

**FOCUS**

# 21대 국회에 바라는 SW산업 이슈

21대 국회 최대 숙제는 'AI·언택트 시대 맞는 SW·디지털 법제도·생태계 만들기'

21대 국회에 바라는 SW산업 이슈

**TREND**

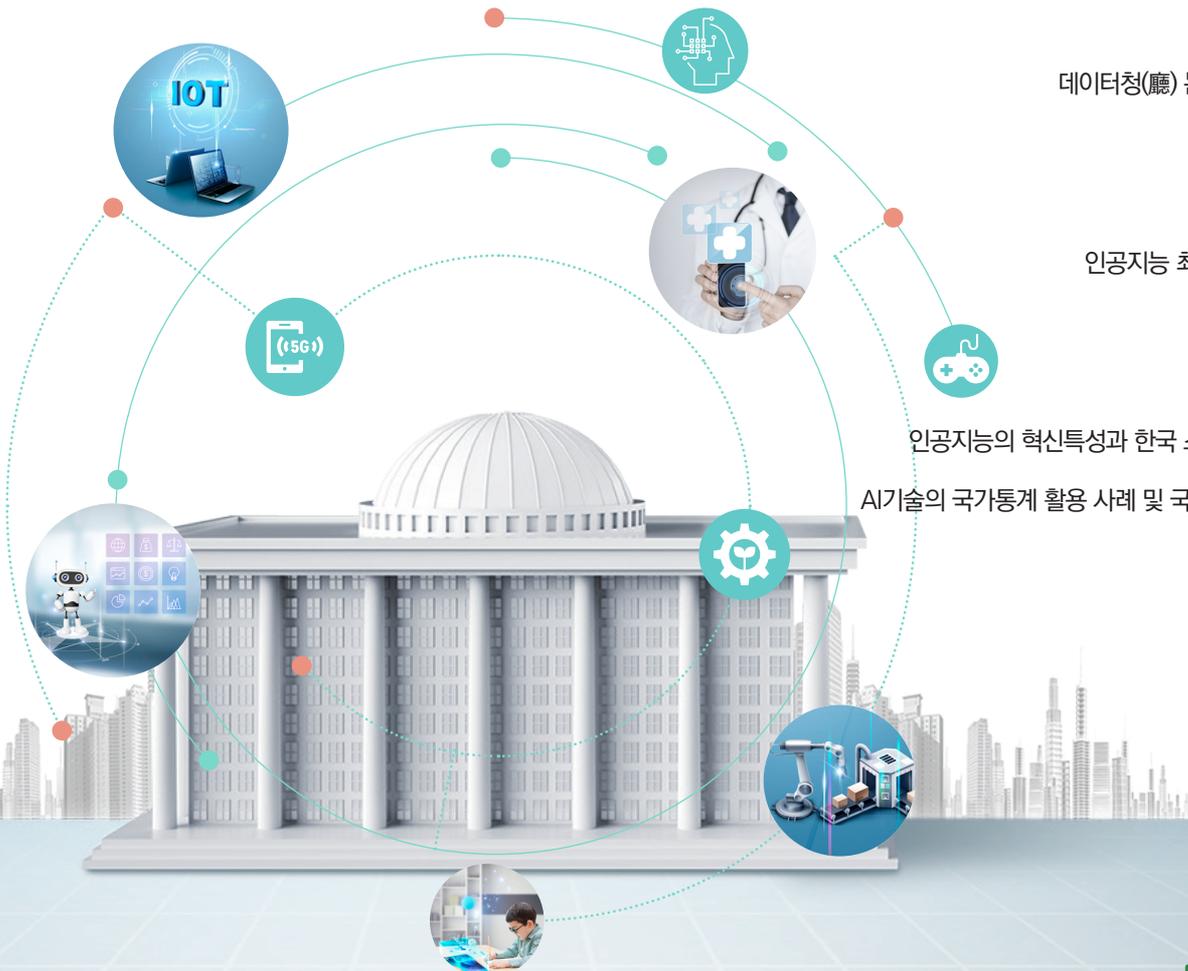
데이터청(廳) 논의에서 선결문제

**AI BRIEF**

인공지능 최신 동향과 시사점

**ISSUE**

인공지능의 혁신특성과 한국 스타트업 사례연구  
 AI기술의 국가통계 활용 사례 및 국내 도입 촉진 방안

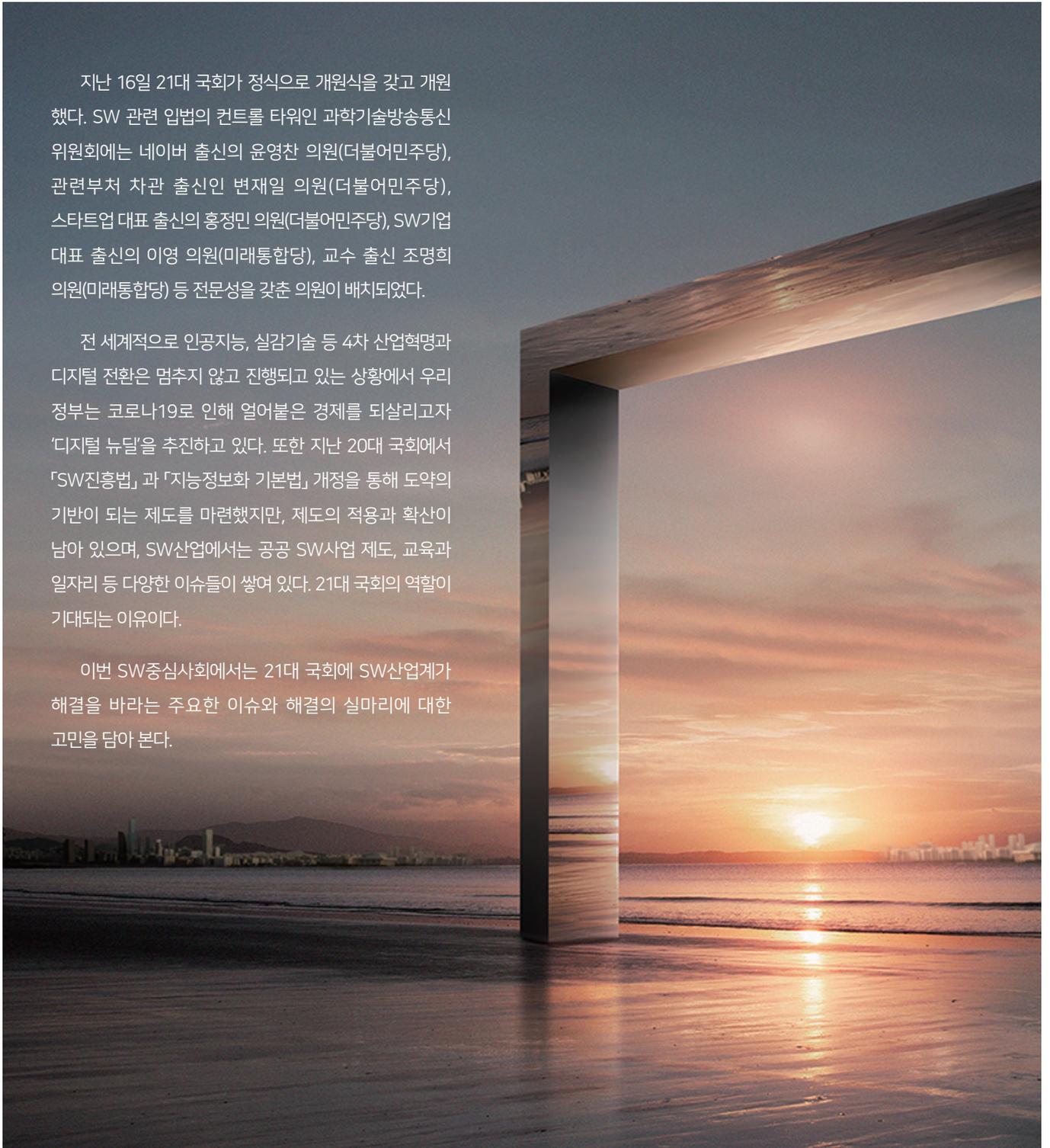


# 21대 국회에 바라는 SW산업 이슈

지난 16일 21대 국회가 정식으로 개원식을 갖고 개원했다. SW 관련 입법의 컨트롤 타워인 과학기술방송통신 위원회에는 네이버 출신의 윤영찬 의원(더불어민주당), 관련부처 차관 출신인 변재일 의원(더불어민주당), 스타트업 대표 출신의 홍정민 의원(더불어민주당), SW기업 대표 출신의 이영 의원(미래통합당), 교수 출신 조명희 의원(미래통합당) 등 전문성을 갖춘 의원이 배치되었다.

전 세계적으로 인공지능, 실감기술 등 4차 산업혁명과 디지털 전환은 멈추지 않고 진행되고 있는 상황에서 우리 정부는 코로나19로 인해 얼어붙은 경제를 되살리고자 '디지털 뉴딜'을 추진하고 있다. 또한 지난 20대 국회에서 「SW진흥법」과 「지능정보화 기본법」 개정을 통해 도약의 기반이 되는 제도를 마련했지만, 제도의 적용과 확산이 남아 있으며, SW산업에서는 공공 SW사업 제도, 교육과 일자리 등 다양한 이슈들이 쌓여 있다. 21대 국회의 역할이 기대되는 이유이다.

이번 SW중심사회에서는 21대 국회에 SW산업계가 해결을 바라는 주요한 이슈와 해결의 실마리에 대한 고민을 담아 본다.



## CONTENTS

### 04 동향 Trend

데이터청(廳) 논의에서 선결문제

On the Discussion of Establishing Data Administration Office

### 12 AI 브리프 AI Brief

인공지능 최신 동향과 시사점

Recent Trends of AI and Implications

### 22 포토 에세이 Photo Essay

사진산책 중간(中間)

Your Moment with Photos : The Stopover

### 24 포커스 Focus : 21대 국회에 바라는 SW산업 이슈

21대 국회 최대 숙제는 'AI·언택트 시대 맞는 SW·디지털 법제도·생태계 만들기'  
 The Biggest Task of 21<sup>st</sup> National Assembly : Establishing Legal & Ecosystem for the AI and Untact

21대 국회에 바라는 SW산업 이슈

SW Industry Issues for 21<sup>st</sup> National Assembly

### 42 이슈 Issue

인공지능의 혁신특성과 한국 스타트업 사례연구

Innovation Patterns of AI and Cases of Startups in Korea

AI기술의 국가통계 활용 사례 및 국내 도입 촉진 방안

Use Cases of AI Technology in National Statistics & Suggestions for Promoting Domestic Adoption



발행인 박현재 소장(소프트웨어정책연구소)  
 편집위원장 김덕현 외래교수(세종사이버대학교)  
 편집위원회 권오병 교수(경희대학교)  
 김숙경 초빙교수(KAIST)  
 김철한 교수(대전대학교)  
 김형석 교수(건국대학교)  
 방은주 기자(지디넷코리아)  
 윤기영 대표(FnS컨설팅)  
 장윤희 교수(단국대학교)  
 정충식 교수(경성대학교)  
 조영임 교수(가천대학교)

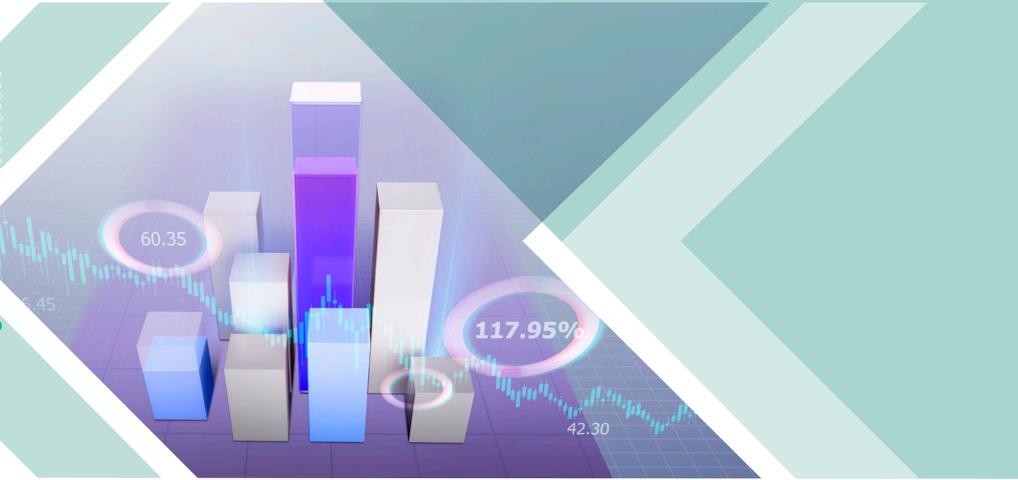
홈페이지 [www.spri.kr](http://www.spri.kr)

전화 031-739-7300

주소 경기도 성남시 분당구 대왕판교로  
 712번길 22 글로벌R&D센터  
 연구동 A 4층, 소프트웨어정책연구소

디자인 제작 (주)늘품플러스 [www.nplplus.co.kr](http://www.nplplus.co.kr)





# 데이터청(廳) 논의에서 선결문제

이종주 | 연구원 | lj@spri.kr

4차 산업혁명 이후 데이터의 중요성은 커져 가고 있다. 국회에서 데이터청 설립 논의가 진행 중이다. 하지만 데이터의 권리에 대한 명확한 정립이 없어서 데이터를 보유한 민간기업이 시장에 데이터를 제공할 동인이 없다. 또한 데이터 산업을 육성하기 위한 국가 기본계획과 그것을 성공적으로 안착시키기 위한 법률적 근거가 없다. 데이터청 설립에 앞서 이러한 문제들을 법/제도 제정이나 정비를 통해 준비할 필요가 있다.

In the domestic industrial market, the importance of data has increased since the 4th industrial revolution. So there is a discussion of the establishment of the Korean Data Property Office at the National Assembly. However, there is no clear definition of the rights of the data, so there is no driving force for private companies with data to provide data to the market. In addition, there is no national basic plan for fostering the data industry, so there is no strategy to foster the data industry. These issues need to be addressed legislatively prior to the establishment of the Korean Data Property Office.

## ② 데이터산업 육성의 중요성

스위스 다보스 포럼에서 4차 산업혁명이 의제로 다루어진지 4년이 지난 지금 사방팔방(四方八方)에서 ‘데이터, 데이터, 데이터’ 하는 소리가 빗발치고 있다. 누군가는 데이터를 미래 산업의 원유라고 칭하기도 하고, 또 다른 누군가는 미래 산업의 쌀(米)이라고 칭하기도 한다. 최근 대통령께서 한국판 디지털 뉴딜을 선언하면서 ‘데이터를 수집, 축적, 활용하는 데이터 인프라 구축을 국가적 사업으로 추진하겠다’고 밝힌 바 있다. 또한 정부는 7월 14일(화) 대통령 주재 한국판 뉴딜 국민보고대회(제7차 비상경제회의)를 개최하여 「한국판 뉴딜 종합계획」을 확정하고 발표하였는데,<sup>1</sup> ‘데이터담’을 설립하기 위해 준비하고 있음을 밝혔다. 최근 이처럼 국가 차원에서 데이터 산업을 육성하고자 하는 가운데 일부 국회의원들은 ‘데이터청(廳)’ 설립을 추진하고 있다.

## ② 데이터청(廳) 논의

최근 국회 여야 양쪽에서 데이터청(廳)에 대한 논의가 진행되고 있다. 이보다 앞서 KAIST 정재승 교수는 모 언론 칼럼을 통해 데이터청 설립 필요성을 주장한 바 있다.<sup>2</sup> “정부가관이 데이터를 전략적으로 모으고, 효율적으로 관리하고, 분쟁을 공정하게 해결하는 시스템을 구축해야 한다.”라고 데이터청의 역할까지 제시하였다. 강원도지사와 민간 연구기관인 ‘여시재’ 원장을 역임한 이광재 국회의원은 당선 전부터 데이터청 설치 필요성을 주장하였고, 당선 직후 데이터청 설립에 대한 당론을 모으고 있다. 야당에서도 데이터청 설립 목소리가 나오고 있어, 국회 차원의 논의가 확대될 것으로 예상된다.

■ [표 1] 데이터청 설립 관련 정치권의 논의내용

일자 (2020년)	소속	성명/직책	발언 내용
6월 8일	더불어민주당	이광재 국회의원	“지난주에 데이터청 설립과 관련해 (당 내부에서) 비공개 토론을 했었는데 의견을 모으고 있는 중” “데이터 관리와 관련해 컨트롤 타워가 필요하다는 점은 컨센선스(합의)를 이뤘다.” <sup>3</sup>
6월 11일	미래통합당	김종인 비상대책위원장	“정부가 갖고 있는 데이터나 민간이 소유한 데이터를 종합 관리해서 데이터가 필요한 기업이나 개인이 효율적으로 활용할 수 있는 기구를 만드는 것이 선결 과제라고 생각해 데이터청 설립을 건의했다.” <sup>4</sup>

1 관계부처합동(2020.7.14.), “「한국판 뉴딜」 종합계획 - 선도국가로 도약하는 대한민국으로 대전환 -”

2 중앙선데이(2019.4.20.), “[정재승 칼럼] 데이터청과 데이터거래소를 설립하자”

3 뉴스토마토(2020.6.8.), “이광재 “데이터청 설립 피할 수 없는 대세””

4 MBC뉴스(2020.6.11.), “통합당, 좌담회 열고 ‘데이터청’ 논의”

일자 (2020년)	소속	성명/직책	발언 내용
6월 12일	미래통합당	원희룡 제주특별자치도지사	“제주도는 미래전략국을 통해 지방정부 차원에서의 데이터 생산과 공유, 활용을 적극 추진해 왔다.” “지방정부가 지역데이터를 국가에 제공하고, 국가와 데이터를 공유하는 데 있어 데이터 자치분권에 대한 고민이 필요하다.” “지방정부에 데이터 전문 직렬 공무원들을 두고, 데이터 기반 행정을 펼 수 있게 해야 한다.” <sup>5</sup>
6월 16일	더불어민주당	김태년 원내대표	“데이터를 제공하고 거래하는 기준을 명확히 정립하고 정보 유통과 활용, 개인정보 보호 영역까지 포괄하는 제도 설계가 필요한 만큼, 정부와 함께 데이터청 설립 방안을 검토하겠다.” <sup>6</sup>
6월 25일	국민의당	안철수 당대표	데이터청 설립에 다소 부정적이나, 그 필요성에 대해서는 인정하였다. “대통령 또는 국무총리 소속으로 해서 직접 모든 부처의 데이터를 공개할 수 있도록 하는 구조가 되어야 그나마 정부부처가 움직일까 말까 한다.” <sup>7</sup>
7월 7일	더불어민주당	이광재 국회의원	“데이터청으로 가면 전 부처를 관할하는 거대한 일을 하기가 쉽지 않기 때문에 청보다 위상이 강한 데이터부가 (해법) 될 수 있을 것”이라며 데이터부의 필요성을 주장함. <sup>8</sup>

※ 자료 : 신문기사 인용하여 작성

데이터청 설립 논의에서 이를 반대하는 목소리도 있다.<sup>9</sup> 데이터청이 설립되면 1) 데이터 산업의 혁신보다는 규제를 강화할 가능성이 높고, 2) 데이터 산업 육성은 데이터청 설치만으로 해결되는 것이 아니고 관계부처 협력을 통한 국가 전략 마련이 더 중요한 과업이므로 그 필요성에 의문을 제기하는 것이다. 더불어 정부와의 충분한 논의가 없었기에 국회의 협의만으로 정부부처가 설치되기는 어려움이 있을 것으로 보인다.

현재 정부는 기관·기업 내부에만 갇혀있던 데이터가 유통·활용될 수 있도록 하여 우리나라의 낙후된 데이터 생태계를 혁신할 목적으로 ‘빅데이터 플랫폼 및 센터 구축 사업’을 추진하고 있다. 이 사업은 공공과 민간이 협업하여 빅데이터 센터 등에서 수집된 데이터를 플랫폼에서 분석·유통하고 혁신 서비스를 발굴·확산하는 등 데이터 기반의 가치 창출 생태계 조성을 지원하고자 한다. 2019년 기준으로 10개의 플랫폼과 100여 개의 센터가 운영 중이다. 이중 플랫폼 및 센터 간 연결로 경제적 가치와 활용도가 높은 데이터를 생성하고 있다. 과기정통부는 단기적으로 2021년도까지 플랫폼의 안정화 및 고도화를 통해 데이터 기반 생태계 육성을 목표로 사업을 추진하고 있다.

<sup>5</sup> 이데일리(2020.6.12.), “원희룡 “김종인 데이터청 제안 환영, 제주 유치 희망””  
<sup>6</sup> YTN(2020.6.16.), “김태년 “디지털 경제 대전환...데이터청·거래소 설립 검토””  
<sup>7</sup> 뉴스시스(2020.6.25.), “안철수 “데이터청 하나만으로는 작은 생색 그칠 수도””  
<sup>8</sup> 전자신문(2020.7.7.), “이광재 의원 “데이터 컨트롤타워 ‘데이터부’ 필요””  
<sup>9</sup> 아시아경제(2020.7.13.), “[데스크칼럼] 데이터청을 반대하는 두 가지 이유”

[그림 1] 빅데이터 플랫폼/센터 구축 사업



※ 자료 : 빅데이터 플랫폼 통합지도 홈페이지(<https://www.bigdata-map.kr>)<sup>10</sup>

다만 정부에서 추진하고 있는 사업 역시 데이터의 활용에만 치중되어 다소 아쉬운 면이 있다. 민간 기업이 보유한 데이터를 시장에 내놓을 동인이 여전히 부족한 상태이다. 또한 다양한 정부부처에 소관 산업별로 분산 보관 및 관리되고 있는 데이터에 대한 전략적 수집, 축적, 활용방안이 마련되지 않고 있다.

## ④ 데이터 기본법 논의

### ○ 데이터와 관련된 법제도 현황

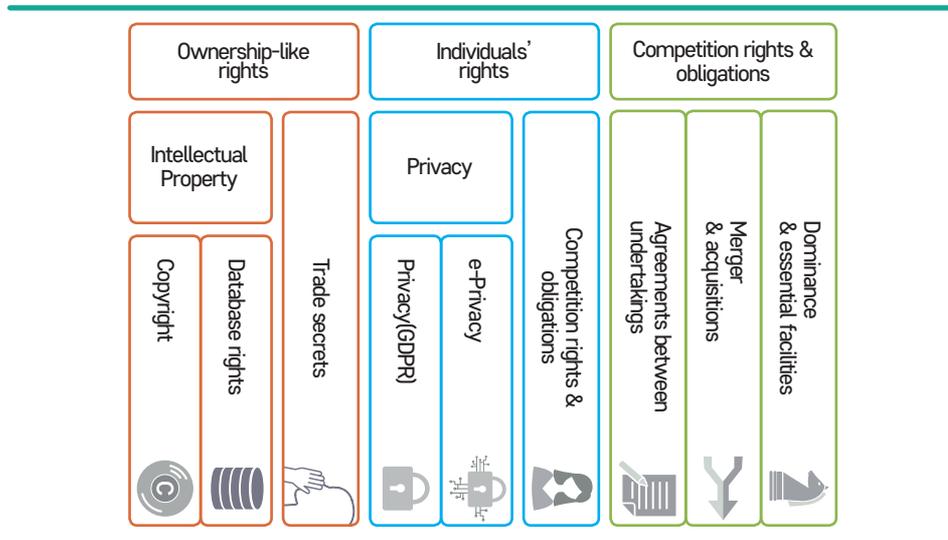
데이터 관련 법제도는 개인정보 영역과 비개인정보 영역으로 접근하고 있다.<sup>11</sup> 우선 개인정보에 대해서는 ‘데이터 3법’이라 일컫는 ‘개인정보보호법’, ‘정보통신망법’, ‘신용정보법’이 올해 8월 5일부터 시행된다. 이번 개정에 따라 ‘마이데이터’ 사업 등 정보주체의 데이터 이동권(Right to Data Portability)이

<sup>10</sup> 10개 분야별로 센터 등에서 구축된 데이터를 한 공간으로 최대한 모아 시장에서 요구하는 데이터로 가공·분석 및 유통하는 플랫폼을 지정하였다. 분야별 플랫폼 10개소와 이와 연계된 기관별 센터 100개소를 구축하는데 3년간 총 1,516억 원을 투입할 계획이며, 올해 총 640억원 규모로 구축 사업을 추진하고 있다.

<sup>11</sup> 이종주(2018), “데이터 소유권 동향”, 소프트웨어중심사회

법적 근거를 갖게 되었다.<sup>12</sup> 또한 가명정보, 익명정보 등을 통해서 개인정보의 비식별정보화를 통해 데이터의 활용 가능성을 높이고 있다. 특히 개인정보보호위원회의 관리·감독 기능 및 여러 법령에서 다루고 있는 개인정보 보호에 관한 유사 규정 또는 중복 규정이 개인정보 보호법으로 일원화되었고, 이에 대한 관리·감독 기구가 개인정보보호위원회로 통합되었다. 이번 법령개정으로 개인정보와 관련된 데이터 활용에서 정보 주체의 엄격한 ‘사전 동의’와 관련된 인격권 문제가 일부나마 정비된 것이다.

■ [그림 2] 데이터 권리



※ 자료 : EU White Paper(2017)

비개인정보 영역의 데이터 이슈와 관련해서는 아직 명확하게 정리되지 않은 상태이다. 올해 12월 10일부터 시행되는 「지능정보화 기본법」 제2조에는 데이터에 대한 정의가 명확하게 정리되었고, 제42조 및 제43조에는 데이터의 생산·관리·유통·활용의 활성화를 위해 조문이 제정되었다. 기본적으로 「지능정보화 기본법」은 데이터에 관한 정부의 정책 수립 계획과 지원방향을 위주로 제정되었다. 하지만 새로운 기술 발전과 함께 형성된 개념인 데이터에 대한 소유권 내지 재산권에 대한 논의가 없어 다소 아쉬움이 있다. 데이터 시장이 활성화되려면 데이터에 대한 권리가 명확해야 한다. 데이터에 대한 권리가 명확하지 않으면 민간기업이 보유한 데이터가 시장에 공급되는데 한계가 있기 때문이다. 또한 위법에 비정형 데이터의 수집, 축적, 활용에 대한 법제도적 접근이 전무한 부분도 아쉽다. 최근 데이터는 구조화된 데이터뿐만 아니라 반정형 또는 비정형 데이터도 디지털 데이터로 활용되고 있고, 실제로

12 한국데이터산업진흥원(2019), “데이터 이동권 도입 방안 연구”

후자가 훨씬 많은 양을 차지하고 있다.<sup>13</sup> 비정형 데이터는 정형 데이터에 비해 높은 차원의 코딩(가공)이 필수적이어서 지식재산권의 보호와 유사한 접근이 요구된다.

데이터에 대한 권리는 경쟁법적 관점도 있다(그림 2 참조). EU White Paper(2017)는 영업비밀을 소유권 유사권리로 규정하고 있다. 이는 부당한 방법으로 영업비밀을 침해하는 경우 발생하는 채권적 지위를 의미하여<sup>14</sup> 민간기업이 데이터를 시장으로 내놓을 동인을 제공하는 것은 아니다.

**지능정보화 기본법**

[시행 2021.6.10.] [법률 제17344호, 2020.6.9. 전부개정]

**제42조(데이터 관련 시책의 마련)** ① 정부는 지능정보화의 효율적 추진과 지능정보서비스의 제공·이용 활성화에 필요한 데이터의 생산·수집 및 유통·활용 등을 촉진하기 위하여 필요한 정책을 추진하여야 한다.

② 과학기술정보통신부장관은 다음 각호의 사항이 포함된 시책을 수립·시행하여야 한다. 다만 공공 데이터에 관한 사항은 「공공데이터의 제공 및 이용 활성화에 관한 법률」에 따른다.

1. 데이터 관련 시책의 기본방향
2. 데이터의 생산·수집 및 유통·활용
3. 데이터 유통 활성화 및 유통체계 구축
4. 데이터의 생산·수집 및 유통·활용에 관한 기술개발의 추진
5. 데이터의 표준화 및 품질제고
6. 데이터 전문인력 양성 및 데이터 전문기업 육성
7. 제2호부터 제6호까지와 관련한 자원의 확보
8. 그 밖에 데이터의 생산·수집 및 유통·활용을 위하여 필요한 사항

③ 과학기술정보통신부장관은 데이터의 효율적인 생산·수집 및 유통·활용을 위하여 표준화를 추진하여야 한다. 다만 「공공데이터의 제공 및 이용 활성화에 관한 법률」, 「산업표준화법」 등 다른 법률에 관련 표준이 있는 경우에는 그 표준을 따라야 한다.

**제43조(데이터의 유통·활용)** ① 정부는 데이터의 효율적인 생산·수집·관리와 원활한 유통·활용을 위하여 국가기관 등, 법인, 기관 및 단체와의 협력체계를 구축하고, 이를 위한 지원을 할 수 있다.

② 정부는 지능정보사회 구현을 위하여 원활한 유통과 활용이 필요한 다음 각호의 데이터를 생산·수집 또는 보유하고 있는 국가기관 등, 법인, 기관 및 단체를 지원할 수 있다. 다만 공공데이터에 관한 사항은 「공공데이터의 제공 및 이용 활성화에 관한 법률」에 따른다.

1. 국가적으로 보존 및 이용 가치가 있는 자료로서 학술, 문화, 과학기술, 행정 등에 관한 디지털화된 자료나 디지털화의 필요성이 인정되는 데이터
2. 국민 생활의 질적 향상과 복리 증진 및 안전을 위하여 필요한 데이터
3. 국가 경제·산업의 발전을 도모하고 국가경쟁력 확보 등을 위하여 필요한 데이터
4. 그 밖에 지능정보화 및 지능정보서비스의 발전을 위하여 필요한 데이터

③ 제2항 각호의 데이터의 생산·수집·유통·활용 등을 지원하기 위하여 지능정보사회원에 데이터통합 지원센터를 설치할 수 있다. 다만 공공데이터에 관한 사항은 「공공데이터의 제공 및 이용 활성화에 관한 법률」에 따른다.

④ 제2항에 따른 지원의 내용 및 방법 등에 관하여 필요한 사항과 제3항에 따른 데이터통합지원센터의 기능·운영 등에 관한 구체적인 사항은 대통령령으로 정한다.

<sup>13</sup> 정보통신정책연구원(2019), “데이터소유권에 관한 법·제도 및 정책 연구”

<sup>14</sup> 황의창·황광연(2009), “부정경쟁방지 및 영업비밀보호법”, 세창출판사

「공공데이터의 제공 및 이용 활성화에 관한 법률」 제1조에 의하면 ‘공공기관이 보유·관리하는 데이터(즉, 공공데이터)의 제공 및 그 이용 활성화에 관한 사항을 규정’하였는데, 이 역시 데이터 산업 육성이나 데이터의 기초적인 권리 창출과 보호 측면보다는 데이터 활용에 중점을 두고 있다. 정보통신정책연구원(2019)은 데이터의 소유권을 인정하기는 어려우나, 배타적 재산권의 인정 가능성을 시사하였다. 데이터를 가장 많이 수집하여 축적하고, 가공하여 활용하는 민간기업 데이터를 시장에 내놓게 하려면 데이터의 배타적 재산권 인정이 필요할 것으로 보인다. 이를 기반으로 데이터 산업과 관련 시장이 형성될 수 있을 것으로 기대한다. 현재 상황은 인공지능, AR/VR 기술로 달려가기 위해서 데이터의 중요성을 인지하여 고속도로를 건설하고 있으나, 그 위를 질주할 데이터 자체가 부족하거나, 고속도로 이용에 따른 한계비용을 감당할 기업이 없는 상태라고 할 수 있다.

○ 데이터와 관련된 타 분야 입법례

데이터와 관련해서 생겨나는 입법적 문제의 양상은 과거 1991년에 시행된 「환경정책기본법」의 논의와 비슷하다. 해당 법령은 다양하고 복잡해져가는 환경문제를 효과적으로 대응하고, 환경과 관련한 주요한 사안별로 개별법령을 제정하기 위해, 프레임을 갖추기 위해 제정되었다. 당시에 환경 문제와 관련된 다수의 부처에서 각기 해결방안을 논의하고, 개별법 제정으로 이를 대응하였기에 환경정책에 대한 교통정리가 요구된 상황이었다. 그래서 「환경정책기본법」은 환경과 관련된 법령의 지도 역할로 제정된 것이다. 위 법령은 기본법으로써 단순히 기본원칙 또는 이념에 대한 선언적 수준에 멈추어 있는 것이 아니라, 환경과 관련한 기본법으로 지위 확인, 관련 규정 정비, 현안 문제 해결 등 변화하는 환경(또는 환경산업)에 발맞추어 개정이 있었다.<sup>15</sup> 현재는 환경과 환경산업에 관한 가장 기본적인 법령으로써 자리매김하고 있다. 환경과 데이터는 그 간격이 적지 않으나, 비경합성과 비배제성으로 인한 공공재적 성격에 유사점이 있다. 또한 시대와 기술발전에 따라 환경문제가 변화를 겪듯이, 데이터 또한 시대와 기술발전에 따라 변화를 겪고 있으므로 기준을 세우고 변화를 대응해야 하는 공통점이 존재한다. 데이터에 관한 입법적 정비가 되어 있지 않은 현시점에 환경 분야의 사례를 참조한다면 보다 충실한 논의가 진행될 것이다.

■ [표 2] 환경정책기본법 목차

장	절	조	내용
제1장		제1조~제11조	총칙
2장	제1절	제12조~제13조	환경기준
	제2절	제14조~제39조	기본적 시책
	제3절	제40조~제41조	자연환경의 보전 및 환경
	제4절	제42조~제44조	분쟁 조정 및 피해 구제
	제5절	제45조~제53조	환경개선특별회계의 설치

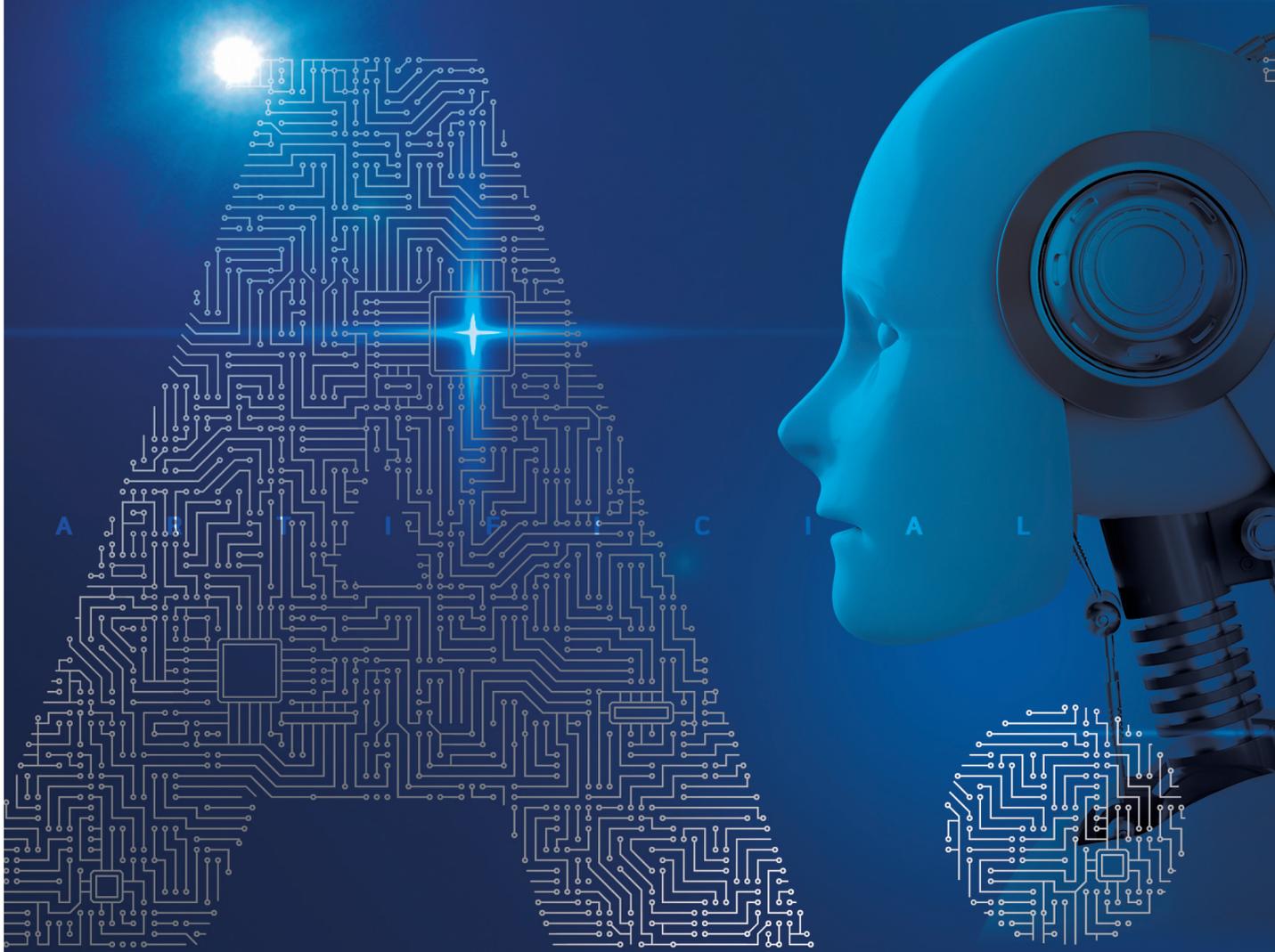
<sup>15</sup> 법제연구원(2009), “환경정책기본법의 체계 정비방안 연구”

장	절	조	내용
제3장		제54조~제57조	법제상 및 재정상의 조치
제4장		제58조~제59조	환경정책위원회
제5장		제60조~제61조	보칙

데이터와 유사하게 최근에 배타적 재산권에 대해 논의가 된 것은 「콘텐츠산업진흥법」이다. 이 법에서는 콘텐츠의 디지털화 및 온라인화에 투하된 비용과 노력에 대한 경제적 가치를 담보하고자 저작권법의 독점 배타적 권리의 부여와는 달리 경쟁법상의 부정경쟁방지 법리를 적용하여 콘텐츠 제작자의 영업상 이익을 보호하고 있다. 즉 콘텐츠 산업을 육성하기 위해서 저작권법상으로 보호받지 못하는 영역에 대해서 해당 법령이 콘텐츠를 보유한 기업의 권리를 보호하고 있는 것이다. 콘텐츠와 데이터 역시 비경합성과 비배제성이라는 공공재적 성격이 유사하며, 기술발전에 따라 생겨난 신기술 영역이기도 하다. 더불어 두 영역 모두 재산권적 성격을 인정할 때 시장에서 활용을 기대할 수 있기도 하다. 저작권법에 따라 '데이터베이스'와 부정경쟁방지 및 영업비밀보호에 관한 법률에 따라 일정한 요건을 갖춘 경우 외에 데이터 보호에 대해서는 무방비 상태이다. 따라서 데이터가 산업으로써 활성화 하기 위해서는 「콘텐츠산업진흥법」 사례를 살펴볼 필요가 있다.

## 🔍 데이터청 설립 논의에서 선결문제

데이터청 나아가 데이터부의 설립 논의가 확대되고 있다. 전 산업에서 흩어져 있는 데이터의 활용 요구에 대한 응답인 것으로 보인다. 새로운 정부부처의 설립으로 데이터의 활발한 활용이 기대될 수도 있으나, 중요한 부분이 빠져 있어 이 부분의 해결이 선결되어야 할 것이다. 우선 데이터 산업 육성을 위해서는 '데이터의 권리'에 대해 접근할 필요가 있다. '데이터 3법의 개정'과 '빅데이터 플랫폼 및 센터 구축 사업'으로 인해 데이터의 활용 활성화의 가능성은 높아졌으나, 여전히 민간기업이 보유한 데이터를 시장에 내놓을 동인은 부족하다. 소유권 등의 배타적 권리를 부여할 것인지 경쟁법 등으로 공정한 데이터 시장 질서를 만들지에 대한 논의의 결론을 내놓아야 할 시기이다. 다음으로 데이터와 관련된 국가전략이 필요하다. 데이터 산업을 육성하고, 이를 통해 신산업을 강화시키려면 법에 근거한 국가전략 수립이 있어야 하기 때문이다. 단순히 새로운 정부부처를 설립하는 것은 법 집행을 위한 기관을 두는 것에 불과하다. 지식재산권과 관련해서 「지식재산기본법」은 정부가 국가지식재산기본계획을 수립하도록 하고 있으며, 「환경정책기본법」은 정부가 국가환경종합계획을 수립할 것을 명시하고 있다. 이와 같은 법령들은 기본(또는 종합)계획을 통해서 변화하는 시대 흐름을 반영하고, 다양한 정부부처의 기능을 조정하고 있다. 데이터와 관련해서도 국가가 법에 근거한 기본계획을 수립하여 나아가간다면, 현재에 존재하는 다양한 문제들이 하나씩 해결되어 갈 것으로 전망된다.



# SPRI AI BRIEF

T E L L I G E N C E

## 인공지능 최신 동향과 시사점 Recent Trends of AI and Implications

### Table of Contents

---

- 이코노미스트紙, “AI 기술의 한계에 대한 냉정한 인식 필요”
- 美, IT공룡들 안면인식 AI 개발 및 공급 중단 발표
- OpenAI, 자연어 처리를 위한 범용 API 공개
- 뉴욕타임즈, 미국의 AI 인력 확보 차질 우려
- 美 AI4K12 이니셔티브, 초중등 AI 교육 가이드라인 업데이트
- 일본, AI 특화 슈퍼컴퓨터 개발 통해 세계 1위 탈환
- AI 연구자 집단, 범죄 예측 AI 연구 논문 철회 요청
- 구글, 새로운 음성추출 AI 개발

## 이코노미스트紙, “AI 기술의 한계에 대한 냉정한 인식 필요”<sup>1</sup>

### ① 영국 경제주간지 이코노미스트는 AI 기술의 비약적 발전에도 불구하고 도전적 과제에서 기대했던 성과를 달성하지 못했다고 언급

- 자율주행차 개발을 주도해 온 구글, 테슬라 및 완성차 업계는 개발 목표 시점을 수차례 수년 이상 연기
  - 구글 자율주행차 개발책임 맡았던 크리스 엄슨(Chris Urmson)은 자율주행차 상용화 시점이 예상보다 30년 이상 미뤄질 것으로 전망<sup>2</sup>
- 코로나19에 대응하기 위해 AI를 활용하는 다양한 아이디어가 논의되었지만 진단, 감염자 관리, 치료제 개발 등에서 실질적인 역할은 미미
  - 감염자 관리는 방역 당국과 의료진이 직접 수행했고, AI를 활용한 혁신 신약 개발보다는 기존 약물의 적용에 초점

### ② 그동안 AI 혁신을 견인해 온 △데이터 축적, △컴퓨팅 기술, △알고리즘의 발전이 이제는 AI의 한계를 야기하고 있음을 지적

- **데이터** AI는 대량의 데이터를 통해 학습을 하여 성능을 고도화하므로 데이터가 없는 대부분의 영역에서는 활용 불가
  - AI가 코로나19 대응에 적극 활용되지 못한 주된 이유는 바이러스에 대한 충분한 데이터가 없고 사람들의 이동 경로를 정밀하게 추적하지 못했기 때문
- **컴퓨터** 데이터를 통한 학습과 추론 과정에서 대규모의 컴퓨팅 자원이 필요하여 AI 도입에 따른 비용이 편익을 상회하는 경우가 대부분
  - \* 벤처투자기업 MMC에 따르면 조사 대상 기업의 40%는 AI를 전혀 사용하지 않음
- **알고리즘** 학습 결과를 다른 분야에 적용하는 능력이 없고 인과관계를 혼동하며 판단의 이유를 설명하지 못하는 등 취약점에 노출

### ③ 현재 AI 기술의 가능성과 한계를 명확히 판단하여 적용 가능 분야에서 적극 활용하는 동시에 기술의 한계를 극복하는 차세대 AI 기술에 투자해야 할 시점

- 현재는 AI 기술에 대한 환상과 거품이 겹치고 성공과 실패가 혼재하는 단계\*로서 기술 수준에 맞춰 성공 가능성이 높은 투자에 집중해야 하는 시점
  - \* 가트너 하이프사이클(2019)에 따르면 딥러닝, 자연어처리, 컴퓨터비전 등 주요 AI 기술은 현재 거품의 정점을 지나 많은 기업이 신중하게 투자하는 ‘환멸의 계곡 단계(Trough of Disillusionment)’에 도달
- 소량의 데이터로 학습할 수 있고 학습 결과를 다양한 영역에 적용할 수 있는 차세대 범용 AI 기술에 대한 선제적 투자도 필요

1 The Economist(2020.1.7.), “An Understanding of AI’s Limitations Is Starting to Sink in”

2 The Economist(2019.10.10.), “Driverless Cars Are Stuck in Jam”

## 美, IT공룡들 안면인식 AI 개발 및 공급 중단 발표<sup>3</sup>

### 1 IBM, Amazon, MS 등 안면인식 AI 기술과 SW의 사용 및 판매 중단 연달아 발표

- **IBM** 범용적인 얼굴 인식 및 분석 SW를 더 이상 만들지 않을 것을 선언(6.9.)
- **Amazon** 자사의 안면인식 SW인 레커그니션(Rekognition)의 판매 중단을 선언하고 경찰 등 감시 기관에 1년간 공급하지 않기로 결정(6.10.)
- **Microsoft** 안면인식 AI 기술을 통제할 국가적 법률이 시행될 때까지 경찰기관에 해당 기술을 공급하지 않기로 결정(6.11.)

### 2 안면인식 AI 기술에 대한 부작용 경계 및 사용 제한 움직임에 대한 대응

- 현재 美 법 집행 기관이 활용 중인 안면인식 AI 시스템의 정확도와 인종 및 성별에 따른 편견 논란\*이 지속적으로 제기
  - \* 美 캘리포니아주의회 의원들 사진을 경찰의 안면인식 AI로 판별한 결과 80명 중 26명이 범죄자로 잘못 판정되었으며 이 중 절반은 유색인종으로 드러나 논란(2019.8.)<sup>4</sup>
- 안면인식 AI에 의한 감시, 인종 분류, 기본권 및 개인정보 침해 등 충분한 법과 절차적 검토를 거치지 않고 오남용될 위험\*에 대한 경각심 확산
  - \* 美 스타트업 Clearview AI는 알고리즘 학습에 활용한 30억 장 이상의 이미지를 사용자 SNS에서 무단으로 수집, 사용자에게 대한 통지나 동의 없이 개인정보를 보유하여 논란(2020.3.)<sup>5</sup>
- 최근 미국에서 발생한 경찰 과잉진압사건을 계기로 안면인식 시스템에 대한 우려 증가\*와 논의 중인 경찰 개혁 법안에 안면인식 AI 기술의 사용을 제한할 가능성 대두<sup>6</sup>
  - \* 흑인 조지 플로이드의 과잉진압 사망사건의 해당 경찰기관인 미네소타주 미니애폴리스시는 Clearview AI사의 안면인식 시스템을 사용 중

### 3 AI 기술 발전에 따른 이익과 반작용에 대한 균형 잡힌 시각, 법·제도적 안전 장치, 윤리적인 AI 활용에 대한 지속적 관심이 요구

- 안면인식 AI 기술로 인하여 범죄 사건을 보다 쉽고 빠르게 해결하는 등 긍정적인 효과도 있지만, 부작용에 대한 법적 안전 장치를 마련하는 등 균형적 활용 방안 필요
- 국가 주도적으로 안면인식을 포함한 시기술의 광범위한 활용을 촉진하는 중국과 시민권 및 윤리 문제로 기술 확산에 제동이 걸린 미국의 향후 AI 기술 패권 경쟁 양상이 주목

<sup>3</sup> Forbes(2020.6.13.), "Facial Recognition Bans : What Do They Mean For AI(Artificial Intelligence)?"  
<sup>4</sup> CNN(2019.8.15.), "Facial-recognition technology flagged 26 California lawmakers as criminals. This bill to ban the tech is headed to the Senate"  
<sup>5</sup> Fortune(2020.3.3.), "What is Clearview AI and why is it raising so many privacy red flags?"  
<sup>6</sup> Wired(2020.6.12.), "A Bill in Congress Would Limit Uses of Facial Recognition"

## OpenAI, 자연어 처리를 위한 범용 API 공개<sup>7</sup>

### ① 그간 자연어 처리 분야는 딥러닝의 부상으로 성능이 대폭 개선되어 왔고, 그 중심에는 AI를 대표하는 기업인 구글과 OpenAI가 지속적인 성과를 산출

- **자연어 처리** 대표적인 자연어 처리의 연구 주제는 특정 문장에서 다음에 올 단어를 예측하는 언어 모델, 질문 문장에 대한 답을 찾아내는 질의응답, 문장이나 문서의 유사성 분류, 언어를 자동으로 번역하는 기계번역 등이 있음
  - **구글** 2017년 어텐션 메커니즘을 바탕으로 한 트랜스포머(Transformer), 2018년 트랜스포머를 기반으로 한 BERT 모델을 공개하여 당시 11개의 자연어처리 과업에서 최고수준의 결과를 달성하고, 2020년에는 계산을 최적화한 리포머(Reformer) 공개
  - **OpenAI** 2018년 생성형 사전학습(Generative Pre-Training, GTP)을 제안하여 다양한 자연어 처리를 위해 학습된 언어 모델을 재학습시키는 접근을 취했고, 2019년에는 개선된 GTP인 GTP-2를 공개
- **그간의 문제점** 범용적인 자연어 처리 모델은 과업에 따라 여전히 미세조정(Fine-Tuning)<sup>\*해야 하는 비용이 필요하여 이것을 점진적으로 개선할 필요성 존재
 
  - \* 이미 학습된 인공신경망을 유사한 과업에 전이(Transfer)하기 위해 가장 보편적으로 사용하는 방법으로 학습하고자 하는 과업에 데이터를 바탕으로 재학습하는 과정</sup>

### ② OpenAI가 공개한 자연어 처리 도구인 OpenAI API는 새로운 방법론인 GPT-3를 적용하여 학습 효율을 큰 폭으로 개선

- **GPT-3<sup>8</sup>** GPT-3은 미세조정이 없는 퓨샷 학습\*을 통해 언어 모델의 학습 효율을 개선
  - \* **Few Shot 학습** 상대적으로 적은 데이터로 인공신경망을 재학습하는 방법으로 하나의 데이터는 원샷, 데이터 없이 바로 새로운 과업을 적용하는 방법은 제로샷 학습이라고 불림
  - **학습 데이터** 인터넷 웹사이트의 데이터를 수집하는 비영리단체인 Common Crawl 데이터를 활용했고, 정제를 거쳐 약 1조 개의 단어로 구성된 데이터로 학습
  - **AI 모델** 트랜스포머 기반의 GPT-2 모델을 기본으로, 학습을 위한 가중치의 수를 조정하여 GPT-3은 기존 언어 모델보다 10배 큰 총 1,750억 개의 모수를 학습
- **API** OpenAI는 GPT-3 기반의 자연어 처리 도구인 모델을 공개하는 대신 베타 버전의 API를 공개하여 쉽게 활용 가능한 고성능 자연어 처리 기능을 제공

### ③ 이번 OpenAI의 API 공개는 AI의 악의적인 활용을 경계하며 인간 친화적인 AI 활용을 추구하는 비영리 연구조직으로서 OpenAI 조직 철학을 보여줌

- OpenAI는 API의 활용을 통해 자연어 처리의 비용 절감, 진입 장벽 저하 등의 긍정적인 활용을 추구하며, 악의적인 활용에 대해서는 즉시 API 활용을 종료하는 기준을 명시

<sup>7</sup> OpenAI(2020.6.11.), "OpenAI API"

<sup>8</sup> OpenAI(2020.6.5.), "Language Models are Few-Shot Learners, arXiv"

## 뉴욕타임즈, 미국의 AI 인력 확보 차질 우려<sup>9</sup>

### 1 뉴욕타임즈紙는 미국 AI 인재 수급 현황 분석 보고서를 인용<sup>10</sup>하며 외국인 근로자 제한 조치로 인해 최고 수준의 AI 인력 확보에 차질을 우려

- Paulson Institute 산하 MacroPolo 연구팀의 보고에서 세계 최고 AI 인력의 60%가 미국 기관 및 기업에서 근무 중으로 나타남
  - 주요 시학회(NeurlPS)의 2018, 2019년 게재 논문 저자\*의 30%가 중국 출신 연구자들이지만 이들 상당수가 미국 내 거주하거나 미국 기관 근무 중
  - \* 저자들은 Google, Microsoft, MIT, CMU 등 미국 내 우수 기업 및 대학에서 근무 중이며 상위 25대 기관 중 중국 기관은 북경대, 칭화대 2개가 유일
- 최근, 미중 갈등이 우수한 중국 기술 인력들의 미국 진출과 체류를 어렵게 하는 정책들로 이어지면서 미국의 AI 고급 인재 확보에 대한 우려 제기

### 2 중국 출신의 AI 인력은 미국 AI 기술을 주도하는 비밀 병기

- MacroPolo는 NeurlPS 2019에 발표된 1,400개의 논문 중 175개를 임의 선정해 저자 671명의 국적, 학위 취득 국가와 현재 취업 현황을 분석
  - 저자들은 학부과정을 중국(29%), 미국(20%), 유럽(18%)에서 마쳤으며, 대학원은 미국(52%), 유럽(16%), 중국(9%)에서, 대학원 졸업 후에는 미국(53%), 유럽(14%), 중국(10%) 순으로 근무 중인 것으로 분석
  - 박사학위를 취득한 중국 국적의 유학생 중 90%는 최소 5년간 미국 내 체류
- 실제 중국 출신 AI 인력들이 현재의 미국 AI 기술 경쟁력에 큰 기여를 한 것으로 평가
  - Li Deng 박사는 Microsoft 연구원 경력을 거쳐 딥러닝 기반의 음성인식 기술을, Fei-Fei Li 스탠포드大 교수는 Google에서 2년간 근무하며 컴퓨터 비전 인식 기술을 획기적으로 개선하는데 기여한 것으로 평가
  - 또한, 구글이 2017년 추진한 시 기반 국방 시스템 프로젝트(Project Maven)의 초기 핵심 개발팀은 구글 AI 엔지니어 12명으로 이 중 5명이 중국 국적이었음에도 진행

### 3 미·중 갈등 확대로 중국 출신 고급 AI 인력의 미국 진출이 어려울 것으로 예상되며 이를 계기로 해외 우수 AI 인력 확보를 위한 각국의 경쟁 가속화 전망

- 트럼프 대통령은 COVID-19 대응의 일환으로 올해 말까지 한시적으로 외국인에 대한 취업 비자를 발급을 전면 중단하는 행정 명령에 서명(6.22.)
  - \* 실리콘밸리의 고숙련 기술직들이 받는 전문직 취업 비자(H-1B, H-2B), 주재원 비자(L-1), 교환 학생 및 연수 비자(J-1) 등 발급 중단 전망
- 중국의 우수 AI 인력 확보를 위한 미국 외 다국적 기업의 유치 경쟁, 미국의 자국 AI 인재 양성 및 동맹국 중심의 해외 인재 확보를 위한 새로운 정책에 관심 필요

<sup>9</sup> Newyork Times(2020.6.9.), "A U.S Secret Weapon in AI : Chinese Talent"

<sup>10</sup> MacroPolo(2020.6.9.), "America's Got AI Talent : US' Big Lead in AI Research Is Built on Importing Researchers"

## 美 AI4K12 이니셔티브, 초중등 AI 교육 가이드라인 업데이트<sup>11</sup>

### ① 미국 초중등 AI 교육을 위해 설립된 ‘AI4K12 이니셔티브’는 2019년 4월 발표한 초중등 AI 교육 가이드라인(5 Big Ideas)을 구체화하여 지난 5월에 공개

- AI4K12 이니셔티브는 △초중등 AI 교육 가이드라인 개발, △AI 교재 개발, △AI 교육자 커뮤니티 육성 등을 목표로 미국 컴퓨터교사협회(CSTA), 인공지능학회(AAAI) 등이 공동 설립
- 이들이 제안한 5 Big Ideas는 인공지능을 이해하기 위해 필요한 5대 학습 분야를 정의하고 학령별 수준에 따른 교육 목표와 내용을 제시

■ [표 1] AI4K12 이니셔티브의 5 Big Ideas 개념 및 교육 목표

주제	1. 인식	2. 표현·추론	3. 학습	4. 상호작용	5. 사회적 영향
개념	AI는 센서를 사용하여 세상을 인식	AI는 세상을 표현하고 구조화해 추론에 사용	AI는 데이터를 통해 학습	AI와 인간의 상호작용에 많은 지식이 필요	AI는 긍정·부정적 효과를 동시에 야기
교육 목표	<ul style="list-style-type: none"> <li>• 인간 감각과 센서의 차이 이해</li> <li>• 컴퓨터 인식의 작동 방식과 한계 이해</li> <li>• 시각, 음성 등 인식 유형 파악</li> </ul>	<ul style="list-style-type: none"> <li>• 표현의 유형 파악</li> <li>• 추론 알고리즘 유형 및 작동원리 이해</li> <li>• 추론 알고리즘의 한계 이해</li> </ul>	<ul style="list-style-type: none"> <li>• 학습 알고리즘 유형</li> <li>• 인공지능경망 기초 개념 이해</li> <li>• 데이터가 학습에 미치는 영향 이해</li> <li>• 기계학습의 한계</li> </ul>	<ul style="list-style-type: none"> <li>• 자연어의 이해</li> <li>• 감성 컴퓨팅 이해</li> <li>• 상식 추론 이해</li> <li>• 인간-로봇의 자연 스런 상호작용</li> </ul>	<ul style="list-style-type: none"> <li>• AI의 산업, 정부 등 영향력 이해</li> <li>• AI의 윤리적 딜레마 및 윤리 표준 마련</li> <li>• AI에 의한 일자리, 업무변화 이해</li> </ul>

### ② 이번 발표는 5대 학습 분야 중 인식(Perception)에 해당하는 △센싱, △처리, △도메인 지식에 대한 학습 콘텐츠를 상세히 기술

- **센싱** 사람과 컴퓨터가 사물을 인지하는지를 방식의 차이에 대해 교육
  - 인간은 세상을 시각, 청각, 촉각, 미각, 후각으로 인지하는 반면 컴퓨터는 카메라로 세상을 ‘보고’, 마이크로폰으로 ‘들음’ 수 있음을 이해
- **처리** 컴퓨터가 어떤 방식으로 사물의 특징을 구분하고 이해하는지를 교육
  - 가령, 얼굴의 특징을 추출해보고, 얼굴 인식이 어떻게 작동하는지 설명하도록 유도
- **도메인 지식** 편향성 등 AI 인식의 한계를 극복하기 위해 다양한 분야의 지식을 교육
  - 사회문화와 인간에 대한 지식이 부족할 경우, 얼굴 인식 시스템 등이 편향성을 가질 수 있음을 이해시킴

### ③ 우리나라의 보편적 AI 교육을 위한 초중등단계의 기초 연구와 교육과정 개발 필요

- 정부는 최근 ‘AI국가전략’(2019.12.), ‘정보교육종합계획’(2020.5.) 등 초중등 AI 보편교육 강화 및 표준 교육과정 개발 등을 추진 중
- 이에, AI4K12 이니셔티브의 5 Big Ideas는 우리나라 초중등 대상의 체계적인 AI 교육을 마련하기 위한 좋은 참고자료가 될 것으로 기대

11 <https://drive.google.com/file/d/1qznNZPfnz9ohh0WZYQcyE9lw9e27ZpS4/view?usp=sharing>

## 일본, AI 특화 슈퍼컴퓨터 개발 통해 세계 1위 탈환<sup>12</sup>

- 1** TOP500이 발표한 전 세계 슈퍼컴퓨터 순위에서 일본의 후가쿠(Fugaku)가 1위 차지(2020.6.)

  - 과학기술 시뮬레이션에서 활용되는 선형대수 알고리즘(HPL, High Performance Linpack)의 성능을 배정밀도\*로 실측하여 순위를 산출
    - \* 배정밀도(Double Precision)는 64-bit의 이진수로 실수를 표현하는 방법
  - 후가쿠는 이화학연구소(RIKEN)와 슈퍼컴퓨터 전문기업 후지쯔가 공동 개발했으며 전통적인 슈퍼컴퓨터 영역뿐 아니라 AI 성능에서도 독보적인 1위를 달성
  
- 2** 이번 순위는 슈퍼컴퓨터의 전 세계 경쟁구도가 재편되고 있으며, AI 분야에서 슈퍼컴퓨터의 중요성이 커지고 있음을 시사

  - **美·中·日 3강구도** 일본의 1위 탈환으로 슈퍼컴퓨터의 전통적인 강자였던 일본\*이 다시 부상하여, 미국과 중국 중심의 양강구도에 도전\*\*
    - \* 일본의 1위 슈퍼컴퓨터 : Numerical Wind Tunnel(1993~1994), Earth Simulator(2002~2004), K Computer(2011~2012), Fugaku(2020~)
    - \*\* 미국과 중국의 10위권 슈퍼컴퓨터 보유대수 현황 : (2018.6.) 美(6대), 中(2대) → (2019.6.) 美(5대), 中(2대) → (2020.6.) 美(4대), 中(2대)
  - **AI 특화 슈퍼컴퓨터 부상** 10위권에 등재된 슈퍼컴퓨터 중 절반 이상이 AI를 지원하는 GPU 가속기를 탑재했으며, TOP500도 AI 성능 측정 방법론을 활용하여 순위를 산출
    - **GPU 탑재** 슈퍼컴퓨터에서의 AI 활용이 강조됨에 따라, AI를 효율적으로 지원하는 GPU 기반의 슈퍼컴퓨터 점유율이 10위권 슈퍼컴퓨터의 절반 이상을 차지\*
      - \* 10위권 GPU 기반 슈퍼컴퓨터 : (2018.6.) 5대 → (2019.6.) 5대 → (2020.6.) 6대
    - **AI 성능지표 개발** TOP500은 美 테네시대의 잭 동가라(Jack Dongarra) 교수 연구진이 개발한 슈퍼컴퓨터의 AI성능지표(HPL-AI<sup>13</sup>)를 활용해 순위\*를 발표
      - \* 후가쿠는 1.42엑사플롭스(초당 100경 번 연산)를 달성하여 1위 등극
  
- 3** 향후 AI는 슈퍼컴퓨터의 주요 수요처로 부상할 것으로 예상되며, AI 인프라 생태계 조성을 위한 AI 특화 슈퍼컴퓨터의 투자가 활성화될 전망

  - 새로 TOP500 순위에 등재된 슈퍼컴퓨터는 AI를 활용할 수 있는 GPU 기반의 시스템이 주류를 이루고 있으며, AI 성능을 측정하기 위한 노력도 이어지고 있음
  - 우리나라 역시 『광주 AI클러스터』에 도입될 슈퍼컴퓨터급 AI인프라 등 관련 정책을 통해 국가 AI 경쟁력 확보와 AI 연구개발 생태계를 고도화할 필요

<sup>12</sup> TOP500(2020.6.22.), "Japan Captures TOP500 Crown with Arm-Powered Supercomputer"

<sup>13</sup> High Performance Linpack - Artificial Intelligence(HPL-AI), HPL-AI Mixed-Precision Benchmark

## AI 연구자 집단, 범죄 예측 AI 연구 논문 철회 요청<sup>14</sup>

### 1 2,400명이 넘는 산·학·연 AI 연구자들이 신경망을 적용한 안면 인식 기반의 범죄 예측 논문의 게재 철회를 공개적으로 요청(2020.6.)<sup>15</sup>

- 美, Harrisburg 과학기술대학의 연구진은 독일 Springer Publishing社를 통해 얼굴 인식 기반의 범죄 예측을 다룬 논문의 게재 예정 발표(2020.5.)
  - \* 논문 제목은 “A Deep Neural Network Model to Predict Criminality Using Image Processing”으로 Springer Nature - Research Book Series: Transactions on Computational Science and Computational Intelligence 게재 예정이라고 소개
- 논문의 게재 예정 소식이 전해지자 지난 6월 22일 2,435명의 산·학·연 AI 연구자들은 해당 논문의 게재 철회를 요청하는 공식 서한을 온라인에 공개

### 2 AI 연구자들은 공개서한을 통해 범죄 예측을 위한 AI 기술 활용의 윤리적 문제와 방법론적 오류에 대한 심각한 우려(Grave Concern)를 표명

- **윤리** 사람의 신체 정보를 바탕으로 한 안면 인식 기술은 정확하지 않은 성능으로 그간 인종 차별 이슈 등 사회적 논란\*을 야기한 바, 매우 신중한 접근이 필요
  - \* 지난 2016년 미국의 주 법원에서 사용하던 재범위험예측 소프트웨어인 COMPAS는 흑인을 고위험군으로 분류했지만 실제 재범율은 백인에 비해 절반 이하에 불과
- **방법론** 신경망 학습에 활용한 각종 범죄자 정보 자체가 사회적 편견의 산물일 뿐만 아니라 사람의 얼굴 특성과 범죄 예측 사이의 인과 관계를 설정하는 것 자체가 오류
  - 유색 인종에 대한 차별이 반영된 기존 형사 시스템에서 생성된 데이터 자체가 왜곡된 것이므로 ‘더 많은 데이터’와 ‘효과적 전처리’ 기술도 근본적인 해결책이 될 수 없음
  - 또한, 기계학습 모델이 사물 인식에서 월등한 성능을 낸다고 해서 그것을 범죄자 얼굴 인식과 예측에 적용하는 것은 잘못된 접근\*
  - \* 뚜렷이 식별 가능한 사물과 달리 ‘범죄자의 얼굴’은 비교·평가할 수 있는 실제 검증 자료(Ground Truth)가 존재하지 않음

### 3 인공지능의 공정성에 대한 사회적 논의 확대 전망

- 이번 AI연구자들의 학술 논문 철회 요구\*는 향후 데이터 편향성, 알고리즘 중립성 및 관련 AI 기술에 대한 규제 필요성을 촉구하는 사회적 논쟁으로 확산될 전망
  - \* 현재 해당 논문은 최종 심사 결과 게재 불가로 저자들에게 통보된 상태(6.22.)
- 기계 학습 연구에 활용하는 데이터가 정치, 사회, 문화, 법·제도 등과 연관성이 높을 경우 데이터 편향성에 대해 보다 면밀한 사전 검토 필요
  - 시기술전문가, 데이터 과학자 그리고 사회학자, 인류학자, 철학자 등 인문사회 연구자들이 협업을 통해 학습 데이터의 사회·문화적 맥락에 따른 오류가능성 확인 필요

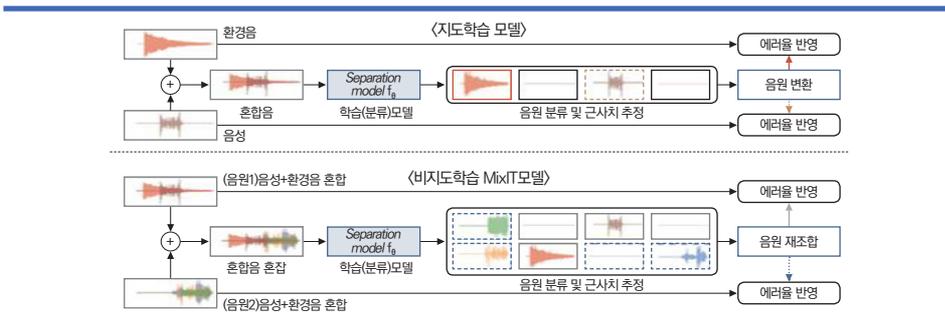
<sup>14</sup> TechCrunch(2020.6.24.), “AI researchers condemn predictive crime software, citing racial bias and flawed methods”

<sup>15</sup> <https://medium.com/@CoalitionForCriticalTechnology/abolish-the-techtoprisonpipeline-9b5b14366b16>

## 구글, 새로운 음성추출 AI 개발<sup>16</sup>

- 1 구글은 혼합된 음향 속에서 원하는 음성을 분리·추출하는 새로운 방식의 MixIT (Mixture Invariant Training)라는 음성추출 AI를 발표
  - 다자간 대화 녹음과 같은 혼합된 음원에서 특정 음성을 분리하고 음질을 향상시키는 방법론을 구현한 논문<sup>17</sup>을 아카이브(arXiv)에 게재(6.23.)
- 2 MixIT는 기존의 음성 추출 AI와 다르게 비지도학습(Unsupervised Learning)방식이며 추출 성능 및 음질은 유사하되, 더욱 다양한 상황에 적용 가능한 범용성을 제공
  - 기존 알고리즘들은 ‘잡음 섞인 소리’와 ‘잡음 없는 원음’을 지도학습(Supervised Learning)하는 방식을 취하며, 음성 추출 시 반향이 있거나 소리 유형이 불일치할 때 효과가 떨어짐
    - 기존의 지도학습 방식의 음성추출 AI는 제한적인 모델링과 가정을 통해 차별적인 훈련이 어렵고, 원본에 대한 추론과 근사치 복원에 있어서 범용성이 떨어졌음
  - MixIT 알고리즘은 방대한 양의 여러 소리가 혼합된 실제 데이터(Wild Data)를 학습\*하였으며, 음성과 주변 소리 분리, 분리된 음질 향상을 통해 지도학습 방식과 동등한 성능 제공
    - \* Google Cloud Tensor 처리장치인 4개의 TPU(16chips)를 활용하였으며, 390시간 이상 추출된 남·녀 음성 녹음 및 주변 소리 등 오픈소스 데이터(WSJ0-2mix, Libri2Mix, FreeSound)를 학습
    - 입력된 오디오 데이터에서 여러 개의 음원을 분리·추출, 근사추정하고 분리된 음원을 재조합하여 실제 음성 데이터에 가깝게 복원

[그림 1] 기존 방식(지도학습)과 MixIT(비지도학습)의 음성추출 방식 비교



- 3 구글은 소리인식 및 추출 AI를 경쟁력 있는 신사업 아이템으로 선정하여 연구 개발에 박차를 가하고 있어 다양한 분야로 응용이 가속화 될 것으로 전망
  - 최근 1만 종류 이상의 새소리를 구분하는 인식 시스템에도 적극 투자<sup>18</sup>하고 있으며, 구글 어시스턴스(Google Assistant)\* 등과 연계 프로젝트도 수행
    - \* 2016년 5월 출시한 구글의 AI 기반 가상 비서 서비스

<sup>16</sup> VentureBeat(2020.6.25.), "Google's MixIT AI Isolates Speakers in Audio Recordings"

<sup>17</sup> Wisdom, Scott, et al.(2020), "Unsupervised Sound Separation Using Mixtures of Mixtures." arXiv preprint arXiv:2006.12701

<sup>18</sup> Voicebot.ai(2020.6.26.), "Voice Match is for the Birds : New Google Competition Seeks Avian Audio AI"



사진산책

# 중간 中間

디지털과 아날로그의 중간.  
일상과 섬의 중간 같은 사진.  
특정 주제나 내용에 치우치지 않는 누구나 공감할 수 있는 사진

---

이호준 Lee, Ho-Joon 언론학박사 [ighwns@hanmail.net](mailto:ighwns@hanmail.net)

우정사업본부에서 근무하고 있으며, 대한항공 여행사진 공모전에서 2회 수상하고,  
세 차례의 개인전과 단체전 3회를 개최했다.



콘크리트 덩어리 속에서 자연 생태계 복원을 논하는 것은 공염불일 가능성이 높다. 멸종한 나무와 풀밭을 갈아엎고 자동차 길 따라 가로수 심어 놓은 걸로 도시의 자연성 회복을 외치는 것과 같다. 그래도 서울은 개발론자들도 어찌할 수 없는 큰 산과 강이 있어 그나마 자연을 말할 여지가 있다. 서울 안에서 생태계 복원을 목격할 수 있는 장소를 꼽으라면 단연 밤섬이 떠오른다. 조선시대부터 강 한복판으로 들어온 사람들이 모여 살던 마을, 그러나 1960년대 한강 개발의 희생양으로 강제로 무인도가 된 사연 많은 섬. 폭파되고 제 살 같은 흙과 돌을 빼앗겨 황무지로 전락한 땅. 그러나 반세기 넘어 하염없이 흐른 강은 토사를 쌓아올려 섬의 모습을 되찾고, 울창한 숲을 품은 아름다운 장소로 거듭나게 해주었다. 이제 사람은 살지 않지만, 대신 새와 물고기들의 안락한 터전으로 서울 생태계의 보물로 자리매김했다. 1,000만 인구가 사는 한가운데에 이런 곳이 있다는 것은 경이로운 일이며, 시민들에게 축복이다. 섬을 가로지르는 서강대교 위에서 바로 보는 사계절 풍경과 일출·일몰 장면은 환상 그 자체다. 아침저녁으로 날아오르는 새들의 집단 이동과 군무 장면을 보노라면 이곳이 과연 도시 속인가 하는 의구심이 들 정도다. 이렇게 밤섬은 상처 치유와 복원이라는 자연의 위대한 능력을 사람들 보란 듯 증언해주고 있다.



# SPRI FOCUS

## 21대 국회에 바라는 SW산업 이슈

- 21대 국회 최대 숙제는 ‘AI·언택트 시대 맞는 SW·디지털 법제도·생태계 만들기’

안경애 디지털타임스 ICT과학부 부장

- 21대 국회에 바라는 SW산업 이슈

조영훈 한국소프트웨어산업협회 산업정책실 실장

\* 이 원고는 필진의 개인 견해이며,  
소프트웨어정책연구소의 공식의견과 다를 수 있습니다.



## 21대 국회 최대 숙제는 ‘AI·언택트 시대 맞는 SW·디지털 법제도·생태계 만들기’

“최소한 수년이 걸릴 디지털 전환이 지난 수개월 만에 이뤄졌다. 위기상황에서 산업과 사회 시스템 유지와 회복, 팬데믹 극복에 디지털 기술이 얼마나 중요한지가 확인됐다. 앞으로 모든 기업과 국가의 경쟁력은 디지털과 데이터, 언택트 수준이 좌우할 것이다.”

코로나19 팬데믹 상황을 경험한 ICT 및 소프트웨어(SW) 전문가들의 공통된 얘기다.

디지털 혁신을 기업성장의 화두로 올려놓고 추진하던 기업들이 코로나19 사태를 계기로 제조현장부터 업무방식, 유통·외식 등 고객접점에까지 자동화와 스마트화를 공격적으로 추진하고 있다. 공장은 물론 매장까지 산업 전 사이클과 현장에 자동화와 AI를 입혀 업무와 소통, 리스크 대응을 돕기 위해 기술투자에 나섰다.

국내 기업들의 혁신 속도도 빨라졌다. 금융·자동차·반도체·통신·제철 등 업종을 가리지 않고 국내 주요 기업들의 공통적인 화두는 데이터, AI, 클라우드다. 여기에 코로나19로 인해 언택트라는 키워드가 추가됐다. 기업들은 내·외부에 흩어져 있던 고객 정보와 재고 데이터, 공급망 정보와 외부 소셜네트워크서비스(SNS) 등 가능한 모든 데이터를 융합해 급변하는 시장과

● ●  
안경애  
디지털타임스 ICT과학부 부장  
naturean@dt.co.kr

고객행동을 실시간으로 읽어내 위기에 대응하고 기회를 잡으려 하고 있다. 각국 정부도 코로나19로 인한 경제충격을 최소화하고 성장기회를 잡기 위해 디지털과 언택트 투자를 확대하고 있다.

코로나19로 인한 충격파를 도약의 디딤돌로 삼으려면 막 활동을 시작한 21대 국회의 역할이 무엇보다 크다. 국가와 산업, 사회 전반의 SW와 AI, 언택트 경쟁력을 높이기 위해서는 해묵은 법·제도를 뜯어고치고, 신산업이 성장할 수 있는 생태계를 만드는 일이 시급하다. 산업계와 전문가들은 미래 지향적인 투자와 사업 확장과 함께 그동안 국내 ICT 산업과 SW 생태계의 글로벌화를 힘들게 한 고질적 문제들을 서둘러 개선해야 한다고 지적한다.

업종을 불문하고 모든 기업에서 데이터와 AI 활용이 생존의 핵심 키워드가 된 만큼, 원격의료·마이데이터 등 신산업이 클 수 있도록 대통령이 힘을 실어주는 강력한 거버넌스 구축과 이해관계 조정 노력도 반드시 필요한 시점이다.

## SW·AI생태계 혁신적 재설계 필요

코로나19 팬데믹 속에 'K방역'이라는 브랜드가 생겨날 정도로 우리나라는 위기를 기회로 만드는 저력을 세계적으로 확인시켰다. 국내 기업들이 글로벌 사업을 추진하는 과정에서 걸림돌이 돼 온 '코리아 디스카운트'가 '코리아 프리미엄'으로 바뀌는 반전의 기회를 맞은 만큼, 이를 토대로 ICT와 SW, 혁신산업이 글로벌 경쟁력을 키울 수 있도록 21대 국회가 법·규제 혁신에 4년간 올인해야 한다고 전문가들은 목소리를 높인다.

세계 각국이 생활·산업·사회 전체를 디지털화하는 작업에 나설 전망이다. 코로나19로 높아진 국가 위상을 기회로 연결하려면 디지털혁신을 통해 국내 생태계부터 혁신 친화적으로 바꿔야 한다는 지적이다. 특히 산업계에서는 국내 기업들이 글로벌 시장에서 제대로 승부하려면 국내 성공사례 확보가 필수인 만큼, 곳곳에 숨은 '갈라파고스적 규

**[표 1] 혁신산업 관련 주요 법제도 이슈**

분야	현재상황	이슈
데이터 산업 육성	데이터 3법 개정안 1월 9일 국회 통과. 8월 5일 시행 앞두고 시행령·고시 등 후속작업 중	<ul style="list-style-type: none"> <li>•개보법 시행령의 가명정보 활용·결함 관련 조항 과도하게 엄격하고 모호한 규정 많아 불확실성 우려</li> <li>•개보법과 신용정보법 시행령의 규정 달라 혼선 예상</li> </ul>
소프트웨어 산업 육성	소프트웨어진흥법 개정안 5월 20일 국회 통과. 11월 시행 앞두고 후속작업 중	<ul style="list-style-type: none"> <li>•정부의 SW 제값주기, 기업의 공공SW사업 결과물 활용 허용, 원격지 개발 허용 등 실질적 제도 정착 필요</li> <li>•정부 예산·조달·계약 등 연관 제도 개선 필요</li> </ul>
국가 R&D 경쟁력·생산성 강화	국가연구개발혁신법 제정안 5월 20일 국회 통과. 내년 1월 시행 앞두고 후속작업 중	<ul style="list-style-type: none"> <li>•검증 없이 부처별로 해온 R&amp;D 기획·관리·평가 기준을 통일하고 통합하는 실질적 시스템·장치 마련 필요</li> </ul>
국가 AI 경쟁력 강화	국가정보화기본법 개정안 5월 20일 국회 통과. 11월 시행 앞두고 후속작업 중	<ul style="list-style-type: none"> <li>•데이터 3법 통한 데이터산업 성장으로 AI 기반 마련 필요</li> <li>•실증사업 나열 아닌 국가 시스템, 일하는 방식, 인력 등 근본적 혁신 필요</li> </ul>
OTT 법제화	넷플릭스 등 글로벌 OTT, 부가통신사업자로 규정돼 콘텐츠에 대한 법적 의무, 규제 없어	<ul style="list-style-type: none"> <li>•편성, 광고, 심의 등 방송사업자에 준하는 규제 필요</li> <li>•공정경쟁, 이용자보호 등 최소한 사전·사후규제 도입 필요</li> </ul>
혁신 모빌리티	여객자동차운수사업법 개정안 3월 국회 통과. 내년 4월 시행 앞두고 후속작업 중	<ul style="list-style-type: none"> <li>•신규 플랫폼 운송사업자, 정해진 택시 대수 한도 내에서 차량 운행 허용, 혁신산업을 기존 산업 틀에 맞춰, 시행령 등 하위법령서 제도 경직성 보완 필요</li> </ul>

제<sup>1</sup>와 시대에 맞지 않는 법·제도를 정비해야 한다고 촉구하고 있다.

당장 발등에 떨어진 불이 데이터 3법, SW진흥법, 국가연구개발혁신법, 국가정보화기본법 등 혁신 관련 법을 제대로 완성하는 것이다.

이들 4개 법은 데이터·기술패권 시대에 승부하기 위한 국가 전략과 철학을 담은 만큼 '국가혁신 4법'이라 할 만하다. 20대 국회 임기 종료를 앞두고 앞서 지난 1월에는 데이터 3법, 4월 20일에는 나머지 3법이 통과되면서 변화를 위한 밑그림은 완성됐다. 이제 소관부처가 주도해 만드는 시행령, 시행규칙에 기존 생태계를 확실히 바꾸는 혁신적 내용이 담기도록 정부 부처와의 공조와 견제가 필수적이다.

구태연 법무법인 린 변호사는 "SW와 ICT 산업은 데이터 주도권과 직결되고, 한발 더 나아가 경제와 정치 주권으로 연결된다"면서 "데이터 산업과 국가경제 차원에서 SW와 ICT 정책 이슈를 들여다 보고, 글로벌 경쟁력을 가지도록 국내 생태계를 만들어야 한다"고 지적했다.

## 데이터 3법 '첫 단추' 잘 꿰어야

특히 데이터 3법<sup>2</sup>에서 첫 단추를 잘 꿰지 못하면 기대를 모아온 데이터 경제 활성화가 헛구호가 될 수 있다는 우려가 큰 상황이다.

데이터 3법 중 핵심인 개인정보보호법 시행령이 지나치게 엄격하거나 모호해 기업의 적극적인 투자와 산업성장이



힘들다는 지적이 잇따르고 있다.

데이터 3법은 특정 개인을 식별할 수 없도록 조치한 '가명정보' 개념을 도입해 데이터 활용도를 높이는 게 골자다. 그런데 개인정보보호법 시행령에서 개인정보 추가 이용·제공을 위해 당초 목적과의 상당한 관련성, 추가 이용 예측 가능성, 제3자 이익침해 방지, 가명처리 의무 등 4가지 조건을 모두 충족하도록 규정, 원래 법 조항보다 더 엄격하다는 평가를 받고 있다.

업계와 전문가들의 지적이 이어지자 정부가 하위 법령의 일부 조항을 삭제하거나 수정하겠다는 계획을 밝혔지만 여전히 데이터의 활용보다 보호에 초점을 두고 있다는 우려가 이어진다. 또 시행령이 개인정보 보호를 강조하고 인권침해 우려를 내세운 시민단체 의견을 반영하는 과정에서, 법에서 규정하지 않은 조항까지 담겨 앞뒤가 맞지 않는다는 지적이 나온다. 각호의 사항을 모두 충족하도록 한 조항은 지나치게 까다로운 만큼 각호의 사항을 '종합적으로 고려하는' 수준으로 보완해야 한다는 의견이 나온다. 4가지 요건을 모두 충족하려면 실제 업무에 적용이 어렵고, 법적 분쟁이 이어질 가능성이 크다는 우려도 크다.

대한상공회의소도 지난 6월 관련 간담회를 열고 '상당한 관련성'의 의미가 불명확하고, 수집한 정황과 처리 관행에 비춰 추가 이용을 예측하도록 한 것은 지나치게 엄격하

1 갈라파고스적 규제는 육지에서 고립돼 독특한 생태계를 이룬 갈라파고스 제도처럼 국제적 흐름이나 변화에 대응하지 못하고 현장과 동떨어진 불합리한 규제를 말한다.

2 데이터 3법은 개인정보보호법·정보통신망법·신용정보법을 말한다. 개인정보보호법 개정안은 가명정보 개념 추가를 통해 통계 작성, 연구 등에 개인정보를 활용할 수 있도록 했다. 정보통신망법 개정안은 개인정보 관련 규정을 개인정보보호법에 이관하는 내용을 담았다. 신용정보법 개정안은 통계 작성, 연구, 공익적 기록 보존 등을 위해 가명정보를 동의 없이 이용·제공 가능케 하는 규정을 포함시켰다. 기존에는 개인정보 활용을 위해 '개인정보 수집 및 이용 동의서'가 반드시 필요했지만, 특정 개인을 알아볼 수 없도록 비식별화한 '가명정보'는 개인동의 없이도 활용할 길이 열렸다.

다는 의견을 제시했다. 행정안전부는 이를 고려해 상당한 관련성에서 '상당한'이라는 단어를 삭제하고 '수집한 정황과 처리 관행'을 '수집한 정황 또는 처리 관행'으로 수정하는 방안에 대해 검토하겠다고 밝혔다. 논란이 된 '제3자의 이익 침해 관련' 문구는 추가로 의견을 수렴하기로 했다. 다만 가명 처리를 하지 않아도 되는 부분까지 가명을 의무화하는 내용은 검토 대상에 포함되지 않았다.

가명정보 결합과 관련해서도 개인정보보호법 시행령은 연계정보 생성기관과 결합전문기관 두 곳을 거치도록 규정해 신용정보법보다 까다롭다.

두 기업이 각각의 가명정보를 결합하려면 개인정보보호위원회 위원장이나 관계 중앙행정기관의 장이 지정하는 전문기관에 결합신청서를 제출한 뒤 승인을 받아 전문기관 내에 마련된 안전한 분석 공간에서 결합된 정보를 분석하는 식이다. 이에 앞서 한국인터넷진흥원 등 연계정보 생성(결합기 관리) 기관을 통해 식별자를 암호화하는 과정을 거친다. 결합된 정보를 외부로 반출하려면 전문기관의 안전성 평가와 승인도 받아야 한다.

법조계 전문가들은 이 절차가 신용정보법에 비해 과도한 규제라고 지적한다. 신용정보법 시행령 개정안은 금융회사가 데이터를 결합하고자 할 경우 금융위원회가 지정한 전문기관에 결합을 신청하고 가명·익명처리, 적정성 평가 등 충분한 안전조치를 거쳐 결합의뢰기관에 데이터를 제공할 수 있다. 개인정보보호법과 비교해 가명정보를 분석하는 장소 제한이 없고 외부 반출에 대한 조건도 간단하다.

법이 개정돼도 시행령·고시 등에 취지와 다른 조항이 담기면 도리어 독소조항으로 작용하는 만큼, 개인정보보호

법 규정을 신용정보법 수준으로 통일해야 한다는 지적이다.

시행령 내 불확실성을 해소하기 위해 가명정보 등에 대한 가이드라인이 필요하다는 목소리도 크지만, 가이드라인의 법적 근거를 확보하지 못하면 자칫 또 다른 리스크가 될 우려가 있다.

이성엽 한국데이터법정책학회(고려대 교수)는 데이터 3법과 관련해 "기업의 개인정보 비식별화 조치 절차에 대해 가이드라인을 만들어 관리하겠다는 게 정부 계획인데, 가이드라인은 법적 근거가 없고 면책이 안 된다"면서 "과거 시민단체의 고발로 기업의 비식별화 시도가 중단된 사례를 재연하지 않으려면 가이드라인에 법적 추정력을 부여하고, 법 시행 1년 정도는 가이드라인을 따랐을 경우, 법적 책임을 안 지우는 유예기간을 두는 것도 검토할 필요가 있다"고 말했다.

또한 이 교수는 신용정보법에 비해 복잡하고 까다로운 개인정보법의 가명정보 결합절차를 개선하고, 개인정보보호위원회의 법 유권해석 기능을 대폭 확대해 기업의 우려를 해소해 줄 필요가 있다고 덧붙였다.

정부가 하위법령 제정 과정에서 법의 원래 취지에 벗어 나지 않도록 국회가 견제 역할을 제대로 하는 동시에, 이번 개정 결과가 미진할 경우 재개정을 서두를 필요가 있다.

## SW진흥법에 혁신성 제대로 담아야

20년 만에 전면 개정된 SW진흥법도 12월 시행을 앞두고, 시행령, 시행규칙, 고시 등이 완성돼야 한다. 공공SW 시장을 비롯한 SW산업 전체의 구조를 바꾸는 상세설계가 속도 있게 이뤄져야 한다.

■ [표 2] 한국의 디지털 경쟁력 순위

지식(11위)			기술(17위)			미래 준비도(4위)		
인재	교육훈련	과학기술	규제개선	자본여건	기술여건	신기술적용도	미래대비	IT통합
30위	5위	6위	26위	29위	7위	4위	5위	21위

※ 자료 : IMD(2019.9.)

산업계와 전문가들은 특히 SW의 가치를 제대로 인정하고 제값을 주는 문화를 공공영역부터 제대로 정착시키는 게 급하다고 지적한다. 구시대적 갑을 관계가 아니라 공공과 산업이 시너지를 얻을 수 있도록 원격지 개발, 지식재산권 활용, 민간 서비스 활용, 상용SW 육성 철학과 방법을 시행령 등에 구체적으로 담는 게 필수다.

특히 불명확한 제안요청서(RFP)와 비밀비재한 과업변경 요구, 과업변경에 따른 추가 대가 지급 관련 체계 부재로 인한 SW 기업들의 피해와 발주기관·수주기업 간 갈등을 해소하기 위한 방안 마련이 시급한 것으로 분석된다.

공공 SW 사업을 발주하는 과정에서 RFP가 불명확한 데서 문제가 시작한다. 심한 경우 원래 기업이 예상한 사업 범위와 결과적으로 수행한 내용이 2배 가까이 차이가 나기도 한다. 이 때문에 대부분의 SW와 시스템통합(SI) 기업들이 발주기관과 법적 분쟁을 통해 갈등을 벌이고 있다. 정부가 문제를 개선하기 위해 2013년 'SW사업 요구사항 분석·적용 가이드'를 내놓았지만 제대로 된 해결책이 되지 못하고 있다.

SW업계와 전문가들은 SW 개발사업의 특성상 사업 수행 과정에서 발주자의 요구사항이 더 구체화하면서 과업 범위가 늘어나는 경우가 많은 만큼 SW진흥법 개정안에 관련 내용을 구체적으로 담아야 한다고 지적한다.

SW사업 과업변경심의위원회가 활성화될 수 있도록 실효성 있는 조치를 담은 한편 폐쇄화된 요구사항의 상세

화 기준을 강화하고 철저한 모니터링 체계를 갖출 필요가 있다.

구태연 변호사는 “코로나를 계기로 엄청난 SW 시장이 열릴 것”이라면서 “국내 기업들이 기회를 제대로 활용할 수 있도록 SW에 대한 인식부터 정부 예산, 조달, 계약 등에서 전면적 변화가 필요하다”고 지적했다. 이어 “공공SW 개발결과물 재활용 허용, 과도한 선급·하자보증제도 개선, SW기술자 경력 중심의 등급제 완전 금지도 필요하다”고 말했다.

이와 함께 조달청의 공공 IT·SW 사업자 선정 시 전문가 평가위원 풀을 확대하고 명단을 비공개해 기업들의 과당 영업경쟁을 막고, 기술평가 점수의 변별력을 높이기 위해 입찰가격 하한선을 현재의 80%에서 90% 이상으로 높여야 한다는 의견이다.

구 변호사는 또한 “공공SW 사업에서 개발SW의 지식재산권은 발주기관과 수행기업의 공동 소유지만 국정원의 보안규정 때문에 기업이 이미 개발해 보유한 SW의 소스코드나 개발 결과물을 반입·반출해 활용할 수 없다 보니 개발 인력을 투입해 처음부터 다시 개발하는 비효율이 반복되고 있다”고 지적했다. 그러면서 “SW사업자가 보유한 지식재산권의 활용과 재활용에 따른 SW 소스코드 반입·반출에 대한 새로운 가이드라인을 만들고, SW임치제도나 저작권 등록, 공유서버 활용 등을 활용할 필요가 있다”고 제안했다.

IT 벤처기업가 출신 이영 미래통합당 의원(비례대표)은 “포스트 코로나 시대에는 ICT와 혁신기술이 뒷받침되지 않

■ [그림 1] 분야별 AI 활용효과(~2022년)



※ 자료 : OECD(2019), 맥킨지(2018) 등



으면 제조업에 의존해온 국가 경쟁력을 유지하기 힘들 수밖에 없다"면서 "20년간 진일보하지 못한 SW 제값 받기부터 시작해 국가 정책에서 SW·ICT 비중을 높이고, 글로벌 경쟁력을 갖춘 혁신기업들이 성장을 이끌도록 제대로 된 국가 어젠다를 만들고 속도감 있게 결과를 만들어야 할 것"이라고 강조했다.

## 공공IT, 용역 일변도 벗어나 민간투자형 사업 확대 필요

SW진흥법 개정안에 민간투자형 공공SW사업 추진 근거 담긴 가운데, 정부가 기업에 정해진 일감을 주는 용역 방식 대신, 민간이 주도해 개발한 솔루션과 디지털 서비스를 도입하는 방식도 늘릴 필요가 있다. 민간투자형 SW사업은 민간의 자본과 기술을 활용해 민간과 공공이 협력해 추진하는 공공SW사업을 의미한다. 정부의 경직된 예산과 사업방식으로는 4차 산업혁명 시대에 맞는 다양성과 창의성, 혁신성을 적기에 발휘해 공공서비스를 기획·운영하는 데 한계가

있을 수밖에 없다. 플랫폼 서비스가 보편화되면서 민간은 이미 서비스를 직접 개발하고 소유하는 대신 서비스 이용으로 전환 중인 만큼 공공도 변화가 필요한 시점이다.

민간투자형 공공SW사업이 확대되면 정부는 예산부담이 줄어들고 기업은 장기적인 신사업 기회를 확보할 수 있는데다 국민은 급변하는 최신 기술이 적용된 공공서비스를 이용할 수 있는 장점이 있다. 사업의 연속성과 안정성이 떨어지는 일회성 SI 프로젝트와 달리 기업이 직접 사업주도권을 갖고 장기적인 투자를 할 수 있어 글로벌 시장 진출에도 유리하다. 제한된 정부예산 내에서 사업을 하지 않아도 되는 만큼 전체 시장 파이가 커지는 효과도 기대된다.

성공적인 민관협력(PPP)형 공공IT사업은 서울시 티머니가 대표적인데, 이명박 시장 시절 시장 역점사업으로 밀어붙인 결과다. LG CNS가 기술과 자본을 대고, 서울시는 버스·지하철·택시를 포괄하는 대중교통 요금징수 사업기회를 제공했다. 지자체가 몇년 단위로 사업자를 선정해 시스템을 업그레이드하는 것과 달리 민간이 신기술을 빠르게 적용하니 이용자들의 만족도가 높고 세계 각국으로 수출하는

성과로 이어졌다. 스마트시티 국가 시범도시 사업에도 PPP 방식이 도입됐다.

이런 방식 사업이 본격화되려면 정부의 예산, 조달, 계약 등 제도 전반의 변화가 필수인 만큼 국회의 역할이 필요하다. 정부는 수년 전부터 공공SW사업에 민간투자를 확대하는 방안을 검토해 왔지만 연관제도 변화나 수요와 연결되지 않고, 실증 수준에 그쳐 한계가 있었다.

과기정통부가 시범사업 형태로 일부 추진했지만 실제 수요부처가 앞서서 나서지 않다 보니 장기적 사업으로 이어지지 않았다. 공정성을 이유로, 처음 아이디어를 제시한 기업에만 사업기회가 제공되는 게 아니라 경쟁입찰 방식으로 사업자를 선정해 기업의 투자 의지를 떨어뜨리기도 했다. 여기에 SW진흥법 개정으로 법제도가 갖춰진 데다 코로나 19로 디지털·언택트 전환 수요가 커진 만큼 환경이 조성된 상황이다.

조미리에 바이티더블유 대표는 “민간투자 방식을 적용하면 용역사업과 달리 기업이 나만의 사업아이템을 가지고 해외 시장으로 나갈 수 있다. 국민은 더 나은 서비스를 이용하고 국가는 재정부담을 줄이는 일거삼득 효과가 있다”면서 “중앙정부 차원에서 관련 사업 아이템을 마련해 적극적으로 확대할 필요가 있다”고 말했다.

행정안전부와 한국정보화진흥원은 민간주도형 디지털 정부서비스 중장기 발전방안을 연내에 수립할 예정이다. 국내외 정책과 적용사례를 조사하고, 앞으로 정부가 지향해야 할 비전과 목표를 수립할 계획이다. 또 국내 공공서비스 중 민간 주도형으로 전환이 필요한 서비스와, 민간 서비스 중 적용 가능한 분야를 발굴할 계획이다. 민간의 디지털 서

비스를 정부가 구매해 국민에게 서비스하는 민간서비스 활용형, 공공서비스 API(애플리케이션 프로그래밍 인터페이스)를 민간에 개방해 국민에게 서비스를 제공하는 공공서비스 개방형 등 다양한 방식을 검토한다. 공공이 민간서비스를 이용·구매하는 방식도 조사한다.

사업이 제대로 확산되려면 관련 법제도 개편이 필수다. 디지털 서비스 전문계약제도 등이 도입될 수 있도록 국가계약법 시행령, 조달사업법 시행령, 예산지침 등이 개정돼야 한다. 국회와 관련 부처 간의 공조와 이해가 반드시 필요하다.

### 현실성 없는 주52시간제-에타제도 보완 서둘러야

코로나19 상황에 문제가 극명히 드러난 주52시간 근로제 보완도 시급하다. 특히 SW 산업은 획일적이고 강제적인 근로시간 규제에서 벗어나 자율적이고 창의적인 환경이 필요한 만큼, 근무시간 제한 특례(예외) 업종으로 지정해야 한다는 목소리가 높다.

구태언 변호사는 “종업원 50~300인 중소기업은 주52시간 근로제의 획일적 시행으로 핵심 기술인재 채용과 육성, 대체가 더욱 어려워지고, 혁신성장 생태계도 후퇴가 우려된다. 중소기업에 대한 제도 적용을 연기하거나 유보해도 대기업·중견기업과의 근무 생태계 격차로 인한 인력난 심화가 현실화될 전망이다”이라면서 “SW·ICT 기술은 개발·오픈, 긴급장애, 시스템 안정화 등 예측하기 힘든 변수가 많아 주말·휴일·야간 등에도 불규칙한 근무가 불가피한 만큼 제도 특례업종에 포함시켜야 한다”고 지적했다.

■ [표 3] 국내외 인공지능 시장 전망

(단위: 세계시장 백만 달러, 국내시장은 십억 원)

구분	2017	2018	2019	2020	2021	2022	2023	CAGR
세계 시장	4,819	7,345	11,283	17,267	25,995	37,987	53,231	49.2%
국내 시장	6,400	7,500	9,100	11,100	13,320	15,984	19,180	20.1%

※ 자료: IITP(2008.12.), “ICT R&D 기술로드맵 2023”, Statista(세계)/과기정통부(국내) 자료 참고

**[표 4] 국내외 빅데이터 시장 전망**

(단위 : 세계시장 백만 달러, 국내시장은 십억 원)

구분	2017	2018	2019	2020	2021	2022	2023	CAGR
세계 시장	53,440	59,040	65,450	72,890	81,380	90,006	99,547	10.6%
국내 시장	244.2	298.7	358.3	446.8	557.2	694.8	866.4	24.7%

※ 자료 : IITP(2008.12.), "ICT R&D 기술로드맵 2023", IDC 자료 참고

또 R&D 직무의 탄력근로제 단위기간을 3개월에서 6개월로 늘리고, 소재·부품·장비분야의 R&D에 한정된 특별 연장근로 인가 범위를 전체 분야로 확대하는 등 전면적인 제도 재설계가 필요하다.

신기술 등장과 기술변화가 시시각각 이뤄지는 ICT, SW 분야의 성격과 맞지 않는 정부 예비타당성 조사 제도도 개선이 필요하다.

무분별한 예산 낭비를 막고 효율성을 높이기 위해 총사업비 500억 원 이상, 국비 지원 300억 원 이상인 신규 사업의 정책적·경제적 타당성을 사전 검증·평가하는 예타 제도를 ICT, SW 연구개발에도 그대로 적용하는 것은 효과보다 부작용이 더 많다. 하루가 다르게 바뀌는 시장에서 10년간 개발내용과 단계별 목표를 제시하고 그에 맞춰 사업을 추진하면 국가 예산낭비로 이어질 수밖에 없다. 예타 대상 사업 규모를 확대하고, 사업 특성에 따라 예타를 면제하거나 유연한 사업내용 변경을 허용하는 등 현실성 있는 제도를 도입할 필요가 있다.

### 국가R&D·원격의료 규제 혁신도 시급

국가R&D와 디지털혁신 체계를 재설계하는 법으로, 각각 내년 1월과 올해 말 시행되는 국가R&D혁신법과 인공지능·정보통신기술법도 제대로 된 상세설계가 필요하다.

국가R&D혁신법은 국가R&D사업 전체의 기준을 담은 최초의 법률로, 부처별로 제각각이어서 생산성과 효율성을 떨어뜨렸던 R&D 기획·관리·평가 기준이 통일된다. 그동안 검증과 크로스체크 없이 정부부처별로 기획하고 집행하던

국가R&D 사업이 통합 관리되고 평가가 일원화되면 R&D 생산성과 성과가 획기적으로 높아질 수 있다. 2018년 기준 660개 사업, 6만 4,000개 과제에 달하는 국가R&D의 틀이 제대로 짜이도록 시행령 등에 발전적인 내용을 포함시켜야 한다. 인공지능·정보통신기술법은 25년간 정보화의 법적 기반이었던 국가정보통신기술법 AI 시대에 맞게 개편하는 것으로, 미·중 등에 뒤처진 AI 경쟁력을 높이기 위한 생태계 확충방안을 시행령 등에 포함해야 한다. 인력양성, 표준화, 데이터 유통, 전문기업 육성 등도 상세설계 해야 한다.

코로나19 상황에서 국가 브랜드 업그레이드에 일조한 'K방역' 효과를 이어가려면 원격의료, 원격모니터링 산업을 열어 의료와 IT 간 융합효과를 극대화하는 것도 서둘러야 한다. 의사협회 등 이해관계자들의 반대가 '타다' 이슈 같이 흘러가는 것을 막으려면 청와대와 주무부처의 확실한 의지가 필요하다. 전문가들은 당뇨·고혈압 등 만성질환 재진을 비롯해 필요한 부분부터 제한적으로 도입한 후 확대하는 방안이 현실적이라는 의견을 내놓는다. 새로운 산업이 성장하면 SW 산업생태계는 그만큼 도약 기회를 얻게 된다.

### 네거티브 규제시스템 전면 도입 필요

지침, 가이드라인 등 법적 효력이 불명확한 각종 기준이 전 분야에서 실질적인 규제로 작동하는 만큼 이 같은 현상규제를 찾아내 혁신하는 노력도 필요하다. 정부부처 홈페이지 등에서 제대로 내용이 공개되지도 않은 채 실제 정책현장에서 강력한 힘을 발휘하는 지침, 가이드라인 등을 전수조사해 개선 로드맵을 수립해야 한다.

궁극적으로는 국가 규제시스템을 포지티브 방식이 아닌 포괄적 네거티브 체계로 서둘러 전환해야 한다. 작년 7월 행정규제법 개정으로 네거티브 규제 전환의 입법적 토대가 마련됐지만 여러 부처에 의한 이중규제 등의 한계 때문에 AI, 헬스케어 등 신기술 분야 산업현장이 체감할 수 있는 효과를 내지 못하고 있다.

전문가들은 빅데이터, 차세대통신, 인공지능, 자율주행차, 드론, 맞춤형 헬스케어, 스마트시티, 가상·증강현실, 지능형로봇, 지능형반도체, 첨단소재, 혁신신약, 신재생에너지 등 정부 13대 혁신성장동력 분야 만이라도 조속한 사전허용, 사후관리 규제 적용 방식을 적용해 신산업을 키워야 한다고 지적한다. 또 13대 혁신성장동력 분야 관련 규제법령을 일괄적으로 네거티브 규제를 적용해 정비할 것을 주문한다.

### ‘국회 쏠림’ 줄이고 정부 거버넌스, 민간 자율성 강화 필요

혁신산업 관련 정책이슈와 업계 간 이해갈등이 있을 때마다 기업도 정부도 국회만 바라보고, 국회가 모든 이슈의 해결사 역할을 하는 구조도 개선이 필요하다. 민간과 행정

부처의 자율성을 강화하고, 대통령이 힘을 실어주는 보다 강한 거버넌스를 구축해 국가 디지털·비대면 경제·사회 전략을 강력하게 추진해야 한다는 목소리가 크다.

이성엽 한국데이터법정책학회장은 “정치가 지나치게 행정과 경제에 관여해 혁신경제와 디지털전환을 막아서는 안 된다”면서 “국회의 역할을 줄이고, 정부 역시 국가 비전을 제시하고 산업계를 지원하는 데 집중해야 한다”고 밝혔다.

전문가들은 기술·산업 이슈가 지나치게 정치적으로 확대돼 민간의 활력을 저해해선 안 된다고 지적한다. ‘타다’ 이슈에서 드러난 문제를 되풀이해서는 안 된다는 것. 합리와 시장논리, 국민 이익을 기준으로 갈등조정을 함으로써 데이터·모빌리티·원격의료 등 미래 산업이 기업 경쟁력과 소비자 이익에 반하도록 설계되지 않도록 중심을 잡는 게 중요하다.

IT와 온라인이 사회변화를 이끌고, 이를 기회로 기업들이 국내·외에서 성장하려면 국회는 과도한 규제를 담은 법·제도를 재설계하는 일에 집중할 필요가 있다. 청와대와 행정부는 대통령이 적극적으로 참여하는 국가 거버넌스와 이해갈등 조정체계를 정비해 원격의료, 혁신 모빌리티, 온라인동영상서비스(OTT) 등을 둘러싼 갈등을 조정하고 국가 전략을 밀어붙여야 한다. 이를 통해 세계 최초로 상용화한

■ [표 5] 개인정보보호법 시행령 개정안 관련 쟁점·보완 사항

조항	내용	업계 의견	행정안전부 입장
개인정보의 추가적인 이용·제공기준 (제14조의2)	<ul style="list-style-type: none"> <li>• 다음 각호의 사항을 모두 충족하는 경우                             <ul style="list-style-type: none"> <li>- 추가처리 목적과 당초 수집목적의 상당한 관련성</li> <li>- 수집한 정황과 처리 관행에 비추어 예측 가능성</li> <li>- 추가 처리가 정보주체나 제3자 이익을 부당하게 침해하지 않을 것</li> <li>- 가명처리로 추가처리 목적을 달성할 수 있는 경우 가명처리할 것</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>- 각호의 사항 ‘모두 충족’ → 각호의 사항 ‘종합적으로 고려’로 수정해야</li> <li>- 상당한 관련성에서 ‘상당한’이 의미하는 바가 불명확</li> <li>- 제3자 범위가 모호</li> </ul>	<ul style="list-style-type: none"> <li>- ‘상당한’ 삭제 검토</li> <li>- ‘수집한 정황과 처리 관행’ → ‘수집한 정황 또는 처리 관행’으로 수정 검토</li> </ul>
가명정보 결합절차 (제29조의2,3,4)	<ul style="list-style-type: none"> <li>- 연계정보 생성(결합키 관리) 기관과 결합전문관 거쳐야 함</li> <li>- 결합전문기관 내 물리적 공간에서만 분석 가능</li> </ul>	<ul style="list-style-type: none"> <li>- 신용정보법과의 형평성 문제</li> </ul>	<ul style="list-style-type: none"> <li>- 가명정보 결합 관련 두 기관 운영은 외부 해킹 공격 대비 목적</li> </ul>
가명정보 안전성 확보조치(제29조의5)	<ul style="list-style-type: none"> <li>- 가명정보 처리 목적 달성되거나 보유 기간 경과하면 지체없이 파기</li> </ul>	<ul style="list-style-type: none"> <li>- 법률상 비식별화된 가명정보 파기의무 면제(삭제 필요성)</li> </ul>	

5G 서비스 위에 원격의료·자율차 등 혁신기술을 연결해 성장기회를 만드는 산업간 협업질서를 만들어내야 한다. 포털과 통신사업 간 협력이 필요하고, 구글·넷플릭스 등 글로벌 기업과 경쟁해 국내 기업들이 성장 발판을 가질 수 있도록 관련 법·제도 정비도 이뤄져야 한다.

## AI 시대 맞는 교육혁신, 디지털 뉴딜 효과 키우기 필요

이미 현실화된 AI 대변혁기를 맞아, 사라지는 일자리가 신생 일자리보다 훨씬 많은 일자리 충격과 노동의 변화에 대비한 교육혁신도 중요한 숙제다. 코로나19가 가져온 위기에 대응한 언택트·원격교육 도입 등 기존 교육시스템 재설계도 서둘러야 한다.

AI·빅데이터 전문가 차상균 서울대 교수(데이터사이언스대학원추진본부장)는 “4차 산업혁명과 AI 확산으로 인한 중하위 계층 일자리 위험에 대한 대응책은 교육을 바꾸는 것밖에 없다”면서 “대학 안에 새로운 혁신의 그릇을 계속 만들고, 지금과는 다른 방식으로 인재 유치에 나서야 한다”고 밝혔다.

전문가들은 국민들이 변화에 적응할 수 있도록 재교육, 평생교육 체계를 전면 개편하고, 변화에 적극적인 대학에 인센티브를 줘야 한다고 지적한다. 기술변혁으로 인한 일자리 상실이 이미 현실로 다가왔는데 우리 산업구조가 전진하지 못하면서 상황이 더 악화되고 있다는 것.

새로운 교육을 이끌 인재 확보가 가장 시급하고 중요한 과제로 꼽힌다. 독일과 프랑스는 대학을 바꾸기 위한 인재 유치에 나서 각각 100개의 교수 자리를 새로 만들었다. 세계 모든 국가가 같은 고민을 하는 만큼 글로벌 협력을 통한 문제 해결이 필요하다고 전문가들은 조언한다. 독일, 프랑스 등과 협력하면서 이미 유럽과 협업 중인 일본과의 관계 개선에도 나설 필요가 있다는 지적이다. 이와 함께 해외 우수 기업과 연구소 유치, 대학 교수의 기업·공기업 겸직 허

용, 해외 대학교수의 국내 대학 겸직 허용·지원, 해외 우수 인재 유치·지원, 글로벌 수준의 문화와 인프라를 갖춘 민간 공동 연구원 설립 등 가능한 방안을 모두 실행에 옮길 필요가 있다. 또 산업현장 재직자들에게 데이터 교육을 시켜 ‘시민 데이터 사이언티스트’를 키우고 디지털 뉴딜의 효과를 양질의 일자리 창출로 이어가야 한다.

김광수 성균관대 교수(SW학과)는 7월 6일 국회에서 열린 ‘국회디지털경제혁신연구포럼’ 출범 행사에서 “디지털 뉴딜 사업을 통해 시작된 단기적인 일자리가 장기적이고 안정적인 일자리로 발전해야 한다”면서 “사업 참여인력에 대한 전문교육을 병행함으로써 기존 AI 산업계에서 데이터를 활용하는 과정에서 필요한 신규 채용과 연계될 수 있도록 해야 한다”고 밝혔다. 이어 “구축된 데이터를 활용해 새로운 AI 서비스를 창출하는 창업 생태계가 활성화될 수 있도록 지원해 새로운 고용 창출이 이뤄지도록 해야 한다”고 제안했다.

전문가들은 디지털 뉴딜 사업이 단기적인 실업 대책이 아니라 궁극적으로 우리 경제 구조를 혁신하고 지속 가능한 일자리 창출을 이뤄내려면 정부 내에 디지털 뉴딜 정책을 강력하게 추진할 수 있는 추진체계를 갖추고, 데이터 구축뿐만 아니라 관련 규제 개선, 산업 생태계 육성, 핵심 인력 양성 등 모든 요소를 함께 추진해야 한다고 한 목소리를 낸다. 사업 추진과정에서 정부가 아닌 민간 부문이 핵심 주체가 되는 것도 반드시 필요하다. AI 학습용 데이터 구축, 5G 인프라 구축 등 디지털 뉴딜 정책의 상당 부분이 민간 부문에 직접 투자할 수밖에 없는데, 이 과정에서 민간 산업 생태계를 왜곡해서는 안 된다는 것.

김광수 교수는 “2000년대 초반 벤처 생태계가 지속적으로 성장하지 못했던 사례나 중소기업 R&D 지원이 좀비 기업을 양산하는 사례 등 정부의 직접 개입이 부작용을 초래한 경우는 흔히 볼 수 있다”면서 “AI 데이터 구축 대상과 활용방법, 네트워크 인프라 구축 방법 등 많은 부분에서 민간이 주체가 되도록 하고 민간의 수요를 최우선적으로 고려해 추진해야 한다”고 지적했다.

## 21대 국회에 바라는 SW산업 이슈

소프트웨어산업진흥법이 18년 만에 『소프트웨어진흥법』으로 전면 개정되었다. 그간 산업계가 줄기차게 요구해 온 해묵은 소프트웨어(SW)산업 선진화 숙제를 해결하기 위해 48개조에서 78개조로 확대되었으며 원격지개발, 과업변경에 따른 계약금액조정, 하도급제한, 상용SW 사용 촉진을 규정했다. 그 이외에 SW안전을 강조(제30조~제31조)하여 관련 사업이 형성될 것으로 보이고, SW진흥단지지정 및 조성(제12조), SW창업 활성화(제14조), SW인력 양성(제22조~제23조), SW기술자 우대(제37조), SW산출물 활용 보장 등 지식재산권 보호 강화(제59조)를 규정하는 등 SW산업 전반이 성장할 수 있도록 장치를 마련하였다.

최근 4차 산업혁명의 핵심이며 모든 산업의 혁신의 도구로써 인식되는 SW는 코로나19로 인한 “뉴노멀”시대의 경쟁력 강화와 체질개선의 수단으로 더욱 각광을 받고 있다. 지난 5월, 20대 마지막 국회에서 『소프트웨어진흥법』 전부개정안이 통과되었으며 7월에는 SW가 중심이 된

● ●  
조영훈  
한국소프트웨어산업협회  
산업정책실 실장  
lv105@sw.or.kr

“한국판 뉴딜”이 발표되어 업계, 학계 등 관련된 분야에서 기대의 목소리가 커지고 있어 개정된 소프트웨어진흥법에 포함된 내용과 그 이외의 우리가 준비해야 할 이슈들을 살펴보고자 한다.

## 국내외 SW산업계 이슈

과학기술정보통신부(이하 과기정통부)의 SW전략은 2018년 9월 발표된 “4차 산업혁명 일자리 창출을 위한 소프트웨어 혁신성장 전략”이며, 핵심적인 내용은 SW분야의 ‘산업환경개선’과 인재양성’으로 정리될 수 있다.

산업환경개선은 그간 SW산업계에서 발주자와 수주자 간, 원사업자와 하도급사업자 간, 고용주와 고용인 간의 불합리한 관행을 개선하여 산업발전과 이를 통해 일자리 창출의 선순환을 이루고자 하는 전략이다. 공공SW사업에서는 법과 제도를 통해 많은 부분이 개선되었지만, 수발주자 간의 불공정 관행으로 적정예산확보(SW제값 받기)가 되지 않으며, 이로 인해 SW품질은 낮아지고 SW산업계의 역량이 발전하지 않고 있다는 문제의식이 늘어나고 있다.

인재양성은 SW교육도입, 즉, 초·중·고교에서 SW교육을 시행하는 것으로 초기 교육부터 SW적인 체계적 사고와 간단한 코딩(Coding)으로 SW에 대한 흥미유발과 관심을 유도한다. 이후 고학년으로 가면 마이스터고등학교, SW중심대학 등에서 보다 전문적인 SW교육을 통해 향후 SW전문인재로 키워 산업계의 중심인재로 활용한다는 것이 주요 내용이다. 흔히 말하는 SW산업은 제조업이나 건설처럼 자본과 시설, 장비보다 사람의 아이디어와 기술력이 더 중요하기 때문에 인재의 수급은 늘 문제로 지적되어 왔다. 기업들이 지적하는 수급의 문제는 첫 번째는 SW산

업이 금융이나 전자와 같이 소위 잘나가는 직종이 아니라는 인식으로 지원이 계속 줄고 있다는 것이다. 실제 대학에서도 얼마 전까지는 SW분야 관련 학과가 축소되거나 지원자가 감소하는 추세였으며 SW업종은 반복되는 야근에 현장출장(정부기관 및 공기업의 지방이전)으로 기피하는 업종으로 인식되었다. 즉, SW분야 전공자들이 사회에 진출했을 때, 고용이 불안정하고 적당한 임금이 보장되지 않으며 SW전공자로 양성된 인재들이 향후에 취업하거나 스타트업, 벤처창업 등의 활로가 보장되지 않는 구조적 문제에 대한 불안이 팽배했다고 생각된다. 두 번째는 국내 SW산업이 그동안 SI(System Integration)중심으로 편중되어 제품을 대량생산하듯이 반복되는 단순 코딩에 대해 부정적 인식이 누적되어 왔다. 게다가 SW제품은 국산을 선호하지 않는 분위기로 인해 뛰어난 인재들에 대한 정당한 대우가 보장되어 있지 않았던 문제가 글로벌 기업과 대비되면서 SW산업은 기피하는 분야로 굳어졌다. 마지막으로 벤처창업을 하더라도 미국이나 유럽처럼 인수/합병을 통해 부를 축적하거나 새로운 분야로 도전할 수 있는 상황이 아니라 어느 정도 안정권에 접어들어도 겨우 회사의 생존을 위한 사업에 불과하여 인재들이 미래지향적 비전을 가질 수 없는 현실이었다.

최근의 상황은 4차 산업혁명, 국외 글로벌 기업의 출현으로 SW에 대한 인식은 긍정적을 넘어서 경제의 핵심이자 유망 직종으로 바뀌었다. 최근 등장하고 있는 유니콘기업(기업가치 10억\$ 이상 비상장기업)의 대다수가 콘텐츠 플랫폼이나 공유경제 플랫폼 비즈니스 모델을 채용하고 있어, 기업의 차별적 경쟁력에 근간을 둔 SW역량이 매우 중요하게 인식된다. 때문에 우리나라 정부에서도 ICBM(IoT, Cloud, Big Data, Mobile)이나 DNA(Data, Network(5G), AI) 등 신기술분야를 계속 지원하고 발전시켜 새로운 시장

창출과 일자리 창출을 위해 노력하고 있다. 하지만 국내외 SW시장규모의 낙관적인 성장 전망에도 불구하고 근로시간단축, 일본의 무역제재, 미·중 간 무역분쟁 등 예상하지 못했던 이슈가 발생하여 SW산업의 획기적인 발전 및 일자리 창출도 아직까진 큰 성과를 체감하지 못하고 있다. 특히 코로나19 감염병으로 인해 비대면(Untact) 비즈니스가 새롭게 부상하고 있고 글로벌 협업보다는 각자도생의 국내자급이 부각되는 등 기존의 방식을 근본적으로 바꾸어 뉴노멀시대를 대비해야 한다는 이슈도 있어 SW진흥법 및 하위법령에서 다루는 사항 이외에 몇 가지 현실적인 산업 현장의 목소리를 전하고자 한다.

## SW 교육

교육부는 교육부 고시 제2015-74호(2015.9.23.)에서 초·중·고교에서의 ICT(Information and Communication Technology)교육을 강화하였다. 초등학교 5~6학년 중 '실

과' 과목 내에 ICT중심의 정보 교육을 SW기초소양 중심으로 개편하여 17시간 이상을 필수로 이수하도록 실습·체험 위주의 내용을 구성하였으며, 중학교는 '정보'과목을 선택에서 필수과목으로 전환하여 34시간 이상 이수하도록 하였다. 고등학교는 심화선택 과목인 '정보'를 일반선택 과목으로 전환하고, 과학계열인 '정보과학'에서 프로그래밍과 알고리즘에 대한 심화 교육이 가능하도록 구성하였다. 이는 2018년도에 중학교 1학년을 시작으로, 단계적으로 2019년에 모든 초등학교 5학년 또는 6학년과 중학교 2학년까지 확대되고, 2020년에는 중학교 3학년까지 확대 운영된다.

초·중·고에서 ICT 교육을 강화하는 것은 일단 반갑다. 2018년과 2019년 SW교육을 시행했는데, 2년간 수행한 결과를 가지고 성과판단을 하는 것은 성급할 수 있다. 그런데 초등학교 5, 6학년의 2년간 17시간, 중학교 3년간 34시간의 수업이 컴퓨팅사고 기반 문제해결 능력을 키우는 데 최소의 시간이라도 될 수 있는지에 대해서는 비판의 여지가 있다. 중학교 3년에 34시간이면 한 학기에 5.6시간에 불과

■ [그림 1] 초·중·고 2015 교육과정 개편 내용

구분	현행	개편안	주요 개편 방향
초등학교 (2019년~)	실과 내 ICT 단원(12시간)	실과 내 SW 기초 교육(17시간 이상)	· 문제해결과정, 알고리즘, 프로그래밍 체험, 정보윤리 의식 함양
중학교 (2018년~)	'정보' 과목 (선택교과)	'정보' 과목 34시간 이상(필수교과)	· 컴퓨팅사고 기반 문제해결 실시 · 간단한 알고리즘, 프로그래밍 개발
고등학교 (2018년~)	'정보' 과목 (심화선택 과목)	'정보' 과목 (일반선택 과목)	· 다양한 분야와 융합하여 알고리즘, 프로그램 설계

하다. 일주일에 한 시간도 되지 않는데 학생들이 SW를 이해하고 적용할 것을 기대하기 어렵다. 성인에게도 이는 어려울 것이다. 일주일에 1시간 남짓 교육으로 무슨 목표를 달성할 수 있을 것이며 교재나 강의내용에 내실을 기하기도 어려울 것이고 교사나 학생의 능력이 뛰어나도 원하는 교육 효과를 얻기 어렵다. 다른 한편으로 SW의무교육이 입시의 한 종류로 인식될 수 있다는 목소리도 아직 있다. SW에 관심이 있어 이를 직업으로 고려하기 보다 진학률이나 취업률이 좋은 마이스터고등학교를 거쳐 SW중심대학에 진학하면 다른 전공자보다 취업에 유리한 조건에 있을 수 있다. 거의 멸종에 가까웠던 SW프로그래밍 전문학원들도 학생들의 관심으로 생겨날 것으로 기대하고 있지만 일부에서는 입시학원화될 것을 우려하고 있다.

SW교육에 대한 개선의견은 첫 번째로 “관심”-“사고(思考)”-“기술”-“활용”의 단계적으로 명확한 목표를 마련해야 한다. 초등학교에서는 학생들이 SW산업이라는 분야도 있고 이런 것이 직업이 될 수 있다는 것만 심어주어도 충분할 것 같다. 초등학교에서는 SW에 대한 관심을 이끌고 중학교에서는 실제 간단한 코딩을 통해 문제를 해결해보며 고등학교나 대학에서 전문적인 교육을 통해 SW전문인력을 배출하여 산업계와의 선순환을 만들어 내는 그림에 맞게 방향수정이 필요하다. 아이들에게 최초의 컴퓨터가 ABC인지 ENIAC인지, 프로그래밍을 잘하려고 이진법과 함수를 중점적으로 가르친다면 관심은 커녕 기피심을 심어주게 될 것이다. 따라서 체계적인 교육보다는 재미있는 교육이 더 중요하며 좀 더 시각적인 교재와 커리큘럼이 필요하다.

학교를 졸업하고 SW에 관심이 많아서 많은 기술이 습득된 사람도 실제 기업이나 현장에 활용되려면 어느 정도의 추가교육은 있어야 한다. 대부분 표준과 범용을 배우는 것인데 기업은 여러 사람이 빠른 이해와 업무의 인수인계

가 필요하기 때문이다. 이 단계가 지나고 현장에서 전문가가 되어 일을 하다 보면 반드시 재교육도 필요하다. SW는 하드웨어와 마찬가지로 3~4년만 지나면 유행이 지나거나 새로운 기술로 대체되어 지속적인 교육이 타 산업보다 더 필요하다. 빠르게 변화하는 SW에 따라 교육 프로그램도 변화해야 한다. 교육 대상도 초·중·고와 대학에 재학 중인 학생에 국한되지 않는다. 취업준비생과 재직자를 대상으로 일종의 보수교육이 진행되어야 하므로, 관련한 교육 프로그램을 꾸준하고 보완하거나 새롭게 만들어야 한다.

두 번째는 정보소외계층이나 중장년층의 SW평생교육이다. 이는 원리보다 활용위주로 지속 교육이 되어야 한다. 이전에 주민센터 단위의 컴퓨터교육이 정보격차 해소에 큰 역할을 했고 이런 교육으로 인해 IT활용능력 향상으로 K-방역을 성공적으로 이끌고 있다고 생각한다. PC사용법, 인터넷접속법, 검색엔진사용법 등의 교육을 현재에 맞게 PC에서 Mobile로, 웹(WEB)에서 앱(App), PC뱅킹에서 모바일뱅킹 등 커리큘럼을 개발하고 보급 및 교육에 힘써야 한다. 또한 SW전문성이 필요한 강사육성(교육)도 충분히 고민해야 하며, SW산업계의 은퇴자 및 경력단절자를 채용하여 이들의 전문지식을 활용, 전 국민에게 평등한 교육을 시행하는 것도 검토해야 한다.

## SW M&A 환경 조성

경제가 발전함에 따라 SW산업계에서는 쉽게 개선되지 않는 문제가 있다. 바로 창업과 M&A이다. 언제부터가 우리사회가 젊은이들에게 “기업가정신(Entrepreneurship)”을 강조하며 창업을 권유하고 있는데 이는 더 생각해 볼 필요가 있다. 혁신적인 아이디어가 성패를 좌우한다며 SW분



야가 창업하기 좋고 성공하기 쉬운 산업이라고 말한다. 물론 제조업, 건설업 등 타 산업보다 자본이나 기술에 대한 의존도가 적고 소수의 뛰어난 인적능력이 좌지우지하는 것은 사실이지만 창업을 하려면 좋은 아이디어만 있어서 되는 것은 아니다. 혁신적인 서비스 모델을 개발해도 기업을 운영하려면 경험이 있어야 한다고 생각한다. 자금을 대한 운영 능력도 필요하고 직원의 능력을 100% 발휘할 수 있는 환경조성과 공감대형성도 필요하고 각종 법·제도도 파악하여 문제가 없도록 해야 한다. 이런 중요한 것을 사회에 첫발을 내딛는 젊은이가 충분히 대비하고 기업을 유지할 수 있을까? 아니 불가능에 가깝다. 설사 소 뒷걸음 치다 쥐 잡는 격으로 초기에 사업이 잘 풀렸다고 치자, 그 다음 기업의 비전과 목표는 어느 수준으로 잡아 진행해야 하며 직원이 몇 배

늘어났을 때 어떻게 운영해야 하는지는 누구도 가르쳐주지 않는다.

벤처기업이 성공적으로 진행된다는 미국에서는 우리와 가장 다른 면이 M&A라고 생각한다. 정말 혁신적인 아이디어로 집 안 차고에서 사업모델을 만들고 실제 가능성을 보여주면 투자자와 쉽게 만날 수 있다. 여기서 더 성공하면 이 기업을 다른 큰 기업이 인수/합병하여 전문 경영인이 비즈니스를 더 크게 만들고, 최초의 아이디어를 냈던 기업은 인수 금액을 통해 다른 비즈니스 아이디어를 만드는 구조이다. 우리나라의 경우 벤처기업의 아이디어를 제 값을 주고 사겠다는 생각은 부족한 것으로 보인다. 카카오가 2015년 '김기사'를 600억원에 인수하였으며, 2019년에는 독일의 DH가 우리나라의 '배달의 민족'을 4조 7500억원에 인수했

다. 이러한 소식은 반갑기는 하나, 일반적인 것이 아니다. 대부분의 경우 우리나라 벤처기업은 외국의 벤처기업에 인수 되는 것을 기다리거나, 우리나라 대기업에게 아이디어를 뺏기지 않도록 경계하며 버티는 것이 할 수 있는 일의 전부일 수 있다. 즉 복싱선수가 링에 올라가서 경기를 하는데 상대방이나 내가 KO되어야만 끝나는 게임을 한다. 한 게임을 이겨도 쉽 없이 다음 상대와 죽을 때까지 경쟁해야 하는 것이 우리나라 SW생태계인 것이다. 그래서 SW산업계에서 중소기업이 대기업으로 성공한 사례를 쉽게 볼 수 없는 것이고 SW기업의 역사가 오래가지 못하는 이유가 된다.

SW의 신기술과 새로운 비즈니스는 대부분 중소기업에서 시작할 확률이 높다. 왜냐하면 중소기업은 대기업이나 시장을 주도하는 기업의 니치마켓을 노리기 마련이고, 더 적은 비용의 서비스를 찾아야 후발주자로서 경쟁력이 있기 때문이다. 하지만 어느 정도 시간이 지나면 이러한 신기술과 비즈니스는 중소기업에서 조용히 사라지고 대기업이 주도하는 경우가 많다. 클라우드(Cloud)가 그랬고, 핀테크(FINTech)가 그랬고, AI도 그럴 것이다. 물론 사업의 성패를 위해 많은 인력과 자본이 필요하므로 대기업이 살아남을 확률이 높겠지만 주된 원인은 기업을 M&A하지 않기 때문이다. 어느 정도 시간만 버티면 중소기업은 망하거나 사업 전환을 하므로 여력이 있는 기업은 굳이 비용을 들여 인수/합병을 하지 않아도 관련 인력과 기술을 쉽게 확보할 수 있기 때문이다.

탈출구가 보이지 않는 기업경영을 죽을 때까지 하라고 강요하는 것은 금해야 한다. 선배기업가들은 M&A를 미덕으로 인식하고 실행하여 후배들에게 좋은 환경을 만들어 주고 창업을 권해야 한다. 즉 중소기업은 혁신적인 아이디어를 실행해보고 이를 대기업이 받아 비즈니스로 더 많은 이익을 창출하는 선순환적인 생태계가 필요하다. 그리고 스타트업에게는 세제나 금융지원 등 여러 가지 지원방법도 좋

겠으나 가장 중요한 것은 혁신적인 아이디어를 인정할 수 있는 제도, 즉 지적재산권 활성화와 강력한 보호시책을 마련해 주고 창업을 권해야 한다.

## 바람직한 SW 생태계 조성

코로나19로 인해 많은 것이 바뀔 것으로 예상하는 전문가가 많다. 기존의 방식으로는 현재와 미래를 대응하기 어렵다. 일하는 방식, 비즈니스 서비스, 관리방법도 새롭게 제시될 것이다. 분명한 것은 어떻게 변화하던 간에 중심은 SW라는 것에 큰 이견이 없는 것으로 알고 있다.

이번에 통과된 SW진흥법은 그간의 불공정한 관행을 개선하고 새로운 SW문화와 기술이 무리 없이 실생활에 녹아들 수 있도록 마련되었고 하위법령 작업도 잘 진행되고 있다. 그간 산업계가 외쳐왔던 제값 받기, 적정예산확보, 기술자처우향상, 지적권보호 등은 대부분 SW진흥법에 담겨 있다. 법도 중요하지만 법을 지킬 수 있도록 장치를 마련하는 것도 우리의 숙제이며 어렵게 맞이한 SW대도약시기에 모두가 합심하여 K-드라마, K-POP, K-방역에 이어 K-SW라는 메가트렌드로 이어지길 바란다.

# 인공지능의 혁신특성과 한국 스타트업 사례연구



## Executive Summary

코로나 팬데믹 상황으로 인해 디지털 전환과 인공지능이 다시 조명을 받고 있다. 그간 인공지능을 중심으로 한 테크 기업들의 거품 논란과 인공지능의 세 번째 겨울이 다시 도래한다는 비관적 견해도 일부 존재해 왔다. 하지만 머신러닝을 중심으로 한 인공지능의 사업적 가치가 이미 현실화되기 시작했다. 제한적인 응용 분야에서 작은 성과들을 내고 있지만 이제 막 본격적으로 꽃을 피우고 있다고 해도 과언이 아니다. 인공지능은 기존 도메인 산업의 가치사슬을 분화 및 해체(Unbundling)하고 이를 재구성(Constellation)하여 단위 프로세스만으로도 다양한 가치를 창출할 수 있는 독특한 세부 산업(Sector)을 형성해 나가고 있다. 자원과 역량이 부족한 신생기업들에게도, 보유한 기술 역량에 맞는 도메인 산업을 잘 선별하여, 전문화된 니치 영역으로 진입할 수 있는 다양한 기회가 앞으로 더 풍부해질 것이다. 이 보고서는 인공지능의 혁신 특성을 정리하고 카테고리 킬러로 자리매김하고 있는 한국 신생기업의 사례들을 살펴본다. 또한, 그들이 겪고 있는 다양한 산업적 이슈와 애로사항을 토대로 그간 주목받지 못했던 기업들의 제도적 역량의 중요성을 강조한다.

● ●  
강승희  
선임연구원  
dellabee@spr.kr

The digital transition and artificial intelligence are on the rise again due to the corona pandemic situation. There have been some pessimistic views that the bubble controversy of tech companies centered on AI and the third winter of AI are coming again. However, the business value of artificial intelligence, centered on machine learning, has already begun to become a reality. Although they are making small achievements in limited applications, it is not an exaggeration to say that they are just starting to bloom. Artificial intelligence is creating a unique sector that can re-create and diversify the value chain of the existing domain industry, unbundling it, and reconstituting it, turning a hidden unit process into an explicit unique value proposition. Even for start-ups that lack resources and competencies, various opportunities to enter the specialized niche area by selecting a domain industry that fits their technological capabilities will become more abundant in the future. This report summarizes the innovation characteristics of artificial intelligence and looks at the cases of Korean startups that are positioned as category killer. It also emphasizes the importance of institutional competencies of companies that have not been noticed based on various industrial issues and difficulties they are experiencing.

## 1 연구 배경

코로나 팬데믹 상황으로 인해 비대면 원격 근로, 교육, 진료 등이 화두다. 이러한 비대면 서비스의 확산은 디지털 기술이 있기에 가능했고, 곧 디지털 전환의 가속화와 맞닿아 있다. 디지털 전환에 대한 요구가 증가할수록 인공지능에 대한 투자도 일반화되고 있다. 최근 대중화된 스트리밍 방송이나 원격 교육의 품질을 좌우하는 카메라·방송 화질의 고도화와 개선이 모두 딥러닝 기반으로 전환되고 있는 점을 예로 들 수 있겠다.

그간 우리는 인공지능을 좁은 의미의 기술로 받아들여 산업·기업의 현장과 사회·개인의 일상에 적용하기에 앞서 이 기술의 경제적 파급효과와 부작용에 대해 걱정 섞인 열띤 토론을 이어왔다. 인공지능이 이뤄내는 인지 노동의 자동화로 인한 노동 시장의 변화, 지식의 생명주기 변화로 인한 조직의 변화, 디지털 전환 갈등, 사이버 보안 이슈, 의사결정 편향과 양극화 등 부작용과 불확실성으로 인한 두려움과 우려에 보다 초점을 맞추었다. 위험에 대한 인지와 대비는 물론 중요하지만, 과도한 우려를 걷어내고 이성적으로 바라보면 기술은 가치 중립적인 도구일 뿐이다. 도구를 어떤 목적으로, 어디에 어떻게 활용하느냐에 따라 우리가 선호하는 미래도, 디스토피아도 모두 현실화될 가능성이 있다. 그렇다면 우리는 보다 나은 미래를 위해 무엇을 어떻게 준비해야 하는가?

돌아보면 과학 기술 발전의 오랜 역사에서 기술 혁신은 필연적으로 자동화와 일자리의 변화를 낳았다. 이러한 기술 궤적, 혹은 진화의 경로 속에서 사회적 갈등이 촉발되기도 하고, 이를 통해 사회 계층이 재결집하고 중심점이 이동하면서 사회가 점진적으로 기술 진보를 받아들이는 과정이 되풀이되었다. 아무리 갈등 상황이 극적으로 치달게 되더라도 사회는 기술 혁신을 받아들이는 쪽으로 불가역적으로 진보해 왔다. 그러나 그 변화의 경로는 기술의 특성에 따라 달랐다.

따라서 이 이슈리포트는 최근 패턴 인식 등의 인지 과정 자동화를 급격하게 가속화시키고 있는 머신러닝을 중심으로 인공지능이라는 기술이 어떤 혁신 특성을 갖고 있는지 먼저 파악해 볼 것이다. 또한, 인공지능 기술을 전문 서비스 분야에 적용·융합한 한국 대표 신생 기업 사례를 통해 후발 기업이 어떻게 하면 생존 및 성장해 나갈 수 있는지, 이슈 사항과 제도적 압력은 무엇인지에 대한 고민을 담아 보고, 시사점 및 결론으로 마무리할 것이다.

## 2 AI 혁신 특성

인공지능(AI, Artificial Intelligence)은 지능적 행위자(Agent)의 설계에 관한 이론 및 시스템(Poole et al, 1998), 혹은 인지와 추론, 행위를 가능하게 하는 계산에 대한 연구(Winston, 1992)라고 정의할 수 있다. 가장 오래된 정의로, 인간의 사고, 그리고 의사결정, 문제 풀기, 학습 등의 활동에 연관시킬 수 있는 활동들의 자동화(Bellman, 1978)라고도 할 수 있다. 이러한 다양한 관점을 아울러, 인공지능이라 하면 일반적으로 시각 인지, 음성 인식, 의사결정 및 언어 번역 등 인간의 지능이 필요한 분야를 실행할 수 있는 이론 및 컴퓨터 시스템을 일컫는다. 최근 AI의 빠른 성장 동인은 모바일과 클라우드의 활용, GPGPU, 다양한 가속기를 포함한 컴퓨팅 성능 향상, 빅데이터 해석 과정의 지능화, 새로운 알고리즘의 발전으로 압축하여 설명할 수 있다(Deloitte, 2019). 인터넷 및 모바일 기기의 높은 활용도는 막대한 양의 데이터와 정보를 인터넷에 연결된 모든 기기에 제공하고 있다. 이는 빅데이터 기반으로 인공지능의 활용 가능한 데이터 총량의 측면에서 급증을 가져왔고, 이에 기초한 확률 추론 통계 모델의 진화를 가져오고 있다. CPU와 GPU의 성능으로 대변할 수 있는 컴퓨팅 성능의 향상 또한 머신 러닝 모델 정교화를 가속화하고 있으며, 새로운 알고리즘을 통해 머신 러닝 시각 인지 기술 등의 기술 발전을 촉진하고 있다. 1986년 제프리 힌튼은 역전파 기술이 둘이나 셋 이상의 계층을 가진 심층 신경망을 훈련시킬 수 있다는 혁신적인 연구를 발표했으며, 이것이 최근 각광을 받고 있는 딥러닝 시대를 열었다.

인공지능을 기술이 아닌 산업(Sector)이라 할 수 있는가? 인공지능은 산업(Industry) 분류 체계 속에서는 SW에 속하지만, 전통적인 SW산업과 달리 각 도메인 산업에 적용하는 방식에 있어 다른 혁신패턴을 보인다. 물론, 인공지능은 SW와 완벽하게 분리될 수는 없는 개념이다. 이메일을 보내는

SW 프로그램과 인공지능으로 구현된 스팸 필터가 함께 상호 보완적으로 존재하듯이 말이다. 그렇다면 전통적인 SW와 인공지능은 어떤 관점에서 서로 다른 특성을 보이는가?

IT서비스로 대표되는 전통 SW산업은 선형적 가치사슬 기반의 전략적 경영방식 속에서 가격과 비용의 최소화라는 목적함수에 의해 순차적이고 선형적인, 프로세스의 연결성과 통합이 중요하다. 규칙 기반의 자동화 과정으로 대표된다고 요약할 수 있겠다. 전통 SW산업이 도메인 산업에 기여하는 방식은 위와 같은 과정에 따라 범위의 경제, 규모의 경제라는 오래된 공식을 적용하는 것이었다. 선진 프로세스라는 이름으로 모범사례(Best Practice)가 강요되기도 했다. 하지만, 인공지능은 기존 도메인 산업의 가치사슬을 분화 및 해체(Unbundling)하고 이를 재구성(Constellation)하여 단위 프로세스만으로도 다양한 가치를 창출할 수 있는 산업(Sector)을 형성해 나가고 있다. 기존 도메인 산업의 프로세스 중 일부에 불과했던, 그간 자동화 및 표준화가 어려웠던 마진을 남기지 않는 단위 프로세스를 떼어내어 산업(Sector)화하고 있는 것이다.

그렇다면 어떤 인공지능의 특성이 이를 가능케 하는가? 전통 SW와 차별되는 인공지능만의 혁신 특성은 패턴 등의 인지 과정을 자동화할 뿐 아니라, 학습에 기초해 스스로 진화한다는 점이다. 이를테면, 구조방정식 등의 개념에 기초하여 정형화되지 않은 입력값을 받고도 비교적 정확한 결과값(예: 분류 카테고리)을 예측할 수 있고, 기존 알고리즘에 포함되어 있지 않은 분류 공간도 새로 추가하여 지속적으로 학습하며 알고리즘, 혹은 분류모델을 진화시킬 수 있다(Rao et al., 2019). 소스 코드를 고치는 것이 아니라 그간 쌓여 있던 데이터를 활용하여, 더 난이도가 높고 복잡한 물리적·정신적 노동집약적 과정을 자동화할 수 있도록 하는 것이다. 브린올프슨 등(2018)은 이를 자동화의 자동화라고 불렀는데, 대규모 학습 인프라가 필요한 아마존, 구글, 마이크로소프트 등에서 핵심 기술을 확보하고 있는 AutoML<sup>1</sup> 사례가 그러하다. 자동화된 알고리즘 작성의 자동화 정도로 해석이 가능하다. 이 AutoML은 머신 러닝 모델의 구조를 자동으로 선택하고 작성 혹은 개선해 준다.

다음 장에서는 이와 같은 인공지능의 혁신 특성에 대한 이해를 바탕으로, 비교적 늦게 인공지능 기술 개발에 뛰어든 한국에서 니치 시장을 공략하여 성공을 거둔 기업들의 사례를 통해 한국에서 가능한 AI 혁신 창출을 위한 경로의 실마리를 찾아본다.

<sup>1</sup> AutoML은 머신러닝 모델 개발 시 시간이 많이 걸리는 반복적인 작업을 자동화하는 프로세스로, 이를 통해 데이터 사이언티스트, 분석가 및 개발자가 모델 품질을 유지하면서도 확장성, 효율성 및 생산성이 높은 머신러닝 모델을 구축할 수 있다. 나아가, 비전문가도 별도의 전문지식 없이 머신러닝 모델과 기술을 활용할 수 있다. AutoML이 적용되는 프로세스는 데이터 사전처리, 특성(Feature) 엔지니어링, 모델 선택, 하이퍼파라미터 최적화, 파이프라인 선택, 평가 지표의 자동 선택과 유효성 검증 절차, 문제 탐지 등이다.

### 3 AI 혁신 사례

#### 3.1 사례기업의 선정

이 보고서에 수록된 AI 혁신 사례는 주요 서비스 산업인 의료, 금융, 마케팅/광고 분야를 선정 하여, 기존 산업 환경을 AI 기술 기반 혁신을 통해 변화시키고 있는 창업 7년 내의 각 분야 대표 스타트업 (카테고리 킬러)들을 인지도와 투자유치, 매출성장을 기준으로 선별하였다. 이 사례 기업들은 기존에는 그 자체로는 부가가치를 창출하지 못했고, 자동화가 어려웠던 단위 프로세스들을 자동화하면서 니치 시장을 공략하고 있다는 공통점이 있다.

각 분야 대표 사례 기업의 목록과 간략한 소개는 다음과 같다.

요소	대표기업(업력)	소개	주요 기술, 제품 및 서비스
의료	A사(6년) - CTO 인터뷰	<ul style="list-style-type: none"> <li>• 누적투자액 170억 (기업가치 1,400억 원)</li> <li>• 의사 포함 90여 명 규모</li> <li>• 국내 허가 의료 AI 기술 3여 개, 임상시험 종료 2개, 임상시험 진행 중 1개 등</li> </ul>	<ul style="list-style-type: none"> <li>• 골연령 진단 보조 기술</li> <li>• 흉부 X-레이 진단 보조기술</li> <li>• 폐암 조기 검진 보조기술</li> <li>• 치매진단 보조기술</li> <li>• 안저영상 분석 기술</li> <li>• 입원환자 대상 실시간 생체신호 기반 응급 질환 예측 기술</li> <li>• 판독용 음성 인식 기술</li> </ul>
금융	B사(4년) - CTO 인터뷰	<ul style="list-style-type: none"> <li>• 누적투자액 120억 이상</li> <li>• 30여 명 직원규모</li> <li>• 인공지능 기반 금융투자, 자산운용 솔루션</li> </ul>	<ul style="list-style-type: none"> <li>• 딥러닝 기반 로보어드바이저 (국내시장 80% 점유)</li> <li>• 인공지능 상장지수펀드(ETF) 뉴욕 거래소 상장</li> <li>• 인공지능 주문 집행(AXE)</li> </ul>
마케팅/광고	C사(7년) - 이사 인터뷰	<ul style="list-style-type: none"> <li>• 누적투자액 약 300억</li> <li>• 100여 명 직원규모</li> <li>• 인공지능 기반 모바일 마케팅, 자동 광고 집행(애드테크) 기업</li> </ul>	<ul style="list-style-type: none"> <li>• 자동광고 집행 퍼포먼스 엔진</li> </ul>

#### 3.2 의료

##### ■ 저출산·고령화·저성장 시대와 의료 패러다임의 변화

한국은 1970년 이후 OECD 국가 중 기대 수명이 가장 획기적으로 개선된 터키(24년↑)에 이어 2위 (20년↑)를 차지했다(James, Devaux and Sassi, 2017). 그 영향 요인 중에서는 교육·소득 수준의 증가, 생활·위생 환경의 개선뿐 아니라 보건 의료 시스템 및 정책의 획기적인 개선이 주목받았다. 특히,

1997년 건강보험의 도입이 주요한 영향으로 작용한 최근 20여 년간 일반 국민들의 의료 서비스에 대한 물리적·경제적 접근성과 보장성이 증대되어 2017년에는 기대 수명이 74.7세에서 82.7세로 늘어났고, 이는 OECD 평균 80.7세를 상회하는 수준이다.

하지만 앞으로의 전망은 그다지 밝지 않다. 전 세계 평균보다 3배 이상 빠른 고령화 속도, 2015~2020년 합계출산율 기준 세계에서 가장 낮은 출산율(1.11명)을 보이는 초고령화·저출산 시대로 돌입하고 있는 것이다(통계청, 2019). 중장기적으로 의료비 지출을 포함한 노년부양비의 부담이 급진적으로 증가하고 있는 데다, 그 짐을 짊어질 주체가 줄어들고 있다. 더구나 이러한 인구구조의 변화, 생산인구 감소에 따른 저성장 추세, 제4차 산업혁명과 인공지능 시대의 도래로 인해 생산·노동환경이 변화하는 등 거시적인 사회경제적 제반 환경 변화의 방향성은 우리가 선호하는 미래 발전 방향과 궤를 달리하고 있다.

이러한 환경 변화 속에서, 20세기 후반에 주를 이루었던 다수의 환자에게 입증된 누적된 통계적 치료법을 우선 적용하는 표준화된 증거 중심의 의학(Evidence-based Medicine) 패러다임도 변화하고 있다. 환자마다 각기 다른 특성을 가진 환경, 유전 요인과 질병 경력, 건강 생활 습관 등을 사전적으로 인지하여 맞춤형 치료 및 복약 지도를 하는 맞춤 의학(Personalized Medicine), 나아가 분자 프로파일링 기술을 임상병리학에 도입한 정밀 의학(Precision Medicine)으로 변모하고 있는 것이다(Lu et al., 2014; 한국보건사회연구원, 2017). 디지털 헬스케어는 디지털화된 맞춤, 정밀 의학을 포괄하는 개념인데, 의료 전달 체계, 전통 의료산업 간 경계의 변화, 지불 시스템의 혁신, 의료 서비스 품질 관리 기법의 발전 등으로 인해 의료인의 역할과 의료 서비스의 양태, 관련 제도의 변화를 설명하기에 적합하다(최윤섭, 2014). 시대의 요구에 맞는 예방 및 조기 관리 중심의 의료 체계 전환은 예방(Preventive), 예측(Predictive), 참여(Participatory), 맞춤(Personalized)의 4P 전략으로 디지털 기술을 중심으로 추진되어, 의료의 영역을 확장 시키고 있다. WHO에서는 국가 수준의 ICT 기반 의료 혁신을 위한 국가 수준의 전략과 정책(National eHealth Strategy Toolkit, 2012)을 이미 강조한 바 있으나, 한국은 보건 의료 분야 ICT 투자 장려 수준이 타 국가보다 낮은 수준이다(김정곤, 이서진, 2016).

## ■ 기존 의료 분야의 가치사슬 생태계 변화

혁신적 사업자의 등장으로 전통 의료사업 모델이 프로세스별로 분화, 연계됨으로써 맞춤 의료 기기(Point of Care), 원격의료·약국, 정밀의료·진단, 예방을 위한 건강관리 등의 새로운 니치가 생성되고 있다. 정밀의료·진단 분야에서는 특히 제3차 병원 중심으로 외국계 SW기업, 국내 SW 및 통신기업, 스타트업의 인공지능 활용 의료 서비스를 도입하고 있다. 한편, 데이터와 관련해서는 관계부처, 공공기관, 유관 학계, 의료계, 전문가, 창업가를 아우르는 보건의료빅데이터 추진단을 통해 2017년부터 다양한 데이터를 표준화, 연계, 비식별화하는 기술 과제를 논의하고 있다.

■ [그림 1] 데이터 기반 의료 서비스 프로세스 변화



※ 자료 : <https://blogs.sap.com/2016/06/30/the-digital-transformation-of-healthcare-part-2/>

### ■ 정밀의료/진단 산업 분야의 환경적 변화와 사례기업의 대응

현재 의료 산업에서 진단 행위는 의료 기기로부터 생성된 의료영상, 생체 신호 등의 데이터를 기반으로 의학적 지식과 임상 경험에 근거한 의료진의 판단으로 이루어지고 있다. 이러한 의학적 판단의 대부분은 의료진의 지식과 경험, 임상 환경에 의존적이고, 따라서 의료진 간의 판단 일치도 뿐 아니라, 동일 의료진의 판단에 대한 일관성이 높지 않은 문제가 있다. 또한 의료 기기의 보급 및 검사 건수의 증가에 비해 의료진의 수가 부족하여 갈수록 의료진의 판독 부담이 가중되고 있으며, 따라서 의료 서비스의 표준적인 품질 관리에 대한 대책 마련이 필요한 상황이다. 한편, 각 진단의 영역에서 숙련도가 높은 전문의 수는 수요에 비해서 늘 모자라며, 공간적으로는 지방이나 섬 등의 외지, 시간적으로는 심야나 주말 등의 수요에 대해서 대응을 하지 못하는 한계가 있다.

정리하자면, 기존의 의료가 개별 증례에 기초한 근거 중심 의학이었다면, 통합된 대량의 데이터에 기초한 데이터 기반의 의료로 변화해 가고 있다. 사례기업은 진단을 보조하는 의료 인공지능 솔루션을 개발하고 있고, 대량의 판독 데이터를 인공지능 모델에 학습시켜 의료진의 일관되고 정확한 의학적 판단을 지원하며, 소프트웨어 의료 기기로서 시간적 공간적 제약이 없이 의료 진단 서비스를 제공하고 고품질 의료 서비스의 접근성을 향상시킨다. 결국, 대량의 의료 데이터에서 가치를 창출할 수 있는 검진, 진단, 치료 등의 여러 영역에서 의사결정 보조 솔루션 개발 및 공급자로 자리하고 있는 것이다.

사례기업의 기술적 역량은 기술 기회가 높고 결합적인 지식기반의 알고리즘을 활용하여, 기술·도메인 전문성을 갖고 데이터를 확보한 후, 모방이 어렵고 복잡도가 높은 상세 모델링에 전문화된 의료 영상 인공지능 기술이 핵심적이라 할 수 있다. 한편, 사업 성장을 위한 전략으로 전문화된 니치로서 향후 플랫폼 기반 시스템 경쟁을 하고자 하는 구체적인 목표를 세우고 있다.

*“초기 시장의 확보를 위해서는 각 의료 단계나 영역별 솔루션으로 전문화하여 진입하였으나, 추후에는 각 단계별 데이터 및 의사결정 체계를 통합하는 지능형 플랫폼으로 진화하여 의료 서비스의 품질과 비용을 최적화하는 운영 플랫폼으로 사업 영역을 넓히는 전략을 구사하고 있다.” - A사 CTO*

## ■ 시도입 관련 이슈와 사례 기업의 대응

의료 영역에서 시도입으로 인한 주요 산업적 이슈는 시솔루션 사용에 대한 지불의 주체와 적정한 서비스의 가격 결정 문제이다. 현재의 의료 시는 의료진을 대체하는 것이 아닌 의료진의 판단을 보조하는 기기로서 위치하고 있으며, 따라서 시활용에 대한 비용 효과성이나 진단 및 치료결과에 기여도 등을 평가하기 어려운 부분이 있다. 그에 따라 수가의 형태로 공공 재정을 이용한 지원에 대한 결정이 어려우며, 이에 따라 병원은 시솔루션 도입을 위해 자체적인 추가 비용을 부담해야 하므로, 의료 시솔루션의 확산 속도가 상대적으로 느리다. 이에, 사례기업은 다양한 자체적으로 수행한 임상 검증 결과 및 문헌적 근거와 선도적으로 도입한 병원의 사례를 중심으로 고객을 설득하고 있고, 병원이 아닌 의료기기 회사와 협업하여 사업화를 진행하고 있다. 또한 궁극적으로 의료 시솔루션이 의료의 질과 효율성을 높임을 입증할 수 있는 근거 마련에 노력하고 있으며, 이에 따라 병원의 전산화, 디지털화에 대한 과거 PACS 수가 보조 사례와 같이, 병원의 지능화에 대한 수가를 보조함으로써 시솔루션 도입을 가속화하기 위한 제도적 혁신을 꾀하기 위해 다양한 협의체, 이해관계자와 협력과 연구를 지속하고 있다.

## ■ 사업 활동과 확장에 관한 애로사항

의료 분야는 대표적인 규제 산업으로서, 기술개발 이외에도 여러 이해관계자와 상충되는 이슈들을 해결하기 위한 많은 노력이 필요하며, 따라서 수익성과 지속성에 대한 불확실성이 높은 산업 분야이다. 특히 보수적인 의료분야에서 시를 임상에 도입하기 위해서는 충분한 검증과 근거가 필요한데, 이러한 데이터를 쌓기 위해서는 역설적으로 많은 임상 현장에 도입되어 활용되어야 한다는 모순이 존재한다. 일부 검증을 위한 정부 과제나 선도적인 병원의 조기 도입으로 인해 근거가 창출되기 시작하고 있으나 여전히 확산의 속도가 더딘 상황이다. 또한 시 도입으로 인한 기존 임상 프로세스의 변화에 대한 저항감과 시에 대한 불신, 혹은 시로 인한 경쟁력 상실에 대한 우려 등, 의료진의 이해도와 신뢰를 제고하는 부분 역시 현실적인 부담으로 작용하고 있다.

끝으로, 의료 시 솔루션 개발을 위해서는 대량의 의료 데이터가 필요한데, 이를 획득하고 처리하기 위한 비용이 높고 절차가 복잡하다. 가장 큰 법적 혹은 제도적 리스크는 시 모델을 개발하고 검증하기 위한 과정에서 필요한 환자 데이터에 대한 접근 및 활용에 관한 불확실한 규제들이며, 특히 데이터 익명화 및 가명화에 관한 정의, 방법과 절차 등에 관한 구체적인 규정들이 미비한 점이다. 따라서 이러한 리스크를 최소화하기 위해 병원들은 환자 데이터 접근을 위해 점차 보다 복잡하고 엄격한 절차를 적용하고 있어 연구개발의 속도를 저해하는 원인이 되고 있으며, 데이터 3법 통과에도 불구하고 환자 데이터의 상업적 연구 활용에 관해서는 시민단체와 정치권에서 참여한 논쟁의 주제가 되고 있어 규제적 불확실성이 상존하고 있다. 또한, 인공지능 의료기기로서 식약처의 승인을 받기 위한 임상시험 진행 및 인허가 획득에 대한 투자비용이 상대적으로 매우 높은 데 반해, 사업화와 수익 발생에는 많은 시간이 소요되는 점도 통제하기 어려운 요소이다.

### 3.3 금융

#### ● 저금리 기조에 따른 휴먼 리스크·비용 절감 트렌드와 가치사슬 변화

저성장, 코로나 팬데믹 상황 등으로 국제적인 저금리 기조가 지속되고 있다. 금융 분야에서는 투자 상품 및 운용에 대한 불신과 우려가 높아지고 있으며, 자산관리 분야의 중요성도 부각되고 있는 상황이다. 휴리스틱 방식에 의존한 액티브 펀드 운용 및 주문 집행의 한계성이 점차 대두되고 있고, 패시브, 시퀀트로 이동하는 패러다임의 전환이 가속화되고 있다. 골드만삭스, 도이치뱅크 등 초대형 금융기관의 대규모 구조조정 및 관련 국내 사업부 철수(주식 세일즈, 트레이딩 부서 등)는 휴먼 리스크 및 비용 절감 추세를 보여주는 대표적인 사례이다.

이러한 트렌드 속에서 인터넷 은행, AI트레이딩 및 자산운용, 지급결제, 송금 등의 틈새 분야와 관련한 혁신적 사업자의 등장으로 전통 금융업 모델이 분화됨으로써 은행이 규모·범위의 경계를 달성한 기반이 약해지고, 고객 정보에 대한 접근성이 약해지면서 예대 마진, 지급결제 수수료 등 전통 수익원이 축소될 가능성이 높아졌다. 또한 탈집중화, 탈중개화를 통해 기존 사업자의 수익성이 약화되고 소비자 입장의 서비스 품질과 접근성은 높아짐으로써 기존 금융시스템의 비율율이 개선되고 자본 배분의 효율성이 높아질 것이라 전망된다.



※ 한국은행, WEF(2015) 참조하여 저자 작성

또한, 테크핀 기업들의 금융 산업 가치사슬 해체를 통한 사업 확장 기세가 거세다. 금융 산업은 국가 차원에서 경제적 안정성을 뒷받침하는 중요한 산업이라는 이유로 규제와 보호를 받아왔기 때문에, 특별히 마케팅을 할 유인도 없었고, 핵심 분야가 아닌 비대면 서비스 등에 필요한 기술은 아웃소싱으로 해결해 왔다. 한편 테크핀 기업들은 무료 서비스와 광고 수익으로 대변되는 플랫폼

서비스에 능하고, 이를 뒷받침할 수 있는 기술적 역량이 뛰어나다. 중국, 미국, 그리고 한국의 테크핀 기업들의 분화된 금융 산업 진출상황을 간략히 비교해 보면 아래와 같다. 기존에는 금융 기업을 통해 종합적으로 이루어졌던 일들이 테크기업들의 진출로 인해 분화되고 전문화되면서, 소비자들은 기업이 아니라 필요한 서비스만을 접하면 되는 구조로 바뀌어 가고 있는 것이다.

■ [표 1] 글로벌 테크핀 기업들의 금융산업 분야별 진출현황

국가	기업	지불결제	예금	대출	투자중개	보험
한국	삼성	삼성페이	-	-	-	-
	카카오	카카오페이	카카오뱅크	-	카카오주식	카카오페이 간편보험
미국	구글	구글페이	구글페이 연동계좌 (시티그룹 제휴)	스탠포드 연방신용조합 신용대출	-	구글컴퍼어 자동차 보험
	아마존	아마존페이	아마존캐시	아마존렌딩	-	아마존 헤븐 건강보험 (JP모건 제휴)
	페이스북	메신저페이	-	클리어뱅크와 소기업 대출 시장 협업	-	-
	애플	애플페이	예금을 위한 애플뱅크	애플 연방 신용 조합 (북버지니아 커뮤니티 신용 조합)	-	애플라이프
중국	알리바바	알리페이	마이뱅크	• 알리바바 파이낸셜 타오바오/텐마오 판매자 대상 신용대출	• 중국국제금융공사 지분투자 • 위어바오 • 자오차이바오 • 안토포춘	중안온라인보험 기업(19.9%)
	바이두	바이두지갑	바이신뱅크 (시틱은행 제휴)		-	바이두보험 (알리안츠 제휴)
	빈센트	텐페이	위뱅크		• 중국국제금융공사 지분투자 • 리차이통	중안온라인보험 기업(15%)

※ 투이컨설팅(2019), 해외 언론 기사 수집을 통해 저자 작성

■ 금융 산업 환경 변화에 대한 사례 기업의 대응

사례 기업은 B2B 사업모델을 가지고 전문화된 영역에 니치로서 진입하여 기존 금융 기업에 자산관리 및 투자 중개 서비스를 제공하고 있다. 사례 기업이 개발한 SI기반의 자산 배분 모델은 로보 어드바이저 형태로 현재 각 은행에 서비스되고 있으며, 고객들은 이를 통해 데이터 기반의 신뢰할 수 있고 검증된 포트폴리오를 구성하고 있다. 또한, SI주문 집행 시스템의 경우 주식 시장의 데이터를 매

시각 정밀하게 분석하여 정해진 기간 동안 가장 낮은 가격에 해당 주식을 매수하는 AI 구매 시스템으로 기존 인력보다 저비용으로 높은 효율을 낸다. 또한, 계량적 투자모형을 AI 기반으로 향상시킨 ETF 3종을 뉴욕증권거래소에 직접 상장하기도 했다. 고비용, 모럴 해저드와 같은 금융권 휴먼 리스크를 주도적으로 해결하고, 노동 집약적인 금융 산업을 시를 통해 질적으로 개선시키면서 상품 투명화 및 안정성 증대, 업무 효율화 등을 이뤄내고 있다.



사례기업은 전 세계를 아우르는 금융 데이터의 전처리 및 금융 AI모델링에 독보적인 기술력을 보유하고 있다. 특히, 상장 폐지 종목을 포함한 글로벌 금융 원천데이터 및 거시경제 데이터를 자체 개발한 API를 통해 AI모델의 학습에 활용하고 있으며, AI학습 모델링에서 가장 중요한 과적합(Overfitting) 문제<sup>2</sup>를 최소화하기 위한 독보적인 기술적 노하우 및 모델들을 보유하고 있다. 해당 기술력을 바탕으로 독자 개발한 AI ETF 3종은 뉴욕증권거래소(NYSE)에 2019년부터 상장하여 지금까지 높은 성과를 내며 운용 중이다. 아래 그림은 사례기업의 ETF가 타 ETF 대비 독보적인 성과를 내고 있음을 보여주는 그래프이다. 2020년 5월 기준 상장코드 QRFT의 1년 수익률은 총 수익기준 15.52%를 기록했다. 같은 기간 4.08% 상승한 S&P 500 지수 대비 11.44%p 초과 수익을 거둔 셈이다. 상장 코드 AMOM 역시 16.19%의 1년 수익률을 보였다. 이 인공지능 ETF들은 인간의 개입 없이 100% 딥러닝 기반으로 운용된다.

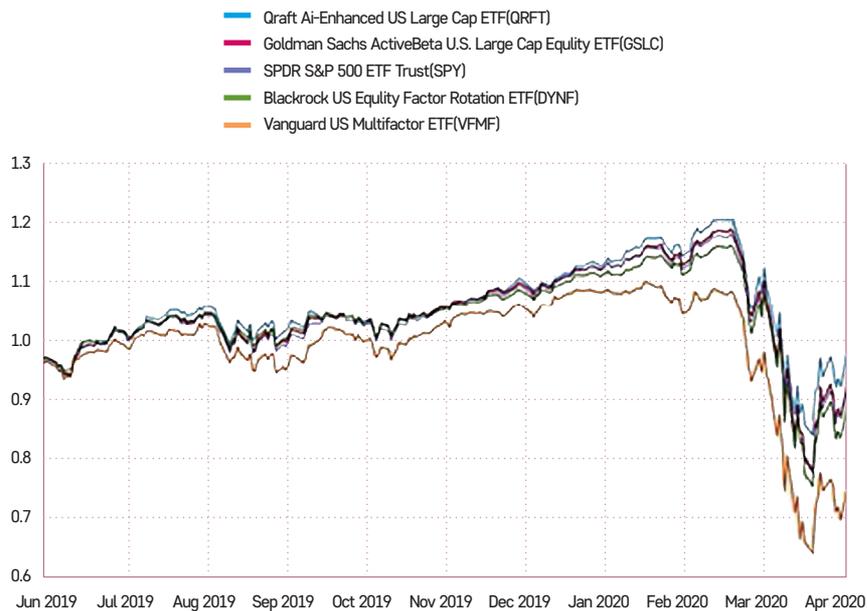
2 <과적합(Overfitting) 문제>

과도하게 훈련데이터에 대해 모델을 학습하여, 새로운 데이터에 대한 예측 정확도가 떨어지는 현상으로, 모델의 복잡도가 너무 높거나 낮으면 부정확한 결과를 얻게 된다. 이를 해결하기 위한 적당한 복잡도를 찾아가는 과정을 정규화(Regularization)라 한다.



\* [https://docs.aws.amazon.com/ko\\_kr/machine-learning/latest/dg/model-fit-underfitting-vs-overfitting.html](https://docs.aws.amazon.com/ko_kr/machine-learning/latest/dg/model-fit-underfitting-vs-overfitting.html)

■ [그림 2] Total Return



※ S&P Global Compustat

### ■ 시도입 관련 이슈와 사례 기업의 대응

시도입으로 가장 우려되는 사회적 이슈는 종사 인력의 감축이다. 사례 기업이 개발한 시상품 및 서비스는 많은 인력을 필요로 하지 않고, 전자동 운영 시스템을 도입시킬 수 있어 산업 내 해당 분야의 인력을 상당수 대체 가능한 것으로 판단된다. 시의 노동력 대체 문제는 오랜 기간 동안 제기되어온 이슈이며, 이해 관계자들의 협력을 통해 해결해 나가야 할 문제이다. 기존 인력은 단순 반복 업무의 빠른 처리 등 노동집약적인 최적화 달성이 목표가 아닌, 보다 창의적이고 새로운 개념이나 비즈니스 모델을 설계·기획하는 업무를 할 수 있도록 지원해야 한다. 업무의 재분배와 인력의 재배치가 필요한 것이다.

### ■ 사업 활동과 확장에 관한 애로사항

금융업은 강력한 규제 산업 중 하나이며 전통적인 인가 기반 산업이다. 특히 국내의 경우 무인가 및 미등록 업체는 금융투자업을 할 수 없다. 즉, 「자본시장과 금융투자업에 관한 법률」에 따른 금융투자업 인가를 받지 않고는 금융투자업이 불가하며, 투자자문업과 투자일임업 또한 등록된 기업만 사업을 할 수 있다. 특히 사례기업과 같은 핀테크 업체가 국내에서 금융 상품을 자유롭게 개발하고 판매하려면 해당 인가 및 등록이 필요한 상황이고, 인가 및 등록을 위해서는 상당한 자본금이 필요하다. 현재도

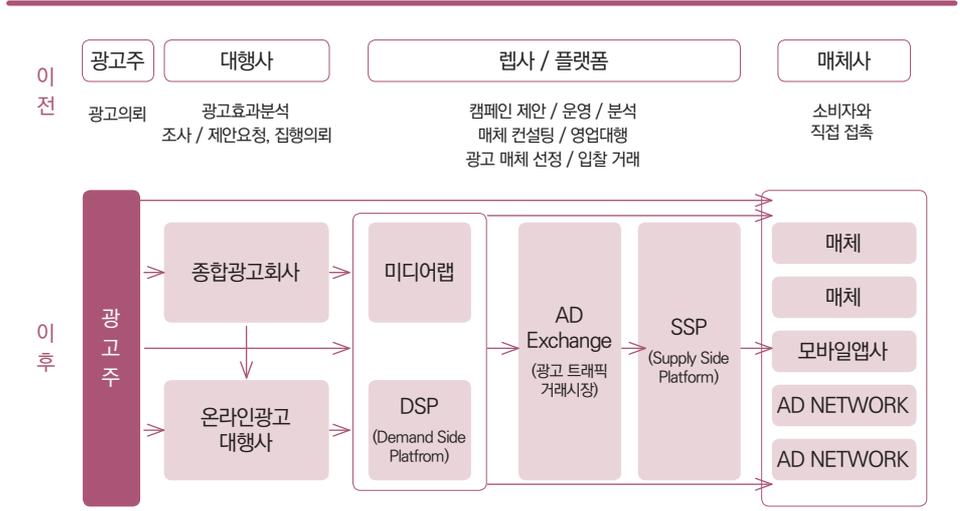
금융 규제 샌드박스와 지정대리인 제도 등 금융당국의 규제변화 움직임이 활발히 이뤄지고 있지만, 향후 핀테크 업체 및 산업이 성숙될 수 있도록 다양한 금융 규제들이 완화 및 개방될 수 있어야 할 것이다.

### 3.4 마케팅/광고

#### ● 모바일, 플랫폼 및 비정형 콘텐츠 확산으로 인한 마케팅 가치사슬 변화

채널별 광고 시장은 오프라인 전통 매체, 온라인 매체로 나뉘는데, 온라인 매체의 경우 배너, 검색, 모바일, 프로그래매틱의 형태로 발전해 왔다. 이전에는 광고주, 대행사, 렙사/플랫폼을 거쳐 매체사로 선형적으로 이동하는 가치사슬이었다면, 이제는 분화된 플랫폼들을 단계별로 필요에 따라 이용하거나, 혹은 광고주가 직접 매체사에 접하는 경우도 많아졌다.

■ [그림 3] 광고 가치사슬 분화 양상



※ 사례기업 이사 인터뷰 토대로 저자 작성

국내는 인지도 자체는 글로벌 수준이라 할 수 있으나 그 활용이나 기술력이 아직 미성숙한 단계이다. 전 세계적으로 전통적인 광고 시장 생태계 참여자인 종합광고회사, 온라인광고대행사의 역할이 축소되거나 사라지고 있으며, 플랫폼 영역과 매체가 분리되고, 모바일 광고 분야는 거의 프로그래매틱 광고로 이동한 상태이다. 생태계 참여자 간 전방위적 확장이 일어나고 있으며, 브로커들의 수수료 구조 및 명목이 변화하여 실제 광고대행사의 수수료가 축소되는 등, 중간자를

건너뛰는 현상이 발생하고 있다. 그러나 한국의 경우 모바일 광고의 대부분이 프로그래매틱으로 이동했다고 하기에는, 네이버와 같은 대형 앱의 직거래 광고가 여전히 존재하고 거래액도 상당히 큰 편이다. 프로그래매틱 광고를 페이스북, 구글의 광고를 포함하는 개념으로 설정한다면, 모바일 광고 시장의 상당수를 포함한다고 이야기할 수도 있을 것이다.

## ■ 마케팅/광고 산업 환경 변화에 대한 사례 기업의 대응

사례 기업은 모바일 앱 전문 DSP(Demand-Side Platform)로서, 프로그래매틱 광고 및 실시간 거래 영역에 속하고 있어, 전통적인 광고 집행 형태, 즉 광고주/대행사가 직접 매체를 대응하던 방식에서 이를 자동으로 집행하는 방식으로 바꾸는 데 일조하고 있다. 무엇보다 앱 광고 지면과 앱 마케팅은 페이스북과 구글과 같은 빅미디어 광고에 집중되어 있거나 혹은 애드 네트워크 중심으로 개발되어 있었는데, 앱 영역에서 특화시킨 DSP로서 독보적인 니치로서 자리매김한 것이다. 경쟁사 대비 답테크 기업으로서의 경쟁우위는 세 가지로 요약할 수 있는데, 머신러닝 기반의 광고 대상 타겟팅과 최적화, 글로벌 광고 집행 플랫폼이다. 기존의 모바일 광고 마케팅은 지면 중심의 타겟팅 방식에 머물러 있었다. 즉, 특정 앱에 자사 서비스를 선호할 고객들이 많을 것으로 예상되면, 해당 앱의 광고 지면에 광고를 내보내는 형태였던 것이다. 이는 직접 지면을 사는 광고 형태뿐만 아니라 프로그래매틱 광고에서도 해당되는 방식이다.

프로그래매틱 광고에서는 애드 네트워크에 속한 특정 지면들을 광고의 속성에 따라 선별하고 해당 지면에 각 지면별 적합한 가격으로 광고가 나가도록 하는 경우가 대부분이었다. 사례 기업은 광고주의 자사 데이터를 이용한 모델을 구축, 매 광고 요청에 대해 반응 확률을 계산할 수 있도록 하였다. 초당 최대 140만 건에 이르는 광고 요청에 대해 실시간으로 대응하여, 각 광고에 반응할 확률이 높은 지면과 광고 대상을 찾아냄으로써 광고 성과를 높일 수 있었다. 이는 제3의 기업 데이터를 이용하여 미리 광고를 내보낼 광고 대상을 선정한 후, 해당 대상에게 정해진 가격의 광고를 집행하도록 하는 기존의 오디언스 타겟팅 보다도 진일보한 형태로, 광고 대상, 집행 지면 및 기타 여러 요소들을 함께 고려하여 가격을 자동으로 책정하고 내보내도록 하고 있다. 결국, 효율적 광고 집행에 영향을 주는 요소들이 실시간 반영된 타겟팅 방식을 선도하고 있는 것이다.

또한, 기존의 국내 프로그래매틱 광고가 한국이라는 지역권 내 광고에 머물러 있었다면, 사례기업은 광대한 글로벌 트래픽과 데이터를 바탕으로 글로벌 광고를 쉽게 집행할 수 있도록 하는 장을 열었다. 기존의 지역 및 국가 기반의 광고시장에서는 타국에 광고를 하기 위해서 타국의 로컬 에이전시를 이용하거나 혹은 해당 국가에 광고를 내보낼 수 있는 애드 네트워크를 찾는 탐색 비용을 들여야만 했다. 사례 기업은 전 세계 다수 국가의 광고 트래픽을 다량 확보하여 광고주의 위치와 무관하게 글로벌 광고를 집행할 수 있도록 하였다. 이러한 발전은 사내 마케팅팀이 글로벌 광고 집행을 모두 컨트롤 할 수 있도록 하고, 로컬 에이전시를 활용하기 어려운 소액으로도 타국에 광고를 집행할 수 있으며,

또한 동시에 여러 국가에 광고 집행 가능하게 하는 등 중소기업 앱 서비스의 글로벌 진출을 도와 동반성장하고 있다. 현재 사례기업은 글로벌하게 30개 이상의 애드 익스체인지/네트워크 등 매체유통 및 거래 네트워크와, 250여 개국으로부터 하루 광고 요청 800억 건을 받고 있는 상태다. 또한 사례기업의 머신러닝 모델은, 일회성으로 모델을 생성하여 자동집행 조건을 만들어내는 것이 아니라, 자동으로 여러 가지 광고 요소들을 조정하며 광고 최적화를 한다. 이는 광고 집행자의 큰 개입 없이도 시간이 지남에 따라 광고 효율을 높이도록 하는 효과를 갖게 한다. 하나하나의 콘텐츠 자산(예 : 광고 소재)들의 소진량과 입찰 가격을 조절하는 수고를 덜어내고 자동화시킴으로써, 광고 집행 시 들어가는 집행 인력 자원 소요를 줄였다.

### ■ AI혁신 특성별 사례기업의 혁신경로

인프라, 모델, 도메인 분야를 중심으로 사례기업의 혁신특성과 경로를 살펴보면, 먼저 인프라 측면에서 프로그래매틱 광고는 대용량 트래픽을 받고 처리할 수 있는 인프라를 필요로 한다. 특히 글로벌 DSP라면 그 처리 용량이 막대하고 광고집행을 위한 트래픽도 막대하여 이를 비용 효율적으로 처리할 수 있는 인프라를 구축하는 것이 사업성에 큰 영향을 끼친다. 무엇보다, 안정적으로 광고를 집행하기 위해서는 시간대별로 크게 차이나는 트래픽을 유동적으로 처리할 수 있어야 하며, 대량의 트래픽이 몰리더라도 광고의 집행의 장애가 없어야 한다. 사례기업은 클라우드 서버를 이용하여 비용 효율적이고 안정적인 인프라를 개발하였으며 이를 통해 타사 대비 높은 기술 장벽을 구축할 수 있었다. 모델 측면에서는 합법적인 데이터를 확보하고 광고 효율을 높일 수 있는 우수한 입찰 모델을 확보하는 것이 해당 분야에서의 엣지가 되고 있다. 제3기업의 데이터, 혹은 그들이 확보한 플랫폼(예 : 유튜브, 페이스북 등)을 통해 얻는 정보들을 적극 활용하는 타 광고플랫폼과 달리, 사례기업은 캠페인의 주체인 광고주의 '광고주 자체 데이터'만을 이용하면서도 효율적인 광고를 집행하고 있다. 사례기업은 높은 수준의 딥러닝 모델을 구축하고 신뢰도가 높은 데이터만을 활용함으로써, 개인정보 활용 등의 데이터 이슈에서 자유로우면서도 높은 퍼포먼스를 내고 있다. 마지막으로 도메인 측면에서 사례기업이 집중하는 분야는 모바일 앱 마케팅 부문이다. 높은 퍼포먼스를 요구하는 마케팅이 발달되어 있고, 특히 실시간 입찰로 광고주의 기대 수준을 만족시키기 어려운 분야에 독보적인 기술력을 바탕으로 진출함으로써 새로운 광고 형태·방식을 개척하고 선도하고 있으며, 이로 인해 해당 분야에서는 글로벌 최상층에 속하는 DSP로 발전하였다.

### ■ 시도입으로 인한 산업적 이슈와 역기능

광고 사기 관련 세계적 이슈가 있었으며 이러한 광고 사기는 지속적으로 발생하고 있다. 광고 효율 등에 따라 자동으로 움직이는 프로그래매틱 광고의 속성을 파고들어 퍼블리셔가 가짜 클릭과 같은 거짓 신호를 발생시키는 경우도 있으며, DSP자체에서 광고 효율을 높이기 위해 거짓 신호를 발생

사기는 경우도 발생하고 있다. 특히 Affiliate Network라고 불리는 형태의 다단계 광고 네트워크라든가, 혹은 퍼포먼스 광고(CPI, CPA, CPS Network으로 불리는) 노출이 아닌 이후 전환이 있을 때에만 광고비를 지급하는 형태의 광고에서 광고 사기는 더 활발하게 이루어졌다. 악성 광고 회사나 퍼블리셔의 경우, 봇 등을 이용하여 마치 실제 유저가 앱을 사용하는 것과 유사한 패턴 거짓 신호를 발생시킴으로써 광고주를 속이기도 한다. 광고주는 이러한 광고 사기의 피해를 최소화하기 위하여 광고비 미지급 조건에 대한 가이드라인을 만들어 매체들과 협의하기도 한다.

사례 기업은 이러한 광고 사기 기업과 차별화하고 높은 신뢰도를 확보하기 위해 지속적으로 퍼블리셔들의 행태를 모니터링하여 거짓 신호 등을 생성하는 저품질 퍼블리셔를 차단하거나, 혹은 모델의 알고리즘을 이용하여 저품질 퍼블리셔를 배제시키고 추가적인 모니터링을 통해 품질 관리를 하고 있다. 또한 다방면으로 분석한 자료들을 시장에 제공하고, 애드텍 분야의 각종 회사와 연합하여 광고 사기를 퇴출시키기 위한 시장 교육과 고발에 앞장서고 있다. 이의 일환으로, 글로벌 마케팅 파트너사들이 연합한 CAAF(Coalition Against Ad Fraud)에 참여하고도 있다. 또한, 광고 집행 데이터를 투명하게 공개함으로써 자신들의 신뢰도를 높일 뿐만 아니라 타 광고매체들의 투명성과 깨끗한 운영을 유도하고 있다. 민간에서는 모바일 앱 데이터를 다양한 회사들이 이에 대해 분석하고 믿을 만한 매체를 정기적으로 발표하기도 하는데, 이 중 Kochava Traffic Index라는 모바일 어트리뷰션 분석 기업에서 해당 사례기업을 탐 매체로 선정하였다. 또한, 전세계 모바일 어트리뷰션을 선도하는 앱스플라이어가 발표한 앱스플라이어 퍼포먼스 인덱스에서도 2018년부터 줄곧 상위권에 이름을 올렸다.

### ■ 시사점 추진시의 애로사항과 법제도적 리스크

광고 시장은 대부분 자율규제에 머물러 있다. 그간 애드 ID는 개인을 특정하지 않기에 개인정보로 취급되고 있지는 않은데, GDPR<sup>3</sup>은 물론이고, 국내에서도 애드 ID를 타깃팅하는 방식이 불법으로 되어 있지는 않다. 하지만 개인정보보호 이슈에 매우 민감한 금융권이나 대기업의 경우 개인정보 보호법을 포괄적이고 보수적으로 해석하여 아주 단순한 모바일 광고에 있어서도 애드 ID의 활용을 꺼리고 있어, 광고 성과 추적을 위한 모바일 마케팅 파트너를 활용하지 못하는 경우가 많으며, 이로 인해 다양한 사업 기회를 놓치고 있다. 이처럼 합법적으로 얻고 활용할 수 있는 다량의 데이터들을 이용하는 것에 있어 제약을 받는 경우가 아직도 많다.

<sup>3</sup> General Data Protection Regulation. 유럽 연합 일반 데이터 보호 규칙은 유럽연합의 법으로써 유럽연합에 속해 있거나 유럽경제지역에 속해있는 모든 인구들의 사생활 보호와 개인정보를 보호해주는 규제이다. 이 법은 유럽연합과 유럽경제지역 이외 지역의 개인정보의 침해 또한 적용이 가능하다.

#### 4 새로운 혁신의 경로와 시사점

AI혁신체제하에서의 경쟁은 인프라, AI알고리즘, 데이터 및 모델링, 도메인 분야를 모두 포괄하는 시스템 경쟁이고, 각 기술 요소에 대해 국가별로 축적한 역량의 수준이 다르기 때문에 다양한 조합의 다면적 시스템 추적이 가능하다. 선발자로서는 초기 투자 비용이 막대한 인프라와 AI알고리즘 분야를 선도 개발하고, 적절한 수준에서 공유하는 공개 플랫폼 생태계를 구축하면서 세계적인 입지를 다지면서 모델, 도메인 분야의 시장 경쟁을 유도하는 것이 유리한 전략이지만, 후발자는 상대적으로 노동 집약적인 모델링 분야에서 입지를 다지고, 이미 구축된 생태계에 편승하거나 모방하는 전략이 일반적인 생존 전략이다. 하지만, 이를 넘어 경로창출형 생태계를 구축하기 위해서는 지역적 특성이 강한 자연어처리나 안면인식과 같은 특정 기술 분야를 세분화하여 정부 개입과 같은 기회의 창을 최대한 활용함으로써, 단계적인 내수 시장에서의 데이터 확보 및 독점, 블록화를 거쳐 독자적인 공개 생태계를 얻을 수도 있다.

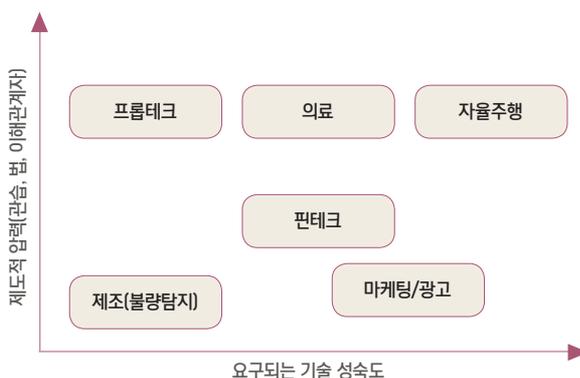
한편, 수직 산업과의 융합이 추가 되는 도메인 분야는 산업의 기존 경쟁력에 기반하여 생산성이나 기능이 증강되는 특성을 가지고 있다. 그중에서도 신수요가 창출되는 의료, 금융, 마케팅/광고와 같은 분야는 초기 시장 확보가 중요하므로 정부의 적절한 개입이 필요하다고 볼 수 있다. 이 중에서도 의료와 금융 분야는 보수적인 전문 서비스 분야로서 정부의 강력한 규제 및 보호를 받고 있다. 이러한 분야는 기술적인 장애 요소 보다는 제도적 이슈가 크다. 이러한 분야에서 AI기반 혁신이 창출되고 다양한 신생 기업이 출현하며 산업의 역동성이 제고되기 위해서는 먼저 가치사슬의 해체(Unbundling)가 선행되어야 한다. 가치사슬의 분화를 통해 세분화된 시장에 신생 기업이 전문화된 니치로서 진입할 수 있는 길이 열려 있으며, 그 대표적인 사례를 3장에서 다루었다.

이 사례들은 후발국 기업으로써의 혁신 경로 개척 사례를 보여주는데, 대부분 모델링 분야에서 특화된 노하우를 확보하고 있는 딥테크 기업들이다. 한편, 글로벌 대표 선도 스타트업으로 자리매김 하여 세계시장을 개척한 사례들은 데이터 특성상 인프라 분야의 독보적인 활용·구성 및 유지관리 역량 구축이 필수였다. 데이터는 도메인 특성에 따라 그 유형과 특성이 다르고, 이것이 혁신경로에도 영향을 미친다. 결국, 정보의 분할과 데이터 특성의 재조합 역시 혁신 특성을 좌우하는 중요한 요소라는 것이다.

기업이 머신러닝과 관련한 초기 기술역량을 가지고 도메인 산업을 선정하는 데 있어, 16개월에 2배씩 계산량 기준, 알고리즘의 효율이 높아졌다고는 하지만(OpenAI, 2020) 아직은 태동기에 있다고 할 수 있는 기술이기에, 적용할 수 있는 분야가 한정적이고 수익모델을 만들기 어렵다는 도전적인 과제를 맞닥뜨리게 된다. 어떤 분야에 응용해야 실현 가능한 상업적 가치를 창출해 낼 수 있을 것인가? 이 보고서는 이를 개념화하여 의사결정을 보조하는 도구로 활용할 수 있는 생각의 틀을 제시해 본다. 관습이나 법제도, 이해관계자 간 조정을 포함하는 제도적 압력의 축과, 요구되는 기술 성숙도 혹은

신뢰도의 축으로 구분하여 살펴보면, 기업이 보유한 역량 대비 관련 도메인의 초기 진입 가능성을 가능할 수 있다. 특히, 그간 깊이 고려하지 않았던 제도적 압력에 대응할 수 있는 역량이 신생 분야(Nascent Sector)에 출현한 신생 기업에게는 특히 중요한 고려 대상임을 강조하고자 한다. 예를 들어 투명 교정시장을 개척한 디지털 치아교정치료 시스템을 제공하는 D사의 경우, 치료 전에 투명 교정의 전·후를 머신러닝 기반으로 모사하여 기대효과를 가시화하여 보여주는 사이트<sup>4</sup>를 의료 관련법 규제로 인해 현재 기준 국내에서만 출시하지 못한 상태다.

■ [그림 4] 초기 진입 도메인 선정을 위한 틀의 예시



다시 정리하면, 인공지능은 기존 도메인 산업의 가치사슬을 분화 및 해체(Unbundling)하고 이를 재구성(Constellation)하여 단위 프로세스만으로도 다양한 가치를 창출할 수 있는 산업(Sector)을 형성해 나가고 있다. 인공지능의 혁신 특성은 인지과정을 자동화하고, 학습에 기초해 스스로 진화(Learning-Based Self-Evolution)한다는 것으로 요약할 수 있지만, 머신러닝의 경우 데이터에서 시작한 정량도구이기 때문에 데이터에 의해 한계, 혹은 활용 범위의 경계가 결정되며,<sup>5</sup> 아직 원천 기술은 초기 발전 단계에 있다. 한국과 같은 인공지능 기술 후발국에서도, 보유한 기술 역량에 맞는 도메인 산업을 잘 선별한다면, 전문화된 니치 영역으로 진입할 수 있는 다양한 기회는 앞으로 더 풍부해 질 것이다.

<sup>4</sup> <https://www.invisalign.com/get-started/smile-assessment>

<sup>5</sup> 머신러닝은 데이터에 기초한 정량적 학습을 토대로 하기 때문에 데이터로 인한 한계가 존재한다. 일반적으로 학습하는 기계는 문제를 스스로 설정할 수 없고, 정량적이고 대량 데이터가 있는 문제만 풀 수 있으며, 결과에 책임을 지지 않는다(야노 가즈오, 2019).

## 참고문헌

### 1. 국내문헌

김정곤, 이서진(2016), “주요국의 ICT융합 의료산업 전략 및 시사점”, 대외경제정책연구원  
 박형곤(2019), “AI의 사업적 적용 및 전개”, 『인더스트리 포커스』, 딜로이트 컨설팅  
 야노 가즈오(2019), “데이터의 보이지 않는 손”  
 최윤섭(2014), “헬스케어 이노베이션”  
 통계청(2019)  
 투이컨설팅(2019)  
 한국보건사회연구원(2017)  
 한국은행(2015)

### 2. 국외문헌

Bellman(1978), “An introduction to artificial intelligence : Can computers think?”  
 Brinjolffsson and Mitchell(2017), “What can machine learning do? Workforce implications”  
 Science, 358(6370), 1530-1534  
 James, Devaux and Sassi(2017), “Inclusive growth and health”  
 Lu et al(2014)  
 National eHealth Strategy Toolkit(2012)  
 OpenAI(2020)  
 Poole et al(1998), “Computational Intelligence”  
 Rao et al(2019)  
 Winston(1992), “Artificial Intelligence”  
 World Economic Forum(2015)

### 자문위원

분야	이름	소속	직함
인공지능 혁신특성	김진형	중앙대학교	석좌교수
	김명호	중앙대학교	조교수
	함명주	삼성리서치	수석연구원
	김휘영	연세대학교	연구조교수
사례연구, 구성검토	이창준	한양대학교	조교수

# AI기술의 국가통계 활용 사례 및 국내 도입 촉진 방안



## Executive Summary

국가통계는 사회통합 기능과 정책수립·평가기능을 갖추고 있는 국가의 중요한 공공재로서 통계법을 통해 엄격히 관리되고 있다. 특히 새로운 국가통계를 생산하기 위해서는 반드시 통계청의 통계작성 승인을 득해야 하므로 승인 기준과 품질 척도의 변화는 통계 생산자 모두에게 중요한 사안으로 볼 수 있다.

최근 국내외를 막론하고 AI 및 빅데이터 기술을 국가통계에 활용할 수 있는가에 관한 논의가 본격적으로 진행되는 추세다. 같은 맥락에서 실제 국가통계에 한시적으로 도입해 보거나 테스트하는 사례 또한 증가하고 있다. 이는 결국 AI 및 빅데이터 기술을 통계 생산에 활용할 시 국가통계로 승인 가능한지에 대한 고려로 확대된다. 국내 통계청 또한 올해 상반기 중 조사통계 심사에 맞추어져 있던 승인 심사 문항을 조사 통계 외 생산 방식에도 적용할 수 있도록 일부 개정하는 방안을 추진하고 있으며, 중장기적으로는 국가의 통계 관리 범위를 확대하기 위한 신규 제도 검토를 예고했다. 국가통계 승인 기준의 변화가 가시화되는 시점이다.

● ●  
김정민  
선임연구원  
jungmink26@spri.kr

● ●  
전이슬  
연구원  
jys@spri.kr

● ●  
황유림  
인턴연구원  
yurim0923@spri.kr

제도적 변화 조짐에도 불구하고 AI 및 빅데이터 기술은 그간 활용했던 통계적 기법과는 방법론의 구조, 평가 기준 등이 상이해 실제 도입에 대한 저항이 존재한다. 그런 연유에서 보고서는 해외의 앞선 논의와 실제 적용 사례를 소개하고 시기술이 통계 생산에 활용될 시 예상되는 기술적 쟁점이 무엇인지 구체적으로 분석해봄으로써, 도입 촉진에 필요한 요소를 발굴하고 이를 달성하기 위한 방안을 제시하였다.

예상 쟁점 도출을 위해 국가통계에 시기술을 도입하는 이슈를 두 가지 시각으로 분리하였다. 조사 통계에서 완전히 탈피해 빅데이터에 기반한 통계를 생산하는 ‘신규 통계 생산’의 시각과 조사통계 생산 과정을 그대로 수용하되 각 생산 과정의 효율성과 성능을 향상시키고자 시를 도입하는 ‘통계 생산 프로세스의 현대화’가 그것이다. 서로 다른 두 시각에서 바라본 예상 쟁점은 다음과 같이 요약된다.

구분	예상 쟁점	시사점
신규 통계 생산	AI·빅데이터 기반의 분석결과에 관한 신뢰가 보편적으로 형성되지 않음	신규 통계 생산 시 활용에 적합한 검증된 알고리즘을 공표하고 이에 대한 세부적인 활용 지침 마련이 요구
	신규 통계의 생산방식은 기존의 통계 생산 프레임워크에 맞추어 해석하기에 적합하지 않음	빅데이터 분석 프로세스를 포괄할 수 있는 통계 프로세스 표준의 개정 고려
	재현 불가능한 비결정론적(Non-deterministic) 알고리즘 기반의 결과를 국가통계로 관리 가능한지 여부	인공지능 기술의 특성 분석 및 표준화된 평가 기준 정립이 요구됨
통계 생산 프로세스의 현대화	기존의 방법 대비 우수성을 객관적으로 비교가능한가	전통적 통계 기법-AI알고리즘의 비교 검증 방안 발굴이 필요
	시기술 도입은 통계 생산부터 공표까지의 제한된 생산 기간을 준수할 수 있는 해법인가	다수의 실증 사업을 통한 도입 적합성 진단 필요

분석된 예상 쟁점들과 그에 따른 시사점을 토대로 AI 및 빅데이터 기술의 국가통계 도입이 촉진되기 위해 필요한 두 가지 요소를 도출하였다. 첫 번째로 시기술의 대표성 부여를 위한 표준화 된 기술 개발 및 품질 평가 기준 마련을 꼽을 수 있겠으며, 두 번째로는 실증 사업 확대를 통해 국가통계 유형별 도입 적합성을 지속적으로 테스트 해볼 수 있는 장이 마련되어야 한다는 점이다.

끝으로 국가통계에 AI 및 빅데이터 기술 도입을 촉진할 수 있는 실질적 방안을 제시하기 위해, UN유럽 경제위원회 「데이터 샌드박스」프로젝트의 일환으로 결성된 아일랜드 통계청-기술연구소 간 협력 사례를 참고하였다. 이를 통해 국내 통계청-AI표준수립기관 간 협력 체계 및 기관별 역할 기능을 제안하였다.

National statistics are important public goods of the country with functions of social integration and policy establishment·evaluation, and are strictly managed through statistical laws. In particular, in order to produce new national statistics, changes in approval criteria and quality measures can be seen as important for both statistical producers, as new national statistics must be approved by the National Statistical Office.

Recently, discussions on whether AI and Big data technologies can be used for national statistics, both at home and abroad, are underway in earnest. In the same vein, there are also increasing number of cases of temporary adoption or testing of actual national statistics. This eventually extends to the consideration of whether AI and Big data technologies can be approved as national statistics when utilizing them for statistical production. In the first half of this year, the National Statistical Office is also pushing to revise some of the approval screening questions that were tailored to the examination of survey statistics so that they can be applied to production methods other than survey statistics, and in the mid- to long-term, it announced a new system review to expand the scope of the nation's statistical management. This is when changes in the national statistical approval criteria are visible.

Despite signs of institutional change, AI and Big data technologies differ from the statistical techniques used so far, so there is resistance to actual adoption. For such a reason, the report introduced advanced discussions and actual application cases abroad and specifically analyzed what technological issues would be expected when AI technology is used for statistical production, thus exploring the necessary factors for promoting the adoption and presenting measures to achieve them.

The issue of adopting AI technology to national statistics has been divided into two perspectives to draw expected issues. They include the view of 'new statistical production' that produces statistics based on big data, completely deviating from survey statistics, and 'modernization of statistical production processes' that adopts AI to accommodate the process of producing survey statistics but to improve the efficiency and performance of each production process. The expected issues from two different perspectives are summarized as follows.

Categorization	Expected issues	Implications
New statistical production	Trust in analysis results based on AI and Big data is not universally formed	When producing new statistics, it is necessary to publish proven algorithms that are suitable for utilization and to provide detailed utilization guidelines.
	The production method of new statistics is not suitable to be interpreted in accordance with the existing statistical production framework	Consider amending statistical process standards that can encompass the Big data analysis process
	Whether results based on non-deterministic algorithms that cannot be reproducible are manageable as national statistics	Characteristic analysis of artificial intelligence technology and establishment of standardized evaluation criteria are required
Modernization of statistical production processes	Is it possible to objectively compare excellence compared to existing methods?	Discovering methods of comparison verification of Traditional statistical techniques-AI algorithms is required
	Is the adoption of AI technology a solution to comply with the limited production period from statistical production to publication?	Need to diagnose suitability of adoption through a number of demonstration projects

Based on the expected issues analyzed and their implications, two factors were derived to facilitate the adoption of national statistics for AI and Big data technology. First, standardized technology development and quality evaluation criteria for the representativeness of AI technology should be considered, and second, there should be a place to continuously test the suitability of adoption by type of national statistics by expanding demonstration business.

Finally, to present practical measures to promote the adoption of AI and Big data technologies to national statistics, we referred to examples of cooperation between the National Statistical Office and the Institute of Technology in Ireland, which was formed as part of the UN Economic Council's Big Data Sandbox project. Through this, the cooperative system and role functions of each institution were proposed between the Korean National Statistical Office and AI Standard Establishment Agency.

## 1 서론

- **국가통계(National Statistics)란, 국가통계 제도를 통해 배포되는 통계를 의미<sup>1</sup>**  
 ※ 비공식 통계임을 배포과정에 명시한 경우는 국가통계 범위에서 제외

  - 국가통계는 사회통합 기능과 정책수립·평가기능을 갖추고 있어<sup>2</sup> 국가에 매우 중요한 가치를 가지는 공공재임
  - 중요한 자산인 만큼 이에 상응해 요구되는 높은 수준의 품질 척도가 존재하며 정기적인 품질진단을 통해 관리되고 있는 상황
- **국내 통계법상 국가통계는, 통계작성지정기관에 의해 생산되는 통계작성 승인을 획득한 통계를 의미<sup>3</sup>**

  - 국가통계는 자료수집의 방법에 따라 조사통계(설문조사 기반), 보고통계(행정자료 기반), 가공통계(1차 통계 기반)으로 구분
  - 미승인 통계의 공표는 통계법에 저촉되므로, 국가통계로의 승인 여부는 새롭게 작성된 통계의 활용도와 존폐여부를 결정하는 주요한 사안임
- **최근 국내외적으로 AI 및 빅데이터 기술의 국가통계 활용이 논의됨에 따라 국내 국가통계 승인 관련 사항들에 변화의 기류가 감지**

  - 2020년 5월 통계청은 심사기준 개정, 시범통계제도 도입 등을 통해 빅데이터 활용 통계를 국가 관리체계 내 포괄하기 위한 실질적 토대 구축을 예고<sup>4</sup>
  - 해당 변화는 AI·빅데이터 등 SW기술과 통계 분야 간 융합을 촉진시키는 계기로 작용될 수 있어, 향후 중요성이 더 강조될 것이라 예상 가능함

그런 이유에서 본 보고서는 AI 및 빅데이터 기술의 국가통계 활용에 관한  
 해외 국가의 선제적 논의들과 도입 사례를 소개하고,  
 도입 시 예상되는 쟁점 및 도입 활성화를 위한 역할 체계를 제안하였음

1 OECD statistics 용어사전의 “official statistics” 참조

2 통계교육원(2015), “국가통계의 이해” 참고

3 통계법 제3장 제15조(통계작성지정기관의 지정), 제17조(지정통계의 지정 및 지정취소), 제4장 제1절 제18조(통계작성의 승인) 부분 참고

4 통계청(2020), 「빅데이터 활용 통계, 등 통계 다양성 확대를 위한 국가통계 승인기준 보완방안(안)

## 2 국가통계의 시기술 도입 논의와 적용 사례

### (1) AI·빅데이터 통계의 도입 논의

*“Just as haute cuisine must incessantly reinvent itself in order to stay at the forefront of gastronomy, official statistics is also confronted with a rapidly changing context and needs.”*

“최고급 식단이 미식계의 선두 자리를 계속 유지하기 위해서는 끊임없는 재창조가 필요한 것처럼, 공식 통계<sup>5</sup> 역시 급속하게 변화하는 상황과 요구사항을 정면으로 마주해야 한다.”

Walter J. Radermacher, 前 Eurostat 통계청장, 2018

#### ■ 유럽 국가들을 중심으로 국가통계의 품질 및 조사환경 개선에 대한 공감대가 형성됨에 따라, 이를 해결할 방안으로서 빅데이터 및 인공지능 기술 도입이 추진 및 검토되고 있음

##### ○ 국가통계 개선이 이슈로 부각된 주요 원인들로 아래와 같은 현상이 거론

- ① 새로운 유형의 산업이 빠르게 등장함에 따른 산업구조 반영의 어려움
- ② 기업 활동의 다양화(지역에 기반하지 않은 산업, 사무실 연락처 부재 등)
- ③ 조사별 응답자 중복에 의한 회수를 악화
- ④ 즉시 활용 가능한 민간 중심의 데이터의 급증 등

##### ○ 공통적인 문제 인식과는 별개로 AI 또는 빅데이터 기술 도입에 대한 가치 판단 및 취지는 국가별로 상이하며, 사안에 접근하는 방식 또한 다양

#### ■ 관련 연구들은 UN유럽경제위원회(UNECE)의 국가통계 현대화 논의<sup>6</sup>가 본격적으로 진행된 2014년 이후 각 국가별로 꾸준히 추진되어 왔음

※ 참고 : 아일랜드는 2014년 UNECE의 국가통계 현대화 일환 프로젝트에서 주도적 역할을 함으로써 타 국가대비 초기 이슈를 선점

##### ○ (아일랜드 : 도입 유형 및 기회요인 분석) 빅데이터 기술을 통계에 활용하는 4가지 유형을 제시\* 하였으며, 관련 된 기회 및 위협을 분석

\* ① 기존 통계의 완전/부분 대체, ② 추가적인 통계 정보 제공, ③ 통계 추정치의 개선, ④ 완전히 새로운 통계 정보 제공

<sup>5</sup> 일반적으로 공식통계(Official Statistics)와 국가통계(National Statistics)는 동일 의미로 혼용됨

<sup>6</sup> UNECE에서는 정기적으로 고위그룹(High Level Group, HLG) 차원의 국가통계의 현대화 논의가 이루어지며, 그 일환으로 빅데이터 활용 국가통계 프로젝트가 추진되어 왔음

**【표 1】 빅데이터를 국가통계에 활용함에 따른 기회 요소(아일랜드)**

범주(Category)	기회 요인(Opportunities)
데이터	<ul style="list-style-type: none"> <li>• 보완, 대체, 개선 및 기존의 데이터에 새로운 요소를 추가</li> <li>• 서로 다른 데이터 간 연결성 개선</li> <li>• 전산사회과학, 데이터 과학 및 데이터 산업들 간의 협업 강화</li> </ul>
품질/기능	<ul style="list-style-type: none"> <li>• 시의성 있는 결과 생산</li> <li>• 품질 및 진실성 향상</li> <li>• 쉬운 관찰(혹은 지역) 간 교차 비교</li> <li>• 새롭고 더 나은 통찰력을 가져다 줄 신규 데이터 분석</li> <li>• 마이크로 레벨 및 미시 분석으로 확장 및 보완</li> <li>• 기존 통계의 구성을 재편</li> </ul>
생산성/효율성	<ul style="list-style-type: none"> <li>• 시민과 기업의 설문 응답 피로도 개선</li> <li>• 업무 최적화 및 생산성 향상</li> <li>• 비용의 절감</li> <li>• 고 부가가치 업무에 직원을 재배치</li> <li>• 국가 통계의 가시성 및 활용 강화</li> </ul>

※ 자료 : IOS(2015), "The opportunities, challenges and risks of big data for official statistics"

- **(스위스 : 도입 필요성을 이론적으로 분석)** 스위스 통계청은 국가통계의 효과적 정책 지원을 위해 조사 통계와 데이터 과학에 기반한 통계를 병행 생산해야 함을 제안
  - 의사결정 과정에서 연역적 추론과 귀납적 추론 결과가 함께 고려되어야 효과적임을 근거로 데이터 과학 기반의 통계 필요성을 역설

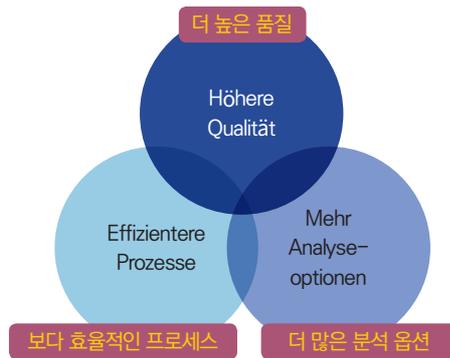
**【표 2】 일반 통계와 데이터 과학의 추론 방식에 따른 구분(스위스)**

통계 구분	목적	추론의 방식
전통적 의미의 통계	이론적 개념을 이용해 기존 개념(가설)의 유효성을 설명하고 확인하기 위한 기초 자료	연역적(가설 우선)
데이터 과학	통제 또는 조사자의 감독 없이 새로운 개념(가설 또는 이론)을 개발하기 위해 수집 및 설계	귀납적(데이터 우선)

※ 자료 : DGINS(2018)

- **(독일 : 프로세스 혁신을 논의)** 독일 연방 통계청은 머신러닝 기술을 기존 통계 생산 프로세스에 도입하는데 주안점을 두고 연구 수행
  - 2017년 국가차원의 디지털 어젠다(Agenda) 59건 중 선도 프로젝트로 '국가통계의 머신러닝 기술 검증'이 채택되어 관련 연구가 활발히 진행
  - 국가통계의 자동화 및 기업통계 영역에서의 시도입이 목표

■ [그림 1] 머신러닝 도입을 통한 기대효과



※ 자료 : 독일 연방 통계청(2019), "Machine Learning im Statistischen Bundesamt"

- **(네덜란드 : 빅데이터의 실질적 활용 방법)** 네덜란드는 빅데이터의 장점을 열거하면서도, 빅데이터를 기존 통계 방법론으로 해석하려는 시각을 견지
  - 새로운 방법론의 활용이 사용자에게 투명하게 공개되는 한, 국가통계청은 신규 모델의 사용을 두려워하지 않아야 함을 제언

■ [표 3] 국가통계에서의 빅데이터 활용 방안(네덜란드)

활용 구분	해석
빅데이터의 불완전성을 그대로 수용	<ul style="list-style-type: none"> <li>• 빅데이터는 연속성·비교가능성을 유지하기 어렵고 모집단의 범위가 실시간으로 변화할 수 있어 시계열 설명이 불가능하고 비약이 발생될 여지</li> <li>• 반면 빅데이터는 존재하는 사실 자체로 사회의 흥미를 유발하는 효과</li> </ul>
통계적 모델링을 통한 데이터 정형화	<ul style="list-style-type: none"> <li>• 최근 수리통계학 및 응용통계학자 간 빅데이터를 통계적으로 해석하기 위한 다양한 방법이 개발되었음</li> <li>• 기계학습 기술과 같은 신규 방법론은 전통적 통계 기법과 함께 고려 가능</li> </ul>

※ 자료 : 네덜란드 통계청(2020)

- **(일본 : 先실험 後평가)** 일본은 국가통계에 빅데이터를 활용 시 시의성·적시성 등에서 이점이 있는 반면 정확성·편중성 측면에서 우려를 제기
  - 2019년 상업통계를 민간 부문 빅데이터를 활용해 대체하는 실험을 추진
  - \* 해당 사례는 UNECE가 권장하는 민간 POS데이터 활용 통계를 벤치마크

**■ [표 4] 민간 부문 빅데이터 활용을 통한 기존 통계 대체가 주는 장단점(일본)**

장 점	단 점
<ul style="list-style-type: none"> <li>• 통계 공표의 가속화</li> <li>• 집계 빈도를 높일 수 있음</li> <li>• 품목 및 행위 기반 데이터를 활용하므로 표준 분류 체계에 비해 유연한 방식으로 집계 가능</li> <li>• 설문 참가자의 부담이 감소</li> <li>• 통계작성 전반의 효율성 향상</li> </ul>	<ul style="list-style-type: none"> <li>• 정확성과 편중성(Bias)을 제어하기 어려움</li> <li>• 기업의 합병 및 파산 등의 가능성으로 인해 데이터의 지속적인 가용성을 보장하기 어려움</li> </ul>

※ 자료 : RIETI(2019), "Can Big Data Change Official Statistics"

**■ 이처럼 다양한 국가에서 도입을 위한 논의가 이루어지는 것과 더불어 실증·시범 사업 등 세계 각국의 발 빠른 실험이 병행되고 있는 상황**

- 이와 관련된 글로벌 단위의 조사로는 독일 연방 통계청 주도로 수행한 「통계 기관에서의 머신러닝 사용에 관한 조사」<sup>7</sup>가 유일
- 상기 조사 대상 국가는 아래와 같으며 본고는 조사 결과 중 도입 및 실증 단계에 근접한 유의미한 적용 사례를 선별 후 분석하였음
  - 독일 : 독일 내 14개 주 통계청 및 18개 통계 생산 기관
  - 해외 : EU 27개 회원국, EFTA 4개국, 비EU 6개국 등

**■ [표 5] 통계 기관에서의 머신러닝 사용에 관한 조사의 대상국(독일 제외)**

구 분	대상 국가
유럽연합(EU)	오스트리아, 벨기에, 캐나다, 크로아티아, 키프로스, 체코, 덴마크, 에스토니아, 핀란드, 프랑스, 그리스, 헝가리, 아일랜드, 이탈리아, 라트비아, 리투아니아, 룩셈부르크, 몰타, 네덜란드, 노르웨이, 폴란드, 포르투갈, 루마니아, 슬로바키아, 슬로베니아, 스페인, 스웨덴, 영국
유럽자유무역연합(EFTA)	아이슬란드, 노르웨이, 스위스, 리히텐슈타인
비EU 국가	호주, 캐나다, 이스라엘, 일본, 뉴질랜드, 미국
기타	유럽연합통계청(Eurostat), OECD

<sup>7</sup> 독일 연방 통계청(2018), "Survey on the use of machine learning in statistical institutions"

## (2) 해외 각국의 AI·빅데이터 통계 주요 도입 사례

### ● (미국) 정형·비정형 설문 응답에 대한 코드 분류를 위해 도입

활용 기관	BLS(노동 통계국)	도입 현황	도입
활용 구분	업무상 상해 및 질병 설문 조사의 근로자 상해 응답에 대한 자동 코드 분류		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 업무상 상해 및 질병 설문조사를 통해 수집되는 서술형 답변은 매년 수만 건으로, 수천 개의 상해 및 질병 특성 코드 중 상응하는 6개 코드와 매핑이 필요</li> <li>• (도입 방법) 설문 응답을 분석해 각 응답에 가장 적합한 코드 6개를 추천해줌으로써 자동 분류하는 방식</li> </ul>		
활용 기술	정규화된 로지스틱 회귀(~2018) → 딥러닝(2019~)	도입 목적	생산성, 정확성

활용 기관	Census(통계청)	도입 현황	테스트 단계
활용 구분	미국 지역사회 설문 조사의 산업 및 직업에 대한 코드 분류		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 미국 지역사회 설문 조사(ACS)는 매년 약 250만 명의 개인 산업 및 직업과 관련된 표준화되지 않은 응답이 수집되며, 이를 산업·직업 분류 코드와 매핑해야 함</li> <li>• (도입 방법) 설문상의 산업 및 직업 관련 응답을 분석해 이에 상응하는 산업·직업 분류 코드로 자동 분류</li> </ul>		
활용 기술	로지스틱 회귀	도입 목적	생산성, 정확성

### ● (독일) 분류, 데이터 연계, 보조 데이터 확보, 이상치 탐지 등 다방면의 테스트가 진행되고 있으며, 분류와 관련된 업무 일부는 실제 도입

활용 기관	연방 통계청	도입 현황	도입
활용 구분	국내 법인에 대한 유럽국민계정체계(ESA)상 기관 분류 할당		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 유럽 연합(EU)에 소속된 국가는 유럽국민계정체계가 활용하는 기관 분류에 맞추어 자국 법인을 분류해야 함</li> <li>• (도입 방법) 법인의 회계 활동 정보에 기반해 유형을 파악하고 기관 분류</li> </ul>		
활용 기술	서포트 벡터 머신(SVM)	도입 목적	생산성, 연결성

활용 기관	연방 통계청	도입 현황	도입
활용 구분	공예 분야 행정 통계의 집계 대상 분류		
세부 내용	<ul style="list-style-type: none"> <li>• <b>(업무 개요)</b> 독일 수공예 분야 통계는 납품 과정에서 발생하는 행정 데이터를 통해 생산되나, 거래 과정에 명시된 모든 기업이 수공예 분야 기업이 아니므로 관계없는 기업을 수기로 찾아 제외시켜야 하는 어려움 존재</li> <li>• <b>(도입 방법)</b> 행정 데이터에 등장하는 모든 법인 정보를 토대로, 제외 후보군을 분류함으로써 과업에 필요한 인력을 축소</li> </ul>		
활용 기술	랜덤 포레스트, 서포트 벡터 머신	도입 목적	정확성, 비용효율

활용 기관	연방 통계청	도입 현황	테스트 단계
활용 구분	데이터 연계를 통한 연방 고용청의 패널 데이터 보강		
세부 내용	<ul style="list-style-type: none"> <li>• <b>(업무 개요)</b> 통계청의 소득 구조 조사(SES)는 시간당 노동자의 임금 정보가 조사되므로 최저 임금에 따른 소득의 영향 정보 도출이 가능한 반면, 연방 고용청에서 제공하는 데이터인 통합 고용이력(IEB)에는 소득 정보가 누락</li> <li>• <b>(도입 방법)</b> SES데이터에 기반해 학습한 머신러닝 모델을 IEB에 적용함으로써 데이터를 연계해 최저 임금 변화에 대한 소득 영향 정보를 IEB에 추가</li> </ul>		
활용 기술	랜덤 포레스트	도입 목적	연결성

활용 기관	연방 통계청	도입 현황	테스트 단계
활용 구분	온라인 구인 공고 분류		
세부 내용	<ul style="list-style-type: none"> <li>• <b>(업무 개요)</b> 기업이 제시하는 임금, 구인 인력의 교육 수준, 채용 담당자의 구인 형태 등 국가 정책 수립에 필요한 정보가 부족한 경우가 많음</li> <li>• <b>(도입 방법)</b> 온라인 구인 공고를 웹 스크래핑 후 텍스트를 분석함으로써 구인 공고의 특성을 분류하고 관련 정보를 추출</li> </ul>		
활용 기술	K 최근접 이웃 클러스터링, 다항분포 나이브 베이즈	도입 목적	부가정보, 인사이트

활용 기관	연방 통계청	도입 현황	테스트 단계
활용 구분	기업 구조 통계 응답결과의 이상값 식별		
세부 내용	<ul style="list-style-type: none"> <li>• <b>(업무 개요)</b> 기업 통계 조사의 응답결과에는 타당성이 부족하거나 극단적인 이상치가 포함되어 있어 검토 과정에 많은 시간이 소요</li> <li>• <b>(도입 방법)</b> 검토 대상을 줄이기 위한 방법으로서 머신러닝을 활용해 검토 대상을 추천</li> </ul>		
활용 기술	격리 포레스트(Isolation Forest)	도입 목적	생산성, 정확성

● (캐나다) 분류, 이상값 검출 등을 위해 도입되었으며, 결측 값 대체를 위한 분류작업에 비지도 학습을 활용한 게 특징

활용 기관	연방 통계청	도입 현황	도입
활용 구분	월별 소매 거래 조사 및 분기별 소매 상품 조사에 스캔 데이터 활용		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 월별 소매 거래 조사는 복미 상품 분류 시스템(NAPCS)의 상품 코드에 기준해 상품 코드별 매출 현황을 통계화하는 작업</li> <li>• (도입 방법) 민간 기업 스캐너 데이터로 수집되는 상품 설명을 텍스트 분석하여 NAPCS코드와 매핑</li> </ul>		
활용 기술	XGBoost(분산 그라디언트 부스팅), BoW(Bag-of-Words), n-gram 모델	도입 목적	기존 통계 대체

활용 기관	연방 통계청	도입 현황	도입
활용 구분	국제 무역 데이터 이상값 검출		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 국제 무역 통계의 데이터 중 단위의 오류, 0과 1로 표기된 이상값을 처리하는 작업을 위해 수기 검토와 변경 승인 프로세스를 운영</li> <li>• (도입 방법) 머신러닝을 통해 검토 프로세스를 자동화</li> </ul>		
활용 기술	XGBoost 트리 모델	도입 목적	생산성, 정확성

활용 기관	연방 통계청	도입 현황	도입
활용 구분	인구 조사를 구성하는 이민 허가 변수에 대한 결측값 대체		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 2016년 인구 조사에 이민 허가와 관련된 데이터가 추가 변수로 지정되었으며, 이는 추가 조사가 아닌 별도 데이터와의 연계를 통해 확보되었음. 일부 개인의 경우에는 데이터가 없어 결측값이 발생하였으며 이를 대체해야 하는 업무가 발생</li> <li>• (도입 방법) 유사 특성을 보이는 조사 데이터를 학습하여 결측 데이터 항목의 대체값을 추정</li> </ul>		
활용 기술	Relief 알고리즘, K 최근접 이웃 클러스터링	도입 목적	기존 통계 대체

활용 기관	연방 통계청	도입 현황	도입(한시적)
활용 구분	인구 조사 신규 콘텐츠 수요 분석		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 2016년 조사된 인구 조사의 신규 콘텐츠 설문결과 약 110만 건에 대한 활용처 모색</li> <li>• (도입 방법) 설문 결과 텍스트의 맥락 및 주제를 해석하여 요약</li> </ul>		
활용 기술	자연어 처리 기술	도입 목적	인사이트

● (네덜란드) 웹 수집 데이터 분석 및 데이터 연계 목적으로 활용

활용 기관	통계청	도입 현황	도입
활용 구분	인구 조사를 구성하는 이민 허가 변수에 대한 결측값 대체		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 유럽 통계 시스템(ESS) 내 빅데이터 프로젝트의 일환으로써, 기업의 인터넷 정보, 활동, 주소 정보, 소유권 구조 등과 같은 기존 정보를 개선하거나 업데이트하기 위해 추진<sup>8</sup></li> <li>• (도입 방법) 구글 맞춤형 검색 시스템, 기업 URL 및 연계된 행정데이터를 활용해 대상 기업과 관련한 웹 정보를 수집한 후, 텍스트 마이닝 및 추론 기술을 통해 기업 정보를 정제</li> </ul>		
활용 기술	의사결정 트리, 랜덤 포레스트 등	도입 목적	부가정보

활용 기관	통계청	도입 현황	도입
활용 구분	보건 실태 조사를 대상으로 한 혼합형 설문 조사의 무응답 대체 기법 보강		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 네덜란드는 설문 조사의 비용절감 및 회수율 악화를 해결하기 위한 대책을 지속적으로 연구해왔으며, 해결책의 일환으로 혼합형 설문 조사(Mixed Mode Survey)<sup>9</sup>를 추진해왔음</li> <li>• (도입 방법) 실태 조사의 실시 이후 발생하는 무응답을 타 조사결과를 통해 대체하는 과정에서 발생 가능한 각종 분류문제(중화, 할당 등)를 머신러닝 기법을 혼용해 해결</li> </ul>		
활용 기술	분류 트리	도입 목적	연결성, 정확성

활용 기관	통계청	도입 현황	시범 도입
활용 구분	인터넷 구매를 통한 자국 소비자의 국경 간 거래 현황 예측		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 유럽연합에 소속된 국가들 사이의 인터넷 구매는 설문 조사의 샘플링 및 조사대상의 언어 문제 등으로 인해 거래 총액을 정확하게 측정하거나 추정하기 어려움</li> <li>• (도입 방법) 유럽연합 소속 법인의 세금 정보와 인터넷 데이터를 연계해 국경 간 인터넷 거래를 추정<sup>10</sup></li> </ul>		
활용 기술	비선형 SVM(RbfSVC), 랜덤 포레스트	도입 목적	부가정보

활용 기관	통계청	도입 현황	테스트 단계
활용 구분	기업 통계 무응답 대체 기법의 개선		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 기업 통계를 위한 무응답 대체 기법의 정확성 문제가 지적되었으며, 신뢰도 높은 추정을 위한 개선이 요구되고 있음</li> <li>• (도입 방법) 머신러닝 기법을 사용해 추정치의 신뢰도 향상 및 자동 무응답 대체를 위한 모델을 구축</li> </ul>		
활용 기술	그래디언트 부스팅 머신(GBM)	도입 목적	정확성, 생산성

<sup>8</sup> Eurostat(2018), "ESSnet Big Data Specific Grant Agreement No 1"(SGA-2)  
<sup>9</sup> 조사 결과 데이터를 확보하는 과정에서 2개 이상의 조사 방법을 채택하는 기법을 의미하며, 상기 사례에서는 1차 실사 결과에서 비롯된 무응답 샘플의 대체에 다른 조사 방식으로 도출된 결과를 활용함으로써 모집단 편향성을 해소하는 접근을 추진  
<sup>10</sup> Q.A.Meertens, C.G.H.Diks(2018), "A Data-Driven Supply-Side Approach for Measuring Cross-Border Internet Purchases"

■ (스위스) 파라데이터(Paradata)<sup>11</sup> 머신러닝 분석 결과를 회수율 제고에 활용

활용 기관	연방 통계청	도입 현황	도입
활용 구분	조사 무응답 매커니즘의 모델링		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 설문 조사의 무응답 사례를 최소화하기 위한 노력</li> <li>• (도입 방법) 무응답자의 행위를 모델링하여 유형을 분류하고 실사에 참고</li> </ul>		
활용 기술	카이-제곱 자동 상호 작용 탐지 (CHAID, 의사결정 트리의 종류 중 하나)	도입 목적	생산성, 비용효율

활용 기관	연방 통계청	도입 현황	테스트 단계
활용 구분	이상치로 의심되는 응답 결과 추천		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 수정 규칙이 없는 조사 항목들에 대한 이상치를 검토</li> <li>• (도입 방법) 머신러닝을 통해 잘못된 응답으로 의심되는 케이스를 감지하고 이를 검토자에게 전달하여 검토 시간을 단축</li> </ul>		
활용 기술	부스티드 결정 트리, 랜덤 포레스트, 신경망, 나이브 베이즈 등	도입 목적	생산성

활용 기관	연방 통계청	도입 현황	테스트 단계
활용 구분	항공 이미지에 기반한 토지 피복 및 토지 사용 코드 분류 자동화		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 지표면의 물리적 상태를 인공위성 촬영 이미지 등을 이용하여 분류 항목별 코드로 구분하는 작업</li> <li>• (도입 방법) 머신러닝을 통해 항공 이미지를 분석하고 각 지형에 상응하는 코드와 자동 매핑</li> </ul>		
활용 기술	합성곱 신경망(CNN)	도입 목적	기존 통계 대체

■ (오스트리아) 특정 통계 조사가 아닌 범용적인 도입에 주안점

활용 기관	통계청	도입 현황	도입
활용 구분	설문 무응답 대체 및 데이터 통합		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 행정 데이터를 포함하는 설문 조사 및 프로젝트의 결측값 대체 이슈 및 공통 변수가 존재하는 데이터 셋의 연계</li> <li>• (도입 방법) 군집화 알고리즘을 통해 결측값을 대체하고 분류 기법을 활용해 데이터 셋 연계</li> </ul>		
활용 기술	K 최근접 이웃 클러스터링(결측값 대체), 랜덤 포레스트(데이터 연계)	도입 목적	정확성, 연결성

<sup>11</sup> 조사과정자료(Paradata, Process Data)는 조사를 진행하는 과정에서 자연적으로 발생하게 되는 파생 자료이며 조사 관리를 위한 보조 자료의 성격을 가짐(임경은, 2012)

활용 기관	통계청	도입 현황	도입(한시적)
활용 구분	행정 데이터와 SILC 융합을 통한 ICT 조사의 소득 문항 대체		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 소득 및 생활 조건 통계(SILC)와 ICT관련 조사에 가계소득을 묻는 문항이 포함되어 있으나 조사 대상이 중복</li> <li>• (도입 방법) 행정자료 및 SILC에 기반해 ICT조사 대상의 가계소득을 추정하는데 머신러닝을 활용하고 ICT조사의 소득 항목 제외</li> </ul>		
활용 기술	랜덤 포레스트	도입 목적	연결성

● (루마니아) 데이터 결합 정확도 향상을 통해 표본의 질을 개선

활용 기관	통계청	도입 현황	도입
활용 구분	EU-SILC 등록 통계 표본의 질 개선		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 유럽연합의 소득 및 생활조건에 관한 연합 통계(EU-SILC)에 등록할 통계는 기타 표본 조사 과정에서 수집된 정보와 행정 정보를 결합해 모집단 및 표본 추출이 수행되어야 함</li> <li>• (도입 방법) 기타 표본 조사 파생 정보와 행정 데이터 간 결합 과정에서 머신러닝을 활용</li> </ul>		
활용 기술	랜덤 포레스트	도입 목적	연결성, 정확성

● (핀란드) 비정형 텍스트 데이터를 토대로 특성을 분류하기 위해 활용

활용 기관	통계청	도입 현황	도입
활용 구분	경찰청 교통사고 보고 문건에 기반해 사고 내역 자동 분류		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 자유 형식으로 작성된 도로 교통사고 보고서에 기반해 상해 유발 사고와 그 외 사고를 분류하는 작업</li> <li>• (도입 방법) 텍스트 마이닝 및 분류 기법을 토대로 비정형 텍스트 기반의 보고서 내용을 분류</li> </ul>		
활용 기술	TF-IDF, 로지스틱 회귀모형	도입 목적	생산성

● (벨기에) 일자리 분석을 위한 코드 분류에 활용

활용 기관	통계청	도입 현황	테스트 단계
활용 구분	구인 공고별 알맞은 경제 활동 통계 분류 코드 예측		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 일자리 구인 현황에 대한 분석 수요</li> <li>• (도입 방법) 벨기에 일자리 포털에 등록된 구인 공고상 직무 정보를 수집해 유로연합 경제 활동 통계 분류(NACE) 코드와 매핑하여 통계 생산</li> </ul>		
활용 기술	서포트 벡터 머신	도입 목적	부가정보

● (룩셈부르크) 설문 응답값 검증 및 분류 부문에 도입

활용 기관	통계청	도입 현황	도입
활용 구분	민간 기업의 스캐너 데이터를 활용해 상품 분류 코드 추천		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 소비자 물가 지수의 소비 기준으로 활용되는 COICOP(목적에 따른 개별 소비 분류) 코드와 실제 소비 현황을 매핑 해야함</li> <li>• (도입 방법) 캐나다의 「월별 소매 거래 조사 및 분기별 소매 상품 조사에 스캔 데이터 활용」 부문과 유사</li> </ul>		
활용 기술	다중 선형 회귀(MLR)	도입 목적	기존 통계 대체

활용 기관	통계청	도입 현황	도입
활용 구분	근로 환경 조사 데이터 검증을 위해 머신러닝 도입		
세부 내용	<ul style="list-style-type: none"> <li>• (업무 개요) 근로 환경 조사는 개인의 사내 R&amp;D활동 수준을 측정하나, 응답값에 대한 검증이 필요하다는 지적</li> <li>• (도입 방법) 현재·과거 설문 조사 및 행정데이터를 결합해 조사 응답자가 조사 기간 내 R&amp;D 활동을 수행했는 여부를 추정하여 조사 결과 검증</li> </ul>		
활용 기술	양상블 모델링(Model Stacking)	도입 목적	정확성, 연결성

3 AI·빅데이터 통계 도입 시 예상 쟁점

● 국가통계 부문의 AI·빅데이터 기술 도입 목적은 (1) 신규 통계 생산, (2) 통계 생산 프로세스의 현대화 총 두 가지 시각으로 구분 가능

- 두 시각은 논의의 출발점이 다소 상이해 목적을 달성하기 위한 국가별 정책 및 제도에 차이가 발생할 것임을 추정해볼 수 있음

■ [표 6] 국가통계의 AI·빅데이터 기술 도입의 두 가지 시각

구분	세부 적용 분야
신규 통계 생산	<ul style="list-style-type: none"> <li>• 대체(Replace) : 기존 통계를 완전 또는 부분 대체하는 수단</li> <li>• 추가(Addition) : 조사 통계 방식으로 도출이 어려운 부문에 대해 데이터 분석에 기반을 둔 우회적 해결</li> <li>• 인사이트(Insight) : 완전히 새로운 분야의 통계 데이터 제시</li> </ul>
통계 생산 프로세스의 현대화	<ul style="list-style-type: none"> <li>• 정확성(Accuracy) : 통계의 정확도 향상</li> <li>• 생산성(Productivity) : 업무 효율화를 위한 자동화 기법</li> <li>• 연결성(Connectivity) : 연결 가능한 데이터들의 상호 보완적 활용</li> <li>• 비용(Price) : 통계 생산의 비용 절감을 위해 활용</li> </ul>

- 3장에서는 AI·빅데이터 기술이 국가통계 제도권 내에 안착하는데 고려되어야 할 예상 쟁점 사항들을 상기 시각으로 구분하여 분석하였음

## (1) 신규 통계 생산

● (조작적 정의) 본고에서의 신규 통계란 조사통계 기법을 활용하지 않고 빅데이터·SI기술을 활용해 생산한 통계(국가통계)로 정의

○ 빅데이터(민간, 행정)를 활용한 기초 통계 : 국가 현황 파악에 기초가 되는 1차 통계로써, 통계 생산을 위해 민간 빅데이터 및 행정 통계를 활용

\* 완전히 새로운 통계를 발굴하는 것과 더불어 이미 존재하는 조사통계 방식의 기초 통계를 빅데이터·SI기술을 통해 대체하는 경우가 이에 해당

○ 빅데이터(민간, 행정)를 활용한 가공 통계 : 국가 정책 지원, 사회 현상의 진단 등의 목적으로 민간·행정 빅데이터를 가공·분석한 2차 통계

\* 소셜 분석, 언론 데이터 분석, 빅데이터·SI기술을 활용한 지수 개발 등 분석 대상과 기술로서 빅데이터 및 SI를 채택한 경우

○ 기초 통계 데이터를 활용한 가공 통계 : 현존하는 국가 통계 데이터를 빅데이터·SI기술을 통해 2차 분석함으로써 도출하는 신규 통계

\* 통계법에 저촉되지 않는 범위 내에서 설문조사 데이터기반 빅데이터·SI분석을 수행해 별도의 부가적인 지표 및 시사점을 도출하는 경우 이에 해당

● (예상 쟁점 #1) 신규 방식을 수용하는 데 있어 방법론의 대표성 및 보편적 신뢰가 형성되지 않은 문제

○ 국가통계는 통계적으로 널리 활용되는 과학적인 작성기법을 사용하여야 하나,<sup>12</sup> 머신러닝(Machine Learning), 딥러닝(Deep Learning) 등 SI기술을 통한 분석 결과의 신뢰가 보편적으로 형성되지 않은 상황

○ 본질적인 원인은 수많은 머신러닝 알고리즘 및 파생된 기법들 중 신규 통계 생성에 적합한 기술이 제시되지 않음에서 기인

→ 신규 통계 생산 시 활용에 적합한 검증된 알고리즘을 정의하고 이에 대한 세부적인 활용 지침 마련이 요구됨

12 국가통계 기본원칙의 [제 2항 : 신뢰성 제고의 실천방안 중 발취

● (예상 쟁점 #2) 신규 통계의 생산방식은 기존의 통계 생산 프레임워크에 맞추어 해석하기에 적합하지 않음

- 국가의 통계업무를 구체화하여 단계별 제시한 국제표준은 GSBPM으로, 해당 모델의 경우 조사 통계에 적합한 프로세스를 준용하고 있어 대표적 데이터 마이닝 업계 표준(CRISP-DM<sup>13</sup>)과는 구조가 상이

\* 한국의 경우 GSBPM을 참고해 국내 환경에 맞추어 재정의한 KSBPM을 준용

- CRISP-DM은 빅데이터의 제어 및 분석 모델의 유효성 검증과 평가에 주안점을 둔 프로세스로서, 기존 통계모델과 통합하기 어려운 지점을 형성

\* 스위스 통계청은 GSBPM이 빅데이터 통계(신규 통계)를 포괄하기에 적합하지 않으며, 서로 다른 프로세스의 개발 또는 상호보완적인 연계 방안을 유럽 통계 시스템(ESS)에서 다룰 것을 강력히 권고

→ 빅데이터 분석 프로세스를 포괄할 수 있는 통계 프로세스 표준의 개정 고려

■ [표 7] CRISP-DM(좌) 및 프로세스 역할 요약(우)

	<ul style="list-style-type: none"> <li>▶ 사업 이해(Business Understanding) - 사업·기술관점 목표수립, 프로젝트 기획 등</li> <hr/> <li>▶ 데이터 이해(Data Understanding) - 데이터 수집, 특성 확인, 목표데이터 도출 등</li> <hr/> <li>▶ 데이터 준비(Data Preparation) - 분석 데이터 선별·정제·검증·구조화·통합 등</li> <hr/> <li>▶ 모델링(Modeling) - 기법선택, 유효성검증, 모델 생성, 순위 지정 등</li> <hr/> <li>▶ 평가(Evaluation) - 모델링결과 평가, 주요항목 도출, 향후단계 결정 등</li> <hr/> <li>▶ 전개(Deployment) - 전개전략 수립, 지속/유지전략 수립, 보고서 작성 등</li> </ul>
--	--

※ 자료 : IBM SPSS Modeler Essentials(2017)

<sup>13</sup> Cross-industry standard process for data mining(CRISP-DM) : 데이터 마이닝 분야에서 가장 널리 사용되는 분석 모델로서, 관련 전문가가 사용하는 일반적인 접근 방식을 정의

■ [그림 2] 한국통계업무프로세스모델(V.2.0)

기획	설계	구축	수집	처리	분석	배포	보관	평가
수요 파악	산출물 설계	수집 도구 구현	자료 수집대상 설정	자료통합	산출물 작성	공표자료 점검 및 직재	자료보관규칙 정의	평가 계획 수립
수요 확정	항목 설정	처리시스템 개발, 개선	자료 수집 준비	분류 및 코딩	산출물 검증	공표 산출물 작성	자료 보존 및 관리	수행 및 보고서 작성
산출물표 수립	수집 방법 설계	배포시스템 개발, 개선	자료 수집 진행	자료검토 및 보완	산출물 해석 및 설명 작성	자료 배포 관리		개선과제도출 및 실행계획수립
통계적 개념 정립	모집단 및 표본설계	업무 절차 설정	수집자료 점검 및 완료	결측치 처리	정보보호 및 공개범위 검토	배포 촉진		
자료 가용성 검토	자료처리 방법 설계	시스템 점검		신규변수(항목) 및 단위 도출	산출물 확정	이용자그림 관리		
통계 작성 계획안 수립	통계 작성 체계 설계	작성 과정 통합 점검		가중치 계산				
		통계 작성 체계 확정		집계				
				자료 집계 및 확정				

※ 자료 : 이의규(2018), "표준 자료처리 모델 GSDEMs의 소개 및 시사점"

■ (예상 쟁점 #3) 재현 불가능한 비결정론적(Non-Deterministic) 알고리즘<sup>14</sup> 기반의 결과를 국가 통계로 관리 가능한지 여부

- 널리 활용되는 빅데이터·시기술 중 결과 도출 과정에서 연구자의 주관적 판단이 반드시 요구되는 알고리즘 존재
  - \* 비지도학습(Unsupervised Learning)으로 분류되는 클러스터링의 경우 분류의 기준이 되는 군집의 수(k)를 연구자 판단을 통해 결정
- 빅데이터·시알고리즘의 일부는 주관적 판단 없이도 동일한 파라미터에 의해 도출된 결과가 확률적으로 다를 수 있음
  - 알고리즘 동작 과정에서 난수(Random Value)를 계산에 활용하므로, 통계의 재현 가능성을 완전히 충족하기 어려운 경우 존재
  - 통계적 모델링의 유효성을 검정력(p)으로 판단할 수 있는 것과 달리 시알고리즘은 유사한 신뢰 기준이 표준화되지 않음
- 한편, 지도학습(Supervised Learning)의 경우 주어진 검증 데이터를 대상으로 학습 모델의 성능 측정이 가능해 상대적으로 활용에 용이할 것으로 예상

→ 인공지능 기술의 특성 분석 및 표준화된 평가 기준 정립이 요구됨

<sup>14</sup> 비결정론적 알고리즘이란 예측한 그대로 동작하지 않는 알고리즘을 의미하는데, 확률적으로 최적의 근사값을 찾는 원리에 기반하므로 동일 알고리즘 시행 시 동일 결과를 100% 보장할 수 없음

■ [표 8] 신규 통계에 잠재적으로 활용 가능한 머신러닝 알고리즘의 종류와 특성

구분	설명	대표적 기술군	활용측면의 성능 분석 가능여부
지도학습 (Supervised Learning)	입력에 대한 정답(레이블)을 반복 학습시켜 정답이 제시되지 않은 입력의 정답을 찾는 알고리즘	분류, 회귀, 신경망 <sup>15</sup>	가능 (검증 데이터 기반)
반지도학습 (Semi-Supervised Learning)	정답을 보유한 소수의 입력과 그렇지 않은 다수 입력을 혼용하여 학습함으로써 성능을 높이는 기법	상동	가능 (검증 데이터 기반)
비지도학습 (Unsupervised Learning)	사람의 지도 없이 컴퓨터가 스스로 데이터의 유형을 학습하는 기술	군집화, 분포 추정 <sup>16</sup> , 분류 <sup>17</sup>	불가능 (모델에 대한 품질 평가는 가능)

※ 머신러닝 기법으로 현재 상태(State)에서 어떤 행동(Action)이 보상(Reward)을 획득하는데 효과적인지를 학습하는 방식인 강화학습(Reinforcement Learning)이 존재하나, 해당 주제에 적합하지 않다고 판단하여 제외

## (2) 통계 생산 프로세스의 현대화

### ■ (조작적 정의) 프로세스 현대화는 기존의 조사통계 프로세스를 수용하되 통계 데이터의 품질 향상, 생산 절차의 효율성 등을 개선하기 위해 빅데이터·AI기술을 도입하는 것을 의미

○ **기능 개선** : 통계 프로세스상 의사결정 요소들의 성능을 높이는 기법들로서 총화, 대체, 보정 등의 현행 통계 방법론을 보완 또는 대체<sup>18</sup>

○ **생산성 증대** : 통계 업무의 비효율을 개선하고 생산성을 높이기 위한 기술 도입으로서 코드화, 이상값(Outlier) 검출, 레코드 연계, 데이터 비식별화 지원 등으로 구성

### ■ (도입 가능 부문) 데이터를 제어하고 처리하는 절차 다수에서 AI기술 도입 가능성 검토가 가능

○ 통계 생산 프로세스 참조 기준으로는 GSBPM(v5.1)을 활용

- GSBPM은 총 8단계로 구성되며 각 단계에는 하위 프로세스가 존재

\* 본 프로세스는 통계 생산 과정에서 유연하게 적용하고 해석할 수 있음<sup>19</sup>

<sup>15</sup> 인공 신경망은 뉴런의 구조에서 착안한 지도학습의 일종으로서, 신경망의 구조가 복잡(다수의 계층으로 구성)하게 설계되어 있는 경우를 심층 신경망(Deep Learning)이라 칭하며, 일반적으로는 분류를 위한 방법론으로 활용됨

<sup>16</sup> 데이터의 분포를 설명하기에 최적인 확률 분포에 매핑하여 해석하는 기법

<sup>17</sup> 비지도학습을 통한 분류는 최근 GPT-2(비지도학습 기반 문장 예측모델)의 성공으로 조명받기 시작한 분야로서, 향후 지도학습의 대체제로 평가되고 있으나 신뢰성을 갖추기 위한 학습 규모, 학습 시간 등을 고려했을 때 통계 생산 목적의 활용 가치는 아직까지 낮음

<sup>18</sup> 유엔유럽경제위원회(2018), "The use of machine learning in official statistics" 참고

<sup>19</sup> statswiki.unece.org, "Understanding the GSBPM" 참고

- 해외 문헌을 통해 머신러닝 활용이 가능하다고 판단된 통계 생산 세부 프로세스와 과업은 아래와 같음

■ [표 9] 머신러닝 활용이 가능한 국가통계 생산 과정 및 예상 기술군(GSBPM 기준)

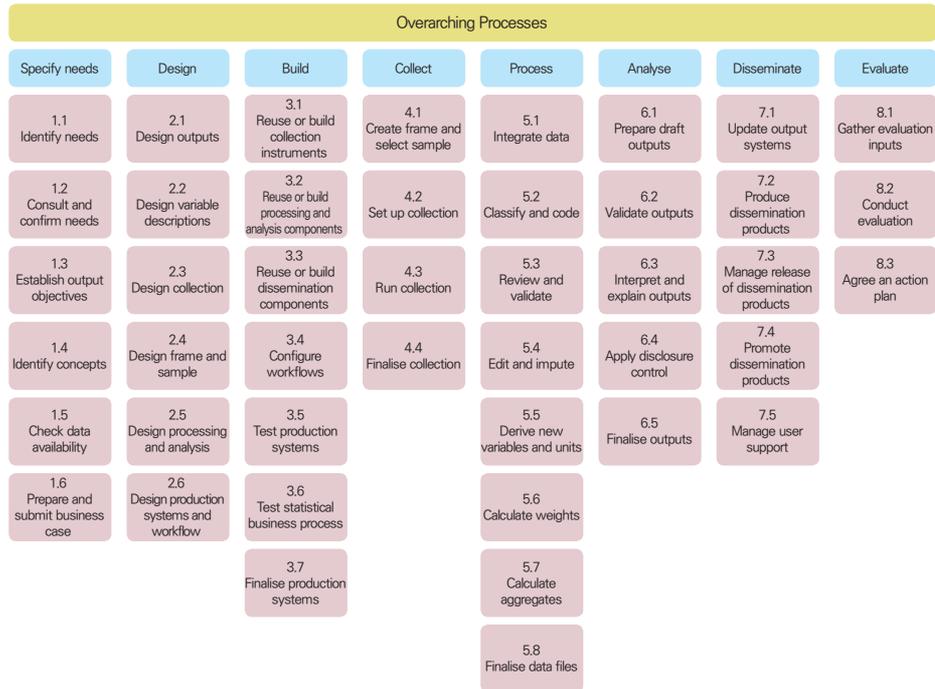
코드 (1레벨)	명 칭	코드 (2레벨)	명 칭	과업 정의	도입 가능 기술군 <sup>20</sup>
2	설계	2.4	모집단 및 표본 설계	<ul style="list-style-type: none"> <li>모집단을 식별 및 지정</li> <li>적합한 샘플링 기준 및 방법론 결정</li> </ul>	<ul style="list-style-type: none"> <li>분류</li> <li>군집화</li> </ul>
4	수집	4.1	모집단 생성 및 샘플링	<ul style="list-style-type: none"> <li>2.4에서 지정된 모집단을 생성</li> <li>2.4에서 지정된 샘플링 수행</li> </ul>	<ul style="list-style-type: none"> <li>분류</li> </ul>
		4.3	수집 진행	<ul style="list-style-type: none"> <li>실사 진행 및 후속조치 점검</li> <li>수기 데이터 입력 및 현장 작업 관리 등</li> </ul>	<ul style="list-style-type: none"> <li>분류</li> <li>군집화</li> <li>회귀</li> </ul>
5	처리	5.1	자료 통합	<ul style="list-style-type: none"> <li>데이터의 출처가 2개 이상일 시 통합</li> <li>검증된 비 통계 데이터로 설문 결과 일부를 대체하는 과정도 이에 해당</li> </ul>	<ul style="list-style-type: none"> <li>군집화</li> </ul>
		5.2	분류 및 코딩	<ul style="list-style-type: none"> <li>설문 결과 데이터의 분류 및 코딩</li> </ul>	<ul style="list-style-type: none"> <li>분류</li> </ul>
		5.3	검토 및 검증	<ul style="list-style-type: none"> <li>데이터를 검사하여 이상치, 항목 무응답, 잘못된 코딩 등 잠재적 문제, 오류, 불일치 사항을 식별</li> </ul>	<ul style="list-style-type: none"> <li>분류</li> <li>군집화</li> </ul>
		5.4	편집 및 대체	<ul style="list-style-type: none"> <li>부정확 데이터 및 누락, 신뢰할 수 없는 데이터를 새로운 값으로 대체 및 제거</li> </ul>	<ul style="list-style-type: none"> <li>분류</li> <li>회귀</li> <li>분포 추정</li> </ul>
		5.6	가중치 계산	<ul style="list-style-type: none"> <li>설계 단계(2)에서 정의한 가중치를 조사결과에 적용함으로써 모집단에 대한 통계 결과를 도출</li> <li>설문 결과의 무응답 조정 및 변수의 정규화 등</li> </ul>	<ul style="list-style-type: none"> <li>분류</li> <li>회귀</li> </ul>
6	분석	6.2	산출물 검증	<ul style="list-style-type: none"> <li>품질 측정 프레임워크에 따라 통계 품질을 검증</li> </ul>	<ul style="list-style-type: none"> <li>군집화</li> </ul>
		6.3	산출물 해석 및 설명 작성	<ul style="list-style-type: none"> <li>결과가 초기 기대치를 얼마나 잘 반영하는지 평가</li> <li>다양한 관점에서 통계 결과를 해석</li> <li>심층 통계 분석 등</li> </ul>	<ul style="list-style-type: none"> <li>(미정)<sup>21</sup></li> </ul>
		6.4	정보보호 및 정보공개 검토	<ul style="list-style-type: none"> <li>데이터 배포 과정의 기밀성 훼손 방지</li> <li>필요에 따라 데이터 비식별화 처리 수행</li> </ul>	<ul style="list-style-type: none"> <li>회귀</li> <li>분류</li> </ul>
7	배포	7.5	이용자 지원 관리	<ul style="list-style-type: none"> <li>통계 서비스에 대한 사용자 질의 및 요청에 대해 합의된 마감일 내에 응답을 제공</li> </ul>	<ul style="list-style-type: none"> <li>(미정)</li> </ul>
8	평가	8.2	평가 실시	<ul style="list-style-type: none"> <li>8.1단계에서 확보한 평가 피드백과 목표한 벤치마킹 대상(만약 존재한다면)과 결과를 비교한 후 평가보고서 작성</li> </ul>	<ul style="list-style-type: none"> <li>(미정)</li> </ul>

※ 자료 : 유엔유럽경제위원회, 독일 통계청

<sup>20</sup> 인공지능경망의 경우, 설계 목적에 따라 모든 기술군을 포괄 가능하므로 별도로 명시하지 않음

<sup>21</sup> 2차 분석의 영역으로서 예상 기술군을 특정하는 것은 부적절하다 판단하였음

■ [그림 3] 일반 통계 비즈니스 프로세스 모델(GSBPM v5.1)



※ 자료 : unece.org

● (예상 쟁점 #1) 기존의 방법 대비 우수성을 객관적으로 비교 가능한가

- 시기술 도입이 현실화되기 위해서는 기존 활용 기법 대비 AI 기술을 활용할 시 품질의 상대적 우위를 증명해야 하나, 품질 진단 기준이 불명확
  - 1차 통계의 경우 과거 데이터 또한 기존 방법론을 통해 생산된 자료일 가능성이 높아 객관적인 방법론 우위 측정에 한계
  - \* 통계 주제 및 특성에 따라 기업 정보, 행정 정보 등 기존 조사 외 객관적 정보를 사후 습득할 수 있는 경우에는 해당 이슈에서 자유로울 수 있음
- 신규 통계의 쟁점(신규 통계 생산-예상 쟁점#3 참조)과 마찬가지로 알고리즘 특성에 의한 신뢰성 문제가 대두될 가능성
  - 해외 사례에서 비지도학습을 국가통계에 활용한 사례가 존재하나, 그럼에도 불구하고 검증 이슈에서 자유로운 것은 아님

→ 전통적 통계 기법-AI알고리즘의 비교 검증 방안 발굴이 필요

● (예상 쟁점 #2) AI기술 도입은 통계 생산부터 공표까지의 제한된 생산 기간을 준수할 수 있는 해법인가

- AI모델링은 방법론의 종류, 학습 데이터의 규모, 최소 성능 기준<sup>22</sup> 등에 의해 최종 모델 산출까지의 시간이 유동적
- 국가 통계는 시의성<sup>23</sup> 및 정시성<sup>24</sup>을 갖추기 위해 필연적으로 생산 기한을 준수해야 하므로, AI학습 알고리즘의 학습시간-품질 간 조율 문제가 발생할 가능성

→ 다수의 실증 사업을 통한 통계 유형별 도입 적합성 진단 필요

※ 참고 : 현행 조사기관 국가통계 또한 제한된 생산 기간 준수를 고려하는 사례가 존재하는데, 항목 무응답 대체 기법으로 명시적 형태의 대체 모형(Explicit Model)<sup>25</sup>이 아닌 내재적 모형(Implicit Model)<sup>26</sup>을 주로 채택하는 경우가 대표적<sup>27</sup>

## 4 국가통계 AI도입 촉진 방안

### (1) 국내 통계 제도의 개편 방향

● 국내 통계청은 단기적으로 현행법 내 승인가준의 일부 보완, 중장기적으로는 신고제도 및 시범통계제도 도입을 검토할 것임을 예고

- 단기적 관점의 승인가준 보완은 기존 통계작성승인 심사표상의 (1) 조사통계에 국한된 심사문항을 범용적인 문구로 변경하고 (2) 기존 통계 대체를 위한 빅데이터 도입의 경우 중복을 허용할 수 있도록 하는 조치

\* 예1) (현행) 조사표 설계나 조사 항목이 동일 또는 유사한가?  
→ (수정) 수집 항목 및 결과표 항목이 동일 또는 유사한가?

\* 예2) (추가) 기존 통계를 보완하는 사항이 있는가?

<sup>22</sup> AI성능 지표로서 민감성, 정밀도, 거짓양성률(FPR) 등이 존재(전자신문, 2019.10.)

<sup>23</sup> 시의성 : 작성 기준시점과 결과공표시점 간의 차이를 나타내는 통계의 현실 반영도와 관련된 개념

<sup>24</sup> 정시성 : 예정된 공표시기를 정확히 준수하는가에 대한 개념

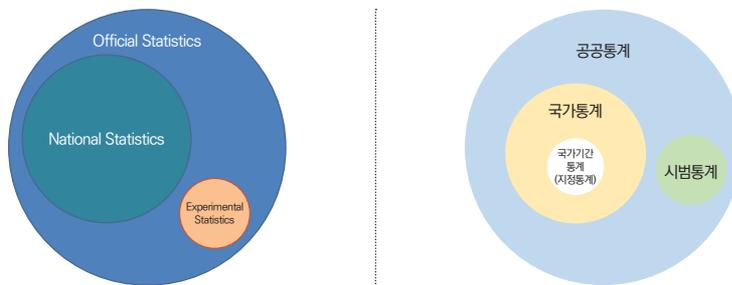
<sup>25</sup> 전체 응답 결과에 상응하는 예상 분포를 추정해 무응답 항목을 보완하는 방법으로서, 모델링 결과가 존재해 수치 검증이 가능한 장점을 가지나, 추정 난이도 및 소요 기간으로 인해 선호되지 않음

<sup>26</sup> 무응답 항목을 추정할 수 있는 별도 데이터를 참조하여 결측값을 대체하는 방법으로서, 추정 난이도가 낮고 소요기간이 짧아 선호되는 기법임. 대표적으로 핫덱, 콜드덱 기법이 존재

<sup>27</sup> 명시적 형태의 대체 모형을 추정해 항목 무응답을 처리하는 기법은 생성 모델의 검증이 가능한 장점이 있는 반면, 모델링 과정의 기간 소요, 추가 예산 등의 현실적 어려움이 존재

- 중장기적 관점에서 고려될 신규 제도 도입은 국가가 관리하는 통계의 범위를 확대하는 방향을 의미하나, 아직 구체화되지 않음(아래 참조)
- (시범통계제도) 통계청에서 언급한 시범통계제도란 영국의 실험통계(Experiment Statistics) 제도에서 착안한 국가통계 관리체계 개편을 의미
- 영국의 실험통계는 새로 개발되거나 혁신적인 공식통계로 정의<sup>28</sup>되어 AI·빅데이터 활용 통계를 국가에서 관리하는데 용이한 제도
  - ※ 영국은 공식통계 범위 내 국가통계와 실험통계를 별도로 두어 관리하나, 국내는 국가에서 관리하는 통계가 국가통계로 일원화
  - ※ 실험통계는 추후 국가통계로서의 품질 조건을 충족할 경우 국가통계로 승격이 가능하며 통계 생산의 지속 및 폐지가 비교적 자유로움
- 해당 제도는 국가통계에 新기술을 활용할 수 있는 시기를 앞당기는데 의의가 있어 일종의 샌드박스(Sandbox) 제도<sup>29</sup> 범주에 해당
- 국내 통계청 또한 국가통계 승인 대비 완화된 기준에 입각한 시범통계 제도 운영을 구상하고 있어 향후 귀추가 주목

■ [그림 4] (좌)영국의 공식통계 개념도와 (우)한국의 시범통계 도입 구상도



※ 자료 : (좌)Government Statistical Service(2019), (우)정보통신정책연구원(2019)

- 이처럼 AI활용 통계의 국가통계 승인 기회를 개방하는 것은 긍정적이나, 실제 AI도입을 유도하고 활성화하는 측면에서는 해결해야 할 쟁점이 산적

28 GSS Guidance on Experimental Statistics(2019.7.) 인용

29 국민의 생명·안전에 위해가 되지 않는 한 마음껏 도전하고 새로운 시도를 해볼 수 있도록 기회를 부여하기 위한 제도를 의미

- 3장에서 제시한 쟁점들은 국내 국가통계 제도의 개편이 완료되더라도 여전히 해소되기 어려운 AI 자체의 본질적 사안으로서 별도의 해법이 요구됨

■ [표 10] AI·빅데이터 활용 통계 도입의 예상쟁점 및 시사점(3장 참조)

구분	예상 쟁점	시사점
신규 통계 생산	AI·빅데이터 기반의 분석결과에 관한 신뢰가 보편적으로 형성되지 않음	신규 통계 생산 시 활용에 적합한 검증된 알고리즘을 공표하고 이에 대한 세부적인 활용 지침 마련이 요구
	신규 통계의 생산방식은 기존의 통계 생산 프레임워크에 맞추어 해석하기에 적합하지 않음	빅데이터 분석 프로세스를 포괄할 수 있는 통계 프로세스 표준의 개정 고려
	재현 불가능한 비결정론적(Non-deterministic) 알고리즘 기반의 결과를 국가 통계로 관리 가능한지 여부	인공지능 기술의 특성 분석 및 표준화된 평가 기준 정립이 요구됨
통계 생산 프로세스의 현대화	기존의 방법 대비 우수성을 객관적으로 비교가능한가	전통적 통계 기법-AI알고리즘의 비교 검증 방안 발굴이 필요
	AI기술 도입은 통계 생산부터 공표까지의 제한된 생산 기간을 준수할 수 있는 해법인가	다수의 실증 사업을 통한 도입 적합성 진단 필요

- 통계 혁신을 위한 제도 개편과 함께, 활용 촉진을 위한 AI·빅데이터 기술의 대표성 확보가 필요한 시점

## (2) 국가통계 시도입 촉진 방안

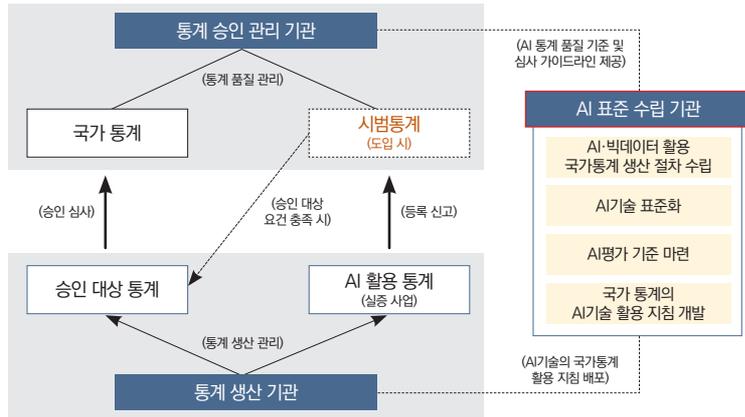
- (제안) 통계청의 정책 변화 방향에 발맞추면서도 실질적인 시도입 성과를 높이는데 필요한 기능과 역할을 수행하는 시표준 수립 기관 지정
- (제안 취지) 향후 국가통계 관리 체계 변화에 대응해 AI기술의 대표성을 부여하고 실증 사업 확대를 유도하기 위함

■ [표 11] 국가통계의 AI기술 도입 촉진을 위해 필요한 요소

구분	시사점	
신규 통계 생산	검증된 AI알고리즘 공표 및 세부적 활용 지침	AI기술의 대표성 부여
	AI·빅데이터 활용 국가통계 생산 절차 수립	
	기술 특성 분석 및 평가 기준 표준화	
통계 생산 프로세스의 현대화	기존 통계기법-AI알고리즘의 비교 검증 방안 발굴	실증 사업의 확대
	다수의 실증 사업을 통한 도입 적합성 진단 필요	

● (실행 방안) 통계승인관리기관-연구기관 간, AI중심의 국가통계 혁신을 위한 협력 체계<sup>30</sup>

■ [그림 5] AI중심의 국가통계 혁신을 위한 협력 체계(안)



※ 자료 : 소프트웨어정책연구소

5 요약 및 시사점

● 유럽 국가들을 중심으로 국가통계의 품질 및 조사환경 개선에 대한 공감대가 형성됨에 따라, 이를 해결할 방안으로서 빅데이터 및 인공지능 기술 도입이 추진 및 검토되고 있음

○ 국가통계의 AI·빅데이터 도입 필요성이 수면 위로 떠오른 주요 원인으로 아래와 같은 사회 현상이 거론

- ① 새로운 유형의 산업이 빠르게 등장함에 따른 산업구조 반영의 어려움
- ② 기업 활동의 다양화(지역에 기반하지 않은 산업, 사무실 연락처 부재 등)
- ③ 조사별 응답자 중복에 의한 회수율 약화
- ④ 즉시 활용 가능한 민간 중심의 데이터의 급증

○ 이를 해결하기 위하여 독일, 네덜란드, 스위스 등 세계 각국의 국가통계 AI·빅데이터 기술 도입 사례가 증가하고 있는 상황임

30 2014년 UNESE 주도의 빅데이터 샌드박스 프로젝트 추진을 위해 아일랜드 통계청(CSO)과 슈퍼컴퓨팅 연구소(CHEC) 간 협력 체계가 구성 된 사례 참고

● 국가통계의 AI·빅데이터 기술 도입은 아래 두 가지 시각으로 구분

- (신규 통계 생산) 조사통계 기법을 활용하지 않고 빅데이터·시기술을 활용해 통계 생산
- (통계 생산 프로세스의 현대화) 기존의 조사통계 프로세스를 수용하되 통계 데이터의 품질 향상, 생산 절차의 효율성 등을 개선하기 위해 빅데이터·시기술을 도입

구분	세부 적용 분야
신규 통계 생산	<ul style="list-style-type: none"> <li>• 대체(Replace) : 기존 통계를 완전 또는 부분 대체하는 수단</li> <li>• 부가(Addition) : 조사 통계 방식으로 도출이 어려운 부문에 대해 데이터 분석에 기반을 둔 우회적 해결</li> <li>• 인사이트(Insight) : 완전히 새로운 분야의 통계 데이터 제시</li> </ul>
통계 생산 프로세스의 현대화	<ul style="list-style-type: none"> <li>• 정확성(Accuracy) : 통계의 정확도 향상</li> <li>• 생산성(Productivity) : 업무 효율화를 위한 자동화 기법</li> <li>• 연결성(Connectivity) : 연결 가능한 데이터들의 상호 보완적 활용</li> <li>• 비용(Price) : 통계 생산의 비용 절감을 위해 활용</li> </ul>

● 국가통계에 시기술 도입 시 예상되는 쟁점들은 아래와 같으며 시기술의 대표성 부여와 실증 사업의 확대 필요성을 시사

구분	예상 쟁점	
신규 통계 생산	AI·빅데이터 기반의 분석결과에 관한 신뢰가 보편적으로 형성되지 않음	시기술의 대표성 부여
	신규 통계의 생산방식은 기존의 통계 생산 프레임워크에 맞추어 해석하기에 적합하지 않음	
	재현 불가능한 비결정론적(Non-Deterministic) 알고리즘 기반의 결과를 국가통계로 관리 가능한지 여부	실증 사업의 확대
통계 생산 프로세스의 현대화	<ul style="list-style-type: none"> <li>기존의 방법 대비 우수성을 객관적으로 비교 가능한가</li> <li>시기술 도입은 통계 생산부터 공표까지의 제한된 생산 기간을 준수할 수 있는 해법인가</li> </ul>	

- (시기술의 대표성 부여) 국가통계 활용 목적의 시기술 개발 및 표준화, 성능 평가기준 마련 등
  - (실증사업의 확대) 다양한 유형의 시활용 통계 생산 실험을 통해 시기술들의 국가통계 도입 적합성 및 실효성 진단
- 국가통계 시도입 촉진을 위해 기술 검증·평가·활용 지침 마련 등을 수행하는 협력 기관 지정을 제안
- 통계승인관리기관·연구기관 간 시중심의 국가통계 혁신을 위한 협력 체계를 구축함으로써, 실질적인 제도 도입의 성과를 상승시키고 국가통계 혁신의 시기를 앞당기는데 기여할 것으로 기대

## 참고문헌

### 1. 국내문헌

- 통계교육원(2015), “국가통계 이해”
- 통계청(2020), “[빅데이터 활용통계] 등 통계 다양성 확대를 위한 국가통계 승인기준 보완방향(안)”
- 통계청(2015), “통계조정업무 매뉴얼”
- 임경은(2012), “조사과정자료 수집 및 활용을 위한 가이드라인 수립 방안”
- 이규근(2018), “표준 자료처리 모델 GSDEMs의 소개 및 시사점”
- 전자신문(2019), “[기고] AI 모델과 AI 기반 시스템, 도대체 어떻게 평가해야 하나”
- 통계청(2019), “빅데이터 활용통계의 국가통계 승인관리방안 연구”

### 2. 국외문헌

- Rademacher, W. J.(2018), “Official statistics in the era of big data opportunities and threats”
- UNECE(2018), “The use of machine learning in official statistics”
- Rob Kitchin(2015), “The opportunities, challenges and risks of big data for official statistics”
- Loison, B., & Kuonen, D.(2018), “Are Current Frameworks in the Official Statistical Production Appropriate for the Usage of Big Data and Trusted Smart Statistics?”
- Martin Beck, Florian Dumpert, Joerg Feuerhake(2018), “Machine Learning in Official Statistics”
- Federal Statistical Office Germany(2019), “Introduction to Big Data in Official Statistics”
- Statistics Netherland(2020), “Discussion paper : Big data in official statistics”
- Eurostat(2017), “ESSnet Big Data Specific Grant Agreement No 1”(SGA-1)
- Edith, desiree de leeuw.(2008), “Mixed mode surveys : When and Why”
- Meertens, Q. A., Diks, C. G. H., van den Herik, H. J., & Takes, F. W.(2018), “A data-driven supply-side approach for measuring cross-border internet purchases”
- Wirth, R., Hipp, J.(2000), “CRISP-DM : Towards a standard process model for data mining”
- Government Statistical Service(2018), “GSS Guidance on Experimental Statistics”
- INFOSTAT Slovakia(2013), “Introducing New Tool for Official Statistics”, Genetic Programming

### 3. 기타

The OECD Glossary of Statistical Terms, <https://stats.oecd.org/glossary/>

통계법, [law.go.kr/법령/통계법](http://law.go.kr/법령/통계법)

Big Data Sandbox, <https://joinup.ec.europa.eu/solution/big-data-sandbox>

RIETI, Can Big Data Change Official Statistics, <https://www.rieti.go.jp/en/events/bbl/19031401.html>

국가통계 기본원칙 전문, [http://kostat.go.kr/portal/korea/kor\\_ko/2/index.action](http://kostat.go.kr/portal/korea/kor_ko/2/index.action)

GPT-2 : 1.5B Release . OpenAI, <https://openai.com/blog/gpt-2-1-5b-release/>

Understanding the GSBPM, <https://statswiki.unece.org/>

Enabling Big Data approaches to gather official statistics, <https://www.ichec.ie/partnerships/public-sector/cso>

Big Data Sandbox, <https://joinup.ec.europa.eu/solution/big-data-sandbox/>



## 사명 Mission

소프트웨어 정책 연구를 통한 국가의 미래전략을 선도함  
Leading Nation's Future Strategy through Research on Software Policy

## 미래상 Vision

국민 행복과 미래사회 준비에 기여하는 소프트웨어 정책 플랫폼  
Software Policy Platform contributing to the public happiness and future society

## 핵심 가치 Core Values

전문성  
Expertise

다양성  
Diversity

신뢰  
Trust

## 역할 Roles

건강한 소프트웨어 산업 생태계 육성  
To build a fair Ecosystem for Software Industry

소프트웨어 융합을 통한 사회 혁신  
To innovate a Society through Software Convergence

국가 소프트웨어 통계 체계의 고도화  
To advance the National Software Statistics System

개방형 소프트웨어 정책 연구 플랫폼 구축  
To establish an Open Research Platform for Software Policy

소프트웨어정책연구소  
Software Policy & Research Institute



이호준 Lee, Ho-Joon  
 언론학박사 ighwns@hanmail.net

우정사업본부에서 근무하고 있으며,  
 대한항공 여행사진 공모전에서 2회  
 수상하고, 세 차례의 개인전과 단체전  
 3회를 개최했다.



경기도 성남시 분당구 대왕판교로 712번길 22 글로벌 R&D센터 연구동(A)

Global R&D Ceneter 4F, 22, Daewangpangyo-ro 712beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do

[www.spri.kr](http://www.spri.kr)



9 772586 502001  
 ISSN 2586-5021