

인공지능과 경제성장

김지희(KAIST 기술경영학부)

인공지능과 시장경제

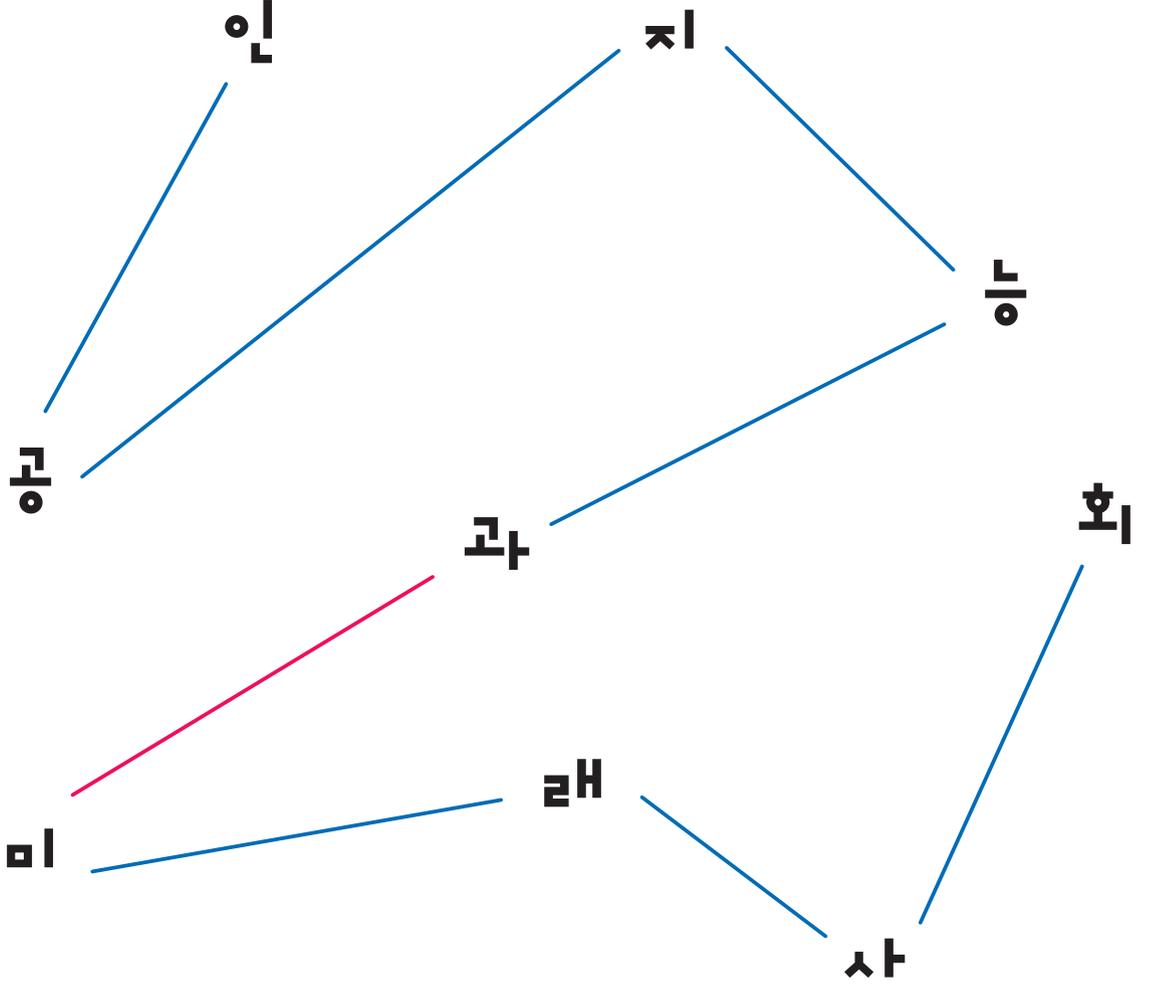
정해빈(서울대학교 법학대학원)
고학수(서울대학교 법학전문대학원)
임용(서울대학교 법학전문대학원)

인공지능과 고용시장의 변화

이수형(서강대학교 경제학부)

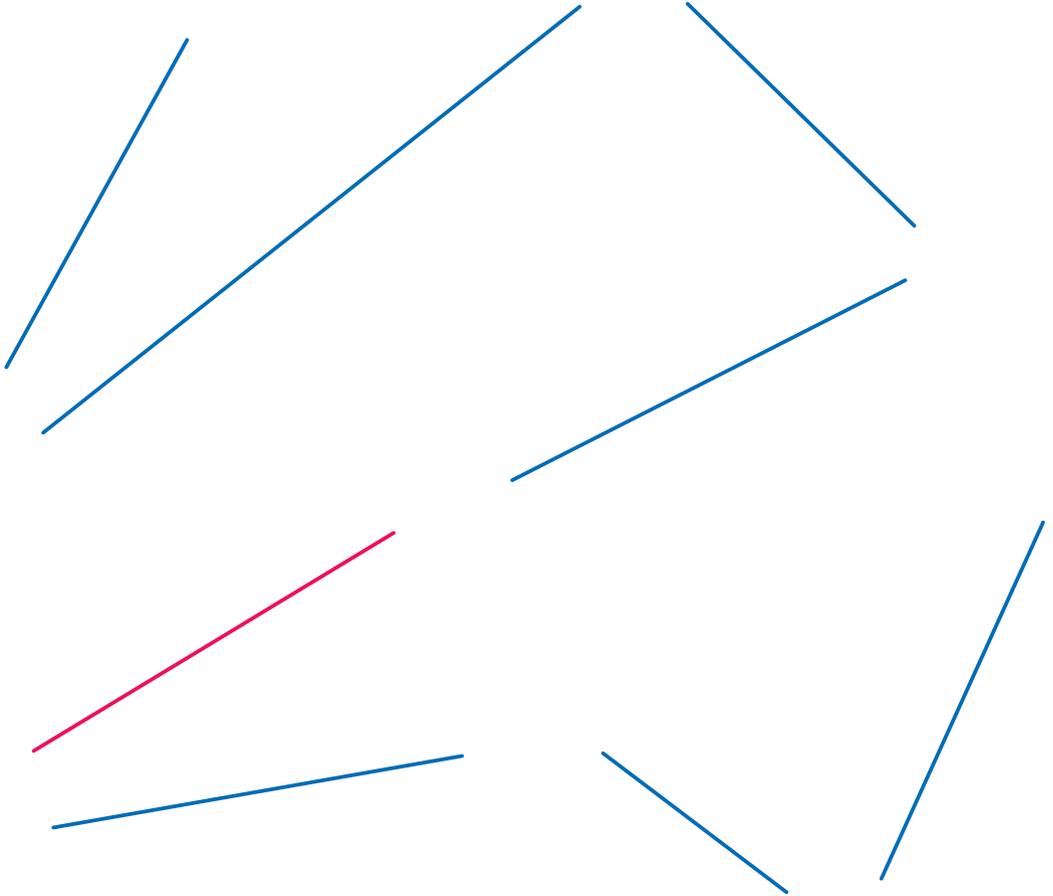
윤리적 인공지능의 실현과 과제

고학수(서울대학교 법학전문대학원)
박도현(서울대학교 법학대학원)
이나래(서울대학교 법학대학원)



목차

04	인공지능과 경제성장 김지희(KAIST 기술경영학부)
20	인공지능과 시장경제 정해빈(서울대학교 법학대학원) 고학수(서울대학교 법학전문대학원) 임용(서울대학교 법학전문대학원)
36	인공지능과 고용시장의 변화 이수형(서강대학교 경제학부)
58	윤리적 인공지능의 실현과 과제 고학수(서울대학교 법학전문대학원) 박도현(서울대학교 법학대학원) 이나래(서울대학교 법학대학원)



인공지능과 경제 성장: 인공지능은 경제 성장의 촉진제가 될 수 있을까

김지희(교수, KAIST 경영대학 기술경영학부)



웹에서 PDF 바로 보기

인류가 이제껏 경험했던 기술 혁신은 폭발적인 성장(성장률의 증가)을 가져온 적은 없지만, 산업혁명 이후 지난 150년동안 지속적인 성장의 원동력이었다. 그런데 국제 금융 위기를 겪고 난 뒤인 약 10여년전부터 미국과 유럽의 경제 선진국들이 기대치 보다 낮은 경제성장률을 보이며 경제 성장에 대한 우려가 시작되었다. 기술 혁신과 그에 따른 생산성의 향상이 예전만 못하기 때문에 이제는 저성장이 당연하다는 의견도 있다. 최근 머신러닝을 기반으로 한 인공지능 기술의 빠른 발전에도 이러한 우려는 사라지지 않는 것으로 보인다. 앞으로도 인공지능 기술의 발전은 이전의 경제 성장세를 회복하기에는 역부족일까. 아니면 저성장을 넘어 폭발적인 경제 성장, 즉 특이점과 같은 전혀 새로운 양상의 경제 성장을 가능하게 할까.

이 글에서는 위의 질문에 대한 경제학 연구와 경제학자들의 의견을 살펴본다. 경제학 이론에 의하면 인공지능 기술이 스스로 발전할 수 있게 되더라도 ‘모든 생산 활동의 완전한 자동화’가 힘들다면 폭발적인 경제 성장은 힘들 것으로 보인다. 보물의 비용 질병 원리(Baumol’s cost disease) 때문에 자동화되지 못한 작업들이 경제 성장에 장애가 되기 때문이다. 경제학자들의 의견을 살펴보면 반 정도가 조심스러운 입장 -‘앞으로의 기술 발전으로 저성장을 벗어날 수 있을지 확실하게 알 수 없다’ -을 취하고 있다. 인공지능 기술을 경제 전반에 응용할 수 있는 환경적 기반이 마련된다면, 인공지능 기술이 이전의 경제 성장세를 회복하고 지속적으로 성장하는 원동력이 될 수 있을 것이라는 의견이 나머지 반에서는 우세하다. 또한 경제학자들은 인공지능 기술의 발전으로 인한 일자리 문제나 불평등 심화 등의 경제적 손실이 생산성 향상과 같은 경제적 이익보다는 크지 않을 것으로 보고 있다.

1. 들어가는 말

이 글에서는 인공지능이 경제 성장의 경로를 어떻게 바꿀 수 있을 지 이야기하기 위해서 먼저 경제 성장과 그 역사에 관해서 생각해 보려고 한다. 그 위에 기술 혁신의 역사와 경제 성장을 엮어보고, 인공지능 기술은 어떤 기술 혁신에 속하는 지 이야기해 본다. 이를 바탕으로 인공지능 기술에 관한 경제학 연구와 경제학자들의 의견을 살펴 보면서 질문에 대한 답을 찾아나가려고 한다. 마지막으로 경제 성장 뿐만 아니라 인공지능 기술이 가져올 일자리와 불평등의 문제에 대해서도 간략하게 살펴본다.

2. '경제 성장' 생각해보기

한국은행은 올해 우리나라 경제 성장률이 2.5%가 될 것이라고 예측했다. 경제 성장률을 얘기할 때 가장 많이 언급하는 1인당 GDP(Gross Domestic Product, 국내총생산) 성장률에 대한 전망이다. 평균적으로 우리나라에 사는 한 사람당 작년보다 올해 버는 돈, 또는 쓰는 돈, 또는 생산하는 가치가 물가상승률을 감안하고도 2.5% 증가할 것이라는 얘기다. 우리 경제는 2017년에는 3.1%, 2018년에는 2.7% 성장하였고, 2018년에 미국은 2.8%, 중국은 6.6% 성장하였다.⁰¹

우리나라는 1970~80년대에 경제 성장률이 10% 이상을 기록하는 해도 많았지만, 지난 국제 금융 위기에서 벗어난 이후에는 2~3% 수준의 경제 성장률을 보이고 있다.(그림 1) 언뜻 보아 과거에 비해 경제 성장이 더뎠다고 걱정할 수도 있겠지만, 경제가 어느 정도 성숙하면 사실이 이상의 경제 성장률은 기대하기 힘들다. 비교적 쉬운 경제 성장의 촉진제들을 이미 다 써버렸기 때문이다. 전쟁으로 폐허가 된 땅에서는 자본 투자를 통해, 평균 교육 수준이 초등 교육에도 미치지 못했을 때는 교육 투자를 통해서 빠른 성장이 가능하지만, 어느 수준에 이르면 빠른 성장세를 지속하기 힘들다. 10만원에서 20만원으로 두 배 수입이 커지는 경우와 1억에서 2억으로 두 배 수입이 커지는 경우를 비교해서 생각해봐도 쉽게 알 수 있다. 우리나라는 2017년 기준 1인당 GDP가 \$29,743로 세계 28위인 꽤 잘 사는 나라다 (일본 25위, 이탈리아 26위, 스페인 31위)⁰². 참고로 G7 국가 평균 경제 성장률은 1970년 이후 지금까지 3.5%를 넘은 적이 여섯 해 밖에 없다. 또한 몇 해 전부터 미국과 유럽의 선진 경제국들은 저성장이 유독 오래 지속되고 있다며 걱정하기 시작했다. 2017년 1인당 GDP 성장률은 28개 유럽연합 국가 평균 2.24%, OECD 국가 평균 1.98%, G7 국가 평균 1.70%였다.⁰³

01 데이터 출처: 한국은 한국은행, 미국은 미 연방은행 경제 데이터(FRED), 중국은 중국 통계청 실질 1인당 GDP 성장률 발표 자료를 인용하였다.

02 데이터 출처: 세계은행(World Bank)

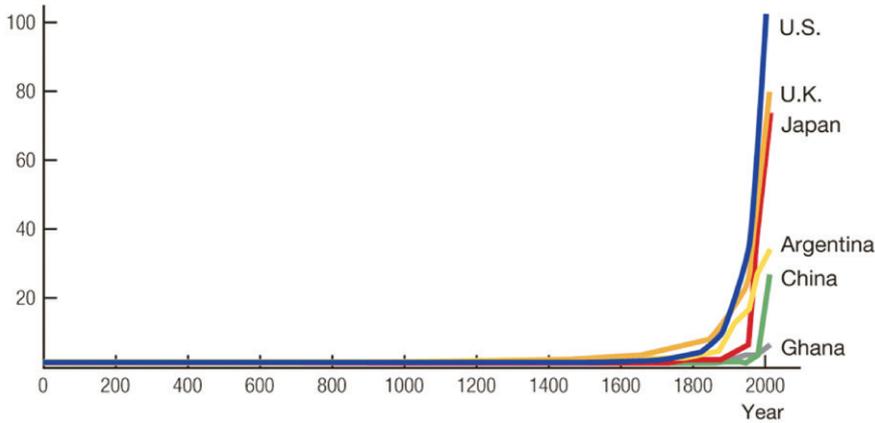
03 데이터 출처: OECD



[그림 1] 1953-2018년 우리나라 1인당 GDP 성장률(%) (출처: 한국은행)

Economic Growth over the Very Long Run in Six Countries

Per capita GDP
(multiple of 300 dollars)



Copyright © 2018 W. W. Norton & Company, Inc.

[그림 2] 6개 국가의 1인당 GDP 변화 (출처: Jones (2017))

요약하면, 우리 경제는 비교적 짧은 시간 동안 빠르게 성장하여 경제 선진국이 되었고, 최근 경제 성장의 속도가 다른 선진 국가들과 비슷한 수준으로 낮아졌다. 이러한 우리 경제 성장의 역사는 한강의 ‘기적’이라고 불릴 만큼 흔하지 않다. 그럼 다른 나라의 경제는 어떻게 성장했을까. 특히나 자본주의 시장 경제 체제를 가장 먼저, 그리고 오래 유지해 온 미국 경제 성장의 역사를 살펴보자.

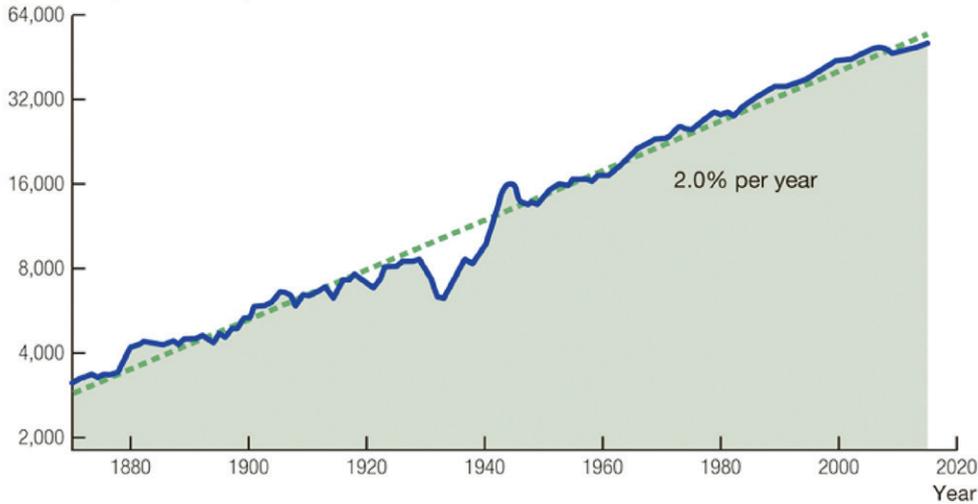
1. 미국 경제 성장의 역사와 기술 발전의 역사 함께 보기

[그림 2]는 기원 후 현재까지 6개 나라의 1인당 GDP 변화를 보여주고 있다. 18세기까지는 6개 나라가 비슷한 경제 수준을 유지하면서 매우 더디게 성장하다가, 미국과 영국은 19세기 전후로 경제 성장세가 급격하게 바뀌었다. 무엇이 이러한 변화를 가져왔을까.

Per Capita GDP in the United States, Ratio Scale

Per capita GDP

(ratio scale, 2009 dollars)

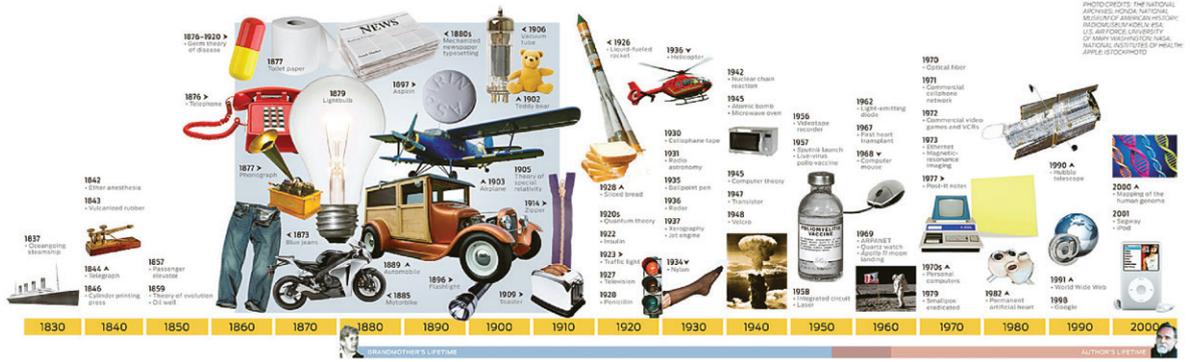


[그림 3] 근현대 미국의 1인당 GDP 변화와 성장률 (출처: Jones (2017))

이 시기에는 1, 2차 산업혁명이 일어났다. 1760년부터 영국에서 먼저 일어난 1차 산업혁명이 경제 성장에 시동을 걸었고, 1870년부터 제 1차 세계대전 발발 전까지 2차 산업혁명은 경제 성장을 빠른 궤도에 올려놓았다.

스탠포드의 경제학자 Chad Jones는 1,2차 산업혁명과 같은 기술 혁신 ‘사건’과 그에 따른 경제 성장은 필연적으로 일어날 수 밖에 없음을 이론적으로 보였다(Jones (2001)), 경제 성장의 원천은 기술 혁신이고, 기술 혁신의 원천은 사람이다. 서서히 증가하던 인구는 인적 자본의 증가로 이어졌고, 인적 자본의 증가는 기술 혁신을 가져왔다. 기술 혁신은 다시 인구 증가에 도움이 되기 때문에 선순환의 고리가 생긴다. 이러한 선순환을 계속 하다보면 언젠가 경제 성장에 시동이 걸리는 시점이 오는 것이다. 추가적으로 재산권을 보호하는 제도의 도입이 산업혁명 시점, 즉 성장 추이가 바뀌는 시점을 200년 정도 앞당겼다고 한다. 재산권이 보호됨으로써 사람들이 기술 개발을 할 동기가 커졌고, 따라서 기술 혁신에 더 속도가 붙게된 것이다.

그렇다면 2차 산업혁명 이후 미국 경제 성장률은 어땠을까. 우리나라 1970-80년대처럼 한 해 10% 이상 빠르게 성장했을까? 1870년 이후 미국의 1인당 GDP 성장을 살펴보자. [그림 3]은 미국의 1인당 GDP 변화를 보여주는데, 그래프의 기울기가 1인당 GDP 성장률이 되도록 Y축을



[그림 4] 기술 발전의 역사 (출처: IEEE Spectrum (2008))

조정한 것이다. 이 그래프에서 볼 수 있는 것처럼 미국의 경제 성장은 우리 경제와 다른 양상을 보여준다. 미국 경제는 세계 대공황, 1,2차 세계대전 등에 의한 경기의 부침에 따라 변동이 있긴 했지만 평균적으로 매년 2% 정도로 성장해왔다. 미국이 10% 이상의 성장률을 보인 것은 경제 위기에서 벗어나는 회복기 때 잠시 뿐이다. 경기 회복기에는 전쟁으로 황폐해진 땅에 건물을 짓는 것처럼 빠른 성장이 비교적 쉽다. 과거에 10% 이상의 성장률을 꽤 자주 경험한 우리에게 2%는 그다지 높은 경제 성장률이 아니라고 느껴질 수도 있겠다. 하지만 미국이 오늘날 경제 대국이 된 것은 다른 국가들보다 일찍, 그래서 상대적으로 오랜 시간동안, 안정적으로 성장을 지속해왔기 때문이다.

요약하면, 제 2차 산업혁명 이후 인류 기술 발전의 최전선에서 기술의 혜택을 알뜰히 챙겼을 미국 경제는 지난 150여년동안 매년 2%정도로 성장을 지속해왔다. 이 시점에서 이런 의문이 들 수 있다. 인류의 눈부신 기술 발전과 혁신이 10%, 20%대의 성장을 가져온 것이 아니라 평균적으로 매년 2%의 성장을 가져왔을 뿐이라고? [그림 4]에서 볼 수 있는 것처럼 현대인들에게는 파급효과를 상상하기 힘든, 아주 오래 전 증기기관의 발명부터, 자동차, 냉장고, 세탁기, 항생제, 비교적 최근의 컴퓨터, 인터넷, 스마트폰 등 기술은 지속적으로 우리 삶의 질을 완전히 바꾸어 놓았는데,

이를 돈으로 바꿔서 환산하면 매년 2% 성장한 것과 같다니, 뭔가 잘못 계산된 것은 아닐까.

이러한 의문을 풀기 위해서 우선 성장률은 복리 이자율과 같다는 점을 생각해야한다. 1인당 GDP가 매년 2%씩 증가하면, 약 35년 후에 두 배, 약 70년 후에 네 배가 되고, 150년 후에는 20배가 커진 셈이 된다. 제 2차 산업혁명 이후 '경제 수준이 20배 향상됐다'와 '경제가 매년 2%씩 성장했다'를 비교하면 사실 같은 말이지만 전자가 더 수긍하기 쉬운 것이다.

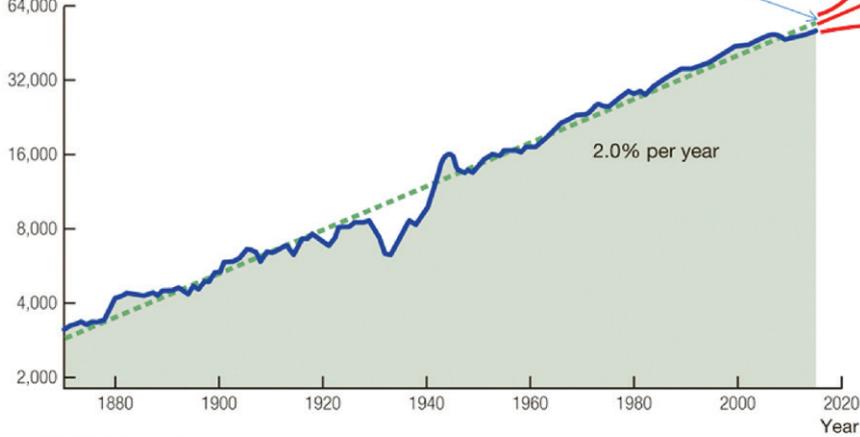
중요한 것은 기술 혁신을 주 원동력으로 성장해 온 미국 경제가 '지속적'으로 안정적인 성장을 유지해왔다는 점이다. 즉, 인류가 이제껏 경험했던 기술 혁신은 폭발적인 성장(성장률의 증가)을 가져온 적은 없지만, 지난 150년동안 지속적인 성장의 원동력이었다. 바꿔서 생각해보면, 그 동안의 기술 혁신이 없었다면 매년 2% 성장을 유지할 수 없었을 것이다. 그리고 앞으로도 이러한 성장세를 유지하려면, 기술 혁신의 추세도 지속되어야 한다.

2. 기술 혁신과 경제 성장의 미래 전망

그러면 앞으로의 기술 혁신과 경제 성장에 대한 전망은 어떨까. 지난 2007~2008년 미국발 국제 금융 위기 이후 경기 회복에 오랜 시간이 걸리면서 경제학자들 사이에 미국 경제 뿐만 아니라 선진 경제 전반에서 장기적 경기 침체(secular stagnation, growth slowdown)를 우려하는 목소리가 나오기 시작했다. 그 원인에 대해서 여러 의견이 있지만, 그 중 하나로 2000년대에 들어 기술 혁신이 예전같지 않다는 점이 종종 언급된다. 경제학자 Tyler Cowen은 혁신의 황금기는 지났다고 얘기한다(Cowen (2011)). 이미 쉽게 개발할 수 있는 기술들은 다 나왔기 때문에 앞으로의 기술 개발은 훨씬 힘들고, 그 파급 효과도 적을 것이라는 것이 그의 주장이다. Northwestern 대학의 경제학자 Robert Gordon, Oxford의 경제사학자 Carl Benedikt Frey, 그리고 장기적 경기 침체를 처음 화두로 만든 Harvard의 경제학자 Larry Summers 등은 디지털 기술을 바탕으로 한 산업은 이전보다 더 적은 인력과 자본을 필요로 하기 때문에 경제 규모가 축소되고, 이는 저성장으로 이어질 수 있다고 얘기한다.

미래의 기술 혁신은 상당 부분 인공지능 기술이 이끌어 가게 될텐데, 인공지능 기술은 경제 성장에 어떤 영향을 줄까. Tyler Cowen의 주장처럼 인공지능 기술로도 장기적 경기 침체의 우려를 벗어나기 힘들까, 아니면 인공지능 기술 덕분에 장기적 경기 침체에서 벗어나 지난 150여년의 성장세를 이어갈 수 있을까, 아니면 인공지능 기술이 폭발적인 성장을 가능하게 하여 우리는 전혀 새로운 경제 성장을 경험하게 될까.

Per Capita GDP in the United States, Ratio Scale
 Per capita GDP
 (ratio scale, 2009 dollars)



Copyright © 2018 W. W. Norton & Company, Inc.

[그림 5] 인공지능 기술은 미래에 성장의 양상을 어떻게 바꿀까

3. '인공지능' 생각해보기

1. 인공지능의 정의

인공지능을 다양하게 정의할 수 있다. 일반적으로 인공지능을 인간이 사고하는 것처럼 사고하는 지능이라고만 생각하기 쉬운데, 실제 학계나 업계에서는 인간의 사고를 모사하는 방식에 국한하지 않고, 어떤 방식으로든 특정 문제를 자체적으로 해결하는 시스템을 통틀어 인공지능이라고 생각한다. 전문가들 사이에서 많이 회자되는 정의로는 Nilsson (2009)의 정의가 있다.⁰⁴ 그는 “인공지능이란 기계에 지능을 부여하는 작업을 일컬으며, 지능이란 주어진 환경에서 적절하게, 그리고 예측적으로 대처하며 기능하는 능력이다”라고 정의하였다. 즉, 인공지능의 ‘지능’을 인간의 지능으로 한정짓지 않는다. 이러한 정의에 의하면 우리에게 어느 정도는 익숙한 자동화 기술 대부분을 인공지능기술로 볼 수 있다. 인공지능

04 원문 정의를 그대로 가져오면 다음과 같다. “Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment.”

기술의 범주를 3차 산업혁명 이후의 많은 기술로 확대할 수 있는 것이다. 흔히 생각하는 스마트폰의 음성 비서 서비스나 알파고(AlphaGo) 뿐만 아니라 인터넷 검색 기술, 문서 작성 소프트웨어에 문법 오류를 자동으로 처리해주는 기술, 공장 자동화 기술 등도 모두 인공지능 기술이라고 생각할 수 있다.

2. 인공지능 기술은 어떤 기술 혁신인가

이번에는 '기술 혁신'에 대해서 구체적으로 생각해보자. 기술 혁신을 다양한 방법으로 분류할 수 있겠지만, 경제 성장 이론에서는 크게 (1) 새로운 상품/서비스를 개발하는 것, (2) 이미 존재하는 상품/서비스의 품질이나 생산성을 향상시키는 것, 이렇게 두가지로 분류한다. 물론 이 분류는 배타적이지는 않다. 예를 들어 인터넷 기술은 SNS, 화상채팅과 같은 새로운 웹 서비스를 가능하게 하는 동시에 기존에 존재했던 TV나 영화 등의 문화 상품을 소비하는 방식을 바꾸기도 했다.

이러한 분류를 바탕으로 인공지능 기술이 어떤 성격을 띄는지 생각해보자. 우선 자동화 기술을 모두 인공지능 기술이라고 생각한다면, 인공지능 기술은 이미 존재하는 상품/서비스의 생산성을 향상하는 기술로 생각할 수 있다. 예를 들어 공장 생산라인에 자동화 기술을 도입함으로써 자본이 일부 노동은 대체하고, 일부 고급 인력의 노동은 보조하여 이미 개발된 상품이나 서비스의 질과 생산성을 높일 수 있다.

또한 인공지능 기술의 발달로 새로운 상품이나 서비스를 개발하는 것이 수월해질 수도 있다. 간접적으로는 자동화 기술이 생산을 위한 노동에 사용하는 시간을 줄여줌으로써 인간이 더 많은 시간을 새로운 상품이나 서비스 개발에 투자할 수 있다. 보다 직접적으로는 인공지능 기술이 기술 혁신 과정 자체에 참여할 수도 있다. 인공지능 기술이 최적화나 예측의 문제를 빠른 시간 안에 높은 정확도로 풀어낸다면, 과학자들이 새로운 기술을 개발하는 데에 직접적인 도움이 될 수 있다. 인공지능 기술이 상품 생산의 생산성뿐만 아니라 기술 혁신의 생산성도 높일 수 있는 것이다. 바로 이 점, 인공지능 기술이 기술 개발 과정 자체에 직접 참여할 수 있다는 점이 이전의 기술혁신과 가장 큰 차이점이다.

3. 인공지능이 이끄는 경제 성장

이제 이 글의 핵심 질문으로 돌아가보자. 인공지능 기술은 경제 성장에 어떤 영향을 줄까. 인공지능 기술로도 장기적 경기 침체의 우려를 벗어나기 힘들까, 아니면 인공지능 기술은 지난 150여년의 성장세를 이어갈 수

있을까, 아니면 인공지능 기술이 폭발적인 성장을 가능하게 할까.

인공지능 기술을 다룬 경제 이론들은 대부분 최근의 경기 침체는 일시적인 것이고 지속적인 경제 성장을 유지할 수 있을 것으로 전망한다. MIT의 경제학자 Daron Acemoglu와 그의 제자 Pascual Restrepo는 앞서 살펴본 두 가지 성격의 기술 혁신을 모형화하여 지속적인 경제 성장을 전망하였다. (Acemoglu and Restrepo (2018)) 이들은 이론 모형을 통해 인간의 노동을 대체하는 자동화 기술이 발전할수록, 자동화 기술을 개발하는 것보다는 인간의 노동을 필요로 하는 새로운 일자리를 만들어 내는 기술을 개발하는 것이 더 이득이 될 수 있다는 것을 보였다. 평형 상태에서는 자동화 기술의 발전 속도와 새로운 일자리를 만들어 내는 기술 발전의 속도가 같아져서 일자리 생성 효과가 노동 대체 효과를 지속적으로 상쇄시키고 장기적으로는 지속적인 성장세를 유지할 수 있음을 보였다.

Aghion, Jones, and Jones (2017)에서는 다양하게 모형의 가정을 바꿔가며 인공지능 기술의 발전과 경제 성장을 전망하였다. 가정에 따라 지속적인 경제 성장도, 성장률의 점진적인 증가도, 또한 성장률 자체가 지속적으로 증가하는 폭발적인 경제 성장도 모두 가능함을 보였다. 하지만 저자들은 현실적으로 성장률의 증가, 특히나 폭발적인 경제 성장은 불가능할 것이라고 주장한다.

저자들의 여러 모형 중 인공지능 기술이 기술 개발 과정 자체에 직접 참여할 수 있다고 가정한 모형에서 폭발적인 경제 성장이 가능할 것이라는 예측이 나온다. 이 때 인공지능 기술은 특이점(Singularity)에 도달하고 경제 성장률 자체가 지속적으로 증가한다. 특이점(Singularity)은 인공지능이 비약적으로 발전해 인간의 지능을 뛰어넘어 자기 스스로 지능이 성장하기 시작하는 때를 이야기한다. 특이점에서는 인공지능이 인간의 도움 없이도 홀로 기술 혁신을 이루어낼 수 있게 된다.

하지만 이러한 인공지능 기술의 특이점이 폭발적인 경제 성장으로 이어지려면 하나의 조건이 더 만족되어야 한다. ‘모든’ 생산과 기술 개발이 ‘완전한’ 자동화, 즉 모든 일들을 인공지능이 할 수 있어야만 폭발적인 경제 성장이 가능해진다. 자동화가 불가능한 작업이 존재하거나, 생산성의 증가가 충분히 빠르지 않은 작업이 있을 경우, 전체 경제의 성장률은 자동화된 작업이 결정하는 것이 아니라 자동화가 쉽지 않은 작업이 결정하게 된다. 보물의 비용 질병(Baumol's cost disease) 원리 때문이다. 여러 상품/서비스가 서로 보완 관계를 가지고 전체 경제가 형성되는데, 이 때 상대적으로 비효율적으로 생산되는 상품/서비스, 즉 기술 발전이 더딘 상품/서비스의 생산 비용이 비싸고, 비싸지만 다른 상품/서비스와 보완 관계이기 때문에 생산을 하지 않을 수는 없고, 그래서 결국은 전체 경제의 발목(?)을 잡는 셈이 된다.

저자들은 ‘모든 것의 완전한 자동화’가 현실적으로 불가능하기 때문에 폭발적인 경제 성장은 힘들 것이라고 얘기한다. 또한 인공지능 전문가들 사이에서도 물리적인 컴퓨팅 자원의 한계와 상대적으로 느린 센서 기술의 발전 등을 이유로 특이점은 가까운 미래에 오지 않을 것이라는 의견이 중론이다.

경제 이론들이 지속적인 경제 성장을 전망한다면, 최근의 장기적 경기 침체에 대한 우려는 어떻게 설명할 수 있을까. MIT의 경제학자 Erik Brynjolfsson와 그의 동료들은 인공지능 기술이 범용 기술(General purpose technology)임을 강조하며, 범용 기술의 특성상 기술의 파급 효과가 경제 성장으로 나타나기에는 시간이 좀 더 걸릴 것이라고 주장한다(Brynjolfsson, Rock, and Syverson (2018)). 범용 기술은 경제 전반에 영향을 줄 수 있는 기술로 증기기관, 반도체, 컴퓨터, 인터넷 등이 이에 해당한다. 범용 기술이 경제 전반에 적용되어 경제 성장으로 이어지려면, 기술을 실제 경제 활동에 적용할 수 있는 교육된 인력, 회사나 사회 전반의 조직 개편 등이 선행되어야 하는데, 인공 지능 기술을 위한 이러한 변화들이 아직 준비되지 않았다는 것이다.

비슷한 맥락에서 Syverson (2013)은 기술이 경제 성장에 반영되는 것은 물결처럼 몇 번에 걸쳐 오는데, 따라서 최근의 저성장은 물결의 저점에 해당되어 일시적인 것일 뿐, 곧 이전의 성장세를 회복할 수 있음을 시사한다. McKinsey 연구원들은 더 적극적으로 에너지, 생산, 교육, 헬스, 건설, 공공분야 등에서 사례를 들며 경제 성장과 연결될 수 있는 기술 혁신의 가능성을 제시했다. (Bailey, Manyika, and Gupta (2013))

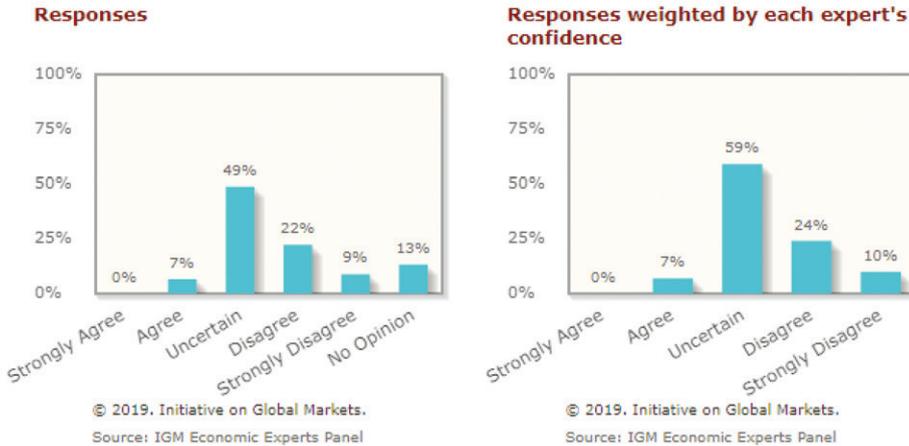
경제학자들 사이의 중론은 Tyler Cowen보다는 좀 더 긍정적이지만 조심스럽다. 2014년에 시카고 경영대에서 저명한 경제학자들 45명을 대상으로 조사한 설문에는 따르면, ‘미래의 혁신은 지난 150여년과 같은 지속적인 경제 성장을 가져오기에 충분히 혁신적이지 않을 것이다’ 라는 주장에 경제학자들은 동의 7%, 불확실 49%, 반대 22%, 매우 반대 9%의 의견을 냈다.⁰⁵

[그림 6]은 ‘미래의 혁신은 미국과 서구 유럽 국가에서 지난 150여년과 같은 수준의 지속적인 경제 성장을 가져오기에는 충분히 혁신적이지 않을 것이다’라는 주장에 얼마나 동의하는지, 답변에 대한 확신은 얼마나 되는지 저명한 경제학자들을 대상으로 설문한 결과를 보여주고 있다. 왼쪽 막대 그래프는 단순히 의견을 종합한 것이고, 오른쪽 막대 그래프는 각 경제학자의 확신도를 가중치로 곱한 후 의견을 종합한 결과를 보여주고 있다.

05 <http://www.igmchicago.org/surveys/innovation-and-growth>

Innovation and Growth

Future innovations worldwide will not be transformational enough to promote sustained per-capita economic growth rates in the U.S. and western Europe over the next century as high as those over the past 150 years.



[그림 6] 기술 혁신과 경제 성장의 미래에 관한 경제학자 대상 설문 결과
(출처: 시카고 경영대학 IGM Economic Experts Panel (2014))

4. 인공지능과 관련한 다른 경제 이슈 - 일자리와 불평등의 문제

사실 인공지능이 우리 경제에 미치는 영향을 생각했을 때 대부분의 사람들은 경제 성장보다는 일자리 문제에 더 관심을 갖는 듯 하다. 대중의 우려는 인공지능 기술이 발전하면서 ‘실업률이 증가할 것이다’, ‘많은 직업군이 사라질 것이다’, ‘인간이 인공지능을 위해 일하게 될 것이다’, ‘인간은 일을 할 필요가 없게 될 것이다 - 그래서 대다수의 인간은 쓸모가 없어질 것이다’, ‘인간은 일을 할 필요가 없게 될 것이다 - 그래서 좋을 것이다’ 등 다양하다.

시카고 경영대에서 경제학자들을 대상으로 조사한 결과, 경제학자들은 대체로 과거의 로봇은 실업률을 높이는데 크게 기여하지는 않았으나 앞으로의 인공지능 기술은 실업률을 높일 수 있고, 그럼에도 불구하고 경제

전반에의 손실보다는 기여가 더 클 것이라는 데에 동의했다.⁰⁶ 인공지능이 가져올 미래를 두려워하는 일반 대중들이 놓치고 있는 것은 (1) 인공지능의 노동대체적인 위협보다 노동보조적인 혜택이 더 클 것이며, (2) 기술의 발전으로 인해 기계로 대체되는 일자리도 있지만, 새로 생겨나는 일자리 또한 많을 것이라는 사실이다. McKinsey Global Institute (2018)에서도 장기적으로 노동 자동화가 가져오는 경제적 이득보다 노동 생산성 향상이나 새로운 상품이나 서비스 개발로 인한 경제적 이득이 더 클 것이라는 분석을 내놓았다.

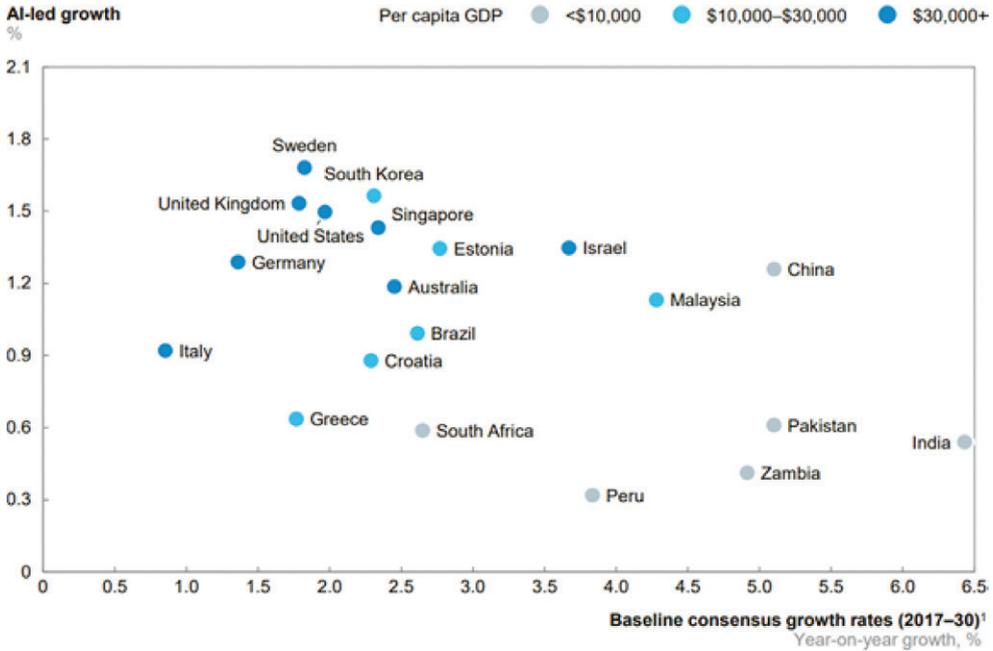
U.C. Berkeley의 경제학자이자 Google의 수석 이코노미스트인 Hal Varian은 좀 더 긍정적인 의견을 내놓았다. 그는 미래에 저출산 등의 인구구조의 변화로 인한 노동 공급의 감소가 자동화로 인한 노동 수요의 감소보다 더 크기 때문에 실업률 문제는 크지 않을 것이라고 주장한다. 또한 한 직업에서 수행하는 업무는 꽤 복잡하며, 업무를 여러 세부 작업으로 쪼개었을 때 인공지능이 업무의 몇몇 ‘작업’을 대체할 수는 있겠지만 이는 보조적 역할을 할 뿐이어서 인공지능이 인간이 행하는 세부 ‘작업’을 없애는 효과보다는 생산성 향상 이득이 더 클 수 있다고 얘기한다.⁰⁷

다만 인공지능 기술이 발전함으로 인해 고학력-저학력 간의 불평등이 커질 수 있다는 점에는 많은 경제학자들이 동의하고 있다. 하지만 이는 과거의 기술 혁신이 가지는 문제와 크게 다르지는 않다. 기술 발전으로 인해 직업을 잃고 피해를 보는 누군가는 항상 존재할 것이다. 결국 불평등의 문제는 우리가 사회적, 정치적, 제도적으로 이 문제를 해결할 의지가 얼마나 있는냐에 달려있다.

마지막으로 인공지능기술을 앞세운 4차 산업혁명에 한국이 뒤처지는 것은 아닌가 하는 우려에 대해서 언급하려고 한다. McKinsey의 예측에 의하면 [그림 7]에서 볼 수 있는 것처럼 미래 한국 경제의 성장에서 인공지능 기술의 기여가 매우 크다고 한다(McKinsey Global Institute 2018). 여러 산업분야의 생산에서 공장 자동화가 많이 되어있는 점 때문에 이러한 예측 결과가 나온 것으로 예상된다. 반면에 서비스 산업에서의 자동화나 새로운 상품과 서비스 개발에서의 인공지능 활용 측면에서는 아직 부족한 면이 많기 때문에 이런 점들이 보완된다면 단기적으로는 인공지능 기술 혁신을 통해 절대적인 성장률이 개선될 수 있는 여지가 있는 것으로 보인다.

06 <http://www.igmchicago.org/surveys/robots>
<http://www.igmchicago.org/surveys/robots-and-artificial-intelligence>
<http://www.igmchicago.org/surveys/robots-and-artificial-intelligence-2>

07 <https://www.gsb.stanford.edu/insights/misplaced-fear-job-stealing-robots>



[그림 7] 국가별 인공지능으로 인한 성장 기여도 (2017-2030)
 (출처: McKinsey Global Institute (2018))

5. 맺음말

인류가 이제껏 경험했던 기술 혁신은 산업혁명 이후 지난 150년동안 지속적인 성장의 원동력이었다. 그런데 국제 금융 위기 이후 약 10여년전부터 미국과 유럽의 경제 선진국들이 국가들이 낮은 경제 성장을 겪으면서, 경제학자들 사이에서 기술 혁신과 그에 따른 생산성의 향상이 예전만 못해서 앞으로는 이전의 성장률을 회복하기 힘들 것이라는 의견들이 나오기 시작했다. 이 글에서는 미래의 기술 혁신을 주도할 핵심 기술 중 하나인 인공지능 기술이 이러한 우려를 없애고 경제 성장의 양상을 바꿀 수 있을지 다양한 경제학 연구와 경제학자들의 의견을 살펴보았다.

‘모든 생산 활동의 완전한 자동화’는 힘들다는 현실을 생각했을 때, 보물의 비용 질병 원리가 적용된다. 즉, 자동화되지 못한, 여전히 인간이 해야 되는 일이 존재하면 전체 경제의 성장률은 인간이 해야 되는 일의 성장 속도가 결정하게 되므로 인공지능 기술이 폭발적인 경제 성장으로 이어지기는 힘들 것으로 보인다. 경제학자들의 의견을 살펴보면 반 정도가 조심스러운 입장 - ‘인공지능 기술이 성장에 어떤 영향을 미칠 지 아직 확실하게 알 수 없다’ - 을 취하고 있다. 나머지 반 중에서는 인공지능

기술을 경제 전반에 응용할 수 있는 환경적 기반이 마련된다면, 인공 지능 기술이 이전의 경제 성장세를 회복하고 지속적으로 성장하는 원동력이 될 수 있을 것이라는 의견이 우세했다.

비록 경제 성장에 관해서는 아직 의견이 모아지지는 않았지만, 경제학자들 대다수가 인공지능 기술이 가져올 경제적 이익은 일자리 감소, 불평등 증가 등에서 비롯 될 수 있는 경제적 손실보다 클 수 있다는 의견에 동의했다.

경제학자들의 말을 믿는다면, '경제적으로는 인공지능 기술에 대한 우려를 크게 하지 않아도 된다'라는 결론을 내릴 수 있겠다. 경제학자들은 관련 기반 기술이나 산업 구조적 변화가 뒷받침된다면 인공지능 기술 덕분에 저성장의 위기를 벗어나 예전의 지속적인 경제 성장세를 회복할 수 있을 것이며, 일자리 손실이 그렇게 크지 않을 수도 있고, 기술의 혜택이 모든 사람에게 공평하지 않을 수는 있겠지만 그래도 득이 실보다 더 클 것으로 보고 있다.

아직 인공지능 기술의 경제 성장의 열매는 열리지 않았다. 우리 경제는 어떤 준비를 해야 할까. 가장 쉽게는 다른 경제 선진국들이 먼저 성장의 열매를 맺기 전에 인공지능 원천 기술 개발 자체에 힘을 쏟는 생각을 할 수 있겠다. 하지만 원천 기술 개발이나 노동 대체적인 인공지능뿐만 아니라 노동 보조적이거나 기술 혁신에 직접 도움이 될 수 있는 인공지능 응용 분야 개척에도 투자를 해야 할 것이다. 뿐만 아니라 관련 기반 기술이나 기술 도입을 위한 제도 정비와 같은 인프라 구축, 통계, 데이터 분석, 컴퓨터 공학 등 인공지능 관련 기초 교육 강화 등을 위한 정책적 고민이 필요하다.

동시에 기술 발전에 혜택보다는 손실을 입은 사람들, 불평등 증가의 피해를 최소화하기 위한 노력, 즉 인공지능이 우리 사회에 가져올 부정적 효과를 최소화 하는 노력도 필요하다. 하지만 시장은 돈이 되지 않는 한, 기술 혁신에 '인간'을 고려하지 않을 것이다. 기술 혁신의 방향은 시장에서 결정한다. 시장 이윤의 논리에 따라 상대적인 인공지능 원천/응용 기술 발전의 속도, 노동대체적/노동보조적/기술혁신보조적 인공지능 기술 발전의 속도가 정해질 것이다. 그래서 인간에게 결과적으로 '나쁜' 인공지능 기술이 많이 개발될 수도 있고, 경제학자들이 생각하는 것보다 많은 일자리가 대체될 수도 있다. 하지만 아직 희망이 있다. 아직 인공지능 기술의 발전 방향은 인류가 주도할 수 있기 때문이다. 다행히 기술 발전 최전선, 시장 논리가 조금은 흐릿한 연구 교육 분야에서 인간 중심적인 인공지능(human-centered AI)을 위한 노력을 하고 있다. 스탠포드 대학은 최근에 인간 중심 AI 센터(Stanford Institute for Human-Centered AI)를 설립하고 컴퓨터공학, 생명공학, 철학, 법학, 경영학, 심리학, 정치학, 역사학 등 다양한 관점에서 인간 중심의 인공지능 기술을 고민하고 있다. MIT 또한 곧

AI 중심 연구 교육 기관으로 인공지능 기술 뿐만 아니라 인공지능 윤리와 인류사회에의 영향을 고민하는 것을 주요 미션으로 하는 단과대학을 설립한다고 한다. 미국의 또 다른 명문 공대를 가지고 있는 카네기멜론 대학에서도 AI 교육 과정에 윤리, 인문, 예술 수업을 포함하고 있다. 우리나라도 큰 세금을 투자해 AI 대학원을 만든다고 한다. 인간에 대한 고민이 얼마나 교육과 연구 과정에 녹아들 지는 의문이다.

특이점이 곧 올 것 같다는 우려를 하던 Tesla의 CEO Elon Musk는 최근 “인간은 과소평가되었다(Humans are underrated)”라며 Tesla의 공장을 지나치게 자동화한 것을 후회했다고 한다. 인간의 능력을 뛰어넘는 인공지능은 쉽지 않을 것이며, 인공지능 기술 개발의 주체도 인간이고, 그 혜택을 받는 것도 인류여야 한다는 사실을 생각하면, 더더욱 인공지능이 가져올 미래를 두려워할 필요가 없겠다.

참고 문헌

- Acemoglu, Daron and Pascual Restrepo, “The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment.” *American Economic Review*, 108(6) (2018)
- Aghion, Philippe, Benjamin F. Jones and Charles I. Jones, “Artificial Intelligence and Economic Growth,” *Working Paper* (2017)
- Baily, Martin Neil, James Manyika and Shalabh Gupta, “U.S. Productivity Growth: An Optimistic Perspective,” *International Productivity Monitor*, Centre for the Study of Living Standards, vol. 25, pages 3–12 (2013)
- Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. “Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics.” In *The economics of artificial intelligence: An agenda*. University of Chicago Press (2018)
- Cowen, Tyler. “The great stagnation: How America ate all the low-hanging fruit of modern history, got sick, and will (eventually) feel better: A Penguin eSpecial from Dutton.” Penguin (2011)
- Hemous, David and Morten Olsen, “The Rise of the Machines: Automation, Horizontal Innovation and Income Inequality,” *Working Paper* (2018)
- Jones, Charles I. “Was an industrial revolution inevitable? Economic growth over the very long run.” *Advances in macroeconomics* 1, no. 2 (2001).
- McKinsey Global Institute, “Modeling the Impact of AI on the World Economy,” *Discussion Paper* September 2018 (2018)
- Nilsson, Nils J., “The Quest for Artificial Intelligence,” Cambridge University Press (2009)
- Syverson, Chad. “Will History Repeat Itself? Comments on ‘Is the Information Technology Revolution Over?’” *International Productivity Monitor*, 25: 37-40. (2013).

인공지능과 시장경제: 정책적 현안과 과제

정해빈(박사과정, 서울대학교 법학대학원)

고학수(교수, 서울대학교 법학전문대학원)

임용(교수, 서울대학교 법학전문대학원)



웹에서 PDF 바로 보기

최근 인공지능은 금융, 의료, 유통, 교육 등 수많은 영역에 광범위하게 침투하면서 시장경제의 지형을 빠르게 변화시키고 있다. 이러한 급속한 변화를 가능하게 만드는 핵심 기술 중 하나가 바로 딥러닝이라는 데이터 학습 기반의 인공지능 기술이다. 그리고 그 배경에는 빅데이터로 알려진 데이터 마이닝 기술의 양적, 질적 개선이 자리잡고 있다.

인공지능을 주어진 문제에 대해 가능한 합리적인 판단을 내리는 체계라고 정의한다면, 딥러닝 인공지능은 연역적으로 도출되거나 사전에 이식된 지식이 아닌, 입력된 문제 상황의 데이터 패턴을 분석하여 습득한 지식을 바탕으로 판단을 내린다는 특징을 지닌다. 이처럼 딥러닝 인공지능 개발에는 학습에 필요한 적정한 데이터의 생성과 활용이 필수적이므로, 인공지능 정책을 설계함에 있어 관련 규제들이 단순히 인공지능이라는 특정 기술에 국한된 것이 아니라, 정보기술 산업 전반에 영향을 미치는 개인정보 및 데이터 규제와 밀접한 관련이 있다는 점에 유념해야 할 것이다. 한편 시장경제와 그 내에서의 경쟁에 대해 인공지능의 발전과 활용이 어떤 영향을 미칠 지에 관하여 다양한 예측과 주장들이 제시되기 시작했고, 그에 대한 실증적 검토와 분석이 필요한 상황이다. 특히 구체적으로 규제의 필요성, 목표, 내용 및 수준을 논함에 있어 미래의 불확실성에 대한 인식, 규제가 해결하고자 하는 우려에 대한 명확한 정의, 사회적 편익과 비용을 모두 고려한 규제 방안의 디자인, 그리고 다층적·다면적으로 이루어질 가능성이 높은 규제들의 누적적·총체적 효과에 대한 신중한 분석과 검토가 요구된다.

1. 논의의 목적

인공지능(artificial intelligence)의 발전에 따른 사회·경제의 변화를 예상한 세계 각국은 인공지능 기술과 응용의 역량 강화를 위한 국가적 차원의 노력을 경주하고 있다.⁰¹ 캐나다는 인공지능에 관한 포괄적인 국가 전략(Pan-Canadian Artificial Intelligence Strategy)을 2017년에 발표하였고, 일본, 중국, 프랑스, 영국 등도 인공지능에 대한 대대적 투자를 준비하고 있다.⁰² 우리나라도 그러한 혁신에 수반되는 경제적 기회와 혜택을 누리기 위해 인공지능 산업을 전폭적으로 지원하거나 관련 규제를 전면적으로 해제하는 등의 국가적 결단이 필요하다는 목소리가 나오고 있고, 인공지능 기술의 빠른 혁신 속도 때문에 그런지 초초함도 없지 않다.

그런데 인공지능 기술의 원리와 그 사회적 영향에 대한 구체적인 이해와 예측 없이 국가적으로 정책 방향을 설정하고 사업을 지원하는 것은 자칫 시장 내 경쟁을 교란하거나 지원대상에 들지 못한 혁신기업들의 성장을 저해할 수도 있다.⁰³ 이 글은 인공지능 혁신과 데이터의 관계를 기초로 오늘날 인공지능의 시장경제에의 파급효과와 관련하여 논란이 되고 있는 주요 현안들을 살펴봄으로써 사회적으로 이로운 인공지능 정책의 설계에 기여하는 것을 목적으로 한다.

2. 논의의 출발

우리는 과학기술의 발전이 그 혜택에 대한 기대와 동시에 사회적 불신과 우려를 야기하는 경우를 종종 목격해 왔다. 인공지능 기술도 다르지 않다. 예를 들어 인공지능이 지적 생명체라는 인류의 특별한 지위를 위협하여 인류의 존속 자체를 어렵게 만드는 중대한 역사적 특이점이 될 것이라는 주장마저 제기되고 있다.⁰⁴ 그러나 인공지능에 막대한 공포와 극단적인 발상들은 기술의 단계적 발전과 사회의 상호작용에 관한 구체적 사고를 자칫 방해할 방해할 수도 있다. 이처럼 정작 필요한 논의가 활발히 이루어지지 못하고 있는 원인에는 인공지능이라는 용어가 매우 다양한 기술과 구현을 애매하고 포괄적으로 지칭하는 식으로 사용되고 있기 때문이기도 한데, 정확한 개념 정의와 논의의 지표 없이 모호하게 인공지능 전반에 관하여 논하는 것은 오늘날 절실히 요구되는 통섭적 고찰에 도움이 되지 않는다.

01 현재 인공지능 기술 연구는 정보를 처리하고 다루는 소프트웨어 측면뿐만 아니라 반도체의 연산능력을 강화하는 등의 하드웨어 측면에도 응용되고 있다(나영식, 조재혁, “인공지능(반도체)”, 「KISTEP(한국과학기술기획평가원) 기술동향브리프», 2019 Vol. 1. (2019), 5-10면; 전승우, “인공지능(AI) 프로세서, 새로운 혁신의 원동력 될까”, 「LG경제연구원 보고서」(2018. 11. 21.), 10면).

02 한국정보화진흥원(NIA), 「AI Network Lab 이슈리포트」 Vol. 3. (2018), 13-24면.

03 Manoj Kewalramani, *China's Quest for AI Leadership: Prospects and Challenges*, Takshashila Working Paper 2018-02 (2018), 17-19면.

04 Robert M. Geraci, *Apocalyptic AI: Religion and the promise of artificial intelligence*, 76(1) *Journal of the American Academy of Religion* (2008), 155-156면.

이 글에서는 현재 빠르게 발전하고 있는 인공지능 기술의 총아 중 하나인 딥러닝(deep learning) 기술에 초점을 맞춰 논의를 진행하고자 한다.⁰⁵ 오늘날 인공지능 혁신을 견인하고 있는 기술이 딥러닝이라는 점에 대해서는 큰 이론이 없을 것이다.⁰⁶ 물론 그 외에도 퍼지 전문가 시스템(fuzzy expert system) 등을 포함한 다양한 인공지능 기술들이 함께 개발·연구되고 있는 것은 사실이나, 이들 기술들은 딥러닝 자체를 대체하기보다는 딥러닝의 부족한 점을 보완하는 측면에서 많이 연구되고 있다.⁰⁷ 그렇다면 딥러닝 기술이 이처럼 인공지능 혁신의 견인차 역할을 하고 있는 이유는 무엇일까? 이 질문에 답하기 위해서는 인공지능이 무엇이고 어떻게 발전 해왔는지에 관하여 간략히 살펴보기로 하겠다.

1. 인공지능의 개념과 전개

인공지능이 무엇인지 설명하기 위하여 역사적으로 “인간처럼(humanly)” 행동하거나 사고하는 것을 포함한 다양한 시도들이 있어왔는데, 오늘날 인공지능 연구는 현실에서 요구되는 실용적 목표에 ‘합리적으로’ 도달하는 것에 역량을 집중하는 경향을 보이고 있다.⁰⁸

주어진 정보에 대해 합리적으로 행동하는 인공지능을 개발하기 위한 노력은 크게 두 갈래로 나뉜다. 첫 번째는 ‘지식베이스(knowledge base)’ 접근 방식이다. 이 방식은 입력된 문제 상황에 대해 답변을 산출하는 판단의 논거로서 사전에 인간에 의해 삽입된 지식을 활용하며, 이를 ‘전문가 시스템(expert system)’이라 부르기도 한다.⁰⁹ 그런데 전문가 시스템은 판단 과정에서 기반지식 내에 상충하는 지점이 있는 경우, 인공지능이 스스로 기반지식 간의 우열을 판단할 수 없기 때문에 외부적으로 인간이 개입해서

05 따라서 미래에 언젠가 등장할지도 모르는 여하한 인공지능 기술과 그 구현이 촉발할지도 모를 어떤 사회적 우려는 이 글의 논의의 대상이 아니다.

06 딥러닝은 20세기 후반부터 연구되어왔던 머신러닝(machine learning) 기술의 연장선상에 있는 기술로서, 학술 데이터베이스 SCOPUS에 등재된 인공지능 관련 연구 중 머신러닝과 확률적 추론 분야에 관한 논문이 2010년에 28%에서 2017년에는 56%에 육박하는 수준으로 증가하였다고 한다(Yoav Shoham et al., *The AI Index 2018 Annual Report* (Stanford, AI Index Steering Committee, Human-Centered AI Initiative at Stanford University, 2018), 9면).

07 오늘날 인공지능 분야는 한 가지 기술만을 단독으로 사용하기보다는 의도된 목표에 맞는 최적의 성과를 낼 수 있도록 다양한 기술을 결합하는 앙상블(ensemble) 기법을 따르고 있다. 예를 들어 Jui-Sheng Chou & Anh-Duc Pham, *Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength*, 49 *Construction and Building Materials* (2013), 558-560면 등 참조.

08 Stuart Russell & Peter Norvig, *Artificial Intelligence: a Modern Approach* (3rd ed., 2009), 2-5면.

09 Michael Negnevitsky, *Artificial Intelligence: a Guide to Intelligent Systems* (2nd ed., 2005), 10-12면.

기반지식을 업데이트해야만 한다는 근본적인 한계가 있고, 그로 인한 개발상의 난점들이 있다.

그 대안으로 등장한 두 번째 방식이 바로 ‘머신러닝(machine learning)’인데, 문제에 관한 데이터의 패턴으로부터 기반지식을 습득하여 활용하는 방식이다.¹⁰ 머신러닝에 따른 기반지식은 인공지능 외부로부터 삽입된 규칙들을 연역적으로 구성한 결과가 아니며, 주어진 데이터의 패턴에 따라 인공지능 내부로부터 경험적으로 구성된 것이다. 이처럼 머신러닝은 데이터로부터 기반지식을 축적하는 관계로, 훈련용으로 사전에 확보한 데이터는 물론 새롭게 입력되는 문제 상황도 기반지식을 업데이트하는 데이터로 받아들일 수 있고, 그 결과 끊임없이 상황에 부합하도록 기반지식을 향상시킬 수 있는 유연함이 장점이다. 시대가 변하고 이에 따라 판단기준이 변할 때마다 인간이 그 기준을 직접 변경시켜줘야 하는 지식베이스 접근방식과 달리, 머신러닝은 변화하는 환경의 데이터에 따라 자연스럽게 판단기준이 변하게 되는 것이다.

그런데, 예를 들어 ‘절도’라는 단일한 개념이 현실에서 벌어질 수 있는 다종다양한 여러 행위들을 포괄하고 있듯이, 하나의 개념(concept)에 해당하는 여러 데이터 사이에는 편차(factors of variation)가 존재할 수밖에 없다. 이 상황에서 인간이 의도한 바에 따라 분류되도록 데이터를 적절하게 표현(representation)하여 알고리즘에 반영하는 것은 쉽지 않고, 머신러닝 기술은 이 지점에서 오랫동안 난관에 봉착하게 된다.¹¹ 이 장기간의 난관을 극복하고 머신러닝의 다음 단계를 열어젖힌 것이 바로 오늘날의 딥러닝 기술이다.

2. 딥러닝 인공지능의 등장

1) 딥러닝과 빅데이터

딥러닝의 등장 배경에는 빅데이터(big data)라 불리는 데이터 공학의 혁신적인 발전이 자리잡고 있다.¹² 빅데이터는 데이터 분석의 양적, 그리고 질적 발전이라는 두 측면에서 바라볼 수 있는데, 양적 측면에서는 다량의 데이터를 한꺼번에 다룰 수 있는 기술적 개선이, 질적 측면에서는 데이터 마이닝(data mining) 기술의 발전에 힘입어 비정형적(unstructured) 자료들을 처리할 수 있는 기술 개발을 그 원동력으로 삼고 있다.

데이터 공학의 개선은 머신러닝 인공지능이 데이터를 처리하는 모델 자체의 개선에도 기여하였다. 예를 들어 인공지능이 감시카메라

10 Ibid. 165-168면.

11 Ian Goodfellow, Yoshua Bengio & Aaron Courville, *Deep Learning* (2016), 2-4면.

12 장병탁, 「장교수의 딥러닝」, (2017), 8면.

영상으로부터 절도의 현행범을 포착하는 작업을 상상해보자. 종래의 머신러닝은 영상에 나온 행동이 절도에 해당하는지 여부를 그 행동 자체로부터 곧장 판단하려 했다면, 딥러닝은 판단 대상인 하나의 행동을 여러 단계로 세분화하여 학습한다는 차이가 있다. 절도라는 중층적인 개념을 걷기, 뛰기 등 간단한 표현으로부터 쌓아 올려 판단함으로써 다종다양한 절도의 형태에 대해 유연하게 판단을 내릴 수 있게 된 것이다. 딥러닝 기술은 세상을 “개념들의 계층(hierarchy of concepts)”으로 표현하는 셈이다. 이 같은 머신러닝 인공지능 모델의 기술적 개선은, 빅데이터로 대표되는 데이터 공학 기술의 발전과 함께 상호작용하며 딥러닝 기술의 발전을 촉진하였다.¹³

2) 딥러닝의 한계

결국 딥러닝 인공지능은 판단에 필요한 지식을 데이터로부터 습득한다는 머신러닝의 기본 골격은 바꾸지 않은 채 데이터 처리기술을 대폭 개선함으로써 성능을 개선시킨 것이라고도 할 수 있다.¹⁴ 따라서 딥러닝 인공지능 기술은 한번 개발되면 모든 종류의 문제해결에 범용적으로 활용 가능한 인공지능(general problem solver)은 아니고, 설정된 목표와 훈련용 데이터에 맞게 최적화되어 작동하되 훈련용 데이터의 종류를 가리지 않는다는 의미에서 범용적 알고리즘 기술이라 할 수 있다.¹⁵

그러므로, 딥러닝 기술이 효과적으로 작동하려면, ① 인공지능에게 판단에 필요한 충분한 지식을 습득하도록 만들 수 있을 정도로 많은 양의 데이터가 적정하게 확보되어야 하고,¹⁶ ② 최적화의 목표가 분명하고 수량적으로 규정 가능해야 하며, ③ 판단 대상이 기존 훈련용 데이터에 비추어 이례적이지 않아야 한다는 조건이 전제되어야 한다.¹⁷ 뿐만 아니라 딥러닝 기술은 인공지능의 기본 알고리즘을 결정하고 데이터의 훈련방법을 설정하는 등의 작업을 해낼 수 있는 숙련된 인공지능 전문가를 필요로 한다. 이와 같은 물적(데이터), 인적(전문가) 조건이 갖춰지지 않을 경우 딥러닝 인공지능의 효과가 충분히 발휘되기 어렵다.

나아가, 딥러닝 인공지능이 판단을 내리는 논거는 데이터로부터 수집한 패턴 그 자체라는 점을 분명하게 인지할 필요가 있다. 예를 들어 딥러닝

13 Daniel E. O'Leary, *Artificial intelligence and big data*, 28(2) IEEE Intelligent Systems (2013), 97면.

14 Goodfellow, Bengio & Courville, 18-20면.

15 추형석, “인공지능, 어디까지 왔나?”, 『SPRi(소프트웨어정책연구소) 이슈리포트』, Vol. 6. (2018), 2-4면. 예를 들어 광고효과를 늘리기 위해 광고 데이터로 학습시킨 딥러닝 인공지능을 기상을 예측하는 데 곧장 활용할 수는 없는 것이다.

16 Forbes Insights, *Inside The R&D Of AI Ethics* (2019. 3. 27.)

17 Gary Marcus, *Deep learning: A critical appraisal*, arXiv preprint arXiv:1801.00631 (2018), 12-14면.

인공지능이 X라는 요소를 지닌 사람이 암이 걸릴 가능성이 높다고 진단을 했다면, 이는 X라는 요소가 암으로 연결되는 역학적 기전에 대한 규명에 근거한 것이 아니라, X라는 요소를 지닌 사람에게 암이 발병한 사례가 많기 때문이라는 통계적 논거에 기초한 진단이다. 딥러닝 인공지능의 판단이 불투명하다고 일컬어지는 이유 역시 인간을 뛰어넘는 현명한 판단 때문이라기보다는, 판단에 쓰이는 통계적 논거들이 비선형적이고 복잡하기 때문에 나타나는 현상이라 할 수 있다. 이처럼 딥러닝 인공지능은 현재 통계적 논거에 따라 조치를 선택하는 데 적절한 석유 탐사와 같은 이학적 판단에 대체로 적합하고, 형량을 위한 가치 판단이나 사회적 합의를 필요로 하는 사회적 문제의 최종 결론을 도출하는 데 바로 활용하는 것에는 한계가 있다.¹⁸

3) 딥러닝에 의한 혁신의 실제

딥러닝 인공지능의 혁신이 데이터 공학에 터잡고 있는 관계로, 현재 인공지능에 기반한 혁신이라 알려진 수 많은 사례들이 공통적으로 (i) 데이터, (ii) 최적화라는 두 가지 요소에 바탕을 두고 있음을 알 수 있다. 이 두 가지 요소에 따라 최근 인공지능에 의한 혁신이라 알려지고 있는 사례들을 크게 아래의 네 가지 유형으로 분류하는 것이 가능하다.

데이터 부족	데이터 풍족	
유형 II (예시) 여신 거래 데이터가 부족한 개발도상국에서 통화기록 등 대안적 데이터로 대출심사	유형 I (예시) 여신 거래 데이터가 풍부한 선진국에서 기존 신용대출 고객들의 부도 가능성 식별	최적화 용이
유형 III 혁신 달성 어려움	유형 IV (예시) 여신 거래 데이터가 풍부한 선진국에서 새로운 금융서비스를 개척	최적화 난해

18 이른바 창의적인 인공지능의 예시로 알려지고 있는 구글의 알파고(Alpha Go)가 바둑에서 우수한 성능을 낼 수 있는 이유 역시 바둑이라는 영역에서는 훈련용 데이터의 확보가 쉽고 달성해야 할 최적화의 목표가 분명하다는 점에 있다. 바둑처럼 게임 내에서 가능한 행동과 승리 조건이라는 목표가 명확한 영역에서는 훈련용 데이터를 엔지니어들이 직접 생성해낼 수 있기 때문이다. 인간이 과거에 두었던 바둑의 기보를 학습하는 것을 넘어서, 인공지능끼리 경기한 기보를 데이터로 생성하고, 이를 학습에 다시금 반영하는 등, 데이터를 대량으로 학습하고 이를 판단기준에 반영하는 피드백에 의해, 바둑에서 딥러닝 인공지능 알고리즘은 인간보다 더 우수한 문제해결 능력을 보일 수 있게 되었다(David Silver, et al., *Mastering the game of Go with deep neural networks and tree search*, 529 Nature (2016), 486면).

- **유형 I** : 데이터가 풍부하고 최적화가 용이한 경우, 딥러닝 기술을 적용하는 것만으로도 인간에 의한 작업에 비해 상당한 생산성 개선을 거둘 수 있다. 보험처리 데이터를 바탕으로 전형적인 상해로 인한 보험금 지급 결정을 자동화하거나,¹⁹ 의료 데이터를 기준으로 환자를 신속하게 진단 분류할 수 있는 의료 인공지능의 개발,²⁰ 물류 데이터를 이용하여 유통비용을 절감하는 인공지능,²¹ 행정정보를 이용하여 청소년의 중퇴가능성을 판단하는 인공지능²² 등의 혁신 사례들은 정형화된 데이터를 추론과 판단에 활용함으로써 생산성 개선을 추구한다는 점에서 모두 이 유형의 혁신에 해당한다고 볼 수 있다.
- **유형 II** : 데이터가 부족하지만 목표가 분명한 경우, 유사한 효과를 내는 다른 데이터를 이용함으로써 역시 생산성 개선을 거둘 수 있다. 예를 들어 전통적 은행업의 미발달로 인해 여신거래 내역 등 고전적인 신용거래 데이터가 부족한 개발도상국에서 통화내역 또는 스마트폰 접속기록 등 새로운 종류의 데이터를 소비자 신용정보 산정에 활용하여 포용적 금융(financial inclusion)을 지원하는 인공지능의 사례나, 납이 함유된 수도관이 매장된 위치를 찾기 위해 주택의 건설시점이나 건축 형태를 분석하는 인공지능²³ 등의 사례가 이에 해당한다.²⁴

19 이응용, “금융 생태계를 변혁하는 인공지능의 확산 및 시사점”, 『KISA(한국인터넷진흥원) REPORT』, 2018 Vol. 9., (2018), 1-9면.

20 의료는 데이터가 풍부하면서도 치료라는 목표가 분명하다는 점에서 딥러닝 인공지능이 효과적으로 응용될 수 있는 분야로 손꼽히고 있다(CADTH, *An Overview of Clinical Applications of Artificial Intelligence*, 174 CADTH issues in emerging health technologies (2018), 5-11면; AHSN Network, *Accelerating Artificial Intelligence in health and care: results from a state of the nation survey* (2018), 24-25, 47면; KPMG, *How the UK can win the AI race* (2018), 12-16면 참조). 이와 관련된 국내 사례로 이영재, “루닛, 미국암학회 장식...AI 분석 항암제 반응 예측”, 의협신문 (2019. 3. 28.)과 김민수, “응급환자 위중도 AI 이용해 판별...뷰노, 첫 연구결과 공개”, 조선비즈 (2018. 10. 16.) 등 참조.

21 Deloitte, *Artificial Intelligence Innovation Report 2018* (2018), 15-25면; Mayor of London, *London: The AI Growth Capital of Europe* (2018), 68-69면.

22 Juan O. Freuler & Carlos Iglesias, *Algorithms and Artificial Intelligence in Latin America: A Study of Implementation by Governments in Argentina and Uruguay* (2018), 15-22면.

23 McKinsey Global Institute, *Notes from the AI Frontier – Applying AI for Social Good* (2018), 18-27면. 관련 기사로 Una Softic, *Artificial intelligence fosters financial inclusion in Asia*, Nikkei Asian Review (2018. 10. 3.) 참조.

24 이 유형에서 말하는 데이터의 부족이란 종래의 판단기준으로 삼아온 전통적 데이터가 부족한 것을 의미하고, 혁신을 위해서는 이와 같이 부족한 전통적 데이터의 자리를 메울 대안적 데이터가 필수적으로 요청된다. 아무런 데이터 없이 인공지능이 관상만 보고 금융소비자의 신용도를 알아 맞출 수는 없다.

- **유형Ⅲ**: 데이터가 부족한데 목표마저 불분명한 경우 핵심적인 과제를 수행함에 있어 딥러닝 인공지능이 인간을 대신하여 적절한 성과를 바로 내기 어렵다. 적용범위나 대상조차 정의되어 있지 않은 전혀 새로운 기술의 개발 또는 새로운 사회적 규범의 정립과 같이 전례를 유사하게 모방하기보다는 창의성과 연역적 논리가 중요한 영역이 그 예다.
- **유형Ⅳ**: 데이터가 풍족하나 목표가 불분명한 경우, 딥러닝 인공지능은 불분명한 목표를 뚜렷한 수량적 목표로 근사하여 처리함으로써 생산성의 개선을 꾀할 수 있다. 예를 들어 딥러닝 기술은 프로 스포츠 선수들의 능력을 판단하고자 종래에는 사용되지 않던 새로운 지표를 개발하는 등의 작업에 사용될 수 있다.²⁵

위와 같은 분류를 통해 우리는 오늘날 딥러닝 인공지능을 통한 혁신에 유효 적절한 데이터의 수급이 중요하다는 점을 알 수 있다. 즉, 현재 인공지능의 혁신은 기본적으로 데이터 의존성을 보이고 있다. 데이터 없이도 혁신을 이뤄냈다는 사례들은 대부분 유형Ⅱ나 유형Ⅳ에 해당하는 것으로, 종래 지표에 활용되었던 전통적인 데이터가 부족하다는 것일 뿐 이를 대체할 데이터 또는 새로운 지표를 가공하는데 필요한 데이터는 풍족했던 상황을 알 수 있다. 이는 향후 (딥러닝) 인공지능 정책에 관한 정책을 설계함에 있어 데이터 규제와의 연관성을 반드시 고려해야 함을 시사해 준다.²⁶

3. 인공지능과 시장경제 정책의 현안

앞서 살펴본 딥러닝 기술의 특징과 이를 통한 혁신 창출에 있어서의 데이터의 역할은 인공지능에 관련된 시장경제 정책을 설계할 때 그 목표와 방향에 관하여 시사점을 제공한다. 아래에서는 현 시점에서 그와 관련하여 문제되고 있는 사회의 현안들을 간략히 살펴보도록 하겠다.

딥러닝 인공지능 기술의 개발과 사용에 필요한 하드웨어 프로세서 가격의 하락과 클라우드 컴퓨팅의 발전에 따른 비용 절감에 더해 인공지능에 의한 생산성 혁신이 가속화되면 개인들은 더 우수한 수준의 상품과 서비스를 더 낮은 가격에 향유할 수 있을 것이다.²⁷ 또 인공지능 기술은 교통사고나 환경오염, 산업재해 등의 위험을 축소시킴으로써 개인의 삶을 개선하는 다양한 편익을 가져다 줄 것으로 기대되고 있다.²⁸

25 Edd Gent, *How AI is helping sports teams scout star players*, NBC (2018. 6. 13.).

26 대한민국 정부 관계부처 합동, “혁신성장 전략투자, 데이터·AI경제 활성화 계획 보도자료” (2019. 1. 16.).

27 CADTH, 5-11면.

28 Little Hoover Commission, *Artificial Intelligence: A Roadmap for California* (Little Hoover Commission, Report 245, 2018), 8면.

하지만 인공지능의 도입으로 인한 편익만 있는 것은 아니다. 인공지능의 도입이 유발할 수 있는 사회적 비용(외부효과 포함)으로는 개인에 대한 가격 차별로 인한 소비자 잉여의 상실, 강화된 프로파일링에 기초한 차별,²⁹ 그리고 궁극적으로 인공지능에 의한 개인의 자율성 침해와 그로 인한 민주주의의 위기의 도래 등이 거론되고 있다.³⁰ 이 중 특히 논란이 되고 있는 것 중 하나는 인공지능이 유발할 수 있는 차별 및 불평등 심화의 문제다. 예를 들어 아마존은 딥러닝 인공지능 기술로 지원자의 자기소개서를 분석하여 면접대상자를 선발하려 했으나, 인공지능이 남성적 어휘를 사용하는 지원자를 우대하는 등 문제를 일으킨 탓에 이를 포기한 바 있다.³¹

이처럼 딥러닝 인공지능이 개인에 관한 데이터를 기반으로 상용화되는 경우 어려운 문제를 유발할 수 있다. 개인에 관한 데이터를 분석함에 있어서는 ① 평가항목 설계의 타당성과 ② 평가자의 의도라는 두 가지 측면에서 왜곡이 잠재되어 있을 가능성이 있는데, 딥러닝 인공지능은 이러한 왜곡을 구분하여 데이터를 학습하기 어려우므로 기존의 훈련용 데이터에 내재되어 있던 왜곡을 더욱 체계적으로 학습할 가능성이 높기 때문이다. 인공지능을 사용한 차별은 단순히 행위자의 개인적 편견에 의존하는 것이 아니라 집합적 데이터를 바탕으로 차별을 재생산한다는 점에서 “경로의존적(path-dependent)”으로 차별에 기여하는 셈이다.³² 과거의 데이터에 차별이 반영되어 있는 경우 외에, 데이터가 상대적으로 희소한 경우에도 딥러닝 인공지능이 부당한 대우를 할 수 있다. 예를 들어 특정 소수집단에 관한 정보가 포함된 데이터로 훈련하는 딥러닝 인공지능이 이러한 정보를 데이터의 패턴으로 인식하기보다는 일시적인 아웃라이어(outlier)로 여겨 소수의 입장에서 부당해 보일 수 있는 다수에 적절한 결과를 그대로 도출 내지 적용할 가능성이 있다. 다만, 이러한 인공지능과 차별의 문제를 제어하기 위해 어떤 종류의 정책적 수단을 도입할 것인지에 관하여는 의견이 갈리고 있으므로, 데이터 규제와 인공지능에 의한 사회경제적 차별의 관계에 대해서는 신중한 접근이 필요하다.³³

29 Giovanni Sartor, *Artificial Intelligence: Challenges for EU Citizens and Consumers*, European Parliament IP/A/IMCO/2018-16, 4-7면.

30 Chris Reed, Elizabeth Kennedy & Sara Silva, *Responsibility, Autonomy and Accountability: Legal Liability for Machine Learning*, Queen Mary School of Law Legal Studies Research Paper No. 243/2016 (2016), 19-21면.

31 Jeffrey Dastin, *Amazon scraps secret AI recruiting tool that showed bias against women*, Reuters (2018. 10. 10.).

32 Betsy Anne Williams, Catherine F. Brooks & Totam Shmargad, *How Algorithms Discriminate Based on Data They Lack*, 8 Journal of Information Policy (2018), 91-95면.

33 고학수, 정해빈, 박도현, “인공지능과 차별”, 『저스티스』 171 (2019), 242-265면.

한편 매킨지는 대한민국에서 인공지능 기술의 도입이 다른 국가들에 비해 빠르게 높은 수준으로 이뤄질 것이며, 이에 따라 인공지능 기술 도입에 의한 경제성장을 감안할 때 대한민국은 인공지능의 혜택을 많이 받는 국가 중의 하나가 될 것이라 예측하고 있다.³⁴ 이처럼 딥러닝 인공지능 기술은 경제적으로 새로운 산업을 창발하고 높은 부가 가치를 생산할 것으로 기대되지만, 동시에 인간의 노동을 대체함으로써 노동시장의 한 쪽에서는 실업을 유발할 것이라는 우려도 있다. 이러한 실업 발생은 사회정치적 불안정과 수요측면의 경제적 침체를 일으키고, 실업자를 재교육하고 재배치하는 과도기적 비용으로 이어질 수 있다. 그러나 실업의 실질적인 모습과 구체적인 규모에 관해서는 예측이 첨예하게 엇갈리고 있다. 인공지능이 자동화된 형태로 업무를 어디까지 대체할 수 있을지에 관하여, 경제협력개발기구(OECD)는 인공지능에 의해 대체될 수 있는 고위험군을 9%로 추정한 반면, PwC는 38%로 추정하고 있다.³⁵ 여기서 유의할 점은 인공지능의 비용 효과에 따라서 국가마다 인공지능 도입에 따른 경제적 충격이 다를 수 있다는 것이다.³⁶ 그러므로 인공지능 시대의 대규모 실업에 대비한다고 하여 로봇세(robot tax)³⁷ 또는 보편적 기본소득(universal basic income)³⁸ 등의 전면적인 정책수단이 반드시 필요하다고 단정하기엔 앞서 분야 및 업무 유형별로 예측되는 실업 발생의 가능성과 규모, 그리고 정책수단의 실행에 소요될 자원과 사회적 비용 등에 대한 선행적 분석을 하는 등 신중하게 접근할 필요가 있다.

한편 현재 데이터와 이를 활용하는 인공지능(알고리즘)의 활용이 시장에서의 경쟁과 소비자 후생에 미칠 영향에 관해서도 다양한 의견들이

34 McKinsey Global Institute, *Notes from the AI frontier - Modeling the impact of AI on the world economy* (2018), 33-38면.

35 김건우, “인공지능에 의한 일자리 위험 진단 - 사무·판매·기계조작 직군 대체 가능성 높아”, 「LG경제연구원 보고서」 (2018. 5. 15.), 3-4면.

36 딥러닝 인공지능이 인간에 비견할 정도로 또는 그보다도 더 우수하게 업무를 수행할 수 있는가의 문제 만큼이나 인간과 비교한 수행 비용이 얼마나 경제적인가도 중요한 문제다. 소프트웨어인 인공지능과 이를 탑재한 하드웨어인 로봇을 도입하는 것이 비용의 측면에서 인간을 채용하여 숙련시키는 것에 비해 적어도 동등한 품질의 노동을 산출하면서도 저렴해야 인공지능에 의한 인간 노동의 대체가 본격적으로 시작될 수 있기 때문이다. 테슬라(Tesla)는 전기자동차 생산공정의 완전한 자동화를 공언하였으나, 현재로서는 로봇 공학과 자동화 기술이 충분히 발전하지 못한 탓에 오히려 인간노동보다 생산 속도가 크게 낮아지는 등 생산비용이 실질적으로 증가함으로써 위기를 겪었다는 보도도 있다(Samuel Gibbs, *Tesla halts Model 3 production as firm scrambles to improve automation*, The Guardian (2018. 4. 17.)).

37 Ryan Abbott & Bret Bogenschneider, *Should Robots Pay Taxes: Tax Policy in the Age of Automation*, 12 Harv. L. & Pol’y Rev. (2018), 168-169면.

38 Anton Korinek & Joseph E. Stiglitz, *Artificial intelligence and its implications for income distribution and unemployment*, NBER No. w24174 (2017), 32-33면.

개진되고 있는데,³⁹ 논란이 되고 있는 주요 이슈로는 ① 디지털 사업자의 특수성, ② 데이터 관련 규제가 시장경쟁에 미치는 영향, ③ 합병이 혁신과 소비자후생에 미치는 영향, ④ 알고리즘에 의한 답함 가능성 등을 들 수 있다.⁴⁰

이 중 특히 디지털 사업자의 특수성과 관련해서는 아마존, 구글, 페이스북 등 이른바 플랫폼(platform) 사업자들이 경쟁법에서 우려할만한 시장력을 실제로 보유하고 행사할 수 있는지 여부가 문제되고 있다. 플랫폼 사업자가 그 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 이른바 ‘네트워크 효과(network effect)’가 발생할 수 있으며, 그 결과 이용자들의 ‘전환 비용(switching cost)’이 증가하고 관련시장에서의 진입장벽이 높아질 수 있기 때문이다. 다만, 기존 사업자들이 서비스 과정에서 이용자들의 데이터를 대량으로 보유한다는 것이 실제로 진입장벽을 형성하고 반경쟁적인 효과를 발생시킬 수 있는지에 대해서는 논란이 있다. 데이터가 많다고 하여 그에 비례하여 서비스의 질이 무조건 개선되는 것은 아니며, 데이터를 관리하고 분석할 수 있는 기술력 등 경쟁에 영향을 미치는 다양한 요소들도 있기 때문이다. 나아가 오늘날 온라인 플랫폼 사업자들이 검색, 이메일, SNS, 클라우드 스토리지 등 수많은 서비스를 무료로 소비자들에게 제공한다는 점에서 소비자들의 후생이 과연 그러한 플랫폼 사업자들에 의해 악화되고 있는 것인지에 대해 의문을 제기하는 견해도 있다.

한편 데이터를 둘러싼 다양한 규제들과 경쟁법상의 규제의 관계에 대해서도 심도 있는 고민이 필요하다. 그 한 예로 데이터 이동(data portability)에 관한 규제를 들 수 있다. 데이터 규제는 프라이버시 보호, 보안 및 관리의 신뢰성을 확보하기 위한 투명성 조치가 결합하여 정보기술 분야에서 독자적인 법 논리를 구축해왔다. 이를 정보주체의 데이터에 대한 접근과 통제가 강조되면서 개인이 자신에 관하여 사업자가 보유하고 있는 데이터를 받아서 다른 사업자에게 이전할 수 있는 데이터 이동권(right to data portability)이 중요한 개념으로 대두되기에 이르렀다. 이러한 권리는 유럽연합(EU)이 2018년부터 시행 중인 일반정보보호규정(General Data

39 Jason Furman & Robert Seamans, *AI and the Economy*, 19(1) Innovation Policy and the Economy (2019), 175-183면.

40 U.K. Digital Competition Expert Panel, *Unlocking digital competition* (2019. 3. 13.), 8-12면.

Protection Regulation: GDPR)에 명문화 되었고,⁴¹ EU는 금융 영역에 관한 PSD 2 지침(Payment Service Directive 2)에서도 이동권에 관한 규정을 삽입하였다.

그러나 개인이 사업자로부터 자신에 관한 데이터를 반출하여 다른 사업자에 이전할 수 있도록 돕는 권리가 실제로 경쟁을 촉진하는데 도움이 될 것인지는 명확하지 않다.⁴² 이를 통해 데이터 플랫폼 사이에 경쟁이 촉진될 가능성도 있지만, 오히려 사용자들이 데이터 보안 또는 개인정보 보호가 불안해 보이는 신생 또는 후발기업으로부터 선두기업으로 데이터를 이전하여 기존 플랫폼 사업자에 대한 고착화(lock-in)가 더욱 가속화될 가능성도 있기 때문이다. 더구나 데이터 이동권의 보장은 사업 현장에서 상당한 비용과 어려움을 초래할 수 있다. 무엇보다 사업자간의 데이터에 대한 표준화 및 상호운용성을 확보하는 과정이 필수적일 텐데, 데이터 기술의 차이, 다양한 이해관계, 비용부담 등 여러 요소를 고려하면 그 과정이 간단치 않을 것임을 쉽게 예상할 수 있다. 다만, 데이터 이동권의 보장이 전환비용을 감소시킴으로써 적어도 이론적으로는 플랫폼의 직간접적인 네트워크 효과로 인한 경쟁 저해의 우려를 약화시킬 여지가 있는 만큼 정확한 판단을 위해서는 실증적인 검토가 필요하다고 하겠다.

4. 추가적 정책 과제 및 결어

인공지능 정책에 관하여 이 글에서 논한 정책적 현안 외에 추가적으로 다뤄져야 할 과제로는 다음과 같은 것을 수 있다.

- (1) **위험의 제재와 규명 절차의 정립:** 딥러닝 인공지능은 ‘인간처럼’ 행동하고 사고하는 독립 주체가 아니라는 점에서 인공지능에 의한 위험의 규율이 인공지능 그 자체에 대한 통제로 논리필연적으로 이어지는 것은 아니다. 그러나 인공지능을 활용하는 인간에 대한 규율은 의도 및 과실을 책임의 중요한 요건으로 여기는 전통적 법체계에서 어려움을 겪고 있으며, 딥러닝 인공지능 기술의 불투명성(opacity)으로 인해 위험의 발생원인과 당사자들의 기여를

41 GDPR Art. 20(1) 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.

42 Heike Mai, *PSD 2, open banking and the value of personal data*, Deutsche Bank Research (2018. 6. 28.), 5면.

사후적으로 재판에서 규명하는 것에 어려움이 많다. 인간이 유발하는 위험은 배상 등 민사적 제재, 허가 등 행정적 통제와 처벌 등 형사적 제재로 규율되어 왔는데, 이러한 제재와 제재를 부과하는 절차 모두가 인공지능이라는 새로운 도전 앞에서 변화가 필요한 것은 아닌지 살펴보아야 할 것이다.

(2) **과세제도의 검토:** 인공지능과 로봇이 생산과정에서 인간을 더 많이 대체할수록, 최종적 부가가치의 생산에 인간이 기여하는 부분은 줄어든다. 이에 대응하여 과세제도가 변화해야 할 필요가 있는지, 설령 이러한 변화가 문제라 하더라도 기존 과세제도를 탄력적으로 운용하는 것만으로도 해결이 가능한 것은 아닌지를 살펴볼 필요가 있다.⁴³

(3) **정책간 조율의 문제:** 인공지능과 데이터 정책에는 인공지능 등 산업적 활용, 차별금지, 사생활 및 개인정보보호, 경쟁정책 등 전혀 다른 쟁점들이 뒤섞여 있으므로, 이러한 다각적인 가치들을 조화시킬 수 있는 정책적 조율이 필요하다. 이 과정에서 어떤 사회적 가치를 어느 범위에서 우선할 것인지에 대한 합의가 반드시 필요하며, 가치 간의 조율을 일관되게 해낼 수 있도록 인공지능과 데이터를 규율할 행정기관이나 심판기관이 새롭게 필요한 것은 아닌지 생각해 볼 필요가 있다.

이처럼 인공지능 관련 정책 및 규제들의 입안, 집행 및 해제 필요성을 판단함에 있어 산업적 기회에 주목하는 편면적 시각을 넘어서 실업, 차별, 독과점, 사회윤리, 형사법 체계 등을 아우르는 융합적인 상상력이 필요하다. 각각의 분야가 기존에 연구해왔던 논리를 인공지능이라는 이슈에 덮어씌우는 것이 아니라, 인공지능이라는 새로운 사회경제의 원동력을 어떠한 원리로 규율해야 사회의 전 분야에 일관되고 통일적인 정책을 수립할 수 있는지를 고민해야 하는 것이다. 그런 관점에서 규제의 필요성, 목표, 내용 및 수준을 구체적으로 논함에 있어 미래의 불확실성에 대한 인식, 규제가 해결하고자 하는 우려에 대한 명확한 정의, 사회적 편익과 비용은 모두 고려한 규제 방안의 디자인, 그리고 다층적·다면적으로 이루어질 가능성이 높은 규제들의 누적적·총체적 효과에 대한 신중한 분석과 검토가 요구된다고 하겠다.

43 Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 UC DL Rev. (2017), 425-7면.

단행본

장병탁, 「장교수의 딥러닝」(2017)
Ian Goodfellow, Yoshua Bengio & Aaron Courville, *Deep Learning* (2016)
Michael Negnevitsky, *Artificial Intelligence: a Guide to Intelligent Systems* (2nd ed., 2005)
Stuart Russell & Peter Norvig, *Artificial Intelligence: a Modern Approach* (3rd ed., 2009)

학술논문

고학수, 정해빈, 박도현. “인공지능과 차별.” 『저스티스』 171 (2019)
Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 UC DL Rev. (2017)
Robert M. Geraci, *Apocalyptic AI: Religion and the promise of artificial intelligence*, 76(1) Journal of the American Academy of Religion (2008)
Gary Marcus, *Deep learning: A critical appraisal*, *arXiv preprint arXiv:1801.00631* (2018)
Daniel E. O’Leary, *Artificial intelligence and big data*, 28(2) IEEE Intelligent Systems (2013)
Jason Furman & Robert Seamans, *AI and the Economy*, 19(1) Innovation Policy and the Economy (2019)
Betsy Anne Williams, Catherine F. Brooks & Totam Shmargad, *How Algorithms Discriminate Based on Data They Lack*, 8 Journal of Information Policy (2018)
Chris Reed, Elizabeth Kennedy & Sara Silva, *Responsibility, Autonomy and Accountability: Legal Liability for Machine Learning*, Queen Mary School of Law Legal Studies Research Paper No. 243/2016 (2016)
David Silver, et al., *Mastering the game of Go with deep neural networks and tree search*, 529 Nature (2016)

보고서

김건우, “인공지능에 의한 일자리 위험 진단 - 사무·판매·기계조작 직군 대체 가능성 높아”, 『LG경제연구원 보고서』(2018. 5. 15.)
나영식, 조재혁, “인공지능(반도체)”, 『KISTEP(한국과학기술기획평가원) 기술동향브리프』, 2019 Vol. 1., (2019)
이용용, “금융 생태계를 변혁하는 인공지능의 확산 및 시사점”, 『KISA(한국인터넷진흥원) REPORT』, 2018 Vol. 9. (2018)
전승우, “인공지능(AI) 프로세서, 새로운 혁신의 원동력 될까”, 『LG경제연구원 보고서』(2018, 11. 21.)
추형석, “인공지능, 어디까지 왔나?”, 『SPRI(소프트웨어정책연구소) 이슈리포트』, 2018 Vol. 6. (2018)
한국정보화진흥원(NIA), 『AI Network Lab 이슈리포트』, Vol. 3. (2018)
AHSN Network, *Accelerating Artificial Intelligence in health and care: results from a state of the nation survey*, (AHSN Network, 2018)
CADTH, *An Overview of Clinical Applications of Artificial Intelligence*, CADTH issues in emerging health technologies 174 (2018)
Deloitte, *Artificial Intelligence Innovation Report 2018* (2018)
Heike Mai, *PSD 2, open banking and the value of personal data*, Deutsche Bank Research (2018. 6. 28.)
Juan O. Freuler & Carlos Iglesias, *Algorithms and Artificial Intelligence in Latin America: A Study of Implementation by Governments in Argentina and Uruguay* (2018)
KPMG, *How the UK can win the AI race* (KPMG UK, 2018)

Little Hoover Commission, *Artificial Intelligence: A Roadmap for California* (Little Hoover Commission, Report 245 (2018))

Manoj Kewalramani, *China's Quest for AI Leadership: Prospects and Challenges*, Takshashila Working Paper 2018-02 (2018)

Mayor of London, *London: The AI Growth Capital of Europe* (2018)

McKinsey Global Institute, *Notes from the AI Frontier – Applying AI for Social Good* (2018)

Giovanni Sartor, *Artificial Intelligence: Challenges for EU Citizens and Consumers*, European Parliament IP/A/IMCO/2018-16

Yoav Shoham et al., *The AI Index 2018 Annual Report* (Stanford, AI Index Steering Committee, Human-Centered AI Initiative at Stanford University, 2018)

U.K. Digital Competition Expert Panel, *Unlocking digital competition* (2019. 3. 13.)

기사 외

Forbes Insights, *Inside The R&D Of AI Ethics* (2019. 3. 27.)

Jeffrey Dastin, *Amazon scraps secret AI recruiting tool that showed bias against women*, Reuters (2018. 10. 10.)

Edd Gent, *How AI is helping sports teams scout star players*, NBC (2018. 6. 13.)

Samuel Gibbs, *Tesla halts Model 3 production as firm scrambles to improve automation*, The Guardian (2018. 4. 17.)

Una Softic, *Artificial intelligence fosters financial inclusion in Asia*, Nikkei Asian Review (2018. 10. 3.)

인공지능과 고용시장의 변화

이수형(교수, 서강대학교 경제학부)



웹에서 PDF 바로 보기

초기의 인공지능(AI)은 소프트웨어가 작업을 수행하는데 필요한 규칙을 시스템화하는 것으로 시작되었으나, 빅데이터의 등장과 컴퓨팅 능력의 증가로 인간이 사전적으로 정한 규칙에 매이지 않고 정확도가 높은 알고리즘을 체득하는 방향으로 발전해왔다(OECD, 2018). 경제학계의 많은 연구에서는 시를 생산활동에서 인간의 개입이 필요하지 않거나, 적게 필요로 하는 자동화(automation)를 가능하게 하는 방법 중 하나로 간주한다.

경제학 이론적으로 자동화의 도입은 기존의 인력의 대체하는 효과를 유발하는 동시에, 기술과 보완적인 관계에 있는 인력에 대한 수요를 늘리는 효과 또한 가지고 있다. 일부 실증 경제 분석에 따르면, 자동화와 보완적 관계에 있는 경우 일자리 증가뿐만 아니라 종사자들의 생산성, 임금 상승 등에도 긍정적인 효과를 가진다고 보고되고 있다(Autor, 2003).

따라서 AI가 일자리에 미치는 영향은 AI 기술이 어떠한 직업을 대체할 것인지 그리고 AI 기술 확대에 의해 어떠한 직종이 더 필요하게 될지 두 가지 측면에서 고려할 필요가 있다. 하지만 후자의 경우 객관적인 예측(prediction)이 어려운 이유로 관련 연구가 상대적으로 드물게 보고되고 있다. 따라서 본 보고서에서는 전자의 문제를 중심으로 연구방법론과 주요 결과를 소개하고자 한다.

본 연구에서 중점적으로 소개할 대표적인 연구는 Frey and Osborne(2013)와 Arntz et al.(2016), 그리고 두 논문의 방법론을 절충한 Nedelkoska and Quintini(2018)이다. AI와 일자리 대체의 선구적인 연구인 Frey and Osborne(2013)는 AI로 대체하기 어려운 직무를 선정 후 미국 직업정보시스템(O*Net)을 이용하여 각 직업별로 해당 직무의 비중을 측정하였다. 이와 달리 Arntz et al.(2016)는 O*Net 대신 국제성인역량조사(PIAAC)을 이용하여 AI의 대체가능성을 연구하였으며, 근로자 단위의 정보(성별, 교육수준, 사업체 크기 등)를 대체가능성을 예측하는 설명 변수로 사용하였다는 차별성을 가진다. 마지막으로 두 연구의 방법론을 절충한 Nedelkoska and Quintini(2018)는 기본적으로 Frey and Osborne(2013)의 방법론을 사용하되, PIAAC을 이용하여 대체가능성을 추정하였다.

그에 따른 주요 결과로 Frey and Osborne(2013)은 미국 일자리의 47%가 AI에 의해 대체될 가능성이 높은 고위험군에 속한다고 추정하였다. 대체가능성이 높은 일자리에선 텔레마케터, 세무사 등이 속해있으며, 대체 확률이 1% 미만인 일자리에선 사회복지사, 놀이 치료사 등이 속하는 것으로 예측되었다. 반면에 Arntz et al.(2016)는 AI에 의해 대체될 일자리가 9%에 불과한 것으로 보고하였으며, Nedelkoska and Quintini(2018)는 14%로 보고하여 두 연구의 중간에 위치한다. 특히 Nedelkoska and Quintini(2018)는 자동화 위험은 모든 근로자에게 동일하게 적용되지는 않으며 주로 저학력이나 기초적인 업무를 요하는 저숙련(low-skilled) 일자리에 큰 영향을 미친다고 분석하였다.

다음으로 AI와 국내 일자리에 대해 살펴보았다. 한국의 경우 2016년을 기준으로 세계에서 가장 높은 로봇 도입률을 보였으며 특히 제조업에서 가장 높은 로봇 밀도를 보여주었다. 그에 비해 서비스 로봇 분야는 제조업 분야의 절반 수준에 그쳤지만, 키오스크와 ATM 등을 필두로 지속적인 성장세를 보이고 있다.

AI 도입에 따른 국내 일자리의 대체효과는 앞서 살펴본 각 방법론에 따라 상이하다. Arntz et al. (2016) 방법론에 따르면 한국은 OECD 국가 중 고위험 일자리 비중이 가장 낮은 국가에 속하나, Nedelkoska and Quintini(2018)에 따르면 중위권의 위험을 지닌다고 보고된다. 국내 연구진의 연구로는 김세움(2015)과 김건우(2018)가 대표적이다. 김세움(2015)는 우리나라 전체 일자리 중 57%가 대체위험성 고위험군에 속해 자동화에 매우 취약하다고 분석하였다. 해당 일자리에는 주로 영업 및 판매 직종이 속하는 것으로 나타났다. 반면에 대체위험성이 낮은 직종은 '관리자'와 '전문가 및 관련 종사자'로 나타났다(김건우, 2018). 하지만 세부 직종 중 '사무종사자'와 '장치, 기계조작 및 조립 종사자'의 경우 대체 확률이 100%에 가까운 높은 값을 보였다. AI에 의해 대체되기 힘든 직업으로는 보건, 교육, 연구 등 상호 의사소통이나 지적능력이 요구되는 직업이 해당되었다. 소득 수준별로는 월평균 소득이 100-300만원인 취업자의 자동화 위험이 높은 것으로 나타나, AI도입이 국내 중산층에 미치는 영향이 클 것임을 알 수 있다(김건우, 2018).

다수의 선행연구를 종합해보면 AI로 인한 직업 손실 위험을 지닌 인구는 보수적으로 계산할 경우에도 약 6%에 달한다. 이는 2018년 기준 국내 실업률이 3.8%임을 감안할 때 매우 높은 수준으로 볼 수 있다. 특히 AI 도입으로 인해 실업뿐만 아니라 실업자가 재취업하는 데에 어려움이 따를 예정이다. 따라서 노동시장에 참여할 학생들뿐만 아니라 AI로 대체될 가능성이 높은 직종에 근무하고 있는 근로자에 대한 사전교육 또한 요구된다. 이러한 구조적 변화에 대응하기 위해서는 국가차원의 인재양성 정책이 도움이 될 수 있다. 한국의 경우 2017년 교육부의 “제4차 산업혁명과 미래 교육 실천방안”을 발표하며 중 장기적 정책 방향을 설정하겠다고 밝혔다. 하지만 초·중·고등학교에 대비해 상대적으로 부족한 대학교육에 대한 투자와 AI 관련 강의 공급 부족 등은 아직 해결해야 할 과제로 남아있다. 또한 대학교육과 노동시장 간의 낮은 연계성과 파이프라인(pipeline) 부재 또한 문제점이다. 이러한 교육시장과 노동시장 정책의 보완을 통해 자동화 위험으로부터 국내 노동시장의 충격을 최소화하는 노력이 필요할 것이다.

1. 들어가는 글

알파고의 사례에서 보듯이 알고리즘과 빅데이터에 기반한 새로운 알고리즘의 활용이 사회 전반에서 급격히 늘어나고 있다. 인공지능(AI), 사물인터넷(IoT), 빅데이터 등을 종합하여 산업혁명에 버금가는 변화가 이미 진행 중이며 이러한 변화를 종종 4차 산업혁명이라고 부르기도 한다(World Economic Forum, 2016A). 4차 산업혁명이라는 용어가 적절한지에는 대해서는 학계의 논란이 존재하나, 전반적으로 새로 개발되고 활용되는 기술이 세계 경제 전반에 큰 영향을 미친다는 점에 대해서는 공통적인 견해를 보인다.

지금까지의 기술의 발전은 비교적 정형화 시키기 쉬운 농업, 광공업, 제조업 등에서 진행되어 왔으며, 이들 산업에서 제공하던 일자리를 줄이는 결과를 가져왔다. 이에 나아가 앞으로의 자동화 기술의 발전은 사무 행정 등 반복적인 업무를 하는 인력들을 빠르게 대체할 것으로 예상된다. 반면에, 데이터 분석가, AI 머신러닝 전문가 등 과학기술을 요구하는 인력에 대한 수요는 높아질 것으로 예상되고 있다(World Economic Forum, 2016B). 이렇듯 AI를 통한 새로운 기술과 보완적인 관계가 있는 경우 일자리 증가뿐만 아니라 종사자들의 생산성, 임금 상승 등에도 긍정적인 효과를 가진다고 보고되고 있다(Autor, 2003). 하지만 후자의 경우 객관적인 예측(prediction)이 어려운 관계로 관련 연구가 상대적으로 드물게 보고되고 있다.

이에 본 고에서는 해외 선행연구를 통하여 AI 도입으로 인한 산업 혹은 직종별 대체가능성을 추정하는 방법론과 분석 결과를 소개하고자 한다. 이와 더불어 국내 선행연구를 통해 AI 도입이 국내 노동시장에 미치는 영향을 살펴보았다. 전자의 경우, 학계에서 선구적인 논문으로 여겨지는 Frey and Osborne(2013)와 이를 주요 OECD 국가로 확장 분석한 Arntz et al.(2016), 그리고 양 논문의 절충적인 방법론을 모든 OECD 국가에 적용한 Nedelkoska and Quintini(2018)를 소개한다. 후자의 경우 앞의 세 가지 논문의 방법론을 한국에 적용한 김세움(2015)과 김건우(2018)를 검토한다. 마지막으로, AI 도입이라는 구조적 변화에 대응하기 위한 국가차원의 노력들에 대하여 소개한다.

주요 결과로, Frey and Osborne(2013)은 미국 일자리의 47%가 AI에 의해 대체될 가능성이 높은 고위험군에 속한다고 추정하였으며, 대체 확률이 1% 미만인 일자리에겐 사회복지사, 놀이 치료사 등이 속하는 것으로 예측했다. 반면에 Arntz et al.(2016)은 AI에 의해 대체될 일자리가 9%에 불과한 것으로 보고하였으며, Nedelkoska and Quintini(2018)는 14%로 보고하였다.

국내의 경우 Arntz et al.(2016) 방법론에 따르면 한국은 OECD 국가 중 고위험 일자리 비중이 가장 낮은 국가에 속하나, Nedelkoska and Quintini(2018)에 따르면 중위권의 위험을 지닌다고 보고된다.

김세움(2015)에 따르면 우리나라 전체 일자리 중 57%가 대체위험성 고위험군에 속해 자동화에 매우 취약하다는 분석이 가능하다. 해당 일자리에선 주로 영업 및 판매 직종이 속하는 것으로 나타났다. 반면에 대체위험성이 낮은 직종은 ‘관리자’와 ‘전문가 및 관련 종사자’로 나타났다(김건우, 2018).

본 보고서는 다음과 같이 구성되었다. 제 II장에서 인공지능과 노동시장에 대해 살펴보고, 해외 선행연구를 통해 시의 노동 대체효과를 다룬다. 제 III장에서는 앞서 살펴본 방법론을 국내에 적용하여 시와 국내의 노동시장을 살펴본다. 마지막으로 IV장에서 이러한 노동시장의 변화에 대응하기 위한 국가차원의 노력들에 대해 논의한다.

2. 인공지능 (Artificial Intelligence)

1. 개관

인공지능 (Artificial Intelligence, AI)은 컴퓨터 공학의 한 분야로서 기계가 지식을 습득, 적용하는 지능적인 행동을 수행하는 능력을 말한다(Manyika, 2017). 초기의 AI는 소프트웨어가 작업을 수행하는데 필요한 규칙을 시스템화하는 것으로 시작되었으나, 정확도가 높지 않는 등 실용화에 난제가 있었다. 그러나, 빅데이터의 등장과 컴퓨팅 능력의 증가로 AI는 Machine learning, Deep learning 등을 통하여 인간이 사전적으로 정한 규칙에 매이지 않고 정확도가 높은 알고리즘을 체득하는 방향으로 발전해 왔다(OECD, 2018).

AI를 언급할 때 자주 사용되는 용어로는 인지 컴퓨팅(Cognitive Computing), 기계 학습(Machine Learning), 지도 학습(Supervised Learning), 딥 러닝(Deep Learning) 등이 있다. 인지 컴퓨팅은 시와 동의어로 주로 사용되고, 기계 학습은 인공 지능의 한 분야로 데이터를 통해 컴퓨터가 스스로 학습할 수 있도록 하는 기술을 개발하는 분야이다. 딥 러닝(Deep Learning), 지도 학습(Supervised Learning) 등은 기계 학습을 구현하는 방법들 중에 하나이다.

2. 시와 일자리

1) 개관

위에서 살펴보았듯이 시의 개념이나 구체적인 실현 방법, 응용 사례들은 시시각각으로 진화하고 있다. 이에 AI 기술이 고용시장에 어떠한 영향을 미치는지 논함에 있어서 우선 시를 어떠한 경제적 요인(Economic Factor)으로 간주해야 할 지에 대한 고려가 필요하다. 경제학계의

연구에서는 AI를 인간의 개입이 필요하지 않거나, 적게 필요로 하는 자동화(automation)를 실현하는 방법의 하나로 간주한다. 학자에 따라서 AI로 인한 자동화를 컴퓨터화(computerization)라는 용어로 사용하기도 한다. 본 논고에서는 두 가지의 용어를 동일하게 간주하며, 일관성을 위해 “자동화”라는 용어를 사용한다.

산업혁명 이후 자동화는 기계, 로봇, 컴퓨터 등을 이용하여 전문분야에 걸쳐 진행되어 왔다. 그러나, 지금까지의 자동화 기술은 비교적 정형화 시키기 용이한 작업들에 적용되어 왔다. 예를 들어 트랙터 등 농업 기계를 이용하여 경작에 필요한 인력을 줄이거나, 제조업 공장에서 산업기계를 이용하여 자동차를 조립한다는 현상이 이에 속한다. 이러한 자동화는 농업, 광공업, 제조업에서 많이 진행되어 왔고, 이들 산업에서 제공하던 일자리를 줄이는 결과를 가져왔다. 이러한 일자리 영향이 중산층의 일자리 기회를 감소시켜 소득 불평등을 심화시켰다는 일련의 연구결과들도 보고되고 있다(Autor, 2003).

시분야에서 최근 급진적인 발전을 이루고 있는 big data, machine learning 등의 최근 기술들은 지금까지 자동화의 영역을 서비스 산업이나, 보다 고차원적인 인지능력을 필요로 하는 분야까지 확장시키고 있다. 예를 들어 소비자들을 상대하는 고객 응대 서비스의 경우 챗봇(Chatbot)이나 음성인식을 기반으로 한 자동화 기술들을 빠르게 도입하고 있다. 이러한 서비스 자동화 기술은 두 가지 단계를 필요로 한다. 우선, 사람들이 고객 응대 직원들과 어떠한 의사소통을 하는지를 알 수 있는 음성 파일들이 축적되어 있어야 한다. 다음으로, 음성 파일에 담겨있는 소비자가 주로 하는 질문들과, 고객 응대 직원들의 대답을 패턴화 시켜서, 실제 고객이 질문을 하였을 때 자동적으로 적합한 대답을 만들어내는 기술이 필요하다. 이러한 기술은 위에 살펴본 기계 학습의 방법을 이용하면 가능하다. 해외의 경우 고객 응대 같은 비교적 급여가 낮은 서비스 직종뿐만 아니라 법률, 회계, 세무 등 다양한 분야에서 서비스가 제공되고 있다. 이러한 예는 AI에 기반한 자동화는 낮은 급여의 직종뿐만 아니라, 회계사, 법무사, 변호사 등 높은 급여를 제공하는 직종까지 영향을 줄 수 있다는 점을 시사한다.

자동화를 가능하게 하는 기술은 그 정의상 인력의 필요를 줄이는 효과를 가지고 있다. 따라서, 자동화가 도입됨에 따라 인력이 기계/컴퓨터로 대체(substitution)되는 효과를 지니게 된다. 그러나 동시에, 그러한 기술과 보완적인 관계(complementarity)가 있는 인력에 대한 수요도 늘리는 효과를 가지고 있다. 특히 자동화와 보완적인 관계에 있는 일자리의 경우, 일자리가 증가할 뿐만 아니라, 종사자들의 생산성 향상, 임금 인상 등 상당한 정도의 긍정적인 효과를 가져온다고 보고되고 있다(Autor, 2003).

따라서, AI가 일자리에 미치는 영향은 우선 지금의 AI 기술이 어떠한 종류의 직업을 대체할 수 있을 것인지 그리고 AI 확대로 인하여 어떠한 직종이 더 필요로 하게 될지 두 가지 측면에서 나누어 고려할 필요가 있다. 이러한 필요성에도 불구하고, 최근의 경제학적인 연구는 AI로 인하여 대체가 될 가능성이 높은 직종에 대한 연구들이 많이 보고되었고, 후자의 경우는 상대적으로 드물게 보고되었다. 그 이유는 전자의 경우 특정한 성격을 가진 직종들이 자동화로 쉽게 대체가 가능하다는 가정하에 비교적 객관적인 분석이 가능하나, 후자의 경우 급격하게 변화하고 있는 AI 기술을 고려할 때 객관적인 예측(prediction)이 어렵기 때문이다. 이에 아래의 소절들에서는 전자의 문제를 중심으로 연구방법론과 주요 결과를 설명하되, 후자의 문제에 대해서는 간략하게 논의 동향을 검토한다.

2) AI의 노동 대체

(1) 연구 방법론

AI에 의하여 어떠한 직업이 대체가 될 가능성이 높은지를 분석하는 연구들은 대체로 다음의 분석과정을 거친다. 첫째, 사전적(ex-ante)으로 AI가 대체하기 쉬운 직업 특성, 혹은 반대로 AI가 대체하기 어려운 직업의 특성을 몇 가지 특정한다. 둘째, 직무(task), 직업(occupation), 산업(industry)별로 얼마나 AI로 대체되기 쉬운(혹은 어려운) 업무로 구성되어 있는지 계산한다. 셋째, 두 번째 단계에서 계산한 지표를 AI의 대체 가능성(probability)으로 전환한다. 필요할 경우 분석 대상에 따라 대체 가능성 결과를 직업, 산업, 혹은 국가 단위로 합산하여 얼마나 AI 대체될 위험이 높은지 수치화 한다.

예를 들어, AI와 일자리 대체에 선구적인 연구인 Frey and Osborne(2013)의 경우 먼저, 감지 및 조작(Perception and Manipulation), 창의적 지능(Creative Intelligence), 사회적 지능(Social Intelligence)이라는 세 개의 직무(task)를 AI로 대체하기 어려운 직무로 간주하였다. [표 1]의 (1)~(3)열을 통해 각 직무를 살펴보자면, 감지 및 조작 업무는 구조화되지 않은 업무 환경과 그와 관련된 업무로 자동화가 어려운 경우를 말한다. 따라서 해당 분류에는 사람이 직접 수동으로 조작하고 수행하는 업무가 포함되게 된다. 다음으로 창의적 지능은 새롭고 가치 있는 아이디어를 내거나 문제해결을 위한 창의적 방법을 개발하는 능력을 말한다. 또한 공예품, 음악, 공연 등을 고안해내는 능력을 포함하여 예술가가 이 분야에 포함된다. 마지막으로 사회적 지능은 협상, 설득, 보살핌과 같은 광범위한 업무에서 중요한 역할을 하는 역량으로, 사회적 상호작용을 수반하는 업무를 말한다.

둘째로, 미국 직업정보시스템(O*Net) 정보를 이용하여 각 직업(occupation) 별로 위의 세가지 직무의 비중이 얼마나 요구되는지를 추출한다. 다음 단계로 분석 대상의 직업(occupation)중 약 10%(70개)의 직업에 대해서 서베이를 진행해 향후 10년동안 시로 대체될 가능성이 얼마나 높은지 의견을 구한 뒤, 두 번째 단계의 지표와 상관관계를 구한다. 마지막으로 이 상관관계를 이용하여 서베이 결과가 없는 나머지 90%의 직업들에 대해 시로의 대체가능성을 예측한다.

미국 외에 여러 국가들을 동시에 비교하는 연구들의 경우 O*Net 대신 국제성인역량조사(PIAAC, The Program for the International Assessment of Adult Competencies)를 이용하여 직업별로 시에의 대체가능성을 연구한다. O*Net과 PIAAC간의 분류에는 큰 차이가 없으며, 아래 [표 1]에서 각 직무 분류의 유사성을 확인할 수 있다.

(2) 시의 노동 대체 효과

시의 노동 대체효과에 대하여 처음 계량화한 연구인 Frey and Osborne (2013)에 따르면, 미국 일자리의 47%가 시에 의하여 대체될 가능성이 높은 고위험군에 속한다고 추정하였다. 저자들은 고위험군에 포함된 직업은 향후 10년에서 20년 사이에 대체될 가능성이 70% 이상인 경우로 정의하였다. 아래 [표 2]는 Frey and Osborne(2013)가 제시한 미국의 702개 직업 중에서 시로 대체가능성이 가장 높은 10개 직업과 대체가능성이 가장 낮은 10개 직업을 보여준다. 예를 들어 텔레마케터(telemarketer)나 세무사(tax preparer)과 같은 직업은 99%의 확률로 10년에서 20년사이에 시로 대체될 것으로 예상하였으나, 놀이 치료사(recreational therapists)나 의료, 사회복지사(healthcare social workers)는 같은 기간 동안 시로 대체될 확률이 1% 미만으로 예측되었다.

한편 Arntz et al.(2016)은 OECD 국가를 대상으로 O*Net대신 국제성인역량조사(PIAAC)를 이용하여 시의 대체가능성을 연구하였다. 이들의 연구결과는 앞선 Frey and Osborne(2013)과 큰 차이를 보인다. 예를 들어, Frey and Osborne(2013)는 미국의 일자리 중 47%가 시에 대체될 가능성이 높다고 예측한 반면, Arntz et al.(2016)은 대체될 일자리가 9%에 불과하다고 분석하였다. 이렇게 큰 차이가 나는 중요한 이유는 후자의 경우 각 국가별로 근로자 단위의 정보(성별, 교육수준, 사업체 크기 등)를 시의 대체가능성을 예측하는 설명 변수로 사용한 반면, 전자의 경우 직업(occupation)별로 시에의 대체가능성을 예측하였다는 점이다. 그러나, 어떠한 방법이 더 낫다고 가늠하기는 어려운데, 그 이유는 다음과 같다. 각 국가, 직종별로 교육 수준 등 근로자의 특징이 직업별로 상이하므로 같은 직업군일지라도 국가별로 시의 대체가능성이 상이할 수 있다. 이러한

가능성은 Arntz et al.(2016)에서 잘 반영하고 있다. 그러나 Arntz et al.(2016)이 택한 방법론의 한계는 각 직업에 근무하는 근로자의 특성이 고정되었다는 가정하에 AI의 대체 가능성을 계산한다는 점이다. 따라서 AI가 발전함에 따라 해당 직업에 종사하는 근로자의 유형이 변화할 경우, Arntz et al.(2016)의 분석결과에 큰 편의(bias)가 발생할 가능성이 높다.

위 두 논문들의 방법론을 절충한 연구로Nedelkoska and Quintini(2018)이 있다. 저자들은 기본적으로Frey and Osborne(2013)의 방법론을 사용하되, Arntz et al.(2016)와 마찬가지로 국제성인역량조사(PIAAC)를 이용하여 AI의 대체가능성을 계산하였다. 그러나 Arntz et al.(2016)와는 상이하게 OECD 국가 중 분석 대상을 32개 국가 전체로 확대하였고, AI에의 대체가능성을 계산함에 있어서 근로자들의 특성을 고려하지는 않았다. 이들의 결과에 따르면, 32개 OECD 국가에서 AI에 의하여 대체될 가능성이 70% 이상인 고위험군의 직업은 14%에 이르며, AI 대체가능성이 50% 에서 70% 사이인 중위험군의 직업은 약 32%에 이른다. 따라서, 이들의 예측 결과는Frey and Osborne(2013)와Arntz et al.(2016)의 중간에 위치하고 있다.

Nedelkoska and Quintini(2018)은 자동화의 위험은 모든 근로자에게 동일하게 적용되지는 않으며, 주로 저학력이나 기초적인 업무를 요하는 직종(제조업, 농업, 일부 서비스업(우편·배달업, 수송업, 요식업))에 큰 영향을 미치고 전문적 훈련이나 고등교육을 요하는 직종은 큰 영향을 받지 않는다고 보고하였다. 아래 [표 3]에 자동화확률이 높은 산업 20개와 낮은 산업 20개를 확인할 수 있다. 자동화가 고속려 직업에 영향을 준다는(adversely affect) 논의가 있으나, Nedelkoska and Quintini(2018)는 이전의 기술 발전이 중간 속려 직업의 고용을 감소시켜 노동시장 양극화를 야기했다면 AI 도입은 저속려 일자리에 더 위협적이라고 주장하였다.

추가로 주목할 만한 결과는 자동화의 위험이 청소년이 속한 일자리(teenage jobs)에서 가장 크다는 것이다. 이는 자동화가 청소년 실업을 유발할 가능성이 높다는 것을 의미한다. 그러나 한편으로는 청소년의 경우 다른 성인 노동자에 비해 직업간 이전이 순조로울 수 있다는 것을 뜻하기도 한다. 대부분의 국가에서 청년들의 능력이 노년층에 비해 높게 나타나며, 이는 그들이 새로운 기술 도입으로 인한 새로운 일자리에 더 쉽게 적응할 수 있음을 뜻하기 때문이다. 따라서 대다수의 고위험 일자리가 학생들의 일자리와 관련되어 있는 만큼 그 외 분야의 인턴제도를 활성화시키는 제도는 학생들의 현재 직업에서 필요한 특수적인 기술(job-specific skill)과 다른 직업에서도 사용할 수 있는 일반적인 기술(general skill) 습득을 촉진시킬 수 있다고 보고한다.

3) AI와 보완적인 노동

앞서 기술 하였듯이, 학술적인 측면에서 AI 기술로 인하여 수요가 창출되거나 증가할 직업에 대한 연구는 상당히 드물고, 몇몇 조사연구들이 존재하는 상황이다. 일례로 맥킨지 보고서(Manyika, 2017)는 육체적이고 정형화된 기술과 단순 인지 기술은 AI 에 의하여 대체되기 쉽지만 동시에 고차원의 인지 기술과 사회적, 감정적 대면 기술, 과학 기술과 관련된 직업은 오히려 2030년이 되면 일자리가 더욱 증가할 것이라고 보았다. 기술이 발전되면서 많은 직업들이 대체되겠지만 동시에 구조적 변화에 따른 새로운 일자리 창출 또한 일어날 것이다. SW 엔지니어, 웹 개발자 등 IT 관련 일자리는 2,000~5,000만개, 가사 서비스 분야는 5,000~9,000개, 헬스케어 산업은 8,000~1억 3,000만개, 소비자-건강-교육 분야에서는 3억에서 3억 6,500만개의 일자리가 새로이 창출될 것이라 예측하였다.

또한 2018년 세계경제포럼 보고서는 AI 시대의 미래 인재에게 요구되는 기술은 크게 ‘기술과 관련된 능력(Technology-related skills)’과 ‘비인지적 능력(non-cognitive soft skills)’이라고 예상하였다. 구체적으로 기술과 관련된 능력에는 분석적 사고, 시스템 분석, 전략 습득, 프로그래밍 등 신기술에 대한 숙련도가 포함된다. 그러나 이러한 기술적 숙련도뿐만 아니라 창의성, 독창성, 협상과 같은 인간과 관련된 비인지적 능력 또한 현재에 비해 수요가 크게 증가할 것으로 예측되었는데, 해당 순위는 아래 [표 4]에서 확인 가능하다. 이에 따라 AI로 수요가 늘어날 직군으로는 데이터분석가, AI 머신러닝 전문가 등의 기술 전문가와 판매 및 마케팅 전문가, 훈련 및 개발 전문가 등 비인지적 능력이 강조되는 직군이 포함되는 것으로 나타났다.

3. AI와 한국의 일자리

1. 한국 일자리의 현황

국제로봇협회에 따르면 한국에는 2016년 기준 10,000명당 631개의 로봇이 도입되었으며 이는 전세계에서 가장 높은 수치이다(Rob, 2018). 또한 2010년 이래로 계속해서 제조업 산업에서 가장 높은 로봇 밀도를 보여주었는데(IFR, 2018), 2016년 기준으로 세계 평균 로봇 도입률의 8배가 넘는 수치를 보였다. 2017년 한 해 산업용 로봇 판매량은 또한 294,000대를 웃돌며 사상 최대치를 기록하였다. 그에 비해 서비스 로봇 분야는 규모 면에서 2016년 기준 73억 달러로 제조업 로봇 시장의 절반 정도 수준에 미쳤다. 하지만 인공지능 및 ICT 기술이 발전되고 최저임금이 인상되면서 서비스 로봇 시장은 계속해서 성장세를 보여주고 있다.

서비스 로봇의 대표적인 예로는 무인 결제 기능을 담은 키오스크와 ATM이다. 키오스크는 2014년부터 시작하여 매년 높은 성장세를 보이고 있으며 ATM은 편의점 설치로 그 비중을 높이고 있다. 해당 로봇들을 통한 인건비 절감액이 커서 이러한 무인화는 앞으로 가속화될 것으로 보인다(한국경제, 2018). 그 예시로 [표 5]에 최근 외식·유통업체들이 도입한 서비스 로봇의 현황이 나열되어 있다.

2. AI의 일자리 대체 효과

AI가 한국의 일자리에 얼마만큼의 대체효과를 낼지에 대해서는 위에 살펴본 방법론 별로 상이한 결과를 지닌다. [표 6]에 각 방법론에 따른 OECD 국가별 자동화 위험률과 평균 및 중간값이 정리되어 있다. 예를 들어, Arntz et al.(2016)의 방법론에 따르면 한국의 경우 OECD 국가중 AI로 대체될 위험이 높은 일자리의 비중이 6%로 가장 낮은 국가 중 하나에 속하나, Nedelkoska and Quintini(2018)의 결과에 따르면 OECD 국가 중 중위권의 위험을 지닌다고 보고되고 있다.

국내 연구진의 연구로는 김세움(2015)와 김건우(2018)가 있다. 두 연구 모두 Frey and Osborne(2013)이 추정된 대체확률을 활용한다는 공통점이 있지만, 국내 직업과 매칭하는 방법에서 차이를 보인다. 김세움 한국노동연구원 부연구위원이 발표한 ‘기술 진보에 따른 노동시장 변화와 대응’보고서는 Frey and Osborne(2013)가 추정된 대체확률을 미국 노동통계국(U.S. Bureau of Labor Statistics)의 직업소재자료와 직종별 고용통계(Occupational Employment Statistics)를 참조하여 우리나라의 2012년도 직업사전의 직업과 매칭하는 방법을 활용하였다. 이와 다르게 김건우(2018)은 추정된 미국의 대체확률을 국제표준직업분류 기준으로 전환한 후, 이를 한국표준직업분류와 연계하는 방법을 사용하였다.

각 결과를 살펴보면, 김세움(2015)는 우리나라 전체 일자리의 57%가 향후 기술진보로 인하여 대체될 가능성이 높은 고위험군에 속한다고 보고한다. 이는 한국표준직업분류 소분류 단위 132개 직종에 대한 분석이며, 총 301개의 세분류 단위로 분석한 경우에도 55%라는 유사한 결과를 보인다. 미국 노동시장의 47%가 고위험군에 속한 것과 비교해보면 우리나라 노동시장이 미국에 비해 더 취약하다는 분석이 가능하다. 직업별 자동화 확률은 중분류만을 제공하고 있는데, 보건 의료 및 교육 관련직, 문화 예술 관련직, 관리직 등은 낮은 자동화 위험을 가지는 반면 판매 종사자의 경우 높은 자동화 위험을 보이는 것으로 나타났다.

이와 달리 김건우(2018)은 총 423개의 세분류 기준을 적용해 자동화 확률을 분석하였고, 전체 일자리의 43%가 자동화 고위험군에 속한다고

보았다. 자동화 위험이 낮은 하위 10개 직업과 상위 10개 직업도 함께 제시하였는데, 다른 나라와 비슷하게 대체확률이 낮은 분류에 속하는 직업들은 주로 ‘관리자’와 ‘전문가 및 관련 종사자’인 것으로 나타났다. 아래 [표 7]에 김세움(2015)과 김건우(2018)의 결과를 함께 제시하였다.

두 연구 모두 ‘관리자’와 ‘전문가’의 자동화 확률이 낮은 것으로 보고하였지만, 세분류 단위로 분석한 김건우(2018)의 경우, ‘전문가 및 관련 종사자’의 하위 분류인 ‘사무 종사자’, ‘장치, 기계조작 및 조립 종사자’의 경우 자동화 확률이 1에 가까운 높은 값을 보인다고 보고하였다. 또한 통신 서비스 판매원, 텔레마케터, 인터넷 판매원 등 과 같이 온라인을 통해 판매를 하는 직업들도 자동화 위험이 높은 분류에 포함되었을 뿐만 아니라 관세사, 회계사, 세무사와 같은 전문직도 자동화 위험이 높은 것으로 나타났다. 반면에, 인공지능에 의해 대체되기 힘든 직업으로는 보건, 교육, 연구 등 상호 의사소통이나 지적 능력이 요구되는 직업이 속해 두 연구 결과가 유사함을 확인할 수 있다. 그 중에서도 영양사, 의사, 교육 관련 전문가, 성직자, 공학 기술자 및 연구원 등이 낮은 자동화 가능성을 보였다.

두 연구를 종합해보면, 자동화될 가능성이 높은 직종은 주로 영업 및 판매 직종이라는 특징을 갖고 있다. 김세움(2015)은 미국의 경우 교육, 법률, 의료 등의 고숙련 서비스 직종의 일자리가 많아 근로자가 대체될 가능성이 낮은 반면 한국은 그의 절반 정도에 그치며, 영업 및 판매직종 비율 또한 높아 미국보다 높은 자동화 확률을 보인다고 해석하였다.

나아가 김건우(2018)는 소득 수준별로 자동화 확률을 분석하였으며, 고위험군 비중이 가장 높은 구간은 월평균 소득 수준이 100~200만원, 200~300만원인 취업자로 비중이 각각 47%였다. 우리나라 전체 취업자 중 60%가 소득 100~300만원 구간에 분포하고 있다는 점을 미루어 보았을 때 인공지능에 의한 자동화의 위험은 중산층에 미치는 영향이 클 것이라고 볼 수 있다. 이는 자동화의 위험이 저학력, 저소득층에 가장 크다고 보고한 Arntz et al.(2016)과 차이를 보여 주목할 만하다.

추가로 국내의 높은 자영업 비율을 고려했을 때, 대표적인 자영업 분야로 알려진 부문이 자동화 위험이 높은 것으로 나타났다. [표 7]에 따르면, 김세움(2015)의 경우 음식서비스 관련직은 74%, 영업 및 판매 관련직은 약 85%의 높은 자동화 위험률을 보였다. 김건우(2018)에서도 숙박 및 음식점업과 도매 및 소매업이 3대 고위험 산업임을 보였고 각각 서비스, 판매 직종의 취업자 비중이 높다고 지적해, 편의점, 치킨집, 카페 등 도소매 판매, 음식 서비스와 관련된 자영업이 큰 영향을 받을 것으로 보인다.

4. 정책적 함의 및 소결

다수의 선행연구를 종합해보면 시로 인한 직업 손실 위험을 지닌 인구는 보수적으로 계산할 경우에도 약 6%에 달한다. 이는 2018년 국내 실업률 수준이 3.8%임을 감안할 때 매우 높은 수준으로 볼 수 있다. 특히 시로 대체될 직업과 시시대에 노동수요가 증가할 직업이 상이함에 따라 실업자가 재취업하는 데에 어려움이 따를 예정이다. 노동자가 직업을 바꿀 경우 인적자본 중 일부만이 새로운 직업에서 사용됨으로 이를 보완하지 않으면 임금이나 재취업 기회에 어려움을 겪는다(Gathmann and Schönberg, 2006). 따라서 이후 노동시장에 참여할 학생들에게도 시에 부합한 교육이 이루어져야 하지만, 시로 대체될 가능성이 높은 직종에 근무하고 있는 근로자 역시 사전적인 교육에 힘써야 할 것이다. 근로자가 비자발적인 실업을 경험하기 전에 이직(job-to-job transition)을 가능하게 하거나, 실직 후 구직기간을 줄임으로써 실직에 따른 소득 감소, 정신적 스트레스, 낙인효과 등을 줄이는 효과를 가질 수 있다.

예시로, 해외 주요국은 이러한 구조적 변화에 대응하기 위한 인재양성 정책을 펼치고 있다(정보통신기술진흥센터, 2018). 대표적으로 미국은 AI 전문가, 연구자 및 데이터 과학자 양성을 위한 방안을 마련하였다. 정부 차원에서 다양한 과학기술 단체를 활용하여 STEM 교육 프로그램을 지원, 운영하려는 노력을 보이고 있는 것이다. 미국 국립학술원(National Academies Press)은 데이터 과학자 양성을 위해 대학의 데이터 과학 교육 강화를 기반으로 하는 교육과정을 개발하고자 권고안을 제시하기도 했다. 결과적으로 미국의 우수한 대학인 버클리 대학, 카네기 멜론 대학, 미시간 대학, 매사추세츠 공과 대학 등에서는 데이터 과학 전공과 관련된 교육 과정을 신설하였다.

일본은 2018년 4월, AI 기술전략 실행계획을 발표하여 ICT 전문가 인력을 육성하기 위한 여러 정책적 지원을 펼치고 있다. 첨단 ICT 인재와 일반 ICT 인재를 구분하여 그에 맞는 학교 기관 기업 차원에서의 지원 방안을 구축하였다. 또한 문부과학성은 과학과 관련된 연구를 하는 박사과정 학생 및 박사 학위 보유자를 모집, 선정하여 지원을 통해 다양한 경력을 실현하게끔 하는 데이터 인재양성 프로그램을 발표하였다. 중국의 중국과학원대학 또한 2017년부터 AI 특화 인재양성을 위한 인공지능 기술 단과대를 운영하고 있다.

한국의 경우 2017년 교육부에서 “제 4차 산업혁명과 미래교육 실천방안”을 발표하였다. 4차 산업혁명에 대비하여 정부는 ‘유연화’, ‘자율화’, ‘개별화’, ‘전문화’, ‘인간화’라는 다섯 가지 핵심 키워드를 설정하였고 이를 반영하는 중 장기적 교육정책 방향을 설정하겠다고 하였다. 하지만 실제로 신기술의 융합적 지식을 교육시킬 수 있는 교육 및 지원 정책에 대한 구체적인 사항은 아직 제시되지 않은 실정이다.

코딩교육등이 초등학교 단계에서 이루어지고는 있으나, 막상 이들 교육이 대학교육까지 연계되어 실제 노동시장에서 필요로 하는 인력으로 양성되는 파이프라인(pipeline)이 형성되어 있지는 않다. 일례로, 조선일보에 따르면 서울대 공대 컴퓨터공학부가 강사와 강의실을 확보하지 못해 소프트웨어 개발 관련 강의에 있어 다른 과 학생의 수강을 제한하기로 했다고 보도되고 있다. 이러한 상황은 4차 산업혁명에 대비하기 위해 특정 교육에 대한 수요는 증가하고 있지만 이에 교육기관이 적절하게 대응하지 못하고 있다고 해석된다. 실제로 컴퓨터공학을 전공 또는 부전공하려는 학생은 2009년 기준 13명에서 2018년 기준 106명으로 크게 증가하였다. 하지만 여전히 정원은 55명에 그치고 있고 변화하는 교육 트렌드에 유연하게 대응하지 못하는 실정이다.

노동시장에 진입 전인 청년층뿐만 아니라 자동화로 직업을 잃을 가능성이 높은 노동자에게도 시에 부합한 교육이 요구되는 바이다. 이는 정부의 직업훈련 지원을 통해 진행될 수 있는데, 국내의 경우 직업훈련 체계의 정립이 미흡하다는 한계가 있다. 일례로 그간의 직업훈련은 산업수요를 잘 반영하지 못한다는 점과 우선적으로 고려되어야 할 대상인 비정규직, 저학력 및 고령자 등의 취약계층이 오히려 훈련으로부터 배제되어 왔다는 지적이 있다(강순희, 2011). 또한 훈련투입 대비 훈련의 성과가 저조하여 효과성이 크지 않다는 한계가 있다. 미시데이터를 활용하여 정부에서 제공하는 공공직업훈련의 인과효과를 살펴본 연구의 경우에도 직업훈련이 취업에 긍정적인 영향을 미치는 증거를 발견하지 못했다(김용성&박우람, 2015; 유경준&이철인, 2008, 강순희 외, 2015). 이에 중·고등교육 과정을 아우르는 교육정책과 노동시장 정책의 보완을 통해 국내 노동시장의 충격을 최소화하는 정책적 노력이 절실한 것으로 보인다.

참고 문헌

- 강순희(2011), 공공직업훈련의 의의와 개선과제. 월간 노동리뷰, 1, pp. 49-54.
- 강순희, 어수봉, 최기성(2015), 미취업자의 직업훈련 참가 결정요인과 고용성과 분석. HRD 연구(구 인력개발연구), 17(2), pp. 267-298.
- 김건우(2018), 「인공지능에 의한 일자리 위험 진단」, LG 경제연구원.
- 김세움(2015), 「기술진보에 따른 노동시장 변화와 대응」, 한국노동연구원.
- 김용성, 박우람(2015), 실업지속의 원인 분석과 직업훈련의 효과 및 개선방안에 관한 연구, KDI Policy Study, 23, pp. 1-88.

남성일(2017), 「쉬운 노동경제학」, 박영사.

유경준, 이철인(2008), 실업자 직업훈련의 효과 추정, 노동경제논집, 31(1), 59-103.

윤종혁(2017), 「제 4차 산업혁명과 미래교육 실천방안」, 교육부, 한국교육개발원.

정보통신기술진흥센터(2018), 「해외 주요국의 4차 산업혁명 대응 인재양성 정책 동향」, 해외 ICT R&D 정책동향.

Acemoglu, D. and Autor, D.H.(2011), Skills, tasks and technologies: implications for employment and earnings. In: Handbook of Labor Economics, vol. 4B, pp. 1043-1171.

Arntz, M., Gregory, M. and Zierahn U.(2016), The risk of automation for jobs in OECD countries. OECD.

Autor, D.H., Levy and F., Murnane, R.J.(2003), The skill content of recent technological change: an empirical exploration. The Quarterly Journal of Economics, vol.118, No.4, pp. 1279-1333.

Bughin, J., Hazan, E., Lund, S., Dahlström, P., Wiesinger, A. and Subramaniam, A.(2018), Skill shift automation and the future of the workforce. McKinsey Global Institute.

Gathmann, C. and Schönberg, U.(2006), How general is specific human capital?, IZA Discussion Paper No.2485.

David H.(2013), The “task approach” to labor markets: an overview(No.w18711). National Bureau of Economic Research

David, H.(2015), Why are there still so many jobs? The history and future of workplace automation. Journal of Economic Perspectives, vo.29, No.3, pp. 3-30.

Frey, C.B. and Osborne, M.(2013), The future of employment : how susceptible are jobs to computerisation? Technological forecasting and social change, vol.114, pp. 254-280.

Manyika, J.(2017), A future that works: automation, employment, and productivity, McKinsey Global Institute.

Nedelkoska, L. and Quintini, G.(2018), Automation, skills use and training, OECD Social, Employment and Migration Working Papers, No.202.

OECD(2018), Transformative technologies and jobs of the future, Background report for the Canadian G7 Innovation Ministers' Meeting.

Rob Smith, "South Korea has the highest density of robot workers in the world", <https://www.weforum.org/agenda/2018/04/countries-with-most-robot-workers-per-human/>, 2018.04.18.

World Economic Forum (2016A), Annual report 2015-2016.

World Economic Forum (2016B), The future of jobs report. 2016.

World Economic Forum (2018), The future of jobs report 2018.

IFR, "Robot density rises globally", 2018.02.07, <https://ifr.org/ifr-press-releases/news/robot-density-rises-globally/>

아시아타임즈, "무인점포, 어디까지 왔나?", 2018.09.25, <http://www.asiatime.co.kr/news/articleView.html?idxno=198916>

조선일보, "최저임금 뛰자... 무인주문기 판매가 2배로 뛰었다", 2018.07.20, http://biz.chosun.com/site/data/html_dir/2018/07/20/2018072000309.html

조선일보, "코딩 배우려는 학생 밀려드는데 교수, 강의실 부족하다며... 서울대 컴퓨터공학부, 수강 제한", 2019.01.29, http://news.chosun.com/site/data/html_dir/2019/01/29/2019012900152.html?fbclid=IwAR12sa6RFX5ePCrrzxk-UY3pMDAQQmFUyG0vXpRVV9xnjVW8jYbxZrZrVE/

한국경제, "시중은행 ATM 대수 2년새 11% 줄어... "돈 찾기 힘들네", 2018.05.09, <http://news.hankyung.com/article/2018050946746/>

Computerisation bottleneck	O*NET Variable	O*NET Description	PIAAC Variable	PIAAC Description
Perception and Manipulation	Finger Dexterity	The ability to make precisely coordinated movements of the fingers of one or both hands to grasp, manipulate, or assemble very small objects.	Fingers	How often- using skill or accuracy with your hands or fingers?
	Manual Dexterity	The ability to quickly move your hand, your hand together with your arm, or your two hands to grasp, manipulate, or assemble objects.		
	Cramped Work Space, Awkward Positions	How often does this job require working in cramped work spaces that requires getting into awkward positions?		
Creative Intelligence	Originality	The ability to come up with unusual or clever ideas about a given topic or situation, or to develop creative ways to solve a problem.	Problem-solving simple	How often- relatively simple problems that take no more than 5 minutes to find a good solution?
	Fine Arts	Knowledge of theory and techniques required to compose, produce, and perform works of music, dance, visual arts, drama, and sculpture.	Problem-solving complex	Problem solving- complex problems that take at least 30 minutes thinking time to find a good solution?
Social Intelligence	Social Perceptiveness	Being aware of others' reactions and understanding why they react as they do.	Teaching	How often- instructing, training or teaching people, individually or in groups?
	Negotiation	Bringing others together and trying to reconcile differences.	Advise	How often- advising people?
	Persuasion	Persuading others to change their minds or behavior.	Plan for others	How often- planning the activities of others?
			Communication	How often- sharing work-related information with co-workers?
Assisting and Caring for Others	Providing personal assistance, medical attention, emotional support, or other personal care to others such as coworkers, customers, or patients.	Negotiable	How often- negotiating with people either inside or outside your firm or organization?	
		Influence	How often- persuading or influencing people?	
		Sell	How often- selling a product or selling a service?	

[표 1] O*NET variables corresponding to identified engineering bottlenecks
출처: Frey and Osborne(2017), Table 1., Ndelkoska and Quintini(2018), Table 4.2.

순위		자동화 확률	직업
하위10위	1	0.0028	Recreational Therapists
	2	0.0030	First-Line Supervisors of Mechanics, Installers, and Repairers
	3	0.0030	Emergency Management Directors
	4	0.0031	Mental Health and Substance Abuse Social Workers
	5	0.0033	Audiologists
	6	0.0035	Occupational Therapists
	7	0.0035	Orthotists and Prosthetists
	8	0.0035	Healthcare Social Workers
	9	0.0036	Oral and Maxillofacial Surgeons
	10	0.0036	First-Line Supervisors of Fire Fighting and Prevention Workers
상위10위	691	0.99	Data Entry Keyers
	692	0.99	Library Technicians
	693	0.99	New Accounts Clerks
	694	0.99	Photographic Process Workers and Processing Machine Operators
	695	0.99	Tax Preparers
	696	0.99	Cargo and Freight Agents
	697	0.99	Watch Repairers
	698	0.99	Insurance Underwriters
	699	0.99	Mathematical Technicians
	700	0.99	Sewers, Hand
	701	0.99	Title Examiners, Abstractors, and Searchers
	702	0.99	Telemarketers

[표 2] 미국 직업별 자동화 확률 순위

출처: Appendix table in Frey and Osborne (2013)

Industry	Probability
Agriculture, hunting	0.57
Postal and courier activities	0.56
Manufacture of wearing apparel	0.56
Fishing and aquaculture	0.55
Food and beverage service activities	0.55
Manufacture of food products	0.54
Manufacture of wood	0.54
Land transport and transport via pipelines	0.54
Other mining and quarrying	0.54
Waste collection, treatment	0.53
Manufacture of textiles	0.53
Printing and reproduction of recorded media	0.53
Services to buildings and landscape	0.53
Manufacture of tobacco products	0.52
Manufacture of fabricated metal products	0.52
Manufacture of paper	0.52
Manufacture of other non-metallic mineral products	0.52
Manufacture of electrical equipment	0.51
Manufacture of motor vehicles, trailers and semi-trailers	0.51
Forestry and logging	0.51
Architectural and engineering, technical testing	0.43
Employment activities	0.43
Real estate	0.43
Extraction of crude petroleum and natural gas	0.43
Residential care	0.43
Human health	0.42
Telecommunications	0.42
Advertising and market research	0.42
Air transport	0.42
Mining support service activities	0.42
Public administration and defense, social security	0.42
Travel agency, tour operators	0.41
Computer programming, consultancy	0.41
Legal and accounting	0.41
Social work without accommodation	0.40
Programming and broadcasting activities	0.40
Information service activities	0.40
Remediation, other waste management	0.38
Head offices, management consultancy	0.34
Education	0.33

[표 3] 산업별 평균 자동화 확률

출처: Nedelkoska and Quintini (2018), p.52, OECD, figure 4.4

주석: 산업분류는 ISIC Rev 4, 2-digit.

Rank	Today, 2018	Trending, 2022	Declining, 2022
1	Analytical thinking and innovation	Analytical thinking and innovation	Manual dexterity, endurance and precision
2	Complex problem-solving	Active learning and learning strategies	Memory, verbal, auditory and spatial abilities
3	Critical thinking and analysis	Creativity, originality and initiative	Management of financial, material resources
4	Active learning and learning strategies	Technology design and programming	Technology installation and maintenance
5	Creativity, originality and initiative	Critical thinking and analysis	Reading, writing, math and active listening
6	Attention to detail, trustworthiness	Complex problem-solving	Management of personnel
7	Emotional intelligence	Leadership and social influence	Quality control and safety awareness
8	Reasoning, problem-solving and ideation	Emotional intelligence	Coordination and time management
9	Leadership and social influence	Reasoning, problem-solving and ideation	Visual, auditory and speech abilities
10	Coordination and time management	Systems analysis and evaluation	Technology use, monitoring and control

[표 4] 노동시장에서 요구되는 기술: 2018년과 2022년 비교

출처: World Economic Forum (2018), The future of jobs report 2018, p.12, table 4

회사명	내용
롯데리아	1350개 매장 중 762개(56%) 매장에서 키오스크 운영
버거킹	현재 67% 매장(313개 중 210개)서 키오스크 운영 중, 전 매장으로 확대 예정
CU	무인 결제 매장 3곳에서 10곳으로 확대
이마트	144개 매장 중 40곳에서 무인 계산대 운영 중
베스킨라빈스	수도권 5개 매장에 자판기 방식 '아이스크림 ATM' 도입
애슐리클래식	손님이 직접 사용 식기를 정리하는 셀프 서비스 매장 도입

[표 5] 외식 유통업체들의 서비스 로봇 도입 현황

출처: 조선일보, 2018.07

	Arntz et al (2016)			Nedelkoska and Quintini(2018)	
	Share of people at high risk	Mean	Median	Mean	Median
Austria	12%	43%	44%	47%	48%
Belgium	7%	38%	35%	48%	49%
Canada	9%	39%	37%	46%	46%
Czech Republic	10%	44%	48%	45%	45%
Denmark	9%	38%	34%	48%	49%
Estonia	6%	36%	32%	45%	44%
Finland	7%	35%	31%	46%	47%
France	9%	38%	36%	43%	41%
Germany	12%	43%	44%	49%	51%
Ireland	8%	36%	32%	52%	54%
Italy	10%	43%	44%	46%	45%
Japan	7%	37%	35%	49%	52%
Korea	6%	35%	32%	51%	53%
Netherlands	10%	40%	39%	46%	47%
Norway	10%	37%	34%	45%	44%
Poland	7%	40%	40%	41%	40%
Russian Federation	2%	29%	26%	50%	52%
Slovak Republic	11%	44%	48%	47%	49%
Spain	12%	38%	35%	57%	62%
Sweden	7%	36%	33%	51%	54%
United Kingdom	10%	39%	37%	44%	43%
United States	9%	38%	35%	43%	41%

[표 6] OECD 국가의 자동화 확률

출처: Arntz et al (2016), p.33, table 4; Nedelkoska and Quintini(2018), p.46, table 4.5

순위	김세움(2015)		김건우(2018)	
	자동화 확률	직업명	자동화 확률	직업명
하위10위	0.238	보건, 의료 관련직	0.004	영양사
	0.254	교육 및 자연과학, 사회과학 연구 관련직	0.004	전문 의사
	0.277	사회복지 및 종교 관련직	0.004	장학관, 연구관 및 교육 관련 전문가
	0.302	정보통신 관련직	0.007	교육 관리자
	0.325	문화, 예술, 디자인 방송 관련직	0.007	보건의료관련 관리자
	0.400	법률, 경찰, 소방 교도 관련직	0.008	중고등학교 교사
	0.424	관리직	0.009	학습지 및 방문 교수
	0.540	미용, 숙박, 여행, 오락, 스포츠 관련직	0.001	컴퓨터시스템 설계 및 분석가
	0.651	경영, 회계, 사무 관련직	0.012	특수교육 교사
	0.666	운전 및 운송 관련직	0.012	약사 및 한약사
상위10위	0.854	섬유 및 의복 관련직	0.990	통신서비스 판매원
	0.847	영업 및 판매 관련직	0.990	텔레마케터
	0.778	경비 및 청소 관련직	0.990	인터넷 판매원
	0.762	식품가공 관련직	0.990	사진인화 및 현상기 조직원
	0.753	기계 관련직	0.985	관세사
	0.740	음식서비스 관련직	0.985	무역 사무원
	0.723	금융, 보험 관련직	0.980	전산 자료 입력원 및 사무 보조원
	0.720	화학 관련직	0.970	경리 사무원
	0.715	재료 관련직	0.970	상품 대여원
	0.715	농업어업 관련직	0.970	표백 및 염색 관련 조직원

[표 7] 국내 직업별 자동화 확률 순위

출처: 김세움(2015) 표 2-3 재정리, 김건우(2018), p.7, 차트 6.

주석: 김세움(2015)은 직종 중분류(총 23개)를 사용한 반면, 김건우(2018)은 세분류(총 2,423개)를 활용하여 분석함.

중·세분류 중 군인은 Frey and Osborne(2013)에서 분석되지 않아 두 분류 모두에서 제외됨.

윤리적 인공지능의 실현과 과제

고학수(교수, 서울대학교 법학전문대학원)

이나래(변호사, 서울대학교 법학대학원 박사과정)

박도현(변호사, 서울대학교 법학대학원 박사과정)



웹에서 PDF 바로 보기

본고는 인공지능 시대에 우리나라가 앞으로 어떠한 형태의 윤리규범 체계를 갖추어야 하는가라는 문제의식에 대한 실마리를 제공하는 것을 목적으로 하는 글이다. 아시모프 ‘로봇 3원칙’ 이래로, 인공지능 윤리규범 이슈는, 첫째, 논의의 주체가 설계자, 제작자, 이용자와 같이 배후에 위치한 인간에게로 확장되었고, 둘째, 책임귀속에 관련된 과거의 대전제가 오늘날에는 통용되지 않게 된 경우가 많아졌으며, 셋째, 인공지능에 대한 대중의 알 권리와 이해관계자의 주체적 참여에 대한 요구가 늘어나면서 책임성(accountability)에 대한 관심이 확장되고 있고, 넷째, 일률적인 규정을 기계적으로 적용하는 방식의 한계가 지적되고 그 대신 인공지능이 활용되는 구체적 맥락(context) 위주의 접근법에 대한 필요성이 강조되는 방향으로 논의가 이루어지고 있다.

인공지능의 윤리적 측면에 관한 국내외의 논의는 최근 2년여의 기간 동안 급속도로 진전이 이루어졌다. 그동안 해외에서 제시된 윤리규범은 대체적으로는, ① 기본원칙을 중심으로 하는 윤리규범 유형, ② 기본원칙에 더하여 주요 이슈에 대해 상세하게 함께 다룬 윤리규범 및 보고서 유형, ③ 윤리규범의 정립방안에 대한 구체적 방법론을 함께 언급한 유형으로 구분지어 파악할 수 있다. 기본원칙 중심의 윤리규범은 OECD 권고안, 일본 총무성 가이드라인, 아실로마 원칙, 그리고 그 이외에 마이크로소프트, 구글과 같은 대기업에서 발표한 규범 등을 들 수 있다. 기본원칙 중심의 윤리규범은 다양한 참여자들의 다양한 관점 및 상이한 이해관계를 종합하기에 용이한 방식인 한편, 추가적인 후속 규범이 마련되지 않으면 규범력이 상대적으로 낮을 가능성이 있다. 다음으로, 기본원칙과 주요 이슈를 함께 다룬 윤리규범 유형으로는 IEEE 보고서, 영국 상원 보고서, 그리고 UNGP & IAPP의 보고서를 들 수 있다. 구체적 논의에 대한 세부사항까지도 포함하는 셋째 유형으로는 유럽의회 결의안, EU 집행위원회 고위급 전문가 그룹의 가이드라인 등을 들 수 있다. 이러한 윤리규범 중에는, 원칙에 대한 선언 그 자체로 의미를 가지는 것도 있고, 후속 작업과 모니터링의 과정을 통해 더 구체화하고 규범력을 확보할 수 있는 장치가 마련된 것도 있다.

우리나라에서 인공지능 윤리 이슈에 대한 논의는 2007년 발표된 ‘로봇윤리헌장 초안’에서 시작되었다. 세계적으로도 빠른 편이다. 그러나 그로부터 10여년이 지난 지금도 논의가 크게 진전되지는 못한 상황이다. 하지만, 알파고에서 촉발된 충격 이후로 인공지능 규범 마련에 대한 관심이 늘어나면서, 로봇기본법안, 지능정보사회 윤리헌장, 국가정보화 기본법 전부개정법률안 등 다양한 형태로 관심이 표출되고 있는 중이다. ‘개발자, 공급자, 이용자 등에 대한 공공성 책무성 통제성 투명성’ 등의 명제로 대표되는 국내의 논의는 해외의 논의와 대체로 일맥상통한 내용을 담고 있다. 그러나 인공지능 규범 논의는 국내의 실정과 부합하는 동시에 국제사회와 발맞추어 진행해나가야 한다는 점, 일견 원론적이고 교과서적인 내용의 단순한 나열로 비쳐지는 경우에도 그 배후에는 이해관계자들 사이에 자신에게 유리한 방향으로 향후 규범논의를 이끌어가려는 치열한 이익대립이 숨겨져 있는 점을 고려하여 관련 이슈들에 대한 면밀한 검토와 분석이 필요하다.

인류가 지향해야 할 궁극의 목표는 인공지능이 인간 존엄성과 기본권 실현을 위한 방향으로 활용되어야 한다는 점이다. 따라서 그러한 목적과 부합하지 않는 여러 관념, 특히 이분법적 관념들은 인공지능 시대와 부합하도록 변화를 모색할 필요가 있다. 가령, 윤리를 비롯한 여타 사회규범 유형을 실정법적 법규범과 엄밀히 구별하는 태도, 기술과 규범을 엄밀하게 분리하는 태도는 이 논의의 맥락에서는 대체로 도움이 되기 어려운 태도이다. 구속력과 강제력은 높지 않지만, 구성원에게 사실상의 행위규범으로 기능하는 ‘연성법(soft law)’이 인공지능 윤리규범 논의의 대안으로 부상하는 것에 주목할 필요가 있다. 또한, 인공지능의 ‘책임’에 관해서도, 법적 책임을 의미하는 ‘liability’ 개념이 많이 언급되지 않고, 그보다 더 폭넓은 의미의 사회적 책임을 의미하는 ‘accountability’ 개념이 좀 더 흔하게 언급되는 것은 중요한 함의를 갖는다.

1. 들어가는 글

인공지능 기술이 급속히 발전하고 널리 확산되면서, 부작용과 피해를 미연에 방지하기 위한 규범적 논의도 점차 늘어가는 중이다. 미국 스탠퍼드 대학에서는 2016년 9월 인공지능 100년 연구 첫 번째 보고서를 발표했고, 오바마 행정부는 2016년 10월과 12월 기술적 연구개발과 별도로 사회 경제적 이슈에 집중한 보고서를 내놓았다. AI Now Institute는 2016년부터 주로 사회 윤리적 이슈에 관한 연례보고서를 발표해왔다. 구글(Google)에서도 올해 초 이와 같은 대열에 동참하여 거버넌스 보고서를 발표했다. 마이크로소프트(Microsoft), 인텔(Intel), 소니(Sony)를 비롯한 글로벌 기업들 또한 보고서, 원칙(principle), 가이드라인, 모범사례(best practice) 등 명칭을 불문하고 윤리적 고려를 반영한 나름대로의 실무관행을 구축해가고 있다.

윤리적 인공지능에 대한 관심은 유럽에서도 높은 수준으로 나타나고 있다. 유럽에서는 인공지능의 윤리적 측면에 대한 관심이 일찍이 나타나기 시작했고, 유럽연합(European Union, EU) 차원에서의 논의를 보면 엄격한 법제를 도입함으로써 신기술 활용과 기본권 보호를 조화하려는 움직임이 좀 더 두드러지게 나타난다. 대표적 예로, 유럽연합에서는 개인정보보호규정(General Data Protection Regulation, 이하 “GDPR”)이 2018년 5월 25일부터 시행되고 있는데, GDPR에는 인공지능의 윤리적 측면에서 중요한 의미를 지니는 조항들이 적지 않게 포함되어 있다. 핵심적으로, GDPR은 개인정보 주체에게 법적 효력이나 그와 유사한 중대한 효과를 미칠 경우에 프로파일링을 포함한 ‘오직 자동화된 의사결정’의 대상이 되지 않을 권리를 인정한다. 정보주체는 또한 ‘설명을 요구할 권리(right to explanation)’와 ‘잊힐 권리(right to be forgotten)’로 널리 알려져 있는 정보 제공권, 열람권, 삭제권, 반대권과 같은 다양한 권리를 행사할 수 있다. GDPR은 유럽연합 역내 모든 구성원에게 적용되므로 해외의 사업자도 경우에 따라 이를 준수할 의무가 부과되고, 이를 위반할 경우 강력한 제재가 동반된다.

이 글은, 우리나라가 제4차 산업혁명 시대에 인공지능의 윤리적 측면에 관하여 어떤 형태의 규범체계를 갖추어야 하는가라는 질문을 염두에 두고, 최근의 국내외 논의 동향은 어떠한지 살펴보고 이로부터 시사점을 모색해 보고자 한다. 이하에서는 먼저 인공지능 윤리규범 논의의 변천사를 개관하면서 그동안 지적되어온 주요한 쟁점을 살펴보도록 한다(II). 다음으로 이런 과정을 거쳐 정립된 윤리규범의 현황을 살펴보기 위하여 해외 주요 윤리규범의 내용을 고찰하고 그 의의와 한계를 도출한다(III). 끝으로 여기서 얻은 시사점을 국내 윤리규범 논의에 적용하고 향후 나아가야 할 방향을 모색한다(IV).

본격적 논의에 들어가기에 앞서, 본고의 몇 가지 전제사항에 대해 밝혀두고자 한다. 첫째로, 본고에서 분석의 대상으로 삼는 ‘인공지능’은 소프트웨어 격에 해당하는 알고리즘뿐만 아니라, 학습의 원천인 (빅)데이터, 나아가 자율주행 자동차와 같은 하드웨어(로봇)가 부착된 경우까지 포괄한 넓은 개념이다. 둘째로, 본고의 ‘인공지능’은 오늘날 국내외 학계와 실무의 연구개발 대상인 유형에 한정하도록 한다. 일부 매체에서는 인류와 같은 방식으로 사고하고 행동하거나, 그 이상의 인지능력을 확보하여 인류의 생존을 위협하는 미래지향적 인공지능의 모습을 그려내고 있기도 하다. 물론 이러한 인공지능에 관해서도 ‘인공지능 윤리’의 측면에 관해 다룰 수는 있지만, 이 글은 현시점을 기준으로 하여 실제로 일반 이용자들을 대상으로 제공되고 있는 (또는 상당히 가까운 장래에 제공될 것으로 예상되는) 유형의 기술을 전제로 논의를 전개한다. 셋째로, 본고에서의 ‘윤리’는 사회를 보다 바람직한 방향으로 이끌기 위하여 사회구성원에게 요구되는 사회규범의 일종으로, 법규범과 비교하여 강제성과 명확성은 다소 낮지만 도덕성과 유연성은 보다 높은 유형으로 간주한다.⁰¹

2. 인공지능 윤리의 변천사와 주요 쟁점

1. 윤리주체의 다변화와 인간을 중시하는 윤리관

인공지능 윤리규범의 시초로 많은 이들은 SF 소설가 아이작 아시모프(Isaac Asimov)가 1942년 런어라운드(Runaround)라는 소설에서 최초로 언급한 ‘로봇 3원칙’을 꼽는다. 3원칙의 내용은 다음과 같다. 첫째, 로봇은 인간에게 해를 끼치거나, 어떠한 행동도 하지 않아 인간에게 해가 가해지도록 하면 안 된다. 둘째, 로봇은 1원칙에 위배되지 않는 한 인간의 명령에 복종하여야 한다. 셋째, 로봇은 1원칙과 2원칙에 위배되지 않는 한 자신을 보호하여야 한다.

로봇 3원칙을 엄격하게 고수하면, 극단적으로는 인간에게 위해의 여지가 있으면 정당방위와 같은 해악을 방지하기 위한 개입조차도 할 수 없다는 반직관적 결론에 다다른다. 아시모프는 1985년 로봇과 제국(Robots and Empire)이라는 소설에서 “로봇은 인류에게 해를 끼치거나, 어떠한 행동도 하지 않음으로써 인류에게 해를 가하도록 하면 안 된다”는 ‘0원칙’을 제시한다. 이는, 3원칙의 ‘인간’ 자리에 ‘인류’를 도입하여 위와 같은 모순적 결론을 방지하려는 것이다.⁰²

01 김중호, “인공지능 시대의 윤리와 법적 과제”, 『과학기술법연구 제24권 제3호』, 2018, 184-187면 참조.

02 고인석, “아시모프의 로봇 3법칙 다시 보기: 윤리적인 로봇 만들기”, 『철학연구 제93집』, 2012, 102면.

로봇 3원칙은 이하에서 살펴볼 2006년 유럽로봇연구 네트워크 로봇윤리 로드맵의 출발점이 되는 등, 인공지능 윤리 논의에서 나름대로의 역할을 담당하였다. 그러나 어디까지나 소설의 일부에 불과한 로봇 3원칙만으로는 윤리적 문제를 해결하는 데 한계가 있었다. 로봇 3원칙은 무엇이 문제였을까? 선언적 원칙이어서 구체성도 부족하였지만 무엇보다도 윤리주체가 인간이 아니라 로봇인 점이 주로 지적되었다. 로봇을 만들고 이용하는 ‘인간’이 준수하여야 할 윤리를 ‘로봇’에 전가할 수 있도록 오해될 여지가 있었기 때문이다.⁰³ 이러한 문제는 최초의 인공지능 윤리규범으로 평가받고 있는, 일본 후쿠오카에서 2004년 발표된 ‘세계로봇선언(World Robot Declaration)’에서도 마찬가지였다.⁰⁴ ‘로봇과 인간의 공존’을 중시하는 세계로봇선언은 로봇이 인간에게 일방적으로 복종하는 관계로 설정한 ‘로봇 3원칙’에 비해 진일보한 것으로 평가될 수 있지만, 세계로봇선언에서도 인간의 윤리에 대한 언급은 부재했다.

다른 한편, 유럽로봇연구 네트워크(European robotics research Network, EURON)은 2003년부터 3년 동안의 연구를 거쳐 2006년에 ‘로봇윤리 로드맵’을 발표했는데, 이 로드맵은 인간의 윤리를 주된 목표로 함을 명시했다.⁰⁵ 3대 이해당사자인 설계자(designer), 제작자(manufacturer), 이용자(user)를 특정하여, 로봇을 제작하거나 활용할 때 중시해야 할 인간의 존엄성과 권리를 비롯한 13대 원칙을 통해 이러한 목표를 구체화했다. 이듬해인 2007년 우리나라 산업자원부도 인간중심, 인간과 로봇의 공존, 인간과 로봇의 윤리라는 그동안의 논의를 종합적으로 반영한 ‘로봇윤리헌장’ 초안을 공개하였다. 이런 과정을 거치면서, 오늘날 인공지능 윤리는 대략적으로 ① 인공지능 자체의 윤리, ② 인공지능 설계자나 제작자의 윤리, ③ 인공지능 이용자의 윤리라는 세 가지 범주로 나누어 볼 수 있게 되었고, 인간과 인공지능의 공존 속에서 인간의 존엄성과 기본권을 도모하는 윤리관이 수립되었다.⁰⁶

그렇다면 ‘인공지능 자체의 윤리’는 어떤 모습으로 발전해왔을까?

03 Robin R. Murphy David D. Woods, “Beyond Asimov: The Three Laws of Responsible Robotics”, IEEE Intelligent Systems Vol. 24, Issue 4, 2009, pp. 14-20은 이를 지적하고, 3원칙의 대안을 제시한다.

04 International Robot Fair 2004, “World Robot Declaration”, <http://prw.kyodonews.jp/prwfile/prdata/0370/release/200402259634/index.html>, 2004. 2. 25.자.

05 Gianmarco Veruggio, EURON Roboethics Roadmap(ver. 1.1), 2006, p. 7. 이는 이듬해 발표된 로드맵 1.2버전에서도 마찬가지이다.

06 정채연, “지능정보사회에서 지능로봇의 윤리화 과제와 전망 - 근대적 윤리담론에 대한 대안적 접근을 중심으로 -”, 『동북아법연구 제12권 제1호』, 2018, 90-93면; Peter M. Asaro, “What Should We Want From a Robot Ethic?”, International Review of Information Ethics Vol. 6, 2006, pp. 9-16 참조.

오늘날 인공지능 자체의 윤리는 로봇 3원칙을 넘어, 인공지능의 의사결정이 인간의 윤리적 직관과 부합하게 조정하는 보다 적극적인 방법론을 가리키는 것으로 이해된다. 윤리적 의사결정의 주체인 인공지능은 통상적으로 ‘인공적 도덕행위자(Artificial Moral Agent, 이하 “AMA”)’로 불린다.⁰⁷ AMA를 만드는 기술적 방식은 윤리적 규칙을 학습시키는 방식(하향식 접근), 윤리적 사례를 학습시키는 방식(상향식 접근), 양자를 혼합한 방식으로 대별된다. 그러나 어떤 방식이건, 가령 자율주행차가 사고에 직면한 상황에서 탑승자와 보행자 중 누구를 우선시할지와 같은 현실적 문제(일명 ‘트롤리 딜레마’)에서 한계를 지적받고 있다. 단순화하여 생각하면, 단순한 공리주의 원칙을 적용한 자율주행차라면 2명 이상의 보행자를 살리기 위하여 탑승자를 사망에 이르는 의사결정을 할 것인데, 소비자들이 그와 같은 자율주행차는 구매하지 않을 것이기 때문에 시장의 성립 자체가 어려울 수 있다는 것이다.⁰⁸

2. 책임귀속의 어려움을 극복하기 위한 방법론 모색

한편, 사회규범의 목적이 구성원들의 행위를 바람직한 방향으로 이끄는 것에 있다고 본다면, 그러한 목적과 부합하지 않은 행위를 했거나, 할 것으로 예상되는 구성원에 대한 조정(coordination) 문제가 사회규범의 핵심적 역할이 된다. 전통적으로 조정기능을 매개하는 수단으로 활용되어온 것이 바로 ‘책임’ 개념이고, 사회규범을 대표하는 윤리와 법 분야에서의 ‘responsibility’와 ‘liability’가 여기에 대응되는 용어이다. 흔히들 말하는 “법은 도덕의 최소한”이라는 격언에 의하면 전자가 후자에 비해 다소 넓은 개념으로 볼 수 있지만, 반드시 그렇지는 않다. 그동안 양자의 본성과 관계에 대하여는 수많은 논의가 있어왔지만, 간단명료한 답이 도출된 것은 아니다. 다만, 양자는 밀접한 관계가 있어, 어떤 경우는 법규범이, 어떤 경우는 도덕이나 윤리규범이 서로를 이끌어가는 상호보완적 관계라고 할 수 있다.⁰⁹ 이들은 오랫동안 비슷한 전제를 공유해왔는데, 어떤 행위는 자율적 주체인 인간의 자유의지에서 비롯된 것으로, 주체는 행위가 낳는 결과를 예견할 수 있고, 주체와 행위 간의 인과관계가 명확하다는 것이다.¹⁰

07 대표적 연구로, 웬델 윌러치 콜린 알렌, 노태복 역, 『왜 로봇의 도덕인가』, 메디치미디어, 2014가 있다.

08 Jean-François Bonnefon Azim Shariff Iyad Rahwan, “The social dilemma of autonomous vehicles”, Science Vol. 354, Issue. 6293, 2016, pp. 1573-1576.

09 김건우, “로봇윤리 vs. 로봇법학: 따로 또 같이”, 『법철학연구 제20권 제2호』, 2017, 33면 이하 참조.

10 Merel Noorman, “Computing and Moral Responsibility”, Stanford Encyclopedia of Philosophy, <https://plato.stanford.edu/entries/computing-responsibility/>, 2018. 2. 16.자.

이러한 관점에서, 인공지능 책임성이란 인공지능의 의사결정에 의해 사회적 문제가 발생했을 때, 원인이 되는 행위를 한 주체에게 귀속되는 도덕적, 법적 책임을 뜻하는 것으로 볼 수 있다. 이는, 이해당사자 누군가에게 책임이 귀속되어야 하고, 책임공백이 있어서는 안 된다는 ‘이해관계자의 책무’ 개념을 암묵적으로 전제하는 것이다. 영국의 공학, 물리학 연구위원회(Engineering and Physical Science Research Council, EPSRC)가 2010년 발표한 로봇윤리 5원칙에서 ‘법적 책임의 인간에 대한 귀속’을 명시하고 ‘안전과 보안’을 특히 강조한 것은 그 일환으로 바라볼 수 있다.¹¹

그렇지만 오늘날 인공지능의 기술적인 특징은 이러한 책임 개념과 자연스럽게 조응하지 않는 면이 있어서, 전통적 책임관을 그대로 적용하면 소위 책임격차(responsibility gap)가 발생하여 문제해결이 어려울 수 있다. 예를 들어, 마이크로소프트가 2016년 3월 대화형 인공지능으로 개발한 챗봇 테이(Tay)가 악성 이용자들로부터 학습한 인종차별적 발언을 트윗을 하여 사회적 논란이 발생한 상황을 보자.¹² 이 경우는 테이에게 혐오표현을 학습시킨 이용자를 제재하여 문제를 해결할 수 있을 것이다. 그러나 인공지능 알고리즘의 훈련 데이터가 과거의 편견을 반영한 경우, 문제를 유발한 주체가 불특정 다수인 과거의 인류라면 책임규명이 어려워질 수 있다. 그밖에 무수히 많은 참여자나 데이터가 개입하는 오늘날 소스코드 제작과정을 생각해보면 책임소재를 가리는 일은 사실상 불가능해질 수 있다(이를 ‘many hands의 문제’라고 한다).¹³

다른 한편, 인공지능 의사결정의 불투명성(opacity)은 새로운 형태의 문제를 낳는 원인이 된다. 여기서 불투명성은 ① 인공지능 보유주체의 지식재산권이나 계약상 특약조항과 같은 ‘제도적 측면’, ② 인공지능경망의 복잡성과 같은 알고리즘의 ‘본질적 측면’, ③ 검증과정의 인적, 물적 비용문제로 인한 ‘현실적 측면’ 등 다종다양한 이유로 인해 발생된다.¹⁴ 이러한 불투명성은 오늘날 인공지능 의사결정의 증대된 자율성(autonomy)과 결부되어, 현실적 문제해결 능력이 미약한 인공지능 그 자체에게 책임을 돌려야만 하는 것처럼 여겨질 수 있다. 결과적으로 책임을 귀속할 행위자가 부재하게 되거나 또는 그 정반대로 너무 많아져서 책임귀속이 힘들어질 수 있다.

11 Engineering and Physical Science Research Council, “Principles of robotics”, <http://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/>, 2010. 9.자.

12 Peter Lee, “Learning from Tays introduction”, <https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/>, 2016. 3. 25.자.

13 Helen Nissenbaum, “Accountability in a Computerized Society”, Science and Engineering Ethics Vol. 2, Issue. 1, 1996, pp. 28-32.

14 고학수 정해빈 박도현, “인공지능과 차별”, 『저스티스 통권 제171호』, 2019, 235-236면.

유럽연합이 2012년부터 3년 여 동안의 ‘로봇법 프로젝트(RoboLaw Project)’ 연구를 거쳐 2014년 발표한 로봇규제 가이드라인은 그 대안으로 3가지의 기준 원칙을 제시한다.¹⁵ 첫째, 가장 기술친화적 원칙으로, 로봇산업의 혁신을 촉진하고 규제비용을 절감하기 위하여 기술적으로 불가피한 책임의 경우는 (적어도 단기적으로는) 면책(immunity)을 인정하는 방안이다. 둘째, 일정한 수준의 자율성을 획득한 로봇에 대하여 일종의 법인격(legal personhood)을 부여하는 방안이다. 이러한 방법론은 이후 로봇에게 법적 책임을 부여할 수 있도록 ‘전자인(electronic person)’ 지위를 부여할 필요가 있다는 유럽연합 의회 결의안¹⁶으로 이어진다. 셋째, 기본권 보호 정도가 가장 높은 방안으로, 민법상 특수불법행위나 제조물책임법과 같은 일부 특별법에 마련되어 있는 무과실책임(strict liability)을 법제화하는 것이다.

3. 투명성의 확보와 사회구성원의 주체성 강화

위와 같은 세 가지 방안은 윤리적 책임(responsibility)의 귀속이 어려워지는 현실적 문제를 적어도 민 형사상의 법적 책임(liability) 측면에서는 어느 정도 해결하였다고 평가할 수 있다. 그런데 최근 들어 인공지능의 자율적 의사결정이 점차 늘어나면서, 사회적으로 새로운 차원의 문제제기가 이루어져왔다. 그것은 앞서 언급한 인공지능의 불투명성으로 인하여 사회구성원의 알 권리와 절차적 참여권이 제대로 보장되지 않는다는 것이었다. 인공지능이 행한 의사결정이 인류의 삶에 중대한 영향력을 행사하고 있음에도 불구하고, 그러한 의사결정이 어떠한 근거로 이루어졌는지, 어떠한 절차를 거쳐서 이익제기를 할 수 있는지에 대한 문제의식은 상대적으로 미약했던 것이다.¹⁷ 예컨대 앞서 본 테이 사례의 당사자의 경우, 민사상 손해배상은 별론으로 하더라도, 사적자치의 원칙에 비추어 기업체에 단지 인공지능의 혐오표현을 근거로 인공지능 사용을 중단하라는 청구권을 행사할 수 있는지는 현행법상 논란이 따른다. 나아가 이와 같은 문제는 향후 법적 책임을 평가함에 있어 ‘증명(proof)’의 현실적 어려움과도 맞닿아 있다.

‘기술영역의 적법절차(technological due process)’로 불리는 일련의 문제제기가 이어지며, 인공지능의 불투명성을 해소하는 규범적 노력이

15 Erica Palmerini et al., “Guidelines on Regulating Robotics”, 2014, pp. 23-24.

16 European Parliament, “European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))”, 2017. 2. 16.자.

17 프랭크 파스칼레는 이러한 상황을 ‘블랙박스 사회(black-box society)’라는 말로 표현한다. 프랭크 파스칼레, 이시은 역, 『블랙박스 사회』, 안티고네, 2016.

‘투명성(transparency) 원칙’이라는 명칭으로 대두되었다.¹⁸ 그리고 여기서 파생된 ‘책임성’ 개념이 바로 ‘accountability’이다(이를 한글로 표현할 때, ‘책임(성)’ 또는 ‘책무(성)’이라고도 하고 다른 유형의 책임과 구분하여 ‘설명책임’이라 하기도 한다).¹⁹ 설명책임은 단지 인공지능 의사결정의 효력이 배후의 인간 행위자에게 귀속되는 차원을 넘어, 그런 의사결정이 어떤 과정을 거쳐 이루어졌는지 설명하도록 요구하는 책임을 의미한다. 이를 현실에서 구현하기 위한 구체적인 방법론으로서, 그동안 인공지능 감사(audit), 검사(review), 영향평가(impact assessment)와 같은 공법적 규제부터, 설명책임의 달성 정도에 따라 면책의 수준을 달리 인정하는 것과 같은 사법적 해석론까지 다양한 논의가 이루어지고 있다.

인공지능 규범체계 관점에서 ‘거버넌스(governance)’라는 용어가 주로 사용되는 이유 역시 같은 맥락에서 설명된다. 주로 책임의 부과에 초점이 맞춰진 하향식(top-down)의 규범체계를 넘어 구성원의 능동적 참여를 포용하는 상향식(bottom-up)의 규범체계를 전제하기 때문이다. 그렇게 보면 ‘accountability’는 윤리적, 법적 책임 개념(responsibility, liability)이 담당해오던 역할에 설명책임과 능동적 참여 기능을 더한 광의의 책임성 개념인 것으로 이해할 수 있을 것이다. 그런 점에서, 진입규제나 형사처벌과 같은 전통적 방식의 규제뿐만 아니라, AMA나 소위 ‘설명가능 인공지능(explainable AI, XAI)’ 등을 통해 알고리즘 의사결정 방식을 설명하는 기술적 접근법 또한 얼마든지 책임성 원칙을 달성하는 적절한 수단이 될 수 있다. GDPR에서 설명을 요구할 권리나 잊힐 권리를 정한 이유도 책임성에 대한 광의의 이해방식을 취할 때 보다 분명한 파악이 가능해진다. 이와 병행하여 구성원들의 인공지능 문해력(literacy)을 증진하는 노력도 장기적으로 요구되며, 인공지능 활용에 따른 혜택을 누리고 위험을 초래한 당사자가 비용을 분담할 수 있는 메커니즘을 구축하는 것이 책임성을 구현할 수 있는 한 가지 방안이다. 그렇다면 인공지능 시대의 책임성(accountability) 원칙이란 인공지능의 활용을 인간과 인공지능의 공존공영으로 이끄는 사회규범의 총체적 활용과정으로 재구성할 수 있다. 인공지능의 활용에 제약을 두는 것이 책임성을 강화하는 전형적 형태지만, 경우에 따라 인공지능을 적극적으로 활용하여 책임성을 강화할 수도 있다는 것이다. 인공지능 책임성 원칙은 오늘날 인공지능 규범 논의에서 가장 기본적이고 핵심적인 원칙의 하나로 인정받고 있다.

18 고학수 정해빈 박도현, “인공지능과 차별”(주 14), 245면 이하 참조.

19 Robyn Caplan et al., “Algorithmic Accountability: A Primer”, Data&Society, 2018, p. 10.

4. 이중효과의 조화와 맥락 중심의 접근방식

다른 한편, 인공지능의 강한 복잡계(complex system)적 특성은 규범적 평가를 일률적으로 행하기 어렵게 만드는 요인이 된다. 예를 들어, 인공지능 의사결정의 차별적 효과를 검증하는 과정에서, 불투명성을 극복하기 위하여 때로는 문제되는 이해당사자의 개인정보나 민감정보를 수집할 필요가 있는데, 이는 프라이버시라는 또 다른 중대한 기본권과 정면으로 충돌한다.²⁰ 기본권 사이의 충돌을 넘어, 단일한 기본권에 대해 유발된 효과가 상충할 수 있다. 예를 들어, 인공지능의 활용은 누군가에게는 노동권의 신장을, 누군가에게는 노동권의 박탈을 낳는다.²¹ 경쟁, 자율성, 삶의 질(well-being), 민주주의 등에 대한 우려도 같은 맥락에 위치한다.

그렇지만 해악의 위험과 혜택의 가능성이라는 일종의 ‘이중효과(double effect)’를 딜레마나 역설의 원천으로만 바라볼 것은 아니다. 얼마든지 양자 사이의 발전적 조화를 구상해 볼 수 있다. 예를 들어, 단순히 인공지능이 부가가치를 창출한다는 등의 일반론적인 이익을 넘어, 자율주행차는 장애인이나 노약자의 이동권을 강화하고, 드론은 물리적 접근성이 낮은 지역의 운송에 기여할 수 있는 등 새로운 기술을 통해서 새로이 구현가능하게 된 기능이 존재한다. 다른 한편, 2018년 우리나라에서 논란이 발생한 바 있는, ‘킬러 로봇’을 포함한 ‘치명적 자율무기(Lethal Autonomous Weapons, LAWS)’의 연구개발 및 사용과 관련된 이슈에 관해서는 국제사회의 논의에 적극적으로 참여하고 주도해 나가도록 노력해야 할 것이다.²² 더 넓게는 인공지능의 복잡계적 특성을 고려하여, 해악의 의도가 없어도 예측하지 못한 악결과가 나타날 수 있음을 인지하고 그에 대비하여야 한다. 그렇다고 하여, 이중효과를 근거로 특정 유형의 신기술에 대한 원천적 금지를 규범적 대안으로 제시하는 접근방식은 신중해야 할 필요가 있다.

이처럼 인공지능에 대한 선형적 판단 대신에 의사결정의 구체적 상황에 따라 규범적 판단이 달라져야 한다는 접근방식을 ‘맥락(context) 중심의 접근방식’이라 부르는데, 이러한 맥락중심의 접근방식은 인공지능 규범논의에서 점차 더욱 강조되고 있다. 인공지능은 자율성을 증진할 수도 억제할 수도, 인류에게 혜택이 될 수도 해가 될 수도 있다는 것을 전제로, 개별 맥락을 고려하여 규범적 또는 정책적 판단을 해야 한다는 것이다.

20 고학수 정해빈 박도현, “인공지능과 차별”(주 14), 255면.

21 U.S. Executive Office of the President, “Artificial Intelligence, Automation, and the Economy”, 2016, p. 10 이하 참조.

22 전문가들의 공개서한으로, Future of Life Institute, “Autonomous Weapons: An Open Letter from AI & Robotics Researchers”, <http://futureoflife.org/open-letter-autonomous-weapons/>, 2015. 7. 28.자 참조.

앞서 본 유럽연합의 로봇규제 가이드라인은 일반론 대신 사례별 접근을 택하고, 규제와 산업발전이 공존 가능성을 전제로 하며, 기술을 바탕으로 한 자율규제, 윤리원칙, 엄격한 법제를 상호 조화한 프레임워크를 제시함으로써 이러한 사고를 반영하였다.²³

지금까지 인공지능 윤리규범의 변천사를 추적해보면서 주요한 쟁점을 살펴보았다. 오늘날의 윤리규범은 큰 틀에서 이와 같은 이해방식을 공유하고 그에 기초하고 있다. 그러나 영역이나 지역 등의 기준 하에 비교해보면, 강조하는 측면이나 구속력의 정도 등에 있어서는 어느 정도 차이가 존재한다. 국제사회에서 이루어지는 논의의 공통점과 차이점을 분명히 파악하였을 때, 비로소 우리나라의 실정에 맞는 특유한 규범을 확립할 수 있다는 측면에서 최신규범의 내용을 고찰할 필요가 특히 강하다. 또한 국가적인 차원에서는 해외의 논의에 적극 참여하고 주도해야 할 필요도 있다. 이런 필요성에 비추어, 이하에서는 해외의 최신 윤리규범 내용을 비교 분석해보도록 한다.

3. 해외의 최신 인공지능 윤리규범 비교 분석

1. 개괄

아래에서는 해외에서의 인공지능 윤리이슈에 대한 주요 논의를 검토하면서 그 시사점을 찾아보고자 한다. 인공지능 기술이 상용화되고 고도화되면서, 미국과 유럽연합을 중심으로 인공지능 윤리이슈에 대한 많은 논의가 이루어지고 있다. 인공지능 기술을 서비스나 플랫폼에 활용하고자 하는 IT 회사들 역시 연구개발과 서비스에 있어 인공지능의 윤리원칙 등 자체적 규범을 발표하는 곳들이 나타나고 있다.

그러나 윤리규범의 구체적 내용을 살펴보면, 다루고 있는 이슈의 범위나 정도에서 적지 않은 차이가 있다. 인공지능을 연구개발하고 활용하는 데 적용되는 일반적이고 기본적인 원칙을 천명한 수준인 경우도 있고, 여기서 더 나아가 주로 인공지능이 이용되고 있거나 가까운 시일 이내에 이용될 것으로 예측되는 분야에서의 주요한 윤리적 이슈까지 함께 다룬 경우도 있다. 또한, 인공지능 규범을 어떻게 정립해야 할지에 대하여 비교적 구체적인 방법론을 언급한 경우도 있다. 이 글에서는 다루어지는 내용의 범위에 따라, ① 기본원칙을 중심으로 하는 윤리규범 유형, ② 기본원칙과 주요 이슈를 함께 (흔히 보고서의 형태로) 다룬 윤리규범 유형, ③ 윤리규범의 정립방안에 대한 구체적 방법론을 함께 언급한 유형으로 나누어

23 Erica Palmerini et al., "Guidelines on Regulating Robotics"(주 15), p. 8 이하; 이원태, "유럽연합(EU)의 로봇법(RoboLaw) 프로젝트", 『KISO Journal Vol. 23』, 2016, 29-32면 참조.

사례를 제시하고자 한다. 매우 다양한 형태의 다양한 문건들이 제시된 바 있는데, 이 중 일부를 선별하여 소개한다. 주요 문건들의 목록은 별도로 정리하여 부록으로 포함하였다.

2. 기본원칙 중심의 윤리규범

기본원칙 중심의 윤리규범은 주로 인공지능 시스템을 설계하고 개발할 때 준수하여야 하는 일반적 사항을 담고 있다. 이런 윤리규범의 예시로, OECD 보고서, 일본 총무성 가이드라인, 아실로마 원칙 및 구글, 마이크로소프트 등 대기업에서 발표한 규범을 소개한다.

1) OECD AIGO, “Draft Recommendation of the Council on Artificial Intelligence” (2019년 5월 최종안 제출 예정)

OECD는 2018년 9월 인공지능 원칙의 방향성을 제시하는 전문가 그룹인 ‘AIGO(AI Expert Group at OECD, 이하 “AIGO”)’를 만들었다. AIGO는 정부, 기업, 노동자, 대중에게 인공지능의 이익을 극대화하고 위험은 최소화할 수 있도록 하는 작업을 하고 있고, 그 일환으로 인공지능 가이드라인의 초안을 마련하였다.²⁴ 2019년 5월경 개최될 예정인 각료이사회(Ministerial Council Meeting)에 현재 초안 상태인 인공지능 가이드라인 권고안이 제출되어 확정될 예정이다.²⁵ 권고안에는, 신뢰할 수 있는 인공지능을 만들어 이를 책임 있게 관리하기 위해 필요한 기본원칙과, 신뢰할 수 있는 인공지능을 만들기 위해 필요한 국가정책, 국제협력의 방향 등에 대한 내용이 담겨있다.

권고안 초안을 보면, 신뢰할 수 있는 인공지능을 만들어 책임 있게 관리하기 위한 원칙으로 특정 집단을 배제하지 않는(inclusive) 포용성, 지속가능한 발전과 삶의 질(well-being), 인간 중심의 가치, 공정성, 투명성 및 설명가능성, 견고함 및 안전 등의 개념이 강조된다. 그리고 인공지능에 대한 국가정책 및 국제협력의 측면에서, 인공지능 연구개발 투자와 디지털 생태계의 조성, 인공지능 혁신이 이루어질 수 있는 정책적 환경을 조성하기 위한 개별국가의 노력 및 국제사회의 협력을 촉구한다.

이와 같은 제언을 바탕으로, AIGO는 이를 실행할 수 있도록 늦어도 2019년 12월 말까지 OECD 디지털 경제정책위원회에 실무 가이드라인을

24 OECD, “OECD creates expert group to foster trust in artificial intelligence”, <http://www.oecd.org/going-digital/ai/oecd-creates-expert-group-to-foster-trust-in-artificial-intelligence.htm>, 2018. 9. 13.자.

25 OECD, “OECD moves forward on developing guidelines for artificial intelligence (AI)”, <http://www.oecd.org/going-digital/ai/oecd-moves-forward-on-developing-guidelines-for-artificial-intelligence.htm>, 2019. 2. 20.자.

제출하고, 다양한 이해관계자를 확보하고 이들이 학제간 대화를 할 수 있도록 인공지능 정책에 대한 정보를 교환하는 포럼을 만들며, 제안사항에 대한 실행을 모니터링하고 이 제안을 채택한 날로부터 5년 이내에 상황을 보고하도록 권고하는 내용을 담았다. 이처럼 OECD 권고안은 권고안의 제시에 그치지 않고, 그 이행을 담보하기 위한 후속조치까지 마련하고 있다. 그리고 이 권고안은 다양한 OECD 회원국은 물론 여러 이해당사자 사이의 협의를 반영하는 것이라는 특징이 있다.

2) 일본 총무성, “Draft AI R&D Guidelines for International Discussions” (2017년 7월)²⁶

일본 총무성은 2017년 7월 인공지능 활용을 통한 이익을 증대하고 위험을 통제하기 위한 목적으로 여러 나라 이해관계자들 사이에 구속력 없는 연성법 형태의 가이드라인을 만드는 데에 도움을 얻기 위하여 “국제적 논의를 위한 인공지능 연구개발 관련 가이드라인 초안(Draft AI R&D Guidelines for International Discussions, 이하 ”가이드라인”)”을 발표하였다. 가이드라인은 인공지능 이용자들의 이익을 보호하고 위험의 확산을 억제하여 인간 중심의 ‘지식네트워크 사회(知連社會)’를 만드는 것을 목표로 제시했다. 가이드라인은 논의의 대상을 현재 사용되고 있는 좁은 인공지능(Narrow AI)으로 상정하고, 자율적 인공지능(autonomous AI) 혹은 일반 인공지능(Artificial General Intelligence, AGI)은 논의 대상에서 제외하였다.

가이드라인은 필요할 경우 계속적으로 가이드라인을 검토하고 유연하게 수정할 것을 전제로 인공지능 연구개발 시 준수해야 할 9가지 원칙을 발표했다. 우선, ① 인공지능 시스템의 상호연결성 및 상호운용성을 보장하도록 하는 ‘협력의 원칙’, 그리고 ② 인공지능 시스템의 입출력에 대한 검증가능성과 의사결정에 대한 설명가능성을 보장하도록 하는 ‘투명성 원칙’이 제시되었다.

다음으로, ③ 사전적으로 인공지능의 행위를 검증하고 유효성을 확인하기 위하여 노력하며, 연구실이나 샌드박스 등 닫힌 공간에서 인공지능 시스템을 실험해 보도록 하는 ‘통제가능성의 원칙’이 있다. 여기에는 사람이나 다른 신뢰 가능한 인공지능 시스템이 인공지능을 모니터링 하거나 경고하는 형태로 감시하고 섯다운, 네트워크 차단, 보수 등의 반대조치를 취하는 것이 효율적인지 여부를 검토하여야 한다는 내용도 포함된다. 같은

26 The Conference toward AI Network Society, “Draft AI R&D Guidelines for International Discussions”, www.soumu.go.jp/main_content/000507517.pdf, 2017. 7. 28.자; 이듬해인 2018년에는 AI의 이용(utilization) 측면에서 비슷한 내용의 원칙이 발표되었다. The Conference toward AI Network Society, “Draft AI Utilization Principles”, www.soumu.go.jp/main_content/000581310.pdf, 2018. 7. 17.자.

맥락에서 ④ 인공지능 시스템이 이용자 또는 제3자의 생명, 신체, 재산을 해하지 않도록 하여야 한다는 ‘안전의 원칙’, 그리고 ⑤ 인공지능 시스템이 의도한 대로 작동하고 승인받지 않은 제3자를 위해 이용되지 않게 하는 안정성과 물리적 공격과 사고에 견딜 수 있도록 하는 견고함을 갖추고 기밀성과 진실성 등을 보장하여야 한다는 ‘보안의 원칙’이 존재한다.

또한 ⑥ 사전에 프라이버시 침해의 위험을 평가하고 프라이버시 영향평가를 진행하도록 하는 ‘프라이버시의 원칙’, ⑦ 인공지능 시스템 연구개발에 있어 인간의 존엄성 및 개인의 자율성을 존중하고 국제인권법 및 국제인도법에 기반한 인간성의 가치를 침해하지 않도록 하는 ‘윤리의 원칙’, ⑧ 이용자가 인공지능 시스템의 도움을 받아 의사결정을 할 때 시의적절하게 인터페이스를 제공하는 등 이용자가 적절하게 이용할 수 있도록 하는 ‘이용자 지원의 원칙’, 그리고 ⑨ 인공지능 시스템을 활용할지 여부를 결정할 수 있도록 이용자에게 기술적 특성에 대해 정보와 설명을 제공하며 다양한 이해관계자들과 대화를 통해 의견을 듣는 방식으로 이해관계자들의 참여를 유도하도록 하는 ‘책임의 원칙’이 있다.

3) Future of Life Institute, “Asilomar AI Principles”(2017년 1월)

민간영역에서도 비정부기구, 연구자, 기술 전문가 등을 포함한 여러 관련당사자들 사이에 인공지능이 제기하는 도전과제에 맞서 책임 있는 인공지능(responsible AI)을 어떻게 구현할 것인지 논의하기 위한 움직임이 계속되었다. 2015년에는 치명적 자율무기(LAWs)가 세계적으로 이슈화되어, 이에 관한 문제제기를 하는 “인공지능에 대한 공개서한(Open Letter on Artificial Intelligence)”에 일론 머스크, 빌 게이츠, 스티븐 호킹 등 유명인사 및 8,000명 이상의 인공지능 연구자들이 서명하기도 했다. 이러한 민간사회의 움직임이 계속되어 미국 캘리포니아 아실로마에서는 2017. 1. 6.부터 3일 동안 ‘Beneficial AI 2017’ 컨퍼런스가 개최되어 인공지능 윤리규범이 논의되었고, 비영리 연구단체인 ‘삶의 미래 연구소(Future of Life Institute)’가 2017. 1. 17. ‘아실로마 원칙(Asilomar Principles)’이라는 이름으로 이를 정리하여 발표하였다.²⁷ 아실로마 원칙은 총 23가지의 사항으로 구성되어 있다. 먼저 유익한 인공지능 연구에 재정지원이 이루어지도록 하여야 하고, 과학-정책 간의 연계를 통한 교류가 있어야 하며, 협력 및 신뢰, 투명성을 중심으로 하는 연구문화를 갖도록 해야 한다는 원칙을 주장한다. 또한 인공지능 시스템의 설계 및 개발 시의 윤리에 대해 안전을 보장하고 투명성을 담보할 것, 인간과 인공지능의 가치가 일치하여야 할 것, 인공지능 시스템에 결정을 위임할지 여부를 사람이 선택할 수 있어야

27 양희태, “인공지능의 위험성에 대한 우려로 제정된 아실로마 인공지능 원칙”, 과학기술정책 제27권 제8호, 2017, 4면.

한다는 ‘통제의 원칙’ 등을 담고 있다. 장기적인 이슈로서 인간의 지성을 초월하는 소위 ‘초지능(super intelligence)’이 등장했을 때를 대비해서 위험을 관리할 수 있도록 하는 원칙에 대해서도 언급되었다.²⁸

4) 민간기업에서 발표한 가이드라인

인공지능 기술을 활용하여 다양한 서비스를 제공하고 있는 몇몇 기업들도 인공지능과 관련된 원칙을 자체적으로 발표하였다. 대표적으로 마이크로소프트는 앞서 언급한 테이 사건이 발생한 이후 몇 개의 가이드라인을 마련하여 발표하였다. 2018년 1월 “The Future Computed” 책자를 발간하면서 인공지능에 관한 원칙을 포함하기도 했고, 개별 상품별로, 2018년 10월에는 대화형 인공지능(챗봇) 개발자를 대상으로 한 10개 가이드라인을 발표한 뒤,²⁹ 2018년 12월에는 안면인식 기술의 개발과 활용에 관한 6가지 원칙을 발표하였다.³⁰ 그 내용으로는 투명성이나 책임의 원칙과 같이 인공지능을 활용함에 있어 기본적으로 준수하여야 하는 원칙이 포함되었다. 특히 대화형 인공지능(챗봇)의 경우 테이 사례를 고려해 인공지능의 능력을 넘어 사람의 판단이 필요한 경우에는 사람이 통제할 수 있도록 하고, 문화적인 규범을 존중하도록 설계하며, 키워드 필터링 방식을 이용해서 인공지능이 이용자의 공격적인 입력에 적절히 대응하도록 하고, 인공지능이 사람을 공정하게 대할 것을 규정했다. 안면인식 인공지능의 경우 인간에 대한 감시의 도구로 악용될 우려가 있다는 점을 고려해서 최소한 정보주체에게 동의를 받고 인공지능을 사용할 것을 권고하였다. 그리고 법집행기관이 안면인식 기술의 사용을 요청한 경우는, 공적인 공간에서 기술을 사용할 수 있는 기준과 한계를 정한 법률이 있고, 법원이 집행기관에 기술의 사용을 허가하였으며, 생명이나 신체에 대한 긴급한 위험이 있는 등 반드시 필요한 경우에 정부의 정당한 이익(legitimate interests)과 개인의 자유(civil liberties) 및 프라이버시의 보호 사이에서의 이익형량을 전제로 하여 제한적으로만 사용할 수 있도록 규정하였다. 원칙을 제시한 수준의 간단한 문건이지만, 이와 같은 문건은 개별 상황에 따라서는 매우 구체적인 현실적인 함의를 가질 수도 있다. 예를 들어, 만일 안면인식 기술의 활용에 있어 동의가 필수적으로 요구된다면, 이 기술의 현실적인 활용가능성은 상당히 제한될 수밖에 없다.

28 Future of Life Institute, “Asilomar AI Principles”, <https://futureoflife.org/ai-principles/?cn-reloaded=1>, 2017. 1. 17.자.

29 Microsoft, “Responsible bots: 10 guidelines for developers of conversational AI”, http://www.microsoft.com/en-us/research/uploads/prod/2018/11/Bot_Guidelines_Nov_2018.pdf, 2018. 11. 4.자.

30 Microsoft, “Six Principles for Developing and Deploying Facial Recognition Technology”, <https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2018/12/MSFT-Principles-on-Facial-Recognition.pdf>, 2018. 12.자.

구글도 2018년 6월 인공지능에 대한 원칙을 발표하였다. 구글의 인공지능이 사회적으로 이익이 되고, 불공정한 편견을 만들어 내거나 강화하지 않아야 한다는 기본원칙 등이 포함된 것이다.³¹ 그리고 이 윤리원칙에 대한 후속 작업으로, 2018년 12월 책임성 있는 인공지능 실무관행(Responsible AI Practices)을 발표하였다.³² 구글은 이러한 원칙을 실제로 거버넌스 구조에 적용하고자, 2019년 3월 외부 인사들로 이루어진 자문위원회(Advanced Technology External Advisory Council)를 구성했다. 다만, 현재 자문위원회는 일부 위원의 탈퇴 및 자격논란 등으로 인해 문제에 봉착한 상황이다.³³

3. 기본원칙 및 세부분야에서의 이슈를 제기하는 윤리규범 유형

인공지능을 설계하고 개발할 때 준수하여야 할 기본원칙을 제시하면서, 동시에 인공지능이 사회에 미치는 영향을 고려할 때 발생할 수 있는 이슈를 찾아 분석하고 해결방안을 검토하는 분석보고서의 성격을 포괄하는 윤리규범 유형도 있다. 이 유형에 해당하는 사례로, IEEE 보고서, 영국 상원 보고서 및 UNGP & IAPP의 보고서에 대해 살펴보기로 한다.

1) IEEE, “Ethically Aligned Design” (2016년 12월, 2017년 12월, 2019년 예정)³⁴

전기전자기술자협회(Institute of Electrical and Electronics Engineers, 이하 “IEEE”)는 세계적인 규모의 기술 전문가 집단인 동시에 중요한 국제표준화기구이기도 하다. 인공지능 및 자율시스템의 윤리적 고려사항에 대한 IEEE 글로벌 이니셔티브에서는 2016년 12월과 2017년 12월 두 차례에 걸쳐서 “Ethically Aligned Design”이라는 이름의 보고서를 발간했고, 이에 대한 의견 수렴을 거쳐 2019년 최종 버전을 발간할 예정에 있다. 이하에서는 2016년 보고서의 내용에 감정 컴퓨팅(affective computing), 섹스 로봇, 시스템 조작, 넛징(nudging) 등의 새로운 영역의

31 Sundar Pichai, “AI at Google: our principles”, <https://www.blog.google/technology/ai/ai-principles/> 2018. 6. 7.자.

32 Google AI, “Responsible AI Practices”, <https://ai.google/education/responsible-ai-practices>, 2018. 12.자.

33 Kent Walker, “An external advisory council to help advance the responsible development of AI”, <https://www.blog.google/technology/ai/external-advisory-council-help-advance-responsible-development-ai/>, 2019. 3. 26.자.

34 Institute of Electrical and Electronics Engineers, “Ethically Aligned Design Version 1 - A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems”, 2016; Institute of Electrical and Electronics Engineers, “Ethically Aligned Design Version 2 - A Vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems”, 2017.

문제를 추가하여 발표한 2017년 보고서에 담긴 내용을 간략히 살펴본다.

이 보고서에서는 인공지능을 설계할 때, 인간이 얻을 수 있는 경제적 이익뿐만 아니라, 인공지능이 사람의 감정적, 정서적 측면에서의 삶의 질에 미치는 영향까지 고려해서 인공지능 시스템을 설계해야 한다는 점을 강조하고 있다. 더 나아가, 인공지능 기술이 더욱 발전한 상황, 즉 자율무기 시스템이 매우 고도화된 상황, 스스로 판단할 수 있는 일반 인공지능(Artificial General Intelligence)이나 초인공지능(Artificial Super Intelligence)이 등장한 상황, 사람의 감정 영역까지 관장할 수 있는 컴퓨팅 기술이 등장한 상황, 가상현실 기술 등의 발전으로 현실과 가상세계의 구분이 되지 않는 상황 등에서 대두될 윤리적인 문제에 대해서도 함께 언급하고 있다.

보고서는 인공지능을 설계, 개발, 실행할 때 적용되어야 할 5가지의 기본원칙을 제시하는데, ① 인공지능이 인권, 자유, 인간의 존엄성 및 문화의 다양성을 존중하는 방향으로 설계되고 운영되어야 한다는 원칙, ② 인공지능이 인간의 삶의 질에 어떤 영향을 미칠지 고려하여 삶의 질을 증진하는 방향으로 개발되고 이용되어야 한다는 원칙, ③ 설계자와 이용자가 인공지능의 결정이나 행위에 대한 책임을 지도록 보장해야 하고 이를 위하여 인공지능에 대한 등록 시스템과 기록보존 시스템을 만들어야 한다는 원칙, ④ 인공지능의 알고리즘과 의사결정 기준을 투명하게 알 수 있도록 하고 투명성에 대한 객관적 평가가 가능하도록 하여야 한다는 원칙, ⑤ 인공지능 시스템이 오용될 가능성이 있다는 것을 의식하고 그 위험을 최소화하도록 하여야 한다는 원칙이 바로 그것이다.

그리고 이 보고서에서는 위 원칙을 준수함에 있어 함께 고려되어야 할 중요한 영역으로 개인정보 분야를 언급하면서 개인정보에 대한 정보주체의 권리를 존중하고 개인정보에 대한 접근을 정보주체가 통제할 수 있어야 한다고 언급하였다. 또한 디지털 세계의 인격(digital personas)이 개인의 인격과 달라질 수 있음을 인지하고, 디지털 세계의 인격도 통제할 수 있도록 해야 하고, 마이데이터 논의에서 볼 수 있는 것과 같이 개인정보의 주체가 자신의 정보를 관리할 수 있는 방향으로 기술발전이 이루어져야 할 것으로 언급되었다. 또한 정보주체의 권리를 보장하려면 인공지능 시스템이 개인의 데이터를 수집하는 경우 데이터가 수집되었다는 사실과 수집의 목적이 무엇인지에 대해 정보주체에게 통지를 해야 하고, 다른 한편 정당하게 주어진 권한을 넘어 데이터에 접근이 이루어질 수도 있는 우회로(backdoor)는 존재하지 않아야 한다고 규정하였다.

그 이외에 이 보고서는 인공지능이 인간에게 피해를 줄 경우의 책임 분배의 문제를 인공지능의 법적 지위와 관련하여 다루고 있다. 책임 문제를 다루는 두 가지 프레임워크로 재산법의 원칙에 따라 보는 방안과 인공지능에

위임되지 않아야 하는 결정을 선별하고 그에 대한 인간의 통제를 보장하는 방안을 제시한다. 보고서는 현재 단계에서 인공지능에 독립적인 법인격(legal personhood)을 부여하는 것에 대하여 부정적인 입장을 취한다. 그 이유로, 인공지능에 독립적인 법인격을 부여하는 경우 인공지능의 의사결정에 대한 사람의 책임을 제한하거나 없앨 수 있고, 인공지능을 설계, 개발, 이용할 때 인공지능의 안전을 보장하려는 인센티브를 낮출 가능성 등이 제시되었다. 그에 따라, 이 보고서는 인공지능의 책임은 현재로서는 재산법 영역의 이슈로 다루어야 한다고 하고, 입법자들과 법집행자들은 인공지능을 이용한다는 것이 손해배상과 같은 법적 책임을 피할 수 있는 수단으로 이용되지 않도록 고려할 필요가 있다고 한다. 그리고 법적 책임 가능성에 따른 배상책임 등 경제적 책임의 이행을 보장하기 위한 보험 등의 제도를 갖출 필요성이 있을 것임을 언급한다.

보고서는 인공지능 시스템의 투명성을 보장하기 위한 구체적인 방안으로서, 당사자, 변호사와 법원으로 하여금 정부 또는 다른 국가기관이 채택하는 알고리즘과 데이터에 대해서 합리적인 수준에서 접근할 수 있도록 허용하는 ‘접근가능성’을 강조한다. 또한, 인공지능 시스템에 내재된 논리와 규칙에 대하여, 관련된 리스크를 평가하고 엄격한 테스트 및 감시가 가능하도록 해야 하며, 인공지능 시스템이 내린 결정 및 해당 결정의 근거를 기록하는 감사추적을 통해 제3자에 의한 검증이 가능하게 해야 한다며 ‘감시 및 검증가능성’을 강조한다. 한편, 대중은 알 권리 차원에서 인공지능 시스템에 대한 윤리적인 결정을 내리거나 지원을 하는 주체가 누구인지를 알 수 있어야 한다고 본다.

보고서는 교육정책 및 의식의 제고가 필요하다는 점도 언급하면서 인공지능 기술의 발전이 사회적으로 어떤 영향을 미치는지에 대하여 대중에게 교육을 할 필요가 있으며, 관련 분야의 전문가를 충분히 양성하여야 한다고 한다. 끝으로 보고서는 인공지능 윤리의 이론적 바탕으로 전통적 윤리를 기반으로 두면서도 인공지능이 인류의 삶의 질에 미칠 영향도 동시에 고려해야 한다고 본다. 구체적으로 자율적인 인공지능 시스템에 어떤 가치가 포함되어야 할지에 관한 ‘가치에 기반을 둔 설계 방법론(value-based design methodologies)’을 도입할 것을 제시한다.

2) 영국 상원, “AI in the UK: ready, willing and able?”
(2018. 4. 16.)³⁵

영국 상원의 인공지능 특별위원회에서는 2018년 4월 영국이 인공지능 관련 산업의 발전을 선도할 수 있는 토대를 만들고 이를 위해 검토하고

35 UK House of Lords, “AI in the UK: ready, willing and able?”, 2018.

준비하여야 할 사항이 어떤 것인지에 대하여 영국 정부에 권고하는 내용을 담아 “AI in the UK: ready, willing and able?”이라는 보고서를 발간했다. 본 보고서에서는 인공지능을 설계하고 개발할 때 준수해야 할 원칙뿐만 아니라, 정부가 인공지능 산업에 어떻게 관여하고 규제해야 할지, 사회 전반적으로 인공지능과 함께 살아가기 위해서는 어떻게 해야 하는지, 인공지능의 위험을 줄이려면 어떻게 해야 하는지 등과 같이 여러 측면에 관한 이슈를 검토하고 있다.

본 보고서는 인공지능이 준수해야 할 중요한 5가지 원칙을 제시한다. ① 인류 공동의 선과 이익을 위해 인공지능을 개발하여야 한다는 원칙, ② 인공지능은 이해가능성(intelligibility)과 공정성(fairness)의 원칙하에 작동하여야 한다는 원칙, ③ 인공지능이 개인, 가족 및 공동체의 개인정보 권리를 줄이거나 프라이버시를 침해하는 방향으로 이용되지 말아야 한다는 원칙, ④ 모든 시민이 인공지능과 함께 정신 감정 경제적 번영을 누릴 수 있도록 교육을 받을 권리를 가진다는 원칙, ⑤ 인공지능에 사람을 해치거나 파괴하거나 속일 수 있도록 자율적인 권한이 부여되어서는 안 된다는 원칙이 그것이다.

또한, 이 보고서에서는 인공지능을 설계하는 과정에서 데이터의 독점과 편견 문제를 비교적 자세히 언급하고 있다. 인공지능 기술이 발전하려면 데이터에 대한 접근과 통제가 필수적인데 큰 IT 기업은 스스로 확보한 데이터의 양도 많고 구매하기도 용이한 반면, 중소기업(SME)의 경우에는 많은 양질의 데이터세트에 접근하기가 쉽지 않을 것이라고 지적하고 있다. 이를 고려하여, 잠재적인 데이터 독점 문제를 적극적으로 검토해야 하는 한편, 정부에서는 공공기관이 가지고 있는 데이터를 인공지능 연구자와 개발자들에게 제공하고, 개인정보를 안전하게 공유할 수 있는 프레임워크를 수립해야 할 필요가 있음을 강조한다. 이 보고서는 또한 편견이 있는 데이터를 학습한 인공지능의 위험에 대하여 언급하면서, 데이터가 다양한 인구집단을 대표하도록 하고 데이터 학습을 통해 사회적 불평등이 고착화되지 않도록 하기 위해 적절한 조치가 필요하다고 한다. 이를 위해 인공지능 알고리즘에 대해서 감사하고 테스트하는 시스템을 만들어야 한다고 주장한다.

한편, 보고서에서는 인공지능의 사용이 활성화 되어 있는 헬스케어 분야에서 중소기업이 국가의 보건의료 시스템 데이터에 접근할 수 있도록 해야 한다는 점을 강조하고, 이를 위하여 데이터를 적절히 익명화해서 환자의 데이터를 공유할 수 있는 프레임워크를 만들 것을 정부에 권고하고 있다.

또한 보고서는 인공지능이 노동시장에 미칠 영향에 대해서도 주목하면서, 영국에서 인공지능의 발전에 따라 노동시장에 미치는 영향에 대해 계속 분석하고 평가하여야 하며 이에 대한 대책을 세워야 한다고

강조한다. 뿐만 아니라, 사회 전반적으로 인공지능이 미칠 영향을 고려해 디지털 및 데이터 문해력(literacy) 교육을 강화하고, 모든 사람이 인공지능이 제공하는 기회에 접근할 수 있도록 하며, 인공지능에 의해 초래될 수 있는 사회적, 지역적 불평등을 해결하는 방안을 만들어야 한다고 주장한다.

인공지능의 사회적 위험과 관련해서 이 보고서가 인공지능의 행위에 대해 어떤 법적 책임을 부여할 것인지를 직접 언급하고 있지는 않다. 다만, 보고서에서는 영국 법사위원회로 하여금 현재의 법률이 인공지능의 법적 책임을 언급하기에 적절한지 여부를 검토하도록 하고, 필요한 경우 정부가 이에 대한 적절한 구제책을 제공하도록 권고한다. 보고서는 인공지능에만 특화된 규제를 도입하는 것은 현 단계에서 적절치 않다고 보고 인공지능이 활용되는 개별 영역 별로 규제자들이 추가적인 규제가 있을 경우의 영향에 대해서 충분히 검토할 것을 권고하고, 다만, 인공지능을 개발하고 이용하는 공적, 사적 기관에서 활용할 수 있는 여러 영역에 교차적으로 적용될 수 있는 윤리적 행위 준칙은 필요할 것으로 보고 있다.

3) UNGP & IAPP, “Building ethics into privacy frameworks for big data and AI” (2018년 10월)³⁶

UNGP(UN Global Pulse)와 IAPP(International Association of Privacy Professionals)은 2017년 5월 “견고한 데이터 프라이버시와 윤리 프로그램의 구축: 이론에서 실제로(Building a strong data privacy and ethics program: from theory to practice)” 포럼을 개최했다. 포럼에서 논의된 내용을 바탕으로 조직에서 데이터 윤리를 실행할 수 있는 방법론과 빅데이터 분석 분야에서 데이터 윤리 및 프라이버시 모범사례를 마련하는데 중점을 두고 2018년 10월 “빅데이터와 인공지능 관련 프라이버시 프레임워크에 윤리를 구축하기(Building ethics into privacy frameworks for big data and AI)”라는 보고서를 출간했다. 이 보고서는 프라이버시 분야의 전문가들을 대표하는 민간의 협회가 UN 기구와 협력하여 인공지능 시스템 개발에 필수적인 데이터 측면을 중심으로 인공지능 윤리를 논의한 결과물을 만들어 냈다는 점에 일차적인 의의가 있다.

이 보고서는 빅데이터와 관련한 주요 윤리적 문제로 사이버보안, 인권 및 프라이버시를 들고, 이러한 문제로 인한 부작용을 최소화하기 위해서는 데이터 수집과 프로젝트 개발 단계에서부터 빅데이터 활용과 관련된 리스크와 프라이버시 및 윤리 문제를 적극적으로 고려해야 한다고 강조한다. 이 보고서에서는 인권 차원에서 프라이버시 권리를 다루는 것이 필요하다고 하면서, 빅데이터나 인공지능의 잘못된 사용(misuse)도 문제일 수 있지만

36 United Nations Global Pulse International Association of Privacy Professionals, “Building ethics into privacy frameworks for big data and AI”, 2018.

이를 사용하지 않음에 따라(missed use) 인권에 미치는 영향도 있을 수 있음을 고려해서 균형 있게 판단해야 한다고 본다. 그리고 이 보고서는 데이터에 대한 분석과 의사결정을 통해 새로운 방식으로 개인에게 영향이 나타날 수 있기 때문에 성문화된 데이터 윤리가 필요하다는 점을 지적하고, 전반적으로 데이터 활용이 가져오는 사회적 가치의 관점에서 윤리 이슈를 볼 필요가 있다고 한다.

그리고 이 보고서에서는 데이터 윤리를 실행하는 방법으로 ‘내부적인’ 프레임워크와 ‘외부적인’ 프레임워크 모두가 필요하다고 본다. 데이터 윤리에 입각해서 운영을 하려면 책임과 투명성이 특히 중요하므로 조직 내부에서 프라이버시와 윤리 프레임워크를 만들고, 데이터 윤리에 대한 리더십을 구축하며, 회사나 조직 차원의 윤리적인 접근에 대하여 지속적으로 평가할 수 있는 데이터 영향평거나 리스크평가 방법을 구축하는 것이 중요하다고 강조한다.

4. 기본원칙과 이슈 제시를 넘어 구체적 논의를 담은 윤리규범 유형

인공지능 설계와 개발에 관한 기본원칙과 이와 관련된 세부적 이슈를 언급하는 것에서부터 한 걸음 더 나아가, 인공지능 관련 사회규범을 어떻게 형성할지에 대해 구체적으로 논의하는 유형의 문건도 있다. 그 대표적 사례로, 유럽의회의 로보틱스에 관한 결의안과 EU 집행위원회 고위급 전문가 그룹의 인공지능 윤리 가이드라인에 관해 살펴본다.

1) 유럽의회, “로보틱스에 관한 민사법적 규율에 관한 위원회 권고 결의안” (2017년 2월)³⁷

인공지능 알고리즘이 발달하면서 주로 로봇의 행위에 대하여 손해배상과 같은 법적 책임과 관련된 민사법적 문제가 대두되자 유럽의회(European Parliament)는 2017. 2. 16. 집행위원회(Commission)에 대한 권고의 내용을 담아 “로보틱스에 관한 민사법적 규율에 관한 권고안”을 결의하였다. 이 결의안은 기존에 논의된 바 있는 로봇윤리 로드맵과 로봇윤리 가이드라인 등을 바탕으로, 유럽의회에서 로봇의 행위에 대한 책임과 법인격 부여와 같은 민사법적 이슈를 구체화하여, 이에 대해 집행위원회로 하여금 후속 조치를 취하도록 권고하는 형태를 띤 것이다. 결의에서는 로봇의 사회적, 의학적, 생명윤리적인 영향을 고려해서 현재 유럽연합의 법적인 프레임워크를 윤리적 원칙에 따라 업데이트하고

37 European Parliament, “European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))”, 2017. 2. 16.자.

보완하여야 할 필요성이 있음을 언급한다. 또한, 투명성의 원칙을 강조해서 로봇(advanced robots)에 블랙박스를 설치해서 해당 로봇이 수행하는 모든 거래(transaction)에 대한 데이터를 기록하고, 의사결정에 이르기까지의 로직을 기록하도록 해야 한다고 본다. 그리고 인공지능 기술을 발전시키기 위해서는 인공지능 시스템 사이의 상호운용성(interoperability)이 핵심일 것으로 보고, EU 집행위원회로 하여금 폐쇄적인 전유 시스템(proprietary system)을 피하고 개방된 기준과 혁신적인 라이선싱 모델에 기초한 개방된 환경을 조성할 것을 요청한다.

유럽의회는 로봇이 자율적으로 의사결정을 할 수 있게 되면, 로봇이 야기한 피해에 대하여 책임 있는 당사자를 찾아 손해에 대한 책임을 지우는 것이 어려울 수 있고, 그 경우 전통적 규율이나 계약법적 책임 방식으로는 문제를 해결하기 어려울 수도 있다는 한계를 인식하였다. 이에 EU 집행위원회로 하여금 이 분야에 대한 사회적 논의를 시작하고 별도 기구를 지정하여 이 문제에 대한 전문성을 확보할 것을 권고하였다. 유럽의회는 로봇의 행위에 의한 법적 책임 문제를 다루기 위한 방안 중 하나로, 로봇에 대한 의무적인 책임보험이 역할을 할 수도 있을 것이라 하고, 또한 보험이 커버되지 않는 범위의 손해에 대비하기 위해 보상 목적의 펀드를 별도로 만들 수도 있을 것임을 언급한다. 그리고 의무 책임보험제도에 대한 전제로, 로봇을 관리기관에 등록하도록 하는 ‘스마트로봇 등록제도’를 만들어 로봇의 등록번호를 통하여 로봇과 보험 또는 펀드 간의 연결 관계를 확인할 수 있는 시스템 구축을 고려하도록 했다. 장기적으로는 로봇과 관련하여 전자인, 전자적 인격 등의 특별한 법적인 지위를 만들 것인지에 대해서 별도로 고려하도록 권고하고 있다.

결의안 첨부(Annex)에서는 로보틱스 헌장(Charter on Robotics)을 제시하여 집행위원회가 이를 고려하도록 하였다. 로보틱스 헌장은, 로봇 엔지니어를 위한 윤리적 행위 준칙(Code of Ethical Conduct for Robotics Engineers)과 연구 윤리위원회 준칙(Code for Research Ethics Committees)를 포함하고 있다. 로봇 엔지니어들에게는, 기본권을 존중하고 책임 있는 태도를 가지며 프라이버시를 존중하고 연구에 리스크가 따를 경우 리스크에 대한 평가 및 관리 프로토콜을 마련할 것을 준칙으로 언급한다. 연구 윤리위원회와 관련해서는, 독립성(independence), 전문성(competence), 투명성 및 책임성(transparency and accountability)을 대전제로 하여, 다양한 배경(multidisciplinary)을 가진 위원으로 구성될 필요가 있음을 언급하고 있다.

2) EU 집행위원회 인공지능 고위급 전문가 그룹(HLEG), “Ethics Guidelines for Trustworthy AI” (2019년 4월)³⁸

유럽연합 집행위원회(European Commission)는 2018년 4월 25일 인공지능 고위급 전문가 그룹(High-Level Expert Group, 이하 “HLEG”)을 발족했다. HLEG은 2018. 12. 18. “신뢰가능한 인공지능을 위한 윤리 가이드라인(Ethics Guidelines for Trustworthy AI)” 초안을 발표하였고, 이후 2019. 4. 8. 최종본을 발표했다. 가이드라인에서는 신뢰가능한 인공지능 시스템은 합법적이고(lawful), 윤리적이며(ethical), 견고한(robust) 인공지능 시스템을 의미한다고 정의하고, 법률적인 측면은 제외하고 주로 윤리적이고 견고한 시스템에 초점을 맞추어 이를 위한 원칙과 이 원칙을 실행하기 위한 요건 및 준수 여부를 평가할 수 있는 기준을 구체화하여 실시하고 있다.

이 가이드라인에서는 신뢰가능한 인공지능을 위하여 기본권에 대한 존중이 핵심요소라고 보고 이를 위한 원칙을 제시한다. 구체적 내용으로는, ① 인공지능 시스템은 사람을 대리하는 에이전트일 뿐이며, 인공지능 시스템에 종속되거나 강압 받지 않을 자유가 보장되어야 하며, 직접적 또는 간접적으로 인공지능의 의사결정의 대상이 될지 여부는 인간이 결정해야 한다는 ‘인간의 자율성에 대한 존중의 원칙(respect for human autonomy)’, ② 인공지능이 어린이, 소수자 등의 취약계층을 포함하여 인간에게 해를 끼치지 않아야 한다는 ‘해악방지의 원칙(prevention of harm)’, ③ 인공지능이 실질적, 절차적 측면 모두에서 공정해야 한다는 ‘공정성의 원칙(fairness)’, ④ 인공지능의 기술적인 투명성과 사업모델의 투명성을 기반으로 하여 설명이 가능해야 한다는 ‘설명가능성의 원칙(explicability)’을 들고 있다.

다음으로 이런 원칙을 실현하기 위한 요구사항으로, ① 자율성에 대한 존중의 원칙에 따라 인공지능 시스템이 인간 이용자의 에이전트로서 인간의 기본권을 증진시키고 인간의 감독을 허용해야 하고, ② 해악방지의 원칙에 따라 인공지능 시스템에 의한 원하지 않은 피해를 예방하기 위해 기술적 견고함과 안전성을 갖추어야 하며, ③ 인공지능 시스템이 프라이버시에 피해를 가져오지 않도록 하기 위해 데이터 처리에 대한 접근 프로토콜과 프라이버시를 보호하는 방향의 데이터 거버넌스(data governance) 구조를 만들어야 하고, ④ 설명가능성의 원칙에 따라 인공지능 시스템이 내린 결정에 관련된 데이터세트와 절차가 추적 가능하고 인공지능 시스템의 기술적 측면과 그에 관한 사람의 결정이 설명 가능하도록 해서 투명성을

38 European Commission High-Level Expert Group, “Ethics Guidelines for Trustworthy AI”, <https://ec.europa.eu/futurium/en/ai-alliance-consultation>, 2019. 4. 8.자.

보장해야 하며, ⑤ 공정성의 원칙에 따라 인공지능 시스템이 다양성을 보장하고 성별, 나이, 능력 등을 불문하고 누구나 접근 가능하도록 보편적 설계(universal design)를 통해서 차별이 없어야 하고, ⑥ 공정성의 원칙과 해악방지의 원칙에 따라서 사회적, 환경적 요소도 고려요소에 포함시켜 지속가능하고 생태계에 대해서도 책임성 있는 시스템을 만들어야 하며, ⑦ 공정성의 원칙에 따라 인공지능 시스템과 시스템이 만들어낸 결과를 감사할 수 있도록 해야 하고 부정적인 영향을 최소화하고 부정적인 영향이 발생할 경우 이를 보고하며 잘못된 점을 바로잡을 수 있도록 해서 책임 보장 메커니즘을 만들 것을 들고 있다. 가이드라인은, 신뢰할 수 있는 인공지능을 구현하기 위해서는 인공지능의 설계와 시스템 구조와 관련된 기술적인 방법뿐만 아니라 규제나 행동강령과 같은 비기술적인 방법 모두 중요하고 함께 고려되어야 한다고 본다.

가이드라인에서는 인공지능을 개발하고, 배치하며, 이용하는 과정에서 신뢰할 수 있는 인공지능에 해당하는지 여부를 평가할 수 있는 항목들을 만들어 예시로 제시하는 한편, 인공지능 시스템의 이해관계자들이 자신의 조직에서 가이드라인에 제시된 평가요소를 조직의 지배구조 메커니즘에 포함시켜 실천하도록 권고하였다. 특히 회사와 같은 조직에서 경영자, 컴플라이언스 및 법무부서, 상품 및 서비스 개발부서 등이 인공지능 시스템의 개발과 관련해서 어떤 역할을 해야 하는지에 대해서도 언급하였다. 다만, 가이드라인은 평가 리스트를 모두 준수했다고 하여 관련 법률을 모두 준수한 것이라고는 볼 수 없고 리스트 내용도 모든 관련된 사항을 망라하고 있지는 않다는 것을 밝히고 있다. 하지만 이 리스트가 인공지능 시스템을 개발하거나 도입하려는 조직에서 중요한 내부적인 참조 평가기준으로 활용될 수는 있을 것이다. EU 집행위원회에서는 이러한 평가목록이 실무에서 활용될 수 있도록 하기 위해서 2019년 6월부터 모든 이해관계자들이 평가목록을 테스트 해보고 피드백을 제공할 수 있도록 하며, AI HLEG를 통하여 공적인 영역과 사적인 영역의 이해관계자들이 모두 모여 가이드라인이 실제로 어떻게 적용될 수 있을지에 대한 심층적 검토를 하도록 할 것이라고 밝혔다.

4. 해외 윤리규범이 국내에 주는 시사점

1. 해외 윤리규범의 특징과 경향성

지금까지 살펴본 다양한 해외 논의의 대략적 특징과 경향성은 다음과 같이 요약될 수 있다. 첫째, 윤리규범의 형성 주체에 따른 차이가 존재한다. 사적 주체가 내놓은 규범의 형태가 일반론적이거나 원론적 내용이 담긴 경우가 많은 반면, 공적 주체를 통해 도출된 규범의 경우는 상대적으로

다루는 분야가 넓은 편이고 규범의 구체성도 높은 경우가 많다. 특히, 유럽에서 발표된 규범에 그런 경향이 더 명확하게 나타나는 편이다. 둘째, 다양한 성격과 이해관계를 가진 주체가 공동으로 내놓은 규범일수록 상대적으로 원론적인 내용 위주이다. 대표적으로 OECD의 경우에는 매우 다양한 이해관계를 가진 주체들 사이의 협의와 협상을 통한 합의를 반영하여, 구체적인 내용이 많이 담겨있지 않다. 수많은 사적 주체가 공동으로 내놓은 아실로마 원칙 또한 참여자들이 관심을 두는 영역이 다르기 때문인지 23가지로 항목의 숫자가 적지 않지만 내용상으로는 단문 형태의 일반적 명제의 진술에 그치고 있다. 다만, 원론적인 내용 위주라고 해서 반드시 규범력이나 강제력이 적은 것은 아니다. 특히 OECD 규범은 후속조치와 정기적인 모니터링에 관한 내용을 담고 있어, 이러한 방식을 통해 실질적인 규범력이 확보될 수 있는 장치를 마련하고 있다.

셋째, 지역적인 문화적 차이나 규제를 바라보는 시각차이도 어느 정도 엿볼 수 있다. 유럽과 미국을 비교하면, 유럽연합의 경우는 시범사업(pilot test) 단계에 돌입할 정도로 윤리규범의 내용이 구체화된 상태이다. 반면, 미국의 경우 대체로 개별 기업들을 통해 규범이 제시되고, 그와 별개로 학계를 통해 활발한 논의가 이루어지고 있는 형편이다. 이런 차이에는 다양한 원인이 있을 수 있지만, 최근의 법제도 동향을 보면 유럽에서는 GDPR과 같은 일반적인 법규범을 마련하여 통일적으로 시행하는 것이 자연스럽게 받아들여지는 한편, 미국에서는 적어도 지금까지는 개별 영역별 접근 그리고 자율규제의 역할을 상대적으로 강조하는 면이 있다. 따라서 이와 같은 경향성을 국내 인공지능 규범 논의에 참조할 때에는, 우리나라가 처해있는 현실적 여건을 종합적으로 고려한 대안을 마련하여야 할 것이다.

2. 우리나라 윤리규범 논의의 현황과 과제

그렇다면 앞서 살펴본 해외의 논의와 비교하여, 우리나라의 인공지능 윤리규범 논의는 어떤 모습으로 변화해왔을까? 먼저 앞서 언급했던 2007년에 발표된 로봇윤리헌장 초안을 살펴보자. 이는 지능형 로봇 개발 및 보급 촉진법 제18조에 법적 근거를 두고, 2016년에는 개선안까지 마련되었지만 아직까지 정식으로 시행되지는 못했다. 그러나 알파고가 사회적으로 큰 파장을 낳은 이후 인공지능 규범 마련에 대한 공감대가 형성되면서 여러 논의가 이어지고 있다. 박영선 의원은 2017. 7. 19. 로봇의 윤리와 책임에 관한 내용을 담은 ‘로봇기본법안’을 대표로 발의하였고, 그해 연말 정보문화포럼과 한국정보화진흥원이 공동으로 공공성(Publicness), 책무성(Accountability), 통제성(Controllability), 투명성(Transparency)이라는 소위 ‘PACT 원칙’을 개발자, 공급자,

이용자에 대해 적용한 ‘지능정보사회 윤리 가이드라인’이 마련되었다. 그리고 카카오는 2018. 1. 31.에 ‘카카오 알고리즘 윤리 헌장’을 마련하여 발표한 바 있다. 다만, 카카오 윤리 헌장은, ① 카카오 알고리즘의 기본 원칙, ② 차별에 대한 경계, ③ 학습 데이터 운영, ④ 알고리즘의 독립성, ⑤ 알고리즘에 대한 설명이라는 다섯 가지 원칙을 원론적으로 설명한 것이어서, 내용이 구체적이지 않다. 같은 해 6월에 발표된 ‘지능정보사회 윤리헌장’에서도 단지 여섯 가지의 선언적 명제를 짚막하게 언급한다.³⁹ 한편, 변재일 의원이 2018. 2. 14. 대표로 발의한 ‘국가정보화 기본법 전부개정법률안’ 제62조는 “공공성 책무성 통제성 투명성 등의 윤리원칙을 담은 지능정보사회윤리”라는 표현을 명시적으로 언급하고 있다.

‘개발자, 공급자, 이용자의 공공성 책무성 통제성 투명성’이라는 명제로 대표되는 국내의 논의는 규범 논의의 주체와 목적을 인간으로 상정한다는 점, 책무성과 투명성이라는 주체성과 책임성을 동시에 고려한 원칙을 수용하고 있다는 점에서 어느 정도 해외의 논의와 일맥상통한 내용을 담고 있다고 평가할 수 있다. 향후에 논의가 계속된다면, 이제부터는 규범의 구체화와 강제력 부여를 비롯한 좀 더 현실적인 사항들에 대한 논의가 이루어질 것으로 예상해 볼 수 있다. 국내에서의 향후 논의는 국내의 실정에 부합하는 동시에 국제사회의 논의에 발맞추어 진행되어야 한다. IEEE나 IAPP의 사례처럼 전문가들이 참여하여 만든 실질적인 국제표준은 실무자나 개발자 등에게는 사실상의 행동지침으로 작용할 수 있다. 다른 한편 OECD나 유럽연합 등을 통한 논의는, 참가 당사국을 통해 후속 논의와 후속 조치를 하는 장치를 마련하고 있어서, 이를 통해 현실적인 집행력이 확보될 것이고, 이는 정책적으로도 적지 않은 파급력을 미칠 것이다.

그런데 인공지능 윤리규범에 관한 논의가 경우에 따라 일견 원론적이고 교과서적 차원의 논의에 그치는 것으로 비쳐질 수도 있지만, 매우 간단한 추상적인 내용만을 담고 있는 규범의 경우에도 그 이면에는 이해관계자 사이에 규범논의를 이끌어가려는 치열한 이익대립이 숨겨져 있는 것이 일반적임을 간과해서는 곤란하다. 윤리규범의 내용이 매우 간략한 경우에, 개별 규범에 담긴 내용이 어떤 것인지 파악하여 분석하는 것이 중요한 것은 물론 규범에 담기지 않은 것은 어떤 것인지에 대하여 파악하고 그 함의를 분석하는 것이 더 중요할 수도 있다. 또는 추상적인 원칙으로부터 분석을 통해 ‘행간’을 읽어낼 필요가 있을 수도 있다. 그 이외에도, 선언적인 내용이, 선언 그 자체에 그치는 경우와 실제로 집행력을 확보하게 되는 경우 사이의 구분도 중요하다. 그러한 면밀한 분석과 판단이 이루어지지 않으면 우리나라의 실정에 부합하지 않거나 부작용을 낳을 수 있는 규범임에도

39 정보문화포럼 한국정보화진흥원, “지능정보사회 윤리 가이드라인”, 2017, 23면.

‘사실상의 국제표준(de facto global standard)’이라는 이유로 억지로 수용하는 결과를 낳을 수도 있다. 국제사회의 ‘규범전쟁(norm war)’ 속에서 주체적으로 목소리를 내고 나아가 논의에서 선도자로서의 지위를 점유하기 위해서는 그와 같은 깊이 있는 분석과 관련된 논의가 지속될 필요가 있다.

5. 윤리적 인공지능의 미래: 결론을 대신하여

이하에서는 결론을 대신하여 윤리적 인공지능의 실현과 정착을 위한 몇 가지 제언을 정리하도록 한다. 지금까지의 논의를 통해 공감대를 형성한 가장 커다란 원칙은, 인류가 지향해야 할 목표는 인공지능이 인간의 존엄성과 권리의 실현을 위한 방향으로 활용되어야 한다는 점이다. 그러한 대원칙에 부합하지 않는 관념들, 특히 아래에서 언급하는 이분법적 관념들은 인공지능 시대와 부합하도록 변화를 모색할 필요가 있다.

우선, 법규범과 윤리를 비롯한 여타 사회규범의 유형을 엄밀히 구별하는 태도를 들 수 있다. 전통적으로 법규범은 강제력을 주요한 근거로 삼아 여타 사회규범 유형과 뚜렷이 구분되어 왔지만, 인공지능의 맥락에서는 그러한 구분이 명확하지 않은 경우도 많고 또한 그러한 구분으로부터의 실익이 크지 않은 경우도 많다. 둘째로 기술과 규범을 엄밀하게 분리하여 파악하려는 태도가 존재한다. 인공지능 윤리의 맥락에서는, 기술과 규범이 서로 보완적인 역할을 할 필요가 있다. 셋째, 법규범 내부에서 해악의 위험성에 대하여 사전적으로 공적, 행정적 규제를 강조하는 영역과, 발생한 해악에 대하여 주로 사후적이고 민사적인 사법판단을 중시하는 영역을 엄밀히 구분하려는 태도가 있다. 이는, 사적 영역은 ‘사적자치의 원칙’이 강조되어 구체적 해악이 초래되지 않는 이상 외부에서의 개입이 정당화되기 어렵고, 소위 ‘위험사회’에 대한 사전규제는 공공기관이 담당해야 할 몫이라는 입장과도 이어지는 것이다. 이 또한 인공지능의 맥락에서는 현실성이 약한 구분이다.

전반적으로, 오늘날 진행되고 있는 국제사회의 인공지능 규범논의는 이와 같은 전통적 이분법의 관념과는 크게 다른 것이다. 무엇보다도 인공지능과 같이 급속도로 발전하는 신기술은 전형적인 변화 패턴을 예상하기 어렵다. 예컨대, 오늘날 인공지능 방법론의 전형이라고 할 만한 것은 딥러닝(Deep Learning) 기술일 텐데, 어떠한 인공지능 시스템도 학습방식을 오직 딥러닝에만 국한하지는 않는다. 딥러닝과 대비되는 규칙기반 전문가 시스템도 여전히 사용되고 있고, 여러 유형의 인공지능 모델이 조합을 이루어 이용되는 경우를 빈번하게 볼 수 있다. 또한 딥러닝 알고리즘도 수많은 변이가 발생 중이고, 딥러닝 패러다임 자체가 전환될 가능성도 배제할 수 없다. 그리고 실제 상용화 과정에서는 흔히 기존의 제품과 서비스에 인공지능 요소를 약간씩 포함시키고, 이를 점차 확대하고

고도화하는 과정을 거치게 된다. 이때, 개별 제품이나 서비스에서 인공지능 요소를 별도로 떼어내어 법규범적 검토를 하는 것은 현실적이지 않다.

인공지능 기술의 발전과 변화의 속도가 빠르고 변화의 방향에 대해서도 예측이 어렵다는 점은 입법자가 인공지능의 단일한 본성을 상정하고 법적 절차를 통하여 사전적 규제 체계를 마련한다거나 사후적 민 형사 책임 위주의 법체계를 입법하는 ‘경성법(hard law)’ 위주의 전통적 접근을 어렵게 한다. 인공지능은 국제사회의 이해관계자들 사이의 첨예한 대립이 빈번하다는 점에서, 구속력과 강제력을 지나치게 강조하는 경우 도리어 무규범(anomie) 상태로 흐르거나 서로 상충되는 법체계가 복잡하게 혼재하는 상태가 초래될 가능성도 있다. 반면, 실정법적인 구속력과 강제력은 없지만, 행위규범의 일종으로 구성원에게 사실상의 영향력을 미치는 ‘연성법(soft law)’을 통해 경성법 체계와 조화를 시도하는 방안을 대안으로 생각해볼 수 있다. 앞서 보았듯 여러 사적, 공적 주체가 가이드라인, 원칙, 행동강령과 같은 다양한 이름으로 만들어내고 있는 연성법은 미래사회의 방향성을 제시하고, 현실과 이상 간의 간극을 메우는 등 보다 유연한 대처가 가능한 장점을 가지기 때문이다.⁴⁰ 연성법은 기술규제, 자율규제와도 공존하는 대목이 있어 인공지능 거버넌스 담론에서 많은 지지를 획득하고 있다. ‘책임’에 관한 논의에 있어서도, 법적 책임을 의미하는 ‘liability’가 아니라 그 보다 포괄적인 의미를 가지는 ‘accountability’ 개념이 주로 언급되는 것에 유의할 필요가 있다.

다만, 연성법 체제를 좀 더 본격적으로 도입하기 위해서는 이를 위한 충분한 준비가 필요하다. 가장 기본적으로 우리나라의 법체계 구조와 관행상, 경성법이 연성법에 비해 높은 예측가능성을 가진다는 점이 지적될 수 있다. 연성법이 경성법의 보완재라기보다는 예측가능성 낮은 ‘추가적 규제’로 무분별하게 남용될 경우, 명확한 경성법 일원론적 규제방식보다도 못한 결과를 낳을 수도 있다. 또한 행정, 사법, 산업 영역에 대한 사회구성원들의 신뢰(trust) 정도도 중요한 변수가 될 수 있다. 신뢰수준이 낮은 영역에 연성법을 통한 독자적 또는 자율적 규율 권한을 과도하게 부여하는 경우 막대한 사회비용이 유발될 수도 있을 것이기 때문이다. 다른 한편, 연성법이 그저 대원칙에 대한 선언으로만 비쳐지고 현실적이고 실질적인 구속력이 확보되지 못한다면, 많은 경우에 이는 불필요한 낭비만 초래할 수도 있다. 그런 면에서, 위에서 살펴본 다양한 규범들도, 실효성이 있는 규범과 실효성의 확보가 어려운 규범으로 나누어 살펴볼 수 있다. 앞으로의 인공지능 윤리담론은 이런 다양한 측면을 함께 고려한 풍부하고 깊이있는 논의가 되어야 할 것이다.

40 최난실현, “연성규범(Soft Law)의 기능과 법적 효력 : EU 경쟁법상의 논의를 중심으로”, 『법학연구 제16집 제2호』, 2013, 96-98면 참조.

참고 문헌

- 김효은, 『인공지능과 윤리』, 커뮤니케이션북스, 2019.
- 라파엘 카푸로 미카엘 나겐보르크, 변순용 송선영 역, 『로봇윤리 - 로봇의 윤리적 문제들 -』, 어문학사, 2013.
- 웬델 윌러치 콜린 알렌, 노태복 역, 『왜 로봇의 도덕인가』, 메디치미디어, 2014.
- 이원태 문정욱 이시직 심우민 강일신, 『지능정보사회의 규범체계 정립을 위한 법 제도 연구』, 정보통신정책연구원, 2016.
- 캐시 오닐, 김정혜 역, 『대량살상 수확무기』, 흐름출판, 2017.
- 프랭크 파스칼레, 이시은 역, 『블랙박스 사회』, 안티고네, 2016.
- 한희원, 『인공지능(AI) 법과 공존윤리』, 박영사, 2018.
- 고인석, “아시모프의 로봇 3법칙 다시 보기: 윤리적인 로봇 만들기”, 『철학연구 제93집』, 2012.
- 고학수, “인공지능 알고리즘과 시장”, 『데이터 이코노미』, 한스미디어, 2017.
- 고학수 정해빈 박도현, “인공지능과 차별”, 『저스티스 통권 제171호』, 2019.
- 김건우, “로봇윤리 vs. 로봇법학: 따로 또 같이”, 『법철학연구 제20권 제2호』, 2017.
- 김중호, “인공지능 시대의 윤리와 법적 과제”, 『과학기술법연구 제24권 제3호』, 2018.
- 신상규, “인공지능 시대의 윤리학”, 『지식의 지평 제21권』, 2016.
- 양천수, “현대 지능정보사회와 인격성의 확장”, 『동북아법연구 제12권 제1호』, 2018.
- 양희태, “인공지능의 위험성에 대한 우려로 제정된 아실로마 인공지능 원칙”, 『과학기술정책 제27권 제8호』, 2017.
- 오요한 홍성욱, “인공지능 알고리즘은 사람을 차별하는가?”, 『과학기술학연구 제18권 제3호』, 2018.
- 윤지영, “인공지능 기술 관련 법적 제도적 논의 현황”, 『법과학을 적용한 형사사법의 선진화 방안(VIII) : 인공지능 기술』, 한국형사정책연구원, 2017.
- 이원태, “유럽연합(EU)의 로봇법(RoboLaw) 프로젝트”, 『KISO Journal Vol. 23』, 2016.
- 이원태, “4차 산업혁명과 지능정보사회의 규범 재정립”, 『KISDI Premium Report 17-10』, 2017.
- 정보문화포럼 한국정보화진흥원, “지능정보사회 윤리 가이드라인”, 2017
- 정재연, “지능정보사회에서 지능로봇의 윤리화 과제와 전망 - 근대적 윤리담론에 대한 대안적 접근을 중심으로 -”, 『동북아법연구 제12권 제1호』, 2018.
- 최난설현, “연성규범(Soft Law)의 기능과 법적 효력 : EU 경쟁법상의 논의를 중심으로”, 『법학연구 제16집 제2호』, 2013.
- 카카오 정책지원팀, “KAKAO AI REPORT”, Vol. 1, 2017.
- 한희원, “인공지능(AI) 치명적자율무기(LAWs)의 법적 윤리적 쟁점에 대한 기초연구”, 『중앙법학 제20집 제1호』, 2018.
- Alex Campolo et al., “AI Now 2017 Report”, 2017.
- Erica Palmerini et al., “Guidelines on Regulating Robotics”, 2014.
- European Commission High-Level Expert Group, “Ethics Guidelines for Trustworthy AI”, 2019.
- Gianmarco Veruggio, “EURON Roboethics Roadmap(ver. 1.1)”, 2006.
- Gianmarco Veruggio, “EURON Roboethics Roadmap(ver. 1.2)”, 2007.
- Google, “Perspectives on Issues in AI Governance”, 2019.
- Helen Nissenbaum, “Accountability in a Computerized Society”, Science and Engineering Ethics Vol. 2, Issue. 1, 1996.

Institute of Electrical and Electronics Engineers, "Ethically Aligned Design Version 1 - A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems", 2016.

Institute of Electrical and Electronics Engineers, "Ethically Aligned Design Version 2 - A Vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems", 2017.

Jean-François Bonnefon·Azim Shariff·Iyad Rahwan, "The social dilemma of autonomous vehicles", Science Vol. 354, Issue. 6293, 2016.

Karl M. Manheim, Lyric Kaplan, "Artificial Intelligence: Risks to Privacy and Democracy", Forthcoming, Yale Journal of Law and Technology, 2018.

Kate Crawford·Meredith Whittaker, "The AI Now Report", 2016.

Meredith Whittaker et al., "AI Now Report 2018", 2018.

Merel Noorman, "Computing and Moral Responsibility", Stanford Encyclopedia of Philosophy, 2018.

Microsoft, "Responsible bots: 10 guidelines for developers of conversational AI", 2018.

Microsoft, "Six Principles for Developing and Deploying Facial Recognition Technology", 2018.

Peter M. Asaro, "What Should We Want From a Robot Ethic?", International Review of Information Ethics Vol. 6, 2006.

Peter Stone et al., "Artificial intelligence and life in 2030", One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, 2016.

Robin R. Murphy · David D. Woods, "Beyond Asimov: The Three Laws of Responsible Robotics", IEEE Intelligent Systems Vol. 24, Issue 4, 2009.

Robyn Caplan et al., "Algorithmic Accountability: A Primer", Data&Society, 2018.

The Conference toward AI Network Society, "Draft AI R&D Guidelines for International Discussions", 2017.

The Conference toward AI Network Society, "Draft AI Utilization Principles", 2018.

UK House of Lords, "AI in the UK: ready, willing and able?", 2018.

United Nations Global Pulse · International Association of Privacy Professionals, "Building ethics into privacy frameworks for big data and AI", 2018.

U.S. Executive Office of the President, "Preparing for the Future of Artificial Intelligence", 2016.

U.S. Executive Office of the President, "Artificial Intelligence, Automation, and the Economy", 2016.

부록

인공지능과 관련한 주요 윤리 규범 목록

(발간일 기준 정렬)

	발간주체	윤리 규범	발간일
1	유럽로봇연구네트워크(EURON)	Roboethics Roadmap Release 1.1 Roboethics Roadmap Release 1.2	2006.7 2007.1
2	산업통상자원부	로봇윤리헌장 초안	2007
3	유럽 로봇법 프로젝트(RoboLaw)	RoboLaw Guidelines on Regulating Robotics	2014.9
4	IEEE (Institute of Electrical and Electronics Engineers)	Ethically Aligned Design	2016.12 발간 (Version I), 2017.12 발간 (Version II), 2019년 최종본 발간 예정
5	Future of Life Institute	Asilomar AI Principles	2017.1
6	ACM (Association for Computing Machinery)	Principles for Algorithmic Transparency and Accountability	2017.1
7	유럽의회 법사위원회(European Parliament Committee on Legal Affairs)	European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics	2017.2
8	일본 인공지능학회	윤리지침	2017.2
9	독일 교통부 윤리위원회 (German Federal Ministry of Transport and Digital Transformation, Ethics Commission)	German Ethics Code for Automated and Connected Driving	2017.6
10	Sage	Ethics of Code: Developing AI for Business with Five Core Principles	2017.6
11	영국 상원(House of Lords)	5 Overarching Principles for an AI	2017.6
12	일본 총무성	Draft AI R&D Guidelines for International Discussions	2017.7
13	UNESCO COMEST (UNESCO 세계과학기술윤리위원회)	Report of COMEST on Robotics Ethics	2017.9
14	정보문화포럼 & 한국정보화진흥원	지능정보사회 윤리 가이드라인	2017.11

	발간주체	윤리 규범	발간일
15	카카오	카카오 알고리즘 윤리현장	2018.1
16	영국 상원(House of Lords)	AI in the UK: ready, willing and able	2018.4
17	Google	AI at Google: Our Principles	2018.6
18	정보문화포럼 & 한국정보화진흥원	지능정보사회 윤리현장	2018.6
19	ACM (Association for Computing Machinery)	Code of Ethics	2018.7
20	Microsoft	Responsible bots: 10 guidelines for developers of conversational AI	2018.10
21	Sony	Sony Group AI Ethics Guidelines	2018.10
22	UNGP & IAPP (UN Global Pulse & International Association of Privacy Professionals)	Building ethics into privacy framework for big data and AI	2018.10
23	Public Voice	Universal Guidelines for Artificial Intelligence	2018.10
24	Microsoft	Six Principles for Developing and Deploying Facial Recognition Technology	2018.12
25	Google	Responsible AI Practices	2018.12
26	UNESCO COMEST (UNESCO 세계과학기술윤리위원회)	Preliminary Study on the Technical and Legal Aspects Relating to the Desirability of a Standard-Setting Instrument on the Ethics of Artificial Intelligence	2019.3
27	EU 집행위원회 인공지능 고위급 전문가 그룹(High-Level Expert Group)	Ethics Guidelines for Trustworthy AI	2019.4
28	OECD AI Expert Group	Draft Recommendations of the Council on Artificial Intelligence	2019.5 예정

서울대학교
인공지능정책
이니셔티브 안내

서울대학교 인공지능정책 이니셔티브는 인공지능과 관련된 다양한 사회경제적, 법적, 정책적 이슈들을 연구하고 논의하기 위해 시작된 서울대학교 법과경제연구센터의 프로그램입니다. ‘소셜랩(Social Lab)’ 개념을 지향하여, 여러 배경과 관심을 가진 분들 사이의 협업과 지속적인 대화를 추구합니다. 서울대학교 법학전문대학원의 고학수 교수와 임용 교수가 함께 이끌고 있습니다.

1. 발간물 안내

서울대학교 인공지능정책 이니셔티브의 주요 발간물은 이슈페이퍼와 워킹페이퍼가 있고, 비정기적으로 발간되는 단행본 및 학술행사 자료집 등이 있습니다. 이슈페이퍼와 워킹페이퍼 등의 자료들은 홈페이지를 통해 다운로드 받으실 수 있습니다.

2. 행사 안내

서울대학교 인공지능정책 이니셔티브의 주요 행사는 이슈페이퍼를 발표하고 논의하는 행사(상반기 및 하반기 각 1회) 그리고 국내외 연구자들을 초빙하여 진행하는 대규모 국제학술대회(연 1회) 등이 있습니다. 그 이외에 비정기적으로 진행하는 행사들도 있습니다.

3. 이슈페이퍼 2019

이번 이슈페이퍼는 서울대학교 인공지능정책 이니셔티브의 첫 이슈페이퍼로, 2019.5.16. D2 Startup Factory에서 열린 행사에 맞춰 준비되었습니다.

Seoul National University
Center for Law & Economics
AI Policy Initiative

이슈페이퍼 2019
“인공지능과 미래사회”