

서울대학교 인공지능 정책 이니셔티브
<http://ai.re.kr/>

2018 국제학술대회 보고서

인공지능의 시대: 기술발전에 따른 책임과 규제
ARTIFICIAL INTELLIGENCE TODAY:
GOVERNANCE AND ACCOUNTABILITY

날짜 2018. 8. 24.(금)

장소 한국프레스센터 국제회의장

주최 SNU 서울대학교 아시아태평양법 연구소 서울대학교 법과경제연구센터
Seoul National University Asia-Pacific Law Institute SNU Center for Law & Economics

후원 **NAVER** **Google** **Microsoft** **facebook** **kakao**

서울대학교 인공지능 정책 이니셔티브
<http://ai.re.kr/>

2018 국제학술대회 보고서

인공지능의 시대: 기술발전에 따른 책임과 규제
**ARTIFICIAL INTELLIGENCE TODAY:
GOVERNANCE AND ACCOUNTABILITY**

서울대학교 인공지능 정책 이니셔티브

(<http://ai.re.kr/>)

2018 국제학술대회 보고서

공동디렉터 : 고학수 (서울대학교) & 임 용 (서울대학교)

보고서 작성 : 박도현 (서울대학교 박사과정)

국제학술대회

인공지능의 시대: 기술 발전에 따른 책임과 규제

(Artificial Intelligence Today: Governance and Accountability)

날 짜 : 2018. 8. 24.(금)

장 소 : 한국프레스센터 국제회의장

주 최 : 서울대학교 아시아태평양법연구소 & 서울대학교 법과경제연구센터

후 원 : NAVER, Google, Microsoft, facebook, kakao

행사 동영상은 다음 사이트를 통해 볼 수 있습니다.

네이버 채널 : <https://tv.naver.com/aipolicyinitiative>

유튜브 : https://www.youtube.com/channel/UCKyxSZOtLB1YvkKM2_Mq8gQ



행사일정

- 전체 진행 : 최경진 (가천대학교)

- 환영사 10:00-10:05
 - 고태수 (서울대학교 법과경제연구센터장)
- 축사 10:05-10:15
 - 장병규 (대통령직속 4차산업혁명위원회 위원장)

- 개회연설 10:15-10:35
 - Julie Brill (마이크로소프트, 전 미국 연방거래위원회 위원)

- 기조연설 I 10:35-10:55
 - Edward Felten (프린스턴대학교)

- 세션 I 11:00-12:30

데이터 거버넌스, 인공지능, 그리고 프라이버시 : 라운드 테이블
Round Table Discussion on Data Governance, AI and Privacy
모더레이터 : 임 용 (서울대학교)

패널

 - 김형배 (공정거래위원회)
 - 송창현 (네이버, 네이버랩스)
 - Jatinder Singh (캠브리지대학교)
 - Julie Brill (마이크로소프트, 전 미국 연방거래위원회 위원)
 - Edward Felten (프린스턴대학교)

- 점심 · 휴식 12:30-14:00



- 기초연설 II** 14:00–14:20
- David Gunning (미 국방부 DARPA, Explainable AI (XAI) Program)
- 세션 II** 14:25–15:55
- 자동화된 의사결정에서의 책임성(accountability)과 윤리
Accountability and Ethics in Autonomous Decision-Making
모더레이터 : 고태수 (서울대학교)
- 패널 • Woodrow Hartzog (노스이스턴대학교)
- Brent Mittelstadt (옥스포드 인터넷연구소)
 - Su Jiang (북경대학교)
 - Jake Lucchi (구글)
 - David Gunning (DARPA)
- 휴 식** 15:55–16:20
- 세션 III** 16:20–17:40
- 블록체인을 포함한 신기술을 어떻게 규제할 것인가
How to Regulate Blockchain and Other Emerging Technologies
모더레이터: 남수진 (한국외국어대학교)
- 패널 • Michèle Finck (막스플랑크연구소)
- Douglas Arner (홍콩대학교)
 - 이상용 (충남대학교)
 - 양현서 (카카오)
- 전체 토론 및 폐회** 17:40–18:00

국제학술대회

인공지능의 시대: 기술 발전에 따른 책임과 규제

(Artificial Intelligence Today: Governance and Accountability)

들어가는 말

현 세대의 인류는 새로운 시대의 출발점에 서 있다. 특히 ‘인공지능(Artificial Intelligence, AI)’과 ‘빅데이터(Big Data)’와 관련된 논의가 구체적인 분야를 막론하고 세계 곳곳에서 한창 진행되고 있다.¹⁾ 서울대학교 아시아태평양법 연구소와 법과경제연구센터는 이러한 국제적인 흐름에 발맞추어 지난 2017년부터 인공지능과 빅데이터를 둘러싼 여러 법적, 정책적 이슈를 논의하기 위하여 국내외 관련 분야의 전문가들을 초빙하여 컨퍼런스를 개최해오고 있다.²⁾ 한편, 서울대학교 법과경제연구센터에서는 2017년부터 인공지능의 지속적 발전과 활용증가로 나타나는 사회적, 경제적, 법적, 윤리적 문제들을 고민하고 정책적 함의를 모색하는 연구 프로그램인 서울대학교 인공지능 정책 이니셔티브(SNU AI Policy Initiative)를 진행 중이다.³⁾

2018년 제2회 인공지능 컨퍼런스에서는, 첨단기술 발전과 더불어 도래한 인공지능 시대에서의 ‘책임’과 ‘규제’를 주제로 삼았다. 제1세션은 주로 규제 측면에 집중하여, 올바른 ‘데이터 거버넌스(Data Governance)’의 구축에 관한 문제를, 제2세션은 주로 책임 측면에

1) 대표적 사례로, 스탠포드 대학에서 2014년부터 진행 중인 ‘인공지능에 대한 100년 연구(One Hundred Year Study on Artificial Intelligence)’와, 오바마 정부 당시 빅데이터와 인공지능에 관한 백악관 보고서, ‘AI Now Institute’에서 2016년부터 진행되고 있는 심포지엄(<https://ainowinstitute.org/> 참조) 등이 있다.

2) 제1회 컨퍼런스에서는 ‘인공지능, 알고리즘, 개인정보보호를 둘러싼 정책적 과제(Policy Issues surrounding AI, Algorithms & Privacy)’라는 제목을 걸고, 오프닝 세션에서 ‘인공지능, 알고리즘, 개인정보보호를 둘러싼 정책적 과제’를 논의한 뒤 개별 세션에서 ‘인공지능 빅데이터와 시장경쟁의 문제’, ‘인공지능 의사결정과 법적 사회적 책임(accountability)’, ‘데이터 비식별화’를 다루었다. KDI에서는 제1회 컨퍼런스를 촬영하여 인터넷에 자막과 함께 동영상을 업로드 한 바 있다. 그 주소는 아래와 같다.
(https://www.youtube.com/playlist?list=PLOP6ilKzhDLQ_a2hMmD0vxsJn0d-aQco8)

3) 서울대학교 법과경제센터에서는 지금까지 많은 연구를 거쳐 『핀테크 시대』, 『개인정보 보호의 법과 정책』, 『개인정보 비식별화 방법론』, 『데이터 이코노미』 등의 연구총서를 발간해왔다.
보다 자세한 내용은 <http://ai.re.kr/who-we-are/> 참조.

집중하여, 점차 현실화되고 있는 많은 ‘자동화된 의사결정(autonomous decision-making)’ 상황에서의 책임성(accountability)과 윤리 문제를 다룬 뒤, 지난 제1회 컨퍼런스와 마찬가지로 마지막 제3세션은 주로 기술적 측면에 주목하여, 작년 말부터 국내에서도 한창 논란이 된 블록체인(blockchain)을 포함한 다양한 신기술을 올바르게 규율하는 방안을 논의하였다.⁴⁾

인공지능과 블록체인을 비롯한 제4차 산업혁명의 이슈는 특성상 국내에만 국한되지 않는 초국적 파급효과를 불러온다는 점에서, 이번 제2회 컨퍼런스에서도 국제적으로 명망 있는 관련 분야 인사들을 초빙하였다. 특히 개최연설자 줄리 브릴(Julie Brill)은 미국 연방거래위원회(Federal Trade Commission, FTC) 위원(Commissioner)으로 재직할 후에 현재는 마이크로소프트(Microsoft)에 재직하고 있어, 국가와 민간기업의 시각을 두루 파악하고 있는 데이터 거버넌스 분야의 세계적 권위자로 손꼽히고 있다. 그밖에 제1 기조 연설자 프린스턴 대학교 교수 에드워드 펠튼(Edward Felten)은 프라이버시와 보안 분야에서 권위자인 동시에 오바마 정부에서는 백악관 Deputy Chief Technology Officer로 재직하면서 IT 정책을 직접 입안하기도 했다. 제2 기조연설자 데이비드 거닝(David Gunning)은 미국 국방부 산하의 방위고등연구계획국(Defense Advanced Research Projects Agency, DARPA)에서 인공지능의 투명성을 높이는 방안으로서 세계적으로 많은 논의가 이루어지고 있는 이른바 ‘설명가능 인공지능(Explainable AI, XAI)’ 프로젝트를 담당하고 있다. 그밖에 미국, 유럽, 아시아 등지에서 인정받고 있는 중진 및 신진 학자들을 고루 초빙하여 균형 잡힌 논의를 이룰 수 있도록 했다. 학계뿐만이 아니라, 대통령 직속 4차산업혁명 위원회 장병규 위원장과 공정거래위원회 김형배 국장과 같이 공직에 몸담고 있는 인사들과, 네이버 송창현 CTO처럼 사기업에 재직 중인 인사들을 고루 초빙하여 컨퍼런스 논의가 각계각층의 시각을 반영할 수 있도록 많은 노력을 기울였다. 이에 법조계뿐만 아니라 공학, 경제학, 철학 등 다양한 분야의 배경을 가진 300명이 넘는 청중들이 컨퍼런스에 참석하여 자리를 빛내주었다.

4) 자세한 내용은 <http://ai.re.kr/공지사항/?uid=10&mod=document&pageid=1> 참조.

I. 제1세션 - 데이터 거버넌스, 인공지능, 그리고 프라이버시 : 라운드 테이블 (Round Table Discussion on Data Governance, AI and Privacy)

1. 개회연설

개회연설자 줄리 브릴이 본 대한민국은 대변혁의 중심에 선 국가이다. 대한민국은 브릴이 30년 전 처음 방문한 당시는 산업화와 민주화라는, 두 번째로 방문한 지금은 제4차 산업혁명이라는 대변혁에 마주하고 있다. 이러한 시대의 흐름에서 대한민국은 어떠한 역할을 할 수 있을까? 브릴은 5년 연속 블룸버그(Bloomberg) 혁신지수 세계 1위,⁵⁾ 근로자 1인당 로봇 수 세계 1위,⁶⁾ GDP 대비 R&D 비율 세계 2위⁷⁾ 인공지능 특허 수 세계 3위⁸⁾ 등의 지표를 언급하면서, 대한민국이 가지고 있는 잠재력을 높이 평가하였다. 그리고 브릴은 대한민국 정부가 지난 5월, 향후 5년 간 인공지능 분야에 약 2.2조원을 투자할 것이라는 발표를 한 사실을 상기하면서,⁹⁾ 정부의 강력한 의지가 필요하다는 점을 덧붙였다.

널리 알려져 있는 것처럼 제4차 산업혁명의 두 축인 인공지능과 빅데이터는 인류의 많은 부분을 뒤바꿔왔다. 여가, 교통, 교육을 비롯한 일상뿐 아니라, 사회복지, 고용, 안보를 비롯한 막대한 파급효과를 낼 수 있는 영역에까지 신기술이 점차 영향력을 뻗어 오고 있다.¹⁰⁾ 이러한 변화를 가능케 한 요인으로는 저장장치의 발달과 클라우드 컴퓨팅 등을 통한 빅데이터 시대의 도래와, GPU와 병렬컴퓨팅으로 대표되는 새로운 하드웨어

5) Michelle Jamrisco · Wei Lu, “The U.S. Drops Out of the Top 10 in Innovation Ranking”, Bloomberg (2018. 1. 23.).

6) 그러나 이 지표는 기계의 노동력 대체에 관련된 부정적인 요인으로 평가되기도 한다. 예를 들어, 구본권, “‘로봇 도입률 1위’ 한국은 자동화에 더 안전할까?”, 한겨레 (2017. 3. 17.) 참조.

7) 2016년 기준으로 우리나라의 GDP 대비 총 R&D 비율은 4.24%로, 4.25%를 차지한 이스라엘에 이은 2위를 기록하였다(한국연구재단, “R&D 통계 핸드북 2018” (2018), 8쪽).

8) 조일구, “인공지능(AI) 기술을 선도하는 미국, 중국, 한국의 최근 특허 및 논문 활동 조사”, 주간기술동향 (2017), 23쪽에 따르면 대한민국의 인공지능 관련 특허 수는 미국, 일본에 따른 3위이다.

9) 과학기술정보통신부, “I-Korea 4.0 실현을 위한 인공지능(AI) R&D 전략” (2018), 39쪽.

10) Peter Stone et al., “Artificial Intelligence and Life in 2030”, One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University, Stanford, CA (2016), p. 18. 이하는 인공지능이 커다란 영향을 끼칠 것으로 예상되는 주요 영역으로, 교통(Transportation), 가정/서비스 로봇(Home/service Robot), 보건의료(Health Care), 교육(Education), 자원이 부족한 지역(Low-resource communities), 공공안전 및 안보(Public safety and Security), 취업과 일자리(Employment and Workplace), 여가(Entertainment) 등 총 8가지를 언급하고 있다. 이 문제에 관한 국내의 논의로서, 관계부처 합동, “제4차 산업혁명에 대응한 지능정보사회 중장기 종합대책” (2017), 15쪽 이하의 의료, 제조, 금융, 유통, 산업 기타, 교통, 도시, 웰니스, 주거 등을 ‘9대 영역’으로 제시하였다.

기술과 알고리즘 효율성을 높여주는 최적화 기술 등을 바탕으로 한 정보처리 속도의 급격한 향상이 꼽힌다.¹¹⁾ 그밖에 수많은 관련기술의 발전을 통해 시지각(vision recognition) 분야에서는 이미 인공지능의 인식률 정확도가 인간의 수준을 능가하기에 이르렀다.¹²⁾ 자율주행 자동차(Autonomous Car)¹³⁾를 비롯한 여러 영역에서, 종래 인간이 담당할 일을 자동화하는 기술이 상용화되었거나, 상용화를 앞두고 있다. 이전의 산업혁명과 제4차 산업혁명을 차별화하는 핵심요소로 꼽히는, 육체적 능력을 넘어선 인지능력의 자동화로 인해 인간은 창의적 업무에 시간과 역량을 전적으로 투입할 수 있게 되었고 보다 자유로워졌다.

브릴은 인공지능이 인간의 삶을 실질적으로 개선한 대표적 사례로 현재 진행 중에 있는 프리모니션(Premonition) 프로젝트를 언급하였다. 이 프로젝트는 지카 바이러스나 에볼라 바이러스 등 치명적 질병의 확산을 방지하기 위한 해법을 찾는 것을 목적으로 한다. 자율주행 드론이 모기를 채집한 뒤, 피를 뽑아 모기가 문 동물 피에 대한 유전정보를 얻어 병원균에 대한 이해를 높여나가게 된다. 이러한 작업을 통해 인류에게 치명적인 질병의 조기경보 시스템을 갖춰나가는 것이 프리모니션 프로젝트의 최종 목표이다.¹⁴⁾ 로체스터 공과대학에서 인공지능 음성인식 솔루션을 활용하여 청각장애인들이 수업을 들을 수 있도록 도움을 준 일도 언급하였다¹⁵⁾.

그렇지만 브릴은 다른 한편으로 신기술에는 늘 명암이 공존해온 점을 인정하였다. 작게는 일자리가 없어지고, 크게는 산업 자체가 사라지는 경우도 있었다. 이런 변화를 통해 경제 전반이 흔들릴 수 있고, 누군가에게는 기회가 될 수 있지만 다른 누군가는 큰 피해를 입을 수도 있다.¹⁶⁾ 기술이 인류를 파멸로 이끌거나, 기술의 혜택을 일부만 누릴

11) Jatinder Singh et al., “Responsibility & Machine Learning: Part of a Process”, (2016), p. 2.

12) 이미지넷(Imagenet)은 2010년부터 ‘대용량 시각인식 경연대회(Large Scale Visual Recognition Challenge, LSVRC)’라는 경연대회를 개최해 이미지 인식능력을 겨루어 왔는데, 인공지능의 오류율은 2010년 28.5%에서 2017년 2.5% 미만까지 감소했다(Yoav Shoham et al., “Artificial Intelligence Index 2017 Annual Report” (2017), p. 26.). 특히 마이크로소프트의 ‘ResNet’은 2015년 약 3.5%의 오류율을 달성하여 인간의 정확도를 추월하는 결과를 얻기도 하였다.

13) 자율주행 자동차에 관한 자세한 내용은, 이상돈 외 6인, 『인공지능과 자율자동차, 그리고 법』, 세창출판사 (2017) 참조.

14) 보다 자세한 정보는 <https://www.microsoft.com/en-us/research/project/project-premonition/> 참조.

15) 보다 자세한 정보는 <https://www.facebook.com/MicrosoftKorea/videos/1750267711677486/> 참조.

16) Executive Office of the President, “Artificial Intelligence, Automation, and the Economy” (2016), pp. 1-2에는 이전의 기술발전이 저숙련 노동자의 생산성 향상을 가져온 반면, 최근의 기술발전은 오히려 고

수 있다면, 그러한 기술발전은 누구에게도 지지되기 어려울 것이다. 브릴은 이를 인정하면서, 마이크로소프트 역시 기본적으로 인공지능에 대한 인간중심적 접근법을 견지해 나가기 위해 다음과 같은 6가지 원칙을 고안하였다고 밝혔다.¹⁷⁾

[마이크로소프트 인공지능 윤리 6원칙]

1. 프라이버시와 보안: 여느 클라우드 기술과 마찬가지로, AI 시스템 역시 데이터의 수집, 사용, 그리고 저장에 관련한 개인정보 보호법을 준수하고, 개인정보가 개인정보 보호 기준에 따라 사용되고 보호받을 수 있도록 해야 한다.
2. 투명성: AI가 미치는 영향력이 늘어나는 만큼, AI가 어떻게 작동하는지에 대한 정보를 이용자들에게 제공해야 한다. AI가 편향된 또는 오류가 있는 의도치 않은 결과물을 내놨을 때, 보다 쉽게 식별할 수 있도록 해야 한다.
3. 공정성: AI가 예를 들어 환자를 위한 치료법, 또는 인력 고용에 관한 의사 결정을 할 때, 비슷한 증상 또는 자격을 가진 모든 이들에게 같은 결과를 제공해야 한다. 공정성을 보장하기 위해, AI 시스템에 편견 또는 성향이 어떤 영향을 주는 지 파악해야 한다.
4. 신뢰성: AI는 뚜렷한 기준에 따라 예기치 못한 상황에서도 문제없이 대응하고, 의도치 않은 방향으로 발달하지 않도록 엄밀한 시험 과정을 거쳐야 한다. AI가 언제 어떻게 사용되어야 하는지를 판단하는 사람들의 역할이 중요하다.
5. 포괄성: AI는 인간의 광범위한 요구와 경험에 충족하는 결과물을 내놓기 위해, 의도치 않게 특정 제품이나 환경에 대한 분석을 제외하는 일이 없도록 해야 한다.
6. 책무성: AI 개발자나 서비스 제공자들은 자신들의 AI 시스템에 대한 책임을 가져야 한다. AI의 개발 단계에서부터 시스템이 운영되는 있는 동안까지도 언제나 지켜져야 한다.

비록 이 6가지 원칙이 모든 윤리원칙을 포괄하는 것도 아니고, 반드시 가장 중요한 원칙이라고 할 수도 없겠지만, 브릴은 6원칙이 최소한 인공지능 윤리에 관한 논의를 시작하는 데 있어 중요한 역할을 담당하였으면 한다는 입장을 밝혔다. 이처럼 기술이 ‘할 수 있는 것’과 ‘하여야 하는 것’ 사이의 상충은 최근 국내에서도 큰 문제가 된 바 있다. 올

숙련 노동자의 생산성 향상을 가져와서 빈부격차를 증대시켰다는 점이 지적된 한편, 앞으로의 인공지능 기술발전은 저숙련, 고숙련 노동자를 가리지 않고 위협이 될 수 있다는 점이 지적된 바 있다.
 17) 마이크로소프트는 지난 2016년 3월 테이(Tay)라는 채팅봇을 출시했다가, 인종차별적 표현을 학습하게 되어 곤욕을 치른 일이 있고, 이후 윤리성에 대한 고민을 확대해왔다.
 마이크로소프트 인공지능 6원칙은 https://news.microsoft.com/ko-kr/features/ai_is_built_on_trust/ 참조.

초 카이스트(KAIST)가 방산업체와 인공지능 자율무기(Lethal Autonomous Weapon System, LAWS)¹⁸⁾ 개발을 연구하려고 계획한다는 기사가 보도된 뒤, 저명 로봇학자들이 보이콧을 선언한 일이 있었다. 총장이 즉각 소위 ‘킬러 로봇’을 개발할 계획이 없다는 입장을 표명 하면서 사태가 일단락되기는 했지만,¹⁹⁾ 기술개발과 윤리가 불가분의 관계에 있다는 점을 깨닫게 한 중요한 사건이었다. 오늘날과 같은 다양성의 세계에서, 과연 ‘인공지능 윤리’는 누구의 어떤 가치관을 말하는 것일까? 브릴은 이와 같은 어려운 물음에 직면하여, 각계 각층의 전문가들이 머리를 맞댈 수 있는 장으로서 이번 인공지능 컨퍼런스의 의의를 부각 하면서 발표를 마무리하였다.

2. 기조연설

기조연설자 에드워드 펠튼의 연설주제는 인공지능의 ‘설명가능성(explainability)’에 관한 것이었다. 인공지능이 유발하는 여러 문제의 근원은 대체로 최근 대두되고 있는 인공 신경망(Artificial Neural Network, ANN) 기반 인공지능 의사결정 방식과, 우리 인간의 그것이 근본적으로 다르기 때문에 비롯되는 것으로 평가된다. 문제는 실제로 법적 문제를 해결하려면 책임소재를 가려야만 하는데, 이와 같은 이른바 인공지능의 불투명성(opacity)²⁰⁾은 책임소재를 가리기 매우 어렵게 만든다는 것이다. 설명가능성 이슈는 인공지능의 의사결정 방식을 인간이 이해할 수 있는 방식으로 설명할 수 있는 방법론에 관한 논의이다. 펠튼은 설명가능성 논의의 맥락을 제대로 이해하려면, 먼저 인공지능의 역사에 대한 간략한 이해가 필요하다고 지적하였다.

인공지능의 시초로는 인간의 두뇌 구조에서 영감을 받아, 이를 수학적으로 모사한 모델을 통해 추론을 할 수 있을 것이라고 바라본 맥컬럭(McCulloch)과 피츠(Pitts)의 1943년 논문이 꼽힌다.²¹⁾ 이후 튜링(Turing)은 1950년 논문에서 “기계도 생각할 수 있는가?”라는 역사에 남는 질문을 던진 뒤, ‘이미테이션 게임(imitation game)’ 혹은 ‘튜링

18) 이른바 ‘치명적 자율무기’로 불리는 인공지능 무기에 관련된 논의로는 한희원, “인공지능(AI) 치명적 자율무기(LAWS)의 법적·윤리적 쟁점에 대한 기초연구”, 중앙법학 제20권 제1호 (2018) 등 참조.

19) 김연지, “‘킬러 로봇 우려’ 해외학자 57인, KAIST 보이콧 선언 ‘철회’”, 노컷뉴스 (2018. 4. 9.).

20) 인공지능의 불투명성에 관련된 논의로, Jenna Burrell, “How the machine ‘thinks’: Understanding opacity in machine learning algorithms”, Big Data & Society (2016) 등 참조.

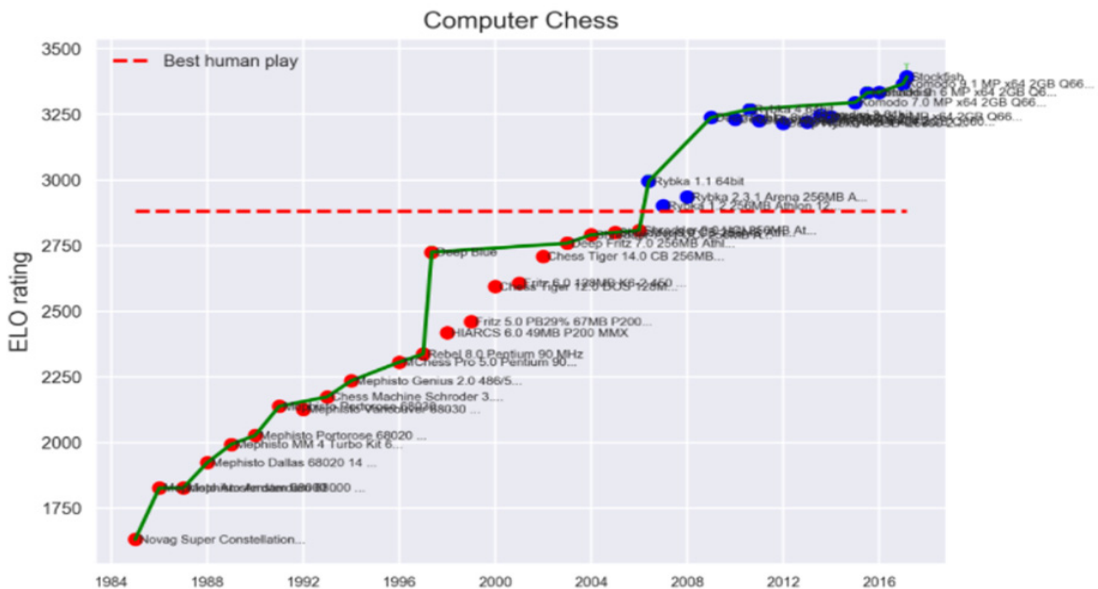
21) Warren S. McCulloch · Walter H. Pitts, “A Logical Calculus of the Ideas Immanent in Nervous Activity”, Bulletin of Mathematical Biophysics, 5 (1943).

테스트(Turing test)'라고 불리는 '생각하는 기계'에 대한 기준을 제시하였다.²²⁾ 그가 제시한 기준이 독특한 점은, 심적 경험 내지는 내적 상태를 배제한 채로, 오직 겉으로 드러난 행태만을 기준으로 지능을 정의했다는 것이다.²³⁾ 펠튼은 1950년대부터 2000년대 까지 부침을 거듭하면서 많은 발전이 있었지만,²⁴⁾ 중요한 점은 결과적으로 2000년이 되었음에도 해결되지 못한 문제가 산더미처럼 많았다는 것이라고 지적한다. 그러다가 2010년 즈음부터 앞서 언급하였듯 데이터, 소프트웨어, 하드웨어 측면에서 양적, 질적으로 급격한 성장이 이루어지면서 현재와 같은 변혁기를 마주하게 되었다.

펠튼은 이러한 역사 속에서 자신이 얻은 교훈을 크게 세 가지 측면에서 제시한다. 첫째로, 인공지능은 단일체가 아니라는 점이다. 마치 영화에서처럼 어떤 자의식 있는 단일체만이 인공지능이라고 착각할 수 있지만, 현실에서 인공지능이라고 할 수 있고 그렇게 불리는 범주는 일률적으로 규정할 수 없을 만큼 넓다. 또한 현재 상용화되고 있는 인공지능은 소위 '약인공지능(weak AI)'으로 불리는 좁은 범위의 작업에 국한된 것이지, '강인공지능(strong AI)' 내지 '일반인공지능(Artificial General Intelligence; AGI)'이 아니다. 데이브 허니(Dave Honey)라는 사람은 “돈을 벌려면 약인공지능을, 영화를 만들려면 강 인공지능을 다루어야 한다”는 말을 하기도 했다. 결과적으로 보면 자동화가 완료되는 시점이나 인간의 능력을 초월하는 시점은 단일하기보다는, 구체적 과제에 따라 다를 것으로 예측된다. 둘째로, 인공지능은 사람처럼 생각하지 않는다는 점이다. 앞서 튜링의 통찰에서 보았듯, 우리의 주된 초점은 지능적 시스템을 제작하는 것이지 인간지능을 고스란히 구현하는 것이 아니다²⁵⁾. 때문에 인공지능은 인간이라면 절대로 풀지 못하는 문제를 풀어내기도 하지만, 인간이라면 절대로 하지 않을 실수를 하기도 한다.²⁶⁾ 이는 그 자체로 문제지만, 인간과 인공지능이 협업을 하는 경우에도 장애물이 된다.²⁷⁾ 셋째로,

22) A. M. Turing, “Computing Machinery and Intelligence”, *Mind*, New Series, Vol. 59, No. 236 (1950).
 23) 사고가 아닌 행동을 기준으로 한 튜링의 정의는 이후 철학자 존 설의 일명 ‘중국어 방 논변(Chinese Room argument)’에 의해 비판을 받기도 한다(John R. Searle, “Minds, brains, and programs”, *Behavioral and Brain Sciences*, 3(3) (1980) 참조).
 24) 이에 관한 자세한 내용은, 마쓰오 유타카, 박기원 역, 『인공지능과 딥러닝』, 동아엠앤비 (2015) 참조.
 25) 스텐퍼트 러셀 · 피터 노빅, 류광 역, 『인공지능: 현대적 접근방식(제3판)』, 제이펍 (2016), 6쪽 참조.
 26) 구글에서 흑인 사진에 ‘고릴라’ 표지를 붙여 문제가 된 사안이 대표적이다. Sophie Curtis, “Google Photos labels black people as ‘gorillas’”, *The Telegraph* (2015. 7. 1.).
 27) 많은 눈자들이 인공지능과 인간의 협업을 강조하고 있는데, 일례로 영상의학 분야에서 림프절 세포를 관찰해 암세포가 포함되었는지를 판별할 때의 오차율이 인공지능은 7.5%, 병리학자는 3.5%인데, 이들이 협동한 결과 0.5%로 크게 줄었다(Executive Office of the President, “Preparing for the future of AI” (2016) p. 11.).

많은 과제에서 대체로 노력을 하면 할수록, 데이터를 많이 투입할수록 성과가 좋아졌다는 점이다. 인공지능은 특정 시점에 제작한 뒤 그것으로 끝나는 산물이 아닌, 지속적인 학습(업데이트)을 통한 성과물이다. 펠튼은 그 증거로 80년대부터 체스 프로그램 능력을 수치화한 지표인 ELO 점수가 꾸준히 증가한 점을 제시했다. 흔히 말하는 것처럼, 경험으로부터 배우는 능력은 직관을 발휘하는 인간이 우수하다고 해도, 인공지능은 배울 수 있는 경험 자체를 인간에 비해 더 많이 할 수 있다는 점에서 지금과 같은 성과가 도출될 수 있었던 것이다.



[그림 1] 체스 프로그램 ELO 점수 증가 추이(펠튼 발표자료, 자료집 39쪽)

이런 시사점을 토대로 설명가능성 문제를 살펴보면, 몇 가지 통찰을 얻을 수 있다. 흔히들 인공지능은 설명가능성이 떨어지는 반면에, 사람은 그렇지 않다고 단정적으로 이야기하지만 과연 그런지는 의문이다. 오히려 기계는 규칙에 따라 정확히 작동하는 반면, 인간은 설명을 내놓기는 해도 그 설명과 실제 동기가 일치하는지 알기 어렵다. 더군다나 인류가 인간 두뇌를 그렇게 많이 파악하고 있지도 않다. 때문에 인공지능의 설명력이 뒤떨어진다는 주장은 다음과 같이 정교하게 구분하여 이해할 필요가 있다.

펠튼에 따르면, 인공지능의 설명가능성이 낮다는 것은 크게 네 가지 의미를 가지고

있다. 첫 번째로는 투명성(transparency)이 낮다는 것일 수 있다. 인공지능이 무언가 하고 있다는 정보는 알 수 있지만 그 정보를 공개하지 않기 때문이다.²⁸⁾ 이런 측면의 불투명성은 대개 영업비밀 내지 안전성(보안) 문제, 혹은 조작(gaming) 위험성 등을 근거로 유발되게 마련이다. 그런데 이는 생각해 보면, 인공지능 기술 그 자체의 문제는 아니다. 오히려 불완전한 데이터 거버넌스 내지 법체계가 불투명성을 유발한 직접적 원인이라고 볼 수 있다. 둘째로는, 설명이 존재하기는 해도, 너무 복잡하다는 것이다. 앞서 언급하였듯, 인공지능 기반의 인공지능은 대체로 인간이 직관적으로 이해하기 어려운 것으로 알려져 있다. 무엇이 입력되어, 어떤 수학적 연산과정을 거쳐서, 어떤 결과를 도출했다는 정도의 서술은 가능하더라도, 이와 같은 서술은 인간의 ‘이해’라는 관점에서 진정한 의미의 ‘설명’이라고 보기는 어렵다. 셋째는 인공지능의 의사결정이 인간이 볼 때 비합리적인 경우이다. 시스템이 이런 행동을 보이고 있고 왜, 무엇을 하고 있는지도 설명할 수 있는데 인간이 결과를 합리적이라고 생각하지 않는 경우다. 예를 들어, 대출의 상환가능성을 예측하는 시스템이 ‘특정한 색의 신발을 신는 사람은 상환을 할 가능성이 높다/낮다’는 등 인간이 보았을 때 상관성이 없는 변수를 가지고 대출의 상환가능성을 예측할 수 있다. 통계적으로 상관관계의 측면에서 보면 유효한 예측일 수도 있지만, 인간 사회에서 이런 설명이 수용될 수 있는지는 논란의 여지가 있다.²⁹⁾ 마지막으로 정당화가 부족한 경우이다. 시스템의 행위가 논리적으로 이해는 되지만, 도출된 결과가 정당하지 않다고 보는 것이다. 이는 설명력이 아니라, 결과에 대한 불만족이라고 말할 수 있다. 인공지능의 의사결정 결과물과 인류가 가지고 있는 도덕적 가치를 이어주는 진정한 의미에서의 ‘설명’을 필요로 하는 것이다.³⁰⁾

결과적으로, 설명가능성의 문제는 전적으로 기술적 문제도, 전적으로 규범적 문제도 아니다. 진정한 설명이 무엇인지를 판단하는 문제는 결국 규범적으로 판단하여야 할 문제

28) 인공지능의 투명성과 관련된 여러 문제점을 지적한 책으로, 프랭크 파스칼레, 이시은 역, 『블랙박스 사회』, 안티고네 (2016) 참조.

29) 일례로 페이스북에 업로드 한 컬리 프라이(curlly fries) 사진에 ‘좋아요(like)’ 버튼을 클릭한 사람과 그들의 지능지수 사이에 높은 상관성이 도출되었다고 하더라도, 이를 채용의 근거로 삼는 것이 정당한지 여부는 전혀 별개의 문제이다(Catherine Tucker, “Privacy, Algorithms and Artificial Intelligence”, *The Economics of Artificial Intelligence: An Agenda* (2017), p. 9. 참조).

30) 이 문제에 관한 대표적 문헌으로, Executive Office of the President, “Big Data: Seizing Opportunities, Preserving Values” (2014); Federal Trade Commission, “Big Data: A Tool for Inclusion or Exclusion” (2016); Executive Office of the President, “Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights” (2016) 등 참조.

이지만, 진정한 설명을 구현하는 작업은 결국 기술에 의존하여야 하기 때문이다. 진정한 ‘설명’을 뒷받침할 수 있는 법적, 기술적 요소가 무엇이고, 양자를 조화할 수 있는 거버넌스 체계는 어떠한지를 살펴보기 위해 향후의 논의가 이어지게 된다.

3. 라운드 테이블(round table)

지금까지 논의한 것처럼, 제4차 산업혁명은 광범위한 기술적 변혁뿐 아니라 사회적 변혁을 함께 수반하고 있다. 그에 따라 개인들은 사용자·소비자의 입장에서 새로운 도전 과제에 직면하게 된다. 이에 대해 자틴더 싱(Jatinder Singh)³¹⁾ 패널은 소비자들이 개인 정보 이용에 대한 동의를 무수히 요구받는 과정에서, 동의가 지닌 중요성에 비해 그에 대한 알 권리를 제대로 누리지 못하여 개인정보자기결정권이 침해당하고 있다고 인식한다는 점을 지적했다. 실질적으로 ‘충분한 정보에 기반을 둔 사전동의(informed consent)’가 이루어질 수 있으려면, 단순히 기술 그 자체에 대한 정보뿐만이 아니라, 동의에 따른 산출물, 기술에 내재된 위험성, 동의를 통해 이득을 보는 사람을 비롯한 관련된 이해 당사자 등의 정보가 제공되어야 한다.³²⁾ 싱은 비록 현재는 소비자들이 별 생각 없이 정보제공에 동의하는 경향이 있지만, 예를 들어 집 문이 잠겨서 외출할 수 없는 상황이 도래한다면 양상이 크게 달라질 것이라고 보았다.

그렇다면 기업 입장에서는 과연 장밋빛 미래만이 존재하는 것일까? 송창현 네이버 CTO는 기업 입장에서 처한 난점을 다음과 같이 요약했다. 네이버는 2011년부터 인공지능에 대한 투자를 시작한 뒤, 챗봇 ‘톡톡’, 번역기 ‘파파고’, 스피커 ‘클로바’ 등을 개발하고 최근에는 네이버랩스라는 자회사를 통해 연구개발을 계속해오고 있다. 소위 ‘앰비언트 지능(ambient intelligence)’으로 불리는 생활환경지능을 구현하기 위하여 인공지능을 물리적 공간과 결합하는 프로젝트를 진행하고 있다. 자율주행과 이를 위한 지도 작성 부분에서는 어느 정도 성과도 거두었다. 그러나 대내적으로 관련 분야의 인재, 컴퓨터 인프라, 데이터 태깅(tagging)에 필요한 인력 등이 부족한데다가, 대외적으로 새로운 사업을 진행하려고 할 때 명확한 법과 정책이 부재한 일도 많다. 정부와 논의

31) 캠브리지대학교 연구원.

32) 정보제공 그 자체가 실질적인 ‘인지와 판단’의 토대까지 제공하지 않는다는 점을 지적하는 권영준, “계약법의 사상적 기초와 그 시사점”, 저스티스 통권 제124호 (2011), 193-194쪽 참조.

하려고 하더라도 컨트롤타워가 부재하여 누구와 상의하여야 하는지, 누구에게 인증을 받아야 하는지 알기 어려운 형편이다. 다른 한편으로는, 데이터 액세스 허용여부를 놓고 신규 기업과 전통적 기업이 갈등을 빚고 있는 상황이라고 하였다³³⁾.

데이터 거버넌스의 마지막 한 축인 규제당국 입장에서는 어떤 도전과제가 부각되고 있을까? 김형배 패널은 규제당국의 입장에서 최근 부각되는 데이터 경제와 인공지능 기술은 그 실체와 방향성을 종잡기 어렵다는 점을 인정하는 한편, 경쟁법과 소비자법 차원에서 어떻게 대처할지에 관하여 학계·산업계와 협업중이라고 밝혔다.³⁴⁾ 당국은 기본적으로 새로운 산업은 국가경제에 이롭다는 입장이되, 다만 그 과정에서 발생한 소비자 해악이나 경쟁질서 침해 행위는 사후적으로 규제해야 한다는 원칙에 입각하고 있다는 점을 덧붙였다.

제4차 산업혁명은 이처럼 소비자, 기업, 규제당국 각각의 입장에서 과거와는 다른 새로운 도전과제를 제시하고 있다. 그동안 바람직한 규제 방식에 관하여 법적 접근과 기술적 접근의 우월성을 놓고 설왕설래가 이어져 왔다. 펠튼은 두 가지 방식 모두가 혼용될 것이라고 바라보면서, 기술이 해결하지 못하는 부분이 분명히 존재하고, 법이 기술발전 속도를 따라가지 못하는 부분이 분명히 있기 때문에 양자를 적절히 조화할 필요가 있다는 결론을 제시하였다. 개발자들은 사회적 요구를 반영한 설계를 하여야 하고, 정책당국은 기술적 이해도를 높이기 위해 전문 인력을 충분히 확보해야 한다는 것이다.

브릴은 규제의 장기적 필요성에 대하여는 공감하면서도, 자동차와 의약품의 사례를 들면서 법은 결국 기술 발전을 사후적으로 쫓아가는 입장일 수밖에 없으므로, 사전에 즉각적으로 개입하는 방식은 다소 성급할 수 있다고 지적하였다. 구체적으로 명확한 원칙부터 수립한 뒤, 기업이 윤리적 개발을 해나가도록 윤리적 프레임워크를 구축할 필요가 있다고 보았다. 그리고 당국은 소비자들이 예컨대 반사실적(counterfactual) 설명과 같은 유용한 정보를 적시에 얻을 수 있도록 실질적 규율을 펼쳐나가야 한다고 주장하였다.³⁵⁾

33) 인공지능 분야에서 국가의 역할에 관하여는 김병운, “인공지능 동향분석과 국가차원 정책제언”, 정보화 정책 제23권 제1호 (2016), 86쪽 이하 등 참조.

34) 제4차 산업혁명의 추세에 맞추어 경쟁법 체계를 개편하여야 한다는 입장으로 공정거래위원회, “공정거래법제 개선 특별위원회 최종 보고서” (2018), 27-28쪽 참조.

35) The White House, “Consumer Data Privacy in a Networked World: A framework for protecting privacy and promoting innovation in the global digital economy” (2012), pp. 5-7 참조.

한편, 기업 입장에서 소비자들이 데이터 이용방식에 대한 의구심을 느끼게 될 경우 산업 전체가 붕괴될 수 있다는 점에서, 법적 규제의 존부와 무관하게 신뢰를 줄 수 있는 윤리적 기술개발을 할 필요가 있다고 보았다. 예를 들어, 브릴은 글로벌 기업인 마이크로소프트는 EU GDPR (General Data Protection Regulation)의 직접 적용 대상이기 때문에 GDPR이 정한 데이터 접근권, 삭제권, 이동권 등을 준수해야 하는 것은 당연하지만, 그와 무관하게 ‘프라이버시 대시보드’라는 툴을 이용하여 이와 같은 권리를 확보하고 있다고 언급하였다. 나아가 기업이 규제당국에 대한 교육을 담당하거나, 정부에 대해 이의를 제기할 수도 있어야 한다고 주장했다. 송창현 패널 또한 지도 서비스나 음성인식 서비스 등에서 문제되는 위치정보를 비롯한 프라이버시 문제에 대해 언급하면서, 소비자들이 개인정보에 관한 통제권을 행사할 수 있도록 해주는 유저 인터페이스(UI)를 개발하고 있다고 덧붙였다. 같은 맥락에서, 개인정보자기결정권이 올바르게 행사되기 위한 전제로 정보격차 내지 정보문맹의 문제가 해소될 필요도 있다. 펠튼은 한편으로는 양과 질 측면에서 과거와 다른 측면이 있음을 인정하면서도 다른 한편으로는 과거에 있었던 문제의 연장선에서 파악할 수 있음을 강조하였다. 교육이나 복지 등을 통한 적절한 정책적 대응을 통해 비록 해결하기 어렵지만, 해결하기 불가능한 일은 아니라는 것이다.³⁶⁾ 김형배 패널은 혁신산업에서는 과잉규제가 소비자 후생을 저해할 수 있어 조심스러운 태도를 취해야 하지만, 소비자 이익을 저해하는 반경쟁적 행위에 대한 규제는 불가피함을 역설했다.

결론적으로 합리적 데이터 거버넌스 모델을 수립함에 있어, 대한민국은 자율규제를 보다 강조하는 미국과, 엄격한 규제를 보다 강조하는 유럽연합 식의 두 가지 모델 중 어느 하나를 준거로 삼든지, 혹은 제3의 길을 택해야 하는 기로에 놓여있다. 브릴은 두 모델에 대하여는 모두 B+ 정도의 점수를 줄 수 있지만, 대한민국은 양자의 장점을 혼용하여 A대 점수를 받을 수 있는 기회가 있다는 점을 강조하였다. 일례로 GDPR은 비식별화에 대해 상대적으로 엄격한 기준을 부여하고 있는 것으로 평가할 수 있는데, 여기에 예를 들어서 FTC가 2012년 내놓은, 제3자에게 비식별 데이터를 제공한 경우 이들에게 재식별화를 하지 않겠다는 명확한 약속을 받는 방안을 결합하면 대한민국은 두 가지 모델의 장점을 모두 누릴 수 있다는 것이다. 데이터 이용과 프라이버시 보호 양자를 조화할 수 있는

36) Executive Office of the President, “Artificial Intelligence, Automation, and the Economy” (2016), p. 26. 이하 참조.

데이터 거버넌스 모델을 확립할 수 있다는 것이다. 이에 대해 펠튼은 원칙적으로 동의 하면서, 실무에서 재식별화 수준이 어느 정도까지 도달했는지 충분한 기술적 파악이 선행 되어야 함을 덧붙였다.

II. 제2세션 - 자동화된 의사결정에서의 책임성과 윤리 (Accountability and Ethics in Autonomous Decision-Making)

1. 기조연설

앞선 세션에서 주로 데이터에 대한 규제 거버넌스 문제를 다루었다면, 이번 세션은 인공지능의 책임성과 윤리성에 초점을 맞추었다. 먼저 기조연설자인 데이비드 거닝은 책임성과 윤리성을 고양하기 위한 대표적 기술적 방법론으로 꼽히는 일명 ‘설명가능 인공지능’의 방법론과 현황을 소개하였다. 미국 방위고등연구계획국이 진행하고 있는 이 프로젝트는 총 4년 간 진행되는 작업으로, 첫째 프로젝트가 완료된 상황이다.³⁷⁾

앞서 설명했듯, 인간의 뇌를 본뜬 인공신경망 기반 인공지능은 인공뉴런과 은닉층 수가 증가함에 따라 정확도가 개선되는 추세에 있다. 인공지능의 정확도가 향상됨에 따라 다양한 응용기술이 출현하고, 종래 인간이 담당하던 많은 업무가 자동화되고 있다. 그러나 이런 장점과는 별개로, 인공지능 알고리즘이 내린 의사결정의 특성으로 인해 종래와는 다른 종류의 부작용이 나타나고 있다. 일견 ‘자율성’이 있는 것처럼 보이는 인공지능의 특성은 책임 공백을 낳을 수 있고,³⁸⁾ 인간의 나쁜 습성이 담긴 데이터를 학습하여 ‘차별 (discrimination)³⁹⁾’과 같은 악습을 되풀이하는 경우도 발견되고 있다. 이런 문제를 사전적으로 예방하거나, 사후적으로 규율하려면, 위법성의 존부와 정도를 판정할 기준이 있어야 하는데, 인공지능의 불투명성은 이를 방해하는 근본적 요인이 된다. 이런 연유로, 대중들에게 인공지능이 의사결정을 내린 근거를 그들이 이해할 수 있게끔 설명해주는 방법론을

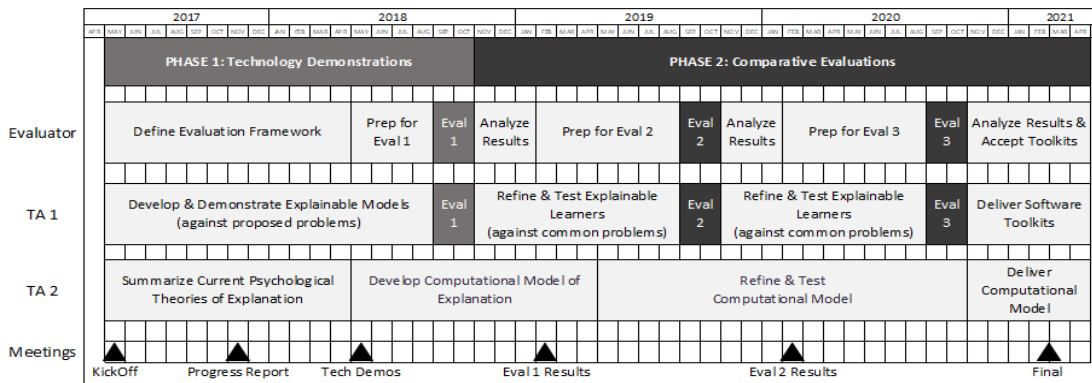
37) <https://www.darpa.mil/program/explainable-artificial-intelligence> 참조.

38) 이상용, “인공지능과 계약법 - 인공 에이전트에 의한 계약과 사적자치의 원칙 -”, 비교사법 제23권 4호 (2016), 1648-1649쪽.

39) 대표적 연구로 Solon Barocas · Andrew D. Selbst, “Big Data’s Disparate Impact”, 104 California Law Review 671 (2016) 등 참조.

마련하려는 ‘설명가능 인공지능’ 연구가 시작된 것이다. 최근 시행된 GDPR의 이른바 ‘설명을 요구할 권리(Right to explanation)’로 인하여, 설명가능 인공지능은 실무적으로도 그 중요성을 널리 인정받고 있다.⁴⁰⁾

설명가능 인공지능 프로젝트는 크게 모델 자체의 설명력을 높이는 작업과, 적절한 설명을 제공하는 유저 인터페이스(UI)를 개발하는 작업으로 두 갈래에 걸쳐 진행되고 있다.⁴¹⁾ 주목하고 있는 영역(attention map)이나 특징(feature) 시각화, 사례나 비유, 반사 실적(counterfactual) 방식을 통한 설명과 같은 다양한 방법론이 연구되고 있다. 설명에 관한 분석뿐만이 아니라, 이를 자율 시스템에 통합하기 위한 방법론, 사용자가 어떤 모델을 보다 더 직관적으로 느끼는지를 심성모델을 동원하여 연구하는 인지과학 방법론 등이 종합적으로 논의되고 있다.⁴²⁾ 최종적인 목표는 일견 상충(trade-off)하는 관계에 놓여있는, ‘모델의 성능’과 ‘설명가능성’ 양자를 모두 드높이려는 것이다.



[그림 2] DARPA 설명가능 인공지능(XAI) 프로젝트 일정(거닝 발표자료, 자료집 82쪽)

40) 적용 대상인 조항의 구체적인 내용에 관하여는 이원태, “EU의 알고리즘 규제 이슈와 정책적 시사점”, KISDI Premium Report 16-12 (2016), 19-20쪽 참조.

41) 이에 대하여 설명하고 있는 국내 문헌으로 한지연·최재식, “설명가능 인공지능”, 소음·진동 제27권 제6호 (2017), 10쪽 이하 참조.

42) David Gunning, “Explainable artificial intelligence (XAI)”, Defense Advanced Research Projects Agency (DARPA) nd Web (2017). 해당 파일의 출처는 아래 사이트 참조.

<http://explainablesystems.comp.nus.edu.sg/wp-content/uploads/2018/03/XAI%20for%20IUI%202018.pdf>

2. 패널 토론

제2세션의 패널은 발표를 맡은 거닝 외에, 우드로우 하트조그(Woodrow Hartzog),⁴³⁾ 브렌트 미텔슈타트(Brent Mittelstadt),⁴⁴⁾ 수 지앙(Su Jiang)⁴⁵⁾ 교수와, 구글 소속 제이크 루치(Jake Lucchi)로 구성되었다. 먼저 하트조그는 불법행위법 교수로서, 인공지능의 책임소재에 대한 난점을 인정한 뒤, 로봇공학자들과 협력하여 한 연구 두 편에 대해 소개하였다.⁴⁶⁾ 우선 하트조그는 맥락에 따른 인공지능의 작동과 행동을 기술하는 데 한계가 존재할 수밖에 없음을 인정하였다. 그렇지만 오류(failure)를 저지르게 마련인 인공지능 시스템 개발 과정에서 기술적 구체화에 대한 책임은 필연적으로 요구되고, 때문에 개발자, 공급자, 이용자들이 상호 검사와 교육에 협력하여야 할 모종의 의무를 상정해볼 수 있다고 주장하였다. 인간의 언어가 지닌 불명확성과, 시스템의 복잡성에 의해 인공지능의 모든 행위양태를 상세히 이해하고 기술한다는 것은 사실상 불가능한 일이다. 인간의 행위는 구문론(syntactic)과 의미론(semantic)적인 특성을 종합적으로 동원할 때 비로소 온전히 기술하고 구현할 수 있다는 점에서, 이를 온전히 극복하지 못한 현재의 자동화된 시스템은 언제나 오류에 노출되어 있고, 예측가능성 측면에서 근본적 한계에 마주하고 있다고 평가할 수 있다.

하트조그는 인공지능 시스템이 범할 수 있는 오류를 크게 구문론적 오류, 의미론적 오류, 검사상의 오류, 경고상의 오류 네 가지로 구분한다. 그리고 앞서 언급한 것처럼 이러한 오류의 가장 근본적 원인은 인공지능 알고리즘 제작과 학습 과정에 참여하는 이해관계자 사이의 교류가 부족하기 때문이고, 이들이 네 가지 오류를 방지하기 위한 검사 과정에 참여하여 상호 교육을 하는 경우 오류를 줄일 수 있다고 보았다. 일례로, 개발자나 공급자가 충분하다고 생각하는 경고 방식이 이용자들이 보기에는 별 도움이 되지 않는 것일 수 있다. 검사와 교육을 통해 표준화된 절차나 규칙을 제정하는 것도 가능한 방안이다. 이와 같은 과정을 거쳐 책임 주체와 범위를 결정하였을 때, 진정한 의미에서의 윤리성과 책임성을 구현한 것으로 볼 수 있을 것이다.

43) 노스이스턴대학교 교수.

44) 옥스퍼드대학교 연구원.

45) 북경대학교 교수.

46) Cindy M. et al., “An education model of reasonable and Good-Faith Effort for autonomous systems” (2017); William D. Smart et al., “An education theory of fault for autonomous systems”, Proceedings of We Robot (2017).

이어 미텔슈타트는 앞서 거닝이 발표하였던 설명가능 인공지능의 방법론 중 하나인 반사실적(counterfactual) 설명에 대하여 발표를 하였다. 미텔슈타트는 현행 GDPR의 설명을 요구할 권리는 초안에서와 달리 구속력을 상실했다는 점을 언급하는 한편,⁴⁷⁾ 설명 설명을 요구할 권리의 구속력을 인정한다고 하더라도 진정한 설명이 무엇인지에 대한 논의가 추가로 필요하다는 점을 지적했다. 미텔슈타트에 따르면, 모델의 일반적 작동원리를 설명한다는 것은 결국 청자를 고려하여야 하는 것이다. 한 연구에 따르면 인간은 대조적 방식의 설명을 잘 이해한다고 한다.⁴⁸⁾ 반사실적 설명이라고 함은 이런 대조적 방식을 활용한 설명이다. 예를 들어, 은행이 어떤 사람의 대출 신청을 거절할 때, 대출이 승인될 수 있었던 조건(연 소득이 1만 달러 더 많았다면, 신청한 대출액이 2천만 원 더 적었다면 기타 등등)을 알려주는 방식이 대표적이다.⁴⁹⁾ 물론 그 조건은 무수히 많겠지만 청자에게 가장 의미 있는, 청자가 자신의 노력으로 뒤바꿀 수 있는 최소의 조건을 알려주는 것이다.

한편, 수 지양은 인공지능이 개인정보나 책임성과 같은 부분적 이슈뿐만이 아니라, 인간의 실존 자체를 위협하게 될 수 있다는 점을 지적하였다. 그리고 어떤 인공지능 프로젝트를 분석해본 결과 인류를 파멸시킬 가능성이 있는 경우에는 아예 프로젝트의 실행 자체를 사전에 금해야 한다고 보았다. 그렇지 않은 경우는, 앞서 언급한 발표를 종합하여 기술 기반 접근방식과 규제기반 접근방식을 혼용할 필요가 있다고 지금까지 논의를 정리하였다. 수 지양은 비행기에 실린 블랙박스과 같은 신기술을 통한 접근법 역시 일리가 있지만, 기본적으로 규제기반 접근방식에 보다 초점이 맞춰져야 한다고 보았다. 그 구체적인 방법론으로서, 수 지양은 최근에 마이크로소프트에서 발간한 『The Future Computed: Artificial Intelligence and its role in society』⁵⁰⁾에서 제안한 방법론을 소개하였다. 윤리적 프레임워크를 기본으로 하되, 규제당국과 각 섹터별로 인공지능 업체들이 긴밀하게 협력하고, 법적 규제는 전면에서 등장하는 대신 이러한 프레임워크를 보완하는 연성법(soft law)일 필요가 있다는 것이다. 인공지능 책임성은 종래와 같이 규제 중심적 하향식(top-down) 방법론으로는 달성하기 어렵고, 산업계, 규제당국, 법조인들의 상호 협력을

47) Sandra Wachter et al., “Why a right to explanation of automated decision-making does not exist in the general data protection regulation”. *International Data Privacy Law*, 7(2) (2017) 참조.

48) ‘좋은 설명’이 되기 위한 요건으로 Tim Miller, “Explanation in Artificial Intelligence: Insights from the Social Sciences”, arXiv:1706.07269v3 (2017), p. 6 참조.

49) Sandra Wachter et al., “Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR”, *Harvard Journal of Law & Technology*, 31(2) (2018) p. 5 이하.

50) <https://blogs.microsoft.com/blog/2018/01/17/future-computed-artificial-intelligence-role-society/> 참조.

통해 구축해가는 것이 되어야 한다는 주장이다.

마지막으로 루치는 기계번역(machine translation)과 이미지 판독과 같은 혁신적 신기술을 언급하는 동시에, 인공지능이 현실에 존재하는 불공평을 정교하게 반영해야 할지, 보정하여야 할지와 같은 어려운 물음이 제기되고 있다는 점을 지적했다. 루치는 구글이 인간과 인공지능이 협업하여 ‘인간 중심적 인공지능’을 구현할 수 있도록 하는 연구를 진행하고 있고, 이를 위한 원칙도 명확히 수립해둔 상황이라고 밝혔다.⁵¹⁾

III. 제3세션 – 블록체인을 포함한 신기술을 어떻게 규제할 것인가

(How to Regulate Blockchain and Other Emerging Technologies)

제3세션은 최근 국내외적으로 엄청난 관심을 끌었던 블록체인과 가상화폐를 비롯한 신기술에 대한 규제에 관련된 논의이다. 제3세션의 패널은 미셸 핑크(Michèle Finck) 막스 플랑크연구소 선임연구원, 더글라스 아너(Douglas Arner) 홍콩대학교 교수, 이상용 충남대학교 교수, 양현서 카카오 이사로 구성되었다. 핑크는 먼저 블록체인 기술과 그 작동 방식에 대해 간단히 소개했다. 블록체인(blockchain)은 추가만 가능한(append-only), 알고리즘의 합의로 유지되고 복수 노드(컴퓨터)에 저장되는 탈중앙화된 데이터베이스를 뜻한다.⁵²⁾ 분산형 구조인 블록체인의 특성상, 블록체인 내에서 이루어지는 작업이나 생성되는 데이터를 법적으로 어떻게 평가할 수 있는지가 쟁점으로 대두되고 있다. GDPR을 비롯한 각국 개인정보 보호법제 상에서는, ‘개인정보’ 요건을 충족하는지 여부에 따라 법적으로 큰 차이가 빚어지게 된다⁵³⁾. 특히 모두에게 개방된 퍼블릭·비허가형 블록체인의 경우에 그렇지 않은 프라이빗·허가형 블록체인에 비해 이와 같은 문제가 더욱 대두된다.

‘데이터의 자유로운 이동’과, ‘개인의 기본권 보호’ 두 가지의 이념을 기치로 내걸고 출범한 GDPR은 기본적으로 데이터의 저장에 중점으로 집중되는 형태를 전제로 삼고 있다.

51) <https://ai.google/research/teams/brain/pair> 참조.

52) ‘블록체인’에 대한 보다 자세한 설명으로는 Michèle Finck, “Blockchain Regulation”, Max Planck Institute for Innovation & Competition Research Paper No. 17-13 (2017), 2쪽 이하 참조.

53) ‘개인정보’ 요건의 핵심인 ‘식별성’을 제거하는 기술적 방법론인 ‘비식별화’에 관한 가이드라인으로 관계 부처 합동, “개인정보 비식별 조치 가이드라인 - 비식별 조치·기준 및 지원 관리체계 안내 -”, (2016) 참조.

그러나 블록체인은 이와 달리 데이터의 수집, 저장, 처리가 탈중앙화 되는 것을 전제로 삼고 있다. 문제는 GDPR에서 바라보는 ‘개인정보’가 매우 광범위하여, 웬만한 정보는 ‘개인정보’ 요건을 구비하게 되고, 비식별화 절차를 거쳐도 비가역적인 상태인 ‘익명화(anonymous)’가 아닌, 가역적인 상태인 ‘가명화(pseudonymous)’ 수준으로만 인정될 것으로 예상된다. 핑크가 볼 때는 거래 데이터(transaction data)와 공개키(public key)는 ‘개인정보’성이 부정되지 않고, 기껏해야 가명화 정보로 인정될 것이라고 한다.⁵⁴⁾ 그렇다면 GDPR에서 요구하는 권리와 의무를 누군가에게 지워야만 할 것인데, 탈중앙화를 표방한 블록체인 기술의 특성상 뚜렷한 ‘데이터 컨트롤러(data controller)’를 상정하기가 어려워졌다. 나아가 GDPR에 규정된 수정권, 잊혀질 권리 등을 기술적으로 구현할 수 있을지도 문제될 수 있다.⁵⁵⁾ 블록체인 기술이 국경을 넘나들며 이용되는 경우 이러한 문제는 보다 심화된다. 핑크는 또한 스마트 계약(smart contract)에서 현실적으로 사람의 개입이 일부 불가피할 수 있다는 점에서 GDPR 상의 ‘완전히 자동화된 의사결정(solely automated decision-making)’이라는 요건의 충족이 어려울 수 있다는 점도 지적하였다. 블록체인과 GDPR이 충돌한다는 주장은 바로 이러한 지점으로부터 기원한다.⁵⁶⁾ 다만, 핑크는 그렇다고 하더라도 블록체인은 정보주체의 통제권을 강화하는 특성을 가지고 있고, 이런 점에서 GDPR과 근본적으로 양립이 가능하다는 점을 강조했다. 그렇다면 블록체인 기술의 유효성을 포기할 것이 아니라, 신기술 발전과 법의 실효성 양자를 조화하는 방안을 강구하는 것이 온당할 것이다.

다음 발표자인 아너는 주로 블록체인이 금융 산업과 규제에 미친 영향에 주목했다. 아너는 지난 10년을 돌아볼 때, 비트코인은 2008년 글로벌 금융위기와 함께 등장한 산물이라는 점을 강조했다. 블록체인, 나아가 소위 ‘핀테크(Fintech)’라고 불리는 산업 전반이 글로벌 금융위기 이후 대안으로 부상하고 있다. 아너는 핀테크의 역사를 크게 4가지 키워드로 분석했다. (1) 고빈도 매매(High Frequent Trading), 지급결제(payment and settlement)를 비롯한 도매 금융업에서의 변화, (2) 스타트업의 부상으로 인한 R&D 활성화,

54) 구체적 논지는 Michèle Finck, “Blockchains and Data Protection in the European Union”, Max Planck Institute for Innovation & Competition Research Paper No. 18-01 (2017), p. 9 이하 참조.

55) 보다 자세한 내용은 박민정 외 2인, “개인정보보호법제 관점에서 본 블록체인의 법적 쟁점 : GDPR 및 국내 개인정보보호법을 바탕으로”, Journal of Information Technology Applications & Management 제25권 제2호 (2018) 등 참조.

56) 블록체인과 GDPR의 상충 여부에 대하여는 아직까지 확립된 견해는 존재하지 않고, 논쟁 중에 있다. 루카스 미리언, “‘블록체인과 GDPR’ 상극일까, 공생 관계일까?”, itworld (2018. 5. 9.). 등 참조.

(3) 수십 억 명에 달하는 인구를 지닌 개발도상국에서의 디지털 변혁, (4) 마지막으로 개인 정보 보호법, 경쟁법, 소비자법과 같은 여러 실정법 영역의 교차점에 위치한다는 점이다.⁵⁷⁾ 그렇다면 핀테크를 바라보는 금융당국의 관점은 어떠할까?

금융규제는 금융위기를 방지하고 금융시장의 안전성을 도모하는 역할도 담당하여야 하지만, 동시에 경제성장과 금융 포용성을 도모하는 역할도 담당하고 있다. 그렇지만 2008년 글로벌 금융위기가 금융당국의 시각을 보수적인 방향으로 급격히 전환시켰고, 이에 따라 핀테크 산업이나 비트코인, 블록체인 등 신기술에 대하여도 부정적 태도가 부각되었다. 심지어 폴 볼커(Paul Volcker) 전 미국 연방준비제도이사회(Board of Governors of the Federal Reserve System, FRB) 의장은 “유일하게 쓸모 있는 금융혁신은 ATM 뿐”이라고 할 정도였다. 아너에 따르면 비트코인을 보안성을 구비한 전자지급 시스템이라고 선전하였다면 강한 규제를 받지 않았겠지만, 중앙은행 역할을 대체할 것이라고 선전하는 바람에 엄청난 규제를 받게 되었다고 한다. 중국이 최근에 암호화폐를 강력하게 규제하는 것도 같은 맥락에서 이해할 수 있다.

그렇다면 블록체인에 대하여는 어떠한 입장을 취해야 할까? 아너는 블록체인이라고 하여 다 같은 맥락에서 활용되는 것이 아니라는 점을 먼저 강조한다. 활용된 유형이 퍼블릭, 프라이빗 중 어느 것인지를 살펴보아야 하고, 어느 용도로 활용되는지 여부도 중요하다. 일례로 ICO (Initial Coin Offering)에 적용되는 블록체인⁵⁸⁾과 그렇지 않은 블록체인에 대한 금융당국의 규제태도가 같을 수는 없을 것이다. 아너는 이러한 점을 고려하여 앞으로도 보다 많은 고민이 필요하다고 역설하면서 온라인 핀테크 강좌⁵⁹⁾를 소개한 뒤 발표를 마무리하였다.

이어 이상용 교수는 작업증명(proof of work)이나 스마트 계약(smart contract)과 같은 핵심 기술을 열거하면서, 블록체인은 과거부터 존재해온 기술들을 조합함으로써 일견 상충된 듯이 보이는 ‘개방성’과 ‘신뢰성’ 두 가지 가치를 동시에 달성할 수 있는 획기적 기술이

57) 핀테크의 역사는 Douglas W. Arner et al., “The Evolution of FinTech: A New Post-Crisis Paradigm?”, UNSW Law Research Paper No. 2016-62 (2016) 참조.

58) ICO에 관한 보다 자세한 내용은 성희환, “가상화폐의 공모(ICO)와 상장에 대한 적정 규제방안”, 상사법 연구 제37권 제1호 (2018) 등 참조.

59) <https://www.edx.org/course/introduction-to-fintech> 참조.

라는 점을 강조하였다. 불변성, 투명성, 분산성과 같은 블록체인의 특성 덕에 ‘신뢰할 수 있는 제3자(trusted third party)’로 불리는 중개자 없이도 당사자는 신뢰를 확보할 수 있게 된다. 이는 결과적으로 거래비용의 감소를 가져온다. 그렇지만 블록체인이 종래 법체계를 완전히 대체할 것이라는 등의 과도한 환상과, 그 반대로 블록체인을 범죄의 수단으로 바라보는 등의 과도한 공포가 확산되는 것을 경계해야 한다고 강조하였다. 이러한 근거 없는 환상이나 공포는 관련 법제도의 정비를 늦추어 법적 불확실성을 높이는 한편, 바람직한 산업 발전을 저해하는 결과를 낳는 것이다.⁶⁰⁾

이상용 교수는 블록체인 기술에 관련된 법적 쟁점을 분석하고 개선방향을 찾아 나아가는 것이 필요하다고 하면서, ‘기본기술 구현’과 ‘서비스 제공’ 두 단계로 구분하면서 논의를 이어나갔다. 먼저 기본기술 구현 단계에서 가장 문제되는 쟁점은 개인정보 및 법적 책임 이슈라고 보았다. 특히 블록체인 기술의 특성상 다수의 당사자가 참여하게 마련이므로 책임의 분배를 어떻게 해야 할지가 향후 중요한 과제로 남게 될 것이라고 바라보았다. 한편, 서비스 제공 단계에서는 ‘비트코인’으로 대표되는 암호화폐(cryptocurrency) 문제를 가장 먼저 거론하였다. 일단 암호화폐의 법적 성질을 어떻게 보는지에 따라서 과세 대상 여부나 어떤 종류의 과세가 적용되는지 여부가 달라지고,⁶¹⁾ 강제집행 대상 여부도 달라진다.⁶²⁾ 이때 암호화폐는 발행자가 존재하지 않으므로, 법적 규제는 주로 거래소에 집중되는 양상을 보인다. 단체법적으로는 DAO, ICO가 주로 문제된다. 탈중앙화된 자율조직(Decentralized Autonomous Organization, DAO)으로 불리는 DAO의 경우, 소위 ‘The DAO’ 사건⁶³⁾에서 문제된 해킹뿐만 아니라 회사법과 유사한 특성에도 불구하고 회사법을 그대로 적용하기는 어렵다는 점에서 법적 공백의 위험이 제기되고 있다. ICO의 경우, 자금조달의 효율성을 극대화할 수 있지만 투자자 보호의 측면에서는 보완이 필요하다고 지적했다. 이상용 교수는 ICO는 발행자가 명확하므로 보다 집중화된 규제가 가능해 ‘자본시장과 금융투자업에 관한 법률’의 조속한 개정이 필요하다고 보았다.

60) ‘혁신’과 ‘이익’ 보호를 겸비한 규제 틀 마련의 필요성에 관하여는, 최경진, “인공지능과 데이터, 그리고 법”, 『데이터 이코노미』, 한스미디어 (2017), 263쪽 이하 참조.

61) 올해 정부의 세법 개정안에 의하면 암호화폐 과세는 연기되었고, 거래소의 중소기업 세액감면은 제외되었다(기획재정부, “2018년 세법개정안”, 9쪽 참조).

62) 참고로 최근 대법원에서는 비트코인은 재산적 가치가 있는 무형의 재산이고, 특정되어 있다는 근거를 들면서 ‘범죄수익은닉의 규제 및 처벌 등에 관한 법률’에서 정한 ‘물수’의 대상이 된다고 판단하였다(대법원 2018. 5. 30. 선고 2018도3619 판결).

63) The DAO 사건의 자세한 내용은 박재현, “암호화폐, 알트코인 이야기”, 지디넷코리아 (2017. 10. 10.) 참조.

또 다른 중요한 서비스인 스마트 계약의 경우 개념 자체는 오래 전부터 있었지만 이더리움에 의해 구체화되어, 오랜 법언인 ‘계약은 반드시 지켜져야 한다(pacta sunt servanda)’는 격언을 보장하는 기술로 부각되고 있다. 인공지능이 계약 체결의 자동화를 이끌고 있다면, 스마트 계약은 계약 이행의 자동화를 이끌고 있다는 것이다. 때문에 분쟁에 대한 초점도 계약의 이행보다는 계약의 성립 단계에 맞추어지게 된다. 한편, 스마트 계약의 출현에 따라 코드 그 자체가 곧바로 계약에 해당하는지, 전통적인 법적 관점에서 계약은 단지 의사의 합치로 이루어지는 것일 뿐인지 등 다양한 이슈에 대하여 견해가 대립하고 있다. 이 문제는 이론적 차원에서 그치지 않고 실질적 차이를 빚게 되는데, 대표적 사례가 앞서 언급한 The DAO 해킹 사건이다. 이더리움 측은 코드 그 자체를 계약으로 보는 대신에 해킹을 사기행위로 판단하였고, 그에 따라 거래를 소급적으로 무효화하는 일명 ‘하드 포크(hard fork)’를 단행하게 된 것이다. 이상용 교수는 보다 현실적 대안으로서 기존과 반대 방향의 소위 ‘역 거래(reverse trading)’를 상정할 수 있지만, 그 사이에 지급불능 상태에 빠질 수 있는 위험성이 있음을 함께 언급하였다. 이행 단계에서도 법적 쟁점은 여전히 존재하는데 예컨대 계속적 거래에서 사정변경의 원칙이 작동하거나, 미국의 유명한 다크 웹(dark web) ‘실크로드(SilkRoad)’ 사건과 같이 공법적 규제가 필요한 사안을 떠올려볼 수 있겠다.⁶⁴⁾

마지막으로 양현서 이사는 국제적으로 널리 인정받고 있는 블록체인의 가능성에도 불구하고, 국내에서는 암호화폐에만 과도한 관심이 몰린 점을 지적하였다. 이에 따라, 대한민국 정부는 블록체인 기술은 장려하면서 일종의 유틸리티 역할을 하는 암호화폐는 규제하는 애매한 정책을 펼치고 있다고 보았다. 양현서 이사는 스마트하지 못한 규제를 고수한다면 기술과 인재, 결과적으로 국부를 해외에 유출하는 결과를 낳게 될 것이라고 지적했다. 대표적 사례로 현재 대한민국 기업이 해외에서 ICO를 진행하면서 해당 국가에 세금을 납부하고 있다는 점을 들었다. 블록체인 산업은 고용 창출효과가 뛰어난 것으로 평가받고 있기에 대한민국 전반을 위협하고 있는 일자리 문제의 해결에 기여할 것이라고 덧붙이기도 하였다. 따라서 정부의 역할은 신기술을 무조건 금지하는 것이 아니라, 시대의 변화를 고려한 네거티브 규제나 규제 샌드박스와 같은 스마트 규제(smart regulation)를 고안하는 것이어야 한다고 주장하였다.⁶⁵⁾

64) 이상의 쟁점에 관한 심층적 연구로는 김제완, “블록체인 기술의 계약법 적용상의 쟁점 - ‘스마트계약(Smart Contract)’을 중심으로 -”, 법조 제67권 제1호 (2018) 등 참조.

Ⅰ 참고문헌

1. 국내문헌

고학수 외, 『핀테크 시대』, 박영사 (2015).

고학수 외 20인, 『개인정보 보호의 법과 정책(개정판)』, 박영사 (2016).

고학수 외 4인, 『개인정보 비식별화 방법론 - 보건의료정보를 중심으로 -』, 박영사 (2017).

고학수 외 9인, 『데이터 이코노미』, 한스미디어 (2017).

마쓰오 유타카, 박기원 역, 『인공지능과 딥러닝』, 동아엠앤비 (2015).

서울대학교 아시아태평양법 연구소 · 서울대학교 법과경제연구센터, 『인공지능, 알고리즘, 개인정보보호를 둘러싼 정책적 과제』 (2017).

서울대학교 아시아태평양법 연구소 · 서울대학교 법과경제연구센터, 『인공지능의 시대: 기술 발전에 따른 책임과 규제』 (2018).

스튜어트 러셀 · 피터 노빅, 류광 역, 『인공지능: 현대적 접근방식(제3판)』, 제이펍 (2016).

이상돈 외 6인, 『인공지능과 자율자동차, 그리고 법』, 세창출판사 (2017).

프랭크 파스칼레, 이시은 역, 『블랙박스 사회』, 안티고네 (2016).

공정거래위원회, “공정거래법제 개선 특별위원회 최종 보고서” (2018).

과학기술정보통신부, “I-Korea 4.0 실현을 위한 인공지능(AI) R&D 전략” (2018).

관계부처 합동, “개인정보 비식별 조치 가이드라인 - 비식별 조치 · 기준 및 지원 관리체계 안내 -” (2016).

관계부처 합동, “제4차 산업혁명에 대응한 지능정보사회 중장기 종합대책” (2017).

구본권, “로봇 도입률 1위 한국은 자동화에 더 안전할까?”, 한겨레 (2017. 3. 17.).

권영준, “계약법의 사상적 기초와 그 시사점”, 저스티스 통권 제124호 (2011).

기획재정부, “2018년 세법개정안” (2018).

김병은, “인공지능 동향분석과 국가차원 정책제언”, 정보화정책 제23권 제1호 (2016).

김연지, ““킬러 로봇 우려” 해외학자 57인, KAIST 보이콧 선언 ‘철회’”, 노컷뉴스 (2018. 4. 9.).

김제완, “블록체인 기술의 계약법 적용상의 쟁점 - ‘스마트계약(Smart Contract)’을 중심으로 -”, 법조 제67권 제1호 (2018).

65) 스마트 금융규제체계에 관한 구체적 내용은 성희환, “4차 산업혁명의 시대에서 「네거티브규제 패러다임」에 따른 금융규제체계의 재구축 방안 연구”, 법과 정책 제24권 제1호 (2018) 등 참조.

- 루카스 미리언, “블록체인과 GDPR’ 상극일까, 공생 관계일까?”, itwolrd (2018. 5. 9.).
- 박민정 외 2인, “개인정보보호법제 관점에서 본 블록체인의 법적 쟁점 : GDPR 및 국내 개인정보보호법을 바탕으로”, Journal of Information Technology Applications & Management 제25권 제2호 (2018).
- 박재현, “암호화폐, 알트코인 이야기”, 지디넷코리아 (2017. 10. 10.).
- 성희활, “4차 산업혁명의 시대에서 「네거티브규제 패러다임」에 따른 금융규제체계의 재구축 방안 연구”, 법과 정책 제24권 제1호 (2018).
- 성희활, “가상화폐의 공모(ICO)와 상장에 대한 걱정 규제방안”, 상사법연구 제37권 제1호 (2018).
- 이상용, “인공지능과 계약법 - 인공 에이전트에 의한 계약과 사적자치의 원칙 -”, 비교사법 제23권 4호 (2016).
- 이원태, “EU의 알고리즘 규제 이슈와 정책적 시사점”, KISDI Premium Report 16-12 (2016).
- 조일규, “인공지능(AI) 기술을 선도하는 미국, 중국, 한국의 최근 특허 및 논문 활동 조사”, 주간기술동향 (2017).
- 한국연구재단, “R&D 통계 핸드북 2018” (2018).
- 한지연·최재식, “설명가능 인공지능”, 소음·진동 제27권 제6호 (2017).
- 한희원, “인공지능(AI) 치명적자율무기(LAWs)의 법적·윤리적 쟁점에 대한 기초연구”, 중앙법학 제20권 제1호 (2018).

2. 해외문헌

- Alex Campolo et al., “AI Now 2017 Report” (2017).
- A. M. Turing, “Computing Machinery and Intelligence”, Mind, New Series, Vol. 59, No. 236 (1950).
- Catherine Tucker, “Privacy, Algorithms and Artificial Intelligence”, The Economics of Artificial Intelligence: An Agenda (2017).
- Cindy M. et al., “An education model of reasonable and Good-Faith Effort for autonomous systems” (2017).
- David Gunning, “Explainable artificial intelligence (XAI)”, Defense Advanced Research Projects Agency (DARPA) nd Web (2017).

- Douglas W. Arner et al., “The Evolution of FinTech: A New Post-Crisis Paradigm?”, UNSW Law Research Paper No. 2016-62 (2016).
- Executive Office of the President, “Big Data: Seizing Opportunities, Preserving Values” (2014).
- Executive Office of the President, “Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights” (2016).
- Executive Office of the President, “Preparing for the future of AI” (2016).
- Executive Office of the President, “Artificial Intelligence, Automation, and the Economy” (2016).
- Federal Trade Commission, “Big Data: A Tool for Inclusion or Exclusion” (2016).
- Jatinder Singh et al., “Responsibility & Machine Learning: Part of a Process”, (2016).
- Jenna Burrell, “How the machine ‘thinks’: Understanding opacity in machine learning algorithms”, *Big Data & Society* (2016).
- John R. Searle, “Minds, brains, and programs”, *Behavioral and Brain Sciences*, 3(3) (1980).
- Kate Crawford · Meredith Whittaker, “The AI Now Report” (2016).
- Michèle Finck, “Blockchain Regulation”, Max Planck Institute for Innovation & Competition Research Paper No. 17-13 (2017).
- Michèle Finck, “Blockchains and Data Protection in the European Union”, Max Planck Institute for Innovation & Competition Research Paper No. 18-01 (2017).
- Michelle Jamrisko · Wei Lu, “The U.S. Drops Out of the Top 10 in Innovation Ranking”, *Bloomberg* (2018. 1. 23.).
- Peter Stone et al., “Artificial Intelligence and Life in 2030”, *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*, Stanford University, Stanford, CA (2016).
- Sandra Wachter et al., “Why a right to explanation of automated decision-making does not exist in the general data protection regulation”. *International Data Privacy Law*, 7(2) (2017).
- Sandra Wachter et al., “Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR”, *Harvard Journal of Law & Technology*, 31(2) (2018).
- Solon Barocas · Andrew D. Selbst, “Big Data’s Disparate Impact”, *104 California Law*

- Review 671 (2016).
- Sophie Curtis, “Google Photos labels black people as ‘gorillas’”, The Telegraph (2015. 7. 1.).
- The White House, “Consumer Data Privacy in a Networked World: A framework for protecting privacy and promoting innovation in the global digital economy” (2012).
- Tim Miller, “Explanation in Artificial Intelligence: Insights from the Social Sciences”, arXiv:1706.07269v3 (2017).
- Warren S. McCulloch · Walter H. Pitts, “A Logical Calculus of the Ideas Immanent in Nervous Activity”, Bulletin of Mathematical Biophysics, 5 (1943).
- William D. Smart et al., “An education theory of fault for autonomous systems”, Proceedings of We Robot (2017).
- Yoav Shoham et al., “Artificial Intelligence Index 2017 Annual Report” (2017).

준비한 사람들 & 도와준 사람들

서울대학교 인공지능 정책 이니셔티브 공동디렉터 : 고학수 & 임 용

행사준비 실무총괄 : 김태훈

행사준비 실무 : 김은수

자료집 정리 및 행사보고서 작성 : 박도현

번역 및 현장 안내 : 장흥원, 김시온(9기),

강성구, 박유준, 강명지, 최정찬, 박제연, 김시온(10기),

유경하, 류상우(이상 서울대학교 법학전문대학원 인공지능법학회),

김혜인, 강승우, 간결(이상 서울대학교 법학대학원)

동영상 제작 : 이지현, 박재승, 윤건희, 조영채(이상 서울대학교 자율전공학부)

동시통역 : 천지은, 손선희

