

## 인공지능과 시장경쟁: 데이터에 대한 규율을 중심으로\*

임 용\*\* · 정 해 빈\*\*\* · 고 학 수\*\*\*\*

### 논 문 초 록

최근 들어 커다란 기술적 발전을 보이고 있는 인공지능, 특히 머신러닝 인공지능 기술은, 유용한 훈련용 데이터의 존재를 전제로 한다는 중요한 특징을 지니고 있다. 유용한 데이터가 존재하는지 그리고 이를 이용할 수 있는지 여부에 따라, 기술의 발전은 물론 시장경쟁의 구도가 크게 달라질 수 있다. 따라서 인공지능과 관련된 시장 경쟁의 양태는 기술의 발전으로부터 영향을 받는 문제일 뿐만 아니라, 개인정보 및 데이터에 관한 규율과도 매우 밀접한 관련이 있다. 또한 데이터 및 이를 기반으로 하여 개발된 인공지능은 차별이나 공정성 등과 관련된 새로운 사회적 과제를 제기하기도 한다. 인공지능과 관련된 맥락에서 새로이 제기될 수 있는 차별이나 공정성 이슈 또한 데이터와 관련이 매우 높은 것이다. 인공지능 기술이 시장경쟁을 포함하여 사회 및 경제에 어떤 구체적인 영향을 미칠지에 관하여는 다양한 가능성이 언급되고 있는 상황이다. 효과적인 인공지능 정책의 수립을 위해서는 데이터 정책과의 연관성을 매우 면밀하게 고려할 필요가 있다.

핵심 주제어: 인공지능, 데이터(data), 알고리즘적 차별, 공정성

경제학문헌목록 주제분류: K3, L4, O3

투고 일자: 2019. 9. 16. 심사 및 수정 일자: 2019. 10. 16. 게재 확정 일자: 2019. 10. 25.

\* 본 논문은 한국경제학회 한국경제포럼편집위원회에서 주최한 “AI의 확산! 한국경제 분야별 영향은?” 심포지엄(2019년 8월 28일) 발표원고를 심사자의 논평을 반영하여 수정 보완한 것이다. 이 글은 ‘인공지능과 시장경쟁’라는 제목으로 2019.5에 발표된 서울대학교 법과경제연구센터 인공지능정책 이니셔티브의 이슈페이퍼(이후 ‘인공지능과 시장경쟁’으로 수정하여 워킹페이퍼로 냈)에 담긴 내용을 수정하고 보완하여 작성하였다.

\*\* 제1저자, 서울대학교 법학전문대학원 부교수, e-mail: yonglim@snu.ac.kr

\*\*\* 제2저자, 서울대학교 박사과정, e-mail: shiningsea@snu.ac.kr

\*\*\*\* 교신저자, 서울대학교 법학전문대학원 교수, e-mail: hsk@snu.ac.kr

## I. 논의의 배경

인공지능 기술은 매우 빠르게 발전하고 있고, 다양한 맥락에서 활용도가 늘어나고 있다. 또한 인공지능 기술은 교통사고나 환경오염, 산업재해 등의 위험을 축소시킴으로써 개인의 삶을 개선하는 등 생산성 증대를 넘어 매우 다양한 편익을 가져다 줄 것으로 기대되고 있다(Little Hoover Commission, 2018, p. 8). 인공지능 기술의 활용을 전제로 하는 경제 패러다임의 변화와 함께, 세계 여러 나라의 정부와 정책당국은 적절한 정책방향의 모색을 위한 많은 시도를 하고 있기도 하다(한국정보화진흥원, 2018, pp. 13-24).

그런데 최근 널리 연구되고 있는 머신러닝 인공지능 기술의 특징을 감안할 때, 인공지능 기술은 데이터와 밀접하게 연관되어 있으므로 인공지능 기술의 발전을 위한 정책적 고려에 있어서는 기술 자체에 대한 고려뿐만 아니라 데이터에 대한 고려 또한 매우 중요하다. 어떤 데이터가 존재하고 이용가능한지(availability), 누가 그런 데이터에 대해 접근권한(access)을 보유하고 있는지 등에 따라서 기술의 발전은 물론, 개인정보보호와 관련된 이슈나 시장경쟁과 관련된 이슈 등에도 상당한 영향이 나타날 수 있다. 또한 어떤 데이터를 이용하여 어떤 인공지능 모형을 구축했는지에 따라, 알고리즘적 차별이나 공정성에 관한 이슈가 제기될 수도 있다. 인공지능 정책에 대한 구상은 이러한 여러 효과에 대한 구체적이고 상세한 분석을 필요로 한다(Manoj Kewalramani, 2018, pp. 17-19).

이 글은 인공지능 기술의 발전을 통한 혁신, 데이터에 대한 법적 규율, 그리고 시장경쟁의 관계에 관하여 정리한 뒤 이에 대한 시사점을 도출하고자 한다.

## II. 인공지능 기술 개요

인공지능의 개념을 정의하기 위하여 “인간처럼(humanly)” 행동하거나 사고하는 것이 인공지능이라는 등 개념정의를 위한 다양한 시도들이 있어왔다. 경우에 따라서는 인간처럼 행동하거나 사고한다는 것을 정의하는 것 자체가 논란의 대상이 되기도 한다. 오늘날의 인공지능 연구들은 현실에서 요구되는 수량적 목표에 가능한 “합리적으로(rationally)” 도달한다는 목표에 역량을 집중하는 경향을 보인다(Russell, Stuart and Peter Norvig, 2009, pp. 2-5).

그렇다면 인공지능은 어떻게 개발되어 상용화되고 현실에 적용되는가? 많은 경우

에 인공지능 기술은 여러 전문가들 사이의 협업을 통해 발전하고 현실에 적용된다. 즉, 문제해결에 범용적으로 적용할 수 있는(general problem solving) 독립적인 지적 체계를 구성해내기 위한 시도도 있기는 하지만, 일반적으로는 인공지능이 해결하고자 하는 특정 영역에 대한 과제에 대하여는 해당 영역의 전문지식(domain knowledge)을 고려하여 구체적인 과제의 해결에 도움이 되는 인공지능을 개발하고 고도화하기 위한 시도가 흔히 이루어진다. 이러한 개발과 고도화의 시도는 크게 두 가지 갈래에서 진행되어 왔다.

## 1. 인공지능 기술의 두 갈래

첫 번째는 ‘지식베이스(knowledge base)’ 방식이라 불리는 접근방식이다. 이 방식은 입력된 문제 상황에 대해 답변을 산출하는 판단의 논거로서 사전에 인간에 의해 삽입된 지식을 활용하는 방식이다. 그러한 특징을 반영하여 이 방식은 ‘전문가 시스템(expert system)’이라 불리기도 한다. 즉, 지식베이스 방식은 문제에 대한 판단을 내림에 있어 사람(전문가)들이 사전에 축조해놓은 지식의 편람을 필요로 하므로 제작 과정에서 해당 분야에 대한 전문가들과의 협업이 필수적으로 요구된다. 그러나 지식베이스 접근방식은 개별 판단의 과정에서 기반지식 사이에서 서로 충돌하는 상황이 발생할 경우에, 인공지능이 스스로 기반지식 간의 우열을 판단할 수 없게 되고 그로 인해 외부적으로 인간이 개입해서 기반지식을 업데이트해야 한다는 한계가 있어 왔다.

두 번째는 ‘머신러닝(machine learning)’ 방식이다. 이 방식은 해결하려는 문제 상황에 대해 답변을 산출하기 위한 판단의 논거로, 종래의 유사한 문제 상황들로부터 데이터의 패턴을 습득하여 활용한다. 머신러닝에서 이용되는 기반지식은 전문가들에 의해 인공지능 외부로부터 삽입된 규칙들을 연역적으로 구성한 결과가 아니며, 훈련용 데이터의 패턴을 인공지능 알고리즘이 학습함으로써 인공지능 내부로부터 경험적으로 구성된 것이다. 이처럼 머신러닝 인공지능 기술은 데이터로부터 기반지식을 축적하게 된다. 훈련용으로 최초로 확보한 데이터는 모형 구축을 위해 필요하지만, 그 이후로 새롭게 입력되는 데이터 또한 기반지식을 업데이트하기 위해 활용될 수 있고, 그 결과 지속적으로 상황의 변화를 반영하여 기반지식을 향상시킬 수 있다는 장점이 있다. 상황이 변화하고 그에 따라 판단기준이 변할 경우 매번 그 기준을 외부의 입력을 통해 작위적으로 변경시켜야 하는 지식베이스 접근방식과 달리, 머신러닝은 데이터의 변화 및 업데이트에 따라 자연스럽게 판단기준 또한 변할 수 있도록 고안된 방

식이다(Negnevitsky, Michael, 2005, pp.10-12, 165-168).

## 2. 머신러닝과 딥러닝

머신러닝 기술의 개발에 있어 핵심적인 난관은 데이터를 어떻게 지식으로 전환시키는가의 문제였다. 예를 들어 인공지능이 감시카메라 영상으로부터 절도의 현행범을 포착하는 작업을 상상해보자. 인공지능이 구체적으로 어떤 영상을 포착했을 때 그 영상을 통해 파악한 행동을 절도라고 판단하여 경보를 울려야 할 것인가? 절도행위의 수많은 복잡다단한 태양을 고려할 때, 인공지능 알고리즘이 인식할 수 있는 절도행위의 수량적인 또는 기계적인 단순한 기준을 정하는 것은 그리 간단치 않은 문제일 수 있음을 직관적으로 파악할 수 있다. 이처럼 ‘절도’라는 단일한 개념 안에 현실에서 벌어질 수 있는 여러 다양한 행위들이 포괄될 수 있듯이, 하나의 개념(concept)에 해당하는 여러 데이터 유닛 사이에는 변화나 편차의 가능성(factors of variation)이 존재할 수밖에 없다. 이러한 변화나 편차에도 불구하고, 개별 개념에 적절하게 포섭되고 지식으로 정리될 수 있도록 데이터를 적절하게 표현(representation)하여 알고리즘에 반영하는 것은 쉽지 않은 문제일 수 있다(Goodfellow, Ian, Yoshua Bengio and Aaron Courville, 2016, pp.2-4).

머신러닝 기술의 발전에 있어, 이 난관은 빅데이터(big data) 시대의 도래와 함께 돌파구가 마련되기 시작했다. 빅데이터는 데이터 분석의 양적 발전, 그리고 질적 발전이라는 두 측면에서 바라볼 수 있는데, 양적 측면에서는 다량의 데이터를 한꺼번에 다룰 수 있는 연산능력의 기술적 개선을 들 수 있으며, 질적 측면에서는 데이터 마이닝(data mining) 기술의 발전에 힘입어 종래에는 분석이 어려웠던 이미지, 동영상 등 비정형적(unstructured) 자료들을 처리할 수 있는 기술의 등장을 언급할 수 있다.

데이터 공학 방법론의 발전은 데이터를 처리하는 모델 자체의 개선을 위한 새로운 기회를 마련해주었다. 예를 들어 종래의 머신러닝 기술은 영상에 나온 행동이 절도에 해당하는지 여부를 그 영상 자체로부터 곧바로 정의하려 했다면, 근래에 활용도가 높아진 딥러닝(deep learning) 기술은 판단 대상인 하나의 행동을 여러 단계로 세분화하여 학습함으로써 단계적으로 정의를 내리는 방법을 이용한다. 절도라는 하나의 개념을 걷기, 뛰기 등을 포함한 좀 더 간단한 개념으로 나누어 여러 단계를 거쳐 정의함으로써 절도의 다양한 양태와 단계를 반영하여 유연하게 판단을 내릴 수 있도록 하는 것이다. 그러한 점에서 딥러닝 기술은 세상을 “개념들의 계층(hierarchy of

concepts)”으로 표현하는 기술이다(O’Leary, Daniel E., 2013, p. 97).

머신러닝 기술의 발전과 함께 관련된 연구도 크게 활성화되고 있다. 예를 들어, 학술 데이터베이스 Scopus에 등재된 인공지능 관련 연구 중에서 머신러닝과 확률적 추론 분야에 관한 논문은 2010년에 28%에서 2017년에는 56%에 육박하는 수준으로 증가하였다(Shoham, Yoav et al., 2018, p. 9).

### 3. 딥러닝 기술의 한계

딥러닝 기술은 활용도가 계속 높아지고 있다. 이 기술은 결국 문제해결 또는 판단에 필요한 지식을 데이터로부터 학습한다는 점에서 머신러닝 기술의 기본적인 골격은 바꾸지 않은 채, 데이터의 학습을 여러 층(layer)을 통해 할 수 있도록 하여, 성능을 크게 개선시킨 기술이라고 할 수 있다.

그런데, 딥러닝 기술을 이용하여 판단이 이루어지는 과정은 결국 데이터로부터 수집하고 분석한 패턴이 가장 핵심적인 것이다. 예를 들어, 어떤 딥러닝 모델이 후두암 환자들과 관련된 의료데이터를 분석한 결과 X라는 요소를 지닌 사람이 후두암이 걸릴 가능성이 높다고 판단을 했다면, 이는 X라는 요소가 후두암으로 연결되는 의학적 메커니즘이나 원인을 규명해낸 것이 아니라, X라는 요소를 지닌 사람에게 후두암이 발병한 사례가 많이 보인다는 통계적 근거에 기초한 판단에 그칠 따름이다. 딥러닝 인공지능의 판단은 흔히 불투명(opaque)하다고 언급되는데, 그 이유는 개별 층(layer)에서 어떤 유형의 작업이나 판단이 이루어지는지 파악하기 어렵다는 점 그리고 판단에 쓰이는 통계적 논거들이 비선형적이고 복잡하다는 점 등에 기인하는 것이다. 딥러닝 기술은, 일반적으로 인과관계(causality)의 분석에는 어려움이 나타날 가능성이 높고, 다른 한편 매우 다양한 독립변수를 통한 통계적 상관관계(statistical correlation)를 분석하는 데에는 매우 유용한 방법론일 수 있다.

개별 영역에서 딥러닝 기술이 좋은 성과를 내기 위해서 중요한 핵심적인 요건은 데이터이다. 대중매체를 통해 많이 알려진 구글의 알파고(Alpha Go)가 바둑에서 우수한 성능을 낼 수 있었던 배경에는, 바둑이라는 영역에서 기존 기보를 활용하여 훈련용 데이터를 확보하고 구축하는 것이 어렵지 않고, 또한 달성해야 할 최적화를 위한 목표가 명확했다는 점 등이 있다. 바둑의 경우, 해당 게임 내에서 허용되는 가능한 선택상황이 구체적으로 정해져 있고 승리 여부에 대한 판단 기준 및 목표가 매우 명확하다. 인공지능 개발의 맥락에서는, 기존 기보를 어렵지 않게 훈련용으로 활용할

수 있을뿐더러, 더 나아가 가상의 기보 데이터를 직접 생성해내는 것도 가능하게 되었고, 이로부터 좋은 성과를 내는 인공지능의 개발이 가능하게 되었다.

기술이 발전함에 따라 과거 바둑 경기의 실제 기보를 학습하는 것을 넘어서, 인공지능끼리 경기하도록 하여 새로운 기보를 추가 데이터로 생성하고, 이를 학습에 다시금 반영하는 등, 데이터를 대량으로 추출하고 이를 학습하여 판단에 반영하는 피드백이 가능해 지기도 하였다. 이러한 과정을 거쳐, 바둑에서의 딥러닝 알고리즘들은 매우 우수한 문제해결 능력을 가질 수 있게 되었다(Silver, David et al., 2016, p.486). 다만, 사회나 경제 이슈와 관련된 데이터는 적지 않은 경우에 사람의 행태에 관한 내용을 다루는 것이어서, 바둑의 경우처럼 데이터 자체를 인공적으로 생성해 내는 방식을 이용하는 데에는 한계가 있다. 인공 데이터는 사람의 실제 행태를 반영하는 것이 아닐뿐더러, 통계적 편향이나 차별 등과 관련된 논란을 불러일으킬 수도 있다.

딥러닝을 이용하여 현실적으로 유용한 결과를 얻기 위해서는 몇 가지 중요한 전제가 충족되어야 한다. 첫째, 인공지능의 판단에 필요한 지식의 습득이 가능하기 위해서는 충분한 데이터가 확보되어야 한다. 이는 양(quantity)의 문제일 뿐 아니라 질(quality)이나 다양성의 문제이기도 하다. 양이 많더라도, 확실적인 정보가 담긴 데이터나 충분히 상세(granular)하지 않은 데이터가 주어져 있다면, 유용한 인공지능 모형의 구축은 어려워 질 수 있다. 예를 들어, 엑스레이 영상 이미지를 이용하여 진단용 인공지능 모형을 구축하고자 하는데, 훈련용으로 이용할 수 있는 이미지 데이터가 모두 건강한 사람의 데이터라면 효과적인 인공지능 모형의 구축은 어려워진다. 인공지능은 건강한 사람의 엑스레이 이미지와 건강하지 않은 사람의 엑스레이 이미지 사이의 차이를 구분하는 과정을 통해 그 기능이 개발되고 고도화되는 것이기 때문이다. 둘째, 최적화의 목표가 분명하고 수량적 내지 수치적으로 규정 가능해야 한다. 사전적으로 목표를 정해놓지 않는 방식의 인공지능도 있지만, 현실적으로 활용도가 높은 것은 목표를 정해놓는 것이고, 특히 사회나 경제적 문제를 대상으로 하는 인공지능은 대체로 최적화의 목표가 명확할 필요가 있다. 셋째, 문제해결 또는 판단의 대상이 보이는 특징이 훈련용 데이터가 일반적으로 보이는 특징에 비추어 크게 이례적이지 않아야 한다(Marcus, Gary, 2018, pp.12-14; Forbes Insight, 2019. 3. 27.).

#### 4. 데이터의 문제

딥러닝 기술이 사회적으로 또는 산업적으로 의미있는 성과를 내기 위해서는 유용한

훈련용 데이터의 존재가 필수적이다. 데이터에 관한 이슈를 좀 더 구체적으로 파악해보면, (i) 훈련용 데이터로 이용할 수 있는 풍부하고 다양한 데이터의 확보 및 이용가능성, 즉 데이터의 가용성(availability)이 중요한 요소가 된다. 또한, (ii) 인공지능을 이용하여 해결하려는 문제 자체가 알고리즘을 이용하여 최적화가 용이한 유형의 문제인지 여부에도 영향을 받게 된다. 이를 고려하면, 인공지능에 기반한 기술개발 및 혁신의 사례들을 아래와 같이 4가지의 유형으로 개념화하여 분류해 볼 수 있다.

데이터가 부족함	데이터가 충분함	
<p><b>유형 II</b></p> <p>문제해결에 필요한 통상적 데이터가 부족한 상황에서 대안적 데이터를 통해 문제해결을 시도</p>	<p><b>유형 I</b></p> <p>문제해결에 필요한 데이터의 가용성(availability)을 확보한 상황에서 유용한 인공지능 모델 도출</p>	<p>목표의 최적화가 용이함</p>
<p><b>유형 III</b></p> <p>혁신 달성 어려움</p>	<p><b>유형 IV</b></p> <p>풍부한 데이터를 바탕으로 문제해결을 위한 대안적 목표를 발굴하여 유용한 결과를 도출하기 위한 시도</p>	<p>목표의 최적화가 어려움</p>

(1) **유형 I** : 해결하려는 문제에 대한 목표의 최적화가 용이하고, 이에 필요한 데이터의 가용성도 충분히 확보되는 경우에, 인공지능 기술을 적용하여 유용한 결과를 도출하는 것이 가능하다. 몇 가지 예시를 들자면, (1) 보험처리 데이터를 바탕으로 전형적인 상해로 인한 보험금 지급 결정을 자동화하는 사례(이응용, 2018, pp.1-9), (2) 의료 데이터를 기준으로 환자를 신속하게 진단·분류할 수 있는 의료 인공지능의 개발(CADTH, 2018, pp.5-11; KPMG, 2018, pp.12-16; AHSN Network, 2018, pp.24-25), (3) 물류 데이터를 이용하여 유통 및 보관비용을 절감하는 인공지능의 도입(Deloitte, 2018, pp.15-25; Mayor of London, 2018, pp.68-69), (4) 행정 데이터를 이용하여 청소년기 학생들의 중퇴가능성을 판단하는 인공지능(Freuler, Juan O., and Carlos Iglesias, 2018, pp.15-25) 등을 들 수 있다. 이러한 사례들은 대체로 정형화된 데이터를 최적화가 용이한 목표에 대한 추론과 판단에 활용하여 예측능력의 개선이나 비용절감 등을 추구한 것이다. 물론, 이 유형의 경우에 데이터가 충분

한 것을 전제로 분류를 하였지만, 실제 개별 상황에서 데이터의 가용성이 확보될 수 있는 것인지는 해당 영역에 적용되는 법제도를 비롯한 다양한 요소에 따라 다를 수 있다.

(2) **유형Ⅱ** : 이 유형은 해결하려는 문제에 대한 목표의 최적화 자체는 용이하지만, 이에 필요한 데이터가 일반적으로 부족하거나 가용하지 않은 경우이다. 이러한 경우에는 유사한 효과를 내는 다른 대안적 데이터를 이용할 수 있는지 여부가 관건이 된다. 예를 들어, 전통적 은행업의 미발달로 인해 여신거래 내역 등 신용거래 데이터가 부족한 경제에서, 통화내역 또는 스마트기기 접속기록 등 새로운 종류의 데이터를 소비자 신용정보에 편입하여 신용도 평가에 활용하는 경우를 생각해 볼 수 있다. 더 넓게, 일반적으로 ‘썬파일러(thin filer)’에 대해, 대안적이거나 보완적인 데이터를 확보하여 좀 더 정확한 신용도 평가를 하는 유형의 시도가 이러한 유형의 사례가 될 것이다. 더욱 구체적인 사례의 또 다른 예로, 기존의 수도관 중에서 납이 함유된 수도관이 매장된 위치가 어디인지를 일일이 발굴하여 조사하고 확인하는데 비용이 상당히 많이 들기 때문에, 이에 대한 보완적인 방법으로 주택의 건설시점이나 건축 형태 등에 대해 인공지능 분석을 하여 확인 후보지를 좁히는 경우를 들 수 있다(McKinsey Global Institute, 2018, pp.18-27).

(3) **유형Ⅲ** : 해결하려는 문제에 대한 목표의 최적화가 난해한 상황에서, 일반적으로 가용한 데이터 또한 부족한 경우, 문제의 해결을 위해 인공지능이 기여를 하기 쉽다. 이 유형의 상황에서는, 적용범위나 대상조차 정의되어 있지 않은 전혀 새로운 기술의 개발과 같이, 기존의 사례로부터 유사한 패턴을 발견하여 찾아가기보다는 창의성과 연역적 논리가 중요한 사례의 발굴을 위한 시도가 이루어질 수 있다.

(4) **유형Ⅳ** : 해결하려는 문제에 대한 목표의 최적화가 난해한 상황에서, 일반적으로 가용한 데이터는 상대적으로 풍부한 경우, 불분명해 보이는 목표를 명확한 수량적 지표로 근사하는 방식을 통해 인공지능 모형을 개발하기 위한 시도를 할 수 있다. 한 예를 들자면, 프로 스포츠 선수들의 연봉협상과정에서 선수들의 능력을 좀 더 정확하게 판단하기 위해 딥러닝 기술을 활용하여 종래에는 사용되지 않던 새로운 지표를 개발하기 위한 시도가 이루어질 수 있다(Gent, Edd, 2018. 6. 13.).



(5) **시사점** : 위와 같은 분류를 통해 몇 가지 시사점을 얻을 수 있다. 첫째, 현실적으로 가능한 기술개발과 가능성이 높지 않은 기술개발을 구분하여, 합리적인 판단을 할 필요가 있다. 인공지능의 성과는 최적화가 용이하고 데이터가 풍부한 유형 I의 영역에서 극대화될 수 있다. 그 반대로 데이터의 가용성이 낮고 최적화 또한 쉽지 않은 유형Ⅲ의 영역에서는 인공지능 기술을 활용하여 기초연구를 할 수는 있겠지만 응용연구를 통해 사회적으로 곧바로 활용될 수 있는 결과를 도출하는 것은 어려울 수 있다. 유형 I의 경우에도, 개별 영역이나 개별 사안에 따라서 데이터의 가용성에 커다란 차이가 있을 수 있어서, 이에 관한 추가적인 고려가 필요할 수 있다.

둘째, 위와 관련하여, 정책적 과제의 측면에서는 필요한 데이터의 가용성을 어떻게 확보할 것인지에 상당한 관심을 기울일 필요가 있다. 필요한 데이터의 확보는 인공지능 기술의 개발을 위해 기본 전제가 되는 핵심적인 과제이다. 다른 한편, 유용한 데이터의 확보가 쉽지 않은 경우를 어렵지 않게 접할 수 있다. 우선, 필요한 데이터 중에서 개인정보 유형의 데이터가 포함된 경우에는, 개인정보 보호법의 규율을 비롯하여, 프라이버시 보호와 관련된 우려로 인해 데이터의 가용성을 확보하는 것이 어려울 수 있다.

개인정보보호와 관련된 이슈와는 별개로, 데이터에 대한 접근에 현실적인 제약이 있는 경우도 흔히 생각할 수 있다. 데이터는 일반적으로 비경합성을 보이는 등 부분적으로 공공재와 유사한 측면도 있지만, 데이터를 보유한 당사자가 이에 대해 제3자의 접근을 허용하지 않을 수도 있다. 이는 개인정보 보호법 등 법적 제약으로 인한 경우도 있을 수 있지만, 그와 관계없이 데이터 보유 자체가 경쟁에 영향을 미칠 것이라는 판단하에 배타적인 보유를 원하기 때문일 수도 있다.

셋째, 유형Ⅱ의 경우 좀 더 면밀한 정책적 검토가 필요할 수 있다. 데이터의 가용성이 충분히 확보되지 않은 상황에서 대안적 데이터를 확보하여 유용한 결과를 도출해 낸 것이 상당히 의미있는 성과라고 볼 수도 있지만, 다른 한편, 만일 데이터의 가용성이 높지 않게 된 것 자체가 개인정보보호를 포함한 법정정책적 판단을 반영한 것이라면, 대안적 데이터를 확보하여 유용한 결과를 도출해 냈다는 것은 그러한 법정정책적 판단을 무력화하는 것일 수도 있다. 따라서, 유형Ⅱ의 경우에는, 기존에 알려져 있지 않았던 데이터를 발굴하여 새로운 결과를 만들어낸 경우도 있을 수 있지만, 사회적인 차원의 정책적 판단을 우회하는 경우도 있을 수 있다. 전자의 경우에는 일반적으로 긍정적인 의미를 부여할 수 있겠지만, 후자의 경우에는 대체로 긍정적인 의미를 부여하기 어려울 것이다.

이러한 쟁점들을 고려할 때 데이터에 대한 규율을 어떻게 할 것인지는 인공지능에 대한 정책적 판단에 있어 핵심적인 영역이 될 수밖에 없다. 아래에서 데이터에 관한 규율을 어떻게 해야 할 것인지 좀 더 구체적으로 살펴본다.

### Ⅲ. 데이터에 관한 규율

데이터에 관한 규율은, 개인정보보호를 포함하여 매우 다양한 사회적, 법적 규범과 복잡하게 연관되어 있다. 어떤 유형의 법제도에 대해 어떤 원칙을 고려하여 규율 체계를 마련하고 개선을 모색할 것인지 본다.

#### 1. 데이터의 가용성

정책적인 측면에서는, 우선 공공재적 성격이 상대적으로 명확한 유형의 데이터에 대해서는 그 공급이 최대한 원활해지도록 하는 법제도를 구상할 수 있다. 한국은행이나 통계청 등 경제관련 데이터를 오랫동안 다루어온 기관의 데이터나 기상청의 날씨 데이터 등 일부 정부나 공공기관을 통한 데이터의 수집과 제공은 정부의 당연한 역할의 일부로 이미 오래전부터 인식되어 왔다. 또한, 공공데이터의 제공 및 이용 활성화에 관한 법률 등 법령을 통해 ‘공공데이터’의 제공이 좀 더 원활해지도록 하는 시도도 있어왔다. 전반적으로는 공공영역에서의 데이터 가용성 제고를 위한 노력을 기울일 필요가 있지만, 다른 한편, 이러한 맥락의 정부의 역할에 관하여 인공지능 기술을 고려할 때 어떤 추가적 고려가 필요할 것인지에 관해서도 생각할 필요가 있다.

첫째, 정부실패(government failure)의 문제가 발생할 수 있다. 별다른 규제나 제한이 없을 경우에 민간의 개별 기업들이 스스로 데이터를 수집하고 활용할 수 있었을 상황이라면, 정부가 개입하여 데이터를 일괄적으로 공유하도록 법제도를 만드는 것은 시장을 왜곡시키는 결과를 가져올 가능성이 있다. 특히, 이 과정에서 어떤 데이터를 공유할 것인지 그 종류, 구체적인 내용, 속성 등에 대해 정부가 일방적으로 결정하여, 정부가 정한 방식이나 데이터 구조를 전제로 수집과 공유를 하도록 한다면, 그러한 왜곡의 가능성은 더욱 커질 수 있다. 또한 이 과정에서 표준화(standardization)이나 상호운용성(inter-operability) 등과 관련한 추가적인 이슈가 발생할 수도 있다.

둘째, 데이터의 품질(quality)이나 유용성과 관련된 문제가 발생할 수도 있다. 공공데이터의 제공을 강제할 경우에, 데이터가 모이는 곳은 개념적으로 공유지

(commons)와 유사한 기능을 수행하게 되면서, ‘공유지의 비극’과 유사한 상황이 발생할 수도 있다. 개별 조직의 입장에서는 ‘데이터 공유지’에 제공하는 데이터는 좋은 품질의 데이터가 아닐 수도 있고 유용성이 높지 않은 데이터일 수도 있다. 오히려 좋은 품질의, 유용성이 높은 데이터는 가급적 조직 내부에 남겨놓고 싶은 유인이 있는 것이 보통일 것이어서, 결국 ‘데이터 공유지’에 취합되는 데이터는 대체적으로 품질도 좋지 않고 유용성도 떨어지는 데이터일 가능성이 있다.

셋째, 공공영역에서의 데이터 가용성 개선은 프라이버시 문제에 대한 고려를 포함한 기존 데이터 규율 체계의 기본원칙과 양립가능한 선에서 가능할 것이어서, 법제도의 여러 제약을 고려한 실질적인 개선책을 찾기 위해서 상당히 복잡한 고려가 필요할 수 있다. 사람에 관한 데이터의 경우에, 기존에 공공영역에서 제공되던 데이터는 인구총조사나 국민계정 등에서 볼 수 있는 것과 같이 총계(aggregate) 데이터인 경우가 많다. 하지만 인공지능 기술의 개발을 위해 요구되는 데이터는 개인 단위의 데이터인 경우가 적지 않다.

개인 단위의 데이터는, 이를 분석하여 다양한 인공지능 기술의 개발로 이어질 수 있다. 공공영역에서는 예를 들어 사법당국의 법집행이나 행형에서부터 민간영역의 다양한 상품 제공에 이르기까지 무궁무진한 잠재적 혁신과 활용을 가져올 수 있다. 하지만, 이 영역에서의 기술개발은 그와 동시에 프라이버시 보호나 차별금지 등 중대한 사회적 이익과 충돌할 가능성이 있기 때문에, 매우 세심한 고려가 필요하게 된다. 이처럼 개인에 관한 데이터의 활용이 가져올 잠재적 유용성을 고려하는 동시에, 그 분야에 관한 또 다른 사회적 이익을 고려하여 반복적으로 조율과 조정의 과정을 거치는 것이 필요할 수 있다. 이러한 유형의 상황은, 현대 금융시스템을 고려할 때 핵심적으로 중요한 데이터 산업 중 하나인 신용평가(credit scoring) 산업, 특히 개인 신용평가제도의 예를 통해 살펴볼 수 있다.

기업에 대한 신용공여와 비교할 때 개인에 대한 소매여신은 채권시장에서 쉽게 거래되지 않고, 실제로 소매여신에 대한 위협의 평가모형은 여신의 시장가치에 대한 추정보다 부도가능성의 측정에 초점을 맞추는 경향이 있다(이균희, 2004, pp.216-217). 문제는 개인의 부도가능성을 어떤 자료들을 근거로 측정할 것인지에 관한 것이다. 개인의 부도가능성 판단에 제약이 있을 수 있는 것은, 측정에 도움이 될 수 있는 데이터의 수집과 확보 자체가 곤란할 수도 있기 때문에 해결이 어려울 수 있다. 이 문제의 해결을 위해 선진국들은 개인신용평가 제도를 도입하였는데, 그 구체적인 방식은 개별 국가에 따라 역사적으로 상당히 상이하게 발전되어 왔다.

미국의 경우를 보면, 개인신용 데이터의 수집과 관리에 대해 국가의 개입을 최소화하고 그 역할을 민간 기업들에게 허용하였다. 이를 통해 여러 신용조회회사(credit bureau, CB)들이 경쟁적으로 개인신용평가를 주도하고 이를 고객들에게 공개해왔다. 이는 경쟁을 통해 가장 정확한 신용도 측정을 유도할 수 있다는 정책적 판단이 암묵적으로 전제된 것이라 할 수 있다. 다만, 공정신용보고법(Fair Credit Reporting Act)를 비롯한 여러 법령과 규제를 통해, 공정성 확보나 차별금지 등의 사회적 가치를 반영하도록 하고 있고, 이는 구체적으로는 성별, 종교, 인종 등 일부 유형의 정보에 대해서는 신용평가의 과정에서 고려하지 못하게 하는 방식을 통해 구현되기도 한다.

유럽은 이와 다른 방식을 택하였다. 전반적으로는 개인신용평가의 역할이 좀 더 제한적이고 더 세밀한 규제의 대상이 된다. 예를 들어, 프랑스는 부실채권에 대한 거시경제적 관점의 관리가 중요한 정책적 목표로 책정되었고, 소매금융의 맥락에서는 개별 금융소비자에 대한 상세하고 차별적인 평가가 강조되지 않는 방향으로 신용정보의 수집과 분석이 이루어지게 되었다. 그로 인해, 미국에서의 소매금융은 개인고객에 대한 분석에 따라 여신제공 여부는 물론 이자율을 포함하여 여신의 형태와 조건이 크게 달라질 수 있지만, 프랑스 소매금융의 경우에 개인고객에 대한 판단은 여신을 제공할 것인지 여부에 대한 판단에 있어 중요할 수는 있지만 구체적인 여신 조건에 대해서는 상세한 분석이나 개인별 차별화된 의사결정이 이루어지지 않을 수도 있다(Rona-Tas, Akos and Alya Guseva, 2018, pp.62-65).

개인신용평가제도에 대한 미국과 프랑스의 간략한 비교는, 데이터 수집 및 처리에 대한 법제도적 판단에 있어 매우 다양한 사회적 가치가 고려될 수 있고, 그 결론이 상당히 달리 나타날 수 있음을 보여준다. 단순화해서 본다면, 개인신용평가 맥락에서의 사회적 가치는 넓게 두 가지 유형의 가치 사이의 사회적 판단을 반영하는 것으로 볼 수 있다. 즉, (i) 가급적 상세하고 풍부한 데이터를 확보하여 개개인에 대한 정확하고 세밀한 신용도 평가를 하고, 이를 금융기관의 의사결정에 적극적이고 정확하게 반영하는 것, 그리고 (ii) 그 과정에서 상세한 개인 수준 데이터를 수집하고 반영하는 것으로 평가될 경우에 나타날 수 있는 프라이버시 보호나 금융시장에서의 차별 논란 등과 관련된 문제의 가능성을 차단하는 것이 그것이다. 이 두 가지 가치는 일반적으로 서로 상충되는 것일 수 있는데, 그 사이에서 어떤 사회적 균형을 모색할 것인지는 개별 국가에 따라 판단이 다를 수 있다.

개인신용평가에 대하여는 개인에 대한 여신시장을 어떻게 관리할 것인가가 정책적

고민의 출발이었듯이, 보건의료, 교육, 사법, 인사, 행정 등 각기 다른 영역에서는 각기 다른 정책적 고민과 고려가 필요할 수밖에 없다. 그 과정에서, 데이터의 확보와 공유에 관한 대원칙을 개별 영역에서의 정책적 목표에 어떻게 투영하여 어떤 결론을 도출할 것인지는 일의적으로 사전에 정할 수 없다.

## 2. 차별과 프라이버시

데이터 규제 of 일반적 원칙을 지탱하는 뼈대는 개인정보와 프라이버시 뿐만이 아니다. 최근 부상하고 있는 또 다른 중대한 사회적 쟁점은 인공지능 알고리즘이 사회적 차별을 강화할 수 있고 공정성을 해칠 수 있는 가능성과 관련된 것이다. 알고리즘적 차별(algorithmic discrimination)이라 불리는 이 새로운 쟁점은, 시장에서의 개인에 대한 가격차별로 인해 발생할 수 있는 소비자 잉여의 이전으로부터, 개인에 대한 강화된 프로파일링에 기초하여 발생할 수 있는 사회적 차별에 이르기까지 폭넓은 범위가 포함될 수 있다(Sartor, Giovanni, 2018, pp. 4-6). 이로부터 더 나아가 궁극적으로는 인공지능에 의한 개인의 자율성 침해와 그로 인한 민주주의의 위기의 도래 등 중대한 사회적 문제의 가능성까지 언급되기도 한다(Reed, Chris, Elizabeth Kennedy and Sara Silva, 2016, pp. 19-21). 이 맥락에서의 구체적인 예로, 미국의 온라인 유통기업 아마존의 고용 인공지능 개발 사례를 들 수 있다. 이 회사는 딥러닝 기술을 개발하여 회사 지원자의 지원서류의 분석을 통해 1단계 판단을 하는 모형을 수 년 간 개발해 왔으나, 인공지능이 여성적 어휘를 사용한 지원자를 차등적으로 판단하는 문제가 지속되어 결국 개발 자체를 포기한 바 있다(Dastin, Jeffrey, 2018, 10. 10.).

이처럼 딥러닝 인공지능이 개인에 관한 데이터를 기반으로 상용화되는 경우에 사회적으로 문제될 수 있는 여러 새로운 상황이 나타날 수 있다. 통계적 왜곡의 가능성 맥락에서는, 개인에 관한 데이터를 수집 및 처리함에 있어서 넓게 두 가지 유형의 문제가 있을 수 있다. 첫째, 구체적인 항목 설계의 타당성, 그리고 둘째, 평가자의 의도에 관한 것이다. 첫째 문제는 어떤 구체적인 항목을 수집할 것인지에 관한 것으로, 여기에는 개별 항목 유형의 선별 및 항목별 상세성(granularity) 수준도 포함될 수 있다. 둘째는 개별 항목에 대하여, 왜 그 항목을 수집하는 것인지의 목적에 관한 것이다. 고용 알고리즘의 예를 들자면, 비효율적인(즉, 평가의 정확성을 높이는 데에 있어 유용한 정보를 제공해주지 않는) 평가항목임에도 불구하고 그것이 조직 내부에서

중요한 성과지표로 인식되는 경우에, 우수한 인력자원들로 하여금 해당 평가항목을 의식적으로 고려하도록 유도할 수 있고, 다른 한편, 평가자가 편향이나 편견을 지니고 특정한 평가항목을 설계한 경우에, 해당 항목의 고려가 불필요하거나 왜곡을 가져올 수 있는 것임에도 불구하고 정상적인 판단지표처럼 작동할 수도 있는 것이다.

그러나 딥러닝 기술은 데이터 그 자체를 판단지식의 근거로 삼는다는 점에서, 훈련용 데이터를 미리 정밀하게 분석하지 않는 이상, 딥러닝을 통한 판단에 왜곡이 있더라도 왜곡이 존재하는지 자체를 인식하는 것이 어려울 수 있고, 오히려 기존의 훈련용 데이터에 내재되어 있던 왜곡을 더욱 체계적으로 학습하게 될 가능성도 있다. 인공지능을 통한 차별은 단순히 행위자의 개인적 편견에 의존하는 것이 아니라 집합적 데이터를 바탕으로 차별을 재생산한다는 점에서 “경로의존적(path-dependent)”으로 차별에 기여하는 것이라 할 수 있다(Williams, Betsy A., Catherine F. Brooks and Totam Shmargad, 2018, pp. 91-95).

데이터의 수집과정에서 이미 차별이 누적적으로 반영되어 있는 경우뿐만 아니라, 차별과 관련된 맥락에서 문제가 될 수 있는 데이터가 훈련 데이터에 희소하게 나타나는 경우에도 인공지능이 부적절한 처리를 할 수 있다. 예를 들어 특정 소수집단에 관한 정보가 포함된 데이터로 훈련하는 인공지능은, 이러한 ‘소수’의 정보를 데이터의 패턴으로 인식하여 판단지식에 반영하기보다는 통계적인 아웃라이어(outlier)로 여겨 무시해버릴 수 있다. 이러한 과정을 통해 개발된 인공지능 알고리즘들은 훈련용 데이터의 다수를 차지하는 사례를 일반화하여 판단을 내릴 것이므로, 결국 소수집단에게 부당한 판단을 강요하는 결과를 가져올 수 있다.

알고리즘적 차별이 나타나게 되는 가장 기본적인 원천은 데이터이다. 우선, 통계적인 모집단(population)에 해당하는 데이터에는 사회적인 또는 역사적인 차별이 이미 반영되어 있을 수 있다. 그러한 경우에는, 훈련용 데이터 구축을 위한 샘플링 과정에서 모집단의 통계적 특성을 정확하게 반영하는 것이 오히려 알고리즘적 차별의 문제로 귀결될 수 있다. 다른 한편, 이러한 문제의 극복을 위해 훈련용 데이터 구축 과정에서 작위적인 판단을 하여 데이터를 구축하거나 수정한다면, 이 또한 논란의 대상이 될 수 있다. 이와 관련하여, 인공지능이 차별적 판단을 할 경우 그 원인이 무엇인지 판단하기 위해, 또는 그러한 차별적 판단의 가능성을 예방하기 위해, 데이터 자체를 사전에 정밀하게 분석하는 것이 대안으로 제시될 수도 있다. 하지만 데이터를 분석하는 것은 많은 경우에 프라이버시 보호의 대원칙에 반하는 것이 될 가능성이 높다. 다른 한편, 개인정보가 포함된 데이터에 대해 비식별화(de-identification) 기법을 적용

하는 것이 언급되기도 하지만, 아직까지는 인공지능 맥락에 있어 비식별화의 역할은 제한적이다(고학수 외, 2017). 개인정보 보호법상 ‘민감정보’라 분류되는 정보나 그 이외에도 인종, 종교 등 민감성이 높은 유형의 정보를 수집하는 것이 차별을 강화할 수도 있다는 의견과 그러한 정보를 수집해야 데이터를 분석하여 사후적으로라도 차별을 억제할 수 있다는 의견 등 다양한 견해들이 제시되고 있다(고학수·정해빈·박도현, 2019, pp. 242-265).

#### IV. 데이터에 관한 규율과 시장경쟁

데이터 수집에는 비용이 소요될 뿐만 아니라, 프라이버시 보호라는 견고한 법적 규율이 존재하고 있다는 사실을 고려한다면, 상업적 활용가능성이 높은 ‘개인에 관한 데이터’는 시장에서 기업간 거래로 취득되기보다는, 일반적으로는 충분한 자본을 가진 기업들이 일정한 비용을 지출하면서 개별 소비자로부터 직접 취득하게 될 가능성이 높다. 오늘날 수많은 인터넷 기업들이 제공하고 있는 무료 이메일과 클라우드 서비스들은 그 비용의 또 다른 면모에 불과하다고 볼 수도 있다.

그렇다면 이러한 경향이 혹시 시장의 과독점화를 촉진하여 시장실패로 이어지는 것은 아닐까? 달리 말해, 프라이버시 보호나 차별 금지 등을 이유로 한 데이터에 대한 다양한 규율이 반경쟁적(anti-competitive)인 요소를 지니고 있는 것은 아닐까? 혹은, 데이터에 대한 개인의 통제력(ownership)을 강화하는 방식을 통해 개인의 권리를 신장하면서도 경쟁을 강화할 수 있지 않을까? 데이터와 경쟁의 문제는 경쟁당국과 산업현장에서는 이미 깊이 있는 분석과 논의를 필요로 하는 심각한 이슈로 받아들여지고 있다.

시장경쟁과 데이터에 관한 쟁점은 이미 미국과 유럽에서 수많은 행정적 처분을 이끌어내는 등 현실적인 법적 쟁점으로 나타나고 있다. 크게는 ① 디지털 사업자의 특수성, ② 데이터 관련 규제가 시장경쟁에 미치는 영향, ③ 합병이 혁신과 소비자 후생에 미치는 영향, ④ 알고리즘에 의한 담합 등 여러 이슈가 논의되고 있다(UK Digital Competition Expert Panel, 2019, pp. 8-12). 여기에서는 그 중에서도 가장 현실적으로 논의가 많이 되고 있는 앞의 두 쟁점을 언급하고자 한다.

## 1. 디지털 사업자의 특수성

디지털 사업자의 특수성에 관해서는 우선 아마존, 구글, 페이스북 등 플랫폼(platform) 사업자들이 경쟁법에서 우려할만한 시장지배력을 실제로 보유하고 행사할 수 있는지 여부가 논란이 되고 있다. 논란이 발생할 수 있는 기본적인 이유는, 플랫폼 사업자가 플랫폼 이용자들의 데이터를 수집하고, 이러한 데이터를 관련 플랫폼의 서비스 향상 및 관련 인공지능 기술의 개발과 최적화에 활용함으로써 이용자들이 느끼는 플랫폼의 매력을 증진시키고 경쟁력을 더욱 강화하는 형태로 ‘네트워크 효과(network effect)’가 발생할 수 있으며, 그 결과 이용자들의 ‘전환 비용(switching cost)’이 증가하고 관련시장에서의 진입장벽이 높아질 수 있기 때문이다.

그러나, 기존 사업자들이 서비스 제공의 과정에서 이용자들의 데이터를 대량으로 보유한다는 것이 실제로 진입장벽을 형성하고 반경쟁적인 효과를 발생시킬 수 있는지에 대해서는 논란이 있다. 데이터센터를 운영하는 것은 비용이 소요되는 것이고, 상업적 활용가능성이 낮은 데이터를 무작정 대량으로 보관하는 것은 회사의 수익성을 저해할 뿐이다. 더구나 대규모 데이터에 대해 그때그때의 필요에 따라 신축적이고 유연한 제공이나 활용이 가능한 형태로 관리를 하는 것은 상당한 자원과 비용이 소요된다. 뿐만 아니라, 데이터가 많다고 하여 그에 비례하여 서비스의 질이 항상 개선되는 것은 아니며, 데이터를 관리하고 분석할 수 있는 기술력 등 경쟁에 영향을 미치는 다양한 요소들을 함께 고려할 필요가 있다. 데이터라는 재료가 아무리 좋아도, 이를 수익모델로 연결시킬 수 있는 경영능력과 이를 뒷받침할 수 있는 인공지능 알고리즘의 기술력이 부족하다면, 보유한 데이터의 양과 질이 상대적으로 뒤떨어지는 경쟁업체에 비해 더 낮은 수준의 서비스를 제공할 가능성도 있다.

나아가 오늘날 온라인 플랫폼 사업자들이 검색엔진, 이메일, 소셜미디어, 클라우드 스토리지 등 수많은 서비스를 무료로 소비자들에게 제공한다는 점에서 소비자들의 후생이 과연 그러한 플랫폼 사업자들에 의해 악화되고 있는 것인지에 대해 의문을 제기하는 견해도 있다. 소비자의 후생을 악화시키는 독과점 상태가 아니라, 소비자의 후생이 유지되거나 오히려 더 개선될 수 있는 경쟁적인 경합 상태에 오늘날의 플랫폼 사업자들이 놓여있을 수도 있기 때문이다.

결국, 대형 온라인 플랫폼 사업자들이 데이터를 매개로 하여 지배력을 확보하고 경쟁을 저해하는 결과가 초래될 것인지에 관해서는 여러 가지의 이론적인 가능성이 있다. 실증적인 연구가 도움이 될 수 있겠지만, 아직까지 뚜렷한 방향이 제시된 것으로



보기는 어려운 상황이다.

## 2. 데이터에 관한 규율과 시장경쟁의 이슈 : 데이터 이동성의 경우

데이터를 둘러싼 다양한 규율이 시장경쟁에 미치는 영향에 대해서도 깊이 있는 고민이 필요하다. 유럽연합에서의 데이터 이동성(data portability)에 관한 논의가 대표적인 예시가 될 수 있다. 데이터 규제는 프라이버시 보호, 보안 및 관리의 신뢰성을 확보하기 위한 투명성 조치가 결합하여 정보기술 분야에서 독자적인 법 논리를 구축해왔다. 이러한 논리는 최근 정보주체의 데이터에 대한 주체성, 통제력 등 오너십(ownership)이 강조되면서, 정보주체의 데이터에 대한 접근과 통제를 촉진하기 위해 개인 이용자가 자신에 관하여 특정 사업자가 보유하고 있는 데이터를 다른 사업자에게 이전할 수 있는 권리까지도 보장해야 한다는 이른바 데이터 이동권(right to data portability)이라는 개념으로 대두되기에 이르렀다. 데이터 이동권은 유럽연합이 2018년부터 시행 중인 일반정보보호규정(General Data Protection Regulation, 'GDPR')에 명문화되어 도입되었으며(Article 20(1)), 지불결제 서비스의 영역에서는 유럽연합이 'PSD 2 지침(Payment Service Directive 2)'을 마련하면서 이동권에 관한 규정을 포함한 바 있다(Mai, Heike, 2018, p. 5).

유럽연합 GDPR의 해당 조항을 보면, 이동권에 관하여 여러 가지의 원칙과 제한이 적용되고 있다.<sup>1)</sup> 해당 조항에 포함된 원칙과 제한은 다음과 같이 정리할 수 있다. 첫

1) GDPR의 해당조항은 다음과 같다.

### Article 20. Right to data portability

1. The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided, where:
  - (a) the processing is based on consent pursuant to point (a) of Article 6(1) or point (a) of Article 9(2) or on a contract pursuant to point (b) of Article 6(1); and
  - (b) the processing is carried out by automated means.
2. In exercising his or her right to data portability pursuant to paragraph 1, the data subject shall have the right to have the personal data transmitted directly from one controller to another, where technically feasible.
3. The exercise of the right referred to in paragraph 1 of this Article shall be without

째, 정보주체는 자신이 데이터 컨트롤러(개인정보 처리자)에게 제공한(provided) 정보에 한하여 이동권을 요구할 수 있다. 분석 결과에 기초하여 마련된 추정 데이터(inferred data)는 이동권 적용의 대상이 되기 어렵다. 둘째, 정보주체는 자신에 관한(concerning him or her) 개인정보에 해당하는 정보에 대해서만 이동권을 요구할 수 있을 뿐 제3자의 개인정보에 대해서는 이동권을 요구할 수 없다. 셋째, 정보 처리가 자동화된 방식(automated means)으로 이루어진 경우에만 이동권을 행사할 수 있다. 넷째, 구조화되고, 통상적으로 이용되며, 기계(컴퓨터)가 읽을 수 있는 포맷(structured, commonly used and machine-readable format)으로만 이동권이 행사될 수 있다. 다섯째, 데이터의 이동에 있어 한 컨트롤러부터 제3의 컨트롤러에게 직접 전송을 요구할 수 있다. 여섯째, 해당 권리는 기술적으로 가능한(technically feasible) 경우에 한하여 행사될 수 있다.

그런데 데이터 이동권이 보장되면 어떤 긍정적인 효과가 발생할 것으로 기대할 수 있는가? 우선, 이용자가 하나의 플랫폼으로부터 또 다른 플랫폼으로 이동함에 있어, '전환비용(switching cost)'이 크게 절감될 수 있다. 이에 따라, '록인(lock-in)' 효과가 줄어들 수 있고, 플랫폼 사이의 경쟁이 촉진될 가능성이 높아질 수 있다. 둘째, 이동권을 통해 정보주체가 자신의 데이터에 관한 통제권을 좀 더 적극적으로 행사하는 계기가 마련될 수 있다. 이에 따라 개인정보자기결정권의 실효성이 높아질 수 있다. 셋째, 이 과정을 통해, 이용자가 플랫폼 사업자에 대해 좀 더 실효성 있는 형태로 협상력을 확보하는 계기가 마련될 수 있다.

그러나 데이터 이동권은 사업 현장에서의 비용 및 기술적 수준을 고려할 때 현실적으로는 구현이 쉽지 않을 수 있다. 개인정보처리자들은 각자의 사업적 필요에 따라 전혀 다른 영역과 항목의 데이터를 수집하고, 이를 자신이 제공하는 서비스에 맞게 데이터베이스를 구축해 놓았을 것인데, 이러한 차이를 극복하고 사업자 간에 데이터의 호환성을 마련하는 것은 매우 복잡한 실무적 문제를 야기할 가능성이 높다. 특히 이를 위해서는 데이터에 대한 표준화 및 상호운용성을 확보하는 과정이 필수적일 것인데, 기존에 이용해 오던 기술이나 표준의 차이에서 오는 문제뿐만 아니라, 새로운

---

prejudice to Article 17. That right shall not apply to processing necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.

4. The right referred to in paragraph 1 shall not adversely affect the rights and freedoms of others.

공통의 표준을 도입하고 실행하는 과정에서 야기될 비용을 누가 어떤 형태로 부담할 것인지 등 관련된 여러 요소를 고려하면 그 과정이 간단치 않을 것임을 쉽게 예상할 수 있다.

GDPR의 경우를 전제로 할 때 나타날 수 있는 데이터 이동권 개념의 한계를 더 구체적으로 정리해 보면 다음을 포함한다. 우선, GDPR의 경우, 권리의 인정은 정보주체가 제공한 정보에 국한되고 컨트롤러가 생성하거나 분석하여 마련한 정보에는 해당되지 않는다. 또한, 기술적으로 가능한 경우에만 이동권이 적용될 수 있다. 이동권 행사를 통해 제3자의 권리가 침해될 가능성이 있으면(예컨대, 여러 명의 인물이 포함된 단체 사진의 경우), 적용에 제약이 발생할 수밖에 없다. 이러한 제약 사항을 고려하면, 제도의 도입을 통해 긍정적인 효과가 얼마나 달성될 수 있을지 불명확하다. 둘째, 기술적인 면에서도 많은 제약이 나타날 가능성이 있다. 예를 들어, 하나의 이메일 서비스에서 또 다른 이메일 서비스로 이동하는 경우를 생각해 보면, 데이터 포맷, 첨부파일 유무, 용량 등등 매우 다양한 요소에 따라 여러 형태의 기술적인 어려움이 나타날 것을 쉽게 예상할 수 있다. 또한 이러한 기술적이고 현실적인 어려움을 어떻게 해결할 것인지에 관해 논의하는 과정은 매우 복잡할 수밖에 없다. 그리고 실제로 데이터 이동을 실행하는 과정에서, 일부 데이터의 유실 및 원활한 서비스 제공의 어려움 등이 일상화될 가능성이 높고, 이에 관하여 다양한 불만과 민원이 제기될 가능성이 높다. 셋째, 이동권의 구현을 위해 요구되는 비용이 매우 높은 수준에 달할 가능성이 있다. 구체적으로 어느 정도 수준의 비용이 소요될지에 대해서는 별도의 추산이 필요하겠지만, 다른 한편, 이 비용을 누가 어떻게 분담하게 될 것인지를 둘러싸고 상당한 갈등이 빚어질 가능성이 있다. 넷째, 이동권에 관한 논의가 구체화될 경우, 데이터의 표준화(standardization) 및 상호운용성(inter-operability) 확보가 핵심 이슈가 될 것이다. 그 경우, 누가 주도하여 어떤 방식의 표준이 정해지고, 어떻게 상호운용성이 확보될 것인지에 따라서, 개별 기업들에는 매우 커다란 직접적인 영향이 나타날 수 있다. 현실적으로는, 표준화 작업 등과 관련하여 풍부한 경험을 가진 글로벌 기업들이 논의를 주도하게 될 가능성이 있다. 다섯째, 이동권이 실제로 구현될 경우, 데이터의 이동과정에서 프라이버시 침해 또는 데이터 유출의 가능성이 높아질 수 있다. 또한, 그 과정에서 누구에게 과실이 있었는지, 관련된 법적 책임 문제는 어떻게 해결할 것인지 등의 문제를 둘러싸고 많은 논란이 발생할 수 있다. 여섯째, 일반적인 이용자 입장에서는, 예를 들어, 특화된 소수의 서비스를 제공하는 플랫폼보다는 다양한 서비스를 제공하는 플랫폼, 보안이나 프라이버시 맥락에서 신뢰도가 높은 플랫폼

등으로 이동하는 경향이 나타나게 될 가능성이 있다. 그렇게 될 경우, 경쟁을 촉진하는 것이 아니라, 정반대로 거대 플랫폼으로의 집중현상이 더욱 촉진될 가능성이 있다.

물론, 이는 생각해 볼 수 있는 다양한 우려와 한계를 정리한 것인데, 이러한 한계와 우려가 실제로 중요한 요인으로 작동하게 될 것인지는 명확하지 않다. 다른 한편, 이와 같은 한계와 우려를 고려하면, 이동권 구현을 통해 긍정적인 효과를 기대해 볼 수 있는 상황을 좀 더 구체적으로 모색해 볼 수 있다. 일반적으로, (1) 이미 데이터에 대한 표준화 및 상호운용성이 상당한 수준으로 달성되어 있는 경우, (2) 거대 플랫폼으로의 집중에 대한 우려가 상대적으로 적은 경우, (3) 관련된 비용의 부담을 둘러싸고 이해관계자 사이에 갈등이 크지 않을 경우 등에는 데이터 이동권에 대하여 좀 더 적극적으로 모색을 해볼 수 있을 것이다. 그러한 경우에 해당될 수 있는 예로 금융산업을 생각해 볼 수 있다. 위에 언급한 것과 같이, 이미 EU에서는 PSD 2에 데이터 이동권에 관한 내용을 포함한 바 있다. 금융산업은 데이터의 표준화가 상대적으로 높은 수준으로 이미 달성되어 있고, 대체로 동질적인 서비스를 제공하는 것으로 인해 데이터 이동권을 도입하더라도 거대 플랫폼으로의 ‘쏠림’ 현상이 두드러지게 나타날 가능성은 다른 산업에 비해 높지 않을 것으로 예상할 수 있다. 또한, 금융영역의 경우, 규제산업으로서의 특징을 고려하면, 소요될 비용의 분담을 둘러싼 논란 또한 상대적으로 적을 가능성이 있다. 다른 한편, 데이터 이동권을 통해 소비자를 둘러싼 금융기관 사이의 경쟁을 부분적으로라도 촉진하는 효과를 기대해 볼 수 있다.

하지만 좀 더 일반적인 상황에서는 개인이 사업자로부터 자신에 관한 데이터를 반출하여 다른 사업자에 이전할 수 있도록 돕는 권리가 실제로 경쟁을 촉진하는데 도움이 될 것인지 여부는 명확하지 않다. 데이터 이동권의 보장이 플랫폼 사업자 간의 경쟁을 촉진할 가능성도 있지만, 오히려 사용자들이 데이터 보안(security) 또는 프라이버시 보호가 불안해 보이는 후발기업으로부터 자신의 데이터를 반출하여 선두기업으로 데이터를 이전함으로써 기존 플랫폼 사업자를 향한 ‘쏠림’ 현상을 더욱 가속화시킬 가능성도 있기 때문이다. 이에 관하여 실증적인 연구를 하는 것이 도움이 될 것이나, 데이터 이동권의 보장이 전환비용을 감소시킴으로써 플랫폼의 직간접적인 네트워크 효과로 인한 경쟁 저해의 우려를 약화시킬 수 있는 가능성이 어느 정도인지에 관하여는 아직 명확한 실증적인 연구가 도출되지 않고 있는 상황이다.

## V. 추가적 정책 과제

이 글을 통해 인공지능과 관련된 정책적 고려에 있어 데이터가 왜 중요한지 정리하고, 또한 프라이버시, 차별, 경쟁정책, 사회윤리 등을 아우르는 매우 다양한 경제적, 사회적 쟁점이 어떤 형태로 대두되고 있는지 정리하고자 하였다. 국가적인 인공지능 정책의 구상을 위해서는, 각각의 개별 산업·학술분야별로 기존에 연구해왔던 논리나 노하우가 인공지능 시대에 와서 어떻게 변모해야 할지 고민하는 동시에, 더 넓게는 인공지능이라는 새로운 패러다임에 맞춰 사회경제가 어떻게 변모하고 있는지에 관심을 기울일 필요가 있다.

이 글에서는 인공지능 정책에서 고려되어야 할 다층적·다면적 쟁점들의 일부만을 살펴보았으나, 그 밖에 논의되어야 할 쟁점 역시 다양하다. 인공지능 정책의 맥락에서 추가적으로 다뤄져야 할 특히 중요한 과제의 예시로 아래의 3가지를 포함할 수 있다.

**(1) 리스크에 대한 파악과 규명을 위한 절차의 정립** : 인공지능은 ‘인간처럼’ 행동하고 사고하는 독립 주체가 아니라는 점에서 인공지능에 의한 리스크의 규율이 인공지능 그 자체에 대한 통제로 당연히 이어지는 것은 아니다. 그러나 인공지능을 활용하는 인간에 대한 규율은 의도 및 과실을 책임의 중요한 요건으로 여기는 전통적 법체계에서 어려움을 겪고 있으며, 특히 딥러닝 기술의 불투명성(opacity)으로 인해 리스크의 원인파 당사자들의 기여 등 인과관계를 사후적으로 법적 절차를 통해 규명하는 것에는 많은 어려움이 예상되고 있다. 인간이 유발하는 리스크는 손해배상 등 민사적 제재, 허가 등 행정적 통제 및 과징금을 포함한 행정 제재, 그리고 형사처벌 등 형사적 제재로 규율되어 왔는데, 이러한 제재의 구체적인 내용 그리고 제재를 부과하는 절차 모두 인공지능 기술의 맥락에서 새로이 상세한 분석이 필요할 것이다.

**(2) 조세제도와 노동법** : 인공지능과 로봇의 활용이 생산성이나 경제성장에 어떤 영향을 미칠 것인지에 대해서는 다양한 견해가 있다. 인공지능과 로봇이 생산과정에서 사람을 더 많이 대체할수록, 최종적 부가가치의 생산에 사람이 기여하는 부분은 줄어들 것이다. 그런데 이는 생산성 하락을 막아주는 긍정적인 역할일 수 있는 한편, 생산과정에서의 노동의 역할에 대한 재고가 필요할 수도 있고, 노동력의 재훈련이나 재배치가 필요한 경우도 늘어날 수 있다. 또한 노동시장의 변화를 고려한 조세 및 보

조금 제도에 대한 사회적 변화가 요구될 수도 있다. 생산현장에서의 변화는 단기간에 급속도로 나타나기 보다는 점진적으로 나타날 가능성이 높는데, 이러한 변화에 대응하여 노동법제와 조세제도가 변화해야 할 필요가 있는지, 변화가 필요하다면 어떤 변화가 필요할 것인지에 대해서는 별도의 분석과 사회적인 논의가 필요하다(Calo, Ryan, 2017, pp.425-427).

(3) **다양한 정책 목표 사이의 조율** : 인공지능 및 데이터 정책에는 상업적 활용 가능성, 차별금지, 프라이버시 및 개인정보보호, 시장경쟁 등 전혀 다른 성격의 여러 쟁점들이 복잡하게 얽혀 있고, 이러한 다양한 가치 사이에 조화를 달성하기 위한 정책적 조율이 필요하다. 이 과정에서 어떠한 사회적 가치를 어느 범위에서 우선하거나 어떤 방식으로 조율할 것인지에 대한 합의가 필요하며, 서로 다른 사회적 가치 사이의 조율을 일관성 있게 달성할 수 있도록 인공지능과 데이터를 규율할 행정적인 메커니즘을 어떻게 마련할 것인지 생각해 볼 필요가 있다.

#### ■ 참 고 문 헌

1. 고학수 외 4인, 『개인정보 비식별화 방법론』, 박영사, 2017.
2. 고학수 · 정해빈 · 박도현, “인공지능과 차별,” 『저스티스』, 제171호, 2019, pp.199-277.
3. 이균희, “바젤2 협약 기반 소매여신의 신용위험 측정 방법론에 대한 연구,” 『경영논총』, 제15-2집, 2004, pp.215-230.
4. 이용용, “금융 생태계를 변혁하는 인공지능의 확산 및 시사점,” 『KISA(한국인터넷진흥원) REPORT』, 2018 Vol. 9., 2018.
5. 장병탁, 『장교수의 딥러닝』, 홍릉과학출판사, 2017.
6. 추형석, “인공지능, 어디까지 왔나?” 『SPRi(소프트웨어정책연구소) 이슈리포트』, Vol. 6, 2018.
7. 한국정보화진흥원(NIA), 『AI Network Lab 이슈리포트』, Vol. 3, 2018.
8. AHSN Network, *Accelerating Artificial Intelligence in Health and Care: Results from a State of the Nation Survey*, AHSN Network, 2018.
9. CADTH, “An Overview of Clinical Applications of Artificial Intelligence,” *CADTH Issues in Emerging Health Technologies*, Vol. 174, 2018.
10. Calo, Ryan, “Artificial Intelligence Policy: A Primer and Roadmap,” *UC Davis Law Review*, Vol. 51, 2017, pp.399-436.
11. Dastin, Jeffrey, *Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women*, Reuters, 2018. 10. 10.
12. Deloitte, *Artificial Intelligence Innovation Report 2018*, 2018.
13. Forbes Insights, *Inside The R&D Of AI Ethics*, 2019. 3. 27.

14. Freuler, Juan O. and Carlos Iglesias, *Algorithms and Artificial Intelligence in Latin America: A Study of Implementation by Governments in Argentina and Uruguay*, World Wide Web Foundation, 2018.
15. Gent, Edd, *How AI is Helping Sports Teams Scout Star Players*, NBC, 2018. 6. 13.
16. Goodfellow, Ian, Yoshua Bengio and Aaron Courville, *Deep Learning*, MIT Press, 2016.
17. Kewalramani, Manoj, *China's Quest for AI Leadership: Prospects and Challenges*, Takshashila Working Paper, 2018-02, 2018.
18. KPMG, *How the UK can win the AI Race*, KPMG UK, 2018.
19. Little Hoover Commission, *Artificial Intelligence: A Roadmap for California*, Little Hoover Commission Report, Vol. 245, 2018.
20. Mai, Heike, *PSD 2, Open Banking and the Value of Personal Data*, Deutsche Bank Research, 2018. 6. 28.
21. Mayor of London, *London: The AI Growth Capital of Europe*, 2018.
22. Marcus, Gary, "Deep Learning: A Critical Appraisal," *arXiv preprint arXiv:1801.00631*, 2018, pp. 1-27.
23. McKinsey Global Institute, *Notes from the AI Frontier - Applying AI for Social Good*, 2018.
24. Negnevitsky, Michael, *Artificial Intelligence: A Guide to Intelligent Systems*, 2<sup>nd</sup> ed., Addison-Wesley, 2005.
25. O'Leary, Daniel E., "Artificial Intelligence and Big Data," *IEEE Intelligent Systems*, Vol. 28, No. 2, 2013, pp. 96-99.
26. Reed, Chris, Elizabeth Kennedy and Sara Silva, "Responsibility, Autonomy and Accountability: Legal Liability for Machine Learning," *Queen Mary School of Law Legal Studies Research Paper*, No. 243/2016, 2016, pp. 1-31.
27. Rona-Tas, Akos and Alya Guseva, "Consumer Credit in Comparative Perspective," *Annual Review of Sociology*, Vol. 44, No. 1, 2018, pp. 55-75.
28. Russell, Stuart and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3<sup>rd</sup> ed., Pearson, 2010.
29. Sartor, Giovanni, *Artificial Intelligence: Challenges for EU Citizens and Consumers*, European Parliament IP/A/IMCO/2018-16, 2018.
30. Shoham, Yoav, et al., *The AI Index 2018 Annual Report*, Stanford AI Index Steering Committee, 2018.
31. Silver, David, et al., "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature*, Vol. 529, 2016, pp. 484-489.
32. U.K. Digital Competition Expert Panel, *Unlocking Digital Competition*, 2019. 3. 13.
33. Williams, Betsy A., Catherine F. Brooks and Totam Shmargad, "How Algorithms Discriminate Based on Data They Lack," *Journal of Information Policy*, Vol. 8, 2018, pp. 78-115.

## Artificial Intelligence and Market Competition: Significance of Data and Relevant Policy Considerations

Yong Lim\* · Hae Bin Jeong\*\* · Haksoo Ko\*\*\*

### Abstract

Artificial intelligence, as it is currently being developed, is a data-driven technology. This can be easily observed in deep learning technology's data dependency, which presupposes the existence of useful training data. As such, the development of the technology and how the market evolves will naturally be impacted by regulation on the collection and use of personal information. The use of artificial intelligence has also begun to raise new social challenges related to discrimination and fairness. There are various predictions about how artificial intelligence technology will impact our society and economy, including market competition. In order to establish optimal and effective artificial intelligence policies, it is imperative to analyze and consider the interconnected effects of data-related policies that will impact the future development of the technology and competition surrounding it.

**Key Words:** artificial intelligence, data, algorithmic discrimination, fairness

**JEL Classification:** K3, L4, O3

---

*Received: Sept. 16, 2019. Revised: Oct. 16, 2019. Accepted: Oct. 25, 2019.*

\* First Author, Associate Professor, Seoul National University School of Law, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Korea, Phone: +82-2-880-7549, e-mail: yonglim@snu.ac.kr

\*\* Second Author, Ph. D. Candidate, Seoul National University School of Law, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Korea, Phone: +82-2-880-2602, e-mail: shiningsea@snu.ac.kr

\*\*\* Corresponding Author, Professor, Seoul National University School of Law, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Korea, Phone: +82-2-880-2602, e-mail: hsk@snu.ac.kr