

다산학 35호 (2019.12) | 209~240

AI와 인간의 공진화共進化와 관련해서 본 다산 철학

정재현 | 서강대 철학과 교수

목차

1. 이끄는 글: 왜 다산 철학인가?
2. 다산철학에 있어서의 주체와 책임의 문제
3. 다산 철학에 있어서의 상관적 관계론
4. AI의 존재방식과 남겨진 과제

1. 이끄는 글: 왜 다산 철학인가? *

1950년 앨런 튜링(Alan Mathison Turing)의 튜링 테스트가 등장한 이래 기계도 인간과 같이 생각할 수 있는 가능성이 끊임없이 제기되었지만 적어도 직관이나 감성의 영역에서는 기계가 끝내 인간을 대체할 수 없으리라는 생각이 있었던 것도 사실이다. 그러나 2016년 인공지능 알파고가 세계적인 프로바둑기사 이세돌을 꺾자, 인간에게만 가능한 것이라고 여겼던 직관이나 감성의 영역도 머지않은 장래에 인공지능(AI)에 의해 구현될 수 있을 것이라는 우려 섞인 전망이 나오게 되었다. 합리적 계산은 물론이고, 직관이나 감성의 영역에서도 인간에 비해 월등한 능력을 지닌 인공지능의 출현이 사람들에게 우려 아니 심지어 공포를 주는 이유는 무엇인가? 그것은 아마도 인공지능이 모든 면에서 인간을 압도했을 경우 벌어질 수 있는 여러 우려스러운 가상적 상황 때문일 것이다. 그것은 단지 비슷한 아니 더 월등한 능력을 가진 존재의 출현으로 인해, 소위 만물의 영장으로서의 인간의 존엄성이 손상되는 것만이 아니라, 그런 존재에 의해 인간이 노예처럼 지배당하거나 멸망당할 수 있다는 우려와 공포라고 할 수 있다. 이런 우려와 공포는 오래 전부터 수많은 공상 과학영화나 소설에 등장하곤 했지만,¹⁾ 2016년 알파고의 등장은 이런 우려가 단지 공상에만 그치지 않을 수 있음을 보여주었다. 이러한 위협적 상황에 대해 인간이 대응할 수 있는 방식은 대체로 두 가지 일 것이다. 하나는 그런 상황이 도래하지 않도록 인공

* 이 논문은 2016년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임.(NRF-2016S1A5A2A03927217)

1) 예컨대, 알렉스 프로야스(Alex Proyas) 감독에 의해 만들어진 공상과학 영화 <아이, 로봇>(I, Robot, 2004)을 들 수 있다. 여기서 인간으로부터 독립해서 자율적으로 생각하게 된 인공지능 로봇이 인간을 위한다는 명목으로 인간에게 제한을 가하는 상황이 그려진다.

지능의 개발을 통제(즉 개발을 지연 혹은 중지) 하는 것이고, 다른 하나는 인간과 그런 우월적 존재와의 공진화共進化를 모색하는 것일 것이다. 이 글은 후자의 입장을 취한다. 그것은 진실을 알려는 인간의 무한한 호기심 앞에 과학의 발전을 지연 혹은 중지하려는 시도는 인류역사를 통하여 늘 성공적이지 않았고, 가능하지도 않았으며, 아마도 이는 미래에도 그러할 것이라고 보기 때문이다.

AI의 위협에 대해 내가 AI와의 공동 번영의 방식을 모색하고, 그 모색의 와중에 동아시아 철학 혹은 다산철학에 주목하는 것은 단순한 일련의 생각들을 통해서이다. 먼저 AI의 출현을 위협으로 느끼게 되는 것은 어떤 배경적 믿음이 개입되어 있기 때문에 그런 것이 아닌가 생각이 들었다. 그리고 만약 그렇다면 AI가 위협이 되지 않을 수 있는 상황을 구현하는 데에는 이런 배경적 믿음을 새로운 철학적 믿음으로 대체함으로써 가능하지 않을까 하는 생각을 하게 되었다. 그리고 마지막으로 그 새로운 철학적 믿음을 형성함에 있어서 동아시아철학이나 다산철학은 도움이 될 수 있다는 생각이 들었다. 좀 더 구체적으로 말하면, 단적으로 나는 AI의 능력이 인간의 생존을 위협할 것이라는 우려는 현대인에게 팽배한 개체 실체론과 본질주의(혹은 위계주의나 기초주의)의 믿음들이 촉발시킨 것이라고 본다. 이러한 믿음들은 너무나 광범위하게 진리로 믿어지고 있기에, 우리는 뛰어난 능력을 지닌 AI의 출현 가능성이 제기되자마자 그것에 대해 적개심내지 경계의 눈길을 보내는 것이라고 생각한다. 즉 세계는 개체들로 이루어져있으며, 개체 사이에는 어떤 능력이나 힘의 관계에 있어서 상하의 위계가 성립할 수밖에 없다는 생각들은 AI를 인간과 별개의, 그리고 월등한 능력의 개체로 바라보게 하며, 이런 인식은 우리로 하여금 초능력을 지닌 인공지능의 개체가 인간 개체들을 제치고 새로운 지구의 주인이 되는 상황을 그리게

한다는 것이다. 나는 이러한 개체주의적 본질론에 맞서서 상관적 관계론을 그 대안적 사고로 제안하려고 한다. 상관적 관계론이란 존재의 기본 단위가 개체가 아니고, 오히려 개체 간의 상관적 관계라는 형이 상학적 주장이다. 따라서 이에 따르면 개체 실체론에서 말하는 개체나 개체의 본질적 속성들이란 것도 그 독립적 존재근거가 없어지게 된다. 사실 지구상에서 개인주의와 자본주의의 팽배와 그로 인한 다양한 모순적 상황에 맞서서 상관적 관계론이 그 대안적 사고로 주목을 받았지만, 나는 여기서 이 상관적 관계론을 AI와의 공존에 필요한 대안적 사고로 제시하는 것이다. 나는 이 상관적 관계론이 동아시아의 철학²⁾ 특히 다산철학에 의해 효과적으로 현대에 재전유될 수 있다고 믿으며, 또한 이 상관적 관계론이 인공지능 전문가들에 의해서도 인공지능의 존재방식으로 인정되어진다고 주장한다. 물론 이런 상관적 관계론을 단순히 이론적 가능성으로 제시하는 것만으로 인공지능과의 공존이 자동적으로 이루어지는 것은 아니고, 이 이론에 대한 헌신(commitment)이 필요하다는 것을 주장하려고 한다. 그리고 이 헌신의 강조가 바로 다산 철학의 특징인 자율성의 강조와 궤를 같이 하는 것이라고 주장하려고 한다. 이러한 주장들을 위해 이 글은 먼저 다산철학이 동아시아 철학 중에서도 왜 효과적인 상관적 관계론인지를 보여주고서, 이러한 상관적 관계론이 인공지능 전문가에 의해서도 인공지능의 존재방식으로 제안된다는 점을 지적할 것이다. 마지막으로 결국 상관적 관계론에 의 헌신이 중요하며, 이것이 다산철학의 자율성 강조에 의해 암시되었다는 것을 지적하는 순서로 구성하려고 한다.

2) 동아시아의 철학적 세계관이 상관적 관계론이라는 주장은 많은 학자들도 공감하는 것이다. 예컨대, 천지일체나 만물일체로 표현되는 이러한 관점은 아이만호(Philip J. Ivanhoe) 등에 의해 Oneness Hypothesis로 일컬어진다. Philip J. Ivanhoe(eds), *The Oneness Hypothesis: Beyond The Boundary of Self*, New York: Columbia University Press, 2018.

2. 다산철학에 있어서의 주체와 책임의 문제

상관적 관계론이 AI와 인간의 공존에 시사를 준다고 해서 상관적 관계론이 만병통치의 약은 아니다. 즉 단순한 상관적 관계론은 무조건적인 대안이 될 수 없고, 또 마냥 반가운 것만은 아님을 지적해야 한다. 상관적 관계론은 자칫하면, 즉 잘못 이해하면 누구에게도 책임을 물을 수 없는 無化의 상황, 즉 자율성이나 주체의 소멸을 함축하는 일종의 신비주의를 의미할 수 있기 때문이다. 나아가 동아시아의 상관적 관계론 즉, 비개체주의적 혹은 비주체주의적 철학이 그것이 의도했던 평등성이 아니라 오히려 상반된 경향, 즉 위계성으로 빠질 수 있는 위험성은 모종삼 같은 이들에 의해 지적되어져 왔다.³⁾ 나는 동아시아 상관적 관계론이 흔히들 생각하듯이 신비주의나 위계주의를 지향한 것은 아니며, 그 증거로 동아시아 상관적 관계론에는 도덕적 책무의 개념이 암묵적으로 있었다고 주장한다. 또한 이런 도덕적 책무의 개념이 동아시아 철학 내에서는 다산의 철학에서 가장 뚜렷하게 보인다고 주장하려 한다.

핑가레트(Herbert Fingarette)는 공자의 철학에서 흥미롭게도 선택이나 자유의지의 개념이 보이지 않는다고 주장하였다.⁴⁾ 그러나 이러한 모습은 어쩌면 공자만이 아니고 상관적 관계론의 세계관을 견지하는 많은 동아시아 철학자에게도 적용할 수 있는 테제일 것이다. 즉 그것은 성리학은 물론이고 다산의 철학에도 해당되는 이야기 일 수 있다. 다산도 공자와 마찬가지로 ‘갈림길’(cross-road)이 없는 길, 즉 하나

3) 모종삼, 『政道與治道』, 《모종삼선생전집》 10, 대만: 학생서국, 신판서 26-32쪽.

4) Herbert Fingarette, *Confucius- The Secular as Sacred*, New York: Harper & Row, Publishers, Inc., 1972, pp. 18~36.

의 옳은 길에 따르든지 아니면 그 길에서 벗어나든지 밖에 할 수 없는, 다시 말하면 선택권이 없는 상황을 생각했다고 볼 수 있다. 다산이 인간을 동물과 구별시켜 주는 것으로 생각했던 도의지성道義之性이 '인간에게 하나의 올바른 길'(道義)을 상징하기 때문이다. 도의지성이 이처럼 여러 길 중의 하나가 아니고, 유일한 길을 상징한다면, 다산이 강조했던 자주지권이나 권형의 개념도 진정한 의미에서의 자유의지나 선택의 개념이 아니라고 할 수 있다. 이것은 윤리적 덕을 인간다움의 길로 생각한 유교 사상가들에게는 불가피한 상황일 것이다. 그러면 공자, 주자, 다산과 같이 상관적 관계론을 지지하는 동아시아 철학자들에게서 자아는 소멸되는 것일까? 아니다. 많은 유가의 철학자들은 확장된 자아의 개념으로 이 문제를 해결하려고 하였다. 유가의 인인仁人이 가리키는 인격상은 바로 이런 확장된 자아, 혹은 획득된 본성을 자신의 본성으로 여기는 사람에 해당할 것이다.⁵⁾ 필자는 이러한 유가의 기본 입장에 기본적으로 동의하면서, 그들의 확장된 자아에 대한 주장들이 다산의 철학을 통해 더욱 객관적으로 보강될 수 있다고 본다. 왜냐하면 다산은 상관적 관계론을 유지하면서도, 즉 개체 사물의 상호관련적 존재성을 주장하면서도, 자아의 책임의 문제에 대해서 어느 동아시아 사상가보다도 뚜렷한 입장을 표했기 때문이다. 다시 말해 단순히 암묵적으로 행위자의 행위에 따른 책임이나 행위에서의 자율적 선택을 말한 것이 아니고, 보다 직접적으로 행위자의 행위에 따른 책임이나 행위에서의 자율적 선택을 거론하고 있다.

다산이 책무나 자율성을 명백하게 드러낸 것은 그가 성性보다는 심心(靈體, 靈明)을 강조하는 데에서 먼저 드러난다. 다산이 지적하듯이,

5) Joel J. Kupperman, *Learning from Asian Philosophy*, New York: Oxford University Press, 1999, pp. 17-23.

성리학 전통에서는 성性を 중심으로 놓은 사람들도 있었고, 심心を 중심으로 놓은 사람들도 있었다. 예컨대, 주자가 성性を 중심에 놓았다면,⁶⁾ 다산은 심心を 중심으로 생각한 사람이라고 할 수 있다. 심心を 중심으로 하였다는 것은 그가 원리의 지각을 원리 자체보다 중요하게 생각하였다는 것이다. 한마디로 주어진 것 보다는 그 주어진 것을 어떻게 바라보아야 하는지, 어떻게 처리해야 하는지에 초점을 두었다고 할 수 있다.

다산의 주체에 대한 강조는 성性の 개념에 대한 그의 시선에서도 보이는데, 그는 주자와는 달리 성性を 주어진 선천적 원리로 보지 않고, 주체의 기호嗜好로 본다. 즉 원리보다는 그 원리의 인지 내지 의식을 강조하는 것이다.

성性이란 사람 마음의 기호嗜好를 말하니, 채소는 똥거름을 좋아하고 연꽃은 물을 좋아하는 것과 같다. ‘사람의 성[人性]’은 선을 좋아하니 선善을 실천하고 의義를 쌓으면 활기차고 당당해지지만, 악을 행하고 마음을 저버리면 막혀서 굼주리게 된다. 예전 유학자들이 성性에 대해 말한 것은, 모두 맹자의 본뜻이 아니다.⁷⁾

천명의 성性 또한 기호嗜好로 말할 수 있다. 사람이 잉태하게 되면, 천이 영명하고 형상이 없는 실체를 부여하는데, 그 실체라는

6) 물론 이것도 정확한 것은 아니다. 주희가 性은 물론이고, 心도 강조했었다는 것은 의심할 수 없기 때문이다. 주희의 철학을 존숭하였던 퇴계를 위시한 조선의 성리학자들이 心을 중심으로 한 일련의 논쟁에 열중했던 것이 이를 잘 보여준다. 다만 다산은 기존의 성리학자들에 비해 상대적으로 더욱 心에 주안점을 두었다고 할 수 있다는 것이다.

7) 『大學講義』, 「傳七章」(『定本』 제6책, 153쪽). “性者, 人心之嗜好也. 如蔬菜之嗜糞, 如芙蕖之嗜水. 人性嗜善, 行善集義則茁壯, 行惡負心則沮餒. 先儒言性, 皆非孟子之本旨也.”

것은 선을 즐거워하고 악을 미워하며 덕을 좋아하고 더러운 것
을 부끄러워하는데, 이를 성性이라 한다.⁸⁾

나아가 다산의 주체성은 일종의 자유의지로 해석이 되는 그의 권형
權衡이나 자주지권自主之權의 개념에서가장 명확하게 드러난다고 할
수 있다. 다산은 권형이나 자주지권을 통해 선善은 물론 악惡도 선택할
수 있다고 보았다고 한다. 하지만 이것은 일반적으로 자유의지가 그러
하듯이 어떤 사람이 악惡을 선택했을 때, 그의 선택에는 아무런 문제가
없고, 따라서 우리는 그가 행한 선의 선택과 마찬가지로 악의 선택도
존중해주어야 함을 의미하지 않는다. 권형이나 자주지권은 이런 의미
에서의 자유의지가 아니다. 권형을 통해 악을 선택할 수 있다고 하는
것은 다만 악의 선택에 의해 발생한 그 결과를 감내해야 함을 의미할
뿐이다.

하늘은 이미 사람에게 선을 할 수도 악을 할 수도 있는 권형權衡
을 주었는데, 이에 그 아래 면에는 선을 하기는 어렵고 악을 하기
는 쉬운 육체(具)를 주었으며, 그 윗면에는 선을 즐거워하고 악을
부끄러워하는 성性을 주었다.⁹⁾

하늘이 영지靈知를 부여할 때 거기에는 재才도 있고 세勢도 있고
성性도 있다. 재才라는 것은 그 능력이요, 그 권형이다. 기린은

8) 『中庸自箴』, 「中庸自箴一」(『定本』 제6책, 228쪽). “天命之性, 亦可以嗜好言, 蓋人之胚胎既成, 天則賦之以靈明無形之體, 而其爲物也, 樂善而惡惡, 好德而恥汚, 斯之謂性也.”

9) 『心經密驗』, 「心性總意」(『定本』 제6책, 200쪽). “天既予人以可善可惡之權衡. 於是就其下面, 又予之以難善易惡之具, 就其上面, 又予之以樂善恥惡之性.”

착한 것으로 정해져 있기 때문에 착한 것이 공이 되지 않고, 사랑은 악한 것으로 정해져 있기 때문에 악한 것이 죄가 되지 않는다. 사람은 그 재才가 선을 할 수도 악을 할 수도 있는데, 능력은 자력에 달려 있고 권형은 자주에 달려있다. 그러므로 선을 하면 그를 칭찬하고 악을 하면 그를 꾸짖는다.¹⁰⁾

하늘이 사람한테 자주지권을 주었다. 가령 선을 하려고 하면 선을 하고 악을 하려고 하면 악을 하여, 향방方向이 유동적이고 정해지지 않아 그 권능이 자신한테 있으며, 금수禽獸가 일정한 마음을 갖고 있는 것과는 같지 않다. 그러므로 선을 행하면 실제로 자신의 공功이 되고 악을 행하면 실제로 자신의 죄가 된다.¹¹⁾

다산의 자율성이나 주체성은 악의 원인에 대한 그의 생각에서도 보인다. 다산은 성리학의 일반적 관념과 같이 악惡이 기질의 편협함, 투박함에 의한 결과라고 생각하지 않는다. 그는 전부 다는 아니지만 인간의 도덕적 악행이 자신의 자율적 판단에 따른 것이라고 보았다. 따라서 인간은 자신이 저지른 악행에 대한 책임을 져야 한다. 만약 악惡의 원인이 성리학자들의 관념과 같이 대부분 타고난 기질氣質 때문이라면, 악의 결과는 우리가 책임지는 것이 아니다. 대체로 인간은 대개 자신이 선택한 것 혹은 자신이 통제 가능한 것에만 책임이 있다고 믿

10) 『매씨서평』, 「閻氏古文疏證抄 一 : 南雷黃宗羲序」(『定本』 제13책, 362쪽). “天之賦靈知也, 有才焉, 有勢焉, 有性焉. 才者, 其能・其權也. 麒麟定於善, 故善不爲功, 豺狼定於惡, 故惡不爲罪. 人則其才可善可惡, 能在乎自力, 權在乎自主, 故善則讚之, 惡則訾之.”

11) 이지형 역주, 『譯註 茶山 孟子要義』, 현대실학사, 1994, 135~136쪽.

어지기 때문이다.¹²⁾ 이런 이유로 다산은 악惡의 원인을 자신의 타고난 기질보다는 자신이 잘 못 마음을 사용한 때문으로 보았다. 한마디로 인간의 악은 기질氣質 때문만이 아니라, 적극적으로 선을 실천하지 않았기 때문이다.¹³⁾

맹자孟子는 성性を 논함에 불선不善을 함닉陷溺에 귀결시켰는데, 송유宋儒들은 성性を 논함에 불선不善을 기질氣質에 귀결시켰다. 함닉은 '자기로부터 말미암는 것'이니 그것을 구제할 수 있는 방법이 있지만, 기질氣質은 '선천적인 것'이니 거기로부터 벗어날 방도가 없다. 그렇다면 사람이 누구인들 자포자기自暴自棄하여 스스로 하류下流의 천한 사람으로 돌아가는 것을 달가워하지 않겠는가?¹⁴⁾

자신을 (악에) 빠뜨리는 것은 혈기의 사욕 때문이거나, 습속의 오염 때문이거나, 외물의 유혹 때문이다. 이 때문에 양심이 없어져 큰 악에 이르게 되니, 어찌 기질의 탓으로만 돌릴 수 있겠는가?¹⁵⁾

12) 물론 본문의 뒤에서 지적하겠지만, 잘 통제되지 않는 곳에 대한 책임도 물을 수 있다. Joel J. Kupperman, *Learning from Asian Philosophy*, New York: Oxford University Press, 1999, pp. 19~20.

13) 다산은 지성의 우월성과 열등성은 기질에 의해 정해지지만 덕성은 그렇지 않다고 보았다.

14) 『맹자요의』(『定本』 제6책, 25쪽). “孟子論性，以不善歸之於陷溺，宋儒論性，以不善歸之於氣質。陷溺由己，其救有術，氣質由天，其脫無路，人孰不自暴自棄，甘自歸於下流之賤乎？”

15) 이지형 역주, 『譯註 茶山 孟子要義』, 현대실학사, 1994, 335쪽. “陷溺之術，或以形氣之私慾，或以習俗之薰染，或以外物之引誘。以此之故，良心陷溺，至於大惡，何得以氣質爲誘乎？”

인간이 인간다운 것은 바로 선택할 수 있고, 또 그에 따라 책임을 질 수 있기 때문이다. 성리학은 만물일체萬物一體의 주장과 같은 관계성은 강하지만 자율성의 측면이 약하다. 반면 다산 철학은 위에서 보듯이 자율성 내지 책임성의 개념이 강하다. 즉 자신이 통제할 수 없는 곳에 책임은 없고, 오직 자신이 선택하고 결정한 것에만 책임이 뒤따른다는 상식적 관념을 보유했다고 할 수 있다. 사실 자신의 의지에 의해 앞으로의 행위가 변화되지 않는다면 그 행위에 대해 칭찬하거나 문책을 하는 것은 의미가 없어 보인다. 이 때문에 상식적인 관점에서는 오직 자발적 행위들에게만 책임을 물을 수 있다고 하는 것이다.¹⁶⁾ 이런 문맥에서 인간과 동물의 차이점은 바로 자신이 주체적으로 선택을 할 수 있느냐로 귀결된다. 다산에 따르면 동물은 본능에 따라서만 살고, 인간만이 선택을 할 수 있다.¹⁷⁾ 인간만이 오직 주체적 존재이다.

개와 소는 먹이를 던져주면 먹고자 할 따름이고, 칼날로써 두렵게 하면 피하고자 할 따름이니, 그것들에게는 단지 기질의 성만 있는 것을 알 수 있습니다. 또 사람은 선악에 대해 모두 스스로 할 수 있어서 그것을 능히 스스로 주체적으로 해나가고, 금수는 선악에 대해 스스로 할 수 없어서, 그 행동이 그렇지 않을 수 없게 된다. 사람은 도둑을 만나면 소리쳐 물리치기도 하고 피를 내

16) Joel J. Kupperman, *Character*, New York: Oxford University Press, 1991, p. 60.

17) 백영선은 톨랜스의 이론을 끌어들이어 이를 다산과 주자의 차이로 드러낸다. 다산은 주자와는 달리 인간은 동물과는 달리 도덕적 행위자 moral agent의 측면이 있다고 하였다. 주자는 품을 그저 자연에 따른 데에서도 성립한다고 하여 동물에게도 도덕성을 인정하였다. 하지만 다산은 오직 도덕을 의지적으로 행한 인간에게만 도덕적 책임을 물을 수 있다고 보았다. Youngsun Back, "Are animals moral?: Zhu Xi and Jeong Yakyong's views on nonhuman animals", *Asian Philosophy*, 28/2, 2018, pp. 97~116.

어 사로잡기도 하지만, 개는 도둑을 만나면 짖어 소리를 낼 수는 있지만 짖지 않고 피를 낼 수는 없으니, 그 능력이 모두 일정하게 정해져 있음을 알 수 있습니다.¹⁸⁾

마지막으로 다산의 주체성은 그의 실천의 강조에서 보여진다. 앞서 성기호설性嗜好說을 통해 원리보다 주체를 강조한 다산의 입장은 또한 원리의 이성적 인지보다 실천적 인지를 강조하게 된다. 도덕성은 실천을 통해 완성되는 것이다.

인仁·의義·예禮·지智의 명칭은 일을 행한 뒤에 이루어지는 것이다. 그러므로 사람을 사랑한 뒤에 그것을 인仁이라고 하느니, 사람을 사랑하기 이전에는 인仁이라는 명칭이 성립되지 않는다. …… 어찌 인仁·의義·예禮·지智의 네 알맹이가 복숭아와 살구의 씨처럼 사람의 마음속에 덩어리로 잠재해 있는 것이겠는가? …… 만약에 사단四端의 이면에 또 다시 이른바 인仁·의義·예禮·지智라는 것이 있어 은연히 잠복하여 주인이 된다면 이것은 맹자의 확충의 공부가 그 근본을 버리고 그 끝을 잡는 격이며, 그 머리를 놓치고 그 꼬리를 잡는 격이다.¹⁹⁾

나는 물론 성리학자들에게 책임의 관념이 없었다고 하는 다산의 비판을 전적으로 수용하지 않는다. 성리학자들에게도 간접적 책임의 개

18) 이지형 역주, 『譯註 茶山 孟子要義』, 현대실학사, 1994, 316쪽. “犬與牛也, 投之以食, 欲食焉而已, 怵之以刃, 欲避焉而已, 可見其單有氣質之性也. 且人之於善惡, 皆能自作, 以其能自主張也, 禽獸之於善惡, 不能自作, 以其爲不得不然也. 人遇盜, 或聲而逐之, 或計而擒之, 犬遇盜, 能吠而聲之, 不能不吠而計之, 可見其能皆定能也.”

19) 이지형 역주, 『譯註 茶山 孟子要義』, 현대실학사, 1994, 91~96쪽.

념이 있다고 우리는 볼 수 있기 때문이다. 예컨대, 우리의 행위는 전적으로 의식적 자아의 선택에 의하지는 않지만, 우리의 노력이나 수양의 일부에 의해 우리의 성품(character)이 변형이 되고 다시 이것이 행위를 낳게 되므로 우리는 우리의 행위에 간접적으로나마 책임이 있다고 할 수 있기 때문이다. 혹은 설사 우리의 성품의 형성에 우리의 의식적 선택이 많은 역할을 하지 않더라도, 즉 우리의 성품이 대체로 비자발적으로 형성된 것이라고 해도 이것이 이 성품으로부터 나오는 행위에 우리가 책임이 없음을 보여주지 않는다.²⁰⁾ 조금이라도 성품의 변형에 우리의 노력이 기여를 하는 바가 있다면, 적어도 우리는 우리의 성품에 어느 정도 책임이 있다. 이렇게 보면 성리학자들에게는 책임의 관념이 암묵적으로 숨어있다고 할 수 있다. 성리학자들이 비록 악을 기질 탓으로 돌렸다하더라도, 그 누구도 선천적 기질 때문에 악한 행위를 했으니, 책임을 면제해 주어야 한다든지 혹은 처벌을 하지 말아야 한다고 주장한 적이 없었던 점도 이런 해석을 지지한다. 성리학자는 물론이고 유가 전통에서는 적어도 우리의 노력에 의해 우리의 성품이 어느 정도 바뀐다는 점을 한 번도 의심하지 않았다. 그러니 당장은 아니더라도 본인의 노력에 의해 우리의 성품이 미래에 바뀔 수 있다면, 성품에서 연유하는 행동의 책임도 자연스럽게 개인에게 물을 수 있을 것이다. 다산의 공헌은 단지 성리학자들의 이런 암묵적 혹은 간접적 책임의 존재를 인정하는 것을 넘어서, 이를 보다 분명히 하여 그 책임을 직접적으로 우리에게 귀속시킨 데 있었다고 볼 수 있다.

앞서 말한 대로, 통제 가능한 자발적 행위에만 책임을 지우는 것은 문제가 있을 수 있다. 분명히 그것이 잘못이라고 알았다하더라도 나의

20) Joel J. Kupperman, *Character*, New York: Oxford University Press, 1991, p. 56.

의지가 약해서 나의 의식적 결심을 행동에 옮기지 못한 것도 불가피한 행동이었다고 볼 수 있기 때문이다. 즉 만약 나의 의식적 행위도 어느 정도까지는 나의 통제를 벗어난 행위라면, 오직 무의식적 행위, 비자발적 행위만이 나의 통제 범위 밖에 있기에 나의 책임이 없다는 주장도 받아들일 수 없을 것이다. 즉 이처럼 내 의식적 행위에 의한 행동으로 보여진 것도 사실은 미리 결정된 행동이었다고 볼 수 있다면, 결국 우리의 책임문제는 단순히 의식적 행위에만 국한되어질 수 없을 것이다. 다시 말해 사실 내가 통제할 수 있는 것과 내가 통제할 수 없는 것의 경계도 그렇게 분명하지 않다는 점이 중요하다. 우리는 끊임없이 나와 내가 아닌 것의 경계에 있기 때문에, 분명히 통제되는 것과 통제되지 않는 것의 경계는 단기간에 나눌 수 있는 것이 아니라, 점진적으로 확보되는 것이다. 나의 통제의 범위는 늘어날 수도, 그리고 혹은 줄어들 수도 있기에, 나의 행동에 대한 책임의 범위는 단지 그 행동이 의식적, 자발적으로 이루어진 것이나 아니냐에 달려 있는 것이 아니다.

3. 다산 철학에 있어서의 상관적 관계론

다산이 비록 자주지권과 권형 등의 개념을 통해 자유의지를 강조한 것처럼 보이지만, 다산의 자유의지는 앞서 말했듯이 개인 단독의 의지가 아니다. 판단과 행위의 주체는 개인적 자아가 아니라 공동체적 자아 혹은 도덕적 자아이다(道義之性). 인간은 선택을 할 수 있는 존재(自主之權)이기에 특별한 존재이지만, 그 선택은 이미 어떤 의미에서는 나 개인에 의해서가 아니라 사회에 의해서 방향 지워졌다고 할 수 있다. 즉 도의지성으로 나아가느냐 앓느냐의 선택만이 있다고 할 수 있다. 앞서

도 말했지만 다산은 인간이 동물과 다른 것은 도의지성道義之性이 있기 때문이라고 보았다. 이런 점에서, 다산의 자유의지가 어떤 의미에 서건 완전한 개인적, 자유주의적 자유의지가 아님은 분명하다. 그것은 그의 자유의지가 어디까지나 도의지성의 한계 안에서 표현되었기 때문이다.

사람은 도의와 기질을 합하여 하나의 性이 되는 것이 본연이다.
금수가 오로지 기질지성만 가지는 것도 또한 본연이다. 어째서
(사람의 본연을) 반드시 기질과 상대해서 말할 필요가 있는가?²¹⁾

이처럼 자유의지가 영명과 자주지권으로 표현되었다면, 다산이 상 관적 관계론의 전통에 서 있는 것은 도의지성으로 표현되었다고 할 수 있다. 그런데, 도의지성은 동물과 다른 인간의 본질적 속성을 말하는 것인데, 이것이 상관적 관계론을 함축한다는 것은 무슨 의미인가? 일반적으로 만물일체론은 일종의 동물과 인간의 연속성을 말하는 반면, 위에서 말하는 도의지성은 일종의 도덕적 본성으로 인간을 동물로부터 구별시켜주는 인간의 본질을 가리키는 것이다. 도대체 인간과 동물이 본질적으로 다르다는 것이 인간과 동물의 연속성을 주장하는 것과 어떻게 연결될 수 있는가? 정반대의 주장이 아닌가? 다산은 실제로 만물일체론에 대해 거부감을 보인다.

만물일체 그것은 옛 경전에서는 절대 나오지 않는 말이다.²²⁾

21) 『맹자요의』(『定本』 제7책, 189쪽). “人之合道義・氣質而爲一性者, 是本然也, 禽獸之單有氣質之性, 亦本然也, 何必與氣質對言之乎?”

22) 『중용강의』(『定本』 제6책, 295쪽). “萬物一體, 其在古經, 絕無此語.”

한자경은 이러한 다산의 언급에 비추어 다산을 외재주의자라고 한다. 그는 성리학과 불교의 입장을 내재주의라고 보고, 다산은 『천주실의』의 영향을 통해 이러한 동아시아전통에서 벗어나게 되었다고 한다. 다산에 있어서 세계에 조화를 가져다주는 것은 만물에 내재되어 있는 원리에 의해서가 아니라, 어디까지나 외적 존재인 신에 의해 주어진 것이라고 보는 것이다. 그녀의 말에 의하면, 다산에게서 일자는 오직 신이며, 도의지심과 같은 것은 인간을 타자와 구별시켜주는 표층적 종차에 불과하다고 한다.²³⁾ 그러나 하나의 존재 원리인 도의지성은 그것이 사물 내부에 있건, 아니면 밖에 있건 만물의 창조와 관련이 있는 신에 의해 주어진 것이기에 이것은 적어도 사물들 사이에는 조화의 원리, 혹은 상호 연관성이 있음을 보여주는 것이다. 성리학의 본연지성이 만물 일체론을 함축하는 것이 사실이라면, 다산의 도의지성도 만물일체론과 무관한 것은 아닐 것이다. 그러므로 도의지성은 단순히 인간을 타 존재와 구별시켜주는 지표의 역할만 하는 것이 아니다.

이런 점에서 우리는 도의지성과 앞서 말한 영명 혹은 자주지권의 관계를 다시 생각해보아야 한다. 어떤 의미에서는 다산에게서는 영명 혹은 자주지권보다 기호로서의 성性, 도의지성으로서의 성性이 중요하다. 앞서 다산은 인간과 동물을 구분하는 근거를 권형이나 자주지권인 것처럼 말했으나, 사실 더 중요한 것은 도의지성이다. 이러한 도덕적 선을 지향하는 본성이 없으면 결코 인간은 도덕적인 삶을 살 수가 없기 때문이다.

만일 선을 즐기고 악을 부끄러워하는 성性を 하늘이 내려주셔서

23) 한자경, 『한국철학의 맥』, 이화여자대학교출판부, 2015, 296~98쪽.

그것으로 하여금 선을 선호하게 하고 의로움을 살찌게 만들지 않았더라면, 세상을 떠날 때 까지 모든 힘을 다할 지라도 지극히 미미하고 조그만 선행도 이루기 어려울 것이다 …… 이 성성이 인간에게 있다는 것은 지극히 귀한 보배이니, 존중하고 받들어 한 시라도 떨어지거나 어긋나지 말아야 할 것이다.²⁴⁾

인간의 주체성을 강조하는 자유의지론, 인간이 관계적 존재임을 나타내는 도의지성, 그리고 이런 도의지성을 완성하는 것이 우리의 노력에 의해 가능하다는 것²⁵⁾이 모두 인간의 마음에 함께 포함되어 있다는 점이 중요하다.

영체(마음) 안에는 세 가지 이치가 있다. 그 性에 대하여 말하면 선을 즐거워하고 악을 부끄러워한다. 이것은 맹자가 말하는 성선性善이다. 그 권형에 대하여 말하면 선할 수도 있고 악할 수도 있다. 이런 측면이 고자의 단수 비유이고 양옹의 선악 혼재설이 나오게 된 이유이다. 그 행사에 대하여 말하면 선해지기는 어렵고 악하기는 쉽다. 이런 측면에 순자의 성악설이 나오게 된 이유이다. 우리 사람의 영체 안에 이 세 이치가 없는 것이 아니다. 하늘이 사람에게 선할 수도 있고 악할 수도 있는 권형을 이미 주었다. 그 아래로는 선하기 어렵고 악하기 쉬운 육체를 주었으며 그

24) 『心經密驗』, 「心性總義」(『定本』 제6책, 199쪽). “就其上面, 又予之以樂善恥惡之性. 若無此性, 吾人從古以來, 無一人能作些微之小善者也. …… 聖人以性爲寶, 罔敢墜失者以此.”

25) 백민정은 그의 논문에서 다산의 자유의지를 이성적 자유의지가 아니라, 수양론적 자유의지라고 하였다. 백민정, 「다산 심성론에서 도덕감정과 자유의지에 관한 연구」, 『한국실학연구』 14, 한국실학학회, 2007, 441쪽.

위로는 선을 즐거워하고 악을 부끄러워하는 성性を 또 주었다.
만약 이 성이 없었다면 우리 인간은 예부터 아주 하찮은 작은 선
이라도 할 수 있는 자가 하나도 없었을 것이다.²⁶⁾

이것은 아마도 『중용』의 가장 첫 구절들 즉 “하늘(天)이 명령한 것을
본성이라고 한다. 그 본성을 따라가는 것을 길이라고 한다. 그 길에 따
라 나아가기 위해 노력하는 것을 가르침이라고 한다.”(天命之謂性, 率性
之謂道, 修道之謂教)라는 말이 보여주는 통찰일 것이다. 다시 말해, 다산
은 인간과 동물을 구분하는 근거를 다음과 같이 두 가지로 든다.

1. 도의지성道義之性이다. 우리의 영명은 도덕적인 것을 선호하
는 기호 즉 성性を 가지고 있다.
2. 권형權衡이다. 권형은 우리의 영명의 자주지권이고 자유의지
이다.

이처럼 다산은 의지와 욕구(기호)를 통해 우리의 마음에 대해 말했
다. 그렇다면 이러한 다산의 도덕 심리학 체계 속에 이성온 별도로 존
재하지 않는 것인가? 아니다. 우리는 의지와 욕구 혹은 경향성을 이성
과 연결시킬 수가 있다. 먼저 위에서 말한 의지와 욕구는 천天으로부터
온 것이라는 점에 주의해야 한다.

26) 『心經密驗』, 「心性總意」(『定本』 제6책, 199쪽). “總之, 靈體之內, 厥有三理. 言乎其性, 則樂
善而恥惡, 此孟子所謂性善也. 言乎其權衡, 則可善可惡, 此告子湍水之喻, 揚雄善惡混之
說所有作也. 言乎其行事, 則難善而易惡, 此荀卿性惡之說所有作也. 荀與 楊也, 認性字本
誤, 其說以差, 非吾人靈體之內, 本無此三理也. 天既予人以可善可惡之權衡. 於是就其下
面, 又予之以難善易惡之具, 就其上面, 又予之樂善恥惡之性. 若無此性, 吾人古以來, 無一
人能作些微之小善者也.”

‘천이 나에게 생명을 부여할 때 덕(선)을 좋아하는 성향과 선을 택하는 능력을 주었다. 이것은 비록 나에게 있는 것이지만 본래 천명이다.’²⁷⁾

나아가 천 뜻을 서구의 신神과 유사하게 볼 수 있다면, 신에 의거한 경향성과 의지가 이성과 연결될 수 있다는 생각은 아퀴나스(Thomas Aquinas)에게서 보인다. 먼저 아퀴나스는 이성을 법과 같은 것으로 보는데, 이것은 이 둘이 다 ‘어떤 것을 다른 것으로 이끌고 명령하는 것’을 본질적 작용으로 갖기 때문이다.²⁸⁾ 아퀴나스는 나아가서 ‘우리의 자연본성에 각인된 경향성에 따라 작용하는 것이, 자연 본성적으로 선한 것이고, 또한 신의 모상으로서, 이성적 존재로서 인간의 이성에 따르는 것이라고 주장한다.’²⁹⁾ 천이 나에게 덕을 좋아하는 성향(경향성 혹은 기호)과 선을 택하는 능력(의지)을 주었다는 것은 아퀴나스의 관점과 비교하면 일종의 이성적 방식을 갖게 되었다는 것이라 할 수 있다.

다산의 도의지성이 개체세계관보다 상관적 세계관, 만물일체론에 더 가깝다고 하는 것은 바로 이 도의지성이 천으로 왔다는 사실에서 기인한다. 앞서 지적했듯이, 흔히 다산이 영향을 받은 것으로 보이는 『천주실의』에는 분명히 만물일체론을 비판하는 부분이 있다. 바로 이 점 때문에 한자경은 다산이 자주지권이나 권형을 주장했을 때, 그것을 다산이 만물일체론을 비판하는 것으로 읽었던 것이다. 실제로 다

27) 『中庸自箴』, 「『中庸自箴一』」(『定本』 제6책, 231쪽). “天賦我性, 授之以好德讀作善之情, 畀之以擇善之能, 此雖在我, 其本天命也.”

28) 이상섭, 「법은 인간적 행위의 외적 원리일 뿐인가」, 『중세철학』 22호, 한국중세철학회, 2016.

29) 앞의 논문.

산은 모든 만물이 공유하는 의미의 본연지성을 거부하고, 인간만의 독특성을 강조하기 위해 인간의 본연지성을 말한다. 물론 이 본연지성에서도 특별히 그 한 측면인 도의지성은 인간의 본질을 잘 드러내 주는 것이라고 할 수 있다. 그러나 도의지성은 인간의 본성으로 볼 수 있지만, 개별자의 개성이라기보다는 인간종의 본성이라고 할 수 있다. 또 한 그 인간종의 본성은 천으로부터 유래하는 하나의 특정한 성질이라는 점에서 도의지성은 하나의 길을 말하는 유가전통에 여전히 기대어 있고, 이 하나의 길을 강조한다는 점에서 만물간의 조화를 강조하는 만물일체론에 가까울 수 있다고 할 수 있다. 마치 순자가 천지인天地人의 차이를 말하면서도, 각각이 각자의 역할을 가지고 같이 참여한다는 점에서 통합된다고 하였듯이, 도의지성은 인간의 본질을 나타내면서 또한 하늘로부터 유래하는 존재의 원리로, 사물의 연결성을 강조하는 이론이라고 할 수 있다. 따라서 다산이 비록 만물일체론을 비판하였지만 그가 비판하는 만물일체론은 인간의 자주성을 부인하는 신비주의적 만물일체론이었던 것이고, 그는 사실 전체적으로 개체실체론보다는 상관적 관계론을 함축하는 만물일체론에 가까운 사상을 펼쳤다고 할 수 있을 것이다.

4. AI의 존재방식과 남겨진 과제

AI는 위계적 존재방식인 피라미드형의 상부를 차지하면서 아래의 열등한 인간에게 명령을 내리는 개체적 존재가 아니고, 인간 개체와 함께 연결되어 있는 관계적 존재이다. 이러한 AI의 존재방식에 대한 관념은 인공지능 전문가라고 할 수 있는 마이크로소프트 연구원(Microsoft

Research) 호비츠(Eric Horvitz)의 다음과 같은 주장에서 볼 수 있다: “인공지능(AI)은 정말로 어떤 개체가 아니다. 그것은 풍부한 하위-지식 분야들, 방법들, 통찰, 지각, 발화와 대화, 결정과 기획, 로봇공학 등등의 집합이다.”(AI is not really any single thing- it is a set of rich sub-disciplines and methods, vision, perception, speech and dialogue, decisions and planning, robotics and so on.)³⁰⁾ 이 말의 의미는 AI는 인간과 별개로 존재하는 개체가 아니라, 개체 인간들에 의해 영향을 받는 관계적 존재라는 의미이다. 그러므로 인간도 AI와 함께 관계적 존재를 이루면서 자율적인 존재로 살아갈 수 있다는 것이다. 인간은 늘 AI로부터 가르침을 받는 열등한 존재가 아니라, AI를 변화시키는 동등한 존재이다. AI는 결코 인간과 별개의 개체로 자신을 규정하지 않을 것이고, 이는 인간도 마찬가지일 것이다. 이것은 마치 뇌의 각 부분들이 각각 나름의 역할을 하면서도 하나의 개체로 여겨지지 않는 것과 마찬가지로, ‘연장된 마음’(extended mind)이라고 볼 수 있는 현재의 컴퓨터와 같이 AI는 인간의 연장이다. 이런 관점에서 과학기술자 라투르(Latour)를 인용하면서 하대청은 다음과 같이 쓰고 있다.

인공지능과 같은 기술이 더 발전하면 할수록 기계의 자율성은 증대되는 것이 아니라 오히려 기계와 인간사이의 연결성은 더욱 커지고 매개는 더욱 깊어질 것이다. 많은 논평가들이 기술의 발전에 따라 인공지능의 자율성은 점점 확대될 것이라고 전망하지만, 그렇게 말하는 ‘자율성’은 수많은 매개와 연결에 의존하고 있으며, 앞으로 인공지능은 인간적, 사회적, 정치적, 물질적 요소

30) <https://www.surveecto.com/blog/how-artificial-intelligence-is-changing-development/>

들과 더욱 복잡하게 매개될 것이다. 온갖 미디어 기기들이 늘어나면서 우리의 기억력과 사고를 기기들이 매개하듯이, 기계들도 수많은 인간노동과 다른 물질적 제도적 장치들에 의해 새롭게 매개되는 것이다. 인간과 비인간의 뒤엎힘은 앞으로 더 커지고, 복잡해지고, 우리의 시야에 잘 보이지도 않고, 자주 인간의 통제를 벗어날 것이다(Latour 1993; 1999).³¹⁾

AI는 위계적 질서의 최상위에 위치하면서 그 밑에 위치하는 인간들에게 아래에서 위로 명령을 내리는 개별자가 아니다. 그것은 오히려 인간을 포함한 수많은 존재들에 의해 연결된 관계적 존재임에 틀림없는 것이다. 이런 상호관계 속에서 AI가 우리의 마음을 잘 알듯이, 우리도 AI의 마음을 잘 알게 될 것이다.

물론 여전히 인간사회가 개체주의와 위계주의에 의해 불평등한 관계로 이루어진 구조를 가지고 있는 것이 사실이다. 또한 이러한 구조 속에 기술의 고도화에 따른 개체적 존재인 사이보그(Cyborg), 트랜스휴먼(Transhuman) 등이 출현할 수 있는 것도 사실이다. 따라서 이런 개체적 존재들이 출현하는 미래의 사회도 여전히 개체주의적이고, 위계적이고, 불평등할 수 있다. 다산의 상관적 관계론이 아무리 매력적이고, 또 인공지능이 아무리 관계적 존재임을 역설한다고 해도, 인간과 인공지능과의 윈윈의 공조는 일종의 희망적 사항일 수 있다. 미래의 현실도 현재와 마찬가지로 얼마든지 위계질서와 차별 속에서 약자가 고통을 받을 가능성이 많다. 그래서 상관적 관계론이나 공진화의 이론을 단순히 이론 속에서 받아들여서는 미래사회가 현재보다 더 나은 사

31) 하대청, 「'휠체어 탄 인공지능' 포스트휴먼 시대의 기술윤리」, 『인공지능의 윤리적 법적 사회적 쟁점들』, 한국포스트휴먼학회 등, 2019년 춘계 연합학술대회, 2019, 113쪽.

회로 도약하리란 기대를 할 수 없다.

그룹에도 많은 과학 공상 영화나 소설에서 그리는 우울한 특정 형태의 미래사회가 그대로 인류 미래의 결론이라고 생각할 필요는 없다. 톰 크루즈(Tom Cruise)가 주연한 영화 <엣지 오브 투모로우>(Edge of Tomorrow, 2014)가 그려내듯이 인간은 다른 인간이 만들어낸 오류를 수정하기 위해서, 그 다른 사람이 죽은 곳에서 다시 시작할 수 있다. 인간의 모든 행위는 더 나은 사회를 위한 인간지식의 자원이 될 것이다. 우리는 우리의 개체주의적 오류 사고를 더 반복하지 않을 수 있는 시기를 인류가 생존하는 한 언젠가는 갖게 될 수 있을 것이다. 물론 이것은 가만히 있어서 도달되는 것이 아니고, 우리의 의지와 노력에 의해 가능하게 될 것이다.

상관적 관계론과 자율성의 강조는 우리가 다산에게서 확연히 볼 수 있는 통찰이다. 하지만 우리가 AI의 공존을 위해 필요로 하는, 동아시아의 상관적 관계론에서 제안하는 이상적 공존의 상태는 그저 노력 없이 가능한 상태가 아니라, 우리의 끊임없는 노력을 통하여 이뤄내야 할 상태이다. 그것은 어쩌면 손쉬운 타협적 공존을 통해서 이루어지지 않고, 치열한 투쟁과 갈등을 통해서 이루어낸 상태일 것이다. 다산이 강조한 자율성이 바로 이 점을 강조하는 것이다. 즉, 상관적 관계론의 믿음은 그저 지적인 이해에서 가능하지 않다. (AI까지 포함하여) 타인을 자신과 연결시키는 것은 우리의 끊임없는 타인에 대한 관심과 배려에 의해 가능해지기 때문이다. 다산은 권형과 자주지권의 개념을 강조하면서 바로 이런 적극적 자유의지의 행사만이 인간과 동물을 구별시킨다고 말하는데, 사실 그것은 그저 우리에게 자유의지가 있다는 사실을 전달하기 위해서라기보다는 상관적 관계론을 실현하기 위해서는 우리의 치열한 노력이 필요함을 보여주는 것이라고 본다. 즉 만물일체의 원리

는 그 원리에 대한 끊임없는 지각과 그 구현에의 노력을 통해 우리가
획득할 수 있는 것이다. 중요한 것은 원리가 아니라, 그 원리에 대한 헌
신(commitment)일 것이다.

(논문투고일 2019년 8월 23일, 심사확정일 2019년 10월 11일, 게재확정일 2019년 10월 11일)

참고문헌

『大學講義』

『梅氏書評』

『心經密驗』

『中庸講義』

『中庸自箴』

모종삼, 『政道與治道』, 《모종삼선생전집》 10, 대만: 학생서국.

백민정, 「다산 심성론에서 도덕감정과 자유의지에 관한 연구」, 『한국실학연구』 14, 한국실학학회, 2007.

이상섭, 「법은 인간적 행위의 외적 원리일 뿐인가」, 『중세철학』 22호, 한국중세철학회, 2016.

이지형 역주, 『譯註 茶山 孟子要義』, 현대실학사, 1994.

하대청, 「‘휠체어 탄 인공지능’ 포스트휴먼 시대의 기술윤리」, 『인공지능의 윤리적 법적 사회 적 쟁점들』, 한국포스트휴먼학회 등, 2019년 춘계 연합학술대회, 2019.

한자경, 『한국철학의 맥』, 이화여자대학교출판부, 2015.

Back, Youngsun, “Are animals moral?: Zhu Xi and Jeong Yakyoung’s views on nonhuman animals”, *Asian Philosophy*, 28/2, 2018.

Fingarette, Herbert, *Confucius- The Secular as Sacred*, New York:

Harper & Row, Publishers, Inc., 1972.

Ivanhoe, Philip J. (eds), *The Oneness Hypothesis: Beyond The Boundary of Self*, New York: Columbia University Press, 2018.

Kupperman, Joel J., *Learning from Asian Philosophy*, New York: Oxford University Press, 1999.

_____, *Character*, New York: Oxford University Press, 1991.

<https://www.surveyccto.com/blog/how-artificial-intelligence-is-changing-development>

기질로 악의 원인을 설명하는 주자학과 잘못된 마음가짐에 악의 원인을 두는 다산 정약용(1762~1836) 중에 누가 수양과 사회 발전에 대한 문제를 해결함에 더 나은 철학인지는 분명하지 않다. 하지만 적어도 책임과 자율성의 문제에 있어서 다산의 철학은 적어도 주자학에 비해 약간의 우월성을 가지는데, 다산은 적어도 우리가 다르게 행동 할 수 있었던 행동들에 대해서만 책임이 있다는 생각을 가지고 있었기 때문이다. 다산은 영명(마음), 자주지권(자유의지) 등의 개념을 통해 자율성을 강조하는 동시에 그것을 도의지성(도덕적 본성)과 연결시켜 자신이 생각하는 자율성이 자유주의의 자율성이 아니라 도덕적 자율성이라는 것을 분명히 한다. 인간의 자율성과 상관적 관계성의 균형을 강조하는 다산의 철학은 미래의 인간과 AI의 공진화 과제에 대한 새로운 관점을 제공할 수 있다. 우리가 수많은 영화와 소설에서 접하는 소위 인공 지능의 위협은 인공 지능이 인간 지능을 넘어서는 초지능에 기초하여 인간을 무시하고 결국 인간을 노예로 만드는 상황을 가리킨다. 하지만 인공 지능으로 인한 이 가상의 위협 상황은 오늘날의 세상을 지배하는 개인주의와 위계적 사고에서 비롯된 것이다. 다시 말해, 인간에 대한 AI의 위협은 AI가 개체로서 생각하고 행동하는 것으로 묘사되고, 그러한 AI가 피라미드 계층의 상위 계층에서 절대적 행위자로 나타나는 상황에서 발생한다. 미래에 실현 될 수 있는 인공 지능과 인간의 평화로운 공존은 다산의 철학에 의해 암시 될 수 있다. 이 모델에 따르면 AI는 개인이 아니라

인간과 AI의 네트워크이며 AI와 인간의 관계는 위계적이 아니라 상호적이다. 인공 지능과 인간 사이의 상호 관계는 우리 인간 두뇌 안의 부분들 사이의 상호 관계라고 생각해 볼 수 있다. 이 모델에서는 AI가 인간 지능을 초월하는 슈퍼 지능을 가지고 있더라도, 그 지능은 인간에게 명령하는 것이 아니라 인간 주체를 통해 얻은 정보로 언제든지 원래의 시각을 수정할 수 있다. 이러한 방식으로 인간의 자율성은 AI에 의해 보존되고 강화 될 수 있다. 따라서 AI는 더 나은 판단을 통해 인간에게 무력감이나 열등감을 주는 개별 실체가 아니라 인간의 판단을 소중한 자원으로 간주하고 인간에게 의존하는 관계적 실체이다. 따라서 인공 지능은 인간에게 자신의 의견을 표현할 수 있는 더 많은 기회를 제공한다. AI가 인간을 이해하는 것처럼 인간은 AI를 이해한다. 인간과 AI는 더 나은 세상을 위해 서로에게 영향을 주고 협력한다. 그러나 인간과 AI의 원인의 공존모델은 단순히 상관적 관계론을 이론적 가능성으로 제시하는 데서 실현되는 것이 아니라, 우리의 끊임없는 노력에 의해 실현가능하다. 이것이 바로 다산이 강조하는 자주지권의 개념이 의미하는 것이다.

주제어 | 다산, 인공지능, 공진화, 개별실체론, 상관관계론.

Abstract

Tasan's Philosophy as a model for AI and human co-evolution

Chong Chae-hyun

Professor of Department of Philosophy, Sogang University

Between the neo-Confucians who blame the cause of evil for their endowed temperament and Tasan(1762-1836), a Korean Confucian scholar, who thinks it as the misguided use of heartmind, I do not know whose philosophy is better for the issues of both self-cultivation and social improvement. However, at least in the matters of responsibility and autonomy, Tasan has an advantage over other neo-Confucians because he holds the idea that we are responsible only for those actions which could be taken differently. Tasan emphasizes autonomy more than any other Confucians through the concepts such as *yeongmyeong*(spiritual brightness) and *jajujiigwon*(moral autonomy). He, however, relating them with the concept of *doujiisong*(the moral nature), clarifies that such autonomy is not autonomy of liberalism, but moral autonomy. Tasan's position, which emphasizes the balance between the autonomy and relationality of humans, provides a new perspective on the task of co-evolution between human and AI in the future. The so-called AI threats which we can now find in numerous films

and novels occur in the situations in which AIs ignore humans based on their super-intelligence beyond human intelligence and finally enslaves humans. But I think these imaginary threatening situations caused by AI come from individualism and hierarchical thinking which dominate today's world. In other words, AI threats to humans occur in the situations where AIs are depicted to be thought and acted as individuals, and where such AIs appear as absolute agents in the upper layers of the pyramid hierarchy. The peaceful coexistence of AI and human being, which can be realized in the future in my view, may be hinted by Tasan 's philosophy. This is what I mean by Tasan's philosophy as a model for AI and human co-evolution. According to this model, AI is not an individual, but a network of humans and AIs and the relationship between AIs and humans is not hierarchical, but mutual. The mutual relationship held between AI and humans can be shown in the mutual relationship among different parts of human brain. In this model, even if AI has super intelligence that transcends human intelligence, its intelligence is not one that commands human beings, but one that can be modified its original view at any time by information obtained through human subjects. Human autonomy in this way can be preserved and strengthened by AI. Therefore, AI is not an individual entity that gives humans a sense of helplessness or inferiority through its better judgment, but a relational entity that considers and relies on human judgment as a valuable resource. AIs, therefore, give humans more chances to express their opinions. Humans understand AIs as much as AIs understand humans. They

are influencing and cooperating each other for a better world. However, the win-win coexistence model of human and AI is not realized by simply presenting the mutual relationality as theoretical possibility, but by our constant efforts. This is what the concept of Tasan's *jajujiigwon*(moral autonomy) meant.

Key Words | Tasan, AI(artificial intelligence), co-evolution, theory of individual substances, correlativism.

