인공지능에 관한 비판적 스케치

A Critical Sketch on Artificial Intelligence

저자 (Authors)
이재현
Jaehyun Lee

출처 (Source)
마르크스주의 연구 13(3), 2016.8, 12-43(32 pages)
MARXISM 13(3), 2016.8, 12-43(32 pages)

발행처 (Publisher)
경상대학교 사회과학연구원(마르크스주의 연구)
Institute for Social Sciences, Gyeongsang National University

URL
http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06761285

APA Style
이재현 (2016). 인공지능에 관한 비판적 스케치. 마르크스주의 연구, 13(3), 12-43

마르크스주의 연구
특집

인공지능에 관한 비판적 스케치

이재현*

이 글은 인공지능의 몇 가지 문제들을 비판적으로 스케치한다. 비판의 영역은 크게 세 가지인데, 그것들은 각기 기계 학습과 통계적 학습이론, 인공지능의 윤리적 문제들, 그리고 인공지능을 둘러싼 정치경제학적 문제들이다. 이 글은 현재 성취된 인공지능 기술을 존중하고 감안한다는 취지에서 '약한 인공지능'과 '사회적으로 신체화된 인지'를 전제하며, 또 부르주아적 방법론적 개인주의에 대해서 비판적 입장을 취한다.

기계 학습 및 통계적 학습 이론은 그 이론적 전제인 베이지언 확률이 주관주의적이라는 점에서 철학적으로 취약한 토대 위에 서 있다. 또 비즈니스 인텔리전스에 의해서 추동되는 데이터 마이닝은 인구의 상당수를 배제하거나 차별하고 있는 데이터에 의존하고 있다는 점에서 큰 문제를 안고 있다.

인공지능의 윤리에 있어서는 만약 인공지능 에이전트가 인간과 같거나 그 이상의 지적 능력을 갖춘다면 인간과 같은 윤리적 지위를 지녀야 한다. 그런데, 정작 경계해야 할 것은 군사적으로 이용되는 지능적 에이전트가 아니라 자본가 역할을 할 수도 있는 지능적 에이전트다. 현재의 지배-종속 구조가 타파되지 않은 채 인공지능 기술이 지배하는 사회가 도래하는 것이야말로 윤리적으로 큰 문제다.

정치경제학적 문제 영역에 있어서 인공지능 기술의 도래에도 불구하고 인간에게 고유한 것으로 남아 있을 가능성이 큰 분야들은 주로 구조화된 데이터나 알고리즘으로 처리되기 힘든 지식과 관련된 노동, 육체적 숙련을 요하는 노동, 감정적 접촉이 중요한 노동이다. 인공지능 기술 등이 마르크스주의에 대해 제기하는 도전적 과제들은 크게 한편으로 가치론, 그리고 다른 한편으로 자본주의의 구조적 위기론으로 나뉠 수 있다.

주요 용어: 약한 인공지능, 베이지언 확률, 개체 발생 비차별, 프로그래밍 노동, 인공지능 비판.

* 한국예술종합학교 영상원 강사, sunanugi@gmail.com

1. 머리말

이 글은 인공지능에 관한 약간의 문제들을 비판적으로 스케치한다. '스케치'라는 다소간에 비유적인 표현을 사용하는 까닭은 인공지능의 활동과 성과 등이 학제적이기 때문이다. 인공지능에 관해서 적절한 학술적 발화를 하기 위해서는, 적어도, 수학, 통계학, 컴퓨터 프로그래밍 각각에 걸친 상당한 지식이 필요하다. 또 때에 따라서는 인지과학, 신경과학, 진화 심리학 및 진화 생물학 등에 관한 지식도 필요하다.

게다가, 인공지능의 활동과 성과는 단순히 과학의 영역에 제한되는 것이 아니다. 그것들은 과학-공학-기술의 복합체이며, 또 그 성과의 상당 부분은 일상생활에 이미 알게 모르게 들어와 있다. 예컨대, 구글이나 아마존 등에서의 검색에서부터 신용 카드의 사용을 거쳐서 자동청소기의 작동에까지 인공지능 기술이 활용되고 있다. 실제로 활용되고 있는 것에서부터 조만간에 활용될 수 있는 것에 이르기까지 인공지능의 공학적이고도 기술적인 영역을 대강이라도 파악한다는 것은 매우 힘들다. 이런 점에서도, '스케치'는 불가피하다.

이 글은 소위 '약한 인공지능'의 입장을 취해서 인공지능 문제에 접근한다. '약한 인공지능'과 '강한 인공지능'의 구분은 미국 철학자 존 설에 의해서 최초로 이루어졌다. 전자는 "마음(mind) 연구에 있어서 컴퓨터의 주된 역할이 아주 강력한 도구를 우리에게 줌으로써, 우리로 하여금 좀 더 엄격하고 정밀한 방식으로 가설들을 정식화하고 검증할 수 있도록 하"려는 것이고, 후자는 단지 도구에 그치는 것이 아니라 오히려 "올바른 프로그램을 가진 컴퓨터들이 말 그대로 여타의 인지적 상태를 **이해**하고 보유한다고 여겨지는 취지에서, 적절하게 프로그램된 컴퓨터[자체]가 말 그대로 진짜로 마음"이라는 입장인 것이다 (Searle, 1980: 417. 강조는 설의 것, 이하 괄호 안 보충은 인용자의 것).

더 쉽게 정리하면, 강한 인공지능은 고차적이고 복합적이며 통합된 차원의 인간 지능과 맞먹는 수준의 기계적 지능을 가리키고, 약한 인공지능은 저차적이고 부분적이며 분리된 지적 기능을 컴퓨터가 수행하게 만드는 것을 가리킨다. 컴퓨터는 아주 방대한 데이터를 극히 빠르게 처리할 수 있는데, 약한 인공지능은 바로 이런 능력을 이용해서, 컴퓨터라는 기계로 하여금, 비록 단편적이

고 제한된 영역과 수준에서일지라도, 인간의 지적 기능 중에서 특정한 것을 자동적으로 처리하게 만드는 것을 목표로 삼는다.

그런데, 러셀과 노빅에 의하면, "인공지능 연구자들 대부분은 약[한] 인공지능 가설을 당연한 것으로 받아들이며, 강[한] 인공지능 가설은 신경 쓰지 않는다"(러셀·노빅, 2016b: 645). 약한 인공지능의 입장을 취한다는 것은, 비판적 논의를 해나가되, 인공지능 분야에서 현재까지 이루어진, 그리고 조만간에 이루어질 가시적 성과들을 가급적 존중하겠다는 뜻이다.

인공지능과 관련해서, 인지과학의 여러 패러다임들 중에서 이 글이 기대고 있는 것은 '신체화된 인지'다. 심광현은 톰슨과 바렐라의 논의에 기대서 인지과학의 역사를 3세대로 구분한다. '계산주의 - 연결주의 - 신체화된 역동주의'가 바로 그것이다(심광현, 2014: 450-463). 심광현이 소개하고 있는 바의 '신체화된 역동주의'는 보통 '신체화된 인지'라고 불린다. '신체화된 인지' 접근법은 인지가 두뇌에서만 일어나는 것이 아니라는 입장에 서 있다. 신체화된 인지 접근법은 인지가 인간의 몸 전체와 그 몸을 둘러싼 환경 사이에서 일어난다고 파악한다.

더 나아가, 나는 기존의 '신체화된 인지'라는 것이 추상적이고 원자론적인 신체를 전제하고 있다는 점에 관해서는 매우 비판적이다. 이 글은, 철저하게 인간의 신체가 본디 사회적, 역사적이라는 점을 중시한다. 내 입장은, 굳이 표현하자면, '사회적으로 신체화된 인지'다. 사회적으로 신체화된 인지의 접근법에서는 인간의 환경을 한편으로 자연사적이고 생태적인 차원으로, 다른 한편으로 사회적, 역사적, 문화적인 차원으로 나누되, 양자의 상호작용에 주목한다. 사회적으로 신체화된 인지의 입장에서는, 예컨대, 계급 관계, 젠더 관계, 네이션-스테이트 속에서의 에스닉한 관계 등을 고려해야 할 중요한 환경적 요인으로 간주한다. 인간의 신체는 바로 이러한 관계들로 이루어진 앙상블이라고 할 수 있다. 당연히, 이러한 관계의 앙상블인 몸 전체에 의해서 이루어지는 인지과정의 바탕에는 무엇보다 인간의 노동이 깔려 있고, 또 동시에 두 차원의 환경을 매개하는 언어 및 담론적 활동이 스며들어가 있다.

이러한 전제들 위에서 인공지능의 문제들을 스케치하되, 영미권에서 '인지과학' '인지과학의/과 철학' '인공지능의/과 철학'이라는 이름 아래 다루어온

많은 것들은 가급적 일부러 비켜가려고 한다. 한편으로, 그것들 대부분은 '신체화된 인지'로까지 나아가지 못한 채, 계산주의 내지는 기능주의나 연결주의 수준에 머물러 있다. 다른 한편으로, 그것들 거의 대부분은 기본적으로 '방법론적 개인주의' 패러다임 안에 갇혀 있다[1)]

'약한 인공지능' '사회적으로 신체화된 인지' '방법론적 개인주의 비판', 이 세 가지를 함께 묶어놓고 보면 상당한 긴장과 혼란, 그리고 알력과 갈등이 예상된다. 왜냐하면, '사회적으로 신체화된 인지' 패러다임은 그 자체를 일관되고 강력하게 주장하다 보면 '강한 인공지능'과 연결될 가능성이 크고, '방법론적 개인주의'를 일관되고 강력하게 비판하는 입장도 마찬가지로 '강한 인공지능'을 함축할 가능성이 농후하기 때문이다. 무슨 말이냐 하면, '인공지능'의 강약에 대한 선호는 '인간 지능'의 강약에 대한 선호와 느슨하게나마 대체로 맞물려 있는데, '사회적으로 신체화된 인지' 및 '방법론적 개인주의 비판'은 '강한 인간 지능'을 선호하는 쪽이기 때문이다.

아무튼, 이 세 가지 전제 모두를 고수하면서, 동시에 인공지능에 관한 실제의 활동 및 성과를 존중하되, 이를 비판적으로 다뤄나간다는 것은 매우 어렵고도 까다로운 일이다. 이 글 제목에 굳이 스케치라는 말을 사용한 또 다른 이유는 바로 이런 어려움과 까다로움 때문이다.

2. 기계 학습과 통계적 학습 이론

오늘날 인공지능의 여러 분야에서 가장 성공적이고도 유력한 성과를 보이고 있는 것은 기계 학습(machine learning)이다. 톰 미첼은 기계 학습에 관해서 다음과 같은 정의를 내렸다.

> 컴퓨터 프로그램은, 과제 T에 있어서 P로 측정된 성과가 경험 E와 더불어 증진되는 경우에, 일정 종류의 과제 T 및 성과 측정 P에 관해서 경험 E를 학습한다

1) 인지과학에서의 방법론적 개인주의에 관한 비판 및 '사회적으로 신체화된 인지'에 논의는 이재현(2016 b: 9-10)을 참조할 수 있다.

고 말할 수 있다(Mitchell, 1997: 2).

미첼은 이러한 조작적 정의를 통해서, 인공지능 및 인지과학의 영역에서 이제까지 제기되어온 철학적인 논쟁을 회피하려고 한다. 즉, "기계는 생각할 수 있는가?"라든가 "기계는 학습할 수 있는가?"와 같은 문제들 말이다. 미첼 정의의 또 다른 특징은 학습의 범위를 '일정한 종류의 과제'에 한정시키고 있다는 점이다. 즉 과제별로 특정적(task-specific)인 알고리즘에 한정해서 인공지능의 학습을 논하고 있는 것이다.

미첼의 정의를 떠나서, 상식적으로 쉽게 표현한다면, 기계 학습은 컴퓨터 시스템을 통해서, 즉 컴퓨터라는 기계를 통해서, 인간의 능력과 같거나 비슷한, 종종 실제로는 인간의 능력을 능가하는 학습을 '자동적으로' 수행하게 하려는 것을 뜻한다. 오늘날 기계 학습은 확률 모델 및 통계학적 방법에 의해서 주로 '빅 데이터'를 처리함으로써 학습을 수행한다. 빅 데이터를 통계학적으로 처리한다는 점에서 기계 학습은 '데이터 마이닝'과 같은 것이라고 말할 수 있다. 따라서 기계 학습과 데이터 마이닝을 구별하는 것은 실무적이고도 직업적인 차원에서만 가능하다. 즉 어느 분야에서 누가 하고 있느냐에 따라서 같은 일이 다르게 불리고 있는 것이다.

위튼은 데이터 마이닝을 '데이터로부터 구조적인 패턴을 찾거나 기술하는 것'으로 정의한 다음에 데이터 마이닝을 위한 기술, 테크닉, 알고리즘 등의 의미로 기계 학습이란 말을 사용한다(Witten, 2011: xxi 이하). 한편, 통계학적 분석을 이용한 데이터 마이닝 활용에 관한 최근의 한 핸드북에서는, 고전적 기계 학습을 제3세대로, 또 통계적 학습 이론을 제4세대로 구분하고 있다(Nisbet et al., 2009: 11-13). 물론, 니스벳 등이 제4세대 통계적 학습 이론의 구체적 사례로 거론하고 있는 것들은, 오늘날 인공지능 분야 일반에서는 그냥 '기계 학습'에 포함되어 다루어진다. 편하게 말한다면, 데이터 마이닝은 비즈니스 인텔리전스가 요구하는 실용적 예측을 상대적으로 더 추구하고 있다고 할 수 있다.

기계 학습과 관련해서 무엇보다도 가장 먼저 지적해두어야 할 것은, 기계 학습이 확률 모델과 통계학적 방법을 사용하는 한에서, 모든 기계 학습은 전통적으로 철학, 특히 인식론 및 진리이론에서 귀납법의 문제 내지 한계들로 지적

해 온 결함들을 불가피하게 품고 있다는 점이다.

좀 더 구체적으로, 우리는 기계 학습 모델 대부분이 밑에 깔고 있는 확률 이론을 문제 삼을 수 있다. 그것은 바로 베이지언 확률 이론의 문제다.

베이지언 확률 모델은 이론적으로 주관주의적이라는 한계를 갖고 있다. 즉 일상적 세계와 자연과학적 세계에서 "많은 문제와 관련되는 확률의 문제는 실제적이며 객관적인 것"임에 비해서 베이지언 확률은 "개인적인 믿음의 정도를 확률과 동일시하"기 때문에 주관주의적 확률 이론에 바탕을 둔 베이지언 모델에 의지해서는 "과학의 토대와 과학의 발전은 설명될 수 없을 것 같다"는 것이다(송하석 1998: 82).

이런 비판의 연장선에서, 길리스는 주관주의적 베이지언 확률 모델로부터 생겨난 베이즈 망(bayesian network)도 당연히 고전적 통계 이론에 의해서 보완되어야 한다고 주장한다(Gilles 2004: 214). 베이즈 망이라는 것은 각각의 확률변수를 노드(node)들로 만들고 각각의 노드들 사이의 부모자식 관계를 화살표로 연결한 것이다. 러셀과 노빅의 표준 교과서는 베이즈 망이 불확실성을 정량화하는 것으로 다루고 있지만(러셀 & 노빅 2016 b: 1-32), 길리스와 같은 포퍼주의자에게는 그렇지 않다. 이러한 비판적 입장은 윌리엄즈도 마찬가지이며, 윌리엄즈는 통계적 학습 모델 전체를 비판하고 있다(Williams 2001: 129-146).

대부분의 논리학자가 명제들로 이루어진 추론 과정의 절차적 타당성만을 문제 삼을 뿐 개별 명제 자체의 건전성(사실 여부)은 도외시하고 있듯이, 대부분의 통계학자 내지는 통계적 방법을 이용하는 연구자나 공학자, 기술자들은 주어진 통계 자료 자체의 사회적 유의미성을 도외시한다. 즉, 주어진 데이터가 사회적으로 불완전하고 결함이 있다는 점은 문제 삼지 않는 경우가 많다.

기계학습과 관련된 통계적 방법이 다루는 통계 자료들은 크게 봐서 정부의 공식통계이거나 아니면 소위 '빅 데이터'다. 그런데, 예컨대, 미국 정부의 실업 통계 중 불완전 취업자를 포함한 실업률(U6)에 문제가 많다는 것은 이미 레이건 정부의 재무 차관을 지낸 경제학자 폴 크레이그 로버츠가 지적한 바 있다(Roberts 2013). 정부 통계의 경우, 조사 방식에 큰 문제가 있는 것이다.

그 다음으로, 빅 데이터에도 큰 문제가 있다. 오늘날 빅 데이터는 민간 부문에서 상품의 생산 과정 및 판매 과정에서 자동적으로 집계된다. 그런데 이 빅

데이터 자체가 물신화되어 있다고 할 수 있다. 민간 부문에서 생산되는 빅 데이터는 정부 통계에 비해서 사회적, 경제적인 배제와 차별의 문제를 더 많이 안고 있다. 배제와 차별의 문제란 이런 것이다. 나처럼 신용카드도 없고 인터넷 쇼핑도 거의 하지 않는 사람은 빅 데이터에서 제외될 가능성이 크다. 물론 나 같은 사람의 상거래도 슈퍼마켓의 계산대를 통과하는 한 계산대의 바코드 리더가 읽어낸다. 그리고 읽어낸 그것은 어딘가의 데이터베이스에 빅 데이터의 일부로서 저장이 될 것이다. 나는 하루 거의 대부분을 컴퓨터 앞에 앉아서 일한다. 검색을 굉장히 많이 하는 편이다. 그런 한에서 구글이나 아마존이 만들어 내는 빅 데이터에는 분명히 내 인터넷 사용 정보가 들어가 있을 것이다. 하지만, 페이스북도 트위터도 인스타그램도 카카오톡도 전혀 하지 않는 사람은 빅 데이터 통계에 잘 잡히지 않을 것이다.

어쨌거나 인터넷이나 스마트폰 사용이 불가능하거나 이런 미디어들에 연결이 차단되어 있거나 접근이 거부되어 있는 사람들은 빅 데이터로부터 배제되어 버리는 게 틀림없는 사실이다. 자본주의 글로벌 체제의 하위에 속하는 사회, 또 어떤 사회에서든지 간에 소득이나 소비 수준이 매우 낮은 인구 집단, 그리고 상대적으로는 인터넷이나 SNS 미디어를 사용하지 않는 사람들은 빅 데이터로부터 제외되거나 배제될 가능성이 크다. 쉽게 말해서 신용 관련 데이터 자체가 크게 부족한 극빈층은 금융 거래의 적격 여부 판정에 있어서 지능형 에이전트 시스템에 의해 부적격으로 찍힐 수밖에 없다. 이러한 사례는 통계 수치라는 게 애당초 사회적으로 구성된 것이라는 점을 잘 보여준다.

백악관의 2014년 및 2016년 보고서, 미국연방거래위원회의 2016년 보고서, 그리고 프랭크 파스콸레의 『블랙박스 사회』에서 공통적으로 지적하고 있는 것이 바로 빅 데이터 자체 및 빅 데이터 처리 과정에 의해서 계급·인종·성별·지역 등에 따라서 차별이 심각하게 행해지고 있다는 점이다(The White House, 2014, 2016; Federal Trade Commission, 2016; 파스콸레, 2016).

통계적 방법을 통한 지식 발견 및 획득에 있어서 결국 중요한 것은 그 지식이 어떠한 것이냐는 물음이다. 빅 데이터로부터의 통계적 처리를 통해서 자본가들이 얻고자 원하는 지적 패턴 내지 지식은, 일단 소비자 프로파일링을 해내고, 그럼으로써 상품이나 서비스를 더 많이 팔 기회를 찾거나 새로운 상품이나

서비스를 개발할 수 있는 기회를 찾는 것이다. 영국의 정보철학 연구자 플로리디는 이러한 현상을 '빅 데이터로부터 스몰 패턴으로'라고 요약하고 있다(Floridi, 2014: 306-308).

자본가의 관점이나 입장을 도외시하고 말한다면, 데이터 마이닝에 의한 비즈니스 인텔리전스가 발견하거나 획득하려는 지식 거의 대부분은 주로 자본가들에게만 필요할 뿐이다. 마르크스주의 및 비판적 인문과학, 사회과학 등이 지속해 온 실증주의 비판은 기계 학습이나 통계적 학습에 대해서 여전히 타당하다. 통계적 방법론 및 이에 의거한 기계 학습으로써 과학적 탐구를 대체하려는 경향은 경계되어야 한다.

통계란 말의 어원이 국가에 관련되어 있다는 점은 매우 의미심장하다. 영어 statistics는 1770년에 독일어 'Statistik'란 말로부터 만들어졌다(http://www.merriam-webster.com/dictionary/statistics). 또 여기서 어미 '-ics'는 '~에 속하는'이란 뜻이다. 결국 'statistics'란 말의 어원적 의미는 '국가의 통치 기술(국가적인 것)에 관련된 것'이다. 이러한 어원적 사태로부터 우리는, 영어 및 다른 유럽어들에서 통계라는 말 자체가, 통치와 연관된 관료주의 내지는 기술 관료의 통치적 합리성과 연관되어 생겨났다는 것을 잘 알 수 있다.

기계 학습에 의한 지식의 발견 및 획득에 관한 논의는 결국 우리가 학습 및 지식이란 개념으로 과연 무엇을 표상하고 추구하고 있는가 하는 문제로 귀결된다. 약한 인공지능의 입장에서 인공지능의 기술적 성과를 긍정적으로 받아들인다고 하더라도, 근미래에 인공지능 기술이 현단계보다 더욱더 범용적인 형태와 수준에서 보급된다면, 그때에 과연 인간은 무엇을 학습하고 어떤 지식을 추구해야 할 것인가 하는 점이 중요한 과제가 되는 것이다. 이것은 현존하는 디지털 디바이드의 문제, 그리고 데이터, 정보, 지식 및 정보통신기술과 관하여 현존하는 계급, 젠더, 인종, 국적 등에서의 불균등성과 비대칭성에 계속 주목하고 이것들로 인한 모순과 한계를 해결해나가는 일의 중요성을 새삼 일깨워준다.

3. 인공지능의 윤리

영미권에서 '인공지능의 윤리'라는 타이틀 아래 수행되고 있는 학술적, 지적 프로젝트에서 다루어지고 있는 문제들의 상당 부분은 아직 성취되고 있지 않은 과학-공학-기술적 차원에 속한다. 그런 점에서 인공지능 윤리의 문제들은 SF에서 재현된 테마나 소재들과 겹친다. 하지만, 그렇다고 해서, 이것들 모두가 아주 먼 미래에 속하는 것만은 아니다. 예컨대, 지능적인 에이전트가 야기하는 문제들에 대한 윤리적, 법적 책임을 누구에게 물을 것인가 하는 문제들의 일부는 이미 현재적이다. 예컨대, 자율 주행차가 빚어내는 교통사고에서 책임은 누구에게 있는가 하는 것이 그 사례다. 게다가, '로봇'이라고 할 때, 무엇인가 기계적이고 물리적인 몸체를 갖는 에이전트만 있는 것이 아니라 소프트웨어 수준과 형태의 에이전트들도 많다는 점을 잊지 말아야 한다.

인공지능 윤리의 문제들은 로봇공학의 성과가 빚어내고 있거나 빚어낼 문제들과 깊은 연관이 있다. 로봇공학의 이런 문제들은 '로봇 윤리(roboethics)'라는 이름의 탐구 활동 아래 포괄되고 있는데, 이 문제들은 주로 지능적 시스템을 갖춘 로봇을 구상하거나 설계하거나 생산하거나 사용하거나 이런 로봇들과 대면하게 되는 인간들이 겪어야 할 윤리적 문제들이다. 이와는 대조적으로, 그러한 지적 능력을 갖춘 에이전트들 자체가 갖거나 겪게 되는 윤리적 문제들을 다루는 영역은 통상 '기계 윤리(machine ethics)'라는 이름으로 포괄되고 있다. 도식적으로 정리한다면, 현재의 통상적 용법상, 영어권에서 로봇 윤리는 인간의 문제를 다루고 기계 윤리는 로봇의 문제를 다룬다.

이러한 통상적인 구분은, 기계 윤리와 로봇 윤리를 책 이름으로 내건 앤솔러지들이나 개인 저작의 구성에도 잘 나타나 있다. 앱니 등이 편집한 『로봇 윤리』(2012)에 수록된 여러 필자의 논문들은 디자인과 프로그래밍, 군사, 법, 심리학과 섹스, 의약과 돌봄, 권리와 윤리 등의 주제를 다룬다(Abney et al., 2012: v-vii). 지능적 로봇이 인간 삶의 여러 영역에서 야기할 것으로 예견되는 문제를 다루는 것이다. 차페스터스의 저작 『로봇 윤리』(2016)는 지능적인 로봇의 용도에 따라 의학, 보조, 사회, 전쟁 등으로 나누어 문제를 고찰한다(Tzafestas, 2016: ix-xiii). 한편, 앤더슨과 앤더슨이 편집한 『기계 윤리』(2011)는 지능적인 로봇,

혹은 컴퓨터 시스템이 과연 사람과 같거나 비슷한 수준의 윤리적인 에이전트일 수 있는가의 문제를 중점적으로 고찰한 뒤, SF적 접근법, 인공지능적 접근법, 심리학적-사회학적 접근법, 철학적 접근법 등으로 범주를 나누어 윤리적 문제들을 다룬다(Anderson & Anderson, 2011: v-viii).

실상, 로봇 윤리와 기계 윤리 사이의 구분은 그리 명쾌하지 않다. 서로 연관되어 있고 서로 매개되어 있다. 이 두 개를 서로 연결하는 질문이 바로, 기계에게 윤리를 가르칠 수 있느냐는 것이다. 이 물음은 다르게 표현하면, 윤리를 프로그래밍할 수 있느냐는 것이 된다. 인간 지적 능력의 총체성 및 상호연관을 강조하는 강한 인공지능 입장에서는 윤리를 프로그래밍할 수 없다고 보겠지만, 약한 인공지능 입장에 선다면 가능하다고 볼 수도 있다. 예컨대 칸트 윤리 체계나 공리주의 윤리 체계 등과 같은 특정한 것을 선택해서 프로그래밍하는 것에 관해서는 이미 상당한 연구가 진행되어 있다. 한편, 한국의 법학 분야에서도, 예컨대 베이즈 망을 이용해서 법적 논증을 처리하는 것을 다룬 논문이 나와 있다(고민조 외, 2014). 법적 추론을 수행하는 인공지능 시스템을 당장 사법 판결 업무에 직접 투입하는 것은 여러모로 무리가 따르겠지만, 법적 의사결정 과정에서 옆에 두고 참조하는 수준이라면 이미 부분적으로 도입되었다. 내 생각에, 윤리적, 혹은 법적 추론의 상당 부분은 일정하게 알고리즘들의 집합으로 대치할 수 있다.

현재 수준에서, 로봇 윤리와 기계 윤리를 통틀어서, 실질적으로 가장 무서운 것은 인간의 뇌와 컴퓨터 시스템을 직접 연결시키는 기술이다. 미 국방성의 방위고등연구계획국(Defense Advanced Research Projects Agency)은 두뇌-컴퓨터 인터페이스(brain–computer interfaces) 분야를 1970년대부터 개척해왔고, 지금도 하고 있다(재미한인과학기술자협회, 2015).

비슷한 맥락에서, 인공지능 기술을 군사적인 용도로 사용하지 말자는 캠페인을 벌인 스티븐 호킹 등의 활동이 주목된다. 이미 다른 영역, 예컨대 의학 분야의 뇌사 판정 문제라든가 안락사 문제, 그리고 유전자 공학 등에서 여러 가지 윤리적 문제에 관해서 실제로 성립해 있는 법적인 제약을 염두에 둔다면, 인공지능 기술의 군사적 이용을 법을 통해 금지할 수 있을지도 모르겠다.

하지만, 이런 금지의 실효적 가능성은 회의적이다. 이제까지 인류가 개발한

모든 중요한 기술은 예외 없이 군사적인 용도로 사용되었으며, 또한, 인터넷이 그러하듯이, 심지어 많은 기술들은 군사적인 목적에서 개발되었다. 실제 효과를 염두에 두고 말한다면, 아직 실현되지 않은 인공지능 군사 무기에 대해서 반대할 것이 아니라, 이미 실현된 화학 무기나 핵무기의 폐기에 반대하는 것이 더 낫다고 여겨진다. 스티븐 호킹 등의 반대는 그 선한 의도에도 불구하고 핵무기를 중심으로 한 미국의 글로벌한 군사적 헤게모니 체제 아래에서는 결과적으로 본의 아니게 지극히 위선적인 것이 되어버리고 만다. 인공지능 기술의 군사 무기화 등보다 오히려 기존의 핵무기, 생화학 무기, 스마트 폭탄을 탑재한 드론 등이 인류에게 더 긴급하고 사활적인 위협이기 때문이다.

그런가 하면, 반대로, 인공지능 분야에서 에이전트에 기반을 둔 모델들이 우리에게 약속하고 있는 바를 근미래의 가상적 공간에 투사해 본다면, 혁명적인 지능형 에이전트나 봉기 전문 지능형 에이전트를 상상하는 것도 얼마든지 가능하다. 우리 인간 대신, 시위를 하고, 고공 농성을 하고, 화염병을 던지고, 물대포를 맞고, 사찰을 당하고, 고문을 당하고, 때로는 테러나 납치를 당하고, 심지어 의문사나 암매장을 당하는 그런 에이전트들 말이다. 물론 이런 에이전트들을 '구입'하는 데에 엄청난 비용이 들게 되겠지만 말이다.

'인공지능의 윤리'에서 가장 핵심이 되는 것은 인간과 같은 지적 수준의 인공지능 로봇, 혹은 지능적 컴퓨터 시스템을 갖춘 에이전트에게 과연 윤리적 지위를 부여할 수 있는가 하는 문제일 것이다. 보스트롬과 유드코우스키는 이 문제와 연관하여 '개체 발생 비차별의 원칙(principle of ontogeny non-discrimination)'을 제출했다. 인간과 비교해서, 동일한 기능 및 동일한 의식적 경험을 갖고 있다면, 설령 발생 과정이 다르다고 하더라도 인간과 동일한 윤리적 지위를 가져야 한다는 것이다(Bostrom & Yudkowsky, 2014: 323). 즉, 기계가 인간의 능력을 갖춘다면 윤리적 영역에서 인간의 대접을 해주어야 한다는 것이다. 나도 그렇게 생각한다.

기계가 생각할 수 있다거나 없다거나, 혹은 기계가 윤리적일 수 있다거나 없다거나 하는 문제 및 물음에 대한 답은 결국 우리가 생각이라든가 윤리를 어떻게 정의하는가에 달렸다. 여기서, 주의해야 할 것은, 우선 첫째로, 인간이 윤리적인 기계를 만들어낼 수 있느냐는 문제와 인간과 동일한 능력을 갖춘 인

공지능 로봇에게 인간과 동일한 윤리적 지위를 보장할 수 있느냐는 문제는 일단 서로 다른 문제라는 점이다. 다음으로 주의해야 할 점은, 우리가 위 문제들 중 후자의 답을 구하는 과정에서 인간에 대한 안이하고 낙관적인 이해 방식, 그리고 부르주아들 특유의 추상적 인간중심주의에 의거한 기계관을 경계해야 한다는 것이다.

간단히 또 냉소적으로 말한다면, 지금까지의 역사 현실을 놓고 볼 때, 인간은 그리 존엄한 존재도 아니고 그리 자율적인 존재도 아니었다. 인간의 존엄성과 자율성은 인류 문명이 확보하고 성취해 낸 귀중한 보편적 이념 가치이기는 하다. 그렇지만 현실에서는 그 이념이 보편적으로 실현되지도 않았으며 또 지금도 실현되고 있지 않다. 기계와 인간 사이의 관계를 거칠게 요약해보자. 자본주의 생산관계 안에서, 그러니까 자본주의적 생산과정 안에서 인간-노동자는 기계-자본와 대면해왔고, 더 나아가서 기계-자본에 종속되어왔다. 쉽게 말해서, 노동자로서의 인간은 기계보다 못한 존재로 취급당해왔다. 자본주의 생산관계에서는 기계가 인간-노동자에 대해서 이미 더 우월하고 더 지배적인 지위를 차지하고 있는 것이다.

이런 조건과 상황에서, 지능적 기계가 과연 인간과 같은 도덕적 지위로 올라설 수 있느냐는 물음 자체는 매우 어리석은 것이다. 굳이 대답해야 한다면, 내 답은 '그렇다'이다. 20년쯤 뒤에는 내가 지금 쓰고 있는 글 정도는 지능적 에이전트, 그러니까 논문 집필용 봇이 대신 써 줄지도 모른다. 만약 그렇게 된다면, 그러한 에이전트를 어떻게 내가 하나의 도덕적 인격체로 대접하지 않을 수 있단 말인가.

소위 특이점(singularity)을 둘러싼 논의도 이런 관점과 발상에서 비판할 수가 있다. 버너 빈지라든가 레이 커즈와일 등은, 미래에 기술 변화의 속도가 급속히 변함으로써 그 영향이 넓어져 인간의 생활이 되돌릴 수 없도록 변화되는 지점을 상정하고 이를 특이점이라고 불렀다. 특이점이 언제 오는가 하는 점에 관해서는 사람마다 주장이 다른데, 그들 나름대로 여러 가지로 '예측'하고 있지만, 그들 주장의 근거를 냉정하고도 비판적으로 음미해보면, 그들 주장은 '예측'이 아니라 '예언'의 영역에 속한다고 할 수 있다. 미래학자들 중에서도 모디스(Modis)와 같은 사람은 특이점이 온다는 것에 대해서 전반적으로 회의적이다

(Modis, 2012).

문제는, 이러한 특이점 담론이 종종 좀 더 희석된 형태로, 그러니까 그 시점을 더 앞당겨서 바로 지금, 소위 '제4차 산업혁명' 담론으로 유포되고 있다는 점이다. 소위 제4차 산업혁명론은 2016년 다보스 포럼의 주제였다. 제4차 산업혁명이란 정보통신기술 융합이 만드는 산업혁명이라는 건데, 인공지능 로봇, 사물인터넷, 모바일, 3D프린터, 무인자동차, 나노·바이오기술을 응용한 새로운 제품들이 등장해서는, 산업과 경제, 사회, 정치, 학습, 생활 방식 등이 혁명적으로 바뀐다는 것이다[2)]

다보스 포럼은 글로벌 독점 자본가들 및 그들을 대변하는 이데올로그들이 자본주의 체제의 현재 및 미래와 관련하여 어젠다를 발굴하고 세팅하는 자리, 그리고 그 어젠다들을 전 세계적으로 선전하고 홍보하는 자리다. 그 어젠다들은 표면적, 구체적으로 무엇이든 간에, 본디 새로운 투자, 새로운 상품, 새로운 수요, 새로운 착취 방식 등에 관한 것에 불과하다.

부르주아 미래학의 핵심 주제는 결국 이노베이션과 경쟁이다. 이노베이션과 경쟁의 문제는 개별 자본가들로서도 늘 어렵고, 낯설고, 두려운 것이다. 미래학은 미래학이 펼쳐 보이는 새로운 세계에 대한 긍정적 이미지, 혹은 부분적으로 양가적이기도 하지만 결국에는 자본가들이 주도하는 세계라는 의미에서 긍정적인 이미지를 통해서, 계급 전체로서의 자본가들에게 약간의 긴장감이 곁들여진 안도감을 부여하고, 나머지 자본주의 인구 대다수에게도 미래학이 제시하는 상상적 세계에 편승할 것을 권유한다.

미래학은 늘 이데올로기의 종언을 말하고, 또 기존의 이념적 좌우 구분이, 즉 계급적 대립과 갈등이 해체되거나 녹아 없어질 것이라고 예언한다. 미래학은 지금 경쟁에서 뒤처지면 혹은 경쟁을 준비하지 못한다면 미래에 살아남지 못한다고 모두를 협박한다. 미래 사회는 모든 분야에서 바뀌므로 이미 시작된 경쟁적인 이노베이션에서 탈락하면 곧 다가올 미래에 누구나 루저가 되고 만다는 게 미래학 담론의 핵심이다. 그것은 결국 변형된 종말론이고 은폐된 최후의 심판론이다.

---

2) 이하 미래학적 담론 비판은 이재현(2016a)을 그대로 옮겨온 것이다.

미래학과 이에 동조하는 일부 이론의 가장 큰 문제는 새로운 미래 사회가 마치 자본주의가 아닌 것처럼 거짓되게 믿게 만든다는 점이다. 예컨대 인지자본주의론 등에서는 오늘날의 자본주의가 과거의 자본주의와 아주 다르다고 주장한다. 하지만, 분업, 협업, 기술 발전에 기초한 거대한 사회적 생산력과 그 성과를 자본가 계급이 독점한다는 점에서, 21세기 자본주의는 18세기 자본주의와 전혀 다르지 않다. 그 중추적 에너지가 증기든, 전기든, 인터넷 및 인공지능 기술과 결합한 재생 에너지든 간에 말이다. 또, 이노베이션과 경쟁의 내용과 형식이 무엇이든 간에 말이다.

나는 위에서 '개체 발생 비차별의 원칙'을 받아들인다고 말했다. 그 말은 이미 우리가 알게 모르게 인공지능의 성과와 더불어 살고 있고 또 근미래에는 더욱더 그러할 것이라는 현실 인식 및 예상을 그 밑에 깔고 있다. 이미 현재 소프트웨어 봇(bot) 등의 지능적 에이전트들은 우리들 각자의 과거 검색 내력을 감안해서 무엇인가 새로운 것을 검색해서 우리에게 억지로 제시한다. 또 우리가 이메일을 사용할 때, 다른 소프트웨어 봇은 스스로 알아서 스팸 메일을 분류한다. 이미 우리는 인공지능 기술에 파묻혀서, 혹은 인공지능 기술에 포위되어서, 혹은 인공지능 기술에 젖은 채 인공지능 기술과 더불어서 살아가고 있는 중이다.

인공지능 기술이 더 발전해서, 소위 '지능의 폭발'이 이뤄지거나 혹은 '인공일반 지능(artificial general intelligence)'이 도래하는 상황이 과연 올지 어떨지는 단정적으로 말할 수 없다. 다만, 설령 그런 상황이 온다고 하더라도, SF에서 자주 표상된 것과 같은 파국적인 결말이 올 건가 아닌가 하는 게 관건은 아니라는 점을 깨달아야 한다. 만약 그러한 단계에서 인공지능 에이전트들이 인간과 같거나 인간을 능가하는 지적 능력이나 윤리적 지위를 갖는다고 가정한다면, 오히려 중대한 문제는 그러한 인공지능 에이전트들이 출현한 상황에서, 인간 고유의 개체성 및 사회성, 역사성 등이 과연 어떻게 변형, 탈각, 승화, 해체, 소멸될 것인가 하는 점이다.

특히, 현재 존속하고 있는 바의 지배-종속 관계가 인공지능 에이전트들의 세계에서는 어떻게 변형될 것인가 하는 점이 중요할 것이다. 자본주의 축적 시스템 자체를 놓고 말한다면, 자본이 인격화되어 있다고 표상되어온 인간 자본가

보다는 인공지능 에이전트 자본가가 자본 축적에 훨씬 더 효율적이고 효과적일 것이라고 나는 예상한다. 일종의 판타지에 바탕을 둔 사고실험에서 말하는 것인데, 근미래의 인공지능 기술은 무엇보다도 먼저 자본가들의 '일자리(?)'를 없앨 것이다. 지능적 에이전트가 욕망과 의지와 코나투스 등을 갖게 된다면, 분명히 인간 자본가들로부터 소유권과 경영권을 박탈해버릴 것이다.

현재의 지배-종속 관계가 타파되지 않은 채, 인공지능 기술이 지배적으로 되는 근미래 사회로 우리가 들어서게 된다면, 그때 우리가 두려워해야 할 것은 로보캅과 같은 군사형 인공지능 로봇이 아니라 바로 그러한 인공지능 에이전트 자본가일 것이다. 아무튼, 온건하게 말한다면, 현재 존속하고 있는 지배-종속 관계에 대해서 근미래의 인공지능 기술이 어떻게 스며들어 올 것인가 혹은 반대로 인공지능 기술이 지배적일 사회에 대해서 현존하는 지배-종속 관계가 어떻게 배어들어 남을 것인가 하는 점이 결정적으로 중요한 문제다.

4. 정치경제학적 궁금증

인공지능 기술이 가져올 폐해 중에서 보통 사람들을 현재 가장 불안하게 만드는 것은 그것이 많은 일자리를 없앨 것이라는 점이다. 주류 부르주아 경제학자들 일부도 이런 입장에 서 있다(Sachs, 2015; Gordon, 2016). 사실 이 문제는 그리 새로운 것은 아니다. 산업혁명 이래 기계 일반의 채용, 그리고 20세기 후반 이래 극소전자기술 및 정보통신기술의 채용이 그러했다. 따라서, 인공지능 기술의 도입이 앞의 두 가지와 어떻게 질적으로 다른가 하는 점이 좀 더 명확히 밝혀진 다음에 인공지능 기술의 폐해가 논의되어야 합당하다. 반면에, 동시에, 현재로서 인공지능 기술은 아직 전기나 인터넷 등처럼 범용적인 기술이 아니라는 점이 명확하게 인식될 필요가 있다. 그러니까, 현단계 및 근미래 자본주의 테크놀로지 지형도 안에 인공지능 기술을 적절하게 배치한 다음에, 이러한 적절한 배치에 바탕을 두고서 탐구하고 논의해나가는 게 필요하다.

실업의 문제에 관해서는, 조만간에 쉽게 없어질 것 같지는 않은 일자리에 관해서 따져보는 게 여러모로 효과적일 듯하다. 현단계 인공지능 기술의 수준

과 관련해서 상식적인 관점에서 볼 때, 지식 발견 내지 획득이라는 점에서 크게 세 가지 범주로 나누어서 접근하는 게 가능할 듯하다. 이 범주란 지식 자체의 성격, 육체적 숙련, 감정적 접촉이다.

우선, 구조화된 데이터로 만들어지기 어려운 지식들이 있다. 아직까지는 인문, 사회과학이 다루는 상당수의 지식들이 그러하다. 언어 문제에 한정해서 본다면, 현단계 인공지능 기술 중에서 자연언어 처리 기술이라든가 시맨틱 웹 등이 이러한 한계를 돌파하기 위해서 시도되고 있기는 하다. 하지만, 아직 충분히 만족스럽지는 않다. 아직까지는 컴퓨터로 처리되지 못하고 있는 지식들, 그러니까 달리 표현하자면, '젖어 있으며 딱딱하지 않은(wet and not robust)' 지식들이 명백히 서로 아주 다른 많은 영역에 걸쳐서 엄존하고 있다. 예컨대, 지능적 시스템을 갖춘 에이전트들이 플라톤 및 아리스토텔레스의 저작을 공자 및 맹자의 저작, 그리고 뉴턴, 다윈 및 마르크스의 저작과 함께 능숙하게 다루는 것은 현재로서는 거의 불가능에 가까운 일이라고 여겨진다. 반면에, 인공지능 기술의 도전에 응해서, 구조화된 데이터로 만들어지기 어려운, 영역 별로 특정한 지식들의 위계 및 위상을 정립하고 그 근거를 제시하는 일은 오늘날 비판적 인문, 사회과학의 주요한 과제다.

그다음으로, 알고리즘들의 집합으로 처리하기 어려운 영역에 속하는 지식들이 있다. 이 지식들은 그 자체로 독립적으로 따로 떼어내기가 어렵거나 혹은 이런 지식을 다루는 지적 활동의 절차나 순서를 분절화내기가 힘들다. 이런 지식들의 일부는 기존에 소위 암묵지라고 불려 왔으며, 또 그것들의 다른 일부는 뚜렷이 그 범위와 깊이를 확정하기 힘든 배경 지식을 요구하는 것이다. 암묵지가 절대적으로 요구되고 전제되는 지적 활동, 혹은 암묵지가 활 여러 위상이나 측면에 걸쳐서 많이 녹아 들어가 있는 지적 활동에 관련된 지식들은 알고리즘으로 처리하기가 불가능하거나 매우 힘들다.

배경 지식과 관련해서 말한다면, 통계적 기술에 의해서 데이터베이스화될 수 있는 자연과학적 지식, 그리고 인문, 사회과학 영역의 일부 지식은 인공지능 기술에 의해 처리되기가 쉽다. 하지만, 생활세계와 관련된 배경지식은 처리되기가 아주 힘들다. 생활세계의 체험과 관련해서 사람들이 보통 가지고 있는 기억 능력 및 연상 능력 등은 고도로 문화적, 역사적이며, 또 그런 한에서 매우

개별적, 구체적이기 때문이다. 또 인간의 인지적 기능들은 각각 어느 정도는 편재적, 모듈적이지만, 반면에 상당 부분의 인지적 기능들은 두뇌 안에서 극히 복잡한 글로벌한 상호작용을 통해서 이루어진다. 개별적, 구체적이란 것은 쉽게 말해서 추상적 규정들의 복합체다. 그런데, 아무리 사소한 수준의 개별적이고도 구체적인 복합체라고 하더라도 여기에는 인간 생활세계에 고유한 '차원의 저주'가 아주 강력하게 작용한다. 햄릿의 대사 "죽느냐 사느냐"는 형식적으로는 1비트 문제로 보이지만 실존적으로는 소위 'NP-완전 문제'만큼이나 어렵다고 할 수 있다. NP-완전 문제란, 쉽게 예를 들자면, 암호를 풀 때 하나씩 대입해가면서 푸는 것 말고는 문제 해결의 특별한 알고리즘이 알려져 있지 않은 경우와 같은 것을 뜻하는 용어다.

마지막으로, 고도로 창조적인 지적 활동이 있다. 인공지능 기술에 의해 소설을 쓴다든가 영화 대본을 만든다든가 하는 게 지금 시도되고 있지만 이것들 역시 충분히 만족스럽지는 않다. 내 주관적 체험에 의거해서 말한다면, 이 중에서도 예컨대 사진을 찍는 일은 지적인 에이전트가 쉽게 할 수 있지만, 유화를 그리는 일은 매우 힘들 것으로 보인다. 마찬가지로, 소설을 쓰는 일은 영화나 TV 드라마의 시나리오를 만드는 일보다 어려워 보인다. 반면에, 시를 쓰는 일은 소설을 쓰는 일보다 더 쉬울 거라고 예측된다. 음악 분야에서도 작곡보다는 연주나 가창이 더 어려울 것이고, 연주에서도 서양 악기 연주보다는 농현(弄絃)에 크게 의지하는 전통 악기 연주가 더 어려울 것으로 보인다.

그다음으로는, 육체적 숙련을 요구하는 일들도 쉽게 사라지기 어려울 것으로 보인다. 이것은 현단계 인공지능 기술 및 로봇공학 기술의 수준 및 한계와 맞물려 있다. 이 점에서 현재 실제로 구현되고 있는 기술 수준은 한편으로는 자동차 생산 라인에서의 로봇들, 다른 한편으로는 세계로봇축구대회에서 경쟁하는 로봇들로 나타나고 있다. 러셀과 노빅의 구분에 따른다면, 전자는 '실세계 문제'에 속하고 후자는 '장난감 문제'에 속한다(러셀·노빅, 2016a: 86-92). 어느 문제 유형에 속하는 것이든 간에 육체적 숙련을 요구하는 과제는 현단계 및 근미래의 인공지능 기술이 쉽게 해결할 수 없다. 예를 들어, 거대 플랜트를 설계하는 일은 인공지능 에이전트에 의해서 성취되고 실현될 가능성이 아주 높지만, 정작 그 플랜트를 건설하는 과정에서 '노가다 잡부'의 일은 현단계 및

근미래 인공지능 기술에 의해서는 대신하기가 아주 힘들다. '노가다 잡부'의 일은, 지각, 운동 및 동작과 관련된 인공지능 로봇 기술의 현단계 성취 수준과 비교해볼 때, 나름 상당한 육체적 숙련에 의존하는 것이며, 또한 매우 범용적이라고 여겨지기 때문이다.

감정적 접촉이 핵심적인 기능이 되고 있는 일도 지능적 에이전트에 의해 대체되기는 어렵다. 무엇보다도, 감정은 한편으로는 우리 육체와 관련이 되어 있고, 다른 한편으로는 고차적이면서도 문화적이고 역사적인 정신 활동과 관련이 되어 있기 때문이다. 감정의 이 두 측면을 인공지능 기술이 제대로 해결하지 못하는 한, 감정 노동의 대부분은 인공적 에이전트가 대신하기가 힘들다. 감정적 접촉에 있어서, 소위 '일라이저 효과'라는 게 있기는 하지만 이것은 지능적 에이전트와 접촉하고 대면하는 인간이 자기의 상대가 인공물이라는 것을 모르는 상태에서 생겨나는 것이다. '일라이저'란 정신과 의사를 챗봇(chatbot)의 형태와 수준에서 시뮬레이션한 초보적 인공지능 프로그램의 이름이다. 많은 사람들은 '일라이저'가 인공물이라는 것을 깨닫지 못한 채 일라이저와 상담 대화를 진행했다. 하지만, 접촉하고 대면하는 상대가 인공물이라는 것을 명백히 알고 있는 경우에는 소위 '불쾌한 골짜기(uncanny valley)' 문제가 발생한다(https://ko.wikipedia.org/wiki/불쾌한_골짜기). 이 문제는 인간이 로봇이나 인간이 아닌 것들에 대해 느끼는 감정에 관련된 가설이다. 즉 인간에 거의 가까운 로봇이 실제로는 인간과는 달리 과도하게 이상한 행동을 보이기 때문에 인간과 로봇 간의 상호작용에 필요한 감정을 이끌어내는 데 실패한다는 것이다.

감정적 접촉이 핵심적인 유형의 노동으로서는 가사 노동과 돌봄 노동을 거론할 수 있다. 그런데, 명심해야 할 것은 이 유형의 노동들이, 인공지능 기술의 문제와는 별개의 역사적 맥락을 갖고 있다는 점이다. 개괄적으로 요약하자면, 이 노동의 영역에서는, 이미 상당한 정도로 상품화가 진행되어왔다는 점, 선진 자본주의 사회에서는 임금이 값싼 이주 노동력이 담당하고 있다는 점, 그리고 어느 사회에서나 이 노동이 제공하는 서비스의 혜택 내지 수용은 계급적으로 차별적으로 이루어져 왔다는 점을 꼽을 수 있다.

교육 노동이나 상담 노동, 그리고 판매 및 세일즈 노동 등도 비슷한 맥락에 속한다고 볼 수 있다. 인공지능 역사에 있어서 고전적인 전문가 시스템에서 제

대로 다룰 수 없었던 과제들을 수행하는 노동의 상당 부분도 이와 같은 맥락에 속한다. 고전적인 전문가 시스템이란 인공지능 역사의 초기 단계에서 성립했던 것인데, 특정 분야의 인간 전문가들이 가지고 있는 전문적인 지식을 정리하고 구조화시켜서 컴퓨터로 처리하게 함으로써, 일반인도 이 전문지식을 이용할 수 있도록 한 시스템이다. 고전적 전문가 시스템이 처리할 수 있는 과제는 예컨대 광물 탐사 등과 같은 영역에 한정되었다. 반면에, 아직 실현되어 나타난 것은 아니지만, 만약 지능적 에이전트가 인간과 같은 윤리적 지위를 갖게 된다고 할 때, 감정 노동을 수행하는 지능적 에이전트가 수용자(patient)로서 겪게 되는 감정적, 윤리적 영향의 문제도 예상해볼 수 있다. 이것은 소위 인공지능의 윤리 영역에서 최근 주요하게 대두되고 있는 화두다(Gunkel, 2012: 93-157).

마르크스주의 경제학 쪽으로 논의의 조명을 집중한다고 하면, 다음과 같은 여러 가지 문제들이 드러난다. 우선 인공 지능과 관련해서 참고의 기반으로 삼아야 할 마르크스 텍스트의 정합적 이해 문제가 있다. 마르크스는 이러한 것과 연관된 문제를 『그룬트리세』, 『1861-63년 경제학 초고』, 『자본』 등에서 논하고 있다. 의미심장한 것은 마르크스가 『1861-1863년 경제학 초고』에서 다음과 같이 서술하고 있다는 점이다. "이런 종류의 스스로 작동하는 자동기계(self-acting automaton) 개념은 예를 들면 프라이스(Richard Price)의 이자 및 복리의 계산의 토대에 놓여 있다"(MECW 33: 71). 이 문장은 자본가의 의식에서 잉여가치가 물신화된 채 단지 이윤의 형식으로만 나타난다는 점을 비판하는 과정에서 나왔다. 또한 마르크스는 "자본주의 생산이 지배하는 사회의 관점에서 자본은 스스로 작동하는 것(SELFACTOR)으로 – 모종의 숨겨진 성질의 결과로, 자기 증식하는 성질을 저절로 보유한 가치로 – 나타난다"고 언급하고 있다(MECW 33: 74. 대문자 강조는 마르크스의 것). 마르크스의 이러한 언급은 자기 증식하는 가치로서의 자본과 자동적으로 지식을 획득하는 기술로서의 인공지능 사이의 관계에 대해서 매우 함축적이다.

『그룬트리세』에서의 '일반 지성' 테마에 관한 마르크스의 논의를 어떻게 이해할 것이냐는 문제도 있다. 이에 관해서는 이미 잘 알려져 있다시피, 하욱의 총괄적 논의가 있었다(Haug, 2010). 그 논문에서 하욱은 『그룬트리세』의 이해에 관한 인지자본주의자 등의 주장을 기각하면서 결론 부분에서 이렇게 말하

고 있다. "'일반지성'이라는 카테고리는 앞을 내다보는, 자본주의를 초월하는 의미만은 아니다. '일반지성'은 이미 작동하는 것으로서 봉쇄되어(blocked) 있다. 그리고 바로 이런 봉쇄 속에서 일반지성은 부정적으로, 그것의 봉쇄가 위기를 야기하는 가운데, 현재화되어 있다"(Haug, 2010: 215). 여기서 봉쇄되어 있다는 말의 의미는 일반지성의 해방적 잠재력이 포스트자본주의사회에서처럼 완전히 구현되어 있는 게 아니라 예컨대 독점 자본의 이익을 위해서 일반지성이 여러 가지 수준과 방식으로 제약되어 있거나 구속되어 있거나 통제되어 있다는 뜻이다. 이런 언급은 인공지능의 과학 및 기술에 관해서도 타당하다.

마르크스는 『자본』 1권의 13장(독일어본 기준) "기계와 대공업"에서 거대한 자동 장치(Automaten) 내지는 자동적 기계 시스템(automatisches Maschinensystem)에 관해서 논의하고 있다. 물론, 마르크스가 여기서 언급하는 자동 장치 내지는 자동 기계 시스템은 19세기 중반의 그것이기는 하다. 하지만, 마르크스의 이런 언급은 오늘날 인공 지능 기술과 관련하여 충분히 참조할 만하다. 문제는, 한국어 문헌들에 제한해서 말하는 한, 『그룬트리세』에서의 '기계류' 및 '일반 지성'에 관한 논의가 『자본』에서의 자동 장치에 관한 논의들과 충분히 제대로 연결되어 정리되어 있지 못하다는 점이다.

마르크스 경제학 이론 체계 전반과 관련하여, 인공지능 기술이 집중적으로 문제가 되는 것은 자동화와 관련된 영역일 것이다. 그런데 그동안 마르크스주의 아카데미 안에서 자동화 문제는 충분히 논의되지 못했다. 이것은 글로벌한 수준의 역사적 정세와 관련이 있는 것으로 보인다. 즉, 자동화 문제가 현실적으로 부각되던 1980년대 말에서 1990년대 초에 이르는 시기에 역사적 사회주의 체제가 몰락해버린 것이다. 예외가 없는 것은 아니지만 대체로 보아서, 이러한 긴박한 정세 상황 자체에 정치적, 이론적으로 대응하다 보니 마르크스주의 아카데미 진영은 자동화 문제와 같은 비교적 작은 영역의 문제를 제대로 섬세하게 다루지 못한 채 체계적인 이론적 검토나 성찰 없이 극소전자기술 및 정보통신기술이 기술적으로 지배하는 단계로 끌려 들어와 버렸다.

한국에서 이와 연관된 논쟁은 소위 '정보재 가치 논쟁'이라는 타이틀 아래 2000년대 들어서 수행되었다. 이 논쟁의 성과는 인공지능 문제와 연결해서 확장, 심화, 발전시킬 필요가 있다고 여겨진다. 나는 경제학 전문가가 아니라서

이런 문제들을 제대로 잘 다룰 수가 없지만 상식적인 수준에서 몇 가지 궁금한 것들을 물어보는 방식과 수준으로 늘어놓는 것은 가능하다고 본다.

우선, 예를 들어, 상품 생산에서 불변자본의 양이 가변자본의 양에 비해서 상대적으로 엄청나게 큰 경우를 상정해볼 수 있다. 이때 불변자본의 거의 대부분을 인공지능 로봇이 차지한다고 하자. '살아 있는 노동'을 중시하는 마르크스주의의 기본 입장에서 보자면 이런 경우라고 하더라도 불변자본 가치량과 가변자본 가치량과 잉여가치량의 절대적 및 상대적 관계는 매우 중요하다. 하지만, 보통의 부르주아 계산법이라면 상대적으로 엄청나게 작은 수치는 무시된 채 처리된다. 심지어, 네오-리카도주의자이며 스라파주의자인 스티드먼 같은 경우에는, 이와 연관해서 극히 임의적인 산술 모형을 제시한 다음에, "잉여가치는 사라져도 이윤은 산출될 수 있다"는 황당한 주장을 했다(Steedman, 1985: 147). 이런 경우는 어떻게 비판하는 게 효과적인가.

마찬가지로 불변자본이 상대적으로 엄청나게 큰 경우이기는 하지만, 물질적 로봇들보다는, R&D 노동 내지는 알고리즘을 발견해내는 소프트웨어적 성격의 노동이 불변자본에서 더 큰 비중을 차지하는 경우도 상정해볼 수 있다. 이런 노동들은, 로봇 등과는 물리적, 사회적 성격이 상당히 다르다. 물론 지능적 에이전트는 '죽어버린 노동'으로만 볼 수는 없고 도리어 그 정의상 특수한 방식으로 일정하게 '살아' 있는 것으로 간주해야 하는 게 '약한 인공지능'의 취지에 맞는 듯하기는 하지만 말이다. 어쨌거나 이런 노동들은 기존의 사물-기계처럼 그저 완전히 죽어버린 노동은 아니다. 어느 정도는 잠재적으로 죽어 있고 또 어느 정도는 잠재적으로 살아 있다고 하는 게 적당할 듯하다. 생산성이 아주 높은 노동 내지는 특별잉여가치를 산출하는 노동을 규정할 때 마르크스는 '제곱된 노동(potenzierte Arbeit)'이라는 표현을 종종 사용했다. 독일어 명사 'Potenz'는 '잠재력'이란 뜻도 갖고 '제곱'이란 뜻도 갖고 있다. 또 참고로 물리학에서 '퍼텐셜 에너지(potential energy)'는 '위치 에너지'로 번역되기도 한다. 아무튼 이런 노동은 또 어떻게 처리해야 하는가.

반면에, 일상적 경험상, 프로그래밍 노동의 어떤 유형은, 특히 유지, 보수 등을 담당하는 경우에는, 고정 불변자본의 감가상각에 해당하는 부분으로 보인다. 혹은 이와는 다르게, 유지, 보수 등을 담당하는 프로그래밍 노동이 유동

불변자본에 속하는 것으로, 즉 예컨대 일종의 에너지라든가 원료와 같은 것으로 처리될 수도 있을 것이다.

그런가 하면, 현실에서 '프로그래밍 노가다'라고 불리고 있는, 프로그래밍 노동의 상대적으로 열등한 유형은 불변자본에 속하는 것이 아니라 명백히 가변자본에 속하는 듯이 보인다. 이 노동의 상당 부분은 오늘날 한국에서 비정규직의 형태로 존재하며, 직관적으로 보아서 엄청나게 '착취'당하고 있다고 할 수 있다. 즉 이 노동 대부분은 노동력 가치 이하의 임금만을 지불받고 있다.

일반적으로 말해서, 인공지능의 과학, 공학 및 기술에 있어서, 인공지능 프로그램 언어 자체를 만들어내는 노동, 새로운 통계 법칙이나 테크닉 등을 발견하는 노동, 이런 테크닉에 의거해서 컴퓨터 알고리즘 집합을 창출하는 노동, 그리고 창출된 알고리즘들에 따라서 프로그램을 직접 짜는 노동 등은 과학적, 공학적, 기술적으로 그 성격이나 수준이 다르고, 또 생산과정이나 가치구성에 있어서 차지하는 위치나 효과도 서로 다르다. 기존의 생산적 노동에 대한 규정과 관련해서 이것들은 어떻게 서로 분별적으로 이해할 수 있는가.

1980년대 중반에 구미에서 자동화 논쟁이 잠깐 이루어진 적이 있었다(Morris-Suzuki, 1984, 1986; Steedman, 1985). 자동화 생산, 즉 결국 인공지능 기술이 관련된 생산에 대한 논쟁에서, 스티드먼과 모리스-스즈키는 각각 소프트웨어적인 것과 하드웨어적인 것 사이의 구분, 혹은 물적 생산에 직접 투입된 것과 생산자를 향한 것과 소비자를 향한 것 사이의 구분 등을 제시했다. 그 두 사람의 범주적 구별은 한국의 정보재 논쟁과 비교해볼 때 매우 불충분하다. 더욱더 정교하게 다듬은 다음에 적절하게 배치해서 논의할 필요가 있는 것으로 보인다.

한편, 내가 이해하고 있는 한, 모리스-스즈키와 스티드먼은 그들의 논쟁에서 플랫폼 문제를 고려하지 못했다. 플랫폼 문제와 관해서 기존의 정보재 논쟁에서는 박지웅의 논의가 주목될 만한데(박지웅, 2011), 박지웅의 성과 및 기여가 인공지능 플랫폼에서도 타당한 것인가는 검토 거리가 된다. 구글 등 여러 글로벌 초국적 독점 자본들은 지금 인공지능 기술과 관련된 혁신 경쟁을 벌이면서 서로 다른 인공지능 플랫폼을 구축하고 있다. PC의 OS에 있어서 마이크로소프트사의 독점과 같은 형태는 인공지능 플랫폼 영역에서 나타나지는 않을 것으로 보인다. 그렇다면, 과점의 형태가 될 텐데 그 귀추가 주목된다.

한국의 정보재 논쟁의 성과 전체를 놓고 말한다면, 위에서 언급한 문제들에 대해, 현재로서 가장 포괄적인 설명은 안현효가 제시한 것으로 보인다(안현효 2012a, 2012b, 2013. 2016). 안현효는 그 이전까지의 정보재 논쟁에서 강남훈의 성과를 발전적, 비판적으로 계승하면서 '정보지대' 개념을 정립해냈다. 안현효는 마르크스 가치론의 핵심 명제들을 해치지 않으려고 애쓰면서도, 인지자본주의론의 문제의식을 비판적으로 수용하려 한다. 나아가 안현효는 이를 기본소득론의 이론적 토대로 삼으려 한다.

이와 같은 안현효의 지적 기획은 상대적으로 매우 뛰어난 성과와 기여를 거두고 있음에는 틀림없지만, 적지 않은 난점을 갖고 있다. 우선 정보지대론의 정책적 함의로 제시되는 기본소득 대안의 문제점은 차치하고라도,[3] 안현효를 포함해서, 기존의 정보재 논쟁 참가자들이 공통적으로 겪는 혼란이 있다. 논쟁자들은 정보재의 한계생산 비용이 제로에 가깝다는 것에 대개 당황스러워 한다. 이것을 어떻게 이론적으로 처리할 것인가에 대해서 고민한다. 하지만 이것은 기본적으로 그다지 큰 문제가 아니다. 마르크스는 다음과 같이 말했다.

> 협업과 분업에서 생겨나는 생산력은 자본에는 한 푼의 비용도 발생시키지 않는다. 그것은 사회적 노동이 만들어 내는 자연력이다. 생산과정에 사용되는 증기·물 등과 같은 자연력도 마찬가지로 아무런 비용이 들지 않는다. [...] 과학도 이 자연력과 마찬가지다. (자본 1권. MEW 23:407)

그 다음으로, 이렇듯 거의 공짜로 얻을 수 있는 자연력과 마찬가지인 지식과 정보를 자본가들이 지적 재산권이나 무역 협정 등을 통해서 독점하는 것에 관해서는 독점 이윤 개념으로써 얼마든지 처리할 수 있다.

강남훈과 안현효는 정보재 가치 논쟁에서, 잉여가치, 특별잉여가치, 독점 이윤, 차액 지대 등과 같은 여러 범주를 동원해서 설명한다. 이런 다원적 설명 방식 자체가 잘못된 것은 아니지만, 차액 지대 및 독점 지대와 같은 범주는 불필요한 것으로 보인다. 지식, 정보, 과학 등은 땅과 같은 자연물과는 다르다.

---

3) 기본소득론에 대한 비판은 우선 김성구(2016)를 참조할 수 있다.

지식, 정보, 과학은 생산될 수 있지만 땅은 애당초 생산될 수가 없다. 또 지식 등은 얼마든지 공유될 수 있으며 재생산될 때 제로에 가까운 비용이 들어간다. 그에 반해 땅은 물리적으로 제한되어 있다. 요컨대, 과학은 아무런 비용이 들지 않는 자연력과 마찬가지인 반면에, 땅은 독점될 수 있는 희소한 자연적 대상이고 또 독점적으로 소유되는 한에서만 자본주의적 지대를 낳는다.

그 다음으로, 사회적 가치가 결정되는 수준에 관해서 말하자면, 지대는 최열등지의 수준에 의해서 결정된다. 한편, 보통의 산업 생산물은 중위(평균적인)의 기술/생산 수준에 의해서 결정된다. 하지만 정보재는 부분적으로는 중위의 기술/생산 수준에 의해서, 때로는 종종 최상위의 기술/생산 수준에 의해서 그 사회적 가치가 결정된다. 이와 비교해 볼 때, 인공지능의 상품화와 관련된 사회적 가치는 거의 대부분이 최상위의 기술/생산수준에 의해서 결정된다고 할 수 있다. 이러한 수준을 염두에 두는 한, 정보재의 가치와 가격 문제를 지대에 비유해서 설명하는 것은 아무래도 크게 무리를 범하는 것이다.

최상위의 기술/생산수준에 의해서 사회적 가치가 결정된다는 것은 새로운 기술/생산방식에 의한 제품이나 서비스가 시장에 나오면, 그 이하의 기술/생산방식에 의한 제품이나 서비스는 곧 바로 거의 폐기되거나, 혹은 시장에서 즉시 거의 다 퇴출당하거나, 아니면 최소한, 엄청난 수준의 '도덕적 가치하락'을 겪는 한에서만 겨우 일부가 살아남는다는 뜻이다. 개별 자본들로서는 사회적 가치와 개별적 가치의 차이를 향유하거나 혹은 이 차이를 서로 넘겨주거나 넘겨받거나 할 기회가 아예 깡그리 없어지는 것이다.

특별잉여가치가 상대적 잉여가치로 변하기 전까지 자본가가 특별잉여가치와 맺는 관계를 마르크스는 비유적으로 '젊은 첫사랑의 시기'(MEW 23:429)라고 표현했다. 일상적으로는, 젊은 세대의 스마트폰 '기(기)변(경)'이 아주 대표적인 사례다. 마르크스의 비유를 연장한다면, 정보재 상당수에서는, 결혼(특별잉여가치의 상대적 잉여가치화)은 결코 없고 단지 첫사랑과 같은 관계들만이 계속 이어지고 있는 것이다.

인공지능 분야의 과학이나 기술은 거의가 다 더욱 더 그러하다. 인공지능 분야에서 새로운 기술은 낡은 기술을 철저히 그리고 재빨리 구축해버린다. 기계학습에 관해서 1200쪽이 넘는 번역본의 원본을 2012년에 내놓은 케빈 머피

는 자신의 책 마지막 28장에서 '딥 러닝(심화학습/심층학습)'을 아주 간략히 다루면서 이렇게 말하고 있다. "딥 러닝의 토픽은 매우 빠르게 발전하는 것을 인지해야 하며, 따라서 28장에서 다룬 내용은 곧 쓸모없는 정보가 될 수도 있다" (머피 2015:1157).

요컨대, 정보재 논쟁에 지대라는 범주를 끌어들이는 것은 설득력이 없는 것으로 보인다. 오캄의 면도날 원리는 마르크스 이론의 재구성에도 적용된다. 즉 지대 범주를 도입해서 설명하기보다 특별잉여가치 같은 범주에 더 비중을 두고 이론적 작업을 해나는 게 적절한 것 같다.

정보재 및 인공지능 분야에서는 특별잉여가치가 상대적 잉여가치로 쉽게 바뀌지 않는다. 그것은 한편으로는 지적 재산권 및 무역 협정 등에 의한 독점 때문이고, 다른 한편으로는 소위 파괴적인 기술 혁신들이 연속적으로 빨리빨리 서로 교체하면서 등장하기 때문이다. 어느 한 유형의 특별잉여가치가 상대적 잉여가치로 바뀌기 전에, 즉 그런 유형의 특별잉여가치를 만들어 낸 바로 그 기술이 동일한 부문 안에서 널리 보급되기도 전에, 다른 유형의 특별잉여가치를 낳는 새 기술이 등장해서 앞의 그 기술을 구축해버린다. 또 심지어 이 과정에서 종종 19세기 방식과 수준의 '부문'과 '부문' 사이의 경계 자체는 흔들리거나 흐려지거나 사라져버린다. 이론적으로는 아주 명백한 평균이윤율 명제에도 불구하고, 여러 가지 현실적 장벽으로 인해서 부문들 사이를 넘나들지 못하던 자본에게는 엄청난 비용 절감 효과가 생긴다.

그런데, 이렇게 파괴적이고도 연속인 기술 혁신이 가능한 것은, 새로 도입되거나 채택된 기술이 비즈니스 수익 모델에 연결되어 성공을 거두기만 한다면, 어마어마한 '창업자 이득'을 가공(의제)자본 영역(주식)에서 얻어낼 수 있기 때문이다.

아무튼, 기존의 정보재 가치 논쟁은, 상당한 혼란과 한계에도 불구하고, 당면 과제인 인공지능의 마르크스주의적 분석을 위해 유용한 이론적 자원을 제공하는 것으로 보인다. 그런데, 이러한 가치론의 문제보다 실천적으로 더 중요한 것은 인공지능 기술에 의한 혁신이 현단계 글로벌 자본주의 체제에 대해서 구조적, 거시적으로 어떠한 변화를 가져올 것인가 하는 점이다. 인공지능 기술의 발전이 불변자본의 가치를 급격하게 감소시킬 것이라는 것은 직관적으로

아주 명확하다. 그렇다면, 이 결과는 중장기적으로 축적 체제 및 이윤율에 어떠한 영향을 미칠 것인가, 또 이러한 영향들은 현재 진행 중인 글로벌 자본주의 체제의 위기와 맞물려 구체적으로 어떻게 변화되어 나타날 것인가 하는 것이 탐구되어야 한다.

그런데 여기서 명심해야 할 포인트는, 이러한 영향과 변화가 틀림없이 경제 권역별로, 네이션-스테이트별로, 혹은 네이션-스테이트 안에서도 부문 및 지역별로 매우 불균등하게 진행될 거라는 점이다. 따라서, 추상적인 탐구보다는 구체적 상황에 대한 구체적 분석이 더 요구된다고 하겠다. 또 아울러서, 정보통신 영역에서 여러 플랫폼들의 사용 가치를 높이고 있지만 지불되고 있지는 않은 노동, 즉 사용자의 일상적, 문화적 활동을 이론적으로 어떻게 다루고 평가할 것인가 하는 점도 중요하다. 인공지능 기술이 지식 발견 및 획득에 있어서 사용하고 있는 빅 데이터의 대부분은 바로 이런 활동에 의해서 산출되고 있기 때문이다. 게다가 이런 활동들은 포스트-자본주의 사회에서의 중요한 인간적 활동 형태라고 예견된다.

5. 맺음말

이제까지 인공지능의 몇 가지 문제들에 관해 개괄적인 스케치를 했다. 이 개괄적인 스케치는 너무 피상적이며, 또한 해당 영역 및 문제에 관한 핵심적 문헌들을 제대로 서베이하지도 못했다. 약간의 문제들을 들추어낸 다음에 그 문제들에 대해서 매우 주관적이고 천박한 평가적 견해를 덧붙였을 뿐이다. 나는 결론 부분에서 '인공지능 비판'이라는 과제를 마르크스주의 내지 진보-좌파 아카데미 진영에 제기하고 싶다. 나는 이전에 다음과 같이 제안했다.

> 나는, 마르크스가 수행한 '정치경제학 비판'에 상응하는 '인지과학/신경과학 비판'이 필요하다는 말로 이 얄팍한 글의 결론을 대신하고자 한다. 우리는 인지과학 및 신경과학을 깊이 있게 알아야 하며, 마르크스가 부르주아 정치경제학을 공부하고, 이해하고, 그런 바탕 위에서 정치경제학 비판을 수행한 만큼 인지과학/

> 신경과학을 제대로 공부하고, 이해하고, 비판해나가는 과정이 필요하다. 한때 인지적 매핑이란 말이 유행한 적이 있는데, 지금은 인지과학 및 신경과학에 대한 철학적 매핑, 정치경제학적 매핑이 시급하다고 할 수 있다(이재현, 2016b: 11).

이제는 위 제안의 '인지과학/신경과학 비판'에 '인공지능'을 덧붙이고 싶다. 즉, '인지과학/신경과학/인공지능 비판'이 필요하다는 것이다.

엄격한 눈으로 본다면 이 글은 본격적인 학술 논문의 수준에 미치지 못할 것이다. 말 그대로 개괄적인 스케치에 불과하다. 레토릭한 에세이에 해당하는 이 글은 구멍들이 너무 크고 많다. 우선, 통계적 학습 이론에 대한 비판이 치밀하지 못하다. 그것은 확률 및 통계에 관한 내 이해가 전적으로 고등학교 수학 수준에 머물러 있기 때문에 그러하다. 기계 학습 전반이 베이지언 확률의 문제를 갖고 있는 것은 결코 아니다. 오늘날 기계 학습 분야는 소위 '나이브한 베이즈 추론' 방식에만 머물러 있는 것이 결코 아니다. 오늘날 기계 학습은, 심층적인 다층 인공 신경망, 유전 알고리즘, 진화 전략, 유전 프로그래밍, 진화 신경망, 적응형 뉴로-퍼지 추론 시스템, 퍼지 진화 시스템 등과 같은 하이브리드 지능 시스템으로 발전해 있다. 이런 하이브리드 지능 시스템은 좀 더 평범한 말로 하면, 소프트 컴퓨팅에 해당한다. 기계 학습 및 통계적 학습 방법에 관해서는 전문가가 나서서 제대로 정교하고도 단단하게 다루어 주기를 바란다.

아무튼, 이런 것들의 치명적인 문제는, 실용적으로야 그럴 듯한 성과를 낳고 있는지는 모르지만, 고전적 전문가 시스템과는 달리, 구체적으로 어떻게 해서 그러한 지식을 발견하거나 획득했는가를 우리에게 설명해 줄 수 없다는 점이다. 이것은 우리가 통상 체험하고 있는 상황이나 수준에서의 학습과 지식과는 전적으로 다르다. 즉, 어떻게 해서 알았는가, 왜 그런가, 혹은 어째서 그러한가를 설명하지 못하는 것이다. 기계 학습의 경우, 지식 발견 및 획득이라는 일은 다양한 신경망에서의 가중치, 혹은 통계적 알고리즘으로 구성된 모형의 파라미터 값을 구하는 것으로 귀착될 뿐이다. 다시 말해서, 과학 철학에서의 고전적인 용어로 표현한다면, '발견의 맥락'에서는 상당한 기술적 성취를 보여왔지만 '정당화의 맥락'의 문제들은 거의 대부분이 묵살된 채 블랙박스 속에 남아 있는 것이다.

인공지능 기술을 옹호하는 사람들은 종종 “새가 날면 되는 것이지, 어떻게 나는가를 새 스스로가 이해하거나 다른 생명체에게 설명할 수 있어야 할 필요는 없다”고 강변한다. 그러나 인간의 지적 역사는 단지 무엇인가를 기능적인 면에서 성공적으로 작동시킬 수 있는 공학이나 기술만을 추구해 온 것이 아니다. 인간의 과학이란 본디 ‘왜’와 ‘어떻게’에 관해서 이해하고 설명하는 것을 주된 사명으로 삼아왔다. 그렇다면, 더욱더 이런 정도의 개괄적인 스케치보다는 좀 더 견실하고 치밀한 인공지능 비판이 요구된다. 그런데 이러한 비판을 수행하기 위해서는 인공지능의 과학-공학-기술에 대해 좀 더 제대로 잘 알아가는 일을 꾸준히 해야 한다. 익히 널리 알려진 격언대로, 악마는 디테일에 있기 때문이다.

(2016년 7월 4일 투고, 7월 20일 심사, 7월 26일 게재 확정)

❒ 참고문헌

고민조·박주용. 2014. 「베이지안 망을 이용한 법적 논증」. ≪서울대학교 법학≫, 제55권 제1호.

구본권. 2015. 『로봇 시대, 인간의 일』. 어크로스.

김성구. 2016. 「기본소득, 참 받고 싶은데요」. ≪워커스≫, 14호(2016.6.15). (사)참세상.

러셀, 스튜어드·노빅, 피터. 2016a. 『인공지능: 현대적 접근방식』 1. 류광 옮김. 제이펍.

_____. 2016b. 『인공지능: 현대적 접근방식』 2. 류광 옮김. 제이펍.

머피, 캐빈. 2015. 『머신 러닝』. 노영찬 옮김. 에이콘.

박지웅. 2011. 「정보재 가치와 플랫폼: 양면시장을 고려한 정보재 가치논쟁의 검토」. ≪경제학연구≫, 59집 1호.

송하석. 1998. 「주관주의 확률이론」. ≪논리연구≫, 2권 0호.

심광현. 2014. 「제3세대 인지과학과 '신체화된 마음의 정치학」. 『맑스와 마음의 정치학: 생산양식과 주체양식의 변증법』. 문화과학사.

안현효. 2012a. 「인지자본주의와 기본소득: 기본소득의 유지 가능성」. ≪마르크스주의 연구≫, 제9권 제1호. 경상대학교 사회과학연구원.

_____. 2012b. 「기본소득의 가치론적 기초: 정보재 가치 논쟁 재론」. ≪마르크스주의 연구≫, 제9권 제4호. 경상대학교 사회과학연구원.

_____. 2013. 「정보지대와 노동가치론: 전희상의 논평에 대한 답변」. ≪마르크스주의 연구≫, 제10권 제4호. 경상대학교 사회과학연구원.

_____. 2016. 「인지에 적용된 공유자원 패러다임: 기본소득의 가치론적 기초」. ≪마르크스주의 연구≫, 제13권 제2호. 경상대학교 사회과학연구원.

이재현. 2016a. 「제4차 산업혁명? 할렐루야!: 미래학이 감추고 있는 것들」. ≪워커스≫, 3호.

_____. 2016b. 「인지과학과 이데올로기」('리딩 맑스' 2106년 4월 발표문). 제8회 맑스코뮤날레 집행위원회.

재미한인과학기술자협회. 2015. 「뇌-컴퓨터 인터페이스 기술에 대한 국내·외 연구개발 동향 조사」. Korean-American Scientists and Engineers Association.

파스콸레, 프랭크. 2016. 『블랙박스 사회』. 이시은 옮김. 안티고네.

Abney, K. et al(eds.). 2012. *Robot Ethics: the Ethical and Social Implications of Robotics*. The MIT Press.

Anderson, M. & S. L. Anderson(eds.). 2011. *Machine Ethics*. Cambridge University Press.

Bostrom, N. & E. Yudkowsky. "The Ethics of Artificial Intelligence." in Keith & Ramsey(2014).

Corfield, D. & J. Williamson(eds.). 2001. *Foundations of Bayesianism*. Springer.

Eden, A. H. et al(eds.). 2012. *Singularity Hypotheses: A Scientific and Philosophical Assessment*. Springer.

Federal Trade Commission. 2016. "Big Data: A Tool for Inclusion Or Exclusion."

Floridi, L. 2014. "Big Data and Information Quality." in Floridi & Illari(2014).

Floridi, L. & P. Illari(eds.). 2014. *The Philosophy of Information Quality*. Springer.

Gillies, D. 2004. "Handling Uncertainty in Artificial Intelligence, and the Bayesian Controversy." in Stadle(2004).

Gordon, R. J. 2016. *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War*. Princeton University Press.

Gunkel, D. J. 2012. *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. The MIT Press.

Haug, W. F. 2010. "General Intellect." *Historical Materialism,* 18. Brill.

Keith, F. & W. M. Ramsey(eds.). 2014. *The Cambridge Handbook of Artificial Intelligence*. Cambridge University Press.

MECW 33. 2010. *Karl Marx Frederick Engels Collected Works*. Volume 33: Marx 1861-63. Lawrence & Wishart.

MEW 23. 1962. *Karl Marx Friedrich Engels Werke*. Band 23. Dietz Verlag.

Mitchell, T. M. 1997. *Machine Learning*. McGraw Hill.

Modis, T. 2012. "Why the Singularity Cannot Happen." in Eden(2012).

Morris-Suzuki, T. 1984. "Robots and Capitalism." *NLR,* I-147.

_____. 1986. "Capitalism in the Computer Age." *NLR,* I-160.

Nisbet, R. et al. 2009. *Handbook of Statistical Analysis and Data Mining Applications*. Academic Press.

Roberts, P. C. 2013. "More Misleading Official Employment Statistics." http://www.paulcraigroberts.org/2013/12/10/misleading-official-employment-statistics-paul-craig-roberts/. 검색일: 2016.6.30.

Sachs, J. 2015. "Robots: Curse or Blessing"(NBER Working Paper No. 21091). The National Bureau of Economic Research.

Stadle, F. et al. 2004. *Induction and Deduction in the Science*s. Springer.

Searle, J. R. 1980. "Minds Brains and Programs." in *Behavioral and Brain Sciences,* 3(3).

Steedman, I. 1985. “Robots and Capitalism: Comment.” *NLR*, I-151.
The White House. 2014. “Big Data: Seizing Opportunities, Preserving Values.”
_____. 2016. “Big Data: A Report on Algorithmic Systems.”
Tzafestas, S. G. 2016. *Roboethics: A Navigating Overview*. Springer.
Williams, P. 2001. “Probabilistic Learning Models.” in Corfield & Williamson(2001).
Witten, I. H. et al. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.

❒ 영문초록

A Critical Sketch on Artificial Intelligence

Jaehyun Lee

This paper considers some problems of artificial intelligence (AI) technology. The assumptions are weak AI, socially-embodied cognition, and methodological holism. Machine learning's weakness comes from the subjective character of Bayesian probability. If an artificial agent has the same functionality and conscious experience as a human, then it has the same human moral status as well. If AI is developed, then in Marxist economics it will be more important to explain the theoretical character of various forms of its digital labor.

Keywords: critical approaches to AI technology, weak artificial intelligence, Bayesian probability, principle of ontogeny non-discrimination, digital labor.