

# 인간과 인공지능의 공존 가능성에 대한 탐색: 책임의 윤리와 문학적 상상력\* \*\*

Exploring the Coexistence of Human and Artificial Intelligence:  
Ethics of Responsibility and Literary Imagination

박소영(중앙대 HK연구교수)\*\*\*

## <목 차>

I. 인공지능 시대의 도래	IV. 인간과 인공지능의 공존의 가
II. 무책임한 신과 책임의 윤리: 『프랑켄슈타인』	능성
III. 자기책임의 윤리: 『안드로이드 는 전기양의 꿈을 꾸는가?』	V. 글을 마치며
	참고문헌
	Abstract

## [국문 요약]

인간은 오랜 기간 동안 문명과 사회를 발전시키기 위해 자신들이 만든 도구 및 기계와 지속적으로 관계를 맺어 왔다. 제 4 차 산업 혁명 시대 또는 포스트휴먼 시대로 접어들면서 인간과 기계와의 관계는 새로운 국면에 접어들었고 이 시대의 가장 중요한 문제 중 하나가 되었다. 인공지능 시대에 우리는 인간과 기계의 경계에 대한 근본적인 혼란에 직면해 있다. 인간을 닮은 기계 인공지능의 출현이 예견되면서 인간과 기계의 공존이라는 주제는 책임의 윤리를 동반하게 된다. 이 연구는 두 편의 공상과학 소설을 중심으로 인공지능과의 공존을 위한 책임의 윤리를 다룬다. 하나는 최초의 공상과학 소설로 잘 알려진 메리 셸리(Mary Shelly)의 『프랑켄슈타인』(*Frankenstein, or Modern Prometheus*)(1818)이고, 다른 하나는 현대 공상과학 소

\* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A6A3A01078538).

\*\* 이 논문은 2018년 인공지능시대 윤리 공동학술대회에서 ‘소설에 재현된 인간과 인공지능의 공존 가능성’이라는 주제로 발표한 원고를 수정, 보완한 것임.

\*\*\* psyong7@cau.ac.kr

설의 대표작 중 하나인 필립 K. 딕(Philip K. Dick)의 『안드로이드는 전기양의 꿈을 꾸는가?』(*Do Androids Dream of Electric Sheep?*)(1967)이다. 『프랑켄슈타인』에서 우리는 인간과 동물의 사체를 접합한 후 전기충격을 가해 생명체를 만들어 내는 제목과 같은 이름의 과학자를 만난다. 『안드로이드는 전기양의 꿈을 꾸는가?』에서 우리는 종말론적 상황에 처한 미래를 만나게 된다. 두 편의 소설을 읽음으로써 우리는 인공지능과의 공존을 위한 책임의 윤리에 대한 문학적 상상력의 가능성을 접하게 된다.

**주제어** : 인공지능, 공존, 책임, 윤리, 공상과학 소설, 문학적 상상력

## I. 인공지능 시대의 도래

“인간이 신을 발명하면서 역사는 시작됐고, 인간이 신이 되면서 역사는 끝난다.” 유발 하라리가 『사피엔스』에서 소개한 이 명제가 본 연구의 단초가 되었다는 것이 조금은 과장이지만 아주 거짓은 아니다. 피조물에서 창조자라는 인간 지위의 변화와 그 결과에 대하여 질문을 던지는 이 문장은 인공지능 시대의 도래에 즈음하여 의미심장하다. 특히 후반부는 더욱 그러하다. “인간이 신이 되면서 역사는 끝난다.”가 의미하는 것은 무엇인가? 인간이 신이 된다는 것은 다른 존재의 창조자가 된다는 것으로 볼 수 있을 것이다. 그런데, 인간이 창조자가 되면 역사가 끝난다는 것은 무엇을 의미하는가? 세계 내 자신의 존재론적 지위를 확보하기 위해, 피조물로서의 자신을 정립하기 위해 절대적 창조자인 신을 상정한 인간이, 절대적 창조자의 자리에 스스로를 위치시키지 못하는 상황에 대한 요약으로도 이 문장을 볼 수 있다. 이로부터 하나의 질문이 제기된다. 인간은 스스로 신이 되고 싶어 만들어낸 자신의 창조물과 양립하는 공존의 길을 걸을 수 있을 것인가?

먼저 인간이 자신의 창조물과 맺어 왔던 관계에 주목할 필요가 있겠다. 인류의 역사와 더불어 인간은 끊임없이 도구 혹은 기계를 만들어 왔고, 그 관계는 기술의 발달과 더불어 새로운 국면을 맞으며 변화해 왔다. “도구 혹은 기술 사용자”로서의 인류는 자신과 주변의 관계를 중심으로 유형무형의 수많은 도구와 기계들을 생산하며 자신들의 길을 걸어온 것이다. 이제 제 4차 산업혁명의 시대 혹은 포스트 휴먼으로 명명되는 시대에 진입하면서, 기계와 인간의 관계의 문제는 다시금 새로운 국면에 접어들었다. 기존의 1차, 2차 그리고 3차 산업혁명으로 명명되는 단계까지 기술에 대한 인간의 주요 관심은 노동력의 증대 및 생산성의 질적인 향상과 같은 양적인 팽창의 문제에 관련된 것이었다. 이제까지는 인간과 기계의 관계가

둘 사이의 경계에 관한 존재론적 혹은 본질적 혼란을 야기하는 문제는 아니었다는 것이다.

소위 포스트휴먼으로 명명되는 시대에 이르러 이제 우리는 인간성 혹은 기계성 사이의 경계, 그리고 인간 존재론에 대한 근본적 혼란에 직면하게 되었다. 단순한 도구의 수준을 뛰어넘어 인간의 지능을 닮은 인공지능의 출현이 예견되거나 혹은 이미 현실화되고 있기 때문이다. 인공지능을 탑재한 포스트휴먼으로서의 그 어떤 존재와 인간 사이의 구분이 모호해지는 미래가 멀지 않아 보인다. 인간과 기계 혹은 기술의 관계가 불분명한 존재론적 경계를 갖는 시대로 진입한 것이다.

인공지능에 대한 장밋빛 기대 못지않게 그로 인한 부작용을 우려하는 목소리 또한 점점 커지고 있다. 인공지능의 발전이 단순히 산업구조의 변화에 국한되지 않고, 인간에 대한 정의 및 경계를 모호하게 할 것이라 예상되기 때문이다. 인공지능 의료의 발달은 단순한 신체 일부의 기능을 향상시키는데 그치지 않을 것이라고 전망된다. 인간 두뇌의 기억 혹은 정보 저장능력과 판단능력을 돕는 인공지능 장치가 인간의 두뇌 혹은 신체에 직접 이식되어 인간이 가진 지능의 한계를 보완할 가능성이 높아지고 있다. 인간의 신체를 닮은 유기체와 정보 형태를 담은 인공지능 확장물이 결합된 사이보그 혹은 안드로이드에 대한 가정은 인공지능이 일상화될 미래에 대한 불안을 자극한다.

인공지능의 출현이 충분히 예견되는 것이 지금의 상황이라 할지라도 인공지능에 대해 지나치게 과장된 불안감을 우리가 가지고 있는 것은 아닌가 하는 생각이 든다. 또한 인간과 인공지능을 이렇게 동일 선상에 놓고 직접 비교하는 것이 과연 가능한가라는 의문도 든다. 왜냐하면, 약한 인공지능과 강한 인공지능의 차이에 대한 논의를 빌어 본다면, 현재 우리가 목도하는 인공지능은 특정 기능 중심의 약한 인공지능(Weak AI)이며, 인간의 능력에 비견하거나 그것을 뛰어넘는 강한 인공지능(Strong AI) 혹은 초인공지능(Super AI)의 출현은 아직은 먼 미래에나 현실화될 것으로 전망되기 때문이다.

강한 인공지능의 출현 자체를 부정적으로 보는 견해도 있다. “중국어방 논쟁”(Chinese Room Argument)으로 알려진 설(John Searle)은 인공지능을 약한 인공지능과 강한 인공지능으로 구분한다. 이 구분에 의하면, 약한 인공지능은 인간의 활동을 지원하는 도구로 활용되지만, 강한 인공지능은 그 자체가 하나의 마음으로 인간과 같은 사고를 할 수 있다고 한다. 그러나 그의 논의에서 주목할 점은 그가 약한 인공지능에서 강한 인공지능으로의 전환을 자연스러운 이행 과정으로 전제하고 있지는 않다는 점이다(Searle, 2004: 90). 인공지능시대가 도래했다고는 하지만 고도로 인간화되었거나 인간과의 구별이 거의 불가능한 강한 인공지

능이라던 인공지능 로봇은 아직은 문학 작품이나 영화 속 상상의 차원에서나 가능하다.

우리에게 익숙해진 인공지능에 대한 상상, 즉 공상과학소설과 영화에 등장하는 성숙한 인공지능의 출현은 어쩌면 한참 지난 미래에나 가능한 일일지도 모른다. 그럼에도 우리는 인공지능 혹은 포스트휴먼과의 공존의 문제를 아포칼립스적 과격으로 미리 가정하고 있다. 그렇다고 해서 기술문명을 포기하는 것은 또한 아니면서도 말이다. 과학기술이 가져오는 아포칼립스적 결말은 공상과학 소설의 흔한 가정 중 하나이다. 공상과학 소설은 인공지능이 등장하는 미래사회를 예측하는 문학적 상상력의 공간이자 인간과 기계의 공존을 위해 요구되는 책임의 윤리를 다루는 공간이기도 하다.

이 연구는 두 편의 공상과학 소설을 중심으로 인공지능과의 공존을 위한 책임의 윤리를 다루고자 한다. 하나는 최초의 공상과학 소설로 잘 알려진 메리 셸리(Mary Shelly)의 『프랑켄슈타인』(*Frankenstein, or Modern Prometheus*) (1818)이고, 다른 하나는 현대 공상과학 소설의 대표작 중 하나인 필립 K. 딕(Philip K. Dick)의 『안드로이드는 전기양의 꿈을 꾸는가?』(*Do Androids Dream of Electric Sheep?*) (1967)이다. 『프랑켄슈타인』에서 우리는 인간과 동물의 사체를 접합한 후 전기충격을 가해 생명을 불어 넣는 제목과 같은 이름의 과학자를 만난다. 『안드로이드 전기양의 꿈을 꾸는가?』에서 우리는 종말론적 상황에 처한 미래의 지구와 인간 그리고 안드로이드를 만난다. 두 편의 소설을 읽음으로써 문학적 상상력을 통해 인공지능과의 공존을 위한 책임의 윤리를 탐색하게 될 것을 필자는 기대한다.

## II. 무책임한 신과 책임의 윤리: 『프랑켄슈타인』

인간은 자신을 모방하여 만들어낸 자기 존재와 욕망의 투사 대상인 인공지능에 대해 책임감을 가질 수 있을 것인가? 이 질문의 제기와 이에 대한 탐색은 최소 200년 전으로 거슬러 올라간다. 메리 셸리(Mary Shelley)는 1818년 소설 『프랑켄슈타인』에서 이 문제를 다룬다. 이 소설은 시체로부터 수집한 각 신체부위들을 연결한 후 전기충격을 가하여 생명을 부여한 과학자 프랑켄슈타인과 괴물의 형상을 한 그의 피조물에 대한 이야기다.

프랑켄슈타인의 생명 창조 행위는 일견 절대자에 의한 인간창조에 비견된다.

그가 생명을 부여한 존재는 자유자재로 움직일 뿐 아니라 주변 환경에 대한 학습을 통해 생존법을 알아낸다. 자신의 존재에 대해 생각하고 사고하며 정체성까지도 탐구한다는 점에서는 인간과 다르지 않아 보인다. 스스로의 감정과 지적인 가능성 그리고 욕망과 꿈을 가지고 있다는 점에서 그는 인간적 특성을 갖는다. 그러나 그의 탄생이 갖는 인공성은 자연스러운 인간 탄생과는 다르며, 이식과 합성의 결과물인 그의 외양은 보는 이에게 혐오감과 두려움을 불러일으킨다. 프랑켄슈타인이 자신의 창조물에 생명을 부여하고 그 결과물을 대면했을 때 보여준 첫 번째 반응에 주목해 보자.

이 재앙에 대한 나의 감정을 어떻게 표현할 수 있을 것인가? 무한한 수고와 주의를 쏟아 부어 만들어내고자 했던 그 빌어먹을 것을 어떻게 표현할 것인가? 그의 사지는 균형이 맞았고, 나는 그의 외양을 아름다운 것으로 선택했다. 아름다움! — 위대한 신이시여! 그의 누런 피부는 그 아래에 놓인 근육과 혈관들을 차마 다 덮지도 못하였다. 그의 머리카락은 윤기 나는 검정색칼을 하고 치렁치렁했다. 그의 치아는 진주처럼 하얀색이었다. 그러나 이 호화로운 요소들은 그의 물기 있는 눈들과 더욱 끔찍한 대비를 만들어 낼 뿐이었다. 그의 눈은 쭈글쭈글한 가죽과 검은 일자 입술 안에 놓인 암갈색의 흰 구덩이처럼 보였다<sup>1)</sup>(Shelly, 2017: 41-42).

자신의 창조물이 눈을 뜨고 자신을 바라보는 생명창조의 중요한 순간을 그는 “재앙”으로 규정하고 있다. 그가 가졌던 애초의 열망과 비교해 볼 때 이러한 반응은 예측을 벗어나는 것이다. 흥미로운 점은 이 부분이 성경의 창세기에 기록된 창조의 순간에 절대자가 자신의 피조물을 대하며 보여주는 태도와 대비를 이룬다는 점이다. “God saw all that he had made, and it was very good”(신은 자신이 만든 모든 것을 보았으니, 그것은 보기에 좋았도다)(Genesis 1: 32). 이러한 태도는 하라리 테제가 보여주는 태도와 차이를 보이면서, 그 의미를 다시 곱씹게 한다. 신을 창조하면서 역사가 시작되었다는 것은 피조물로서의 자신의 존재가 시작되었음을 의미한다. 그 피조물을 대하는 성경 속 창조자의 태도는 보다시피 긍정

1) 본문에서 인용한 부분의 영어 원문은 다음과 같다.

“How can I describe my emotions at this catastrophe, or how delineate the wretch whom with such infinite pains and care I had endeavoured to form? His limbs were in proportion, and I had selected his features as beautiful. Beautiful! — Great God! His yellow skin scarcely covered the work of muscles and arteries beneath; his hair was of a lustrous black, and flowing; his teeth of a pearly whiteness; but these luxuriances only formed a more horrid contrast with his watery eyes, that seemed almost of the same colour as the dun white sockets in which they were set, his shrivelled complexion, and straight black lips.”(본문의 우리말 번역은 필자의 것임)

적인 것으로 표현되어 있다. 이는 자기 존재의 기원에 대해 인간이 기대하는 긍정적인 책임의 윤리와 다를 바 아니다. 하라리 명제와 비교해볼 때, 우리는 이 부분에서 창조와 그것에 관련된 책임의 윤리에 대해 우리가 어떤 기대를 갖는지 추출해 낼 수 있다. 그 기대와 비교할 때 자신의 모든 것을 쏟아 부은 창조물을 대면하자마자 그것을 거부하는 프랑켄슈타인의 태도에서 우리는 윤리적 책임의 가능성을 찾아보기 힘들다고 할 것이다. 오히려 책임지지 않으려는 무책임한 모습까지도 엿보게 된다.

이후 창조자로 인해 괴물로 낙인찍힌 생명체는 자신의 정체성을 추구하며 자신을 만들어 놓고는 책임을 지지 않으려는 프랑켄슈타인을 벌주기 위해 그의 가족과 주변 인물들을 하나씩 둘씩 해하며 그의 삶을 파괴하게 된다. 이러한 상황은 프랑켄슈타인이 미처 예상치 못한 것이다. 괴물은 그에게 공존의 책임을 요구하는 것이다. 그가 자신의 창조물에 대해 취하는 태도의 핵심인 이 공존의 책임이야말로 이 소설의 기저에 흐르는 주제 중 하나라고 할 것이다. 주인공이자 창조자인 프랑켄슈타인은 공존에 관련한 적어도 두 가지 종류의 책임에 직면한다고 볼 수 있다. 첫째는 자신의 창조물이 야기한 파괴적 상황에 대한 책임, 둘째 자신이 창조한 괴물에 대한 책임이라고 할 것이다. 여기서 우리는 “우리의 기술은 ... 스스로 무엇을 원하는지도 모르는 채 불만스러워하는 무책임한 신들, 이보다 더 위험한 존재가 또 있을까?”(하라리, 2015: 588)라는 또 다른 언급을 떠올리게 된다. 무책임한 신은 창조자로서의 자신과 피조물 그리고 세상을 모두 파괴시킬 수 있는 것이다. 우리는 무책임한 신이 될 것인가?

### Ⅲ. 자기책임의 윤리: 『안드로이드는 전기양의 꿈을 꾸는가?』

프랑켄슈타인과 괴물에 관련된 책임감의 문제는 비단 유기물의 형태를 취한 창조물에 대해서만 유효하게 적용되는 것은 아니다. 인간이 만들어낸 모든 것들에 대한 책임감 혹은 공존의 관계 맺음의 문제는 인간과 기계의 문제로 확대된다. 아직까지 인간이 만들어낼 수 있는 기계 중 최고의 기술 집약체는 인조인간 혹은 인공지능 로봇이라 할 수 있다. 공상과학 소설과 미래 영화 등은 인조인간과 로봇들이 가득한 가상의 세계를 보여준다.

그렇다면, SF 소설이나 미래 영화의 화면을 가득 채우고 있는 인간과 구분이 거의 되지 않으면서 인류에게 파국적 결말을 초래하는 ‘사악한’ 인공지능이나 로

봇에 대한 상상은 어디서 오는 것인가? 제임슨(Fredric Jameson)이 자신의 SF론 모음집 『미래의 고고학: 유토피아라 불리는 욕망 그리고 다른 SF 소설들』(*Archeologies of Future: the desire called utopia and other science fictions*)에서 지적한 바 있듯이, 인간을 닮은 사악한 인공지능에 대한 상상과 그것이 불러올 파국적 결말에 대한 두려움은 “our incapacity to imagine the future”(미래를 상상하지 못하는 우리의 무능함)(Jameson, 2007: 188-89)에서 기인하는지도 모르겠다.

제임슨이 지적한 무능함이란 프랑켄슈타인이 보여주었던 상상의 무능 및 공존에 대한 무책임과 다를 바 없다. 그는 자신의 창조물 속에서 무엇을 대면할 것인지를 상상하지 못했다. 우드슨(Stephani Etheridge Woodson)의 설명에 기대자면, 아름다움(beauty)과 선함(goodness) 그리고 지각(perception)에 대한 탐색이기도 한 소설 『프랑켄슈타인』에서 자신의 창조물에 대한 프랑켄슈타인의 부정적 정의는 피조물 자체의 정체성 보다는 인간 자신의 정체성에 의존한다고 할 것이다(Shelly, 2017: 41).

프랑켄슈타인은 창조물 속에서 자아로서의 자신의 모습을 기대하였으나 정작 그가 대면한 것은 자기 안의 타자의 모습이었다. 그는 자신과 동근원적 존재로서의 창조물을 기대했으나 정작 마주친 것은 자신 안의 타자인 추한 괴물의 모습이었던 것이다. 그가 대면한 괴물은 자신의 욕망과 꿈을 모두 쏟아 부어 투사한 존재로서 결국 자기 자신에 다름 아닌 것이다. 여기서 우리는 피조물과의 공존을 위한 세 번째의 책임을 접하게 된다. 그것은 자신 안에 깃들인 동근원적 존재로서의 타자에 대한 책임이라고 할 수 있다.

이야기를 현실 속 인공지능에 적용시켜 보자. 미래에 등장할 인공지능 로봇의 모습은 인간을 닮아 있으며, 인간을 위협하여 파국적 결말을 가져오는 무시무시한 존재로 가정된다. 프랑켄슈타인이 대면했던 인조 생명체와 다를 바 없다. 그러한 가정이 가능한 것은 타자를 배척하는 지극히 인간적인 선악관, 사회관 등을 그대로 투영한 결과라 할 것이다. 인공지능을 타자로 만들고 인간을 우월한 위치에 두고 있다는 점에서는 기존의 인간중심적 휴머니즘 사상을 그대로 반영하는 가정이다. 그러한 인간중심적 휴머니즘은 트랜스휴머니즘이 표방하는 로봇 관에 기초한 서사구조를 대변한다고도 할 수 있다.

커즈와일(Ray Kurzweil)이 주창하듯 기술 발달에 힘입어 인류를 인위적으로 변형시키고 인류의 약점을 극복할 수 있다는 트랜스휴머니즘의 전망은 기술의 기하급수적인 발전 속도에 힘입어 공상 과학 소설이나 영화와 같은 허구적 서사에서 현실적 사실로 이행되고 있다. 혹은 그러한 이행이 당연한 것으로 간주되는 것이

이즈음의 경향이다. 1960년대 말에 출간된 덕의 소설 『안드로이드는 전기양의 꿈을 꾸는가?』는 기술발달 및 인공지능이 가져올 미래사회를 그리고 있다. 장밋빛 전망과는 거리가 먼 아포칼립스적 상황을 그려내고 있는 이 작품에서, 인간과 인공지능을 탑재한 인조인간 안드로이드의 관계 및 공존의 가능성은 이야기의 중심에 있다.

이 소설의 작가 필립 K. 덕에 대해서는 많은 수식어가 따라다닌다. 가장 대표적인 것으로, “현재가 가진 불안의 요소들을 가장 잘 다루고 있는 대가들 중 한 명”(one of the masters of present day discontents)이라는 앨디스(Brian Aldiss)의 것을 인용할 수 있겠다(Dick, 1996: vii 재인용). 이 소설은 기술발달의 결과를 제어하거나 통제하지 못하는 인간의 불안과 공포에 대한 은유로 읽히는 것이다.

소설의 무대는 핵전쟁의 결과 폐허로 변해버린 미래의 지구이다. 유전자 체계에 이상이 생기고 그 결과 지능에 문제가 발생한 여러 층위로 인간이 분류된다. 물질적 여건을 갖춘 정상적인 지능의 인간들은 대부분 우주의 식민 행성으로 이주를 하고 소위 “닭대가리”(chicken head)로 지칭되는, 지능에 문제가 생긴 인간들과 소수의 정상인들만이 낙진과 폐허 속에 남아있다. 식민 행성에서 인간의 편의를 돕기 위한 도구적 존재로 만들어진 인공지능 탑재 안드로이드들이 그들의 주인을 살해(kill)하고 지구로 잠입하였는데, 이들을 추적하여 회수(retire)하는 현상금 사냥꾼(bounty hunter) 데커드(Rick Deckard)가 하루 동안 겪는 사건들이 주요 줄거리이다. 인간을 위해 봉사하는 도구로서의 로봇이라는 트랜스휴머니즘 로봇관이 갖는 문제와 인간의 안위에 도움이 되지 않는 로봇을 어떻게 처리해야 할지의 문제가 첨예하게 드러난다는 점에서 이 소설은 인간중심적 사고를 비판하는 윤리적 측면을 분명히 보여준다.

인공지능을 둘러싼 윤리에 대한 논의는 이미 우리 주변에 존재한다. 화학자이자 공상과학 소설가인 아시모프(Issac Asimov)가 정의한 로봇학 3원칙이 대표적이다. 구체적인 내용을 보면 다음과 같다. 첫째, 로봇은 인간에 해를 가하거나, 혹은 행동을 하지 않음으로써 인간에게 해가 가도록 해서는 안 된다. 둘째, 로봇은 인간이 내리는 명령들에 복종해야만 하며, 단 이러한 명령들이 첫 번째 법칙에 위배될 때에는 예외로 한다. 셋째, 로봇은 자신의 존재를 보호해야만 하며, 단 그러한 보호가 첫 번째와 두 번째 법칙에 위배될 때에는 예외로 한다(변순용, 2018: 238 재인용). 조심스럽기는 하지만 아시모프의 원칙은 트랜스휴머니즘이 표방하는 로봇 관을 반영하는 윤리적 원칙이라고 할 수 있다. 그런 의미에서 지극히 인간중심적 윤리라고 할 것이다.



이 소설에 등장하는 탈주 안드로이드들이 처음부터 인간을 파괴하려는 의도를 가진 것은 아닌 것 같다는 점에서, 이들은 아시모프의 로봇 원칙에 충실한 존재들이다. 또한 애초에 만들어진 의도대로 임무를 충실히 이행한다는 점에서 수행성이나 기능성에 있어 이상이 없는 존재들이다. 이런 점에서 이들은 트랜스휴머니즘의 로봇 관 너머로 탈주하는 존재들이라고 할 것이다. 이들이 탈주한 원인은 스스로의 정체성에 의문을 품고 존재로서의 자신에 대한 책임을 다하기 위한 동기로부터 비롯된다. 이러한 점에서 이들은 지극히 인간을 닮은 존재들이다. 생물학적으로도 이 안드로이드들은 전기 충전기와 같은 에너지 충전 없이 스스로 음식을 섭취하고 배설하는 생물학적 자기생성을 실행하는 자율적인 체계 혹은 “자기생성”(autopoiesis)의 존재들이다.

자기생성은 생명을 규정하는 가장 기초적인 설명으로 그 전형적인 예는 살아있는 세포다. 이 경우 세포를 이루는 과정들은 화학적인 자기조직과 자기폐쇄의 반복이다. 유기체는 스스로를 생성하기 위해 막을 형성하여 외부 환경에 대해 자신을 폐쇄하면서 동시에 스스로를 조직해 나가는 것이다. 이 자기폐쇄와 자기조직 과정들의 반복적인 상호의존은 그 자체의 막을 또한 생산하면서, 자기 생산하는 신진대사적 네트워크의 형태를 취한다. 이 네트워크는 생화학적인 영역에서 체계를 통일체로 구성하고, 환경과 가능한 상호작용의 영역을 결정하며 자기생성으로 알려져 있는 것이다(Thompson, 2005: 101).

툼슨의 설명을 따른다면, 이 소설의 안드로이드들은 자기 생성적 존재로 유기체적 생명체이며, 게다가 겉으로 보기에 인간과의 구별이 어려운 외양을 하고 있다. 그러나 진짜 인간이라는 것이 너무나 중요한 의미를 갖는 이 텍스트에서는 이들은 정당한 생명체로 인정받지 못한다. 단적으로 누군가 그들을 없애더라도 그것은 “살해”(kill)가 아니라 “회수”(retire)하는 것으로 표현됨에 유의하자.

포스트휴먼 이론가 헤일즈(Katherine Hayles)는 이러한 지점을 포착하고 있다. 그는 이 소설의 작가 필립 K. 딕에 대해 “Working along apparently independent line of thought, Dick understood that how boundaries are constituted would be a central issue in deciding what counts as living in the late twentieth century”(딕은 분명히 독립적인 사상을 가지고 있었고, 20세기 후반에 생명체의 정의를 결정할 때 경계를 구성하는 것이 중심문제가 되리라는 사실을 알았다)라고 평가한다(Hayles, 1999: 161). 또한, 이 소설에서 다루어지는 “안드로이드는 사회적 존재인 것처럼 잘못 취급되는 물건이 아니라 물건인 것처럼 잘못 취급되는 사회적 존재”(Hayles, 1999: 162)라는 포스트휴머니즘 견해를 내놓는다. 이 소설의 안드로이드들은 진짜 인간과 가짜 인간의 경계 선상에서 끊임

없이 밖으로 내몰리는 존재들이라는 것이다.

실제로 경험하지 않고서도 사고에 의해 지각하고 설명하는 견해를 우리는 “사변적 윤리”라고 부른다. 문학적 상상력의 공간은 실제로 존재하지 않는 가상의 세계를 통해 사변적 윤리의 가능성을 제시한다. 이렇게 본다면, 문학과 윤리는 상호보완의 관계를 맺는다고 할 수 있다. 문학은 직접적으로 언급하고 있지는 않지만, 상상력의 공간을 통해 윤리가 가야하는 방향을 은유와 허구적 서사로 제시하고 있는 것이다.

자기 생성적 존재, 사회적 존재로서 레이첼 로즌 및 그녀와 동일한 인공지능 장치 Nexus-6을 장착한 도주 안드로이드들은 인간과 거의 구분되지 않으며 사회속에서 저마다 생산 목적에 맞는 기능을 수행하고 있다. 전문적인 현상금 사냥꾼 테커드가 자신의 검사 대상인 레이첼 로즌이라는 이름의 여성 혹은 “그것”이 안드로이드란 사실을 처음에는 알아차리지 못함에 주목하자. 다음은 테커드가 레이첼 로즌을 대상으로 안드로이드 여부를 알아내는 검사를 실시하는 부분이다.

“제 서류가방 말인데요.” 릭은 보이드 캠프 질문지를 찾으려고 가방을 뒤지며 말했다.

“멋지죠. 안 그래요? 경찰서에서 지급한 거죠.”

“그래요. 그러네요.” 레이첼이 심드렁하게 말했다.

“갓난아기 가죽으로 만든 거예요.” 릭이 말했다. 그는 서류가방의 검은 가죽 표면을 손으로 쓰다듬었다. “1백 퍼센트 진품 갓난아기 가죽으로요.” 그는 두 개의 눈금판 표시기에서 바늘이 미친 듯이 빙빙 도는 것을 보았다. 하지만 그 전에 잠시 머뭇거림이 있었다. 반응이 나오기는 나왔지만 너무 늦게 나온 것이다. 그는 1초도 되지 않는 반응 시간까지도 속속들이 알고 있었다. 즉 정확한 반응 시간을 알고 있었던 것이다. 원래는 이런 지연이 없어야만 했다<sup>2)</sup>(Dick, 1996: 58-59).

어린 시절의 기억 등을 포함한 인공지능을 탑재하고 있어서 스스로를 진짜 인간으로 확신하는 안드로이드와 인간을 구분하는 방법으로 이 작품이 제시하는 방법은 감정이입 검사뿐이다. 이 부분에 제시된 Voigt-Kampff 테스트라는 명칭은 감정이입 가능성 유무를 묻는 검사법으로 튜링 테스트에 그 기초를 둔다(Persson,

2) “My briefcase,” Rick said as he rummaged for the Voigt-Kampff forms. “Nice, isn’t it? department issue.”

“Well, well,” Rachael said remotely.

“Babyhide,” Rick said. He stroked the black leather surface of the briefcase. “One hundred percent genuine human babyhide.” He saw the two dial indicators gyrate frantically. But only after a pause. The reaction had come, but too late. He knew the reaction period down to a fraction of a second, the correct reaction period: there should have been none.(본문의 우리말 번역은 필자의 것임)

2018).

소설의 서사에 따르면, 수정란의 형태로부터 계통발생과 개체발생을 반복하여 인간으로 탄생한 생물학적 자연 인간과 달리 수정란의 접합자 단계에서 DNS(Digital Nervous System)이라는 일종의 디지털 신경체계 인자를 이식하여 만들어낸 안드로이드는 공포나 혐오와 같은 본원적 감정이 없다. 다만 이식된 기억만이 존재할 뿐이다. 안드로이드에게 인공지능의 이식은 가능하지만, 감정을 주입하는 것은 매우 복잡하고 정교한 기술력을 필요로 하는 것이며 아직 실현되지 않은 것으로 텍스트는 기술한다(Dick, 1996: 61-69). 이들에게 본원적 감정과 정서는 경험에 의한 것이 아니라 사실상 학습된 두뇌 작용이므로, 자극과 반응 사이에 저장된 기억을 소환하는데 필요한 약간의 시간 차이가 발생한다. 소설 속의 감정검사는 그 시간 차이를 감지하여 인간과 안드로이드를 구분하고 있다.

여기서 주목할 점은 이 작품에 등장하는 안드로이드들이 프랑켄슈타인의 괴물만큼이나 인간과 동근원적인 존재들이라는 것이다. 접합자라는 명칭을 쓰고 있지만 이들의 기원은 여전히 인간의 수정란이라는 점에서 인간과 동일한 유전자를 보유한다. 안드로이드 혹은 전자 동물에 대비하여, 진짜 인간 그리고 진짜 생명을 가진 동물에 대해서만 존재론적 의미를 부여하는 이 작품의 내러티브에서 정작 진짜 인간의 특징이 무엇인지는 매우 모호하게 그려진다. 안드로이드를 구별하는 표식으로 감정이입의 부재를 들고 있지만, 인간 역시 자신의 감정 조절을 위해 인공물인 “기분 조절 오르간(mood organ)”에 의존한다는 점에서 감정에 관한 한 취약하다. 인간의 감정만이 갖는 본질적 능력으로 제시된 감정이입 혹은 공감(empathy)조차도 “머서교(Mercerism)”라는 종교 형태의 사회적 감정이입 기계에 접속함으로써 가능하니 말이다.

이러한 점에서 인간과 안드로이드로 구성된 이 소설의 등장인물들이 보여주는 모습은 흥미롭다. 인간이면서도 기계에 의존하지 않고는 자신의 감정을 조절하거나 통제하지 못하는 정상적인 인간들이 있다. 반면 정상적인 인간으로 분류되지 못하지만 고도의 감정이입을 보여주는 닭대가리라 불리는 저능인간도 있다. 감정이 부재하는 사이코패스 안드로이드가 있는 반면, 감정의 가능성을 보여주는 존재도 있고, 인간보다 더 인간적인 감정을 보여주는 안드로이드도 등장한다. 감정 반응 속도를 가지고 인간과 안드로이드를 구별하려는 시도는 적어도 이 소설의 서사에 따르면 불합리해 보인다(Hayles, 1999: 163-172).

결국, 소설에 등장하는 안드로이드들은 인간과 동일한 기원으로부터 만들어진 인간의 욕망이 투사된 존재라고 볼 수 있다. 이들은 인간과 비인간 사이에 존재하는 경계의 “면도날”(Blade) 위에서 끊임없이 경계선 밖으로 내몰리는 존재들이다.

그 경계로부터 인간 또한 자유로워 보이지는 않는다. 인간과 비인간의 경계에 선 존재는 안드로이드뿐 아니라 인간도 마찬가지이며, 이들의 공존 가능성을 묻는 문제는 또 다른 자아로서의 자신, 타자로서의 자신에 대한 책임을 묻는 문제에 다름 아니다.

#### IV. 인간과 인공지능의 공존의 가능성

인간과 인공지능의 공존을 문제 삼는데 있어 전제되어야 할 인간의 속성 가운데 하나는 무엇일까? 쿨(Diana Coole)과 프로스트(Samantha Frost)는 그것을 신체(corporeality) 혹은 몸(body)에 대한 강조로 제시한다. 그들은 편역서 『신사물론: 온톨로지, 행위자, 그리고 정치』(*New Materialisms: Ontology, Agency, and Politics*)에서 신체(corporeality) 혹은 몸(body)이라고 하는 것을 주변의 자연환경, 다른 신체들, 그리고 사회경제적 요소들이 복합적으로 구성하는 하나의 구조물로 정의한다(Coole and Frost, 2010: 20). 살아있는 신체는 몸짓을 통해 다른 신체들과 의사소통하고 본능적 반응들을 초래하는 등의 기능을 수행하며, 이 기능을 수행함에 있어서 의식적 자각을 반드시 통할 필요가 없는 형태의 판단들이 관여한다는 것이 그들의 생각이다.

신체 혹은 몸에 대한 강조가 비단 이들만의 것은 아니다. 인간의 육체를 강조하는 견해의 역사는 생각보다 길다. 다수의 페미니스트들이 이미 남성과 여성의 불평등에 대한 기원으로 몸이 가진 차이를 지적한 바 있다. 길리언 로즈(Gillian Rose)는 “몸이란 자연적이지 않고 권력과 정체성의 지형도, 또는 권력과 정체성의 관계에 대한 지형도이다”라 설명한 바 있다(Rose, 1998: 175). 남성 백인 남성이라는 생물학적 신체, 사물성, 혹은 물질성을 중심에 두고 연령, 인종, 성별, 종별을 두고 전개되어 왔던 오랜 권력 투쟁의 과정을 생각해 보더라도 이 점은 분명하다.

신체성을 강조하는 것은, 신사물론자들(new materialist) 혹은 포스트휴머니스트들이 주창하듯이, “인간중심적인 휴머니즘 기계”(the anthropological machine of humanism)에 대한 거부와 맥을 같이한다. 이들은 인간중심주의를 거부하는데 그치지 않고 인간 신체에 대한 특권 또한 인정하지 않는다. 여성에 대한 남성의 신체, 흑인에 대한 황인종에 대한 백인의 신체, 미성년에 대한 성인의 신체에 대한 특권이 폐기되었다. 이와 같이 동물들의 신체 혹은 인공지능을 탑재한 인공적인 구조물들의 신체를 포함하여 모든 신체들이 행위자(agent)의 특정한 역량을 보여

준다는 의견을 주장하는 이들의 목소리가 점점 커지고 있다.

그 결과, 인간을 자연과 구별하기 위해 사용하였던 인류라는 종의 개념 자체, 자기-성찰, 자기-인식과 같이 합리성을 포함하여 인간만의 속성으로 간주되던 특성들이 진화론 혹은 우주론이라는 더 넓은 범위의 체계 안에서는 우발적이거나 일시적은 형태 혹은 과정에 불과한 것으로 간주되고 있다. 그리하여 인간성의 완성이란 인간성의 구원과 같은 용어들은 더 이상 역사의 필연적 운명으로 이해되지 않으며, 진화의 목표라는 인간중심적 개념 또한 힘을 잃어 가고 있다(Coole and Frost, 2010: 22). 이제 인간만의 전유물이라고 간주되던 자기-성찰과 사고 혹은 인지적 역량들이 더 이상 그 가치를 인정받지 않게 된다. 대신 이러한 역량이 인간이외에도 다양한 형태의 생명체들과 존재들 혹은 사물들에게서 서로 다른 정도로 나타날 수 있는 보편적인 역량들로 간주되기 시작한다.

포스트휴머니즘 논의 혹은 신사물론 논의는 이런 식으로 전통적인 윤리적 범주들을 시대에 뒤진 것으로 만들고 있다. 예컨대, 과학자들이 동물과 인간의 생명을 인공적으로 창조하거나 확장하고 있고, 유전자를 조작하거나 합성하여 새로운 생명 형태를 창조하며 인위적으로 새로운 종들을 만들어 내면서, 우리의 기존 윤리와 정치적 사고의 기반이 되는 개념들과 경계들이 교란되고 있다. 예컨대, 인공지능을 탑재한 형태의 일종의 인조 혹은 합성 생명체들은 개인이라고 하는 우리의 개념에 도전장을 던진다. 지적 생명체와 그렇지 않은 생명체 사이의 차이들은 인간을 다른 동물들과 구분하고 인간이 모든 물질적 자원을 인간을 위한 도구로 전용하는데 정당성을 부여하는 중심 개념이었기 때문이다.

이러한 논의로부터 우리는 다음과 같은 사변적 윤리를 도출하게 된다. 인공지능 기술 혹은 과학의 발달로 인해 물질로부터 생명체를 창조해 낼 역량을 우리가 충분히 갖추게 된다면, 그리고 그러한 생명체의 형태들이 특정한 임무를 수행할 수 있는 지적인 형태의 행위자의 모습을 갖추게 된다면, 기존에 인간과 다른 존재들을 구분하는데 필수적인 차이들로 간주되던 것들은 그 효력을 잃게 될 뿐 아니라, 그러한 차이들로부터 도출한 규범 혹은 기준들(norms)의 확실성 또한 모호해지게 된다는 것이다. 이러한 문제는 생명 형태들 자체에 관련된 문제들을 제기한다.

필자는 문학이야말로 그러한 문제들을 제기하고, 인공 혹은 인조 생명체들에 대한 새로운 윤리적 가치와 기준을 탐색할 수 있는 가능성을 제공한다고 본다. 문학은 인간을 이해하려는 노력의 총체이자 인간의 미래를 예측하려는 노력으로 가득한 상상력의 실체이자 산물이다. 다양한 문학작품을 통해 인간은 자신들의 미래를 예측한다. 물론 그 예측이 언제나 그대로 실현될 가능성은 희박하지만, 적어도 현재의 인간들이 가진 문제와 고민을 보여주곤 한다. 다시 말해 과학소설은 미래

를 예측하는 것 이상의 무엇인가를 수행한다는 점에서, 현재와 미래에 영향을 미치는 윤리적 책임의 문제를 도출해 낸다고 볼 수 있다.

우리가 읽는 과학소설은 대중의 상상력을 자극하며 그들의 기억과 더불어 반향을 불러일으킨다. 많은 과학소설들이 출판과 더불어 곧 잊히기도 하지만, 유난히 오래 기억되는 것들이 있다. 소설 『안드로이드는 전기양의 꿈을 꾸는가?』는 그 대표적인 경우라고 할 것이다. 이 이야기가 대중의 상상력을 사로잡는다는 사실과 여기에서 예측하는 미래가 실현되느냐의 여부 사이에 정확한 대응관계는 없다고 할 것이다. 미래를 예측한다고 해서 그것이 반드시 실현되지는 않기 때문이다. 또한 미래 예측의 현실화 여부가 이 소설을 읽는 독자의 관건도 아니다. 미래에 대한 예측과 그것의 실현 여부 보다는 현재 우리가 무엇을 생각하고 있는가에 대해서 고민하게 하는 것이 이 소설에서 더 중요한 관건일 수도 있기 때문이다. 인간과 인공지능의 공존 가능성에 대한 탐색도 그 하나라고 할 것이다.

소설 속 서사의 과정에서 인간이 비인간으로 판정된 안드로이드들을 가차 없이 살해하는 과정, 레이저 건으로 회수(retire)하는 과정에서 이 가능성은 모습을 드러낸다. 화성 식민지에서 자유를 얻기 위해 인간을 살해하고 탈주한 안드로이드들을 회수하는 목적이 인간의 생명을 보호하기 위한 것이라고 말하는 목소리가 서사의 한 편에 제시된다. 그러나 그러한 목적은 더 이상 인간에게 유용한 도구로만, 한낱 기계로만 취급되고 소모되기를 거부한 안드로이드들을 제거하는 것에 대한 웅색한 정당화에 불과하다고 말하는 또 다른 목소리가 있고, 그 목소리에 소설의 서사는 더 큰 힘을 실어준다. 인간의 기준을 정당화하려는 이 과정에서 역설적으로 발생하는 역효과 혹은 부작용은 안드로이드를 부정하고 폭력을 행사하는 과정에서 정작 인간성의 일부가 떨어져나가는 것임(박인찬, 2015: 137)을 극명하게 보여주기 때문이다. 결국 인간과 비인간의 경계, 누가 더 인간다운가의 경계는 서사가 진행될수록 모호해지고 만다.

이 소설에는 인간과 인공지능의 공존 가능성을 제시하는 인물이 제시된다. 그는 안드로이드를 회수하고 머리수대로 돈을 받는 현상금 사냥꾼 테커드의 대척점에 해당하는 위치를 점하는 인물 이지도어(Isidore)이다. 이지도어는 인간을 분류하는 기준에 의하면 “답대거리”로 분류되는 소위 저능한 인간이다. 그러나 흥미로운 것은 그가 살고 있는 아파트는 소설 내내 끊임없이 묘사되는 것처럼, 모든 것이 떨어지며 모든 것을 죽음으로 몰고 가는 엔트로피 과정, 혹은 폐허화의 과정, 이른바 “키플화”의 과정이 멈추는 지점이라는 것이다. 그나마 인간이 인간답게 사는 것은 공간, 인간이건 안드로이드이건 서로를 이해하는 공간을 점하고 있는 인물 이지도어를 통해 공존의 가능성이 제시되는 것이다.

이 소설의 서사가 갖는 윤리적 가능성의 정점은 주인공 데커드가 이지도어가 가진 가능성을 공유하는 인물로 전환되는 지점이다. 소설의 결말부분에 이르러 사냥꾼 데커드는 모든 탈주 안드로이드들을 회수하고 난 후 오레곤의 황무지로 날아가게 된다. 이 부분에서 그는 인간 이외의 다른 존재, 특히 인공 생명체에 대해서도 윤리적 책임을 가진 주체로서의 전환 가능성을 보여준다.

“머서가 그 닭대가리, 이지도어한테 준 거미. 어쩌면 그것도 인공품일지 몰라. 하지만 상관없어. 전기 제품도 제 나름의 생명을 갖고 있으니까. 그런 생명이야 하찮기는 해도 말이야”<sup>3)</sup>(Dick, 1996: 241).

오레곤의 황무지로부터 집으로 돌아오는 길에 데커드는 두꺼비 한 마리를 발견한다. 그는 그것이 진짜 생명이라고 생각했고 집으로 가져간다. 후에 그것이 전기 두꺼비로 밝혀지지만 그는 더 이상 그것을 신경 쓰지 않게 된다. 왜냐하면 그는 인간뿐만 아니라 전기형태의 생명들도 나름대로의 삶을 살아갈 수 있다는 것을 안드로이드에 대한 그간의 경험을 통해 깨달았기 때문이다.

다소간 모호하게 처리되었기에 해석의 불분명함이나 논란의 여지는 있을 수 있겠지만 필자는 이 장면으로부터 인간과 인공지능의 공존 가능성을 본다. 그리고 타자에 의한 윤리 주체 형성이라는 주디스 버틀러의 개념(버틀러, 2013: 38-39)을 떠올린다. 이 개념에 따르면, 복수와 처벌과 같은 폭력에 바탕을 둔 규범의 가치는 결코 윤리적일 수 없다. 윤리나 정의를 복수나 처벌에 의해 형성되는 것으로 설명하려는 시도는 바람직하지 않으며, 오히려 사회적 도덕적 폭력성의 일부일 뿐이기 때문이다. 경험을 통해 타자에게서 자기책임의 윤리를 수용하는 주체의 생성이야말로 문학이 상상력의 공간을 통해 제시하는 윤리적 태도라고 할 수 있을 것이다.

## V. 글을 마치며

이 논문은 인간과 인공지능의 공존에 대한 문제로 창조자로서 인간이 자신의 피조물에 대해 갖는 책임의 윤리로 제시하였다. 피조물 자체에 대한 책임, 피조물

3) “The spider Mercer gave the chickenhead, Isidore; it probably was artificial, too. But it doesn’t matter. The electric things have their lives, too. Paltry as those lives are.”(본문의 우리말 번역은 필자의 것임)

로 인해 파괴되거나 혼란이 야기된 세상에 대한 책임, 그리고 피조물 안에 깃든 타자로서의 자신에 대한 책임이 그러한 윤리의 구체적 형태들이다. 이러한 책임의 형태들은 인공지능을 상정함에 있어서 우리가 가져야할 자세를 제시한다.

인공지능에 대한 책임감의 문제는 또한 인간이 인간에 대해 갖는 자세와 책임의 문제로 연결된다. 우리는 우리가 가지는 문제를 다른 인간 혹은 우리와 다른 타자에게 투사한다. 우리 안의 일부를 도려내어 그것을 타자에 투사함으로써 스스로에 대한 책임감으로부터 도망가려는 것일지도 모른다. 인간과 인공지능의 공존 가능성의 문제는 어쩌면 인간과 인간의 공존 가능성의 문제를 표현하는 다른 방식일지도 모르겠다.

타자로서의 인공지능 안에서 우리 자신이 투사된 모습을 발견하고 인정할 때 우리는 인간과 인공지능의 공존 가능성에 대한 단초를 발견할 수 있지 않을까? 지젝이 말하는 바와 같이, “장애로 보이는 것은 가능성의 긍정적 조건이다. 그리하여 정신분석에서 일어나는 것은 증상의 해소가 아니라 불가능성의 조건을 가능성의 조건으로 뒤집는 관점의 변화이다”(지젝, 2010: 279). 우리는 인공지능과의 불화라는 부끄러운 증상 앞에서 새로운 가능성의 지점들을 탐색해야만 한다.

필자는 그 출발점을 공상과학 소설이라는 보다 대중적인 형태의 상상적인 예술 노동에서 찾아보려 하였다. 안토니오 네그리는 예술 노동을 인간 존재의 초과를 생산해내는 창조력으로 보면서 예술은 본질적으로 집단적이며, 그것이 집단적일 때 비로소 해방적인 것이 될 수 있다고 주장한 바 있다. “예술가 안에서 집단적인 것은 존재의 초과를 비로소 해방시키고 존재의 초과를 특이한 것으로 만듭니다”(네그리, 2010: 115). 문학이라는 상상력의 공간, 그 중에서도 공상과학 소설이라고 하는 매우 대중적인 예술의 공간이야말로 인공지능과의 공존의 문제라는 집단 윤리의 문제를 탐색하기에 적합한 장르가 아닌가라는 제안을 하며 글을 마친다.

□ 투고(접수)일 : 2019년 2월 20일 / 심사(수정)일 3월 21일 / 게재확정일 3월 27일



## 참 고 문 헌

- 김건우(2016), “포스트휴먼의 개념적, 규범학적 의의”, 『포스트휴먼시대의 휴먼』, 서울: 아카넷.
- 네그리, 안토니오(2010), 『예술과 다중』, 심세광 옮김, 서울: 갈무리.
- 딕, 필립 K.(2017), 『안드로이드는 전기양의 꿈을 꾸는가?』, 박중서 옮김, 서울: 폴라북스.
- 박인찬(2015), “필립 딕의 키치 세상: 『안드로이드는 전기 양을 꿈꾸는가?』를 통해 본 인간, 기계, 그리고 과학소설”, 『현대영미소설』, 22(3).
- 버틀러, 주디스(2013), 『윤리적 폭력비판: 자기 자신을 사랑하기』, 양효실 옮김, 서울: 인간사랑
- 변순용(2018), “인공지능로봇을 위한 윤리 가이드라인 연구-인공지능로봇윤리의 4원칙을 중심으로”, 『윤리교육연구』, 47.
- 지젝, 슬라보예(2010), “이웃들과 괴물들”, 『이웃』, 정혁현 옮김, 서울: 도서출판b.
- 하라리, 유탄(2015), 『사피엔스』, 조현욱 옮김, 서울: 김영사.
- Coole, Diana & Samantha Frost ed.(2010), *New Materialisms: Ontology, Agency, and Politics*, New York: Duke University Press.
- Dick, Philip K.(1996), *Do Androids Dream of Electric Sheep?*, New York: Ballantine Books.
- Harman, Graham(2002), *Tool-Being: Heidegger and the Metaphysics of Objects*, New York: Open Court.
- Hayles, N. Katherine(1999), *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, Chicago: Chicago University Press.
- Jameson, Fredric(2007), *Archaeologies of the future: the Desire Called Utopia and Other Science Fiction*, New York: Verso.
- Perkowitz, Sidney(2005), “Digital People: From Bionic Humans to Androids”, *Technology and culture*, 46(2).
- Persson, Hans(2018), “Do Androids Dream of Being Human?”,

[http://www.lysator.liu.se/lisff/mb-nr27/Do\\_Androids\\_Dream\\_of\\_Being\\_Human.html](http://www.lysator.liu.se/lisff/mb-nr27/Do_Androids_Dream_of_Being_Human.html) (검색일: 2018. 09. 20)

Rose, Gillian(1998), "Feminism and Geography", *Dangerous Women: Gender and Korean Nationalism*, Ed. Elaine H. Kim and Chungmoo Choi, New York: Routledge.

Searle, John(2004), *Mind: A Brief Introduction*, London: Oxford University Press.

Shelley, Mary(2017), *Frankenstein, or the Modern Prometheus*, New York: The MIT Press.

Thompson, Evan(2006), *Mind in Life*, New York: Harvard University Press.

## &lt;Abstract&gt;

Exploring the Coexistence of Human and Artificial Intelligence:  
Ethics of Responsibility and Literary Imagination

Park, Soyoung

Over the long period of time, humans have continually made relationship with the tools and machines we have made to develop our civilization and society. As we are entering into the era of the Fourth Industrial Revolution or the Post-human era, our relationship with machines has become one of the most important issues. In the era of the Fourth Industrial Revolution or the Post-human era, more specifically in the age of Artificial Intelligence(AI), we are facing a fundamental confusion about the boundary between human and machine.

In this paper, I will contemplate the question of the coexistence of human and machine, more specifically human and AI. For my research, I choose two science fictions. One is *Frankenstein, or the Modern Prometheus*(1818) by Mary Shelly, and the other is *Do Androids Dream of Electric Sheep?*(1967) by P. K. Dick. In *Frankenstein*, we meet a scientist by the same name of the title who creates and brings life to a manlike monster that turns on him and eventually destroys his life into dystopia. In *Do Androids Dream of Electric Sheep?*, we encounter future world living with androids in the earth of apocalyptic situation. By reading the two novels, I would like to explore the possibility of literary imagination of ethical responsibility for coexistence with artificial intelligence.

key words : Artificial Intelligence, Coexistence, Responsibility, Ethics,  
Science Fiction, Literary Imagination