

포스트휴머니즘의 저편*

-인공지능인문학 개념 정립을 위한 시론-

김형주**, 이찬규***

주제분류 포스트휴머니즘, 인공지능인문학, 근대철학, 역사철학

주요어 인공지능인문학, 인공지능, 인문학, 포스트휴머니즘, 세계해석학, 반성학

요약문

이 연구는 기초인문학의 토대 위에서 포스트휴먼 시대를 이끌어 갈 인공지능인문학의 개념을 정립하는 것을 목표로 한다. 이를 위해 첫째, ‘인공지능인문학’을 구성하고 있는 세 가지 요소로 ‘인공지능’, ‘인문학의 내용’, ‘인문학의 방법’을 추출한다. 둘째, 소위 4차 산업혁명시대라 불리는 인공지능 시대의 휴머니즘으로부터 휴머니즘의 본령인 서구 근대 르네상스의 인문학 개념과 그것의 전사(前史)를 살펴본다. 이를 통해 인공지능 시대를 읽는 포스트휴머니스트들의 시각을 쫓는 한편, 그들이 부지불식간에 전제하고 있는 전통인문학의 내용을 드러낸다. 마지막으로 인문학의 방법론으로서 인문학의 본질적인 사명 중의 하나인 시대와 자기에 대한 반성적 역할을 앞에서 추출한 내용에 적용하여 전통인문학 분과외에 인공지능시대의 변화를 수용하여 재구조화된 인공지능인문학 5개 연구영역인 인공지능 기술비평학, 인공지능 관계·소통학, 인공지능 사회·문화학, 인공지능 윤리·규범학, 인공지능 인문데이터해석학을 도출한다.

* 이 연구는 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임을 밝힙니다.(NRF-2017S1A6A3A01078538)

** 제1저자, 중앙대학교

*** 교신저자, 중앙대학교

1. 들어가는 말: 자기모순의 변증법

헤브라이즘과 헬레니즘의 합수머리인 기원후 2세기, 교회사 최초의 교부 테르툴리아누스(Tertullianus)는 “예루살렘과 아테네와 무슨 관계가 있는가?”¹⁾라는 질문을 던지면서 기독교와 철학, 나아가 신앙과 이성의 이질성을 주장했다. 그에 따르면 기독교철학은 등근 사각형과 같은 형용모순이다. 이러한 그의 주장이 무색하게도, 실제로는 헤브라이즘과 헬레니즘은 서양정신을 이끌어 온 두 축이며, 예루살렘과 아테네의 변주곡은 서양지성사의 국가(國歌)라 해도 과언이 아니다. 이 둘은 끊임없이 서로를 견제하고 때로는 화해하는 대대(對待)관계 속에서 각자의 지평을 확장하면서 궁극적으로는 서양정신이라는 일자를 형성해왔고 지금도 그러하다. 감성과 이성의 대립과 합치 논쟁은 몸과 마음의 동근원성 논제로, 근대의 경험론, 합리론 논쟁은 21세기 현재 신경생리학을 무기로 장착한 환원주의 철학으로 이어지면서 여전히 생활세계의 이면에서 세력다툼을 벌이고 있는 형국이다. 일견 모순은 반복을 일삼으며 끊임없이 차이를 생성한다.

“인공지능과 인문학이 무슨 관계가 있는가?” 지금의 학제에 따르면 인공지능은 컴퓨터 공학의 한 연구영역이다. 전 세계 대부분 대학의 공과대학에서는 인공지능 교과를 필수과목으로 지정하여 운영하고 있다. 이를테면 러셀과 노르빅의 *Artificial Intelligence: A modern Approach* 는 우리나라는 물론 세계 각국의 인공지능 교과에서 가장 많이 사용하는 교재다. 현 대학의 학제 구분은 물론, 인문학의 태생적 목적과 발전사에 대한 우리의 상식에 따르면, 학문으로서의 인문학과 인공지능은 매우 다르다. 위

1) Tertullianus, *De Praescriptione Haereticorum* 7.
(http://www.tertullian.org/works/de_praescriptione_haereticorum.htm)

질문이 내포하고 있는 답은 인공지능과 인문학은 마치 아테네와 예루살렘처럼 서로 어울릴 수 없으며, 그렇기 때문에 하나의 학문으로서의 ‘인공지능인문학’은 자기모순을 내재하고 있다는 것일 것이다. 그러나 ‘인공지능’이라는 말이 유행처럼 회자되고 있는 지금, 학문이란 무엇인가에 대한 본질적인 고민을 하지 않는다면 위 질문에 대한 긍정적인 답을 내놓는 것은 그리 어렵지 않다. 애플의 고(故) 스티브 잡스가 말한 “기술과 인문학의 결합”은 공학과 인문학의 융합에 대한 무한긍정의 아이콘이다. 이러한 입장을 취하면 ‘인공지능인문학’은 인공지능에 인문적 사고력을 탑재하여 인공지능 기술을 인간 친화적으로 만드는 활동 등을 의미할 수 있다. 그러나 이는 우리가 던진 질문에 대한 답이 되기엔 여전히 부족하다. 학문으로서의 인공지능, 학문으로서의 인문학이 어떻게 관계할 수 있는가? “인공지능인문학”²⁾은 가능한가?

이 연구는 기초인문학의 토대 위에서 포스트휴먼 시대를 이끌어 갈 인공지능인문학의 개념을 정립하는 것을 목표로 한다. 덧붙여 말하자면 ‘인공지능인문학’이라는 일종의 모순 속에 잠재하는 긴장의 에너지가 자신 밖으로 표출될 때 발현되는 양상을 기술하는 것이 이 연구의 종착점이다.

이를 위해 첫째, ‘인공지능인문학’을 구성하고 있는 세 가지 요소로 ‘인공지능’, ‘인문학의 내용’, ‘인문학의 방법’을 추출한다. 둘째, 소위 4차 산업혁명시대라 불리는 인공지능 시대의 휴머니즘으로부터 휴머니즘의 본령인 서구 근대 르네상스의 인문학 개념과 그것의 전사(前史)를 살펴보고자 한다. 이를 통해 인공지능 시대를 읽는 포스트휴머니스트들의 시각을 쫓는 한편, 그들이 부지불식간에 전제하고 있는 전통인문학의 내용을 드러낸다. 마지막으로 인문학의 방법론으로서 인문학의 본질적인 사명 중의 하나인 시대와 자기에 대한 반성적 역할을 앞에서 추출한 내용에 적용하

2) 중앙대학교 인문콘텐츠연구소는 2017년부터 인문기초학문분야 HK+사업으로 “포스트휴먼 시대, 인문학 가치 고양을 위한 인공지능인문학 구축”이라는 아젠다로 연구를 수행하고 있다.

여 전통인문학분과위에 인공지능시대의 변화를 수용하여 재구조화된 인공지능인문학 5개 연구영역을 도출하겠다.

2. 인공지능인문학의 배경

2.1 ‘인공지능’, 그리고 ‘세계해석학’과 ‘반성학’

어떤 것을 새로 만든다는 것은 있는 것들을 조합하는 것을 뜻하기도 하지만, 자신이 행한 대담한 시도를 설명하는 과정을 의미하기도 한다. 인공지능인문학의 의미에 대한 논의는 뒤로 미루고, 우선 이 개념에 대한 평면적인 분석을 시도하면, ‘인공지능’과 ‘인문학’이 구성요소로 도출된다. 본 연구의 목표가 기초인문학으로서의 인공지능인문학의 개념정립이므로 ‘인공지능’이라는 학문의 의미를 규정하기 전에 인문학의 정초작업과 체계화작업이 활발하였던 서양근대의 르네상스 전후에 주목하여 이 신설 기초인문학이 전승해야 할 학문의 본질적 내용과 방법을 찾는데 주목하고자 한다.

근대정신의 문을 연 데카르트(R. Descartes)의 코기토 테제는 지금까지 서양의 정신사에 “매력”적인 “마술”³⁾을 부리고 있다. 그가 촉발한 주체성 논쟁은 자연과 인간의 이분법을 낳았고, 이는 고대철학의 획기적인 유산인 형상과 진료의 구분을 자양분으로 근대 학문구축의 방법론적 토대인 내용과 형식의 이분법으로 발전되었다. 이 전통에 충실하여 소위 계몽주의 인문정신의 완성을 기한 철학자로 우리는 칸트(I. Kant)를 꼽는다.⁴⁾ 이를테면 그는 『도덕형이상학 기초논기』(Grundlegung zur Metaphysik der Sitten)에서 지금 우리가 그러하듯 ‘도덕형이상학’이라는 새로운 학문

3) H. Brands, *Cogito ergo sum*, Karl Alber, Freiburg, 1982, S. 10.

4) M. Wundt, *die deutsche Schulphilosophie im Zeitalter der Aufklärung*, Olms, Tübingen, 1945, S. 1 참조.

을 구축하기 위하여 자연학, 윤리학(도덕철학), 논리학을 고대 그리스 철학의 세 가지 연구영역으로 제시⁵⁾하고 이에 내용과 형식이라는 도식을 대입하여 논리학을 순수한 형식적인 학문, 자연학과 윤리학을 내용을 갖는 학문으로 따라서 경험적인 것을 가질 수 있는 학문으로 구분한다. 다시 윤리학에 형식과 내용의 도식을 적용하여 내용을 갖는 윤리학으로 인간학을 포함하는 실천철학을, 순전히 형식적인 윤리학으로 도덕형이상학을 추출하기에 이른다.⁶⁾ 우리는 이 태도를 차용하여 인문학의 본질적 내용으로서 그것을 구성하여 왔던 연구영역학문들의 향상적 유용성을 논증할 필요성을 제기하며, ‘학문이란 무엇인가’에 대한 문제에, 다시 말해 학문의 정초작업에 골몰한 자칭 칸트의 충실한 계승자 피히테(Fichte)의 학문정초작업의 원칙인 반성적 태도에 주목한다. 그의 초기저작인 『학문의 개념에 대하여』(Ueber den Begriff der Wissenschaftslehre), 『전체 학문의 기초』(Grundlage der gesammten Wissenschaftslehre)⁷⁾ 등은 그의 주관적 관념론을 집약적으로 나타내는 표어 ‘나는 나다’로 대표되는 자기반성의 원리에 대한 정당화 작업이라 할 수 있다. 우리는 이 반성의 원리를 인공지능인문학 개념정립의 방법론으로 삼는다.

요컨대 우리는 인문학의 학문성 물음이 본격화되고 심화된 서양 근대 인문주의의 시작과 끝을 관통하는 학문적 태도에 대한 분석을 통해 인문학의 내용과 인문학의 방법을 분리하고 이를 토대로 인문학의 내용을 세계해석학으로, 그 방법을 반성학으로 규정한다.

5) 이 세 학문분야가 그리스 철학의 확고한 분과였다는 사실은 하이데거(M. Heidegger)의 “휴머니즘이란 무엇인가(Über den Humanismus)”에서도 확인할 수 있다; M. Heidegger, *Über den Humanismus*, Klostermann, 2000, S. 8.

6) I. Kant, *Grundlegung zur Metaphysik der Sitten(AA4)* in: *Kants gesammelte Schriften* (Sog. *Akademie-Ausgabe*), Berlin: Walter de Gruyter, 1900 ff. S. 387 - 390 참조.

7) 이하 인용 시 각각 *UBW*, *GLW*로 약칭.

2.2 포스트휴먼시대 속의 인문학: 세계해석학으로서의 인문학

두 번의 빙하기를 맞은 ‘인공지능’은 4차 산업이라는 더 큰 파도를 타고 다시 찾아왔다. 인공지능 기술이 문명인의 삶 곳곳에 스며들었다는 뉴스가 식상해진지 오래다. 자연스럽게 학계의 관심도 이러한 현상에 쏠리고 있다. 공학은 인공지능 기술의 질적 발전을 끊임없이 추구하고 있고 사회과학에서는 그 연구 성과를 재빨리 흡수하여 자신의 연구영역에 적용하고 있다. 황혼녘에 날아오르는 미네르바의 부엉이로 상징되는, 다른 학문들에 비해 새로움에 대한 즉각적인 반응에 신중한 인문학도 인공지능 현상에 많은 관심을 기울이고 있다. 끊임없는 자기반성을 필연적 속성으로 갖는 인문학은 최근 스스로에게 ‘포스트휴머니즘’이라는 이름을 붙이기도 한다. 인간에 대한 기존의 통념에 대한 해체와 수정을 주장하는 포스트휴머니즘을 자기전복을 일삼아 발전하는 인문학 자체의 역사의 일부로 간주해야 할 수도 있지만, 인문학이 시대비평의 첨병이라는 사실을 생각해 보면, 이 이론의 확산에는 인공지능 기술의 발달과 이로 인해 고찰되는 새로운 인간상에 대한 생각들이 그 저변에 놓여 있다고 할 수 있다. 포스트휴머니즘은 이러한 이유에서, 다시 말해 포스트휴먼을 둘러싼 현상과 담론들을 인문학적 성찰의 대상으로 삼는다는 의미에서 “포스트휴먼 인문학(Posthuman Humanities)”⁸⁾이라 불릴 수 있다.

포스트휴머니즘의 착상의 근원을 찾자면, 논의는 서양의 근대로 거슬러 올라간다. 괴테의 『파우스트』에는 “훌륭한 사유 능력을 지닌 두뇌까지도/장래에는 사유하는 인간이 만들 것”이라는 말이 등장한다.⁹⁾ “올리버 크뤼거(O. Krüger)는 포스트휴먼이라는 용어가 1656년 토마스 블런트의 글로소그래피아에 이미 등장하였다”¹⁰⁾고 주장한다. 한편 이 개념의 의미

8) R. Braidotti, *The Posthuman*, Polity, 2013, 143쪽.

9) 임석원, 「비판적 포스트휴머니즘의 기획: 배타적인 인간중심주의 극복」, 『인간과 포스트휴머니즘』, 이화여자대학교출판부, 2013, 61쪽 참조.

10) 슈테판 헤어브레히터, 김연순, 김웅준 옮김, 『포스트휴머니즘』, 성균관대학교 출

범주를 확장적으로 해석하면 라이프니츠가 이미 수 세기 앞서 언급한 “생각하고 감정을 느끼고 지각할 수 있는 기계”을 포스트휴먼을 구성하는 하나의 조건으로 간주할 수도 있다. 한편 포스트휴머니즘에 대한 본격적인 논의는 1990년대 후반부터 시작되었다. 미국의 문화이론가인 이합 핫산(I. Hassan)을 시작으로 케서린 헤일즈(K. Hayles), 캐리 울프(Cary Wolfe), 닐 배드밍턴(N. Badmington)등이 포스트휴먼 담론의 확산에 많은 기여를 하였다. 닉 보스트롬(N. Bostrom)은 기술의 진보가 포스트휴먼의 도래를 가능하게 할 수 있고 그래야한다는 입장을 피력하는 한편 『역사의 종말』의 저자 프랜시스 후쿠야마(F. Fukuyama)는 이러한 생각이야말로 가장 위험한 생각¹¹⁾이라면서 인류의 기계의존적 진화와 발전에 대한 보스트롬의 제안을 일축한다.

모든 사조가 그렇듯, 포스트휴머니즘 담론도 분화되어 발전하고 있기 때문에 이에 대한 일면적인 정의를 내리는 것은 다소 무리가 있지만, 이 이론이 르네상스 이후 구축된 전통적인 인간개념이 어떠한 형태로는 수정되어야 한다는 대전제로부터 출발한다는 사실 만큼은 의심의 여지가 없다. 기존의 논의를 토대로 포스트휴머니즘을 간략하게 정의하면 다음과 같다.

포스트휴머니즘은 모든 종류의 이분법은 해체되어야 할 폭력적 허상이라는 전제를 토대로 좁게는 육체와 정신의 이원론적 대립을 해체하고 “물질적인 것과 비물질적인 것, 동물과 인간 나아가 유기체와 기계”¹²⁾의 경계를 허물어 “인간적인 것과 인간적이지 않은 것 사이의 이중대립”¹³⁾을 해체한다. 이러한 입장에서 이는 바이오테크놀로

관부, 2012, 54쪽.

11) F. Fukuyama, “Transhumanism”, *Foreign Policy* 144쪽 참조.

12) D. Haraway, in Wolfe, 1995, 36쪽. 『포스트휴머니즘』 190쪽 재인용.

13) N. Badmington, “Pool Alighty!; Or, Humanism, Posthumanism, and Strange Case of Invasion of the Body Snatchers”, *Textual Practice* 15.1, 5-22쪽.

지, 인공지능 등의 과학기술의 성과를 이용하여 생물학적 육체의 한계를 극복하는 포스트휴먼의 생성과 발전¹⁴⁾을 추구한다.

‘포스트’는 일차적으로는 시간, 공간, 순서상 어떤 것의 다음을 의미하지만 이 일차적인 의미는 앞선 것에 대한 반대의 의미를 생성하기도 한다. 그러나 이에 대한 반대는 앞선 것에 어떠한 형태로는 의존하고 있을 수밖에 없기 때문에 앞 선 것에 대한 초월과 극복, 지양(Aufhebung)에 대한 의미도 담고 있다. 이러한 의미에서 ‘포스트’는 시간적인 후속이 아닌 질적인 고양, 즉 넘어선다는 의미를 담고 있는 ‘트랜스(trans)’를 함축한다. 마찬가지로 포스트휴머니즘이 트랜스휴머니즘을 함축한다고 할 수 있는데, 왜냐하면 트랜스휴머니즘이 궁극적으로 추구하는 바는 생물학적 인간이 기계, 기술 등 외부의 도움을 통해 육체와 정신이 증강된 포스트휴먼이기 때문이다. 이러한 이해를 바탕으로 우리는 트랜스휴머니즘을 기계 결정론에 입각한 강력한 포스트휴먼 운동으로 간주한다.¹⁵⁾

포스트휴머니즘이 ‘작용과 반작용’이라는 반복의 철학사에서 등장한 포스트모더니즘과 접두사를 공유한다는 점에서 예측할 수 있듯이, 이는 포스트모더니즘의 사상적 토양에서 발전하였다. 실제로 독일적 전통에서 포스트휴머니즘 담론을 이끌고 있는 헤어브레이터(S. Herbrechter), 조르그너(S. Sorgner)등 많은 학자들은 포스트휴머니즘의 철학적 단초는 포스트모더니즘의 선구자 니체의 철학에 있다고 말한다.¹⁶⁾

14) R. Kurzweil, “The Coming Merging of Mind and Machine”, *Scientific American Report Special Edition*, 2008.

15) 트랜스휴머니즘을 포스트휴머니즘과 분리하여 생각하는 예로는 토머스 필백, 「포스트휴먼 자아: 혼합체로의 도전」, 『인간과 포스트휴머니즘』 참조.

16) 슈테판 헤어브레이터, 50-53쪽, 임석원, 65쪽, S. Sorgner, “Nietzsche, the Overhuman, and Transhumanism”, *Journal of Evolution & Technology*, 20(1), 2009, 29-42쪽 참조. 실제로 현대 기술윤리담론에 관한 학술지 *Journal of Evolution & Technology*는 20호에서 니체가 트랜스휴머니즘에 미친 영향을 주제로 특집호를 발간하였다. 이에 반해 보스트롬은 니체의 위버멘쉬 개념은 트랜스

미래의 인간은 지금까지의 이성으로부터도 우리를 구원해주며, 그 이상에서 성장할 수밖에 없는 것 즉 격렬한 구도에서, 허무를 향하는 의지에서, 허무주의에서 우리를 구원해주게 된다. 이러한 정오와 위대한 결단의 종소리는 의지를 다시금 자유롭게 만들며, 대지에는 목표, 인간에게는 희망을 돌려준다. 안티크리스트이자 반(反)허무주의자, 신과 허무를 초극한 이 자, - 그는 언젠가 올 수밖에 없다....¹⁷⁾

위버멘쉬(Übermensch)를 기다리는 니체의 철학에는 여전히 근대정신에 잔영이 남아있다. “고귀한 이상 그 자체는 육화”되었고 “비인간(Unmensch)과 위버멘쉬의 종합은 나폴레옹”이다.¹⁸⁾ 우리(인간)에 대한 통렬한 비판은 여전히 더 낡은 우리를 은연중에 염두에 두고 있다. 자기를 부정하여 다른 자기를 만들려는 열망은 여전히 자기에 대한 의식에 의존하고 있다. 자기부정을 통한 자기완성은 황금시대를 기다리는 진보적 계몽사관의 변형적 답습이다. 이러한 부정-긍정의 양가성은 포스트휴머니즘이 갖고 있는 자기 파괴적 색채가 드러나면 드러날수록 점점 더 확연히 드러난다. 인간개념의 통념에 대한 적극적인 부정을 통한 인류행복의 최대화의 청사진을 제시하는 트랜스휴머니즘의 주장을 살펴보자. 세계 트랜스휴머니스트 협회(The World Transhumanist Association)을 설립한 보스트롬(N. Bostrom)은 이를 “인간의 노화를 제거하고 인간의 지성적, 육체적, 심리적 능력을 향상시키는 기술을 개발하고 확대함으로써 인간 조건을 근본적으로 향상시키는 것의 가능성과 그 바람직함을 긍정하는 지적, 문화적 운동”¹⁹⁾으로 정의한다. 트랜스휴머니즘은 기계와 생물학적 신체의 경계의

휴머니즘과 단지 “현상적 차원의 유사성”(Surface-Level similarities) 만을 지닐 뿐이라고 하면서 그 영향력을 축소해석한다. N. Bostrom, “A history of transhumanist thought”, *Journal of Evolution & Technology* 14(1), 2005. 4쪽. (<https://nickbostrom.com/papers/history.pdf>)

17) 니체, 김정현 옮김, 『도덕의 계보학』, 2006, 책세상, 448쪽.

18) 니체, 『도덕의 계보학』, 389쪽.

19) Transhumanist FAQ(<https://nickbostrom.com>)

불필요성을 근거로 신체보형물, 보조기를 통한 인간조건의 향상을 주장한다. 이 연장선상에서 이들은 마음 업로딩(mind-uploading)을 통한 탈체화(disembodiment) 까지도 추구한다. 그러나 거꾸로 생각해 보면 이러한 발상은 정신과 육체의 이분법이라는 근대적 발상을 그대로 답습하고 있다고 할 수 있다. 실제로 보스트롬은 “포스트휴머니즘은 완벽한 인간의 완전성, 이성성으로부터 직접 도출된 것이며 르네상스 휴머니즘과 계몽주의를 유산으로 물려받았다고 고백하면서” 근대정신의 완결 개념이라고도 할 수 있는 자율적 개인(autonomous individual)의 자율성(autonomy)이 그것의 핵심 가치(high value)인 반면, 인공지능 기술과 인터페이스 기술의 결합은 이를 위한 도구라고 말한다.²⁰⁾ 이처럼 반휴머니즘을 기치로 전개된 포스트휴머니즘은 자신의 색깔을 드러내면 드러낼수록 휴머니즘으로 회귀한다. 그들이 말하는 “미래의 인간”은 “우리”는 아닐지라도 여전히 인간이며, 포스트휴머니즘은 과거의 휴머니즘은 아닐지라도 여전히 휴머니즘이다. 우리는 이러한 주장의 근거를 인문학 형성의 역사 속에서 찾고자 한다.

지금 통용되고 있는 ‘인문학’이라는 말의 서양적 어원으로 키케로(Cicero)에 의해 처음 사용되었다고 알려진 ‘Studia Humanitas’를 꼽는 것이 일반적이다.²¹⁾ 후마니타스가 가장 처음 언급되었을 때, 이는 “호의적인 인간적인 행동양식(습관)”²²⁾을 뜻하였다. 여기서 우리는 인간적인 것(Humanitas)을 ‘자신의 말(Logos)을 사용할 수 있는 자’로서의 자유인의 덕(Arete)으로 본 것에는 고대 희랍적 전통이 녹아있다고 추측할 수 있다. 실제로 이는 인간성과 탁월함에 대한 교육을 목표로 하는 파이테

20) Transhumanist FAQ.

21) 백종현, 「한국 인문학 진흥의 한 길」, 『지식의 지평』 제 2호, 아카넷, 2007, 126쪽 참조.

22) Werner Conze und Reinhart Koselleck(Hg.), “Menschheit, Humanität, Humanismus”, *Geschichtliche Grundbegriffe: Historisches Lexikon zur politisch-sozialen Sprache in Deutschland*, Bd. 3, Klett-Cotta, Stuttgart, 1982 S. 1064.

이아(paideia)의 의미를 전승받았다. 쉽게 말해 ‘스투디아 후마니타스’는 희랍어 파이데이아의 라틴어 번역이라 해도 무방할 것이다.²³⁾ 키케로는 이를 “모든 자유 학예, 즉 자유인의 교육에 적절한 주제들을 포함하는 용어”로 규정하였다. 이는 7예학 즉 문법, 논리, 수사라는 3학(Trivium)과 기하학, 대수학, 천문학, 화성학이라는 4학(Quadrivium)을 이루는 근본이념이 되었다. 중세에 이르러 3학은 신학과 철학을 배우기 위한 수단적 학문으로 다루어졌다. 그러나 중세의 철학은 이미 신학을 위한 설명적 도구 이상의 역할을 할 수 없었다. 이러한 의미에서 좁은 의미의 중세인문학으로 규정할 수 있는 3학은 부정적인 의미로 사용되었다. 피조물로서 인간이 추구해야 할 것은 신성한 것에 대한 학문, 스투디아 디비니타티스(Studia Divinitatis)일 뿐, 후마니타스는 부정되어야 할 대상으로 간주되었다. 그렇기 때문에 3학(trivium)은 하찮은(trivial) 것이 되었고²⁴⁾ 후마니타스는 단지 육신적인 것(Fleischliches)을 의미하기에 이르렀다. 그러나 잘 알려져 있듯 인문학은 르네상스기에 고대의 인문정신을 되찾으며 명실상부한 학문분야로서 체계를 갖춘다. 문헌학, 문법을 포괄하는 의미에서의 언어학과 수사학, 시학, 역사학, 그리고 정치학을 포함하는 도덕철학이 인문학의 핵심 분야로 자리 잡고 당시 대학의 교과과정을 형성하였다.²⁵⁾ 시간이 지남에 따라 신성에 대한 절대적 추구에 대한 반작용으로 등장한 인간성 고양의 이념은 점차 ‘현세적 잘살’에 대한 추구로 변하였고 이는 일정 부분 인문학의 무게중심을 실용성으로 이동시키는 계기를 마련하였다.²⁶⁾ 소크라테스와 소피스트가 동시대에 살았던 것과 마찬가지로

23) 그러나 후마니타스의 의미가 파이데이아와 꼭 일치하는 것이 아니라는 언급에 대해서는 “Menschheit, Humanität, Humanismus” 1066쪽 참조

24) 강상진, 「서양 고중세의 인문정신과 인문학」, 『지식의 지평』 제 2호, 아카넷, 2007, 81쪽 각주 10번 참조.

25) B. Kimball, *B. A. Orators & Philosophers: A History of the Idea of Liberal Education*, Teachers College Press, 1986, 78쪽, F. Hiorth, *Humanismus - Genau Betrachtet*, Angelika Lenz Verlag, 1996, 57쪽 참조.

로 오늘날 대학에는 ‘최고경영자를 위한 철학노트’와 ‘형이상학의 근본문제들’ 교과가 공존한다.

한편, 고대 로마의 9자유학문에서 의학과 건축학이 제외되면서 7과목이 중세 대학으로 이어졌고 그 중 문법학, 수사학, 논리학이 3학이 따로 추려지고 이에 또다시 몇몇 학문 들이 추가되면서 인문학의 체계가 확립되는 과정과 유사한 역사는 현대에도 반복된다. 독일적 전통에서의 인문학. 인문학(Humanitas)은 정신과학(Geistwissenschaft(en))으로 번역된다. 이 말은 현대철학자 딜타이(W. Dilthey)가 밝혔듯 쉘(J. Schiel)이 밀(J. St. Mill)의 도덕 과학(Moral Science)에 대한 번역어다. 그러나 연원을 거슬러 올라가면, 이 단어는 이미 18세기에 등장하였다. 이 개념은 무명으로 발표된 누가 계몽가인가?(Wer sind die Aufklärer?)라는 논문에서 신에 대한 연구와 구분되는 어떤 것으로 언급된다. 당시 이는 철학과 동의어로 사용된 세계지(Weltweisheit)의 한 부분을 이루는 것으로 이성론, 자연론, 법론, 정치론 등으로 구성되는 것으로 언급되었다. 이후 19세기 초에 칼리닉(E. A. E. Calinich)이 “자연과학과 정신과학의 방법”이라는 책을 발표함을 기점으로, 딜타이에 의한 양자의 구별이 고착화 되면서 정신과학의 범위에서 자연과학은 자연스럽게 배제되었다. 그러나 괴테가 “인문학적 교양이 이제 더 이상 인간의 심정을 움직이지 못한 지 벌써 한 세기가 되었다”고 토로하듯, 인문학은 다시 발달한 자연과학의 영향으로 인간의 모든 행위와 정서적 행동을 물리적, 생리적 실체로 환원될 수 있다는 실증주의적 경향을 띠게 되었다.²⁷⁾ 이러한 움직임은 오늘날까지 지속되고 있으며 앞에서 다룬 포스트휴머니즘도 그 영향 아래 있다고 할

26) A. Grafton and L. Jardine, *From Humanism to the Humanities: Education and the Liberal Arts in Fifteenth- and Sixteenth Century Europe*, Harvard University Press, 1986.

27) 최성환, 「서양분과학문 속의 인문학의 전개과정」, 『지식의 지평』 제 2호, 아카넷, 2007, 107쪽 참조.

수 있다. 나아가 20세기에 들어 대표적으로는 니콜라이 하르트만(N. Hartmann)의 학문을 예로 들 수 있듯이 문화학이 그 범주에 들어왔고 마르크스철학에서 사회정신과학이라는 용어가 등장함으로써 사회학 또한 인문학의 영역에 속하게 되었다.²⁸⁾

이상에서 살펴본 바와 같이 인문학의 역사는 자기정립과 자기부정의 평풍게임이다. 하이데거는 니체가 말한 신의 죽음을 형이상학의 종말로 고쳐 읽었다. 푸코는 『사물의 질서』에서 ‘인간의 종말’을 역설한다. 카시러(E. Cassirer)는 계몽의 완성자라 평가받는 칸트를 형이상학의 파괴자라 평가하였다. 한 시점의 끝이 또 다른 시작이 되고, 한 편에서는 해체의 선구자가 다른 편에서는 이념의 완성자가 되는 일은 인문학의 역사를 살펴보면 그리 드물게 일어나는 사건이 아니다. 이렇게 본다면 인공지능시대에 발맞추어 기존의 인문정신의 전복을 주장하는 포스트휴머니즘도 그다지 과격적이거나 새롭게 여겨지지 않는다. 거칠게 말하자면, 포스트휴머니즘은 여전히 인문학의 품 안에 있다. 그리고 이것이 우리가 다음 장에서 포스트휴머니즘의 저편을 살펴보기 위해 다시금 근대철학의 모토를 되짚어 보고자 하는 이유다. 인공지능시대의 인문학을 대표하는 포스트휴먼 인문학이 자신의 모든 사상적 토양을 부정하는 전혀 다른 새로운 학문의 탄생을 표방한다고 하는 것은 유구한 인문학의 역사에 있어 항상 외쳐져 왔던 구호에 불과한 것일 수도 있다. 이러한 입장에서, 우리는 앞에서 살펴본 언어학, 문학, 역사학, 윤리학, 사회학, 문화학은 인공지능의 확산으로 인한 전복과 폐기의 대상이 아닌 새로운 옷을 갖춰 입어야 할 인문학의 본질적 내용이라 주장할 것이다.

28) A. Diemer, Historische Wörterbuch der Philosophie online. (https://www.schwabeonline.ch/schwabe-xaveropp/elibrary/start.xav?start=%2F%2F%5B%40attr_id%3D%27verw.geisteswissenschaften%27%5D#_elibrary_%2F%2F%5B%40attr_id%3D%27verw.geisteswissenschaften%27%5D_1550933851671)

2.3 근대철학의 기획: 반성학으로서의 인문학

우리는 앞에서 학문의 정초작업을 시도한 대표적인 학자로 칸트와 피히테를 꼽았다. 칸트에 따라 인문학의 내용과 방법을 나는 다음,²⁹⁾ 그것의 내용으로 시대가 변함에도 외모를 달리하며 지속되어온 인문학 연구 영역을 그 구성내용으로 언급하였다. 이번 장에서는 피히테의 『전체 학론의 기초』에서 제시하는 동일적 자아의 자기반성이 모든 학문의 근본원리라는 테제를 근거로 서양형이상학의 발원지이지 영원한 수원지인 데카르트와 코기토 테제에 대한 칸트의 해석을 살펴보면서 인공지능인문학 5개 연구영역 도출의 초석을 마련하고자 한다.

학문은 무엇이며, 그것은 어떤 조건을 갖추고 있어야 하는가하는 물음에 천착한 피히테는 그것의 최초의 절대적 조건으로 ‘나는 있다’를 제시한다. 그는 그 자체 자기 충전적인 것으로서의 학문의 근본조건을 캐묻는데 그 원리는 생각보다 상식적이다. 그에 따르면, 세상의 모든 존재자는 자신이 의존하고 있는 원인으로서의 다른 존재를 갖는다는 사고방식과 마찬가지로, “하나의 학문은 체계적인 형식을 갖는데, 이를 구성하는 모든 명제들은 하나의 근본명제와 관계한다.”³⁰⁾ 각자를 지탱하고 있는 근본명제를 갖는 개별학문들 또한 서로 의존관계에 놓여 있다. 결국 피히테의 고민은 학문에 관한 학문, 즉 학문론은 어떠한 것인가에 대한 문제로 귀결되고 나아가 이 근원적인 학문을 지탱하고 있는 근본명제란 무엇인가 하는 질문으로 귀결된다. 이윽고 그는 이 근본명제란 ‘그것이 맺

29) 한편 지금 다루고자 하는 피히테 역시 이러한 전통에 서있다는 사실은 학문 정초작업의 선구자로 피히테를 꼽고자 하는 우리의 선택에 설득력을 더해준다. 그는 『학론의 개념에 대하여』에서 이성적 판단을 형식적인(formalite) 것으로, 이에 따른 결과를 내용적인(materialite) 것으로 나누어 고찰한다. (예: *UBW in: Fichtes Werk I*, 1971, Walter du Gruyter, Berlin, S. 78) 이는 칸트적 전통을 그대로 이어 받은 것이라 할 수 있고 이렇게 학문을 대하는 태도는 본고가 초지일관 취하는 방법과 동일하다.

30) G. W. Fichte, *UBW*, S. 38.

고 있는 여러 의존관계들 이전에 그것들로부터 독립적으로 확실한 명제³¹⁾고 학문 일반에 대한 학문인 “‘학문론’은 순전히 형식적인 학문인 논리학으로부터 근거지워진다”³²⁾고 말한다.

이러한 전제 위에서 그는 논리학의 제일 명제이자 모든 사고의 근본법칙인 동일률을 꺼내든다. 예로든 동일률 ‘A는 A다’의 의미층위를 ‘A다’와 ‘A가 있다’가 아닌 조건문 ‘만약 A가 있다면, A가 있다’로 제한한 피히테는 이 조건문에서 전건과 후건의 필연적 관계에 주목한다. 그리고 이 관계의 필연성을 X라 칭하고 이 필연성은 판단주체가 자아인 한, 최소한 ‘내’ 안에서 주어져야한다고 추론한다. 그렇다면 ‘A는 A다’에서 전자 A와 후자A는 X에 의해 필연적 관계를 맺고 이 필연적 관계는 “만약 자아 안에 있는 A가 정립된다면, A는 정립된다.”³³⁾로 표현된다. 이로부터 “A는 판단하는 자아의 정립함을 통해 자아 일반에 대자적으로 있다”는 명제가 도출되고 이는 ‘나는 나다’라는 명제를 낳는다. 그리고 ‘나는 있다’는 ‘나는 나다’에 함축되어 있다. 바로 이 ‘나는 있다’가 모든 학문론의 “첫번째 단적인 무제약적 근본명제”³⁴⁾다.

사고의 근본법칙인 동일률(A=A)로부터 ‘나는 있다’를 거쳐 추론된 이 조건은 한눈에 알 수 있듯이 그 유명한 데카르트의 테제 ‘나는 생각한다 그러므로 존재한다’에 기인한다.³⁵⁾ 요컨대 ‘나는 있다’는 데카르트의 테제에 대한 피히테의 해석이다. 데카르트의 이 언명에는 두 개의 다른 자아가 등장하는데, 전자는 생각의 주체로서의 ‘나’이고 후자는 생각의 대상으로서의 ‘나’이다. 다시 말해 ‘나는 생각한다’의 ‘나’는 생각의 주체이고 ‘나는 존재한다’의 ‘나’는 ‘생각된 나’다. 그리고 이 두 문장은 추론을

31) G. W. Fichte, *UBW*, S. 41 참조.

32) G. W. Fichte, *UBW*, S. 46.

33) G. W. Fichte, *GLW* in: *Fichtes Werk I*, 1971, Walter de Gruyter, Berlin, S. 91.

34) G. W. Fichte, *GLW*, S. 91.

35) G. W. Fichte, *GLW*, S. 99 참조.

나타내는 접속어 ‘그러므로(ergo)’로 연결되어 있기 때문에 이 양자는 원인과 결과의 관계로 연결되어 있다. 즉 자아의 생각하는 능력은 자아의 존재의 원인이다. 즉 전자는 후자의 필요조건이다. 그러나 ‘나는 있다’를 학문론의 근본명제로 여기는 피히테는 이러한 데카르트의 주장을 뒤집는다. 그의 주장을 축약하면 ‘나는 있다, 그러므로 생각한다’로 표현된다. 이렇게 본다면 “‘생각’이라는 것은 존재하는 내가 갖고 있는 하나의 특별한 속성이다.”³⁶⁾ 이러한 전제 위에서 ‘나는 있다 그러므로 나는 있다’가 도출되고 이는 결국 “첫 번째 나와 두 번째 나의 의미는 전적으로 같은”³⁷⁾ “나는 나다”에 이른다. 이러한 그의 주장은 후에 상이론(Bildtheorie)으로 체계화된다. 그는 “사실은 단지 내가 하나의 상(Bild)이라는 것 뿐 인데 이 상은 생각하는 나(Bild eines denkenden Ich)외에 다른 것이 아니다”³⁸⁾라고 말한다. 생각하는 나의 자기반성의 결과, 즉 사고행위의 산물은 바로 동일한 자기의 상(Bild)이며 이는 다름 아닌 나의 존재라는 것이 피히테의 주장이다. 이처럼 피히테는 철저하고 더 이상의 전제를 깨물 수 없을 때 까지 계속되는 철저하고 엄밀한 반성의 원리를 학문정립의 최초조건으로 보았다.

인공지능인문학 개념정립을 위한 이 연구는 이러한 반성적 태도에 주의를 기울인다. 우리는 앞에서 인문학의 내용으로 언어학, 문학, 역사학, 윤리학, 사회학, 문화학이 ‘포스트휴머니즘’이라는 말로 대표되는 인공지능 시대에도 여전히 유효하다고 주장하였다. 자기반성의 원리를 이에 대입하면, 학문정초의 방법으로서의 반성학이란 언어학, 문학, 역사학, 윤리학, 사회문화학이란 무엇이며 어떠한가에 대한 탐구라 할 수 있다. 그러나 이는 기존의 인문연구영역에 대한 반성일 뿐 ‘인공지능’라는 새로

36) G. W. Fichte, *GLW*, S. 100 참조.

37) *GLW*, S. 96.

38) G. W. Fichte, TB in: *Fichtes Werk II*, 1971, Walter de Gruyter, Berlin, S. 484 참조.

운 질료를 가지는 인공지능인문학 정립이라는 연구목적에 비추어 보면, 여전히 부족하다. 왜냐하면 ‘나는 나다’라는 피히테의 명제는 철저한 반성의 원리에 입각해 있기는 하지만 반성의 범위는 여전히 자기 자신에게만 있기때문에 자기 자신 밖으로 나아가 지금의 세계를 조우해야하는 새로운 인문학 정립에는 부족하기 때문이다. 그렇기 때문에 우리는 그를 통해 데카르트로부터 이어진 학문정립의 원리로서의 반성적 태도만을 취한 후, 데카르트적 반성을 다른 시각에서 해석한 입장으로 주위를 둘러보자.

피히테가 주목한 데카르트적 반성에 대해 칸트는 같으면서도 다른 해석을 내놓는다. 칸트에 따르면 데카르트의 이 논증은 여러모로 불완전하다. 첫째, ‘나는 생각한다 그러므로 존재한다’는 대전제인 ‘모든 생각하는 존재는 존재한다’가 부재하기 때문에 형식상 불완전하다. 둘째, 설령 대전제를 우리가 인위적으로 삽입한다 할지라도 생각으로부터 존재가 도출되는 이러한 자아는 칸트의 입장에서 볼 때 앞에서 살펴보았듯 자기 생각만을 곱씹는 고립된 자아이다. 이러한 생각의 배경에는 데카르트의 테제 ‘나는 생각한다’에 대한 칸트의 독특한 해석이 놓여있다. 칸트에 따르면 데카르트가 천명한 ‘나는 생각한다’는 ‘나는 주장한다’와 같은 목적어가 빠져 있는 불완전한 문장이다. 이 목적어의 자리에 다시 ‘내’가 들어와 ‘나는 나를 생각한다’라는 문장을 구성할 수 있지만, 그렇다면 이는 피히테의 주관적 관념론과 다를 바 없게 된다. 칸트가 이를 불완전한 문장으로 여긴 것은 그가 생각할 때 우리의 사고의 대상은 항상 나의 의식 밖으로 향해 있기 때문이다. 목적어에는 생각하는 나 자신만을 대상으로 삼는 내용이 텅 빈 반성이 아니라 세상을 지각한 경험적 자아가 자리한다.³⁹⁾ 이제 반성의 축수는 내성이 아닌 지금 여기에 마주하고 있는 생활 세계에 뻗친다. 이를 다시 우리의 연구에 대입한다면, 인문학이 지금 경

39) H. Kim, *Zur Empirizität des "Ich denke" in Kants Kritik der reinen Vernunft*, 2016, Könighausen & Neumann, 136 f. 참조.

험하고 있는 세계와 인문학간의 관계맺음에 대한 반성이 학문의 정초작업의 근본명제로 도출된다. 이로써 ‘인공지능인문학’의 첫 번째 구성요소인 ‘인공지능’이 논의의 범주로 진입할 수 있는 통로가 생겼다. 인공지능인문학의 주체는 인문학이고 주체가 경험하고 있는 대상은 인공지능이다. 주체인 인문학은 스스로를 반성하여 자신의 항구적 존재성을 항시 명확히 인식하는 동시에 이미 자신의 경험의 대상에 놓여 있는 인공지능과의 관계를 통해 변모하고 있는 자신의 모습에 대한 재정립을 시도한다.

3. ‘인공지능인문학’과 인공지능인문학의 연구영역

나는 앞에서 ‘인공지능인문학’의 세 가지 구성 요소를 도출함에 있어 첫째로 ‘인공지능’과 ‘인문학’을 구분하였고 나아가 ‘인문학’을 내용과 형식의 구분이라는 인문학 전통의 접근법에 따라 ‘세계해석학’, ‘반성학’으로 나누었다. 그리고 인문학사에 대한 약속을 통해 세계해석학으로서 인문학 연구영역의 지속적 유효성을, 그 내용을 담는 형식인 방법론으로서 ‘반성’의 원리를 도출하였다. 반성학으로서의 인문학은 첫째, 반성의 시선이 안쪽으로 즉 자기 자신으로 향할 때, 그 본연의 역할을 끊임없이 곱씹게 된다. 이를 통해 우리는 변화하는 시대에 맞추어 자신을 변모하면서도 본질적인 내용을 지속할 수 있는 인문학의 본연의 힘의 근원은 반성력에 있다는 사실을 논증하였다. 앞에서 언급했듯이, 언어학, 문학, 역사학, 윤리학, 사회학, 문화학 등으로 분류되는 전통인문학의 내용은 오늘날에도 여전히 그 유용하다는 주장을 뒷받침한다. 둘째, 반성의 범위가 인문학이 현재 마주하고 있는 세계에 자신을 노출시켰을 때 발생하는 내면의 변화와 세계와 향하게 되면 ‘인공지능에 대한 인문학적 반성’의 길이 열리게 된다. 이상을 토대로 인공지능인문학의 자기정립과정을 정리하면 다음과 같다.

1단계: ‘인공지능인문학’의 내용: 인공지능

2단계: ‘인문학’의 내용 - 세계해석학으로서의 인문학: 언어학, 문학, 역사학, 윤리학, 사회학, 문화학

‘인문학’의 방법 - 반성학으로서의 인문학: 언어학, 문학, 역사학, 윤리학, 사회학, 문화학의 자기정립

3단계: ‘인공지능인문학’의 방법 - ‘인공지능’에 대한 인문학적 반성

4단계: ‘인공지능인문학’의 내용: 인공지능을 반성의 대상(내용)으로 하는 언어학, 문학, 역사학, 윤리학, 사회학, 문화학

이상에서 알 수 있듯이 내용과 형식, 대상과 방법이라는 틀로 분석되고 다듬어진 ‘인공지능인문학’은 인공지능 시대에 기초인문학의 내용으로서 지속하고 있는 언어학, 문학, 역사학, 윤리학, 사회학, 문화학의 본 모습을 간직하는 동시에 ‘인공지능’이라는 학문적 대상이 자신 안에 침투함으로써 감당해야 하는 쇄신의 모습을 나타낸다. 인공지능인문학의 연구영역 체계는 이러한 자기보존과 자기쇄신의 모습을 담고 있다.

첫째, 우리는 인공지능을 대상으로 하는 문학, 역사학 중심의 인문학 연구를 인공지능 기술비평학으로 이른다. 문학 특유의 인문학적 상상력은 현재 포스트휴머니즘 담론을 이끌고 있으며 인공지능을 주제로 하는 여러 SF영화들 역시 진화된 문학의 작품 영역이라 할 수 있다. 인간에 의해 창조된 인간의 모습이 투영된 초인간, 비인간, 반(半)인간에 대한 서술, 그들의 등장에 따른 새로운 사회상 제시, 그들을 바라보는 생물학적 인간의 내면에 대한 묘사 등은 문학의 중요한 과업으로 문학의 역사와 함께 하였다. 그리스 신화에 등장하는 오토마톤(자동기계) 탈로스를 시작으로 “과학적 신화”⁴⁰⁾의 대명사 프랑켄슈타인, 최근 쏟아져 나오는 인공지능 영화의 인공지능 인물들은 문학적 상상력에 의한 잠재적 현실

40) 유발 하라리, 조현욱 옮김, 『사피엔스』, 김영사, 2017, 582쪽.

태다. 한편 인물사, 시대사, 지역사 중심의 미시적 역사연구풍토와 더불어 현재는 빅히스토리과 같이 자연 내 인간종의 지위에 대한 본질적인 질문을 던지는 통합적 역사연구도 활발하다. 선사 시대 이전에 대한 역사적 추측과 마찬가지로 인공지능의 진화에 따른 미래상 역시 역사연구의 한 축을 형성하고 있다. 인공지능 기술비평학은 이러한 인문학적 상상력에 기술의 본질에 대한 사실적 성찰과 인공지능 기술에 대한 실증적 분석을 토대를 제공하여 ‘있음직한 이야기로서의 문학작품과 역사적 추측’에 ‘현실성’이라는 무게 추를 매달고자 한다.

이를 통해 첫째, 인공지능시대의 인간의 본질에 따른 기술 발달의 특성과 인공지능기술 속 인간성의 변화 양상에 대한 연구를 진행한다. 이를 위해 인공지능 발달사에 대한 분석을 수행하고 기계가 대체할 수 없는 인간 지능의 본질 규명을 시도하는 한편, 산업생산성의 프레임을 벗어난 문화적 차원에서의 인공지능기술의 가능성에 대해 연구한다. 둘째, 초인공지능 시대 도래의 역사적 가능성을 타진하고 만약 도래할 경우, 인간의 존엄성 보전을 위한 기술연구 가능성을 탐색한다. 구체적으로 인간의 생애주기변동에 따른 새로운 인간상에 대한 추측 등에 대한 연구를 토대로 인간과 자연, 인공지능의 바람직한 공존을 위한 에코시스템 구축을 위한 이론적 토대를 마련한다.

둘째, 인공지능을 대상으로 하는 언어학 중심의 인문학 연구를 인공지능 관계·소통학으로 규정하고 인공지능의 등장과 발전에 따른 세계 내 존재자들의 관계 재설정과 소통의 변화양상을 연구한다. 앞 장에서 서술했듯 데카르트적 코기토의 위세는 인간의 형이상학적 지위뿐 아니라 인간본질의 이해 방향을 결정한다. 그것의 영향력이 커질 때 자연 안에서의 인간의 지위는 높아지고, 이성과 의식을 본질로 하는 인간의 존재론적 특성의 색채는 짙어진다. 반대의 경우, 인간과 비인간 존재자의 구별은 유연해지는 동시에 이성과 의식은 인간의 본질이 아닌 인간의 특성에 대한 추후적 설명으로 여겨진다. 인간중심주의와 해체주의의 반복사는 데

카르트적 코기토의 크기에 따른다. 데카르트적 코기토가 토대주의 휴머니즘의 근간을 형성하는 반면, 포스트휴머니즘은 이를 해체해야 할 원인으로 본다.⁴¹⁾

인공지능 관계·소통학은 첫째, 언어학의 궁극적 아포리아라 할 수 있는 토대주의에 근간한 보편 문법의 가능성을 역설적으로 지금 포스트휴먼시대에 진단한다. 이러한 시도가 가능한 이유는 인공지능 기술의 발달로 인해 빅데이터의 집적이 가능해졌기 때문이다. 우리는 이를 이용하여 보편 문법과 유형론 구축의 이론적 가능성 확인을 시도한다. 둘째, 인공지능 기술의 발전에 따른 관계, 소통 채널의 다양화에 주목하여 각 채널에 알맞은 의사소통 모델에 대한 연구를 진행한다. 예를 들어 챗봇의 등장은 컴퓨터 메신저를 더 이상 인간 간 소통을 위한 ‘메신저’로만 치부할 수 없게 만들었다. 만약 우리가 반려로봇이나 챗봇을 하나의 소통주체로 인정한다면, 소통과 관계의 채널의 지형도는 급격하게 변하게 된다. 소통의 채널은 인간 대 인간의 소통에서 인간과 인간, 인간과 기계로 확장될 것이며, 나아가 기계와의 소통에 익숙한 인간들의 의사소통 모델도 중요한 연구주제가 될 것이다. 예를 들어 인간과 인공지능의 감성적 소통 연구로서 인공지능의 출력감성과 인간의 교감 양상 연구를, 인공지능과 인간 소통 연구로서 딥러닝 방식을 통한 개별어 습득과 자동 통번역의 현황과 미래 연구를, 인공지능들 간의 소통 가능성에 대한 연구를 진행한다. 이는 소통 방식의 변화에 따른 소통주체의 정체성 변화에 대한 계보학적 연구를 동반한다.

인공지능인문학의 세 번째 연구영역은 인공지능 사회·문화학이다. 인공지능 사회·문화학은 인공지능의 출현에 따른 사회, 문화 현상을 분석하고 그것이 4차 산업혁명기로 분류되는 현재의 사회 문화에 미치는 영

41) 보편 문법과 의식의 연역적 체계의 관계에 대한 논의는 N. Chomsky, *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*, Cambridge University Press, 2009 참조.

향 및 쟁점을 검토한다. 인공지능 기술비평학의 포커스가 인공지능 기술의 발달에 따른 개별 주체의 인격성의 변화 양상에 맞춰져 있다면 인공지능 사회문화학은 기술비평학의 논의위에 개별 문화권, 사회구조, 집단, 체계의 변화 양상에 대한 거시적인 안목에서의 담론 형성을 시도한다. 이를 위해 첫째, 인공지능이 가져온 각 사회, 문화 분야의 변화와 이전 산업혁명이 가져온 변화와의 차이에 대한 연구를 토대로, 전쟁, 자본주의, 산업화와 같이 급변기의 세계사회에 대한 이론적 진단을 감행한 기존 사회학의 작업과 같이 인공지능 상품의 보편화와 상용화에 따른 사회, 문화영역의 쟁점을 들추어내는 한편, 이를 담론화할 수 있는 새로운 연구 방법론을 모색한다. 이를테면 섹스로봇의 등장과 상용화가 성적 주체로서의 개인의 내면과 행위방식에 초래할 수 있는 변화양상을 기술비평학에서 추적한다면, 사회·문화학에서는 이로 인한 남녀 관계, 가족 관계와 개념의 변화, 나아가 그 기술 자본을 소유하고 생산할 수 있는 집단과 그렇지 못할 집단 사이에서 벌어질 수 있는 변화와 문화적 갈등 요소를 예상한다. 둘째, 이를 토대로 인공지능 시대의 사회, 문화 변동이 초래할 수 있는 부정적인 영향에 대한 인간의 대응, 존엄성 보존 방안 연구의 기초 자료를 축적한다. 이 연구는 네 번째 연구영역인 인공지능 윤리·규범학으로 이어진다.

인공지능 윤리·규범학은 위 연구영역학문들이 제시하는 연구 성과와 연동하여 궁극적으로 인공지능 시대에 적합한 실질적인 규범을 제시하는 것을 목적으로 한다. 첫째, 이를 위해 인공지능 개념과 실재간의 차이, 인공지능 개념이 함유하고 있는 다양한 층 차와 범주를 명확히 구분하여 제시한다. 인공지능 산업의 발달과 더불어 ‘인공지능’이라는 단어의 사용 빈도는 급격히 증가하고 있다. 이를테면 수년 전에는 단순한 가전제품이 었지만 최근에는 인공지능 제품이라는 광고가 붙기도 하고, 인공지능을 소재로 한 소설과 영화, 드라마는 기하급수적으로 많아져 우리는 더 이상 이러한 소재에 신선함을 느끼지 못하기도 한다. 이는 인공지능의 실

체, 개념 범주 규정의 문제를 야기한다. 실현 가능한 인공지능과 상상 가능한 인공지능, 기술로서의 인공지능, 지능적 존재자로서의 인공지능에 대한 엄밀한 철학적 규정은 활용 가능한 인공지능 실천적 규범의 선결조건 중 하나다. 넓은 범주에서 철학 연구에 속하는 윤리·규범학은 이러한 개념적 연구를 일차적으로 수행한다. 이는 인공지능 존재자의 존재론적 지위 규정연구로 이어진다. 강 인공지능(Strong AI), 초 인공지능(Super AI), 범용 인공지능(General AI) 개념의 혼재, 특이점(Singularity) 도래의 현실성 논의 등은 근접하게는 개념에 대한 분석적 탐구에 관계하지만, 넓게 보면 인공지능의 본질에 관한 문제, 나아가 지능의 본질에 관한 문제로 이어진다. 포스트휴먼 시대, 인문학 가치 고양을 위한 인공지능인문학 구축을 연구목표로 하는 우리는 인공지능을 도구적 존재로 한정한다는 목적 아래, 인공지능에 관한 윤리 이론을 정립한다. 존재론적 지위 규정에 관한 연구가 ‘의식’의 본질규명과 같은 심리철학, 형이상학적 연구를 기반으로 한다면 윤리 이론 정립은 존재론적 지위 정립을 바탕으로 인공지능과 인간의 관계 변화에 따른 인공지능 대우 문제에 주안점을 둔다. 전자는 설명 가능한 인공지능(explainable AI) 개발과 후자는 인공지능으로 인한 윤리적 진공상황 해소와 관계한다. 인공지능 윤리는 결과적으로 인공지능 거버넌스와 관련한 실천적이고 구체적인 규범 제작과 관계한다. 우리는 로봇윤리현장과 같은 범윤리적 차원의 규범제작과 더불어 생산되는 헬스케어 로봇, 킬러로봇, 반려로봇 등 인공지능 로봇 각각에 대해 생산자, 소비자, 판매자를 위한 윤리 가이드라인을 제시하여 사회적 안전망 구축을 위한 건전한 담론 형성에 참여할 것이다.

이제 마지막으로 ‘내용과 방법’의 프레임을 ‘인공지능인문학’에 다시 적용하면 이로부터 앞에서 서술한 인공지능을 내용으로 갖는 인문학, 즉 인공지능에 대한 인문학 외에 인공지능기술을 방법으로 사용하는 인문학, 즉 인공지능에 의한 인문학이 범주에 들어오게 된다. 방법론으로서의 반성의 대상이 ‘인공지능’(기술)이 된다면 인공지능기술을 인문학연구의 방

법으로 삼는 인문학, 다시 말해 인공지능기술이라는 그릇에 인문콘텐츠라는 내용을 담은 인문학인 디지털인문학이 인공지능 인문데이터해석학이라는 이름으로 자리매김하게 된다.⁴²⁾ 인공지능 인문데이터해석학은 첫째, 인문데이터의 전산화(computing), 분석, 시각화를 주제로 하는 디지털인문학의 성과를 활용한 ‘인간성’ 빅데이터 구축을 수행한다. 그 첫 번째 단계로 다양한 인문콘텐츠에서 재현된 인간의 감정 표현 양상들을 데이터화하여 입체적인 방식으로 구현하는 감정 온톨로지 구축을 시도한다. 인공지능 윤리·규범학이 인공지능인문학의 제반 연구영역의 성과를 인공지능 규범 제작이라는 통합적 성과로 발산한다면, 인공지능 인문데이터해석학은 각 연구영역에서 접근 가능한 인문콘텐츠의 체계적 데이터화를 위한 방법적 토대를 제공한다. 둘째, 좁게는 디지털 기술을 활용한 “문학, 예술작품 등에 나타난 인공지능의 재현양상” 분석, 넓게는 인문고전 자료의 현대적 해석을 연구의 주제로 삼는다. 기존의 인문학 연구가 “직관, 체험, 표현, 이해, 해석”이라는 전통적인 인문학 방법론에 의한 활자텍스트에서 활자텍스트로의 이행인데 비해, 인공지능 인문데이터해석학이 표방하는 인문학 전산화(Humanities Computing), 데이터 마이닝(Data Mining), 데이터 시각화 등의 방법론을 이용하여 산출되는 인문고전텍스트의 결과물은 활자텍스트의 범위로부터 자유로운 결과를 산출한다. 요컨대 인공지능 인문데이터해석학은 ‘인공지능이 등장하는 (넓은 의미의 인문데이터인) 인문고전콘텐츠에 대한 현대적 해석’, 넓은 의미의 인공지능 기술인 디지털인문학 방법론을 이용한 인문고전의 현대적 해석 그리고 인공지능 시대의 인간성 재발견을 위한 디지털 기술을 이용한 인간성 데이터의 입

42) 디지털 인문학이 “전통적인 인문학의 주제를 계승하면서 연구 방법 면에서 디지털 기술을 활용하는 연구”(김현 외 2인, 『디지털 인문학 입문』, Huebook, 2016, 17쪽.)를 의미한다는 것을 본 연구의 맥락에서 재해석하면, 그것의 ‘내용’은 전통 인문학의 주제이고, ‘방법’은 디지털 기술이라 말할 수 있다. 이러한 의미에서 우리는 인문학 연구의 방법론으로 정의한 인공지능 인문데이터해석학이 디지털 인문학을 함축한다고 이해한다.

체적 집적을 위한 방법론적 토대를 의미한다.

이상의 내용을 요약하면 다음과 같다. 인공지능인문학의 내용은 인공지능 기술비평학, 인공지능 관계·소통학, 인공지능 사회·문화학, 인공지능 규범·윤리학으로 구성되고, 방법론은 인공지능 인문데이터해석학이다. 인공지능인문학의 내용을 구성하는 네 개 연구영역은 유기적 체계성을 갖는다. 인공지능 기술비평학이 감행하는 인공지능 기술발전 추적과 기술발전에 따른 개별주체의 인격성과 행위 양식 변화는 다른 세 개 연구영역 연구의 주춧돌을 놓는다. 인공지능 관계·소통학은 그 위에서 인공지능 기술과 인격 주체의 탐구 범위를 언어와 소통으로 한정하여 연구를 진행하는 한편, 인공지능 사회·문화학은 사회의 집단 문화의 변화양상에 주목하며 인공지능 기술비평학의 근간에 발을 딛고 있는 동시에, 연구대상에 있어서는 그것과 대척점을 형성한다. 인공지능 윤리·규범학은 앞의 세 연구영역의 성과를 수렴하여 인공지능 시대의 인간 존엄성 고양을 위한 철학을 수립한다. 인공지능인문학 방법론으로서 인공지능 인문데이터해석학은 디지털 기술을 이용하여 스스로의 학문체계를 구축하는 동시에 앞의 네 연구영역의 연구를 견인한다.

4. 나가는 말: 포스트-포스트휴머니즘(Post-Posthumanism), 코휴머니즘(Co-Humanism)

우리는 위에서 ‘인공지능인문학’의 학문적 정체성과 그 요소들의 내적 연관성을 수립하기 위해서 ‘인문학’이 변화하는 시대에 맞추어 외형을 변화시키면서도 자신의 본질적 특성을 지켜왔다는 사실을 근거로 새로운 인문학 정립을 위해서 내용과 형식을 구분하는 전통인문학의 하나의 접근법이 유효할 수 있다는 것을 논의의 출발점으로 삼았다. 이를 토대로 인문학의 최신 조류 중 하나인 포스트휴머니즘의 특징으로부터 중세, 근

대를 아우르는 역사적 어원학적 탐구를 하여 전통인문학의 연구영역학문을 추려내었고 이로부터 인공지능인문학의 5개 연구영역을 추론하였다. 이 연구의 대전제는 포스트휴머니즘의 저편에는, 그것이 인문학이고 그것의 연구 주체가 인간인 이상, 여전히 인본주의적 태도가 자리잡고 있(어야한)다는 것, 그렇기때문에 인공지능 시대를 포함하여 우리가 맞이하는 미래가 어떠한지 인문학은 인간의 입장에서 상생의 인문학이 되어야 한다는 것이다. 인공지능인문학은 우리가 인공지능과 공존하기 위한 ‘인문학’(Co-Humanism)이다.

우리가 정립해 가고자 하는 이 인문학을 가장 간단히 정의하면 ‘인공지능 시대에 인간은 무엇을 할 것이며, 인간의 가치는 어떻게 규정되어야 하는가’에 대한 답변이라 할 수 있다. 인공지능이 인간의 사고와 활동의 많은 부분을 대신할 것이라는 전망 속에 수만년동안 지속되어 온 ‘생산자로서의 인간’, ‘노동하는 인간’이라는 인간의 기본적인 정체성의 변화도 어렵잖지만 이미 가시권에 들어왔다. 우리는 이러한 변화의 시기에 우리 인간은 무엇을 하고 있을 것인지, 또 무엇을 해야 할 것인지, 또 그 과정은 어떻게 진행될 것인지, 그 과정이 어떻게 진행되어야 바람직할 것인지에 대해 실천적 설득력을 가진 답을 내놓는 연구를 계속해서 진행하고자 한다.

참고문헌

- 강상진, 「서양 고중세의 인문정신과 인문학」, 『지식의 지평』 제 2호, 아카넷 2007.
- 김현 외 2인, 『디지털 인문학 입문』, Huebook, 2016.
- 니체, 김정현 옮김, 『도덕의 계보학』, 책세상, 2006.
- 백중현, 「한국 인문학 진흥의 한 길」, 『지식의 지평』 제 2호, 아카넷, 2007.
- 유발 하라리, 『사피엔스』, 김영사, 2017.
- 임석원, 「비판적 포스트휴머니즘의 기획: 배타적인 인간중심주의 극복」, 『인간과 포스트휴머니즘』, 이화여자대학교출판문화원, 2013.
- 슈테판 헤어브레히터, 김연순, 김웅준 옮김, 『포스트휴머니즘』, 성균관대학교출판부, 2012.
- 최성환, 「서양분과학문 속의 인문학의 전개과정」, 『지식의 지평』 제 2호, 아카넷, 2007.
- A. Diemer, Historische Wörterbuch der Philosophie (https://www.schwabeonline.ch/schwabe-xaveropp/elibrary/start.xav?start=%2F%2F*%5B%40attr_id%3D%27verw.geisteswissenschaften%27%5D#_elibrary_%2F%2F*%5B%40attr_id%3D%27verw.geisteswissenschaften%27%5D__1550933851671)
- A. Grafton and L. Jardine, *From Humanism to the Humanities: Education and the Liberal Arts in Fifteenth- and Sixteenth Century Europe*, Harvard University Press, 1986.
- B. Kimball, B. A, *Orators & Philosophers: A History of the Idea of Liberal Education*, Teachers College Press, 1986.
- F. Fukuyama, “Transhumanism”, *Foreign Policy* 144, 2004.
- F. Hiorth, *Humanismus - Genau Betrachtet*, Angelika Lenz Verlag, 1996.
- G. W. Fichte, *UBW in: Fichtes Werk I*, Walter du Gruyter, Berlin, 1971.

- G. W. Fichte, *GLW in: Fichtes Werk I*, Walter du Gruyter, Berlin, 1971.
- G. W. Fichte, *TB in: Fichtes Werk II*, Walter du Gruyter, Berlin, 1971.
- H. Brands, *Cogito ergo sum*, Karl Alber, Freiburg, 1982.
- H. Kim, *Zur Empirizität des “Ich denke” in Kants Kritik der reinen Vernunft*, Königshausen & Neumann, 2016.
- I. Kant, *Grundlegung zur Metaphysik der Sitten(AA4)* in: *Kants gesammelte Schriften (Sog. Akademie-Ausgabe)*, Berlin: Walter de Gruyter, 1900 ff.
- M. Heidegger, *Über den Humanismus*, Klostermann, 2000.
- N. Badmington, “Pool Alighty!; Or, Humanism, Posthumanism, and Strange Case of Invasion of the Body Snatchers”, *Textual Practice* 15.1, 2001.
- N. Bostrom, Transhumanist FAQ: (<https://nickbostrom.com>)
- N. Bostrom, “A history of transhumanist thought”, *Journal of Evolution & Technology* 14(1), 2005. (<https://nickbostrom.com/papers/history.pdf>)
- R. Braidotti, *The Posthuman*, Polity, 2013.
- R. Kurzweil, “The Coming Merging of Mind and Machine”, *Scientific American Report Special Edition*, 2008.
- Tertullianus, *De Praescriptione Haereticorum*(http://www.tertullian.org/works/de_praescriptione_haereticorum.htm)
- S. Sorgner, “Nietzsche, the Overhuman, and Transhumanism”, *Journal of Evolution & Technology*, Vol. 20(1), 2009.
- W. Conze und R. Koselleck(Hg.),“Menschheit, Humanität, Humanismus”, *Geschichtliche Grundbegriffe: Historischers Lexikon zur politisch-sozialen Sprache in Deutschland*, Bd. 3, Klett-Cotta, Stuttgart, 1982.

Beyond the Posthumanism

- A Try to Laying a Conceptional Foundation Artificial Intelligence Humanities -

Kim, Hyeongjoo · Lee, ChanKyu (Chung-Ang Univ.)

This research aims to establish a kind of foundation for the establishment of “artificial intelligence humanities”. To do this, we first classify the three elements of ‘artificial intelligence’, ‘artificial intelligence’, ‘contents of humanities’, and ‘methods of humanities’. Second, we will examine the history of the concept of humanities from the humanism of the era of artificial intelligence, called the fourth industrial revolution era, to the Renaissance and modern times. Through this, they pursue the viewpoint of post-humanists who read the age of artificial intelligence and extract the contents of the traditional humanities that they are presupposing. Lastly, we apply the reflective role as a methode of humanities, which is one of the essential mission of that, to the content extracted from the previous one. Through this, we derive five parts of artificial intelligence, which are restructured by accepting the changes of the artificial intelligence age on the base of the traditional parts of humanities.

Key words: Artificial Intelligence Humanities, Artificial Intelligence, Humanities, Post-humanism, World-Hermeneutics, Reflective Charater of Humanities

철학탐구 제53집

김형주 e-mail: godwithhj@cau.ac.kr

이찬규 e-mail: leeck@cau.ac.kr

투 고 일	2019년 01월 22일
심 사 일	2019년 01월 29일
계재확정	2019년 02월 21일