GHR 2020

AI and Humanism

Jungwon Lee
(University of Seoul)

근대 휴먼 시대

근대혁명(17~18세기)

- 근대적 개인, 근대적 사회, 인본주의(휴머니즘, 개인의 주체성 존엄성 중시) 탄생
- 인간 본성의 핵심요소-이성, 감성, 도덕성, 가치, 자의식, 자유의지 등-에 대한 본격적인 철학 논의 시작
- 인간과 신, 인간과 인간, 인간과 자연, 인간과 기계의 관계도 인간중심적 관점에서 논의
- 인간중심주의 : 인간과 인간이 아닌 다른 것들 사이에 경계 명확, 다른 모든 것들은 주체인 인간을 중심으로 그 주위에 객체로서 대상화

근·현대 과학기술문명

- 인간은 객체인 자연을 탐구하는 주체(인간중심주의)
  - (과학) 자연의 모든 정보들은 인간이 설계한 관측 장치나 실험도구에 의해 인간이 감지할 수 있는 정보 형태로 수집・분석 ; 자연의 모든 법칙과 현상들은 인간이 만들어 낸 언어 및 개념체계에 의해 규정・해석
  - (기술) 도구, 개발의 주목적은 인간 생활의 풍요로움과 윤택함

- 인간과 기계의 상보적 결합(확장된 인간중심주의)
  - 기계가 없다면 인간과 세계는 단절, 인간은 더 이상 세계를 탐구하는 인식주체가 될 수 없음
  - 나의 신체 일부이자 자아를 확장시켜 주는 장치들

    : 감각지각 능력 향상(망원경, 센서 등), 나의 기억 보존 또는 확장 (메모리, 스마트폰 등)
  - 인간이 도달할 수 없는 미시세계와 초거시 우주에 대한 정보 제공 장치들

    : 이의 해석을 통해서만 세계에 대한 지식 획득 가능

포스트휴먼 시대

21세기 과학기술문명

- 인간의 육체적 활동을 뛰어 넘어 인간의 정신적 활동까지 대신하는 인공지능 발전 (현재)
  - 영국 레딩대의 튜링 테스트(2014.06) : 컴퓨터 프로그램('유진구스트만' : 유진이라는 13세 소년 모의)이 세계 최초 튜링테스트 통과. 컴퓨터 프로그램이 사람처럼 생각할 수 있는가?
  - 알파고 제로(Zero)버전(2017) : 비지도자율학습(기보 학습 하지 않음), 알파고 리(Lee)에 100전 100승, 알파고 마스터(Master)에 100전 89승
  - AI 판결시대
    - ✓미국 대법원의 AI 판사 컴퍼스(Compas) : 재범가능성 판단 및 형량 결정에 활용
    - ✓AI 변호사 로스(Ross) : 2016년, 미국 뉴욕 대형 로펌(베이커 & 호스테틀러) 취직 (초당 10억건 법률문서 검토 분석 등)

■ 21세기 과학기술문명

- IBM Watson For Oncology와 Watson for Genomics
  - ✓환자 개별 특성에 맞춘 암 진단 및 치료 서비스 제공
  - ✓2019년 ASCO(America Society of Clinical Oncology, 미국임상종양학회)' : 인도 마니팔 병원, 왓슨 포 온콜로지가 제공한 정보에 기반해 1,000건의 유방암, 폐, 대장암 환자 사례 중 13.6%의 치료방법을 변경

▪ 인공지능 로봇 사례들(미래)

- 자율주행 자동차
- 인공지능형 군사 킬러로봇 : 기계 자율적 판단에 의한 살인 용인
- 인공지능형 섹스로봇
- 인공지능형 감성로봇 : 인간의 감정을 읽고 표현, 더이상 낯선 타자가 아님, 돌봄로봇, 연인로봇 등으로 활용가능

21세기, 포스트 휴먼 시대

- 기계의 인간화 경향
  - ‘로보 사피엔스’의 등장 가능성(미래) :페이스 달루이시오(Faith D’aluisio)와 피터 멘젤(Peter Menzel),『새로운 종의 진화, 로보사피엔스(2000)』
  - 루발 하라리(Yuval Noah Harari) :『사피엔스』(2014) (유인원에서 사이보그까지)
  - 강인공지능의 등장 : 인간은 아니지만 전반적으로 인간처럼 생각하고 느끼고 말하고 행동하는 인공지능 로봇 ➡ 인간 같은 능동적 행위자로서의 비인간적 인격체
- 인간의 기계화 경향
  - 신체의 일부를 인공장기 혹은 로봇 팔·다리 등으로 대체하는 신체기술, 유전자가위를 활용한 생명편집기술, 강화된 인공세포나 혈액을 생산하는 나노기술 등을 이용한 인간의 능력증강(Human Enhancement) 혹은 증강인간의 출현 가능성
  - 인간의 신체 일부가 기계로 대체된 사이보그 등장

21세기, 포스트 휴먼 시대

- 포스트 휴먼 시대 : 인간과 기계의 탈 경계 시대
  - 그 동안 인간에게만 고유한 것으로 인식됐던 능력들(감성, 이성, 행동 등)이 인간이 아닌 기계에서도 구현 가능한 시대(기계의 인간화 경향)
  - 기계가 인간 외부에서 객체인 도구로 머무는 것이 아니라 인간의 몸과 마음의 일부로서 주체가 되는 시대(인간의 기계화 경향)
  - 인간은 비인간적인 다양한 요소들과 더 이상 분리되지 않고 광범위한 연결을 통해 형성된 관계적 존재
    - ✓인간은 사물·기계들과 서로 연결·교차하면서 상호작용하는 관계 네트워크의 한 노드
    - ✓인간은 다른 생물학적 행위자 및 기술적 행위자들과 상호 연결되어 존재하면서 이들과 함께 정보권(inforsphere)을 공유하는 정보적 유기체로서의 인포그(inforg) (정보철학자 플로리디(L. Floridi))

21세기, 포스트 휴먼 시대

- 우리는 이를 어떻게 이해하고 받아들여야 할까
  - 인간과 기계의 탈경계화는 기계를 바라보는 인간의 시각을 바꾸고, 인간과 기계의 관계를 새롭게 정립하며 궁극적으로 휴머니즘에 변화를 야기할 것으로 예상
  - 지금까지 인간중심주의 관점을 견지해 온 철학적 견해들은 이런 포스트 휴먼 시대를 올바로 이해하고 판단하는데 많은 어려움
    - ✓특히 인공지능(로봇)과 같은 인공적 기계 존재자를 단순한 도구로 보는 인간중심적 관점 때문
    - ✓ 탈 인간중심주의 관점에서 하나의 인격체로 보는 철학적 성찰 필요
  - 탈 인간중심주의 관점에서 휴머니즘 재정립 필요

: 인간중심주의 휴머니즘 → 탈인간중심주의 휴머니즘 → 포스트 휴머니즘

비인간 인격체

왜 고민하는가

- 점점 인간을 닮은 인공지능(로봇)의 등장 : 인간의 끝없는 욕망, 인공지능(로봇)에 대한 인간의 의인화 경향 (예, 반려견의 사례)
  - AI의 정체가 무엇인가, 스스로 학습하여 똑똑해 지고 인간처럼 생각하고 행동하는 이들을 우리는 어떤 존재자로 봐야 할 것인가 ➡ AI의 존재적 지위 고민
- 자율적인 사회적 행위자로서 AI 로봇의 등장 (예, AI 의사/판사, 돌봄로봇 등)
  - 인간 생활세계는 어떻게 달라질 것인가, 달라진 생활 세계 어떤 윤리적 문제들,사회적 문제들이 발생할 것인가 ➡ 과거에 전혀 경험하지 못했던 새로운 유형의 다양한 윤리적·사회적 문제 직면, 사회제도(법률 등)적 측면에서 중요
  - 사회 구성원으로서 어떻게 받아들일 것인가, 인간은 이들과 어떤 관계를 맺을 것인가, 우리는 이들과 어떻게 공존할 것인가 ➡ 인간과 인공지능의 공존이라는 새로운 시대적 과제 부상

■ 무엇을 고민할 것인가

- AI 로봇의 존재론적 본성과 지위 : ‘전자 인격체’
  - ‘전자 인격체’(Electronic persons) : 2016년 5월 31일, 유럽의회 법률위원회, 로봇에 관한 민법규정 초안 제안, 가장 정교한 자율형 로봇의 잠재적 법적 지위("전자 인격성”(electronic personality))
- 인격성에 대한 분석 과 이해
  - 인격성(이성, 감성, 자율성, 도덕성, 자유의지 등)에 대한 과학적 해명 : 인지과학/인공지능/뇌과학, 정도/수준/등급에 따라 인격성 세분화
  - 인격성의 다양한 층위에 대한 철학적 규정 : 개념의 정교화 및 확장, 이를 토대로 다양한 인격체 인정 및 포용
- 인격성의 관계적 형성
  - 자신과 타자와의 상호의존적 관계로 형성, 사회화 과정에서 개인 정체성 형성
  - 고립·분리 대신 연결, 일방성 대신 상호성, 내재적 상태(Being)보다는 역동적인 변화 과정(Becoming) 강조.
- 다가올 AI 시대의 인간의 정체성 및 휴머니즘의 변화

탈인간중심주의 인격성

- 존 로크(J. Locke), 인격체(person)와 인간(human being) 개념 구분
  - 인간은 생물학적 종개념
  - 인격체는 이성에 기반 한 반성적 사고능력, 타자와 구분되는 자아(self) 형성능력, 자아 동일성을 확인하는 기억능력을 갖춘 지적 존재자
  - AI의 경우, 머신러닝으로 인간의 오랜 진화 및 경험 과정에서 축적된 빅데이터에 대한 자율적 학습 능력 : 경험론적 관점에서 존 로크의 조건 실현 가능(“자아는 백지상태에서 출발하여 무수한 경험을 통해 형성”)
- 피터 싱어와 같은 동물 윤리학자 : 고통을 느낄 줄 아는 동물도 인격성을 갖고 있음. 고통을 느끼는 능력(쾌고감수 능력)이 인격성의 중요한 기준, 오랑우탄 ‘산드라’ 사례 (2014년 12월 아르헨티나 법정 판결)
- 법조계의 ‘법 인격’ 개념 : 회사 법인과 같은 인공적인 조직에 법 인격성 부여, 이의 근거는 자율성이나 자유의지가 아닌 자기 소유권

탈인간중심주의 인격성

- 자율성의 문제
  - 세계는 인간이 예측하기 어려운 상황 전개로 언제나 열려 있는 만큼, 로봇에게 변화 상황에 적절히 대응하도록 자발적 학습 능력 부여 : 다양한 수준에서 스스로 무언가를 결정할 수 있는 “자율성” 부여
  - 알파고 제로 버전 : 기보 빅데이터에 대한 학습 없이 알고리즘 그 자체만으로 바둑의 간단한 규칙만 인지, 자체 피드백을 통한 강화학습(비지도학습) 수행, 판단과 결정과정 알 수 없음
  - 딥러닝이 인공지능에게 어떤 방식으로 자율적 판단과 행동의 가능성을 제공하는가? 인공지능에 부여된 자율성의 진정한 의미는? 인간이 아닌 존재자에게도 적용가능한 자율성 개념의 다양성은?
  - 자율성을 기능적 측면에서 수준/정도/유형에 따라 차등화 유형화(가령, 자동성(automaticity), 준자율성(semi-autonomy), 완전한 자율성(fully-autonomy))

■ 탈인간중심주의 인격성

- 도덕성의 문제
  - AI의 자율성 문제 ➡ AI 행위의 윤리 문제 ➡ AI의 도덕적 지위 문제
  - 소극적 의미의 AI윤리(인간의 윤리, AI 관련 윤리 지침들) vs 적극적 의미의 AI윤리(AI 로봇의 윤리, 도덕적 AI로봇의 출현)
    - ✓무어(J.Moor)의 도덕행위자 구분 : 윤리적 영향 행위자(ethical impact agent), 암묵적 윤리 행위자(implicit ethical agent), 명시적 윤리 행위자(explicit ethical agent), 완전한 윤리 행위자(full ethical agent)
  - 도덕성 개념의 세분화와 확장 가능성
    - ✓ 유형/정도/수준에 따라 도덕적 사고능력 세분화 : 가령 조작적 도덕성(설계), 기능적 도덕성(도덕적 추론), 완전한 도덕성(자기정당화). 인공적 도덕 행위자(artificial moral agent, AMA) project
    - ✓관계적 도덕성 개념 : 어떤 도덕성을 지녔는가➡ 인간과 어떤 도덕적 관계에 있는가. 도덕적 피동행위자 지위 부여 (능동행위자(agent) vs 피동행위자(patient))
    - ✓도덕성에 관한 성리학의 관점 : 성리학은 우주와 인간에 대한 통일적 세계관 제시. 사람, 사물 '인의예지신'(仁義禮智信)의 '오상'(五常)을 타고났지만 사물은 기질에 치우쳐 타고난 본성을 인간처럼 온전히 발현하지 못함 주장

탈인간중심주의 휴머니즘

인간중심주의 휴머니즘에서 탈인간중심주의 휴머니즘으로

- AI 로봇의 등장은 인간중심주의 휴머니즘에서, 탈인간중심주의 휴머니즘으로의 전환을 조심스레 요청
- 과도한 인간중심주의로부터 벗어난 휴머니즘
  - 인격성의 정도/수준/유형 등 다양한 층위 인정, 다양한 비인간 인격체의 존재 인정
  - 인간만이 아니라 비인간 인격체 모두를 복잡한 사회 관계망 안에서 다양한 상호 관계를 맺고 있는 사회적 행위자로, 사회의 구성원으로 간주
  - 그러면서도 인간과 비인간 인격체 간의 존재적 위상은 차등화 : 행위의 능동성에 따라 행위자 사이에 여러 층위를 두고, 인간이 비인간 인격체보다는 높은 존재적 지위를 지닌 것으로 간주.
  - 이러한 조건하에서 휴머니즘 강조

인간 의식의 지향성(intentionality) 사례들

- 주체(나)와 객체(외부 세계)의 구분, 몸과 의식의 이분법적 구분을 부정하고 나의 지향적 관점에서 세계를 봄
- 반려견의 사례
  - 인간 역사에서 개의 존재 : 야생늑대에서, 집 개로, 애완견으로, 우리의 삶의 동반자인 반려견으로 그 존재적 지위가 끊임없이 진화
  - 각 존재적 지위에 대응하는 개의 내적 속성이 있고 이것이 생물학적으로 진화한 결과가 아니라, 인간과 개의 사회적 관계, 즉 개가 인간의 생활세계에 끼친 영향으로 인해 인간이 개에게 특별하게 부여한 사회적 지위 및 관계가 끊임없이 변화한 결과
  - 개의 내적 속성이 무엇이든, 개의 존재가 인간의 삶에 어떤 가치와 의미를 형성하면서 영향을 준다면, 인간은 개에게 그 가치와 의미에 상응하는 존재적 지위 부여
- AI 로봇의 사례
  - 로봇 강아지 '아이보'의 장례식과 신사 안치 사건(일본), 섹스 리얼돌과 결혼한 사건(일본), 휴머노이드 로봇 '소피아'에게 시민권 부여 사건(사우디 아라비아), 로봇과의 성관계와 인간과의 성관계의 동등성 논란 등

■ 인간과 기계의 관계에 대한 생각의 전환

- 기계에 대한 도구적 관점
  - 기계는 외부로부터 각인된 목적에 따라 단순히 기능을 수행하는 도구 혹은 수단, 인공지능 기술을 제4차 산업혁명의 관점에서 보는 시선
- 기계에 대한 실존적 관점
  - AI 로봇처럼 기계는 이제 스스로 어떤 지향이나 목적을 갖고 세계를 구성할 뿐 아니라 인간의 실존 자체를 변화시키는 실존적 존재자
- 도구적 관점에서 실존적 관점으로의 전환
  - 기계는 단순히 인간 사회를 떠 받쳐주는 물리적 기반 ➡ 인간과 복잡하게 연결된 관계망 속에서 사회를 구성하는 하나의 행위자(actor 혹은 agency)
  - 기계는 더 이상 인간의 외부에 존재하는 객체가 아니라, 인간처럼 주체화가 가능한 존재 (예, 연인 로봇과의 결혼, AI 로봇의 시민권 등)

■ 탈인간중심주의 철학 담론

- 인격체에 관한 새로운 존재론적 담론
  - 철학, 인지과학, AI 공학, 뇌과학 상호간 통섭 연구
  - 철학에서 주장해 왔던 인격 요소들(이성, 감성, 자율성, 도덕성, 자유의지 등)을 과학적으로 설명. 다시 과학적 사실들에 바탕 하여 인격성 요소들의 본성에 대한 보다 심화되고 정교한 철학적 분석 및 규명(개념의 세련화 및 확장)
  - 인격 요소들의 정도/수준/유형 등 다양한 층위 인정, 이를 토대로 다양한 인격체 포용

- 행위자 중심의 새로운 윤리 체계와 이론 정립
  - 인간,기계를 모두 망라하는 행위자(agency 혹은 actor)의 윤리로 확장
  - AI가 인간을 대신해 의사결정을 했을 때 그 결과에 대한 책임은 누가 져야 하는가?
    - ✓ 인간? : 의사결정을 AI에게 맡기자고 결정한 사람? 아니면 AI가 그런 의사결정을 내리도록 알고리즘 구조를 설계한 사람? 아니면 제3자?
    - ✓ 인공지능? : 어떻게 책임을 질 것인가?, “설명가능한 인공지능”(XAI)과 책무성(Accountability) 부각

■ 포스트 휴머니즘에 대한 전망

- 현재 포스트 휴머니즘이나 트랜스 휴머니즘은 그 의미 자체가 너무나 폭넓고 다양
  - 철학자 페란도(F. Ferrando)의 구분 : 1)반인간주의(Antihumanism), 2)문화적 포스트휴머니즘(Cultural posthumanism), 3)철학적 포스트휴머니즘(Philosophical posthumanism), 4)포스트휴먼 조건(Posthuman condition), 5)트랜스휴머니즘(Transhumanism), 6)AI에 의한 대체(AI takeover), 7)자발적인 인간 멸종(Voluntary Human Extinction)
- 포스트 휴머니즘은 미래의 큰 흐름 혹은 경향성을 표현하고 있어 현실적 대안으로서 아직 적절치 않음 : 포스트휴먼 시대에 AI 로봇의 인간화 경향을 이해하고 해석할 수 있는 현실에 적합한 포스트 휴머니즘의 담론은 아직 미흡