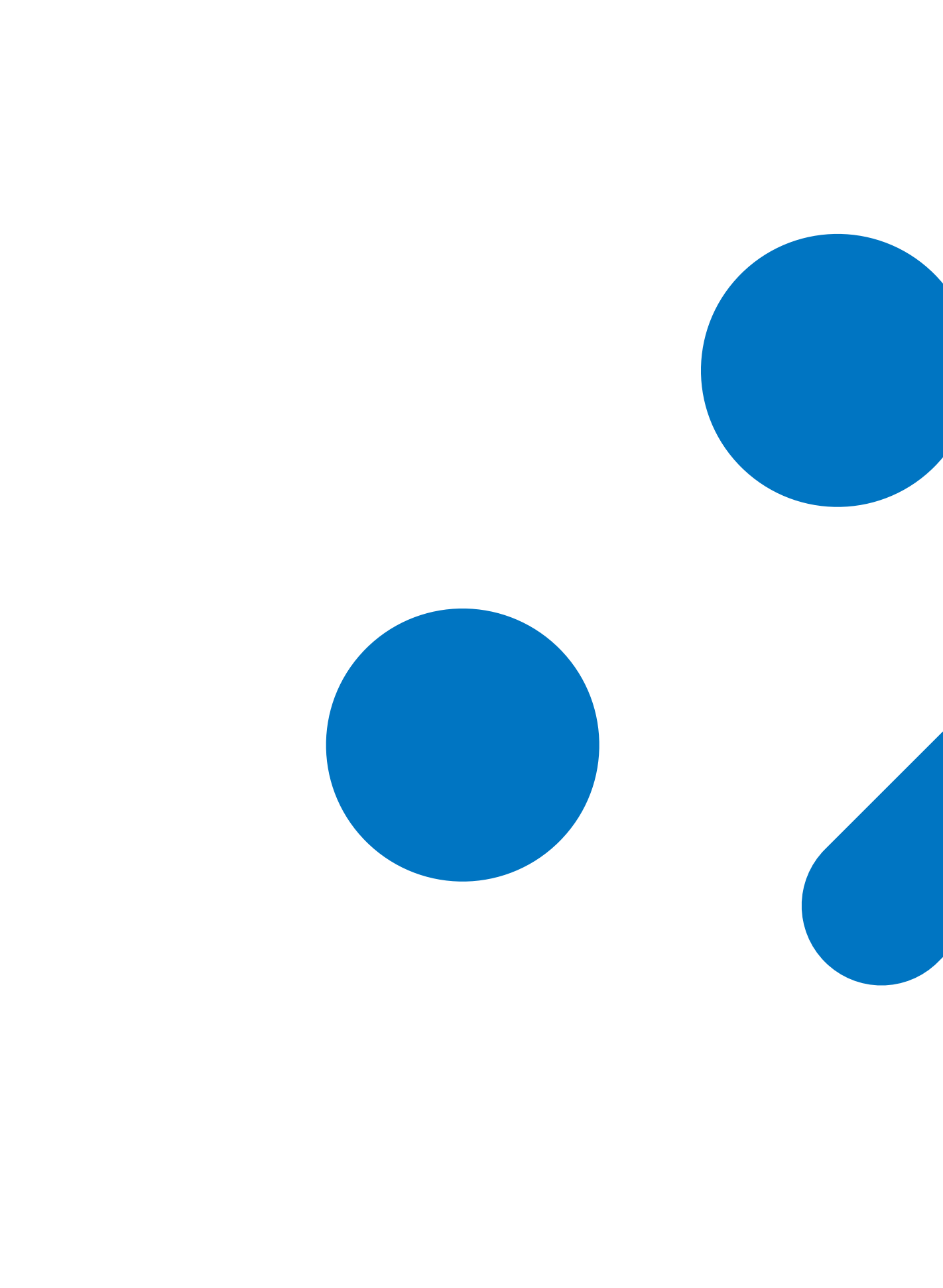
AI and Market Dynamics

2020 SEOUL
AI POLICY CONFERENCE
– AI and Market Dynamics

AI와 시장

목차



2020 SEOUL AI POLICY CONFERENCE – AI and Market Dynamics

머리말

투명하고 공정한 시장의 운용이 혁신의 원천으로 기능한다는 것은 두말할 나위가 없는 사실이다. 불투명하고 불공정한 시장은 경쟁을 저해하고 권력의 비대칭성을 심화하여 경제주체가 창조적 도전에 나설 의지를 상실케 하기 때문이다. 여기서 제기해볼 수 있는 한 가지 물음은 인공지능(Artificial Intelligence, AI)이라는 신기술의 출현이 시장의 투명성과 공정성에 어떤 영향을 유발할 것인가 하는 것이다. 만일 인공지능이 시장의 투명성과 공정성에 부정적 영향을 미칠 것이 예상된다면, 우리는 서둘러 인공지능 시대에 시장이 제 기능을 담당할 수 있도록 도와주는 법정책적 대안을 모색해야 한다.

이러한 문제의식에서, 서울대학교 인공지능 정책 이니셔티브(SNU AI Policy Initiative, SAPI)는 지난 2017년부터 인공지능 분야의 전문가를 초빙하여 법정책적 이슈를 논의하는 컨퍼런스를 매년 개최해오고 있다.[1] 지난 컨퍼런스 모두 한글자막을 첨부한 영상이 제작되었고, 2018년 제2회 컨퍼런스부터는 현재와 같은 형태의 보고서도 제공하고 있다.[2] 서울대학교 인공지능 정책 이니셔티브는 이외에도 인공지능과 관련된 다양한 논의를 융합적 관점에서 수행하고

1 과거 컨퍼런스의 대주제는 다음과 같다.
제1회 컨퍼런스: “인공지능, 알고리즘, 개인정보보호를 둘러싼 정책적 과제 (Policy Issues surrounding AI, Algorithms & Privacy)”
제2회 컨퍼런스: “인공지능의 시대: 기술 발전에 따른 책임과 규제 (Artificial Intelligence Today: Governance and Accountability)”
제3회 컨퍼런스: “미래를 향한 인공지능 정책: 우리는 AI를 신뢰할 수 있을까? (AI Policy for the Future: Can We Trust AI?)”

2 과거 컨퍼런스 영상의 주소는 다음과 같다.
제1회 컨퍼런스: http://www.youtube.com/playlist?list=PLOP6ilKzhDLQ_a2hMmD0vxsJn0d-aQco8
제2회 컨퍼런스 이후: https://www.youtube.com/channel/UCKyxSZOtLB1YvkKM2_Mq8gQ/featured
과거 컨퍼런스 보고서의 주소는 다음과 같다: http://sapi.co.kr/workshops/

있다.[3] 시장경쟁의 문제는 워낙 중요한 주제라서 이미 지난 컨퍼런스에서도 부분적으로 논의가 이루어졌지만, 올해는 컨퍼런스의 대주제로 삼아 본격적으로 논의를 진행하였다.

예년과 달리, 2020년 제4회 컨퍼런스에서는 형식상의 변동이 있었다. 올 초부터 세계를 뒤흔들고 있는 코로나19 바이러스로 인해 오프라인 행사가 불가능하였기 때문이다. 이에 따라, 서울대학교 인공지능 정책 이니셔티브는 온라인 행사를 추진하면서 세션의 개수를 8개로 대폭 확대하고 매 세션을 다른 일자에 1시간 이상 배치하여 기조강연 이외에도 모든 세션에 걸쳐 충분한 논의가 이루어지도록 하였다. 동시에 시장의 공정성과 투명성이라는 대주제에 모든 세션에서의 논의가 하나로 수렴할 수 있도록 만전을 기하였다.

형식상의 변동에도 불구하고, 기조강연을 포함한 모든 세션의 발표자를 해외의 관련 분야 권위자를 초빙하여 예년과 마찬가지로 풍부하고 깊이 있는 논의가 이루어질 수 있었다. 기조강연을 담당한 Cynthia Dwork 교수는 개인정보 보호와 데이터 분석을 조화하는 선도적 기술로 널리 알려진 차등 프라이버시(differential privacy)의 창시자로 유명하다. 이외에도 개별 세션을 담당한 모든 발표자들 역시 유럽연합과 미국의 명망 있는 연구자들로 구성되었다. 특히 이번 컨퍼런스에서는 AI Now Institute, Partnership on AI처럼 세계적으로 널리 알려진 기관에 소속된 연구자들이 참여하여 논의의 폭을 넓혀주는 데 기여하였다. 나아가 인공지능 분야를 선도하는 기업인 Google에서도 발표를 담당하여 산업계의 시각도 포괄할 수 있게 하였다. 이에 여러 날에 걸친 온라인 행사임에도 불구하고 매 세션마다 수백여 명의 참석자들이 자리를 빛내주었다.

---

3 서울대학교 정책 이니셔티브는 2019년부터 5월과 11월 두 차례에 걸쳐 이슈페이퍼를 발간하고 있고, 이외에도 비정기적으로 다양한 학술행사를 개최하고 있다. 자세한 내용은 http://sapi.co.kr/

제1세션

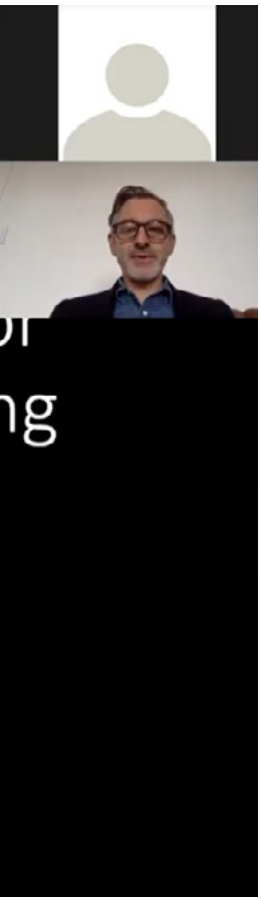
인공지능 정책의 디자인: 도전과 기회

Challenges and Opportunities for AI Policymaking

발표자

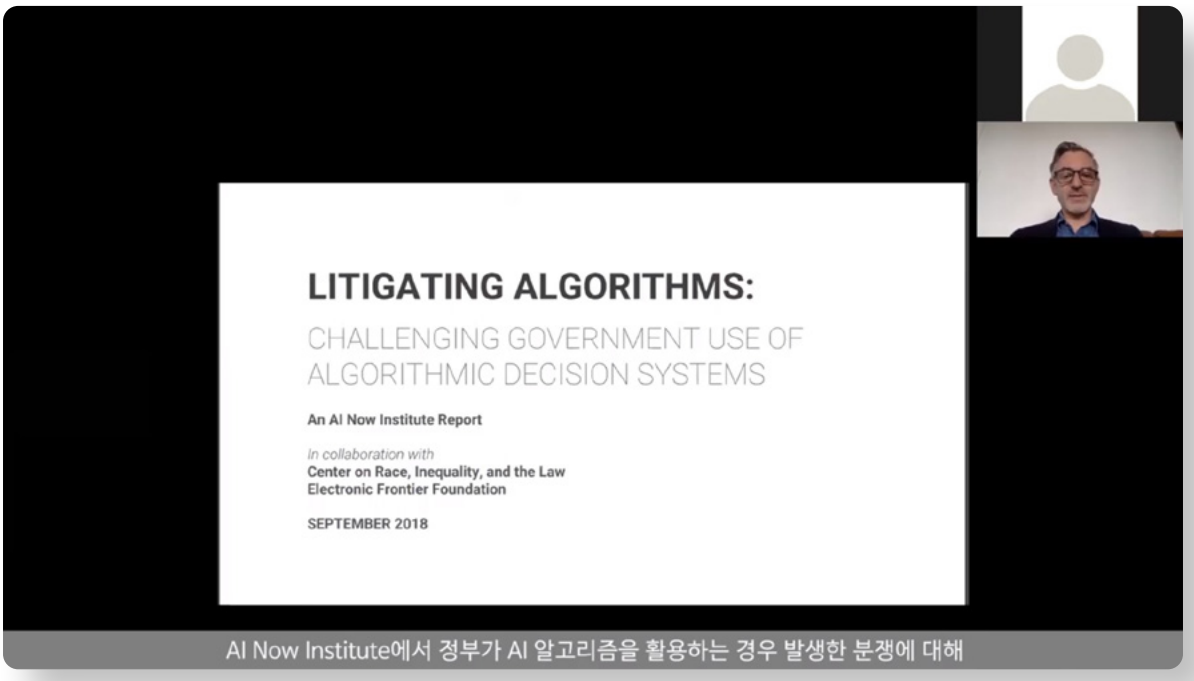
Jason Schultz, New York University


NYU | LAW
Challenges and Opportunities fo
Artificial Intelligence Policymakir
Professor Jason M. Schultz
September 9, 2020



최근 인공지능 기술을 둘러싼 한 가지 중요한 화두는 기술의 발전에 의해 대두되고 있는 안전과 책임 문제를 해소하기 위한 법정책적 접근에 대한 것이다. Jason Schultz 교수는 이러한 문제의식을 제시하고, 인공지능이란 무엇인가에 대한 설명으로 세미나를 시작하였다. 그는 법정책적 관점에서 중요한 인공지능 기술을 크게 두 가지 유형으로 제시하였다. 하나는 아마존 물류창고에서 활용되는 자동분류 기계처럼 하드웨어와 결부되어 움직이는 인공지능 로봇이고, 다른 하나는 길 찾기나 채용여부 결정처럼 의사결정 과정에서 활용되는 질의응답 인공지능이다. 이러한 인공지능은 인간에게 커다란 편익을 줄 수 있지만, 사회적 차별을 고착화하는 등의 해악(harm)을 끼칠 수도 있고, 이에 따라 법정책적 관점에서 인공지능 기술은 새로운 도전이 될 수도, 기회가 될 수도 있다.

Jason Schultz 교수는 인공지능 기술 중 머신러닝(machine learning)과 딥러닝(deep learning)의 개념을 설명하면서, 그 활용 예시로 이메일 스팸필터(spam filter), 개와 고양이를 구분하는 이미지 분류(image classification) 사례를 제시하였다. 머신러닝과 딥러닝 기술은 꾸준히 학습을 하더라도 분류 결과에 오류가 발생할 수 있다. 특히 합성곱 신경망(Convolutional Neural Network, CNN)을 활용한 이미지 분류기술은 특정 이미지를 특정 클래스(class)로 분류하는 과정에서 훈련 데이터의 편향성과 같은 문제로 치명적인 오류를 범하기도 하며, 인간이 그 원인을 알 수 없어 과연 중요한 문제에 이런 기술을 활용할 수 있는지 의문이 제기되고는 한다.



세션1 캡쳐(자막)

Jason Schultz 교수는 인공지능이 법정책적으로 유의미한 문제를 발생시키는 대표적인 경우로 편향성(bias)과 공정성(fairness)의 문제를 지적하면서 중국에서 진행된 연구결과를 소개하였다. 여기서는 범죄자의 얼굴사진으로 학습한 모델을 기반으로 어떤 사람이 범죄자가 될 확률이 높은지를 사진만을 가지고 판별하였다. Jason Schultz 교수는 이러한 시도가 설령 효과적인 결론을 도출할 수 있다 할지라도 적법절차나 무죄추정원칙 등 다양한 법률적 문제를 야기할 수 있음을 지적하였다. 이는 범죄자와 사진 속 몇 가지 특징이 가지는 상관관계에 불과하기 때문이다. 유사한 방법론을 채택한 아마존의 안면인식 모델은 미국 국회의원 28명의 범죄 위험이 높다고 판단한 일이 알려지면서 결국 사용이 중단되기도 하였다.

보다 현실적인 예시는 아마존이 회사에서 성공적인 커리어를 쌓은 직원들의 이력서를 기초로 개발한 인공지능 채용 시스템이다.[1] 본래 인공지능 채용 시스템은 채용 과정에 드는 비용은 물론이고 인간의 자의적 판단도 크게 절감해줄 것으로 기대되었다. 하지만 이 채용 시스템은 여성 지원자에 대해 일괄적으로 낮은 점수를 주는 것으로 판명되어 아마존 스스로 철회하게 되었다. 이러한 결과는 여성에 대한 차별이 만연한 과거의 사회적 분위기가 반영된 데이터를 모델이 학습한 영향 때문이라고 할 수 있다.

또 다른 예시는 COMPAS라는 알고리즘의 문제점을 지적한 ProPublica의 2016년 보도이다.[2] COMPAS는 피고인의 범죄 참여, 생활 방식, 성격과 태도, 가족과 사회적 배제 등의 변수를 점수로 환산해 재범 가능성을 계산해 판사에게 석방 여부 등을 추천하는 알고리즘이다. COMPAS는 인종을 변수로 포함하고 있지 않음에도 흑인에게는 실제와 달리 높은 재범 가능성을 추정하고, 백인에게는 실제와 달리 낮은 재범 가능성을 추정하는 경향성을 보여 커다란 사회적 문제가 되었다.

서류상의 정보를 넘어 인간의 표정을 입력 값으로 하는 분류 작업에도 인공지능 기술이 동원되고 있다. Jason Schultz 교수는 채용 과정에서 인터뷰 영상을 분석하여 지원자의 적합성을 판단하는 서비스를 제공하는 기업을 소개하였다.[3] 나아가 사진이나 동영상을 통해 인간의 감정을 인식할 수 있다고도 한다. 또한 중국에서는 사회적 신용지수를 만들어 각 시민이

1 Dastin, J. (2018) *"Amazon scraps secret AI recruiting tool that showed bias against women"* Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

2 Angwin, J., Larson, J., Surya, M., and Kirchner, L., (2016) *"Machine Bias",* ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

3 HireVue, https://www.hirevue.com/

"선량한" 시민인지를 국가에서 판단할 수 있는 시스템을 도입하고 있다.[4] 사회적 신용지수는 채용, 대출심사, 여행허가 등에 활용되며 각 시민의 구매내역, 온라인 SNS 포스팅 내역, 친구목록 등을 분석하고, 불법주차를 하거나 정부를 비판할 경우 사회적 신용지수가 하락하는 식으로 작동한다. 문제는 이러한 다양한 의사결정 과정에서 인공지능이 오류를 저지를 수 있고, 인간이 이러한 작동과정을 온전히 이해할 수 없다는 점이다.

실제로 인공지능 기술의 활용에 대한 법정책적 대응이 이루어지고 있기도 하다. 예를 들어, 일리노이 주는 인공지능 기술을 활용한 면접평가를 하는 경우 면접자는 피면접자에게 인공지능을 활용한 평가가 진행되고 있다는 사실을 고지할 의무와 인공지능이 어떠한 방법으로 피면접자를 평가하는지를 설명할 의무를 부여하였다.[5] 다만 설령 이러한 설명의무가 부여된다고 하여도 인공지능의 설명가능성에 대한 다양한 한계로 인하여 기업들이 이를 준수할 수 있을지는 불명확하다고 Jason Schultz 교수는 지적하였다.

Jason Schultz 교수는 인공지능 알고리즘에 관한 분쟁해결 과정에서 영업비밀(trade secret)이나 설명의 문제가 대두된 두 가지 사례를 소개하였다.[6] 첫 번째 사례는, 아칸소 주에 거주하는 뇌성마비환자 Tammy Dobbs에 대한 것이다. Tammy Dobbs는 알고리즘에 의해 산정된 간병시간이 너무 적다는 이유로 주정부를 상대로 설명을 요구했으나, 알고리즘을 사용한 주정부는 명확한 답변을 제시하지 못했다. 재판에서 아칸소

4 Marr, B., (2019) *"Chinese Social Credit Score: Utopian Big Data Bliss or Black Mirror on Steroids?",* Forbes. https://www.forbes.com/sites/bernardmarr/2019/01/21/chinese-social-credit-score-utopian-big-data-bliss-or-black-mirror-on-steroids/#5dfbf0e748b8

5 820 ILCS 42/ Artificial Intelligence Video Interview Act

6 Richardson, R., Schultz, J. & Southerland, V. (2019) *"Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems",* AI Now Institute, Available at https://ainowinstitute.org/litigatingalgorithms-2019-us.html

주정부는 알고리즘은 영업비밀이므로 공개할 수 없다고 주장하였으나, 법원은 알고리즘 공개를 명령하였다. Jason Schultz 교수는 그럼에도 불구하고 인간에 비해 컴퓨터 코드는 간병시간 삭감 이유에 대한 적절한 설명을 제시하기 어렵다는 점을 지적하였다.

두 번째 사례는 알고리즘에 따른 직원평가 결과로 해고당한 공립학교 교사를 대신해 노동조합이 교육청을 상대로 제기한 소송이다. 피고측은 역시 알고리즘은 영업비밀에 해당하기 때문에 공개할 수 없다고 주장하였으나, 법원은 중요한 인사결정에 당사자에게 공개할 수 없는 영업비밀인 알고리즘이 사용되었다면 이는 최소한의 적법절차를 지키지 않은 것이고, 그렇다면 적절한 해결방안은 영업비밀은 보호하되 해당 해고결정을 취소하는 것이라고 판단하였다.[7] Jason Schultz 교수는 이처럼 공정성, 투명성에 대한 문제가 해소되지 않는 한 중요한 의사결정을 알고리즘에 의존하는 것은 지양해야 한다고 지적하였다.

7 "When a public agency adopts a policy of making high stakes employment decisions based on secret algorithms incompatible with minimum due process, the proper remedy is to overturn the policy, while leaving the trade secrets intact." (HFT v. HISD, 251 F.Supp.3d 1168 (S.D. Tex. 2017)

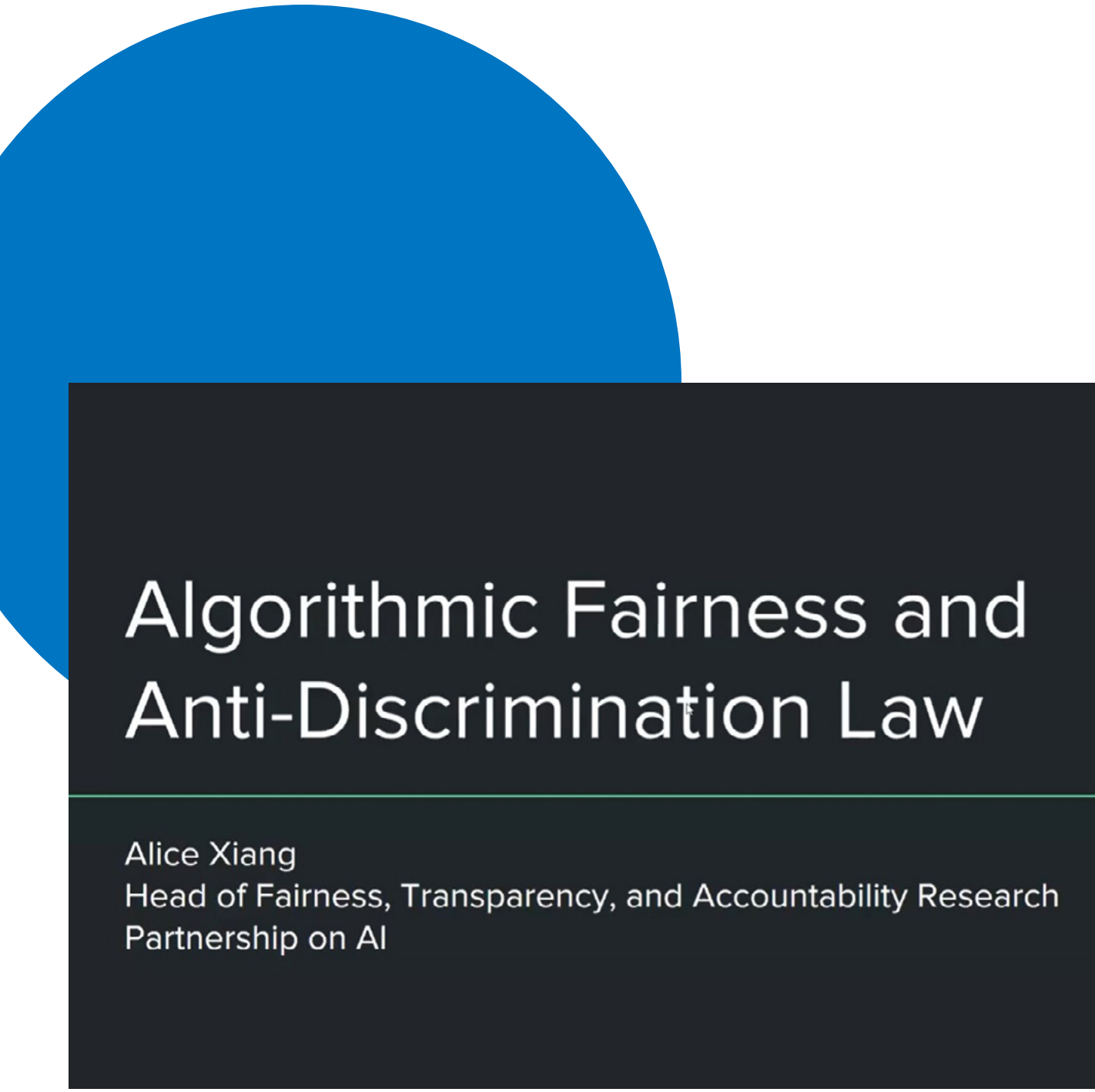
제2세션

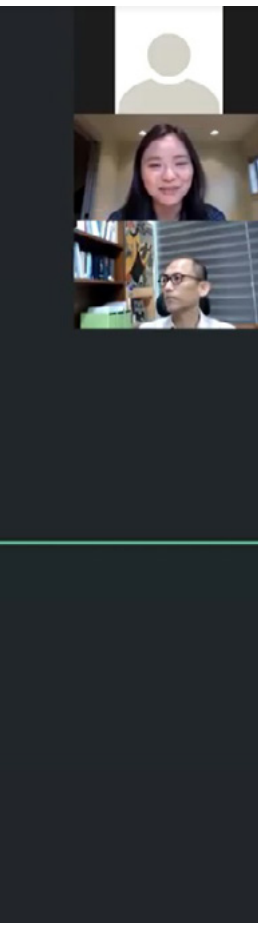
알고리즘과 공정성: 차별금지법이 필요한가?

Algorithmic Fairness and Anti-Discrimination Law

발표자

Alice Xiang, Partnership on AI


Algorithmic Fairness and Anti-Discrimination Law
Alice Xiang
Head of Fairness, Transparency, and Accountability Research
Partnership on AI



제2세션은 계속해서 법정책적 논의를 다루면서, 주로 알고리즘의 공정성과 차별금지법이라는 주제에 주목하였다. Alice Xiang 변호사는 제1세션에도 언급된 COMPAS 재범예측 알고리즘이 흑인을 차별한 사례와 아마존의 채용 알고리즘이 여성을 차별한 사례를 제시하면서 세미나를 시작하였다. 이어서 집단의 대표성이 왜곡되어 해악(representational harm)을 유발한 사례로, 구글에 "CEO"라는 검색어를 입력하면 대부분 남성 CEO의 사진이 나오며 최상위에 노출되는 여성 이미지는 CEO로 표현된 바비인형이라는 점을 제시하였다. 이들 사례의 공통점은 인공지능 알고리즘의 학습 데이터셋이 과거의 편향을 반영한 것이라는 점이다.

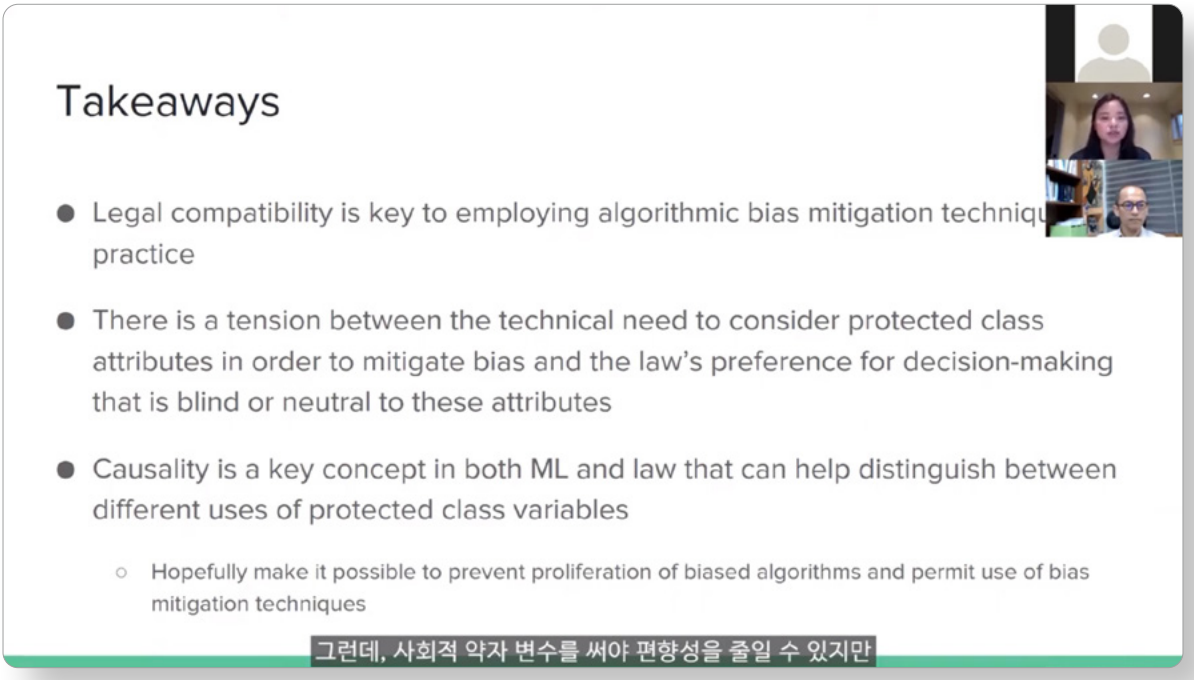
다음으로 Alice Xiang 변호사는 알고리즘 편향(algorithmic bias)이란 무엇인가에 대한 논의로 나아갔다. Alice Xiang 변호사는 (1)알고리즘 의사결정 과정이 어떻게 특정 소집단에게 체계적으로 나쁜 결과를 유발하는지를 중심으로 한 접근법과, (2)사회적으로 문제시되는 인구통계학적 혹은 여타 특성으로 인해 발현된 격차(disparity)를 중심으로 한 접근법을 제시하였다. 이러한 알고리즘 편향에 대한 기술적 정의는 기존의 차별금지법과 조화를 이루어야 한다. 그래야만 알고리즘 편향을 판별하는 지표가 법적 의미의 차별의 근거로 활용될 수 있을 것이며 편향을 감소시키는 방법을 고민할 수 있기 때문이다.

이러한 관점에서 Alice Xiang 변호사는 미국 차별금지법제상 보호받는 변수(protected class variable)의 개념을 설명하였다. 미국에서 인종, 성별, 나이, 장애여부, 출신국가, 종교 등에 따른 차별은 법적으로 금지된다.[1] 이러한 법적인 제약을 준수하는 가장 직관적인 방법은 모델 학습 시에 이들 변수를 제거하는 것이다. Alice Xiang 변호사는 이를 순진한(naïve) 접근법이라고 지적하였다. 이들 변수들을 대체하는 다른 변수(proxy)들을 조합하여도 보호받는 변수를 포함시킨 경우와 동일한 알고리즘 편향이 발생할 수 있기 때문이다. 가령 채용절차에서 지원자의 거주지와 소득수준을 기반으로 해당 지원자의 인종을 예측할 수 있는 식이다. 실제 COMPAS 알고리즘에서 인종은 변수로 입력되지 않았지만 흑인과 백인 간 재범률 평가에서 통계학적으로 유의미한 차이를 보인 바 있다. Alice Xiang 변호사는 그럼에도 불구하고 "무지함을 통한 공정성(fairness through unawareness)"은 여전히 가장 널리 통용되는 접근법임을 지적하였다. 가령 미국 주택도시개발부(U.S. Department of Housing and Urban Development)는 보호받는 변수를 사용하지 않는 알고리즘에 대해서는 차별로 인한 책임을 면제하는 규정을 제정하기도 하였다.[2]

그렇다면 알고리즘 편향을 감소시키는 방법은 무엇인가? Alice Xiang 변호사는 효과적으로 알고리즘 편향을 감소시키기 위해서는 오히려 보호받는 변수를 맥락에 따라 활용해야 하며, 이를 통해 알고리즘의 공정성과 정확성을 동시에 높일 수 있다고 주장하였다. 이러한 접근법은 기존 대학 입학사정 등에서 활용되었던 적극적 우대조치(Affirmative Action)와 그 기본논리를 공유한다. 적극적 우대조치란 역사적으로 차별을 경험한 집단

1 가령 Civil Rights Act of 1964의 Title VII는 고용 맥락에서 특정 인구통계학적 변수에 따른 차별을 금지한다. 42 U.S.C § 2000e-2(a).

2 Fair Housing Act, 미국 주택도시개발부(HUD)의 Disparate Impact Standard, 84 Fed. Reg. 42854 (2019. 8. 19.)


Takeaways
Legal compatibility is key to employing algorithmic bias mitigation techniqu
practice
There is a tension between the technical need to consider protected class attributes in order to mitigate bias and the law's preference for decision-making that is blind or neutral to these attributes
Causality is a key concept in both ML and law that can help distinguish between different uses of protected class variables
Hopefully make it possible to prevent proliferation of biased algorithms and permit use of bias mitigation techniques
그런데, 사회적 약자 변수를 써야 편향성을 줄일 수 있지만

세션2 캡쳐(자막)

을 우대하는 정책을 총체적으로 이르는 말로, 주로 소수인종과 여성을 배려하기 위한 정책으로 활용되었다.

그런데 적극적 우대조치에 대한 Bakke 미국 연방대법원 판결은 인종은 대학입시에서 고려될 수 있는 요소이나 특정 인종의 비율을 정해놓은 할당제는 위헌임을 선언함과 동시에, 인종에 대한 고려는 역사적 불평등의 치유가 아니라 입학생의 다양성이라는 교육적 목적을 위한 것이어야 한다고 판시하였다.[3] Alice Xiang 변호사는 보호받는 변수를 활용한 알고리즘 편향 감소방법은 비율을 정하는 할당제와 유사하고 역사적 불평등을 치유하는데 초점을 맞춘 것이라고 할 수 있어서 Bakke 판결과는 일부 상충되는 지점이 있다고 지적하였다. 2003년 Grutter v. Bollinger과 Gratz v. Bollinger 후속판결에서 미국 연방대법원은 개별 지원자를 평가함에 있어 인종을 고려하는 것은 허용되나 소수인종이라는 사실만으로 가산점을 부여하는 것은 허용되지 않는다고 판시하여 유사한 관점을 이어갔다.[4]

이러한 미국 연방대법원의 적극적 우대조치에 대한 판결례는 알고리즘 편향을 해소하는 방법을 고안함에 있어 제약으로 작용한다. 특정 비율을 보장하는 할당제나 가산점 부여방식이 허용되지 않는다면 소수인종이나 여성에 대한 알고리즘 편향을 어떻게 보정할 수 있을지가 문제되기 때문이다. Alice Xiang 변호사는 차별금지법의 차등적 대우(disparate treatment) 법리에서 중요한 요건으로 작용하는 “인과성(causality)”을 활

3 Regents of the University of California v. Bakke, 438 U.S. 265 (1978)

4 같은 날 내려진 판결로, Grutter v. Bollinger, 539 U.S. 306 (2003) 사건과 Gratz v. Bollinger, 539 U.S. 244 (2003) 사건이다. Grutter 사건에서는 미시건대학교 로스쿨의 입학사정이 문제되었고 Gratz 사건에서는 미시건대학교 학부 입학사정이 문제되었다. 미시건대학교 학부 입학시스템은 지원자를 150점 만점으로 평가하였는데 소수인종은 일괄적으로 20점의 가산점을 부여받았다.

용할 수 있다고 주장하였다.[5] 즉, 알고리즘이 인종이나 성별과 같은 보호받는 변수를 활용하는 행위 자체를 무조건 금지하는 대신, 보호받는 변수와 편향적 결과 사이의 인과성 여부를 구체적으로 평가하자는 것이다. 이에 따르면 동일하게 보호받는 변수를 알고리즘에 활용하더라도 편향적인 결정으로의 인과성을 증가시키는 방향으로 보호받는 변수가 활용되는지, 또는 인과성을 감소시키는 방향으로 활용되는지에 따라 법적 평가를 달리할 수 있게 된다.

Alice Xiang 변호사는 보호받는 변수를 원천적으로 고려하지 못하도록 하는 차별금지법 법리를 알고리즘 편향에 그대로 적용할 경우 오히려 불합리한 차별을 고착시키는 결과를 초래할 수 있음을 지적하는 동시에 인과성에 기초한 평가를 하나의 해법으로 제안하면서 세미나를 마무리하였다.[6]

5 Texas Dept. of Housing and Community Affairs v. Inclusive Communities Project, Inc., 576 U.S. 519 (2015)

6 본 발표에 대한 보다 자세한 내용은 연사의 논문인 Alice Xiang. (2021). *"Reconciling Legal and Technical Approaches to Algorithmic Bias"*, Tennessee Law Review, Vol. 88, No.3 (forthcoming)을 참조.

제3세션

알고리즘과 경쟁: 진화하는 법과 정책

Competition in the Era of Algorithms: Evolving Law & Policy

발표자

Niamh Dunne,
The London School of Economics and Political Science

Competition in the Era of Algorith
Evolving Law & Policy

Niamh Dunne
N.M.Dunne@lse.ac.uk

Seoul National University, 16 September 2020

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

제3세션은 인공지능이 시장경쟁에 미치는 영향력에 대한 법정책적 논의를 주제로 다루었다. 발표자인 Niamh Dunne 교수는 인공지능 알고리즘의 비즈니스적 사용이 시장에서 어떠한 문제를 일으키고 특히 경쟁법의 관점에서 어떠한 이슈로 연결되는지, 그리고 경쟁법이 이러한 문제에 대한 해결책을 제시하기 위하여 어떻게 진화할 필요가 있는지를 화두로 던지면서 논의를 시작하였다. Niamh Dunne 교수는 이번 발표는 유럽연합의 경쟁법을 주된 대상으로 삼았지만, 여타 관할권에서의 논의에도 많은 시사점을 제공할 수 있다는 점을 강조하였다.

유럽연합 경쟁법 당국은 시장에서 알고리즘의 사용이 늘어나는 상황과 관련하여 크게 4가지 핵심 시나리오를 그리고 이에 대응해오고 있다고 한다. 4가지 시나리오는 (1)“로봇 카르텔”이라는 디스토피아적 시나리오, (2) 알고리즘이 시장의 기정사실로 존재하는 시나리오, (3)알고리즘이 시장경쟁적 해악을 가중시키는 시나리오, (4)알고리즘 층위의 경쟁이 그 자체로 장점에 기초한 경쟁(competition on the merits)으로 자리 잡게 되는 시나리오를 가리킨다.

첫 번째 시나리오는 다수의 시장 참여자가 알고리즘을 적용하여 결국에는 사람의 개입 없이 즉흥적으로 로봇이 담합을 할 수 있다는 내용이다. 이에 대한 다양한 연구가 있지만, 대표적으로 “Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy” 책에서 제시한 로봇 카르텔과 경쟁법의 한계에 대하여 찬반양론이 널리 대립하고 있다고 한다. 다만, 로봇 카르텔 시나리오가 현실에서 실제로 문제가 된 사례는 없기에 Niamh Dunne 교수는 인간이 인공지능을 담합의 수단으로

악용한 두 가지 사례를 언급하였다. 하나는 가격설정 방식을 포함한 마케팅용 소프트웨어가 여러 사업자 사이의 명시적 합의를 유도한 *Trod/GB* 사례[1]이고, 다른 하나는 아마존에서 포스터 사업자간 담합의 수단으로 알고리즘을 사용한 *Topkins* 사례[2]이다. 최근에는 실험실 내로 한정된 연구이지만, 알고리즘들이 가격을 높게 책정하도록 하는 합의를 학습한 결과[3]가 발표되기도 하였다고 한다.

두 번째 시나리오는 알고리즘이 시장에서 기정사실로 자리 잡는다는 내용이다. 여기서 알고리즘은 남용의 도구가 아닌 시장의 일부로 기능한다. 이와 관련하여 Niamh Dunne 교수는 두 가지 사례를 소개하였다. 먼저 *Guess* 사건[4]은 의류생산업자인 Guess가 공식 유통사업자와 계약을 체결하면서 Google Ad의 키워드(keywords) 입찰 참여를 금지하는 수직적 합의를 하여 유럽집행위원회(European Commission, 이하 "EC")에서 유럽연합기능조약(Treaty on the Functioning of the European Union, 이하 "TFEU") 제101조[5]에 위배된다고 결정한 사건이다. 다음으로 *Google Search(AdSense)* 사건[6]은 Google이 온라인 언론사 등에 대해 자사의 광고가 가장 수익성이 있는 자리에 위치해야 한다는 배타적인

1 UK competition authority decision in Trod Ltd: posters and frames, 21 July 2016, Case 50223

2 *United States v. Topkins,* No. 15-201 (N.D. Cal. Apr. 30, 2015)

3 Calvano, E., Calzolari, G., Denicolo, V., and Pastorello, S. (2019) *"Artificial Intelligence Algorithmic Pricing and Collusion"*. Available at http://dx.doi.org/10.2139/ssrn.3304991

4 Guess (Case AT.40428 — Guess) European Commission Decision C/2018/8455 [2018] OJ C 47, 6.2.2019, p. 5, Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.C_.2019.047.01.0005.01.ENG&toc=OJ:C:2019:047:TOC

5 TFEU 제101조는 국내 "독점규제 및 공정거래에 관한 법률" 제19조 '부당한 공동행위'에 대응하는 조항으로, 사업주체의 담합(카르텔) 협의를 금지하는 내용을 포함하고 있다.

6 Google Search(AdSense) (Case 40411) European Commission Decision of 20 March 2019 (not yet published)

거래 정책을 펼쳐 EC에서 TFEU 제102조[7]에 위배된다고 본 사건이다.

두 사건 모두 시장에서 어떻게 반경쟁적 행위가 이루어졌는지 이해하려면 알고리즘의 작동원리를 알아야 한다. 예를 들어, *Google Search(AdSense)* 사건에서 Google이 가장 노출이 잘 되는 위치의 광고를 자신에게만 할당하도록 한 행위가 반경쟁적인지 이해하기 위해서는, 온라인 광고의 원리에 대하여 이해할 수 있어야 판단할 수 있다. 위 사건들에서의 행위는 온라인 상에서 알고리즘을 사용해서만 발생하는 것이 아니라 다른 시장에서 알고리즘 없이도 일어날 수 있기에 알고리즘을 사용한 것 자체나 디지털 시장인지 여부는 문제가 되지 않는다.

세 번째 시나리오는 알고리즘의 활용으로 인하여 경쟁에의 제한이 가중되는 형태이다. 대표적 사례가 생산사업자가 유통사업자에게 계약을 체결하면서 일정한 금액 이하로 판매할 수 없다는 제한을 두는 재판매가격유지행위(Resale Price Maintenance, RPM)이다. 유럽연합은 이를 1980년대 *Binon* 사건[8]부터 "목적상(by object) 경쟁제한행위"로 보고 TFEU 제101조에서 금지해왔다. 그러나 이러한 행위의 효과가 불분명하다는 문제의식으로 2004년 유럽연합 경쟁법 집행의 분권화(decentralisation)[9]가 이루어진 이후에 EC에서는 재판매가격유지행위는 다루어지지 않았다. 그러나 이러한 재판매가격유지행위가 최근 온라인 시장에서 문제가 되면서 2018년에 EC가 Asus, Denon & Marantz,

7 TFEU 제102조는 국내 "독점규제 및 공정거래에 관한 법률" 제3조의2 '시장지배적지위의 남용금지'에 대응하는 조항이다.

8 Case 243/83 *SA Binon & Cie v SA Agence et messageries de la pressem* [1985] ECR 2015

9 TFEU 제101조(구 제81조), 제102조(구 제82조)의 시행규칙이 2004. 5. 1. 을 기점으로 R 17/62에서 R 1/2003으로 대체되면서, EC에 집중되어 있던 경쟁법의 집행에 대한 권한을 각 회원국의 경쟁당국 및 법원으로 분산시켜 유럽연합 공동체 경쟁법규의 자율적 시행을 규정하였다.

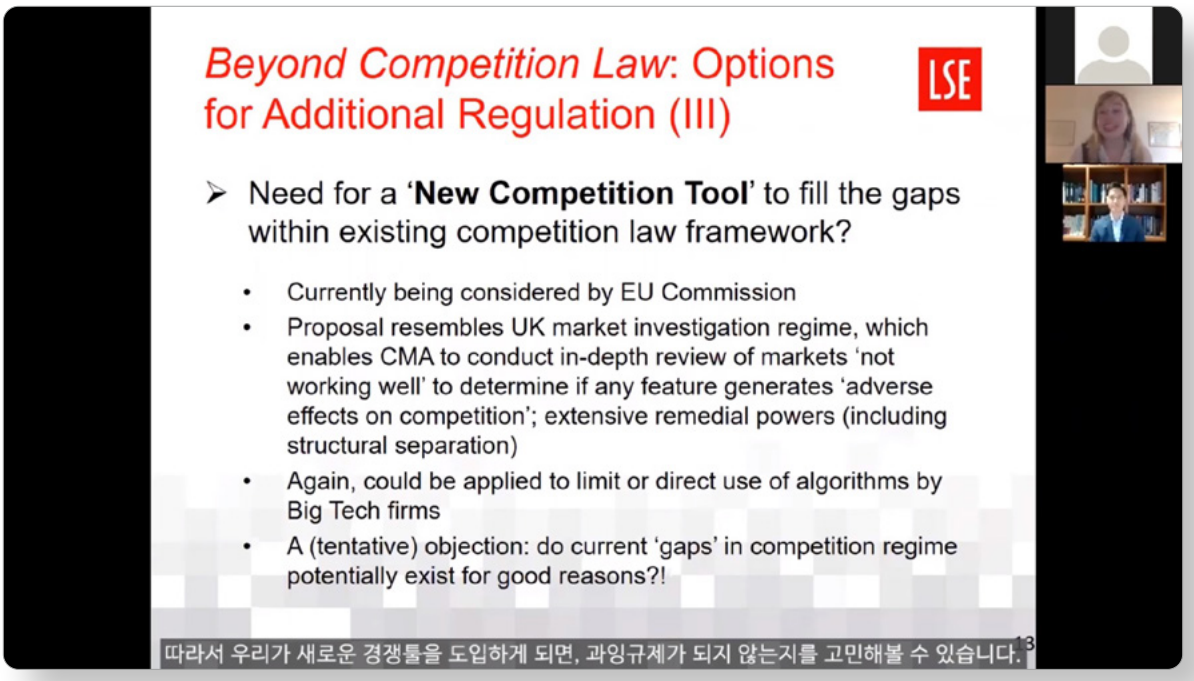
Philips, Pioneer, Guess의 재판매가격유지행위에 "목적상 경쟁제한행위" 법리를 적용하여 경쟁법 위반 결정을 내렸다.[10]

EC에서 10년 넘게 적용하지 않았던 법리가 다시 소환된 이유로, Niamh Dunne 교수는 가격을 추적(price-tracking)하고 결정(price-setting)하는 알고리즘의 사용이 증가되었다는 점을 지적하였다. 상품 생산업자 입장에서는 이러한 알고리즘을 통해 재판매가격을 유통사업자가 준수하는지에 대한 감시가 용이하기 때문에 재판매가격 정책을 유지하도록 하는 압력이 강화되었다는 것이다. 게다가 유통사업자는 알고리즘을 통하여 유통시장에서 경쟁관계에 있는 사업자들의 가격을 쉽게 볼 수 있기 때문에, 재판매가격유지의 집행 역시 용이해질 수 있게 되었다. 기존에 존재하던 재판매가격유지행위에 알고리즘이라는 신기술이 적용되면서 그 피해가 가중되었고, 이에 따라 과거의 법리가 부활하게 되었다는 말이다.

네 번째 시나리오는 알고리즘이 자체로 장점에 기초한 경쟁(competition on the merits)의 일부가 된다는 내용이다. 이와 관련하여, Niamh Dunne 교수는 2017년에 TFEU 제102조에 의해 EC 역사상 가장 큰 금액의 과징금이 부과된 *Google Search(Shopping)*사례[11]를 소개하였다. Google은 "일반검색시장(general searching market)"에서 당시 유럽시장의 85% 점유율을 가지고 있는 독점적 지위의 사업자였고, 뛰어난 검색 알고리즘을 통하여 소비자에 대한 서비스를 향상시킨 부분 또한 존재하였다. 그러나 별개의 쇼핑검색엔진을 활용한 "쇼핑 비교 시장(shopping comparison market)"에서는 Google이 고전을 면치 못하고 있었는데, 여기서 Google은 더 나은 알고리즘을 개발하기 보다는 쇼핑 검색 결과 자사

10 Case AT. 40465 – Asus, Case AT. 40469 – Denon & Marantz, Case AT. 40181 – Philips, Case AT. 40182 – Pioneer

11 Google Search(Shopping) (Case AT. 39740 – Google Search (Shopping)) European Commission Decision of 27.6.2017.



세션3 캡쳐(자막)

의 제품이 상위에 노출되고 경쟁사 제품은 낮은 순위로 조정하여 자사 제품을 우대(self-preferencing)하도록 검색 알고리즘을 수정 및 변형하였다. Niamh Dunne 교수는 디스토피아 시나리오의 전망과 달리 역설적으로 알고리즘이 시장경쟁을 촉진한 결과로 이러한 문제가 대두되었다고 지적하면서, 만일 이러한 왜곡을 방지하고 소비자의 진정한 욕구를 반영할 수 있다면 도리어 알고리즘이 경쟁을 강화할 수 있다고 보았다.

지금까지의 논의를 통해 알 수 있듯, 오늘날 알고리즘은 시장을 이루는 일부분이 되었고, 따라서 알고리즘이 경쟁법에 미치는 영향을 알기 위해서는 알고리즘에 대한 이해가 요구된다. 특히 개발자가 어떻게 알고리즘을 코딩하였는지에 대한 모든 기술적인 부분을 파악하려 하기 보다는,

알고리즘이 어떻게 경쟁에 영향을 미치는지를 파악하는 것이 중요하다. Niamh Dunne 교수는 그 출발점으로 사업자가 자유롭게 작용하는 알고리즘에 반경쟁적인 영향을 야기하기 위하여 고의적으로 개입하지 못하도록 해야 한다고 언급하였다. 그리고 알고리즘도 사업자의 행위의 일부로 보고 알고리즘을 활용한 그 행위가 반경쟁적인지에 대한 판단에 초점을 둔다면, 고도의 기술적인 분야인 알고리즘에는 경쟁법을 적용하기 어렵다고 했던 과거의 의견은 더 이상 받아들이기 어렵다고 보았다.

한편 Niamh Dunne 교수는 오늘날 경쟁법을 디지털 시대에 부합하도록 변화를 주어야 할 필요성에 대해서도 인정하면서 몇 가지 선택지를 제시하였다. 첫 번째 사례는 유럽연합이 이미 수용하고 있는 "온라인 중개 서비스의 사업자 이용자를 위한 공정성 및 투명성 강화에 관한 규칙 2019/1150(Regulation 2019/1150 on Promoting fairness and transparency for business users of online intermediation services)[12]"이다. 본 규칙은 온라인 중개 서비스를 제공하는 사업자들에게 특히 공정한 계약 조건을 제공할 의무를 부과하고, 알고리즘을 활용한 순위 기준 결정 기술에 대한 투명성의 향상에 대해서도 규정한다. 다만, Niamh Dunne 교수는 일반 이용자를 대상으로 하지 않고 집행체계가 약하여 실효성이 취약할 수 있다고 지적하였다.

두 번째로는 사전 행동강령(Ex ante Code of Conduct)을 디지털 사업주체에게 부과하는 것으로, 호주와 영국을 비롯하여 유럽에서도 논의되는 방식이다. 특히 유럽연합 내에서는 대형 기술기업을 대상으로 하고, 디지털 서비스 법안(Digital Service Act Package)을 논의하고 있다. 여기서 특기할 만한 사항은 적용범위를 TFEU 제102조에 규정된 "시장력

12 Avaliable at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019R1150

(market power)"이 아닌 "경제력(economic power)"을 기준으로 산정한다는 것이다. 다만, Niamh Dunne 교수는 예측을 할 수 없는 5년 후, 10년 후의 디지털 시장의 모습에 영향을 미칠 현재의 규범들을 과연 규제당국이 적절하게 규정할 수 있을지 고민해야 한다고 덧붙였다.

세 번째로는 기존의 경쟁법 체계보다는 새로운 디지털 시대에 부합하는 "새로운 경쟁법규(New Competition Tool)"를 마련하는 방안이다. TFEU 제101조, 제102조 이외에 새로운 제도를 도입해야 한다는 의견으로, 현재 영국은 알고리즘이 제대로 작동하는지를 조사하는 시장 조사 제도(UK market investigation regime)과 유사한 제도를 도입하려는 방안을 검토 중이라고 한다.

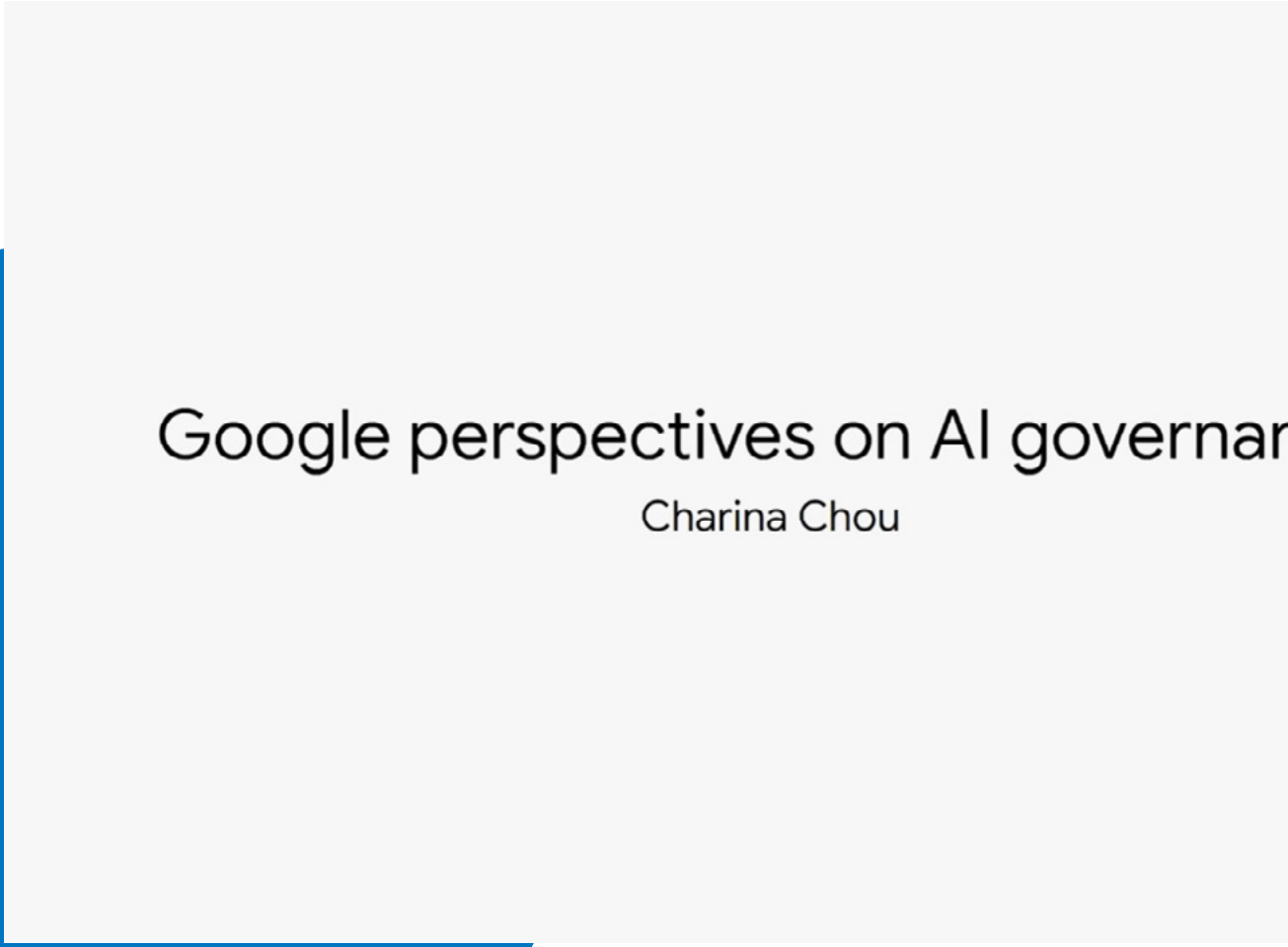
제4세션

인공지능 거버넌스에 대한 구글의 시각

Google Perspectives on AI Governance

발표자

Charina Chou, Google


Google perspectives on AI governar
Charina Chou

제4세션은 개별기업 입장에서의 거버넌스 논의로 초점을 옮겨, 구체적으로 구글의 시각을 주제로 다루었다. Charina Chou 박사에 따르면 구글의 목표는 전 세계의 정보를 체계화하여 모두가 편리하게 이용할 수 있도록 하는 것이다.[1] 인공지능은 이러한 구글의 목표를 달성하기 위해 매우 중요한 역할을 하고 있다. 가령 구글의 대표적인 상품인 검색엔진을 보면, 인터넷에 존재하는 정보가 텍스트에서 이미지, 영상으로 발전하면서 20년 전 구글이 검색 서비스를 처음 선보였을 때의 알고리즘 만으로는 더 이상 효과적인 검색이 불가능하게 되었고, 머신러닝 기술을 접목한 형태로 발전하였다. 인공지능 기술이 제공하는 구글의 기술적 발전은 검색에만 국한되지 않고 가령 암을 조기에 식별하는 의료 이미지 프로세싱(medical image processing)에도 활용되고 있다.

하지만 인공지능 기술의 활용은 새로운 문제와 위험성을 야기하고 있다. Charina Chou 박사는 구글이 2년 전 영국 의료기관의 요청으로 무성증(Aphonia)[2] 환자를 위해 개발을 검토한 독순술(Lip reading)[3] 알고리즘을 예시로 들었다. 무성증 환자의 의사소통을 도울 수 있다는 효용에도 불구하고 독순술 알고리즘이 개발되고 널리 배포될 경우 프라이버시 침해를 비롯한 악용의 소지가 많기 때문이다. 구글은 이처럼 인공지능 기술에 관한 윤리적 의사결정이 필요할 것으로 판단하여 2018년 구글 인공지능 원

1 구글의 Official Mission Statement: “To organize the world’s information and make it universally accessible and useful.”

2 후두의 질환 또는 후두근을 지배하는 조직의 장애로 일어나는 음성의 상실증이다.

3 독순술은 입술, 얼굴, 혀의 움직임을 보고 말을 알아내는 기술이다.

칙(Google AI Principles)을 작성 및 배포하였다.[4] 구글은 독순술 알고리즘의 개발여부와 같은 윤리적 판단이 필요한 지점에서 이러한 원칙에 입각해 판단하고 있다.

Charina Chou 박사는 인공지능 원칙의 집행에 있어 기술적인 조치가 중요한 역할을 수행한다고 강조하였다. 구글은 마치 식품의 영양성분표와 같이 인공지능 모델을 학습하기 위해 사용하는 데이터셋에 관해서도 주요 정보를 요약해서 제공하고 있다. 가령 구글이 기계학습용으로 공개한 Open Images Extended – Crowdsourced 데이터셋의 경우 데이터 활용 목적, 데이터 출처 등이 상세하게 공개되어 있으므로 연구자 및 개발자가 해당 데이터셋의 특성을 알고 활용할 수 있게 되는 효과가 있다. 나아가 구글은 보다 대표성 높은 데이터셋을 구축하기 위해 크라우드소싱(crowdsourcing) 방식을 활용하고 있다. 이는 기존 데이터셋이 대부분 서양의 문화와 관습에 기초하여 구축되었기 때문에 다른 문화권의 데이터가 적절히 대표되고 있지 않다는 지적에서 비롯된 것이다. Charina Chou 박사는 결혼식이나 신부 등의 이미지 검색을 실행하면 하얀 드레스 사진이 검색결과로 제시되는데 이는 서양의 관습에 따른 것이고 다른 나라(가령 인도)의 결혼식 풍경과는 다를 것이라는 점을 지적하였다. 이러한 기술적 노력은 모델 학습 이전 단계에서 데이터셋의 완결성과 다양성을 높이려는 시도이다.

4 주요 내용은 다음과 같다: AI applications must (1) be socially beneficial, (2) avoid creating or reinforcing unfair bias, (3) be built and tested for safety, (4) be accountable to people, (5) incorporate privacy design principles, (6) uphold high standards of scientific excellence, (7) be made available for uses that accord with these principles. AI applications will not pursue (1) technologies that cause overall harm, (2) technologies whose principal purpose is to cause injury to people, (3) surveillance violating international norms, (4) technologies whose purpose violate international law and human rights. https://ai.google/principles/

Charina Chou 박사는 모델 학습 이후 단계에서의 기술적 조치도 중요하다는 점을 강조하면서, 인공지능 모델에 "공정성(fairness)" 제약을 가하는 방안을 소개하였다. 가령 터키어는 대명사에 성별 구분이 없는데, 구글 번역(Google Translate)을 통해 "그는 의사이다(o bir doctor)"라는 터키어를 영어로 번역하면 "he is a doctor"라는 결과값이 도출되는 현상이 발견되었다. 이에 구글은 모델에 사후적 제약을 가하여 사용자가 "he is a doctor"와 "she is a doctor" 중 원하는 결과값을 선택할 수 있도록 하였다. Charina Chou 박사는 또한 인공지능의 "해석가능성(interpretability)"에 대한 기술적 발전으로 TCAV(Testing with Concept Activation Vectors) 방법론을 소개하였다. 인공지능 모델이 입력값을 특정 클래스(class)로 분류하는데 있어 유저가 정의한 개념이 어느 정도 기여하는지를 방향 도함수(directional derivative)로 표현하는 방식이다.[5] 구글은 이러한 정보를 종합해 "AI 활용 실무(General recommended practices for AI)"에 관한 페이지를 운용하고 있기도 하다.[6] 해당 페이지는 인공지능 활용 실무에 대해 공정성(fairness), 해석가능성(interpretability), 프라이버시(privacy), 보안(security) 측면의 자료를 제공한다.

Charina Chou 박사는 인공지능 기술 활용에 있어 기술적 측면뿐 아니라 기업의 운영(operational processes) 차원에서도 중요한 지점들이 많다고 지적하였다. 이와 관련하여 Charina Chou 박사는 크게 네 가지 사항을 지적하였다. 먼저 사내문화적으로 인공지능 기술을 활용하는 과

5 Kim, B., Wattenberg, M., Gilmer, J., Cai, C., Wexler, J., Vlegas, F., & savres, R. (2018). *"Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV)",* Proceedings of the 35th International Conference on Machine Learning, in PMLR

6 https://ai.google/responsibilities/responsible-ai-practices/

정에 각 담당자(stakeholder)가 관련 지식을 숙지하고 있으면서 언제든지 자유롭게 의견을 개진할 수 있는 환경이어야 한다. 실제로 구글은 이를 실천하기 위해 다양한 교육 프로그램을 운영하고 있다고 한다. 둘째, 앞서 설명한 기술적 장치들을 단순히 개발만 할 것이 아니라 기술전문가가 아닌 사람도 쉽게 이해하고 활용할 수 있도록 하여야 한다. 셋째, 독순술 알고리즘과 같은 개별 사례들에 관한 논의내용을 사후적으로 기록, 검토 및 공유하고 이로부터 인공지능 원칙을 발전시키는 업데이트 절차가 있어야 한다. 마지막으로 각종 컨퍼런스, 보고서 등을 통해 사외 커뮤니티와 인공지능 기술의 활용에 관하여 지속적으로 소통하여야 한다.

Charina Chou 박사는 인공지능에 관한 공공정책 역시 중요한 역할을 해야 한다고 주장하였다. 먼저 인공지능 기술 활용 촉진 차원에서 공공기관이 보유하고 있는 데이터셋을 공유하여 민간의 인공지능 기술 활용을 촉진할 수 있다고 지적하였다. 가령 공공기관이 제공하는 많은 데이터는 기계가 읽고 이해하기 어려운 정제되지 않은 형태로 되어있는데, 이러한 부분을 개선하기 위한 정부의 투자가 필요하다는 것이다. 그 외에 데이터의 특성에 대한 정보를 정리하고 제공하는 방안, 인공지능 기술에 대한 교육을 강화하는 방안, 공공분야에서 실제로 인공지능을 활용하는 방안 등도 필요하다고 강조하였다.

나아가 Charina Chou 박사는 인공지능 기술 활용에 대한 논란을 줄이기 위해 인공지능 모델에 대한 “설명가능성(explainability)” 기준을 정립하는 것도 정부의 역할이라고 지적하였다. 인간과 인공지능의 협업(human-AI collaboration) 관련하여 인적 개입을 요청할 권리 등에 관한 논쟁도 공공정책의 영역이다. 결국 중요한 것은 각 분야마다 인공지능 기술의 활용으로 새롭게 대두되는 문제들이 발생할 것이고 규범체계가 이러한 새로운 문제들에 얼마나 잘 대응할 수 있느냐는 것이다. 기존의 규범체계가 상정하지 않은 문제들에 대해서는 법령 등을 신속히 개정할 필

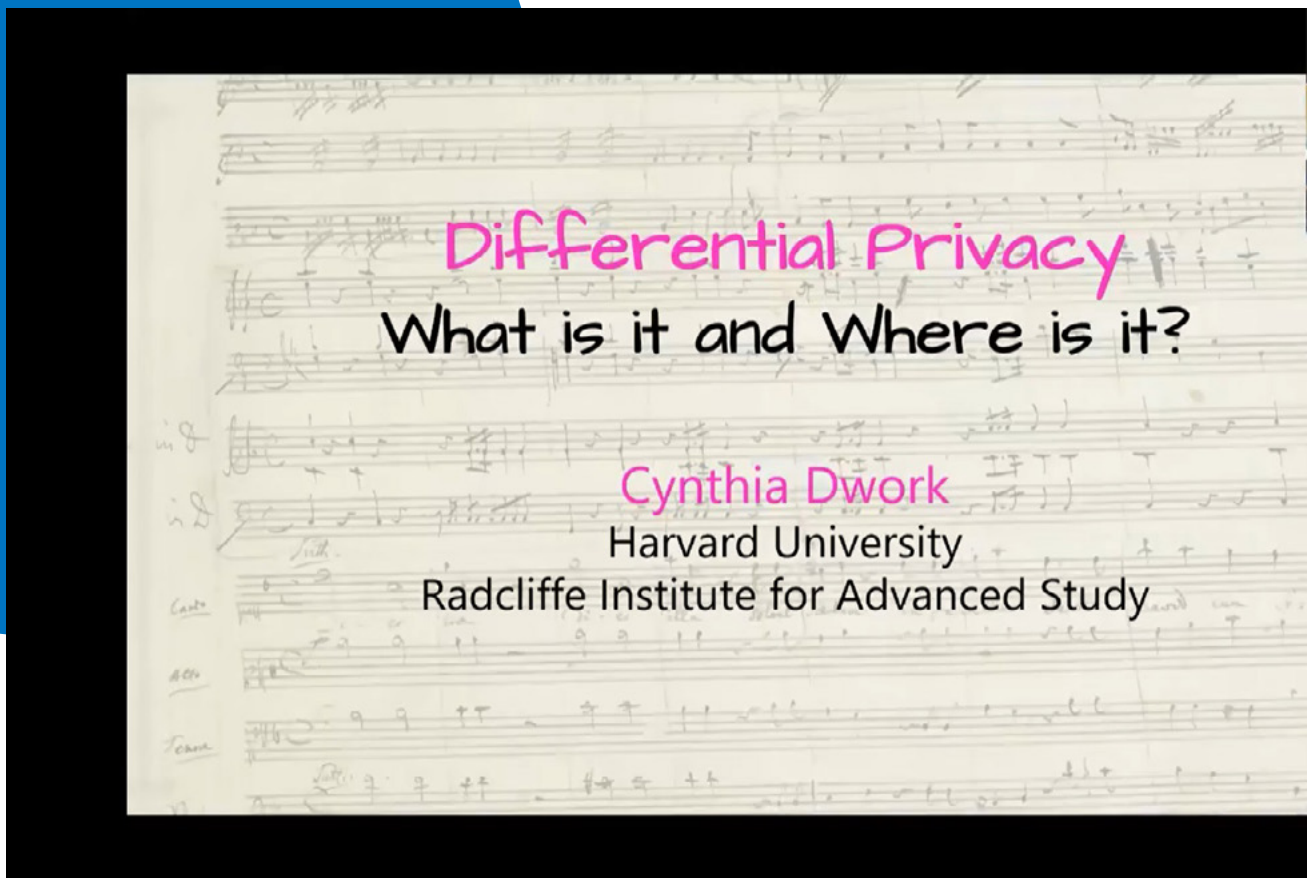
요가 있다. 특히 Charina Chou 박사는 인공지능 기술 활용을 통해 취할 수 있을 것으로 예상되는 사회적 효용을 규제가 저해하는 경우, 이러한 규제를 제거해주는 정부의 역할이 중요하다고 지적하였다. 끝으로 Charina Chou 박사는 구글이 진행하고 있는 기초연구, 응용연구, 인간-인공지능 상호작용 연구와 같은 미래지향적 차원의 논의에 대한 설명을 덧붙이면서 논의를 마무리하였다.

빅데이터와 개인정보 보호:
차등 프라이버시란 무엇인가?

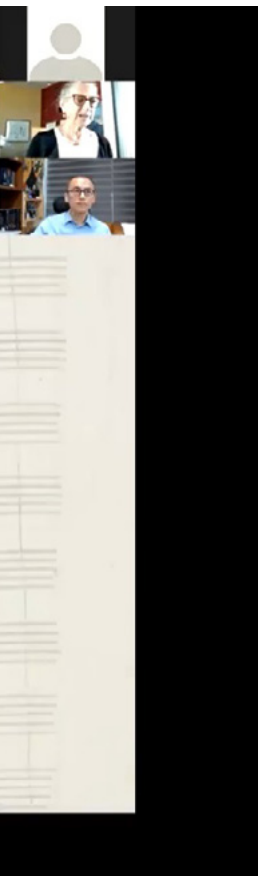
Differential Privacy:
What is it and Where is it?

발표자

Cynthia Dwork, Harvard University


Differential Privacy
What is it and Where is it?
Cynthia Dwork
Harvard University
Radcliffe Institute for Advanced Study

Differential Privacy: What is it and Where is it?

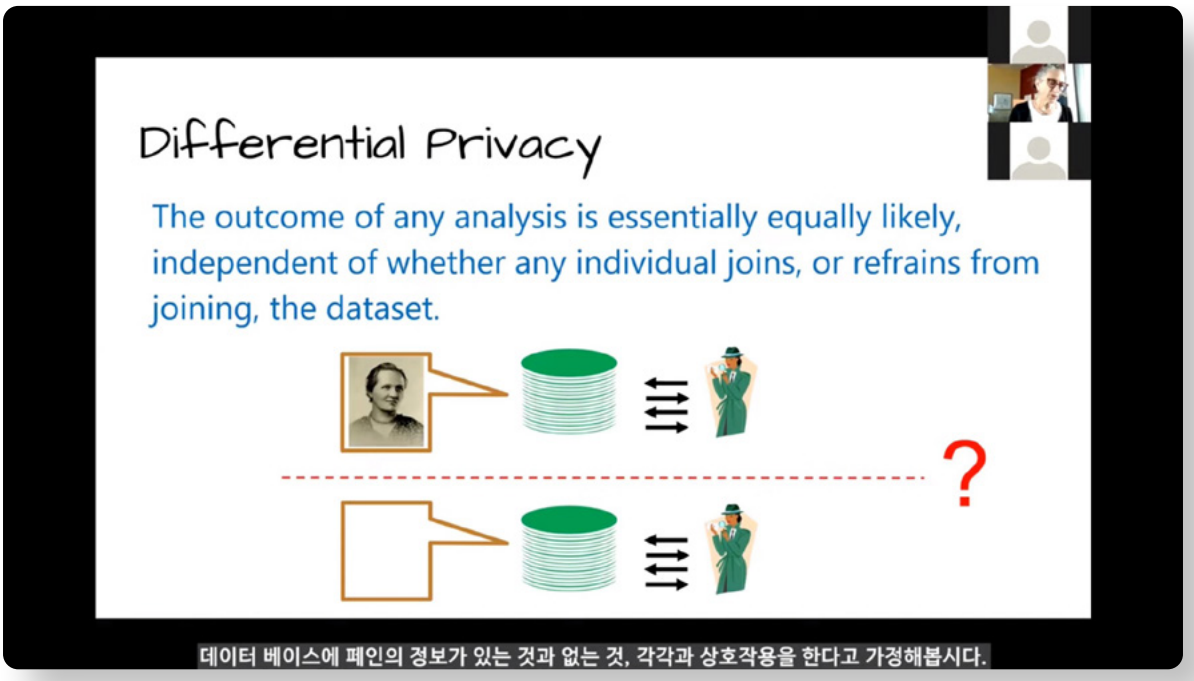


제5세션은 빅데이터 시대에 데이터 분석과 개인정보 보호를 조화하는 차세대 기술로 많은 관심을 받고 있는 차등 프라이버시(differential privacy)를 주제로 논의를 진행하였다. 차등 프라이버시 개념을 창안한 발표자 Cynthia Dwork 교수는 모집단 전체에 대한 통계적 데이터 처리와 전체 중 특정 일부를 분석하는 작업의 차이를 설명하면서 차등 프라이버시는 집단 전체에 대한 통계처리 과정에서의 프라이버시 보호 수단임을 강조하면서 세미나를 시작하였다. 이러한 통계분석의 과정은 일견 그 자체로 프라이버시를 보호하는 것처럼 느껴진다. 표본을 통해 모집단의 특성을 일관되게 추론할 수 있으면서도 표본값은 개인에 대한 확정적인 정보를 전달하지 않기 때문에, 개인은 언제든 "나는 저 통계에 포함되지 않았다"고 주장하여 자신의 정보를 숨길 수 있기 때문이다.

차등 프라이버시는 통계학의 오래된 문제인 "프라이버시를 보존하는 데이터분석 방법(privacy-preserving data analysis)"을 찾고자 하는 시도에서, 미국 인구조사(census)에 활용하고자 개발되었다. 기본적인 분석의 틀은 데이터베이스에 대하여 분석자가 질의(query)를 하면 데이터베이스로부터 이에 대한 응답(answer)을 받는 형태를 취한다. 문제는 응답이 너무 정확하고(overly accurate), 많을수록(too many) 데이터베이스에 포함된 개인의 프라이버시가 침해될 소지가 높다는 것이다.[1] Cynthia Dwork 교수는 이것은 수학적 법칙에 해당한다고 강조하였다.

그렇다면 프라이버시를 보존하는 데이터분석 방법이라고 알려진 차등

1 Cynthia Dwork 교수가 제시한 예시로는 가령 "차량을 보유한 노벨물리학상 수상자수"와 "차량을 보유한 남성 노벨물리학상 수상자수"라는 통계자료가 있다면 유일한 여성 노벨물리학상 수상자가 차량을 보유했는지를 알 수 있는 경우이다.



세션5 캡쳐(자막)

프라이버시는 어떻게 정의되는가? 가장 쉽게 생각할 수 있는 것은 질의 응답을 통해 데이터베이스에 포함된 특정인에 대해 "새로운" 정보를 얻을 수 없으면 프라이버시가 보존된 것이라는 정의이다.[2] 가령 데이터베이스로부터 특정인에 대하여 웹상에 공개된 정보를 얻을 수 있다 하더라도 이는 "새로운" 정보가 아니므로 프라이버시가 침해되는 것은 아니다.[3] 이러한 정의의 문제점은 특정인과 다른 참여자들이 어떤 특성을 공유한다면,

2 DALENIUS, T. (1977), *Towards a methodology for statistical disclosure control.* Statistik Tidskrift, Vol. 15, 429-444, 2–1

3 Cynthia Dwork 교수는 이 정의가 암호학의 Semantic Security와 같은 개념이라고 설명하였다. Semantically secure한 암호체계는 암호문으로부터 원문에 관한 무시가능한(negligible) 정보만을 추출할 수 있는 암호체계를 의미한다.

다른 참여자를 관찰하여 얻은 결과가 특정인에 대한 “새로운” 정보가 되어 본 정의에 따르면 특정인의 프라이버시가 침해된 것으로 인정된다는 점이다. 극단적으로 말해, 세상 모든 사람이 공유하고 있는 특성이 알려지는 상황에서는 모든 개인의 프라이버시가 침해된다는 것인데, 이는 직관적으로 받아들이기 어려운 결론이다.

차등 프라이버시는 이 문제를 해결하기 위해 특정인이 다른 사람으로 대체된 상태에서 질의응답이 이루어져도 특정인에 대하여 같은 “새로운” 정보를 얻을 수 있다면 프라이버시는 보존되는 것이라고 정의하였다. 즉, 어떠한 특정인이 샘플에 포함되든 포함되지 않든 분석의 결과가 유사한 경우, 그 데이터셋은 차등 프라이버시를 만족하는(differentially private) 것이다. 분석의 결과가 유사하다는 것은 확률분포가 유사한 경우를 의미한다. 가령 앞면이 나올 확률이 50%인 동전과 50.1%인 동전이 있다면 이 동전을 반복적으로 던져 얻는 확률분포함수는 거의 유사하기 때문에 1회 또는 수회의 동전 던지기 결과값 만으로는 어떤 동전을 던진 것인지 알 수 없는 것과 같은 원리이다. 이러한 데이터베이스의 특성을 안정성(stability)이라고 표현하는데[4] 이는 데이터셋에 포함된 자의 프라이버시를 보존할 뿐만 아니라 머신러닝 맥락에서는 과적합(over-fitting)을 방지하는 것으로 알려져 있다.[5]

Cynthia Dwork 교수는 차등 프라이버시의 수학적 정의에 대해서도 설명하였다. 간략히 요약하자면 모델 M은 모든 인접한 데이터셋 x, y[6]와

4 데이터셋에 특정인을 포함시키는 경우와 그렇지 않은 경우 결과값이 “안정적으로” 유사하다는 취지이다.

5 즉, 학습 데이터셋(training set)과 시험 데이터셋(test set)이 유사하다면(안정적이라면) 학습 데이터셋을 이용해 학습한 모델이 시험 데이터셋에서 잘 작동할 것이다. 그렇지 않은 경우 모델이 학습 데이터셋에 과적합된 것이다.

6 이상에서 살핀 바와 같이 특정인이 포함되었는지 여부만을 제외하면 완벽히 동일한 데이터셋을 상정한 것이다.

모든 결과값 S에 대하여 아래의 조건을 만족하면 ε-privacy를 만족한다.[7]

Pr[M(x)에서 S를 관찰] ≤ e^{ε}Pr[M(y)에서 S를 관찰]

이상의 정의에서 ε은 프라이버시 손실(privacy loss)을 의미하며 차등 프라이버시의 정도를 의미한다. 위와 같은 차등 프라이버시의 정의는 모델 M의 특성을 나타낸 것이다. 어떠한 분석에 대해서도 이상의 부등식을 만족하는 알고리즘(모델) M을 ε-privacy를 만족한다고 표현하는 것이다.

Cynthia Dwork 교수는 나아가 차등 프라이버시의 특성을 설명하였다. 우선 차등 프라이버시는 미래에도 보전된다(future-proof). 이는 정의 자체에서 나오는 특성으로 미래에 추가정보가 제시되는 등 변화가 일어나도 차등 프라이버시는 유지된다. 둘째, 차등 프라이버시가 적용된 모델의 경우도 개별 질의응답을 통해 미세한 프라이버시 손실이 발생하는데 이러한 손실은 연산을 거듭함에 따라 누적되는 성격을 갖는다.

Cynthia Dwork 교수는 이어서 차등 프라이버시를 구현하는 구체적인 방법인 라플라스 메커니즘(Laplace Noise Addition)을 설명하였다. 라플라스 메커니즘은 분산이 개인정보의 민감도(sensitivity)/ε 상수에 비례하도록 설정된 라플라스 분포로부터 생성한 임의의 노이즈를 원본값에 더하여 변조된 결과를 이용자에게 제공하는 기법을 가리킨다. 이러한 노이즈를 어느 시점에 더할 것인지에 따라 지역적 모델(local model)과 중앙화 모델(centralized model)을 구분할 수 있다. 지역적 모델은 클라이언트 디바이스에서 변조된 개인정보가 서버로 수집된다는 것이 중요한 특징이다. 그러한 점에서 정보주체의 개인정보가 그대로 서버로 수집된 뒤

7 Dwork C., McSherry F., Nissim K., Smith A. (2006) *Calibrating Noise to Sensitivity in Private Data Analysis.* In: Halevi S., Rabin T. (eds) Theory of Cryptography. TCC 2006. Lecture Notes in Computer Science, Vol. 3876. Springer, Berlin, Heidelberg.

변조의 과정을 거치는 중앙화 모델과 대비된다. 가령 구글과 페이스북은 COVID-19에 대응하기 위해 공개한 이동 데이터(mobility data)에 중앙화 모델을 사용하였고, 마이크로소프트는 Windows의 오류 리포팅과 Office의 텍스트 예측 모형에 중앙화 모델을 사용하고 있다. 지역적 모델은 구글이 2014년 적용한 RAPPOR(Randomized Aggregatable Privacy Preserving Ordinal Response) 기법에 차등 프라이버시가 익명처리의 기법으로 사용된다고 명시적으로 설명한 것이 대표적 사례이다.[8]

나아가 차등 프라이버시는 2020년 미국 인구조사(Census) 결과를 공개하는데 적용될 예정이다. 미국 통계국(US Census Bureau) Chief Scientist인 John Abowd는 2010년 인구조사[9]에 대하여 "기술 발전이 기존 방식의 문제점을 노출시켰다. 기존에 프라이버시를 보존한다고 알려졌던 공개자료를 활용해 프라이버시 침해 소지가 있는 정보를 재조합할 수 있게 되었다"라고 하였다. Cynthia Dwork 교수에 따르면 일부 연구자는 인구조사 결과에 차등 프라이버시를 적용하는 것이 데이터베이스의 효용(utility)을 저해한다는 이유로 이에 비판적이기도 하다.

Cynthia Dwork 교수는 데이터베이스의 효용과 프라이버시의 관계가 상충한다는 통념적인 생각을 비판하였다. 데이터를 활용하는 기업, 정부, 연구자와 정보주체 간에는 효용과 프라이버시의 관계에 대한 선호가 다를 수 있다. 일반적으로 프라이버시가 전혀 인정되지 않는 사회에서 데이터 활용자의 효용은 가장 높은 상태일 것이고, 프라이버시에 대한 보호가 강화될수록 데이터 활용자의 효용은 감소할 것이다. 반면 정보주체 입장에서는 "숨길 것이 없다"고 생각하는 개인은 프라이버시 보호가 없는 경우에

8 Google, "Google에서 데이터를 익명화하는 방법", 개인정보 보호 및 약관 https://policies.google.com/technologies/anonymization?hl=ko

9 미국은 헌법 Article 1 Section 2에 따라 매 10년 인구조사를 실시한다.

도 자신의 정보를 사실대로 제공하겠지만, 그렇지 않은 정보주체는 사회를 신뢰하지 못하고 사실대로 정보를 제공하지 않을 수 있다. 그렇다면 사회적 총효용과 프라이버시의 관계는 위로 볼록한 역U자 모양이 될 것이고, 이에 근거해 Cynthia Dwork 교수는 프라이버시 보호가 너무 적거나 많은 경우 데이터베이스의 효용이 낮아질 것이라고 주장하였다.

Cynthia Dwork 교수는 데이터베이스의 효용과 프라이버시의 관계에 기초해 프라이버시 보호수준을 설정하는 것은 결국 정책적 문제이며 차등 프라이버시 체계 하에서는 그러한 정책적 결정은 ε값의 조정을 통해 기술적으로 구현할 수 있다고 지적하였다. 이는 ε값을 누가 어떻게 정할 것인지, 이로 인한 프라이버시 손실이 누적될 때 어떤 질의(query)를 우선시할 것인지에 관한 문제를 남긴다. 가령 이러한 문제에 대하여 Abowd & Schmutte는 "프라이버시 손실을 감안한 데이터 분석 정확도에 대한 지불의사(willingness to pay for data accuracy with increased privacy loss)"를 측정하는 방식으로 최적의 사회적 해법을 찾을 수 있다고 주장하였다.[10]

Cynthia Dwork 교수는 차등 프라이버시는 데이터베이스에 속한 특정인에 관한 정보는 알지 못하게 하면서도 데이터베이스 분석을 가능하게 한 접근법이라고 정리하며 세미나를 마무리하였다.

10 Abowd, J., Schmutte, I. (2019) *"An economic analysis of privacy protection and statistical accuracy as social choices",* American Economic Review, Vol. 109, No.1, pp 171, 194-197.

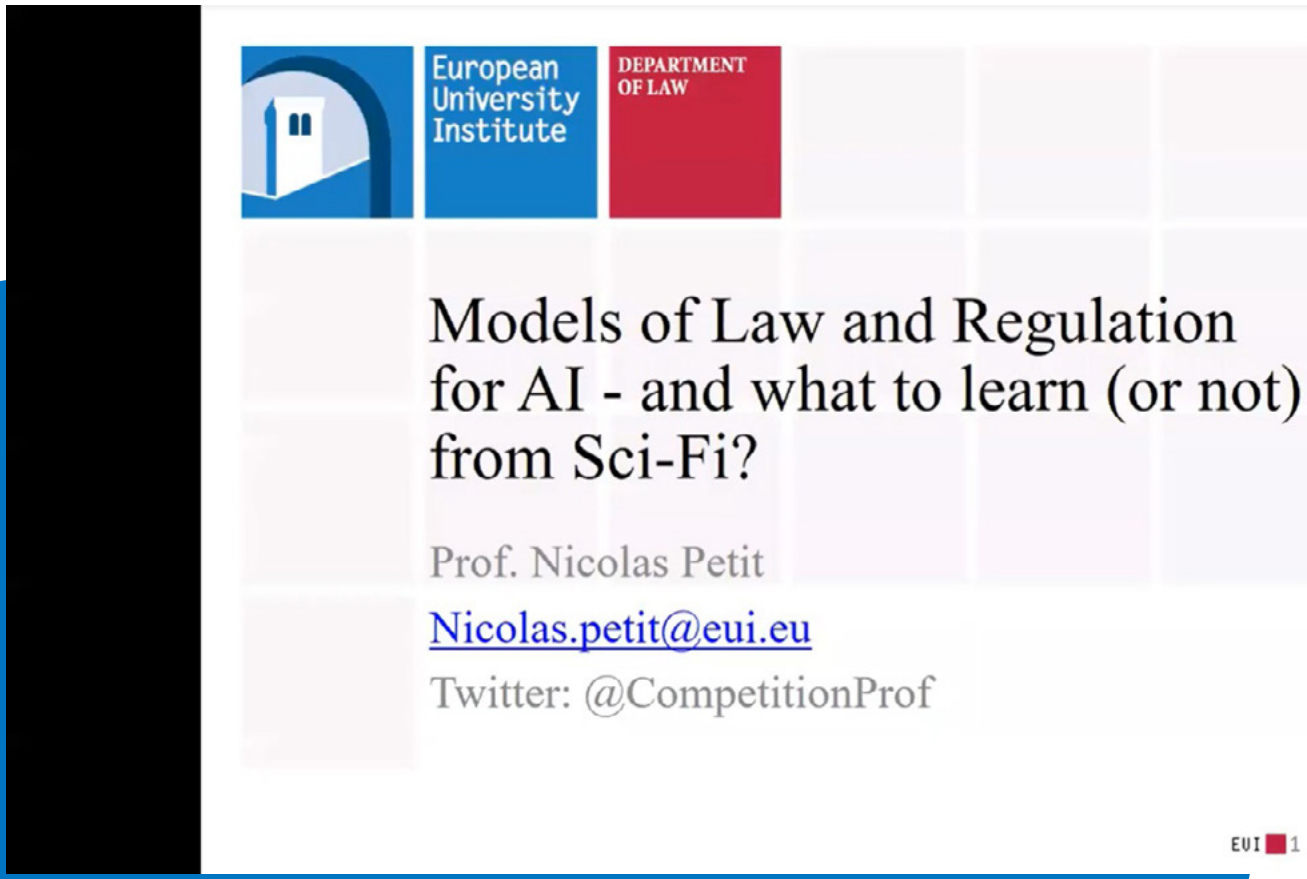
제6세션

인공지능 규제:
법률가를 위한 아시모프

Asimov for Lawyer: What Sci Fi can(not) tell us about the future of AI regulation

발표자

Nicolas Petit,
European University Institute


European University Institute
DEPARTMENT OF LAW
Models of Law and Regulation for AI - and what to learn (or not) from Sci-Fi?
Prof. Nicolas Petit
Nicolas.petit@eui.eu
Twitter: @CompetitionProf
EUI 1

제6세션은 인공지능 기술에 대한 규제를 주제로 논의가 이루어졌다. 발표자 Nicolas Petit 교수는 현재 인공지능 기술에 대한 규제 담론이 전 세계적으로 활발히 진행되고 있다고 언급하면서, 공상과학 소설로부터 향후의 논의가 나아가야 할 방향의 교훈을 얻어보고자 한다며 발표를 시작하였다.

우선 인공지능이란 무엇인가에 대하여, 이는 Church-Turing 논제(thesis)에서 나오는, 사고의 과정을 포함한 일체의 물리적 과정은 컴퓨터 알고리즘으로 모델링할 수 있다는 생각에서 출발한다. 나아가 인공지능은 인간처럼 경험을 기반으로 학습하여 그 기능을 더욱 개선할 수 있다. 이러한 인공지능 분야는 지난 수십 년 간 열망과 실망의 시기를 모두 지나왔고, 대략 지난 15년 전부터 컴퓨터 기술의 연산력 향상, 빅데이터의 축적 등에 힘입어 다시 기대가 높아지고 있는 상황이다. 심지어 최근 학자들은 "이론의 종식(end of theory)"이라는 표현을 통해 인공지능의 모든 프로세스를 이론적으로 설명할 수도 없고, 설명할 필요도 없는 단계에 접어들고 있다고 말하기도 한다. 오늘날 인공지능은 딥러닝(deep learning), 신경망 네트워크 등 방식의 학습을 하여 온전히 설명할 수 없는 결과를 도출하면서, 자율주행, 사법판단 결과의 예측, 법집행, 교육 등 사회 곳곳에서 활용되고 있는 추세이다.

Nicolas Petit 교수는 이러한 인공지능 기술에 적용될 만한 규제 모델

을 4가지로 유형화하여 소개하였다. 첫째는 Black letter law 모델[1]로, 이에 따르면 분쟁이 발생하였을 때는 기존의 법체계에서 명확하게 구축된 해당 사안에 적합한 특정한 영역의 개별법을 법원(source)으로 삼아 해결 방안을 모색하게 된다. 가령 인공지능이 작곡한 음악이 문제될 경우, 현행 저작권법이 규정하는 "저작자에 의한 창작물"의 요건에 부합하는지 여부가 논의의 핵심이 된다. 현행법은 일반적으로 인공지능 기술을 염두에 두고 만들어진 것이 아니므로, 법원과 입법자는 이러한 법령이 애초에 가졌던 목적이 무엇이었는지에 주목하여 한계 사안의 문제해결을 도모하게 된다. 예를 들어, "법인"이라는 개념이 도입된 이유는 "경제적 교류의 활성화"라는 목적을 위한 것이었으므로, 인공지능의 법인격이 문제되는 맥락에서는 이러한 목적의 달성과 결부하여 논의가 진행되어야 한다는 것이다.

두 번째 모델은 신생 현상 모델(Emergent phenomena model)로, 새롭게 나타나는 현상을 규율하는 법규범을 모색하는 모델이다. 이 모델은 인공지능 기술이 기존에 없던 경제적, 과학적 이슈를 만들어낼 것이어서 새로운 입법이 필요하다는 생각에 기초한다. 최근의 예로 드론법, 로봇법 등에 대한 논의들을 살펴볼 수 있다. 법률 전문가가 주로 논의를 이끄는 첫 번째 모델과 달리, 두 번째 모델에서는 공학 전문가와 같은 비법률가가 주로 논의과정에 참여한다는 차이가 있다. 또한 첫 번째 모델과 달리, 특정 영역에 국한하지 않고 통합적 규율을 강조하고 보다 규범적인(normative) 논의에 초점을 맞춘다. 기술 전문가들은 "법이 하는 것(does)"보다 인공지능 기술에 어떻게 법을 적용하여야 할지에 관한 "법이 해야할 것(should)" 대하여 더 초점을 맞추는 경향이 있기 때문이다.

세 번째 모델은 윤리 모델(Ethical model)로, 응용윤리와 결부되어 인

1 오랫동안 축적된 기초법 원칙에 의한 구체적 법규범의 형태를 이르는 표현이다.

공지능에 선과 악을 구분하는 규범을 제공해야 한다는 입장이다. 윤리 모델이 법과 규제에 대한 논의를 명시적으로 규정하고 있지는 않지만, 암묵적으로 이러한 생각이 전제되어 있다. 덕 윤리(virtue ethics), 의무론적 윤리(deontological ethics), 결과주의 윤리(consequentialism)가 윤리 모델을 구성하는 대표적 사례이다. 덕 윤리의 선(good)이란 일상에서의 도덕적 행위에 기초하고, Nicolas Petit 교수에 따르면 인공지능 맥락에서는 "투명성(transparency)"으로 이해된다고 한다. 한편 의무론적 윤리는 인공지능이 결과와 무관하게 특정한 지시사항에 따라야 한다는 점을 강조하고, 결과주의 윤리는 의무와 무관하게 결과가 선해야 한다는 점을 강조하게 된다. 위와 같은 윤리적 모델들은 과거 생명윤리 등 다른 기술 분야에 적용된 바 있고, 현재 우리가 인공지능 기술에 적용하려는 내용과 크게 다르지 않다고 한다.

네 번째 모델로는 위험 규제(Risk regulation) 모델로, 결과주의와 유사하면서도 해악 자체가 아닌 해악이 발현될 확률을 기술적으로 낮추는 데 초점을 둔다. 유럽연합의 "인공지능 백서"[2]에 나타난 바와 같이 인공지능 기술이 가지는 잠재적 위험성을 경감하는 것이 핵심적 사항이 된다. 이러한 접근방식은 사후조치에 초점을 둔 결과주의와 달리 사전예방에 초점을 두고, 이는 통계학적 근거에 기반(evidence-based approach)을 둔다는 특징을 가진다. 따라서 확률로 환산할 수 있는 데이터가 있어야 한다는 한계를 가진다. 인공지능 분야에서는 특정 분야(예: 무기)에 대한 금지 조치, 또는 위험성이 큰 분야(예: 얼굴인식 기술)에서의 예방 조치 등으로 나타날 수 있다.

Nicolas Petit 교수는 이어서 지금까지 언급한 모델에 적용될 수 있는

2 European Commission. (2020). *"White paper on artificial intelligence–a European approach to excellence and trust"* Brussels.

4가지 가능한 오류(fallacies)에 대해 설명하였다. 이는 (1)무관한 법의 역설(The paradox of irrelevant law), (2)중복적인 법의 문제(The problem of redundant law), (3)선의의 실패(The failure of good intentions) 그리고 (4)반사적 규제(Knee jerk regulation)를 가리킨다.

먼저 "무관한 법의 역설"은 특히 명시적 규범을 전제로 하는 첫 번째 모델과 관련이 있다. 법률가들이 현재의 기술을 바탕으로 미래를 상상하기 때문에 실제 발생할 상황과 무관한 법이 등장할 수 있다는 것이다. 만일 법률가들이 과거부터 "자율주행" 차가 아닌, "날아다니는" 차량에 대해 들어왔다면 지금쯤 자율주행차가 아닌 날아다니는 차에 대한 상세한 법을 가지고 있을지도 모른다. 중복적인 법의 문제는 두 번째 모델과 관련된 것으로, 신기술이 유발하는 현상이 기존과 전혀 다른 새로운 것이라는 생각에서 만든 법이 기존의 법과 중복되는 부분을 규율할 때 생기는 문제를 가리킨다. Nicolas Petit 교수는 현실적으로 우리가 마주하는 것이 "새로운 문제"가 아니라 "새로운 방법으로 야기되는 기존의 문제"일 수 있다고 지적하였다. 세 번째 오류인 선의의 실패는 특히 윤리 모델과 관련된 것으로, 선한 의도로 기술에 적용한 윤리가 현실적으로 부작용을 낳는 현상을 의미한다. 윤리에 지나치게 방점을 두게 되면 정작 법적 규제가 필요한 시점에 제대로 대처할 수 없다는 문제가 대표적 사례이다. 게다가 전 세계적으로 공통되는 윤리관은 존재하지 않기 때문에, 트롤리 딜레마와 같은 현실적 사례에서 어떻게 대응해야 할지 난처해질 수 있다. 마지막 오류인 반사적 규제의 문제는 발생할 수 있는 위험에 대하여 과잉의 예방적 규제를 도입하는 현상을 말한다. 후쿠시마 원전 사태에 대응하여 화석연료로 회귀하거나, 자율주행 운행이 인간의 사고율보다 낮은데도 사고 자체를 문제 삼고 전면적 금지를 하는 행태가 과잉규제의 대표적 사례에 해당한다.

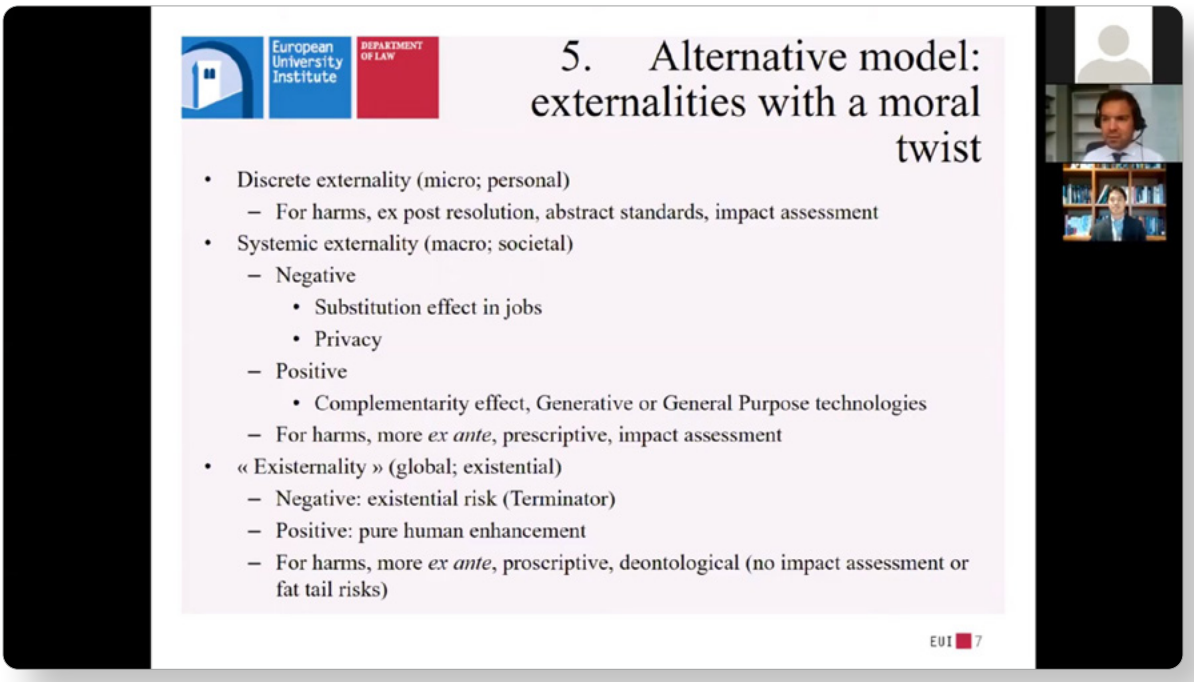
그 다음으로 Nicolas Petit 교수는 인공지능 기술에 대한 공공정책의 측면에서, 기술의 혁신을 기존의 법체계 틀 안에서 규제할 것인지, 아니면

새로운 규제를 준비할 것인지라는 새로운 화두를 제시하였다. 새로운 형태의 규제에 반대하는 입장에서는 (1) 규제와 혁신의 상충 문제 (예: 유럽연합의 GDPR[3]에서 요구하는 데이터 최소화(data minimizing)), (2) 규제 포획(captured regulation)의 문제 (예: 자율주행차 규제 도입 시 자동차보험 업계와 공익 사이의 충돌), (3) 규제를 도입하기 위해서는 기술에 대한 정보를 충분히 축적해야 하지만, 기술이 너무 발전하게 되면 이미 이를 되돌릴 수 없는 단계에 이르러 있을 수 있다는 딜레마의 문제를 들고 있다.

이에 대하여 Nicolas Petit 교수는 기존 법규범을 바탕으로 문제를 해결하는 것은 혁신에 도움이 되지 않는다고 지적하였다. 오히려 규제를 마련함으로써 기술자, 발명가들에게 신호를 보내 올바른 방향으로 연구를 할 수 있도록 할 수 있다는 것이다. 이러한 관점에서, Nicolas Petit 교수는 다섯 번째의 규제 모델로 대체 모델(Alternative model: externalities with a moral twist)을 제시하였다. 대체 모델은 고려되어야 하는 세 가지 효과로, (1)개별적 외부효과(discrete externality), (2)체계적 외부효과(systemic externality), (3)존재론적 외부효과(existerniality)를 제시한다. 개별적 외부효과는 미시적, 개인적 관점에서 고려해야 할 사항으로 기존의 법체계로 대응할 수 있고, 체계적 외부효과는 거시적, 사회적 관점에서 고려해야 할 사항으로 고용이나 개인정보 문제처럼 보다 높은 수준의 사전적 규제가 필요하고, 존재론적 외부효과는 인류 전체의 관점에서 고려해야 할 사항으로 상당히 높은 수준의 규제가 필요하다고 한다.

발표를 마무리하면서 Nicolas Petit 교수는 공상과학이 현실성이 떨어진다는 주장에 대하여, 그 안에 담긴 상상력을 통하여 법과 규제를 인공

3 유럽 일반 개인정보 보호규정으로 공식 표현은 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) 이다.


European University Institute
DEPARTMENT OF LAW
5. Alternative model: externalities with a moral twist
• Discrete externality (micro; personal)
– For harms, ex post resolution, abstract standards, impact assessment
• Systemic externality (macro; societal)
– Negative
• Substitution effect in jobs
• Privacy
– Positive
• Complementarity effect, Generative or General Purpose technologies
– For harms, more ex ante, prescriptive, impact assessment
• « Existernality » (global; existential)
– Negative: existential risk (Terminator)
– Positive: pure human enhancement
– For harms, more ex ante, proscriptive, deontological (no impact assessment or fat tail risks)
EUI 7

세션6 캡쳐(자막)

지능 영역에 어떻게 적용할 수 있을지에 대한 통찰력을 얻을 수 있다고 반박하였다. 특히, 아시모프의 소설을 보면 변화를 예측하고 기술 발전에 따른 인간의 행동 변화를 생각해 볼 수 있다고 하면서, 법률가에게도 도움이 된다고 보았다. 그와 동시에 아시모프가 기술에 대해 낙관적이지도, 비관적이지도 않은 중립적 입장을 취하였다는 점을 강조하였다. 아시모프가 "Runaround"라는 소설에서 주장하는 "로봇 3원칙[4]"이 법의 약점을 보완하기 위한 도구였듯, 인간이 만든 것은 비단 기술 뿐만이 아니라 법정책에도 내재적인 오류가 있을 수 있을 수 있음을 지적하면서 논의를 마무리하였다.

4 로봇 3원칙의 내용은 다음과 같다.
1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

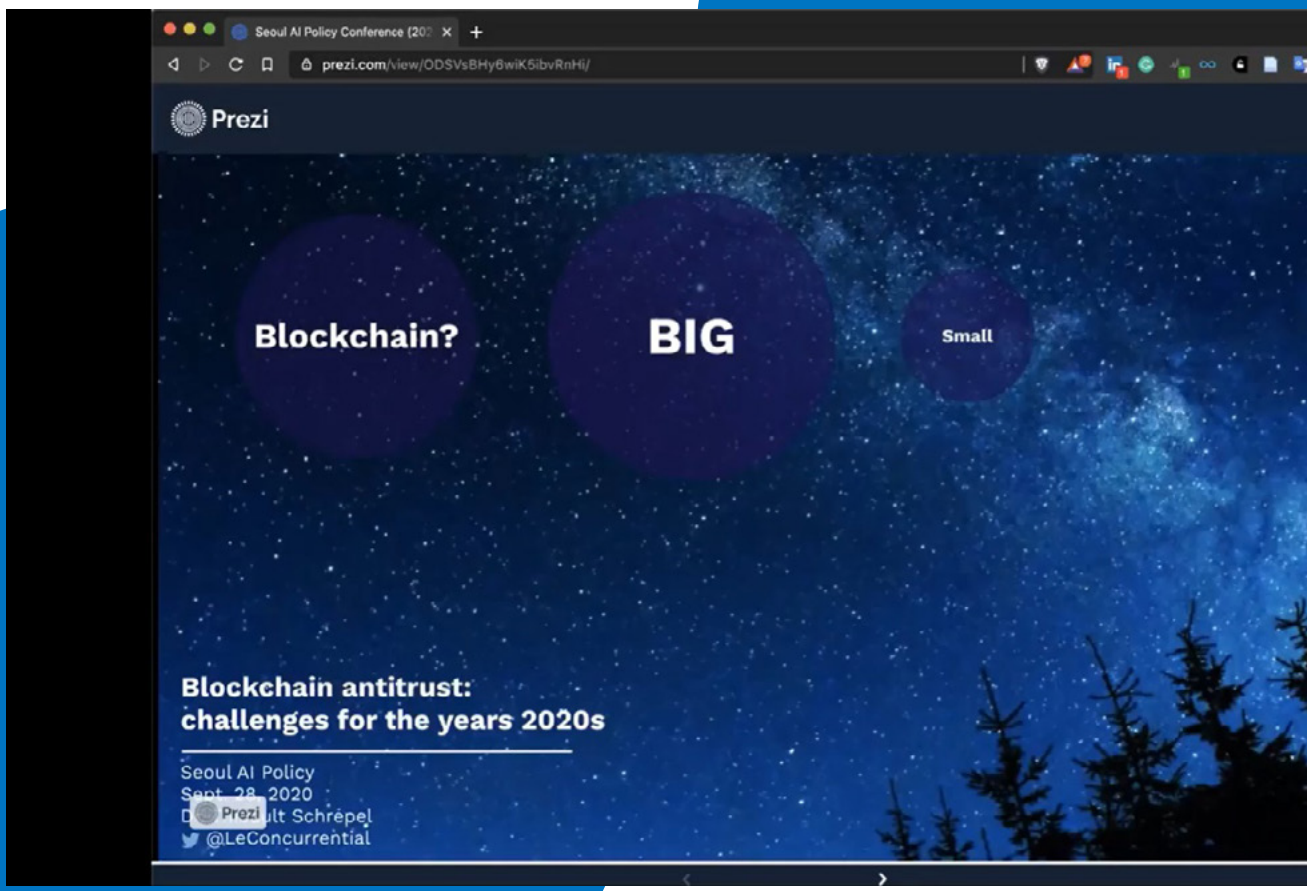
제7세션

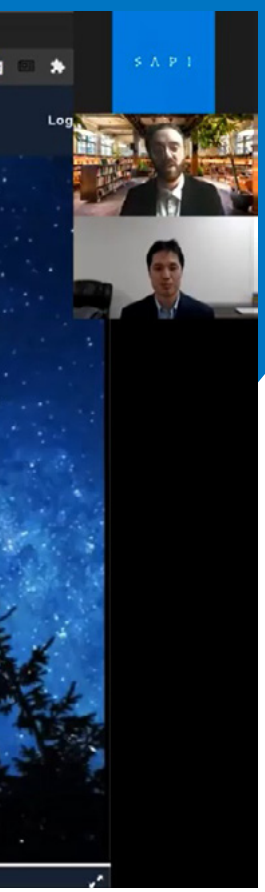
알고리즘과 블록체인: 경쟁법의 새로운 쟁점

Blockchain Antitrust: Challenges and Opportunities

발표자

Thibault Schrepel, Utrecht University


Seoul AI Policy Conference (202
prezi.com/view/ODSVsBHy6wiK5ibvRnHi/
Prezi
Blockchain?
BIG
Small
Blockchain antitrust:
challenges for the years 2020s
Seoul AI Policy
Sept. 28, 2020
Prezi
@LeConcurrential



제7세션은 최근 많은 논의가 이루어지는 블록체인 기술로 대상을 전환하여 논의를 이어갔다. 발표자 Thibault Schrepel 교수는 법정책적 논의를 본격적으로 진행하기에 앞서 블록체인이 무엇인지에 대한 간략한 설명을 제시하면서 발표를 시작하였다. 블록체인은 일종의 아키텍처(Architecture)로 여러 가지 방식의 응용이 가능한 기반기술의 성격을 가진다. 공식적인 정의가 존재하지는 않지만 블록체인이란 일반적으로 "수동적이든 자동적이든, 사용자 간의 모든 종류의 거래를 기록할 수 있는 공개되고(open) 분산된(distributed) 장부(ledger)"를 가리킨다.

블록체인에는 두 가지의 레이어(layer)가 있는데 레이어1은 엑셀 스프레드시트와 같이 정보를 기입한 장부가 공유되는 영역으로, 일반적으로 블록체인이라고 할 때는 레이어1을 가리킨다. 이는 개방된 분산형 데이터베이스로, 엑셀 시트와 같은 형태로 생각해 볼 수 있다. 레이어1에 레이어2를 추가할 수 있는데, 두 번째 계층인 레이어2[1]의 종류에 따라 다음 세 가지 종류로 블록체인 소프트웨어를 분류할 수 있다. 즉, 블록체인은 ①가상화폐(crypto-curriencies), ②스마트 계약(smart contracts), ③그 외 다른 종류의 모든 어플리케이션(other types of applications, 예: 우버)으로 분류된다.

Thibault Schrepel 교수는 이중 특히 경쟁법과 관련하여 중요하게 다루어야 할 개념인 스마트 계약이 체결되는 과정을 동영상으로 시연하면서 소개하였다. 홈페이지에서 당사자의 이름을 입력하고, 어떠한 법률의 적용을 받을 것인지 또는 각자에게 부과되는 의무는 어떠한 것인지를 적용하고, 양 당사자의 이메일 주소를 입력하면 자동적으로 당사자의 이메일 주소로 해당 계약서가 송부된다. 이때 공개 키(public key)를 통해 식별된 사용자가 송부된 스마트 계약서에 전자서명을 하게 되면 거래가 완료되

1 레이어 2(Layer 2)란 기존 블록체인 시스템 위에 얹어서 운영되는 어플리케이션을 의미한다.

고, 계약서에 명시된 의무를 일방 당사자가 이행하게 되면 대금이 지불된다.

다음으로 Thibault Schrepel 교수는 법률가와 경제학자들이 제대로 이해해야 하는 4가지의 개념적 도구(toolbox)를 설명하였다. ① 블록체인은 개인을 식별하는 데 실제 신원 대신 공개 키로 표시되는데, 현재 기술로는 공개 키를 실제 신원으로 변환하는 것이 불가능하고 개인 키(private key)가 있어야 정보에의 접근이 가능한 익명성(pseudonymity)을 가진다. ② 블록체인은 분산되어 저장되기에 특정 중앙서버에서 기록을 통제하거나 삭제할 수 없고 모든 사람이 정보에 접근할 수 있다는 탈중앙화(decentralization)를 특성으로 가진다. ③ 블록체인은 거래가 추가되면서 정보가 변경되면 해시(hash) 값이 완전히 달라진다는 점에서 정보의 변조가 불가능하다는 불변성(immutability)을 가진다. ④ 블록체인의 거래 과정은 자동적으로 진행되기 때문에 예를 들어 스마트 계약 상에 특정한 조건으로 중지 조항을 삽입하지 않는 한 저지할 수 없다는(unstoppable code) 특성을 가진다.

Thibault Schrepel 교수는 블록체인 기술에 대한 개관을 마치고, 본격적으로 경쟁법에 대한 논의를 시작하였다. Thibault Schrepel 교수에 따르면 블록체인 기술과 경쟁법은 궁극적으로 경제적 강제력으로부터 자유로운 거래를 실현하는 것을 목적으로 삼는다고 한다. 결국 블록체인 기술의 목적인 반독점적 경제를 실현하려면 경쟁법이 필요하고, 역으로 집행당국이 반경쟁적 행위를 적발하는 비율이 매우 낮은 현실에서 블록체인 기술이 경쟁법의 취지를 구현하는 데 도움이 될 수 있다고 주장하였다.

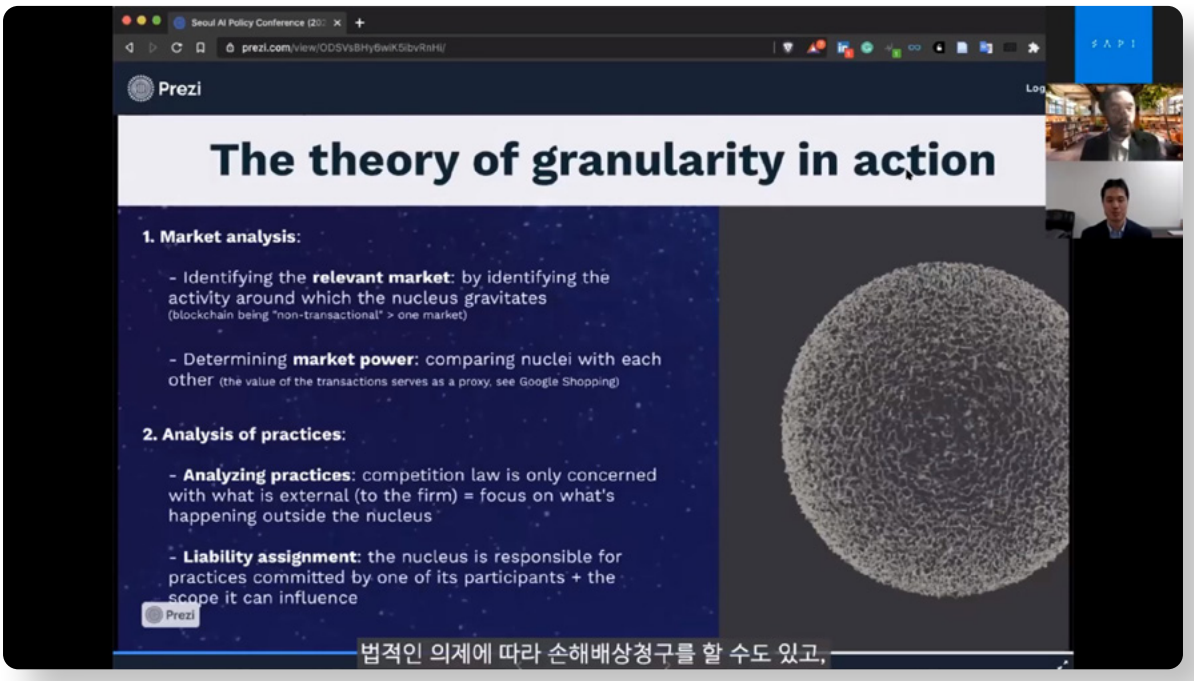
그런데 Thibault Schrepel 교수에 따르면, 경쟁법의 기저에 놓인 기업이론의 관점에서 블록체인 기술의 특성이 난점을 유발한다고 지적하였다. 미국과 유럽의 경쟁법의 적용 대상인 기업을 정의할 때 기반이 되는 Ronald Coase의 기업 이론(Theory of Firm)에 따르면, 기업이란 일

체의 경제활동에 수반되는 의사결정을 지배권을 가진 주체의 하향식(top-down) 의사결정 과정으로 대체하여 거래비용을 절감하려는 목적으로부터 탄생한 산물이다.[2] 설령 별개의 법인이라고 하더라도 지배주체가 동일하다면 양자 사이에서는 담합이 이루어질 수 없다. 누구도 자기 자신과 담합할 수는 없기 때문이다. 이렇게 본다면 기업의 핵심은 "법인격"이 아니라 "지배권"에 있는 셈이다. 문제는 블록체인 내의 핵심 개발자(core developer), 채굴자(miner), 사용자(user)라는 세 가지 주체 누구도 서로를 통제하지 않고 있기에 전통적 기업의 하향식 통제 모델이 블록체인에는 적용될 수 없다는 것이다.

Thibault Schrepel 교수는 이러한 문제를 해소할 필요가 있다는 문제의식에서 경쟁법을 재구성한 자신의 논문, *The Theory of Granularity: A Path for Antitrust in Blockchain Ecosystem*[3]의 내용을 소개하였다. 논문은 첫째, 현재 블록체인에서 핵심 개발자, 채굴자, 사용자 간의 역학관계가 만들어내는 거시적 차원의 현상에 대하여 분석하였다. 우선 ①핵심 개발자는 블록체인을 디자인하고 법칙과 규칙을 정하므로 나름의 힘을 가지지만 개발을 하게 되면 더 이상 통제권이 없고, 이후에 규칙을 바꾸고자 하면 사용자나 채굴자의 동의가 요구된다. ②사용자는 어떤 블록체인을 사용할지 선택할 수는 있지만 블록체인 자체의 디자인에는 관여하기 어렵다. 이들은 대체로는 개별적으로 활동하지만 조직을 형성하여 활동하기도 하는데 이로 인하여 기업과 비슷한 경쟁저해적인 모습도 나타난다. ③ 채굴자는 블록체인에 올라가는 거래의 정보의 프로토콜을 검증하는 역할을 수행하지만, 다른 주체가 하는 역할에는 관여하지 못한다.

2 Coase, R. H., (1937), *"The Nature of the Firm"*. Economica 4.16: 386-405.

3 Thibault, S., (2020), *"The Theory of Granularity: A Path for Antitrust in Blockchain Ecosystems"*. Available at SSRN: https://ssrn.com/abstract=3519032 or http://dx.doi.org/10.2139/ssrn.3519032



세션7 캡쳐(자막)

이러한 참여자들 사이의 역학관계는 법적 의제를 새롭게 형성하게 되는데 이러한 상황을 분석하고 비교하여 경쟁법의 개념과 논리를 적용하는 것을 가리켜 Thibault Schrepel 교수는 자신의 논문 제목에 따라 The theory of granularity라고 지칭하였다. 즉, 경쟁법의 초점을 기업 대신 참여자들 사이에서 형성되는 역학관계에 따른 지배력에 맞추자는 것이다. 이를 통해 블록체인 생태계에 기업이 존재하지 않으나 경쟁법을 적용할 수 있어 더 많은 사람의 참여가 이루어지고 사회적인 효용도 커질 것이라고 보았다.

다음으로 Thibault Schrepel 교수는 블록체인과 관련하여 담합과 시장지배적지위의 남용에 대한 미시적 차원의 논의를 이어갔다. 먼저 담합의 차원에서, 스마트 계약의 기본적인 작동원리 덕분에 합의의 이행을 신뢰할 수 있어 담합이 용이해질 수 있다. 예를 들면, 담합의 합의를 하였는데 한 사업자가 가격을 5% 낮추는 경우 스마트 계약이 자동적으로 그에게 처벌 조항을 적용하는 것이다. 그리고 이때 스마트 계약은 절대 삭제할 수가 없는데, 사업자들이 이를 우려한다면 스마트 계약으로 담합을 하지 않겠지만 코드화 되어 있어 적발되지 않을 것이라고 생각하면 담합에 스마트 계약을 활용할 것이라고 한다. 즉, 스마트 계약의 등장으로 인해 종래의 비협조게임이 협조게임으로 전환되는 셈이다.

시장지배적지위의 남용 문제에 관하여는, 퍼블릭(public) 블록체인과 프라이빗(private) 블록체인을 각각 살펴보아야 한다면서, 특히 프라이빗 블록체인에서 문제가 될 소지가 크다고 하였다. 참여자 모두가 관여하는 퍼블릭 블록체인과 달리 프라이빗 블록체인은 개인이 규칙을 자의적으로 바꿀 수 있고 접근을 통제할 수도 있기 때문이다. 이러한 논의는 끼워팔기와 같은 다른 경쟁법 사례나 지식재산권법과 같은 다른 법제에도 공통적으로 적용될 수 있다고 한다.

발표를 마무리하면서 Thibault Schrepel 교수는 두 가지 주요 과제를 제시하였다. 한 가지는 전문성(expertise)으로, 블록체인을 이해하고 그 원리를 충분히 알고 있어야 법을 집행할 수 있다는 것이다. 블록체인의 기술적 특성으로 인한 기존과는 다른 새로운 문제가 발생할 수 있기 때문이다. 다음으로는 집행과 관련하여 블록체인에 법을 적용함에 있어 블록체인 산업은 기존의 법적, 사회적 규제에 얽매이지 말자는 것을 기본 철학으로 하고 있어 법 적용이 쉽지 않을 것으로 보이나, 그에 따른 인센티브를 제공할 수 있다면 달라질 수 있다고 하였다.

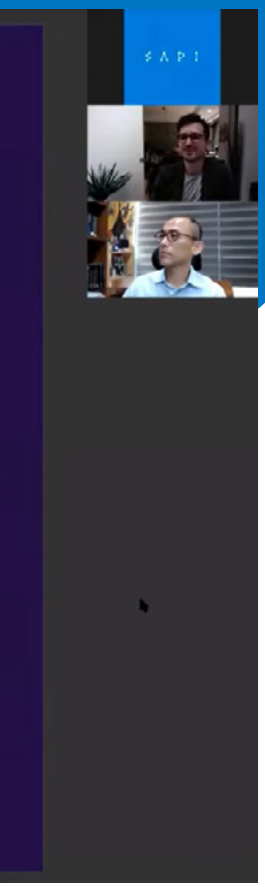
알고리즘과 투명성: 게이밍(Gaming)의 문제(Ignacio Cofone)

When Does Gaming Justify Algorithmic Secrecy?

발표자

Ignacio Cofone, McGil University


When is Algorithmic Secrecy Justified?
Ignacio N. Cofone (McGill University)
Katherine J. Strandburg (New York University)
Montreal-Seoul, 29 September 2020

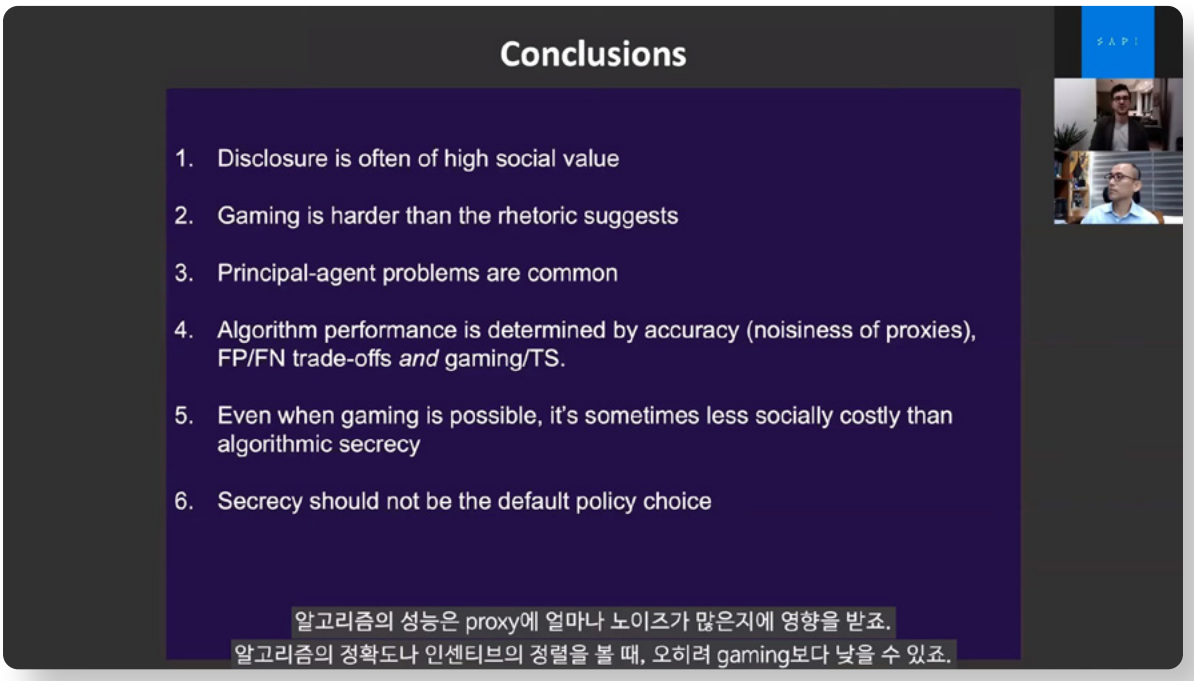


마지막 세션은 인공지능과 투명성을 대주제로 논의를 진행하였다. 발표자 Ignacio Cofone 교수는 19세기에 존재했던 체스 인공지능 “The Turk” 그림을 제시하면서 세미나를 시작하였다. 아마존의 크라우드워킹 플랫폼 Mechanical Turk의 모티브가 된 “The Turk”가 실제로는 그 안에 사람이 들어가 조종한 속임수였듯이, 오늘날 우리가 접하는 인공지능과의 상호작용도 실제로는 인공지능을 프로그래밍하고 디자인하는 사람과의 상호작용으로 환원해서 생각할 수 있음을 상징적으로 보여주는 사례이다.

오늘날 인공지능을 통한 자동화된 결정(algorithmic decisions)은 채용, 금융, 형사절차 등 다양한 분야에서 활용되고 있다. 이것의 작동방식을 간단히 요약하자면, 다양한 피처(feature)를 특정 결과값에 대한 대체변수(proxy)로 사용해 의사결정자의 의사결정 방식을 모방하는 것이다. 가령 소득을 기초로 대출상환 여부를 판단한다면, 소득이라는 피처가 대출상환 여부라는 결과값의 대체변수로 사용된 것이다. 알고리즘 투명성에 관한 논의는 주로 이러한 자동화된 결정에서 사용된 피처와 대체변수에 대한 정보를 공개할지 여부 및 그 범위에 관한 논의이다.

그렇다면 알고리즘 투명성과 관련해서 고려해야 할 상충관계(trade-off)는 무엇인가? 우선 알고리즘 투명성은 컴플라이언스(compliance)[1]를 증진시키며 알고리즘이 가지고 있는 오류(error)와 편향(bias)의 수정을 용이하게 해주는 이점을 가진다. 또한 자동화된 결정을 받는 개인들의 절차적 권리를 보장해주기도 한다. 반면 산업계에서는 알고리즘의 비공개를 주장하는 목소리가 강한데, 알고리즘이 투명하게 공개될 경우 이용자가

1 여기서 컴플라이언스(compliance)란, 이용자가 자동화된 의사결정의 작동방식을 안다는 전제하에 원하는 결과를 달성하기 위해 하는 행위를 의미한다. 가령 카드대금을 연체하지 않는 것이 개인 신용도 평가에 긍정적으로 작용한다는 사실이 공개되었을 때 이용자가 카드대금을 연체하지 않고자 노력하는 행위가 여기에 해당한다. 이하에서 볼 게이밍(gaming)도 이와 유사하기는 하지만, 설계자가 의도한 목적과 일치하는 행위인 컴플라이언스와 달리 게이밍은, 설계자의 본래 목적을 거슬러 이를 이용하는 행위라는 차이가 있다.



세션8 캡쳐(자막)

이를 부당하게 이용하는 게이밍(gaming)이나 경쟁자가 영업비밀(trade secret)에 무임승차하는 문제가 발생할 수 있다는 점을 대표적 근거로 제시한다. 따라서 Ignacio Cofone 교수는 알고리즘의 공개여부 및 공개정도[2]를 결정함에 있어서는 공개로 인한 사회적 비용(게이밍과 영업비밀의 무임승차 문제)과 공개로 인한 사회적 효용(컴플라이언스, 오류와 편향의 개선, 이용자의 권리보장)을 형량해야 한다고 주장하였다.

2 Ignacio Cofone 교수는 알고리즘의 공개도 다양한 층위가 있을 수 있음을 지적한다. (1) 학습 데이터셋, (2) 학습 데이터셋의 출처, (3) 코드, (4) 모델, (5) 피처 및 레이블, (6) 피처 및 레이블의 비중, (7) 아웃풋 값의 종류, (8) 알고리즘의 최종적 목표 등 공개대상이 되는 정보는 다양하다.

민간사업자는 이러한 형량요소를 적절히 반영할 수 있는 주체인가? Ignacio Cofone 교수는 아니라고 답하였다. 민간사업자는 알고리즘 공개로 인한 컴플라이언스가 주는 사회적 가치나 알고리즘의 오류나 편향으로 인한 사회적 비용을 내부화(internalize)하지 않으며, 오히려 규제나 분쟁에 휘말리지 않기 위해 알고리즘을 숨길 인센티브가 있기 때문이다. 그렇다면 판사나 규제당국이 알고리즘 공개 여부를 결정하는 것이 보다 적절하다고 Ignacio Cofone 교수는 지적하였다.

Ignacio Cofone 교수는 구체적 사안에서 알고리즘의 공개여부를 결정할 때 크게 3가지 요소를 검토해야 한다고 제시하였다. 첫째, 알고리즘을 공개할 경우 실제 게이밍, 영업비밀 유출 등 사회적 비용이 발생하는지 여부이다. 코드 전체를 공개할 경우 경쟁사업자로의 영업비밀 유출이 우려될 수 있으나 피처를 공개하는 것은 그럴 우려가 적다. 이용자가 쉽게 바꿀 수 없는 피처[3]가 자동화된 결정에서 중요한 역할을 수행하는 경우 설령 이러한 사실이 공개되더라도 게이밍이 발생할 가능성은 높지 않다. 또한 설령 이용자가 자신의 피처를 바꾸어 알고리즘의 의사결정을 바꾸는 경우에도 실제로 긍정적인 변화를 만들어내 알고리즘의 결정을 바꾼 것이라면 이는 게이밍의 문제가 아니다.[4]

둘째, 알고리즘 공개로 인한 예측성능 저하 등 손실이 예상되는지 여부의 문제이다. 알고리즘의 정확도가 높아 자동화된 의사결정이 잘 작동하고 있다면 이를 공개해 게이밍의 여지가 발생하는 것은 사회적 해악의 우려가 크다. 반면 알고리즘의 정확도가 낮은 상황이라면 게이밍으로 인한 손실은 상대적으로 적을 것으로 예상되며 오히려 알고리즘 공개로 오

3 가령 개인의 키, 주소 등이 대표적이다.

4 이는 앞서 언급한 컴플라이언스의 문제다. 게이밍이 해로운 경우는 실제 객관적인 결과값(가령 개인신용도)은 개선되지 않았음에도 알고리즘의 허점을 이용해서 마치 결과값이 향상된 것처럼 보이게 만드는 경우이다.

류 수정 효과를 기대할 수도 있다. 이러한 부정확성의 개선은 분배적 불평등의 개선을 위해서도 중요한 일이다.

셋째, 알고리즘을 설계하는 자와 사회적 인센티브가 같은 방향으로 작용하는지를 검토해야 한다. 가령 가석방 대상자의 재범위험률을 판단하는 알고리즘의 경우 알고리즘 설계자는 부정오류(false negative)[5]를 최소화하고자 할 가능성이 높으나, 사회적 선호는 헌법원리에 입각해 긍정오류(false positive)[6]를 최소화하는 방향의 설계를 요구한다. Ignacio Cofone 교수는 알고리즘 설계자와 사회적 선호가 일치한다면 알고리즘을 공개할 효용이 적을 것이나 만일 둘의 인센티브가 반대방향으로 작용하는 재범위험률 판단 알고리즘과 같은 경우 알고리즘 공개가 사회적 효용을 높일 수 있다고 지적하였다.

Ignacio Cofone 교수는 알고리즘 공개가 사회적 효용을 증진시키는 경우가 많이 있으며 어떠한 경우에 알고리즘이 공개되어야 하는지에 대한 판단기준을 제시한 것이 이번 연구의 성과라고 하였다. 그러면서 게이밍은 생각보다 현실에서 작동하기 어렵고 설령 게이밍이 발생하더라도 알고리즘의 비공개에 따른 비용이 더 클 수 있으므로, 알고리즘이 비공개인 상황이 디폴트값(default)이 되어서는 안 된다고 강조하면서 세미나를 마무리하였다.

5 실제로는 참(True)인 것이 거짓(Negative)이라고 잘못 판단되는 오류. 실제로는 재범위험률이 높은 사람이지만 가석방이 되는 경우가 이에 해당한다. 자동화된 의사결정에 의해 석방된 사람이 범죄를 일으킬 경우 알고리즘의 신뢰성에 문제가 제기될 수 있으나 석방되지 않은 사람은 범죄를 일으킬 수 없으므로 알고리즘 설계자는 재범위험률이 높은 석방자를 최소화할 인센티브가 있다.

6 반대로 실제로는 거짓(Negative)인 것이 참(True)이라고 잘못 판단되는 오류. 실제로는 재범위험률이 낮은 사람이지만 가석방이 불허되는 경우로써 COMPAS 사건의 경우 흑인의 긍정오류 비율이 백인보다 높게 나타나 문제되었다.

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

Seoul National University
AI Policy Initiative

인공지능 정책 국제 컨퍼런스

SNU AI POLICY INITIATIVE