

## 2019 한국인터넷거버넌스포럼(KrIGF) 워크숍 보고서

작성자 : 최은창(프린터넷프로젝트)

세션명	튜토리얼2. 인공지능의 활용과 거버넌스			
일시	2019.7.5.(금) 15:15~16:45		장소	세종대학교 컨벤션센터
참석자	사회		발제	최은창
	패널			
플로어	약 43명 참여 (국책 연구소, 연구자, 포털 사이트, IT 업계 등이 참석)			

제안내용	<p>AI 활용이 확대되면서 자동화된 판단의 경제적 효과가 경쟁력의 확보가 주목을 받고 있다. AI는 산업적 활용뿐만 아니라 국가 경쟁력, 안보를 강화시키려는 도구가 되고 있다. AI 시스템은 콘텐츠 큐레이션, 상품 추천, 채용 과정의 스크리닝, 인터뷰, 전자상거래의 개별화된 가격 조정, 질병의 진단, 예방적 방범활동, 대량의 가짜뉴스를 자동으로 생산하는 봇(bots) 등은 선거와 정치과정에도 영향을 미친다.</p> <p>그러나 AI의 유용성, 혁신, 비즈니스 기회에는 알고리즘 편향, 로봇의 결함, 오작동 등 위험과 사회적 비용이 수반된다. 오류나 부적절한 판단에 대한 우려도 커지고 있다. AI의 판단은 단순한 노동을 대체하는 것뿐만 아니라 취업 지원자의 인터뷰, 범죄자 양형, 사회적 자원의 배분에도 사용되기 시작했다. AI 알고리즘들은 축적된 데이터를 학습하므로 인종·성별·종교에 따라 편향된 판단을 내릴 수 있다.</p> <p>그렇지만 AI는 사회에 통용되는 상식, 가치체계, 윤리, 형평성에 무지하므로 개인들의 삶에 악영향을 주는 잘못된 판단을 내릴 수 있다. 이러한 가치체계를 설계하여 짜넣는 작업 자체가 매우 어려운 과제이다. 그러나 빅데이터와 컴퓨터 알고리즘으로 얻어지는 유용성에 대한 맹신이 강하다면 그 편향이나 오류 등의 문제는 무시된다.</p> <p>AI 알고리즘 또는 자동화 시스템이 점차 사회적 제도 속으로 진입하여 결정을 내리는 영역이 넓어지고 있으므로 알고리즘 설계, 데이터 세트에 숨어있는 편향은 사회적 약자의 부당한 차별이나 배제를 확대를 키울 수 있는 위험성이 있다. 그러므로 AI 개발에 있어서 윤리 및 거버넌스 논의도 함께 병행될 필요가 있다. 시민사회, 기업, 이용자, 정부 등 다중이해관계자들이 AI 윤리 정책과 거버넌스의 공론화가 필요하다.</p>
	<p>-AI 활용의 유용사례: 빅데이터를 활용한 분석은 지구촌 빈곤 지역의 상황을 파악하고, 저개발국가의 공중보건 활동, 지속가능한 발전목표에도 사용되고 있음</p> <p>① '유익한 인공지능 컨퍼런스(Beneficial AI 2017 conference)'는 아실로마 원칙을 발표하여 AI를 활용하는데 필요한 윤리적 전제조건을 논의하였음</p> <p>② 국제정보통신연합(ITU)이 2016년부터 시작한 AI for Good 은 기술계, 학계, 정부간 기구 등의 네트워크를 연계하여 '지속가능한 발전목표'(SDGs)를 위해 활용하도록 이른바 '공유재로서 인공지능(AI for Commons)을 제안.</p> <p>③ 유엔 글로벌 펄스(Global Pulse) 우간다 지부는 빅데이터 분석을 통해 난민의 유입, 재난 발생 파악, 공중보건 위기에 대한 대처에 활용하고 있음</p> <p>④ 말라리아 진단, 기후변화 파악과 대응, 난민 캠프의 이재민들에게 일자리를 찾아주는 작업에 사용될 수 있음(Stanford, ETH의 연구로 개발된 알고리즘)</p> <p>-AI 활용의 해악사례: ① 해악적 AI사용 사례: 디지털 안보(머신러닝에 의한 해킹, 적대적 머신러닝에 의한 인식오류) 물리적 안보(테러리스트의 드론 공격) 정치적 안보(개인정보를 파괴하는 프로파일링, 감시, 자동화된 허위정보 캠페인)</p> <p>② 자동화된 전투무기의 개발 : 구글의 메이븐(Maven) 프로젝트에 대한 반발.</p> <p>③ 얼굴인식 소프트웨어 남용: 중국의 Skynet 사례, 이용자들의 명시적 동의 없는</p>

논의세부내용	<p>페이스북의 사진 속 얼굴 인식, 아마존 리코그니션(Amazon Rekognition)</p> <p>④ 범죄자에 대한 가석방 결정 시 사용된 알고리즘의 편향성(COMPAS Recidivism Algorithm)에 반발한 Loomis v. Wisconsin 위스콘신주 소송사건,</p> <p>⑤ 예방적 방범에 사용되는 프레드폴 알고리즘의 잘못된 데이터.</p> <p>-블랙박스 문제: AI에 거는 기대와 신뢰도에 대한 믿음이 지나치게 크기 때문에 오류, 부적절한 데이터셋, 편향성, 오작동, 남용 등의 현실적 문제가 부각되지 않고 있음 AI 설계에 수많은 파라미터와 데이터가 사용되지만 그 작동방식은 불투명하다는 문제. 1) 알고리즘 설계의 복잡성 2) 관련성이 없는 데이터, 잘못된 데이터를 입력하는 관행은 판단의 정확성에 문제를 일으킴. AI 시스템을 설계한 주체도 그 작동방식을 파악하지 못하지만 효용성만이 부각되는 역설적 상황 발생</p> <p>-AI 알고리즘의 책무성, 공정성, 투명성, 포섭이 필요한 이유는? : 인간의 안전, 공정성, 정확성 편향성이 보장되도록 AI 시스템을 설계할 필요. 알고리즘의 작은 오류가 반복되면 사회 전체의 해악으로 확대될 우려가 있음</p> <p>-AI 윤리 원칙을 위한 국제적 움직임: ① OECD는 2019년에 회원국 정부들의 규제가 될 수 있는 AI 원칙을 발표, ② EU연합은 윤리적 원칙들을 규범화하여 AI 기술 분야의 선진국들을 견제하려는 의도를 보이고 있음. AI윤리는 이용자 안전이라는 목적인 아니라 국제적 AI 경쟁에서 상대국의 견제라는 성격도 있음. ③ Partnership on AI는 자율규제를 위한 기업들의 연합체로 AI의 안전성, AI의 공정성·투명성·책임성, 인간과 AI의 협력AI와 사회적 영향과 공익 등을 연구함.</p> <p>-데이터와 AI 윤리 원칙과의 관계: 데이터는 AI 산업을 위한 원유이며, 알고리즘은 엔진에 해당되므로 데이터의 확보는 매우 중요. 중국과 같이 개인정보 보호가 특별히 없는 국가라면 데이터의 공급이 쉬워짐. AI 윤리를 위해서는 1) 데이터 거버넌스와 2) 데이터 프로비던스의 확보 3) 설명가능성이 필요.</p> <p>① 데이터 거버넌스: 데이터 주체의 동의를 받고 데이터를 사용하고 '신원'의 익명화, 가명화 절차가 반드시 필요. GDPR '적정성 기준'의 충족이 가능한가?</p> <p>② 데이터 프로비던스: 정확한 판단을 위해 신뢰도 높은 데이터 확보가 필수적. 데이터의 근원 정보를 기록하도록 설계된 구조화된 메타 데이터가 투명해야 문제가 생겨도 그 원인을 찾거나 수정을 통해 바로 잡을 수 있음.</p> <p>③ 설명가능한 AI: 알고리즘이 어떻게 그 판단에 이르게 되었는가를 이해하도록 설명할 수 있어야 함. 그러나 기계언어와 인간의 인지능력 사이의 격차는 상당</p>
	<p>- AI는 인터넷 거버넌스 포럼에서 가장 주목받는 주제로 부상하고 있음</p> <p>- 해외 정부, 기업(private initiatives), ITU 등 국제기구, 세계경제포럼(WEF) OECD, 연구기관, 기업들은 관점이 다른 이유는? : 각자의 이해관계가 다르기 때문. AI 윤리가 자율규제로 충분하다는 그룹과 강제법규가 마련되어야 한다는 그룹으로 양분되고 있음. 대립적 양상은 상당기간 지속될 것으로 보임.</p> <p>- 국제사회에서 논의되는 인공지능 윤리(ethics)와 규제(regulations)에 대한 흐름을 살펴보면 그 시사점을 국내 4차 산업혁명 정책 방향에 참조할 필요</p> <p>- AI 윤리는 누구를 위한 것일까에 대한 고민이 국내에서도 병행되어야 한다.</p> <p>- AI 규제정책은 국가마다의 데이터 관련 규제, 활용영역, 알고리즘 개발 등 기술적 여건, 기술 수준 등 특수성을 감안하여야 하며 신중한 접근이 필요</p> <p>- AI 윤리 정책과 거버넌스는 국제 흐름을 주시하고 인간 중심AI(Human-Centered AI)을 위해 준수할 조건들에 대해 각 분야 이해관계자들이 적극적으로 자신의 이해관계와 주장을 발언할 수 있는 기회를 제공할 필요가 있음.</p>