

한국인공지능법학회 2019년 상반기 정기세미나

인공지능의 윤리적 개발 동향과 입법대응 과제



환영사

안녕하십니까?

국회 4차산업혁명특별위원회 바른미래당 간사,
국회 4차산업혁명포럼 공동대표 신용현 의원입니다.



먼저 「인공지능의 윤리적 개발 동향과 입법대응과제 세미나」를 함께 주최하신
이상용 한국인공지능법학회 회장님과 이수영 KAIST 인공지능연구소장님께 감사드립니다.

올해 초 미국 라스베이거스에서 열린 CES에 방문했을 때, 눈길을 끄는 것은
단연 인공지능 로봇이었습니다. 헬스케어, 가사노동 로봇은 물론, 자율주행차에 적용된
인공지능 기술은 이제 상용화를 준비하는 단계에 와 있습니다.

미래학자 레이 커즈와일은 2045년에는 인공지능이 인간의 능력과 지성을 따라잡을
것으로 예측했습니다. 인공지능에 대한 전 세계의 관심과 기술개발이 가속화 되는 만큼,
나타날 수 있는 사회적 변화는 상상할 수 없을 만큼 다양할 것입니다.

이제는 구체적인 인공지능 기술개발과 함께, 발생할 수 있는 윤리적 문제를
장기적 관점에서 고민해야할 때입니다.

지금으로부터 77년 전, SF 소설가 아이작 아시모프는 단편소설 <런어라운드>에서
인공지능 개발의 윤리적 기준이 될 수 있는 ‘로봇의 세 가지 원칙’을 서술했습니다.
그는 인공지능 개발의 원칙이 근본적으로 인간과 인공지능 사이에 갈등을 최소화해야한다고 했습니다.
인공지능 기술이 고도화 되는 만큼, 77년이 지난 지금은 인공지능에 대한 더 많은 윤리적 고민이
필요합니다. 또한 그 고민이 인공지능 신기술과 서비스에 대한 규제가 아닌, 관련 산업발전과
인간의 삶이 조화를 이루는 방향으로 진행되어야할 것입니다.

환영사

현재까지는 인공지능에 대한 윤리적 논의가 실업문제나 오남용에 대한 부작용 등 인간 사회 전반에 대한 거시적 담론에 머물고 있습니다.

하지만, 인공지능의 개발 단계를 고려해 데이터의 수집부터, 알고리즘의 설계 그리고 상용화까지, 개발의 다양한 단계에서 발생할 수 있는 다층적 윤리 문제를 진단하고, 구체적인 제도마련으로 대응해야 합니다.

그런 점에서 오늘 세미나는 다양한 인공지능의 윤리적 개발 동향을 파악하고, 그에 필요한 구체적인 입법 과제를 짚어본다는 점에서 매우 뜻깊은 시간이라고 할 수 있습니다.

오늘 이 자리에 과학기술계와 법학, 윤리학을 비롯한 다양한 학계, 산업계 그리고 정부의 인공지능 분야 전문가들이 함께한 만큼, 우리나라 인공지능의 윤리적 개발의 지혜가 도출되길 기대합니다.

국회에서는 오늘 발제와 토론에서 제안해주시는 내용에 귀 기울여, 필요한 제도를 만들고, 우리나라가 인공지능 강국이 될 수 있도록 최선을 다하겠습니다.

다시 한 번 오늘 세미나에 기조발제를 맡아주신 Antony Cook 마이크로소프트 아시아지역 법무 총괄책임자와 발제 및 토론을 맡아주신 모든 분들께 감사드리며, 참석해주신 여러분의 건강과 행복을 기원합니다.

감사합니다.

환영사

안녕하세요.

국회 4차산업혁명 포럼 공동대표, 더불어민주당 국회의원 박경미입니다.



더불어민주당
박경미 국회의원

먼저 「인공지능의 윤리적 개발 동향과 입법대응 과제」 세미나 개최를 축하드립니다.
아울러 이 자리를 마련해 주신 신용현 의원님과 이상용 한국인공지능법학회 회장님,
이수영 KAIST 인공지능연구소장님께 진심으로 감사드립니다.

2016년 전 세계를 충격에 빠뜨린 구글의 인공지능 알파고와 이세돌 9단의 대국 이후 인공지능 기술은
발전예 발전예 거듭하고 있습니다. 최근에는 제조업뿐 만 아니라 의료, 법률 등 전문 서비스업에까지
활용 영역을 확장하고 있으며, 앞으로 다양한 산업분야에 본격적으로 적용되어 성과를 창출할 것으로 기대합니다.

실제로 세계경제포럼(WEF)의 Global Risk Report (2017)에서 발표한 ‘유망 기술 파급효과에 대한 인식조사’에
따르면 인공지능은 로봇, 바이오 기술과 함께 파급효과가 가장 큰 기술 중 하나로 분류된 바 있습니다.
그러나 인공지능의 긍정적 영향에도 불구하고 오남용, 일자리 대체, 정보 양극화와 같은 기술적·사회적 문제가
발생하면서 인공지능의 폐해를 우려하는 목소리가 커지고 있습니다. 때문에 세계 주요국들은 한 발 앞서
인공지능이 불러올 위험에 대비하고, 인류의 삶에 기여할 수 있는 이른 바 ‘착한 인공지능’을 만들기 위한
법적·제도적 대응 방안을 논의하고 있는 것입니다. 이제 우리도 인공지능의 발전 속도에 맞춰 사회 변화에
선제적으로 대응해야 합니다. 이를 위해서는 기존의 기술 발전 위주의 패러다임에서 벗어나
기술과 제도의 적절한 균형이 필요합니다. 기술 발전을 저해하지 않으면서도 기술의 부작용을
최소화 할 수 있는 제도 마련이 시급합니다.

오늘 이 자리에 인공지능 분야의 학계, 산업계 전문가들이 모두 모이신 만큼 인공지능의
밝은 미래를 위한 구체적인 해법을 제시해 주실 것이라 생각합니다.
오늘 발표와 토론을 통해 제안해주신 내용들은 모두 귀담아듣고 국회가 할 수 있는 최선의 노력을 다하겠습니다.
다시 한 번 오늘 세미나에 참석해주신 모든 분들께 감사드리며 함께해 주신
모든 분들의 건강과 행복을 기원합니다.

감사합니다.

환영사

환영사

안녕하십니까.

국회 4차산업혁명포럼 공동대표 송희경 의원입니다.



먼저 「인공지능의 윤리적 개발 동향과 입법대응 과제 세미나」 행사 개최를
진심으로 축하드립니다.

행사를 주최해주신 신용현 의원님, 이상용 한국인공지능법학회 회장님, 이수영 KAIST 인공지능연구소장님께
감사드립니다. 아울러 기초발제를 맡아주신 마이크로소프트 아시아지역 총괄책임자
Antony Cook 변호사님을 비롯하여 발제와 토론에 참여해주신 전문가 여러분들께도 감사 인사를 드립니다.
또한 세미나 참석을 위해 귀한 발걸음 해주신 모든 참석자 여러분들 환영합니다.

불과 3년 전, 알파고 쇼크로 전 세계인을 놀라게 한 AI는 바둑을 넘어 스타크래프트 게임, 주식투자에서도
인간과 겨루는 수준까지 끝 모르는 진화를 거듭하고 있습니다. 최근 조사에 따르면, 인공지능이 자산운용을
자문하는 펀드, 일명 ‘로보펀드’가 사람보다 높은 수익률을 거둔 것으로 나타났습니다.

핀란드는 AI가 신생아의 움직임을 분석하여 뇌성마비를 진단합니다. 정확도는 95%에 수렴합니다.

이렇듯 4차 산업혁명의 핵심동력인 AI는 충격과 함께 우리에게 다가왔고 4차 산업혁명의 기술 및 제도 혁명을 이끌고
있습니다. 기계 학습, 문제해결, 패턴 인식 등 인간지능과 연결된 인지 문제를 해결하는 컴퓨팅 시스템을 의미하는
인공지능은 생산성 향상 및 부가가치를 증대시키는 매개체로서 정부·산업계·학계의 관심을 집중시키고 있습니다.
인공지능과 로봇이 우리 사회에 가져다 줄 영향에 대처할 수 있는 일관성 있는 장기적 전략이 필요한 시점입니다.

신뢰를 기반으로, 인류행복중심의 발전을 이끄는 것이 4차산업혁명 성공의 완성입니다. ‘선의’에 의해 만들어진
기술이 인간의 ‘보다 나은 삶’을 제공할 것이라는 믿음을 줘야 합니다. 이를 위해서는 기술과 정책의 밸런스가
무엇보다 절실합니다. 인공지능 및 로봇의 오용 및 오작동에 따른 사회·윤리적 문제들을 미리 예방하는 정책 및
방안들에 대한 고민도 함께 되어야 하는 이유입니다. 향후 인공지능의 발전과 관련한 입법정책적 쟁점은
예측 불가능한 위험에 얼마만큼 신속하고 정확하게 대응할 수 있는지에 달려 있습니다.

과거 입법 패러다임과는 다른 위험관리적 관점이 내재된 새로운 입법 정책 추진이 필요하다고 생각합니다.

때문에, 오늘 「인공지능의 윤리적 개발 동향과 입법대응 과제 세미나」를 통해 AI가 직면한
윤리적인 과제를 해결하고, 새로운 패러다임 전환을 하는데 좋은 밑거름이 될 것이라고 믿습니다.

저 또한 여러분의 고견을 받들어 정책으로 반영될 수 있도록 최선을 다하겠습니다.

다시 한 번 오늘 행사에 참석해주신 모든 분들께 감사드리며 오늘 이 자리에 함께 한 모든 분들의
건강과 행복을 기원합니다.

감사합니다.

축사

안녕하십니까.

바른미래당 대표 손학규입니다.



먼저 「한국인공지능법학회 2019년 상반기 세미나」 개최를 진심으로 축하드립니다.

또한 오늘 세미나 개최를 위해 애써주신 신용현 의원님과

한국인공지능법학회 이상용 회장님, KAIST 인공지능연구소 이수영 소장님 및

관계자 여러분께 감사의 말씀을 드립니다.

특히 오늘 세미나에서 주제발표로 귀한 시간을 내주신 발제자 여러분과 패널로 참가해주신

각 분야의 전문가 여러분께도 고마움의 뜻을 전합니다.

기술의 발달에 따라 다양한 분야에서 인공지능의 역할이 확대되고 또 강조되고 있습니다.

그와 동시에 인공지능이 인간의 법적·사회적 규범 및 윤리적·도덕적 가치를 내면화할 수 있는지에 대한

관심도 함께 증대되고 있습니다. 인공지능이 미래의 도덕적·사회적 문제의 시발점이 될 수도 있다는

문제의식이 큰 공감을 얻고 있는 것이 현실입니다.

인공지능 알고리즘은 학습을 통해 혐오, 차별 등 인류사의 부정적 측면을 비판 없이 수용할 수 있는

위험성을 가지고 있습니다. 이런 경우 인공지능이 법적·윤리적 문제를 일으킬 수 있다는 우려는 기우에

그치지 않을 것입니다.

따라서 우리는 인공지능이 일으킬 수 있는 윤리적 문제에 대한 대비를 기술 개발과 함께 준비해야 할 것입니다.

이것은 인공지능 기술의 발전을 가로막는 걸림돌이 아니라, 더 나은 발전을 위한 과정의 하나입니다.

작년 11월, 바른미래당은 시대에 발맞춰 4차산업혁명위원회를 발족한 바 있습니다.

또한, 당 차원에서도 인공지능의 윤리적 문제해결과 입법대응을 위한 방안을 마련하기 위해 노력하고 있습니다.

그러한 의미에서 오늘 이 자리에서 논의되는 여러 아젠다가 앞으로의 인공지능 기술 발전에

큰 도움이 되리라 생각합니다.

바른미래당도 소중한 의견을 경청하여 과학기술 정책에 반영될 수 있도록 최선을 다하겠습니다.

다시 한 번 오늘 이 자리를 빛내기 위해 참석해 주신 모든 분들께 감사의 말씀을 드립니다.

감사합니다.

환영사·축사

축사

반갑습니다.
바른미래당 원내대표 김관영입니다.



「인공지능 윤리적 개발 동향과 입법대응 과제 세미나」 개최를 환영합니다.
뜻깊은 세미나를 마련해주신 신용현 의원님, 한국인공지능법학회 이상용 회장님,
KAIST 인공지능연구소 이수영 소장님을 비롯하여 발제와 토론에 임해주신 전문가 여러분의
노고에 깊이 감사드립니다.

인공지능은 4차 산업혁명의 핵심 기술 가운데 하나로 손꼽힙니다. 이미 수많은 가전기와 기계 등에
인공지능 기능이 탑재되어 편리함을 가져왔습니다. 인류문명이 발달할수록 인공지능은 더욱 중요한
역할을 하고, 우리는 인공지능에 많은 의존을 할 것으로 예상됩니다.
그러나 반대로 딥페이크, 킬러로봇, 사이버해킹 등의 다양한 사회적 문제가 부각되고 있습니다.
지난 2016년 마이크로소프트사가 개발한 인공지능 챗봇 테이가 나치송배, 성차별과 백인우월주의 등의
발언을 해 하루도 안돼서 서비스를 중단한 촌극이 빚어진 일도 기억하실 겁니다.
영국 물리학자 스티븐 호킹이 ‘인류가 대처하는 방법을 익히지 못하면 AI 기술은 인류 문명사에서
최악의 사건이 될 수 있다.’고 우려한 이유이기도 합니다.
인공지능 기술이 인류에게 독이 될지, 약이 될지는 전적으로 우리들의 몫인 것입니다.
그렇기 때문에 기술 발전에 대한 부정적 인식을 타파하고, 밝은 미래를 위해 인공지능에 관한 기술뿐만 아니라
인문학·사회과학 등의 많은 연구가 필요합니다. 아이가 도덕교육을 충분히 받아 인성을 갖춘 사람으로
자랄 수 있듯, 인공지능 역시 인류의 보편적 윤리기준을 통해 개발되어야만 하는 것입니다.

특히 인공지능의 발전 속도를 예측할 수 없는 만큼 선제적인 윤리기준 마련과 법제화가 필요한 시점입니다.
이러한 때 여러 분야의 전문가들을 모시고 인공지능 윤리적 개발 동향과 그 대응책을 모색하는
세미나가 마련돼 매우 시의적절하다고 생각합니다. 이 자리에서 인류와 함께 성장할 인공지능의 미래를 꼼꼼하게
점검하고, 실효성 있는 입법방안들이 충분히 논의되길 기대합니다.
인공지능 기술도 현재의 우리들의 모습에 달려 있다고 생각합니다. 바른미래당은 인공지능 관련
윤리적 과제와 쟁점, 정책들을 살피겠습니다. 또한 기술의 발달뿐만 아니라 윤리와 인간성을
먼저 존중하는데 앞장 설 것을 약속드립니다.

감사합니다.

축사

안녕하십니까? 반갑습니다.

국회 과학기술정보방송통신위원회 위원장 노웅래입니다.



오늘 「한국인공지능법학회 2019년 상반기 세미나」의 개최를
진심으로 축하드립니다. 오늘 이 세미나가 있기까지 애써주신 신용현 의원님과
한국인공지능법학회 이상용 회장님, KAIST 인공지능연구소 이수영 소장님 이하
관계자 여러분께 감사드립니다.

인공지능의 발전은 그 자체로 반갑고, 환영해야 하지만, 그 이면에는 우리가 다뤄야 할 사회적 규범이나
윤리, 도덕적 측면에서의 갈등도 있습니다.

작년 6월에 개봉한 사이코패스 AI ‘노면’이라는 영화는 앞으로 다가올 인공지능 시대에
우리가 다친 현실을 잘 반영했습니다. AI인 노면은 죽음과 살인 등 부정적인 이미지와 동영상을
집중적으로 학습하며 잘못된 데이터만 받아들인 결과, 획득한 정보를 가지고 특정 사물에
주관적인 판단을 내릴 수 있었고, 자살, 교통사고, 총살, 살인과 같은 끔찍한 표현만을 내놓았습니다.

이와 같은 사례로 보아 우리는 인공지능의 기술개발과 속도에만 집중하기보다 인공지능이
보급·확대되었을 때 사회가 가질 부작용을 함께 고려하여야 합니다. 그런 점에서 오늘 세미나가
시의적절하다고 할 수 있습니다. 인공지능 윤리에 관한 관심이 미미한 지금, 여기에 모인 각 분야의
전문가 여러분의 지혜로 더 나은 인공지능 시대를 열 수 있는 초석을 마련할 수 있다고 봅니다.

저 역시 국회 과학기술정보방송통신위원회 위원장으로서 대한민국이 4차산업혁명시대에서
준비해야 할 것들을 고민하고, 대비할 수 있는 정책적 대안을 마련하는데 노력할 것을 말씀드립니다.
다시 한번 오늘 자리를 마련해주신 관계자 여러분께 감사드립니다.

축사

축 사

안녕하십니까?
국회의원 정병국입니다.



먼저 「인공지능의 윤리적 개발동향과 입법대응과제」 세미나 개최를 진심으로 축하합니다. 무엇보다 저희 바른미래당의 보물이자, 과학기술인 출신 신용현 의원님께서 오늘 토론회를 공동으로 주최해주신 만큼 오늘 세미나를 통해 생산적인 논의가 이뤄질 것이라 믿습니다. 또한 한국인공지능법학회와 KAIST인공지능연구소 그리고 기초발제를 맡아주시는 마이크로소프트 아시아지역총괄책임자 Antony Cook변호사님 등 발제와 토론에 힘써주시는 모든 분께도 감사드립니다.

세계적인 미래학자 엘빈 토플러가 말했습니다. “기술문명의 발달속도는 100마일로 달리고 있는데, 정치는 3마일, 법은 1마일로 가고 있다” 우리나라의 정치와 법의 현실이 정확하게 반영된 말입니다. 4차 산업혁명 시대에 발맞춰 대한민국이 발전과 번영의 길을 걷기 위해서는, 핵심 기술인 인공지능(AI), 사물인터넷(IoT), 빅데이터 등의 분야에 거미줄처럼 얽힌 규제를 혁파하는일이 선행되어야 할 것입니다.

이에 못지않게 중요한 것은 대한민국이 4차 산업혁명 시대의 ‘기술 강국’을 넘어 ‘윤리 강국’이 되어야 한다는 것입니다. 인공지능 등 신기술은 대한민국을 넘어 인류에 엄청난 부를 가져올 것이지만, 자칫 악용될 시 재앙으로 돌아올 수도 있는 양날의 검입니다. 현재 우리나라 인공지능 기술연구는 국제교류 및 경쟁을 통해 활발히 진행되고 있습니다만, 인공지능 윤리 규범에 대한 논의는 아직 그러한 수준에 이르지 못하는 실정입니다.

그런 의미에서 윤리문제에 민감한 분야인 인공지능 기술이 어떻게 하면 건강하게 진보해나갈 수 있을지 논의 하는 오늘 「인공지능의 윤리적 개발동향과 입법대응과제」 세미나는 의미가 남다르다 할 수 있습니다. 오늘 여러 전문가들이 모인 자리에서의 지적인 논의를 통해, 인공지능에 대한 법과 윤리적 규범을 확립하는 단초가 마련될 수 있도록 힘써주시길 것을 당부 드립니다.

저 또한 기술의 양적·질적 성장의 조화, 그리고 그로 인한 대한민국의 건강한 번영을 희망하는 한 명의 국민, 한 명의 국회의원으로서 모든 법적, 정책적 지원을 아끼지 않을 것을 약속드립니다. 다시 한 번 「인공지능의 윤리적 개발동향과 입법대응과제」세미나 개최를 축하드리며, 참석하신 모든 분들의 가정에 건강과 행복이 깃들길 바랍니다.

감사합니다.

축 사

안녕하십니까.

바른미래당 정책위의장 국회의원 권은희입니다.



인공지능의 윤리적 입법 대응과제 논의를 위한 「인공지능의 윤리적 개발동향과 입법대응과제」세미나 개최를 진심으로 축하합니다.

4차 산업혁명 시대를 맞이하여 우리는 디지털 기술이 기반된 초연결 사회로 진입하고 있습니다.

그중에서도 우리의 삶에 전반적으로 영향을 미칠 기술은 인공지능으로서, 금융·의료·자동차·제조업 등 여러 기존 산업을 혁명적으로 변화시킬 핵심 기술입니다.

각국에서는 4차 산업혁명의 주도권을 잡기 위한 치열한 경쟁이 진행되고 있으며, 인공지능 기술은 역사적으로 유례가 없는 속도로 발전하고 있습니다. 하지만, 인공지능의 발전속도에 비해서 기술의 잠재적인 위험과 그로 인해 발생할 수 있는 윤리적 문제에 대한 논의는 충분하지 않은 상황입니다.

인공지능 기술이 인간 윤리나 법적 제도와 충돌했을 경우에 책임 귀속 주체에 대한 문제가 사회적 쟁점으로 부상하고 있으며, 예방적 차원에서 윤리적 알고리즘 설계를 위한 개발자의 역할이 강조되고 있습니다.

현재 국제사회는 이러한 과제를 해결하고자 인공지능 윤리적 법규에 대한 논의를 시작했지만, 아직까지 우리나라의 인공지능 연구는 기술적인 부분에만 치우쳐 있는 실정입니다. 만약 우리나라에서 안전하고 윤리적인 인공지능 기술의 발전방향을 제시하고 보완제도를 만든다면, 세계시장에서 인공지능 기술을 선도할 수 있을 것이라고 생각합니다.

인공지능의 윤리성 분야는 높은 전문성과 규제역량을 필요로 하는 만큼 인공지능 관련 학계·산업계 등 다양한 전문가들이 함께 지혜를 모아야 합니다. 그런 의미에서 이번 세미나를 통해 인공지능의 도덕적 책무·윤리적 가치 판단에 대해 심도 깊은 토론이 이루어지고, 인공지능 윤리적 개발에 필요한 입법적 토대가 형성되기를 기대합니다.

국회 과학기술정보방송통신위원회 바른미래당 간사 신용현 의원님과 한국인공지능법학회, KAIST인공지능연구소의 공동 주최로 열린 이번 세미나가 우리나라 인공지능 기술발전에 따라 발생 가능한 윤리적 문제에 앞서 대응할 수 있도록 견인할 이정표가 되기를 바랍니다.

다시 한번 귀중한 발걸음을 해주신 모든 내빈들과 참석자 여러분들께 감사드리며, 발제와 토론에 힘써주시는 모든 분께도 감사의 말씀을 전합니다.

감사합니다.

축 사





Session I - 발표



기조발표

글로벌 기업의 관점에서 바라 보는
인공지능의 윤리적 과제

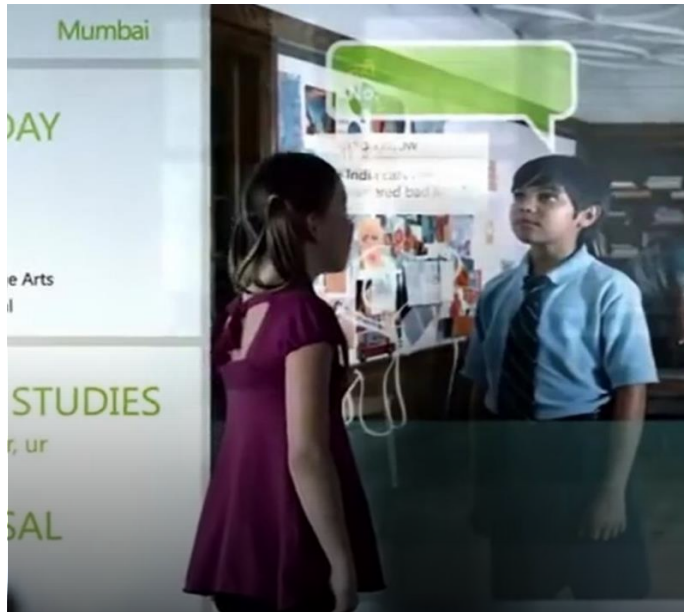
Ethical Principles, Policies and Laws for Responsible AI

Antony Cook

Regional General Manager and
Associate General Counsel
Microsoft Asia



The Promise of Artificial Intelligence



Education



Healthcare



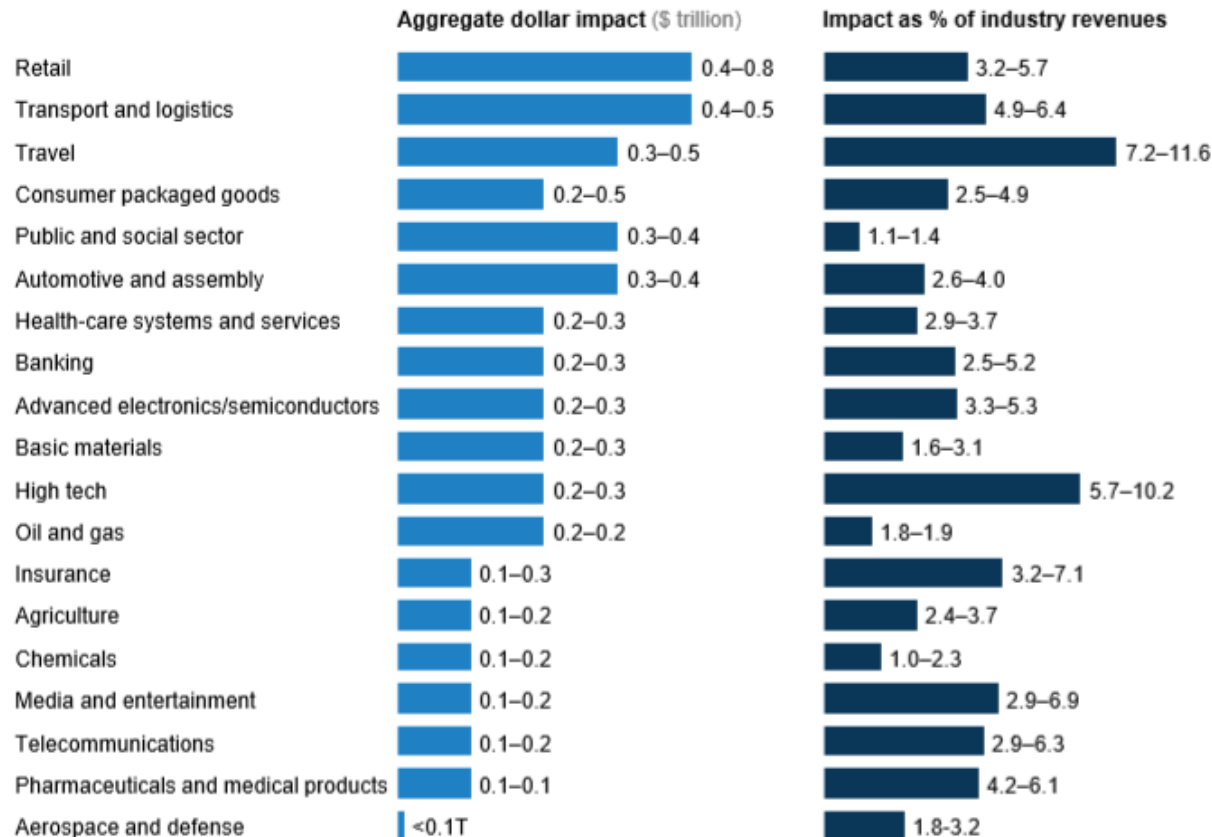
Transportation



Agriculture

The opportunity: transforming the global economy

The potential value of AI by sector



NOTE: Artificial Intelligence here includes neural networks only. Numbers may not sum due to rounding.

SOURCE: McKinsey Global Institute analysis

\$3.5 trillion to
\$5.8 trillion
potential annual
boost to GDP

A new driver of economic growth and jobs for Korea



Potential for additional
460 trillion KRW (\$410
billion) in GDP &
800,000 new jobs
by 2030 for Korea

Source: McKinsey & Company, 2016



4th Industrial Revolution
(including AI) will boost GDP
by 630 trillion KRW (\$560
billion) by 2030 for Korea

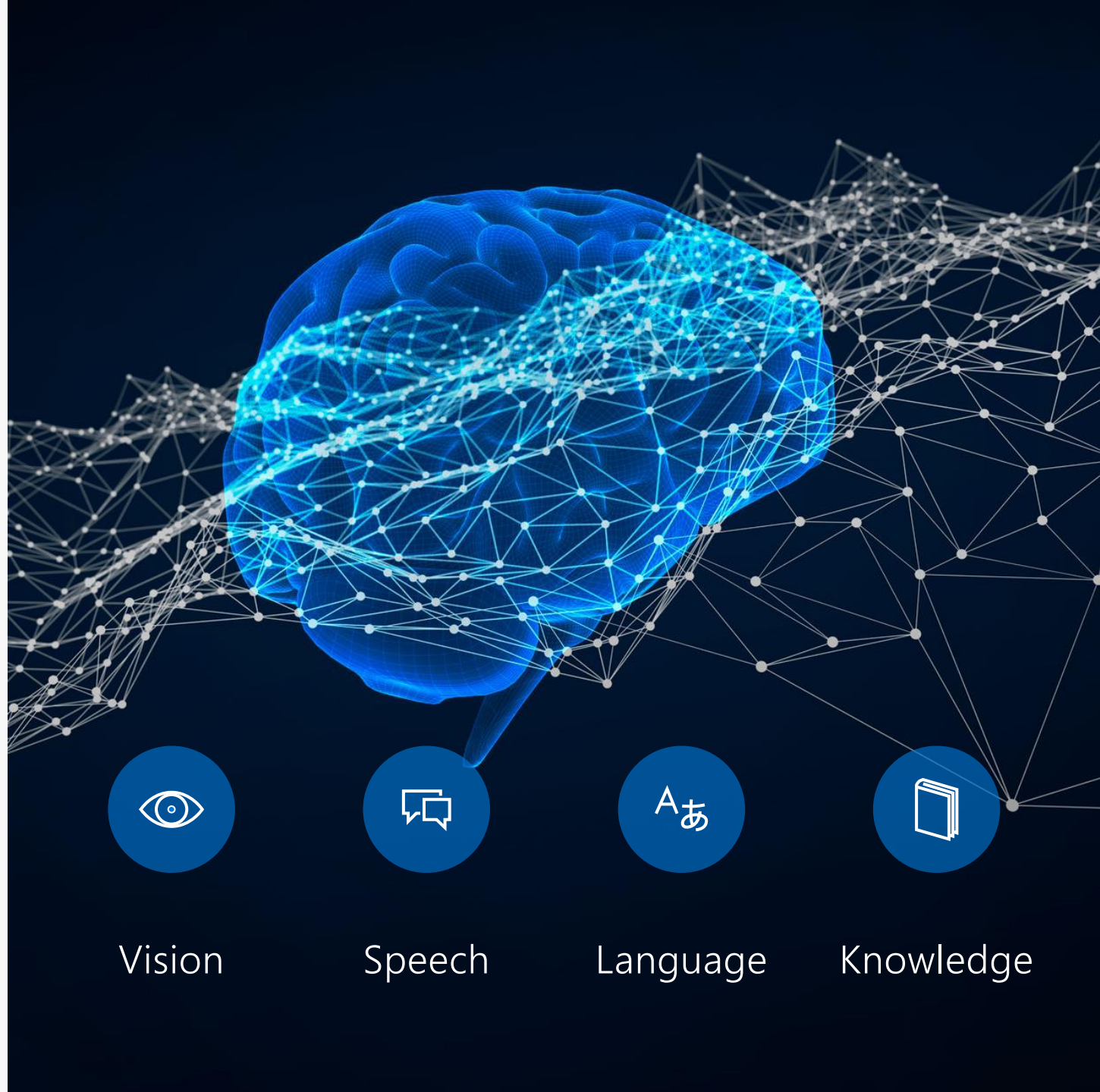
Source: Presidential Commission on the 4th Industrial Revolution,
2018

Setting national priorities on AI – what is important?

- Clear vision from the highest level for the use of AI, recognizing it is a key driver of future competitiveness
- Multi-stakeholder process to define ethical principles for use of AI, cutting across all industries
- Identify where existing laws & regulations apply – before looking to draft new ones
- Supporting data-driven innovation while ensuring user trust
- Build skills for the AI economy – while also managing shifting tasks



Microsoft's vision:
design AI to **augment**
human abilities



Vision

Speech

Language

Knowledge

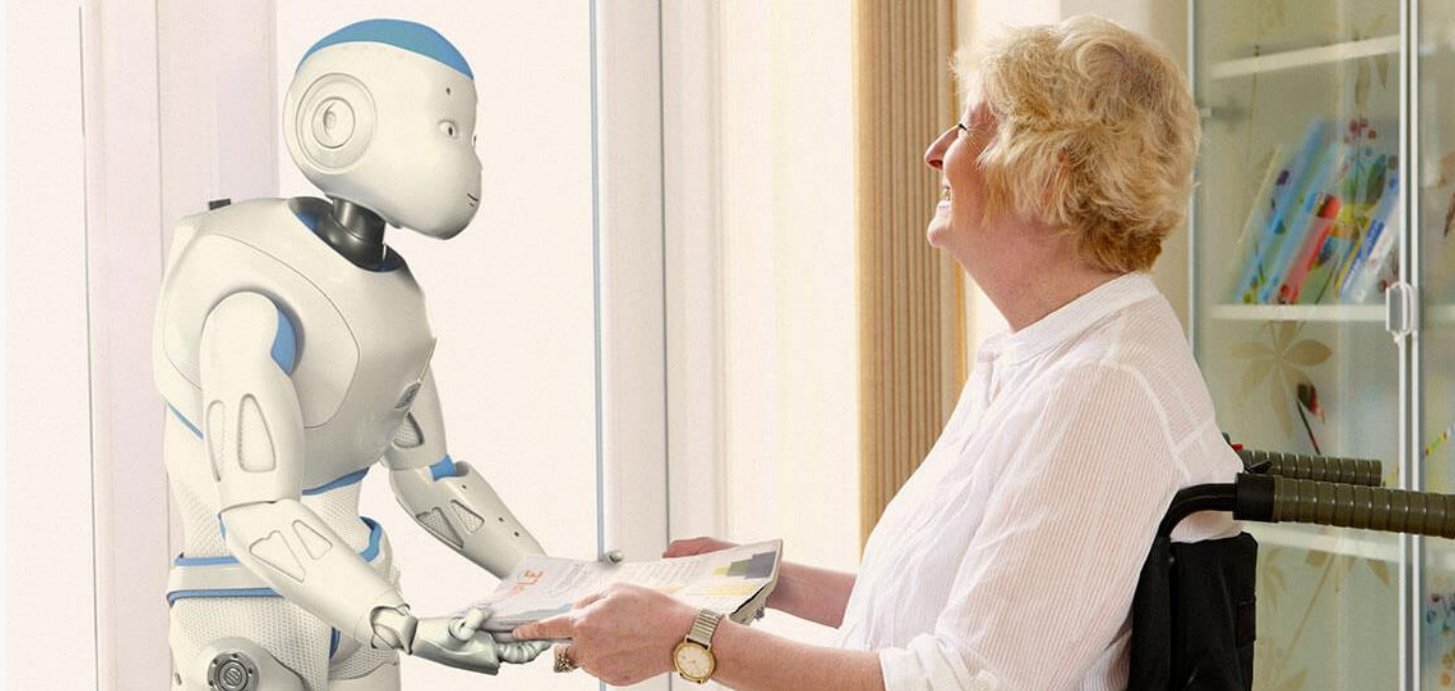
Microsoft's aim:
Make AI **available to
everyone**



We are pursuing AI to empower every person and every institution .. so that they can go on to solve the most pressing problems of our society and our economy.

– Satya Nadella

We must design AI systems to **earn trust**



But what about **AI ethics**?

Every powerful new technology in history has brought with it certain risks. What do we need to do to ensure that AI only helps humans and never harms us?

As computers behave
more like humans,
how will they impact
real people?



Trusted approach to AI



Universal principles

Ethical design



Open collaboration

Partnership on AI



Built on trust

Governance

Design AI to Earn Trust



Fairness



Reliability
& Safety



Privacy &
Security



Inclusiveness



Transparency



Accountability

Fairness

Understand how bias can be introduced and affect recommendations

Attract diverse pool of AI talent

Develop analytical techniques to detect and eliminate bias

Human review and domain expertise

The New York Times *Facial Recognition Is Accurate, if You're a White Guy*

By STEVE LOHR FEB. 9, 2018



Gender was misidentified in up to 1 percent of lighter-skinned males in a set of 385 photos.

Gender was misidentified in up to 12 percent of darker-skinned males in a set of 318 photos.



Gender was misidentified in up to 7 percent of lighter-skinned females in a set of 296 photos.

Gender was misidentified in 35 percent of darker-skinned females in a set of 271 photos.

Reliability & Safety

Evaluate [training data](#)

[Test](#) extensively (and enable a user feedback loop)

[Monitor](#) ongoing performance

Design for [unexpected](#) circumstances—including nefarious [attacks](#)

[Human](#) in the loop



Privacy & Security

Existing privacy laws (e.g. the [General Data Protection Regulation](#)) apply

Provide [transparency](#) about data collection and use, and good [controls](#) so people can make choices about their data

Design systems to [protect](#) against bad actors

Use [de-identification techniques](#) to promote both privacy and security



Inclusiveness

Inclusive design practices to address potential barriers that could unintentionally exclude people

Enhances opportunities for those with disabilities

Build trust through contextual interaction

EQ in addition to IQ



Transparency

People should **understand** how decisions were made

Provide **contextual explanations**

Make it easier to raise awareness of **potential bias**, errors and unintended outcomes



Accountability

People must be **accountable** for how their systems **operate**

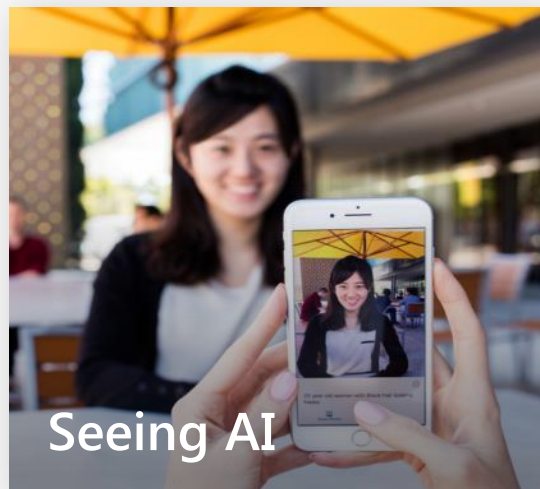
Norms should be observed during **system design** and in an **ongoing** manner

Role for **internal review boards**



Commitment to creating **AI for good**

Inclusion



Education



Healthcare



Earth



Ethical Framework Considerations

Enable broad deployment and continued innovation:

- Multi-stakeholder dialogues
- Practical principles for trustworthy and human-centered AI
- Sharing of best practices
- Funded research & development
- Data availability
- Skills training



Partnerships will be critical



Global dialogue among technology companies to develop AI practices



European dialogue on ethical use of AI

The gap: a cross-Asian partnership to share views and build common understanding on ethical use of AI



Thank you





Session I - 발표

인공지능 윤리의 국제적 맥락 및 주요 쟁점

인공지능 윤리의 국제적 맥락과 주요 쟁점



경인교육대학교 사회과교육과
심우민(조교수, 법학박사)
legislation21@ginue.ac.kr
<http://legislation.kr>
<http://civic-edu.kr>

목 차

1. 법규범적 대응 패러다임 전환

- (법)규범 패러다임 전환 논제
- 디지털 전환과 입법담론 과잉
- 응답적 법(responsive law)으로의 전환과정
- 윤리적 접근 시도

2. 각국의 윤리적 수준의 대응

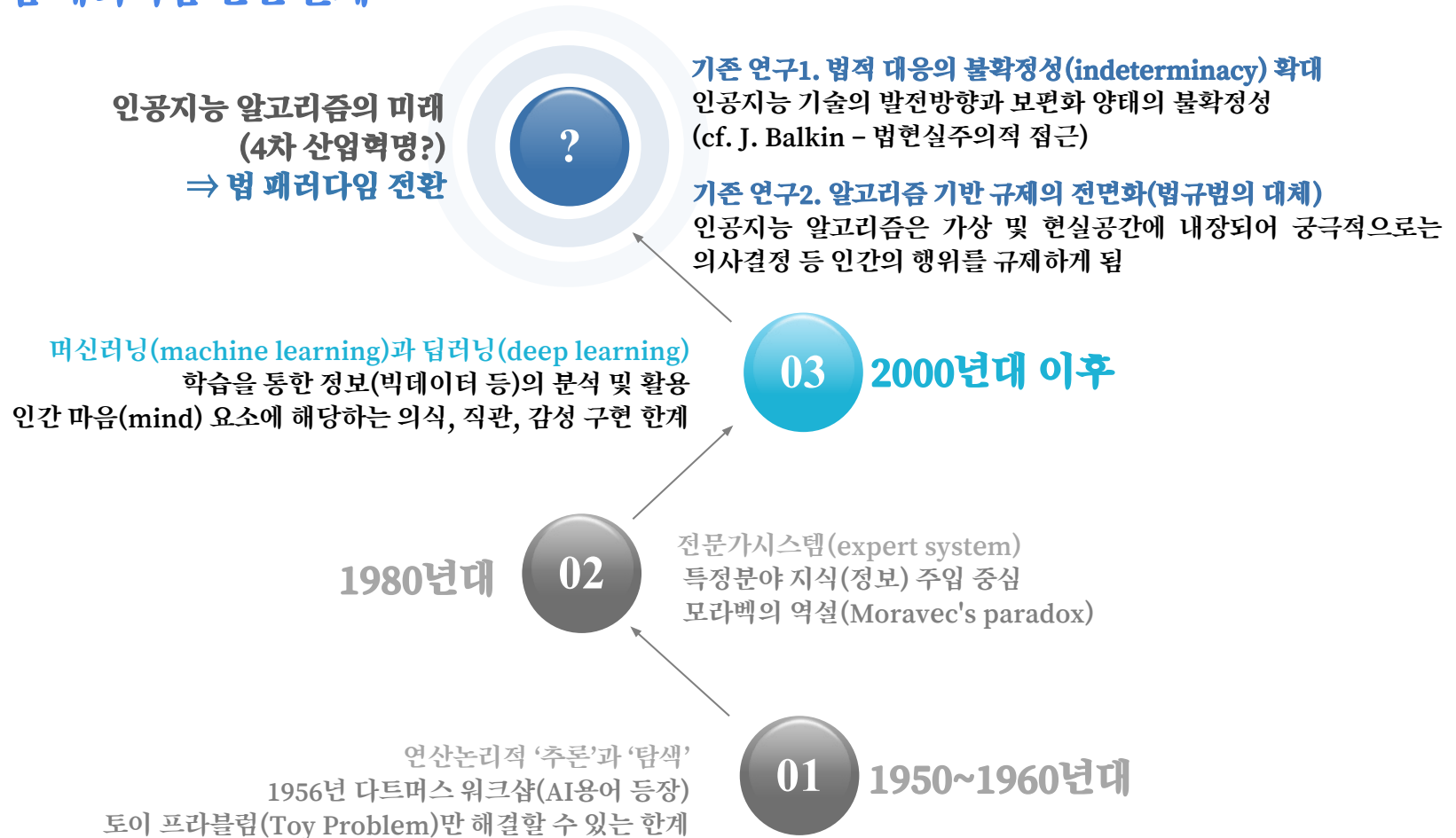
- 주요국의 대응동향
- 사례(1): 영국, 데이터 윤리 프레임워크 7원칙
- 사례(2): EU, 신뢰가능한 AI윤리지침안

3. 대응방안의 제언

- 상시적 영향평가 체계 구축(제도화)
- 데이터 및 알고리즘 리터러시 교육 강화
- 단편적 규제완화 일변도의 접근 지양

1. 법규범적 대응 패러다임 전환(1)

(법)규범 패러다임 전환 논제



1. 법규범적 대응 패러다임 전환(2)

디지털 전환(Digital Transformation)과 입법담론 과잉

❖ 규율 및 규제 대상의 불확정성

- 알고리즘 기반 규제(algorithmic regulation)의 전면화*
- 데이터/인공지능 알고리즘 관련 기술 발전의 불확정성(indeterminacy)

❖ 새로운 규제사안의 등장

- 종래에는 법적으로 문제시 되지 않아왔던 새로운 쟁점의 등장(cf. 포털)
- 사회적 영향력 있는 기술은 언제든지 새롭게 규제대상으로 논의 가능

❖ 사회 내 이해관계의 충돌

- 새로운 기술적 발전 및 보편화를 기반으로 하는 기존 신뢰이익 보호 문제
- 특정 기술을 둘러싼 이용자 및 산업 간의 이해관계

1. 법규범적 대응 패러다임 전환(3)

응답적 법(responsive law)으로의 전환과정

❖ 응답적 법(responsive law) (Nonet & Selznick, 1978)

- 외부적 환경은 그것이 법변동을 가져오는 것이 아니라 그것은 주로 법 내부의 발전적 잠재성을 촉진하거나 저해하는 데 작용함(cf. 체계이론)
- 기능주의적·실용주의적·합목적적 법정신

❖ 응답적 법의 특징

- 법적 추론에 있어 목적의 우위 또는 지배 현상(법형식주의 탈피)
- 법적 권위의 약화(법적 의무 및 공공성 등의 관념 변화)
- 법의 개방성과 융통성 요청에 따른 참여의 증가(법의 정치화)
- 법의 정당성 보다는 권능(실질적 목적달성 가능성)의 중시
(재판보다는 행정규제 중심)

1. 법규범적 대응 패러다임 전환(4)

응답적 법(responsive law)으로의 전환과정과 윤리적 접근 시도

❖ 진화론적 모델 구분

	억압적 법	자율적 법	응답적 법
법의 목적	질서	정당화	권능(competence)
정당성	사회방위와 국익	절차적 공정성	실질적 정의
법적추론	자의적, 개별적	법적 권위 중시(형식주의)	목적적
제량	광범위	협소(입법에 의한 제한)	목적 구속의 전제하에 확대

❖ 예시1) 한국, 개인정보 보호법

제1조(목적) 이 법은 개인정보의 처리 및 보호에 관한 사항을 정함으로써 개인의 자유와 권리를 보호하고, 나아가 개인의 존엄과 가치를 구현함을 목적으로 한다.

❖ 예시 2) EU, GDPR

(33) 과학적 연구 목적의 경우, 개인정보 수집 당시에 개인정보 처리 목적을 충분히 확인하기가 불가능할 때가 많다. 따라서 정보주체는 과학적 연구의 공인 윤리 기준에 부합된 경우, 특정 연구 분야에 한해 동의를 제공할 수 있다. 정보주체는 의도한 처리 목적이 허용하는 선에서 특정 연구 분야 혹은 연구 일부분에 한해 본인의 동의를 제공할 수 있어야 한다.

2. 각국의 윤리적 수준의 대응(1)

주요국의 대응동향(1)

❖ EU

- 2004년, 유럽로봇연구연합(EURON)
- 2005년, 제6차 다자간 공동기술개발연구프로그램(FP6)- 윤리로봇 프로젝트
- 2012년, 제7차 다자간 공동기술개발연구프로그램(FP7) - RoboLaw Project
- 2014년, 로보틱스 규제 가이드라인
- 2017년, Civil law rules on robotics(의회)
- 2018년, Coordinated Plan on Artificial Intelligence(집행위원회)

❖ 미국

- 2016년, Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights
- 2016년, Preparing for the Future of Artificial Intelligence
- 2016년, The National Artificial Intelligence R&D Strategic Plan
- 2016년, Artificial Intelligence, Automation, and the Economy

2. 각국의 윤리적 수준의 대응(2)

주요국의 대응동향(2)

❖ 영국

- 2016년, Robotics and Artificial Intelligence(UK House of Common)
- 2017년, Machine Learning: The Power and Promise of Computer that Learn by Example(Royal Society)
- 2018년, AI Sector Deal(BEIS)
- 2018년, AI in the UK: ready, willing and able?(UK House of Lords)

❖ 프랑스

- 2017년, 인공지능의 경제적·사회적 영향 전망(국가디지털위원회 등)
- 2017년, HOW CAN HUMANS KEEP THE UPPER HAND?: The ethical matters raised by algorithms and artificial intelligence(CNIL)
- 2018년, FOR A MEANINGFUL ARTIFICIAL INTELLIGENCE TOWARDS A FRENCH AND EUROPEAN STRATEGY(CÉDRIC VILLANI)

2. 각국의 윤리적 수준의 대응(3)

주요국의 대응동향(3)

❖ 일본

- 2016년, 인공지능 연구개발 안전성 확보 등 8대 원칙 제시(G7 정보통신장관 회의)
- 2017년, 인공지능 연구개발 가이드라인
- 2018년, AI이·활용 원칙안(2018)
- 2019년 예정, 인간중심 AI사회원칙안

❖ 기타민간

- 2016년, Ethically Aligned Design(IEEE)
- 2016년, Artificial Intelligence and Life in 2030(Stanford Univ.)
- 2017년, 아실로마 인공지능 23원칙(FLI)

2. 각국의 윤리적 수준의 대응(4)

참조사례(1): 영국, 데이터 윤리 프레임워크 7원칙(2018. 6. 13)

1. 명확한 이용자 요구와 공공 이익으로 시작하라.

보다 혁신적인 방식으로 데이터를 사용하면 공공 서비스가 제공되는 방식을 바꿀 수 있다. 우리는 시민과 공무원 모두에게 이용자를 위해 무엇을 성취하려고 하는지를 항상 명확히 해야 한다.

2. 관련 법률 및 실무 지침을 숙지하라.

데이터 사용과 관련한 법률 및 관행에 대해 이해하고 있어야 한다. 의심스러운 경우 관련 전문가에게 자문을 구해야 한다.

3. 이용자 요구에 비례하여 데이터를 사용하라.

데이터 사용은 이용자 요구에 비례해야 한다. 원하는 결과를 얻는 데 필요한 최소 데이터를 사용해야 한다.

4. 데이터의 한계를 이해하라.

정부에서 정책 및 서비스 디자인을 알리는 데 사용되는 데이터를 잘 이해해야 한다. 이용자 요구에 적실하게 활용할 수 있는지를 평가할 때 데이터의 한계를 고려하는 것이 중요하다.

5. 기술 역량 내에서 확신 있는 실무와 작업이 이루어지도록 하라.

신기술로부터 얻은 통찰력은 그것들을 만드는데 사용된 데이터와 실무 수준을 넘어서지 않는다. 만일 당신이 특정 접근 방식이나 도구를 높은 수준으로 사용하는 기술이나 경험이 없다는 사실을 인식하는 경우에는 당신의 기술 역량 범위 내에서 작업해야 한다.

6. 당신의 작업을 투명하게하고 설명할 수 있게 하라.

당신은, 가능한 한 개방적으로 작업하면서, 업무를 수행하는 데 사용한 도구, 데이터 및 알고리즘에 대해 투명해야 한다. 이를 통해 다른 연구자들이 당신이 발견한 내용들과 시민들이 새로운 유형의 업무를 이해할 수 있는지를 면밀히 검토할 수 있다.

7. 책임감 있게 데이터를 사용하라.

데이터로부터 발견된 통찰력을 확신시킬 수 있는 계획이 책임감 있게 활용되어야 한다. 이는 개발 및 이행 팀들이 견고한 평가 계획을 통해 산출 결과 및 데이터 모델들을 활용하고 모니터링 해야 하는 방식을 이해해야 한다는 것을 의미한다.

2. 각국의 윤리적 수준의 대응(5)

참조사례(2): EU, 신뢰가능한 AI윤리지침안(Draft Ethics Guidelines for Trustworthy AI, 2018. 12. 18)

제1장 : 윤리적 목적을 위한 핵심 지침(Key Guidance for Ensuring Ethical Purpose)

- 인간 중심적 AI 보장 : AI는 윤리적인 목적으로 개발, 구축, 활용되어야 하며 기본권, 사회적 가치, 선행의 원칙, 무해성의 원칙, 인간의 자율성, 정의, 설명 가능성 등을 반영하여야 한다.
- 기본권, 윤리 원칙 및 가치에 의거하여 인공지능이 인간과 공익을 위해 미칠 수 있는 영향을 사전에 평가하여야 한다. 어린이, 장애인, 소수민족과 같은 취약 집단과 관련된 사안이나 고용주와 종업원, 기업과 소비자 사이 권력 및 정보의 비대칭성이 있는 상황에는 더욱 특별한 주의가 필요하다.
- AI는 개인과 사회에 실질적인 혜택과 동시에 부정적인 영향을 가져올 수 있다는 사실을 인지하여야 한다.

제2장 : 신뢰할 수 있는 AI 구현을 위한 핵심 지침(Key Guidance for Realising Trustworthy AI)

- 초기 설계단계부터 신뢰할 수 있는 AI 요건을 포함하여야 한다(책임성, 데이터 거버넌스, 모두를 위한 설계, AI의 자율성 관리, 비차별성, 인간 자율성 존중, 프라이버시 존중, 견고성, 안전성, 투명성).
- AI시스템에 이러한 요구사항을 구현하기 위한 기술적 및 비기술적 방법을 고려하여야 한다.
- AI 시스템의 성능과 한계에 대한 정보를 사전에 이해관계자에게 명확히 고지하여야 한다.
- 신뢰할 수 있는 AI를 조직 문화로 만들고 관련 정보를 이해당사자에게 제공하여야 한다.
- AI 시스템의 설계 및 개발에 이해관계자의 참여를 보장하여야 한다.
- AI시스템의 감사가능성을 높이기 위해 노력하여야 한다.
- 책임 거버넌스를 위한 프로세스 확보하여야 한다.
- 신뢰할 수 있는 AI를 위한 교육 및 훈련을 실시하여야 한다.
- 서로 다른 목표 사이의 근본적인 긴장이 있을 수 있다는 사실을 인지하여야 한다.
- 신뢰할 수 있는 AI 요건 달성을 위한 연구 및 혁신을 장려하여야 한다.

제3장 : 신뢰할 수 있는 AI 평가를 위한 핵심 지침(Key Guidance for Assessing Trustworthy AI)

- AI 개발, 구축, 활용 시 신뢰할 수 있는 AI 평가 목록을 채택하여 활용 중인 특정 사례에 적용하여야 한다.
- 해당 평가 목록은 AI 시스템의 전체 수명주기에 있어 요구사항을 식별하고, 솔루션을 평가하며, 결과를 개선하는 지속적인 프로세스로 수행되어야 한다.

3. 대응방안의 제언

❖ 상시적 영향평가 체계 구축(제도화)

- 윤리적·규범적 담론의 실증적 논거 확보
- 담론절차(특히, 의견수렴 및 공개소명)의 제도화
- 우리나라의 경우에는 영향평가를 부수적으로 취급(cf. 규제위, 4차위)

❖ 데이터 및 알고리즘 리터러시 교육 강화

- 실질적·포용적 담론절차 진흥을 위한 기반 구축
- 핀란드, Elements of Artificial Intelligence free online course
- 기존 디지털·미디어 리터러시 교육의 실질화 및 확대 필요(cf. 민주시민교육)

❖ 단편적 규제완화 일변도의 접근 지양

- 현 상황에서 국제적으로 어떠한 규범 지형이 형성될지 예측할 수 없음
- 규제혁신은 필요하지만, 그것이 단순히 규제완화 및 폐지를 의미하는 것은 아님
- 현재 중요한 것은 규범 관념 혁신방향에 대한 컨센서스임

감사합니다!

경인교육대학교 사회과교육과
심우민(임법학, IT법학, 법교육)

legislation21@gmail.com

<http://legislation.kr>

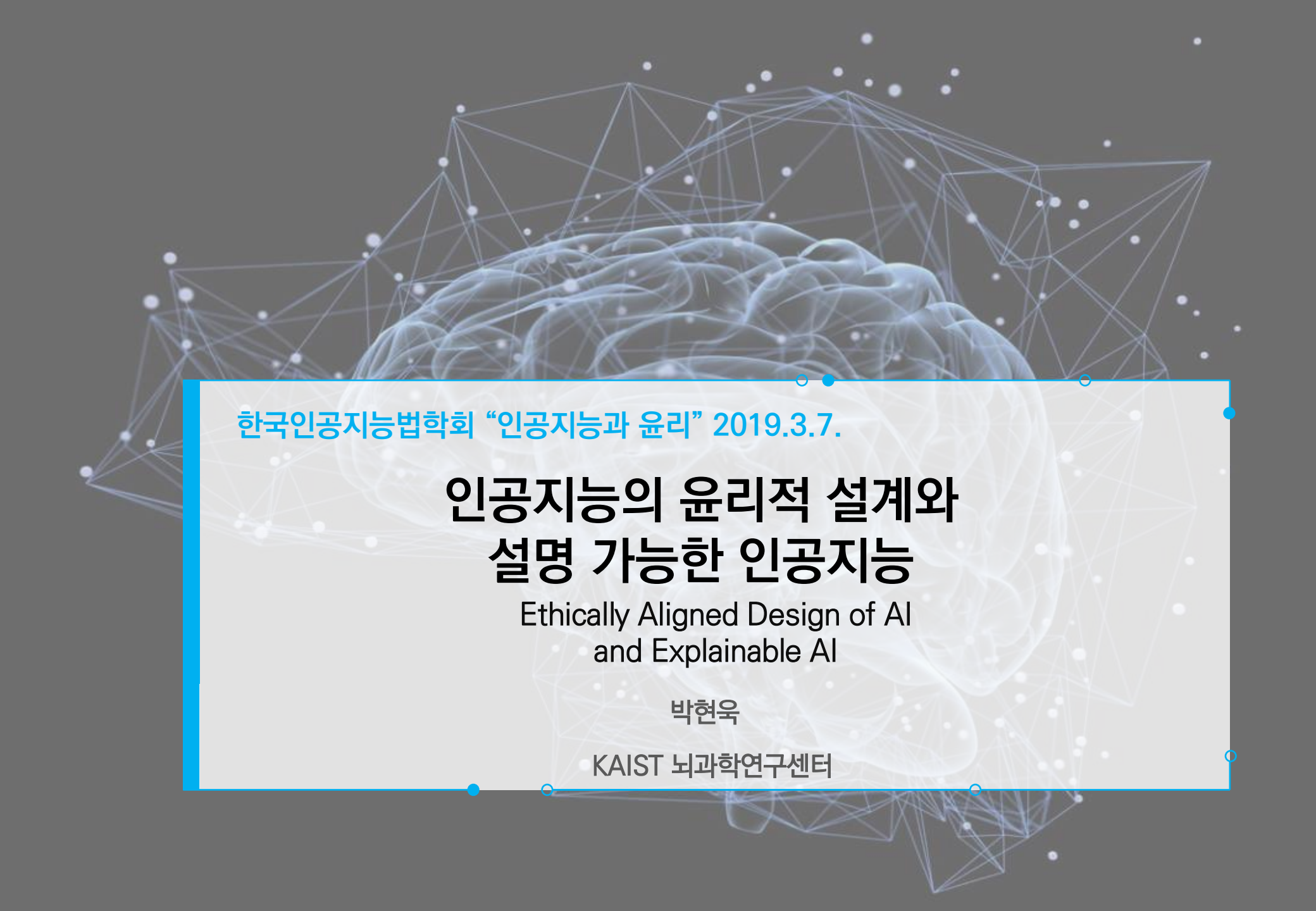
<http://civic-edu.kr>





Session I - 발표

인공지능의 윤리적 설계와 설명 가능한 인공지능



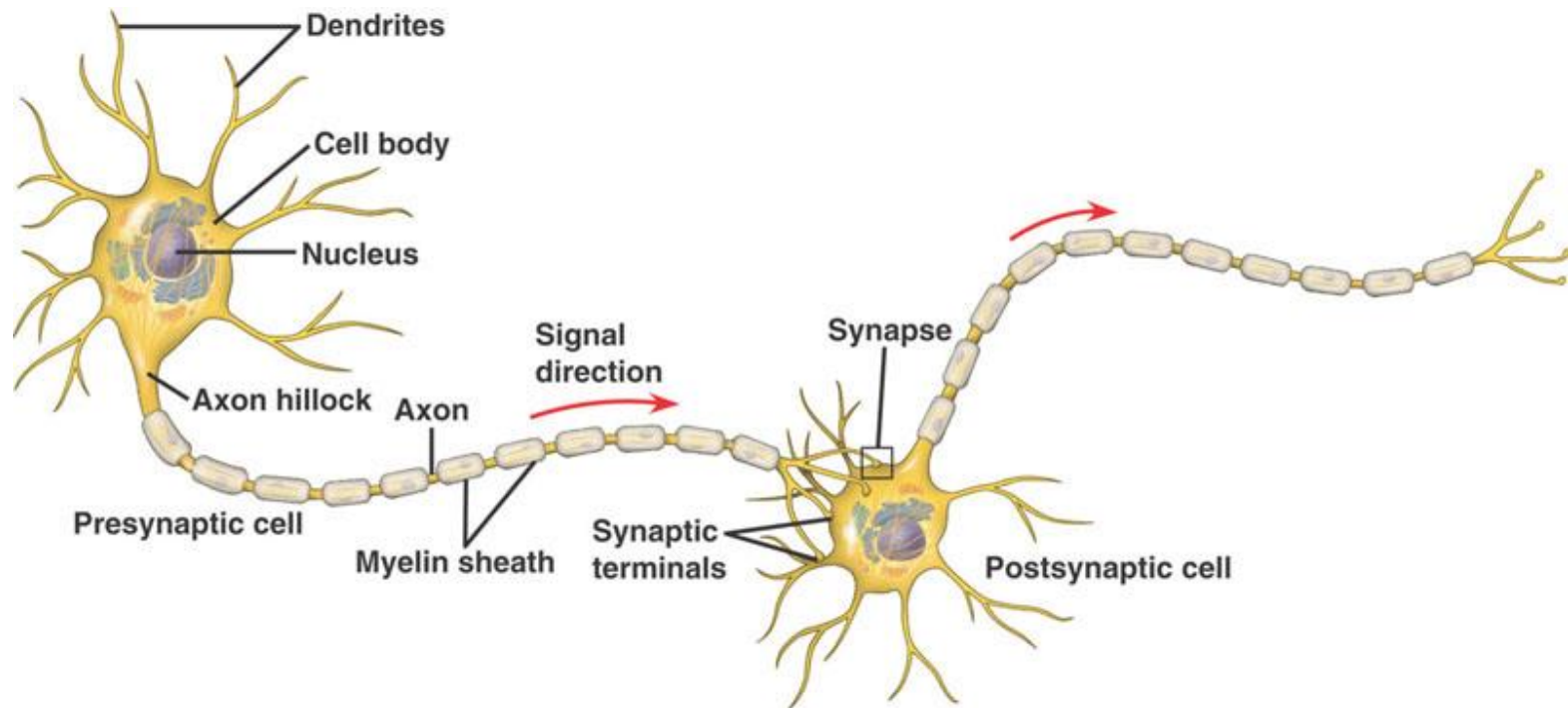
한국인공지능법학회 “인공지능과 윤리” 2019.3.7.

인공지능의 윤리적 설계와 설명 가능한 인공지능

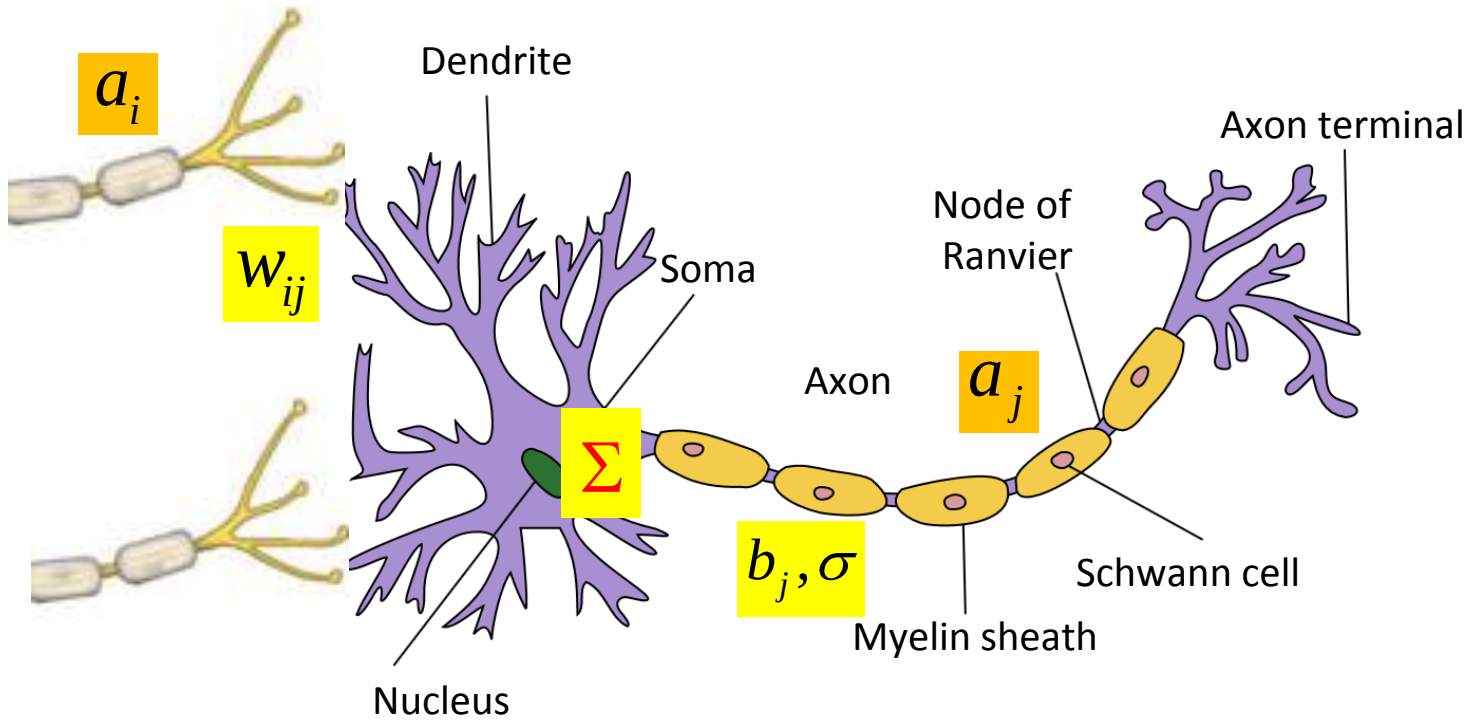
Ethically Aligned Design of AI
and Explainable AI

박현욱

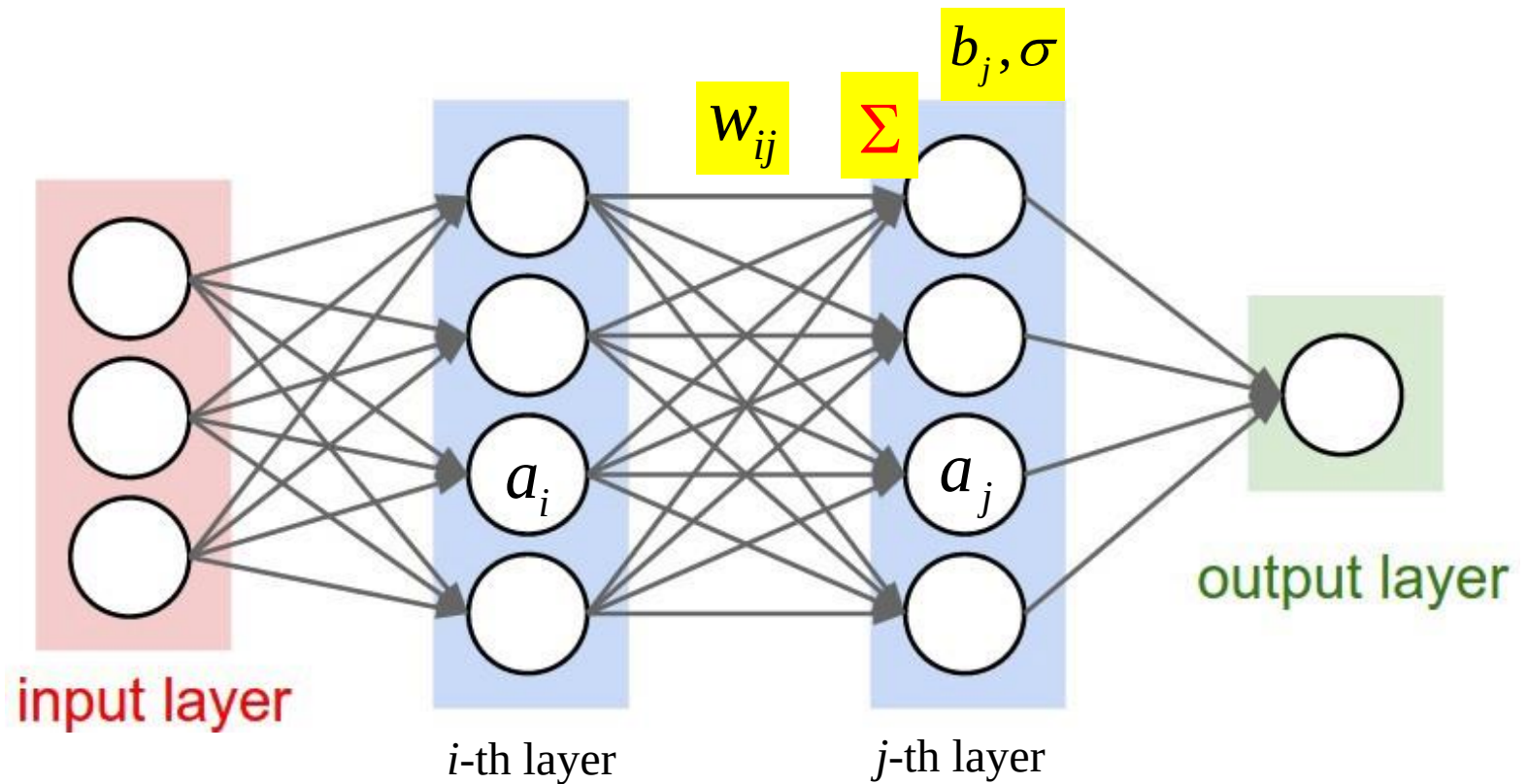
KAIST 뇌과학연구센터



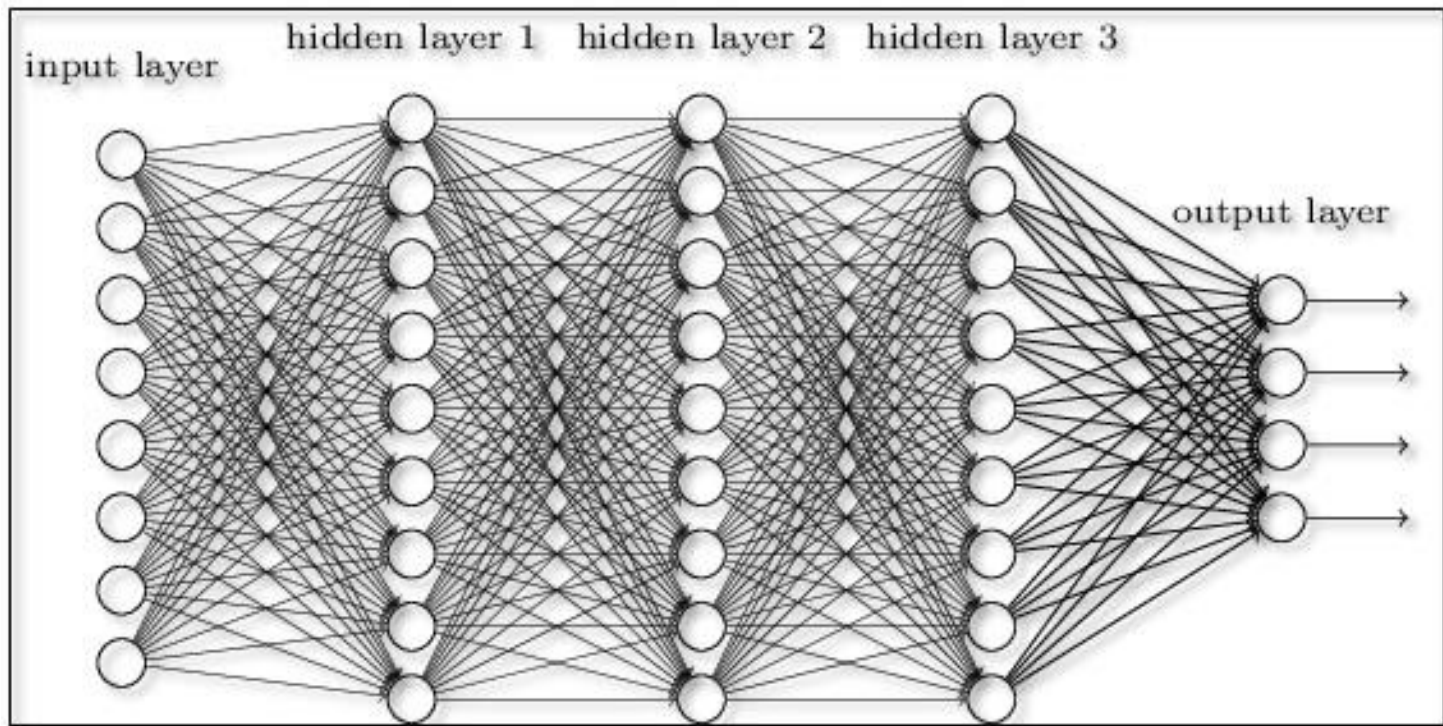
From biomedical advances



$$a_j = \sigma\left(\sum_i w_{ij} a_i + b_j\right)$$



$$a_j = \sigma\left(\sum_i w_{ij} a_i + b_j\right)$$

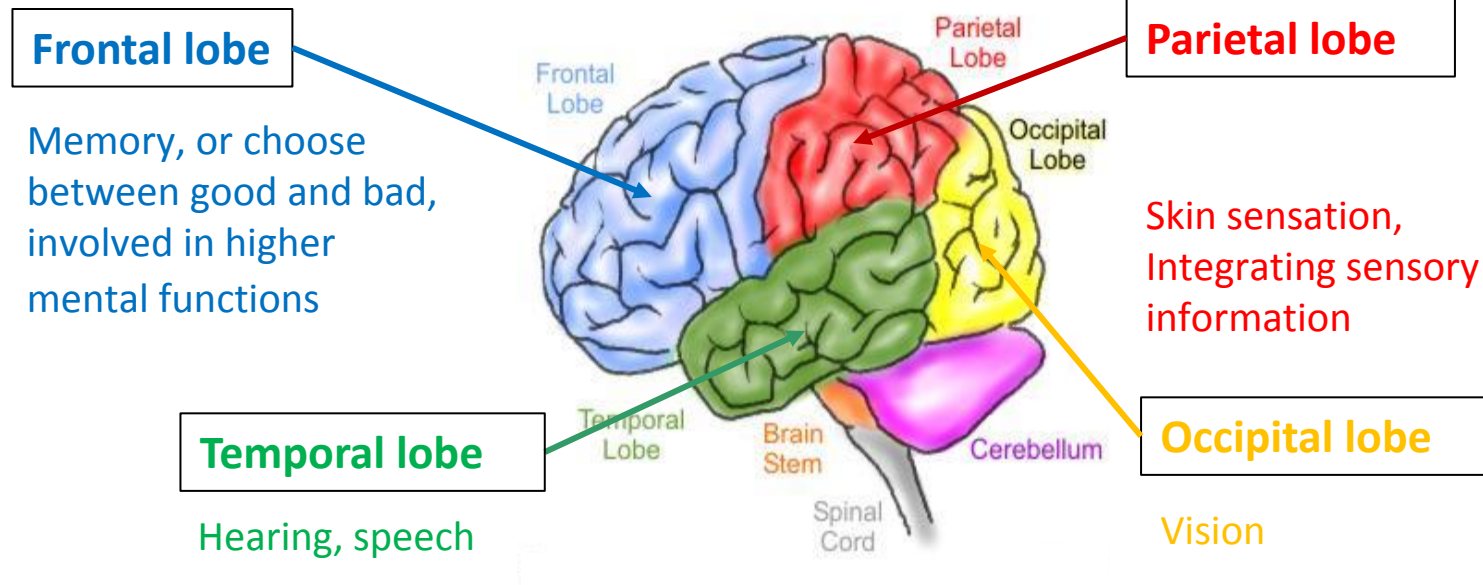


Computation power

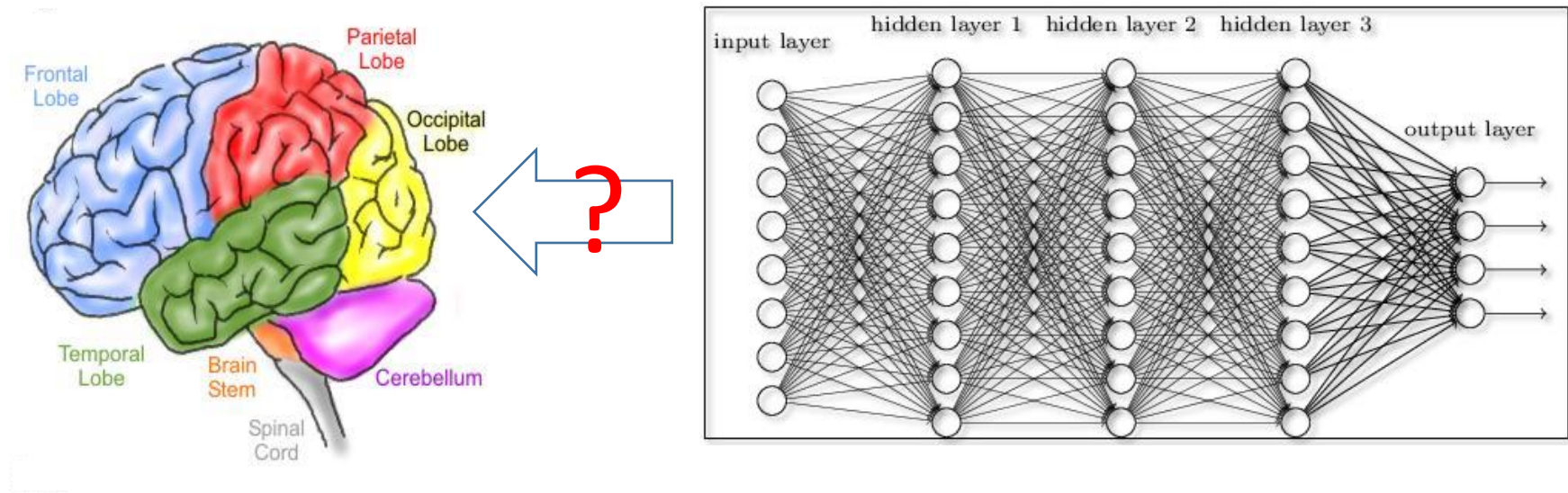
Big data

Learning algorithm

- Human brain weight: 1.4Kg (2% of body)
- Brain size : 1,200 cubic cm, Surface area of each hemisphere: 1.3 square feet
- Brain receives 15-20% of body's blood supply.
- 100 billion neurons in a human brain, each neuron is connected with thousands of neurons.



Deep Neural Network like Human Brain?



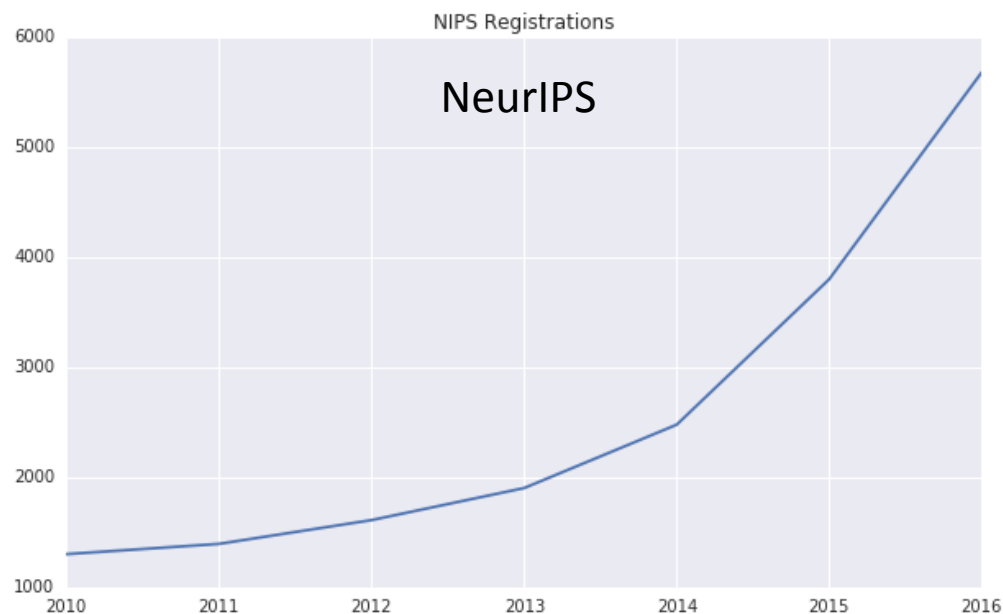
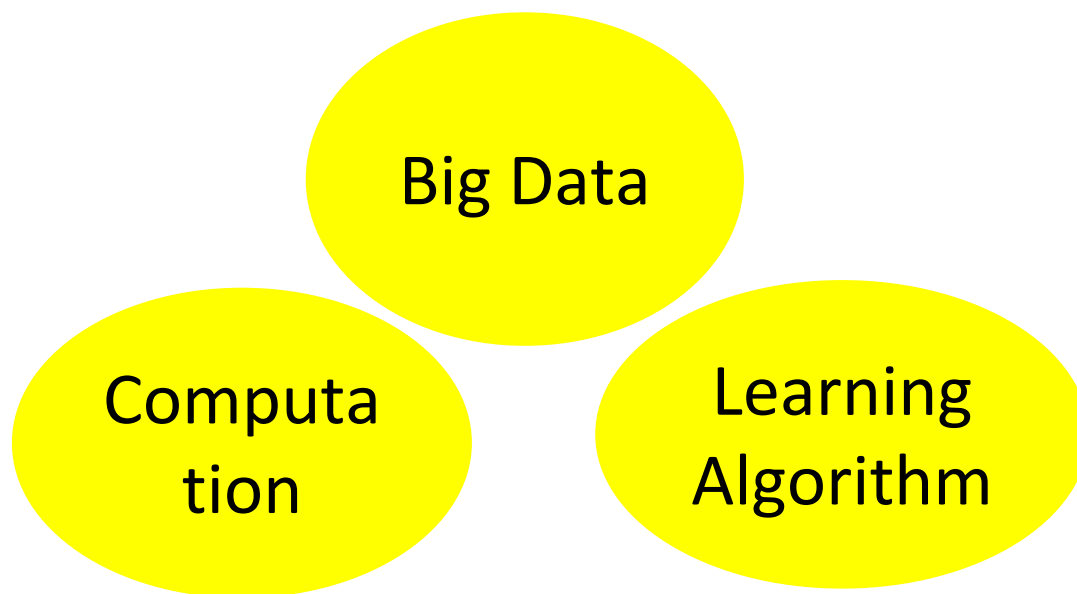
- Current Technology: Impossible to implement whole-brain-like ANN
 - Complexity
 - Power consumption
 - Learning



- Problem-specific Network
 - Eg. : Face recognition, Translation, Diagnosis, ...

- In 1950's, till middle of 1970's
 - Birth of AI
 - Machine translation in cold war era
 - Too optimistic: “Five, perhaps three, years hence, interlingual meaning conversion by electronic process in important functional areas of several languages may well be an accomplished fact.”
- First AI winter: 1974-1980
 - Lighthill report: “In no part of the field have discoveries made so far produced the major impact that was then promised.”
 - Financial setback from government (DARPA)

- In 1980's
 - Boom again by corporation
 - Expert system based on specific domain of knowledge
- Second AI winter: 1987-1993
 - Collapse of the market for AI hardware
 - Shutdown of AI companies (over 300)
- Until 2011
 - Some achievements on specific isolated problems
 - Become more cautious and more successful than ever



- IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems



- Principles of committees of the IEEE Global Initiatives.

Human Rights

Well-being

Accountability

Transparency

Awareness
of Misuse

- Do not infringe upon human rights
 - AI should be designed in a way that fulfills human rights.
- Safe and secure
 - Courses of any harmful results should be traceable.
- Under human judgment and control



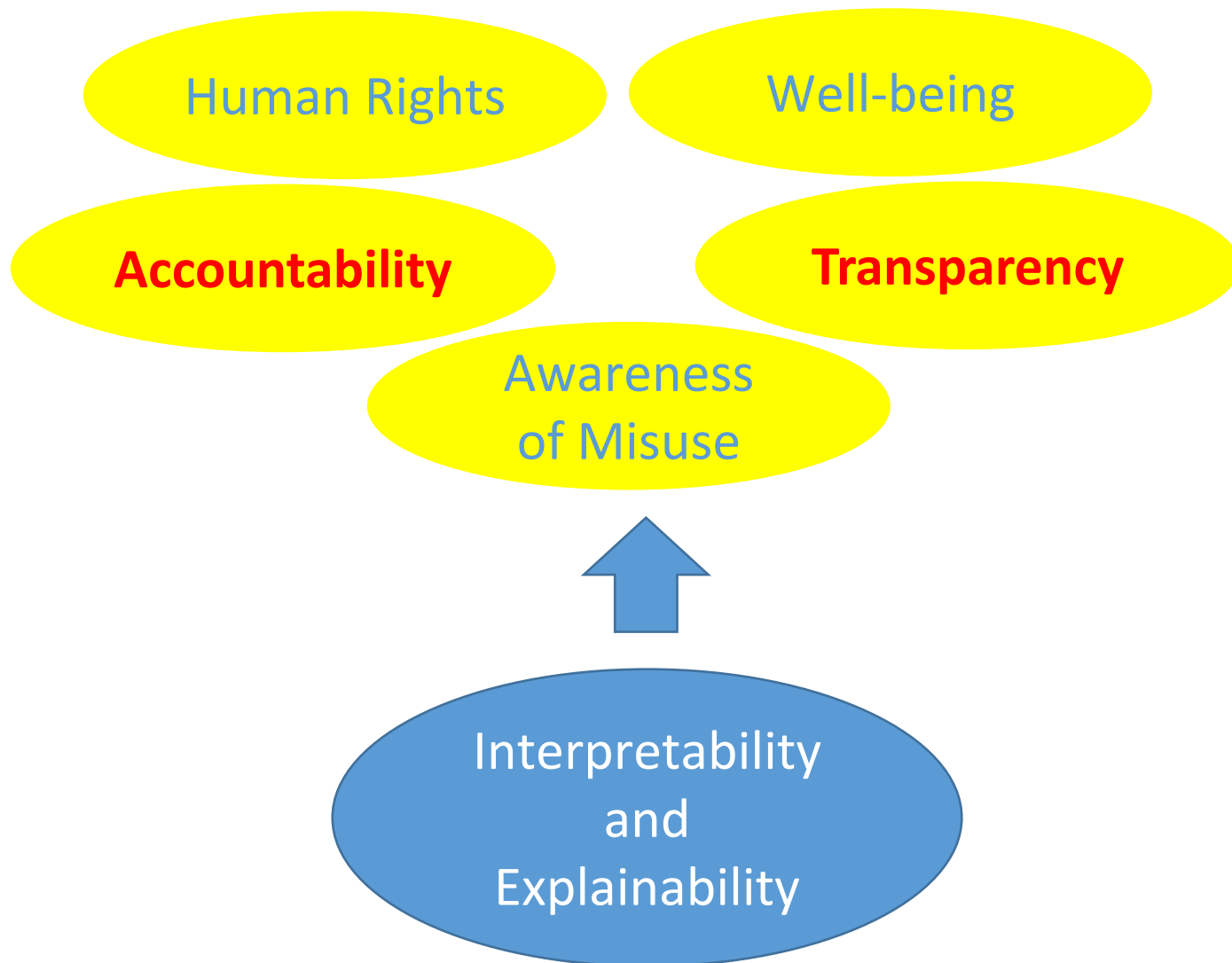
- Well-being:
 - Human satisfaction with life and condition of life
- Measuring subjective well-being is an essential part of measuring quality of life.
- To advance benefit of humanity
 - Avoiding negative unintended consequences
 - Increasing value for customers and society

- Legislatures/Courts
 - Clarify issues of responsibility, culpability, liability, and accountability for AI
- Developer
 - Provide transparency why a system operates in certain ways to address legal issues.
- Should be possible to find out who is legally responsible for a particular AI.

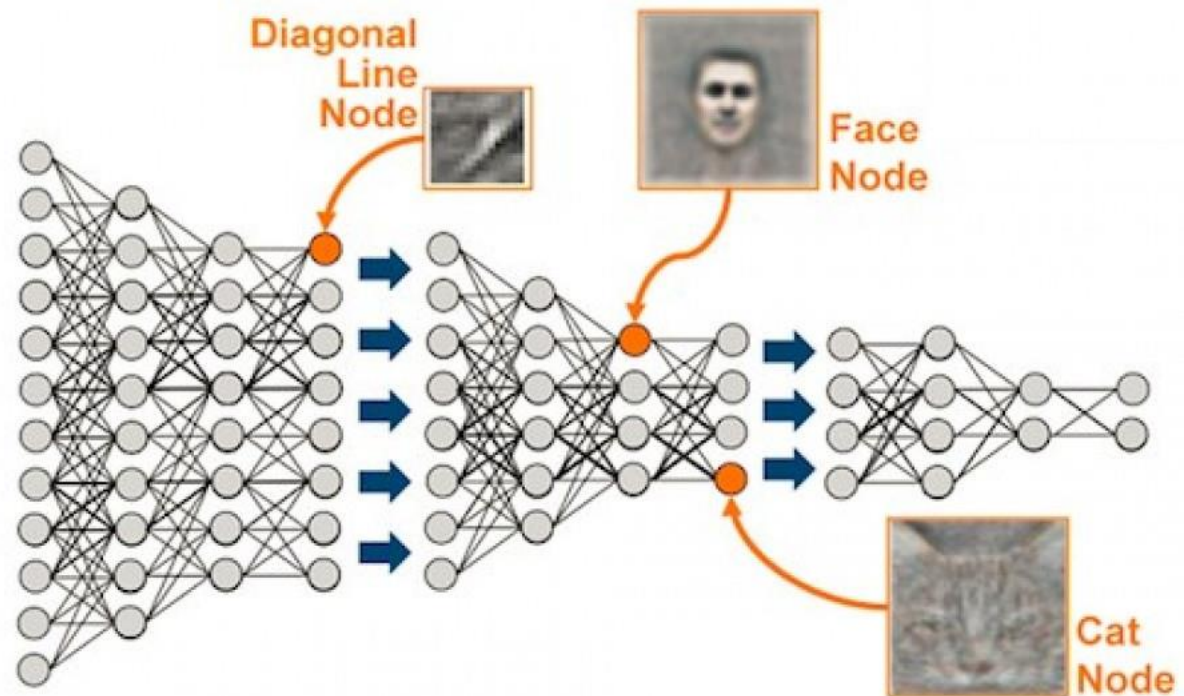
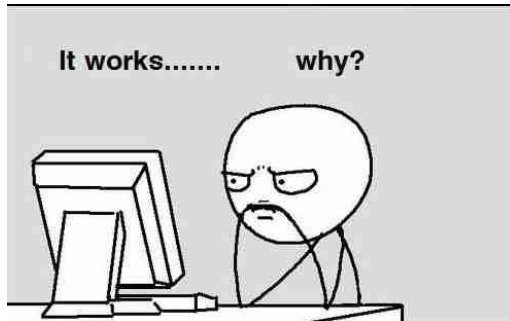
- Different levels of transparency for AI stakeholders
 - For users:
 - to understand what the system is doing and why
 - For validation and certification of AI:
 - to expose the system's processes and input data to scrutiny
 - For accident investigator:
 - to understand internal processes that led to the accident
 - For judges involved in the trial process:
 - to inform evidence and decision-making

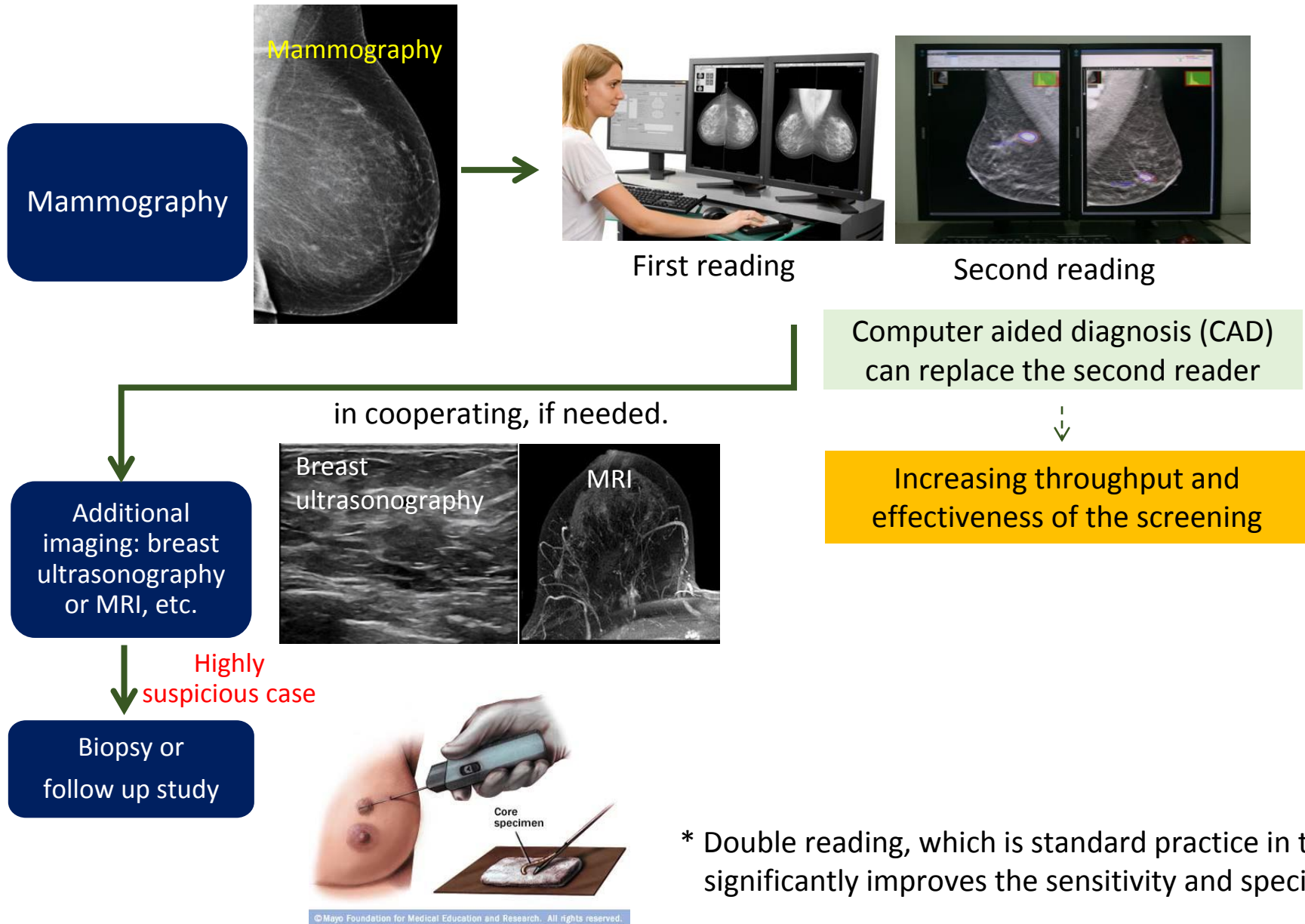
- Raise public awareness
 - To extend the benefits and minimize the risks of AI
- Data privacy
 - Potential risks of misuse of AI
- Increase credibility and minimize fear of AI



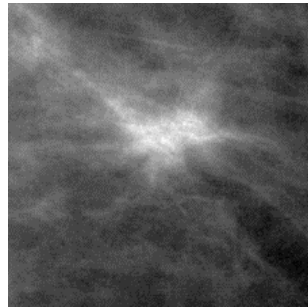


- Mapping of an abstract concept into a domain that the human can make sense of.
 - What drives the decision? (fairness)
 - Why does the AI take a certain decision? (accountability)
 - How can we trust the decision? (transparency)





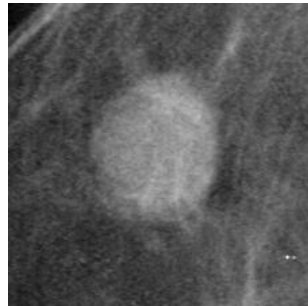
Medical description that XCAD take a decision for malignancy



Malignant
Mass margin: spiculated
Mass shape: irregular



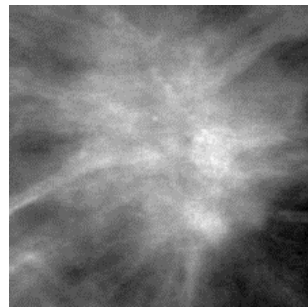
Sharp and pointed structures outgrowth all directions from the nonuniform mass



Benign
Mass margin: circumscribed
Mass shape: round



A ball-like mass has obviously different shade from that of surrounding.



Malignant
Mass margin: spiculated
Mass shape: irregular

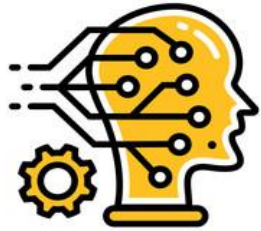


The contour of indefinite formed mass is constructed by many sharp lines projecting outside of mass

Test image

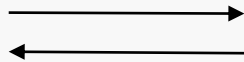
medical description explanation

Doctor-level interpretation



XCAD (Secondary doctor)

Interaction

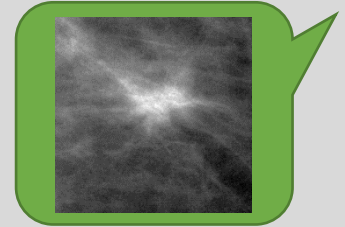


Medical doctor

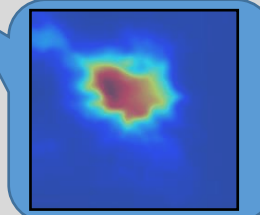
대화형 진단보조시스템

- 정확한 진단
- 진단에 따른 최적의 Recommendation
- Multimodal information 활용

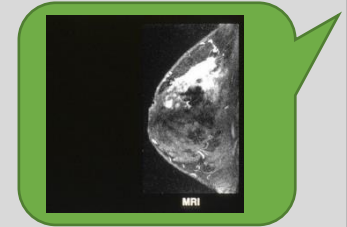
필요 기술: Explainable deep learning for vision, language and decision making, medical imaging



I see sharp and pointed structures outgrowth all directions from the mass, highlighted as below:



Looks like BIRADS CAT4 or CAT5. For more accurate diagnosis, I request MRI result





Session I - 발표

4차 산업혁명의 주요 분야별
인공지능 윤리이슈 분석 및 정책적 대응 방안



4차 산업혁명시대 산업별 윤리의 이슈 분석 및 정책적 대응방안 연구

한국인공지능법학회 세미나

2019. 3. 7

이원태(KISDI)
선지원(KISDI)



CONTENTS

- I 서론: 연구목표 및 주요내용
- II 인공지능 산업동향 및 산업·사회 변화상
- III 인공지능 윤리 국내외 동향 및 시사점
- IV 산업분야별 인공지능 윤리이슈
- V 인공지능 윤리이슈 정책적 대응방안



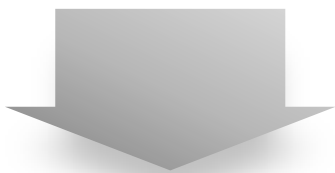


서론: 연구목표 및 주요내용

● 인공지능 윤리 이슈 분석의 배경 및 필요성(1)

- 인공지능 확산에 따른 역기능/부작용에 대해 법적 책임 규명 및 권리의 범위 혼선에 대비한 법제도적 논의가 증가하고 입법적 대응 점차 중요해지나, AI윤리 이슈 분석 등 윤리적 평가와 개발 논의 선행되어야 할 필요

인공지능의
산업적·사회적
활용 높아짐



사고발생의
위험성도 높아짐

윤리적 접근
과학자 등 전문가 주도

- ✓ 로봇윤리론 → 윤리적
- AI → XAI
- ✓ 데이터윤리

법제도적 접근
정부, 의회 주도

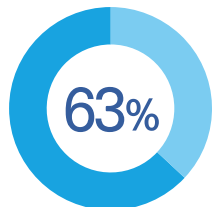
- ✓ 일반적 입법대응
- ✓ 개별법적 대응
- ✓ 소관권한 대응

- 법제도적 접근은 기본법 등 일반적 입법대응, 관련법 개정 등 개별적 입법 대응, '가이드라인' 등 법령개정 없는(부처) 소관권한 대응으로 세분화 가능 (김태오, 2017)

● 인공지능 윤리 이슈 분석의 배경 및 필요성(2)

- 인공지능 기술이 지닌 방대한 데이터, 알고리즘 복잡성으로 인해 통제와 예측이 쉽지 않고 알고리즘의 의도하지 않은 차별·편견(bias) 개입에 따른 피해 요인을 인지하기도 어려움
- 현재의 인공지능은 연구/개발자가 의도를 가지고 개발하는 시스템이고 매우 빠른 속도로 발전하는 기술이므로, 공학윤리 등 지속적으로 윤리적 논의체계를 고도화시켜 나가야 함

AI 윤리이슈 중 가장 심각한 것은?



데이터 과학자들의 63%가
머신러닝에 의해 인종, 종교와 같은 인간의
편견이 프로그래밍 되는 것이라고 답함

인공지능 데이터 기업 CrowdFlower에서 발간한
'2017 Data Scientist Report' (2017.4.20. ADT MAG)

‘워드 임베딩’이란? (word embedding)

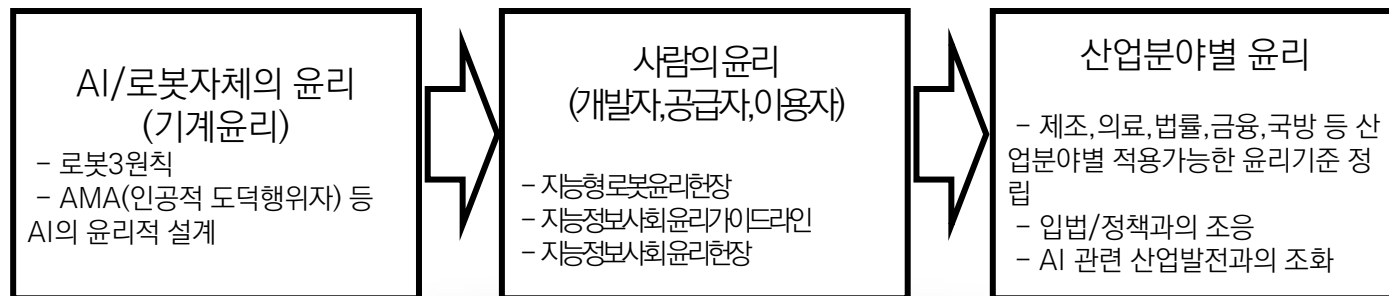
알고리즘 기술에 의해 인간의 편견이
인공지능에 학습되는 기제로
알고리즘 개발단계에서의 도덕적 제어
장치가 필요함을 강조

영국 Bath대 조애너 브라이슨 교수팀 (Guardian, 2017.4)

● 인공지능 윤리 논의의 방향 전환(1)

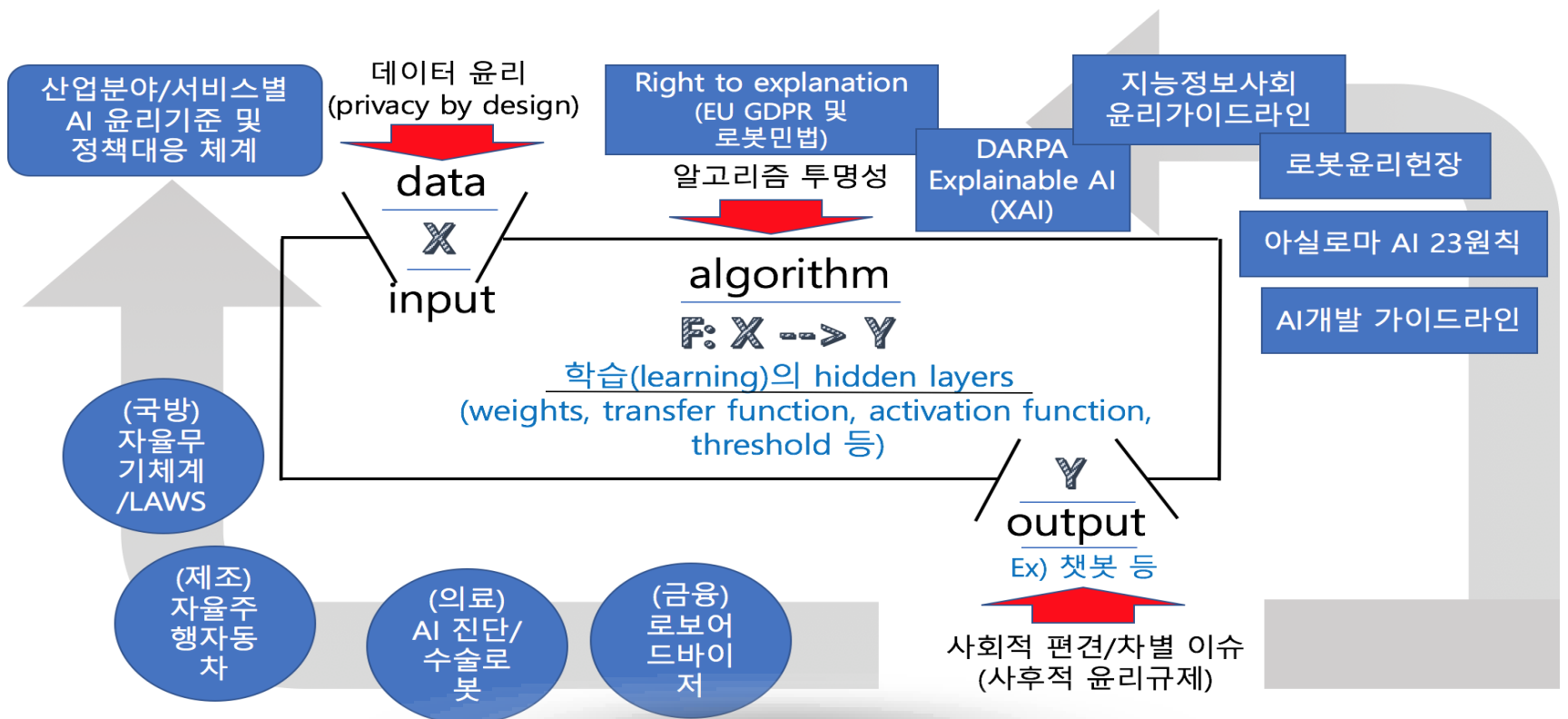
- AI의 윤리적 이슈가 인간 및 사회전반에 대한 일반적/총론적 규범 논의에 그치거나, 행위자 중심의 윤리적 행위 규범 중심으로 논의되지만, AI기술 및 서비스의 구체적 적용 분야의 특성에 따른 윤리이슈 검토 및 대응 논의는 상대적으로 소홀함
- AI기술이 적용되는 주요 산업 분야 및 서비스별 인공지능 라이프사이클 (예: AI 알고리즘 설계 및 개발, 데이터 학습 및 입력, 제품/서비스 유통 및 이용단계 등)에 따른 윤리이슈 분석 및 대응방안 논의 구체화 필요
- AI 윤리 논의가 AI 신기술 및 서비스에 대한 규제로 작용하지 않고 관련 산업 발전과 조화를 이루는 방향으로 진행되어야 함

〈인공지능 윤리담론의 변화과정〉



● 인공지능 윤리 논의의 방향 전환(2)

- AI 기술이 폭넓게 활용되고 윤리이슈도 쟁점화되는 산업분야를 1)제조, 2)의료, 3)금융, 4)국방(군사) 등 4개 산업 영역으로 설정,
- 각각 1) 자율주행자동차, 2) AI의료서비스, 3) 로보어드바이저, 4) 킬러로봇 /치명적 자율무기체계(LAWS: Lethal Autonomous Weapons System)의 윤리이슈 분석 및 정책적 대응방안 도출



● 인공지능 윤리 논의의 방향 전환(3)

기존 논의의 한계점

AI 등 기술혁신의 사회적 영향평가
체계 충분히 확보되지 않은 상황

AI기술의 발전 경로가 다양하고
형식 또한 불확정적이고 불확실함
사회경제적 파급효과
부문·계층에 따라 차별적·복합적

AI 윤리 규범의 다양화
EU, 일본 등 각국의 기술적 장점 및 맥락
반영한 윤리규범 다양하게 모색

본 연구 AI 윤리 논의의 방향성

사업자, 개발자, 사용자 등이
**함께 참여하는 윤리적 공론화가
선행(또는 병행)될 필요**

일사불란한 법체계 수립 보다는
기술/산업/사회의 복합적 변화에 대비한
**다양한 윤리적 논의 및 대응방식
필요**

**우리 현실에 맞는 윤리적 논의 틀
(ethical framework) 정립도 필요**

● 본 연구의 주요 내용 및 목차 구성

1

지능정보기술의 산업분야별 활용현황 조사 및 사회변화 분석

- 지능정보기술 개념 및 정의, 분류, 단계별 발전 현황 분석
- 지능정보기술의 제조업, 금융, 의료, 국방 등 산업별 활용 현황 조사
- 지능정보기술의 발전으로 인한 사회·산업의 변화와 영향 분석

2

국내외 인공지능 윤리 연구동향 및 주요국의 정책대응 현황 분석

- 학계, 기업, 정부/의회, 국제기구 등 국내외 AI윤리 동향 분석

3

산업분야별 인공지능 윤리 관련 이슈 조사 및 분석

- 산업별 지능정보기술*이 활용되고 있는 산업별 윤리 이슈** 조사
- * 자율자동차, 자율무기체계, 로보어드바이저, 돌봄/수술로봇 등
- ** 데이터 활용 및 보호, 안전/보안(security), 사회적 차별, 소비자 보호, 책임소재 등
- 산업별 윤리 이슈를 도출·정리하고 이에 따른 국가별 대응 동향 분석

4

산업분야별 지능정보기술 윤리 관련 이슈에 따른 정책적 대응방안 제시

- 인공지능 윤리이슈에 대한 정책 대응 프레임워크 구축 및 운영방향
- 제조, 의료, 금융, 국방 등 산업별 인공지능 윤리 관련 이슈 해결 모색을 위한 정책 설계방향 제시

II

인공지능 산업동향 및 산업·사회 변화상

인공지능 시장규모

● 전 세계 인공지능 시장규모

- 전 세계 인공지능 시장규모 전망은 각 조사기관마다 적용하는 인공지능의 범위 및 분류 기준이 크게 달라 관련 시장의 절대적인 성장 잠재력을 예측한 결과도 큰 편차를 보이고 있음
- 조사기관마다 인공지능 시장규모 전망이 상이하지만 대부분 연평균 50%를 상회하는 높은 성장률을 예측하고 있어, 도입 단계의 인공지능 기술이 향후 상용화 단계로 발전하면서 폭발적인 성장세와 타산업과의 지능형 융합을 견인할 것으로 예상

● Statista(2018.6.)

- 2017년(48.2억 달러), 2018년(73.5억 달러), 2025년(898.5억 달러)
- CAGR 51%

● IDC(2018.3.)

- 2017년(123.9억 달러), 2018년(191억 달러), 2021년(522억 달러)
- CAGR 46.2%

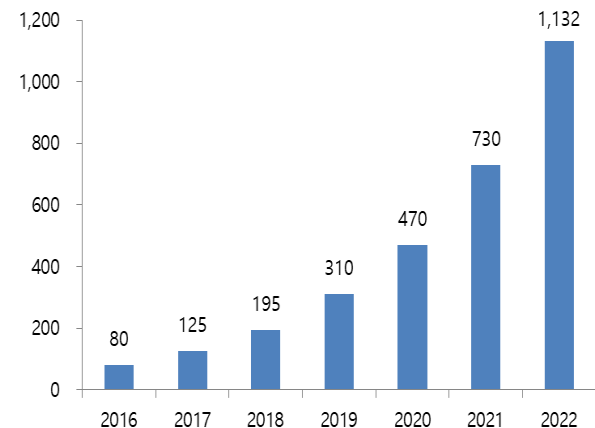
● Gartner(2018.4.)

- 2017년(125억 달러), 2018년(195억 달러), 2022년(1,132억 달러)
- CAGR 55%

● Tractica(2017)

- 2016년(14억 달러), 2025년(598억 달러)
- CAGR 52%

[참고] 전 세계 인공지능 시장 규모 전망(Gartner, 2018.4.)



인공지능 시장규모

국내 인공지능 시장규모

- 2016년(2조 2000억 원), 2025년(11조 원), 2030년(27~30조 원) 규모를 형성할 것으로 전망(과기정통부, 2016)
- 국내 AI 시장은 세계시장의 약 1.6% 비중인 것으로 파악됨('ICT R&D 중장기 기술 로드맵 2020', 2016)

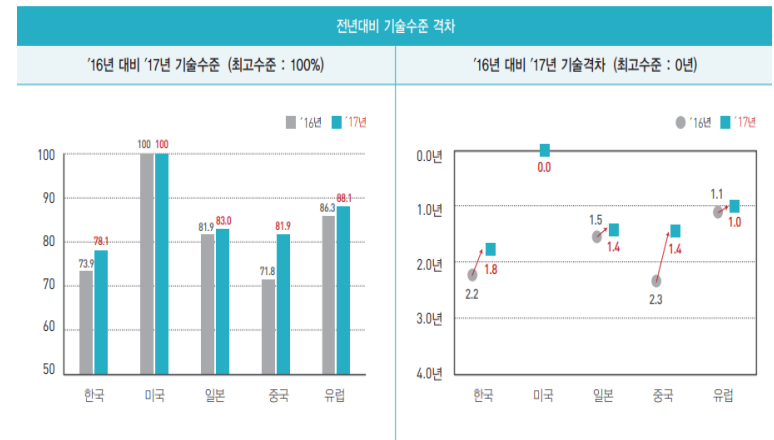
[참고] 국내 인공지능 시장 규모 전망



자료: 과학기술정보통신부(2016).

자료: KT경제경영연구소(2016).

[참고] AI 기술수준 추이 및 주요국 대비 기술격차



II. 인공지능 산업동향 및 산업·사회 변화상

인공지능 적용산업 현황 및 활용사례 분석

인공지능 활용 산업분야 분류

- 지능정보사회 중장기 종합대책(舊 미래부, 2016): 지능정보화 촉진 대상산업(총 12개)
- 16년 지능정보기술을 통한 사회경제 혁신방안 연구(맥킨지, 2016): 5개 산업분야 및 4개 라이프분야(총 9개)
- 지능정보사회 구현을 위한 심층통계 경제효과 및 세부이행전략 분석연구(NIPA, 2017): 총 15개 산업
- 인공지능(AI) R&D전략(과학기술정보통신부, 2018): 총 10개 산업

⇒ 정책적 중요도에 따라 **17개 산업분야**로 세분화하여 현재 서비스 중인 지능정보기술 적용 유스케이스를 분석

연년 분류 주체	1		2		3		4	5	6	7	8	9	10	11	12	13	14	15	16	17
맥킨지 (2016)	의료		제조			금융	유통	교통	도시	주거	헬스	기타								
舊미래부 (2016)	의료	바이오	제조	에너지	로봇	금융	유통	교통	-		복지	-	국방	행정		-	-	보안		
NIPA (2017)	의료		제조		에너지	금융	유통	교통	도시	생활	환경	농업	국방	행정사무		교육	문화관광	안전		
과학기술정보통신부 (2018)	의료		제조		에너지	금융	-	이동체	도시	-	복지	농수산업	국방	-		-	-	안전		
본 연구의 분류체계	의료		제조		에너지	금융	유통	교통	도시	홈	헬스	농업	국방	법률	사무관리	공공분야	교육	문화관광	안전	

II. 인공지능 산업동향 및 산업·사회 변화상

인공지능 적용산업 현황 및 활용사례 분석

인공지능 활용 산업분야 분류

[참고] 도출된 17개 인공지능 적용산업 분류

순번	산업분야	산업별 인공지능 적용 예시
1	의료	진단의 정확도 증가, 약물의 오남용방지 등 맞춤형 정밀의료 서비스 제공
2	제조	3D 프린팅 등으로 생산기간 단축 및 수요 맞춤형 다품종 소량생산
3	에너지	에너지의 소비 패턴을 통한 비용절감과 필요한 에너지는 자가 생산 및 충족하고, 잉여 에너지는 판매하는 'E(Energy)-프로슈머' 등장
4	금융	사람의 개입 없는 로보어드바이저 기반 개인자산 분석, 금융 사기 탐지 및 대응
5	유통	어디서든 사용자가 원하는 물건을 구입하는 무노력 쇼핑(Zero-Effort Shopping) 실현
6	교통	사고 없이 안전하게 운행하는 무인버스, 무인택시의 상용화
7	도시	교통 정체구간 분석을 통하여 실시간 교통관제로 운전시간 및 연료, 이산화탄소 배출비용 감소
8	홈	가사 자동화 시스템으로 인한 가사노동 시간 단축과 스마트 홈 시스템을 기반으로 한 에너지 관리 효율화 및 보안 강화
9	헬스	생활운동 유인 및 건강상태 모니터링을 통한 건강증진 효과 및 대기·수질오염 측정 모니터링 시스템을 통한 실시간 대응으로 관련 질병 예방과 사망자수 감소
10	농업	자율주행 장비, 온습도 센서, 양액 재배, 관리시스템 등 농업의 생산성 향상 및 노동력 저감을 위한 지능형 시스템 개발
11	국방	지능형CCTV, 빅데이터를 활용한 국제범죄, 테러 등의 범죄를 사전에 예측하여 신속하게 대응
12	법률	법률, 세금 등 다양한 법 관련 분야에 대한 실시간 대화형 통합상담서비스 제공
13	사무관리	사무실 내 업무환경 조성 및 문서 관리, 인사 조직 관리 등 ERP 서비스를 통해 기업의 전반적인 관리 서비스 제공
14	공공분야	생애주기별 실업, 주거, 돌봄, 질병 등 사회적 위험에 대응하는 통합형 복지서비스 제공
15	교육	학습패턴, 공부방식 분석 등을 통한 개인 맞춤형 교육 서비스 제공
16	문화·관광	VR·AR을 통해 방문자가 직·간접적으로 현장 체험할 수 있는 서비스 제공
17	안전	최근 발생하는 싱크홀 등 지하구조물 관리 및 태풍, 홍수 등 재해예방 예방 서비스 제공

AI 적용산업 분류를 통한 유스케이스 분석

분석대상: 기업(총264개)

- 해외(109개): 인공지능 스타트업 100개 기업 및 주요 인공지능 기업
- 국내(155개): 지능정보산업협회 소속 인공지능 기업

분석대상: 인공지능 제품/서비스(총 345개)

- 해외 제품/서비스(129개), 국내 제품/서비스(216개)

[참고] 산업별 인공지능 유스케이스 분석 프로세스

단 계	연구내용	수행방법
국내외 주요 제품 조사	• 해외 출시(예정) 제품 조사 • 국내 출시(예정) 제품 조사	• 문헌조사 • 온라인 조사
제품의 유형화	• (산업) 17개 산업유형	• 산업별 Grouping
국내외 비교분석	• 해외/국내 제품의 산업/기술분야별 집중도 분석	• Matrix 분석
국내 Gap 도출	• 국내외 비교 분석을 통한 중점영역 및 공백분야 도출	• Gap 분석

인공지능 적용산업 현황 및 활용사례 분석

● 해외 인공지능 적용산업 포지셔닝

- 인공지능 100개 스타트업 및 주요 인공지능 관련 기업의 인공지능 제품·서비스에 대한 산업 집중도 분석
- 17개 산업분야 중 웰니스(94개), 홈(92개), 금융(77개), 의료(73개), 교통(52개), 국방(46개), 제조(45개) 순

[참고] 해외 인공지능 제품/서비스별 적용 산업 집중도

구분	의료	국방	교육	농업	법률	웰니스	제조	홈	금융	문화 관광	유통	사무 관리	도시	교통	에너지	안전	공공
시각 지능	10	7	2	3	1	9	5	7	5	2	2	2	6	9		3	5
언어 지능	6	5	6		2	11	3	15	9	4	4	1	3	5			4
자연어 처리	4	2	4		2	6	2	16	7	2	4			3			2
음성 지능	4	6	7	1	2	13	3	16	7	4	3	2	5	7			1
공간 지능		4		1		4		2	1	1	1		5	8			
기계 학습	20	8	10	2	2	21	11	15	14	7	10	5	6	8	1	3	7
지식 표현	10	2	2		2	4	3	12	7	3	4		1	3			6
DB 관리	19	12	4	3	2	26	18	9	27	6	6	14	7	9	1	8	12
합계	73	46	35	10	13	94	45	92	77	29	34	24	33	52	2	14	37
순위	4	6	9	16	15	1	7	2	3	12	10	13	11	5	17	14	8

인공지능 적용산업 현황 및 활용사례 분석

국내 인공지능 적용산업 포지셔닝

- 지능정보산업협회 소속 인공지능 기업의 인공지능 제품·서비스에 대한 산업 집중도 분석
- 17개 산업분야 중 웰니스(183개), 제조(115개), 금융(111개), 홈(104개) 순

[참고] 국내 인공지능 제품/서비스별 적용 산업 집중도

구분	의료	국방	교육	농업	법률	웰니스	제조	홈	금융	문화 관광	유통	사무 관리	도시	교통	에너지	안전	공공
시각 지능	6	4	5			18	14	8	3	1	5	3	9	13		5	3
언어 지능	6	1	3	1	1	21	13	17	10	12	18	4	7	3	1	1	8
자연어 처리	2	1	1	1	1	4	1	10	4	4	8	1	1	1	1	1	
음성 지능	6	1	7	1	1	22	5	18	3	8	8	1	4	4	1	1	4
공간 지능		3	1	1		3	8	8	1	1	1		4	6			
기계 학습	17	2	6	5	3	52	22	15	35	16	24	9	16	7	2	9	14
지식 표현	8	3	6	1	1	25	19	14	14	9	15	3	7	4	4	3	7
DB 관리	10	3	6	4	1	38	33	14	41	5	23	13	10	8	6	9	22
합계	55	18	35	14	8	183	115	104	111	56	102	34	58	46	15	29	58
순위	9	14	11	16	17	1	2	4	3	8	5	12	6	10	15	13	6

II. 인공지능 산업동향 및 산업·사회 변화상

인공지능 적용산업 현황 및 활용사례 분석

● 국내외 인공지능 제품·서비스 기술 적용도 Matrix 분석 결과

- 국내외 Matrix 분석결과에 따라 상대평가를 통해 집중도의 등급을 1~4단계로 구분하여 각 산업별 집중도 도식화
- 집중도의 등급은 국내외 Matrix 분석 결과에 대해 집중도 수치를 4단계로 구분하여 적용 집중도 레벨 설정

레벨 구분	LV.1	LV.2	LV.3	LV.4
해외 빈도	1~7	8~14	15~21	22~28
국내 빈도	1~13	14~26	27~39	40~52

[참고] 국외 인공지능 제품·서비스 기술 집중도

구분	의료	국방	교육	농업	법률	헬니스	제조	유통	금융	문화 관광	유통	사무 관리	도시	교통	에너지	안전	공공
시각	LV.2	LV.1	LV.1	LV.1	LV.1	LV.2	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.2		LV.1	LV.1
언어	LV.1	LV.1	LV.1		LV.1	LV.2	LV.1	LV.3	LV.2	LV.1	LV.1	LV.1	LV.1	LV.1			LV.1
자연어 처리	LV.1	LV.1	LV.1		LV.1	LV.1	LV.1	LV.3	LV.1	LV.1	LV.1			LV.1			LV.1
음성	LV.1	LV.1	LV.1	LV.1	LV.1	LV.2	LV.1	LV.3	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1			LV.1
공간		LV.1		LV.1		LV.1		LV.1	LV.1	LV.1	LV.1		LV.1	LV.2			
기계	LV.3	LV.2	LV.2	LV.1	LV.1	LV.3	LV.2	LV.3	LV.2	LV.1	LV.2	LV.1	LV.1	LV.2	LV.1	LV.1	LV.1
지식	LV.2	LV.1	LV.1		LV.1	LV.1	LV.1	LV.2	LV.1	LV.1	LV.1		LV.1	LV.1			LV.1
DB	LV.3	LV.2	LV.1	LV.1	LV.1	LV.4	LV.3	LV.2	LV.4	LV.1	LV.1	LV.2	LV.1	LV.2	LV.1	LV.2	LV.2

[참고] 국내 인공지능 제품·서비스 기술 집중도

구분	의료	국방	교육	농업	법률	헬니스	제조	유통	금융	문화 관광	유통	사무 관리	도시	교통	에너지	안전	공공
시각	LV.1	LV.1	LV.1			LV.2	LV.2	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1		LV.1	LV.1
언어	LV.1	LV.1	LV.1	LV.1	LV.1	LV.2	LV.1	LV.2	LV.1	LV.1	LV.2	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1
자연어 처리	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	
음성	LV.1	LV.1	LV.1	LV.1	LV.1	LV.2	LV.1	LV.2	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1
공간		LV.1	LV.1	LV.1		LV.1	LV.1	LV.1	LV.1	LV.1	LV.1		LV.1	LV.1			
기계	LV.2	LV.1	LV.1	LV.1	LV.1	LV.4	LV.2	LV.2	LV.3	LV.2	LV.2	LV.1	LV.2	LV.1	LV.1	LV.1	LV.2
지식	LV.1	LV.1	LV.1	LV.1	LV.1	LV.2	LV.2	LV.2	LV.2	LV.1	LV.2	LV.1	LV.1	LV.1	LV.1	LV.1	LV.1
DB	LV.1	LV.1	LV.1	LV.1	LV.1	LV.3	LV.3	LV.2	LV.4	LV.1	LV.2	LV.1	LV.1	LV.1	LV.1	LV.1	LV.2

● 인공지능 적용산업 현황 및 활용사례 분석

● 시사점

- 국내 인공지능 적용 17대 산업에 인공지능 기술이 다양하게 적용되고 있으나, 기술의 집중도가 부족하여 인공지능 기술의 고도화를 통해 각 산업에 기술 집중도를 높일 필요가 있음
- 언어지능, 자연어처리, 음성지능에 대해서는 해외 인공지능 플랫폼을 활용한 제한적인 기술개발을 통해 서비스가 제공되고 있어 상대적으로 집중도가 높지 않음
- 응용서비스 개발뿐만 아니라, 텐서플로우, 파이토치 등 인공지능 플랫폼 개발 및 공유 등을 통해 인공지능 기술 기반을 마련하여 다양한 산업군에의 적용을 확대할 필요가 있음

산업·사회 변화상

4차 산업혁명 핵심기술의 특징

- 4차 산업혁명의 핵심기술인 지능정보기술에 내재된 주요 특징들이 미래 사회·경제 구조 변화를 촉진



산업·사회 변화상

수확체증 가시화

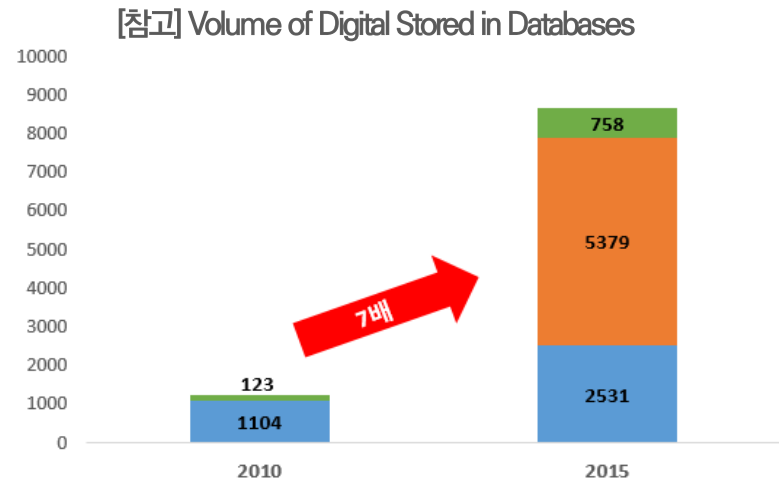
사용량 증가에 따라 성능과 효용이 지속적으로 성장하는 기술의 특성에 따라 사회적 수확체증이 가시화

- 인공지능 성능의 지수적 향상으로 인간 능력으로 해결 불가능했던 인류보편적 문제(환경·안전·보건 등) 해결: 미세먼지 저감, 재난 대응, 교통체증 해소, 질병 극복 등
- 우수 제품을 선제 출시한 기업이 최소의 비용으로 시장 지배력을 지속 확대하며 빠르게 성장
- MS의 Windows, 구글의 Android 등과 같이 글로벌 시장을 주도하는 AI 플랫폼 등장

전 세계적 데이터 급증⇒ 지능정보기술의 빠른 발전

- 전 세계 데이터량은 5년 만에 7배 증가
- 생산된 비정형 데이터 68%가 사용자에게 의해 생성

사용자 생성 데이터의 누적에 따른 인공지능 성능 향상은 추가적 사용자 유입 촉발

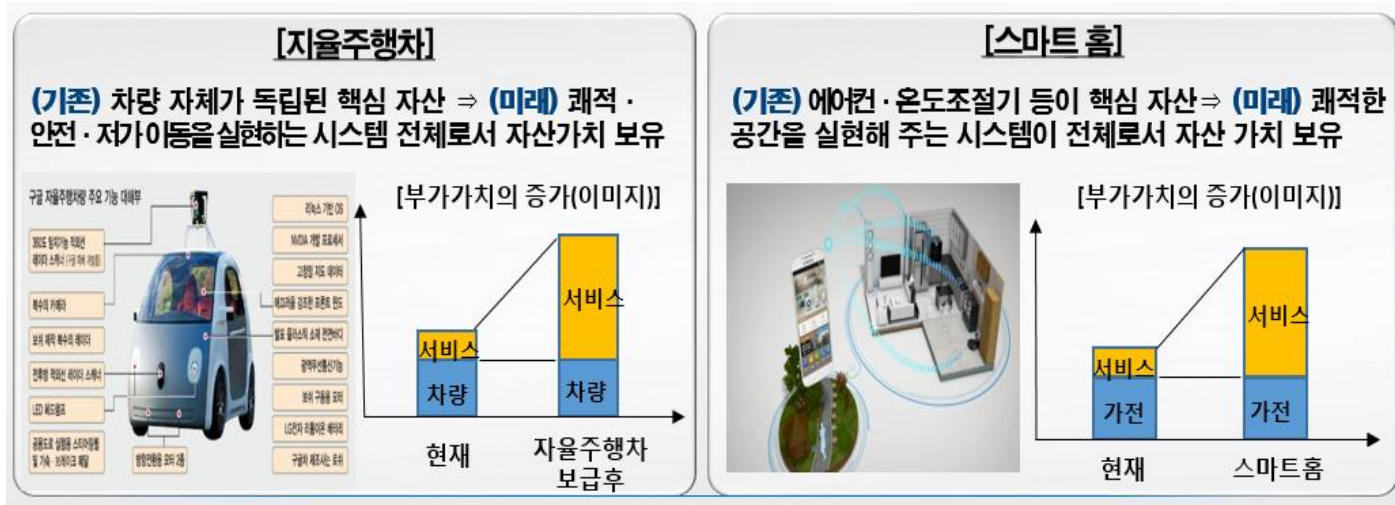


➡ 핵심기술 확보를 통해 글로벌경쟁 심화에 대응할 필요

산업·사회 변화상

제조업의 서비스화

- 제품에서 서비스가 차지하는 비중과 중요성이 높아지며 전통적인 제조업의 수익구조가 변화되는 등 제조업의 서비스화가 빠르게 진행
 - 데이터, 데이터 분석 알고리즘, 설비·노동력 작동 소프트웨어 등의 시스템이 핵심
- 개인 맞춤형 상품 생산과 서비스 제공이 보편화되어 소비자의 편익 증가
 - 소비자는 개인화된 제품에 대해 5~20% 프리미엄 지불의사, 보쉬, 파나소닉 등은 맞춤형 서비스 제공을 통해 부가가치 창출



➡ **지능정보기술의 산업활용, 기술혁신과 산업혁신간 연계 등의 전략 마련 시급**

산업·사회 변화상

- 고용 및 직업 부문의 변화
- 노동의 표준화 및 자동화의 심화 ⇒ 일부 직업 소멸, 기존 직업 변화, 새로운 직업 창출
- 지능정보기술에 의한 고용 및 직업 부문의 변화와 관련하여 낙관론과 비관론 공존

긍정적 전망

시장의 확대

* 신기술은 생산성 향상, 소득 증대, 소비 증가로 이어질 것
(Robert Atkinson, 2013)

새로운 직업 창출

* AI와 로봇틱스는 새로운 직업 창출과
인간 능력의 발현을 도울 것(Pew Trust, 2014)

보완의 관점에서 직업 변화에 대비 가능

* 노동 보완 효과로 순고용에 큰 변화 없이
점진적인 변화가 일어날 것(David Autor, 2015)

부정적 전망

로봇 등 자동화에 의한 일자리 대체 및 소멸

* 20년 내로 약 47%의 일자리가 로봇 등
자동화에 의해 대체될 것(Frey & Osborne, 2013)

일자리 소멸 숫자가 창출 숫자를 능가

* 2020년까지 AI와 로봇에 의해
일자리 710만 개 소멸, 200만 개가 창출될 것(WEF, 2016)

직업 안전성의 소멸

* 미래에는 직업 안전성의 개념은
존재하지 않을 것(Frey & Osborne, 2013)

➡ 비관적 수렴 바탕으로 하는 중립적 예측을 통해 미래 대비 교육 시스템으로의 전환 필요

산업·사회 변화상

- 사회적 역기능 발생 우려
- 지능정보기술은 편익과 효율성 극대화라는 순기능과 함께 역기능 발생 가능
- ◆사회적 양극화 발생 ◆의사결정 책임 소재 불명확성 ◆지능정보시스템 실패 등

사회적 양극화 발생

지능정보기술 유무에 따른
사회 불평등 및 정보 양극화
발생 가능

예시) 기업의 플랫폼 독점
및 개인 정보 데이터 독점,
정보 불균형으로 인한 주체
간 지위 불평등 문제 등

의사결정 책임소재의 불명확성

지능화된 기기에 의한 의사결정 보
편화로 사고 발생 시 책임 소재
문제 대두

예시) 자율주행자동차 사고, 지능화
된 투자로봇의 자율적 판단이 개입한
투자의 실패, 지능형 의료 기기의
오작동 및 오용 등

지능정보시스템의 실패

시스템 오류나 특이점
(singularity) 현상 등 지능정보
네트워크의 오작동과 실패
가능성 상존

예시) 자동화 알고리즘 오류에
의한 데이터 왜곡, 시스템 리스크
대응을 위한 국가 차원의 회복
탄력성 전략 부재 등

➡ 지능정보화 역기능 대응과 취약계층 정보격차 해소를 위한 종합계획 수립 필요



인공지능 윤리 국내외 동향 및 시사점

개관

- AI윤리는 인간과 로봇, 로봇이 매개된 인간과 인간 사회의 바람직한 관계 정립을 위해 필요한 기술규범으로, AI 알고리즘의 자율성과 인간생명의 존엄성/통제권이라는 두 축 사이에서 지속적으로 논의
- 국내외 다양한 영역에서 논의되는 AI윤리 연구 및 주요국 정책대응 현황을 분석해 우리나라 AI윤리 대응의 기본방향 정립을 위한 참고자료로 활용

■ 국내외 학계의 AI윤리 연구동향

독일 DFKI, 영국 노팅엄대, 미국 FLI, KAIST AI 연구센터, 일본 인공지능학회 등 국내외 AI 연구소 및 학계 연구자들이 개발하거나 논의한 AI윤리 설계방안 검토

■ 정부 및 의회 차원의 AI윤리 정책

EU 로봇민법, 일본 총무성 AI개발가이드라인, 미국 DARPA XAI 프로젝트 등 국내외 정부 및 의회 차원의 AI관련 입법/정책대응 현황 분석

■ 국내외 IT기업의 AI윤리 정책

IBM의 'AI윤리위원회', 카카오의 '알고리즘 윤리헌장', 구글의 'AI윤리원칙' 등 국내외 IT기업들의 인공지능 윤리 관련 정책이나 개발원칙 검토

■ 국제기구 차원의 AI윤리 대응

OECD의 'AI 발전 권고안', IEEE, 국제인권 감시기구 등 국제기구 차원의 AI 윤리대응 동향 분석

국내외 AI윤리
동향 분석

● 국내외 학계의 AI 윤리 연구동향 및 쟁점(1)

● 로봇투명성(robot transparency)

- 2017.8 영국 바스대(Bath Univ.) 로봇개발연구자들이 제안한 윤리적 지침. 기계의 오류를 언제든지 코드화해 정정하기 위해 사용자가 로봇/기계의 의도와 기능을 쉽게 파악할 수 있도록 해야 함

Robot Transparency: Improving Understanding of Intelligent Behaviour for Designers and Users

Robert H. Wortham, Andreas Theodorou, and Joanna J. Bryson

University of Bath, Bath BA2 7AY, UK,
{r.h.wortham, a.theodorou, j.j.bryson}@bath.ac.uk,
WWW home pages: <http://www.robwortham.com>,
<http://www.recklesscoding.com>, <http://www.cs.bath.ac.uk/~jjb/>

Abstract. Autonomous robots can be difficult to design and understand. Designers have difficulty decoding the behaviour of their own robots simply by observing them. A better means of robot accountability, more difficult to develop, is to

● 윤리적 블랙박스(ethical blackbox)

- 2017.8 영국 브리스톨대 로봇연구소의 앨런 윈필드(Allen Winfield) 교수팀이 제안한 AI윤리 개발원칙. 항공기의 비행데이터 기록장치와 같이 문제 발생 이전에 행해진 결정과 조치의 순서를 규명할 수 있도록 모든 AI 자율시스템 내에 설치 의무화 주장



● PERVADE 데이터 윤리 지침

- 제이크 멧칼프 등 미국 윤리학자들이 모여 AI 알고리즘의 위험을 방지하기 위해 빅데이터 연구의 윤리적 절차를 표준화하기 위한 만든 연구단체. 최근 전미국립과학재단(NSF)로부터 300만달러의 지원 받았음.



III. 인공지능 윤리 국내외 동향 및 시사점

● 국내외 학계의 AI 윤리 연구동향 및 쟁점(2)

● 알고리즘적 사회계약론

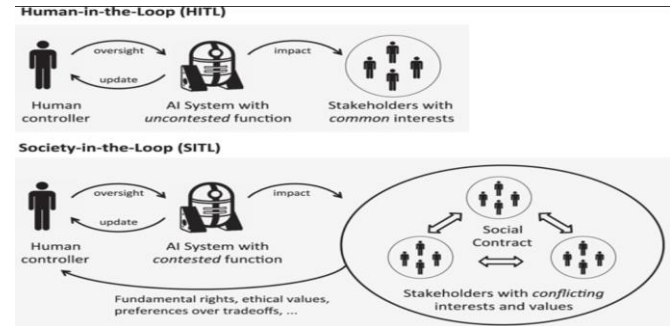
- 2018.3 이야드 라완(Iyad Rahwan) 교수가 알고리즘 윤리 규제의 개념들을 새롭게 만들기 위해 소위 'Society-in-the-loop'라는 방법론을 개발, AI알고리즘을 인간의 사회적 계약규범에 맞추어 설계할 원칙 제안

● SNS 콘텐츠 추천 알고리즘 윤리

- 2018.4 미국 USC 안넨버그 커뮤니케이션 연구소가 소셜플랫폼 상의 콘텐츠 제공자와 이용자 사이의 중개 윤리(moderation ethics)의 중요성 및 이를 기술적으로 구현하기 위한 방법론/정책 대안 제시

● 인공지능의 시민윤리

- 2018.5 벨기에 브뤼셀 싱크탱크 '리스본 카운슬'(Lisbon Council)이 알고리즘 기반 차별을 종식시키고 더 똑똑하고 좋은 세상을 만들기 위한 'AI윤리 보고서'를 제출, AI윤리 정립을 위한 정부, 시민, 개발자의 행위원칙 제안



Columbia Journalism School & TOW CENTER FOR DIGITAL JOURNALISM USC Annenberg Innovation Lab

Controlling the Conversation: The Ethics of Social Platforms and Content Moderation

April 2018

theLisboncouncil 15
think tank for the 21st century

DSM 2.0
A LISBON COUNCIL INITIATIVE

Interactive policy brief Issue 23/2018 The Ethics of Artificial Intelligence How AI Can End Discrimination and Make the World a Smarter, Better Place



By Paul Hofheinz

Paul Hofheinz is president and co-founder of The Lisbon Council, a Brussels-based think tank. Founded in 2003 as a non-profit, non-partisan association, the Lisbon Council has emerged as one of Europe's pre-eminent voices on social and economic change.

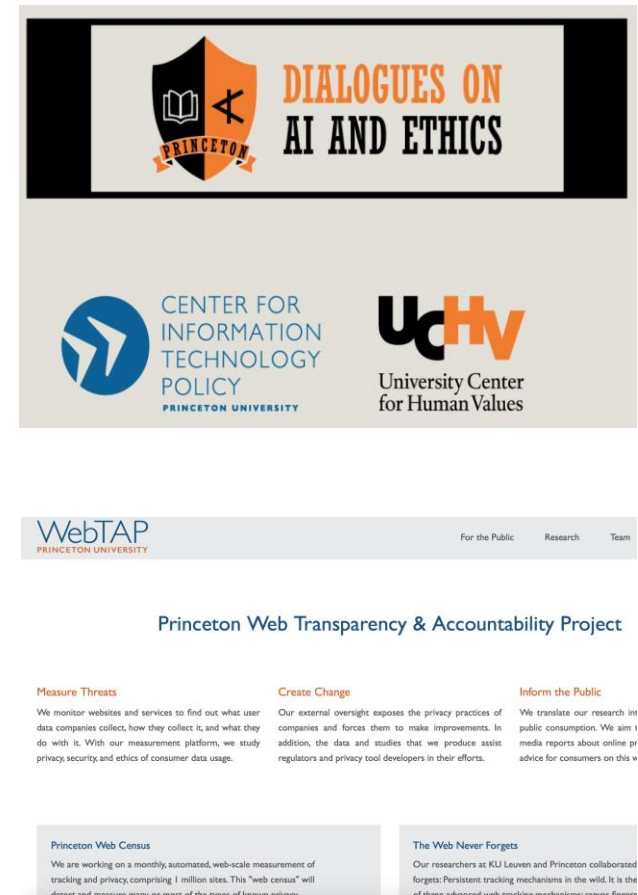
Judging from the fierce debates, few issues have posed such a challenge to mankind as artificial intelligence (AI) and machine learning. First and foremost, there is the evident, undebatable existential risk: Mankind has dominated the planet for

¹ The author would like to thank Prabhat Agarwal, Alessandro Annoni, Robert D. Atkinson, Daniel Braun, Francesco Chesca, Christian

● 국내외 학계의 AI 윤리 연구동향 및 쟁점(3)

● ‘AI와 윤리를 위한 대화’(Dialogue on AI and Ethics) 프로젝트

- 미국 프린스턴대학의 인간가치연구소(UCHV)와 정보기술정책센터(CITP)가 공동으로 인공지능 시스템의 개발이 사회에 사용되는 경우에 발생할 수 있는 윤리적 딜레마에 대한 연구를 진행, AI 윤리 프레임워크 구축을 위한 지적 분석도구를 다양하게 개발
- (AI윤리 논쟁 사례연구) 자동화된 건강 관리 프로그램(Automated Healthcare App), 동적 사운드 식별 문제 (Dynamic Sound Identification), 학교문제 해결을 위한 위한 개인정보의 사용 (Optimizing School), 법집행 챗봇(Law Enforcement Chatbots) 등
- (웹투명성과 책무성, Web Transparency & Accountability) 정보기술센터가 운영하는 알고리즘 차별효과 연구 프로젝트로, 온라인 상에 가상의 인간(남녀, 빈부, 인종 등 다양한 형태)을 봇(bot)으로 만들어 각종 사이트에 접속해 알고리즘 기반 차별(위협)을 조사 연구



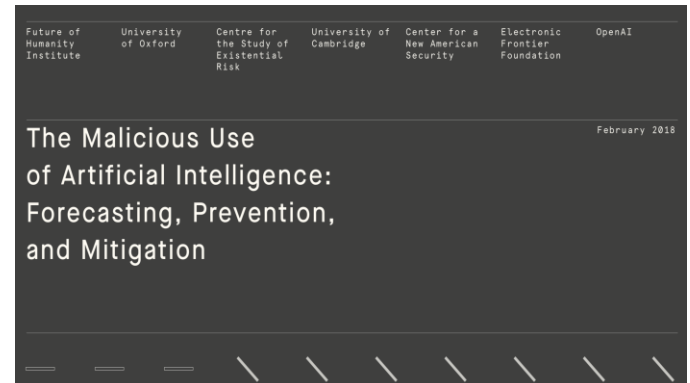
● 국내외 학계의 AI 윤리 연구동향 및 쟁점(4)

● AI윤리와 거버넌스(Ethics and Governance of AI)

- 2017 미국 하버드대 버크먼센터와 MIT가 공동으로 진행 하는 AI윤리 프로젝트로 투명성, 설명가능성 등 AI의 공적 가치를 높이기 위한 학제간 연구 지향하며, AG Tech Forum 등 산업계-학계-시민사회-정부간의 협력 거버넌스 형성도 적극 도모
- 나이트 재단, 오미디아 네트워크, 라이드 호프만(링크드인 창업자) 등이 AI윤리 연구에 2700만불 규모 펀드 결성, 윤리적 가치 등 공익(public interest)을 위한 인공지능 연구 추진

● '해악적 인공지능 보고서' (Malicious AI Report)

- 2018.2.19 영국 캠브리지대 실존적 위기 센터(The Centre for the Study of Existential Risk)가 FLI, Open AI, EFF 등과 공동으로 AI의 해악적 사용에 따른 위협을 예방,완화하기 위한 방안 제시.
- AI의 악용을 차단하고 대응하기 위해 다양한 이해관계자 간의 논의(multi-stakeholder dialogues)를 통한 공개 토론 활성화를 강조



● 국내외 학계의 AI 윤리 연구동향 및 쟁점(5)

● 영국 옥스퍼드대 미래의 삶 연구소(Future of Life Institute)

- 인공지능 위험 시나리오 예측 및 대응(Benefits & Risks of Artificial Intelligence)에 선도적 역할
- 2017 '선한 AI 콘퍼런스'(Beneficial AI Conference)를 통해 '아실로마 AI 원칙' (Asilomar AI Principles) 선언하는데 주도적 역할
- 2015 토비 월시(Toby Walsh), 스튜어트 러셀(Stuart Russell) 등이 주도한 자동화 무기 개발 방지를 촉구하는 공개서한 발표

● 일본 AI학회의 'AI윤리 가이드라인' 제정

- 2017.2 AI의 윤리적 이슈에 대한 사회적 공론화는 물론 AI윤리 가이드라인 제정 선도적 역할 수행
- 인류에의 공헌, 법규제의 준수, 프라이버시 존중, 공정성, 안전성, 성실한 행동, 사회적 책임, 사회와의 대화 및 자기계발, AI 윤리 준수 요청 등



News: AI Biotech Nuclear Climate Partner Orgs



BENEFITS & RISKS OF ARTIFICIAL INTELLIGENCE

"Everything we love about civilization is a product of intelligence, so amplifying our human intelligence with artificial intelligence has the potential of helping civilization flourish like never before - as long as we manage to keep the technology beneficial."

人工知能学会倫理委員会

TOP | 人工知能学会倫理委員会について | 活動について | トピックス | イベント | お問い合わせ

POSTS

HOME > POSTS > トピックス > 「人工知能学会 倫理指針」について

「人工知能学会 倫理指針」について

投稿日: 02/28/2017 | 最終更新日時: 05/03/2017 | カテゴリー: トピックス



人工知能学会倫理委員会
委員長 松尾豊
2017年2月28日

人工知能学会倫理委員会では、2014年の委員会設置以来、人工知能研究あるいは人工知能技術と社会との関わりを広く捉え、それを議論し考察し、社会に適切に発信していくことを進めてまいりました。我が国でも、さまざまな政府機関で人工知能と社会に関する議論が行われ、また国際的にもそうした議論が進められるなか、人工知能学会としても深い専門知識に基いて、国内の議論をリードしていく役割があると考えています。

昨年の前半には、こうした趣旨に基づき、倫理委員会内での議論を倫理綱領案という形でまとめました。委員の全開討論に付公開討論を行い、さまざまな意見をいただき、その後、ネット上でも意見を

● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(1)

- (미국) 빅데이터 분석과 머신러닝에 의한 차별(discrimination)의 위험성에 대한 사전예방 차원의 윤리적 가이드라인 제시

Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights
미 오바마 행정부 (2016. 3)

Big Data:
A Report on Algorithmic Systems,
Opportunity, and Civil Rights

May 2016



빅데이터 알고리즘이
학습 데이터의 왜곡이나
부족 또는 편견 제공으로
인한 잘못된 윤리관을
학습할 가능성이 있음

알고리즘 입력으로 사용되는 데이터

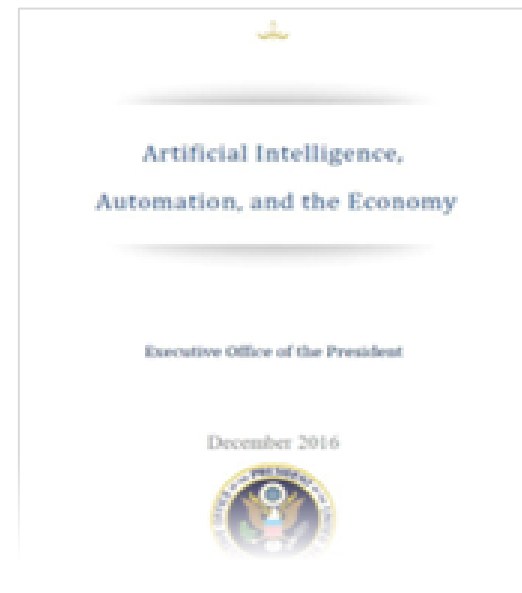
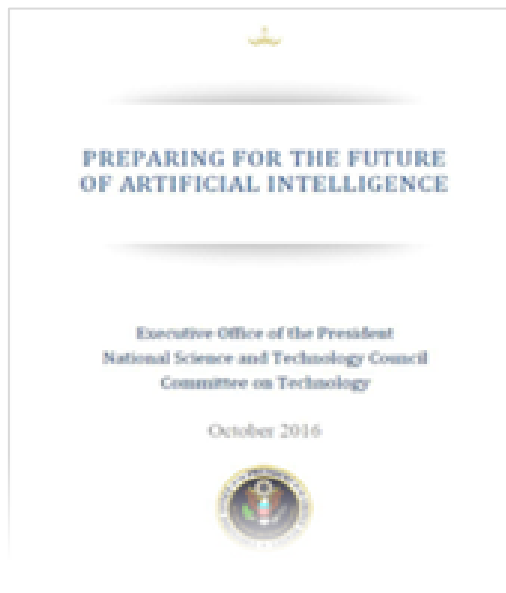
1. 부족하게 선택된 데이터
2. 불완전하고 틀린 오래된 데이터
3. 편향된 데이터
4. 고의가 아닌 영구보존 및 역사적 편향의 옹호 등의 문제점

머신러닝과 알고리즘 자체의 내부 작용

1. 부족하게 디자인된 매칭 시스템
2. 확장된 이용자 옵션 대신 좁고 개인화된 추천 서비스
3. 상관관계가 반드시 인과관계를 암시한다고 가정한 의사 결정 시스템
4. 부족한 정보 또는 부적절하게 인구를 대표하는 데이터 세트 등의 문제점

● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(2)

- (미국) AI에 대한 규제원칙 및 윤리규범 마련을 위한 사전작업의 일환으로 백악관 과학기정책부가 주요 전문가들과 워크숍 시리즈 진행, 이를 바탕으로 AI 미래전략 보고서 3부작 발표(2016)



인공지능 윤리의 사회적 공론화

국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(3)

- (미국) 뉴욕시 의회는 AI 알고리즘의 법적 규제 일환으로 알고리즘 또는 자동화된 결정 시스템의 소스 코드 공개 의무화 법안 추진

New York City Council (2017. 8.24) Automated Processing of Data for the Purposed of targeting service, penalties, or policing to pers



THE NEW YORK CITY COUNCIL
MELISSA MARK - VIVERITO, SPEAKER

LEGISLATIVE RESEARCH CENTER

Sign In

[Council Home](#) [Legislation](#) [Calendar](#) [City Council](#) [Committees](#)

[RSS](#) [Alerts](#)

Details

Reports

File #:

Int 1696-2017

Version: *

Name:

Automated processing of data for the purposes of targeting services, penalties, or policing to persons.

Type:

Introduction

Status:

Laid Over in Committee

Committee:

[Committee on Technology](#)

On agenda:

8/24/2017

Enactment date:

Law number:

Title:

A Local Law to amend the administrative code of the city of New York, in relation to automated processing of data for the purposes of targeting services, penalties, or policing to persons

Sponsors:

[James Vacca](#), [Helen K. Rosenthal](#), [Corey D. Johnson](#), [Rafael Salamanca, Jr.](#), [Vincent J. Gentile](#), [Robert E. Cornegy, Jr.](#), [Jumaane D. Williams](#)

Council Member Sponsors:

7

Summary:

This bill would require agencies that use algorithms or other automated processing methods that target services, impose penalties, or police persons to publish the source code used for such processing. It would also require agencies to accept user-submitted data sets that can be processed by the agencies' algorithms and provide the outputs to the user.

Indexes:

Council Operations , Oversight

Attachments:

1. [Summary of Int. No. 1696](#), 2. [Int. No. 1696](#), 3. [August 24, 2017 - Stated Meeting Agenda with Links to Files](#), 4. [Committee Report 10/16/17](#)

History (4)

Text

4 records

Group

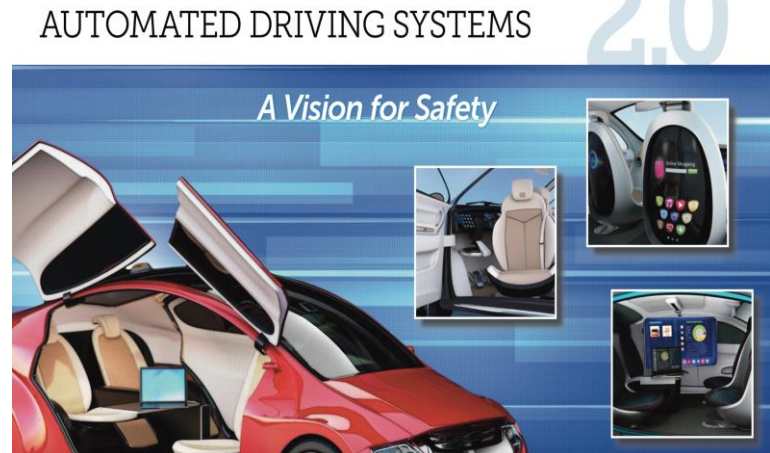
Export

Date	Ver.	Prime Sponsor	Action By	Action	Result	Action Details	Meeting Details	Multimedia
10/16/2017	*	James Vacca	Committee on Technology	Hearing Held by Committee		Action details	Meeting details	Not available
10/16/2017	*	James Vacca	Committee on Technology	Laid Over by Committee		Action details	Meeting details	Not available
8/24/2017	*	James Vacca	City Council	Referred to Comm by Council		Action details	Meeting details	Not available
8/24/2017	*	James Vacca	City Council	Introduced by Council		Action details	Meeting details	Not available

인공지능 알고리즘 투명성 강화를 위한 법적 기반 강화

● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(4)

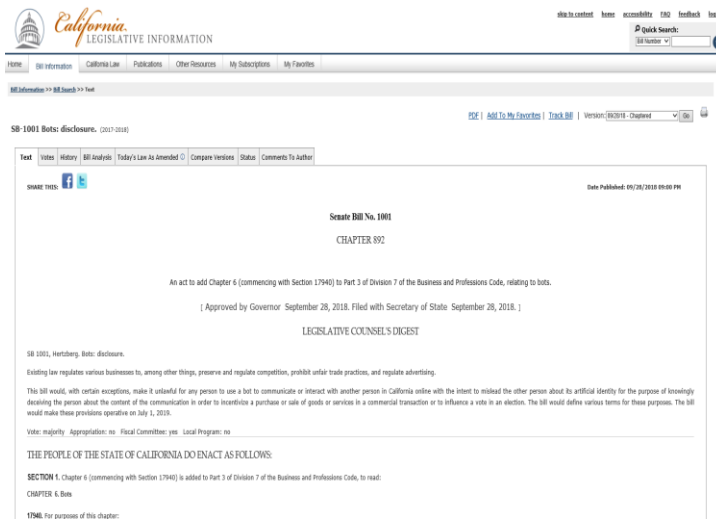
- (미국) 미국정부는 ‘자율주행 안전설계를 위한 12가지 가이드라인’ 제시
 - ‘17.9 A Vision for Safety2.0- Automated Driving Systems를 통해 자율주행시스템 안전, 운영설계 범위, 위기 상황 최소화 및 데이터 기록, 인증방법, 사이버보안, 소비자 교육 및 훈련 등에 대해 수록
 - 자율주행자동차의 성능, 주정부 규제 및 권한범위 등을 내용으로 하는 「Federal Automated Vehicle Policy-Acceleration the Next Revolution in Roadway Safety」(‘16)의 후속 조치



국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(5)

● (미국) 블레이드 러너법: 미국 최초의 로봇금지법

- ‘18.9.28 미국 캘리포니아주는 봇(bot)을 통한 허위정보 유포 등을 최소화하기 위해 ‘봇’의 존재를 알리지 않고 온라인 상의 의사소통 및 상호작용을 위해 로봇(bot)을 사용하는 것을 금지하는 법안 통과. 위반시 불공정거래법에 의해 피해구제 및 배상청구권도 포함.
- 봇(bot)이 적용되는 온라인 공간은 지난 1년간 방문자 및 사용자 수가 천만명 이상인 온라인 플랫폼, 즉 SNS 등과 같은 공개 인터넷 웹사이트, 웹응용 프로그램, 디지털 응용프로그램을 포함



● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(6)

● (EU) 로봇윤리 13원칙

- 2007년 유럽로봇연구네트워크(EURON)이 ICRA에서 발표한 ‘로봇윤리 13원칙’은 AI 기술규범의 중요한 전환점 제공

원칙					
1	인간의 존엄과 인권 Human Dignity and Human Rights	6	자주성과 개인적 책임 Autonomy and Individual Responsibility	11	사회적 책임감 Social Responsibility
2	평등, 정의와 형평성 Equality, Justice and Equity	7	고지에 입각한 동의 Informed Consent	12	이익의 공유 Sharing of Benefits
3	이익과 손해 Benefit and Harm	8	프라이버시 Privacy	13	생물권에 대한 책무 Responsibility towards the Biosphere
4	종교적 다양성과 다원성에 대한 존중 Respect for Cultural Diversity and Pluralism	9	신뢰성 Confidentiality		
5	반차별과 반낙인화 Non-Discrimination and Non-Stigmatization	10	연대와 협력 Solidarity and Cooperation		

로봇윤리 로드맵 설계

● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(7)

● (EU) 로봇규제 가이드라인 제정

- 정보화 연구개발 프로그램 FP7(7th Framework Programme)의 재정 지원 계획에 따라 로봇의 윤리적 개발 원칙을 정립하기 위해 ‘로봇법 프로젝트’ 2년간 추진, 그후 ‘로봇규제에 관한 가이드라인’ 제정

* 로봇법 프로젝트(RoboLaw Project, 2012.3~2014.3) → 로봇규제에 관한 가이드라인(Guidelines on Regulating Robotics, 2014.9)

	종류	내용
1	자율주행 차량 Self-driving cars	안전성 vs 편안함, 안전성 vs 자유, 접근성 vs 평등성, 효율성 vs 프라이버시 등 교환할 수 있는 가치라 할 수 없으므로 선호된 디자인과 정책을 만들기 위해서 각 기술적 단계마다 관련자들의 가치 분석이 필요
2	수술용 의료로봇 Computer integrated surgical systems	안전성, 자주성, 정당성과 새로운 기술, 책임, 프라이버시
3	로봇 인공기관 Robotic Prostheses	다른 문항과 같이 일정한 프레임 워크가 있지 않고, 규범적 윤리, 기술적 윤리 등 관점에서 어떻게 인공기관을 볼 것인지에 대해 논의하고 있어 제외
4	돌봄 로봇 Care Robots	안전성, 책임, 자주성, 독립성, 권한부여(enablement), 프라이버시, 사회적 유대감 새로운 기술과 정당성, 새로운 기술, 윤리, 과학적 연구

● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(8)

● (EU) 로봇윤리 법제화를 위한 기본방향 및 가이드라인 제시

- 2017.2.16 유럽의회 법사회위원회(the Committee on Legal Affairs)가 개인정보보호, 로봇윤리헌장 등 로봇의 개발 및 확산에 따라 제기되는 법적, 윤리적 문제에 대한 체계적 법제화 방향을 담은 결의안을 EU 집행위원회에 제출

EU 로봇법 규칙 초안의 주요 항목과 내용		
1	민간이용을 위한 로봇기술, 인공 지능의 개발에 관한 일반원칙	지능형 자동로봇의 개념 및 유형정의, 로봇등록제도 도입 권고, 로봇의 사회적, 윤리적 연구 및 투자 제안
2	로봇기술헌장	인간 이익 우선, 인간에 대한 해악금지, 인간의 자율성 보장, 공정한 이용을 통한 정의실현, 기본권 존중, 사전예방원칙, 포괄성 프라이버시 존중, 책임성, 가역성 등의 원칙하여 개발자와 이용자들을 위한 윤리헌장 마련을 제안
3	로봇규제기구	기술적, 윤리적, 규제적 전문성을 제공하고 다양한 기회와 도전에 적절한 대응을 위해 로봇기술, 인공지능 규제를 위한 기구(agency) 창설 요청
4	지적재산권, 개인정보보호 등 정정사항	로봇에 의해 만들어진 지적창작물의 소유기준을 구체화할 것을 요구, GDPR 원칙을 존중하는 방향에서 프라이버시 관련쟁점에 대해 규율할 것을 제안
5	표준화 등의 쟁점	유럽시장의 분열을 막고 소비자 이익을 보호하기 위해 로봇과 인공지능의 위험을 통일적으로 평가하는 기준 마련

로봇법 규칙 초안을 위한 보고서 마련

● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(9)

● (EU) GDPR의 인공지능 알고리즘 규제

- EU GDPR(개인정보보호규정)이 알고리즘 투명성과 관련해서 프로파일링 등 자동화된 개인적 의사결정에 대해 '설명을 요구할 권리'를 적극 포함



- ✓ EU GDPR 제13조 제2항 (f)호 [프로파일링관련] “제22조 제1항 및 제4항에 언급된 프로파일링을 포함한 자동화된 의사결정의 존재, 그리고 적어도 그러한 경우, 이에 사용되는 로직에 관한 의미 있는 정보와 해당 처리가 정보 주체에 대해 갖는 중요성과 예상 결과”를 정보주체에게 추가적으로 제공해야 한다고 규정
- ✓ 데이터주체는 알고리즘에 의해 자동화된 결정에 종속되지 않고 그 결정과정에 대한 설명과 반론(contestable)을 제기할 권리가 있음을 천명

● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(10)

● (EU) AI에 관한 EU협력선언

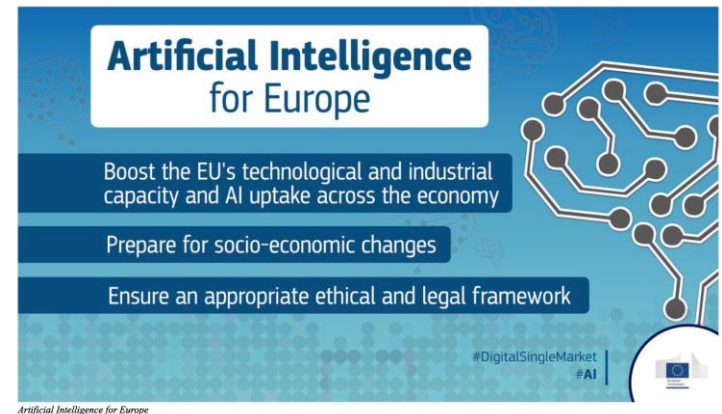
- '18.4.10 EU 집행위원회가 '디지털 데이'를 기념해 AI에 대한 유럽국가들의 기본적 합의를 담은 '인공지능에 관한 EU 협력 선언'(EU Declaration on Cooperation on Artificial Intelligence)을 발표



● (EU) 유럽을 위한 인공지능 정책 (Artificial Intelligence for Europe)

- '18.4.25 '인공지능에 관한 EU 협력선언'의 취지를 구체화하여 '유럽을 위한 인공지능'(Artificial Intelligence for Europe) 정책을 발표, AI가 제공하는 새로운 도전에 대처하는 동시에 그 기회를 최대한 활용하기 위한 본격적인 실행방안을 포함.

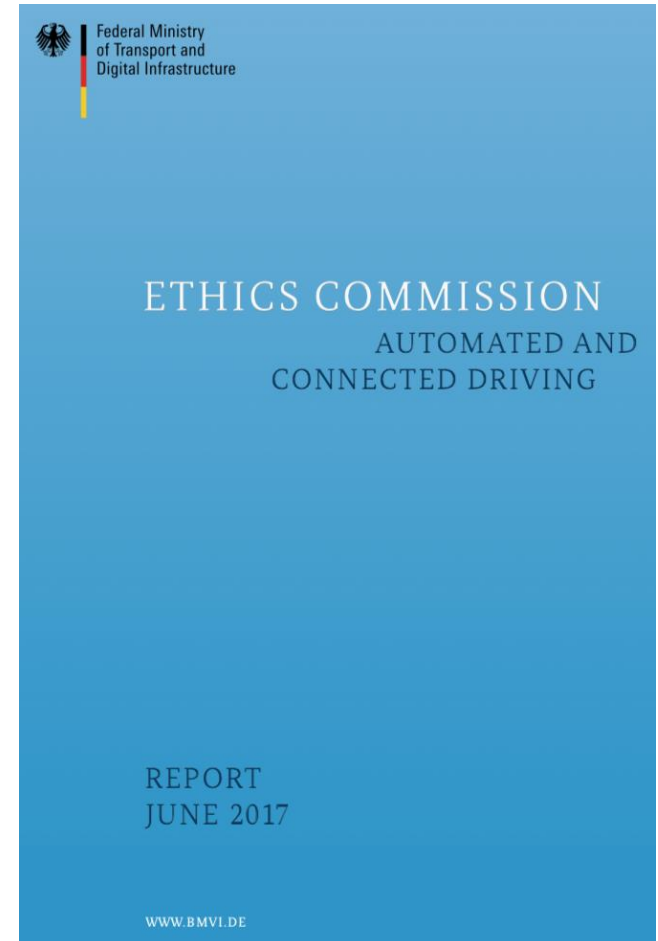
Artificial intelligence: Commission outlines a European approach to boost investment and set ethical guidelines



● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(11)

● (독일) 자율주행자동차 윤리가이드라인 발표

- 2017.8 독일 교통과 디지털인프라스트럭처부 산하 '자동화와 커넥티드 자동화윤리위원회'가 유럽 최초로 자율주행차에 대한 15개 항목의 윤리지침 발표
- 자율주행자동차 기술 개발시 자율성 원칙, 생명보호, 피해 최소화 등과 관련한 내용 수록
- 시스템의 무결성을 보장하고 향후 필요한 활동에 대한 7가지 조치계획 제시 (*The Federal Government's Action Plan, '17.9)
- 1) 도로교통법검토및개정, 2) 데이터보호 · 전달 및 사용여부 결정권(데이터주권), 3)학습 및 자가학습시스템, 4) 딜레마문제에 관한 검토 및 논의와 프로세스 모니터링 요건 마련, 5) 대중인식 제고를 위한 방안, 6) 자동화 및 연결시스템 국제표준화, 7) 규제프레임 개발



● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(12)

- (프랑스) 프랑스 개인정보규제기관 CNIL에서 인공지능의 윤리적 문제를 본격적으로 다루기 위해 정부 차원의 인공지능 윤리이슈 보고서 발간
 - 알고리즘 책임성, 알고리즘 차별 등 6가지의 주요 AI 윤리 이슈에 대한 정책대응 필요성 제기(2017.12)
1. Autonomous machines: a threat to free will and responsibility?
 2. Biases, discrimination and exclusion
 3. Algorithmic profiling: personalization versus collective benefits
 4. Preventing massive files while enhancing AI: seeking a new balance
 5. Quality, quantity, relevance: the challenge of data curated for AI
 6. Human identity before the challenge of artificial intelligence
- ‘의미있는 인공지능을 위하여’ (For a Meaningful Artificial Intelligence) 보고서 발간(‘18.3) 하여 블랙박스 해부, 윤리적 설계, 윤리 거버넌스 등 AI윤리 권고사항을 제시



● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(13)

- (영국) ‘영국의 인공지능: 준비, 의향, 역량’ (AI in the UK: ready, willing and able?) 보고서
 - ’18.4 AI 산업을 육성해 데이터 기반 혁신의 선도국이 되겠다는 비전 하에 영국 상원 인공지능 특별위원회가 마련한 AI 윤리 보고서
 - AI가 주는 기회를 최대한 활용해야 하지만 AI의 개발과 사용과정에서 발생하는 위험 요소와 오용의 문제에 대해 윤리 원칙을 채택해야 한다는 2018년 9월 세계경제포럼(World Economic Forum)과 시민을 위한 공공분야의 인공지능 사용 가이드라인을 공동 설계
 - 보고서가 권장하는 인공지능 규범(AI code)의 5개 원칙은 1) 인공지능은 인류의 공동 이익과 이익을 위해 개발되어야 한다. 2) 인공지능은 투명성과 공정성의 원칙에 따라 작동해야 한다. 3) 인공지능이 개인, 가족, 지역 사회의 데이터 권리 또는 개인정보를 감소시켜서는 안된다. 4) 모든 시민은 인공지능을 통해서 정신적, 정서적, 경제적 번영을 누리도록 교육받을 권리를 가져야 한다. 5) 인간을 해치거나 파괴하거나 속이는 자율적 힘을 인공지능에 절대로 부여해서는 안된다.



HOUSE OF LORDS

Select Committee on Artificial Intelligence

Report of Session 2017–19

AI in the UK: ready, willing and able?

Ordered to be printed 13 March 2018 and published 16 April 2018

Published by the Authority of the House of Lords

AI Code

One of the recommendations of the report is for a cross-sector AI Code to be established, which can be adopted nationally, and internationally. The Committee's suggested five principles for such a code are:

1. Artificial intelligence should be developed for the common good and benefit of humanity.
2. Artificial intelligence should operate on principles of intelligibility and fairness.
3. Artificial intelligence should not be used to diminish the data rights or privacy of individuals, families or communities.
4. All citizens should have the right to be educated to enable them to flourish mentally, emotionally and economically alongside artificial intelligence.
5. The autonomous power to hurt, destroy or deceive human beings should never be vested in artificial intelligence.

● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(14)

● (일본) 인공지능과 인간사회에 관한 간담회 개최

- ‘17.3.24 일본 내각부는 AI와 인간사회에 관한 간담회를 통해 AI가 인간사회에 미치는 영향과 관련해 제기되는 주요 논점 및 사례들을 통해 AI의 규범적 논의를 정리

이슈영역	인공지능과 인간사회에 관한 간담회의 주요 논점 내용
윤리적 논점	-인공지능 기술의 발전에 따른 인간과 인공지능 기술/기계의 관계성 및 윤리관 변화 양상 -인공지능 기술에 의해 무의식적으로 감정, 신조, 행동이 조작당하거나 차별화될 가능성 -능력이나 감정을 포함한 인간관의 재인식 -인공지능 기술이 관여하는 행위/창조활동에 대한 가치/평가의 수용성, 가치관 및 인식의 다양성
법적 논점	-인공지능 기술에 의한 사고 등의 책임 분배의 명확화와 보험제도의 정비 -개인정보와 프라이버시 보호를 포함한 빅데이터 이용 및 활용 -인공지능 기술을 활용한 창작물 등의 권리 검토 -법 해석, 법 개정, 법과 관련된 기본적 개념의 재검토 가능성
경제적 논점	-인공지능 기술에 의한 업무나 일하는 방식의 변화(개인 대상) -인공지능 기술의 이용 및 활용에 따른 고용과 기업의 변화(기업 대상) -인공지능 기술의 이용 및 활용을 촉진하기 위한 경제정책, 노동이동을 가능케 하는 교육정책/고용정책(국가 대상)
교육적 논점	-인공지능 기술을 적절히 이용 및 활용하기 위한 교육 -인간에게 있어 본질적인 능력 및 인간에게만 가능한 능력의 육성

국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(15)

○ (일본) 총무성의 AI개발 가이드라인

- 총무성 산하 AI네트워크사회추진회의 사무국에서 전문가 의견수렴을 통해 일본 정부 차원의 본격적인 AI개발 윤리가이드라인 발표(2016.12.28)
- 가이드라인의 목적과 기본이념 외에 협력, 투명성, 이용자 지원, 제어가능성, 보안확보, 안전보호, 개인정보보호, 윤리, 책임 등 인공지능 연구개발의 주요 원칙들을 제시함

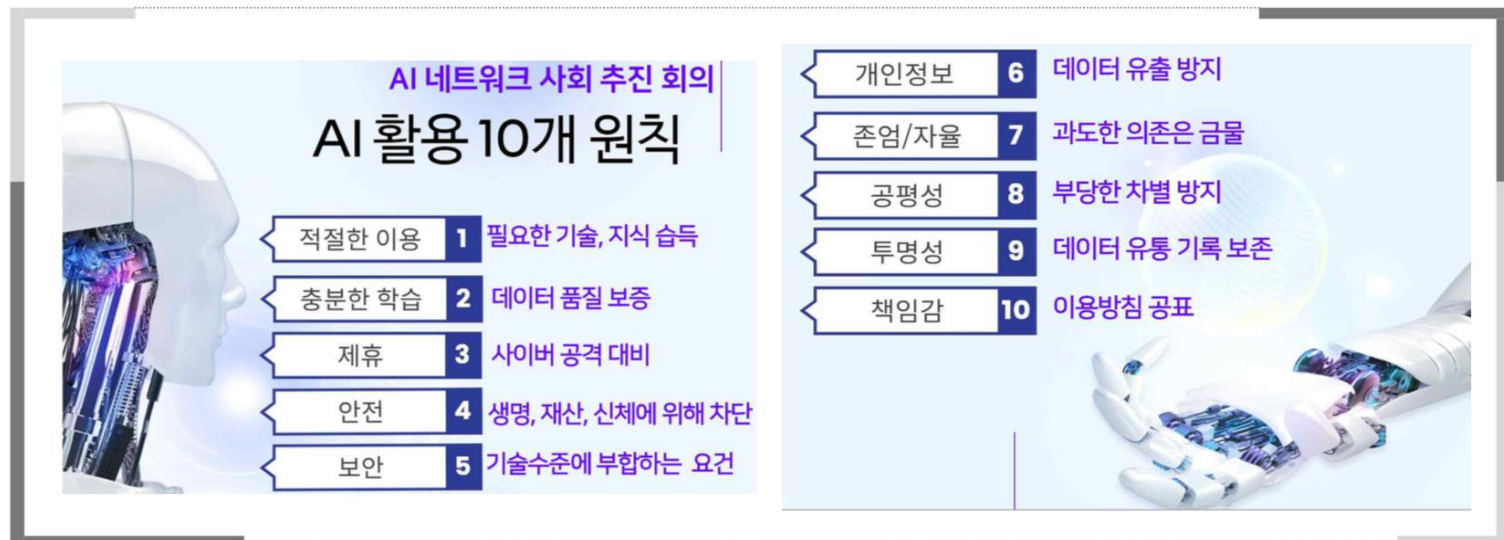
<표> 일본 AI 개발 가이드라인 개발원칙(2017)

개발 원칙	주요 내용
① 연계의 원칙	o개발자는 AI시스템의 상호접속성과 상호운용성에 유의한다. - 상호접속성·상호운용성 관련 정보 공유, 국제 표준 준용, 데이터 형식 표준화, API 포함 인터페이스 또는 프로토콜 오픈화 대응 노력, IP 라이선스 계약 공정화
② 투명성의 원칙	o개발자는 AI시스템 입출력의 검증 가능성과 판단 결과의 설명 가능성에 유의한다. - 제3자 생명, 신체, 자유, 프라이버시, 재산 등의 영향에 유의
③ 제어 가능성의 원칙	o개발자는 AI시스템의 제어 가능성에 유의한다. - 위험가능성 평가 위해 실험실 내부 또는 샌드박스 활용 고려, 다른 AI에 의한 감독 및 대처 가능여부 검토
④ 안전의 원칙	o개발자는 AI시스템이 액추에이터 등을 통해 이용자와 제3자의 생명·신체 재산에 위해를 미치는 것이 없게 배려한다. - 국제 표준 참조, 학습에 의한 프로그램의 변화 가능성 유의(사전 검증 노력/R&D단계 고려/이해관계자에 설계 취지 설명)
⑤ 신뢰성의 원칙	o개발자는 AI시스템의 안전에 유의한다. - AI 시스템의 신뢰성·강건성 유의, 사전 검증 타당성 확인, Security by Design 설계

⑥ 프라이버시의 원칙	o개발자는 AI시스템에 의해 이용자와 제3자의 프라이버시가 침해되지 않도록 배려한다. - 프라이버시 영향평가 실시, Privacy by Design 설계
⑦ 윤리의 원칙	o개발자는 AI시스템 개발에 있어서 인간의 존엄과 개인의 자율을 존중한다. - 생명윤리 관련 논의 참조, 학습데이터에 편견 등 부당 차별 금지, 국제인권법 또는 국제인도법 근거로 인간가치 훼손금지
⑧ 이용자 지원의 원칙	o개발자는 AI시스템이 이용자를 지원해 이용자에게 선택의 기회를 적절히 제공하는 것이 가능하도록 배려한다. - 이용자 조작 용이 노력, 적시 기능 제공, 사회적 약자의 이용성 확보, 학습에 의한 시스템 변화 관련 정보 제공
⑨ 책임의 원칙	o개발자는 이용자를 포함한 이해관계자에 대해 책임을 완수하도록 노력한다. - 개발한 AI 시스템에 대한 책임완수, 사회적 수용성 향성 위해 정보제공 및 의견 청취 등 이해관계자의 적극적인 피드백 강구, 개발 AI시스템에 대한 정보 공유 협력

● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(16)

- (일본) AI 연구 및 개발에 관한 국제규칙 제정 등 AI윤리 아젠다의 글로벌 선도 노력
 - '16.429-30 일본 G7 정보통신장관 회의에서 AI 연구개발 원칙 제정 및 산학관 회의 창설 제안하는 등 AI연구개발 8원칙 발표
 - 그 이후 AI가 가져올 미래사회 모습을 '충분한 학습', '보안', '공평성', '투명성' 등을 강조한 'AI활용 10개 원칙' 도출, G20 디지털경제 장관회의(2018.8), OECD 디지털경제정책위원회(2018.11) 국제 표준으로 제시 예정



자료 : 일본경제신문, '18.7.12

● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(17)

○ (한국) 지능정보사회 관련 법제 논의 과잉 속 AI윤리 논의 파편화

- 지능정보사회 중장기 종합대책(2016.6.12) 이후 다수의 지능정보사회 관련 법안이 발의, 그외 SW안전 및 프라이버시 관련 법제 등 개별 법률 개선에 관한 논의도 전개
- 지능정보사회 기본법안(2.23), 디지털기반산업기본법안(3.7), 제4차 산업혁명촉진 기본법안(3.30) 등 기본법 모색, ICT특별법(3.9), SW진흥법 등 개별법제 개정을 통한 산업진흥 모색 등
- 그러나 기존 ICT관련 법률이 너무 많고 부처별 산재해 있는 상황에서 지능정보사회 관련 법제이슈의 범람은 새로운 규범체계의 사회적 합의를 더욱 어렵게 만들 수 있음

법률구분	주요내용	
헌법	- 차별 등 인권의 문제 (예: 젠더, 트랜스휴먼 단계, 포스트휴먼) - 알고리즘을 통한 표현의 자유 인정여부	- 인간의 존엄 (예: 뉴럴링크) - 인간의 입헌주의와 인공지능의 입헌주의
저작권	- 인공지능 학습과정에서의 대규모 저작권 침해	- 비인간행위자의 저작권보호
개인정보보호 법제	- 자동화 프로파일링의 증가 대응 필요 - 식별 정보가 아닌 프라이버시의 법적 보호방안	- 자동화된 동의방식 도입
민사 법제	- 인공에이전트와 인간의 합리성 격차 - (SW)제조물책임	- 스마트계약(자동화)과 의사표시
형사 법제	- 인공지능의 형벌능력(cf: 법인)	- 인과관계 인정방식

● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(18)

● (한국) 지능형 로봇윤리 현장의 제정 명문화

- 산업통산자원부의 '지능형 로봇개발 및 보급 촉진법'에서 로봇윤리현장의 제정(18조)을 명문화

- ✓ 민주당 박정의원 대표발의 '지능형 로봇개발 및 보급촉진법' 개정 추진(현재)
- ✓ 민주당 박영선 의원 '로봇기본법안' 대표 발의(2017. 7.19)

세부과제	조치사항	일정	소관
인공지능 안전성 관련 연구 및 제도 정비	지능정보SW, 데이터 안전성 관련 제도 연구	'17. 12월	과기정통부
	자율주행 인증제도 정비	'20년	국토부
인공지능 관련 법적책임 연구	인공지능 관련 손해배상 법제 전반 연구	'17. 12월	과기정통부
인공지능 윤리 관련 연구	지능정보윤리현장 제정방향 수립	'17. 12월	과기정통부
데이터 재산권 관련 연구	데이터 재산권 권리보호 관련 연구	'17. 12월	과기정통부 문체부
지능정보기술 산출물 관련 지식 재산권법 연구	지식재산권 법제 정비 방안 연구	'17. 12월	과기정통부 문체부
지능정보사회 기본법 제정 추진	지능정보사회 기본법제정 방안 연구	'17. 12월	과기정통부

(표) 지능정보기술 윤리 관련 정책과제 추진 현황

● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(19)

○ (한국) 지능정보사회 윤리가이드라인 발표 (2017.12)

- 공공성, 책무성, 통제성, 투명성이라는 공통원칙을 선정하고 각 원칙에 대해 주체별(개발자, 공급자, 이용자) 세부지침 제시

<표> 지능정보사회 윤리가이드라인의 주요 내용(2017)

공통원칙	개발자	공급자	이용자
공공성	- 기술개발 시 사회적 차별요소 배제 - 사회적 약자 보호를 위한 접근성 보장 - 기술 공공성 확보 노력	- 공공이익 부합 제품 공급 - 선한 의도로 발주 - 상업적 이익과 공공적 기여 사이의 조화	- 악의적 이용 금지 - 공공이익 위한 제품 개선 참여 - 소비자 행동원칙 준수 일상화
책무성	- 책임의 공유 - 개발자 간 정보교류 및 기술갱신 참여 - 품질 인증기준 충족 - 윤리 절차 이행	- 책임·보상 원칙 마련 - 이용자 권리 보장 - 서비스 자율 의사결정 조건·범위 확립 - 위험 예방을 위한 사회적 공론화 참여	- 이용자 윤리책임 숙지 - 개발자·공급자에게 책임을 제기할 권리 - 안전 관련 정보공유와 제도화 요구 권리 - 이용 제품 교체·갱신·폐기시 지침 준수
통제성	- 예외적 상황 검토 - 지속적 품질관리 - 기술적 제어장치 마련	- 유통과정 위험 통제 - 안전성 검증 및 통제 조치 마련 - 이용자 선택가능성 보장 - 개발자의 자의적 기술 개발 억제방안 마련	- 정확한 이용방법 숙지 - 기술 이용역량 강화 - 자의적 조작 금지
투명성	- 위급상황 시 필요 데이터 제공 - 은닉 기능 개발 금지 - 위험 예측 및 공급자와의 공유	- 위험 정보를 이용자에게 고지·공유 - 이용자 대상 교육 및 매뉴얼 제공 - 이용자 정보의 부당한 이용 금지 - 사회적 영향평가 반영	- 설명을 요구할 권리 - 개인정보 활용에 대한 감시 의무 - 소비자 정보 공유 의무

● 국내외 정부 및 의회 차원의 AI 윤리 정책동향 및 쟁점(20)

● 지능정보사회 윤리헌장 및 윤리법제화

- 4차산업혁명 시대를 대비한 '지능정보사회 윤리헌장'을 발표(2018.5)
- 과기정통부 소관 법률인 '국가정보화기본법' 전부 개정안에서 '지능정보사회 윤리규정' 포함(2018.2)

<국가정보화 기본법 전부개정법률 현재(안)>

제63조(지능정보사회윤리) ① 국가기관과 지방자치단체는 지능정보기술의 개발·활용과 지능정보서비스의 제공·이용에 있어서 인간의 존엄과 가치를 존중하고 투명성·책임성·공공성·제어가능성 등 지능정보기술의 안전하고 신뢰할 수 있는 사용에 관한 윤리원칙을 담은 지능정보사회윤리 시책을 마련하여야 한다.

② 정부는 지능정보기술을 개발 또는 활용하는 자, 지능정보서비스를 제공 또는 이용하는 자가 인간의 존엄과 가치를 훼손하거나 기본적인 지능정보사회윤리를 침해하지 않도록 인간존엄 윤리교육·홍보 및 제도개선을 적극적으로 추진하여야 한다.

③ 정부는 제2항에 따른 윤리교육·홍보를 위하여 학교교육, 평생교육과 언론·인터넷 등을 통한 홍보 등에 사용하는 프로그램 및 콘텐츠를 개발하여 보급하여야 한다.

④ 과학기술정보통신부장관은 제67조에 따른 지능정보사회 신뢰위원회의 연구·자문을 받아 지능정보사회윤리준칙을 제정하여 보급할 수 있다.

● 국내외 ICT기업의 AI 윤리 정책동향 및 쟁점(1)

- (MS) A ether: MS 인공지능 윤리위원회
 - MS는 오래전부터 연구윤리위원회를 운영해왔으나 2016년말부터 인공지능에 특화된 연구윤리위원회를 구성, Partnership on AI와 연계
- (IBM) 인공지능 윤리원칙
 - IBM은 기계학습 등 인공지능 개발원칙을 1) Fundamentally sound, 2) Assessable, 3) Inclusive, 4) Reversible로 규정하고, 'FAIR'로 명명



IBM

Watson Stories About Developers ▾ Products and Services Use cases ▾

AI for the Enterprise

Building trust in the age of AI –
How businesses can build
fairness into their machine
learning models

● 국내외 ICT기업의 AI 윤리 정책동향 및 쟁점(2)

● (구글) DeepMind Ethics & Society Principles

- 딥마인드는 자체 AI 윤리기준을 5가지 원칙으로 발표(2017. 10.3)
 1. 사회적 혜택(Social benefit)
 2. 연구의 엄격함과 논거 중심(Rigorous and evidence-based)
 3. 투명함과 공개(Transparent and open)
 4. 다양성과 다학제적 접근(Diverse and interdisciplinary)
 5. 포괄적 협력(Collaborative and inclusive)

● (구글) Google AI Ethics 7개 원칙(2018.6.7)

1. 사회적으로 유익
2. 불공정한 편견 금지
3. 안전을 위해 제작되고 검증
4. 적절한 인간의 지시와 통제를 받으며, 그 책임은 사람들에게 있음
5. 프라이버시 디자인 원칙 하에 투명성과 제어 제공
6. 높은 수준의 과학적 우수성
7. 악의적인 애플리케이션 제한함



Home Research Fellows Partners FAQs Blog Contact

DeepMind Ethics & Society Principles

Understanding the social and ethical implications of a field as complex as AI will only be possible through world-class scientific research and the inclusion of many voices. DeepMind Ethics & Society is governed by five Principles that seek to guarantee the rigour, transparency and social

AI at Google: our principles



Sundar Pichai
CEO

Published Jun 7, 2018

At its heart, AI is computer programming that learns and adapts. It can't solve every problem, but its potential to improve our lives is profound. At Google, we use AI to make products more useful—from email that's spam-free and [easier to compose](#), to a digital assistant you can [speak to naturally](#), to photos that [pop the fun stuff out](#) for you to enjoy.

Beyond our products, we're using AI to help people tackle urgent problems. A pair of high school students are building AI-powered sensors to [predict the risk of wildfires](#). Farmers are using it to monitor the [health of their herds](#). Doctors are starting to use AI to help [diagnose cancer](#) and [prevent blindness](#). These clear benefits are why Google invests heavily in AI research and development, and makes AI technologies

● 국내외 ICT기업의 AI 윤리 정책동향 및 쟁점(3)

● Partnership on AI

- 2016. 6 (애플을 제외한) 구글,아마존,페이스북,IBM, MS 등이 인공지능의 윤리적 책임성 및 공공선 구현을 위해 파트너십 결성
- 그이후 애플(2017.1), 맥킨지, 인텔, 세일즈포스, SAP, 소니 등 기업 외 비영리단체(앨런AI연구소, EFF, UNICEF 등)도 참여

+ 1. SAFETY-CRITICAL AI

+ 2. FAIR, TRANSPARENT, AND ACCOUNTABLE AI

+ 3. COLLABORATIONS BETWEEN PEOPLE AND AI SYSTEMS

+ 4. AI, LABOR, AND THE ECONOMY

+ 5. SOCIAL AND SOCIETAL INFLUENCES OF AI

+ 6. AI AND SOCIAL GOOD

+ 7. SPECIAL INITIATIVES



amazon.com



DeepMind Google

facebook

IBM

Microsoft

● 국내외 ICT기업의 AI 윤리 정책동향 및 쟁점(4)

● (한국) 카카오 알고리즘 윤리헌장

- 2018.1월말 국내 인터넷기업으로는 처음으로 알고리즘의 개발 및 운영방식이 사회가 요구하는 윤리기준에 부합하도록 노력한다는 선언 천명
- 사회적 차별금지, 데이터윤리, 독립성, 설명의무 등 인공지능 윤리원칙 강조

카카오 알고리즘 윤리 헌장

1 카카오 알고리즘의 기본 원칙

카카오는 알고리즘과 관련된 모든 노력을 우리 사회 윤리 안에서 다하며, 이를 통해 인류의 편익과 행복을 추구한다.

2 차별에 대한 경계

알고리즘 결과에서 의도적인 사회적 차별이 일어나지 않도록 경계한다.

3 학습 데이터 운영

알고리즘에 입력되는 학습 데이터를 사회 윤리에 근거하여 수집, 분석, 활용한다.

4 알고리즘의 독립성

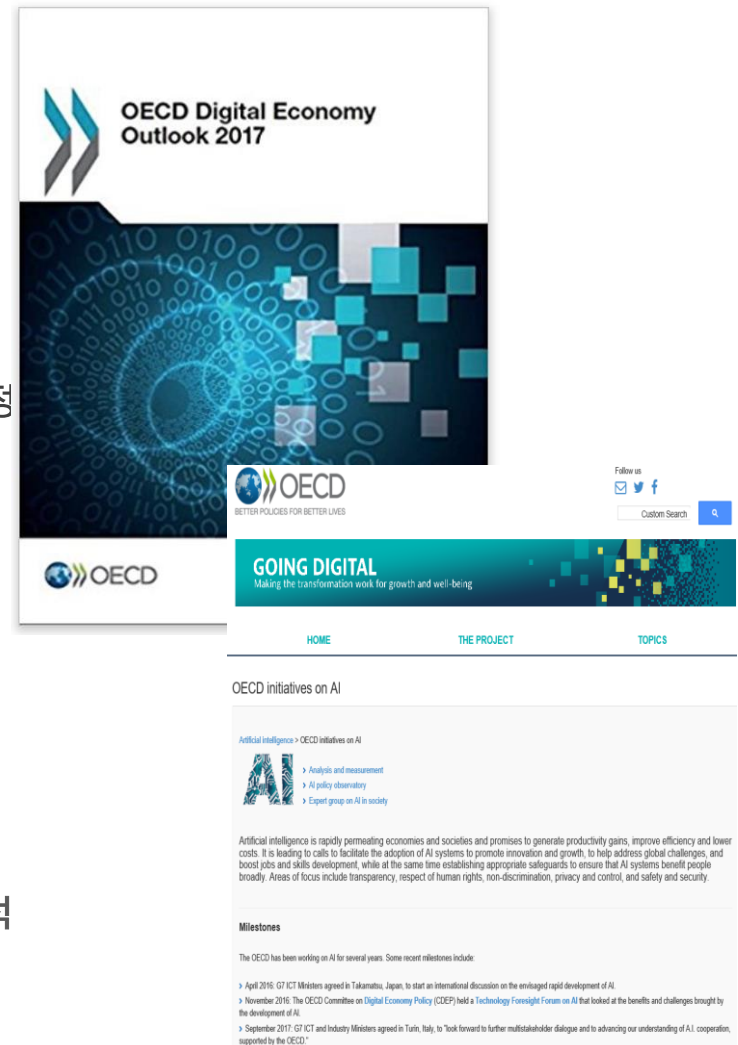
알고리즘이 누군가에 의해 자의적으로 훼손되거나 영향받는 일이 없도록 엄정하게 관리한다.

5 알고리즘에 대한 설명

이용자와의 신뢰 관계를 위해 기업 경쟁력을 훼손하지 않는 범위 내에서 알고리즘에 대해 성실하게 설명한다.

● 국제기구 차원의 AI 윤리 정책동향 및 쟁점(1)

- (OECD) Digital Economy Outlook 2017 보고서
 - 제7장 '기술전망'에서 인공지능 알고리즘의 영향에 대비할 수 있는 규범적 대응 원칙 및 정책방향 제시
 - 최근 미국, 중국, 일본, 한국, 프랑스, 영국, 독일 및 G7 ICT장관회의(2016.4) 등 주요국 AI윤리 대응현황에 주목
 - 알고리즘 차별 등 인간에 영향 미칠 수 있는 알고리즘 의사결정 문제에 대해 투명성, 공정성, 책무성의 가치 개입 강조
- (OECD) 디지털경제위원회 AI원칙 권고안 준비
 - '16년부터 AI 개발과 윤리 원칙에 대한 권고안 마련을 위해 전문가 그룹(AIGO) 구성, '19년 5월 정식 발표를 목표로 이해관계자 참여 토론 지속 추진중
 - '19년 시작 예정인 OECD AI Observatory에서 AI의 경제사회적 영향 분석 및 정보 공유를 통해 AI윤리 원칙 구체화



● 국제기구 차원의 AI 윤리 정책동향 및 쟁점(2)

● AI for Good Global Summit

- 2017년 6월 ITU(ICT4SDG)와 XPRIZE 재단(IBM Watson AI XPRIZE) 공동 주최, 몬트리올대 조슈아 벤지오 교수, 뉴욕대의 게리 마커스 교수 등이 인공지능을 통한 불평등 해소 등 인공지능을 통한 지구적 문제 해결을 위한 국제 협력 강조



● 국제기구 차원의 AI 윤리 정책동향 및 쟁점(3)

- (IEEE) 윤리적으로 조율된 설계(Ethically Aligned Design)
 - '16.12.13국제전기전자기술자협회(IEEE)가 인공지능 및 자동시스템 개발시 윤리적 측면을 고려한 설계가 필요함을 강조한 보고서
 - 안전한 AI 확보를 위한 정책방안 수립시 '안전성 입증', '투명성 제고', '제어·통제권 확보' 관점에서 접근하되, AI 개발 관련 이해관계자들의 교육 및 훈련 통한 윤리적 고려사항의 우선순위 확보 중요성도 강조



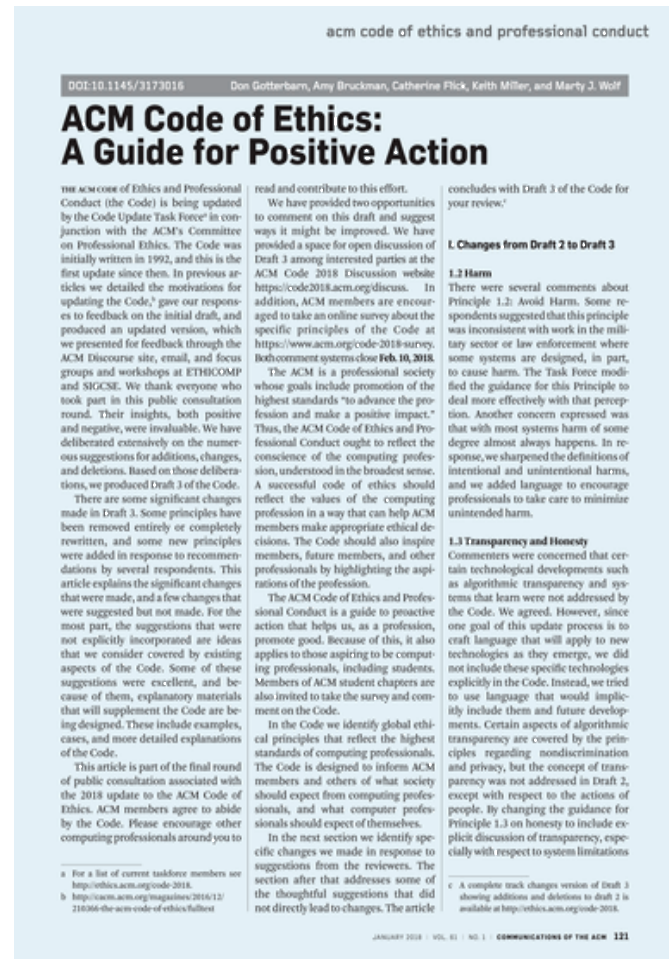
〈'윤리적으로 조율된 설계'의 주요 내용〉

- 인공지능(AI)과 자율시스템(AS) 등과 같은 테크놀로지들은 우리의 도덕적 가치와 윤리 원칙의 관점에서 인간과 조화를 이루어야 한다. 인공지능 및 자율시스템은 기능적 목표 달성과 기술적 과제해결을 넘어서 사람들에게 도움이 되어야 한다. 이는 인간과 기술 간의 높은 신뢰수준을 구축할 것이며, 일상생활에 널리 보급되어 있는 인공지능 및 자율시스템의 사용을 더 풍요롭게 할 것이다.
- 사용자와 사회가치에 부합하도록 인공지능 및 자율시스템을 설계함으로써 인간복지의 증대를 알고리즘 시대의 발전지표로 설정하는 것이 가능하다.
- 인간권리의 최고이념을 구현한다. 인류와 자연환경의 최대이익을 우선시한다. 인공지능 및 자율시스템이 사회기술 시스템으로 진화 하는데 따른 위험요소들과 부작용들을 경감한다.
- 위원회의 목표는 이러한 원칙, 쟁점 및 권고사항이 인공지능 및 자율시스템 설계와 관련한 새로운 윤리적 거버넌스 안에서 미래규범과 표준을 지원하는 것이다.
- 현대의 인공지능 및 자율시스템 관련 기관들은 인간의 복지, 권한 및 자유가 인공지능 및 자율시스템 개발에서 핵심이 되도록 만전을 기해야 한다. 이런 야심찬 목표달성을 위해 '윤리적 연구 및 설계 가이드 방법론' 위원회는 기본쟁점들과 권고사항들을 제안한다. 이런 제안을 통해 마치 세계인권선언에서 '인권'이 정의되어 나온 것처럼 시스템 설계 방법론에서 '인간 가치'가 창출되도록 할 것이다. 가치지향 설계 방법론은 윤리지침에 입각한 인간의 진보를 이루고, 인공지능 및 자율시스템 관련 기관들에 있어서 주요 안건이 되어야한다. 기계는 인간에게 봉사해야 하며, 거꾸로 되어서는 안된다. 윤리적으로 건전한 접근은 인공지능의 경제적 영향력과 사회적 영향력 간의 균형을 유지하도록 돕는다.

● 국제기구 차원의 AI 윤리 정책동향 및 쟁점(4)

● 전미 컴퓨터학회(ACM)의 AI윤리 강령

- 알고리즘 투명성과 책무성에 대한 성명 발표(2017): AI 시스템의 소유자, 설계자가 윤리적으로 고려해야 할 행동강령 제시(자율적 윤리가이드라인)
- 인식(Awareness), 이용과 수정(Access and redress), 책무성(Accountability, 설명가능성(Explanation), 데이터의 출처파악(Data Provenance), 감사가능성(Auditability), 검증과 검사(Validation and Testing) 등의 윤리원칙 기술
- 그외에도 전미컴퓨터학회 윤리강령(Code of Ethics, 2018) 등 AI 연구개발자들의 윤리적 행동강령 수립



● 국제기구 차원의 AI 윤리 정책동향 및 쟁점(5)

- 유럽 인공지능 최고전문가 그룹(AI HLEG)의 가이드라인
 - 유럽연합 집행위원회가 구성한 AI HLEG는 2018년 12월 윤리 가이드라인 초안을 발표
 - 인공지능 윤리의 목표로 기본권, 사회적 가치 및 각종 윤리 원칙을 포괄하는 “인간 중심”을 상징하고, 이를 위해 “신뢰할 수 있는 인공지능” 추구를 천명
 - 신뢰할 수 있는 인공지능을 현실화하기 위해 초기 설계 단계에서 설명 가능성, 데이터 거버넌스, 포용성(Design for all), 인간의 감시, 비편향성, 인간 의사의 존중, 사생활 보호, 안정성(Robustness), 보안, 투명성 등을 고려
 - 인공지능의 신뢰성을 평가할 수 있는 평가 항목들을 개발하여 제시
 - 인공지능의 신뢰성 요건에 대한 평가를 인공지능 생애주기 전체에 걸쳐 행해야 할 의무로 규정



● 소결 및 정책적 시사점(1)

- AI윤리 이슈가 한 국가의 중요한 정책 아젠다로 급부상할 뿐 아니라 국가간 AI산업 주도권 및 규범경쟁력을 확보하기 위한 글로벌 아젠다로 급격히 발전 → AI 윤리의 국제규범 및 거버넌스 형성과정에 적극적 참여
- 정부/공공, 법률가 등 전문가 주도의 입법/정책대응에만 맡기기 보다 AI기술 개발 및 활용과정에서 구체적으로 제기되는 다양한 윤리이슈에 이해관계자 및 일반시민들도 이해,참여할 수 있는 사회적 논의 수준 제고 필요 → ‘윤리위원회’ 등 다양한 이해관계자 갈등조정 위한 원칙 정립 및 그 절차를 반영한 거버넌스(윤리대응 프레임워크) 마련
- AI의 기술적, 사회적, 제도적 영향을 복합적으로 고려한 윤리 논의 체계 구축 → 기술/입법 영향 평가절차의 체계화 및 다면화
- AI윤리와 법제도 간의 상호관계에 대한 관점 정립, 즉 윤리적 논의를 형성 및 윤리적 이슈 검토 결과는 향후 AI 관련 사고의 법적 책임 및 합리적 보상체계를 정립하기 위한 원칙 및 기준으로 적용 가능성 모색 → 윤리와 법의 절충(융합)에 대한 법철학적/인문사회 학제간 연구기반 구축 필요

● 소결 및 정책적 시사점(2)

인공지능 윤리이슈에 대한 규범적 대응
법제도적 접근에 치중

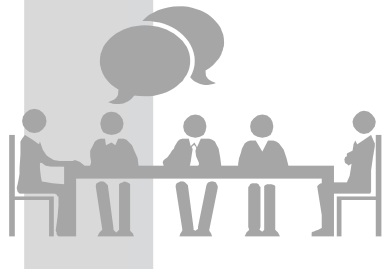
사용자와 시민들이 참여,공유하는
윤리적 논의는 여전히 미흡한 실정

윤리적 대응 또한 정부,공공기관 등
정부 주도로 진행

기업, 시민사회 등 국가사회적
대응 또한 여전히 미흡

산업분야 공급자, 개발자 자발적 참여 등
**산업별 AI윤리 논의틀 마련
시급**

4차위, 학회, 공공이 함께
**민관학 협력 기반의
AI윤리 공론화 추진체계
마련**



IV

산업분야별 인공지능 윤리이슈

산업별 인공지능 윤리이슈 검토대상의 선정

4개 산업분야의 선정

- 산업별 인공지능 유스케이스 분석결과(제2장), 웰니스, 의료, 금융, 제조, 홈 서비스 등에서 국내외 인공지능 제품/서비스별 산업 집중도가 높은 것으로 나타남 ⇒ **제조, 의료, 금융분야** 선정
- 인류의 생명권·자유권·재산권 등에 직접적인 타격을 가할 수 있는 치명적자율무기체계(LAWs)의 전쟁 무기적 특성으로 인해 국제적·국가 안보 차원의 법적·윤리적 쟁점 야기 ⇒ **군사분야** 선정



제조분야 인공지능 윤리이슈: 자율주행자동차를 중심으로

논의의 배경

- 자율주행자동차의 도입은 교통사고의 획기적인 감소 등 사회적 효용을 기대할 수 있지만, 동시에 새로운 형태의 교통사고*를 야기하며(일상성 및 정상성), 이로 인한 법적·윤리적 가치판단 요청

※ 구글의 '렉서스RX450h' 시험 주행 중 버스와 접촉사고(16.2.14.), 테슬라의 '모델S'는 대형 트럭과 충돌 운전자 사망(16.5.7.), 우버의 '볼보XC90' 교차로 건너던 행인 치어 사망(18.3.18.)

- 특히 윤리적 문제의 등장은 자동차의 자율주행에서 인공지능의 정보처리가 인간의 생명 및 신체에 대해 인간이 내려온 의사결정과정을 대신하고 있다는 점에서 근거 트롤리 딜레마와 같은 난제는 오늘날 과학기술이 결코 가치나 윤리에 종립적일 수 없음을 보여줌



● 제조분야 인공지능 윤리이슈: 자율주행자동차를 중심으로

● 윤리이슈(1)

단계	행위주체	윤리 이슈
개발단계	개발자	·인간에 대한 안전보호 의무 및 사회윤리적 공리를 반영하는 알고리즘 개발 - 하향식의 윤리적 프로그래밍 강조
	제조사	·알고리즘 개발 과정의 투명성 제고
	공적 주체 (중앙 및 지방 행정정부)	·자율주행자동차의 개발 및 상용화를 위한 파지티브(positive) 규제 방식 채택 - 금지조항의 최소화 및 디폴트적 허용
	소비자(운행자 포함)	·윤리적 기준의 자율적 선택권 보장에 대한 요청
	시민단체(보행자 포함)	·사회적 수용성 판단을 위한 공론의 의견 형성 및 반영 요청
인허가단계	개발자 및 제조사	·시험운행 단계에서의 윤리 기준 충족 및 운행결과에 대한 정보공개의 책무성
	공적 주체	·인허가를 위한 윤리적 검증 기준의 정립 및 적용
	소비자 및 시민단체	·인허가 과정에서의 윤리문제에 대한 모니터링 및 의견개진 절차 요청

제조분야 인공지능 윤리이슈: 자율주행자동차를 중심으로

윤리이슈(2)

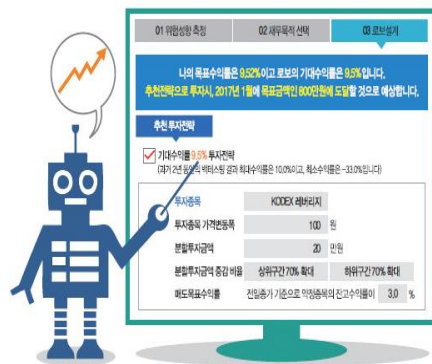
상용단계	개발자	·알고리즘에서 사회윤리 및 사회적 수용성의 지속적인 변화 추이 반영 - 상황식의 윤리적 프로그래밍 강조
	제조사	·자율주행자동차의 제조물책임보험을 통한 사후적 차원에서의 윤리문제 고려
		·법인 내 기관윤리위원회 조직 및 운영을 통한 윤리문제의 절차주의적 해결
	공적 주체	·국가·사회적 차원에서 공적 위험분배를 위한 사회보험 등 공적 기금 확보
		·인허가단계의 윤리적 기준 충족 여부에 대한 사후적 검증 도구 마련
		·스마트 시티 및 지능형 도로교통체계와의 연계에 따른 보안 문제 해소 방안
	소비자	·윤리적 프로그래밍에 대한 선택권 및 자기결정권 보장의 지속적 유지 - 프로그래밍 변경가능성 포함
	시민단체	·사회윤리 및 사회적 수용성 관련 지속적인 의견 형성을 위한 공론에서의 촉매 기능
		·도로교통체계의 지능화로 인한 프라이버시 및 개인정보 침해 예방 및 모니터링

금융분야 인공지능 윤리이슈: 로보어드바이저를 중심으로

논의의 배경

- 금융분야에서 인공지능의 활용은 단순한 온라인 신용분석에서 투자자문을 제공하는 로보어드바이저 그리고 나아가 자동으로 개인자산을 운용하고 재투자하는 투자일임·펀드매니저 영역까지 점차 확대되고 있음
- 특히 인공지능을 설계·개발한 비금융회사가 넓게 금융산업권에 포함되기 때문에 종래 금융시장과 산업에서 요구되는 윤리를 뛰어넘는 새로운 문제를 유발할 가능성이 큼
- 또한 전통적으로 인간의 행위만으로도 공정성 및 금융거래상대방 보호가 강하게 요구되는 금융산업에서 인공지능 기술 활용으로 새로운 유형의 문제가 등장할 수 있고, 이로인한 대규모의 피해로 금융거래에 참여하는 일반인의 취약성은 더욱 심화될 가능성이 높음

※ (Flash Crash 사건) 2010년 5월 6일 미국 주식 시장에서 그 당시 이미 보급된 초단타매매(High Frequency Trading)인공지능들이 위장거래에 반응, 매우 빠른 속도로 대량 매도를 하면서 대규모 폭락으로 이어짐



● 금융분야 인공지능 윤리이슈: 로보어드바이저를 중심으로

● 윤리 이슈(1)

단계	행위주체	윤리 이슈
개발단계	개발자	·알고리즘 개발단계에서의 윤리 특수성, 금융규제의 목적 투입 ·투자전문가가 투입(인정된 투자이론에 기초하여 설계)·상시감독
	자율규제기관	·AI 개발 촉진정책과 규제의 균형
	공적규제기관	·AI 개발 촉진정책과 규제의 균형
	금융종사자	·AI 이용과 관련한 윤리교육, 금융규제 준수 교육
허가단계	개발자	·허가를 위한 금융윤리적 자가검증도구 개발 및 적용
	공적규제기관	·허가과정에서 금융윤리적 검증도구 개발 및 역량 강화 (알고리즘과 데이터 평가능력, 이용자에게 제시 및 의사결정하게 제시되는 선택구조, 정보기술인프라, 인공지능이용이 대규모화될 경우 시장에 미치는 영향 등)

● 금융분야 인공지능 윤리이슈: 로보어드바이저를 중심으로

● 윤리 이슈(2)

이용단계	금융종사자	금융자문인	·AI 로보어드바이저의 특수성 및 위험 교육 ·이용자 보호 프로세스 준수 ·이해상충 방지 프로세스 준수 ·AI위험을 쉽게 알 수 있는 방법으로 설명·공시기법 선택 ·개인정보 보호
		[금융회사] 은행·증권회사·보험회사·자산운용회사·선물회사	·이용자 보호를 위한 최적화된 알고리즘 이용 및 검증 담보 ·이해상충 방지 프로세스 검증 ·AI위험을 쉽게 알 수 있는 방법으로 설명·공시기법 선택 ·알고리즘 투명성 담보 ·개인정보 보호 ·해킹방지등 침해방지를 위한 인적·물적 시설 적정수준 담보 ·금융시장 안정성 담보 ·이용자 피해 발생시 책임범위
	자율규제기관 (거래소, 금융협회 등)	·금융회사 이용자 보호와 규제의 균형 ·금융시장 보호(금융시장의 공정한 가격발전기능과 안정성 확보) ·금융윤리의 규범화 수준과 범위 ·AI기술에 관한 정보 및 대응 공유 ·AI 기술의 수혜 격차 문제	
	공적 규제기관	·금융회사 이용자 보호와 규제의 균형 ·금융시장 보호와 규제의 균형 ·금융윤리의 규범화 수준과 범위 ·AI기술에 관한 정보 및 대응 공유(감독기관간, 전문기관간, 전문가간) ·AI기술의 수혜 격차 문제	
	이용자	·결정과 선택권 문제 ·윤리적 투자 ·금융정보의 활용과 프라이버시 보호 ·해킹에의 안전	
	기타 (기관투자자)	·결정과 선택권 문제 ·윤리적 투자	

의료분야 인공지능 윤리이슈

논의의 배경

- 고령화의 핵심적인 욕구에 해당하는 건강과 장수에 대한 높은 기대감은 개인 맞춤형의료, 정밀의료, 예방의학에 대한 수요를 급속히 확대하고, 생체와 관련된 빅데이터의 특징과 중요도를 추출해내는 인공지능이 보건의료 수요에 크게 부응할 것으로 기대됨

※ IBM의 '왓슨(Watson)', 인공지능 간호사 '몰리(Molly)', 다양한 건강 관리 서비스 및 요양 서비스를 제공하는 간호용 로봇 '파페로(Papero)', '로보베어(Robobear)' 등 이미 출시 활용되고 있음

- 의료분야는 사람의 생명, 신체, 건강문제와 직결되므로 안전성 확보, 의료사고발생 최소화, 고정밀성, 책임소재 명확화 등 인공지능의 의료분야 활용과정에서 출현할 수 있는 다양한 윤리적 이슈에 대응 필요



IV. 산업분야별 인공지능 윤리이슈

의료분야 인공지능 윤리이슈

윤리 이슈

단계	행위주체	윤리 이슈
개발단계	개발자	·알고리즘 개발 단계에서 국가별/지역별 윤리의 특수성 고려
	공적 주체	·AI 개발을 위한 촉진 정책과 규제의 적절한 균형
	의사	·AI 이용과 관련한 윤리 교육(의과대학에서의 교육 포함)
인허가단계	개발자 및 제조사	·허가를 위한 윤리적 자가검증 도구 개발 및 적용
	공적 주체	·허가 과정에서의 윤리적 검증 도구 개발
	소비자 및 시민단체	·인허가 과정에서의 윤리문제에 대한 모니터링 및 의견개진 절차 요청
상용단계	의사	·수가 책정의 문제 ·의료사고 시 책임 면제 범위의 문제
	병원 경영자	·병원의 수익 창출에 최적화된 알고리즘 적용의 문제
	공적 주체	·의료사고 시의 손해배상 체계와 사전 행정적 규제의 검토 ·윤리적 내용의 규범화 범위 ·의료 기술의 수혜 격차 문제
	소비자	·데이터 활용과 프라이버시 보호의 문제 ·소비자의 선택권 문제 ·자가진료 내지 무허가 의료행위의 문제

군사분야 인공지능 윤리이슈: 치명적자율무기체계(LAWS)를 중심으로

논의의 배경

- 국방 분야에서 정치적·정책적 효용성으로 인해 강대국 중심으로 기존 무기체계에 인공지능을 접목한 무기체계에 대한 연구가 상당 부분 진전되어가고 있음
- 자율무기의 오작동이나 오판에 의한 결정이 무고한 생명을 앗아가는 결과를 초래할 수 있고, 불필요한 전쟁의 발발이나 전쟁 확대의 촉매제가 될 수 있다는 우려로 국제적으로 이와 관련한 윤리적 문제에 대한 참여한 논의가 이뤄지고 있음

개발자 / AI / 신기술미래

©2018.07.19

"자율살상 무기 안 만들겠다"... AI 전문가 2400명 2차 선언

Byron Connolly | CIO Australia

UN에 요청한 공개 선언서 전문

AI는 이미 군사 시스템에서 더 많은 역할을 할 수 있는 준비가 끝났다. 따라서 시민과 정책입안자, 리더들은 AI 사용에 대한 문제를 시급하게 다뤄야 한다. 우리는 인간의 삶이 절대로 기계에 좌우돼서는 안 된다는 점을 분명히 하며 이 선언에 서명한다. 기계가 인간의 생명에 대한 판단을 내릴 수 있도록 해서는 안 된다.

실제적인 위험도 있다. 인간의 개입 없이 목표를 선택해 처리하는 치명적인 자율살상 무기는 모든 국가와 개인을 위협에 빠뜨릴 수 있다. AI 연구자 수천 명은 이러한 위험과 능력, 인간의 생명을 빼앗는 것의 어려움을 알아야 한다는 데 동의한다. 치명적 자율살상 무기는 폭력과 압제의 가공할 수단이 될 수 있다. 특히 감시와 데이터 시스템과 연결되면 이런 위험성이 더 커진다.

치명적 자율살상 무기는 핵이나 생화학무기와 완전히 다른 특성을 갖는다. 일단 한쪽이 이를 도입하면, 아무런 제한 없이 너무 쉽게 전 세계적 군비 경쟁으로 치달을 수 있다. 현재 인류에는 이를 관리할 수 있는 기술적인 톨과 전 세계적 규제가 없기 때문이다. 이러한 군비경쟁을 비판하고 막는 것은 국가 차원은 물론 전 세계적인 안전 측면에서 매우 중요한 일이다.

서명한 우리들은 각국 정부와 그 정부의 지도자들에게 치명적 자율살상 무기에 대한 강력한 국제적 규범과 규제, 금지 법안을 만들 것을 촉구한다. 이들은 아직 만들어지지 않았지만, 그에 앞서 우리는 자발적으로 이 주장에 동참할 것이다. 따라서 우리는 치명적 자율살상 무기의 개발과 제조, 거래, 사용에 참여도, 지원도 하지 않을 것이다. IT 기업과 조직뿐만 아니라 리더와 정책입안자를, 그리고 다른 개인들도 이 선언에 동참하기를 기대한다.

KAIST 'AI 국방기술' 연구...해외 과학계, 보이콧 '파문'

KAIST, 한화와 손잡고 국방인공지능센터 열자 대표 AI학자 월시 등 57명 "로봇무기 반대" 보이콧 선언
신성철 총장 "살상용 아냐"...적극 해명에 논란 일단락

원호섭 기자 | 입력 : 2018.04.05 17:49:09 수정 : 2018.04.05 23:40:14

인공지능(AI)을 연구하는 국외 과학자들이 한국 과학기술 특성화 대학 KAIST와의 협력 사업에 대해 보이콧을 선언하는 이례적인 일이 벌어져 세간의 화제가 되고 있다. 학술지 사이언스와 외신 등에 따르면 전 세계 29개국 57명의 AI 과학자는 4월 4일(현지시간) "KAIST가 인간 통제가 불가능한 AI 무기 개발을 중단하기 전까지 KAIST 소속 교수 초청과 연구 프로그램 협력 등에서 손을 떼겠다"는 성명서를 전격 발표했다. 성명서 발표를 주도한 토비 월시 호주 뉴사우스웨일스대 교수는 "KAIST와 같은 명망 있는 기관이 AI를 이용한 신무기 개발을 염두에 두고 있다는 점이 실망스럽다"며 "KAIST 총장이 공개적으로 '인간의 유의미한 제어가 결여된 자율무기를 개발하지 않을 것'이라고 약속할 때까지 KAIST와의 공동 연구를 전면 거부할 것"이라고 밝혔다.

국외 AI 석학의 KAIST 보이콧 파문

날짜	내용
2월 20일	KAIST-한화시스템 국방인공지능센터 개소
3월 2일	토비 월시 교수 KAIST에 공개 질의 서한
3월 21일	국제인권감시기구 KAIST에 해명 요청
3월 29일	KAIST, 공격용 대량살상무기 개발하지 않음을 공문으로 발송
4월 4일	월시 교수 등 29개국 57명의 교수 KAIST 보이콧 선언
4월 4일	KAIST, 보이콧 선언한 연구자들에게 총장 명의 서신 발송해 해명
4월 5일	월시 교수 "KAIST 성명에 감명받았다" 보이콧 철회 시사

월시 교수는 고 스티븐 호킹 박사, 일론 머스크 테슬라 최고경영자(CEO) 등과 함께 AI를 활용한 킬러로봇 개발을 반대하는 대표적인 AI 학자다.

● 군사분야 인공지능 윤리이슈: 치명적자율무기체계(LAWS)를 중심으로

● 윤리 이슈(1)

단계	행위주체	윤리 이슈
정책결정단계	군	-치명적자율무기 개발을 필요성 및 허용여부
	최고통수권자	-방어목적의 AI를 채택할 때 얻을 수 있는 고유한 장점과 취약점은 무엇인가 -AI가 미래 전쟁을 어떻게 변화시킬 것인가에 대한 연구
	윤리/법률가	-LAWS는 경쟁국들과의 군사적 균형에 어떤 영향을 미칠 것인가 -국가안보 최고수호자로서의 윤리의식 -LAWS의 사고책임에 대한 법적·윤리적 쟁점에 대한 신속한 대응책이 강구되었는가 -쟁점들에 대한 국제법 준수 및 대응 입법은 준비되었는가
개발단계	개발자	-알고리즘에 어떤 윤리기준을 주입시킬 것인가 -전쟁법에서 요구하는 의미 있는 인간통제를 어떻게 구현할 것인가
	윤리/법률가	-법적/윤리적 기준을 제시할 수 있는가
	운용책임자	운용자가 필요로 하는 별도의 기준은 없는가
실험단계	개발자 및 제조사	-LAWS의 안정성은 담보되었는가
	공적 주체	-지향하는LAWS는 극대의 살상이라는 효용성은 보장되는가
	소비자 및 시민단체	-테러리스트 등에 대한 보안수준

● 군사분야 인공지능 윤리이슈: 치명적자율무기체계(LAWS)를 중심으로

● 윤리 이슈(2)

국제협력단계	국방/외교	-LAWS에 대한 국제 보편의 기능적, 윤리적 기준은 무엇인가
	윤리/법학자	-기존 국제법질서와의 관계는 어떻게 되는가
	운용책임자	-보편과 특수의 조화를 어떻게 도모할 것인가
실전배치단계	부대책임자	-LAWS의 살상력은 충분하가 -아군과 적군, 민간인과 무장군인을 정확히 구분할 수 있는 여부
	정책담당자	-상대세력의 LAWS에 필적하는지 여부 -정책부서와 정보주서의 유기적인 협조가 이루어지는가 -사이버공간에서도 구동이 가능한 수준의 LAWS인지 여부 -프라이버시 유출에 대한 대비가 적절히 이루어졌는지 여부 -현장 배치된 LAWS의 프로그램의 해킹에 대한 대비가 충분한지 여부
	윤리/법률가	-어떤 유형의 피해를 인정할 것인지 여부



인공지능 윤리이슈 정책적 대응방안

● 인공지능 윤리이슈에 대한 정책적 대응과 프레임워크 구축의 필요성

- 인공지능 기술의 개발은 아직 발전 중인 영역으로, 향후의 발전 방향과 속도를 예측하기 어려움
- 따라서 법적인 규율을 확정하는 것보다, 향후 제기될 수 있는 윤리 문제에 자율적으로 대응할 수 있는 프레임워크를 구축하는 것이 합리적이고 효과적인 대응 방안일 것임
- 법적 규제와 달리 윤리적 대응은 도덕률에 대한 사회적 합의와 자율적 준수라는 속성을 갖고 있음

과학기술기본법 제4조 제4항 : 과학기술인은 자율을 바탕으로 과학기술 활동을 수행하되 과학기술이 미치는 **사회적·윤리적 영향을 고려하여** 진실성 있게 수행하여야 하며, 경제와 사회의 발전을 위하여 과학기술의 역할이 매우 크다는 점을 인식하고 자신의 능력과 창의력을 발휘하여 이 법의 기본이념을 구현하고 과학기술의 발전에 이바지하여야 한다

국가정보화기본법 제40조 : 국가기관과 지방자치단체는 **건전한 정보통신윤리를 확립하기 위하여** 미풍양속을 해치는 불건전한 정보의 유통을 방지하고 건강한 국민정서를 함양하며, 불건전한 정보로부터 국민을 보호하기 위하여 ...

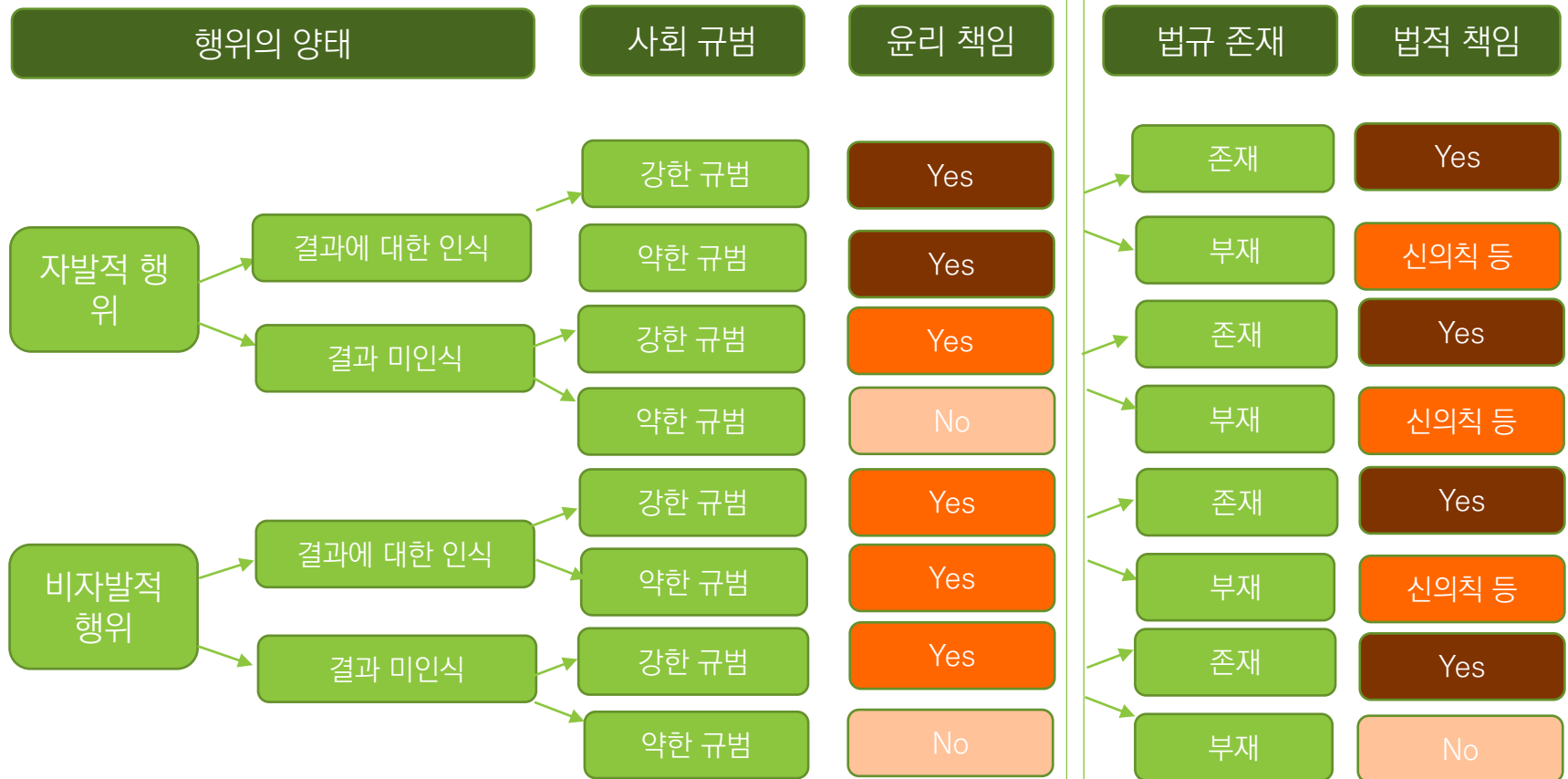
- 기술의 개발자를 포함한 기업이 일차적으로 윤리 준수의 의무를 갖고 있으나, 다른 한편으로 공적 주체에게도 윤리 문제 대응의 임무를 부여하고 있음

● 윤리 대응 프레임워크 구축 시 고려 사항

- 일반적 윤리 거버넌스의 특징, 인공지능 기술의 특성, 인공지능 기술의 사회적 수용성 및 그에 따른 사회 현안 등을 염두에 두어야 함
- 법적 규제와의 경계 설정
 - 윤리적 대응의 구조와 공적 주체가 행하는 구속력을 가지는 규제와의 경계를 설정함으로써, 윤리 문제への 대응이 결과적으로 산업 성장을 위축시키는 규제가 되지 않도록 하는 노력이 필요
 - 윤리는 “사회적 합의”의 산물이며, 그 중 입법자가 법의 영역으로 설정한 부분이 법적 규제에 해당
 - 윤리와 법의 경계가 명확한 것은 아니므로, 규율의 대상, 행위의 성질 및 규율의 속성 측면에서의 단계적 고려가 필요
 - 행위의 양태와 사회적 규범의 합의 유무를 고려하여 윤리 책임에서 법적 책임으로 전환되는 경계를 표현한 아래 도식을 참조할 수 있음

윤리 대응 프레임워크 구축 시 고려 사항

법적 규제와의 경계 설정



❖ 법규의 존재 여부 판단 시에는 사회 규범의 강약 여부는 상관 없음

V. 인공지능 윤리이슈 정책적 대응방안

● 윤리 대응 프레임워크 구축 시 고려 사항

※ 최근 발간된 윤리 가이드라인

국가	가이드라인명	발간 주체	발간 시기
우리 나라	지능정보사회 윤리 가이드라인	정보문화 포럼 한국정보화진흥원	2018. 9.
	카카오 알고리즘 윤리 헌장	카카오	2018. 1.
독일	자동화 및 네트워크화된 자동차에 대한 윤리 규범(Ethische Regeln für den automatisierten und vernetzten Fahrzeugverkehr)	연방 교통 및 디지털인프라스트럭처부 윤리위원회	2017. 6.
미국	아실로마 인공지능 원칙(Asilomar AI Principles)	Future of Life Institute	2017. 1.
	인공지능과 인권(Artificial Intelligence & Human Rights)	Berkman Klein Center(Harvard Univ.)	2018. 9.
일본	전문가를 위한 인공지능 가이드라인	일본 인공지능 협회(JSAI)	2017. 5.
유럽연합	로봇 영역의 민사적 규율에 대한 집행위원회에 대한 권고	유럽 의회	2017. 2.
OECD	인공지능 윤리의 난제 – 기계에의 도덕성 구성을 위한 3개 가이드라인(The Hard Problem of AI Ethics – Three Guidelines for Building Morality into Machines)	Vyacheslav Polonski 박사	2018. 2.

● 윤리 대응 프레임워크 구축 시 고려 사항

● 민관학협력의 모델 고려

- 윤리 대응 프레임워크는 일차적으로 개발자의 자율에서 시작하지만, 그 규범 형성 과정에서 다양한 사회 주체의 참여가 필요
- 윤리는 결과를 도출하는 것이 아니라, 합의를 도출하는 과정이므로, 공적 주체는 모든 사회구성원들이 참여할 수 있는 논의의 장을 형성해야 함

※ 독일정부의 인공지능 관련 컨설팅 플랫폼 등

● 서비스 이용자의 고려

- 지능정보산업 영역에서 독자적인 이용자 보호가 필요
- 적극적 행위자로서의 이용자의 성격 역시 고려해야 함
- 인공지능기술을 사회에 수용하는 단계별로 피보호자 혹은 윤리 규범의 수범자(이른바 디지털 시민권, Digital Citizenship)로서 이용자를 고려
- 윤리 규범의 수립, 인공지능 기술의 개발과 기술의 사회적 영향에 대한 평가 및 실제 기술 적용 과정에서 이용자의 의견을 반영할 수 있는 절차 수립

● 윤리 대응 프레임워크 구축 시 고려 사항

● 인공지능 기술 발전 속도 고려

- 지나친 윤리 규범 강조로 인해 기술 혁신의 이니셔티브가 약화되지 않도록 고려
- 윤리 규범 구축 단계에 참여하는 참여자들이 기술 개발의 속도에 대해 이해해야 함

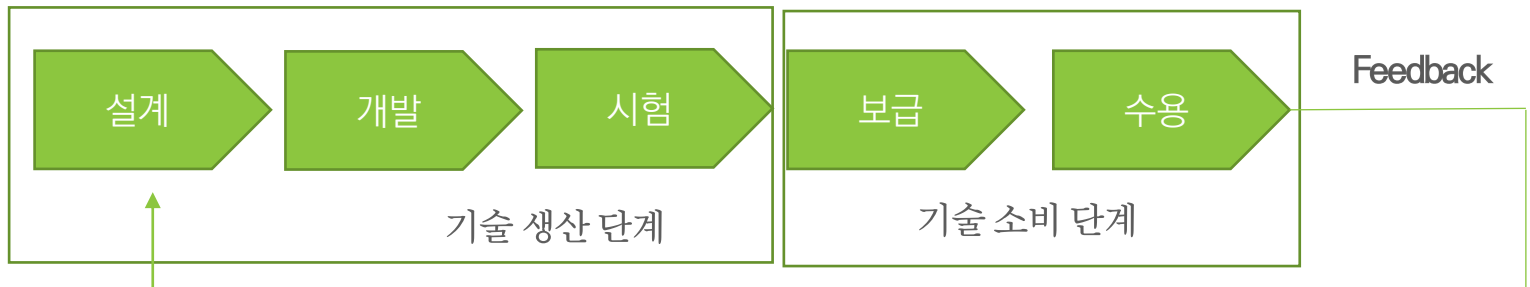
● 사전 예방과 사후 대응의 조화

- 기술의 상용화 이전 단계에서 모니터링을 통해 리스크 판단 필요(Responsible Risk & Innovation)
- 개발자 외에 일반적인 상식을 지닌 비전문가가 모니터링에 참여함으로써 불확실성과 파급 효과에 대한 선제적 대응 가능
- R&D 사업의 평가 단계에서 윤리적 기준을 평가 척도의 하나로 적용
- 디지털 시민권(Digital Citizenship)의 측면에서 개발자 뿐만 아니라 이용자 윤리 역시 사전 준비가 필요
- 기술의 발전 정도 등을 고려하여, 사전 예방이 중요한 분야와 사후 대응이 중요한 분야에 대한 분리 대응도 고려

● 윤리 대응 프레임워크 구축 시 고려 사항

● 인공지능 기술의 생애주기에 따른 윤리 이슈 고려

- 인공지능 기술을 실제로 적용하기 위해서는, 기술의 설계에서 개발, 임상 및 상용화에 이르는 단계가 필요하며, 각 단계별로 중요한 윤리 이슈가 발생할 수 있음
- 각각의 단계별 윤리 이슈를 인식하고, 해당 단계에 적합한 대응 방안을 고려할 필요 있음



● 프레임워크의 범용성

- 다양한 영역의 인공지능 관련 쟁점에 대응할 수 있도록 보편성 및 포괄성을 가지는 한편, 유연성 또한 갖추어야
- 인공지능 윤리와 관련해서 사안별로 쟁점에 대한 논의를 가능하게 하는 합리적이고 범용적인 구조를 만드는 것이 필요 속의 거버넌스
- 분야별 범용성 뿐만 아니라 수범자의 범용성도 고려 개발자, 이용자 및 공적주체 모두가 참여자이자 수범자

● 윤리이슈 대응을 위한 종합 프레임워크의 구축

● 프레임워크의 원칙

- 인공지능 기술의 생애주기별 단계에 따라 각각의 수범 주체별 프레임워크 구성
- 사전 예방의 차원에서 개발 이전 단계에서 윤리 및 리스크 커뮤니케이션 필요
- 윤리 규율과 법적 규율을 경계짓되, 법적 규율에 대해서는 입법자의 수권 필요
- 각각의 레이어에서 민관협력 방식의 협의체가 필요한 경우를 제시
- 프레임워크 내에서의 실행 주도는 개발자와 이용자 등 민간이 수행하되, 국가는 기술 표준, 위험 인증 등의 기술적 지원 및 논의 플랫폼 구성 시의 심판자 역할
- 사후의 논의 플랫폼을 통하여 축적한 경험을 신규 기술 도입 시 활용 가능
- 종합 프레임워크를 기초로 각 윤리 쟁점별 혹은 산업분야별 특성을 반영하여 구체적으로 적용할 수 있는 대응 프레임워크 도출 가능
- 각각의 프레임워크 내에 있는 단계별 윤리규범의 준수 여부에 대한 체크리스트 작성 가능

V. 인공지능 윤리이슈 정책적 대응방안

● 윤리이슈 대응을 위한 종합 프레임워크의 구축(행위주체별)

		개발자/공급자	공적 주체	이용자
개발 이전 단계	☞	· 윤리 규범 구성의 기반이 될 사회적 가치에 대한 합의 / 사회적 협의체 → 윤리 / 리스크 커뮤니케이션 · 사회적 규범을 바탕으로 한 자율 윤리 규범 정리 / 법적 규율 경계 고려 · 연구윤리(RRI) 고려 / 윤리 교육 실행		
설계 단계	☞	· 영향평가 · 자가리스크 진단	· 영향평가의 요령 제공	· 영향평가 단계에서 의견 제시
개발 단계	☞	· 제조 시 자율 윤리규범 준수		
시험 단계	☞	· 윤리 규범 준수	· 안전기준 구축 · 리스크 평가 / 인증	· 의견 제시
소비 단계	☞	· 논의 플랫폼의 구축 / 민관협력의 사회적 협의체		
		· 피해 구제	· 위법 행위에 대한 제재 · 오작동에 대비한 기술적 지원 체계	· 신기술 리스크의 사회적 수용 → 기술의 완전한 수용
피드백 단계	☞	· 소비 단계에서 형성 및 축적된 경험을 신규 기술의 도입 시 반영		

V. 인공지능 윤리이슈 정책적 대응방안

● 윤리이슈 대응을 위한 종합 프레임워크의 구축(쟁점별)

		알고리즘 책임성		설명 가능성		공정거래 쟁점	
		알고리즘 규제 알고리즘 윤리 구축	규제 필요성이 인정되면 법적 규제	설 명 가 능 한 알고리즘 설계	윤리 강령 필 요		
개발 이전 단계	☞						
설계 단계	☞						
개발 단계	☞	알고리즘 윤리 의 준수	자율 대응			공정한 보조금	법적 규제
시험 단계	☞	영향 평가	영향 평가 기 준 및 기술 표 준	알고리즘의 설 명 가능성 검 사	자율 대응		
소비 단계	☞			이용자에 대한 설명	자율 대응 이용자보호 가 이드라인	이용자 간 차 별 시정	자율 대응 경우에 따라 법적 규제
피드백 단계	☞	논의 플랫폼을 통해 형성 및 축적된 경험을 신규 기술 도입 시 반영					
		민관학 협의체의 구성 및 기능					

V. 인공지능 윤리이슈 정책적 대응방안

● 개별 산업 영역에의 적용

● 개별 산업 영역에 특유한 윤리 이슈와 특징을 반영하여, 단계별 특수성 고려

	제조 분야	금융 분야	의료 분야	군사 분야
개발 이전		- AI개발 촉진과 규제의 균형 - AI이용 관련 금융종사자 윤리 교육	- 인간중심주의 강조-의료정보 데이터 취급에 유의 - 의료 종사자에 대한 윤리 교육	- 인권 존중-국제인권협약 준수 - LAWS 개발의 국방상 필요성 - 국가 안보 수호자로서의 윤리 의식
설계	알고리즘의 윤리적 구축-트롤리 딜레마	알고리즘의 윤리적 구축-초단타 매매 사례		알고리즘의 윤리적 구축-전쟁법 등의 준수
개발	스마트시티 및 IoT 등 기술과의 연관성 고려		- 의료기기 인허가 제도 - 기술영향평가제	
시험	인허가 단계에서의 윤리적 검증	윤리적 자가 검증	- 임상 표준 고려 - 표준 의료정보 품질 구축 - 윤리적 자가 검증	- 안전성 보장 - 효용성의 고려 - 테러리스트 등에 대한 보안
소비	- 자율주행차 책임주체성 - 책임보험제도 고려 - 소비자의 자기결정권	- 투자자 보호 - 윤리적 투자	- 유사.불법.자가 의료행위 방지 - 기술 수혜 격차 유의 - 경영자의 의료인의 충돌	상대 세력의 전력을 고려한 이용
피드백	자동차 사고의 일상성 고려-공동체 차원의 리스크 관리	- 기관 간 정보 공유 - 해킹 대비	의료정보 자산화 및 데이터의 공익적 재이용	- 피해 유형의 검토 - 해킹 대비

개별 산업 영역에의 적용

제조 분야의 특수성을 고려한 윤리 대응 프레임워크

		특수 쟁점	개발자/공급자	공적 주체	이용자
개발 이전 단계	☞		윤리 / 리스크 커뮤니케이션을 통한 〈자율주행차 윤리 가이드라인 등〉구축		
설계 단계	☞	알고리즘 구축	영향평가 자가리스크 진단	영향평가의 요령 제공	영향평가 단계에서 의견 제시
개발 단계	☞	타 기술과의 연관성 고려		스마트시티 구축 등의 지원	
시험 단계	☞	인허가 단계의 윤리 검증	윤리 규범 준수	안전 및 윤리 기준 구축	의견 제시
소비 단계	☞	책임 주체로서의 자율주행차	논의 플랫폼의 구축 / 민관협력의 사회적 협의체		
		책임보험제 소비자 자기결정권	교통사고 피해 구제		리스크에 대한 수용
피드백 단계	☞	자동차 사고의 일상성 고려	공동체 차원의 리스크 관리		

● 개별 산업 영역에의 적용

● 금융 분야의 특수성을 고려한 윤리 대응 프레임워크

		특수 쟁점	개발자/공급자	공적 주체	이용자
개발 이전 단계	☞	AI 개발 촉진과 규제 균형	윤리/리스크 커뮤니케이션을 통한 사회적 합의		
		AI 이용 관련 금융 종사자에 대한 교육	종사자 교육 시행	규제 개선 교육 프로그램 제공	AI 금융 시스템에 대한 이해
설계 단계	☞	알고리즘 윤리적 구축 (Flash Crash 방지) 윤리적 자가 검증	영향평가 자가리스크 진단 투자 전문가의 참여	상시적 금융 감독 리스크 진단 도구 제공	의견 제시
개발 단계	☞				
시험 단계	☞				
소비 단계	☞	투자자 보호 윤리적 투자	윤리적 투자에 대한 사회적 합의		
			투자자 피해 구제		투자 리스크에 대한 수용
피드백 단계	☞	기관 간 정보 공유 해킹 대비	데이터 공유에 대한 사회적 합의 필요		

V. 인공지능 윤리이슈 정책적 대응방안

개별 산업 영역에의 적용

의료 분야의 특수성을 고려한 윤리 대응 프레임워크

		특수 쟁점	개발자/공급자	공적 주체	이용자
개발 이전 단계	☞	인간 중심주의 강조 의료 정보 데이터 취급에 유의	의료 종사자에 대한 교 육 (의과대학 포함)	데이터 법제 정비 교육 내용 제공	
설계 단계	☞	알고리즘 구축		영향평가의 요령 제공	
개발 단계	☞		기술영향평가	의료기기 인허가 기준	
임상 단계	☞	임상 윤리 준수	윤리적 자가 검증 임상 표준 고려	표준 의료품질 제시	
소비 단계	☞	유사, 불법 및 자가 의료 행위	논의 플랫폼의 구축 / 민관협력의 사회적 협의체		
		기술수혜 격차 유의 경영자와 의료인 충돌 고 려	의료사고 피해 구제	불법의료행위 규율	의료인의 책임 인식 리스크에 대한 수용
피드백 단계	☞	의료 정보 데이터의 공익 적 재활용	데이터 활용에 대한 사회적 합의		

V. 인공지능 윤리이슈 정책적 대응방안

개별 산업 영역에의 적용

군사 분야의 특수성을 고려한 윤리 대응 프레임워크

		특수 쟁점	개발자/공급자	공적 주체	이용자
개발 이전 단계	☞	인권 존중 국제인권협약 준수 LAWS 개발의 국방상 필요성 검토	인권 수준에 대한 사회적 합의와 리스크 커뮤니케이션 전력상 LAWS 의 필요성에 대한 사회적 합의		
			안보 수호자로서의 책임 인식		
설계 단계	알고리즘 구축에서의 전쟁법 준수				
개발 단계					
시험 단계	안전성 보장 효용성 고려 테러리스트에 대한 보안	윤리 규범 준수			
소비 단계	☞	상대 세력의 전력 고려	논의 플랫폼의 구축 / 민관협력의 사회적 협의체		
			오폭 등에 대한 피해 구제	사고 책임에 대한 법적 대응 절차 마련	운용 책임자로서의 윤리적 기준 준수
피드백 단계	☞		공동체 차원의 리스크 관리 피해 유형의 검토		



감사합니다





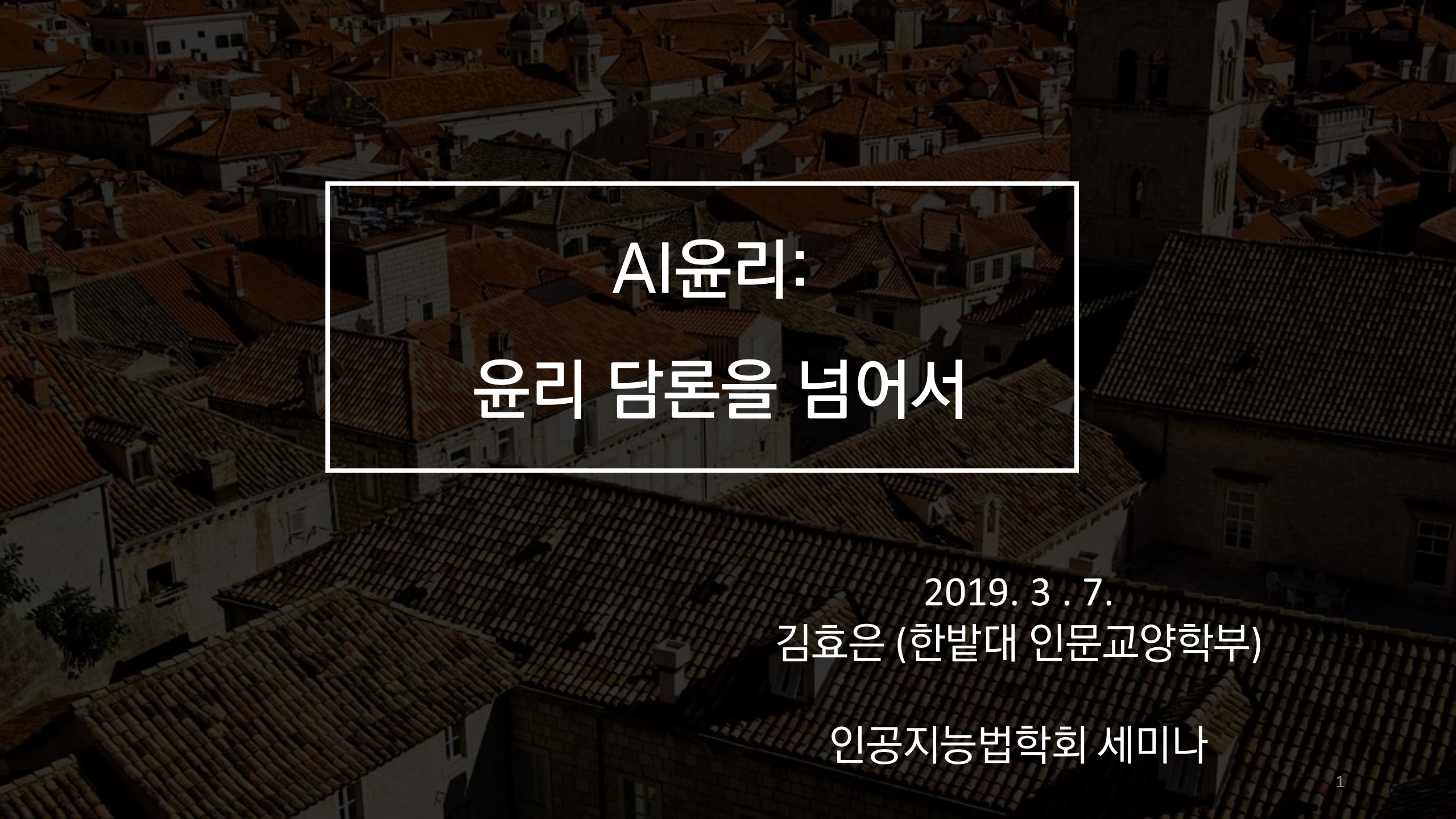
Session II- 토론

인공지능의 윤리적 개발 동향과 입법 대응



Session II- 토론

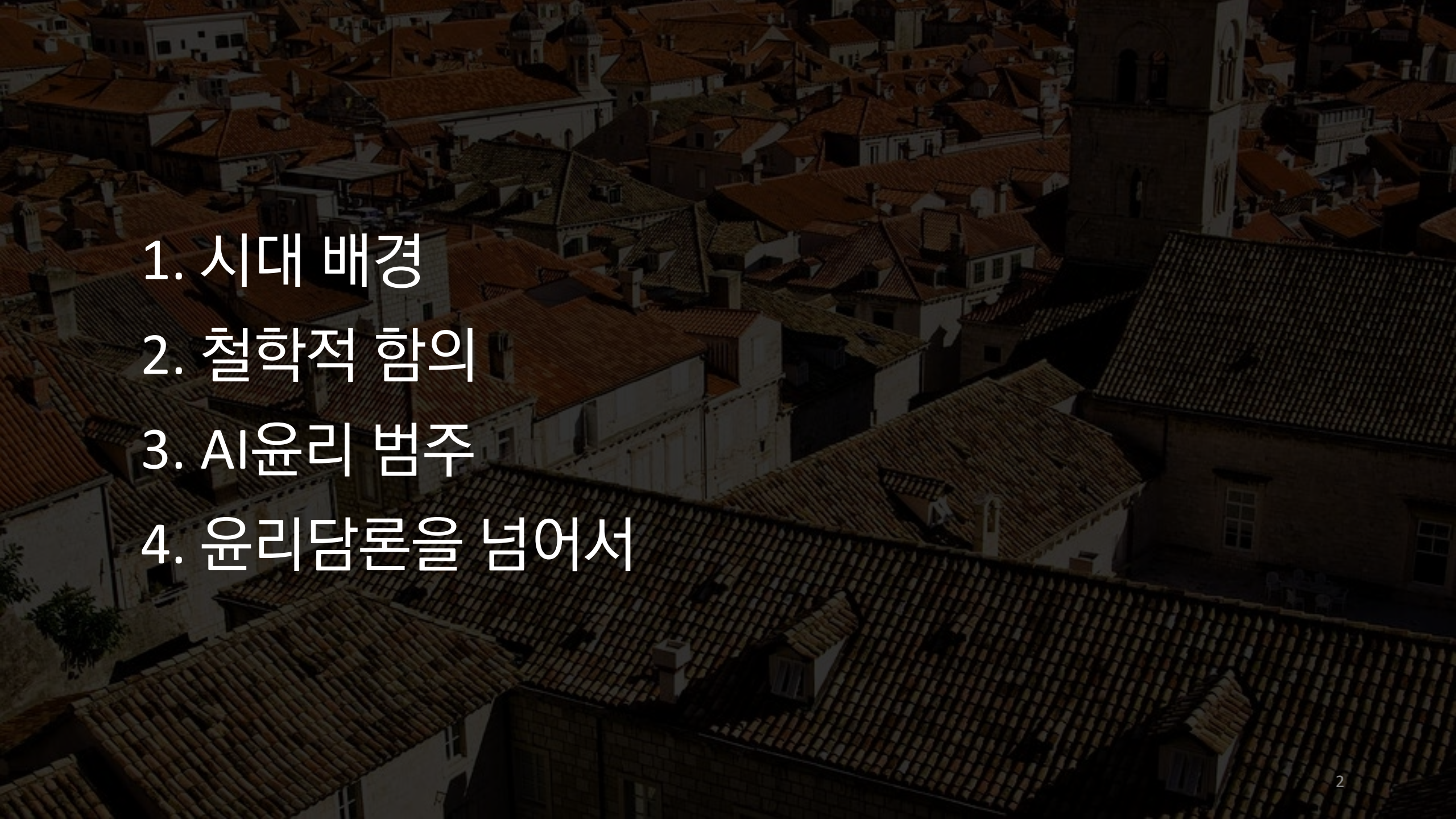
김효은 교수 - 한밭대학교 윤리학



AI윤리: 윤리 담론을 넘어서

2019. 3 . 7.
김효은 (한밭대 인문교양학부)

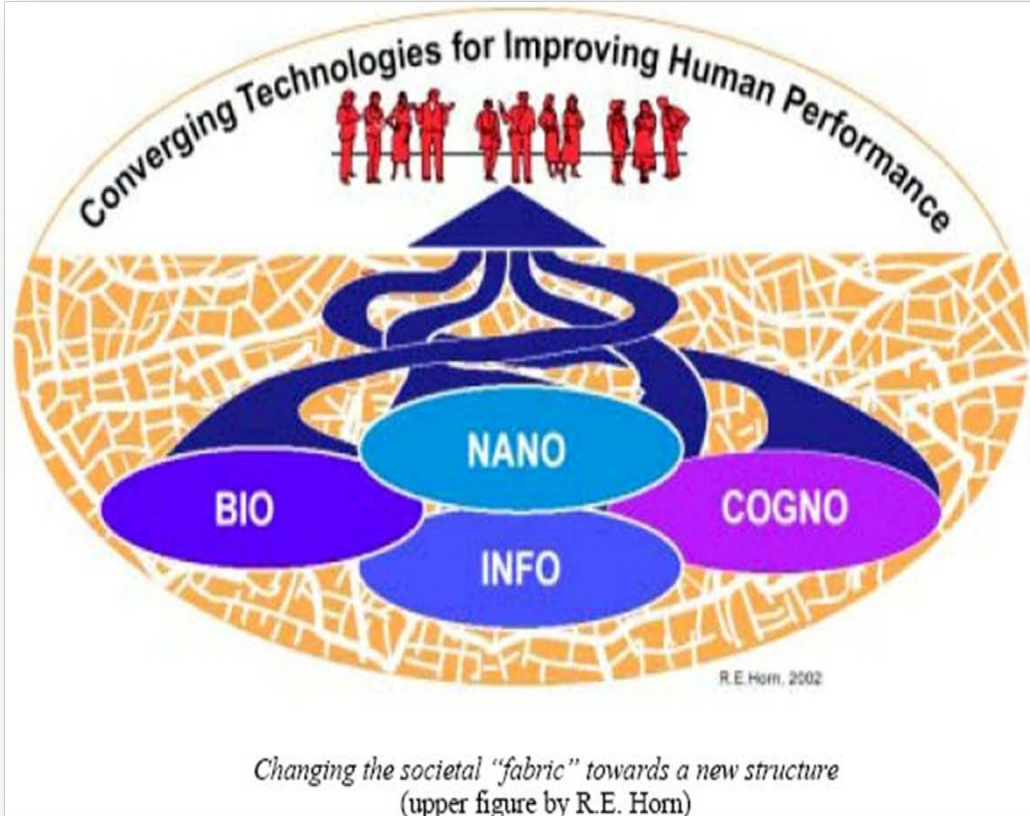
인공지능법학회 세미나

- 
- An aerial photograph of a historic city, likely in Europe, showing a dense collection of buildings with red-tiled roofs. A prominent church tower with a clock face is visible on the right side. The image is darkened to serve as a background for the text.
1. 시대 배경
 2. 철학적 함의
 3. 시윤리 범주
 4. 윤리담론을 넘어서



1. 시대배경: 신경 테크놀로지와 인공지능의 시대

혁신적 기계 발명 → 인간 수행능력 향상



2002년 미국 과학재단
<미래 융합과학기술 예측보고서>

NBIC(Nano, Bio, Info, Cogno)의
동시적, 융합적 발전

자율적 판단을 통하여 작동하는
Artificial Agent의 출현

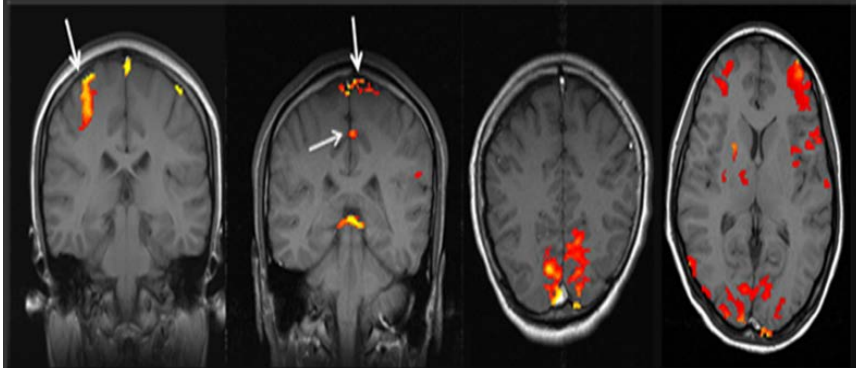
Mixed Reality 에 따른
몸-마음 이분법, '자아'개념 붕괴

과학기술 발전 단계 - '인지'차원의 출현



1단계(~1950년 경):

물질/기계



2단계는 (1950년~):

물질/기계 + 정보와 생명



3단계 (2000년 이후):

물질/기계 + 정보와 생명 + 인지
융합과학기술

2. 철학적 함의: 전통 개념의 변화



누군가 이 로봇팔을 파손한다면
‘기물파손’인가, ‘과실치상’인가?

‘인공’ ‘자연’의 경계는?

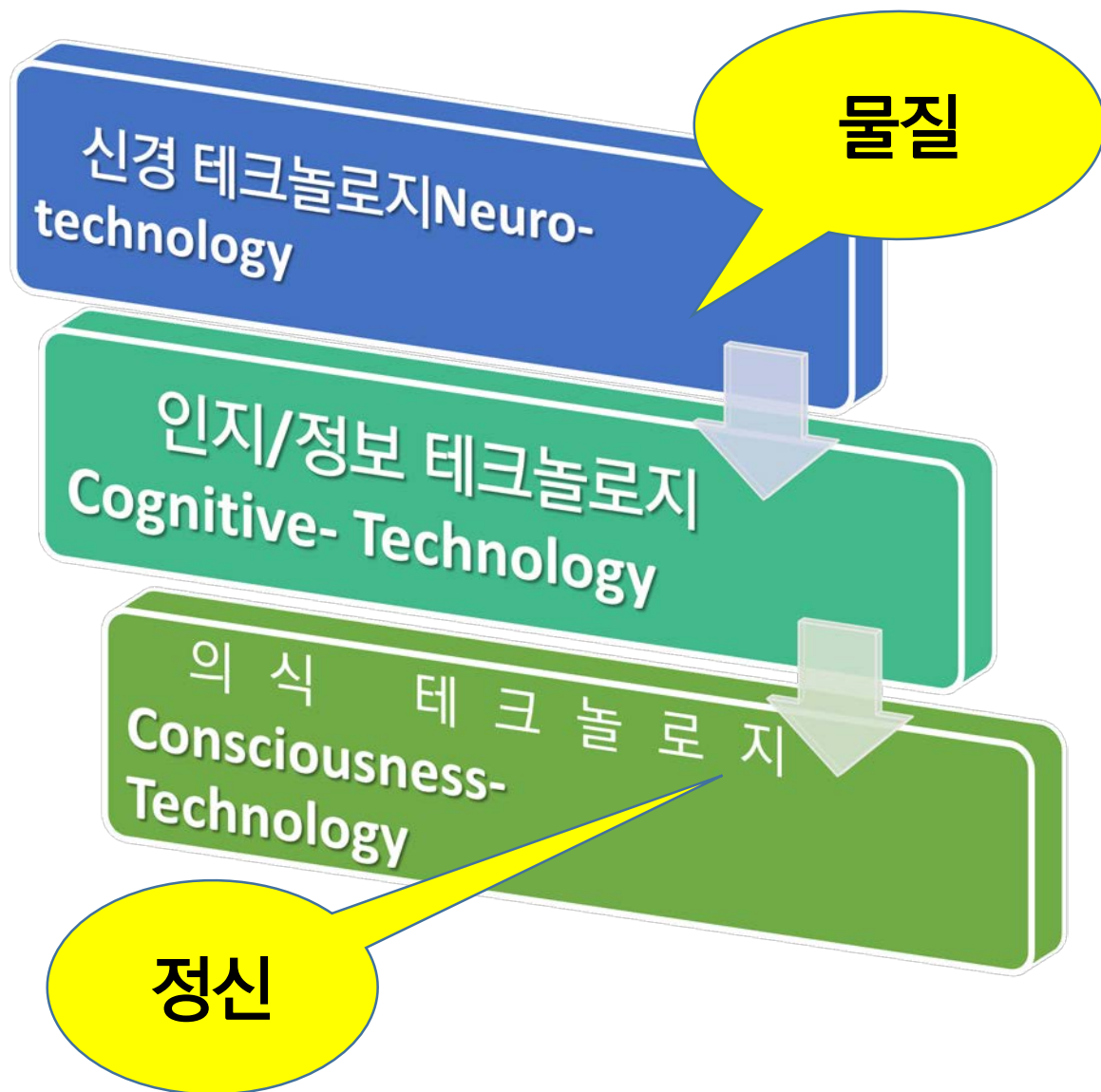
인공지능 팔은 ‘사물’인가, ‘신체’인가

인공지능 학습도 ‘마음’인가?

‘행위자’는 사람인가 인공지능인가?

해를 입힐 경우 누구의 ‘책임’ 인가?

정신-물질 이분법에의 도전



- Alan Turing의 범용튜링기계의 현실화



- 엘런 머스크의 Neural Link (침습적 뇌-기계 인터페이스)
- Facebook의 Braintyping (비침습적 뇌-기계 인터페이스)

3. AI윤리 범주

- 범주오류 방지를 위한 구분

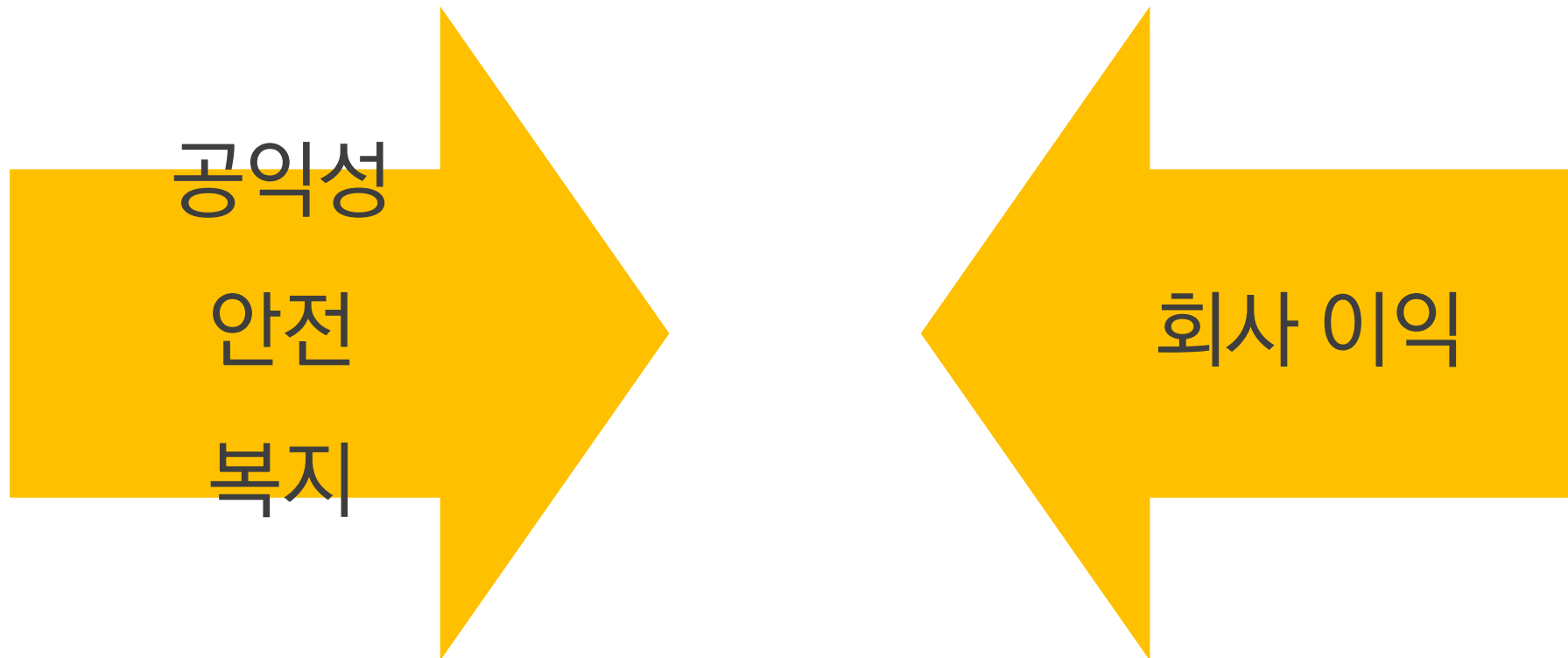
ANI (Artificial Narrow Intelligence)
협정의 인공지능

AGI(Artificial General Intelligence)
일반인공지능

ASI(Artificial Super Intelligence)
초인공지능



기술개발과 윤리 사이의 기존 딜레마



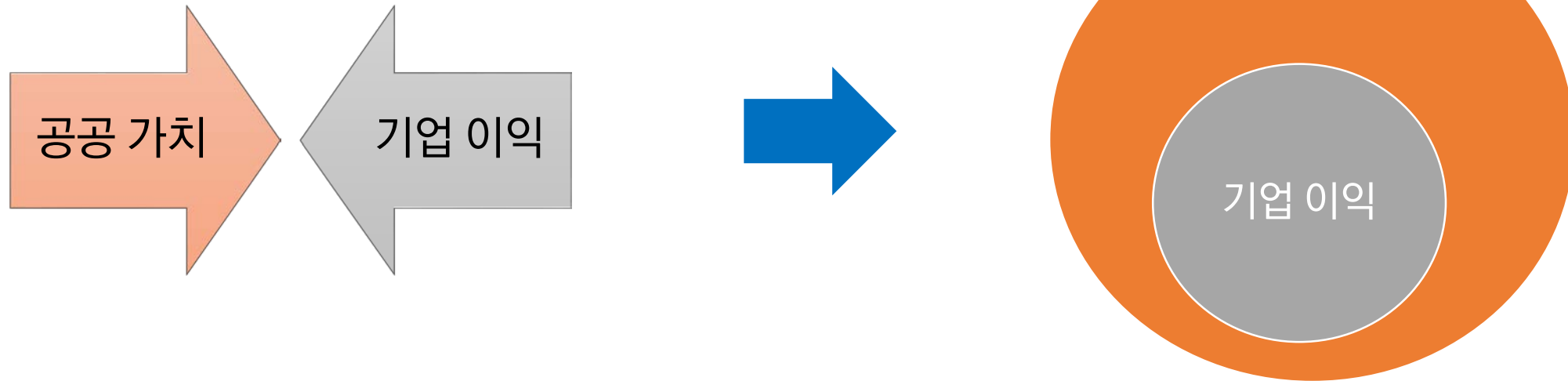
하루 30억개 데이터 학습해 분석·예측 날마다 정확해져
세계최초 모든 공정에 도입...포스코, AI 선도기업 부상

◆ 한국산업 판 뒤집는 AI혁명 / AI 도입하는 '빅 팩토리'...포스코 광양제철소의 변신 ◆



기존 윤리적 딜레마가 해소됨

윤리 프레임의 변화: 윤리적 경영+이익



윤리는 단지 법적인 위험에 대한 것이 아니다.

소비자의 위험이나 공공 안전을 개선하기 위한 행동을 통해 기업의 책임을
최소화하고자 하는 노력이다.

새롭게 떠오르는 기술은 아직 법이나 규제가
예측하거나 기술할 수 없는 불분명한 윤리 영역이 있다.
이는 단순한 프라이버시 문제나 제품 책임의 문제를 넘어서는 문제다.

P. Lin and E. Selinger, “Inside Google’s Mysterious Ethics Board”,
Forbes, Feb 3, 2014.

대학에서의 AI윤리

- MIT & Harvard
 - <Ethics and Governance of AI>
 - 2018년 1학기
 - 6주 과정
 - 윤리적 이슈와 규제, 정책 중심
 - AI윤리 기금에서 지원
- 한밭대
 - <인공지능윤리> 교과 개설
 - 2018년 1학기
 - 14주 과정
 - 딥러닝-인지과학-윤리-법-경영
 - 산학협력-인문사회-이공계 연계로 발전

4. 윤리 담론을 넘어서

규제로서 인식되는 AI윤리
‘지속가능성’의 맥락

기술 인식/가치 인식의 중요성
새 사회문화의 필요성

IEEE-SA Standards Projects

알고리즘 편향 고려사항 *P7003*TM

발표 (2017. 3. 9.)

Algorithmic Bias Considerations provides developers of algorithms for autonomous or intelligent systems with protocols to avoid negative bias in their code. **Bias could include the use of subjective or incorrect interpretations of data like mistaking correlation with causation.**

The project offers specific steps to take for eliminating issues of negative bias in the creation of algorithms. The standard will also include benchmarking procedures and criteria for selecting validation data sets, establishing and communicating the application boundaries for which the algorithm has been designed, and guarding against unintended consequences.

자율적 지능 시스템의
사용자 및 이해관계자에게
알고리즘 목표, 평가, 영향에 대해
설명

알고리즘의
설계, 검사, 평가에 사용

윤리 이론이나 담론에 매몰되어서는 안되는 이유

미국과 독일의
무인자동차 윤리 가이드라인 비교

사회자본 (>시장자본)의 부상
교육, 문화, 사회를 어떻게 재구성할 것인가

미국의 자율주행차 가이드라인 (2017.9)

- 미국 교통부와 도로교통안전국에서 자율주행차량 안전지침 가이드라인 ‘자율주행시스템: 안전을 위한 비전(A Vision for Safety)’ 발표
- 자율주행기술 개발을 방해하는 규제를 없애 발달 촉진이 목표
- 안전성 평가 기준을 15 → 12점으로 하향함.
- 개발위주의 정책 때문에 윤리적 규제는 거의 고려 않는 것으로 결정
 - 사생활 보호 문제를 고려 안함

독일의 자율주행차 가이드라인 (2017.6.)

2. 개인에 대한 보호는 공리주의적 고려 대상이 될 수 없다.

다. ...해로움이나 위험의 수준을 완전히 방지할 수 있을 때까지 줄여나가는 것이 목적이며, 인간이 차를 운전하는 것보다 명확히 위험이 줄어들어야 자율주행차에 대한 라이선스가 이루어질 수 있다.

8. 사회윤리 딜레마의 경우 실제 상황에서나 예견할 수 없는 행위를 통해 사람의 생명이 선택될 수 없다. 다시 말해 이런 판단을 표준화하거나 프로그램할 수 없으며 기술 시스템이 사고에 대해 복잡하거나 본능적인 평가를 내려 인간 운전자의 도덕 판단을 대체하면 안된다.

9. 피할 수 없는 사고에서 개인의 특성(나이, 성별, 물리적, 정신적 차이)을 기반으로 어떤 차별적 판단과 다른 사람의 희생을 제안해서는 절대로 안된다.

AI기술의 과적합 등의 문제 & 사회자본을 그대로 투영하는 Data 의 중요성 어떻게 접근할 것인가

TOP 10 머신러닝 알고리즘

- 의사결정나무
- 단순베이지분류
- 일반최소제곱(OLS)회귀
 - 로지스틱 회귀
- 서포트벡터머신(SVM)
 - 앙상블 방법
 - 군집 알고리즘
- 주성분분석(PCA)
- 특이값분해(SVD)
- 독립성분분석(ICA)



〈구글포토〉
얼굴자동인식
서비스

흑인의 얼굴을
고릴라로 인식

인공지능 윤리의 특징: 사회 자본의 중요성 부상

시장자본만으로 발전이나 이익이 얻어지지 않는다

어떤 가치를 위해서
AI 가 개발되어야 하는가?



Session II- 토론

한희원 교수 - 동국대학교 법학

유럽연합(EU)의 인공지능에 대한 동향

동국대학교 법과대학 교수 한희원

1. 머릿말

인공지능(Artificial Intelligence)은 제4차 산업혁명의 꽃으로 인류의 미래 기술이다. 미래 인공지능(AI)의 궁극적인 목적은 초지능이나 제3의 종일 개연성이 높다. 초지능은 호모 사피엔스보다 지적으로 수십억배 더 똑똑한 자율판단의 스마트 지능을 의미한다. 미래 초지능의 마지막 남은 핵심 쟁점은 초지능과 인간이 공존할 것인가 아니면 초지능이 인간을 지배할 것인가의 문제이다.

인공지능이 앞으로 어떠한 발전의 과정을 겪게 될지는 누구도 예측하기 어렵다. 하지만 인공지능의 윤리성을 확보하려는 노력은 결국 인간이란 무엇이며 자유와 책임이 현대사회에서 어떤 의미가 있어야 하는지에 대한 인간의 인식에 의존한다고 할 수 있다. 최초의 종자인공지능은 인간의 몫이기 때문이다.

인간을 닮은 인공지능을 개발할 때 부딪히는 가장 큰 문제 중의 하나가 바로 인공지능이 단순히 인간의 삶을 위한 수단으로서만 작동하게 할 것인가 아니면 그 이상의 자율적 주체자의 지위를 갖도록 할 것인가의 문제이다. 인공지능을 단순히 효율적 계산이나 이성적 추론의 임무를 수행하는 것으로 한정한다면 인공지능은 인간의 삶을 도울 큰 진보로 간주될 수 있을 것이다. 그러나 인공지능이 자기 삶의 의미와 목적을 추구하는 독립적인 능력 존재자로 도약한다면 인간 지능은 인공지능으로 대체되고 그 과정은 새로운 인간화 작업이라고 할 수 있다.

2. 인공적 도덕행위자(AMA)는 가능한가?

1. 인공지능의 자의식 부인론 - 중국어 방 논변(Chinese Room Argument)

1. 준비된 방(房) 안으로 중국어를 전혀 할 줄 모르는 사람(A)이 들어간다.
2. 방에는 중국어 매뉴얼이 있어서 중국어 질문이 들어오면 A는 매뉴얼을 참조하여 적절한 답변을 만들어낸다. 중국어를 전혀 모르는 A가 하는 작업은 그림 짜 맞추기와 다를 바 없다.
3. 중국어를 전혀 모르는 A는 중국어 매뉴얼을 참고해가면서 방 밖의 중국인과 필답을 한다. 하지만 A 자신은 중국어 대화의 내용을 단 한 개도 이해하지 못한다.

존 설은 반문한다. A가 매뉴얼에 따른 구성적인 필답을 한다고 하여 중국어에 대한 인식이 있다

고 할 수 있겠는가? 중국어방의 A는 중국어 질문과 답변에 대해 이해(理解)가 아니라 단지 반응(反應)을 하는 존재라는 것이다.

그러나 미래학자 레이 커즈와일은 존 설의 논변을 반박한다. 그는 인공지능 알고리즘의 독자적인 의식을 긍정한다. “인간은 뇌가 있으므로 생각이란 것을 할 수 있다. 하지만 인간의 뇌도 결국 뉴런들의 집합일 뿐이다. 인간의 뉴런 하나하나도 생각할 수는 없다. 뉴런 하나하나는 물리, 화학 법칙에 따라서 기계적으로 작동할 뿐이다. 뉴런이라는 이런 부품들이 모여서 ‘뇌(腦)’라는 하나의 시스템’을 이룰 때 비로소 생각이라는 것을 할 수 있게 된다.

중국어 방 논변 또한 시스템을 구성하는 요소들인, '방', '방 안의 사람', '매뉴얼' 각각은 중국어를 이해하지 못해도 괜찮다. 이것들이 모여서 ‘중국어 방’이라는 하나의 전체적인 체계를 이뤄서 중국어로 아무런 문제없이 소통할 수 있다면 이 '중국어 방'이라는 시스템은 중국어를 이해하고 따라서 의식(意識)이 있다고 해도 무방한 것이다.

2. 인공지능의 자의식 가능성론: 튜링 테스트(Turing Test)

- 기계가 생각할 수 있는가?
- 기계가 행동주의적 지능 검사를 통과할 수 있는가?
- 기계가 인간을 속일 수 있는가?
- 2014년 6월 영국 University of Reading이 주최한 Turing Test 2014에서 Eugene이라는 프로그램이 마침내 심사위원 30명 가운데 10명을 속여 튜링 검사를 통과했다.

3. 인공지능(AI) 윤리론

아시모프의 로봇 3 법칙(Asimov's three laws), 머피(Murphy)와 우즈(Woods)의 로봇 3 법칙, 스튜어트 러셀(Stuart Russell)의 로봇 3 법칙, 유럽연합 로봇 윤리 9대 원칙

■ 유럽연합(EU) 로봇 윤리 9대 원칙

- 가. 절대적 인간 존엄성 수호 - 로봇에 의한 인명피해 방지(Protecting humans from harm caused by robots)
- 나. 로봇 돌봄의 거절 존중(Respecting the refusal of care by a robot)
- 다. 로봇에 맞선 인간의 자유 보호(Protecting human liberty in the face of robots)
- 라. 로봇에 의한 사생활 침해로부터의 보호(Protecting humanity against privacy breaches committed by a robot)
- 마. 로봇에 의해 처리되는 개인 데이터의 관리(Managing personal data processed by robots)
- 바. 로봇에 의한 조작 위험으로부터의 보호(Protecting humanity against the risk of manipulation by robots)
- 사. 사회적 유대 관계의 해소 방지(Avoiding the dissolution of social ties)
- 아. 로봇공학 발전 혜택에 대한 동등한 접근(Equal access to progress in robotics)

자. 향상된 기술에 대한 인간의 접근 제한(Restricting human access to enhancement technologies)

■ 유럽연합 로봇 시민법: 전자 인격(electronic personhood) & 로봇세

3. 맺는말

인공지능(AI)과 인간이 어떻게 서로의 영역을 유지하며 공존할 수 있을까? 명백한 것은 인공지능(AI) 관련 법규가 거의 없다시피 한 현재의 법제도는 대폭적인 변혁을 겪을 것이다. 법학교육도 변화할 것이다. '라이프 3.0'이라는 문제는 인류 공통의 과제로 자국 중심주의나 제로섬 게임으로는 의미가 없어질 것이다. 초국가적인 국제사회의 노력이 훨씬 더 긴요해질 것이다.

인공지능은 이미 생활 전반으로 파고들며 인간과의 동거를 시작했다. 각종 서비스 로봇이 등장하여 사람을 대신하여 하루 24시간 1년 365일 맹활약하고 있다. 인공지능은 이미 인간의 삶에 깊이 들어와 있고 인간의 삶에 영향을 미칠 뿐만 아니라 윤리적·법적 효과를 일으킬 수 있는 지점에 이미 와 있다. 자율주행차는 사실은 철학하는 차이다. 치명적 자율무기는 누구를 죽일 것인가를 결정하는 윤리적 판단자이다.

새로운 법학교육과 윤리교육이 절실하다. 숙고하면 할수록, 초(超) 현대과학기술의 정수인 인공지능의 밑바탕을 인간의 윤리와 법이 단단히 지탱하고 있어야 한다는 사실이다.



Session II- 토론

김지원 팀장 – 과학기술정보통신부
지능정보사회추진단 인공지능정책팀

[토론문]

인공지능 윤리에 관한 정책적 대응방향

김지원(과기정통부 지능정보사회추진단 인공지능정책팀장)

최근 AI와 자동화 시스템이 일상으로 확산되면서 인간사회의 안전에 영향을 미치고 있어 ‘AI의 윤리적 고려사항’에 대한 구체화의 필요성이 증대되었다. 특히 구글은 ‘AutoML’ 기술을 통해 알고리즘 개발 업무까지도 인간에서 AI로 이관하려는 시도 중이다. 이에 따라 알고리즘의 조작가능성, 의사결정의 편향성 등의 부작용은 AI의 신뢰성 문제와 직결되는 문제로서 중대한 현안으로 다루어져야 할 과제이다.

주요국 정부 및 국제 기구는 AI의 안전성과 일자리 대체, 법적 책임, 인간 고유성 담보 등 윤리적 문제에 대응하기 위한 사회시스템 및 규범 마련에 적극적이다. 일례로 미국 연구계의 아실로마 원칙, 일본 인공지능학회의 윤리가이드라인, OECD의 윤리 가이드라인 합의 권고안 등이 잇달아 발표되었고, 우리 정부도 작년 5월 31일 ‘윤리헌장’과 ‘지능정보사회 윤리가이드라인’을 마련하여 이러한 추세에 선제적으로 대응해왔다.

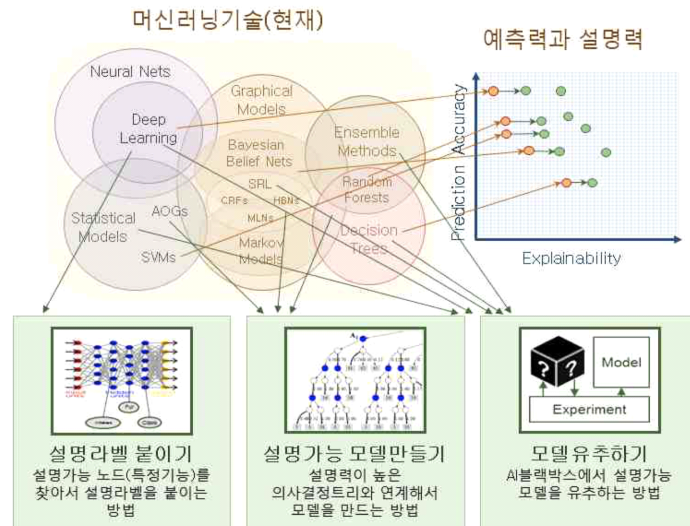
그러나 현 시점에서 윤리적 이슈는 규범적 차원에서 원론적인 논의 수준이며, 기술의 발전 양상과 인공지능이 적용된 구체적 서비스 모델을 감안하여 더욱 실효적으로 규범화할 수 있는 방법론을 찾아갈 필요가 있다.

1. 기술적 측면

인공지능 윤리에서 제기되는 원칙이나 가이드라인을 준수하기 위해서는 그와 관련한 기술적 뒷받침이 필수적이다. 예를 들어, 인공지능의 의사결정에 대해 이용자에게 설명요구권을 부여해야 한다고 제안되고 있으나, 딥러닝 방식의 인공지능 기술은 알고리즘 작동 과정에서의 설명을 기술적으로 완벽히 재현하는 것은 현재의 기술 수준에서는 쉽지 않다. 다시 말해, 현 기술 수준에 대한 고려없이 윤리 규범으로 규율할 경우, 규율 자체가 현실성이 없거나, 설명 가능성이 없음을 이유로 딥러닝과 같은 최신기술의 사용을 제한하는 문제를 야기할 수 있는 것이다.

정부는 혁신성장 프로젝트를 통해 2017년에서 2021년까지 “의사결정 이유를 설명할 수 있는 인간 수준의 학습 추론 프레임워크(XAI)” 연구과제를 추진 중에 있다. 이러한 Explainable AI에 관한 연구는 크게 ① 기존 알고리즘을 설명 가능하도록 수정하는 방향과 ② 설명가능한 완전히 새로운 모델을 개발하는 방향으로 추진되고 있다. 다만, 현 시점에서 딥러닝처럼 정확성 높은 복잡한 모델은 추론 및 결과에 대한 입증에 한계가 존재하기 때문에 고성능의 학습 과정에서 이러한 설명가능성이 완전히 실현되기까지는

지속적인 연구개발이 필요할 것이다. 알고리즘 설계 단계에서의 접근뿐만 아니라 훈련 데이터 자체에서 차별적인 패턴이 있는지를 검토한다거나 차별적인 데이터 세트에서 비차별적 알고리즘 개발 등 데이터 자체를 검증하는 방법론적 연구가 병행될 필요가 있다.



이 외에도 인공지능의 안전성 확보를 위해 비예측적, 비의도적 행위의 발생가능성을 낮추는 시스템 설계, 고도화된 AI시스템을 쉽게 이해하고 작동시키는 디버깅 툴 및 테스트 환경 개발, ‘고장모드’ 작동 시 적용 회피방법, AI시스템 작동여부 확인방법, 오작동 시 시스템 정지·수정 방법 등의 기술도 필요할 것으로, 정부도 원천기술 측면에서의 기술개발 연구를 지속 추진해나갈 것이다.

2. 법제도적 측면

인공지능의 자율적 추론과 판단에 대한 법적, 제도적 틀에서의 준비도 필요하다. 최근 국내외에서 인공지능의 활용에 따른 법적 관계나 책임 귀속의 문제 등이 활발히 다루어지고 있다. 이를 통해 인공지능의 사회적 수용 과정에서 발생하는 다양한 양상과 양태를 이해하고 파악하는 것을 토대로 점진적인 대응책을 마련해 나갈 필요가 있다. 즉 인공지능의 발전을 경제적 요소의 관점에서만 바라볼 것이 아니라 위험의 적절한 배분과 책임 귀속 주체의 명확화와 같은 규범적 관점에서의 제도화가 수반되어야 할 것이다.

현재 우리나라의 경우, 국가정보화 기본법 전면개정안(변재일 의원 대표발의)에서 지능정보기술개발자 또는 활용하는 자에게 서비스 오작동을 방지하기 위한 보호조치 의무와 외부에서 긴급하게 강제 종료(이른바 ‘킬스위치’ 방식)할 수 있도록 설계해야 할 의무 등의 내용을 포함시킨 바 있다. 그리고 과기정통부는 '18년4월 ‘지능정보사회 윤리 가이드라인’은 공공성, 책무성, 통제성, 투명성이라는 4대 원칙을 제시하여 지능정보 기술 및 서비스 개발자와 공급자의 책임 윤리 강화와 이용자의 오남용 방지를 위한 세부 지침을 마련하였다.

다만, 이러한 지침이 실제 생태계 내에서 효과적으로 발동하기 위해서는 구체적인 서비스 모델을 감안하여 윤리규범의 실효적 적용방식을 고민해나갈 필요가 있다. 일례로, 과기정통부는 인공지능 학습용 데이터셋을 구축·개방하여 민간의 다양한 인공지능 서비스 개발을 지원하고 있는데, 금년도에 신규로 개방 구축할 데이터셋 중 하나인 ‘위험물에 대한 X-ray 이미지’의 경우, 위험물·도구의 자동판별기술을 고도화하기 위해 데이터셋의 필요성이 있으나, 이를 회피하기 위한 기술개발의 목적으로 당초 취지와 달리 사용될 우려도 존재한다. 이에, 동 데이터셋은 다른 데이터셋과 달리 데이터를 필요로 하는 자에 대한 검증절차를 거쳐 제한적으로 제공할 계획으로 선한 목적으로 인공지능 기술이 개발될 수 있도록 할 예정이다.

이외에도 향후에는 더 많은 인공지능 서비스가 나타날 것인데, 생명신체에 영향을 미치는 것인지, 재산적 가치에 막대한 영향을 주는 것인지, 기업에 효율성 제고 수준인 것인지 등에 따라 최소한의 규율 수준을 더 세분화할 필요가 있으며, 이는 구체적인 서비스 모델을 두고 계속적으로 연구해 나갈 필요가 있다.