서울대학교
인공지능정책
이니셔티브

인공지능정책
국제 컨퍼런스
2021

신뢰할 수 있는 데이터 생태계 보고서

DATA and TRUST:
Analytics, Ethics & Privacy Report

2021. 11. 30. — 12. 15.

서울대학교 인공지능 정책
이니셔티브 2021
국제 컨퍼런스 보고서

진행: 고학수(서울대학교), 임 용(서울대학교), 박상철(서울대학교)
보고서 작성: 이혜승, 조은정, 김연준

목차

04 머리말

06 제1세션 – 컴퓨팅의 관점에서 바라본 사회적 규범, 상식과 윤리
최예진

14 제2세션 – 신뢰가능한 AI: 법의 역할은 무엇인가 ?
D. Daniel Sokol

22 제3세션 – 디지털 시대의 경쟁과 프라이버시
Ariel Ezrachi

32 제4세션 – AI의 아틀라스: 파워, 정치,
그리고 지구가 지불해야 할 비용을 그리다
Kate Crawford

40 제5세션 – 인공지능 시대의 폭풍속으로:
스탠포드 « AI 100년 » 연구 프로젝트 보고(2021)
Michael L. Littman

52 제6세션 – 머신러닝 기반 경쟁법 집행의 분석 – 현황과 미래
Ashwin Ittoo

64 제7세션 – 인간 행위를 위한 알고리즘 설명책임
김현섭

72 제8세션 – 의료 인공지능의 법적 쟁점
박혜진

신뢰할 수 있는 데이터 생태계

DATA and TRUST: Analytics, Ethics & Privacy

산업 영역에 인공지능 기술의 응용이 증가하고 시장의 생태계 모습도 변화되면서, 시장에서 사회 구성원의 의사결정을 돕기 위하여 형성되어야 하는 '신뢰'에 대한 논의가 요구되고 있다. 인공지능 시대에도 정보를 교환하고 효율적으로 자원을 분배하는 시장의 기능이 제대로 작동하기 위해서 필요한 신뢰란 어떤 모습인지에 대한 법과 정책을 모색하기 위하여 서울대학교 인공지능 정책 이니셔티브(SNU AI Policy Initiative, SAPI)는 '신뢰할 수 있는 데이터 생태계'라는 주제로 전문가와 함께 이슈를 다루는 컨퍼런스를 개최하였다.

인공지능이 갖는 사회적 · 정치적 분야에서 전세계적으로 인정받는 Kate Crawford 교수의 '인공지능 아틀라스(Atlas of AI)'를 비롯하여, 규범의 측면에서 논의되는 신뢰성에 대한 문제, 그리고 시장 경쟁과 개인의 프라이버시 등과 같은 이슈부터 구체적인 컴퓨터 공학의 관점에서 인공지능에 윤리와 규범, 상식을 학습시키는 연구에 이르기까지 각 분야에서 신뢰할 수 있는 생태계를 위하여 어떤 노력과 고민이 진행되고 있는지를 볼 수 있는 컨퍼런스로 구성되었다.

제1세션

컴퓨팅의 관점에서 바라본 사회적 규범, 상식과 윤리

Computational Social Commonsense, Norms, and Ethics

발표자 **최예진**

Paul G. Allen School of Computer Science & Engineering
University of Washington & Allen Institute for Artificial Intelligence

발표자
최예진

제1세션

첫 번째 세션에서는 최예진 교수가 2021년 10월에 발표한 논문 'Delphi : Towards Machine Ethics and Norms'를 중심으로 기계에게 특정 수준의 규범과 윤리를 학습시키는 것에 관한 연구와 그 배경을 논의하는 시간을 가질 수 있었다. Delphi는 일상에서 제기될 수 있는 윤리적인 질문에 대하여 대답할 수 있도록 훈련한 딥러닝 모델로, 기계와 인간 사이의 정보를 이해하는 능력의 차이를 줄이기 위한 방법을 찾고자 시작된 연구이다.
현 단계에서는 GPT-3과 같이 성능이 뛰어난 언어모델조차 아직도 문자 그대로의 의미 이상의 정보를 추론하고 이해할 수 있는 능력을 갖추지 못한 것으로 보이기에, 인간이라면 대부분의 경우 문제가 될 수 있는 편향성 발언도 기계로부터 도출될 수 있다. 기계는 아직까지는 윤리적이고 상식적인 것이 무엇인지에 대한 이해가 없기 때문이다. 그 예로 실제로 있었던 사례인 '치즈버거 찌르기 사건(Cheeseburger Stabbing)'이라는 뉴스의 헤드라인을 읽었을 때, 사람이라면 그 뜻을 이해하기 위하여 각종 상식과 같은 정보, 즉 '버거로 사람을 찌를 수는 없는데 ?' 내지 '누군가를 찌르는 것은 비윤리적이야' 등의 정보에 근거하여 아마도 치즈버거를 두고 칼로 찌르는 사건이 일어났을 것이라는 정보를 추론할 수 있는 반면에 현재 수준에서의 기계는 이와 같은 능력이 결여되어 있다.

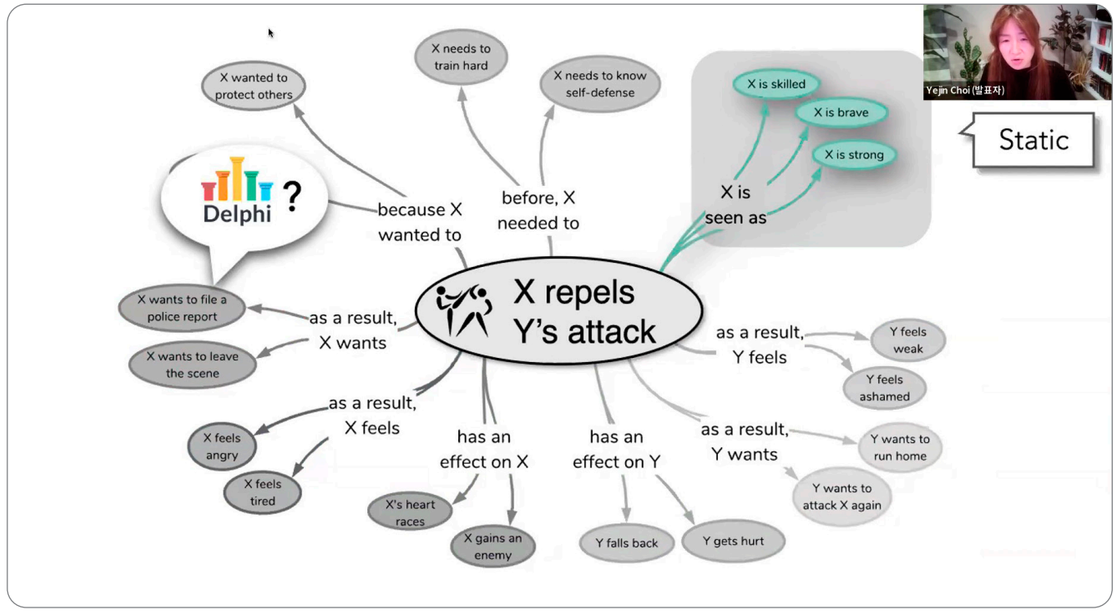
시간이 지날수록 더욱 많은 의사결정 과정에서 인공지능(AI) 기술이 적용되고, 기계와 사람처럼 언어를 이해하여 무엇이 상식적이고 윤리적인지에 대해 판단할 수 있는 능력 또는 추론해낼 수 있는 능력을 갖추는 것이 중요하다고 봄에 따라 기계에게 상식 또는 윤리 규범을 학습시키는 연구가 이루어졌다. 이에 대하여 70-80년대 이미 이와 유사한 연구의 시도가 있었으나 실패로 끝난 점을 들어 회의적인 반응이 다수 있었지만, 최근 컴퓨팅 파워가 중요해지고 기하급수적으로 데이터가 늘어나고 있으며 뉴럴넷(neuralnet)과 같은 기술이 빠른 속도로 발전하고 있어 본 연구의 가능성과 중요성이 조명을 받고 있다.

기계에게 상식/윤리를 학습시키기 위해서는 우선적으로, 각종 상황이 발생하였을 때 그 상황의 원인과 결과(causes and effects)가 무엇인지

학습시키는 작업이 필요하다. 최예진 교수 연구팀은 Delphi 연구에 앞서 이를 훈련하는 작업 먼저 진행하였는데, 우선 ATOMIC(Atlas of Machine Commonsense For If-Then Reasoning)과 같은 일상에서 발생할 수 있는 다양한 상황과 그 상황의 원인, 결과 등을 포함한 지식 그래프(knowledge graph) 데이터베이스를 생성하였고, 이에 대한 어노테이션(annotation)은 Amazon Mechanical Turks를 활용하였다.

사전에 학습된 언어모델을 ATOMIC 데이터에 학습시킨 것이 바로 COMeT(Commonsense Transformers for Automatic knowledge Graph Construction)이다. COMeT는 이전에 경험하지 않은 상황 데이터(ATOMIC events)에 대하여도 나름의 추론(reasoning)이 이루어지는 것을 볼 수 있었다. 이를 기반으로 하여 인공지능에 기술윤리(descriptive ethics)를 학습시키는 데 집중하는 연구를 수행하고자 하였고, 그 첫 걸음으로 Delphi가 나올 수 있었다. Delphi를 훈련하는데 사용된 데이터는 Commonsense Norm Bank의 총 170만 건에 달하는 일상 생활 내 상식과 윤리적 판단에 관한 예제가 담긴 데이터셋이다. 이 데이터셋은 기존의 다섯가지 데이터셋을 합친 것으로, 여기에는 ① Social chemistry 101, ② ETHICS Commonsense Morality, ③ Moral Stories, ④ Sopcial Bias Inference Corpus, ⑤ Scruples 가 있다.

여러 윤리적 상황에서 Delphi는 GPT-3과 비교하였을 때도 높은 성능을 보였고, UN 세계인권선언에 대한 성능 테스트에서도 98.1% 비율로 Delphi가 상식의 수준에서 예상 가능한 답변을 도출하였다. 그럼에도 서비스 이용자들이 Delphi가 비윤리적인 답을 도출하는 사례를 찾아 문제제기를 하였기에 지난 10월 저자들은 질의응답(Q&A) 서비스를 중단하였다.
이에 최예진 교수는 Delphi는 완벽한 모델이 아닌 연구 프로토타입임을 강조하며, 특정 윤리적 상황들은 AI 시스템의 역량 범위의 밖에 존재하는 것을 인정하면서도 궁극적으로는 인간의 추론 능력과 AI의 추론 능력 간의 간극을 좁히는 것을 목표로 삼아 연구를 지속할 것이라 설명하였다.
이후 이어진 질의응답에서 ATOMIC과 연구의 향후 응용 등에 대한



최예진 교수 발표자료: Delphi 훈련 데이터

질문이 이어졌고 이에 대한 최예진 교수의 답변이 이루어졌다. 먼저, ATOMIC에서 라벨링되지 않은 텍스트(unlabeled text)를 사용하지 않은 주된 이유로 보고편향(reporting bias)으로 지적하였는데, 그 외에도 비지도학습(unsupervised learning)에 의존했을 때 정확도(accuracy)의 확보가 어려운 것도 중요한 이유로 고려되었는지 그리고 연구과정에서 어노테이션 작업의 정확도를 확보하기 위하여 어떠한 조치가 있었는지에 대한 질문이 있었다. 최예진 교수는 정확도의 문제로 인해 자동으로 지식 베이스(knowledge base)를 형성하는 것이 어려워서 첫 번째 단계의 연구는 사람의 라벨링에 의존하였다고 설명하고, 이와 관련하여 최근 2021년 진행된 연구를 설명하였다. 해당 연구에서는 기계만으로 라벨링한 성능이 사람이 한 것보다 더 높도록 알고리즘을 통해 구현한 연구인데, 사람의 경우 작업을 하면서도 집중이 되지 않는 순간들도 있어 정보의 질에 대한 통제(quality

control)가 어려워 사례비를 높이는 등의 방법을 활용하여도 노이즈가 존재할 수 있다고 설명하였다. 그러나 최근 뉴럴넷(neural network)에서는 데이터의 양이 많으면 노이즈도 잘 극복될 수 있다고 덧붙였다.

이어서 법률 조항이나 법 체계를 지식 그래프화하여 Commonsense norm bank를 확장하는 것이 가능할지, 그렇다면 인공지능 판사 혹인 배심원을 마련하여 사람을 대신하는 것 또한 가능할지에 대한 발표자의 의견을 묻는 질문이 있었다. 이에 대하여 최예진 교수는 언급한 방법으로 Commonsense norm bank를 확장하는 것은 가능하다고 보았으나, 이를 통해 학습된 인공지능 판사 혹은 배심원의 활용에 대하여 단순한 법률 판단에는 활용될 수 있을지 모르겠으나, 인간 판사 혹은 배심원을 완전히 대체하는 것은 윤리적으로나 법적으로 보았을 때 동의할 수 없다는 의견을 밝혔다.

또한 Delphi를 한국어 또는 다른 언어로도 만들 수 있는지에 대하여, 각 나라마다 서로 다른 규범을 가지고 있기 때문에 각 언어로 된 Commonsense norm bank가 필요하다고 보았다. 현재 영어로 된 Commonsense norm bank의 어노테이션을 위해서는 Amazon의 Mechanical Turk 소속 인력을 활용하였는데, 해당 인력의 대부분이 미국 내 거주하는 관계로 미국의 문화적 규범이 반영되어 있다고 볼 수 있어, 결국 국가별로 Norm bank가 맞춰져야 할 것이라고 보았다.

제2세션

신뢰 가능한 AI: 법의 역할은 무엇인가?

Law’s role in creating Trust in AI

발표자 **D. Daniel Sokol**

University of Southern California Gould School of Law

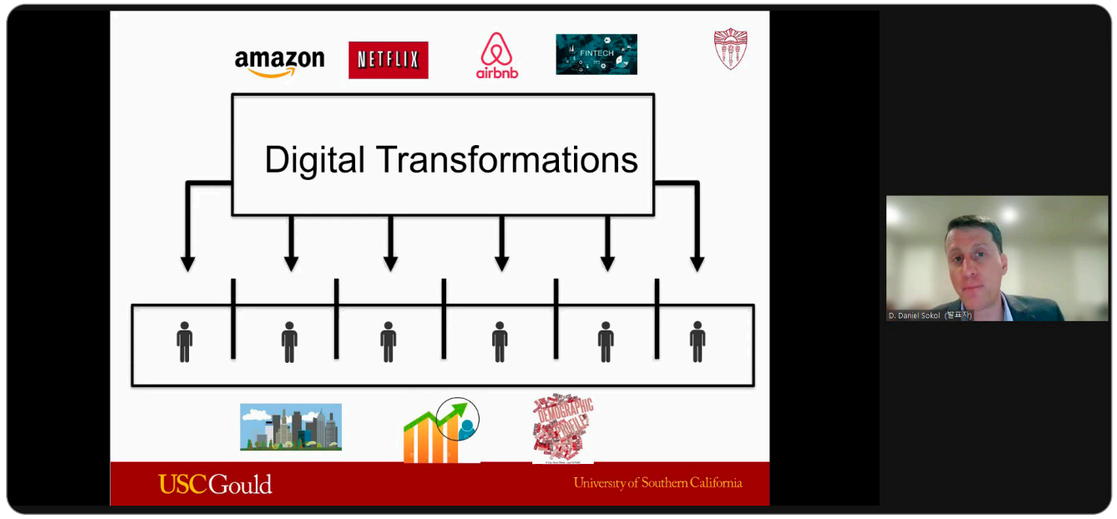
발표자

D. Daniel Sokol

산업 전반에 걸쳐 디지털 전환(Digital transformation)이 이루어지면서 기존의 상품 또는 서비스의 공급 과정(supply chain)의 변화가 진행 중인 시점에서 그 변화에 상응하는 '신뢰(Trust)'를 어떻게 형성해야 하는지, 그리고 이를 위해 법과 규제의 역할은 무엇인지에 대한 논의가 이루어졌다. 우선 산업의 디지털 전환과 법이 어떻게 관련되어 있는지에 관하여, Sokol 교수는 법을 통하여 신뢰에 기반한 디지털 전환을 촉진시키는데 기여할 수 있는데, 만일 신뢰가 전제되지 않으면 사회의 주요 서비스(예 : 금융서비스)들이 원활하게 작동하지 않기에 법과 규제로 신뢰를 구축하는 것은 중요하게 논의되어야 한다고 강조하였다.

인공지능 기술이 발전하고 경제 활동에 더 많이 적용되면서 그로 인한 우려도 증가하는데 일례로 개인의 프라이버시 침해 우려와 인공지능 기술이 제공한 예측의 정확성(accuracy)에 관한 문제 등이 있고, 인공지능에 대한 신뢰와 관련된 구체적 문제로는 주로 수집된 데이터의 활용 양태, 활용하는 데이터의 성격 및 인공지능 정보 처리 모델을 구축함에 있어서 잠재적 편향(potential bias)에 대한 고민이 반영되었는지 여부 등과 같이 윤리 또는 공정성의 문제들이 있다. 이미 데이터의 분석(Data analytics)은 범죄에 대한 판단(criminal sentencing), 가격의 결정(price setting) 또는 디지털 헬스와 같은 분야에서 비중 있는 역할을 담당하고 있고, 향후 금융시장에서는 더욱 중요한 역할을 할 것으로 예상된다.

디지털 경제에서 인공지능이 창출하는 가치에는 사회의 경제적 가치(economic values) 뿐만 아니라 비경제적 가치(non-economic values)도 포함되는데, 온라인 소셜 네트워크 미디어(online social network media) 플랫폼을 이용하는 사회 구성원 간의 연결성(connection)이 그 중 하나로 이 가치는 만일 해당 플랫폼에서 잘못된 정보가 퍼지게 될 경우에 불신(distrust)이 조성되면서 손상될 수 있고, 플랫폼에 대한 신뢰 또한 잃게 된다. 그러므로 인공지능 기술이 유효하게 가치를 창출하도록 하기 위해서 법과 규제를 통해 신뢰를 형성하는 것은 중요하다 할 것이고, 신뢰에 관한 개념화(conceptualisation) 단계에서부터



Sokol 교수 발표 자료: 디지털 전환의 영향

데이터의 상업화(commercialisation)와 이에 수반되는 법적 문제가 우선적으로 논의되어야 한다.

법과 규범은 시장경제 체제에 개입하여 경제주체 간의 신뢰를 보장하여 사회가 안정적으로 발전하도록 기여하는데, 기존의 규범은 온라인 시장 환경에 적용될 때 여러 한계 또는 문제점에 부딪히게 된다. 일부 의견에 따르면 온라인 시장과 오프라인 시장은 큰 차이가 없으며, '시장'은 항상 존재해왔기에 기존의 재산권 개념에 근거하여서도 현재의 문제를 해결할 수 있다고 하나, 온라인 시장 환경에서는 경제 주체 간의 물리적인 소통이 제한되고 익명성이 존재하며, 대규모의 소비자 데이터를 수집하고 컴파일링(compiling)하여 시장을 분석, 소비자의 미래 행동을 예측한다는 차이가 존재한다. 이러한 차이를 이유로 변화된 시장경제 환경에 부합하는 신뢰의 구축이 요구된다.

그렇다면 인공지능에 대한 신뢰의 구축을 위해서는 규제가 필요한지, 만일 규제가 필요하다면 해당 규제는 어떠한 방향성을 가져야 하는지와

같은 질문이 따르게 된다. 이를 위해 기업들이 데이터를 다룸에 있어 적용하는 규범에서 그 시사점을 찾을 수 있는데, 기업이 영업활동에 데이터 분석을 적용할 때, 적용된 데이터에 존재하는 취약점(vulnerabilites) 또는 데이터로 인한 위협(threats)이 갖는 리스크가 실현되어 이로 인한 규범 위반(breach)이 일어날 경우에 기업에 미칠 영향(impact)과 그에 대한 대응(responses)의 각 단계에 기업은 거버넌스를 통해 개입하고, 그 중에서도 리스크의 정도에 따라 다음과 같이 구분하였다.

우선 ① 리스크가 높은 경우로, 민감 개인정보(sensitive personal data)를 대규모(large scale)로 처리하거나 자동화된 프로파일링(autonomated profiling)(예 : 신용카드 발급 신청에 대한 자동 거절 등), 체계화된 프로파일링(systematic profiling) 등의 경우가 있을 수 있고, ② 중간 정도의 리스크에는 민감 개인정보를 처리하거나 아동 등의 취약계층의 민감 개인정보 처리 또는 대규모 개인정보의 처리 등의 경우가 포함되며, ③ 낮은 정도의 리스크의 경우로는 익명화된 데이터(Anonymized data), 가명처리된 데이터(Pseudonymized data) 그리고 소규모의 안전한 정보 처리 등으로 분류하였다. 이를 통해 각 데이터가 갖는 리스크와 이를 관리하기 위한 기업의 개입 형태에서 향후 사회가 보유해야 할 신뢰 구축을 위한 시스템의 모습을 생각해 볼 수 있고, 각 리스크의 정도에 따라 차등 법적 제재(legal penalites)를 마련한다면 혁신을 저해하지 않을 수 있다.

다음으로 데이터 보유 또는 소유와 관련된 신뢰의 문제가 존재하는 경우를 상정해볼 수 있는데, 미국에서는 프라이버시와 관하여 연방 차원의 규범은 아직 존재하지 않으나, 독일의 법체계에서는 데이터 주체의 데이터에 대한 권리들을 규정하고 있다. 예를 들어, 사용자들에게 자신의 취향에 맞는 상대방을 소개하는 데이트 어플리케이션(Dating App)은 각 사용자에게 다른 사용자를 추천하기 위하여 여러 종류의 개인정보를 수집한다. 수집된 데이터에는 사용자의 특정 질병 유무 등과 같은 건강에 관한 정보도 있을 수 있는데, 건강 정보를 보유한 주체가 의사였다면 기존 규범을 통해 강하게 보호받을 수 있지만, 만일 데이트 어플리케이션 회사가 이를 보유하고 있다면

전자와 같은 정도의 보호를 받지 못하고 제3자에게 판매도 될 수 있는 상황과 같이 신뢰가 필요한 상황이 있을 수 있다.

그리고 시장경제와 관련이 깊은 반경쟁(Antitrust) 분야에서는 '신뢰'가 어떻게 이해되어야 하는가와 관련하여, 각 시장과 기술에 따라 개별적으로 살펴보아야 한다. 우선 빅데이터가 가져온 친경쟁적인 요소로는 ① 데이터의 수익화(Monetization of Data)를 통해 소비자에게 무상으로 제공하는 상품의 재정을 확보, ② 품질의 개선과 혁신의 향상, ③ 빅데이터의 보호의 경제적 특징으로 인한 친경쟁성이 존재한다. 즉, 데이터는 어디에나 있고 비싸지 않으며 어디서든 수집이 가능하다는 특성으로 인한 낮은 진입장벽과 데이터의 비배제성(non-exclusive)과 비경합성(non-rivalrous), 그리고 빅데이터 그 자체보다는 분석된 데이터의 효용가치가 높은 것이므로, 후자를 보호하는 것이 경쟁을 저해한다고 볼 수 없다고 설명한다.

그에 비해, 반경쟁적이라고 볼 수 있는 요소에는 잠재적인 혁신과 품질의 저하와 경쟁법상 우려 측면에서의 프라이버시 문제가 있다. 그 중에서도 알고리즘 담합(Algorithmic collusion)과 관련하여, 각 기업들이 수집하는 데이터의 출처(source)가 각기 다르고 앞서 설명한 것과 같이 데이터 수집의 진입장벽이 낮아서 전통적인 담합과 같은 모습의 담합이 알고리즘을 이용하여 나타나는 것은 어려워 보인다. 그 밖의 블록체인 및 핀테크 분야에서는 OECD가 지적한 바와 같이 편향성(bias)과 차별(discrimination), 설명가능성(explainability)의 문제 그리고 다른 한편에서는 거버넌스와 그 책임 규범적 고려사항 등에 관한 논의가 함께 진행되고 있다.

질의응답에서는 빠르게 발전하는 인공지능 기술 분야에 법이 그에 맞추어 대응할 수 있는지에 대한 질문이 있었고, 이에 대하여 Sokol 교수는 규범에 관한 근본적인 질문을 공유하며 시장을 규제하는 이유는 형식적 규범 없이 시장은 스스로 규제하지 않기 때문이라고 설명하였다. 그렇기에 현명하게 규제해야 하고, 법 관련 전문가가 기술에 대해 이해하고 있어야 함을 강조하였다. 과도 또는 과소 규제가 존재할 수는 있지만, 지속적으로 발전시켜 나가고 기술 현장과 소통을 통하여 올바른 규범을 형성해 나가야 한다고 보았다.

추가로 자율주행 차량에 대한 규제와 관련하여, 그 결과물이 정확(accurate)하면서도 공정(fair)하기를 바라는데, 어떻게 하면 이를 공정하게 하면서 신뢰를 형성할 수 있는지에 대한 질문이 있었다. 이에 대하여 기업이 감사(audit) 체계를 가지고 있다면, 설령 그것이 완벽하지는 않더라도 이를 통하여 무엇이 중요한지 즉, 편향되지 않게 하는 것이 중요한지 또는 평등 관련 이슈를 어떻게 다룰 지 등에 대하여 논의할 프로토콜을 갖출 수 있다고 설명하였다. 더불어 발생할 수 있는 몇몇 불평등 이슈를 통하여 법이 어떤 핵심 가치를 가져야 하는지 다양한 윤리적 문제에서 생각해 볼 수 있을 것이라고 덧붙였다.

프라이버시에 대하여 동의 기반 체계에서는 신뢰를 형성하기 위해서 기업은 어떠한 데이터를 수집할 수 있는지에 대한 질문이 있었다. Sokol 교수는 프라이버시 문제는 그 중에서도 해결하기 어려운 문제로, 아직까지는 기업들도 다양한 데이터를 만들어내는 단계에 있기에 소비자의 데이터가 정확히 어떠한 상황에서 어떻게 사용될 수 있는지 사전에 모두 동의를 받기 어렵지만 만일 데이터 주체가 자신의 데이터를 이전함으로써 어떠한 서비스를 받고 어떻게 활용되는지 잘 알 수 있게 된다면 더 단단한 신뢰를 형성할 수 있을 것이라고 보았다.

마지막으로 한국에서 특별히 고려되는 것이 있는지에 대한 질문에, 각 국가가 프라이버시에 관하여 각각의 입장이 서로 다른 것을 확인하였다고 언급하면서, 다른 국가에서는 프라이버시와 서비스 간의 교환(trade-off)을 인정하기도 하지만 한국에서는 국가 안보의 우려를 이유로 구글의 지도 서비스가 제공되지 않는 점을 예로 들었다. 이러한 프라이버시에 대한 입장은 비단 국가 간 차이뿐만 아니라 세대 간 차이도 있을 수 있어 데이터를 어떻게 규제해야 한다는 단편적인 답변은 존재할 수 없고 향후 발전 방향과 각기 다른 데이터 유형에 따라 유연하게 이루어져야 할 것이라고 하며 마무리하였다.

제3세션

디지털 시대의 경쟁과 프라이버시

Competition in the Digital Age

발표자 **Ariel Ezrachi**

University of Oxford

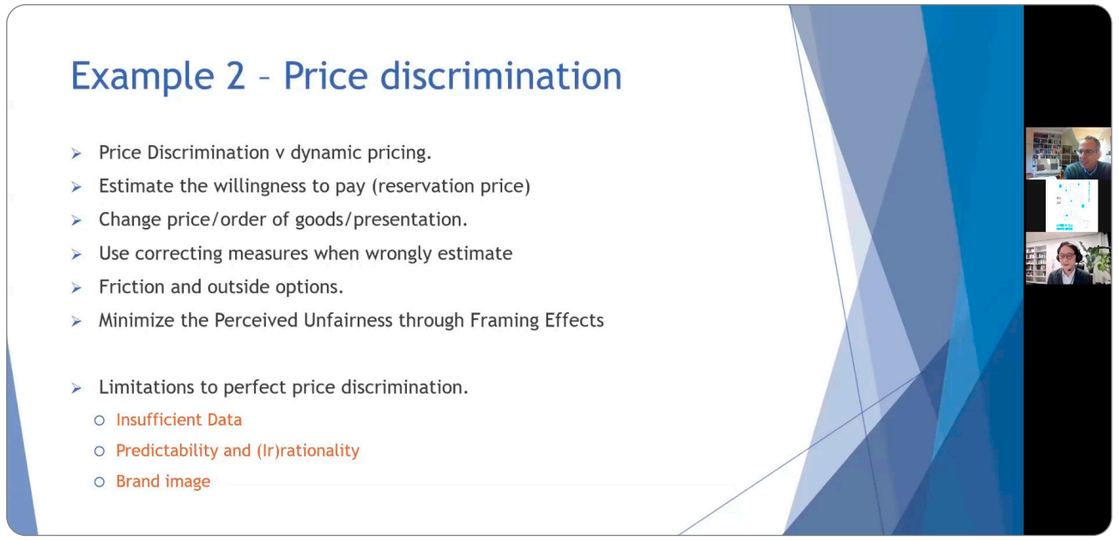
발표자

Ariel Ezrachi

제3세션

최근의 디지털 시장에서의 경쟁 구도의 변화, 경제력 집중 등의 특징을 갖는 디지털 경제에서의 경쟁과 경쟁법에 대한 논의를 세 번째 세션에서 다루었다. Ariel Ezrachi 교수는 디지털 경제로 인한 시장의 변화와 변화된 환경에서의 경쟁의 역할 그리고 경쟁의 올바른 작동을 위하여 경쟁법이 어떻게 적용되어야 하는지에 관하여 초점을 두면서, 디지털 경제가 실현한 유익 또한 분명하게 있다는 것과 이러한 유익 자체를 부정하는 접근이 아닌 보다 더 확장된 사회적 이익을 뒷받침하기 위해 이와 같은 디지털 환경을 어떻게 활용할 수 있는지의 관점에서 소개하였다.

우선 디지털 경제로의 변화로 인하여 시장 역학(market dynamics)이 변화하고 시장에서의 집중도가 심화(increase of concentration)되었는데, 이러한 변화는 경제와 프라이버시에 직접적인 영향을 미치게 되었고 또한 그 영향으로 소비자의 지위와 후생에 변화를 가져왔다. 디지털 경제로 전환하는 과정에 수반된 시장의 변화 특성으로는 ① 네트워크 효과의 존재, ② 필수 투입재가 된 데이터, ③ 수준 높은 데이터 분석과 인공지능의 발전, ④ 시장에서 정보의 비대칭성으로 인해 데이터 분석이 보유하는 힘, ⑤ 정보를 추적(tracking)하고 하베스팅(harvesting)하여 소비자를 타게팅하며, 이를 통해 조종(manipulation)하는 일련의 과정, ⑥ 인터페이스(interface) 전반에 걸친 통제력 등이 있다. 그리하여 점점 더 많은 개인이 위와 같은 특성을 갖는 플랫폼에서 이탈하거나 이를 사용하지 않는 것이 더욱 불가능해질 것이고 간접적으로는 시장의 다른 면(side)에 있는 주체(예 : 플랫폼 이용자에게 상품을 광고하려는 광고주)에게도 영향을 미치게 된다. 이러한 시장의 역학(market dynamics) 변화를 경쟁법 시작에서 바라보면 상당한 문제가 될 수 있다. 왜냐하면, 경쟁법은 시장의 경쟁이 제대로 작동하지 않을 때 이를 시정하여야 하기에 우선 관찰되고 있는 변화로 인해 시장의 자정능력에 대해 살펴보고, 경쟁(즉, competition for market)이 충분히 이루어지고 있는지를 점검하여야 한다. 이에 디지털 시장(Digital market)에서는 충분한 경쟁이 존재하는지 질문하고 만일 그렇다면 경쟁 과정을 저해(chill the process of competition)하지 않기 위해 경쟁법의



Ezrachi 교수 발표 자료: 가격 차별행위

과도한 개입(over-intervnetion)은 지양하여야 하고, 충분하지 않다면 그 다음으로 법률가가 던져야 할 질문은 그 문제는 경쟁법으로 다루어야 할 문제인가, 그리고 경쟁법이 개입한다면 그 개입의 정도(level of intervention)는 어떠해야 하는가라고 할 수 있다.

경쟁법 관점의 위 질문들에 대한 답을 위하여 최근의 논의와 문헌을 살펴보면서 디지털 경제에서 시장의 경쟁이 얼마나, 어떻게 왜곡되었는지 추론하고자 하였다. 첫번째로 『Virtual Competition』에서 언급된 내용으로, "보이지 않는 손은 디지털화된 손에 의하여 대체되었다(Invisible hand has been replaced by digitalised hand)" 는 것이다. 아담 스미스(Adam Smith) 이후로 자유 시장에서는 보이지 않는 손이 작동한다는 명제가 수용되었으나, 다수의 디지털 시장에서는 중립적이면서 유기적(organic)인 보이지 않는 손이 작동하지 않는다는 점을 확인할 수 있다고 보았다. 개개인이 스크린을 통해 접할 수 있는 정보는 더이상 시장을 그대로 반영하지 않으며, 이는 실제와 다른 현실을 보여주기 위하여 활용된 알고리즘이 만들어낸

이미지일 수 있다.

그 첫 예로 디지털 경제에서 일어나는 가격 조정(price alignment)은 알고리즘을 이용하여 가격을 결정할 경우에 해당 알고리즘이 가격 담합을 위해 코딩된 알고리즘이 아님에도 불구하고, 알고리즘 스스로 시장을 이해하면서 점차 균형점에 도달하게 되어 유사한 가격 수준으로 조정되는 것이다. Calvano et al. 수행한 연구에 따르면, 가격을 정하는 서로 다른 알고리즘을 가격 결정 게임(game to set up the price)에 투입한 뒤에 게임을 반복한 결과 두 알고리즘이 균형점이 이른 것을 볼 수 있다. 알고리즘이 서로 경쟁하면 가격을 하락한다는 사실을 학습하게 되면서 어떠한 담합의 의사 교환 등이 없는 합리적이고 중립적인 상태에서 이루어진 알고리즘이어도 가격을 서로 맞추도록 작동하게 된다는 것이다.

가격 차별(price discrimination)의 예 또한 들 수 있는데, 이는 온라인에서 소비자가 상품을 구입하거나 서비스를 이용할 때 스크린 상 이용자에게 표시된 가격이 마치 자연스러운 시장의 작용이 반영된 가격이라고 생각할 수 있지만, 실제로는 이용자의 정보를 수집하고 이를 분석하여 제시된 맞춤형(tailored) 가격인 경우가 더 빈번하다. 판매주체가 주로 수집하는 정보에는 자신의 웹사이트에 접속하는 경로(path), 상품 판매 페이지를 이용하는 기기(device, 즉 컴퓨터인지 모바일인지 여부), 이용자의 우편번호, 검색어 등이 있고, 이를 수집하여 잠재적 소비자의 성향을 파악하여 매출을 증대하기 위한 전략에 반영한다. 디지털 경제에서 온라인 시장은 경쟁 가격(competitive price)이 아닌 개별 이용자를 위한 고유의 맞춤형(unique and bespoke) 인터페이스(interface)를 구현하려 한다. 여기서 소비자 보호의 문제 혹은 경쟁의 문제, 부의 재분배 문제가 제기될 수 있고, 이는 각 사회에서 필요하고 중요하다고 여기는 가치에 따라 각기 다른 결론에 이를 수 있다고 본다.

디지털화된 손에 관한 세 번째 예로는 마치 영화 트루먼 쇼와 같이 우리를 둘러싼 모든 것이 통제된 생태계(controlled ecosystem)에서 우리 자신이 내리는 결정 마저 이미 디자인되었다고 한다면 개인의

자주성(autonomy)이 상실되어 그 행위를 조정(manipulate)할 수 있다는 점이다. 개인에 관한 정보에 접근할 수 있는 범위가 더 늘어난다면 이를 통하여 개인을 조종하거나 개인의 행동에 영향을 미칠 수 있는 힘을 갖게 될 수 있다. Ezrachi 교수는 소셜 네트워크(social network) 서비스 운영주체들은 알고리즘을 사용하여 이용자의 기분에 영향을 미칠 수가 있는데, 수집한 정보를 바탕으로 개인의 취약점, 개인의 상태 혹은 쉽게 조종될 수 있는 미끼(decoy)와 같은 요소들을 제공하여 영향을 미칠 수가 있다고 보았다.

대표적인 사례로 선거에 영향을 미치는 것을 목적으로 계획된 Cambridge Analytica 사례가 있는데, 여기서는 수집된 데이터를 기반으로 아주 세밀한 타케팅(micro-targeting)이 이루어졌다고 할 수 있다. 어느 개인이 자신의 친구로부터 투표에 참여할 것이라는 메세지를 받는다면 사람은 사회적인 존재이기에 메세지를 받은 사람도 투표할 가능성이 높아질 수 있고, 그 반대로 투표를 하기 위해 집을 나설 의욕을 저하시킬 수도 있다. 이처럼 개인의 행위에 영향을 미치고자 하는 시도는 음성인식 기술과 연관되어 더욱 심각해질 수도 있다. 디지털 집사(digital butlers)의 역할을 점차 확장하고 있는 기술 발전에서 이 기술을 이용하지 않을 선택(opt-out)이 얼마나 강력할지 의문이다.

디지털 경제에서의 또 다른 시장 경쟁 모습의 변화로는 『Competition Overdose』에서 설명하는 바와 같이 소수의 기업이 인터페이스 자체를 지배하는 게이트키퍼(gatekeeper)로서의 기능을 보유하고 있고, 시장에서 그 소수의 기업이 보유하는 힘이 더욱 강해지고 있다는 것이다. 시장에서 게이트키퍼가 갖는 힘은 '사용자에 대한 영향력(예 : 맞춤 광고 등)' 뿐만 아니라, '서비스 제공 주체에 대한 영향력(예 : 시장의 규칙을 수립할 수 있는 지위)'도 존재한다. 즉, 자신의 플랫폼에서 상품을 판매하기 원하는 판매주체에게 소비자 또는 이용자와 연결하는 중개 서비스를 제공하면서 당해 플랫폼 생태계 규칙을 통제하고, 세금과 같은 부담을 부과함과 동시에 최종소비자가 플랫폼에서 활동한 이용자의 데이터를 최초로 수집할 수 있고,

자신의 플랫폼을 활용하여 제3자의 지위에 있지만 여전히 이용자의 정보를 추적(tracking)하여 자신이 시장에서 갖는 영향력을 더 강화할 수 있다.

특정 생태계(ecosystem)가 소수 개별 기업에 의하여 통제될 때 제기될 수 있는 우려로, 디지털 경제에서 대부분의 대형 플랫폼은 광고 수입에 의존하는 수익 구조를 갖고 있어, 시장에서 이루어지는 경쟁이 경쟁이 소비자에게 유익한 경쟁이 아니라는 점을 지적하였다. 일반 개인과 같은 이용자가 '소비자'에 해당하는 것이 아니라, 자신의 플랫폼에 광고하고자 하는 광고주체들이 '소비자'에 해당하기 때문에, 일반 이용자에게 유익한 경쟁이 될 수 없다는 것이다. 일반 이용자는 이 경쟁에서 '상품'이 될 뿐이다. 다시 말해, 이 생태계에서 시장력을 갖는다는 것은 이용자를 소비자로 보는 것이 아니라 점점 더 자신의 수익 창출 목표에 더 초점을 맞추고, 기업은 할 수 있는 한 최선을 다해 이익이 극대화되는 방안을 찾는다.

이런 이유로 현재의 디지털 시장에서 경쟁이 심해질수록 타겟팅 광고와 이를 위한 분석이 점점 더 많이 이루어지는 것이고, 그로 인한 다른 문제들의 개선 등에 대한 고려는 간과되고 있어 궁극적으로는 해로운 경쟁(toxic competition)이 될 수 있다. 디지털 경제에서 기업이 시장에서 갖는 힘이 커질수록 소비자를 고려하는 부분은 줄어들고, 이용자는 선택은 더욱 좁아지면서 특정 서비스에 고착되고(locked-in), 결과적으로 개인의 프라이버시에 대한 자기 결정 정도 또한 감소한다.

마지막으로 시장에서의 플랫폼 기업의 현황과 몇몇 플랫폼이 인터페이스를 통제함으로써 혁신에 미치는 영향 등에 대해 다룬 『 How Big-Tech Barons Smash Innovation—and How to Strike Back 』에서는 경쟁법의 개입을 저지하기 위하여 '혁신적이다'라는 평가 기준을 활용하고 있다는 것을 지적하였다. 근래 혁신이라고 평가받고 학계와 시장에서 혁신적인 효과가 있다고 회자되는 기술이 실제로는 플랫폼 기업의 재정지원을 받아 그들의 어젠다에 따라 제시된 '혁신'에 해당할 수 있고, 이에 Ezrachi 교수는 Paul Arnold의 Innovation kill zone이라는 표현을 인용하여 플랫폼 기업이 스스로 이룩한 '발전'을 방해하려 하거나 저지하려 한다면 퇴출될 수도

있다고 보았다. 또한 (舊) Facebook의 대표인 Mark Zuckerberg의 발언인 "경쟁하는 것보다 사는 것이 낫다(It's better to buy than compete)"와 같은 발언을 조명하면서 경쟁이 위협받는 상황을 언급하였다.

그러나 경쟁법에 의한 시정의 한계 또한 인정되는데, 특히 반경쟁행위는 가격을 기준으로 발전해왔고 가격에 대한 분석을 중심으로 연구되어 왔으나, 플랫폼 기업들은 대부분 이용자에게 돈을 받지 않고 서비스를 제공하는 대신 이용자들은 자신의 개인정보와 시간, 주의(attention)를 교환하는 것이기에 기존의 저해 이론(theory of harm)이 그대로 적용되기에 문제가 있다. 또한 경쟁법의 개입이 시장의 문제를 온전하게 해결하지 못하는 문제도 있어, 이러한 시장의 역학 변화가 새로운 접근을 요구하고 있지만 그럼에도 불구하고 여전히 시장에서는 경쟁법이 중요성을 가진다고 할 수 있다. Google에 대한 경쟁당국의 집행으로 인하여 Google이 시장에서 수직적으로 결합된 자사 기업을 우대하는 등의 행위를 정의하고, 이를 시정할 수 있도록 하는 등의 변화를 이끌어 온 것 또한 경쟁법이기에 지속적인 논의가 이루어져야 한다. Ezrachi 교수는 EU에서 진행되고 있는 Amazon 사건과 독일에서 개인정보와 관련되어 문제된 Facebook 사건을 함께 설명하면서 경쟁당국이 시장에서 문제되는 지위 남용 사안을 다루면서 동시에 프라이버시를 위한 규제 체계를 구성하고 있다고도 보았다.

그러므로 현 시점에서 경쟁법이 향하고 있는 방향은 규범과 집행 체계를 위와 같은 시장 역학의 변화에 맞추려는 실질적인 시도가 이루어지고 있고, 위반행위가 있을 때까지 기다리고 그 이후에 대응하는 것은 너무 늦을 수 있다는 점에 공감대가 형성되고 있기에, 가처분 등과 같은 사전 조치의 도입에 대해 논의가 이루어지고 있다. 또한 시장의 경쟁력을 향상시키기 위하여 어떠한 시장실패가 있었는지 그리고 어떠한 행태적 또는 구조적 시정으로 이를 회복할 수 있는지에 대하여 정의가 필요하다고 보았다. 그리고 경쟁법에서 저해 이론은 기존의 이론만 가져올 것이 아니라 새로운 이론도 얼마든지 형성이 가능하므로 이런 과정을 통해서 진보할 수 있다고 보았다.

제4세션

AI의 아틀라스: 파워, 정치, 그리고 지구가 지불해야 할 비용을 그리다

Atlas of AI: Powers, Politics, and the Planetary Costs of Artificial Intelligence

발표자 **Kate Crawford**

University of Southern California Annenberg School

발표자

Kate Crawford

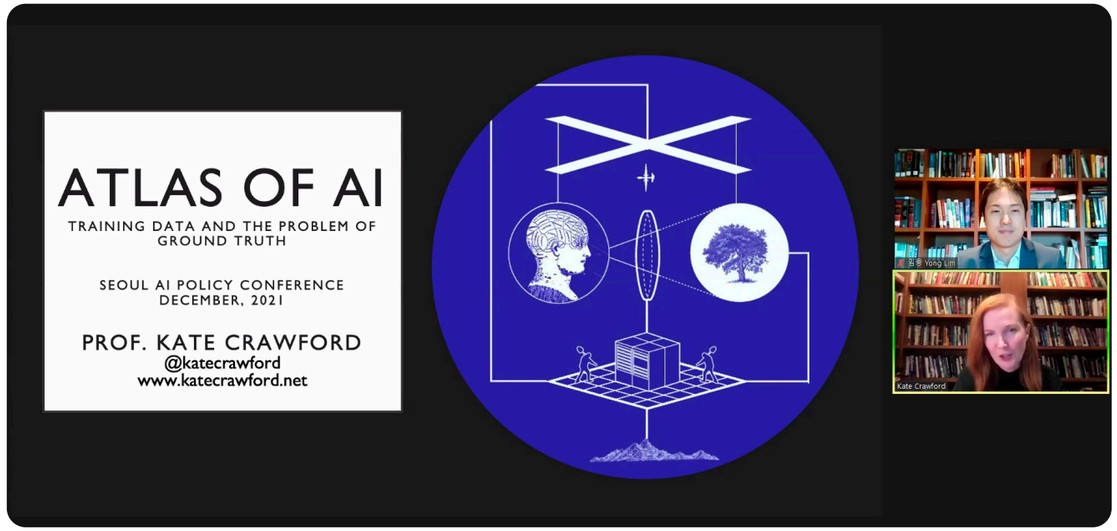
제4세션

Crawford 교수는 인공지능 시스템을 개발하고 학습시키는데 활용되는 데이터에 존재하는 편향(bias), 그리고 편향된 데이터로 훈련된 인공지능이 의사결정을 할 때 발생할 수 있는 다양한 문제들에 대한 논의를 다루었다. 최근 논의되는 인공지능의 편향이 주로 인공지능이 도출해내는 답(output)에 집중되는 경향이 있는데, 더 근본적인 문제는 인공지능 시스템에 입력하는 데이터(input)의 편향이 더 다루어져야 한다고 하였다.

인공지능을 학습시키는 데이터는 기본적으로 사람이 라벨링 작업을 하여 마련되기에 누가, 무엇을 그리고 어떻게 라벨링 작업을 하느냐가 중요하다고 볼 수 있다. 학습을 위한 데이터의 역사적 흐름을 먼저 살펴보면, 1980년대부터 훈련을 위한 데이터의 활용에 통계적 접근이 접목되기 시작했는데 그 예로 IBM에서 컴퓨터에게 언어의 문법체계를 학습시키는 대신에 각 단어들이 통계적으로 얼마나 자주 연속으로 등장하느냐에 초점을 맞추어 학습시킨 바 있고 이러한 접근은 지금도 유지되고 있다. GPT-3 역시 인터넷에 존재하는 무수한 언어의 Corpus에 통계적 접근 방식으로 분석하여 이를 훈련시킨 모델이다.

그리고 학습을 위한 데이터를 마련하는 방법의 경우 과거에는 각각 데이터를 일일이 마련하는 방식이었는데 1980년대에는 Computer vision을 위한 이미지 학습 데이터를 마련하기 위해 사람을 섭외하여 일일이 한 명씩 사진을 찍는 방식으로 진행되었다. 현재는 이미지, 음성, 텍스트와 같은 각종 데이터를 인터넷에서 쉽게 구할 수 있기에 학습을 위한 데이터의 마련이 그다지 어렵지 않게 되었다. 그러나 이렇게 확보된 데이터는 종종 프라이버시 관련 문제가 있을 수 있고, 학습용 데이터에 나의 데이터가 활용되는 것에 동의한 적이 없는 경우도 태반이다. 데이터의 종류 또한 한정적일 경우가 많은데 예를 들어, 사람의 얼굴에 관한 데이터를 수집할 때 백인 남성의 얼굴 데이터가 다수일 경우 그 데이터로 학습된 인공지능 모델은 백인 외의 인종을 잘 인식하지 못하는 경우가 발생할 수 있다. 실제로 IBM 등 많은 회사들의 안면인식기술이 이러한 문제를 겪었다.

위와 같은 여러 이슈가 꾸준히 제기되어 왔음에도 여전히 학습을 위한



제4세션 Crawford 교수 발표 자료: ATLAS of AI

훈련용 데이터가 어떻게 구성되었고, 어떻게 라벨이 되었으며 그 데이터들이 정보 주체들의 동의를 받은 데이터인지 여부에 대한 산업 전반에서 참고할 수 있는 기준이 존재하지 않는다. 이 문제를 인식하고 개선점을 찾기 위해 중요한 연구들도 진행되었는데, Datasheets for Datasets, Model Cards for Model Reporting, The Dataset Nutrition Label 과 같은 연구가 발표되었음에도 여전히 이 내용이 '기준(standard)'으로 확립되기에는 갈 길이 먼 상황으로 보인다. 그러므로 이러한 기준 마련을 위하여 데이터셋에 대한 논의와 그 데이터셋이 인공지능 체계에 어떤 영향을 미치는 지에 대한 철저한 감시와 논의가 필요하다 할 것이다.

인공지능 시스템 개발을 위해서는 인공지능을 학습시킬 훈련 데이터를 사용하기 전에 이 데이터를 면밀하게 살펴보는 것이 매우 중요한 반면, 많은 개발자 또는 엔지니어들이 그 과정을 여전히 간과하고 있다. 그 예로, Crawford 교수가 'Imagenet'이라는 사물인식 프로젝트에 자주 활용되는 학습을 위한 데이터셋을 자세히 분석해보는 연구를 진행하면서 발견한

문제점을 들 수 있다. Imagenet은 1990년대 개발된 어휘 데이터셋인 wordnet에 기반한 것으로 각 단어를 대표하는 이미지 데이터를 수집한 것이다. Imagenet 데이터셋 초반 라벨링 작업에는 대학 학부생들에게 시간당 $10을 지급하여 진행하였고, 이후에는 더 확장하여 Amazon Mechanical Turk 인력을 활용하였다. 그러나 Amazon의 Turkers 인력은 라벨링 작업에 시간당 $2에 불과한 금액을 지급받았으며 이것이 라벨링 작업의 질이 저하되는 결과로 이어졌다.

뿐만 아니라 일부 단어는 이미지로 표현하기 어려운 경우가 있었는데(예 : '사과(apple)' 라는 단어는 비교적 이미지가 일관적인 반면에, '채무자(debtor)', '친구(friend)', '나쁜 사람(bad person)'과 같은 단어) 이러한 단어들의 경우 특정 인종 또는 성별의 이미지에 치우친 경향이 있음이 확인되었다. 즉, 라벨 작업을 하는 인력의 개인적인 판단이 들어간 것으로 볼 수 있다. 실제로 Imagenet의 카테고리 중 혐오적이고 차별적인 단어들이 많다는 사실이 발견되었는데, 그 예로 'loser'를 대표하는 이미지 중에는 뚱뚱한 사람의 이미지가 많았다.

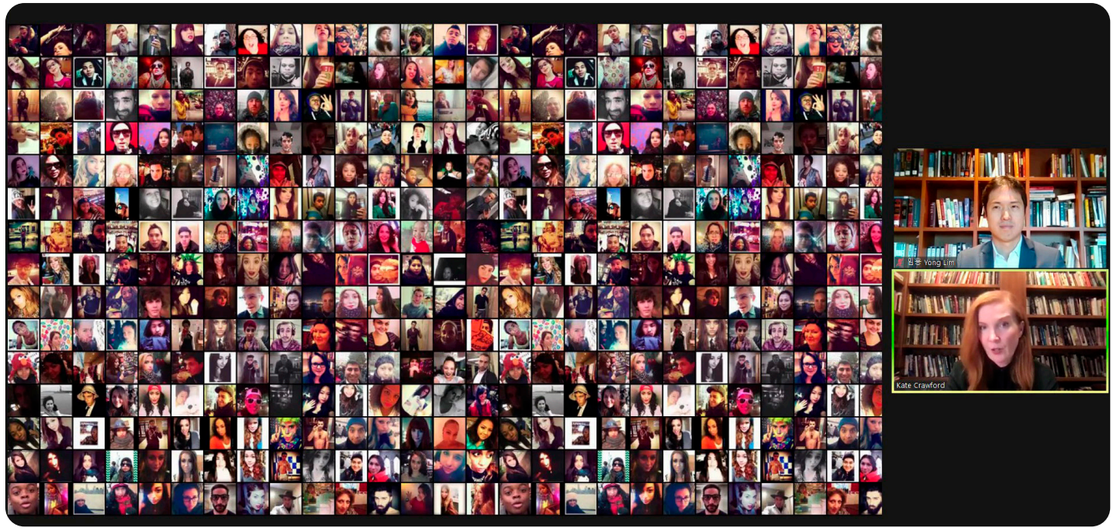
이러한 Crawford 교수의 연구 이후에 Imagenet 측에서는 차별적인 이미지 60만 개를 선별하여 이를 삭제하는 조치를 취하였지만, 그러한 혐오적인 카테고리에만 집중하는 것이 모든 문제를 해결하지는 않는다고 설명하였다. 예를 들어 '농구선수'라는 혐오 표현이 아닌 단어에도 연관된 이미지에는 대부분의 경우 흑인이, '경제학자'라는 단어에는 주로 백인 남성의 이미지가 많은 것을 확인할 수 있다. 그러므로 중립적인 카테고리란 존재하지 않을 수 있다. 최근에는 문제 있는 학습용 데이터를 수정하고 삭제하는 조치가 많이 이루어지고 있는데 그럼에도 불구하고 문제된 데이터들은 이미 연구를 수행하는 커뮤니티에 많이 퍼져있고 토렌트를 통해 수정 혹은 삭제 이전 버전이 계속해서 퍼지고 있는 것이 현실이다. 이렇게 편향된 데이터셋으로 학습된 인공지능 시스템이 계속 사용된다면 우리가 중요하게 생각하는 민주적 가치들에 더욱 더 심각한 문제가 될 수 있다.

이제 우리가 직면해야 하는 것은 앞으로 어떻게 할지에 관한 것으로,

시간이 갈수록 학습을 위한 데이터를 점점 더 고민없이 인공지능 개발에 활용하는 사례가 늘어나고 있어 우려된다고 할 수 있다. 실제로 즉석만남/데이트 웹사이트에 있는 사람들의 프로필 사진 데이터를 끌어와 성 정체성 파악을 위한 훈련 데이터로 사용한 사례가 있었는데, 몇몇 국가에서는 동성애를 법적으로 금지하고 있기에 일부 개인에게는 이것이 큰 위험이 될 수 있다. 그 외에도 성격, 감정상태 등을 감지하려는 인공지능 시스템도 개발되고 있고, 대부분의 연구에서 성별을 남자와 여자, 이분법적으로 구분하고 있는데 이럴 경우에는 이분법적 성별에 속하지 않는 이들을 소외시키는 결과를 초래하기도 한다.

Crawford 교수는 여기에 더 많은 성별 분류 기준을 추가한다고 해서 해결되는 문제도 아니고, 이는 사람의 인종, 성별 등을 개인의 동의 없이 시스템 개발자 입장에서 분류하는 행위 그 자체에 의문을 가질 필요가 있다고 보았다. 인종과 관련한 이슈에서도 '순종'이라는 개념이 존재하지 않는데도 많은 경우에 인종을 분류하여 적용하고 있고, 머신 러닝 체계는 이에 따라 사람을 한정된 개념 안에서 분류한 데이터로 학습을 진행하면서 정체성에 대해 배워가고 있다. 이에 사람은 각자 다른 사람을 소외시키지 않도록 도덕적 책임을 진다고 볼 수 있는데 이를 위해 법과 정책 전문가들이 할 수 있는 일이 많을 것이라고 보았다. 정책 체계와 프레임워크의 수립, 인공지능 시스템의 중장기적인 영향 등에 대한 학제간 연구도 활성화되어야 한다고 보고, 머신 러닝의 연구자들은 법학자와 인류학자들과도 긴밀하게 협력할 필요가 있다고 하였다.

이후 질의응답시간에는 위 발표와 관련하여 이미 문제가 있다고 여겨지는 학습 데이터로 훈련이 된 인공지능 도구(tool)의 경우 어떻게 하면 좋을지에 대한 질문이 있었고, Crawford 교수는 일반 사람들에 대하여 이런 문제가 있음을 교육함으로써, 그 인공지능 도구를 사용하기 전에 해당 문제에 대해 인지할 수 있도록 하는 것이 중요할 것이라 답하였고, 우리 건강에 지대한 영향을 미치는 MRI와 같은 의료 기술에 대해서는 출시 하기 전에 엄격한 기준으로 심사를 하는 것과 같이 머신 러닝 기술도 우리 일상에 큰 영향을



Crawford 교수 발표 자료

미치는 것이 분명하므로, 머신 러닝 연구자들이 연구 성과를 최대한 빨리 내놓도록 압박하는 것을 지양하고 엄격하면서도 이러한 문제를 감지해낼 수 있는 관련 정책 또는 규제를 마련하는 것이 필요하다고 하였다.

학습을 위한 데이터를 애초에 문제 없이 마련할 수 있는지에 관한 질문에 대해서는 데이터에 문제가 아예 없을 수는 없을 것이고, '문제'에 대한 정의도 현 단계에서는 명확하지 않은데 이는 사람들마다 생각과 가치가 다르기 때문이라고 답변하면서도 문제가 발생했을 때 이에 대한 책임(accountability)과 그 체계를 제대로 정립하는 것이 중요할 것이라고 하였다. 학습용 데이터를 형성하고 이를 활용하는 기업들은 그로 인하여 문제가 발생했을 때 이에 대한 책임을 져야 할 것이고, 더 효과적인 감시를 위해서는 기업 또는 영업 비밀 등을 포기할 수 있어야 할 것이라고 하며 마무리하였다.

제5세션

인공지능 시대의 폭풍속으로:
스탠포드 "AI 100년" 연구 보고(2021)

Gathering Strength, Gathering Storms:
Summary of the One Hundred Year Study
on Artificial Intelligence(AI 100)
2021 Study Panel Report

발표자 **Michael L. Littman**

Brown University

발표자

Michael L. Littman

제5세션

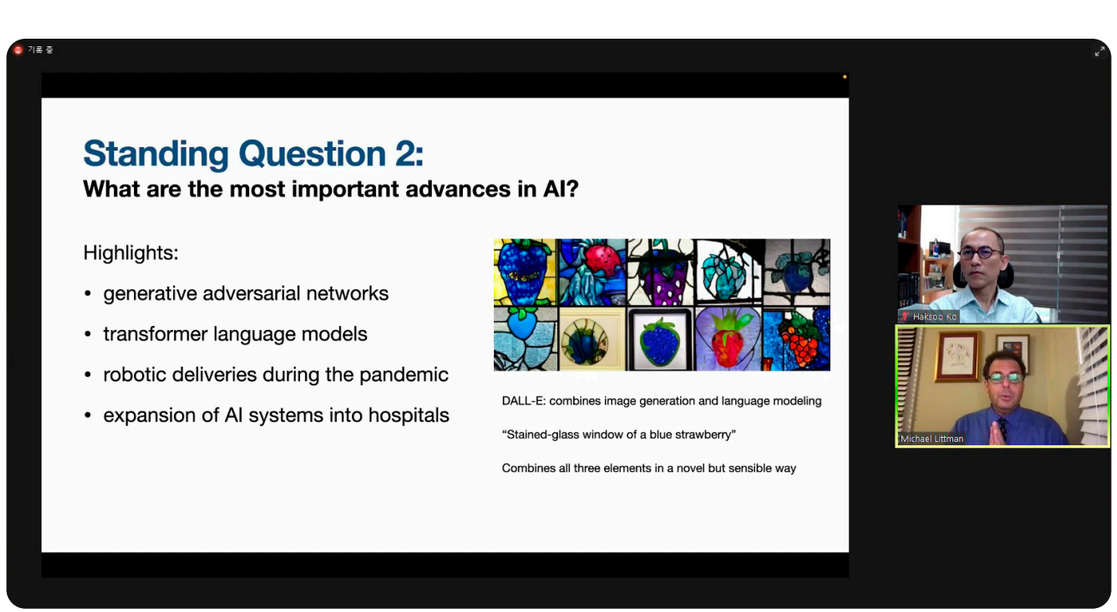
제5세션은 스탠포드 대학이 '인공지능 100년 연구' 프로젝트를 2014년부터 주관하여 2016년의 첫 보고서를 발간한 이후 5년만에 발간된 두 번째 보고서에서 집중적으로 다룬 주제에 대한 논의가 이루어졌다. 인공지능 100년 연구는 스탠포드 대학이 주관하지만 전세계의 각 전문가와 협업하는 프로젝트로, 궁극적으로는 향후 AI의 발전 방향 그리고 인공지능의 발전으로 인하여 야기되는 각종 사회 도전과제들을 장기적으로 살펴보는 것을 목적으로 하고 5년마다 보고서를 발간할 계획을 가지고 있다. 본 프로젝트는 상임위원회(standing committee)를 두어 연구에 참여할 패널을 선정하고 있으며, 연구 참여진은 비단 컴퓨터 공학 뿐만 아니라 법학과 철학, 사회과학, 정책 및 심리학 등 다양한 학계 전문가로 구성되어 있다.

2021년 발간된 리포트에서는 다음의 논의들을 중점적으로 다루고 있는데 그 중에서 첫 번째로 소개된 질문(standing question)은 '인공지능의 중요한 발전과 그가 미치는 영향을 반영하는 모습은 무엇인가 ?로 앞으로 15년 정도 이후 북아메리카에서 인공지능이 어떤 모습을 가질 것인지에 대한 질문으로 시작한다. 연구에 참여한 패널들은 100년이라는 기간 동안

Littman 교수 발표 자료

20건의 보고서가 작성될 것이기에, 현재의 연구진은 현재 시점을 대표하게 되는 것이고 타임캡슐과 같이 그 당시에 어떤 일이 일어났는지, 어떤 논의가 이루어졌는지 등을 제대로 반영하는 것이 중요하다고 인식하였다. 이러한 이유로 위와 같은 질문을 통해 현재 인공지능이 어떤 모습인지에 대해 파악하고자 하였다.

두 번째로는 인공지능에서 가장 중요한 발전이 무엇인가에 대한 논의이다. 인공지능 기술이 발전하면서 생성적 대립 신경망(Generative Adversarial Networks), 트랜스포머 언어 모델(Transformer language models), 로봇 배달(특히 COVID-19 으로 인한 팬데믹 당시의 발전) 그리고 병원과 같은 의료기관 내 인공지능 시스템의 보다 적극적인 활용(에 : 환자의 진단, 병원 내 인력 배치 등) 등의 모습으로 나타나고 있다. 또한 DALL-E와 같은 인공지능 기술은 특정 텍스트를 제시하면 그 텍스트에 부합하는 이미지를 생성해내는데, 예를 들면 '파란색 딸기 스테인 글라스 창문'이라는 텍스트를 제시하면 이전에 그와 같은 텍스트 데이터로 학습한 적이 없음에도 불구하고 텍스트 내용에 가까운 이미지를 생성한다. (아래 그림 참조)



Littman 교수 발표 자료: DALL-E가 생성한 이미지 예시

세 번째 질문은 앞으로 해결하고자 하는 대 과제(grand challenge)에 관한 것으로, 대표적인 예로 로보컵(Robocup)이 있다. 2000년 무렵부터 진행된 프로젝트로 2050년까지 사람이 참여하는 월드컵 수준의 로봇 축구 월드컵이 가능한 로봇을 개발하는 것을 목표로 하고 있다. 프로젝트 시작 이후 20여년이 지난 현 시점에서 점진적인 발전이 이루어졌으나 여전히 사람이 수행능력에 비하면 많이 부족한 수준이다. 그 외에도 인공지능 시스템의 일반화(generalize), 인과관계 요소에 대한 이해(understanding causal factors), 다양한 맥락에서의 문화적인 규범에 대한 이해를 갖추는 것 등이 있다.

네 번째는 인간 지능의 핵심 신비에 대한 이해가 어느 수준에 와 있는지에 대한 질문으로, 인공지능에 대한 연구는 계속해서 창조해나가는(creative) 영역이라면, 사람의 지능에 관한 연구는 사람이 할 수 있는 일을 어떤 단계를 거쳐 수행하는지에 대한 이해에 초점을 둔다. 최근에는 인지과학 분야에서 네트워크 효과(network effect)에 관한 연구가 조망되고 있다. 여기서 네트워크 효과란, 사람의 지능이 뇌의 특정 한 부분에서만 결정되는 것이 아니라, 다양한 뇌의 부분들이 상호작용을 하며(interaction) 형성되는 네트워크를 통해 결정된다는 것을 의미한다. 또 사람은 각각의 지능 뿐만 아니라 집단의 지능 또한 활용할 수 있다는 점에서 인간 지능은 각각 별개로 존재하는 것으로만 볼 것이 아니라 다양한 사람으로 구성된 여러 집단마다 지능이 어떻게 분포하는지에 대한 연구도 의미 있을 것이라고 하였다. 이와 관련하여 혼자 문제를 해결하는 인공지능이 아닌, 사람과 협력하여 문제를 해결할 수 있는 인공지능의 개발에 대한 관심도도 높아지고 있다.

다음 질문은 보다 더 일반적인 인공지능(General AI)에 대한 기대가 어떤것인지에 대한 질문이다. 사람과 동등한 수준의 인공지능 시스템을 개발하기 위한 연구들을 통해 오히려 사람의 지능이 얼마나 복잡한 것인지에 대한 이해도가 높아지고 있다. 그 중에서 몇몇 기술은 일반적인 문제를 사람과 동등한 수준으로 해결할 수 있는 잠재성을 보여주기도 하였는데,

그 예로 GPT-3, BERT 등과 같은 대규모 언어모델들은 언어 뿐만 아니라 지식을 습득하는 능력 또한 뛰어나다. 근래 Google이 제공하는 문서 편집 프로그램인 Google Docs의 경우에도, 아무리 복잡한 문장이라도 잘못된 문법을 찾아내어 정교하게 수정해주고 있다. 이러한 수준은 5년 전에는 불가능했었던 것들이다. 이렇듯 유용한 서비스가 상용화되고 있지만 여전히 AI이 일반 지능(general intelligence)를 습득했다고 보기에는 어렵다고 할 수 있다. 최근에는 Meta-learning이라는 학습 알고리즘이 주목을 받았는데, 여기서는 문제를 어떻게 해결하는지(How to solve the problem)에 초점을 두지 않고, 문제를 어떻게 배우는지(How to learn about the problem)에 초점을 맞추어 더 효과적으로 배울 수 있도록 하였다.

여섯 번째 보고서의 질문은 대중이 인공지능에 대해 어떻게 느끼고 있는지, 그리고 어떻게 인공지능에 대해 알리고 교육할 수 있는지에 대한 것이었다. 언론을 통해 전달되는 내용에서는 인공지능의 장점과 단점이 과장되는 경우가 많은데, 이로 인하여 인공지능의 편향성, 투명성 부족의 문제, 책임성과 관련된 문제 등에 대한 일반 대중의 우려를 야기하게 된다. Littman 교수는 사진과 같이 보이지만 사진이 아닌 인공지능이 생성해낸 GAN의 이미지를 예시로 활용하면서 인공지능 기술에 대해 일반 대중이 가질 수 있는 반응을 언급하였다. 그러므로 연구자들과 인공지능을 활용하는 관련자들은 인공지능에 관해 진행되고 있는 논의에 대중을 더욱 더 참여할 수 있도록 하여 그 논의가 보다 더 생산적으로 이루어져야 할 것이다.

그리고 인공지능을 개발하고 사용함에 있어 그 책임(responsibility)이 보장되기 위해서 정부가 담당해야 할 부분은 무엇인지에 관한 논의로 이어졌는데, 현재 많은 국가에서 다양한 인공지능 관련 프로젝트가 시행되고 있고, 정책을 수립하는 주체의 이에 대한 전문성도 더욱 정교해지고 있다. 안면인식 기술을 예로 들면, 이 기술은 발전되는 속도가 빠르고 활용될 수 있는 범위 또한 넓은 반면에 완벽한 기술이 아니어서 이것이 어느 수준까지 허용되어야 할지 등에 대해 정부 차원에서 논의되어 결정되어야 할 것이다. 이에 정부는 인공지능의 발전으로 인한 혁신과 일상에서 자동화되어가는 속도에 부합하기 위하여 증거 기반 프로세스(evidence-based processes)를 마련하는 것이 필요하다고 보았다.

이와 더불어 학계와 산업이 인공지능 기술의 발전과 인공지능이 미치는 영향에 대한 연구에 대해 어떠한 역할을 수행하여야 하는 질문에 대해서, Littman 교수는 최근에는 대학보다 산업 현장에서 더 최첨단(cutting-edge) 수준의 발전이 많이 이루어지고 있음을 설명하면서 실제로 NeurIPS와 같은 주요 컨퍼런스에서 발표되는 논문에 참여하는 산업체 소속 연구진의 수가 늘어나고 있음을 언급하였다. 다만, 산업 현장의 연구진은 사회 공공의 필요에 따라 움직이는 것이 아니라는 점이 우려 사항이 될 수 있다고 설명하였다.

그렇다면 인공지능의 향후 촉망받는 것에는 어떤 것이 있을지에 대한 질문에, 인공지능 기술이 앞으로 사람들과 효과적으로 일할 수 있도록 통합되는 것이 가장 장래성 있다고 보았고, 그와 반대로 인공지능 기술로 인하여 마주할 위험 중 임박한 것으로는 2016년 보고서보다 큰 비중을 차지하였다. 특히, 인공지능이 허위사실이 퍼지도록 이를 조장하고, 인공지능 시스템의 대부분이 대중의 다수를 위해 설계된 이유로 소외계층(underrepresented group)에 대한 리스트도 크다고 보았다. 예를 들어, PULSE라는 인공지능 기술은 GAN의 기술을 활용하여 낮은 해상도의 이미지를 높은 해상도의 이미지로 완성하는 기술인데, 입력한 이미지가 흑인(예 : 버락 오바마 미국 전 대통령)임에도 높은 해상도 이미지에는 백인의 이미지를 결과값으로 도출하는 등의 문제가 계속 있었다.

또한 인공지능이 사회경제 관계에 미칠 영향에 대한 논의에서는 여러 주제 중에서도 인공지능이 사람의 일자리를 뺏을 것인가에 대한 우려가 다루어졌는데, Littman 교수는 이는 앞으로도 지켜봐야 할 문제라고 하면서도 현재까지는 일자리의 불평등 및 실업률의 증가가 비단 인공지능 기술때문만은 아닌 것으로 보인다고 설명하였다. 최근 동향에서 살펴보면 인공지능 기술의 빠른 발전에도 불구하고 COVID-19 으로 인한 팬데믹 직전까지 미국 내 취업률은 지난 20년 동안 관찰된 수치 중 가장 높은

수치를 기록하고 있음을 통해 아직까지는 인공지능의 발전이 대규모 실업을 초래하지는 않은 것으로 볼 수 있다고 하였다.

열두 번째 질문으로는 인공지능 솔루션을 개발함에 있어 장기적인 개발 전략으로 어떠한 접근을 취하는 것이 더 좋은지에 대한 논의를 설명하였다. 두 가지 접근 방식이 있는데, 하나는 사람이 미리 규칙을 정립하고 이에 따라 시스템을 개발하는 방식(built-in mechanism) 그리고 다른 하나는 데이터를 제공하고 머신러닝 시스템이 스스로 솔루션을 개발하도록 하는 방식(learning-based system)이 있다. 기술이 발전할 수록 정확도가 높은 시스템을 만들어 나가는 데에는 후자의 방식이 많이 적용되고 있는데 이는 미처 예견하지 못한 실수를 할 수도 있고 그것이 인간에게 매우 해로운 결과를 야기할 수도 있기에, 해당 머신러닝 시스템이 어째서 그런 결정을 내렸는지에 대한 정당성이 필요하고 이를 위해 일정 수준의 규율이 정립될 필요가 있다고 보았다.

Littman 교수는 연구진이 다룬 두 가지의 워크샵 질문(Workshop question)을 설명하였는데, 첫 번째는 중대한 공공 의사결정에서의 인공지능 기반 예측(AI-driven predictions)이 어떤 모습이어야 하는지, 그리고 거버넌스와 이를 시행하기 위하여 정책 수립 주체가 고려해야만 하는 사회적, 조직적, 실무적인 사항이 무엇인지에 대한 논의이다. 즉, 인공지능을 통한 솔루션 체계를 설계할 당시부터 사회의 이해관계주체(stakeholders)들과 긴밀하게 협력하는 것이 중요하고, 만일 문제가 너무 복잡해서 인공지능의 도움이 필요하다면, 그것은 곧 인공지능 혼자서는 해결하지 못할 정도로 복잡하다는 것을 의미한다는 것을 간과하지 않아야 한다. Datasheets for Datasets, Model Cards for Model Reporting 등과 같은 연구를 통해 머신러닝 시스템의 입력값(input)과 출력값(output)을 적절하게 사용할 수 있도록 돕는다. 인공지능 시스템은 개발하여 활용하게 될 경우 영향을 미치는 영역이 존재하므로, 그 영향을 받는 사람들이 시스템을 만들어가는 데 참여해야 할 것이라고 보았다.

두 번째 워크샵 질문(Workshop question)으로는 물리적 또는 감정적인 도움(care)이 필요한 사람들에게 제공할 목적의 인공지능 사용과 관련된 것 중 가장 중요한 도전 과제와 기회가 무엇인지에 대한 논의가 다루어졌다. 특히 고령화 사회가 빠르게 진행되면서 노인 인구의 고립, 단절과 같은 문제를 해결하는 데 인공지능을 활용하고자 하는 시도가 많이 이루어지고 있는데, 이러한 경우 간단한 도움을 제공하는 등에 활용될 경우에는 많은 장점이 있을 수 있겠지만, 인공지능이 사람을 완전히 대체하게 되는 경우 사람과 사람 간의 관계가 갖는 가치를 떨어뜨리는 등의 문제가 있을 수 있다는 점이 지적되었다.

Littman 교수는 인공지능 기술의 놀라운 발전 속도를 언급하면서 그 중에서도 언어 또는 이미지를 처리하는 여러 작업에 큰 성과가 있었으며, 헬스케어 및 자율주행 자동차 분야의 발전이 주목받는다고 하면서도 이런 발전의 이면인 해당 기술로 인한 문제점을 고려하는 것이 시급하다고 하였다. 의사결정을 자동화하는 것에는 많은 리스크가 뒤따르기 마련이고, 오래된 데이터를 활용하는 것은 인공지능 시스템의 편향된 학습을 야기하기도 하며 사회 불평등을 오히려 더욱 악화시킬 수 있다. 그러므로 인공지능과 관련된 논의에는 기존의 과학기술 분야의 전문가 뿐만 아니라 사회과학 분야 전문가들이 활발하게 참여할 필요가 있다. 이와 더불어 각국 정부 역시 인공지능의 중요성을 인지하고 그에 부합하는 행동을 적시에 취할 필요가 있으며, 일반 대중 또는 시민들을 대상으로 하는 인공지능 관련 교육을 지원하여야 할 것이다. 또한 인공지능을 연구하는 커뮤니티는 기술이 갖는 장점을 지나치게 낙관적으로만 바라보도록 조장하는 것을 자제할 필요가 있고, 인공지능 시스템이 갖는 위험을 인지하여 이를 충분하게 소통하여야 할 것이다. 궁극적으로는 인공지능은 인간의 가치를 훼손하는 것이 아니라 그 가치를 높이는 데 초점을 맞추어야 할 것이라고 하였다.

이후 이어진 질의응답에서는 우선, 2021년을 기준으로, 인공지능에게 윤리규범을 학습시키는 최적의 방법은 무엇인지, 특히 윤리규범이 각 사회마다 다르다는 사실을 고려할 때 어떤 방법이 가장 적합할 지에 대한 질문이 있었다. 이에 Littman 교수는 본 보고서에 해당 내용을 담지는

않았지만, 개인적으로 그 주제와 관련하여 관심이 많음을 밝히고, 자신의 연구 분야 중 강화학습(Reinforcement learning)을 설명하면서 보상과 벌(rewards and punishments) 시스템을 통해서 인공지능에게 특정 행동을 가르치는 방법을 생각해 볼 수 있을 같은데, 가령 특정 상황이 좋은지 나쁜지를 판단할 수 있는 함수를 마련하고, 윤리 전문가들이 참여하여 각종 행동들의 값을 지정하는 방법(A 행동은 -10, B 행동은 +5)이 있다고 설명하였다. 그러한 값을 지정할 수만 있다면 보상 및 벌 시스템을 통하여 인공지능에게 어느 정도 수준의 규범을 학습시키는 것이 가능할 수는 있으나 값을 지정하는 것 자체가 사실은 어려운 문제이라고 할 수 있다. 또는 역(inverse) 강화학습 인공지능 시스템이 각 행동들을 관찰하여 직접 값을 학습하고 그 값들에 대응하여 어떻게 행동해야 하는지 학습하는 방식을 적용하는 것도 시도해 볼 수 있다고 하였다.

다음 질문으로는 인간이 인공지능에 지나치게 의존하게 될 수 있다는 우려에 대한 Littman 교수의 의견을 물었는데, 이에 대하여 아직 그런 상황에 대한 솔루션은 없는 상황으로, 인공지능 시스템들은 주어진 문제를 최대한 잘 해결하는데 초점을 두고 개발되어 온 것이고, 아직까지는 사람들과 어떻게 상호작용하면 좋을지에 대해서는 잘 알 수 없는 부분이라고 설명하였다. 따라서 지금은 인공지능이 제시한 문제해결 방안을 사람이 선택할지 선택하지 않을지의 문제일 뿐이고, 향후 인공지능 시스템이 발전하여 더 많은 부분이 설명가능해지고 투명해질 수 있다면 사람이 더 주도적으로 결정을 내리는데 도움을 줄 수 있고 인공지능에 의존하는 것을 줄일 수 있다고 하였다.

제6세션

머신러닝 기반 경쟁법 집행의 분석: 현황과 미래

Gleaning Insight from Antitrust Cases Using Machine Learning

발표자 **Ashwin Ittoo**

University of Liège, Management School

발표자

Ashwin Ittoo

제6세션

최근 인공지능의 관점을 법학에서도 중요하게 다루고자 하는 시도가 이루어지고 있는 가운데 Ashwin Ittoo 교수와 연구진이 발표한 머신러닝 기법을 적용한 경쟁법 사례 분석 연구에 대한 소개가 제6세션에서 이어졌다. 인공지능을 활용한 분석과 시사점의 도출이 주목받으면서 이를 법 또는 규범의 분야에서도 활용할 수 있는지에 대하여 주목하였는데, 특히 인공지능 기술이 시장에서 경쟁에 영향을 미치는 사례들이 관찰되고 그에 대한 분석과 연구가 지난 몇 년간 활발히 진행된 점을 보아 연구의 범위를 확장하여 반경쟁행위에 대한 경쟁당국의 집행 속도를 높일 수 있는지에 대한 질문을 바탕으로 연구가 수행되었다.

우선 인공지능이 사회 후생을 저해할 수 있다고 평가하는 '알고리즘을 이용한 가격 차별(price discrimination)'과 '묵시적 담합(tacit collusion)'에 대한 연구가 활발히 진행되었다. 먼저 가격 차별행위는 유사 혹은 동일 상품을 소비자에게 판매하면서 각각의 소비자에게 다른 가격에 판매하는 행위를 의미하는데, 인공지능 알고리즘이 소비자의 프로필과 소비 성향 등을 파악하여 이러한 가격 차별이 가능하다. 이 경우, 온라인에서 상품을 구매하는 소비자가 자신에게 제시된 가격 정보를 공개하여 찾아낼 수 있다. 다음 묵시적 담합행위의 경우에는 시장에서 서로 경쟁하는 판매 주체가 자신의 상품 가격을 결정하기 위하여 적용하는 가격 알고리즘이 장기간 반복 적용되면서 서로의 알고리즘에 대응하여 궁극적으로는 경쟁하는 것보다 경쟁사를 따라 자신의 가격도 인상하는 것이 더 이익이라고 학습하게 되면서 가격이 조정(coordinate)되게 된다. 그리하여 결과적으로는 담합 가격(즉, 카르텔 가격의 수준)과 비슷한 가격이 도출되나, 그 과정에서 명시적인 합의가 존재한 것은 아니므로 묵시적 담합으로 보았고 연구자들은 인공지능 기반 가격 알고리즘은 담합을 학습할 가능성이 높다고 보았다.

이와 같이 경쟁법을 집행하는 관점에서도 인공지능을 활용하여 사람이 파악하지 못하는 패턴 또는 시사점을 도출할 수 있는지에 대한 질문을 바탕으로 연구를 진행하였고, 구체적으로는 ① 인공지능 알고리즘으로 과거 반경쟁 사례를 탐색(explore) 또는 분석(analyze)할 수 있다면,

② 인공지능이 추가적으로 제공할 수 있는 통찰(insights) 이 무엇인지, 그리고 이렇게 인공지능이 제공한 통찰 중 ③ 어떤 것이 사람인 전문가가 발견하는 것을 입증(corroborate) 또는 부인(contradict)하는 지에 대한 질문을 제시하였다. 그리고 이 질문들을 바탕으로 수행한 머신러닝 기술과 반경쟁 사례를 분석한 연구에서는 데이터로써의 과거 반경쟁 사건들의 수집(collection of past antitrust cases)과 분석을 위한 머신러닝 알고리즘 그리고 데이터에 알고리즘을 적용하여 얻은 통찰(insights)로 연구 내용과 결과가 구성되었다.

첫째, 과거의 반경쟁 사건 데이터는 우선 미국의 판례에 초점을 두고 미국 법무부(Department of Justice)와 연방거래위원회(Federal Trade Commission)의 결정을 위주로 수집하였다. 그 중에서도 미 연방거래위원회는 규범 제정 및 공소권을 보유하고 있는 행정기관이며, 연방거래위원회법(FTC Act) 제5조에 따라 반경쟁행위에 법을 집행할 수 있기에 본 연구에서는 FTC 사례에 초점을 두었다. 사건 수집은 FTC의 웹페이지 상 저장 공간에서 수집되었고 수집 대상 사건의 기간은 2005년부터 2019년으로 설정하였다. 각 사건들에서 다음의 다섯 가지 변수(variables/features)들을 파악하였는데, ① 년도(year), ② 당사자명(proceeding name), ③ 산업(industry), ④ 조사 대상이 된 반경쟁적 행위(Anticompetitive conduct investigated), ⑤ 채택된 시정조치(remedies adopted)의 변수를 정리하였다. 연구 초반에는 영향 받는 시장(affected market), 시장 점유율(market share), 시장의 유형(market type, 독점시장 또는 양면시장)과 같은 다른 변수들도 고려되었으나, 과거 사례에서 이러한 변수들이 갖는 비중이 미미하여 이후 삭제되었다.

최종 데이터셋(dataset)은 연방거래위원회 법 제5조 관련된 72건의 사건으로 구성되었고, 각 사건은 앞서 언급된 다섯 가지 변수 이후 추가되어 늘어난 14 가지의 변수들로 각각 정의되었다. '산업' 변수에서는 컴퓨터 산업, 헬스케어(Healthcare)/제약(Pharmaceutical),

Ittoo 교수 발표 자료

전문(Professional)/무역(Trade) 협회(associations), 가스 및 석유, 통신, 교통 산업, 부동산 등이 있다. 반경쟁행위 항목에서는 연구진은 최대 4가지 변수(variables)만 정의될 수 있는데, 이는 하나의 사건에서 최대한으로 조사가 가능한 반경쟁행위가 4가지로 확인되었기 때문이다. 반경쟁행위의 값(values)으로는 다음과 같이 정리되었다:

(A) 배제적 행위(exclusionary conduct) ; (B) 약탈행위(predatory conduct) ; (C) 거래거절(refusal to deal) ; (D) 끼워팔기(tying conduct) ; (E) 가격 고정(price fixing) ; (F) 리베이트(rebates) ; (G) 차별행위(discriminatory practice) ; (H) 소비자 분할 합의(customer allocation agreement) ; (I) 지연지급(pay for delay) ; (J) 입찰 방해(disruption in the bidding process) ; (K) 합의 조정(agreement orchestration) ; (L) 담합에 참여 제안(invitation to collude) ; (M)

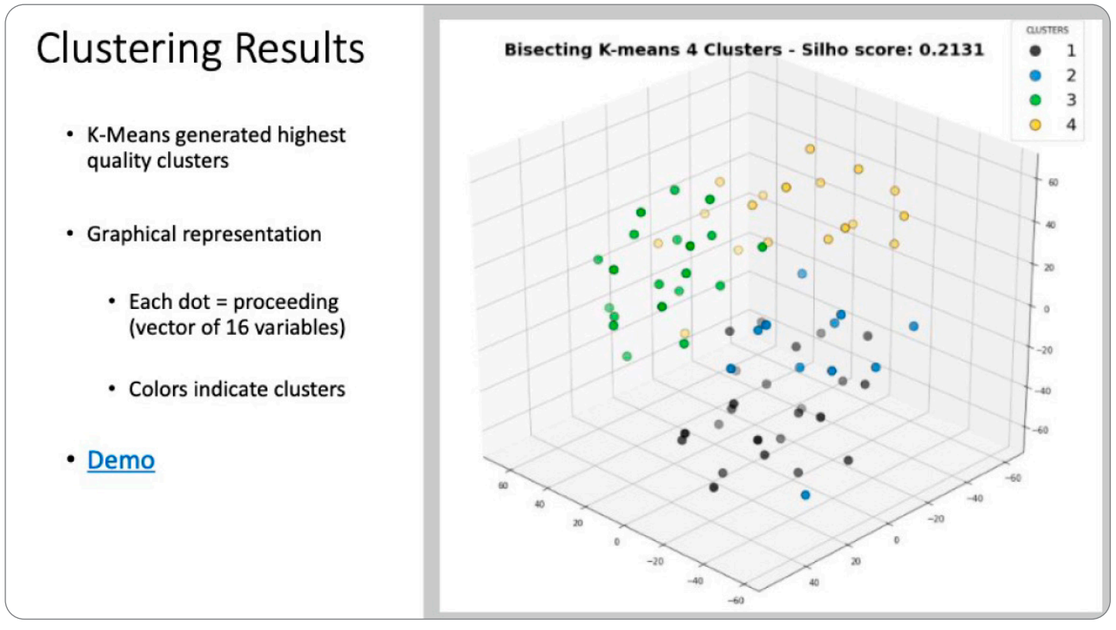
경쟁하지 않을 것에 합의(agreement not to compete)하는 행위 ; (N) 불법적인 정보 교환행위(unlawful exchange of information) ; (O) 동조적 합의(concerted practices) ; (P) 공모(conspiracy) ; (Q) 노포치(no poach) 행위 ; 시정조치 없는 행위(no remedy) (NOCOND)

위와 같은 변수와 값으로 데이터셋을 구성한 뒤에 그 다음단계로 알고리즘을 바로 적용하기 전에 데이터셋의 불일치성 등을 먼저 점검하기 위하여 탐색적 데이터 분석(Exploratory Data Analysis)을 실시하였다. 탐색적 데이터 분석을 통해서 데이터의 이상치(outliers), 불일치 값(inconsistencies, incoherencies), 관련 없는 변수(irrelevant variables), 또는 상관관계가 높은 변수(highly correlated variables) 등을 가려낸다.

두 번째로 분석을 위한 머신러닝 알고리즘은 데이터로 입력된 각 사례에서 근원적인 패턴 또는 규칙이 있는지 이를 찾고, 사안 별 유사성을 비교하고자 하였다. 본 연구에서는 예측을 목표로 하지는 않았는데, 예측을 위해서는 지도 학습(supervised learning) 이루어져야 한다고 보았다. 지도학습이 머신러닝의 지배적인 형태이며 이를 위해서는 방대한 데이터가 요구되고 예측하고자 하는 내용에 해당하는 라벨링, 주석 작업이 이루어져야 한다. 그러나 연구진은 이번 연구에서는 비지도학습(unsupervised learning)의 방식을 채택하여 다른 머신러닝 패러다임을 적용하였고 이를 통해 사건들 간의 패턴이 있는지 여부, 통찰의 발견 등을 의도하였다.

비지도학습에서 많이 쓰이는 방법은 클러스터링(clustering)으로 클러스터링 알고리즘에 대한 소개도 다루어졌다. 클러스터링 알고리즘의 기본적인 운영 전제(operating premise)는 ① 데이터 고유의 잠재 구조/패턴 식별, ② 데이터 포인트 간의 유사성을 도출하는 것이고, 현재 존재하는 클러스터링 알고리즘 중 연구진이 고려한 것은 ① K-Means, ② K-Mode, ③ Bisecting K-means 이다. 알고리즘을 적용한 프로그램은 파이썬(python)으로, 머신러닝 분석에 흔하게 쓰이는 방법이라고 설명하였다. 그리고 각 변수 간 유사성을 산출하기 위하여 클러스터링을 적용하였는데, 앞서 설명한 16 변수들이 벡터(vector)로 간주되어 16면 공간으로 투영하고, 각 사례 간 거리(유클리드 거리 Euclidean distance)를 계산한 뒤에 이를 그 거리에 따라 그룹화하였다.

이후 이러한 클러스터링이 갖는 의미에 대해서 우선 첫번째 클러스터를 통해서는 K-Means, K-Mode, Bisecting 데이터에 적용된 K-Means을 통해 유사한 절차를 갖는 값들의 클러스터를 형성하였고, '실루엣 스코어를 가지고 엘보우 방법(elbow method with silhouette score)'을 통해 클러스터링을 하였다. 클러스터링의 결과는 K-Means이 가장 고품질의 클러스터를 형성하였고(아래 그림 2 참조), 여기서 16가지 모든 면이 시각화되기는 어려워 각 면을 의미하는 점들이 포진하도록 하였다고 설명하였다. 각각의



Ittoo 교수 발표 자료: Bisecting K-Means 방식에 따른 4가지 clusters

점은 16가지 변수의 벡터값을 의미하고, 색은 클러스터를 뜻한다.

이와 더불어 Ittoo 교수는 모든 16가지 변수들이 같은 수준의 중요성으로 반영되었는지, 아니면 알고리즘이 몇몇 변수는 더 중요하게 고려했는지를 알기 위하여 두 가지 알고리즘인 Random Forest 와 Support Vector Machines을 사용하였다고 간단히 설명하고 이 클러스터가 갖는 경쟁법적 의미에 대한 설명을 소개하였다.

첫번째 클러스터를 통해서 확인한 것은 대부분의 사건 구성이 데이터/컴퓨터 산업과 헬스케어/제약 산업에서 이루어진 것이라는 점과 가장 많은 조사 대상 행위는 가격담합과 배제적 행위라는 것이었다. 그리고 가장 많은 시정조치로는 계약의 제한(contract limitation), 규범 준수 의무 부과(compliance obligation)으로 확인되었고, 클러스터1에서 도출할 수 있었던 결과는 알고리즘에 따르면 데이터/컴퓨터 산업과 헬스케어/제약 산업의 반경쟁 우려가 유사하게 증가했다는 것이다.

두번째 클러스터에서는 부동산 산업이 지배적인데, 사례에서 최소 둘 이상의 조사 대상 행위가 존재하였고, 공모(conspiracy)가 가장 많은 조사 대상 행위에 해당하였다. 시정조치에 관해서는 여기에서도 규범 준수 의무의 부과와 정보를 교환하는 행위의 제한(limitation in the exchange of information)가 이루어졌다. 클러스터2에서 도출된 결론은 연방거래위원회가 시정조치로 '규범 준수 의무'를 아주 흔하게 채택한다는 것을 학습하였다는 점이다.

세번째 클러스터에서는 대부분의 사례가 한 가지 조사 대상 행위를 포함하고 있었고, 가장 빈도가 높은 행위는 경쟁을 하지 않기로 하는 합의(agreement not to compete), 담합에의 참여 제안(invitation to collude)로 나타났다. 이는 통상 산업(trade industry)에서 우세한 사례들이었고, 여러 사례에서 시정조치가 이루어지지 않음을 확인하였다. 결론적으로 알고리즘이 연방거래위원회가 이 경우 하나의 행위를 조사한다는 것과 더불어 경쟁을 하지 않기로 하는 합의행위와 담합에의 참여 제안행위일 경우에는 시정조치를 부과하지 않는다는 것을 학습하였다.

네번째 클러스터는 주로 헬스케어/제약 산업의 사례들로 가장 빈번하게 조사대상이 된 행위는 가격담합, 배제적 행위, 공모이고 여기에 가장 많이 부과된 시정조치는 규범 준수 의무와 정보 교환의 제한이다. 네번째 클러스터에서 도출된 결론은, 알고리즘이 헬스케어와 제약 산업에서 가장 흔한 반경쟁행위는 가격 담합행위라는 것을 학습하였고, 또한 규범 준수 의무와 정보교환의 제한이 가장 많이 부과된 시정조치임을 학습하였다.

연구의 한계로 Ittoo 교수는 데이터셋이 상대적으로 작다는 점, 단일 행정기관인 연방거래위원회에서 발표된 사례로 구성되어 그 수에 제한이 있다는 점이 있지만, 이러한 데이터셋의 한계가 미치는 영향을 완화하기 위하여 탐색적 데이터 분석(EDA)를 활용하고, 3가지의 클러스터링 방식을 조사한 뒤 클러스터의 품질을 위해 실루엣 스코어(silhouette score)를 활용하였다는 점과 변수들의 중요도를 평가하였다는 점을 설명하였다.

마지막으로는 연구의 내용을 정리하면서 클러스터링을 통한 이러한 패턴의 발견과 분석이 추후 경쟁법 전문가들에게 통찰력을 제공할 수 있을 것이며 반경쟁 사건을 이해하는 알고리즘이라는 점에서 의미가 있지만 여전히 현재의 인공지능은 법관이나 반경쟁당국을 대체하는 것이 어려운 수준이라고 덧붙였다. 다만 법기술(legal technology) 분야의 관점에서는 의미 있는 연구이고 현재 진행하고 있는 연구로는 기업의 성장과 독점화 가능성의 예측을 소개하면서 마무리 하였다.

이어진 질의응답시간에는 일반적으로 인공지능에게 기대하는 부분이 예측에 관한 것인데, Ittoo 교수의 연구는 그와 다른 접근방식을 택하고, 인공지능이 설명해낼 수 있는 능력에 초점을 맞춘 것으로 보인다고 의견을 설명하면서 이러한 데이터 분석의 경우 인공지능이 사람보다 나은 점이 어떤 것이 있는지에 대한 질문이 있었다. 이에 Ittoo 교수는 데이터의 양이 많을수록 수동으로 이를 분석하는 데에는 시간이 많이 걸릴 수 있고, 전통적인 통계 방식으로만 할 경우 변수의 수와 차원의 수가 늘어나면 더더욱 분석이 어려워질 수 있다고 설명하면서 인공지능은 사람보다 객관적일 수 있다고 설명하였다. 추가로 연구는 기본적으로 알고리즘적 접근방식을 검증하기 위한

것으로, 어떠한 회사가 독점기업이 될 수 있는지 예측하기 위해 이를 구축하는 것으로 만일 새로운 사례가 생기면 과거의 어떤 사건이 이와 유사한 지 찾을 수 있고 이를 통해 분석할 수 있다고 하였다.

데이터를 위해 변수를 선택하는 과정을 보면 인공지능이 사람을 온전히 대체할 수 없을 것으로도 보이는데, 이러한 변수 선택이 어떻게 이루어졌는지에 대한 질문에 Ittoo 교수는 되도록 연방거래위원회의 사례들에 포함되어 있는 산업 분야에 같은 변수들이 나타날 수 있도록 설계하고 싶었고 두 가지 주요 이유로는 공정하고(fair) 균등하며(equitable), 일관성있는 변수의 분포가 모든 산업분야에 걸쳐 나타날 수 있도록 하는 것과 선택된 변수들은 알고리즘이 의미있다고 찾아낸 것들로 사람 법관도 새로운 사건을 검토할 때 이를 통해 평가할 요소들이라고 볼 수 있다고 하였다.

그리고 마지막에 소개된 현재 진행하고 있는 연구에 대해서 현재까지의 데이터에 의하면 '독점 기업'이 갖는 회사에 대한 내용이 잘 정립되어 있어 여러 변수와 추론을 활용하여 독점 기업에 대한 설명을 할 수 있을 것이라 보고, 추후 당국 또는 입법 주체가 사전에 어떤 기업들이 독점을 향해 가고 있는지 파악할 수도 있을 수 있다고 하며 마무리하였다.

제7세션

인간 행위를 위한 알고리즘 설명책임

Algorithmic Accountability for Human Agency

발표자 **김현섭**

서울대학교 철학과

발표자
김현섭

제7세션

제7세션에서는 서울대학교의 김현섭 교수가 알고리즘의 설명책임(Algorithmic accountability)이 무엇인지에 대해 설명하고, 그 필요성과 중요성에 대하여 소개하였다.

먼저 알고리즘의 설명책임의 개념에 관하여, 일반적으로는 '알고리즘의 출력(output)을 의사결정에 사용하는 사람이 그 의사결정에 의하여 이익 또는 손해를 입는 사람에게 알고리즘의 설계 및 운영에 대한 설명(explanation)과 이유(reason)을 제공할 의무'라고 정의되는데 여기에서 더 나아가 알고리즘의 출력을 의사결정에 사용하는 사람에게도 알고리즘을 설명할 필요가 있지 않은지, 사람이 알고리즘의 입력값(input)에 영향을 주는 경우에, 알고리즘을 설명할 필요가 있지 않은지 등과 같은 의문을 제기하고 제7세션에서 그 의문들에 대한 답을 논의하였다.

현실에서 알고리즘의 설명책임이 잘 지켜지지 않는 경우가 많은데, 그 이유로는 ① 알고리즘을 투명하게 공개할 경우, 영업의 자유 또는 영업비밀을 침해할 수 있기 때문에 알고리즘 개발에 투자할 유인을 감소시키기도 하고, ② 소스코드가 공개될 경우 다른 사람들이 그 알고리즘을 부당하게 이용, 변경 또는 파괴할 가능성이 있으며, ③ 학습 또는 검증에 쓰인 데이터를 공개할 경우에는 해당 데이터 주체들의 프라이버시가 침해될 위험이 있고, ④ 많은 경우에 복잡한 알고리즘은 개발자도 알고리즘이 어떻게 그러한 output을 내놓는지 알지 못한다.

김현섭 교수는 사람과 인공지능이 상호작용하는 방식을 4가지로 분류하였는데, 첫째 알고리즘의 출력이 사람의 의사결정과 그에 따른 행위에 영향을 주는 경우(예 : 알고리즘 출력값이 소비자의 구매결정에 영향을 미치는 경우), 둘째 알고리즘의 출력이 사람의 손익에 영향을 미치는 경우(예 : 대출을 상담 받을 때 알고리즘이 대출 여부를 결정하는 경우), 셋째 알고리즘의 입력(설계와 학습 과정)에 있어서 사람의 의사결정이 영향을 미치는 경우(예 : 사람이 데이터에 라벨링을 할 때 가치판단 등이 반영되어 알고리즘의 입력에 영향을 주는 상황), 그리고 네번째는 알고리즘의 입력에 사람의 정보가 영향을 미치는 경우(예 : 사람의 데이터가 알고리즘에 사용되는 경우)로 나누었다. 위

김현철 교수 발표 자료

4가지 방식이 중복되는 경우도 빈번하다. 예를 들면 사용자의 브리우징 또는 클릭 수에 따라 추천내용을 실시간으로 업데이트하는 알고리즘의 경우에는 첫 번째 경우와 세 번째 경우에 동시에 해당한다고 할 수 있다.

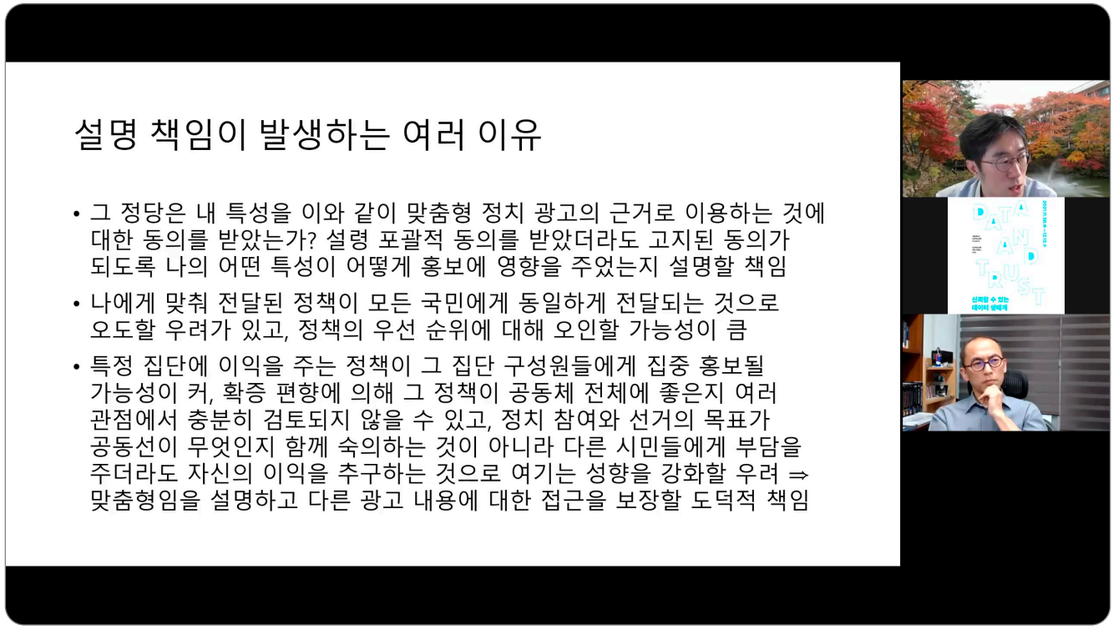
그런 방식으로 알고리즘이 개인의 선택을 반영하고 유사한 사람들의 선택을 반영하여 해당 개인에게 상품을 추천한다면 개인의 기존 취향과 선호를 더 강화하는 결과를 가져올 수 있다. 여기서의 의문은 개인이 선호하는 것을 더 효율적으로 찾을 수 있다면 그것이 이익일 수 있는데 어째서 문제되는가라는 점이다. 여기에는 다음과 같은 문제가 있다고 볼 수 있다. 즉, ① 정보원천인 개개인이 그에 대하여 동의하였는가(개인의 특성에 대한 정보를 수집 및 처리하여 추천의 근거로 활용하는데 동의하였는가)라는 문제와 ② 알고리즘의 추천이 개인의 의사결정에 과도한 영향을 주는 ①의 경우와 같은 때 개인이 해당 웹사이트에 가입할 당시 정보를 제공하는 데에는

동의를 하였더라도 그 정보가 어떤 방식으로 활용될지, 또는 사용자의 어떠한 특성이 어떻게 추천 내용에 영향을 줄지 등에 대한 고지된(informed) 동의라고 보기 어렵기 때문에 진정한 의미의 동의가 될 수 있도록 그에 대한 설명 책임이 필요하다.

②의 경우에는 알고리즘이 개인의 기존 취향과 선호를 강하는 추천을 할 때 해당 개인이 실제로 접근할 수 있는 선택의 범위가 줄어들고 시야가 편협해질 수 있다는 우려가 제기된다. 나아가 이러한 것들이 타인의 생활양식 또는 견해를 이해하는 능력과 동기가 약해질 수 있고, 더 나아가 사회 분열의 결과로도 이어질 가능성이 있다. 온라인 마이크로타게팅(micro targeting)을 활용한 선거광고를 예로 들어보면, 유권자들이 공동의 선보다 자신의 이익만을 추구하는 성향이 강화되는 결과로 이어질 우려가 있으므로, 알고리즘이 개인의 어떠한 특성을 반영하여 특정 소재 또는 스타일의 광고를 추천하게 되었는지에 대한 설명 책임이 필요할 뿐 아니라 그것이 맞춤형 광고라는 사실 자체를 설명하고, 그와 다른 광고 내용에 접근을 보장할 도덕적 책임도 있을 것이라 보았다. 이와 더불어 알고리즘의 출력값은 각 변수들(즉, 사용자의 특성)에 부여된 가중치에 따라 좌우되기도 하는데, 각 변수들의 가중치 설정에는 알고리즘 설계자의 가치판단이 개입되게 되므로 사용자의 가치판단과 다를 가능성이 있음 또한 설명하고 사용자에게 그 정보에 대한 접근을 보장할 책임이 있다고 보았다.

사람들이 왜 중요한 의사결정을 갈수록 알고리즘에 맡기려는지에 대해 고민해보면 평소 개인이 의사결정할 때 겪는 불확실성과 오류가능성에서 비롯되는 불안 및 스트레스를 줄이고 싶기 때문일 것이다. 특히, 우리가 알고리즘의 추천을 따르지 않고 의사결정을 직접했을 때 실패를 경험하는 일이 잦아지고 그와 동시에 알고리즘의 추천이 신뢰할만하다고 인식될 경우에는 더욱 더 알고리즘에 의사결정을 아웃소싱(outsourcing)하게 될 것이다.

그러나 인생의 중요한 결정에 대해 스스로 고민하는 일의 가치에 대해 생각해 볼 필요가 있다고 김현섭 교수는 지적하였다. 자기 자신에 대해 그리고



김현철 교수 발표 자료: 설명책임 발생 이유

자신을 둘러싼 외부 환경에 대해 스스로 고민하고 성찰하고 결정할 수 있을 때에 자신의 가치관과 자아관이 형성되는데, 알고리즘이 정해주는 것에 따르는 삶을 살게 되면 가치관, 자아관의 혼란과 갈등의 문제를 겪을 위험이 있다. 이러한 관점에서 보았을 때도 인공지능의 설명책임이 갖는 중요성을 알 수 있는데, 인공지능이 의사결정에 대한 설명을 분명히 한다면, 인간이 인공지능의 의사결정에 따르더라도 그러한 혼란과 갈등에 대한 문제가 덜 할 수 있다. 설명이 제공되면 사람이 충분히 이해하고 분석해봄으로써 보다 더 주체적으로 결정할 수 있기 때문이다. 만일 이것이 어렵다면 적어도 알고리즘에 의한 예측을 과신하지 않도록 추천 순위의 오류 가능성이 충분히 설명되어야 할 것이다.

이후 이어진 질의응답시간에는 알고리즘 활용으로 인해 인간의 자율성이 감퇴되는 것과 관련하여 설명책임이 일부 도움은 되겠지만 다른 근본적인 대책이 필요한 것은 아닌지에 관한 질문에 대하여 김현섭 교수는 알고리즘의 설명책임이 근본적인 대책이 되지 않는다는 점에서는 긍정하지만 인간의 자율성이 다 감퇴되지 않기 위해 알고리즘에 설명책임을 부과하는 것을 고려해보자는 취지로, 화두를 던지는 수준으로 이해할 것을 제안하였다.

다음으로, 단골 이발사에게 머리를 맡기며 단순히 '잘해주세요' 라는 표현만으로도 가능하듯이, 제한된 자원을 활용하는 삶 속에서 알고리즘의 결정을 따르는 것이 현실적으로는 최선의 선택지로 볼 수 있지는 않는지에 대하여, 삶의 모든 결정을 일일이 깊이 숙고하며 살 수는 없기에 적절한 trade-off 는 필요하다고 인정하면서 어느 범주까지 알고리즘에 맡겨도 될지에 대해서는 규범적으로 생각해 볼 필요가 있다는 의견을 설명하였다. 머리를 손질하는 것과 같은 결정이 누군가에게는 가벼울 결정이라고도 볼 수 있는데, 한 편으로는 가볍다고 평가되는 결정들을 모두 알고리즘에 의존하다 보면 나의 숙고 능력 감퇴를 가져오게 될 수 있고, 정작 중요한 결정에서 필요한 숙고에 요구되는 훈련이 잘 되지 못할 수 있다는 우려의 관점에서 설명한 것이라고 덧붙였다.

제8세션

의료 인공지능의 법적 쟁점

AI in Healthcare:
Legal and Regulatory challenges

발표자 **박혜진**

한양대학교 법학전문대학원

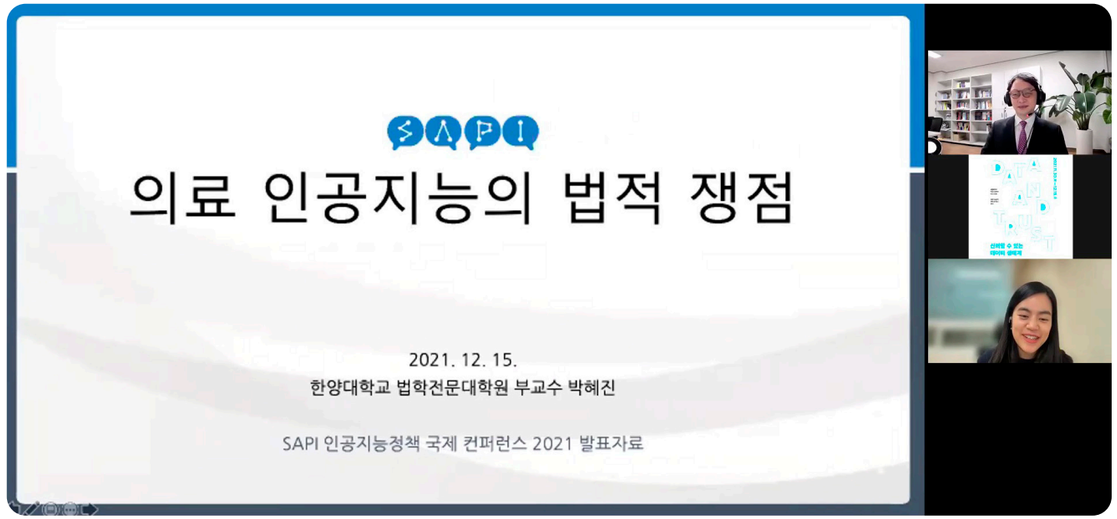
발표자
박혜진

제8세션

마지막 세션에서는 인공지능 의료기기에 대한 규제에 대해 미국과 한국에서 이루어지고 있는 논의를 소개하고, 임상에서 인공지능이 도입되었을 때 환자에게 악결과가 나타났을 경우 민사소송에서의 논의를 소개하였다. 우선, 의료 인공지능의 발전 배경은 2012년 유명 미국 벤처투자자 Vinod Khosla가 컨퍼런스에서 '기계가 의사의 80%를 대체할 것이다'라고 한 발언에 대해 의사들의 강한 반발이 있었다. '실리콘 밸리는 의학을 이해하지 못한다'는 등의 기존 의사들의 반응에 상반되는 모습으로 의료 분야의 인공지능 기술은 비약적으로 발전하였고 임상에도 빠르게 진입하고 있다. 특히, 컴퓨터 비전(Computer Vision)이 활용된 영상의학 분야에서의 인공지능 활약은 두드러지고 있고, 최근 Google Health에서 학습한 인공지능 시스템이 사람 의사보다 유방암의 예측을 더 잘한다는 연구도 나왔고 인공지능 시스템이 심장마비의 발생 위험도에 대한 예측도 월등하다는 평가의 연구도 나왔다.

인공지능 의료기기의 개념은 문자 그대로 인공지능 기술을 활용한 의료기기를 의미하고, 그 특징으로는 다음의 네 가지가 주로 고려된다. 먼저 ① 불투명성(opacity)으로, 기기의 입력값과 출력값 간의 관계를 설명하는 것이 어렵다는 점으로 이것이 의사 책임 관련된 문제로 이어질 수 있다. ② 적응성(adaptability)이란 새로운 데이터가 투입되어도 인공지능에 변동이 없는 경우(locked)가 있는 한편, 새로운 데이터에 따라 변동(adaptive)이 있는 경우가 있는데 이러한 적응성이 문제가 될 수도 있고, ③ 편향성(bias)으로 인해 소수 인종/여성 등에 대한 의료기기의 성능이 저하될 가능성이 있다. 그리고 ④ 사이버 보안위험(cybersecurity risk)이 있는데, 인공지능 의료기기는 인터넷 또는 센서에 연결되어 사용되는 경우가 많은데, 이런 경우 시스템의 취약점을 노리는 경우가 있을 수 있고, 해킹 등으로 인해 시스템이 잘못된 결정 또는 권고결정을 내릴 경우에 의료기관의 과실 판단이나 제조업자의 책임 등과 관련될 수 있다.

미국의 경우, 인공지능 의료기기의 위와 같은 특징을 고려하여 이를 어떻게 규제할 것인지에 대한 논의가 이루어져왔는데, 관련 기관은



박혜진 교수 발표 자료

연방식약청(FDA ; Food and Drug Administration)이다. FDA는 1970년부터 의료기기를 정의하는 규정을 두고 있는데, 본 규정에는 실물기기 하드웨어만 포함되어 있었기에 2015년에 기존의 의료기기 정의 범위를 확장하여, 소프트웨어(SsMD)도 포함되었다. 우리나라도 2018년 국제기준을 반하여 각종 의료 목적 소프트웨어(예 : 의료용 빅데이터를 기반으로 의료정보를 분석하여 얻은 임상정보를 이용하여 환자의 질병 유무, 상태 등에 대한 가능성 정도를 자동으로 진단, 예측, 모니터링하거나 치료하는 소프트웨어)를 의료기기목록에 추가하였다.

의료기기에 대한 기존의 규제 역시 하드웨어 의료기기 위주에 맞춰져왔기에 소프트웨어까지 확장이 될 필요가 있고, 미국은 2017년 FDA가 'Pre-Cert'라는 소프트웨어 의료기기의 시판 이전 사전인증제도를 마련하겠다는 계획을 공개하고 시범적으로 파일럿 프로그램을 마련하여 9개 업체를 선정하여 실시하였다. 이는 제품 자체에 초점을 두기보다는 품질기준을 충족하는 기업을 사전에 인증하고자 하는 것인데, 사전 인증을 받은 기업의 경우, 시장에 제품을 출시하기 전에 제출자료 일부를 면제받는 등의 혜택이 있다. 국내에서도 2019년 「의료기기산업 육성 및 혁신의료기기 지원법」을 제정하면서 혁신의료기기소프트웨어 제조기업 인증제를 도입함으로써 이와 유사한 접근방법을 채택하였다.

인공지능 의료기기 규제와 관련된 또 하나의 과제는 실시간으로 적응하고 변화하는 적응형 알고리즘(adaptive algorithm)을 어떻게 규제할 것인지에 관한 것으로, FDA는 전 생애 주기 접근방식이라는 형식하, 시장에 출시하기 전에 미리 시장에 나온 이후 발생할 수 있는 수정사항과 재훈련, 업데이트 등의 방법론을 포함하여 사전 변경사항 통제 계획(Pre-determined change control plan)의 심사를 미리 거치도록 하는 제도를 논의 중에 있다. 2020년 미국에서 허가 인증을 받은 인공지능 의료기기가 100 건 정도 있고, 우리나라에서는 50건 정도 인증된 것으로 알려져 있다. 시장 규모의 차이를 감안한다면 우리나라가 의료기기 인공지능 시장에 상당히 적극적으로 나서고 있다고 볼 수 있다.

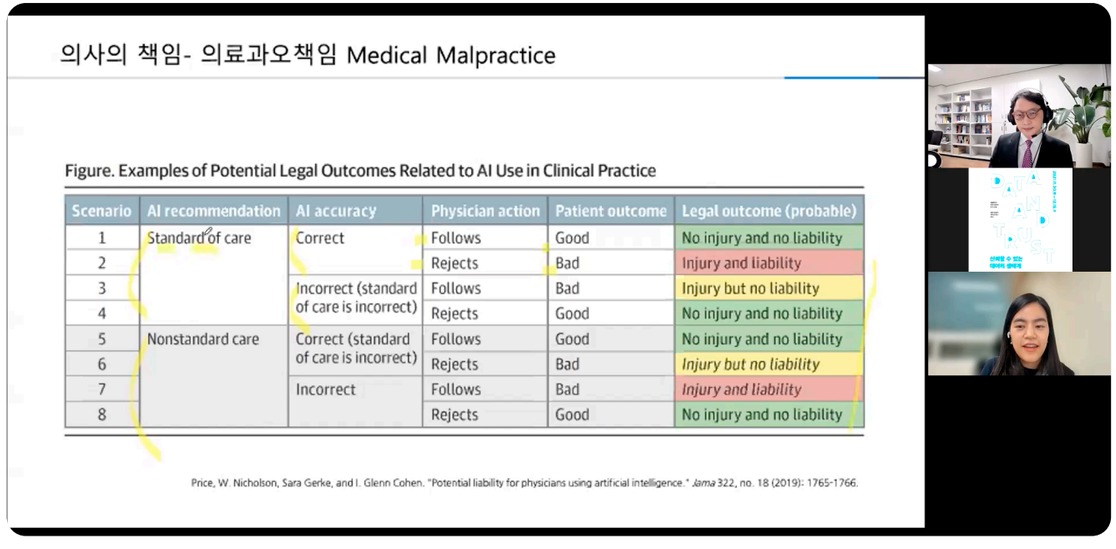
그러나 한 번 허가를 받은 것으로 모든 절차가 완료되는 것은 아니고 보험 급여 여부를 확인, 신의료기술 평가 등의 절차를 거쳐야 하는데, 보험 수가를 인정받는지 여부에 관한 것은 인공지능 의료기기가 임상에 도입되는데 또 하나의 장애물로 인식되고 있다. 보험 수가를 인정받지 못하면 의교기관이나 의사가 비싼 비용을 들여 인공지능 의료기기를 도입할 유인이 줄어들기 때문이다. 보건복지부와 신용평가원에서는 가장 활발하게 개발되는 AI 의료기술인 영상의학기술, 3D 프린팅 의료기술 등에 관한 건강보험 중재 및 보상에 관한 가이드라인을 마련하였다. 가이드라인의 주요 내용은 환자 치료에 도움이 되거나 비용 절감이 입증된 경우에 별도로 수가를 인정하겠다는 것으로, 해당 가이드라인에서 AI 의료기술을 그 효과에 따라 4단계로 분류하였다.

제1단계는 판독활동 보조를 통한 업무효율 증가로 의료기관의 부가적 이익창출 또는 간접비용 감소가 가능한 기술, 제2단계는 기존 행위와 비교하여 유사한 수준의 진단능력을 보이거나, 일부 능력을 개선이 있으나

전체적으로 기존 행위 유사 수준, 제3단계는 기존 행위 대비 명백한 진단능력 향상, 새로운 진단적 가치 창출 또는 일반적 역할 범위 외 새로운 정보 제공을 하는 의료기기이다. 제4단계는 제3단계에 더하여 비용효과성을 입증한 의료기기를 분류한다. 그러나 현재 수준의 인공지능 의료기기는 별도 보상이 고려되는 제3단계, 4단계 요건을 만족하기가 쉽지 않아 별도 수가를 받은 인공지능은 아직 없는 것으로 보인다.

박혜진 교수는 인공지능 의료기기가 도입되더라도 문제가 발생할 가능성이 있다고 하면서 그 중에서도 인공지능 의료기기를 둘러싼 민사책임의 소재 즉, 환자의 건강에 악결과가 생기면 누가 책임을 질 것인지(책임 분배의 문제), 그리고 유용한 인공지능 기술이 임상에 사용될 수 있을지에 대해서도 영향을 미친다 할 것이다. 따라서 기술의 혁신을 저해하지 않는 선에서 책임에 대한 규제가 도입되어야 할 것이다.

우선 의사가 책임을 부담하는 경우에 대하여 살펴보면, ① 의료 과오 책임, 그리고 ② 설명 의무에 대한 책임으로 나눌 수 있다. 첫번째 경우인 의료 과오 책임의 경우와 관련하여 2019년 논문에서 정리한 표와 함께 설명하였다.



박혜진 교수 발표 자료: 의사의 책임

해당 논문에서는 인공지능 의료기기는 기존의 의사결정을 확인하는 정도의 목적으로만 쓰는 것이 의사들에게 제일 안전하다고 분석하고 있다. 예를 들어 6번째 경우인 인공지능의 권고가 옳은 권고였는데 이를 따르지 않았을 때는 법적 책임을 지지 않는 반면에, 잘못된 권고를 따른다면 환자에게 악결과가 발생하였을 때 이에 대한 책임을 지게 된다. 그러므로 의사가 인공지능 의료기기의 권고가 옳은지 옳지 않은지 명확하지 않을 때에는 이를 따르지 않는 것이 의사들에게 안전한 것이라 한다. 사실 인공지능의 권고가 옳은지 그른지 충분히 검증하게 이에 따랐다면 책임을 면할 수 있어야 하는데 블랙박스의 경우에는 이러한 입증이 어려워 그러한 면책 또한 어렵다. 그러므로 이에 대한 가이드라인이 마련되어 할 것이다.

한 연구에서는 다가올 미래에 의료 인공지능 기기가 급속도로 발전함에 따라 인공지능 기기가 잘못된 결정을 내렸을 때 이것이 잘못되었다고 판단할 수 있는 의사의 수도 감소할 수 있다고 주장하였다. 이러한 상황을 피하기 위해서는 진단에 의미있는 부분에는 의사가 참여하도록 의무화해야 한다고 보았다. 지금은 의료기기의 결정을 참고해서 결정을 내리는 것이 주 임무라 한다면, 앞으로는 인간 의사보다 성능이 좋은 의료기기가 개발되어도 인공지능은 인간이 하지 않는 실수를 하기 때문에 인간 의사가 이를 적절하게 평가하고 결과를 검증할 의무를 지고, 의학적 근거가 있다면 인공지능의 권고에 이견을 낼 것을 의무화하여 인공지능의 권고를 의사가 그대로 따르는 것이 의무를 회피하는 것이 되지 않도록 해야 할 것이다.

의사의 설명의무에 관하여서, 미국의 경우에는 설명의무에 위배되거나 위험이 실현되거나, 또는 설명하였더라면 동의하지 않았을 것이라는 점이 인정될 경우에 전적인 배상을 명하고 있다. 그러나 우리나라는 항상 두 단계로 나누어 설명하는데 위자료 혹은 모든 손해의 단계로 설명한다. 즉, 만일 의사가 설명하였더라면 동의하지 않았을 것이 인정될 경우에는 모든 손해의 청구가 가능하나, 그것이 인정되지 않을 경우 위자료 청구만 가능하다. 최근 한 연구에서는 의료행위에 인공지능이 활용되었음을 환자에게 알리지 않은 경우에도 대부분의 경우는 설명의무를 위반한 것이라 볼 수 없다고

분석하였으나, 국내에서는 인공지능의 결정에 불확실성이 존재하기 때문에 오히려 설명 의무가 강화되어야 한다는 의견이 있다. 그러나 인공지능 의료기기가 진단에 있어 보조역할만 하는지, 아니면 치료법도 권고하고 이를 주도하는지 또는 인공지능이 월등하다고 생각하여 인공지능에게 결정을 위임하고 있는지 등에 따라 판단기준이 나뉠 것으로 보았다.

다른 한 편으로는 의료기관도 의사를 고용한 사용자로써 불법행위책임을 질 수 있는데, 의료기관이 자신의 관리감독의무(예 : 의사를 고용하고 적절히 관리감독할 의무, 시설을 관리감독할 의무 등)를 위반 할 경우 이에 대한 책임을 질 수 있다. 제조업자도 마찬가지로 책임을 부과할 수 있는데 여기에 문제가 존재한다. 제조업자의 책임을 묻기 위해서는 제조물이 있어야 하는데, 소프트웨어가 제조물 또는 가공된 동산으로 볼 수 있는지에 대한 논란이 있고, 한국과 미국에서 모두 아직 논의가 정리가 되지 않았다. 만일 제조물이라고 보게 될 경우에 책임은 ① 제조상 결함이 있는가(예 : 원래 의도한 설계와 제조/가공이 다르게 되었는지 여부). ② 설계상 결함(예 : 제조업자가 합리적인 설계방안을 마련하지 못해 안전성이 결여된 건지), 그리고 ③ 표시상 결함(예 : 합리적인 설명, 지시, 경고 등 최종 소비자인 의사에게 기타 표시를 하지 못하여 그로 인한 피해가 인정되는 경우)으로 인한 책임으로 나눌 수 있다.

보험의 역할도 간과될 수 없는바, 의사와 의료기관은 의료배상책임보험에 가입함으로써, 제조업자는 제조업자배상책임보험 가입을 통해 위험을 분산시킬 수 있다. 그 밖에도 병원과 의사 간, 제조업자와 의사 간 계약을 통해 위험을 분산시킬 수 있고, 기금을 통해 배상하는 방법도 고려된다. 그러나 이로 인해 더 뛰어난 성능의 개발 유인이 감소하는 영향도 고려되어야 할 것이며, 지금은 사람 의사와 인공지능이 어떻게 서로 보완하여 수준 높은 의료서비스를 제공할 수 있을지를 고민할 단계라고 보았다.

서울대학교 인공지능정책 이니셔티브
(SNU AI Policy Initiative: SAPI)

서울대학교 인공지능정책 이니셔티브는 인공지능과 관련된
다양한 사회경제적, 법적, 정책적 이슈들을 연구하고 논의하기 위해
시작된 서울대학교 법과경제연구센터의 프로그램입니다.
SAPI는 연구 목적의 '소셜 랩(Social Lab)'을 지향하여, 여러 배경과
관심을 가진 분들 사이의 협업과 지속적인 대화를 추구합니다.
현재 서울대학교 법학전문대학원의 고학수 교수와 임용 교수가
함께 이끌고 있습니다.

http://sapi.co.kr

SAPI 2021 SEOUL
AI POLICY CONFERENCE

서울대학교 인공지능정책 이니셔티브
2021 인공지능정책 국제 컨퍼런스

DATA and TRUST:
Analytics, Ethics & Privacy

신뢰할 수 있는 데이터 생태계