과급효과를 실현할 것으로 전망되고 있다.<sup>2)</sup> 또한, 국내에서는 AI의 도입이 제조생산성을 향상시키고 국민의 복지를 증대시키는 등 2030년까지 약 536조원 규모의 가치를 창출할 것으로 전망된다(NIPA, 2017. 11). 미국과 일본을 비롯한 여러 선도국들은 이미 AI의 높은 잠재효과에 주목하고 AI 기술 경쟁력을 선점하기 위한 국가 R&D 정책을 추진 중에 있다. 이와 더불어 민간 글로벌 기업의 치열한 기술 개발 및 사업화 경쟁의 결과 최근 자율주행자동차, 가상비서 등 AI 응용 제품과 서비스가 상용화 단계에 이르렀다. 국내에서는 2016년 3월에 있었던 알파고 대국 이후에 AI에 대한 국민적 관심과 기대가 고조되면서 ICT 기업 및 스타트업의 AI R&D가 활성화 추세에 있다.

하지만, 이러한 세계적인 흐름과 국내 기업들의 관심에도 불구하고 아직까지 국내 기술 수준은 선도국에 비해 부족하고<sup>3)</sup> AI 관련 산업의 생태계도 미약한 실정이다. 또한, 폐쇄적 R&D 모형의 고착화로 개방 및 협력 기반으로 빠르게 성장 중인 글로벌 R&D 생태계의 발전 속도를 추격하는데 한계가 있다.

AI이 국가경제의 혁신 동력으로 작용하여 산업구조뿐만 아니라 전반적인 사회의 변화를

---

2) 2035년까지 미국 경제에서 AI에 의한 경제적 부가가치효과는 AI이 도입되지 않았을 때 (2.6%)와 비교하여 약 1.5배 더 클 것으로 전망된다(엑센추어, 2016. 12).

3) 국내의 AI 기술 수준은 세계 최고 기술 수준(美) 대비 81.3%로 약 2.3년의 기술격차가 존재한다(IITP, 2017. 7).

견인하는 지능화 혁명을 촉발시킬 것이라는 점을 감안하면 우리나라가 처한 낮은 수준의 AI 기술 경쟁력은 하루 빨리 개선해 나가야 할 점이라고 판단된다.

AI 기술은 전통적인 ICT 산업 부문을 넘어 전통산업을 포괄하는 매우 광범위한 영역으로 확대되어 가고 있다. 이러한 과정에서 단순히 AI 기술 자체뿐만 아니라 응용대상 영역에서 얻어지는 데이터 및 사용자의 피드백 등이 점차 중요해지고 있다. 이러한 점은 AI 기술의 보급 및 확산을 위해서는 개방형 혁신 생태계를 통하여 다양한 분야에서 AI 기술을 응용하는 주체들이 모두 협력하여야 한다는 것을 말한다. 따라서 AI 알고리즘의 표준화 및 오픈소스 라이브러리화 작업이 필요하며, AI 기술의 개발에 필수적인 테스트 및 트레이닝용 데이터 셋의 확보가 요구된다. 그러나 데이터의 확보에는 많은 비용이 소요되어, 규모가 작은 스타트업의 입장에서 이러한 데이터의 확보는 쉽지 않은 일이다.

정부 및 공공부문의 역할은 이러한 AI 기술 생태계에서 훈련용 데이터셋과 오픈소스 라이브러리, 기술평가기준 정립과 같은 공공재의 공급을 촉진하는 것에 있다. 특히 정부가 축적한 방대한 양의 데이터를 민간 스타트업 생태계와 공유하는 것이 필요하다. 즉, 민관 협력 기반의 AI R&D 전략 수립을 통해 개방 및 협력 중심의 R&D 생태계를 조성하고 AI 기술력을 조기에 확보하는 것이 필요한 시점이다.

## 제 2 절 연구의 목표 및 주요내용

본 연구는 4차 산업혁명의 핵심 동인으로 인식되고 있는 AI 기술의 확산을 위해 선도형 R&D의 촉진 및 R&D 성과의 산업 확산을 위한 지원체계 및 제도 개선 방안 등의 정책 방안 마련을 목표로 하고 있다. 특히, 선도형 R&D를 촉진시키기 위해 개방형 R&D 생태계를 조성하는 방안을 마련하고, 이에 대한 세부 전략을 제시한다. 이를 위해 AI 기술의 단계별 발전 현황을 살펴보고, 국내외 AI 관련 R&D 정책과 글로벌 기업 동향을 파악한다. 또한, 선도형 R&D 생태계를 조성하기 위한 핵심 방안으로 오픈 사이언스 체계를 구축하는 것을 제시하면서 우리나라를 포함한 세계 각국의 오픈 사이언스 관련 정책을 분석하고, 이를 통해 시사점을 도출한다. 도출된 시사점을 통해 향후 우리나라에 오픈 사이언스 체계를 구축하기 위한 세부 전략을 제시한다.

연구의 주요 내용과 구성은 다음과 같다. 제 2장에서는 먼저 AI 기술의 개념 및 분류를 살펴보고, 타기관의 기술 수준 분석 결과와 AI 관련 특허 출원 수 분석 등을 통해 국내 AI 기술 수준을 고찰하였다. 또한, AI의 단계별 발전 시나리오를 제시하고, 발전 현황을 분석하였다. 이와 더불어 국내외 AI 기술 및 R&D 동향 파악을 통해 현재 국내 AI 기술이 가진 문제점을 진단하였다.

제3장에서는 AI 기술 개발을 위한 필수 요소로 부각되고 있는 학습데이터와 관련하여 국내외 정책·산업 동향을 파악하고, 전문가 인터뷰 및 AI 공급사를 대상으로 한 설문조사를 통해 AI 기술 발전에 필요한 데이터 셋 수요조사를 수행하였다. 구체적으로, 전문가 인터뷰를 통해 기존 문헌에서 제시하고 있는 기계학습을 위한 데이터 셋 유형 및 분류체계를 검토하였으며, 데이터 수요조사 분석 결과를 기초로 각 기술 또는 적용분야별로 데이터 셋 집중도를 살펴보았다. 또한, 매출, 종사자 수, 사업 분야 등의 기업 특성과 필요한 데이터 수요 간 교차분석을 수행하였다.

제4장에서는 선도형 R&D를 촉진시키기 위한 방안으로 개방형 R&D 생태계 조성 방안을 제시하였다. 개방형 R&D 생태계를 조성하기 위한 하나의 전략으로 오픈 액세스, 오픈 데이터, 오픈 콜라보레이션으로 구성되는 오픈 사이언스 구축 방안을 제시하였다. 이를 위해 먼저 오픈 사이언스의 개념을 정리하고, 개방형 R&D에 있어 선도국이라 할 수 있는 미국을 포함한 각국의 오픈 사이언스 정책 추진 현황 및 이행 실태를 분석하고, 현재 우리나라의 오픈 사이언스 관련 정책과 비교 분석을 통해 시사점을 도출하였다. 또한, 도출된 시사점을 통해 오픈 사이언스를 활성화하기 위한 방안을 오픈 액세스, 오픈 데이터, 오픈 콜라보레이션 각 측면에서 제시하였다.

마지막으로 제5장에서는 선도형 R&D를 촉진시키기 위한 전반적인 정책방향과 세부 정책과제들을 제시하고 있다. 주요 내용은 아래와 같다.

우선 AI 기술의 핵심 경쟁 원천인 데이터와 관련하여 수요자의 니즈에 맞는 데이터 셋 확보, 데이터 활용기반 조성, 그리고 양적·질적으로 향상된 데이터 시장 활성화의 필요성을 설명하였다. 특히, 데이터 활용기반 조성과 관련해서는 ① 연구데이터를 관리·공유하는 것을 제도화해야 하고, ② 데이터 공유·활용 확산을 위한 연구커뮤니티를 형성해야 하며, ③ 국가 연구데이터 플랫폼을 구축하고 이를 제공하는 서비스를 지원해야 하고, ④ 데

이티와 컴퓨팅 활용을 위한 연구개발 인력을 양성해야 할 것을 제안하였다.

또한, AI 분야의 선도형 R&D를 촉진시키고 성과 확산을 유도하기 위한 오픈 사이언스 문화 확산 방안을 ① 국가연구개발 사업 관련 연구성과물 공개 관련 법제도 정비, ② 연구자들로 하여금 연구 성과물 또는 데이터를 자발적으로 공개하게 만드는 적절한 보상 체계 마련, ③ 오픈 사이언스를 위한 지원 조직 및 플랫폼을 구축하여 연구자들의 오픈 사이언스에 대한 진입 장벽 완화, 이렇게 세 가지 측면에서 제시하였다. 마지막으로 오픈 사이언스 도입을 위해서는 결정적으로 공유·개방에 대한 인식의 제고가 필요하며 이를 위해 연구 성과물 및 데이터 공유에 대한 공론화 및 공감대 형성의 필요성을 설명하였다.

## 제2장 지능정보기술 및 R&D 동향

### 제1절 AI 기술 개요

#### 1. AI SW 기술 분류

지능정보기술의 발달은 초연결과 지능화를 기반으로 디지털 대변혁(Digital Transformation)을 야기하며 빠른 속도로 우리 사회를 변화시키고 있다. AI, 빅데이터, IoT 등으로 대변되는 지능정보기술은 생산성을 고도로 향상시키고 산업 전반에 걸쳐 효율성을 극대화시키면서 산업 및 고용 구조 전반에 변화를 가속화시킬 것으로 전망되고 있다. 본 절에서는 이러한 핵심기술 중 AI 기술에 초점을 맞춰 기술 수준과 특허 현황 등을 파악하고자 한다.

서론에서 살펴본 바와 같이, 지능정보기술은 AI 기술의 개발을 통해 미래 지능정보 서비스를 실현하는 관점에서 4대 대분류 및 17대 세부기술로 분류된다. AI SW, AI HW, AI연계 Data·Network, 기초과학으로 분류되는 4대 지능 분야 중에서도 AI SW기술을 통해 인간의 지능을 기계로 구현 한다. 과학기술정보통신부와 정보통신기술진흥센터(IITP)는 이러한 AI 기술을 ‘학습 및 추론’, ‘상황이해’, ‘언어이해’, ‘시각이해’, ‘인지 및 인식 기술’등 총 5개의 핵심기술로 구분하였다.

학습 및 추론 기술은 데이터에 내재된 패턴, 규칙, 의미 등을 알고리즘 기반 학습을 통해 새롭게 입력되는 데이터에 대한 결과를 예측하도록 하는 기술이다. 학습 및 추론 기술의 세부 기술은 지식표현과 지식베이스가 있다. 지식표현 기술은 지식을 컴퓨터가 이해가 가능하도록 표현하는 기술이고, 지식베이스 기술은 지식·규칙·사실 등을 기반 데이터 베이스로 구축하고 이를 관리하는 기술이다. 과학기술정보통신부는 ‘SW컴퓨팅원천기술 개발사업’의 일환으로 2014년부터 2018년까지 일반 사용자도 학습 및 추론 기술을 쉽게 접근·활용할 수 있도록 기계학습 SW ‘SMILE’을 개발하고 있다. 해외에서는 글로벌 기업을 중심으로 기술 개발이 진행되고 있으며, 대표적으로 구글, 페이스북, IBM 등에서 영상, 음성, 텍스트 인식 분야의 학습·추론 기반 AI 플랫폼을 개발하고 있다.

상황이해 기술은 주변 환경에서 발생하는 데이터를 종합적으로 이해하고, 맥락 분석과 판단을 제공하여 주변 환경 및 사람의 감정 인지를 가능토록 만드는 기술이다. 상황이해 기술의 세부기술로는 감정이해, 공간이해, 협력지능 및 자가이해 기술이 있다. 감정이해 기술은 사람의 감정을 인식하고 일정한 기준에 의해 감정을 구분하는 기술이며, 공간이해는 시공간적 세계를 인지하여 3차원의 세계를 올바르게 변형시킬 수 있는 기술이다. 협력지능 기술은 각각의 객체와 교류·이해가 가능하게 만들고, 행동 분석에 따라 대처하는 기술이다. 자가이해 기술은 심리적·정신적 특성을 이해·인지할 수 있는 기술이다. 이러한 상황이해 기술과 관련하여 국내 대기업인 삼성전자, LG전자, SK텔레콤, KT 등을 중심으로 AI 기반 스마트 홈 제어 기술을 개발하고 있으며, Amazon(Alexa), Automated Insight (Wordsmith) 등에서 상황이해 기반 AI SW를 개발하고 있다.

언어이해 기술이란 사람의 말과 글을 사람처럼 이해하고 수행할 수 있도록 자연어를 이해·지식화하는 기술로, 자연어 질의응답, 언어 통번역 및 언어 기반 창작 등이 가능한 언어 지능 기반 기술이다. 언어이해 기술의 세부기술인 자연어처리는 다시 형태소 분석, 개체명인식, 구문분석, 의미분석으로 구분된다. 질문에 대한 답변을 제시하는 '질의응답'과 디지털 음성신호를 컴퓨터에서 처리 가능한 언어로 변환하는 '음성처리' 기술도 언어이해의 세부기술에 해당한다. 마지막으로 자동통번역 기술은 한 언어에서 다른 언어로 자동번역·통역하는 기술이다. 국내에서는 ETRI, 솔트룩스, KAIST, 포항공대 등에서 언어를 통해 지식 학습이 가능한 '엑소브레인 SW'를 개발하고 있다. IBM은 Watson 답 QA 기술을 헬스케어 및 파이낸스 분야에 적용하기 위한 프로젝트를 진행 중에 있으며, 일본 후지쯔는 도쿄대·일본국립정보학연구소와 공동으로 대학입시 합격을 목표로 하는 '토다이 로봇' 프로젝트를 추진하고 있다.

시각이해 기술은 영상의 내용 및 상황을 이해하여 미래를 예측하는 기술로, 영상내용이해, 시각 지식 생성, 내용기반 영상 검색, 비디오 분석 및 예측이 가능하다. 국내 ETRI, GIST 및 포항공대를 중심으로 대규모 영상분석 통해 영상의 의미를 이해하는 '딥뷰W'를 개발하고 있다. 내용 기반 영상 검색, 행동인식, 시각지식 기술이 시각이해 기술의 세부기술에 해당한다. 내용 기반 영상 검색은 영상 데이터에서 색, 모양, 질감 등의 특징을 추출하여 색인과 검색을 수행하는 기술이며, 행동인식 기술은 동영상 내 사물의 행동을 인식하는 기술이다. 마지막으로 시각지식 기술은 행동인식, 영상이해, 배경인식 등을 이용하여

영상데이터로부터 지식정보를 추출·생성하는 기술이다. 미국 DARPA(Defense Advanced Research Projects Agency, 방위고등연구계획국)는 군사용 비디오에서 원거리 행동을 인식하는 원격 행동분석 기술을 개발 중이다. 또한 구글과 마이크로소프트는 딥러닝 기술을 기반으로 영상에서 일상 속 많은 사물을 구별하는 ‘고성인 인식 기술’을 개발하고 있다.

인지 및 인식기술, 즉 인지컴퓨팅 기술은 주변 환경 지각인지, 학습 적용, 지식 추론, 행위 생성 등 사람의 인지구조를 모방하여 통합함으로써 지능형 서비스 개발을 지원하는 기술이다. 휴먼라이프 이해 기술과 인지 아키텍처 기술이 이에 해당한다. 휴먼라이프 이해는 개인의 경력, 건강, 대인관계, 재무 등을 관리하기 위해 지능적인 도움을 제공하는 이해 기술이며, 인지 아키텍처 기술은 사람의 마음 구조를 컴퓨터 모델화 하는 기술이다. 국내 인텔리콘 메타 연구소에서 법률 검색 서비스에 적용하기 위한 지능형 법률 정보 시스템(i-LIS)<sup>4)</sup>를 개발하고 있다. 해외에서는 IBM, 애플, 마이크로소프트, Saffron Tech, Maluuba 등을 중심으로 대규모 데이터 분석 및 처리를 위한 인지컴퓨팅 기술이 개발되고 있다. 다음 <표 2-1>은 AI의 5개 핵심기술 분류 및 세부기술이 정리된 자료이다.

<표 2-1> AI 5대 핵심기술 및 15대 세부기술 분류

AI 핵심 기술 및 세부 기술 분류		
핵심기술	세부기술	기술개요
학습 및 추론 기술 (Learning and Reasoning)	지식표현	분석된 지식을 컴퓨터가 이해할 수 있는 언어로 표현하는 기술
	지식베이스	축적한 전문지식, 문제 해결에 필요한 사실과 규칙이 저장된 데이터 베이스로 구축, 관리하는 기술
상황이해 기술 (Context Understanding)	감정이해	사람의 기분, 감정을 인식, 구분할 수 있는 기술
	공간이해	시공간적 세계를 정확하게 인지하고, 3차원의 세계를 잘 변형시키는 기술
	협력지능	다른 에이전트와 교류하고, 이해하며, 그들의 행동을 해석하고, 효율적으로 대체하는 기술
	자기가해	자기 자신(개성, 정신적 심리적 특성)을 이해하고, 느낄 수 있는 인지적 기술

4) 해당 시스템은 법류의 세계에 수학, 통계학, 컴퓨터 공학, 뇌과학, 인지 과학, 감성 공학 등을 융합한 지능형 법률 정보 서비스이다.

핵심기술	세부기술	기술개요
언어이해 기술 (Language Understanding)	자연어 처리 (형태소분석, 개체명인식, 구문분석, 의미분석)	인간의 자연적 언어를 형태소 분석, 개체명인식, 구문분석, 의미분석하는 기술
	질의응답	질문에 대한 답변을 제시하는 기술
	음성처리	디지털 음성신호를 컴퓨터에서 처리 가능한 언어로 변환하는 기술
	자동통번역	한 언어에서 다른 언어로 자동으로 번역하거나 통역하는 기술
시각이해 기술 (Visual Understanding)	내용 기반 영상 검색	영상 데이터 자체의 특징정보인 색광과 모양, 질감 등 영상 데이터의 내용을 대표할 수 있는 특징들을 추출하고 이를 기반으로 색인과 검색을 수행하는 기술
	행동인식	동영상에서 움직이는 사물의 행동을 인식하는 기술
	시각지식	행동인식, 영상이해, 배경인식 등을 이용하여 영상 데이터로부터 지식정보를 추출, 생성하는 기술
인식 및 인지 기술	휴먼라이프 이해	개인 경력관리, 건강, 대인관계, 재무관리 등 일상생활에서의 지능적 도움을 제공하기 위해 사람의 생활을 이해하는 기술
	인지 아키텍처	인지심리학 측면에서의 사람의 마음 구조를 컴퓨팅 모델화하는 기술

자료: 한국지식재산연구원(2016) “AI(AD) 기술 및 정책 동향”

## 2. AI 기술수준 및 기술격차

2000년대 후반부터 빅데이터를 통한 기계학습과 딥러닝의 발전으로 AI 기술에 대한 관심이 높아지고, 산업으로의 적용이 활발히 진행되기 시작했다. 딥러닝 기술은 얼굴 인식, 사물 인식, 자연어 처리 등을 이용하여 제조, 금융, 의료, 교육 등의 산업에 전반적으로 적용됨에 따라 새로운 비즈니스 기회를 창출하고 있다.

특히 언어 번역, 음성 인식, 추론 등 핵심기술은 IBM Bluemix, 구글의 Tensorflow 등이 세계적으로 선도를 하고 있고, 국내는 Naver Line의 언어 번역 기능, 삼성전자 Galaxy의 음성 인식 기능 등이 제품에 탑재되어 활용되고 있다. 또한 AI 슈퍼컴퓨터 ‘왓슨(Watson)’의 기술을 보유한 미국 IBM과 국내 IT(정보기술) 서비스 업체 SK C&C가 손잡고 2017년부터 한국어로 된 왓슨 서비스를 선보였으며, IBM의 왓슨 기술과 SK C&C의 한국어 기능 탑재를 통해 국내기업 및 공공기관 등에 제공했다. 더불어 기계학습·딥러닝 알고리즘의 진

화로 데이터 분석의 정확도가 향상됨에 따라 공공정책, 산업, 문화 등의 분야에서 AI 기반 빅데이터 분석 기술의 적용이 확대되고 있다.

해외뿐만 아니라 국내 AI 기술의 적용이 확대되는 한편, 국가별로 AI 기술의 역량이 상이하다. 우선, AI 핵심기술 측면에서 AI SW분야는 미국(100%) > 유럽(86.3%) > 일본(81.9%) > 한국(73.9%) > 중국(71.8%) 순으로 기술 수준에 차이가 있다. 선진국 대비 한국은 2.2년의 기술격차를 보이며, 중국과 비슷한 기술수준을 보유한 것으로 조사되었다. 구체적으로 기초연구 분야에서 26.4%p, 응용·개발 분야에서 25.6%p, 사업화 측면에서 26.6%p의 기술수준 격차가 발생했다. 다음 <표 2-2>은 국가별 AI SW 기술격차를 나타낸 자료이다.

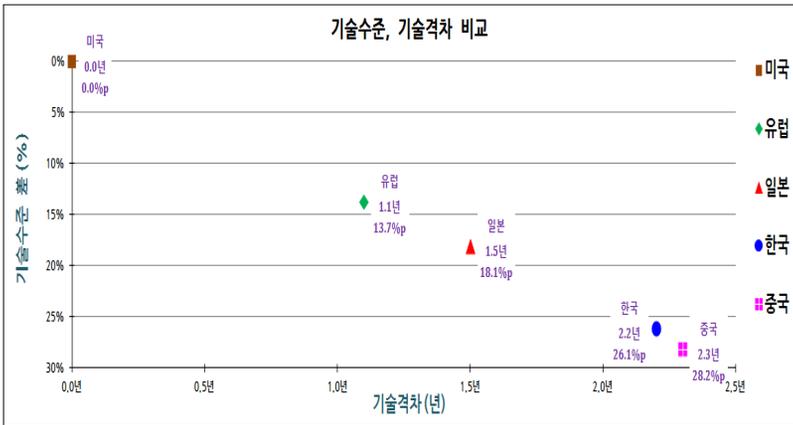
<표 2-2> 2016년 AI SW 기술수준

AI	상대수준(100%)															기술격차(0년)					
	한국			미국			일본			중국			유럽			한국	미국	일본	중국	유럽	
	기초	응용	사업화	기초	응용	사업화	기초	응용	사업화	기초	응용	사업화	기초	응용	사업화						
소분류	학습 및 추론 기술	71.9	72	71.1	100	100	100	81.4	82.8	82.2	70.8	70.6	70.3	87.3	87.7	86.8	2.4	0	1.5	2.5	1.2
	상황이해 기술	72.2	74.1	73.5	100	100	100	80.3	81.8	80.4	69.6	71.5	70.7	84.8	84.8	84.8	2.1	0	1.5	2.2	1.1
	언어이해 기술	78.7	79.4	78	100	100	100	83.7	82.8	83.5	74.8	75.2	75.1	88.4	88.4	87.6	1.8	0	1.2	2.1	0.8
	시각이해 기술	76.5	77.1	75.4	100	100	100	83.7	82.6	82.8	71.9	74.1	74.3	87.3	87.4	86.4	2	0	1.4	2.1	1.2
	인지컴퓨팅 기술	70.4	71.4	71.4	100	100	100	79.9	79.7	78.8	69.5	70.1	68.8	84.2	84.4	83.9	2.4	0	1.7	2.5	1.3
	기타 AI 기술	71.6	72.2	71.3	100	100	100	83.2	82.8	81.2	71.2	71.7	70.9	87.1	86.2	84.6	2.2	0	1.6	2.4	1.1
상대수준 및 기술격차	평균	73.6	74.4	73.4	100	100	100	82	82.1	81.5	71.3	72.2	71.7	86.5	86.5	85.7					
	중분류	73.9			100			81.9			71.8			86.3			2.2	0	1.5	2.3	1.1

자료: IITP(2017), “2016년도 정보통신 기술수준조사보고서”

특히 중국의 AI 분야에 대한 기술수준은 2015년(66.1%) 대비 5.7%p 상승하였고, 기술격차는 0.5년을 단축하면서 한국을 빠르게 추격하고 있다. [그림 2-1]은 미국, 유럽, 일본, 한국, 중국의 기술격차를 나타낸 그래프이다.

[그림 2-1] AI 기술 격차

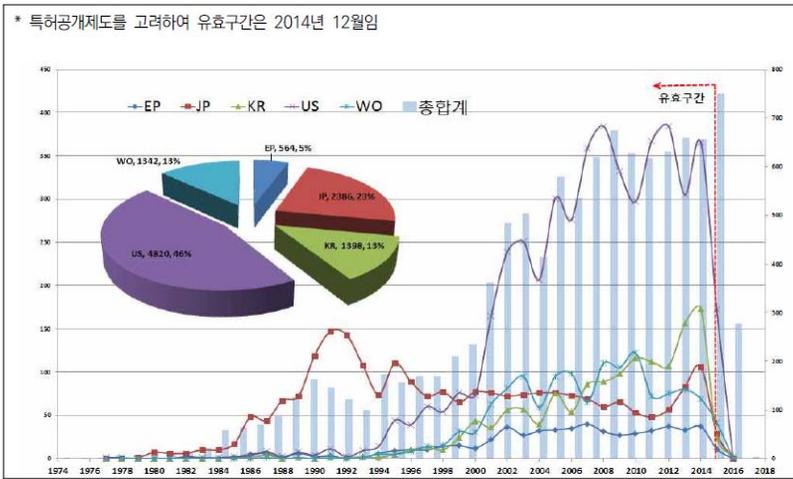


자료: IITP(2017), “2016년도 정보통신 기술수준조사보고서”

### 3. AI 관련 특허 출원

미국과 일본을 비롯한 여러 AI 기술 선도국은 국가 R&D 정책 추진을 통해 기술 경쟁력을 선점하기 위해 노력해왔다. 이러한 노력의 결과로 미국 특허(공개 및 등록특허)는 4,860건으로 전체의 46%를 차지했으며, 일본이 그 다음으로 2,386(23%), 한국이 1,398건(13%), 유럽이 564건 순으로 특허 출원이 진행 됐다(2016년 기준). IITP(2017)에 따르면 한국의 AI 관련 특허는 2017년 기준으로 세계에서 3번째로 많은 특허 출원 수를 기록했다. 해당 특허 수 조사는 2005년 1월부터 2017년 9월까지의 전 세계 특허 출원량을 기준으로 이루어졌다. AI 기술 분야의 선도국인 미국과 일본이 한국에 앞선 1위와 2위를 차지했으며, 기타 국가 중 독일>중국>프랑스>네덜란드>캐나다>영국>대만 순으로 특허 출원량에 차이를 보였다. 다음 [그림 2-2]은 연도별 AI 기술 특허출원 동향을 나타낸 그래프이다

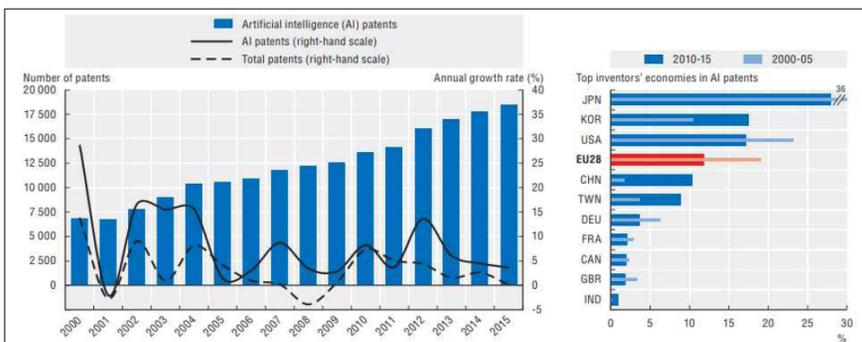
[그림 2-2] AI 분야 국가 연도별 특허출원 동향



자료: 한국지식재산전략원(2016), "이슈페이퍼: AI(AI) 기술 및 정책 동향"

더불어 OECD(2017)에 따르면, 최근 한국의 AI 기술 특허 출원 수의 비중이 크게 늘어났다. 2000년부터 2005년까지 한국의 AI 기술 특허 출원 수에 비해 2010년부터 2015년까지의 한국의 AI 기술 특허 출원 수 비중이 증가했다. 또한 한국, 미국, 일본의 경우 2010년부터 2015년까지 5년 동안 AI 기술 특허가 전체의 약 62%를 차지했다. [그림 2-3]은 2005년부터 2015년까지의 기간 동안 AI 기술 특허 출원 수 추이를 보여준다.

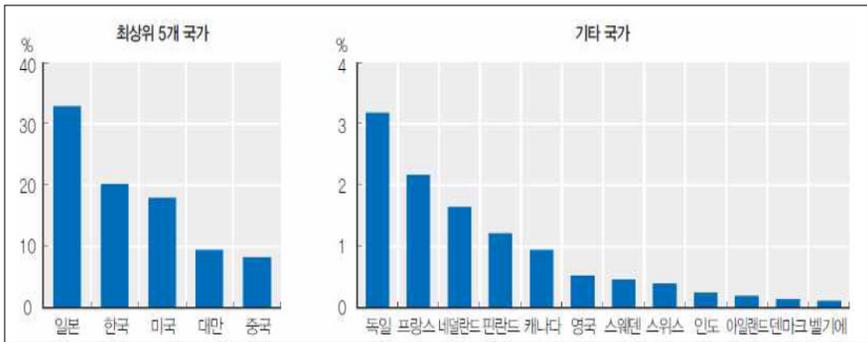
[그림 2-3] AI 기술 특허출원(2000-2015)



자료: OECD(2017), STI Micro-data Lab: Intellectual Property Database, <http://oe.cd/ipstats>

이는 아시아 국가 중에서도 한국의 AI 기술력이 향상되고 있음을 짐작할 수 있는 수치이다. 또한 OECD(2017)에 따르면 2012년부터 2014년까지 본사가 위치한 국가를 기준으로 IP5<sup>5)</sup> 특허를 보유한 50개 기업의 본사 위치 국가를 살펴봤을 때, 한국의 기업 수가 미국의 기업수를 초과한다. 국가별 IP5 보유 기업의 비중은 일본 약 30%, 한국 20%, 미국은 18%를 기록했다. [그림 2-4]은 전 세계 상위 AI 기술 특허 보유 기업의 국가별 분포 추이를 보여준다.

[그림 2-4] 세계 상위 AI 특허 보유 R&D 기업의 국가별 비중(본사 위치 기준) (2012-2014)



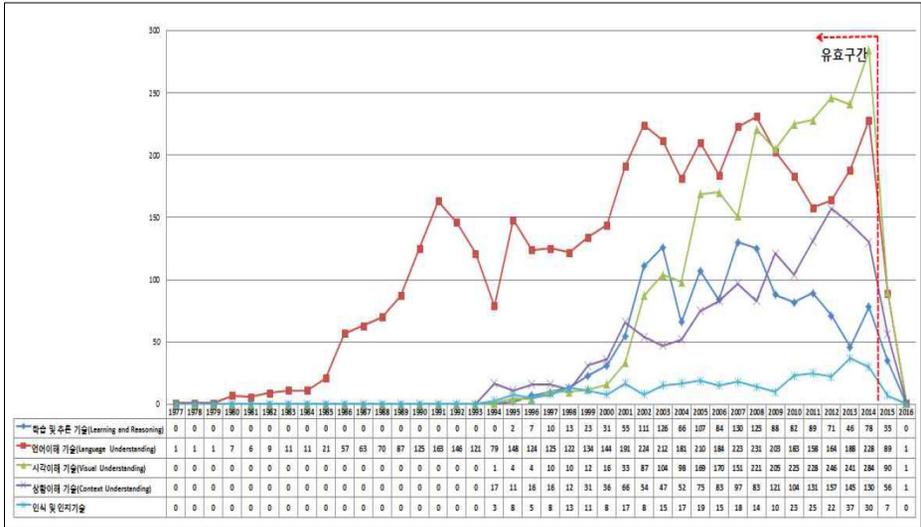
자료: STI Scoreboard(2017) 그림 25 Artificial Intelligence patents by top R&D companies, by headquarter's location, 2012-14.; 이명화·최용인(2017)

주: (JRC-OECD, COR&DIP© Database v.1. and OECD, STI Micro-data Lab: Intellectual Property Database, <http://oe.cd/ipstats>, July 2017)에 기반하여 작성

위의 자료로 알 수 있듯이 AI 기술은 국가별로 특허 출원 수에 차이를 보였으며, 국가별 기술 역량이 빠르게 변화하고 있다. 더불어 한국지식재산전략원(2016)에 따르면 핵심 기술별 특허 출원 수도 상이함을 알 수 있다. AI 핵심 기술별 특허 출원 비율 중 언어이해 기술 특허 출원이 전체의 44%를 차지하여 가장 많은 특허 출원 수를 기록했고, 그 다음이 시각이해 기술(25%) > 상황이해 기술(15%) > 학습 및 추론 기술(13%) > 인식 및 인지 기술(3%) 순으로 특허 출원이 진행됐다. 다음 [그림 2-5]는 연도별 AI 핵심 기술별 특허 출원 추이를 나타낸 자료이다.

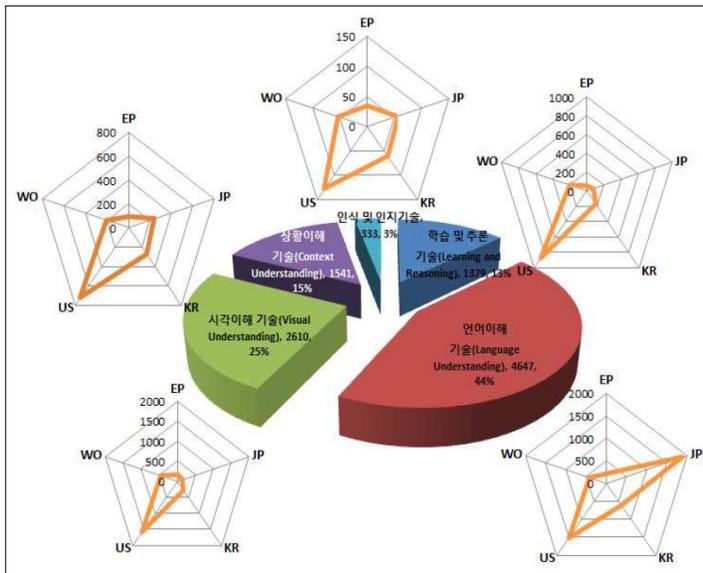
5) 유명 특허 EPO, JPO, KIPO, SIPO, USPTO를 나타냄

[그림 2-5] 핵심기술 연도별 특허 출원 동향



자료: 한국지식재산전략원(2016), “이슈페이퍼: AI(AD) 기술 및 정책 동향”

[그림 2-6] 핵심기술 국가별 특허출원 동향



자료: 한국지식재산전략원(2016), “이슈페이퍼: AI(AD) 기술 및 정책 동향”

국가별 핵심 기술 특허 출원 수를 살펴보면 언어이해 기술 분야는 일본에서 가장 많은 특허 출원이 이루어졌으며, 그 다음으로 미국이 다(多)출원을 기록했다. 일본이 자동 통·번역 기술 분야에서 독보적으로 많은 특허를 보유하고 있으며 시각이해, 학습 및 추론, 상황이해 기술 분야는 미국이 타 국가에 비해 주도적으로 기술을 이끌고 있다. 다음 [그림 2-6]는 국가별 핵심 기술 특허 출원 수·비율을 나타낸 자료이다.

IITP(2017)에 따르면 국내 AI 특허 출원 실적이 가장 우수한 기관은 삼성전자로 총 3,188개의 특허를 보유하고 있다. 두 번째 국방부 직할부대 및 기관은 LG전자로 899개의 특허를 보유하고 있으며, 삼성디스플레이(207개), 삼성전기(207개), 현대자동차(13), KAIST(91개), SK텔레콤(75개) 등이 그 다음으로 많은 특허 출원을 진행했다.

## 제 2 절 국내 지능정보기술 및 R&D 동향

### 1. AI 발전 현황

컴퓨팅시스템의 고성능화·자동화·산업융합화로 인해 AI 기술이 초지능·초연결·초실감 지능형 서비스로 발전하고 있다. 이미 AI 분야의 시장 잠재력은 높이 평가되고 있는데, 각 국가 및 기업별로 AI 관련 분야(번역, 메신저, 머신러닝 등)에 대한 역량을 집중적으로 강화하고 있다.

특히 인간의 생물학적 특성인 뉴런의 연결 구조에서 영감을 받아 초기 AI 기술 ‘인공 신경망(artificial neural network)’이 만들어지고, 거둬 발전되면서 ‘딥러닝(Deep learning)’이 AI의 효과적인 구현을 위한 기술로 자리 잡고 있다. 딥러닝은 입력 데이터를 통해 거둬 ‘학습’하고, 학습된 결과를 토대로 스스로 진화한다는 점에서 지금까지의 기술과 차별된다. 딥러닝을 포함한 여러 AI의 구조와 분석의 토대가 되는 알고리즘은 다양하다. DBN(Deep Belief Network), Autoencoder, CNN(Convolutional Neural Network) 등의 여러 AI 알고리즘은 데이터, 상황(context) 및 풀고자 하는 문제(Task)에 따라 상이하게 적용해야 한다.

AI 기술 중 스스로 학습하는 딥러닝 기술은 여태까지의 알고리즘과 차별되는 ‘일반화(generalization)’ 능력 덕분에, 학습 할 때 사용되는 학습용 데이터 외에 새로운 데이터에 대해서도 높은 해석력을 가질 수 있다. 결국 알고리즘과 더불어 양질의 데이터에 대한 입

력과 학습이 스스로 더 똑똑해 지도록 하는 요소이다.

딥러닝 기술은 학습을 통한 '예측'이 가능하기 때문에, 산업 전역에서 활발히 이용되고 있다. 물론 현재까지의 AI 기술은 인간의 의사결정을 완전히 대체하는 것이 아니라 'Narrow AI' 개념에 포함되는 수준으로, 이미지 분류, 데이터 패턴 인식 등과 같은 작업을 인간보다 더 잘 해낼 수 있는 수준이다.

AI SW 분야에서 현재 특정 분야의 일정한 작업에 대해 우수한 성능을 내는 수준(Narrow AI)에서 점차 다양한 분야의 가변적인 작업에도 대응 가능한 범용 AI 기술(General AI)로 발전해 나갈 전망이다. AI HW분야는 인공지능망 학습에 필요한 대규모 연산을 고속·저지연·저전력으로 구현하기 위한 기술을 중심으로 개발되고 있다. 이는 초기 시스템 반도체 형태에서 점차 생체모사소재 반도체로 발전할 것으로 예상된다. AI 연계 데이터·네트워크 기술 측면에서는 X(ICBM) for AI 또는 AI for X 형태로 AI SW기술과 긴밀하게 연계될 것으로 전망된다. 또한 최신 AI 기술(딥 러닝 등)의 이론적 한계를 극복하기 위하여 기초과학 중 뇌과학·산업수학 연구에 보다 집중할 것으로 예상되며, 중장기적으로는 인간·기계 협업 지원을 위한 뇌-기계 인터페이스(BCI) 분야가 특히 발전할 전망이다.

## 2. 지능정보기술 R&D 동향 (AI HW, AI SW, 기초과학)

우리나라는 AI 기술 시장으로의 진입이 상대적으로 늦었기 때문에 AI 기술에 대한 연구 개발 투자가 늦어졌다. 하지만 최근 4차 산업혁명의 중요성이 더욱 부각되고 국내외적 관심도가 높아짐에 따라 정부 및 민간의 AI 관련 R&D가 확대·강화되고 있는 추세이다. 정부는 '지능정보산업 발전전략(2016. 3월)'과 '지능정보사회 중장기 종합대책(2016. 12월)'에 따라 2017년 AI 기술 관련 R&D를 추진 중에 있다.

AI R&D는 알파고와 이세돌 간의 바둑 대결(2016. 3. 9~3. 15) 이전부터 정부와 민간에서 추진되어 왔는데 알파고로 인해 AI 분야로의 관심이 고조되면서 AIRI 등의 연구소를 설립하여 시각지능, 언어지능 분야 위주의 R&D 투자가 진행되고 있다. 민간의 AI 플랫폼 관련 기업으로 '솔트룩스', '와이즈넷'이 사업을 진행하고 있으며, AI 의료 영상진단 스타트업 '루닛', '뷰노' 등이, 기계 번역 분야의 '시스트란' 등이 활동 중이다. 다음 <표 2-3>은 AI 기술 분야별 발전되는 양상과 지원 예산 등을 정리한 표이다.

〈표 2-3〉 2017년 AI 연구 분야별 주요 내용 및 예산

구분	내용
AI SW분야 원천기술 개발 확대	<ul style="list-style-type: none"> <li>- 산업적 수요가 높은 언어·시각·음성 지능 분야의 원천기술 개발 (엑소브레인·딥뷰·지니톡 등)을 고도화·발전시키고, 중장기적 기술 우위 확보를 위해 선도적으로 차세대 학습·추론 등 AI 기술을 연구하는 'AI 국가전략프로젝트'를 추진 (17년 239억원)</li> <li>- 또한, AI기술 응용 제품·서비스 창출을 촉진하기 위해 가상비서 플랫폼 원천기술을 개발하는 '플래그십 프로젝트*(17년 145억원)'와 노인 돌보미·무인 경계로봇 등을 개발하는 'AI-로봇 융합사업 (17년 100억원)' 등도 추진</li> </ul>
AI HW 기반 확보: AI SW에 특화된 AI HW 기반 확보도 추진	<ul style="list-style-type: none"> <li>- 고성능 컴퓨팅 인프라를 통해 고용량 AI SW의 원활한 실행이 가능토록 관련 슈퍼컴퓨팅 기술개발(17년 62억원)에 투자</li> <li>- 칩·소자 단위에서 기계학습 SW 실행을 최적화하는 지능형반도체·뇌신경모방칩(뉴로모픽칩) 원천기술 개발(17년 196억원)도 추진</li> </ul>
기초기술 분야에 안정적 투자	<ul style="list-style-type: none"> <li>- AI 기술 수준을 혁신적으로 도약시킬 차세대 기술 창출 기반을 조성하기 위해 기초기술 분야에도 투자할 계획</li> <li>- 뇌의 구조, 활동원리 등 뇌과학 연구를 확대('16년 507억원 → '17년 618억원)하고, 대학 산업수학 센터 지정 등 산업수학을 본격 지원(17년 15억원)</li> </ul>

자료: 한국지식재산연구원(2016) "AI(AI) 기술 및 정책 동향"

정부는 지능정보기술 역량 확충으로 5대 과제를 추진하면서 향후 5년간 총 1조원을 투자해 2016년 5월에 '지능정보사회 민관합동 추진협의회'를 구성한 '지능정보사회 중장기 종합 대책수립'을 추진했다. 또한 2017년 지능정보기술 핵심 분야인 AI SW에 대한 연구를 본격화하고 그 외 AI HW, 기초기술을 포함한 3개 분야에 총 1,630억원 규모의 투자를 진행할 예정이다. 2017년 AI기술 R&D 예산은 AI SW(739억원), AI HW(258억원), 기초기술(633억원) 등 총 1,630억원으로, 이는 2016년 관련 예산 1,106억원에 비해 47% 증가된 규모이다. 응용 서비스확산(200억원)과 인력양성 등 산업기반구축(427억원)을 포함한 AI 관련 예산을 고려한다면 2017년 총 2,257억원 규모로 2016년 1,388억원 대비 63% 증가한 셈이다.

〈표 2-4〉 AI기술 R&D 예산 현황

(단위: 억원)

기술 구분	'16년(A)	'17년(B)	증감(B-A)	증감률(%)	'17년 주요 사업
AI SW	407	739	332	81.6	AI 국가전략 프로젝트(239억원) 플래그십 프로젝트(145억원) AI로봇 융합(100억원) 기타 AI R&D 주요 과제 (255억원)
AI HW	192	258	66	34.4	지능형반도체(166억원) 슈퍼컴퓨팅(62억원) 뉴로모픽칩(30억원)
기초	507	633	111	24.9	뇌인지·뇌공학 분야(618억원) 산업수학(15억원)
계	1,106	1,630	524	47.4	

자료: 과기정통부(2017), “2017년 지능정보사회 선도를 위한 AI R&D”

정부는 민간이 주도하고 정부가 기초·원천기술 R&D를 수행하는 구조가 바람직하다는 판단에 따라 AI 분야에 대한 전략적 R&D를 추진하여 AI R&D 결과물을 민간분야에서 쉽게 활용할 수 있도록 할 계획이다. 이를 위해 엑소브레인의 언어처리 주요 요소기술을 산·학·연이 자유롭게 활용할 수 있도록 API 형태로 공개(2017. 9, 한국어 언어처리 7종<sup>6)</sup>)하고, 법률·특허·금융 분야의 사업화 기술개발(2017. 3-2020. 1)도 시작할 예정이다.

한편, 최근 SK, 네이버, 카카오 등의 대기업들도 AI 기술에 대한 투자 계획 발표를 이어 가는 등 AI 기술 관련 R&D가 활발해지는 추세이다. SKT는 AI 전담 조직인 'T 브레인(2016. 10)'을 설립하고 AI 등 신산업 분야에 3년간 5조원 투자 계획이고, 네이버는 미래기술 개발 전담 조직인 '네이버랩스(2017. 1)', 'J팀(2016. 10)' 등을 설립하고 관련 분야에 5년간 5,000 억원 투자 계획이라고 밝힌 바 있다. 카카오는 AI 생태계 발전 및 차세대 성장 동력 확보를 위해 200억원을 투자해 AI 전문 자회사 '카카오 브레인(2017. 2)' 설립했다. 또한 〈표 2-5〉와 같이 엑소브레인, 딥뷰, 차세대 머신러닝 등 舊 미래부(現 과학기술정보통신부)를 중심으로 주요 AI 핵심기술을 개발해 AI 분야관련 정책개발, 연구개발, 사업화 방안 등을 포함하는 통합적 R&D 정책수립을 위한 다방면의 노력을 추진한 바 있다.

6) 한국어 분석 7종 API: 형태소, 개체명 인식, 다의어 분석, 의존 구문 분석, 의미역 인식, 어휘 간 유사도 분석, 동음이의어 분석

〈표 2-5〉 국내 주요 AI R&D 과제 현황

AI명	내용
엑소브레인	- 지식공유지능 진화가 가능한 AI SW 개발이 목표 - 2013부터 23년간 1,070억원 투자
딥뷰	- 대규모 실시간 영상 이해 기반의 시각 지능 플랫폼 개발을 목표
AI 관련 빅데이터	- 초고성능 빅데이터 에코시스템 개발이 목적 - 2015년부터 4년간 129억원 지원

자료: KCA “국내 AI 기술개발의 현주소”, AIIA 재구성

### 제 3 절 해외 지능정보기술 R&D 동향

#### 1. AI 활용을 위한 요소 기술

##### 가. 빅데이터(AI 학습용 데이터)

AI(기계학습)이 기업에게 있어 가장 중요한 혁신의 도구이자 자원으로 등장하고 있다. 이러한 AI를 혁신적인 도구이자 자원으로 활용(application)하고, 적용(adoption)하기 위해서는 알고리즘, 데이터 및 클라우드 컴퓨팅이 필수적이다.

일반적으로 AI(기계학습)의 활용에 있어서 알고리즘과 데이터에 의해서 그 성과가 달라지는데, 알고리즘은 현재 오픈소스를 통해 공유되고 있다. 현재 구글(TensorFlow), 페이스북(Bigsur), 마이크로소프트(DMTK, CNTK), OpenAI, IBM(System ML) 등 글로벌 선도 ICT 기업들은 지능정보의 핵심기술 소스코드 및 API, 트레이닝 및 테스트베드 환경 등 개발 플랫폼을 전 세계에 무료로 공개하며 지능화를 가속하고 있다. 특히 알고리즘을 소프트웨어로 구현하여 모듈화한 라이브러리, 그리고 이를 사용할 수 있게 하는 인터페이스 등으로 구성된 AI 플랫폼이 등장하고 있다. 따라서 AI(기계학습) 기술은 그 기본적인 기법이나 원리가 오픈소스로 공개되어 있으며, 광범위하게 다양한 분야에서 융합될 수 있는 상황이다.

AI 차별화의 기초는 고객·파트너·기업 내부로부터 발생하는 데이터를 수집·저장하여 AI로 분석할 수 있도록 지원하는 것에 존재한다. AI(기계학습)의 활용과 적용은 결국 데이터 구축 및 관리에 기반을 둘 수밖에 없으며, 양질의 데이터는 AI(기계학습)의 성능을 좌우하는 가장 중요한 요소이다. 특히 이러한 AI(기계학습)용 데이터 구축을 위해서는 데이

터 태그 표준화 및 데이터 레이블링(labeling) 등이 반드시 필요하다.

예를 들어 시각지능 서비스의 개발에는 일차적으로 ImageNet, 마이크로소프트 coco, PASCAL VOC, UCF-101, CIFAR100 등의 구조화되고 태그 정보가 부착된 공개 학습 데이터셋이 활용되고 있다. 현재 ICT 선도기업과 스탠퍼드, 토론토 대학 등 연구기관이 협력하여 사진, 동영상, 텍스트 등 AI 연구개발 생태계를 조성하고 확산시키기 위한 대규모 데이터셋 공개하고 있다. 기업들은 이러한 공개 데이터셋을 활용하여 AI 서비스를 만든 후 필요시 자체 데이터셋을 구축하기도 한다.

AI를 활용하기 위해서는 AI 학습을 위한 대용량의 데이터가 필요한 동시에 이러한 빅데이터를 처리하기 위해서 고성능의 컴퓨팅 인프라인 클라우드 컴퓨팅이 필수적이다. 때문에 빅데이터 및 클라우드 시장 규모가 성장하고 있는 추세이다. IDC(2017)에 따르면 5개 산업 분야(뱅킹, 조립제조, 공정제조, 연방 및 중앙정부, 전문 서비스)가 2017년에 빅데이터 및 분석 솔루션에 742억 달러 투자할 것으로 전망된다. 또한 2020년에는 빅데이터 및 분석 솔루션 투자금액이 1015억 달러 규모에 이를 것으로 전망되며, 가장 빠른 성장세는 뱅킹 산업에서 연평균 성장률 13.3%에 이르며 그 다음으로 헬스케어, 보험, 증권 및 자본투자중개업이 빠르게 성장할 것으로 전망된다. 451 Research(2016)에 따르면 전 세계 클라우드 시장 규모 또한 점차적으로 성장하여 2020년에 440억 달러에 달할 것으로 보인다. 다음 <표 2-6>과 <그림 2-7>은 빅데이터 시장 규모를, <표 2-7>은 클라우드 시장 규모를 나타낸 자료이다.

<표 2-6> 세계 빅데이터 시장 규모

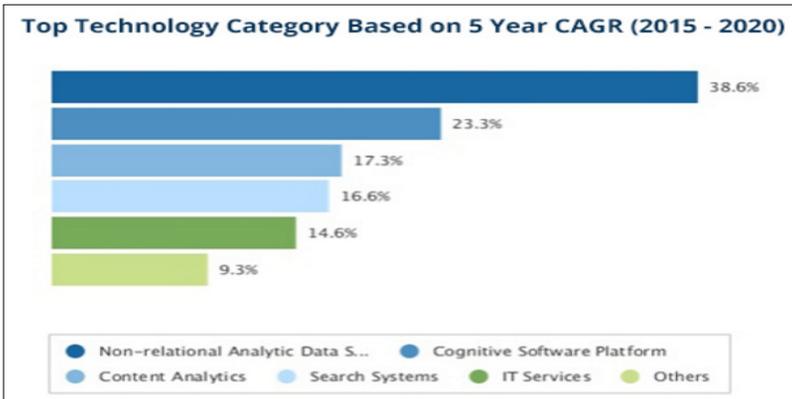
(단위: 십억 달러)

구 분	2013	2014	2015	2016e	2017e	2018e	2019e	2020e	2021e
빅데이터	12.3	19.6	18.3	22.6	27.3	33.5	40.8	49.0	57.3

자료: Wikibon(2016. 7)

주:e는 전망치를 의미

[그림 2-7] 빅데이터 세계 시장 규모



자료: IDC(2016), "World Wide Semiannual Big Data and Analytics Spending Guide"

해외 유수의 기업들은 데이터를 활용해 기업의 경쟁력을 높이고 있으며 다양한 형태와 빠른 속도로 데이터를 가치 있게 활용하고 있다. 최근 데이터가 경제, 사회 전반적으로 혁신을 주도하는 거대한 플랫폼으로 자리 잡아 4차 산업 혁명의 핵심요소로 작용하는데 도움을 주고 있다.

<표 2-7> 데이터를 활용한 해외 비즈니스 성공 사례

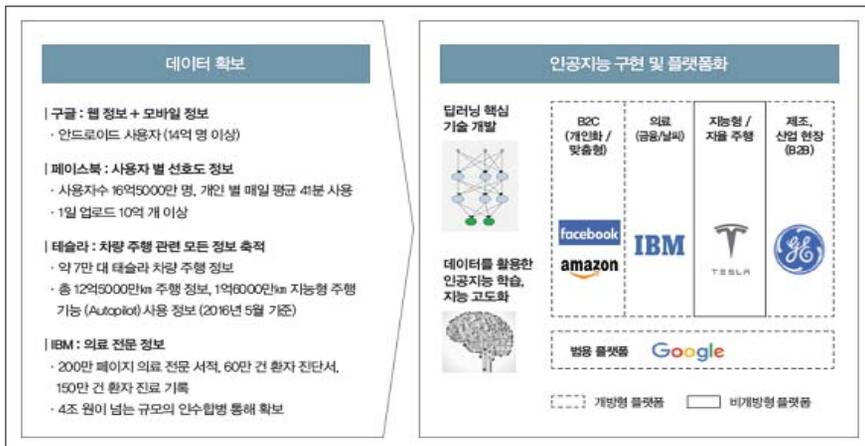
기업	주요사례
	기업명: 인텔 (미국) 주요사례: 칩 제조 과정에서 발생하는 데이터 분석 제조비용: 300만달러 (한화 약32억)
	기업명: 몬산토 (미국) 주요사례: 날씨·토질·온도·위치정보 등 농장에서 수집할 수 있는 수백가지 데이터 분석을 기반으로 최적화 된 작업 순서부터 예상 수확일자, 수확량까지 파악 (2013년에 빅데이터 분석업체 'Climate Corporation'을 10억 달러에 인수)
	기업명: Zillow.com (미국) 주요사례: GIS, 인구통계, 학군, 부동산 데이터 등 부동산 매매에 필요한 데이터를 통합 제공해 창업 6년 만에 기업 가치 10억달러(한화 1조원)

기업	주요사례
<b>SIEMENS</b>	기업명: 지멘스 (독일) 주요사례: 제조 설비에서 발생한 데이터 분석. 생산 라인을 재조정해 생산량 8배 증가
<b>TESCO</b>	기업명: 테스코 (영국) 주요사례: 냉장 데이터를 분석해 영국과 아일랜드 3,000개 점포에서 냉장비용 연 20% 절감
 岐阜大学 GIFU UNIVERSITY	기업명: 기후대학교 부속병원 (일본) 주요사례: 50만 명에 달하는 환자 진료 기록과 연간 4,000건의 수술 데이터를 분석해 재수술 비율 3% 감소, 약제비 최적화로 연간 2억 엔의 의료비 절감
<b>KOMATSU</b>	기업명: 코마츠 (일본) 주요사례: 약 40만 대의 건설 장비 가동 상황 데이터를 GPS 원격 관리 시스템으로 수집, 수요 예측과 수리, 중고차 가격 결정 등에 활용해 10% 이상의 영업이익률 제고

자료: 한국정보화진흥원 『새로운 기술, 새로운 세상 지능 정보사회』(2016. 12) AIIA 재구성

이와 같이 해외에서는 AI 플랫폼을 선점하기 위한 기업의 전쟁이 시작되면서 기업들은 본격적인 AI 기술과 지능학습을 위한 데이터 확보 측면에서 사업을 확장하는데 노력을 기울이고 있다.

〔그림 2-8〕 주요 글로벌 기업의 보유 데이터와 AI 플랫폼화 현황



자료: Tech M 『플랫폼 승자가 모든 산업 생태계 주도한다』(2017. 2)

구글의 경우 인터넷 등의 환경에서 많은 양의 데이터를 확보하고 있으며 페이스북과 아  
마존은 개별 사용자의 성향을 정교하게 분석이 가능한 개인 데이터를 확보하고 있다. 구  
글이 보유한 상당 부분의 정보가 공공데이터인 반면, 페이스북과 아마존이 수년간의 서비  
스를 제공하면서 수집한 정보들은 개별 사용자의 성향을 분석하는 민간 데이터가 중심인  
것으로 전해졌다.

AI의 성능이 기계학습 과정에서 주어지는 데이터에 의해 상당 부분 결정되는 것을 감안  
하면 개인 맞춤형 서비스 제공을 위한 AI 플랫폼은 페이스북과 아마존이 구글 보다 우세  
할 가능성이 있다.

〈표 2-8〉 미국 기업의 AI-데이터 활용 사항

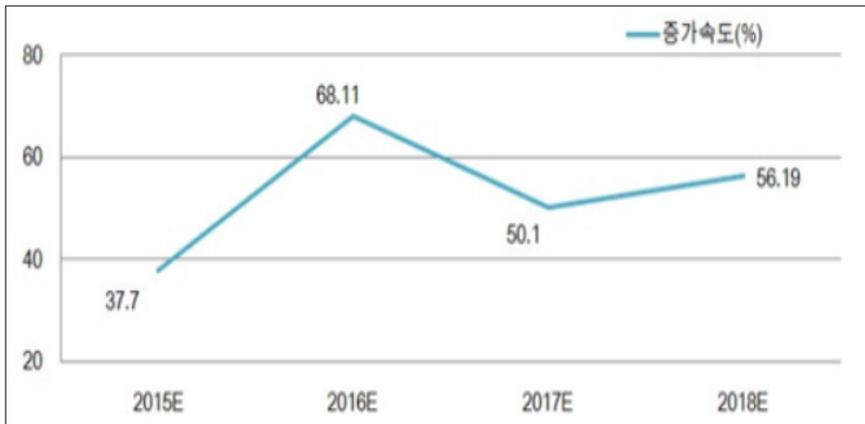
기업	활용사항
	<ul style="list-style-type: none"> <li>- 고도화된 AI를 구현하는데 필요한 핵심요소인 알고리즘, 데이터에 대해 최고 수준의 역량을 확보하고 있어 다양한 산업에 적용가능하다는 이점을 가지고 있음</li> <li>- 높은 다양성을 갖는 데이터를 AI의 기계학습 과정에 활용이 가능해 넓은 분야를 포괄하면서 높은 수준의 AI를 고도화 시키는데 주력</li> </ul>
	<ul style="list-style-type: none"> <li>- 최근 AI 고도화를 위한 데이터 확보에 박차를 가하고 있으면서 사용자들의 성향, 특성을 유추 할 수 있는 데이터를 집중 확보해 세분화시켜 축적함</li> <li>- 데이터 영역에서 구글에 비해 상대적으로 경쟁력이 낮지만 개별 사용자의 특성을 정확히 분석 가능해 AI를 통한 개인화, 맞춤화된 가치 제공 가능 전망</li> </ul>
	<ul style="list-style-type: none"> <li>- 약 7만대의 테슬라 차량 주행 정보와 총 12억 5,000만km 주행 정보, 1억 6,000만km 지능형 주행 등 차량 주행 관련 모든 정보를 축적</li> </ul>
	<ul style="list-style-type: none"> <li>- 200만 페이지 의료 전문 서적, 60만 건 환자진단서, 150만 건의 환자 진료 기록과 4조원이 넘는 규모의 인수 합병을 통해 의료 전문 정보를 확보</li> </ul>

자료: Tech M 『플랫폼 승자가 모든 산업 생태계 주도한다』 AIA 재구성(2017. 2)

중국에서는 막대한 인구가 쏟아내는 정보의 양을 통해 AI 구축의 핵심인 빅데이터 형성 측면에서의 월등한 경쟁력을 강점으로 AI 기술 개발에 노력을 기울이고 있다. 인터넷 통제 강화로 데이터 활용 여부가 다소 어렵지만 많은 양의 데이터 생성을 가능하게 해 AI 기반 형성에 기여하고 있다.

[그림 2-9]에서 보는 바와 같이 중국은 현재 전 세계 데이터의 약 13%를 생성해 내고 있으며, 골드만 삭스는 2020년까지 그 비중이 약 20~25%로 증가할 것이라고 전망<sup>7)</sup>했다. 더불어 중국은 미국, 유럽 등의 선진국과 비교하여 개인정보 보호 수준이 낮은 편이어서 기업이 필요한 정보를 수집하기에 용이하다.

[그림 2-9] 중국 빅데이터 산업규모 증가율



자료: 초이스경제 『중국 '2020년 세계 빅데이터 중심' 노리나』(2016. 6)

[그림 2-10]에서 보는 것과 같이 2016년 중국 AI 산업 규모는 이미 100억 위안(약 1조 6천억원)에 이르렀으며 지속적인 성장을 통해 올해는 300억위안(약 5조) 돌파를 예상하고 있다.

7) <https://www.cnbc.com/2017/09/01/goldman-says-china-has-talent-data-and-infra-structure-to-embrace-ai.html>

[그림 2-10] 2014~2019년 중국 AI 산업규모



자료: KOTRA 『중국 AI산업 G2로 부상』 (2017. 12)

아마존·구글·페이스북 등 미국 소프트웨어 기업들이 주도하는 빅데이터 분야의 도전자는 중국으로서 그 선두에는 중국 3대 IT 기업인 바이두, 알리바바, 텐센트가 존재하고 있다.

또한, 중국은 자국의 스마트폰과 인터넷 사용자들이 쏟아내는 빅데이터를 수집 및 활용해 급성장 중에 있으며 데이터센터 투자는 매년 30% 증가해 기업 가치 1조원이 넘는 빅데이터 관련 스타트업들도 쏟아져 나오고 있다.

<표 2-9> 미국·중국 간 주요 IT기업 데이터 경쟁현황

기업	주요내용
	- 데이터센터 수: 15개(미국, 칠레, 대만, 싱가포르 등) - 구글 클라우드(가상 저장 공간 서비스), 자율 주행 자동차, 검색, 모바일OS, 지도 등 AI 비서
	- 데이터센터 수: 100개(미국, 브라질, 남아공 등) - 기업용 클라우드 서비스, MS오피스 365, 윈도우OS
	- 데이터센터 수: 44개(미국, 독일, 영국 등) - 아마존웹서비스의 클라우드, AI 비서

기업	주요내용
	<ul style="list-style-type: none"> <li>- 데이터센터 수: 14개(중국, 미국)</li> <li>- 클라우드, 자율 주행차, 모바일 상거래, AI 비서</li> </ul>
	<ul style="list-style-type: none"> <li>- 데이터센터 수: 20개(중국, 미국 등)</li> <li>- 메신저 '위챗' 서비스, 텐센트 클라우드, 온라인 게임</li> </ul>

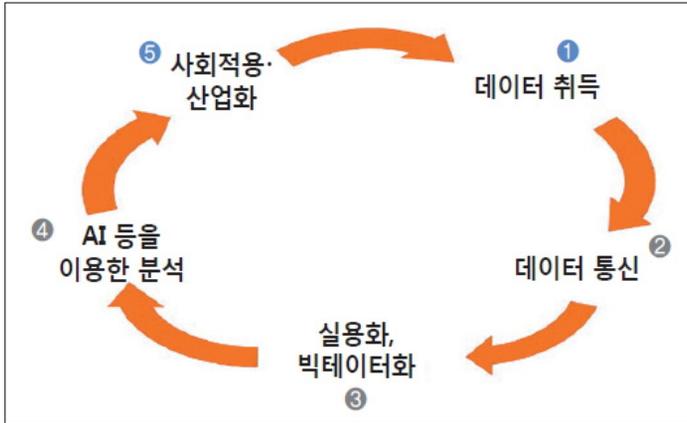
자료: 조선비즈 『폐북은...(중략).. '빅데이터 장악작전』(2017. 7), AIIA 재구성

일본은 일본만의 강점 분야를 적극 활용하고, 데이터 기반 선순환형 비즈니스 모델을 추구해 로봇, 센서 디바이스, 네트워크 인프라, 현실 데이터, 컴퓨터 개발능력 등에서 세계 최고 수준에 이르고 있다. 그리고 기술 및 비즈니스 모델 개발 시 데이터의 취득, 공유, 실용화(빅데이터화), 분석, 활용하는 과정에서 각 단계별로 문제점을 파악하여 해결하고 있다.

일본에서는 사이버 공간과 물리적 공간을 고도로 융합하는 매개체이자 경제성장과 사회 문제 해결이라는 목표를 달성하는 핵심수단이 데이터임을 강조하고 있다. 보건의료 등 다양한 분야에서 비즈니스를 창출하기 위해 현실 데이터 플랫폼 구축과 인간을 지원하는 다양한 형태의 로봇틱스 활용을 결합하고 있다.

이를 바탕으로 데이터의 이용을 위한 데이터 사이클상에서 공장이나 의료보험 등의 현실 데이터 수집과 산업화에도 강점을 가지고 있다. 일본은 세계 최고 수준의 로봇, 센서 등이 존재하면서 공장자동화 등의 제조현장 관련 데이터, 의료보험, 교통 등의 데이터가 풍부해 자동차 등의 높은 세계시장 점유율, 고품질 제품을 이해하고 평가하는 소비자 존재도 강점으로 꼽히고 있다.

[그림 2-11] 데이터 이용 및 활용을 위한 기본 사이클



자료: KIET 『일본의 4차 산업혁명 대응실태와 시사점』(2017. 1)

하지만 일본은 빅데이터 처리 규모가 작아 상대적으로 구축에 한계가 있으며, AI 등을 활용한 분석이 취약하다. 일본의 주요 전자업체들은 미국 기업과의 경쟁에서 생존하기 위해 일본을 대표하는 8개 전자업체가 2019년까지 총 3,000억엔(약 2조 9천억원)을 AI 관련 사업에 투자할 계획이다.

<표 2-10> 일본 8대 전자기업의 AI 활용 현황

구 분	내 용
<b>HITACHI</b>	- 3년 동안 약 1,000억엔을 AI 관련 연구에 투자
<b>SONY</b>	- 1999년 애완로봇 아이보(AIBO)와 2003년 휴머노이드 로봇 QRIO(Quest for Curiosity)를 제작 일본 AI의 선구자 역할을 해옴 - 2016년 5월 실제 세계와 상호작용하며 지속적으로 자율적인 학습이 가능한 AI를 개발 중인 미국의 스타트업 코기타이(Cogitai)에 출자 - 연구개발을 멈추지 않고 있어 코기타이와 협력해 파생된 제품 혹은 서비스를 출시할 계획
<b>Panasonic</b>	- CTO 요시유키 미야베: 2017년 4월, 기업의 매출 감소를 막기 위해 5년 내로 AI와 IoT 관련 전문가 1,000명 모집 선포

구 분	내 용
	<ul style="list-style-type: none"> <li>- 2018년까지 500억엔을 투자, AI 관련 직원을 두배로 늘려 1,500명 고용 계획</li> <li>- AI 관련 기초 연구 투자비용까지 합치면 1,000억엔에 이를 것으로 추정</li> <li>- 2015년 1월 GPU에 특화된 오픈소스의 딥러닝 모듈 토치 (Torch) 공개</li> </ul>
	<ul style="list-style-type: none"> <li>- 인간 전문가의 도움을 받고 있는 기계학습 수준에서 AI에 기반한 완전한 자동화 공정으로, 만나질 걸리던 작업을 30분으로 줄이는 등 AI 개발에 성과가 있다고 발표 (2017. 5)</li> </ul>
	<ul style="list-style-type: none"> <li>- 얼굴 인식 기술에 적극적으로 투자</li> </ul>
	<ul style="list-style-type: none"> <li>- 지금은 쇠락했다는 평가를 내리고 있음</li> <li>- 세계 500대 기업에 속한 경험을 기반으로 AI 관련 연구에 적극적으로 투자 중</li> </ul>

자료: 뉴스 Y 『일본 8대 전자기업들의 AI 투자 현황』 자료 취합, AIA 재구성(2017. 10)

#### 나. 클라우드

고성능 및 대용량의 컴퓨팅 자원이 요구되는 AI 분야의 경우, 전문적인 공급업체가 서비스 지원을 제공하고 있다. 최근에는 IBM(Bluemix), 마이크로소프트(Azure Machine Learning), 아마존(Web service) 구글(Cloud Platform) 등 대용량의 컴퓨팅 자원과 데이터 센터를 보유하고 있는 글로벌 ICT 기업들이 AI 지원 클라우드 서비스를 제공하고 있다.

AI 관련 개발 플랫폼·개발도구와 데이터가 개방·공유되고, 이를 지원하는 클라우드 서비스가 주요 ICT 기업에 의해 제공되고 있다. 따라서 기업들이 AI을 활용하고 적용할 수 있는 환경이 확대되고 있다. 기본적으로 모든 기업들이 AI을 외부(시장)에서 조달하여 활용할 수 있게 될 전망이다.

〈표 2-11〉 세계 클라우드 시장규모

(단위: 십억 달러, %)

구 분	2014	2015	2016e	2017e	2018e	2019e	2020e
클라우드서비스	12.4	16.9	21.9	27.1	32.5	38.1	44.1
PaaS	2.3	3.1	4.0	4.9	5.9	6.9	8.1
SaaS	3.1	4.1	5.4	6.8	8.4	10.0	11.6
IaaS	7.0	9.7	12.5	15.4	18.2	21.2	24.4

자료: 451 Research(2016)

주: e는 전망치를 의미

이러한 AI 기술의 보편성으로 인하여 이를 자원으로 활용하는 기업들은 경쟁 우위를 점하기가 어려울 전망이다. AI 기술의 활용은 모든 기업에게 있어서 필수적이지만 이를 적용하여 차별화된 역량을 발전시키는 것은 매우 어려운 일이다.

## 2. 해외 AI 시장 규모 및 전망

AI 기술과 연관이 깊은 요소기술의 발전으로 AI 기술의 실현이 크게 변화하였다. 이에 따라 AI 시장 규모도 급격히 확대되고 있다. BCC Research(2015)에 따르면 세계 AI 시장규모는 급속히 성장하고 있으며, 매년 20.6%의 성장률로 성장할 것으로 전망된다. 다음 <표 2-12>는 세계 AI 시장 규모 예측을 나타낸 자료이다.

<표 2-12> 세계 AI 시장 규모

(단위: 백만 달러)

구 분	2013	2014	2019P	2024P	CAGR (14~24)
AI	5,339	6,292	15,279	41,215	20.6

자료: BCC Research(2015); Siemens webpage

주: 1) p는 잠정치를 의미

2) AI 기술에 포함된 세부 기술 분류는 Expert Systems, Autonomous robots, Digital assistants, Embedded systems, Neurocomputers를 포함

2011년부터 AI 개인비서 관련 서비스가 활발히 출시되었으며, 해당 시장의 지속적인 성장이 전망된다. 2011년 애플 Siri(시리), 2012년 구글 Now(나우), 2014년 마이크로소프트 Cortana(코타나), 2015년 페이스북 M(엠) 등 글로벌 기업에서 음성인식 및 개인비서 기술 상용화를 진행했다. 시장조사업체 Strategy Analytics(2017)은 2017년 3분기 세계 AI 스피커 출하량이 전년 동기 90만 대에서 7배 급증한 740만대라고 보고했다. 아마존 Alexa(알렉사)를 탑재한 AI 스피커(전체 AI스피커 판매량의 약 67%)가 가장 많이 판매되었으며, 그 다음으로 많이 판매된 AI 스피커는 구글의 스피커(약 25%)로 집계되었다. 다음 <표 2-13>은 음성인식 비서 디바이스 출하량을 나타낸 표이다.

〈표 2-13〉 세계 음성인식 비서 디바이스 출하량

(단위: 만 대)

구 분	2015	2016	2017P	2018P	2019P	2020P	CAGR (‘15~’20)
음성인식 비서 디바이스	110	180	290	520	890	1510	68.9

자료: Strategy Analytics(2017)

주: p는 잠정치를 의미

AI 기술이 적용된 자율주행자동차의 시장규모 또한 향후 20년 간 2,900억 달러(한화 약 324억 원)에 육박할 것으로 전망된다. 골드만삭스에 따르면 첨단 운전지원시스템 및 자율주행자동차의 시장규모는 2015년 약 30억 달러(3조 원), 2025년 960억 달러(107조 원)이 될 것으로 보인다. 다음 〈표 2-14〉는 2025년까지 BI Intelligence(2016)이 전망한 세계 자율주행자동차 출하량을 나타낸 표이다.

〈표 2-14〉 세계 자율주행자동차 출하량

(단위: 천 대)

구 분	2018P	2019P	2020P	2021P	2022P	2023P	2024P	2025P
자율주행 자동차	64	107	149	213	320	426	512	569

자료: BI Intelligence(2016); Business Insider(2017)

주: p는 잠정치를 의미

세계 AI 관련 시장은 급격히 확대되고 있으며, 확대 추세가 가까운 미래까지 계속될 전망이다. 이러한 상품 및 서비스에 활용되는 기술 개별 시장 규모 또한 지속적으로 확대되고 있으며, 개별 규모 변화 추이는 상이하다. 특히 딥러닝 기술이 AI 기술 중 가장 높은 시장 성장률을 보이고 있으며, 그 다음으로는 시각 인식 분야, 자연어 처리 순으로 시장 규모가 증가하고 있다.

〈표 2-15〉 AI 기술 분야별 시장 규모 전망

Technology	2015	2016	2017	2018	2019	2020	2021	2022	CAGR
Cognitive Computing	10.9	111.6	12.4	13.2	14.2	15.0	15.9	16.9	6.5%
Machine Learning	13.6	14.8	16.2	17.9	19.8	21.4	23.6	25.8	9.6%
Deep Learning	108.5	221.1	453.1	929.5	1904.6	3884.9	7935.8	16228.7	104.5%
Predictive APIs	20.3	22.9	26.4	30.8	36.0	43.4	50.5	58.8	16.4%
Natural Language Processing	7.8	10.1	13.0	16.6	21.1	26.3	33.5	42.7	27.5%
Image Recognition	24.4	32.6	44.7	61.7	85.7	119.1	163.5	224.4	37.3%
Speech Recognition	14.6	16.6	19.3	22.6	26.4	30.9	35.8	41.5	16.1%
Other	2.3	2.4	2.6	2.8	3.1	3.5	3.9	4.2	9.1%
Total	202.5	332.1	587.5	1095.1	2111.0	4144.7	8262.4	16643.1	87.7%

자료: Tratica(2015), "Artificial Intelligence for Enterprise Applications", ETRI(2015. 10) "AI 기술과 산업의 가능성", IITP(2016), "ICT R&D 중장기 기술로드맵 2022"

### 3. 국가별 지능정보기술 R&D 전략 및 정책

현재 해외 주요정부들은 AI에 활용될 데이터의 선순환 생태계 구축을 위해 데이터 구축, 유통, 활용 등 균형 잡힌 역할<sup>8)</sup>을 하기 위해 노력 중이다. 이에 해외 주요 정부에서는 민간에 필요한 AI 분야 공공데이터를 지속적으로 가공하고 개방시켜 양질의 데이터 개방 확대와 다양한 민간기업 육성을 위해 정부의 균형 잡힌 역할 수행을 하는데 중점을 두고 있다. 특히 미국, 일본, 중국 등 주요국에서는 AI의 중요성을 인지하고, 정부는 발 빠르게 관련 정책을 마련하고 적극적인 투자에 집중하고 있다.

#### 가. 미국

##### 1) 과학 및 공학

미국의 민간 부문 R&D 투자는 구글, 아마존, IBM 등의 글로벌 기업을 중심으로 AI 기술 혁신을 견인하고, 정부는 기초·원천 분야 장기 투자를 실행 중이다. 미국 연방정부의 과학 및 공학에 대한 지원 총액은 매년 감소하고 있는 추세이며, 2014년 1,003개 학술기관이

8) 세계은행 '경제성장을 위한 오픈 데이터' 보고서: 세계적으로 데이터의 활성화를 위한 정부의 균형 잡힌 역할을 강조

311억 달러를 지원받았던 것에서 2% 감소한 금액인 305억 달러를 총 1,016개의 학술기관이 지원 받았다. 전년 대비 2015년 국립과학재단(NSF)의 지원은 약 2억 달러(4%) 증가했으며, 보건복지부(HHS), 국립과학재단(NSF), 국방부(DoD) 3개의 연방기관 지원 금액이 2015년 전체 85%를 차지했다.

[그림 2-12] (미국) 연방정부 지원 예산('12~'15)

Fiscal year	All agencies	HHS	NSF	DOD	USDA	DOE	NASA	Other agencies*
Current \$millions								
2012	31,003	17,939	5,181	3,653	1,221	1,037	984	989
2013	29,353	16,881	5,023	3,384	1,136	1,077	1,004	846
2014	31,095	17,517	5,092	3,714	1,301	1,527	1,087	857
2015	30,494	17,008	5,295	3,501	1,424	1,272	1,135	859
Constant FY 2009 \$millions								
2012	29,580	17,116	4,943	3,485	1,165	989	939	944
2013	27,533	15,834	4,712	3,174	1,066	1,010	942	795
2014	28,677	16,155	4,696	3,425	1,200	1,408	1,002	790
2015	27,747	15,476	4,818	3,186	1,296	1,157	1,033	782

DOE = Department of Energy; HHS = Department of Health and Human Services; NASA = National Aeronautics and Space Administration; NSF = National Science Foundation; USDA = Department of Agriculture.

자료: IITP(2015), "2015 연방정부의 과학/공학 분야 지원 통계"

<표 2-16> (미국) 과학 및 공학 지원 내용

구분	세부 내용
R&D 지원	보조금, 계약 또는 협력을 위한 개인 및 조직 연구 개발 수행에 필요한 직/간접적 비용 지원
R&D Plant	R&D 수행 위한 시설, 토지, 고정 장비 건설, 취득, 수리, 대여 등을 지원
교육용 시설 및 장비	교육에 사용되는 시설, 토지, 고정 장비 건설, 취득, 수리, 대여 지원
Fellowship, 연수회, 트레이닝 지원	인력 개발 및 유지를 위한 프로그램 지원
일반 지원	자금 및 특정분야 활동 지원
기타 과학 및 공학 지원	학술 과학 및 공학 지원, 기술회의, 교육기관, 예비 학부생 프로그램 지원

자료: IITP(2016), "미국의 빅데이터 R&D 전략계획"

2) 빅데이터(AI 학습용 데이터)

미국 정부는 범부처 ICT R&D 프로그램 중 일부로 머신러닝 및 AI 분과위원회(NITRD)가 발표한 ‘빅데이터 R&D 전략 계획’을 수립하였다. 인력, 사이버인프라, 윤리 등에 대한 정책을 세부적으로 나눠 이행할 것으로 보인다. 다음 <표 2-17>은 미국 정부의 빅데이터 관련 전략 계획을 정리한 표이다.

<표 2-17> (미국) 빅데이터 R&D 전략계획 세부 내용

전략	세부과제
미래 빅데이터 특성을 반영한 기술 개발로 차세대 능력 함양	데이터 크기, 전달·처리 속도, 복잡성을 고려한 기술 개발 새로운 빅데이터 기술 방법론 개발
데이터의 신뢰성 및 더 나은 빅데이터 기반 의사결정을 위한 R&D 지원	데이터의 신뢰성과 타당성 제고 데이터 기반 의사결정 지원 도구 개발
빅데이터 혁신을 가능하게 하는 사이버 인프라 구축 및 강화	국가 데이터 인프라 강화 빅데이터 응용과학 사이버 인프라 역량 강화 유연하고 다양한 인프라 지원
데이터 공유 및 관리를 촉진하는 정책을 통한 데이터 가치를 향상	메타데이터의 모범사례 개발 데이터 자산에 효율적·지속적·안전한 접근 제공
개인정보 보호, 보안 및 빅데이터의 수집·공유·활용의 윤리적 측면의 이해	올바른 개인정보 보호 안전한 빅데이터 사이버공간 구축 데이터 거버넌스 위한 정보윤리 이해
국가의 빅데이터 교육 및 훈련 환경 개선, 폭 넓은 인력 확충	데이터 과학자 양성 데이터 영역 전문가 커뮤니티 확장 데이터 사용 가능한 인력 확충 공공 데이터 활용 역량 개선
정부기관, 대학, 기업, 비영리 단체와의 협력에 의한 빅데이터 혁신 생태계 지원	기관 간 빅데이터 협력 장려 빠른 대응 및 영향력 측정 가능한 정책과 정책추진 프레임워크 구축

자료: NITRD(2016); IITP(2016), “미국의 빅데이터 R&D 전략계획”

이 중 데이터 공유 및 관리를 촉진하는 정책의 일환으로 데이터 가치를 향상시키고, 데이터의 투명성과 효율성을 증가시키는 메타데이터의 모범사례를 발굴하여 국가 데이터 생태계에 지속적으로 제공하고 있다. 또한 미국의 국가과학기술위원회(NSTC) 보고서<sup>9)</sup>에서는 훈련용 데이터와 공공기관의 데이터 공개 등에서 정부의 역할을 명시하고, 공공 전산시스템에 AI을 도입하고 자연어인식 기반 인터페이스를 공공기관 사이트에 도입했다고 설명했다. 데이터 공개와 투명성, 안정성 등에 대해 정부가 주도하여 표준을 설정하고 있으나 트럼프 정부 출범 후 정부 기관 AI 관련 예산을 축소하는 등 연구개발 중심을 민간 기업으로 이동하려는 움직임을 보이고 있다.

### 3) AI

과학기술위원회(NSTC) 산하에 AI을 통한 생활, 업무, 학습, 발견, 소통방법에 대변혁을 목표로 머신러닝 및 AI 분과위원회(NITRD) 신설했다(2016. 5). 2015년 미국 정부가 기밀로 취급하지 않는 AI 기술 R&D 투자 금액은 약 11억 달러였다. 2016년 10월 전략적 연구목표를 다루고 연방정부의 투자 확대, AI R&D 인재 확충 및 유지를 위한 파이프라인구축을 위해 국가 AI R&D 전략계획 설립됐다.

[그림 2-13]과 같이 미국에서는 정부 주도의 AI 로드맵 및 R&D프로젝트를 추진해 연구기관, 학계, 기업을 전 방위적으로 포섭하여 생애주기형 장기 R&D 프로젝트를 추진 중에 있다. AI 활용 우선순위 영역을 제조업, 물류, 금융, 교통, 농업, 마케팅, 통신, 과학기술로 분류하여 영역별로 상이한 R&D 정책을 수립할 것으로 보인다. 또한 국가 AI 전략계획 이외에도 “AI 자동화 그리고 경제(2016. 12)” 등을 통해 AI 연구개발 확대와 시장 활성화 지원책 마련에 힘쓰고 있다. 또한, 첨단 데이터 분석을 위한 머신러닝 알고리즘, 지각능력 향상, 범용·확장형·인간형 AI, 유능한 로봇, 첨단 하드웨어 개발, AI 알고리즘 개선 등 차세대 AI 분야에 대한 기초·장기적 투자를 우선 실천할 계획을 선정했다.

---

9) NSTC(2016), PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE, Executive Office of the PresidentNational Science and Technology Council

[그림 2-13] AI R&D 전략계획 구조



자료: IITP(2016), “해의 ICT R&D 정책동향”을 재구성

주: (붉은색으로 표시된 하부의 줄) 교차 R&D 토대의 결합은 모든 AI 연구에서 중요합니다. (밝은 파란색 및 약간 짙은 파란색으로 표시된) 기본적인 AI R&D의 많은 영역은 (상부의 짙은 파란색으로 표시된) 다양한 사회적 응용프로그램에 영향을 주기 위하여 이와 같은 교차 토대를 기반으로 할 수 있습니다. (괄호 안의 작은 숫자들은 본 계획에서 각각의 주제를 추가 개발하는 전략의 수를 나타냅니다. 이러한 전략들의 순서가 중요성의 우선 과제를 나타내는 것은 아닙니다.)

<표 2-18>과 같이 미국의 계획은 경제적 번영, 삶의 질 개선, 국가 안보 강화 등 국가적 목표달성을 위한 7가지 액션플랜과 AI 분야에 대한 기초·장기적 투자를 제시하며 정부의 AI R&D에 대한 전략 방향을 제안하고 있다.

<표 2-18> (미국) 7대 AI R&D 전략

구분	세부 내용
차세대 AI 투자	지식 발견 위한 데이터 중심의 방법론 AI 시스템의 지각능력 향상 AI의 이론적 역량과 한계에 대한 이해 범용 AI에 대한 연구 추구 확장 가능한 AI 시스템 개발 인간형 AI 연구 촉진 유능하고 신뢰성 있는 로봇 개발 개선된 AI위한 첨단 하드웨어 개선된 하드웨어에 적합한 AI 창출

구 분	세부 내용
인간-AI 협업	인간인지 AI 위한 새로운 알고리즘 탐색 증강인간 위한 AI 기술 개발 시각화 및 인간-AI 인터페이스 기술 개발 효율적 언어처리 시스템 개발
윤리적·법적·사회적 영향 이해 및 대응	설계를 통한 공정성·투명성·책임성 개선 윤리적 AI 구축 윤리적 AI에 적합한 아키텍처 설계
안전 및 보안 보장	설명성 및 투명성 개선 신뢰성 구축 검증 및 확인 강화 공격에 대한 모안 장기적인 AI 안전과 가치 확보
공공 Dataset 환경 개발	다양한 분야의 수요 충족시킬 광범위한 데이터셋 개발 및 접근성 강화 상업적·공익적 이익에 부응하는 훈련·테스트 자원 제작 오픈소스 SW라이브러리 및 툴킷 개발
기술 측정 및 평가	폭넓은 AI 표준 개발 AI 기술 벤치마크 설정 AI 테스트베드의 유용성 확대 AI 커뮤니티에 의한 표준 및 벤치마크 채택
R&D 인력 수요	AI 전문가 확보 및 R&D 인력 수요-공급 분석

자료: IITP(2016), “해외 ICT R&D 정책동향”을 재구성

하지만 이전까지의 AI 관련 정책·전략에 비해 트럼프 정부 출범 후 정부 기관의 AI 관련 예산을 축소하는 등 연구개발 중심을 민간 기업으로 이동하려는 움직임을 보이고 있다.

## 나. 일본

### 1) 로봇

일본 정부는 ‘로봇 신전략’(2015. 1)을 발표하고, 로봇 창출력의 근본적 강화, 로봇 활용 보급, 로봇혁명 전개 및 발전을 목표로 로봇 연구개발 체계를 세웠다. 로봇 연구개발을 일본 사회의 저출산·고령화 문제의 해결책으로 보고 AI 기술 산업 정책을 추진하고 있다. 기존의 로봇 산업에 AI 기술을 접목하는 정책을 활용할 계획이며, 2020년까지 제조, 서비스, 간호·의료, 인프라·재난대응·건설, 농림수산업·식품산업 등 5개 분야를 집중적으로 지원할 예정이다. 또한 정책적 지원 외에 민간투자를 확대하여 1,000억엔 규모의 로봇 프

로젝트를 추진하고 있다. 이러한 일본의 로봇 연구개발 투자를 통해 시장이 성장하고 있으며, 일본 내 상용화가 가시화 된 AI 적용 제품은 휴머노이드, 자율주행차, 산업용 로봇으로 나타났다.

2) 빅데이터

일본은 AI 기술 수준에 따라 3단계 로드맵을 제시하고, 사회 모든 영역이 복합적으로 연결 및 융합되는 AI 에코시스템을 구축할 예정이다. 그리고 데이터 공유 및 활용 플랫폼을 마련해 2020년까지 센서를 통해 수집한 현장 데이터를 가지고 제조현장의 강점을 가지고 있는 독일과의 국제 표준화를 추진할 계획이다.

일본 정부에서 추진한 ‘미래투자전략 2017’을 통해 건강수명연장, 이동혁명실현, 공급망 첨단화, 쾌적한 인프라 도시 만들기, 핀테크를 5대 신성장 전략으로 육성하겠다는 추진계획을 발표하면서 5가지 신성장 전략 분야에 데이터를 활용한 기반구축이라는 일관된 정책을 내세웠다(〈표 2-19〉 참조).

〈표 2-19〉 미래투자전략 2017 실현을 위한 과제

시점	주요 내용
기초연구강화	데이터 통합, 빅데이터 클러스터 리소스 스케줄링, 분산 파일시스템, 스트리밍 컴퓨팅 엔진 기술 AI 컴퓨팅, 인지 컴퓨팅, 블록체인, 가상현실 분야 기술 혁신 산업 애플리케이션, 이해/예측 및 의사 결정 지원 서비스
분산컴퓨팅	분산컴퓨팅 기반의 산업의 데이터 분석 솔루션과 서비스의 형성 차세대 비즈니스 지능, 데이터 마이닝, 데이터 시각화 시맨틱 검색 및 기타 소프트웨어 제품의 개발을 촉진 자원 관리 플랫폼 제품을 의미 이해를 바탕으로 하는 SW
서비스 기반 빅데이터 활용	고속 데이터 처리를 위한 컴퓨팅 자원 활용 제고 (비용절감) 클라우드 기반의 빅데이터 및 지능형 서비스 촉진 공공 클라우드 서비스 강화

자료: 한국정보화진흥원 『빅데이터 동향과 이슈』(2017. 2)

이 정책에서는 4차 산업혁명의 핵심기술을 빅데이터, AI 등으로 규정하고, 부문별(금융, 의학, 생산 등) 기술 및 데이터와 결합한 새로운 제품과 서비스가 만들어지는 현상으로 데이터와 이를 활용할 수 있게 하는 비즈니스 모델의 역할도 강조했다.

[그림 2-14] 공통 기반기술과 데이터 결합을 통한 제품·서비스 산출



자료: 과학기술정책연구원 『일본의 제4차 산업혁명 대응정책과 시사점』 (2017. 5)

일본정부는 데이터를 현실 데이터<sup>10)</sup>와 가상 데이터<sup>11)</sup>로 구분하여 대응 정책을 수립하고 있는데 일본이 경쟁력을 가질 수 있는 분야는 현실 데이터라고 판단해 기술뿐만 아니라 데이터 활용<sup>12)</sup>을 위한 전체 과정을 통합 제공하는 비즈니스 모델을 강조했다. 가상 데이터는 해외 IT 기업이 플랫폼을 지배(구글, 아마존, 애플 등)한다고 하며 현실 데이터는 건강 정보, 주행 데이터, 공장설비 가동 데이터 등 개인과 기업이 생산한 데이터로 적절히 대응한다면 일본에서 플랫폼을 획득할 가능성이 존재하고 있다고 설명했다.

### 3) AI

일본의 AI 관련 R&D 전략은 AI의 발전단계에 맞춰 세부적인 계획을 수립했다. [그림 2-15]와 같이 일본의 전략은 총무성, 문부과학성, 경제산업성 등 정부 부처가 연계하여 AI R&D를 위한 사업을 집중 지원하는 시스템이다.

10) SNS 등 온라인 공간에서 발생하는 데이터

11) 개인과 기업이 생산한 데이터

12) 데이터 활용 과정: 취득 → 교환·공유 → 실용화(빅데이터화) → 분석 → 활용·산업화

[그림 2-15] 일본 AI 연구기관 및 센터

주관부처	연구기관	센터명
문부과학성	이화학연구소	AIP센터
	과학기술진흥기구	인공지능프로젝트
경제산업성	산업기술종합연구소	인공지능연구센터
	신에너지 사업기술총합개발기구	인공지능프로젝트
총무성	정보통신연구기구	뇌정보통신융합연구센터

자료: IITP(2017), “일본의 AI(AI) 정책 동향과 실행전략”

〈표 2-20〉과 같이 일본은 두 가지 주요 국가 정책을 통해 AI 기술 진흥을 추진할 계획이다. 일본재흥전략 및 후속조치를 통해 자국 내 신산업 진흥 정책을 추진할 예정이며, 과학기술기본계획은 ‘초스마트사회’를 강조하여 사회의 전반적인 기술 역량 강화에 집중하고 있다.

〈표 2-20〉 (일본) 과학기술 국가의 실현을 위한 2개의 주요 정책 및 계획

구 분	방안
일본재흥전략 및 후속조치	<ul style="list-style-type: none"> <li>- 미래투자회의, 규제개혁추진회의, 미래 투자를 위한 민관대화를 주최하여 4차 산업혁명을 추진하는 민관 협력 장치를 마련</li> <li>- 2016년 재흥전략은 사회전반의 혁신성 향상, 지속적인 성장을 통한 GDP 600조엔 달성을 목표로 AI·사물인터넷·빅데이터 통한 신산업 창출을 강조</li> </ul>
과학기술기본계획	<ul style="list-style-type: none"> <li>- 지속적인 성장, 국민 안전 확보, 세계 발전에 기여, 지식 자산의 지속적 창출하는 국가 실현을 위해 ①미래 산업 창출과 사회 변혁 ②경제·사회적 과제에 대한 대응 ③기초적인 힘의 강화 ④인재·지식·재원의 선순환 시스템 구축 제시</li> <li>도전적인 R&amp;D와 새로운 가치가 연결되는 일본 사회 미래상을 ‘초스마트사회(society 5.0)’으로 제시(15. 12)</li> </ul>
재흥전략 및 과학기술기본계획 이후 AI 지원 정책	<ul style="list-style-type: none"> <li>- 총무성, 문부과학성, 경제산업성이 연계하여 AI 집중 지원 위한 컨트롤타워 ‘AI기술전략회의’(16. 4)설치</li> <li>- AI기술전략회의에서 액션플랜으로서 ‘AI 산업화 로드맵’ 발표(17. 1)</li> </ul>

자료: IITP(2017), “일본의 AI 정책 동향과 실행전략”을 재구성

정부는 세 부처가 연계된 컨트롤타워 ‘AI기술전략회의’(2016. 4)를 설치하였으며, 2017년 ‘AI 산업화 로드맵’을 발표하여 AI 발전을 위한 장·단기적 계획을 수립했다. 이를 통해 사회 모든 영역이 복합적으로 연결·융합되는 AI 에코시스템을 구축할 예정이다. <표 2-21>과 같이 일본은 AI 산업화 로드맵의 세부 내용을 3단계로 추진할 계획이다.

<표 2-21> (일본) AI 산업화 로드맵 세부 내용

구 분	1단계	2단계	3단계
생산성	AI가 새로운 서비스·제품 개발을 지원	분야를 뛰어넘는 다양한 서비스·제품 창출	인간의 창조력을 증폭시키고 기존개념을 초월한 서비스·제품이 속속 탄생하는 사회 구축
	스마트 공장	제조·유통·서비스 융합이 진전	산업 간 근간이 사라지고, 궁극적인 에코시스템(낭비제로 사회)을 구축
건강	AI를 활용한 건강관리	센서 활용을 통한 예방의료 보급	모든 사람에게 센서를 부착해, 일상적으로 예방의료를 실현함으로써 건강장수산업대국 달성
	건강·의료데이터 정비 /AI가 의사를 보조	모든 수술을 AI가 보조. 재택 검진·진료 보급	의사의 감시 하에 고도의료를 간편하게 제공하는 의료 서비스 실현
	센서 등을 활용한 간병 시설	인간이 관리하는 간병로봇이 간병서비스 제공	일상생활의 다양한 장면에서 범용 로봇이 서비스 제공/신체기능을 대체·보조하는 로봇으로 기능적인 신체를 디자인
공간 이동	장애물 인식 및 위험감지 등을 통해 운전 보조	완전 자동운전 실현을 통해 이동시간·공간을 업무나 취미에 활용	사이버 공간과 피지컬(physical)공간이 융합해 이동 자체가 부가가치화되고, 이동기기의 자동 버전업그레이드 등 주변산업이 발전
	텔레위크 진전	피지컬(physical)공간에 가까운 사이버 공간 환경 실현	

자료: IITP (2017), “일본의 인공지능(AI) 정책 동향과 실행전략”

일본 정부의 AI 진흥 정책은 기술을 활용할 수 있는 인재 육성 위한 노력으로, AI 연구 개발과 사회 이관이란 관점에서 본 인재육성 논의를 단기(즉시전력 육성), 중기(학교교육·직업훈련 등), 장기(학문으로서의 자세) 3단계로 나눠 계획되었다. 다음 <표 2-22>는 일본 정부의 기간별 AI 기술 육성 방안 내용을 정리한 표이다.

<표 2-22> 일본의 AI 기간별 육성 방안

구 분	육성 방안
단기	정부·연구기관 등이 지금까지 진행한 대처를 더욱 충실하게 수행 • 산관학 연계 가이드라인(2025년까지 기업에서 대학·국립연구개발법인에 대한 투자 3배 확대 실현) • NICT를 통한 연구자 수용, 인적 교류 • AI챌린지 콘테스트 • NEDO 특별강좌, TCP • 데이터 관련 인재육성 프로그램 • 산업기술종합연구소(産総研) AI기술 컨소시엄 • 대학 등의 수리·데이터 사이언스 교육 강화 • JST펀딩을 통한 청년인재 육성 • 성장분야를 뒷받침할 정보기술인재 육성거점 형성(enPiT) 등
단기(신규)	즉시전력 육성을 위한 교육 프로그램 구상·실시(신규) • AI와 관련된 사회인을 대상으로 업무상 필요한 분야의 최첨단 지식 및 AI의 체계적인 지식 습득, Real Common Data 강습을 통한 가치창조력 향상을 지향
단~중기	대학과 산업계의 공동연구·인재육성 추진 • 대학과 산업계의 공동연구, OJT를 통한 인재육성 등 개별적인 노력을 '점(点)'에서 '면(面)'으로 전개해 나가는 장치 마련
장기	별도 내용 없음

자료: IITP (2017), “일본의 인공지능(AI) 정책 동향과 실행전략”

#### 다. 중국

##### 1) 빅데이터

중국에서는 정부의 전폭적 지원과 IT 주요기업들의 투자를 통해 AI 기술 개발을 견인하고 있으며, 인터넷 플러스<sup>13)</sup> 정책 발표 이후 지속적으로 AI 기술 개발과 관련된 정책을 추

13) 인터넷 플러스: 인터넷 플랫폼, 정보통신 기술을 활용해 인터넷을 전 산업과 융합시켜 새로운 경제발전 생태계를 창조하는 전략을 의미

가하며 산업 성장을 촉진하는데 주력하고 있다. 2015년에 중국은 국무원이 제조·금융 등 일반산업과 인터넷의 융합을 목적으로 발표한 인터넷 플러스 전략<sup>14)</sup>에 별도로 AI을 중점 분야로 추가했다.

올해 ‘차세대 AI 발전 계획<sup>15)</sup>’을 발표하며 2030년까지 단계별로 AI 기술개발·상업화를 추진해 세계 AI 분야의 선도를 선언한 바 있다.

[그림 2-16]과 같이 중국은 빅데이터 산업 발전 계획(2016-2020)을 통해 빅데이터 산업 성장을 위한 7대 핵심 과제를 발표하였다. 이를 통해 2020년 기술 응용SW의 활성화를 통해 빅데이터 관련 제품 및 서비스 시장의 규모가 1조 위안(168조원)을 상회하고 연평균 성장률은 30%를 유지할 것으로 전망했다.

[그림 2-16] 중국의 빅데이터 산업 성장의 7대 핵심 과제



자료: 한국정보화진흥원 『빅데이터 동향과 이슈』(2017. 2)

빅데이터 분야의 핵심 SW 개발을 지원하고, 특히 주요 제품을 브랜드화하여 보다 높은 수준의 지원을 추진할 예정으로서 산업별 특화 제품과 데이터 유통 및 분석 서비스 기반 회사를 선정해 홍보 및 마케팅을 강력하게 지원하는 정책을 <표 2-23>과 같이 추진하고 있다.

14) [http://www.gov.cn/xinwen/2016-05/23/content\\_5075944.htm](http://www.gov.cn/xinwen/2016-05/23/content_5075944.htm)

15) 차세대 AI 발전계획: 2030년까지 AI 핵심산업을 1,500억위안 (약 25조원)

〈표 2-23〉 중국 빅데이터 산업발전계획 실행원칙

원칙	주요 내용
구동발전	기업을 주체로 빅데이터의 핵심기술을 획득 및 발전하고 관련 상품의 연구 개발을 추진하여 빅데이터 기술, 응용과 상업모델의 구동발전을 강화하는 것을 원칙으로 한다.
응용의 주도화	중국의 방대한 시장규모와 수요(Needs)의 우세를 발휘하여 국가전략, 인민의 수요, 시장수요를 원동력으로 빅데이터 기술상품의 연구개발과 각 업종, 분야에서의 응용을 추진하고, 또한 업종간의 협력과 융합을 실현하여 협동조합형 산업발전을 구축해 나아가는 것을 원칙으로 한다.
개방 및 공유	글로벌 빅데이터 선진기술, 인재와 자금 등 요소자원을 결합하여 자아발전과 외부협력이 동시에 발전하는 개방식 빅데이터 산업발전의 길을 가야 한다. 데이터 개방과 공유의 개념을 확실히 설립하고 관련된 제도를 보완하여 데이터 자원의 개방 공유와 정보의 유통을 추진하는 것을 원칙으로 한다.
융합 및 조율	기업은 빅데이터 산업창조에서의 주도역할을 발휘하고, 정부는 정책적 지지와 가이드 역할을 강화하여, 양호한 정책규범 환경의 구축을 통해 정부, 산업, 학술, 연구, 응용 등 다분야의 융합과 조율체계를 만들고, 나아가서 중앙정부, 각 부서, 지방정부의 빅데이터 발전정책의 연결을 통한 협력발전을 추구하는 것을 원칙으로 한다.
안전규범	안전은 발전의 전제조건이고 발전은 안전의 보장이다. 발전과 안전을 동시에 중요시하고, 정보안전기술의 보장 능력을 강화하고 안전방어체계를 구축하며 정보안전과 개인정보보호를 보장하여야 한다. 또한 업종의 자율성 강화, 규제 완화 등을 통해 데이터 자원의 질서 있는 유통과 규범적인 사용이 이루어지는 것을 원칙으로 한다.

자료: 중국산업규획망 『중국 빅데이터 산업 발전 현황』

## 2) AI

2017년 양회에서 AI가 주요 키워드로 부상하면서 다양한 산업분야에 AI 기술을 활용하고 관련 데이터를 공유하는 등 AI 관련 기업과 타 산업 간의 협력을 통해 경쟁력을 강화해야 한다고 주장했다. 이에 대해 바이두 CEO 리엔홍은 중국 양회에서 '차이나 브레인 프로젝트'<sup>16)</sup>를 제안하여 중국이 AI 분야에서 최고가 되겠다는 계획을 수립하겠다고 전한 바 있다.

16) 민간과 군사용으로 빅데이터(대용량 데이터) 분석 도구, 자율주행차, 스마트 의료 진단, 스마트 드론(무인비행기) 개발

또한, 중국은 AI 기술 측면에서 미국과 양강 구도를 구축하고 있으며, 10년 안에 미국을 제치고 전 세계 AI 기술을 선도할 것으로 전망되고 있다. 또한 미국에 이어 세계 2위의 AI 특허 강국이지만, 특허 등록 수의 성장률은 미국의 7배에 달한 것으로 나타났다. 하지만 중국의 인권의식이 성장하고 정보 통제가 강화되면서 AI 기술의 근간이 될 대용량 데이터의 수집과 활용이 어떻게 변화할지 지켜볼 필요가 있다. 예를 들어 중국의 안면인식 기술은 무단촬영 등 공공질서 위반자, 범죄자 색출 등에 이용되고 있다. 이에 대한 인권침해 이슈가 제기되고 있으나 중국 정부의 인터넷 통제 강화로 중국 외부 데이터의 축적이 제한되면서 AI 기술의 글로벌 적용도가 떨어지는 결과를 초래할 가능성이 있다. 이에 대해 전문기관 골드만삭스<sup>17)</sup>는 중국이 인적자원, 인프라, 산업정책에 힘입어 향후 미국의 AI 기술을 추월할 것으로 전망했다.

#### 4. 해외 AI 기술 동향과 주요 사례

다양한 국가의 정책·전략과 그에 의한 투자·연구개발 노력을 통해 국내외 기술별·산업별 AI 기술의 적용 사례가 확대되고 있다. 또한 정부의 지원과 전략에 따른 민간 기업의 기술 개발 및 투자 또한 기술 견인에 중요한 요인이 되었다. 구글, 페이스북, IBM, 마이크로소프트 등 글로벌 ICT 기업의 투자 및 연구개발은 학습 및 추론, 상황이해, 언어이해 등 각 AI 핵심 기술을 상품 및 서비스에 탑재하기 위해 개발·발전되고 있다. 또한 글로벌 기업은 데이터를 활용한 AI 산업 생태계 장악 및 플랫폼 경쟁을 위해 원천기술을 보유한 AI 기업에 대한 공격적인 M&A를 진행을 하고 있다. 이에 독자적 플랫폼을 구축하여 AI 생태계 장악을 위한 노력을 가속화<sup>18)</sup>하고 있다.

---

17) Lant, K.(2017), Goldman Sachs Report: China's AI Sector is Catching Up to the U.S, Futurism,

<https://futurism.com/goldman-sachs-report-chinas-ai-sector-is-catching-up-to-the-u-s/>

18) 아마존 'Alexa', 구글 'Google Assistant', 애플 'Siri', 마이크로소프트 'Cortana', 바이두 'Duer', 삼성 'Bixby' 등 글로벌 기업들은 AI 생태계 장악을 위해 독자적인 SW 개발

〈표 2-24〉 AI 5개 핵심기술 분류에 따른 해외 기술 동향

5대 핵심기술	기술 동향 내용
학습 및 추론	글로벌 기업(구글, 페이스북, IBM 등)을 중심으로 영상, 음성, 텍스트 인식 분야의 학습 및 추론 기반 AI 플랫폼이 개발 중
상황이해	아마존(Alexa), 오토메이티드 인사이트(위드스미스)등에서 상황이해 기반 AI SW 개발 중
언어이해	IBM의 Watson 딥 QA 기술을 헬스케어 및 파이낸스 분야에 적용하는 프로젝트를 진행 중
	후지쯔는 도쿄대 및 일본국립정보학연구소와 공동으로 대학입시 합격을 목표로 제작된 '토다이 로봇 프로젝트'를 추진 중
시각이해	미국 국방 고등 연구 기획청(DAPPA)는 군사용 비디오에서 원거리 행동을 인식하는 '원격 행동분석 기술'을 개발 중
	구글, 마이크로소프트는 딥러닝 기술을 기반으로 영상에서 사물을 구별하는 '고성인인식기술' 개발 중
인지컴퓨팅	IBM, 애플, 마이크로소프트, Saffron Tech, 말루바, 익스펙트랩스 등을 중심으로 한 기업들이 대규모 데이터 분석 및 처리 위한 인지컴퓨팅 기술을 개발 중

자료: 한국지식재산연구원(2016), "AI(AI) 기술 및 정책 동향"을 재구성

공급 측면에서 알파고는 향후 알고리즘을 활용해 다양한 서비스에 적용하도록 범용으로 개발하는 등 각 기업들이 개발한 플랫폼은 전 산업에 공급되어 다양한 제품·서비스 개발하고 있다. 아마존의 경우 Alexa를 기반 음성비서 스피커 Echo, 삼성 로봇청소기, GE LED, 레노버 뉴글래스 등 다양한 제품 및 서비스에 적용하고 있다.

〈표 2-25〉와 같이 수요 측면에서 향후 AI 시스템이 인지컴퓨팅을 통해 영상 인식, 금융 거래, 헬스케어 데이터 관리 등 거의 모든 산업 분야에 적용 될 것<sup>19)</sup>으로 전망했다.

19) 현재 제조, 서비스, 유통, 쇼핑, 건설, 항공, 자동차, 회계, 번역, 변호, 금융 등 다양한 산업분야에 적용을 위한 기술개발 추진

〈표 2-25〉 산업별 해외 AI 기술 동향

산업 영역	기술 동향 내용
제조업	데이터 분석 능력과 시각지능 통한 공정 효율화 및 휴머노이드형 노동 로봇 활용한 스마트 공장 추세가 가속화 됨
	실시간 생산 모니터링, 이상 상태 감시 및 의사결정을 위한 스케줄링, 예지보전, 불량 예방, 자원관리, 안전관리가 가능하도록 기술 발전
	독일과 미국이 스마트 공장 표준화 경쟁 중에 있으며, 중국은 스마트 공장 구현을 통해 인구 고령화, 노동인구의 공장 근무 기피 현상에 대비
의료 및 바이오 기술	웨어러블 센서를 통한 건강정보 데이터와 이미지 인식 기술을 활용한 진단 데이터로 최적화된 스마트 헬스 솔루션 제공이 가능
	맞춤형 진단 및 건강 가이드가 가능하고, 신약 개발, 유전체/단백체 분석이 가능
국방 기술	무인기와 AI를 결합하여 조종사 육성비용을 절감하고, 전투력 향상과 조종사 보호, 군수장비의 자동조작이 가능해짐
	안보 위협요인 감시 및 검출, 조기 경보 및 위게임 시뮬레이션, 전투 로봇 등이 가능
교육 서비스	학습패턴과 교육 방향성 판단하고 적절한 학습방법 제시 가능할 것으로 기대됨
자동차 기술	자율주행차의 주변 인지 기능을 향상시켰으며 각종 센서를 통한 데이터를 정확하게 처리하여 안전한 경로를 판단하는 역할 수행할 것
광고마케팅, 금융, 법무, 회계 등 기술	빅데이터 분석에 기반한 맞춤형 마케팅 가능하며 일부는 AI 기술로 대체될 것

자료: 한국지식재산연구원(2016), “AI(AI) 기술 및 정책 동향”을 재구성

가. 해외 기업의 AI 기술 활용 현황

구글, IBM, 마이크로소프트, 페이스북 등 글로벌 ICT 기업은 거대 투자를 통해 딥러닝 기술을 개발하고, 이를 활용하여 상품·서비스에 적용하고 있다.

1) 구글

구글은 자율주행차(구글 카), 음성인식(구글 나우), 이미지 인식(구글 포토) 등의 사업을 위해 유망 AI 관련 신생기업을 인수하거나 기술 개발에 힘써왔다. 동시에 딥러닝 패키지 ‘텐서플로우(TensorFlow)’를 공개한 후에 소프트웨어의 개발과 더불어 하드웨어 개발에도

힘쓰고 있다. 2017년 5월, 기계학습 및 텐서플로우의 구동을 뒷받침하기 위한 전용 프로세서 '클라우드TPU'를 공개했다. 기존 TPU는 학습 연산을 구동하지 못했지만, 공개된 클라우드TPU는 학습 연산에도 활용이 가능하게 됐다.

## 2) IBM

IBM은 1990년대 슈퍼컴퓨터 '딥 블루' 이후 자연어 소통 슈퍼컴퓨터인 '왓슨(Watson)'에 약 1조 2,000억원을 투자해 개발했다. 왓슨은 특히 의료 분야에서 주목 받고 있는데, '왓슨 포 온콜로지(Watson for Oncology)'가 왓슨의 한국 사업권을 보유한 SK C&C를 통해 주요 대형 병원에 도입됐다. 이 기술은 의학 논문 및 자료를 분석하여 환자에게 적합하다고 판단되는 치료 방안을 제시하도록 고안되었으며, 암 진료 분야에서 90% 이상 일치하는 진단을 낸다고 알려졌다. 또한 금융 상담 업무와 관련하여 왓슨 기술의 적용을 테스트 하는 등 의료 분야 외에도 검색과 평가, 문서 변환, 번역, 이미지 인식, 성향 분석 등의 기술을 적용하려는 시도가 진행되고 있다. 더불어 효율적인 인지컴퓨팅 기술을 실현할 수 있도록 클라우드 서비스 'Bluemix'를 출시하여 지난 1년간 158억 달러에 이르는 관련 매출을 기록하였다.

## 3) 마이크로소프트

마이크로소프트는 2017년 초 인공신경망(Neural network) 학습 소프트웨어 '코그니티브 툴킷(CNTK)'의 새로운 버전인 '마이크로소프트 코그니티브툴킷(Microsoft Cognitive Toolkit)'을 오픈소스로 공개했다. 베타 버전으로 출시되었던 코그니티브툴킷은 사용자가 인공신경망을 정의하고 입력 데이터를 통해 학습시키는 소프트웨어이다. 이 기술을 통해 중국 스타트업 '에어닥(Airdoc)'과 협력하여 당뇨병성 망막증을 진단하는 등 AI 기술의 적용 범위가 확대되고 있다. 또한 클라우드 서비스 애저(Azure)를 제공하여 AI 기술 구현을 지원하고 있다.

## 4) 페이스북

페이스북은 2015년부터 오픈소스 기반 딥러닝 모듈 '토치(Torch)', AI 서버 '빅서(Bir Sur)' 등을 공개하며 딥러닝 기술 개발을 추진했다. 딥러닝 기술 기반의 얼굴 인식, 번역 등을 페이스북 페이지에 적용해 왔으며, 2017년에는 챗봇(chatbot) 기능의 상용화가 추진되었다. 미국 마스터카드와 협력하여 페이스북 메신저 챗봇 기능이 결합된 전자지갑 '마스터패스'를 공개하여, 메신저 창을 통한 제품 주문 시스템을 제공하고 있다. 또한 2017년 5월, 페이

스북은 기존 AI 번역 기술보다 9배 빠르고 정확한 번역 시스템을 오픈소스로 공개했다. 이번 공개된 번역 시스템은 병렬 처리가 가능한 CNN(Convolution Neural Network)을 이용하여 기존 순환신경망(RNN)을 통해 단어를 하나씩 번역하던 시스템을 벗어나게 됐다.

#### 나. 해외 제조업 기업의 AI 기술 융합 사례

위에서 살펴보았듯이, 글로벌 IT관련 기업 외에 제조업을 기반으로 성장하였던 기업이 ICT 기술, 특히 AI 기술을 융합하여 새로운 비즈니스 모델을 창출한 경우도 있으며, 예시로 General Electric(이하 GE)와 Siemens(이하 지멘스)의 사례를 자세히 살펴보고자 한다.

##### 1) General Electric, GE

GE는 1897년 Thomas Edison이 공동 창립하였다. 120년이 지난 지금, GE는 다우 존스 산업 평균 지수(Dow Jones Industrial Average)에서 가장 오래 지속된 기업으로 획기적인 혁신을 추진하고 있다. 현재 GE는 기존 산업용(소비자) 제품, 금융 서비스 기업에서 산업 인터넷(Industrial Internet)에 중점을 두는 소프트웨어 회사로 탈바꿈하여, 2016년 소프트웨어 매출 70억 달러를 기록하는 디지털 기업으로 변신했다.

GE는 대부분 산업용 자산(설비, 기계, 기기) 판매와 정비 서비스에서 수익을 올려왔다. 하지만 2000년 후반부터 주요 ICT 기업인 IBM, SAP를 비롯한 데이터 분석 전문기업들이 산업용 자산에서 발생하는 데이터를 분석하여 새로운 솔루션을 제공함에 따라 GE와 같은 산업용 자산을 생산·판매하는 업체들에게 위기감이 도래하였다. 이러한 변화에 대응하여 GE는 빅데이터를 기반으로 AI(기계학습)를 활용하는 디지털 기업으로 변환을 추진하였다.

GE는 산업용 자산(설비, 기계, 기기)에서 수집되는 빅데이터를 분석할 수 있도록 대표적인 AI 소프트웨어 제품인 'Predix'를 개발하여 제공하고 있다. 소프트웨어 및 분석 기업으로서 GE는 산업용 자산을 인터넷으로 연결하고, AI을 이러한 산업용 자산에 내장시키고 있다. 이를 통해 해당 기계 및 장치에서 생성된 데이터를 수집·분석하여, 산업용 자산을 보다 효율적으로 운영하고 최적화할 수 있는 통찰력을 제공하고 있다.

이와 같이 GE는 빅데이터와 IoT를 활용할 수 있도록 지원하는 접근 방식으로서 AI의 중요성을 강조하고 있다. AI은 구체적으로 “디지털 쌍둥이(digital twin)”개념을 구체화하는데 기반이 되고 있다. 이러한 디지털 쌍둥이는 산업용 자산의 디지털 복제본(digital replica) 또는 데이터 기반 표현(data-based representation)이라고 할 수 있다.

예를 들어 i) 제트 엔진, 가스 터빈 또는 풍차터빈의 센서가 기계의 속성 (열, 진동, 소음 등)에 대한 데이터를 모으면, ii) 해당 데이터가 클라우드를 통해 수집되고, iii) 기계를 가상으로 복제하여 성능을 분석할 수 있는 모델인 『디지털 쌍둥이』를 구성한다. iv) 이러한 디지털 쌍둥이 모델을 사용하면 산업용 자산의 오류를 진단하고, 유지·관리를 수행할 수 있다.

최종적으로는 AI 기반의 디지털 쌍둥이 모델을 통해 해당 시스템의 예기치 않은 중단 시간을 줄이거나 제거할 수 있다. 더 나아가 디지털 쌍둥이 개념은 플랜트 또는 선단 수준으로 산업용 자산의 집합으로 확장될 수 있다.

특히 다양한 디지털 변수로 구성되는 디지털 쌍둥이 모델로 지속적으로 데이터가 유입되면 산업용 자산에 요구되는 유지·관리를 가장 잘 예측하는 필요 변수와 모델이 시간이 지남에 따라 변경될 수 있다. AI은 새로운 데이터로부터 학습하고, 시간이 지남에 따라 예측 모델을 수정하는 것이 가능하기 때문에 이러한 요구 사항을 해결하기 위한 최적의 기술적인 접근 방식으로 AI을 활용할 수 있다. 즉 AI을 통해 기계 성능의 이상, 특징 및 추세를 파악하고 동작 패턴에 대한 이해를 강화할 수 있다. 또한 AI을 적용하여 특정 시스템 내 효율성을 확인하고, 이를 다른 시스템의 모범 사례로도 사용할 수 있다. GE는 현재 약 750,000쌍의 디지털 쌍둥이를 보유하고 있으며 이를 지속적으로 추가하고 있다.

## 2) Siemens

독일계 제조 기업 지멘스는 1847년에 설립되었다. 독일 정부가 제창한 혁신 전략인 'Industry 4.0'에 따라 ICT를 제조업과 융·복합하여, 전 세계적으로 200여 개 국가에서 사업을 진행하고 있는 거대 복합 기업으로 성장했다. 2014년 기준으로 총 매출액 7,739억 유로로 집계되었으며, 에너지, 공정, 헬스케어, 금융 등 다양한 분야에서 사업을 전개 중이다. 디지털 기술을 융합한 제조업 역량 강화를 위해 지멘스는 약 30년 전부터 품질, 프로세스, 효율을 위해 인공지능경망을 포함한 AI, 기계학습 등의 기술 개발에 힘써 왔다. 1990년대부터 인공지능경망 기술을 제강 및 제지 공장에 적용하기 시작하여 현재는 200명가량의 데이터 분석 및 인공지능경망 전문가를 고용할 정도로 AI에 대한 자원과 역량을 보유하고 있다.

지멘스의 이러한 성장은 동작, 인식, 제어, 정보, 서비스, 예상고객·가치제공 파트 등 다양한 영역에서 데이터를 수집·분석하는 것에 중점을 두고 있기 때문에 가능했다. 제조업 혁신을 위해 소프트웨어를 내재화하는 조직을 설립하여 자체적으로 ICT 분야를 강화했다. 이를 위해 총 연구개발 비용의 1/2을 소프트웨어 부문에 투자하고, 2010년 이후부터 소프트

웨어 전문기업을 활발히 인수하여 자체적인 소프트웨어 역량 확대에 집중하고 있다. UGS, 비스타지, TESIS PLM웨어 등과 같은 ICT 기업들을 인수하여 AI 기술을 개발하고, 이를 통해 제조 공정 데이터를 분석하여 공정 개선이 가능한 서비스를 제공할 수 있게 되었다.

지멘스의 디지털 역량은 AI 기술을 위한 소프트웨어에 IoT, 클라우드, 데이터라는 요소가 뒷받침되고 있다. 이러한 측면에서 지멘스는 'self-optimizing 가스 터빈' 등 딥러닝 및 강화 학습을 이용한 제조 공정의 최적화 관련 특허를 50개 이상 보유하게 되었다. 예를 들어 i) '스마트 박스(smart box)'에 내재된 센서를 통해 기계의 구동 데이터를 생성하고, ii) 커뮤니케이션 인터페이스를 거쳐 데이터를 전환 및 전송하며, iii) AI 시스템에 의해 이러한 데이터를 분석하는 체계로, 기계의 상태와 오작동을 예측할 수 있다.

지멘스는 소프트웨어, 빅데이터, AI 등의 ICT·지능정보기술을 제공하는 전문기업은 아니지만, 기존 기계·설비 제조에서 더 나아가 디지털 서비스를 제공하는 영역으로 확장했다. 제조업 분야의 상품 생명주기 관리 소프트웨어(PLM, Product Lifecycle Management Software), 디지털 쌍둥이 소프트웨어(Digital Twin Software), MindSphere IoT 플랫폼 등을 통하여 서비스를 제공하고 있다.

## 5. 사례 분석을 통한 시사점

1990년대 제조업의 가치원천이 과거 하드웨어 중심에서 소프트웨어 중심으로 변화됨에 따라 소프트웨어 개발·개선으로 품질이 향상되고 고객에 대한 새로운 가치를 제공할 수 있게 되었다.

하드웨어에 대한 직접적인 개선보다는 하드웨어를 관리하는 소프트웨어의 개선으로 성능향상이나 새로운 고객가치를 창출하였다. 문제해결을 위해 소프트웨어를 개발·개선하는 단계로 이해할 수 있다.

최근에는 제조업의 가치원천이 데이터 기반의 AI 중심으로 변화하고 있다. 데이터가 축적되고 이를 활용하기 시작하면서 데이터 기반의 혁신이 강화되었다. 즉 데이터를 수집·가공·분석하는 능력이 성과를 높이는데 크게 기여하고 있다. 최근에는 더 나아가 AI이 현업의 데이터 수집과 분석을 대체하고, 데이터를 자동으로 학습하고 있다. AI이 데이터에 숨어있는 로직을 찾아서 필요한 코드를 만들어내는 단계로 진화하고 있는 것이다.

가치의 원천이 AI 중심으로 변화함에 따라, AI을 활용한 경쟁력 확보에 지속적인 노력이 요구된다. AI의 기본적인 기법·원리가 오픈소스로 공개되어 광범위하게 활용될 수 있으나, 구체적인 응용 및 문제해결에 있어서는 해당 산업의 세부 영역에 대한 지식(domain knowledge)이 필요하다. 즉 특정 업무(Task) 대한 사전적인 지식(프로세스, 제품 및 서비스)과 축적된 데이터가 요구된다. 이와 같이 AI의 활용과 적용은 보유 데이터에 기반을 두기 때문에, 자신들의 요구사항(아이디어)를 명확하게 정의하고 연구된 기존의 AI 모델을 변형(Transfer Learning)하는 것이 요구된다.

제조업 분야의 기업이 AI을 성공적으로 활용·적용하면서 발생하는 결과는 프로세스 및 제품의 혁신에서부터 비즈니스 모델의 변화로 나타난다. 프로세스 및 제품 혁신을 위해 생산설비의 상태, 작업 중인 부품의 품질 정보, 작업자 정보, 생산 환경 정보 등의 데이터를 AI으로 수집·분석·예측하여 생산성과 품질을 높이는 동시에 원가도 절감할 수 있다. 더 나아가서는 제품에 AI 기술을 적용할 수 있다. 제품의 사용 주체, 활용 상태, 기능 및 작동 상태, 사용량 등에 대한 데이터를 실시간으로 수집·분석하여 고객에게 최적화된 서비스를 제공할 수 있다.

또한 AI이 새로운 가치를 학습하여 창출하는 경우, 비즈니스 모델의 변화를 가져올 수 있다. 예를 들어 AI을 통해 제품 관련 최적 활용 서비스 및 유지·관리 서비스를 제공하는 경우, 제품 자체를 제조·판매하는 모델에서 벗어나 제품을 사용하면서 얻는 효과에 대하여 대가를 지불하는 모델로 변화할 수 있다. 더 나아가 제조영역의 도메인 지식과 데이터 기반으로 AI 서비스(컨설팅 및 소프트웨어)를 제공할 수 있다.

제조기업이 AI이라는 자원을 활용·적용하여 기업의 경쟁우위를 확보하는 방법은 크게 두 가지로 나누어 볼 수 있다. 첫 번째로, AI은 기업 운영의 효율성 향상을 위한 가장 강력한 혁신 도구로, 실시간 데이터 수집·분석을 통해 거의 모든 기업의 가치사슬 활동을 개선한다. AI 기술을 통한 제조(생산) 현장의 개선(자동화, 정보화, 자율화)으로 원가(인력)절감을 가능하게 해준다. 하지만 대부분의 기업들이 AI을 자원으로 활용할 수 있어 이러한 운영의 효율성은 재빠르게 모방되어 확산됨에 따라 지속적인 경쟁우위로 유지하기가 어렵다.

두 번째로, AI은 기업 운영의 효율성 향상과 더불어 새로운 제품(서비스)과 프로세스를 통한 시장 창출과 비즈니스 모델의 변화를 가능하게 한다. 새로운 가치를 창출하는 것으로 AI에 기반을 둔 디지털화된 제품(서비스) 제공 및 AI을 이용한 프로세스의 혁신으로 새

로운 시장 창출과 새로운 비즈니스 모델을 창출한다. 이러한 혁신은 경쟁자의 모방이나 확산이 힘들어 지속적인 경쟁우위를 유지할 수 있다.

결국, AI이라는 자원을 활용하여 새로운 비즈니스 모델을 창조(변화)시키고 새로운 기회를 창출해야 AI이 기업의 핵심적인 역량으로 강화될 수 있다.

## 제4절 소결

본 장에서는 AI의 국내외 정책 및 산업동향을 살펴보게 되면서 AI 기술 확보가 미래 국가 경쟁력을 좌우할 것으로 인식하고 있는 것을 정책과 산업 동향을 통해 인지하게 되었다.

미국의 애플, 구글 등은 이러한 AI 기술을 시리, 답마인드 등의 제품에 응용하기 시작하였으며, 기존 스마트폰 생태계의 장악력을 지속하기 위해 AI 플랫폼으로의 확장에 박차를 가하고 있다. 일본의 경우 AI의 개발에 역량을 집중하기 위해 총리 산하의 범부처 컨트론타워를 설치하고 AI 산업을 가속화해 정책을 추진 중이다. 중국은 중국 내에서 수집된 데이터를 기반으로 중국시장 특성에 맞는 서비스를 주로 개발하면서 세계 최고의 AI 선도국가라는 목표를 가지고 있다. 중국은 이러한 목표를 통해 국내 시장뿐만 아니라 전 세계 소비자까지도 만족시키는데 주력하고 기존의 인터넷 규제를 개선하여 글로벌 협업과 오픈리서치를 활성화하고 있다.

반면, 우리나라는 현재 AI 기술과 관련한 역량이 부족하고, 낮은 R&D 투자 규모와 국가 차원의 지원 체계가 미흡해 효율적인 R&D 프로세스 관리를 위한 시스템 구축이 필요한 상황이다. 국내 기업의 경우, 선진국에 대응하고자 AI에 대한 연구개발을 통해 기술을 확보하기 위해 적극적인 투자를 하고 있다. 이에 정부 측에서는 정부-민간의 역할 분담을 통해 연구방향의 타당성을 반영한 후속사업의 기획을 제시해야 된다.

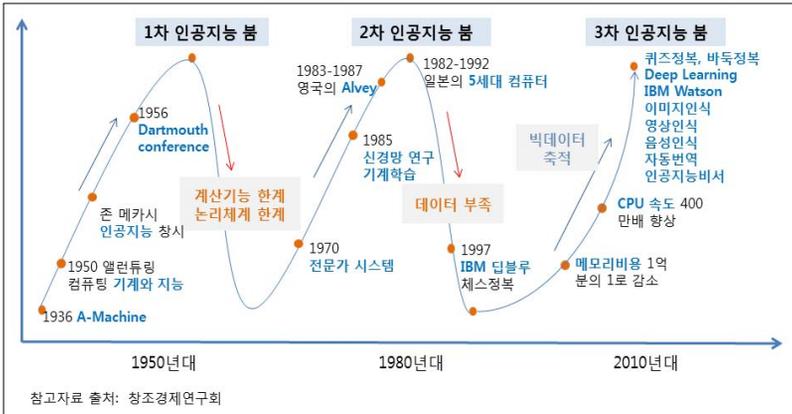
AI 기술 발전에 따른 변화를 전망하고 정부의 공공서비스 역시 기술 발전에 따라 구체적인 계획을 수립해 추진해야 하는 근본적인 이유는 AI 발전의 중요한 기회를 잡는데 있다. 적극적인 계획 수립과 방향 설정을 통해 기회를 선점하여 AI 발전의 새로운 흐름을 선도하고 경제사회 발전에 기여함과 동시에 국가 경쟁력 향상과 전반적인 발전을 이끌어야 한다. 그리고 AI 기술의 확보 기회를 놓치지 않기 위해서는 국내 AI 분야의 경쟁력 제고 및 활성화 정책 제언의 강화가 필요하다.

# 제3장 AI 기업 학습데이터

## 제1절 AI 기술과 데이터

1950년대에 1차 AI 붐을 맞이했으나, HW 자원의 부족으로 그 기대에 맞는 서비스 개발에 어려움을 겪었으며, 2차 AI 붐이 일었던 1970년대에는 HW 자원의 성장에도 불구하고 데이터의 부족으로 인해 AI 기술 개발에 어려움이 발생했다.

[그림 3-1] AI의 발전과정



자료: KEIT(2016), 『현실 속으로 확장하고 있는 인공지능』

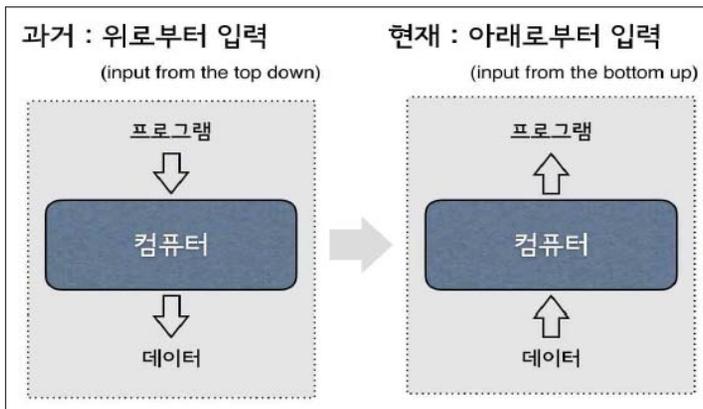
하지만 최근 AI 기술을 구현할 수 있는 슈퍼컴퓨터 및 GPU 등 HW 자원의 폭발적 성장과 머신러닝과 딥러닝으로 대표되는 AI 알고리즘의 눈부신 발전으로 인해 AI 기술이 다시 한 번 주목받고 있다. AI 기술은 기계의 두뇌 역할을 담당하는 알고리즘의 발전과 컴퓨팅 능력 증대, 빅데이터, 클라우드 등 연관 기술의 발달에 힘입어 발전이 가속화되고 있다. 이를 활용해 빠른 시간 안에 기존 성능을 대폭 향상시키는 가능성을 확인할 수 있다.

데이터는 AI 개발을 위한 필수 요소 중 하나로서 AI 패러다임 진화에 기여하고 있다.

[그림 3-2]와 같이 기존 AI 기술 개발은 Top-down 방식을 따라 알고리즘을 개발하여 적용 하였으나, 최근 AI 기술방향은 컴퓨터에 데이터를 제공하면 스스로 알고리즘을 개발하는 Bottom-up방식이 보편화되고 있다. 다시 말해 이와 같은 기술의 발전은 급증하는 지식 기반 데이터를 기반으로 기계가 능동적으로 외부와 반응하고 지식을 습득함으로써 스스로 진화할 수 있는 기본적인 토대를 제공한다.

그리고 인간이 축적해온 방대한 양의 데이터와 지식은 컴퓨터가 이해 및 학습 가능한 데이터로 전환해 AI의 활용을 가능하게 한다. 또한 차후 추론 및 인지에 이르는 수직적인 모델로 발전 중인 인지 영역의 발전을 기대하게 된다. 최근 비정형화된 데이터들로 실시간 지식 습득 및 학습이 가능해지면서 AI 기술의 급속한 발전과 인간의 능력을 증강시키는 방향으로 기술 진화가 예상된다.

[그림 3-2] AI 개발 패러다임의 변화



자료: 한국정보화진흥원 『AI시대의 정부-AI이 어떻게 정부를 변화시킬 것인가』 (2017. 6)

향후 AI 데이터의 중요성은 더욱 부각될 수밖에 없으며, 세계 각국은 AI 데이터 확보 및 개발을 위해 다양한 노력을 기울이고 있다. 이러한 흐름을 이어 받아 국내에서도 AI 데이터 확보를 위한 생태계가 조성 되어야 한다.

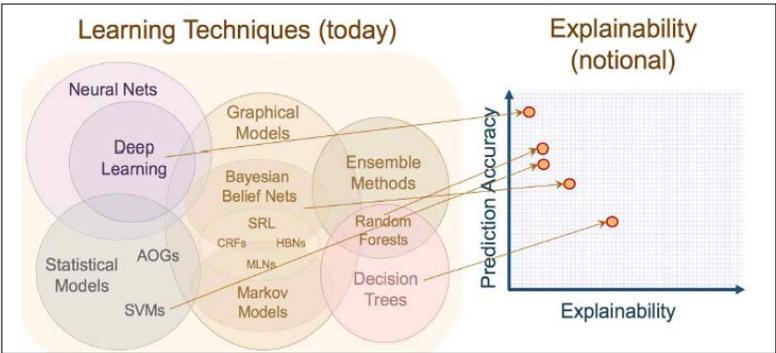
미국의 경우 국가과학기술위원회(NSTC) 보고서<sup>20)</sup>에서 민간기업의 데이터셋 제공을 권유하는 한편, AI 기술의 트레이닝 및 테스트를 위한 데이터셋과 인프라는 공공재로 간주하

여, 공공 부문의 참여를 촉구하고 있다. 하지만 구글 등 ICT 분야 민간 선도 기업들이 개발 플랫폼은 공개하고 있지만 데이터는 공개하고 있지 않으며, 이는 스타트업에게 진입장벽으로 작용할 수 있다. 국내에서도 실정에 맞는 특화된 AI 데이터 확보를 위한 개발 및 공유 촉진 정책이 필요하며 이를 중심으로 혁신주체 간 유기적 협력이 이뤄져야 한다.

데이터의 범위와 종류가 많아질수록 AI 기술의 신뢰도가 향상될 수 있으나 현재 데이터 수준 미흡으로 AI이 내리는 판단에 대해 이유를 알지 못하는 문제가 발생하고 있다.

[그림 3-3]과 같이 David Gunning(美 DARPA, 프로그램 개발자)은 데이터를 처리하는 알고리즘의 정확도가 높아질수록 인간이 이해하는 정도가 낮아진다고 주장하고 있으며 향후 AI을 정책결정과정에 접목하는데 장애요인 중 하나로 작용할 전망이다<sup>21)</sup> 했다. 이에 데이터를 쌓는 것도 중요하지만 여러 데이터들이 상호 연계되도록 개방적인 환경 구현이 필요하다고 할 수 있다.

[그림 3-3] AI의 예측정확도와 설명력 관계<sup>21)</sup>



자료: 美 DARPA 『Explainable Artificial Intelligence, Performance vs. Explainability』 부분 인용 (2017. 6)

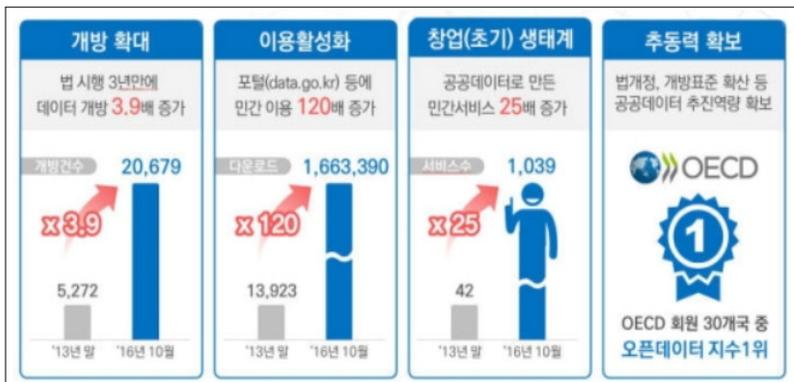
전 세계적으로 이슈가 되고 있는 4차 산업혁명에 대비하여 현 정부는 공공서비스에 적

20) NSTC(2016), PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE, Executive Office of the President National Science and Technology Council  
 21) David Gunning, Explainable Artificial Intelligence, DARPA/I20.  
[http://www.cc.gatech.edu/~alanwags/DLAI2016/\(Gunning\)%20IJCAI-16%20DLAI%20WS.pdf](http://www.cc.gatech.edu/~alanwags/DLAI2016/(Gunning)%20IJCAI-16%20DLAI%20WS.pdf)

용할 AI 기반 공공 데이터 확보를 위해 노력하고 있으나 데이터 구조와 품질 등이 고르지 못해 국민의 기대치에 못 미치는 상황이다. 이에 국가 주도의 데이터 제공 구조에서 벗어나 국민이 직접 데이터 생산에 참여하는 생태계를 구축하여 데이터의 품질을 우수하게 유지할 수 있는 노력이 필요하다.

우리나라 정부는 데이터 상호 연계에 기반 한 AI 산업 진흥 정책을 마련하고 있으나 선진국에 비해 활용도가 떨어지는 게 사실이다. 이에 따른 문제점을 개선하고자 정부에서는 이용하기 쉽고 활용 잠재력이 높은 양질의 데이터를 확대시키기 위해 AI을 활용할 공공데이터 개방의 효과를 확산<sup>22)</sup>토록 계획 중이다.

[그림 3-4] 데이터 개방 및 활용 현황



자료: 전자신문 『공공데이터로 융합 신산업 창출..(중략)..기본계획 발표』(2016. 12)

[그림 3-4]와 같이 행정안전부에서는 국내에서의 공공데이터 확대를 시행하였고, 2013년 이후 법 시행 3년 만에 데이터 개방이 3.9배로 증가한 20,679건이 되면서 민간 이용이 120배가 증가했다. 창업 생태계로부터 공공데이터로 만든 민간서비스가 25배가 증가되면서 OECD 30개국 중 오픈 데이터 지수 1위를 기록한 바 있다.

이러한 정부의 데이터 확대 및 활성화로 데이터 생애주기별 품질관리 범위 확대 및 기

22) AI형 전자정부 기반을 구축해 민원·건강 등 다양한 행정 분야에서 AI을 활용해 빅데이터 분석 사례 발굴, 빅데이터 공통기반 시스템 이용환경 개선 및 컨설팅 지원 등 빅데이터 기반의 행정 활성화를 할 예정

업이 요구하는 품질 수준의 데이터를 제공해 공공데이터 품질관리 수준평가제를 확산할 예정이다. 이에 대해 민간데이터와 융합이 쉬운 공공데이터를 우선 개방, 민간과 기관간의 데이터 결합 등의 지원이 필요하다. 활용가치가 높은 민간 데이터는 개방 및 유통 여건이 열악하기 때문에 융합형 및 지능형 데이터를 통한 신성장 동력을 지원하고 고품질데이터 지원과 신산업 육성을 위한 데이터 활용 생태계를 조성하는데 노력을 기울이고 있다.

AI 기술 발전에 따라 공공기관의 수요도 늘고 있지만, 이에 적용할 수 있는 표준화된 데이터가 없는 상황으로써 민간데이터와 융합이 쉬운 공공데이터를 우선 개방해 데이터를 결합하는 지원이 필요하다. 그리고 <표 3-1>와 같이 국내 정부에서는 지능정보사회의 발전에 맞춰 필요한 새로운 유형의 데이터 전략 수립이 필요한 시점으로 유망 신산업별로 산·학·연 및 부처 간 협업을 통해 데이터 개방 및 활용 기반 조성이 필요한 상황이다.

<표 3-1> 공공/민간데이터 정책 추진 현황

공공	<ul style="list-style-type: none"> <li>- 행정안전부에서는 융합형·지능형 고품질 데이터 개방을 확대해 민간데이터와 융합이 가능한 공공데이터 우선 개방 및 새로운 유형(IoT, AI 등)의 공공데이터 개방전략을 수립<sup>23)</sup></li> <li>- 민간 데이터 및 서비스의 공공활용을 확대시켜 앱 개발 환경 제공을 통한 국민 주도 앱 서비스 개발을 활성화해 양질의 민간데이터를 공동 활용, 우수 민간서비스 및 플랫폼 활용을 강화할 계획</li> </ul>
민간	<ul style="list-style-type: none"> <li>- 프라이버시 침해 없이 데이터 유통과 활용 활성화 추진해 통계데이터 센터 활용 허브화 추진</li> <li>- 비식별정보로서 개인정보의 안전성을 확보 할 수 있도록 비 식별화되어 타 정보와 결합해 부가가치 창출을 지원할 필요가 있음</li> <li>- 특정기관이 보유한 개인정보를 이용한 개인의 동의하에 다른 기관에 제공하여 다양한 서비스 개발에 활용되는 지원이 필요</li> <li>- 민간 빅데이터와 공공데이터간 연계 및 비식별화 지원을 위한 통계데이터 센터 구축 및 운영</li> </ul>

본 장에서는 AI 학습데이터의 분류체계를 연구해 산업체의 개발 기술 또는 산업 영역에 따른 데이터의 중요도를 판단하여 수요 및 분류체계 조사의 객관성을 확보하고자 한다.

23) 신산업 육성을 위한 데이터 활용 생태계를 조성해 AI 등 고성장이 전망되는 신산업 공공데이터 지원 계획을 수립 예정

또한 AI 학습데이터의 우선순위 발굴을 통한 데이터의 양과 품질 수요에 따른 유망 기술 분야 및 적용 산업 분야 별 우선순위를 선정 및 분석한다.

또한 AI 학습데이터 조사 결과를 고려한 지능정보 선도형 R&D 지원 체계를 마련하고자 한다. 기술적 측면으로 학습데이터 품질 측정기법 개발 등을 통한 학습데이터 활용성을 확보하고 정책적인 측면에서는 양질의 정부 및 민간 데이터 공개 등의 사항에 대해 설명할 것이다.

그리고 AI 데이터 분류체계 마련을 위한 전문가(기업) 인터뷰를 통해 기존 문헌 분석 결과와 전문가 인터뷰 결과를 종합하여 데이터 수요를 파악하기 위한 데이터 분류체계를 마련하고자 한다. 이러한 분석은 AI 데이터 분류체계에 따른 조사로 정립된 데이터 수요를 파악하는데 있다.

마지막으로 조사 결과 분석 및 제언을 통해 데이터 수요 파악을 위한 기술-적용분야 매트릭스 분석으로 각 영역별 데이터 집중도를 분석하고자 한다. 이러한 분석을 바탕으로 데이터 수요와 기업 현황(매출, 종사자수, 사업분야 및 영역) 간 교차 분석을 통해 입체적 분석 결과 및 시사점을 도출하기로 한다.

## 제 2 절 전문가 활용 인터뷰

### 1. 데이터 수요조사

기업체가 요구하는 AI 기술 발전에 필요한 데이터<sup>24)</sup>를 파악하기 위해 분류기준 및 이용 목적과 기타 의견들을 청취하기 위하여 AI 전문 기업의 임원들을 대상으로 인터뷰를 시행하였다. 전문가 인터뷰의 목적은 AI 데이터의 분류 체계 파악을 통해 개선점을 도출시켜 데이터 활성화 방안을 마련하는데 있다. 또한, AI 데이터에 따른 분류 기준 마련과 산업 경쟁력이 높은 영역의 의견을 제시하고자 한다.

본 연구에서는 <표 3-2>와 같이 비즈니스 수요가 있는 데이터를 정의하고 분류 기준을 표준화하여 사업 환경 및 채널을 제공할 수 있는 기반을 구축하고자 한다. 그리고 AI 데이

---

24) 데이터의 분류기준은 1) AI SW 기술에 따른 기준과 2) 산업 분류 기준과 3) 학문적인 발전을 위한 대학의 공공데이터 셋으로 분류한다.

터 전문가를 대상으로 AI 데이터 분류 체계 파악 및 개선점 도출해 담당자와 협의 이후 시행한 내용으로 구성하였다.

〈표 3-2〉 AI 데이터 분류 체계 마련에 대한 전문가 인터뷰 질문사항(안)

Dataset 분류	Dataset 생성, 도입	Dataset 이용목적, 역량 수준	Dataset 발전 방향
<ul style="list-style-type: none"> <li>■ AI SW 활용 Dataset 분류기준에 대한 의견 제시</li> <li>■ Dataset 중 가장 경쟁력 있는 영역 등 의견</li> </ul>	<ul style="list-style-type: none"> <li>■ 품질 양이 AI에 활용 가능한 외부 Dataset 수집 및 도입형태 의견</li> <li>■ 기존 빅데이터의 한계를 해결하는 가공 프로세스에 대한 의견</li> </ul>	<ul style="list-style-type: none"> <li>■ 수준별 필요역량 인프라에 대한               <ul style="list-style-type: none"> <li>• 기존 프로세스 효율성</li> <li>• 기존 비즈니스 모델 강화 및 변화</li> <li>• 데이터의 수익 원천화 전통기업, 창업 및 벤처기업</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>■ 산업군 내 기준 Dataset 구축</li> <li>■ AI Dataset 생산 등 생태계 미성숙에 대한 개선의견</li> <li>■ 융복합 Dataset 표준 구축 촉진 기반 조성에 대한 건의</li> </ul>

자료: 전문가 인터뷰 실시계획 (안), AIIA 재구성

AI 데이터 분류체계 도출을 위한 수요조사를 통해 국내 지능정보기술 업체의 needs를 파악하고 정형화 된 질문을 통해 전문가델파이 조사를 실시하였다. 그리고 유형별 우선순위를 설정해 데이터 수요 조사 방안에 대한 조사단계를 〈표 3-2〉와 같이 3단계로 나눠 진행했다.

데이터 조사 대상 기업의 수요를 구체화하기 위해 브레인스토밍을 통해 사전에 질문을 정형화시켜 항목을 유형화했다. 유형별로 우선순위를 설정하고 기업을 대상으로 상위 20개 유형을 설문으로 보완하였다.

〈표 3-3〉 데이터 수요 조사 방안 3단계

①단계 수요 유형화(~8月中旬)		②단계 유형별 우선순위 설정(9月 초)		③단계 수요 조사(10~11月)	
방법	브레인스토밍	방법	델파이설문	방법	설문 조사
대상	전문가 6인+연구진	대상	전문가20인 내외	대상	200여개 기업
결과물 (예상)	50개 유형	결과물 (예상)	50개 유형 → 20개 선정	결과물 (예상)	- 20개 유형별 수요 - 그 외 수요(주관식)

본 인터뷰는 2달(2017. 8~9) 간 AI 산학연 데이터 전문가를 대상으로 실시하였으며 해당 사항은 〈표 3-4〉와 같다.

〈표 3-4〉 전문가 인터뷰 결과 요약

구분	소속기관	전문분야 및 특징	주요 내용
산	솔트룩스	(전문분야) AI데이터, 복합지능 (특징) 국가 다수 AI R&D 참여 중, 자체 데이터센터 운영 등	<ul style="list-style-type: none"> <li>- 머신러닝과 딥러닝에 필요한 데이터는 많은 차이가 존재하며 머신러닝을 위해서는 Labeling이 우수한 데이터가 필수</li> <li>- 정부 입장에서 산업 Domain별로 데이터는 주는 것도 의미가 있으나, 기업들의 수요를 파악하여 집중투자를 하는 것이 타당</li> <li>- 정부 AI R&amp;D사업을 통한 데이터는 매우 유용한 재료이지만 연구 후 활용 확산은 주관기관의 역량에 따라 상이함. 정부 AI R&amp;D사업 선정시 주관기관에 대한 확실한 역량평가 필요</li> </ul>
	플리토	(전문분야) AI기반 통번역 서비스 (특징) 2017년 초 미래부 주관으로 언어처리 확산을 위한 DB(코퍼스) 구축 수요조사 응답	<ul style="list-style-type: none"> <li>- Top-down 방식의 데이터 분류는 스타트업 현실에서는 수요-공급의 괴리 이어질 수 있음</li> <li>- Bottom-up 형식의 데이터 분류가 될 수 있도록 생태계 조성이 더 중요하다고 판단</li> <li>- 직접 데이터 제공보다는 민간 데이터 구입을 위한 보조금 지원이 타당할 수 있음</li> </ul>
	카카오	(전문분야) 음성, 언어지능 기반 AI서비스 (특징) 카카오 플랫폼과 결합하여 다양한 서비스 준비 중	<ul style="list-style-type: none"> <li>- 카카오미니의 경우 자연어처리가 주된 요소 기술로 데이터 분류체계도 기술요소로 구분(ex) 형태소 분석, 개체명 인식, 구문분석 등</li> <li>- 현재 카카오는 Domain을 “문어체, 구어체”, “포털 메뉴” 등에 따라 분류하고 있지만, 고정적인 것은 아님</li> </ul>

구분	소속기관	전문분야 및 특징	주요 내용
산	뷰노 코리아	(전문분야) 딥러닝 기반 의료 시각지능 AI시스템 (특징) 의료분야에 특화하여 “뷰노 메드”, “본에지” 등 개발	<ul style="list-style-type: none"> <li>- AI을 위한 데이터는 ①labeling의 양과 정확성 ②데이터의 양 ③데이터의 연속성 ④데이터의 공개성 등이 보장되어야 함</li> <li>- 딥러닝의 경우 데이터의 양이 매우 중요하기 때문에 정부 차원에서 의료분야 등 민감데이터에 대한 공개가 확장되었으면 함</li> </ul>
연	NIA	(전문분야) 공공분야 데이터 공개 포털 운영 (특징) 의료, 자율형주행차 등에 데이터 연결성을 강조한 AI 데이터 구축 중	<ul style="list-style-type: none"> <li>- 데이터의 양도 중요하지만, 데이터 간의 연결성이 있어야 활용성이 높아질 것이라 판단</li> <li>- 공공데이터포털에 활용되는 산업 Domain을 통해 AI 데이터를 제공하는 것보다 AI 산업체에서 요구하는 데이터 수요를 조사하여 제공하는 것이 타당</li> </ul>
	KETI	(전문분야) 언어지능, 시각지능 등 복합지능을 가진 멀티모달 생활형 AI 서비스 개발 (특징) AI플래그십 과제인 “디지털 동반자” 솔루션 개발 중	<ul style="list-style-type: none"> <li>- Task by Task로 데이터가 정리되어야 머신러닝을 위한 데이터로 활용 가능</li> <li>- 산업 Domain별로 데이터를 공급하는 것을 고려할 것이 아니라 머신러닝이 활용되는 분야의 Task를 구분하여 실질적이 데이터를 제공하는 것이 중요</li> <li>※ 세계 주요 AI데이터 센터들은 Task로 구분하여 제공 중</li> </ul>

위 내용에서 알 수 있듯이 산학연 전문가들은 AI 기술 발전을 위해 데이터의 중요성에 대해 깊이 공감하고 있다. 하지만 수요 데이터의 종류와 제공 방법에 대해서는 각 기관의 성격에 따라 상이한 측면이 있다. 이는 국내 AI 데이터 제공 체계가 확립되어 있지 않고 데이터 활용 방법에 대한 이견이 존재하기 때문이다. 이에 현 시점에서 AI 데이터 활용의 제한사항 등을 분석하여 차후 연구방향 설정에 활용되어야 할 필요성이 있다.

## 2. 데이터 활용 제한사항

AI에 활용할 수 있는 데이터를 재분류 할 수 있는 프레임은 존재하지 않아 AI 기술 분류(정형, 비정형, 반정형/외부, 내부/공공, 기업)에 따라 활용할 수 있는 데이터는 더욱 정하기가 어렵다.

외국의 경우, 머신러닝을 위한 데이터를 명확히 분류하기 보다는 산업 분류 중심으로 활용할 수 있는 데이터 셋을 구분하고 있어 산업별, 기술별, 기술요소별이 아닌 랜덤한 방식으로 구성하고 있다.

〈표 3-5〉 UCI Machine Learning Repository 분야별 데이터셋 리스트

분야 (개수)	데이터셋 예시
Life Sciences (87)	Breast Cancer, Diabetes, Plants, Protein Data, Zoo
Physical Sciences (45)	Function Finding, Ozone Level Detection, Wine
CS / Engineering (121)	Amazon Commerce reviews set, 3D Road Network (North Jutland, Denmark), Blogger, Buzz in social media, Energy efficiency, Mechanical Analysis, Twitter Data set for Arabic Sentiment Analysis, Facebook Comment Volume Dataset
Social Sciences (23)	Census Income, Communities and Crime, Farm Ads
Business (23)	Perfume Data, Restaurant & consumer data, YouTube Spam Collection
Game (10)	Chess (Domain Theories), Poker Hand
Other (67)	Automobile, Legal Case Reports, Tennis Major Tournament Match Statistics

자료: UC 얼바인 『UC Irvine Machine Learning Repository』(<https://archive.ics.uci.edu/ml/index.php>)

산업체의 개발 기술 또는 산업 분류에 따라 수요 데이터의 중요도를 판단할 수 있기 때문에 Sampling이 선행되지 않으면 수요 및 분류체계<sup>25)</sup> 조사의 객관성 확보가 어렵다. 그리고 산업분야, AI 기술요소 별 구분도 의미가 있을 수 있으나 각 기술공급 업체의 요구사항에 따라 획일화하여 구분되기는 어렵다.

25) 머신러닝 개발을 위한 Task별 분류체계에 대한 요구사항이 다수 존재하며, 이에 대한 고려가 필요

### 3. 전문가 인터뷰 결과 활용 방안

본 인터뷰를 통해 AI 관련 산업 및 기술에 따른 수요조사를 위해 <표 3-6>과 같이 Matrix 형태로 세부 기술요소별, 산업별 우선순위를 도출하였다. 그리고 전문가 집단 및 다수의 기업 의견을 수집하여 현재 AI 데이터 제공 환경을 파악하였다.

요구사항을 정량적으로 수요조사 항목에 반영하기 위해 전문가들에게 기술별, 산업별 데이터 셋의 보기를 제시하여 수요 집중도를 조사하여 분석하였다. 또한 유망 기술 분야와 적용산업분야별 우선순위를 선정하고 기업의 데이터 수요와 기업 규모·종사자 수의 상관관계를 분석해 데이터의 양과 품질에 따른 AI 기술 및 산업의 현재 위치와 잠재 위치를 파악하였다.

<표 3-6> 각 산업별 수요데이터 양식

기술분류 산업분류	추론 및 기계학습	지식표현/ 언어지능	청각지능	시각지능	복합지능	...
의료						
국방						
농업						
법률						
웰니스						
제조						
홈						
금융						
문화관광						
유통						
...						

또한 산·학·연의 AI 기술에 활용될 학습 데이터의 현실적인 요구사항을 반영하기 위해 주관식 답변<sup>26)</sup>으로 조사하고 세부요소별로 데이터의 수준 및 품질에 대한 조사를 병행하였다.

26) 주관식 답변에 대해 텍스트 마이닝에 따라 유형화 분류 및 분석

(ex) 1-1. 위 수요에 조사한 데이터의 세부 사항을 자유롭게 서술해주세요  
 1.1.1 세부 사항으로 답변하신 데이터의 수준 및 품질 요구사항은?

[그림 3-5] 데이터 셋 유형화 분류 및 분석



### 제 3 절 조사통계

#### 1. 연구조사대상

지능정보산업협회(2017)에서 수행한 “지능정보산업 생태계 발전을 위한 기업 간 효율적 협력구조 연구”를 통해 진행된 인공지능산업 실태 조사 시 「데이터셋 수요조사」 내용을 추가하여 조사하였으며, 문헌<sup>27)</sup> 및 협회사 명부를 기반으로 모집단(166개사)을 구축하였다. 회수된 100개 기업 중 데이터셋 수요 문항에 답변한 64개사으로서, <표 3-7>과 같은 항목으로 조사·분석하였다.

<표 3-7> 데이터 셋 연구조사 현황

구 분	내 용
조 사 목 적	AI 기업의 머신러닝을 위한 데이터셋 수요 파악
조 사 대 상	2016년 AI산업 관련 제품 생산 또는 서비스 제공 사업체
표 본 배 분	전수조사(166개 사업체)

27) IITP(2015), “국내 인공지능 실태조사”; NIA(2017), “우리나라 AI 기업 현황 조사”

구 분	내 용
조 사 방 법	산업체 조사 전문 면접원에 의한 산업체 방문면접조사
주요조사내용	데이터 확보의 어려움, 데이터셋 수요, 데이터셋 우선 순위 등

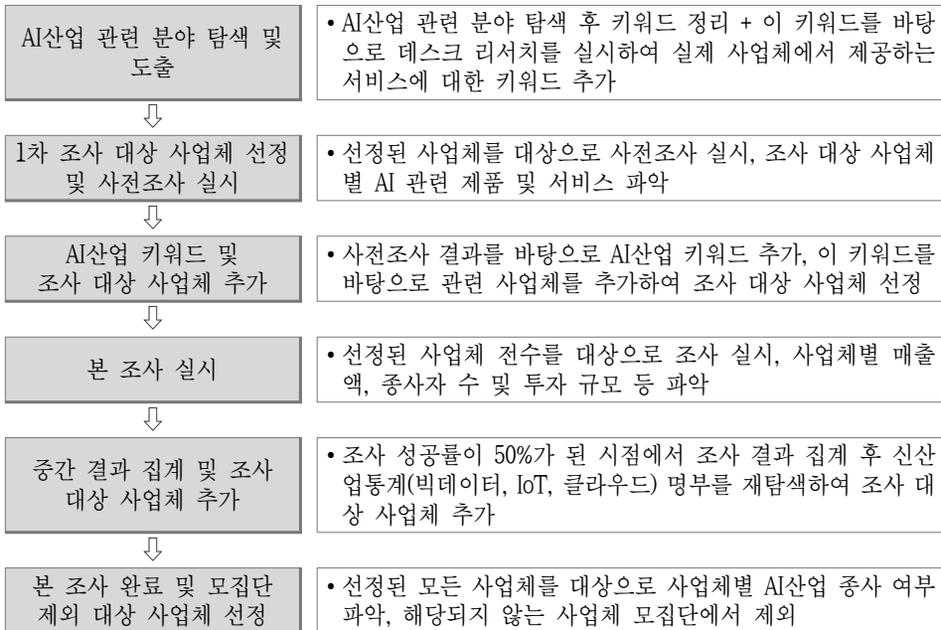
자료: 지능정보산업협회(2017), “지능정보산업 생태계 발전을 위한 기업 간 효율적 협력구조 연구”를 통해 진행된 인공지능 산업 실태 조사 시 「데이터셋 수요조사」 내용을 추가하여 조사 수행

## 2. 연구방법론

### 가. 모집단 구축

목표모집단은 AI 기술 개발 및 AI 적용 제품·서비스·플랫폼의 생산, 유통, 활용, 부가서비스(조사/분석, 컨설팅, 중개) 과정에서 가치를 창출하는 기업으로 설정하였으며, AI산업 모집단 명부 구축 절차는 다음과 같다.

〈표 3-8〉 AI산업 모집단 구축 절차



자료: 지능정보산업협회(2017), “지능정보산업 생태계 발전을 위한 기업 간 효율적 협력구조 연구”를 통해 진행된 인공지능 산업 실태 조사 시 「데이터셋 수요조사」 내용을 추가하여 조사 수행



### 3. 연구내용 결과

#### 가. 데이터 확보의 어려움

데이터 수요 조사에 참여한 AI 기업 64개사 중 36개사(56.3%)가 AI SW개발에 있어 데이터 확보가 어렵다고 응답하였다. 보유 AI 기술 분야별로는 「시각지능」 분야가 가장 많은 어려움을 느끼고 있었으며, 다음으로는 「지식표현 및 언어지능», 「지능형에이전트」 등의 순으로 나타났다.

〈표 3-9〉 AI SW개발을 위한 데이터 확보의 어려움 여부

구분	전체		예		아니오		
	건	%	건	%	건	%	
■ 전 체 ■	64	100.0%	36	56.3%	28	43.8%	
AI 분야별	추론 및 기계학습	51	100.0%	30	58.8%	21	41.2%
	지식표현 및 언어지능	24	100.0%	17	70.8%	7	29.2%
	청각 지능	9	100.0%	6	66.7%	3	33.3%
	시각 지능	15	100.0%	11	73.3%	4	26.7%
	복합 지능	8	100.0%	5	62.5%	3	37.5%
	지능형 에이전트	13	100.0%	9	69.2%	4	30.8%
	인간-기계 협업	7	100.0%	3	42.9%	4	57.1%
	지능형 반도체	2	100.0%	-	0.0%	2	100.0%
	뉴로모픽칩	2	100.0%	1	50.0%	1	50.0%
	고성능컴퓨팅	6	100.0%	2	33.3%	4	66.7%
	양자컴퓨팅	-	100.0%	-	-	-	-

자료: 지능정보산업협회(2017), “지능정보산업 생태계 발전을 위한 기업 간 효율적 협력구조 연구”를 통해 진행된 인공지능 산업 실태 조사 시 「데이터셋 수요조사」 내용을 추가하여 조사 수행

#### 나. 데이터 수요

AI 기술 분야별로는 「추론 및 기계학습(148건)」, AI 적용 분야별로는 「의료(52건)」 분야에 대한 데이터 셋에 대한 수요가 많은 것으로 나타났다.

적용 분야-AI 기술을 영역별로 살펴보면 「의료-추론 및 기계학습」 영역이 32건으로 데이터 수요가 가장 많은 것으로 나타났으며, 다음으로는 「금융-추론 및 기계학습」 15건, 「제조-추론 및 기계학습」 11건 등의 순서로 나타났다.

〈표 3-10〉 AI 기술-적용 분야 데이터셋 수요 영역별 현황

(단위: 건, 중복응답)

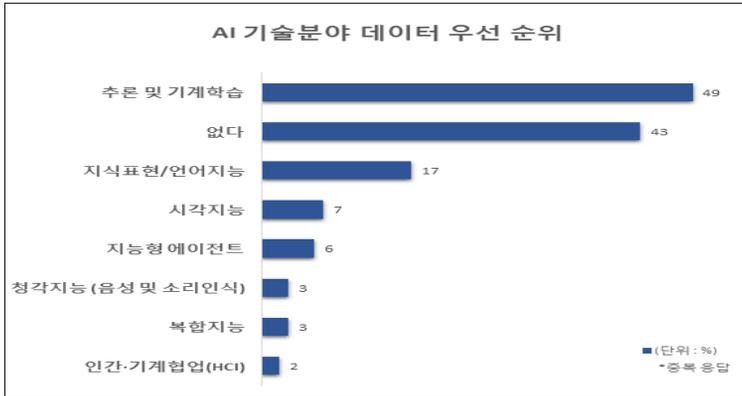
	추론 및 기계학습	지식표현/언어지능	시각 지능	청각지능 (음성 및 소리인식)	복합 지능	지능형 에이전트	인간·기계 협업 (HCI)	합계
① 의료	32	10	3	2	2	2	1	52
② 국방	10	4	2	-	-	1	-	17
③ 교육	9	4	2	-	-	2	-	17
④ 농·축산·어업	4	2	-	-	-	2	-	8
⑤ 법률	5	3	-	-	-	2	1	11
⑥ 웰니스	7	4	2	1	2	2	-	18
⑦ 제조	11	3	1	-	-	2	2	19
⑧ 홈	5	5	1	2	1	2	-	16
⑨ 금융	15	5	1	-	1	5	-	27
⑩ 문화관광	3	2	1	-	-	1	-	7
⑪ 유통	9	4	1	1	1	3	1	20
⑫ 사무관리	8	3	-	-	-	2	-	13
⑬ 도시	2	1	-	-	-	1	-	4
⑭ 교통	6	3	1	1	-	1	-	12
⑮ 에너지	5	-	-	-	1	2	1	9
⑯ 안전	7	1	1	1	1	1	-	12
⑰ 공공서비스 및 복지	10	4	2	-	1	2	-	19
합계	148	58	18	8	10	33	6	281

자료: 지능정보산업협회(2017), “지능정보산업 생태계 발전을 위한 기업 간 효율적 협력구조 연구”를 통해 진행된 인공지능 산업 실태 조사 시 「데이터셋 수요조사」 내용을 추가하여 조사 수행

다. 데이터 우선 순위

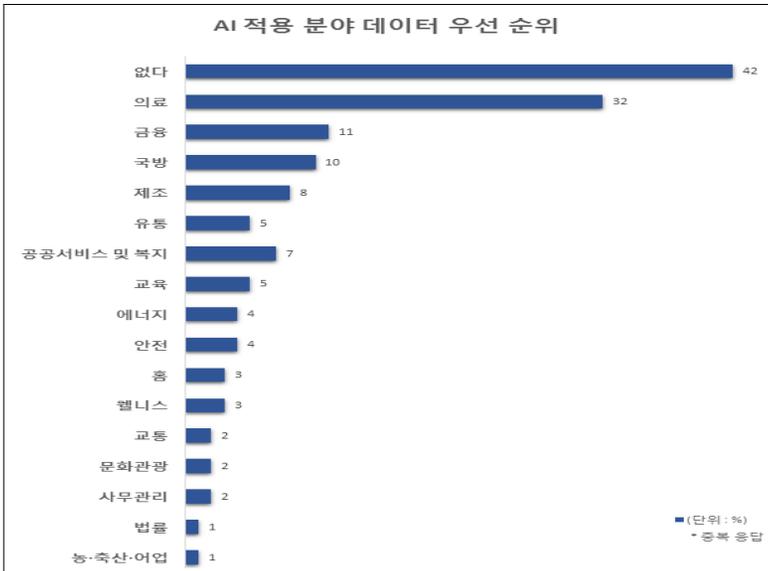
AI 기술 분야별 데이터 우선순위는 「추론 및 기계학습」이 가장 높은 것으로 나타났으며, 다음으로는 「지식표현/언어지능», 「시각지능», 「지능형에이전트」 등의 순서로 나타났다.

[그림 3-7] AI 기술분야 데이터 우선순위



자료: 지능정보산업협회(2017), “지능정보산업 생태계 발전을 위한 기업 간 효율적 협력구조 연구”를 통해 진행된 인공지능 산업 실태 조사 시 「데이터셋 수요조사」 내용을 추가하여 조사 수행

[그림 3-8] AI 적용 분야 데이터 우선 순위



자료: 지능정보산업협회(2017), “지능정보산업 생태계 발전을 위한 기업 간 효율적 협력구조 연구”를 통해 진행된 인공지능 산업 실태 조사 시 「데이터셋 수요조사」 내용을 추가하여 조사 수행

AI 적용 분야별 데이터 우선순위는 「의료」가 가장 높게 나타났으며, 다음으로는 「금융」, 「국방」, 「제조」, 「공공서비스 및 복지」, 「유통」, 「교육」 등의 순서로 나타났다.

## 제4절 소결

AI 사회로 가기 위해 준비해야 할 요소 중 우선순위로 데이터 축적 및 유통을 꼽아 데이터 기반을 확충하고 활용 역량을 키워 제도를 개선해나가야 된다. 현재 AI 등에 활용할 데이터가 절대적으로 부족한 점, 규제로 인해 데이터 유통이 원활하지 못한 점 등을 들어 무엇보다 먼저 데이터 기반을 확충해야 한다고 강조해야 된다. 데이터가 축적되더라도 이를 활용하고 가치를 창출하기 위해서는 데이터 마인드 및 활용역량을 함께 키워야 한다.

또한 데이터의 축적 및 유통을 유도하기 위해서는 개인정보보호법 등의 제도 개선이 필요한 만큼 규제개혁 등 혁신문화와 제도 유연성을 먼저 개선해 나가야 된다.

AI의 지능정보화 준비를 위한 지향 국가로서 우리나라가 롤 모델로 여길만한 나라로는 미국과 영국을 꼽을 수 있다. 미국은 지능정보 분야에서 가장 앞선 나라이기 때문에 배워야 할 점이 많다고 판단되며 영국은 자신만의 장점을 살려서 지능정보사회를 선도하는 국가로 한국이 벤치마킹할 필요가 있다.

〈표 3-11〉 국가별 지능정보화 지향 이유

국가	내용
	<ul style="list-style-type: none"> <li>• 외향적으로 자유분방한 분위기가 있어 개인정보 활용에 대해 엄격할 것 같지만 기업이 사업목적으로 활용하는 것에 대해 관대함 - 구글: AI 분야에서 앞서갈 수 있는 이유도 사업목적에 대해 관대함을 가진 환경이 뒷받침되고 있음</li> <li>• 한국과는 다르게 오픈소스 개념이 확립되어 있어 개발 목적이 명확해 발전에 토대가 됨</li> </ul>
	<ul style="list-style-type: none"> <li>• 미국보다 기술이 뒤쳐져 있지만, 영국만의 장점을 가질 수 있는 분야 (영상, 게임 등 콘텐츠분야)에 특화하여 개발하고 있음</li> <li>• 활용 측면에서도 핀테크와 AI 기술을 융합하여 빅데이터 분석에 적극적으로 활용하고 있으며, 교육 제도 또한 벤치마킹 필요함</li> </ul>

국가	내용
	<ul style="list-style-type: none"> <li>• 로봇, AI에 대한 사회적, 인문적, 문화적 지지기반 형성</li> <li>• 문화 콘텐츠를 통해 국민들이 로봇 등에 대해 친숙함을 느끼고 있음</li> </ul>
	<ul style="list-style-type: none"> <li>• S/W 전문인력 양성을 위해 전문학교에 연 100억원 씩 집중 지원하여 매년 많은 전문가를 양성하고 있으며, 현재 약 10만 명을 양성</li> </ul>

자료: 한국정보화진흥원(2016. 12) 『새로운 기술, 새로운 세상 지능정보사회』 AIIA 재구성(2017. 3)

데이터 수요 조사 결과 56.3%의 기업이 데이터 확보에 어려움을 겪고 있다고 응답하였다. 매출액이 크고, 종사자 수가 많을수록 데이터 확보에 어려움을 겪고 있다고 응답한 경향이 많았는데 이는 소규모 기업에서는 AI SW개발보다는 외부 플랫폼을 활용하는데 기인한다.

AI은 데이터를 통한 미래 예측, 복잡한 의사결정 등 인간만이 가능했던 고도의 지식처리 분야에 도전해 데이터를 통하여 학습하고 스스로 성능을 향상시키는 작업을 수행하고 있다. 빅데이터 기반의 지식처리 능력이 매우 빠르게 향상되면서 향후 인간의 개입 없이도 중요한 결정을 정확히 내릴 수 있는 의사결정능력을 보유하게 된다면 산업을 넘어 사회 전체에 거대한 영향을 끼치게 될 것이다. 학습능력과 더불어 인간의 감정을 읽고 반응하는 감성능력도 AI의 중요한 기능으로 형성시켜야 된다.

AI 제품 및 서비스는 사용자의 특성 및 취향, 기분을 파악하여 인간처럼 섬세한 서비스를 선제적으로 제공하는 것을 가능하게 만든다. AI 기계는 이전보다 똑똑한 기계에서 나아가 인간처럼 보고, 느끼고, 배우며 자연스럽게 상호작용 할 수 있는 기계로 발전하고 있다.

AI은 다양한 산업과 실생활에 빠른 속도로 파고들며 영역을 확장 중이며, 미래 모든 산업 분야에 적용될 것으로 예측돼 AI 기술이 적용된 지능형 어플리케이션이 시장에 출시되고 있다. 이러한 발전 가능성에 대한 기존의 우려가 기대와 믿음으로 변화시켜 새로운 시장과 혁신기회를 창출하는 주요 기술로 거듭나고 있다. AI 시장의 영향력이 확대돼 향후 기술수준을 기업경쟁력의 척도로서 인식시켜야 한다.

## 제 4 장 개방형 R&D 생태계 조성 방안

### 제 1 절 4세대 연구 패러다임과 오픈 사이언스(Open Science) 개념

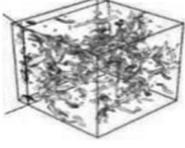
과학기술이 발전하면서 이를 수행하는 방식이 기존의 이론 및 실험 중심의 1 또는 2세대 연구 방식에서 데이터 및 컴퓨팅·사이버인프라 활용하는 3세대 또는 4세대 방식으로 진화하고 있다. 특히 4세대 연구 패러다임은 변화는 최근 첨단 연구 장비, 센서 및 데이터 처리 기술 등 디지털 기술의 발달로 데이터가 폭발적으로 증가하면서 등장하였다. 4세대 연구 패러다임이란 『데이터 분석을 통해 새로운 이론이나 현상을 발견하는 데이터 기반 과학(Data-intensive science)』<sup>28)</sup>를 말한다. 특히, 실험·관측·측정과 관련된 첨단 연구 장비의 출현은 관련 데이터를 크게 증가시키는 한편, 네트워크의 발달은 많은 양의 데이터를 유통 및 전달을 촉진하고 있다. 따라서 이러한 방대한 양의 데이터를 통해 새로운 과학기술의 발전을 추구하는 4세대 연구방식이 도래하였다.

최근의 과학기술의 발전은 엄격한 프로세스의 가설 테스트만을 통해서만 이루어지고 있지 않다. 방대한 양의 데이터, 복잡하고 찾기 힘든 관계, 다양한 학문 분야 간의 협력, 새로운 유형의 실시간 출판 등은 과학적 연구 방법의 패턴을 변화시키고 있다. 특히, 다양한 과학기술 분야에서 데이터 집약적인 연구 방식과 차세대 ICT 기능의 활용하는 방식 등 새로운 연구 방식이 점차 확산되어 가고 있다.

---

28) Hey, T., Tansley, S., & Tolle, K. M.(2009). The fourth paradigm: data-intensive scientific discovery (Vol. 1). Redmond, WA: Microsoft research.

[그림 4-1] 연구 패러다임의 변화

1세대 연구	2세대 연구	3세대 연구	4세대 연구
실험 중심연구	이론 중심연구	컴퓨팅자원 중심연구	데이터 중심연구
	$\left(\frac{a}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$		

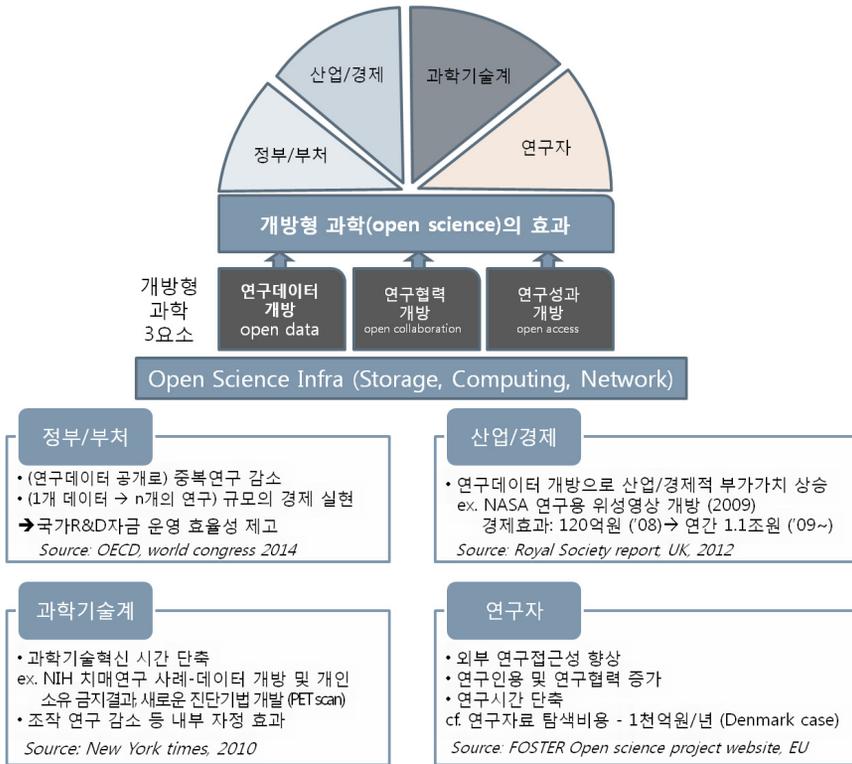
자료: Data Intensive Scientific Discovery(2009), "The Forth Paradigm"

이러한 4세대 연구 방식은 과학기술의 환경 변화에 따라 OECD의 정책 의제로도 등장하고 있다. 또한 4세대 연구 방식은 연구 결과 및 과정을 개방하고, 공유하는 『오픈 사이언스(Open Science)』 개념과 같은 맥락으로 파악할 수 있다. OECD(2015)에 의하면 오픈 사이언스(Open Science)란 『모든 종류의 과학적 지식(논문, 데이터, 방법론, 교육자료 등)을 디지털화하여 개방 및 공유하고 공동연구를 활성화함으로써 과학 연구를 발전시키고자 하는 개념으로, 공공자금으로 지원된 연구 성과 및 연구 과정의 데이터를 디지털 포맷으로 공개하여 연구 커뮤니티뿐만 아니라 일반 국민도 보다 쉽게 접근하고 활용할 수 있도록 하는 메커니즘』이다.<sup>29)</sup>

이러한 오픈 사이언스는 3대 체계, 즉 ① 연구 성과의 개방(Open Access, 이하 오픈 액세스), ② 연구 데이터(Open data, 이하 오픈 데이터: 연구 과정에서 발생하는 모든 형태의 제반 데이터), ③ 연구 협력(Open Collaboration, 이하 오픈 콜라보레이션)으로 나눌 수 있다.

29) 오픈 사이언스 개념이 대두되면서, OECD 과학기술장관회의('15. 19)에서는 대전선언문을 통해 '개방형 과학 촉진을 위한 정책개선 지원'을 정부에 요청하였음

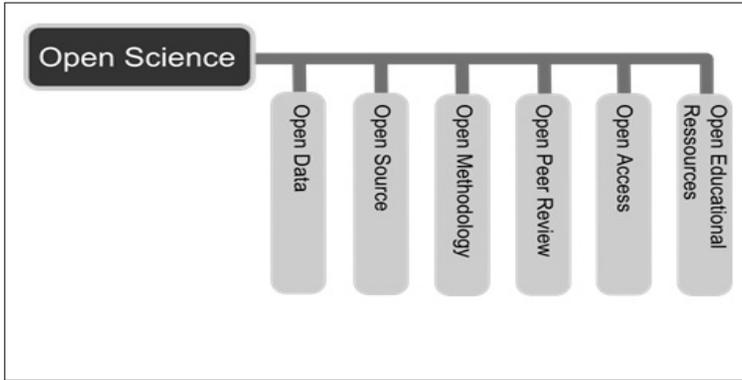
[그림 4-2] 오픈 사이언스의 개념과 이해



자료: 오픈사이언스 포럼 발표자료 (2016)

오픈 사이언스의 범위는 기초과학뿐 만 아니라 인문·사회과학을 포함한 모든 유형의 연구 활동을 포함하며, 개방·공유의 대상을 [그림 4-3] 또는 <표 4-1>과 같이 6가지로 정의하고 있다. 또한 오픈 사이언스는 공공자금으로 부터 펀딩을 받은 연구의 경우 기본적으로 오픈 액세스를 수행하고, 장비와 데이터를 개방·공유하며, 오픈 사이버 인프라에 지속적인 투자를 진행한다는 원칙을 따르도록 권유한다.

[그림 4-3] 6가지 오픈 사이언스 공유 대상



자료: Open Science ASAP 웹페이지

<표 4-1> 6가지 Open Science 공유 대상의 내용

대상	내용
Open Access	연구 활동의 결과물인 논문에 대한 자유로운 접근 허용
Open Data	연구데이터의 개방과 공유
Open Source	하드웨어 및 소프트웨어의 코드 공개 및 자유로운 이용 허용
Open Methodology	연구에 이용된 방법론 공개
Open Peer Review	논문심사의 투명성을 위해 심사자의 리뷰결과 공개
Open Educational Resources	학교에서 활용되는 교육 자료의 공유

자료: Kraker 외(2011), "The Case for an Open Science in Technology Enhanced Learning"

주: 개방형 동료 리뷰 (Open Peer Review)와 공개 교육 자료 (Open Educational Resources)는 과학의 두 가지 중요한 측면

4세대 연구 방식의 데이터 기반 과학은 방대한 데이터 수집과 과학기술 분야 사이의 융·복합 연구가 요구되기 때문에 데이터 공유 및 협력 또한 필요하다. 따라서 다양한 과학기술 분야 간 융합 연구 및 국제 공동연구가 지속적으로 확대되고 있는 추세이며, 이러한 4세대 연구 패러다임의 변화와 흐름이 과학기술 분야의 개방형 혁신인 오픈 사이언스로 이어지고 있다.

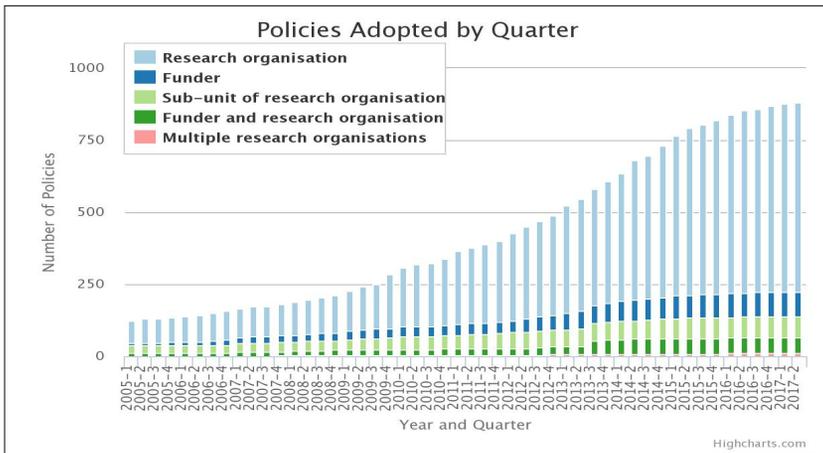
## 제 2 절 오픈 사이언스(Open Science) 동향 및 정책적 시사점

### 1. 오픈 사이언스 정책의 확산

오픈 사이언스는 모든 종류의 과학기술 분야의 지식을 개방하고 공유하는 것으로 네트워크 발달과 함께 필요성이 대두되었으며, 연구개발과정 개방화에 대한 사회적 요청으로 확산되고 있다. 특히 오픈 사이언스는 기술 진보에 따른 필연적 결과가 아닌 사회·정책적 선택으로 해석할 수 있다. 즉, 공적 자금의 지원에 의한 연구 성과(출판물)를 공개하여 연구투자에 대한 사회·경제적 편익을 제고해야 한다는 인식의 확산과 더불어, 학문의 발전을 위해 연구 성과를 공유해야 한다는 주장이 확대된 결과이다.

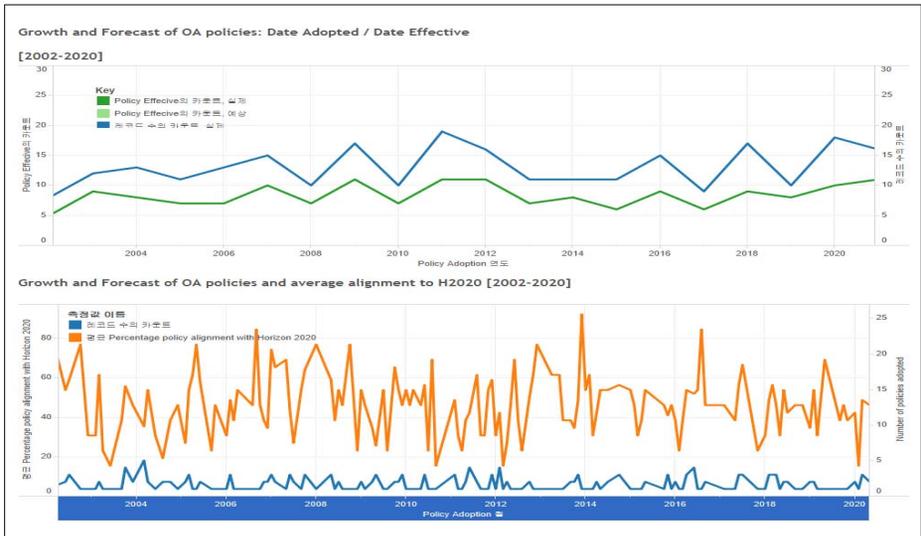
오픈 사이언스정책은 2000년대 이후 확대·도입되었으며, 학계 및 공공연구기관 등에서 자체적으로 연구 성과를 공개하도록 권고하는 등의 사례가 증가하며 2010년 이후 급격하게 늘어나고 있다.

[그림 4-4] 오픈 액세스정책의 확대(2005년~2017년)



자료: Roarmap(2017)

[그림 4-5] 오픈 액세스정책 도입 전망(2002년~2020년)

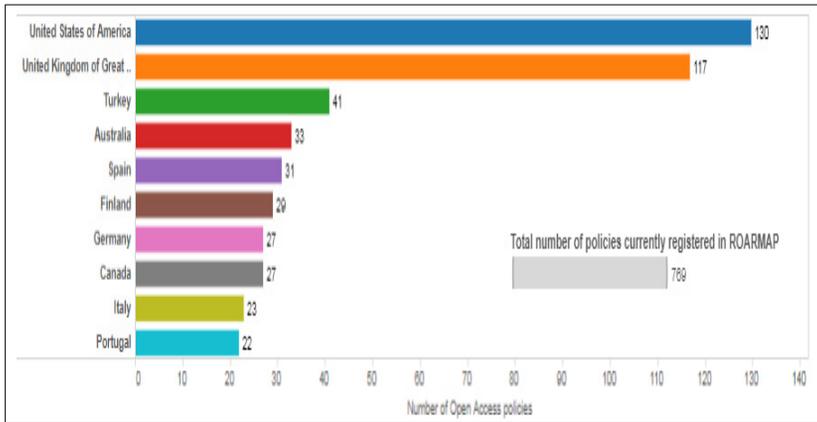


자료: Roarmap(2017)

주: 상단 그래프의 파란 추세선은 'Policy Effective의 카운트', 초록 추세선은 '예상 카운트'를 의미. 하단 그래프의 주황 추세선은 '평균 Percentage Policy alignment with Horizon 2020', 파란 추세선은 '레코드 수의 카운트'를 의미함

미주·유럽지역을 시작으로 오픈 액세스 및 오픈 데이터 관련 정책이 급격히 확대 도입 되었으며, 아시아지역으로 확산되고 있다. 국가별 오픈 액세스 및 오픈 데이터 관련 정책의 수는 상이하며, 유관 정책의 양상 또한 다양하게 나타나고 있다. 주요 국가별 연구 성과의 개방(Open Access)을 위한 정책 수는 [그림 4-6]과 같이 나타나고 있으며, 연구 성과(출판물)의 개방을 위한 정책 수는 미국, 영국, 터키, 호주, 스페인, 핀란드, 독일, 캐나다, 이탈리아, 포르투갈 순으로 높았다.

[그림 4-6] 상위 10개 국가별 연구 성과의 개방(Open Access) 정책 수



자료: Roarmap(2017)

## 2. 오픈 사이언스 정책 추진 현황 및 이행 실태

오픈 사이언스 정책은 세부 체계로 『오픈 액세스/오픈 데이터/오픈 콜라보레이션』으로 분류되어 각 국가별 및 지역별로 도입·이행되고 있다. 현재 오픈 사이언스 정책의 세부 분류 중 주로 오픈 액세스와 오픈 데이터 위주로 정책이 이행되고 있다. 전 세계적으로 오픈 사이언스 정책이 확산 추세에 접어들었으나, 국내에서 오픈 사이언스 정책에 관한 논의는 아직 초기 단계이다.

### 가. 해외

국제기구와 미국, 유럽연합, 유럽국가 등에서 새로운 지식과 가치 창출을 위해 적극적으로 오픈 사이언스 정책을 추진 중이다. 주요 국가별 오픈 사이언스의 도입과 이에 따른 관련 정책의 이행 현황은 <표 4-2>와 같다.

〈표 4-2〉 오픈 사이언스 이행 및 도입

지역	체계	추진	도입
유럽 연합	오픈 액세스	OpenAIRE	<ul style="list-style-type: none"> <li>○ 2008년부터 연구 출판물의 공개 장려하였으며 공공 레파지토리 OpenAIRE를 통해 연구 성과를 공개하도록 함</li> <li>○ Horizon2020*가 지원하는 연구프로젝트 성과 공개를 의무화함</li> <li>○ 공적 지원을 받는 연구의 성과가 오픈 액세스 저널로 출판되거나 공공 레파지토리를 통해 일반인이 열람할 수 있도록 함</li> </ul>
	오픈 데이터	Science 2.0	<ul style="list-style-type: none"> <li>○ 2014년부터 유럽연합이 시행한 오픈 사이언스 관련 정책 Science 2.0 이니셔티브는 과학과 사회의 변화로 인한 새로운 과학지식생산 메커니즘을 예측하고 이에 대응하기 위한 프로젝트임(Crouzier, 2015; STEPI, 2016)</li> <li>○ 2015년부터 유럽 오픈 사이언스 클라우드 파일럿(European Open Science Cloud Pilot)을 시작하여, 유럽 전역에 연구데이터 클라우드 서비스를 제공 추진 중</li> <li>○ 2016년부터 고위급 전문가 자문그룹 '오픈사이언스 정책플랫폼(OSPP)을 운영하여 유럽연합의 오픈 사이언스 정책 추진 방향과 전략에 대한 자문을 제공함</li> </ul>
	오픈 데이터	ORD Pilot	<ul style="list-style-type: none"> <li>○ 2014년부터 2016년까지 Horizon 2020의 일부를 제한하여 공개하는 오픈 데이터 파일럿 사업 시행하였으며 2017년 부터 전 분야로 확대 적용 중임</li> <li>○ 현재 데이터관리계획 마련이 과제로 남아있으며, 데이터 공개가 의무화 되지 않은 상태로, 필요시 비공개로 전환할 수 있는 환경임</li> </ul>
네덜란드	오픈 액세스	과학 출판물 오픈	<ul style="list-style-type: none"> <li>○ Vision on the Future of Science 정책의 일부로 오픈 액세스 및 오픈 데이터 정책이 마련됨</li> <li>○ 2013년 교육문화과학부(The State Secretary for Education, Culture and Science)는 오픈 액세스 정책 방안을 의회에 제출 하여, 5년 안에 60%의 과학 출판물을, 10년 안에 100%의 과학 출판물을 공개할 계획(Open Science and Research, 2017)</li> <li>○ 2020년까지 공공 연구 학술논문 전체를 오픈 액세스로 제공하기 위해 오픈 액세스 정책 이행을 독려 중임(DASTI, 2014)</li> </ul>

지역	체계	추진	도입
네덜란드	오픈 액세스	과학 출판물 오픈	○ 몇 과학 출판(학술지)는 자체 오픈 액세스 비즈니스 모델을 가지고 있으며 네덜란드 정부에 인증 받은 비즈니스 모델로 계획 중임
덴마크	오픈 액세스	학술논문 오픈	○ 2022년까지 학술논문 전체를 오픈 액세스 형태로 서비스하고자 민관협력체제 구축 중임(DASTI, 2014)
독일	오픈 액세스	연구 성과 공개	○ 2003년 베를린 오픈 액세스 선언 이후 독일연구재단(DFG)이 2006년 이후 지원 받은 연구 성과가 공개 및 활용될 수 있도록 요청하고 있음(DFG, 2017)
	오픈 데이터	연구데이터 관리	○ 2008년 독일주요연구기관연합(Alliance of German Science Organizations)가 제안한 공공연구 출판물 및 연구 데이터 공개 이니셔티브 'Digital Information'을 바탕으로 2010년 연구데이터 관리 지침을 마련하여 연구데이터 관리를 요청하고 있음
영국	오픈 액세스	학계 연구 데이터 오픈	○ 연구 관련 오픈 액세스정책 추진 및 오픈 액세스 출판 위한 비용 지원
	오픈 데이터	공공연구데이터 공개	○ 2014년 출판법 개정으로 공개 데이터의 비상업적 이용 및 데이터마이닝을 허용 ○ 2004년부터 DGG(Digital Curation Center)를 운영하여 연구데이터의 체계적 관리 및 공개하도록 지원
프랑스	오픈 액세스	공공연구 학술논문 오픈	○ 연구개발청(ANR)은 2007년부터 오픈 액세스 정책을 추진하여 2016년 법 개정을 통해 공공연구 학술논문의 공개 및 배포에 참여를 독려함
	오픈 데이터	공공데이터 활용	○ 2001년부터 Etalab으로 불리는 정부와 시민의 협력을 도모하는 플랫폼을 마련하여 50만 유로의 투자를 한 바 있음 ○ 오픈 데이터 웹사이트(data.gouv.fr)를 통해 프랑스 시민이 데이터를 공유 및 발견이 가능하도록 하고, 정부 기관과 지방 단체, 대학, 연구기관, 비즈니스 등 1만3천여 개의 데이터 셋이 공유되고 있음(2017년 9월 기준, The Innovation Policy Platform, 2017)
핀란드	오픈 데이터	연구 데이터의 관리	○ 2014년부터 교육문화부(The ministry of Education and Culture)가 오픈 사이언스와 연구 이니셔티브를 발표하고, 과학 연구의 선도 역할을 수행하기 위한 로드맵을 수립함(Open Science and Research, 2017) ○ 연구 데이터 관리 역량의 중요성이 대두되면서 교육기관의 오픈 사이언스의식 교육프로그램 운영
미국	오픈 액세스	연구기관 중심의 오픈	○ 2013년 오바마 정부의 과학기술정책국 메모(OSTP Memorandum) 이후 확대

지역	체계	추진	도입
미국	오픈 액세스	연구기관 중심의 오픈	<ul style="list-style-type: none"> <li>○ 공공연구 성과가 디지털 형태로 연구자 및 기업·사회에게 공개</li> <li>○ 오픈사이언스작업반(Interagency Working Group on Open Science)가 구성되어 한시적으로 연방정부차원의 오픈 사이언스정책 방안에 조연함</li> <li>○ 22개의 연방정부기관에서 오픈 액세스 정책 채택하여, 그 중 18개 기관이 지원받은 연구 출판물에 대한 접근을 요구하고 있고, 14개 기관이 연구데이터 관리 계획을 연구제안서 작성단계에서 요청함</li> </ul>
	오픈 데이터	공공연구기관 연구데이터 개방	<ul style="list-style-type: none"> <li>○ 2013년 '오픈 데이터 거버넌스 프로젝트' 등 일련의 이니셔티브 추진하고, 오픈 거버먼트 데이터 플랫폼(Data.gov)을 구축</li> <li>○ 2014년 '오픈 데이터 국가실천계획'을 발표하여 공공연구기관의 연구데이터, 관측 및 조사데이터의 개방의지를 나타냄</li> </ul>
	오픈 콜라보레이션	콜라보레이션 플랫폼	<ul style="list-style-type: none"> <li>○ IT 프로젝트에 별도의 개발 비용 없이 Social coding 웹사이트 'Github'와 같은 플랫폼을 통해 외부 개발자가 코드 및 API를 변형·업데이트 할 수 있도록 함</li> <li>○ Crowdsourcing 온라인 플랫폼을 통해 국민이 수거되지 않은 생활쓰레기를 알리는 등 정부 서비스 향상을 위해 오픈 콜라보레이션 시행</li> <li>○ OpenStreetMap은 커뮤니티 crowdsourcing 프로세스를 통해 시민이 온라인 매핑 프로젝트에 기여할 수 있도록 하는 웹사이트</li> </ul>
일본	오픈 액세스	공공연구 성과 오픈	<ul style="list-style-type: none"> <li>○ 2013년 이후 일본과학기술원(JST)와 일본학술진흥원(JSPS)가 지원된 공공연구 성과 공개 하도록 함</li> <li>○ 일본과학기술원(JST)의 레파지토리 J_Stage는 현재 80% 이상의 일본 학술논문을 공개하고 있음</li> </ul>

자료: 과학기술정책연구원(2017); 신은정 외(2016); Open Science and research(2017); The Innovation Policy Platform(2017); Clark, Brudney & Jang(2013); Budhathoki & Haythornthwaite(2013); Mergel(2015)를 재구성

#### 나. 국내<sup>30)</sup>

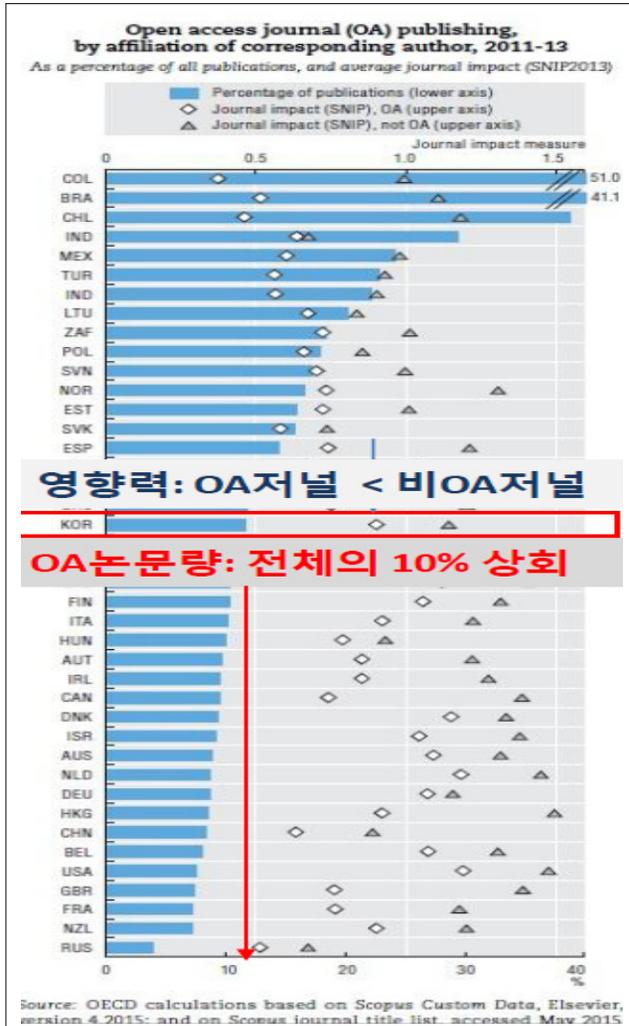
- 오픈 액세스 부문

국내 오픈 액세스 저널(Open Access Journal)이 양적으로 증가하는 추세이다. 국내 연구

30) 신은정(2017), 오픈사이언스 정책 환경 진단 및 시사점

개발 활동 및 논문 생산량이 세계 15위로 증가하였는데, 오픈 액세스 저널의 논문은 전체 논문 수의 약 11%를 차지하고 있다. 특히 국내의 경우에는 오픈 액세스 저널보다 비-오픈 액세스 저널의 영향력이 더 높은 것으로 나타나고, 오픈 액세스 저널보다 기관 레파지토리(Institutional Repository)의 이용 비중이 높은 상황이다.

[그림 4-7] 주요 국가별 오픈 액세스 저널(Open Access Journal) 현황 비교



자료: 신은정(2017), “오픈사이언스 정책 환경 진단 및 시사점”

- 오픈 데이터 부문

연구 데이터의 양과 종류가 증가하고 있지만, 국내 데이터 레파지토리(Data Repository)의 수가 적어 해외의 Data Base를 실시간 이용하는 비중이 큰 상황이다. 특히, 연구 데이터에 있어 국내 연구자들을 대상으로 설문한 결과, 약 63%는 공개한 적 없다고 응답하고 있다. 이러한 배경에는 연구경쟁, 데이터신뢰성, IP, 프라이버시, 업무부담 및 보상 부족 등이 제한요인으로 존재한다고 조사되었다. 국내 연구 데이터 관리 역시 선진국 대비 낮은 수준이다. 연구 데이터 관리와 관련한 업무와 책임이 연구자 개인에게 과중되어 단순 저장 및 관리에 그치는 경향이 있으며, 연구 데이터 관리에 국가적인 지원이 부족한 실정이다. 특히 이러한 데이터 관리 계획(Data Management Plan)에 대한 이해 및 이행 체계 부재로 연구 데이터의 효율적 활용이 어려운 상황이다.

- 오픈 콜로버레이션 부문

ICT 및 디지털 기술을 활용한 글로벌 연구 협력은 기존 연구커뮤니티와 관계망 중심으로 전체 연구의 20%를 차지하고 있다. 최근 20년간 국제협력연구가 정체되었으며, 연구 협력을 위한 온라인 플랫폼 이용이 저조한 실정이다. 이러한 원인으로 ICT·디지털 기술에 대한 문제 보다, 연구협력에 대한 니즈가 부족한 상황과 더불어, 과학기술계가 기업 및 민사회와의 소통이 시작 단계에 있다는 점 등을 지적할 수 있다. 특히 설문 결과 전체 연구 협력 과정 중 온라인 플랫폼을 전혀 이용하지 않는 연구는 전체의 80%로 나타났으며, 온라인 플랫폼 이용하는 20%는 공동 작업과 분석을 위한 이용보다 성과 확산 및 상호 연락을 위해 이용하는 것으로 파악되었다.

- 오픈 사이언스 유관 정책 현황

한국과학기술정보연구원(KISTI)을 중심으로 국가 연구 성과물 등록기탁제도, OAK, NDSL, NTIS 등 연구 성과 및 자료 공유를 위한 정책들이 <표 3-3>과 같이 추진되고 있다. 특히 AI 분야의 경우 <표 3-4>와 같이 민간 부문에서 인터넷 기업들이 주축이 되어 AI 서비스 제공과 관련된 알고리즘 및 데이터를 개방할 수 있도록 준비하고 있다.

〈표 4-3〉 오픈 사이언스 유관 국내 사업 및 법제도 동향

사업	사업 내용	부처/기관
<p>국가연구성과물 등록기탁제도</p>	<ul style="list-style-type: none"> <li>○ 국가연구개발사업에서 산출된 연구 성과물을 국가 지정 DB 및 레파지토리에 등록·기탁, 공개토록 하는 제도</li> <li>- 근거법: 과학기술기본법/국가R&amp;D사업 공동관리규정</li> <li>* 「국가연구개발사업의 관리 등에 관한 규정」 제18조·제25조에 따르면 국가연구개발사업의 9대 주요 성과물**을 미래창조과학부가 지정한 연구 성과물 전담 관리기관에 등록 및 기탁하여 공유하여야 한다.</li> <li>** 논문, 보고서, 특허, 기술정보, 화합물, 소프트웨어, 생명자원(생명정보/생물자원), 신종(농업용 정보/실물), 연구시설 장비</li> </ul>	<p>과학기술 정보통신부/ KISTI 등</p>
<p>OAK(Open Access Korea)사업 (www.oak.go.kr)</p>	<ul style="list-style-type: none"> <li>○ 국내 오픈 액세스를 지원하는 기관 리포지터리에 대한 통합 포털 구축 및 제공</li> <li>- '09. 3 문화체육관광부 국가도서관 지식콘텐츠 창조적 관리 및 확산 사업을 통하여 시작</li> <li>- '14. 4 국립중앙도서관이 인수, OAK(Open Access Korea) 국가지식 정보 구축 및 확산 사업으로 추진</li> <li>다양한 포맷(PDF, HWP, DOC 등 60여개 전자원문 포맷)으로 지식자산화하며, 구글을 포함한 검색포털, OpenDOAR, 기관홈페이지 등을 통해 배포·공유·확산이 가능</li> <li>※ OAK Portal: 국내 기관리포지터리 지식정보의 통합검색 서비스</li> <li>※ OAK Central: 국내 오픈 액세스 학술지의 아카이빙 서비스</li> <li>※ KJCI(Korea Journal Copyright Information): 국내 학회별 학술지 저작권 정책에 대한 정보 제공 시스템</li> </ul>	<p>문체부/ 국립중앙 도서관</p>
<p>국가 R&amp;D 성과물 등록 포털</p>	<ul style="list-style-type: none"> <li>○ 성과물 전담기관의 연구 성과물 등록·기탁 서비스를 종합적으로 안내하고, 등록된 모든 연구 성과물을 한 번에 파악할 수 있는 기능 제공</li> <li>연구자가 자신의 연구과제에서 발생된 국가 R&amp;D 성과물(논문, 특허, 보고서원문, 연구시설·</li> </ul>	<p>과학기술 정보통신부/ KISTI 등</p>

사업	사업 내용	부처/기관
국가 R&D 성과물 등록 포털	장비, 기술요약정보, 화합물, 소프트웨어, 생물 정보, 신제품) 정보를 한곳에서 편리하게 등록· 관리하는 시스템 근거법: 국가연구개발사업의 관리 등에 관한 규 정 제 18조	과학기술 정보통신부/ KISTI 등
NDSL (National Digital Science Library) 사업 (www.ndsl.kr)	○ 국내외 과학기술 관련 논문, 특허, 보고서, 동향 정보 등을 통합 제공하는 플랫폼 구축 및 운영 - 논문·특허·보고서·동향·저널/프로시딩·연 구자·연구기관 등 약 1억 건 이상의 콘텐츠에 대한 검색 및 콘텐츠 유형별 전문검색 서비스 를 제공 국내외 400여개의 협력망 구축을 하여 KISTI 의 기관 소장 정보 자원을 공동으로 활용 가능 전체 1억 건 이상의 국내외 논문, 특허, 보고 서, 동향 등의 콘텐츠를 제공 받을 수 있음 (2017. 10. 23. 기준)	과학기술 정보통신부/ KISTI
NTIS(National Science & Technology Information Service) 사업 (www.ntis.go.kr)	○ 산학연 연구자에게 정보 제공 및 국가 과학기술 혁신 기여 위해 개발된 과학기술정보 서비스 플 랫폼 - 국가연구개발사업 과정에서 산출된 성과물 및 사업관리정보에 대한 정보 DB 구축하여 공개 제공 사업, 과제, 인력, 연구시설장비, 성과 등 국가 연구개발사업에 대한 정보를 한 곳에서 서비스 - 근거법: 과학기술기본법 제 26조, 국가연구개 발사업의 관리등에 대한 규정 제 25조 등 기대효과: ① 국가 R&D 사업 중복 투자 방지 및 투자 효율성 향상 ② R&D 고가장비 중복 구매 최소화 및 국가 차원의 장비 활용도 제 고 ③ 후속연구 및 기술이전·실용화연구·사 업화 추진 기반 제공	과학기술 정보통신부/ KISTI·KI STEP
GSDC사업 (기초연구 실험데이터 글로벌 허브 센터)	○ 국내 글로벌 실험데이터 허브 구축 및 운영을 통해 국내 연구자들이 거대연구시설장비에서 생산된 대용량 데이터를 원활하게 전송, 활용할 수 있도록 지원 - 선진국의 첨단 연구시설의 대형 실험데이터 를 확보·공유함으로써 국내연구자의 데이터 집약형 기초과학연구에 이용할 수 있도록 지원	과학기술 정보통신부/ KISTI

사업	사업 내용	부처/기관
GSDC사업 (기초연구 실험데이터 글로벌 허브 센터)	- 연간 수조원대의 많은 예산이 소요되고 물리적으로 접근이 어려운 선진국의 첨단연구 시설의 실험 데이터 공동 활용 허브 구축을 통한 기초과학분야의 글로벌 경쟁력 확보	과학기술 정보통신부/ KISTI
공공데이터 개방 정책	○ 공공기관 보유, 관리 데이터를 민간에 제공, 활용할 수 있도록 지원하도록 하는 일련의 공공정책 - 근거법: 공공데이터의 제공·이용 활성화에 관한 법률 데이터셋·활용사례·기업탐방 인터뷰·기업 지원 정책정보·개발자 네트워크 등을 제공	안행부/ NIA 등

자료: 신은정(2017) 재구성

〈표 4-4〉 오픈 사이언스 유관 국내 기업 동향

스켈터랩스의 AI플랫폼	○ 데이터 수집 및 AI 기반 플랫폼 출시 예정 <sup>31)</sup> * 카카오브레인과 케이큐브벤처스의 공동투자 (2017. 5. 15 발표) - AI 기반 개인화 플랫폼 기업인 ‘스켈터랩스’가 공공데이터 수집 및 유관 어플리케이션 개발 등을 통한 데이터를 취합 작업 진행 중 기계학습과 자연어처리(NLP) 기술 기반 대화형 AI(챗봇) 플랫폼 개발 중 고객맞춤형 항공권 예약 서비스·주문·지불 등 대규모 데이터 분석이 가능한 플랫폼 기반 POS 시스템 출시 예정	카카오브레인/ 케이큐브벤처스
네이버 딥러닝 데이터센터	○ 딥러닝 학습용 데이터센터 구축하여 알고리즘·사용방법·서비스 적용 결과 등을 통합 제공할 예정 (2017. 6 발표) <sup>32)</sup> 네이버 검색 서비스 통해 축적한 빅데이터와 AI 기술 통합하여 TF(태스크포스)를 구축·프로젝트 진행해 온 경험을 바탕으로 AI 연구개발 생태계 조성 자율주행, 쇼핑검색, 사물인식 등의 AI 서비스 확대	네이버

31) <http://platum.kr/archives/80919>

32) <http://www.greened.kr/news/articleView.html?idxno=30845>

### 3. 국내외 오픈 사이언스 추진 비교

- 오픈 액세스

선진국의 경우 논문 공유 서비스 제공, 오픈 액세스 저널(Open Access Journal), 오픈 액세스 레파지토리(Open Access Repository) 와 함께 오픈 액세스 정책지원 및 촉진을 위한 서비스 및 프로젝트 등이 비교적 활발히 운영되고 있다. 이에 반해 국내의 경우 유력 오픈 액세스 저널이 부재하고, 다양한 오픈 액세스를 촉진하는 정책이 미비하다. 또한 오픈 액세스 저널보다 비-오픈 액세스 저널의 영향력이 더 높은 것으로 분석된다. 특히 오픈 액세스 저널보다 기관 레파지토리(Repository)의 이용 비중이 높다.

- 오픈 데이터

세계적으로 연구데이터 중심 R&D가 활성화되어 데이터 분석이 연구개발의 주요 추진 동력으로 부각되고, 이를 활용한 다분야융합·공동연구가 활성화됨에 따라 연구데이터 공유·활용체계의 필요성이 점증하는 추세다. 이러한 측면에서 연구데이터 개방의 양상은 연구 분야, 연구방법, 관측 장비, 실험장비 등에 따라 다양하며, 규모와 형태적인 측면에서도 매우 다양하다.

주요 선진국은 연구 분야별로 데이터를 체계적으로 관리하는 인프라가 구축되어있다. 예를 들어, 미국의 Big Science, 입자물리, 천체물리, 유전체 등 6개 기초과학분야는 독자적인 연구 데이터 저장 및 분석(독립적인 데이터센터 보유)가 가능하며, Small Science도 연구자 규모 및 연구개발예산 규모가 커서 분야별 대용량 연구 데이터가 존재한다.

또한 선진국의 경우 연구데이터를 공유·활용하는 체계 구축을 위해 노력하고 있다. 미국, 영국 등 주요국들은 논문, 특허 등 최종성과물의 공유를 넘어 국가적 차원에서 연구데이터를 공유·활용하는 체계 구축을 위해 노력하고 있다. 미국 백악관 과학기술정책실 ‘연구데이터 관리 및 공유에 대한 지침’(2013년)에서 원칙을 제시하고, 2016년 빅데이터 연구개발 계획을 발표함에 따라 구체적인 지침을 통해 연구기관 및 연구자들이 개방 및 협력에 참여할 때 따르는 어려움을 해소하고자 했다. 또한 미국의 경우 새로운 발견과 효율적 연구를 위해서 연구데이터의 공동 활용 및 협력을 활발히 진행하고 있다. 대형연구장비 데이터 공유 및 공동연구를 통해 영역 간 융복합의 긍정적인 결과물을 내놓고 있는데, 대표적인 예로 2013년도의 힉스입자 발견과 2016년 중력과 관측의 성공을 들 수 있다. 이외

에도 미국 소재게놈전략(Material Genome Initiative)을 통해 소재 개발기간 및 비용의 1/2 감축을 목표로 하는 등 데이터 및 인프라의 공유를 통해 개방형 연구를 진행하고 있다. 영국은 정보자유법에서 공개 정보의 범위에 데이터셋을 포함시키고, 영국연구위원회의 공동 원칙을 기반으로 공유데이터센터 및 데이터큐레이션 센터를 운영하고 있다. 호주는 '책임 있는 연구수행을 위한 국가지침'에서 연구산출물 관리 규정을 마련하고, 호주국가데이터 서비스를 통해 연구기관들의 메타 데이터를 수집·제공하고 있다

이에 반해 국내에서는 연구데이터 관리·공유와 관련된 체계 및 제도가 부족한 실정이다. 법적 규제 혹은 지침 등으로 연구자의 오픈 사이언스 참여를 강제 혹은 독려하는 체계 역시 미비한 상황이다. 예를 들어 국가 연구 성과물 등록기탁 제도의 대상인 국가 R&D 사업의 9대 주요 성과물에 연구 데이터가 포함되어 있지 않다. 따라서 연구 성과물 공개 항목 중 연구 데이터가 포함되지 않아 데이터의 재활용 가능성, 양질의 연구 성과물 개방 등의 측면에서 부정적인 영향을 미친다. 또한 연구자 규모와 연구개발예산 규모가 상대적으로 작아 연구데이터 인프라의 효율적 운영이 불가능하다. 이에 따라 연구 데이터 인프라, 특히 데이터 레파지토리의 수가 작고, 이로 인하여 연구자 개인이 데이터를 공유하는 과정에 어려움을 겪고 있다.

- 오픈 콜라보레이션

선진국의 연구자들은 공적 자금으로 편당을 받는 R&D 결과물에 대한 소유권을 개인의 자산으로 인식하지 않는 데 비해, 국내 연구자는 연구 데이터를 개인자산으로 인식하여 개방 및 공유에 소극적인 것으로 보인다. 또한 선진국은 연구 문화 측면에서 글로벌 연구 협력이 일상화되어 있으며, 글로벌 협력을 위한 ICT 및 디지털 기술을 적극 활용하고 있다. 이에 반해 국내의 국제 연구협력 네트워크는 저조한 수준이며, 연구자의 온라인 플랫폼 이용이 매우 낮은 상황이다. 이외에도 협력의 측면에서 연구 공동체 형성이 미흡한 것으로 보인다.

#### 4. 국내 오픈 사이언스 정책 관련 시사점

- 오픈 액세스

연구자의 연구 성과물 공개를 강제 혹은 독려하는 체계를 수립·강화해야 한다. 예를 들

어 기관 레파지토리 중심으로 인센티브를 제공하여 연구자 혹은 연구팀의 자발적인 참여를 유도할 수 있다. 기본적으로 연구 분야에 적절한 오픈 액세스 저널, 오픈 액세스 레파지토리가 존재해야 하고, 여기에 기관 레파지토리를 중심으로 인프라와 인센티브를 강화한다면 동일한 분야의 연구 성과물 관람 및 참고가 가능하여 각 연구에 긍정적인 영향을 미칠 것이다. 특히 오픈 액세스 기여에 대한 경력인정 및 보상 지원 체계를 수립하여 적절한 인센티브를 제공한다면, 연구자들의 자발적 참여를 유도할 수 있을 것으로 전망된다.

- 오픈 데이터

국가적 차원에서 연구데이터 공유 집단 또는 연구 공동체의 형성을 촉진하기 위해서는 인프라 지원 및 국가 연구데이터 허브의 구축이 요구된다. 다양한 소규모 연구 그룹, 즉 분야별 연구공동체 형성을 위한 방안을 마련하고 이와 동시에 클라우드 방식의 연구 데이터 관리 지원을 통해 소규모 연구그룹의 연구 데이터 생성·공유·활용·협력을 활성화해야 한다. 또한 다양한 연구 데이터의 관리 수요에 대응하여 연구자 및 연구 공동체의 관리 부담을 최소화하기 위한 연구 데이터 허브가 필요하다.

연구 데이터 관리 측면에서는 ‘연구 데이터 관리 계획(Data Management Plan)’을 마련하고, 연구 데이터 공개 인센티브 제공, DOI(Digital Object Identifier)<sup>33)</sup> 등의 확대 적용과 더불어 후속 활용을 지원하는 데이터 품질관리 서비스도 확대 지원해야 한다. 또한 데이터 및 문헌정보 관리기관이 분산되어 있어 효율적 관리가 어렵기 때문에 관리기관에 대한 정책 혹은 지침 등도 필요할 것으로 보인다. 더 나아가 연구개발(R&D)의 성공과 실패를 포함한 연구데이터를 빅데이터화하여 AI을 적용·활용함으로써 R&D 성과를 제고하는 노력도 필요하다.

이와 관련하여 최근 과학기술정보통신부는 2017년부터 후반부터 과학기술 분야의 혁신 성장 촉진을 위한 「연구데이터 공유·활용 전략」수립을 준비 중에 있다. 이러한 「연구데이터 공유·활용 전략」은 국가 R&D 추진 과정에서 축적되는 연구 데이터를 체계적으로 관리·공유하고 연구자가 새로운 지식과 가치 창출에 연구데이터를 적극 활용토록 지원하기 위해 마련됐다. 과학기술정보통신부는 「연구데이터 공유·활용 전략」을 통해 ① 연구데이

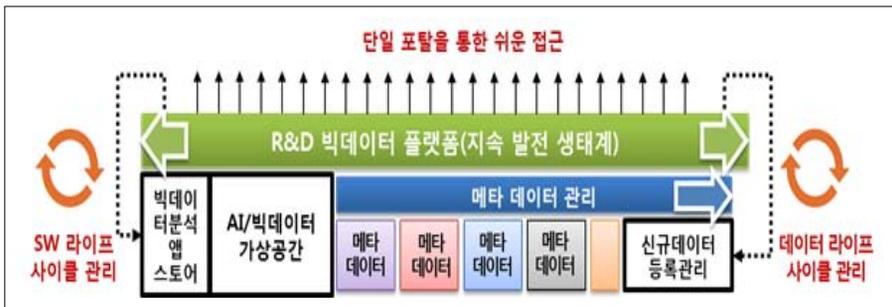
---

33) 객체에 부여되는 고유한 디지털 식별자로서 편리하게 항구적인 접근과 인용을 보장하는 수단

터 관리체계 구축 및 커뮤니티 형성 촉진, ② 국가연구데이터 플랫폼 구축 및 서비스 제공, ③ 데이터 및 컴퓨팅 활용 인재 성장 지원, ④ 연구데이터 공유·활용에 관한 법·제도 마련, ⑤ 연구데이터의 산업적 활용 및 일자리 창출 등 과제를 추진할 계획이다.

이중 연구데이터 관리체계 구축 및 공유 커뮤니티 형성 촉진과 국가연구데이터플랫폼 구축 및 서비스 제공에 대하여 구체적으로 살펴보면 다음과 같다. 우선 연구데이터 관리체계 구축 및 공유 커뮤니티 형성 촉진의 경우 첫째, 관리체계 측면에서 연구데이터의 체계적인 관리를 위해 국가연구데이터센터 및 과학기술 분야별(대분야·소분야) 전문센터<sup>34)</sup> 체계를 구축한다. 국가연구데이터센터는 관리체계 총괄로서 국가연구데이터플랫폼 운영 담당, 분야별 전문센터는 데이터 취합·관리 및 활용촉진 등을 담당한다.

[그림 4-8] 국가 연구데이터플랫폼 개념도



자료: 과학기술정보통신부(2017), “연구개발(R&D) 성과제고를 위한 연구데이터 공유방안”

둘째, 커뮤니티 형성 촉진 측면에서는 대규모 연구데이터를 구축·활용하는 집단연구 과제를 신설하고, 기 구축된 데이터를 활용한 소규모 R&D 과제도 지원한다. 신설되는 집단연구과제의 수행기관으로는 연구데이터 전문센터를 지정한다.

국가연구데이터플랫폼 구축 및 서비스 제공의 경우, 첫째, 통합 플랫폼 측면에서 연구데이터 공유·활용체계 전반의 연구데이터를 윈스톱으로 검색하고 활용할 수 있는 통합 플랫폼을 구축한다. 이러한 통합플랫폼을 통해 사용자에게 연구 데이터 융합 및 분석을 실

34) 바이오, 미래소재, 대형연구장비, AI 등 데이터 집약형 연구 분야부터 우선 전문센터 지정을 추진하고 단계적으로 확산

시할 수 있는 가상 작업환경과 데이터 분석도구를 제공한다. 둘째, 인프라 서비스 측면에서는 전문센터에 클라우드 형태로 분야별 컴퓨팅 및 데이터 인프라를 구축하고, 개별 연구자에게 클라우드를 통해 데이터관리가 가능한 환경을 제공한다.

[그림 4-9] 계층 구조의 데이터 공유·활용 생태계 개념도



자료: 과학기술정보통신부(2017), “연구개발(R&D) 성과제고를 위한 연구데이터 공유방안”

### 제 3 절 오픈 사이언스 기반의 AI 활성화 방안

#### 1. 오픈 사이언스와 AI

사회가 요구하는 과학기술이 점차 복잡해지고 어려워짐에 따라 효율적인 연구개발과 혁신적인 발견을 위해서 연구 성과물의 공개·공유와 확산·활용이 꼭 필요해지고 있고, 이에 더하여 연구 성과물을 공동으로 활용하고 협력하여 연구하는 것이 중요해지고 있다.

주요 선진국은 이와 같은 변화에 주목하고 연구 성과와 과정을 개방화하는 오픈 사이언스 정책을 본격 추진하고 있으며, 바이오, 소재, 천문, 물리·우주 등 다양한 과학기술 분야에서 관련 관리체계 및 커뮤니티가 형성되어 이를 기반으로 생태계가 활성화되고 있다.

최근에는 이 중에서 AI 분야가 가장 활발하게 관리체계 및 커뮤니티가 형성되고 있으며, 이에 따라 연구 성과물이 빠르게 공개·공유되어 확산·활용되고 있다. AI 분야의 연구 성과가 빠르게 공개·공유·확산·활용되는 원인은 글로벌 선도 ICT 기업, 공공조직·대학들을 중심으로 지능정보 핵심기술인 알고리즘의 오픈소스화, 공개 학습 데이터, 트레이닝 및 테스트베드 환경 등을 제공하고 있기 때문이다.

이는 AI 분야에서 기존 오픈 사이언스의 개념이 적극적으로 실현되고 있음을 말해준다.

오픈 사이언스의 오픈 소스는 AI 분야에서 알고리즘의 오픈소스화를 통해 혁신 생태계 조성에 기여하고 있으며, 오픈 데이터의 경우에는 AI에 활용될 수 있는 공개 학습 데이터 셋의 공유 및 공개를 통해 개발 생태계 확산에 기여하고 있다.

〈표 4-5〉 오픈 사이언스와 AI 생태계 비교

체계	Science community	AI ecosystem
Open Source	하드웨어 및 소프트웨어의 코드 공개 및 자유로운 이용 허용	알고리즘을 소프트웨어적으로 구현하여 플랫폼화(모듈화)한 라이브러리 및 이를 사용할 수 있게 하는 인터페이스 등으로 구성되어 있는 AI 플랫폼을 개방
Open Data	연구데이터의 개방과 공유	AI 관련 다양한 이해관계자가 쉽게 활용할 수 있는 접근과 사용이 용이한 대규모 데이터셋 공유·공개
Open Collaboration	개방형 연구의 소통과 협력	AI 관련 다양한 이해관계자가 공동 협업을 통한 혁신 추구

결론적으로 오픈 사이언스의 개념과 오픈 사이언스가 추진하고자 하는 방향은, AI 생태계가 활성화되면서 이루어지고 있는 모습과 일치하며, 동일한 맥락으로 파악될 수 있다.

• AI 플랫폼 개방 및 알고리즘 오픈소스

현재 글로벌 ICT 기업인 구글 뿐 아니라 페이스북, 마이크로소프트, OpenAI, IBM 등이 AI 플랫폼을 오픈소스로 공개하면서 자사 주도의 AI 혁신 생태계를 조성하고자 노력하고 있다.

〈표 4-6〉 주요 글로벌 ICT 기업의 AI 알고리즘 공개 현황

기업	주요내용
구글	<ul style="list-style-type: none"> <li>○ 딥러닝(Deep Learning)을 위한 AI SW 라이브러리 텐서플로우(Tensorflow)를 2015년에 아파치 라이선시(APL)의 오픈소스로 공개</li> <li>○ 기계학습에 최적화된 연산처리장치인 TPU(Tensor Processing Units)를 발표하였으며, 세계 최대 규모의 데이터 센터 운영</li> </ul>
마이크로소프트	<ul style="list-style-type: none"> <li>○ 인공신경망(Neural network) 학습 소프트웨어 '코그니티브툴킷(CNTK)'의 새로운 버전인 '마이크로소프트 코그니티브툴킷(Microsoft Cognitive Toolkit)'을 오픈소스로 공개</li> </ul>

예를 들어 현재 가장 주목받는 AI 기술은 기계학습(머신러닝: 딥러닝, 강화학습)으로 이를 오픈 플랫폼으로 누구나 활용할 수 있도록 개방하고 상호 활용하는 추세이다. 알고리즘을 소프트웨어적으로 구현하여 플랫폼화(모듈화)한 라이브러리 및 이를 사용할 수 있게 하는 인터페이스 등으로 구성되어 있는 AI 플랫폼을 개방하여 제공하고 있다. 이와 같이 기계학습(머신러닝)의 복잡한 수학적·통계학적 알고리즘을 소프트웨어 플랫폼(모듈)으로 구현하여 개방하면, 이를 사용하는 이용자가 그 자세한 내부 로직을 알지 못해도 직관적으로 사용이 가능하다.

현재 딥러닝 알고리즘 구현 및 학습을 위해 Caffe, TensorFlow, Torch 등 오픈소스 프레임워크(개발툴킷)가 제공되고 있다. 이러한 각 프레임워크는 파이선 지원여부, 구현 난이도, 추상화 정도, 속도, 학습모델 제공여부, 학습데이터 제공여부, GPU 지원여부 등 기능 및 성능 측면에서 장단점이 있으므로 목적에 알맞은 개발툴을 활용하는 것이 중요하다.

• AI 알고리즘 오픈 데이터

AI의 성능은 활발히 공개되는 알고리즘보다 데이터의 양질에 더 의존적이다. AI(기계학습)의 활용과 적용은 결국 이러한 데이터 구축 및 관리에 기반을 둘 수밖에 없으며, 양질의 데이터는 AI(기계학습)의 성능을 좌우하는 가장 중요한 요소이다. 이러한 학습용 데이터의 경우에도 구글, 페이스북 등 선도기업을 중심으로 스탠퍼드, 토론토 대학 등 연구기관과 협력하여 동영상 이해, 텍스트 이해 등 AI 연구개발 생태계 확산을 목적으로 대규모 데이터 셋을 공개하고 있다.

〈표 4-7〉 AI 데이터셋 공개 현황

구분	내용
MNIST Handwritten Digit (손글씨 이미지)	0~9까지 수기 숫자 (70,000장)
ImageNet (사물 이미지)	천만장 이상의 웹상 이미지에 대한 레이블링 데이터셋
Open Street Map (지도 데이터)	전 세계의 지도 정보를 파일로 제공 (785GB)
Amazon Co-Purchasing (제품 구매 데이터)	아마존 웹사이트의 구매 정보 50만건, 함께 구매한 상품 리스트 및 제품 구매 정보 제공
LibriSpeech (음성 데이터)	100시간 분량의 음성 데이터 / 자막 데이터 제공

## 2. 오픈 사이언스 관점의 AI 활성화 방향

국내 AI 산업은 시장형성 단계로 일부 대기업 및 ICT 기업에서 AI 연구에 투자하고 있으나 아직은 초기 단계에 이르고 있다. 또한 연구개발 투자규모는 선진국 대비 상대적으로 매우 낮은 수준이다. 이외에도 AI 생태계 구성원들은 데이터 기반의 부족을 느끼고 있으며, 데이터 공급-수요 불균형을 주요 이슈로 제기하고 있다. 특히 국내 중소·벤처 AI 업체들은 데이터를 잘 갖추지 못하고 있을 뿐만 아니라 기업 간 데이터 패권 경쟁 심화, 개인정보보호법 등에 의해 추가적인 데이터를 확보하기가 어려운 현실이다.

이러한 국내 AI의 문제점을 오픈 사이언스 관점에서 살펴보고, 이를 통해 국내 AI 생태계를 활성화하기 위한 방향을 모색해 보면 다음과 같다.

첫째, 법적, 경제적, 기술적 장벽 없이 AI 연구 성과물을 다양한 이해관계자가 공유할 수 있는 환경 구축을 구축한다면, AI 기술의 확장과 활용에 기여할 것으로 보인다. 구체적으로는 경제적 장벽(라이선스 요금, 이용 요금, 회비 등)인 접근 비용 문제에 관련된 노력이 필요하다. AI 오픈 액세스 저널 혹은 AI 분야의 데이터 레파지토리를 구축하여 스타트업 및 개인 개발자에게 제공할 수 있다.

동시에 법률적 문제(저작권, 라이선스 등)인 이용권한 문제 해결에 노력을 통해 사용자들의 이용 장벽을 낮출 수 있다. 개인정보 등이 포함된 데이터에 대한 제한 완화, 민감정

보 이용 확대를 위한 규제 완화 등이 필요하다. 또한 기술적 장벽(영구적인 접근 보장, 표준화 등)을 낮춰 접근성 문제 해결을 위한 노력이 필요하다. CKAN<sup>35)</sup>을 활용하여 데이터 셋 공유, 연계, 관리, 활용을 위한 표준 지원 등을 할 수 있다.

둘째, AI 관련 다양한 이해관계자가 쉽게 활용할 수 있는 접근·사용이 용이한 오픈 데이터 셋의 공유 및 기반 인프라를 제공해야 한다. 사용자 중심의 오픈 데이터 공개 및 공유를 위해 사용자 친화적인 플랫폼 혹은 공유 기반 인프라 제공이 중요하다. 특히 다양한 이해관계자들이 필요로 하는 데이터 셋을 찾아주거나, 활용하기 적합한 형태로 변환(분석 및 가공)할 수 있는 환경을 제공해야 한다. 또한 공유되는 데이터 셋 포맷(CVS, API, Linked open data 등)을 기계 판독성이 높은 형식으로 제공한다면 사용자가 분석 및 가공 과정에서 도움이 될 것이다. 더불어 허브 및 플랫폼 사용자 그룹이 협력할 수 있는 환경이 제공된다면 데이터 셋의 품질 향상의 측면에도 기여할 것으로 보인다. 공동 협업으로 혁신을 추구할 수 있는 환경 구축이 되어 challenge.gov, Kaggle 등과 같은 공동 협업 연구 또는 비즈모델도 기술 및 비즈니스 혁신 장려 측면에서 긍정적인 영향을 미칠 것이다.

특히 오픈 데이터의 확대 적용은 4차 산업혁명에 대응하여 AI 기술 분야에 중장기적인 영향을 미칠 것으로 보인다. AI 분야는 알고리즘과 데이터의 양질에 의해 기술이 완성되는데, 오픈데이터 이행을 통해 과학자 중심의 학계는 물론 AI 관련 스타트업·글로벌 ICT 기업 등 다양한 이해관계자로 참여 범위를 확대하여 해당 산업계의 지원도 가능할 것으로 보인다.

### 3. AI 활성화 방안으로써 디지털 플랫폼

오픈 사이언스 관점에서 AI 생태계 활성화를 위한 구체적 방안으로, AI 기술 개발 지원을 위한 디지털 플랫폼(조직 및 디지털 구성요소 기술을 포함)을 제시할 수 있다. 오픈 사이언스를 AI 분야에 적용하여 AI 생태계를 활성화하기 위해서는 오픈소스 및 오픈 데이터를 통한 공유·제공·지원 등의 다양한 활동을 수행할 수 있는 디지털 플랫폼이 절대적으로 필요하다.

---

35) CKAN(Comprehensive Knowledge )이란 데이터 개방을 위한 오픈소스 데이터 플랫폼이다. 영국, EU 뿐만 아니라 세계 각국 공통 분야의 데이터 개방을 지원한다. 사용자가 오픈소스를 활용해 데이터 개방 위한 시스템을 자체적으로 구축할 수 있다.

특히, 국내 AI 분야는 민간 디지털 플랫폼 공급자의 수가 충분하지 않기 때문에 중소·창업(벤처)기업이 AI 기술을 개발할 수 있는 환경이 열악한 실정이다. 따라서 AI 개발을 위한 오픈소스 및 공공 데이터 공유의 지원 사항(기술적인 구성요소, 가이드 및 양식 제공, 기술 및 품질 지원, 교육 등)을 제공하는 디지털 플랫폼이 요구된다. 특히, 보유한 공공 데이터를 빠르고 쉽게 공유할 수 있는 온라인 디지털 플랫폼이 필수적으로 요구된다.

- 디지털 플랫폼의 이해

최근 Gartner(2016)는 기존 플랫폼과 비교되는 디지털 플랫폼을 “파트너, 공급 업체 및 고객 커뮤니티가 사업적인 이익을 위해 디지털 프로세스 및 역량(digital processes and capabilities)을 공유·개선·확장할 수 있는(share·enhance·extend) 비즈니스 중심 프레임 워크”로 디지털 플랫폼을 정의하고 있다. 디지털 플랫폼이 실제 비즈니스에 있어 가지는 의미는 『디지털 역량(Digital Capability)을 바탕으로 변화의 동인(Enabler)을 포착하여 고객의 문제(Customer)를 해결하는 것』에 있다. 즉 새롭게 성장하는 디지털 플랫폼은 디지털 역량, 변화의 동인, 고객의 문제 등 3가지 영역의 교집합에 존재한다는 것이다.<sup>36)</sup>

우선 디지털 플랫폼은 AI, 클라우드, 빅데이터, IoT, 모바일 같은 새로운 디지털 요소기술(Enabler)에 대한 전문성과 더불어 해당 산업의 제품(서비스)·프로세스에 대한 전문성도 요구된다. 또한 이를 바탕으로 하는 차별화된 디지털 프로세스 구축으로 디지털 전환(Digital Transformation)을 수행할 수 있는 디지털 역량이 확보되어야 한다.

- AI 디지털 플랫폼으로서 『AI 오픈 이노베이션 허브』 방안

국내는 중소·창업/벤처기업이 다양한 AI 응용서비스를 만들어 낼 수 있는 AI 개발 환경이 부족하여, AI 개발을 지원할 수 있는 자원의 제공이 필요하다. 이에 과학기술정보통신부 2017년 말 국내 AI 중소·창업(벤처)기업을 대상으로 AI 개발 핵심 인프라를 지원하는 ‘AI 오픈 이노베이션 허브’를 준비하고 개방하였다.

AI 오픈 이노베이션 허브는 AI의 응용서비스 개발에 필수적인 ① 다양한 양질의 AI 학습용 데이터<sup>37)</sup>, ② AI 오픈 API(응용프로그램 인터페이스)<sup>38)</sup>, ③ 고성능 컴퓨팅 파워 등을 국

---

36) Laurie, Donald L., Yves L. Doz, and Claude P. Sheer. “Creating new growth platforms.” harvard business review 84.5 (2006): 80-90. 내용을 재구성)

37) 머신러닝/딥러닝에 활용할 수 있는(Machine Readable) 형태의 데이터 필요

내 중소기업이 활용할 수 있도록 AI 개발 환경을 지원하는 인프라이다.

우선 AI 경쟁력의 핵심인 데이터 셋은 중소기업에서 구축하기에는 많은 시간과 비용이 소요되어 정부에서 민간 활용도가 높은 법률, 특허, 한국어 위키백과, 한국인의 얼굴 및 한국 음식 이미지 등 총 4종의 데이터 셋을 우선 구축·개방하였다. 또한 국내 민간 및 공공분야에서 보유한 데이터들에 대한 관련 소재 정보(제공사이트, 종류, 규격, 메타정보 등) 1,000여개를 제공하여 AI 데이터로의 활용을 촉진한다.

〈표 4-8〉 AI 학습용 데이터 셋 구축 현황

분야	활용 데이터 종류	예상 서비스
법률	국가법령 중 교통사고, 층간소음, 창업 인허가 분야의 관련 법령, 조문, 판례, 법률 용어 등 11여만건	법령정보 추천 서비스, 법률상담 챗봇 등
특허	국내 출원·등록된 전기·전자분야의 특허정보, 특허도면, 심사정보, 특허전문기술용어 등 110여만건	특허 유망기술 예측, 특허가치 평가 등
일반상식	한국어 위키백과에서 활용도가 높은 일반상식 15여만건	AI 비서, 교육용 챗봇 등
한국 이미지	한국인 안면 이미지 600여만장(200명), 한국음식 이미지 15여만장(150종)	본인확인 기술, 음식검색·식단추천 등

자료: 과학기술정보통신부 (2018), “인공지능(AI) 개발 핵심 인프라를 중소기업에게 지원”

또한 정부 국책과제로 지원하여 개발된 AI 기술을 보다 쉽고 효율적으로 AI 제품·서비스 개발에 활용할 수 있도록 AI SW 결과물을 오픈 API(응용프로그램 인터페이스) 형태로 제공한다. 한국어 분석(11종), 음성처리(3종) 등 총 14종의 오픈 API를 개방하였고, 2018년 하반기에는 한국어 질의응답(1종) 및 대화처리(1종), 외국어 음성인식(6종), 영상이해(1종) 등 총 9종을 추가 공개할 예정이다.

이외에도 컴퓨팅 환경이 상대적으로 열악한 중소기업에게 AI의 핵심 기술인 딥러닝과 같은 방대한 계산을 빠르게 처리하여 신속한 제품·서비스를 개발할 수 있도록 GPU(그래픽처리장치)<sup>39)</sup> 기반의 고성능 클라우드 컴퓨팅 자원도 지원한다.

38) API(Application Programming Interface): 특정 프로그램을 레고 블럭처럼 재조합하여 활용할 수 있도록 제공하는 프로그래밍 인터페이스

39) GPU(Graphic Processor Unit)는 수천 개의 소형 코어로 구성되어 딥러닝과 같은 분야에 방대한 데이터를 효율적으로 병렬처리가 가능

## 제5장 결론 및 정책 제언

최근 급속도로 진행되고 있는 지능정보기술의 발달은 과거 증기기관이나 전기와 같이 사회 전반에 걸쳐 광범위한 파급력을 가지고 혁신을 유발하면서 4차 산업혁명을 이끌 핵심 기술로 받아들여지고 있다. 지능정보기술은 알고리즘의 변형과 확장 및 다양한 유형의 데이터 학습을 통하여 적용 분야가 지속적으로 확대되고 있으며, 범용기술의 지위를 가지고 다양한 산업과 융합하여 국가경제의 혁신 동력으로 작용하여 산업구조의 변화와 사회·제도의 변화를 견인하고 있다. 특히, 지능정보기술 중 AI SW 기술은 지능화를 이끄는 핵심기술로서 생산성과 효율성을 획기적으로 높이는 코어(core) 역할을 하며, 산업·기업·기술 간 경계를 소멸시키는 융합화를 주도하고 있다. 이에 글로벌 기업은 AI 기술의 핵심역량을 확보하는데 주력하고 있으며, 관련 시장을 선점하기 위해 서비스 혁신과 생태계를 구축하는데 많은 노력을 기울이고 있다. 또한, 각국 정부는 국가 주도의 AI R&D 전략을 추진하는 등 AI 연구개발을 확대하고 시장을 활성화하기 위한 지원책 마련에 노력을 기울이고 있는 실정이다. 하지만 국내의 경우 AI을 4차 산업혁명의 핵심기술로 인지하고 있으나 선도국과 비교하여 AI R&D의 투자규모가 부족하고 차세대 전략분야를 발굴하고 투자하는데 미흡하다는 평가를 받고 있다. 이에 본 장에서는 AI R&D의 선도형(first mover) 연구 촉진과 성과 확산을 가속화하기 위해 필요한 정책 방안을 제시하기로 한다.

AI 기술은 대규모 데이터에 대한 자가 학습을 통해 지속적으로 알고리즘 성능을 강화하므로 데이터가 산업의 주요 경쟁 원천이라고 할 수 있으며, AI 기술의 발전을 위해서는 무엇보다 우선해서 데이터를 확보하는 것이 중요하다고 할 수 있다. 하지만, 국내의 경우 AI 관련 데이터의 제공 체계가 확립되어 있지 않고, 연구데이터의 관리·공유·활용 체계 및 제도가 갖춰져 있지 않은 실정이다. 먼저, 효율적이고 효과적인 AI R&D를 위해서 데이터 셋의 확보가 필요하다.

이를 위해서는 기술별·분야별로 필요한 데이터 셋을 파악하고 우선순위에 따라 전략적인 투자를 통해 데이터 셋을 확보해야 한다. 본 연구에서 실시한 기계학습 데이터 수요조사에 따르면 AI 기술 분야별 데이터 셋 우선순위는 추론 및 기계학습, 지식표현/언어지능,

시각지능, 지능형에이전트의 순서로 나타났으며, 분야별로는 의료, 금융, 국방, 제조, 공공 서비스 및 복지 등의 순서로 나타났다. 보다 정밀한 수요조사를 통해 산업계의 니즈를 파악하고, 이를 충족시킬 수 있도록 우선순위 따라 전략적으로 데이터 셋을 수집·가공하여 제공할 수 있는 방안 마련이 시급하다.

국내의 경우 데이터 셋을 확보하는 것뿐만 아니라 데이터와 관련한 제도, 인프라, 활용의 전 측면에서 그 기반이 매우 약하다고 할 수 있다. 현재 연구데이터는 연구성과나 산출물로 관리되고 있지 않으며, 관련 규정 또한 존재하지 않는 등 연구데이터의 관리·공유·활용 체계가 미비하고, 데이터를 분석할 수 있는 연구그룹 및 인력 양성에 대한 정책 또한 부족한 편이다. 따라서 AI R&D의 성과를 제고하기 위해서는 다음과 같은 방안이 필요하다.

첫째, 연구데이터를 효율적으로 관리하고, 이를 공유하는 것을 제도화해야 한다. 이를 위해서는 연구데이터의 관리, 공유, 활용, 지원 등을 위한 규정을 정비하는 등 단계적인 입법을 추진할 필요가 있다. 단기적으로는 국가연구개발사업 등의 성과평가 및 성과관리에 관한 법률 혹은 공동관리규정 등에 연구데이터를 연구성과물로 공식 지정하여 국가 차원의 체계적인 데이터 수집·관리 및 공유·활용 체계 기반을 구축하여야 한다. 또한, 장기적인 관점에서 국가 연구데이터와 관련한 법제를 개선하고 정비해야 한다.

둘째, 분야별로 데이터를 공유하고 활용하는 연구커뮤니티를 형성해야 한다. 앞서 제한한 것과 같이 데이터셋의 우선순위를 파악하여 연구데이터가 실질적으로 공유·활용되는 소분야의 연구 공동체를 선정하고, 공통적으로 사용되는 데이터의 관리 및 분석 SW 등을 개발하는 것을 지원해야 한다. 뿐만 아니라 성과 확산을 위해 현재 가용 가능한 데이터를 활용하는 소규모 연구를 지원하고, 일정기간 동안 예산과 컴퓨팅 파워 및 네트워크 자원을 함께 지원하는 것이 필요하다.

셋째, 국가 연구데이터 플랫폼을 구축하고 이를 제공하는 서비스를 지원해야 한다. 현재 이용가능한 공공 데이터의 경우 여러 기관에서 각기 다른 플랫폼 하에 데이터가 제공되는 등 제공 체계가 상이하여 활용이 용이하지 못하다. 이를 해결하기 위해 여러 종류의 연구데이터를 동일한 방법으로 접근할 수 있는 연구데이터 플랫폼을 구축하고 운영할 필요가 있다. 연구데이터 플랫폼은 데이터의 제약 없는 접근이 가능해야 하며, 신규 데이터의 등록 및 데이터 간 융합 및 활용이 가능한 환경을 제공해야 한다. 또한, 데이터를 관리하고 분석할 수 있는 표준화된 SW 개발 등의 기술지원 및 연구자간 협력 등을 장려하는 것이 중요하다.

마지막으로 데이터와 컴퓨팅 활용을 위한 연구개발인력을 양성해야 한다. 대학생·대학원생 및 현직 연구자들을 대상으로 전문분야에서의 데이터와 컴퓨팅 활용을 위한 교육 프로그램을 제공하고, 인력육성 인프라 및 원격연구 환경으로서 분야별 데이터·컴퓨팅 활용 플랫폼과 교육용 데이터 뱅크를 구축해야 한다. 또한, 데이터 기반 연구 참여를 통한 고급 인력 양성을 추진하고, 데이터 기반 연구 및 전문센터의 증가와 함께 수요가 크게 증가할 것으로 예상되는 ICT 기반 관리·운영·서비스 인력을 양성해야 한다.

AI 기술의 발전에는 다량의 데이터를 활용한 분석이 필수적인 요소로 작용하기 때문에 앞서 제안한 여러 가지 방안들을 통해 양적·질적으로 향상된 데이터 시장의 활성화가 매우 중요하다고 할 수 있다.

AI 분야의 선도형 R&D를 촉진시키고 성과 확산을 유도하기 위해서는 데이터 시장의 활성화 외에도 우리나라 연구개발 분야의 보다 근본적인 문제점인 폐쇄적 R&D 체계를 우선해서 해결해야만 한다. 국내 AI R&D 생태계는 연구 성과·데이터의 공개·공유를 통해 연구개발의 효율성을 높이는 오픈 사이언스 문화의 확산이 절대적으로 부족한 편이다. 국내 정보통신분야 연구자 중 오픈 액세스 저널을 인지하지 못하거나 활용하지 않는 비율이 63%이고, 연구데이터의 공개 경험이 없는 비중이 72%에 달한다(신은정·정원교, 2017). 또한 AI 기술과 관련한 국가 R&D 사업의 산출물인 데이터 및 소스코드의 공유·공개가 부족하고 경쟁형 R&D나 오픈소스 공동개발 등의 연구 협업은 미흡한 편이다. 국내와는 달리 해외는 연구 성과 및 데이터 공개 제도, 도전형 R&D 등 개방·협업 R&D를 촉진하는 제도<sup>40)</sup>를 운영 중이나 국내는 관련 제도가 미비하다.<sup>41)</sup>

이러한 문제점을 개선할 수 있는 방안은 오픈 사이언스 문화의 확산을 장려하는 것이다. 국내의 경우 현재로서는 아직 오픈 사이언스를 위한 공유 체계가 미흡하다. 가장 부족한 부분은 오픈 사이언스를 위한 법제도와 정책이다. 일부 분야(생명, 화합물 등)의 연구 데이터는 이미 기탁하도록 되어있으나 실제 공개율이 저조하고 대상 분야가 한정적이다. 따라서 국가 연구개발사업 수행 후에 생산된 연구 데이터를 제출해야 하는 근거법이 다양한 분야

---

40) 대표적인 예로 미국의 국방성 산하기관인 DARPA의 협업형 R&D 체계와 도전형 R&D를 추구하는 DARPA Challenge가 있다. 이와 관련한 자세한 내용은 참고자료를 참조

41) 과학기술정보통신부에서 도전형 R&D를 촉진하기 위해 2017년에 AI Challenge를 시행

로 확대되어야 한다. 최근 연구 성과물 공개를 요청하는 사회적 요구와 이를 반영하는 세 부 사업 규정들의 증가로 인해 연구자들은 정부로부터 재정지원을 받아 연구개발사업을 수행하는 경우에 별도의 사업단 혹은 레파지토리 등에 성과물을 등록하고 있다. 하지만 공동관리규정에 의거한 연구 성과물 공개 정책과 개별 부처 및 기관에서 추진되는 성과물 공개 정책이 독립적으로 진행되어 상호 연계되지 못하고, 성과물을 관리하는 기관과 DB도 산재되어 있는 등의 문제점이 여전히 존재한다. 이러한 분산적 관리 체계는 연구 성과물을 관리에 있어 비효율적이며, 성과물을 이용하는 연구자들이 공개된 연구 성과물을 활용하는 것을 어렵게 만든다. 따라서 현재 운영되는 국가 연구개발사업 관련 연구 성과물 공개 정책을 재정비하고, 각 부처별, 기관별로 독립적으로 운영되고 있는 레파지토리나 DB 등을 연계시킬 수 있는 방안 마련이 시급하다.

이뿐만 아니라 연구 데이터에 대한 소유권은 누가 가지는 것인지 명확한 기준이 마련되고, 데이터 공개에 대한 보상체계 등 정책이 수립되어야 한다. 이를 통해 연구자들이 연구 데이터를 관리하고 공유할 동기가 부여된다. 현재 연구 성과물을 공개할 경우, 연구자들이 체감할만한 인센티브는 부족한 상황이다. 이러한 상황에서 연구자들의 성과물 공개를 강제할 경우, 공유·공개되는 성과물의 양은 늘릴 수 있으나 질은 현저히 낮아질 수밖에 없다. 따라서 적절한 보상체계를 마련하여 자발적으로 연구 성과물을 공개할 수 있도록 해야 한다.

다음으로 필요한 사항은 오픈 사이언스를 위한 지원 조직과 플랫폼이다. 연구자들은 공공지금으로 수행되는 연구개발 과제에서 이미 수많은 시간을 행정 처리에 할애하고 있다. 앞으로 오픈 사이언스를 위해서는 공유 계획서 작성, 공유 대상 연구 데이터 선별 및 제출 등 다양한 활동을 수행하여야 하는데 그 또한 연구자들에게 부담이 될 수 있다. 따라서 공공 연구 데이터 공유를 위해 필요사항(가이드 및 양식 제공, 기술 및 품질 지원, 교육 등)을 지원하는 조직을 만들어야 한다. 또한 빠르고 쉽게 보유 데이터를 공유할 수 있는 온라인 플랫폼도 필수적이다.

사실 오픈 사이언스 도입을 위해 무엇보다 필요한 것은 인식의 제고이다. 훌륭한 법 제도, 지원조직, 플랫폼이 마련된다 하여도 연구자어나 국가 R&D 의사결정자들이 오픈 사이언스의 필요성과 중요성을 인지하지 못하면 이는 활성화될 수 없다. 연구자가 생성한 논문이나 보고서뿐만 아니라 활용가치가 월등히 높은 연구 데이터 공유에 대한 공론화 및 공감대 형성이 필요하다.

[참고 1]

# 데이터 유통 현황

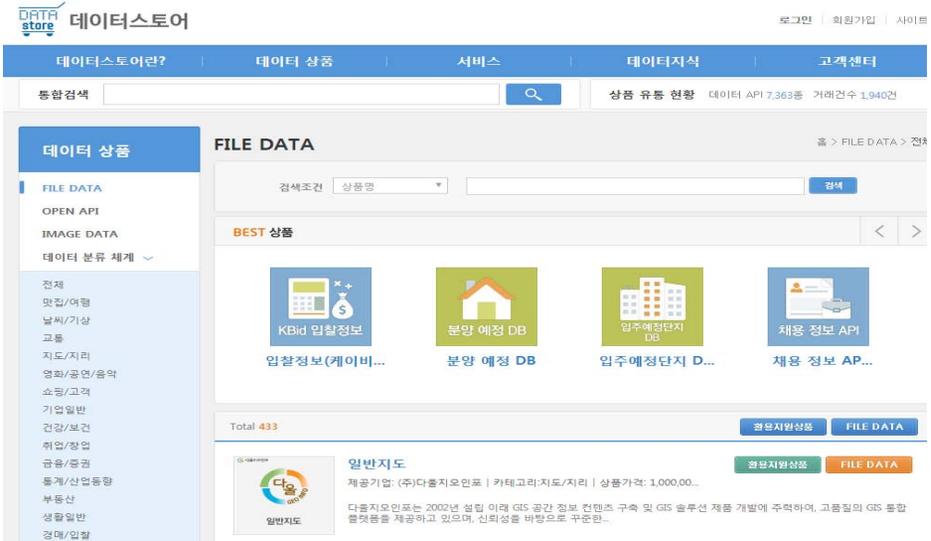
◇ 국내 민간 및 공공 데이터 유통 플랫폼 서비스를 통해 데이터 개방 및 공유 활성화 추진

※ 본 자료는 데이터스토어 웹사이트(<https://www.datastore.or.kr/intro.do>), 공공데이터포털(<http://data.go.kr/>)을 참고하였음

## □ (민간 데이터 유통 플랫폼) 데이터스토어

### ○ 사업 개요

- 데이터 유통 및 이용 활성화를 위한 과학기술정보통신부 지원 사업으로, 한국데이터진흥원이 운영하는 데이터 오픈 마켓
- 데이터 판매·거래, 중개, 가격 산정, API 판매대행 등을 통해 데이터 비즈니스 진흥을 지원



○ 제공 서비스

- 데이터 활용지원
- 데이터 매칭거래
- 데이터 가격산정
- API 무료개발지원
- 데이터상품 카탈로그 서비스

▶ 판매자	DATA store 데이터스토어	▶ 구매자
<p><b>기업 회원</b></p> <ul style="list-style-type: none"> <li>· 영리를 목적으로 DB 판매를 원하는 기업</li> <li>· API 판매 또는 API를 활용한 제사업을 추진하는 기업</li> </ul>	<p><b>FILE DATA 중개</b></p> <ul style="list-style-type: none"> <li>· 온라인 상품 전시</li> <li>· 상품 견적 제공</li> </ul> <p><b>API판매 대행</b></p> <ul style="list-style-type: none"> <li>· Proxy 서비스</li> <li>· 온라인 결제 기능 제공</li> <li>· 매출 정산 서비스</li> <li>· 활용사례 발굴 지원</li> </ul> <p><b>API활용정보 제공</b></p> <ul style="list-style-type: none"> <li>· API소재정보</li> <li>· 예제소스</li> <li>· 포럼게시판</li> </ul> <p><b>온라인 법률상담 서비스</b></p> <ul style="list-style-type: none"> <li>· 거래, 계약서 검토</li> <li>· 데이터 저작권 등</li> </ul> <p style="text-align: center; background-color: #555; color: white; padding: 5px;">온라인 법률상담</p>	<p><b>빅데이터 분석</b></p> <ul style="list-style-type: none"> <li>· 빅데이터 분석을 통한 사업 적용</li> </ul> <p><b>서비스 개발</b></p> <ul style="list-style-type: none"> <li>· 모바일, 웹 서비스 등 개발</li> </ul> <p><b>사업전략 수립</b></p> <ul style="list-style-type: none"> <li>· 사업 추진을 위한 근거자료</li> </ul>
<p><b>기관 회원</b></p> <ul style="list-style-type: none"> <li>· 데이터 또는 API를 유통하여 활용사례 발굴을 원하는 기관</li> </ul>		

○ 데이터 활용 진흥 활성화

- 데이터 스토리: 2017년 8월 7일 기준 총 284개의 데이터스토리를 제공하여 ‘데이터 분석 결과를 활용하는 건 결국 사람’, ‘고객의 소비패턴, 알고 계신가요?’등의 내용으로 서비스 이용자가 데이터 분석에 대해 정보를 쉽게 얻을 수 있도록 도움
- API 기술정보: API 개발자들을 위해 가이드 및 예제 소스 등을 게시하여 개발 환경에서 문제 해결 방법을 제공하고, 질문을 게시할 수 있도록 하여 제공되는 API 활용 개발 시 어려움을 해소할 수 있도록 도움

- 데이터 디렉토리: 국내에서 제작, 유통 중인 DB서비스 및 오픈API에 대한 전반적 현황과 상세정보를 교육/훈련, 경제/금융, 경영/비즈니스, 언론/뉴스 등 9가지의 대분류와 하위 세부 분류로 나눠 제공
- 우수 사례 게시: 데이터스토어의 데이터를 활용하여 만든 어플리케이션을 포함한 서비스를 발굴하여 웹사이트에 게시함으로써 이용자의 시각을 넓히고 데이터 활용에 대한 정보 제공

## □ (공공 데이터 유통 플랫폼) 공공데이터포털

### ○ 사업 개요

- 공공데이터법에 따라 접근 자격에 차별 없이, 모든 국민이 사용 가능하도록 서비스 하는 데이터 공유 플랫폼

**DATA** 공공데이터포털 .GO.KR

데이터셋 | 활용사례 | 참여마당 | 정보공유

검색어를 입력하세요.

**FILE DATA**    **OPEN API**    **STANDARD DATA**

교육, 국토관리, 공공행정, 재정금융, 산업고용, 사회복지, 식량건강, 문화관광, 보건의료, 재난안전, 교통물류, 환경기상, 과학기술, 농축수산, 통일외교안보, 법률

**국가 중점개발 데이터**  
국민의 손으로 직접 선정한 '국가 중점개발 데이터' 36대 분야를 다용량 데이터로 개방합니다.

**공공데이터 활용사례**  
HUPPLE  
휴플(Hupple)  
고속도로 휴게소 음식, 간식, 편의시설, 입장매장 정보 및 우유소 가격 정보를 제공하는 어플리케이션입니다. 개...

**기업탐방 인터뷰**  
속박 산업과 눈이문화 결합, 다 같이 아~~ 놀자!  
기업명 : ㈜아놀자 | 활용사례 : 아놀자맨션

**인기 데이터**

파일데이터	오픈 API
1 전국 전주전선화번호(전주번호)	
2 승차차인원 통계	
3 낙뢰관측 정보	
4 송변전설비 점검항목 정보	

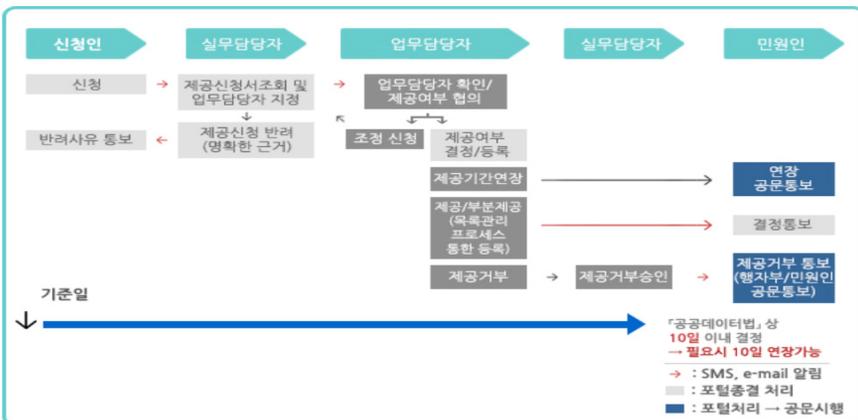
○ 공공데이터활용지원센터

- 공공데이터포털을 통해 공공기관의 데이터 개방과 민간의 이용 활성화를 지원하는 실무 조직
- 공공데이터 활용팀과 공공데이터 개방팀으로 나뉘져 업무 담당

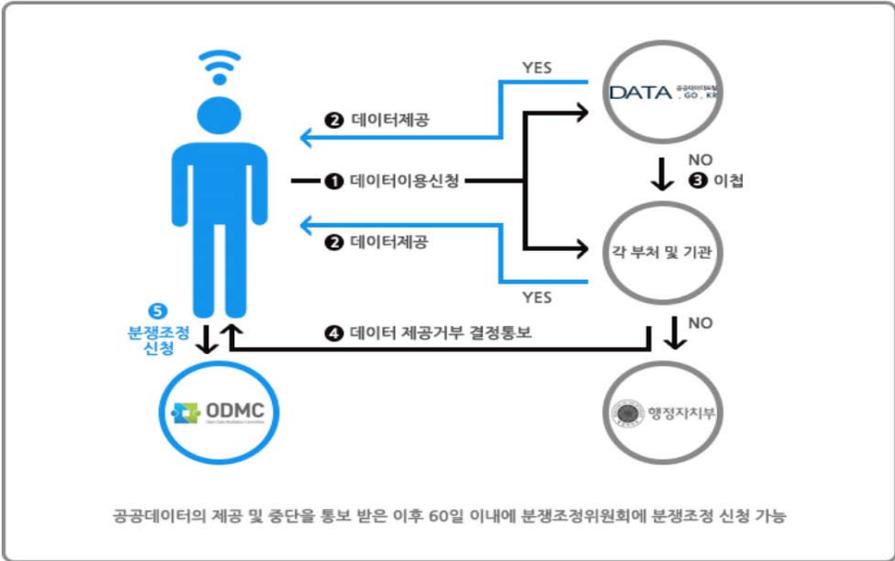


○ 서비스 이용

- 데이터 유형은 파일데이터, 오픈API, 표준데이터 등으로 제공됨
- 별도 신청 절차 없이 각 공공기관의 홈페이지 혹은 공공데이터포털에서 데이터 목록 확인 가능
- 공공데이터포털에서 제공하지 않는 데이터는 제공 신청을 통해 이용 가능하며 제외대상 정보가 포함된 데이터는 제공이 불가능



- 제공 불가능 데이터는 이용자가 공공데이터제공분쟁조정위원회를 통해 조정을 신청 가능



○ 주요 서비스

- 활용 사례 게시: 국내외 데이터 활용 사례를 제공하며 가공된 데이터를 공유하여 데이터 활용에 대한 이용자의 이해를 도움
- 기업탐방 인터뷰 제공: 공공데이터를 활용한 기업의 인터뷰 기사를 제공하여 실제 비즈니스에서 데이터 활용된 사례를 확인할 수 있음
- 기업지원 정책정보 제공: 창업포럼, 핀테크 해커톤 정보 등 각종 기업 지원 정보를 제공하여 기업이 참여할 수 있는 데이터 관련 행사 정보 확인 가능
- 개발자 네트워크 제공: 공공데이터 관련 개발 지식을 공유할 수 있는 커뮤니티 구성

○ 데이터 분류 체계(17. 8. 7 기준)

데이터스토어 (28개 분류)		공공데이터포털 (16개 분류)	
분류	데이터 건수	분류	데이터 건수
맛집/여행	265	문화관광	4,105
날씨/기상	128	환경기상	1,324
교통	247		
지도/지리	972	국토관리	1,057
영화/공연/음악	160		
쇼핑/고객	122		
기업일반	845		
건강/보건	244	식품건강	388
취업/창업	50	산업고용	1,744
금융/증권	75	재정금융	458
통계/산업동향	228		
부동산	76		
생활일반	2,182		
경매/입찰	36		
과학기술/자연	274	과학기술	940
논문/도서	92		
사진/미술	36		
스포츠/레저	77		
여성/육아	79	사회복지	1,568
오락/게임	15		
인문/사회	335	공공행정	3,828
자격증/시험	86		
전통문화	92		
특허/신안	198		
유학/연수	0		
물류/운송	92	교통물류	1,942
교육	182	교육	1,308
		보건의료	1,596
		재난안전	1,678
		농축수산	993
		통일의외교안보	242
		법률	42
<b>총</b>	<b>7,193</b>	<b>총</b>	<b>23,217</b>

자료: 데이터스토어(<https://www.datastore.or.kr/intro.do>), 공공데이터포털(<http://data.go.kr/main.do>)

주: 데이터 건수는 서비스 중인 모든 종류의 데이터 유형을 포함하는 수치임

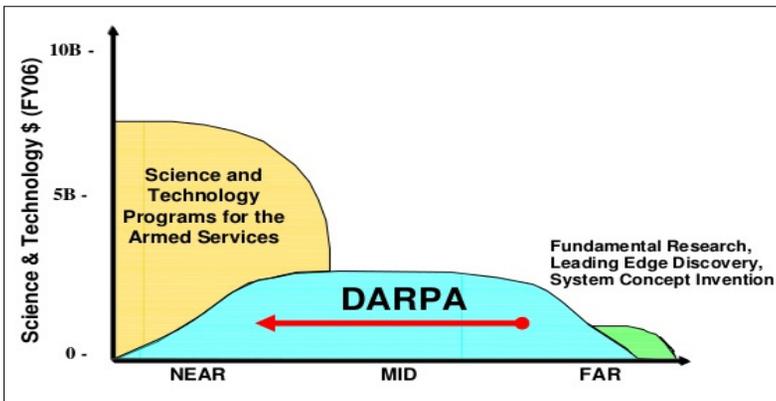
[참고 2]

# 1. DARPA(Defense Advanced Research Projects Agency)

## □ 개요

- 1958년 미국방성(Department of Defense) 산하에 설립된 연구개발 기구
  - 20여 명의 고위 기술 관리들(technical program managers)을 포함한 약 250여 명으로 구성되며, 약 30억 달러('17년 기준)의 예산으로 운영됨
  - 자체 연구개발 시설 없이 모든 연구개발 활동은 철저히 외주를 통해 수행
- 현존하는 최첨단 과학지식을 신속하게 응용하여 군사적 수요와의 간극을 메우는 기술 개발을 추구
  - 현재 알려진 기술적 수요가 아닌 중·장기적 미래에 필요로 하는 혁명적 개념의 국방기술에 투자(High-Risk, High-Return)

### [DARPA의 역할]



자료: DARPA(2009)

## □ 성공 사례

- 설립 이후 국방과 관련된 일련의 연구개발 활동에서 탁월한 기술적 성과를 창출

〈DARPA 기술개발 성공 사례 예시〉

예시	주요 내용
<p style="text-align: center;">인터넷</p> 	<p>목적: 냉전시대에 소련의 핵공격에 대비한 컴퓨터 보호</p> <p>과정 : ① 1962년 JCR Licklider의 연구에 “on-line man computer communication”에 대해 언급                  ② 1969년 10월 알파넷(ARPANET)을 통해 각각 다른 지역에 위치한 세 개의 컴퓨터를 연결하는데 성공하여 network을 만들                  ③ 1969년 말 UCLA, UCSB, UofU, Stanford 네 개의 대학을 연결하면서 인터넷이 시작됨</p>
<p style="text-align: center;">구글 맵스</p> 	<p>목적: 군인들이 낯선 지역에서 작전 수행할 때 지역에 익숙해 질 수 있도록 돕기 위한</p> <p>과정 : ① 1976년 이스라엘 공군특공대가 우간다 엔테베 국제공항에서 테러리스트들이 납치한 인질 구출 작전을 수행하는데 성공함                  ② 이 사건 이후 Aspen Movie Map 프로젝트 만들                  ③ 1979년 DARPA가 MIT Architecture Machine Group 학생들에게 연구비를 지원하였고 Aspen Movie Map을 선보임                  ④ Aspen Movie Map은 상호적 지도로 사용자가 도시를 여행하고 건물 안 투시 또는 시간여행을 통해 과거의 역사적 건물의 모습 재현</p>
<p style="text-align: center;">시리</p> 	<p>목적: 군인들의 정보 습득을 돕기 위한</p> <p>과정: ① 정보소통기술인 (Cognitive Assistant that Learns and Organizes, CALO)가 만들어짐                  ② 2003년 음성인식과 AI를 통합한 개념인 음성 개인 비서인(Personalized Assistant that Leans, PAL)프로그램으로 발전                  ③ 2007년 SIRI라는 벤처기업 설립                  ④ 2010년 애플에서 시리인식 기술을 도입함</p>
<p style="text-align: center;">GPS</p> 	<p>목적: 핵잠수함 미사일의 정확한 위치 추적</p> <p>과정: ① 1957년 Johns Hopkins의 응용물리학자들이 스푸트니크 위성이 지구에 보내는 주파수 추적함                  ② DARPA가 이 연구원들을 통해 ‘도플러항법’ 이용한 위치추적 위성을 이용하여 핵잠수함 정확한 위치 파악 기술을 얻어냄                  ③ 1964년 최초의 위성 탐색 시스템이 생방송 되면서 GPS가 생김</p>

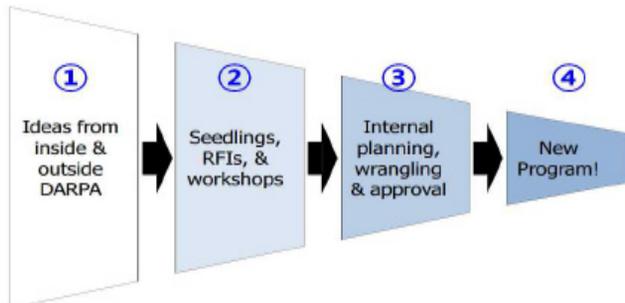
자료: 한국연구재단 정책연구혁신센터

## □ 사업공고

- BAA(Broad Agency Announcement)라고 하며, 특정 연구분야와 해결해야 할 문제를 제시하고 연구자들로부터 제안서를 받는 방식
  - Office-Wide BAA\*와 Program BAA\*\* 두 종류로 구분
  - \* 담당 부문의 기술분야에 해당하는 어떠한 주제에 대해서도 포괄적으로 연구제안을 받는 것으로 Program Manager가 신규 프로그램 아이디어를 수집하기 위해 사용됨
  - \*\* 특정 연구목적용 지정하여 그 목적에 부합하는 연구제안만 받는 것
- 사업공고는 평균 45~75일 동안 추진되며, 제출된 제안서 평가는 평균 14~30일 소요, 수행기관 선정 후 협약까지 평균 60~120일 소요됨

## □ 프로그램 기획

### 〈DARPA 프로그램 기획 절차〉



- DARPA의 프로그램 기획 절차는 크게 4단계로 구분됨
  - DARPA 내·외부로부터 아이디어 수집 → 예비연구(seedling)\*, 공개정보요청(RFI)\*\*, 워크숍\*\*\* → DARPA 내부 기획·토론·승인 → 신규 프로그램 신설
  - \* Office-Wide BAA로 접수된 연구제안서 중 Program Manager가 관심이 있는 것을 선정, 제안자에게 6~9개월 동안 소액의 연구비를 지원하고 제안한 연구의 가능성에 대한 조사를 시키는 것

\*\* DARPA가 민간으로부터 정보를 수집할 때 사용(기술 현장의 문제점 및 존재 여부 파악)

\*\*\* Program Manager가 민간 전문가들로부터 새로운 프로그램 아이디어 등을 수집하기 위함

○ 문제를 인지하고 정의함에 있어서 top-down 구조를 따르고, 정의된 문제의 기술적 아이디어를 찾아 사업을 구성하는데 있어서는 모든 Program Manager가 참여하는 bottom-up 방식을 따름

- Program Manager들은 실패의 확률이 높지만 성공했을 경우 파급효과가 크고, 혁신적 아이디어에 기반하며 기초연구와 단기적 개발수요 간 간극을 메워줄 수 있는 연구주제 제안

- Program Manager는 자신이 제안한 연구주제가 이 세 가지 원칙에 부합한다는 것을 동료 Program Manager와 Senior Technical Manager 그리고 Director 및 Deputy Director에게 증명

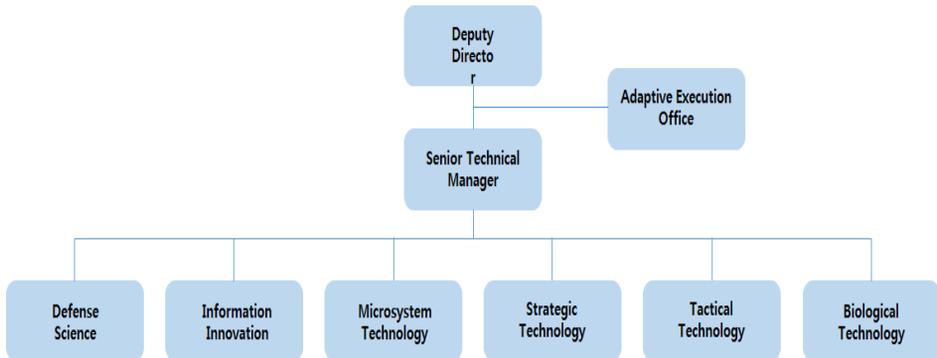
- 각 프로젝트는 4년 동안, 1,000~4,000만 달러의 예산으로 대학을 포함한 5~10개 내외의 연구조직과의 계약 형태로 수행

- 각 프로젝트는 기술적 내용과 예산집행에 대한 전권을 행사하는 단 한 명의 Program Manager에 의해 관리됨(지원 요원은 각 사업의 예산에 의해 배정)

## □ 조직 운영

- 전체 운영을 책임지는 Deputy Director, 6개 분야 별 오피스를 담당하는 Senior Technical Manager, 각 프로젝트를 담당하는 Program Manager\* 들로 구성

[DARPA 조직도]



- \* Program Manager는 산·학·연 등 각 분야로부터 기술적 우수성과 혁신성을 기준으로 선발되며, 3~5년의 계약직으로 고용됨

## □ 예산

- DARPA의 예산은 美 연방정부의 예산결정 과정에 따라 국방부의 예산 내에서 할당됨

### <美 연방정부의 예산결정 과정>

대통령 예산안을 의회에 제출(행정부) → 예산결의안 작성(의회) → 예산안 수립(의회, 연방정부기관 청문회) → 최종 예산안 의결(의회) → 대통령 서명(행정부, 거부권) → 예산법 발효

### <美 정부의 예산 종류>

\* 재량지출(Discretionary Spending): 행정부에서 재량을 가지고 집행이 가능한 예산으로, **R&D 예산은 대부분 재량지출로 책정**되며 의회의 예산(안) 승인이 필요

\* 의무지출(Mandatory Spending)은 주로 사회보장, 의료보험과 같이 법으로 정해져 반드시 집행이 필요한 지출로 직접지출(Direct Spending)으로 불리며 의회의 승인 불필요

## □ 성공요인

- 통상적인 연구개발 프로그램이 채택하고 있는 전형적 운영방식과 접근방법을 거부하고 DARPA만의 독특한 관리 조직 및 운영 구조 채택(DARPA, 2005; Van Atta, 2008)
  - 소규모의 유연성 있는 조직(small and flexible), 계층관료제를 거부하는 평면조직(flat organization), 실질적인 자치권과 관료적 장애로부터의 자유(substantial autonomy and freedom from bureaucratic impediments)
- 간결한 평면조직을 통해 관료제도의 폐해를 최소화하고, 신속적 의사결정과 새로운 수요 및 혁신적 아이디어에 대한 신속한 대응이 가능하도록 함
- Peer Review의 절차를 전면 거부, Program Manager의 의지에 따라 언제든지 사업을 시작할 수 있음
  - peer review를 통해 사업수행상의 오류를 방지할 수 있지만, 이는 길고 지루한 선정과정으로 인하여 긴급하고 혁신적인 연구개발 사업의 수행을 불가능하게 할 뿐만 아니라 보수적인 사업이 채택될 가능성을 높임

## □ 시사점

- 최고의 연구개발 인재를 확보할 수 있는 방안 마련 시급
  - 인재들이 자유롭게 다른 영역(산·학·연)과 순환고용이 될 수 있도록 보장해야 함
  - 순환고용을 보장하기 위한 다양한 인센티브와 제도적 장치 마련 필요
- 발탁된 인재들이 Program manager로서 최대한의 자율권과 권한을 행사할 수 있는 구조 필요
- 관료제적 폐해를 최대한 피하여 신속하고 유연한 의사결정을 담보할 수 있는 가볍고 평면적인 조직 구조를 만들어야 함
- 신속한 의사결정을 위해 Peer Review 절차를 간소화 또는 생략 가능하게 만들어야 함
- In-house 연구시설을 가지지 않고 산·학·연 등 다양한 섹터에 산재해 있는 연구자원을 활용할 수 있어야 함

## 2. DARPA Challenge

□ (분야) 민간에서 아직 본격 연구개발이 시작되지 않고 있는 기술 분야를 선택해 도전적인 목표를 제시하고, 상금을 내걸어 많은 사람들의 기술개발 추진을 유도

- 일반적인 R&D 프로젝트와는 달리 성공에 대한 위험부담 없이 보다 도전적인 목표를 설정할 수 있도록 해주고, 다양한 분야의 참여를 이끌 수 있음
- \* 기존 R&D 프로젝트 방식은 누가 어떤 방식으로 추진하게 하여 성공시킬 것인지를 고려해야하기 때문에 도전적인 목표 설정이 어려움
- \* DARPA 상금보다 훨씬 더 많은 예산이 민간 자체적으로 R&D에 투자하게 됨

<DARPA Challenge 주요 사례>

- \* Grand Challenge('04~'05, 2회): 사막지역에서의 자율주행자동차 경진대회
- \* Urban Challenge('07): 도심지역에서의 자율주행자동차 경진대회
- \* Robotics Challenge('15): 휴머노이드形 재난구조로봇 개발
- \* Cyber Grand Challenge('16): AI을 활용한 사이버 해킹 및 보안기술 개발
- \* Spectrum Collaboration Challenge('19): 무선 주파수의 효과적인 활용을 위한 기술개발

Challenge	기간	Track	예산 규모
Grand Challenge	2004~2005	Open	예산 지원 없음
Urban Challenge	2007	Track A	1단계 \$1,000,000
		Track B	예산 지원 없음
Robotics Challenge	2015	Track A/B	최대 2단계 \$4,000,000
		Track C/D	예산 지원 없음
Cyber Grand Challenge	2016	Funded	최대 2단계 \$1,500,000
		Open	예산 지원 없음
Spectrum Collaboration Challenge	2019	Proposal	최대 3단계 \$2,000,000
		Open	예산 지원 없음

□ (종류) DARPA로부터 예산을 지원받는 “Funded Track”과 예산지원을 받지 않는 “Open Track”으로 추진됨

- (Funded Track) DARPA는 챌린지와 관련된 기술개발 지원 사업을 기획·공고하고, 복수의 수행기관들을 선정해 예산을 지원
  - \* BAA(Broad Agency Announcement)라고 불리는 방식으로 공고가 되는데, 이는 연구주제 및 목표를 제시하고 이를 추진할 사업 계획들을 모집하는 방식
    - 프로젝트는 보통 2~3단계로 지원되며, 단계마다 최대 지원 규모가 정해져 있음
    - 단계 종료 후에는 Preliminary Event와 같은 단계평가를 거쳐, 일정 기준을 충족하는 수행자만 차기 단계를 계속 지원받을 수 있음
    - 수행기관은 연구결과물에 대한 지식재산권을 소유하고 직접 상용화 등을 추진할 수 있으나, DARPA에게 연구결과물의 사용 권한을 제한 없이 부여해야 함
  - \* DARPA에 부여하는 권한의 범위는 예산지원 규모에 비례

Funded Track의 Intellectual Property 관련 조항

Proposers should note that the Government does not own the intellectual property of noncommercial, technical data/computer software developed with Government funds; it instead acquires the right to use the technical data/computer software. Regardless of the scope of the Government's rights, performers may freely use their same data/software for their own commercial purposes.

- DARPA의 해당연도 예산 상황에 따라 Funded Track에서 지원하는 팀의 수와 지원하는 출연금 규모가 달라짐
- (Open Track)DARPA의 예산지원 없이 연구자들이 스스로 기술개발을 추진해 참가하는 트랙
  - 예선이나 결선경기에 참가하기 위해서는 DARPA에서 정해놓은 일정 기술 기준을 충족해야 함
    - \* 챌린지에 따라 동영상을 찍어 제공하거나, 직접 DARPA가 개최하는 행사에 참여해서 기술을 직접 시연함으로써 기준을 충족함을 증명해야함
    - \* Funded Track 참가팀들은 평가를 통해 선정되었으므로 기술수준을 충족했다고 인정

- 기술개발 결과물의 보호를 위해 Open Track을 활용하는 기업들도 있음
- \* 로보틱스 챌린지에서 日 샤프트社는 Funded Track으로 참가하다가 구글에 인수된 이후 Open Track으로 전환하여 참가
- ※ Funded/Open Track 참가팀들은 함께 예선·결선 경기를 치르게 되며, 1~3등을 차지한 참가 팀들이 상금\*을 수여받게 됨
- \* 일반적으로 1등 \$2,000,000, 2등 \$1,000,000, 3등 \$750,000

## □ (기대 성과) 기술의 사업화 촉진 및 우수 연구팀 발굴

- 챌린지 이후 해당 주제에 대한 민간의 관심이 높아지고 기술개발도 본격적으로 추진되면서 궁극적으로 기술의 사업화를 촉진
- \* 그랜드/어번 챌린지는 민간의 자율주행차에 대한 관심을 높이는 계기가 되었음
- 우수 기술력을 보유한 산·학·연 연구팀들을 발굴할 수 있는 기회가 되고 있으며, 정부·민간기업과 챌린지 참가팀 간 연계할 수 있는 기회를 제공하기도 함
- \* 구글은 그랜드/어번 챌린지의 연구자들을 스카우트해서 '09년부터 자율주행차를 개발, 로보틱스 챌린지의 예선경기에서 우수한 성적을 거둔 팀들(日샤프트社를 포함한 상위 8개 팀)을 인수하여 자사의 로봇 개발에 활용

## □ (차별성) 기존 정부 기술개발 사업과 다양한 측면에서 차별성 존재

- 다수의 연구팀이 동일한 연구주제에 대해 기술개발을 추진
- 정부 출연금을 기술개발 이후 상금 형태로 받음
- 보다 도전적인 연구목표 설정 가능
  - 목표를 달성했을 경우에만 상금을 지급하기 때문에 예산 집행에 대한 부담이 없기 때문이기도 함
- 비용 대비 효과성이 높음
  - 총 상금 대비 참여자 수(참여자들이 스스로 비용을 지불하고 개발에 참여) 및 홍보효과

□ (시사점) 챌린지를 추진하는데 있어서 도전적인 연구목표 설정, 대회 운영규칙 및 기술 평가 기준 등을 제정함에 있어서 정밀하게 추진할 필요가 있음

- 도전적인 목표를 설정하되, 달성이 불가능한 목표가 되지 않도록 해야 하며, 이를 위해 현재 기술수준이나 기술발전 속도 등에 대한 정확한 판단이 요구됨
- 정밀한 평가 기준을 정하고, 운영규칙을 설계하고 이를 사전 공지할 필요가 있음
- 대회를 운영하는데 충분한 기간을 보장해야 함
  - DARPA의 경우 기술개발에 필요한 기간을 제공하기 위해 챌린지 기간으로 최소 3년을 잡고 추진

[부 록]

## 통 계 표

〈AI SW 개발을 위한 데이터 확보에 어려움〉

(단위: 개, %)

구 분	사례수	예	아니오	모름/무응답	계	
전 체	(100)	36.0	28.0	36.0	100.0	
인공 지능 분야별	추론 및 기계학습	(71)	42.3	29.6	28.2	100.0
	지식표현 및 언어지능	(31)	54.8	22.6	22.6	100.0
	청각 지능	(12)	50.0	25.0	25.0	100.0
	시각 지능	(22)	50.0	18.2	31.8	100.0
	복합 지능	(14)	35.7	21.4	42.9	100.0
	지능형 에이전트	(14)	64.3	28.6	7.1	100.0
	인간-기계 협업	(9)	33.3	44.4	22.2	100.0
	지능형 반도체	(3)	0.0	66.7	33.3	100.0
	뉴로모픽칩	(3)	33.3	33.3	33.3	100.0
	고성능컴퓨팅	(8)	25.0	50.0	25.0	100.0
	양자컴퓨팅	(1)	0.0	0.0	100.0	100.0
	모름/무응답	(7)	0.0	14.3	85.7	100.0

〈01. AI SW 개발을 위한 데이터 확보가 필요한 분야—의료〉

(단위: 개, %)

구 분		사례 수	없다	추론 및 기계 학습	지식 표현/ 언어 지능	시각 지능	청각 지능 (음성 및 소리 인식)	복합 지능	지능형 에이전트	인간·기계 협업 (HCI)
전 체		(100)	65.0	32.0	10.0	3.0	2.0	2.0	2.0	1.0
인공 지능 분야별	추론 및 기계 학습	(71)	59.2	38.0	9.9	4.2	2.8	2.8	2.8	1.4
	지식 표현 및 언어 지능	(31)	64.5	29.0	12.9	3.2	3.2	3.2	6.5	3.2
	청각 지능	(12)	50.0	50.0	8.3	16.7	8.3	0.0	0.0	0.0
	시각 지능	(22)	68.2	27.3	9.1	13.6	4.5	0.0	0.0	0.0
	복합 지능	(14)	78.6	21.4	0.0	0.0	0.0	7.1	0.0	0.0
	지능형 에이전트	(14)	64.3	28.6	7.1	14.3	0.0	7.1	7.1	0.0
	인간-기계 협업	(9)	77.8	11.1	11.1	0.0	0.0	0.0	0.0	0.0
	지능형 반도체	(3)	66.7	33.3	0.0	0.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	37.5	50.0	0.0	25.0	0.0	0.0	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
모름/무응답	(7)	85.7	14.3	0.0	0.0	0.0	0.0	0.0	0.0	

<02. AI SW 개발을 위한 데이터 확보가 필요한 분야—국방>

(단위 : 개, %)

구 분		사례수	없다	추론 및 기계학습	지식표현/ 언어지능	시각지능	지능형 에이전트
전 체		(100)	87.0	10.0	4.0	2.0	1.0
인공 지능 분야별	추론 및 기계학습	(71)	83.1	14.1	5.6	1.4	1.4
	지식표현 및 언어지능	(31)	80.6	12.9	6.5	3.2	3.2
	청각 지능	(12)	83.3	8.3	0.0	8.3	0.0
	시각 지능	(22)	81.8	9.1	0.0	9.1	0.0
	복합 지능	(14)	100.0	0.0	0.0	0.0	0.0
	지능형 에이전트	(14)	78.6	14.3	7.1	7.1	7.1
	인간-기계 협업	(9)	88.9	0.0	11.1	0.0	0.0
	지능형 반도체	(3)	100.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	62.5	25.0	0.0	12.5	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0	0.0

<03. AI SW 개발을 위한 데이터 확보가 필요한 분야-교육>

(단위: 개, %)

구 분		사례수	없다	추론 및 기계학습	지식표현/언어지능	시각지능	지능형 에이전트
전 체		(100)	90.0	9.0	4.0	2.0	2.0
인공 지능 분야별	추론 및 기계학습	(71)	85.9	12.7	5.6	2.8	2.8
	지식표현 및 언어지능	(31)	83.9	16.1	12.9	3.2	6.5
	청각 지능	(12)	100.0	0.0	0.0	0.0	0.0
	시각 지능	(22)	81.8	13.6	9.1	9.1	4.5
	복합 지능	(14)	100.0	0.0	0.0	0.0	0.0
	지능형 에이전트	(14)	78.6	14.3	14.3	14.3	14.3
	인간-기계 협업	(9)	100.0	0.0	0.0	0.0	0.0
	지능형 반도체	(3)	100.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	87.5	0.0	0.0	12.5	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0	0.0

〈04. AI SW 개발을 위한 데이터 확보가 필요한 분야-농·축산·어업〉

(단위: 개, %)

구 분		사례수	없다	추론 및 기계학습	지식표현/ 언어지능	지능형 에이전트
전 체		(100)	96.0	4.0	2.0	2.0
인공 지능 분야별	추론 및 기계학습	(71)	95.8	4.2	2.8	2.8
	지식표현 및 언어지능	(31)	90.3	9.7	6.5	6.5
	청각 지능	(12)	100.0	0.0	0.0	0.0
	시각 지능	(22)	95.5	4.5	0.0	0.0
	복합 지능	(14)	100.0	0.0	0.0	0.0
	지능형 에이전트	(14)	92.9	7.1	7.1	7.1
	인간-기계 협업	(9)	100.0	0.0	0.0	0.0
	지능형 반도체	(3)	100.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	66.7	33.3	0.0	0.0
	고성능컴퓨팅	(8)	100.0	0.0	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0

〈05. AI SW 개발을 위한 데이터 확보가 필요한 분야—법률〉

(단위: 개, %)

구 분		사례수	없다	추론 및 기계학습	지식표현/ 언어지능	지능형 에이전트	인간·기계 협업(HCI)
전 체		(100)	93.0	5.0	3.0	2.0	1.0
인공 지능 분야별	추론 및 기계학습	(71)	90.1	7.0	4.2	2.8	1.4
	지식표현 및 언어지능	(31)	77.4	16.1	9.7	6.5	3.2
	청각 지능	(12)	91.7	0.0	8.3	0.0	0.0
	시각 지능	(22)	90.9	4.5	4.5	0.0	0.0
	복합 지능	(14)	100.0	0.0	0.0	0.0	0.0
	지능형 에이전트	(14)	85.7	7.1	14.3	7.1	0.0
	인간-기계 협업	(9)	88.9	0.0	0.0	0.0	11.1
	지능형 반도체	(3)	100.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	87.5	0.0	12.5	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0	0.0

〈06. AI SW 개발을 위한 데이터 확보가 필요한 분야－웰니스〉

(단위: 개, %)

구 분	사례수	없다	추론 및 기계학습	지식 표현/ 언어 지능	시각 지능	지능형 에이전트	복합 지능	청각지능 (음성 및 소리인식)	
전 체	(100)	92.0	7.0	4.0	2.0	2.0	2.0	1.0	
인공 지능 분야별	추론 및 기계학습	(71)	88.7	9.9	5.6	2.8	2.8	2.8	1.4
	지식표현 및 언어지능	(31)	77.4	19.4	12.9	6.5	6.5	6.5	3.2
	청각 지능	(12)	91.7	8.3	8.3	8.3	8.3	8.3	8.3
	시각 지능	(22)	86.4	9.1	4.5	9.1	4.5	4.5	4.5
	복합 지능	(14)	92.9	7.1	7.1	7.1	7.1	7.1	7.1
	지능형 에이전트	(14)	85.7	14.3	14.3	7.1	14.3	7.1	7.1
	인간-기계 협업	(9)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	지능형 반도체	(3)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0	0.0	0.0	0.0

<07. AI SW 개발을 위한 데이터 확보가 필요한 분야-제조>

(단위: 개, %)

구 분		사례수	없다	추론 및 기계학습	지식표현/ 언어지능	지능형 에이전트	인간· 기계협업 (HCI)	시각지능
전 체		(100)	88.0	11.0	3.0	2.0	2.0	1.0
인공 지능 분야별	추론 및 기계학습	(71)	84.5	14.1	4.2	2.8	2.8	1.4
	지식표현 및 언어지능	(31)	77.4	19.4	9.7	6.5	6.5	3.2
	청각 지능	(12)	91.7	8.3	0.0	0.0	0.0	0.0
	시각 지능	(22)	86.4	13.6	4.5	4.5	0.0	4.5
	복합 지능	(14)	100.0	0.0	0.0	0.0	0.0	0.0
	지능형 에이전트	(14)	85.7	14.3	7.1	7.1	0.0	7.1
	인간-기계 협업	(9)	88.9	0.0	0.0	0.0	0.0	11.1
	지능형 반도체	(3)	66.7	33.3	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	62.5	37.5	0.0	0.0	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0	0.0	0.0

〈08. AI SW 개발을 위한 데이터 확보가 필요한 분야-흠〉

(단위: 개, %)

구 분		사례수	없다	추론 및 기계학습	지식표현/언어지능	청각지능 (음성 및 소리인식)	지능형 에이전트	시각 지능	복합 지능
전 체		(100)	93.0	5.0	5.0	2.0	2.0	1.0	1.0
인공 지능 분야별	추론 및 기계학습	(71)	91.5	7.0	5.6	1.4	2.8	1.4	1.4
	지식표현 및 언어지능	(31)	77.4	16.1	16.1	6.5	6.5	3.2	3.2
	청각 지능	(12)	83.3	8.3	16.7	16.7	8.3	8.3	8.3
	시각 지능	(22)	86.4	9.1	9.1	9.1	4.5	4.5	4.5
	복합 지능	(14)	92.9	7.1	7.1	7.1	7.1	7.1	7.1
	지능형 에이전트	(14)	85.7	14.3	14.3	7.1	14.3	7.1	7.1
	인간-기계 협업	(9)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	지능형 반도체	(3)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0	0.0	0.0	0.0

<09. AI SW 개발을 위한 데이터 확보가 필요한 분야—금융>

(단위: 개, %)

구 분		사례수	없다	추론 및 기계학습	지능형 에이전트	지식표현/ 언어지능	시각 지능	복합 지능
전 체		(100)	85.0	15.0	5.0	5.0	1.0	1.0
인공 지능 분야별	추론 및 기계학습	(71)	78.9	21.1	7.0	7.0	1.4	1.4
	지식표현 및 언어지능	(31)	74.2	25.8	12.9	16.1	3.2	3.2
	청각 지능	(12)	91.7	8.3	8.3	0.0	0.0	0.0
	시각 지능	(22)	81.8	18.2	9.1	9.1	4.5	0.0
	복합 지능	(14)	100.0	0.0	0.0	0.0	0.0	0.0
	지능형 에이전트	(14)	71.4	28.6	28.6	14.3	7.1	0.0
	인간-기계 협업	(9)	100.0	0.0	0.0	0.0	0.0	0.0
	지능형 반도체	(3)	100.0	0.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	75.0	25.0	12.5	0.0	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0	0.0	0.0

〈10. AI SW 개발을 위한 데이터 확보가 필요한 분야-문화관광〉

(단위: 개, %)

구 분		사례수	없다	추론 및 기계학습	지식표현/언어지능	시각지능	지능형 에이전트
전 체		(100)	96.0	3.0	2.0	1.0	1.0
인공 지능 분야별	추론 및 기계학습	(71)	94.4	4.2	2.8	1.4	1.4
	지식표현 및 언어지능	(31)	87.1	9.7	6.5	3.2	3.2
	청각 지능	(12)	100.0	0.0	0.0	0.0	0.0
	시각 지능	(22)	90.9	4.5	0.0	4.5	0.0
	복합 지능	(14)	100.0	0.0	0.0	0.0	0.0
	지능형 에이전트	(14)	92.9	7.1	7.1	0.0	7.1
	인간-기계 협업	(9)	100.0	0.0	0.0	0.0	0.0
	지능형 반도체	(3)	100.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	100.0	0.0	0.0	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0	0.0

〈11. AI SW 개발을 위한 데이터 확보가 필요한 분야－유통〉

(단위: 개, %)

구 분	사례수	없다	추론 및 기계 학습	지식 표현/ 언어 지능	지능형 에이 전트	청각 지능 (음성 및 소리 인식)	복합 지능	시각 지능	인간· 기계 협업 (HCI)
전 체	(100)	89.0	9.0	4.0	3.0	1.0	1.0	1.0	1.0
인공 지능 분야별	추론 및 기계학습	(71)	85.9	11.3	5.6	4.2	1.4	1.4	1.4
	지식표현 및 언어지능	(31)	80.6	16.1	12.9	9.7	3.2	3.2	0.0
	청각 지능	(12)	83.3	16.7	16.7	8.3	8.3	8.3	0.0
	시각 지능	(22)	77.3	18.2	13.6	4.5	4.5	4.5	4.5
	복합 지능	(14)	92.9	7.1	7.1	0.0	7.1	7.1	0.0
	지능형 에이전트	(14)	71.4	21.4	21.4	14.3	7.1	7.1	7.1
	인간-기계 협업	(9)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	지능형 반도체	(3)	66.7	33.3	0.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	50.0	37.5	12.5	12.5	0.0	0.0	12.5
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
모름/무응답	(7)	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

〈12. AI SW 개발을 위한 데이터 확보가 필요한 분야-사무관리〉

(단위: 개, %)

구 분		사례수	없다	추론 및 기계학습	지식표현/언어지능	지능형 에이전트
전 체		(100)	92.0	8.0	3.0	2.0
인공 지능 분야별	추론 및 기계학습	(71)	90.1	9.9	4.2	2.8
	지식표현 및 언어지능	(31)	87.1	12.9	9.7	6.5
	청각 지능	(12)	100.0	0.0	0.0	0.0
	시각 지능	(22)	90.9	9.1	4.5	0.0
	복합 지능	(14)	100.0	0.0	0.0	0.0
	지능형 에이전트	(14)	92.9	7.1	7.1	7.1
	인간-기계 협업	(9)	100.0	0.0	0.0	0.0
	지능형 반도체	(3)	66.7	33.3	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	87.5	12.5	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0

〈13. AI SW 개발을 위한 데이터 확보가 필요한 분야—도시〉

(단위: 개, %)

구 분		사례수	없다	추론 및 기계학습	지식표현/언어지능	지능형 에이전트
전 체		(100)	98.0	2.0	1.0	1.0
인공 지능 분야별	추론 및 기계학습	(71)	97.2	2.8	1.4	1.4
	지식표현 및 언어지능	(31)	93.5	6.5	3.2	3.2
	청각 지능	(12)	100.0	0.0	0.0	0.0
	시각 지능	(22)	95.5	4.5	0.0	0.0
	복합 지능	(14)	100.0	0.0	0.0	0.0
	지능형 에이전트	(14)	92.9	7.1	7.1	7.1
	인간-기계 협업	(9)	100.0	0.0	0.0	0.0
	지능형 반도체	(3)	100.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	100.0	0.0	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0

<14. AI SW 개발을 위한 데이터 확보가 필요한 분야-교통>

(단위: 개, %)

구 분		사례수	없다	추론 및 기계학습	지식표현/언어지능	청각지능 (음성 및 소리인식)	시각지능	지능형 에이전트
전 체		(100)	94.0	6.0	3.0	1.0	1.0	1.0
인공 지능 분야별	추론 및 기계학습	(71)	91.5	8.5	4.2	1.4	1.4	1.4
	지식표현 및 언어지능	(31)	87.1	12.9	9.7	3.2	0.0	3.2
	청각 지능	(12)	91.7	8.3	8.3	8.3	0.0	0.0
	시각 지능	(22)	81.8	18.2	9.1	4.5	4.5	0.0
	복합 지능	(14)	92.9	7.1	7.1	7.1	0.0	0.0
	지능형 에이전트	(14)	78.6	21.4	14.3	7.1	7.1	7.1
	인간-기계 협업	(9)	100.0	0.0	0.0	0.0	0.0	0.0
	지능형 반도체	(3)	100.0	0.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	100.0	0.0	0.0	0.0	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0	0.0	0.0

<15. AI SW 개발을 위한 데이터 확보가 필요한 분야-에너지>

(단위: 개, %)

구 분		사례수	없다	추론 및 기계학습	지능형 에이전트	복합지능	인간·기계 협업(HCI)
전 체		(100)	94.0	5.0	2.0	1.0	1.0
인공 지능 분야별	추론 및 기계학습	(71)	94.4	5.6	1.4	0.0	0.0
	지식표현 및 언어지능	(31)	93.5	6.5	3.2	0.0	0.0
	청각 지능	(12)	91.7	8.3	8.3	0.0	0.0
	시각 지능	(22)	86.4	13.6	4.5	4.5	4.5
	복합 지능	(14)	92.9	7.1	0.0	7.1	7.1
	지능형 에이전트	(14)	85.7	7.1	14.3	0.0	0.0
	인간-기계 협업	(9)	77.8	11.1	11.1	11.1	11.1
	지능형 반도체	(3)	100.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	87.5	12.5	12.5	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0	0.0

〈16. AI SW 개발을 위한 데이터 확보가 필요한 분야—안전〉

(단위: 개, %)

구 분		사례수	없다	추론 및 기계학습	청각지능 (음성 및 소리인식)	시각 지능	복합 지능	지식표현/ 언어지능	지능형 에이전트
전 체		(100)	92.0	7.0	1.0	1.0	1.0	1.0	1.0
인공 지능 분야별	추론 및 기계학습	(71)	88.7	9.9	1.4	1.4	1.4	1.4	1.4
	지식표현 및 언어지능	(31)	87.1	9.7	3.2	3.2	3.2	3.2	3.2
	청각 지능	(12)	91.7	0.0	8.3	8.3	8.3	0.0	0.0
	시각 지능	(22)	90.9	4.5	4.5	4.5	4.5	0.0	0.0
	복합 지능	(14)	92.9	0.0	7.1	7.1	7.1	0.0	0.0
	지능형 에이전트	(14)	85.7	7.1	7.1	7.1	7.1	7.1	7.1
	인간-기계 협업	(9)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	지능형 반도체	(3)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0	0.0	0.0	0.0

〈17. AI SW 개발을 위한 데이터 확보가 필요한 분야—공공서비스 및 복지〉

(단위: 개, %)

구 분		사례수	없다	추론 및 기계학습	지식표현/ 언어지능	시각 지능	지능형 에이전트	복합 지능
전 체		(100)	87.0	10.0	4.0	2.0	2.0	1.0
인공 지능 분야별	추론 및 기계학습	(71)	84.5	14.1	4.2	2.8	2.8	0.0
	지식표현 및 언어지능	(31)	77.4	19.4	12.9	3.2	6.5	0.0
	청각 지능	(12)	100.0	0.0	0.0	0.0	0.0	0.0
	시각 지능	(22)	86.4	9.1	4.5	9.1	4.5	0.0
	복합 지능	(14)	92.9	0.0	0.0	0.0	0.0	7.1
	지능형 에이전트	(14)	78.6	14.3	14.3	14.3	14.3	0.0
	인간-기계 협업	(9)	100.0	0.0	0.0	0.0	0.0	0.0
	지능형 반도체	(3)	100.0	0.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	100.0	0.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	75.0	12.5	0.0	12.5	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0	0.0
	모름/무응답	(7)	100.0	0.0	0.0	0.0	0.0	0.0

〈18. AI SW 개발을 위한 데이터 확보가 필요한 분야-1) AI SW〉

(단위: 개, %)

구 분	사례수	추론 및 기계 학습	없다	지식 표현/ 언어 지능	시각 지능	지능형 에이 전트	청각 지능 (음성 및 소리 인식)	복합 지능	인간· 기계 협업 (HCI)	
전 체	(100)	49.0	43.0	17.0	7.0	6.0	3.0	3.0	2.0	
인공 지능 분야별	추론 및 기계학습	(71)	59.2	36.6	16.9	8.5	7.0	2.8	1.4	1.4
	지식표현 및 언어지능	(31)	41.9	41.9	38.7	12.9	9.7	6.5	0.0	3.2
	청각 지능	(12)	41.7	33.3	33.3	33.3	8.3	25.0	0.0	0.0
	시각 지능	(22)	40.9	40.9	31.8	31.8	13.6	13.6	4.5	4.5
	복합 지능	(14)	28.6	57.1	7.1	7.1	0.0	7.1	21.4	7.1
	지능형 에이전트	(14)	42.9	28.6	28.6	28.6	42.9	7.1	7.1	0.0
	인간-기계 협업	(9)	33.3	55.6	11.1	0.0	11.1	0.0	11.1	22.2
	지능형 반도체	(3)	33.3	66.7	0.0	0.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	33.3	66.7	0.0	0.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	37.5	37.5	12.5	25.0	12.5	0.0	0.0	0.0
	양자컴퓨팅	(1)	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
모름/무응답	(7)	14.3	85.7	0.0	0.0	0.0	0.0	0.0	0.0	

〈19. AI SW 개발을 위한 데이터 확보가 필요한 분야-2) 적용 분야〉

(단위: 개, %)

구 분		사례수	없다	의료	금융	국방	제조	유통	공공 서비스 및 복지	교육	에너지
전 체		(100)	42.0	32.0	11.0	10.0	8.0	5.0	7.0	5.0	4.0
인공 지능 분야별	추론 및 기계학습	(71)	35.2	35.2	15.5	11.3	11.3	5.6	7.0	7.0	2.8
	지식표현 및 언어지능	(31)	38.7	29.0	16.1	16.1	12.9	6.5	9.7	9.7	0.0
	청각 지능	(12)	33.3	50.0	8.3	8.3	0.0	16.7	0.0	0.0	0.0
	시각 지능	(22)	40.9	36.4	13.6	13.6	4.5	13.6	0.0	9.1	9.1
	복합 지능	(14)	57.1	28.6	0.0	7.1	0.0	7.1	7.1	0.0	7.1
	지능형 에이전트	(14)	28.6	35.7	21.4	7.1	7.1	21.4	7.1	7.1	14.3
	인간-기계 협업	(9)	55.6	33.3	0.0	22.2	11.1	0.0	0.0	0.0	22.2
	지능형 반도체	(3)	66.7	33.3	0.0	0.0	0.0	33.3	0.0	0.0	0.0
	뉴로모픽칩	(3)	66.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	고성능컴퓨팅	(8)	37.5	50.0	25.0	12.5	12.5	37.5	12.5	0.0	0.0
	양자컴퓨팅	(1)	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
모름/무응답	(7)	85.7	14.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

〈20. AI SW 개발을 위한 데이터 확보가 필요한 분야-2) 적용 분야(계속)〉

(단위: 개, %)

구 분		사례수	안전	홈	웰니스	교통	문화 관광	사무 관리	법률	농·축산· 어업
전 체		(100)	4.0	3.0	3.0	2.0	2.0	2.0	1.0	1.0
인공 지능 분야별	추론 및 기계학습	(71)	5.6	2.8	4.2	2.8	2.8	2.8	1.4	0.0
	지식표현 및 언어지능	(31)	0.0	9.7	6.5	3.2	6.5	0.0	3.2	0.0
	청각 지능	(12)	0.0	16.7	0.0	8.3	0.0	0.0	0.0	0.0
	시각 지능	(22)	9.1	9.1	4.5	9.1	4.5	0.0	0.0	0.0
	복합 지능	(14)	0.0	7.1	0.0	7.1	0.0	0.0	0.0	0.0
	지능형 에이전트	(14)	14.3	7.1	0.0	14.3	0.0	0.0	0.0	0.0
	인간-기계 협업	(9)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	지능형 반도체	(3)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	뉴로모픽칩	(3)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	33.3
	고성능컴퓨팅	(8)	12.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	양자컴퓨팅	(1)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	모름/무응답	(7)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

# 참 고 문 헌

## [국내문헌]

IIPT(2017), “일본의 인공지능(AI) 정책 동향과 실행전략”.

과학기술정보통신부(2017), “2017년 지능정보사회 선도를 위한 AI R&D”. 《보도자료》.

\_\_\_\_\_ (2017), “연구개발(R&D) 성과제고를 위한 연구데이터 공유방안”. 《보도자료》.

\_\_\_\_\_ (2018), “인공지능(AI) 개발 핵심 인프라를 중소·벤처기업에게 지원”. 《보도자료》.

김준연(2016), 전통 산업의 디지털 전환: 기회인가, 위기인가? 디지털 전환의 개념, 유형 그리고 조건, SPRI 컬럼.

김민식·최주환(2017), 산업 혁신의 관점에서 바라보는 제4차 산업혁명에 대한 이해, 정보통신방송정책, 제29권 8호 통권 645호.

\_\_\_\_\_ (2016), 제4차 산업혁명과 Industrial IoT·Industrial Internet의 이해, 정보통신방송정책, 제28권 12호 통권 626호.

김민식·손가녕(2017), 제4차 산업 혁명과 디지털 트랜스포메이션 (Digital Transformation)의 이해, 정보통신방송정책, 제28권 12호 통권 626호.

김민식(2013), 플랫폼으로 진화하는 카카오의 사례 분석과 시사점, 방송통신정책지. 제25권 13호 통권 558호.

김인숙·남유선(2016), 4차 산업혁명, 새로운 미래의 물결, 호이테북스.

김호인(2017), “스마트팩토리, AI으로 날개를 달다”, 포스코경영연구원.

박태준(2017), Open Access: Open Access 소개, Postech 박태준학술정보관, <http://libguides.postech.ac.kr/openaccess>.

신은정 외(2016. 8), 오픈 사이언스를 위한 연구 성과물 공개정책과 과제, 과학기술정책연구원.

신은정·정원교(2017), 오픈사이언스정책의 확산과 시사점. STEPI Insight, (216), 1-39.

신은정(2017), 오픈사이언스 정책 환경 진단 및 시사점, 과학기술정책연구원.  
 오픈사이언스 포럼 발표자료 (2016).  
 이지효(2016), 「대담한 디지털 시대」, 알에이치코리아.  
 조병호(2013). “디지털 큐레이션 서비스 동향”. 정보통신산업진흥원.  
 최근우·송기선 & 강요셉(2016), “딥러닝 기술의 이해와 연구개발 정책과제”, 한국과학기술  
 기획평가원, Issue paper 2016-08.  
 최병삼·김창욱·조원영(2014), 「플랫폼, 경영을 바꾼다」, 삼성경제연구소.  
 한국지식재산연구원(2016), “AI(AD) 기술 및 정책 동향”, 《이슈페이퍼》.  
 한겨레 신문(2017. 1. 16), AI·빅데이터 활용 ‘스마트공장’ 제조업 생산방식 혁명 이끈다.  
 ETRI(2015. 10) “AI 기술과 산업의 가능성”, 《이슈리포트》.  
 IITP(2015), “미국의 AI(AD) 기술 R&D추진 동향.  
 \_\_\_\_ (2015), “웨어러블스마트디바이스 예비타당성 사업기획보고서”, 《ICT 디바이스 과제기  
 획보고서》.  
 \_\_\_\_ (2016), “미국의 빅데이터 전략계획”, 《해외 ICT R&D 정책동향》, 2016-12호.  
 \_\_\_\_ (2016), “2015 연방정부의 과학/공학 분야 지원 통계”, 《KIAT Washington DC》, GT 5차  
 년 21호.  
 \_\_\_\_ (2016), “ICT R&D 중장기 기술로드맵 2022”.  
 \_\_\_\_ (2017), “2016년도 정보통신 기술수준조사보고서”.  
 \_\_\_\_ (2017), “일본의 AI(A) 정책 동향과 실행전략”, 《해외 ICT R&D 정책동향》, 2017-01호.  
 NIPA(2013), “EU Horizon 2020 정책 분석”, 《신성장동력 웹진》.  
 포스코경영연구원(2017. 12), 『중국 AI 기술, 미국을 추월할까?』.  
 코트라(2017. 12), 『중국 AI산업 G2로 부상』.  
 뉴스 Y(2017. 10), 『일본 8대 전자 기업들의 AI 투자 현황』.  
 조선비즈(2017. 7), 『폐복은..(중략).. ‘빅데이터 장악작전’』.  
 과학기술정책연구원(2017. 5), 『일본의 제4차 산업혁명 대응 정책과 시사점』.  
 Tech M(2017. 2), 『플랫폼 승자가 모든 산업 생태계 주도한다』.  
 KIET(2017. 1), 『일본의 4차 산업혁명 대응 실태와 시사점』.  
 \_\_\_\_ (2016), 『현실 속으로 확장하고 있는 인공지능』.

정보통신기술진흥센터(2017. 1), 『일본의 AI 정책동향과 실행전략』.  
 \_\_\_\_\_(2015), 『국내 인공지능 실태조사』.  
 \_\_\_\_\_(2016. 10), 『디지털컨텐츠 ICT디바이스 AI 정보보호분야 R&D 중장  
 기 기술 로드맵 2022』.  
 한국과학기술기획평가원(2017. 1), 『지능정보사회 선도 AI 프로젝트』.  
 한국정보화진흥원(2016. 12), 『새로운 기술, 새로운 세상 지능정보사회』.  
 \_\_\_\_\_(2017), 『우리나라 AI 기업 현황 조사』.  
 \_\_\_\_\_(2017. 2), 『빅데이터 동향과 이슈』.  
 \_\_\_\_\_(2017. 6), 『AI시대의 정부- AI이 어떻게 정부를 변화시킬 것인가』.  
 \_\_\_\_\_(2017), 제 4차 산업혁명 시대, 개방형 데이터 플랫폼 진화 방향에 대한  
 제언, Global Open Data, Now Issue Paper Vol. 01.

**[해의문헌]**

A. T. Kearney(2016). Digital Transformation 방법론.  
 BCC Research(2015). Smart Machines: Technologies and Global Markets, Anadre McWilliams,  
 IAS094A.  
 BI Intelligence(2017). The Fully Autonomous Car Update, <https://www.businessinsider.com>.  
 Budhathoki, N. R., & Haythornthwaite, C. (2013). Motivation for open collaboration: Crowd  
 and community models and the case of OpenStreetMap. American Behavioral  
 Scientist, 57(5), 548-575.  
 Capgemini(2016). Travelling to the future with Digital Platforms.Clark, B. Y., Brudney, J.  
 L., & Jang, S. G. (2013). Coproduction of government services and the new  
 information technology: Investigating the distributional biases. Public Administration  
 Review, 73(5), 687-70.  
 Crouzier, T.(2015). "Science Ecosystem 2.0: How will Change Occur?", European Commission.  
 DASTI(2014). Denmark's National Strategy for Open Access, Danish Agency for Science.  
 Forbes(2017). How AI And Machine Learning Are Helping Drive The GE Digital

Transformation.

Gartner(2016a). Building a Digital Business Technology Platform.

\_\_\_\_\_ (2016b). Every Organization Needs a Digital Platforms Strategy.

\_\_\_\_\_ (2016C). Three Styles of Digital Business Platforms.

GE report(2017). I Machine, You Human: How AI Is Helping GE Build A Powerhouse Of Knowledge.

IDC(2015). Digital Transformation(DX): An Opportunity and an Imperative. IDC, <http://www.idc.com/prodserv/3rd-platform/>.

IBM(2011). Digital transformation Creating new business models where digital meets physical.

Kraker, P., Leony, D., Reinhardt, W., & Beham, G.(2011). The case for an open science in technology enhanced learning. *International Journal of Technology Enhanced Learning*, 3(6), 643-654.

Laurie, Donald L., Yves L. Doz, and Claude P. Sheer(2006). "Creating new growth platforms." *harvard business review* 84.5 (2006): 80-90.

Mergel, I(2015). Open Collaboration in the public sector: The case of social coding on GitHub, *Government Information Quarterly*.

MIT business review(2017). GE build an AI workforce.

NITRD(2016). "The Federal BIG DATA Research and Development Strategic Plan".

NSTC(2016), PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE, Executive Office of the President National Science and Technology Council.

OECD(2011). "Open Science: Policy Challenges and Opportunities", internal working document, Country Studies and Outlook Division, Directorate for Science, Technology and Industry, OECD, Paris.

Parker, G. G., Van Alstyne, M. W., & Choudary, S. P.(2016). Platform revolution: How networked markets are transforming the economy and how to make them work for you. WW Norton & Comp.

Siemens(2014). "Facts and Forecasts: Boom for Learning Systems".

Tansley, S., & Tolle, K. M.(2009). The fourth paradigm: data-intensive scientific discovery

(Vol. 1). T. Hey (Ed.). Redmond, WA: Microsoft research.

Tractica(2015). Artificial Intelligence for Enterprise Applications: Deep Learning, Predictive Computing, Image Recognition, Speech Recognition, and Other AI Technologies for Enterprise Markets- Global Market Analysis and Forecasts.

WEF(2015). Industrial Internet of Things: Unleashing the Potential of Connected Products and Services.

\_\_\_\_\_(2016). The fourth Industrial Revolution: what it means, how to respond. <http://blog.goodelearning.com/it4it/it4it-and-digital>.

Mckinsey&Company(2017). 『Artificial intelligence: Implications for China』.

미국 국가과학기술위원회(2016. 10). 『AI 국가 연구 개발 전략 계획』.

중국 중앙인민정부 『“인터넷+” AI 3개년 행동실시방안』.

경제산업성(2016), 『새로운 산업 구조 비전: 제4차 산업 혁명을 선도하는 일본의 전략』.

인민일보(2015. 3), 『기업리더, ‘인터넷+’ 열띤 토론 벌여』.

중국산업규획망 『중국 빅데이터 산업 발전 현황』.

White House, Executive Office of the President, National Science and Technology Council, Committe on Technology(2016. 10). 『PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE』.

CNBC 『China’s artificial intelligence technology is fast catching up to the US, Goldman Sachs says』.

451 Research(2016), Cloud Trends and Expectations, Black & White Paper, <https://www.vmware.com>, 2017. 12. 31. 검색.

### [신문기사]

박수형(2017), “제조업, AI 서비스 덧붙여야 산다”, ZDNet Korea, [http://www.zdnet.co.kr/news/news\\_view.asp?artice\\_id=20170328102621](http://www.zdnet.co.kr/news/news_view.asp?artice_id=20170328102621), 2017. 12. 17. 검색.

송고시간(2017), “AI ‘왓슨’진료 자리 잡나… 국내 병원 잇따라 도입(종합)”, 연합뉴스, <http://www.yonhapnews.co.kr/bulletin/2017/03/15/0200000000AKR2017031518075106>

3.HTML, 2017. 12. 17. 검색.

송고시간(2018), 국내 AI 스피커 100만 시대 ‘활짝’…진짜 승부는 올해, 연합뉴스, <http://www.yonhapnews.co.kr/bulletin/2018/01/10/0200000000AKR20180110185400017.HTML>, 2018. 3. 11. 검색.

이완(2017), “AI·빅데이터 활용 ‘스마트공장’ …제조업 생산방식 혁명 이끈다”, 한겨레, [http://www.hani.co.kr/arti/economy/economy\\_general/778984.html](http://www.hani.co.kr/arti/economy/economy_general/778984.html), 2018. 1. 2. 검색.

임민철(2017), “구글, 2세대 머신러닝 전용칩 ‘클라우드TPU’ 공개”, ZDNet Korea, [http://www.zdnet.co.kr/news/news\\_view.asp?article\\_id=20170518092656](http://www.zdnet.co.kr/news/news_view.asp?article_id=20170518092656), 2017. 12. 17. 검색.

\_\_\_\_\_(2017), “MS, 딥러닝 툴킷 새버전 오픈소스로 공개”, ZDNet Korea, [http://www.zdnet.co.kr/news/news\\_view.asp?article\\_id=20170602102011](http://www.zdnet.co.kr/news/news_view.asp?article_id=20170602102011), 2017.12.18. 검색.

임유경(2017), “페이스북, 번역 속도 9배 높인 딥러닝 기술 공개”, ZDNet Korea, [http://www.zdnet.co.kr/news/news\\_view.asp?article\\_id=20170510065745](http://www.zdnet.co.kr/news/news_view.asp?article_id=20170510065745), 2017. 12. 17. 검색.

조권형(2017), “금융권에 부는 ‘왓슨’ 바람”, 서울경제, <http://www.seaily.com/NewsView/1OERCQE MF6>, 2017. 12. 17. 검색.

전자신문(2016. 12), 『공공데이터로 융합 신산업 창출..(중략)..기본계획 발표』.

### [웹페이지]

Business Insider(2017), Audi is laying the groundwork for drivers to be media consumers, <http://www.businessinsider.com/>, 2018. 3. 6. 검색.

EOSC(2017), EOSCpilot Brief, <https://eoscipilot.eu.>, 2018. 3. 6. 검색.

DFG(2017). <http://www.dfg.de>, 2017. 9. 12. 검색.

Human Brain Project, “Short Overview of the Human Brain Project”, <https://www.humanbrainproject.eu>, 2017. 11. 28일 검색.

IDC(2016). “World Wide Semiannual Big Data and Analytics Spending Guide”, <https://www.idc.com>, 2017. 9. 12. 검색.

Internet of Things Institute(2017). “Siemens’ AI Work Delivers Competitive Advantage in IIoT”, <http://www.ioti.com>, 2017. 12. 31. 검색.

Kotra(2017), “상용화를 목전에 앞둔 미국 증강인간기술”, <http://news.kotra.or.kr>, 2017. 9. 12 검색.

Lant, K.(2017), Goldman Sachs Report: China’s AI Sector is Catching Up to the U.S, Futurism, <https://futurism.com>, 2017. 9. 12 검색.

NVIDIA(n.d.), “비즈니스의 혁신을 일으키는 ‘딥러닝’ 첫번째 이야기”, <http://blogs.nvidia.co.kr>, 2017. 12. 17. 검색.

OpenAIRE, Project factsheets: The OpenAIRE2020 Project, <https://www.openaire.eu>, 2017. 10. 17일 검색.

OECD(2017). STI Micro-data Lab: Intellectual Property Database, <http://oe.cd>.

Open Science and Research, Open Science in Finland, <http://openscience.fi>, 2017. 11. 1 검색.

Open Science ASAP, <http://openscienceasap.org>, 2017. 11. 29 검색.

Re3data(2017. 5. 16). <http://www.re3data.org>, 2017. 11. 1 검색.

Roarmap, “Policies Adopted by Quarter”, <http://roarmap.eprints.org>, 2017. 9. 12 검색.

Roarmap, “Individual Elements of Horizon 2020: Humber of policies which align to each section of the H2020 OA Policy” <http://roarmap.eprints.org>, 2017. 9. 12 검색.

Siemens(2017). “Artificial Intelligence: Optimizing Industrial Operations” <https://www.siemens.com>, 2017. 9. 12 검색.

Siemens(2014). “Facts and Forecasts: Boom for Learning Systems”. <https://www.siemens.com>, 2017. 9. 12. 검색.

The Innovation Policy Platform(2017). France &#8211; Open Science Country note, <https://www.innovationpolicyplatform.org>, 2017. 9. 12 검색.

The Innovation Policy Platform(2017). Netherlands - Open Science Country note, <https://www.innovationpolicyplatform.org>, 2017. 9. 12 검색.

Strategy Analytics(2017), Global Smart Speaker Vendor & Voice OS Market Share by Region Q2 2017, 보고서, <https://www.strategyanalytics.com>, 2017. 12. 31 검색.

IDC(2017). Big Data and Business Analytics Revenues Forecast to Reach \$150.8 Billion This Year, Led by Banking and Manufacturing Investments, According to IDC, <https://www.idc.com>, 2017. 12. 31 검색.

## ● 저 자 소 개 ●

---

### 최 계 영

- 서울대학교 국제경제학 학사
- Univ. of California, Davis 경제학 박사
- 현 정보통신정책연구원 선임연구위원

### 이 학 기

- 경희대학교 경제학 학사
- University of Pittsburgh 경제학 박사
- 현 정보통신정책연구원 부연구위원

### 김 경 훈

- KAIST 경영공학 학사
- KAIST 경영공학 박사
- 현 정보통신정책연구원 부연구위원

### 김 민 식

- 고려대학교 경제학/경영학 학사
- 한국과학기술원 경영학 석사
- 고려대학교 과학기술학 박사과정 수료
- 현 정보통신정책연구원 부연구위원

### 정 원 준

- 성균관대학교 법학 학사
- 고려대학교 법학 석사
- 현 정보통신정책연구원 연구원

### 양 수 연

- 서울대학교 소비자학 학사
- 서울대학교 환경계획학 석사
- 현 정보통신정책연구원 연구원

### 이 가 희

- Univ. of Canberra 학사
- 울산과학기술원 비즈니스 분석 석사
- 현 정보통신정책연구원 연구원

방송통신정책연구 17-방통-45

지능정보기술 R&D의 선도형(First Mover) 연구  
촉진과 성과 확산 가속화를 위한 지원 체계 및  
제도 개선 방안 연구

(A Study on Improving the Information Technology  
R&D to promote for First Mover Research and  
Resulting Diffusion)

---

2018년 1월 일 인쇄

2018년 1월 일 발행

발행인 과학기술정보통신부 장관

발행처 과학기술정보통신부

경기도 과천시 관문로 47 정부과천청사

Homepage: [www.msit.go.kr](http://www.msit.go.kr)

인 쇄 인 성 문 화

---