도구주의를 넘어서 –
시몽동과 라투르를 통한 AI 윤리의 패러다임 전환

김재희 (을지대학교 교양학부, 조교수)

〈목 차〉

Ⅰ. 서론: AI 윤리의 도구주의적 한계와 존재론적 전환의 필요성
Ⅱ. 시몽동의 발생적 기술윤리: 기술적 실천과 사회적 가치의 관계 조율
Ⅲ. 라투르의 매개적 기술윤리: 기술의 블랙박스화에 저항하는 도덕적 반복
Ⅳ. 결론: 발생과 매개의 교차, 역동적 AI 윤리 거버넌스를 향하여

국문초록

본 논문은 현행 AI 윤리의 선언적·사후적 규제 모델이 지닌 도구주의적 한계를 비판하고, 시몽동과 라투르의 기술철학을 교차하여 AI 윤리의 역동적 모델을 모색한다. 시몽동의 발생론은 AI를 연합 환경과의 관계 속에서 구체화되는 과정적 개체로 파악하며 기술적 실천과 사회적 가치를 매개하는 기술교양의 문화적 확립을 요구한다. 라투르의 매개론은 AI를 목적과 행위를 변형하는 우회와 위임의 네트워크로 보며 블랙박스화된 연쇄에 마찰을 일으키는 도덕적 반복을 윤리적 실천으로 제시한다. 이 두 관점의 융합을 통해 AI 윤리를 정적인 규범 적용에서 벗어나 기술 생애주기 전반의 발생 계통과 위임 경로를 가시화하고 지속해서 점검·수정하는 역동적 거버넌스 실천으로 전환할 것을 제안한다.

키워드 : AI 윤리, 도구주의, 구체화, 연합환경, 기술교양, 우회, 위임, 기술적 매개, 도덕적 반복, 시몽동, 라투르

www.kci.go.kr

ⓒ 2026 이화인문과학원

Ⅰ. 서론: AI 윤리의 도구주의적 한계와 존재론적 전환의 필요성

현행 'AI 윤리'는 통상 사회에서 인공지능(Artificial Intelligence)의 윤리적 사용을 확립하기 위한 틀과 지침을 마련하는데 초점을 둔다. 최근 AI를 둘러싼 윤리적·정책적 논의는 공정성(fairness), 투명성(transparency), 설명가능성(explainability), 책임성(accountability)과 같은 주요 원칙의 나열을 넘어서 구체적인 위험관리와 거버넌스 체계로 확장되고 있다. 유네스코의 「인공지능 윤리에 관한 권고」(2021)는 인간의 존엄과 권리를 중심에 두면서 투명성, 공정성, 인간의 감독, 지속가능성 등을 제시하고 이를 정책 실행의 영역까지 구체화한다. 미국 국립표준기술연구소(NIST)의 「AI 위험관리 프레임워크(AI RMF)」(2023)는 거버넌스(govern), 매핑(map), 측정(measure), 관리(manage)의 네 기능을 중심으로 조직이 AI 시스템을 개발, 배포 및 사용할 때 발생할 수 있는 위험을 지속적으로 관리하도록 한다. 유럽연합의 「인공지능법(Artificial Intelligence Act)」(2024)은 2024년 8월 발효되어 2026년 8월 전면 시행에 들어간 세계 최초의 포괄적 AI 규제법으로서, AI의 위험도를 4단계(수용불가위험, 고위험, 제한된 위험, 최소위험)로 분류하여, 인간의 기본권을 침해하는 기술은 원천 금지하고 법안을 위반하는 기업에는 전 세계 연간 매출액의 최대 7%에 달하는 강력한 과징금 제재를 부과한다.[1] 이러한 흐름에 발맞추어 한국 정부 또한 2026년 8월 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」 제27조에 따라 3대 가치와 7대 원칙을 담은 「대한민국 인공지능 윤리원칙」을 발표하였다.[2] 이 원칙은 법적 구속력이 없는 자율규

1) UNESCO (2021), *Recommendation on the Ethics of Artificial Intelligence*, UNESCO; European Parliament and Council of the European Union (2024), *Regulation (EU) 2024/1689 (Artificial Intelligence Act)*, Official Journal of the European Union; Elham Tabassi (2023), *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, National Institute of Standards and Technology(NIST) AI 100-1.

2) 과학기술정보통신부, 「대한민국 인공지능 윤리원칙」, 인공지능 윤리 소통채널, 2026. (https://ai.kisdi.re.kr/aieth/main/contents.do?menuNo=400054, 2026. 9. 2. 접속). 2020년 12월에 발표되었던 『인공지능(AI) 윤리기준』에서는 3대 기본원칙 (인간 존

www.kci.go.kr

범으로서, 기존의 AI 공급자 및 개발자 중심에서 나아가 AI를 활용하는 이용자 기업과 개인, 정부, 사회 각 주체의 역할과 책임을 명시함으로써 AI 생애주기 전반의 안전과 신뢰를 확보하려는 국내의 규범적 실천을 보여준다.

그러나 AI 윤리 지침에 대한 비교 연구들은 이 원칙들의 외형적 수렴 이면에 상당한 해석적 간극이 있음을 보여준다. 안나 조빈(Anna Jobin) 연구진이 전 세계 84개의 비법률적 AI 윤리 문서를 분석한 바에 따르면, 투명성, 공정성, 무해성, 책임성, 프라이버시 같은 핵심 윤리 원칙이 광범위하게 반복되지만 그 구체적 의미와 실행 방식에서는 상당한 편차와 지정학적 불균형이 존재한다고 지적한다.[3] 데이터 윤리학자 브렌트 미텔슈타트(Brent Mittelstadt)는 고수준 원칙에 대한 합의 자체가 실제 윤리적 실천을 보장하지 않는다고 지적하며, AI 윤리 원칙들이 실질적인 효과를 거두지 못하고 선언에 그치는 것은 의료윤리와 달리 AI 분야에는 원칙을 구체적 실천으로 번역할 공유된 역사와 제도적 장치가 상대적으로 결여되어 있기 때문이라고 진단한다.[4] 최근 등장한 생성형 AI에 대한 논문들은 추상적 원칙의 무용성만이 아니라, 생성형 AI의 일상 침투로 인한 인지적 변화, 정치경제학적 구조적 모순도 비판한다. 이리나 슈피겔(Irina Spiegel)은 자율성, 도덕성, 윤리 같은 핵심 개념들이 AI와 고도화된 자율 에이전트가 만들어내는 복잡한 윤리적 문제를 다루기에는 점점 더 불충분해지고 있다고 지적하고, 고정된 규칙 중심의 윤리가 아닌, 가변적이고 네트워크화된 AI 생태계에 맞는 새로운 규범적 기준의 필요성을 제시한다.[5] 로버트 지아니(Robert Gianni)는 AI 윤리의 초점을 기술이 초래

엄성의 원칙, 사회의 공공선 원칙, 기술의 합목적성 원칙)과 10대 핵심요건 (인권 보장, 프라이버시 보호, 다양성 존중, 침해금지, 공공성, 연대성, 데이터 관리, 책임성, 안전성, 투명성)이 제시되었다면, 2026년 8월 발표된 『대한민국 인공지능 윤리원칙』에서는 3대 가치 (인간의 존엄성, 사회의 공공선, 인류의 지속가능성)와 7대 원칙 (인간중심성, 프라이버시 보호, 공정성·포용성, 책임성, 안전성, 신뢰성, 투명성)이 제시되었다.

3) Anna Jobin, Marcello Ienca, Effy Vayena (2019), "The Global Landscape of AI Ethics Guidelines," *Nature Machine Intelligence*, 1, pp. 389~399.

4) Brent Mittelstadt (2019), "Principles Alone Cannot Guarantee Ethical AI," *Nature Machine Intelligence*, 1, pp. 501~507.

www.kci.go.kr

하는 '결과(Consequences)' 중심에서 일상에 스며들어 인간의 습관과 인지를 서서히 바꾸는 '효과(Effects)'에 중심으로 이동해야 하며, 사후 규제나 개발 단계의 기술적 보정만으로는 윤리적 공백을 메울 수 없다고 경고한다.[6] 조엘 화이트(Joel White)는 AI 윤리에서 다루는 공정성, 가치정렬, 책임성 등의 논의가 자본주의적 생산 양식과 거대 테크 기업의 독점 권력 구조를 기정사실화할 뿐이라고 지적하고, 구조적 모순이라는 근본적 원인을 외면한 채 기술적 차원의 '편향 완화'에만 몰두하는 것은 무의미하다고 비판한다.[7]

현행 AI 윤리 지침이나 AI 윤리를 둘러싼 이러한 비판적 흐름에서 여전히 간과되고 있다고 내가 주목하는 것은, AI를 규제하기 위한 윤리 원칙의 실효성 없는 추상성이 아니라, 그 원칙들이 적용되는 방식을 암묵적으로 규정하는 존재론적 전제, 즉 기술을 인간이 미리 설정한 목적과 가치를 구현하고 관리해야 할 중립적 수단으로 파악하는 '도구주의적 경향'이다. 여기에는 기술이란 이미 완성된 객체이고, 윤리는 그 기술의 사용과 결과를 외부에서 평가·통제하는 규범 체계라는 전제가 깔려있다. 이는 AI를 어떻게 사용해야 하는지 사후적으로 규정하는 규칙 찾는 일로 윤리를 축소시키고, 기술적 대상의 발생 과정이나 인간-기술 관계의 변형 문제는 윤리적 성찰에서 밀려날 수 있다.

AI 시스템은 단순한 계산 기계도 아니고, 그렇다고 인간과 동일한 행위자도 아니다. 그것은 단일한 실체라기보다, 데이터 수집의 관행, 모델 설계의 선택, 연산 인프라의 구성, 조직의 목표, 사용자의 행태, 제도적 승인과 규제 등이 서로 얽히며 지속적으로 변형되는 일종의 과정적 실재다. 윤리적 문제는 따라서 기술이 사회에 적용된 이후에만 발생하는 것이 아니라, 기술이 어떤 방식으로 구성되고 어떤 환경 속에 배치되는가라는 발

5) Irina Spiegel (2024), "Criticizing Ethics According to Artificial Intelligence," *arXiv*, arXiv:2408.04609.

6) Robert Gianni (2026), "AI ethics after generative AI: from consequences to effects," *AI & SOCIETY*. Advance online publication.

7) Joel White (2026), "The Questions Concerning the Ethics of AI," *Triple Ampersand Journal* (&&&). (https://tripleampersand.org/the-questions-concerning-the-ethics-of-ai/)

www.kci.go.kr

생적 단계에서 이미 형성된다. 책임 귀속 문제도 간단하지 않다. 알고리즘과 플랫폼 기술이 매개하는 행위는 단일한 행위자에게 환원되지 않으며, 다양한 인간 및 비인간 요소들 사이에서 분산되고 재구성된다. 어떤 알고리즘이 특정 집단에 불리한 결과를 생산한다는 사실이 확인되더라도, 설계한 엔지니어도, 배치한 기업도, 허용한 제도도 저마다 책임을 분산시키고 아무도 명확하게 책임지지 않을 수 있다.

도구주의(instrumentalism)는 인간의 의도나 사회적 합의에 의해 미리 주어진 목적을 효율적으로 실현하는 중립적인 도구가 기술이라고 생각한다.[8] 이에 기초한 규범 적용형 윤리 모델은 기술로 인한 위험을 관리하는 데 일정한 기여를 하지만, 윤리적 문제가 기술의 발생과 매개 단계에서 이미 형성된다는 점을 포착하지 못한다는 점에서 한계를 지닌다. 설명 불가능하고 통제 불가능해 보이는 AI의 발전과 더불어 우리에게 필요한 것은 기술에 적용해야 할 규범 목록의 확장이 아니라, 기술의 존재양식 및 기술과 윤리의 관계를 근본적으로 다시 사유하는 전환적 태도일지 모른다.

질베르 시몽동(Gilbert Simondon)과 브뤼노 라투르(Bruno Latour)는 각기 다른 방향에서 이런 전환의 가능성을 보여준다. 두 철학자는 공통적으로 기술의 고유한 존재 양식에 주목하고, 기술을 단순한 수단이나 가치중립적 도구로 보지 않으며, 따라서 윤리/도덕을 기술 외부의 규범 체계로 이해하지 않는다. 시몽동은 기술적 대상을 고유한 발생과 개체화의 과정을 지닌 존재로 파악하며, 기술적 존재의 개체화가 사회적 의미와 단절되지 않도록 매개하는 문화적 성숙을 윤리의 가능 조건으로 제시한다. 라

8) 이 글에서 '도구주의'는 다음과 같은 테제들의 결합으로 이해한다. 첫째, 기술은 인간이 설정한 목적을 실현하는 수단이다(수단성). 둘째, 목적은 기술의 설계·사용과 무관하게 미리 결정된다(목적의 선행성). 셋째, 기술은 이미 완성된 객체이며 그 발생 과정은 윤리적 평가와 무관하다(완성성). 넷째, 기술 자체는 선하지도 악하지도 않으며, 가치는 오직 사용 목적과 사용자에게서 나온다(중립성). 다섯째, 윤리는 이렇게 주어진 기술의 사용과 결과를 외부에서 평가·통제하는 규범 체계다(외재성). 그러나 이 테제들이 논리적으로 반드시 서로 함축한다고 보지는 않는다. 가령, 첫째와 넷째의 경우, 수단성을 인정하면서도 설계에 가치가 내장된다고 볼 수도 있기 때문이다.

www.kci.go.kr

투르는 인간과 비인간이 결합된 네트워크 속에서 행위를 성립시키는 매개의 연쇄로 기술을 파악하며, 분산된 존재자들의 배치가 정당한지를 끊임없이 재검토하는 비판적 실천으로 도덕을 재정의한다. 두 철학자의 관점에서 비판되는 것은 낙관적 기술도구주의와 비관적 기술디스토피아주의이다. 얼핏 대립하는 것처럼 보이는 이 두 입장은 실은 동일한 이원론, 곧 인간과 기계를 서로 외재적인 두 항으로 놓고 어느 한쪽이 다른 한쪽을 지배한다는 틀을 공유한다. 시몽동과 라투르는 각기 다른 방식으로 이 이원론 자체를 해체하고자 했으며, 바로 이 지점에서 윤리를 기술에 외부적으로 적용되는 규범이 아니라 기술의 존재방식 안에서 이미 발생하는 물음으로 재구성할 수 있는 자원을 제공한다.

이 글에서 필자는 현행 AI 윤리에 암묵적으로 전제되어있는 도구주의적 한계가 시몽동의 발생적 기술론과 라투르의 매개적 기술론을 통해 어떻게 보완될 수 있는지 살펴보고자 한다. 이를 위해 '기술의 존재양식'에 주목한 두 텍스트를 비교 분석의 중심으로 삼는다. 시몽동의 『기술적 대상들의 존재양식에 대하여』(1958/2011)와 라투르의 『존재양식의 탐구』(2013/2023)는 서로 다른 방법론적 탐구에 기초하지만 공통적으로 도구주의적 기술관을 비판하고 기술-윤리 관계를 존재론적으로 재정립하려는 기획을 품고 있다. 이 연구의 목표는 시몽동의 발생적 분석과 라투르의 매개적 분석을 융합하여 기존 인간 중심의 도구주의적 윤리관이 지닌 한계를 보완할 수 있는 AI 윤리의 역동적 모델을 제안하는 데 있다.

II. 시몽동의 발생적 기술윤리: 기술적 실천과 사회적 가치의 관계 조율

1. 기술의 구체화, 연합환경, 그리고 내재적 규범성

시몽동은 기술적 대상을 인간 의도의 단순한 투영이나 외재적 수단으로 취급하는 도구주의를 거부한다. 『기술적 대상들의 존재양식에 대하여』 1부에 따르면, 기술적 대상은 고정된 물체가 아니라, 추상적 상태에서 구

www.kci.go.kr

체적 상태로 발전해 나가는 "구체화(concrétisation)" 과정을 겪는 존재다.[9] 초기 단계의 기술적 대상은 각 기능이 병렬적으로 결합된 상태에 머물러 외부적 보정이나 지속적 관리를 필요로 하지만, 구체화가 진전될수록 기술적 대상은 자신의 작동 조건을 내부화하고 구성 요소들이 상호의존적 체계를 이루어 보다 안정적인 작동을 하게 된다. 이런 구체화는 단순한 효율 향상이 아니라, 기술적 대상이 자기-조절적 체계로 발전하면서 개체로서의 고유한 존재 양식을 획득해 가는 과정이다. 이런 기술적 구체화의 방향은 인간에 의해 임의로 정해지는 것이 아니다. 어떤 구조가 보다 통합적이고 적합한지, 어떤 방향의 구체화가 기술의 내적 논리에 부합하는지 구별하는 기준은 기술 외부가 아니라 기술적 개체화 과정 자체에서 생성된다. 기술적 대상은 기술 내적 문제를 해결해 나가면서 더 안정적인 기능적 작동과 더 정합적인 구조를 향해 나아가는 내적 방향성을 지닌다.

그러나 기술적 개체의 자율적 진화가 곧 인간과 독립된 완전 자동화를 함축하는 것은 아니다. 기술적 구체화는 기술 내부의 정합성만이 아니라 "연합 환경(milieu associé)"[10]과의 관계를 근본 조건으로 한다. 연합환경은 기술 대상이 정상적으로 작동하기 위한 필수 조건인 동시에, 기술 대상의 작동 그 자체에 의해 비로소 지속적으로 조성되고 유지되는 재귀적·자율적 환경이다. 시몽동이 즐겨 드는 갱발(Guimbal) 터빈의 경우, 터빈이 잠겨 있는 바닷물은 단지 외부 자연 환경에 그치는 것이 아니라, 에너지 공급원이면서 동시에 냉각제이자 절연체로서 다기능적으로 작동하는 터빈의 존재 조건이다. 기술적 개체에 연합된 환경은 기술 바깥의 단순한 배경이 아니라, 기술의 구조와 기능을 함께 성립시키는 조건들이다. 기술은 환경을 필요로 하고, 환경은 기술에 의해 다시 조직된다. 양자는 분리된 채 존재하지 않고, 재귀적 인과관계 속에서 함께 형성된다.

이처럼 기술적 대상이 구체화하면서 내부 요소들의 구조적 결합과 외부

9) 질베르 시몽동 (2011), 『기술적 대상들의 존재양식에 대하여』, 김재희(역), 그린비, 22~30쪽.
10) 질베르 시몽동 (2011), 86쪽.

www.kci.go.kr

연합 환경과의 재귀적 균형을 이루어가는 본질적 성질을 시몽동은 "기술성(technicité)"이라 부른다.[11] 기술적 대상이 특정 기능을 수행하는 물리적 존재자라면, 기술성은 그 대상들 안에 침전되어 있는 발명의 역량이다. 발명가가 새로운 기술적 대상을 창조할 때, 그 창조 행위 속에서 작동하는 것은 단순한 목적-수단 계산이 아니라, 기술적 잠재성들을 감지하고 그것들을 새로운 통합으로 이끄는 '기술적 감각'이다. 이 기술적 감각이 인간에게 내면화된 것이 '기술적 정신'이라면, 그것이 물질적 구조 안에 결정화된 것이 기술적 대상인 것이다.

2. 기술과 윤리의 발생론

『기술적 대상들의 존재양식에 대하여』 3부의 발생적 문화론에 따르면, 시몽동은 기술을 근대적 문명의 산물이나 인간 합리성의 도구적 표현으로 보는 관점을 거부하고, 세계-내-존재로서의 인간 실존 양식의 분화(상전이: déphasage)에서 '발생'하는 것으로 이해한다. 인간과 세계의 관계는 고정된 것이 아니라 '준안정적 시스템'과 같아서 과포화된 문제를 해결하기 위해 상전이하는 방식으로 달라진다. 기술, 종교, 과학, 윤리, 미학, 철학 등으로 표현되는 '인간의 세계에 대한 관계 양식'은 이러한 상전이로 분화된 상들(phases)이다. 시몽동은 이 상들의 상호 공존과 상호 협력을 통해 균형 있게 조절되는 문화를 지향하며, 기술을 배제하던 문화의 불균형을 비판하면서 기술을 다른 상들과의 관계 속에서 적합하게 인식하는 균형잡힌 문화의 설립을 주장한다.[12]

시몽동의 발생적 가설에 따르자면, 가장 본원적인 인간과 세계의 관계 양식은 '마술적 단일성'이다. 이는 일체의 구조가 결여된 무차별적 혼돈의 단일체는 아니고, 생명체로서의 인간과 환경으로서의 세계가 특권화된 시공간적 지점들(숲의 심장, 산 정상, 경계 지역 같은 공간적 요충지들, 또는 탄생, 죽음, 잠 등의 시작과 끝을 표시하는 시간적 순간들)의 연결망으

11) 질베르 시몽동 (2011), 109쪽.
12) 김재희 (2017), 『시몽동의 기술철학: 포스트휴먼 사회의 청사진』, 아카넷, 171쪽.

www.kci.go.kr

로 이미 구조화되어 있는 것이다.[13] 그런데 이 원초적 단일성이, 마치 물이 특정 조건에서 얼음과 수증기의 상으로 상전이 하듯이, 특이점들의 망상(網狀) 구조가 과포화되어 '모양과 바탕'으로 분리되면서 '기술'과 '종교'로 분화(상전이)하게 된다. 기술은 세계의 특수하고 국지적인 모양(figures)들에 작용하는 극으로, 종교는 세계의 전체적이고 보편적인 바탕(fond)을 담지하는 극으로 동시 분할된다. 기술이 세계와 접하는 실천적 유용성을 맡는다면, 종교는 세계의 보편적 의미와 총체성을 담당하게 되는 것이다. 기술은 바탕에서 모양을 떼어내어 대상화하고, 종교는 바탕의 권력과 총체성을 유지한다. 마술적 자연은 한편으로는 연장이나 도구의 형태로 운반 가능한 기술적 대상들로 파편화되고, 다른 한편으로는 신적이고 성스러운 종교적 특성들로 추상화·보편화된다. 원초적 단일성의 이러한 양분화는 그러나 동시에 결핍을 낳는다. 기술적 극은 세계 전체의 의미로부터 절단된 조작 가능성에 머물고, 종교적 극은 구체적 실천 없는 의미의 총체성에 머문다.

이 1차 분화로 인해 발생한 불균형과 결핍을 치유하기 위해, 2차 분화로 발생하는 것이 '과학'과 '윤리'다. 우선, 기술과 종교는 각각 이론적 양식과 실천적 양식으로 분화한다. 즉 기술은 실천(세계에 직접 개입하는 구체적 조작과 생산의 실천)과 이론(그 개입의 원리를 탐구하는 추상적 인식)으로 분화하고, 종교는 이론(실재에 대한 체계적이고 총체적인 표상을 요구하는 신학적 교리)과 실천(특수한 명령보다 총체성의 이름으로 정당화되는 절대적인 행위 규범들)으로 분화한다. 그리고 기술적 이론과 종교적 이론 사이의 거리에서 '과학'이, 또 기술적 실천과 종교적 실천 사이의 거리에서 '윤리'가 발생한다. 과학이 기술적 모양들과 종교적 바탕 사이의 이론적·지식적 매개를 수행한다면, 윤리는 기술적 행위의 국지적 실천과 종교적 의미의 보편적 가치를 삶의 전체성 속에서 실천적으로 통합하는 매개체로 등장한다. 다시 말해, 시몽동에게 윤리는 기술을 외면하거나 억압하는 외부적 규범이 아니다. 윤리는 인간과 세계의 원초적 관계가

13) 김재희 (2017), 172쪽.

www.kci.go.kr

분화되면서 발생한 기술적 실천(상대적 소망들)과 종교적 의미(보편적 가치) 사이의 긴장을 조정하고 양립가능한 행위의 최적 조건을 모색하는 동력으로 기능한다.[14] 기술이 가져오는 세계 개입의 힘을 종교적 의미의 차원과 다시 결합하는 것이 윤리의 존재 이유인 것이다.

3. 기술과 윤리를 관계 짓는 기술 정신과 기술 교양

기술과 윤리의 이러한 발생적 관계는 현대 사회에서 '기술 교양(culture technique)'과 '기술적 정신(mentalité technique)'의 확립을 요구한다. 기술적 대상의 내재적 규범성(구체화의 방향성, 연합 환경과의 정합성)이 자동으로 사회적 가치와 일치하지는 않는다. 기술적 규범성과 사회적 규범성 사이의 간극을 메우는 것이 시몽동적 윤리의 과제이다. 윤리는 이 두 규범성이 서로를 무화하지 않고 공명할 수 있는 조건을 마련하는 실천이다. 다시 말해 시몽동의 윤리는 기술에 규범을 부과하는 것도, 기술의 내재적 발전에 사회가 복종하는 것도 아니다. 그것은 두 규범성 사이의 매개이자 조율이다.

시몽동의 「기술적 정신」(1961/2014)에 따르면, 기술적 정신은 기술적 대상들을 단순한 도구나 소비재로 취급하는 것이 아니라, 그것들의 고유한 존재양식, 즉 발생 역사, 내재적 논리, 연합 환경과의 관계 등을 이해하는 능력이자 태도이다.[15] 한편에는 기술을 단순히 사용하고 소비하는 태도, 즉 기술의 존재를 그 기능적 효용으로만 파악하는 태도가 있고, 다른 한편에는 기술을 두려워하고 적대시하는 태도, 즉 기술을 인간성의 위협이나 소외의 원인으로 보는 태도가 있다면, 시몽동은 양자 모두 기술적 대상의 고유한 존재양식을 이해하지 못하는 데서 비롯된다고 본다. 기술적 대상들을 문화의 영역으로 끌어들이지 못할 때, 즉 그것들을 인간 정신의 표현이자 세계와의 관계 방식으로 이해하지 못할 때, 기술적 대상들

14) 질베르 시몽동 (2011), 303쪽.
15) Gilbert Simondon (1961/2014), "La mentalité technique," *Sur la technique(1953-1983)*, PUF. pp. 295~313.

www.kci.go.kr

은 인간에게 소외된 힘으로서 군림한다. 기술의 문화적 소외가 지속되는 한, 기술적 규범성과 사회적 가치 사이의 매개는 불가능하다. 오늘날 AI를 둘러싼 대중적·정책적 담론에서도 유사한 이분법이 재현된다. 한쪽에서는 생성형 AI를 입력된 명령에 반응하는 고도화된 도구로 축소하고, 다른 한쪽에서는 이를 인간의 통제를 벗어날 잠재적 자율체로 과장한다. 시몽동적 관점에서 중요한 것은 어느 한 서사를 선택하는 것이 아니라, 두 서사 모두가 데이터 선별, 아키텍처, 학습, 미세조정, 평가, 배포 환경의 상호작용이라는 기술적 발생 과정을 충분히 설명하지 못한다는 점이다.

따라서 시몽동은 "기술교양(culture technique)" 교육을 강조한다.[16] 기술교양은 단순한 기술 지식이 아니다. 기술적 대상의 존재 양식을 이해하고 그것과 의미 있는 관계를 맺을 수 있는 집단적 역량이다. 기술적 사유의 핵심은 특정 기술 장치의 사용법을 아는 데 있는 것이 아니라, 그 기술적 도식을 통해 사물들 사이의 관계와 작동 방식을 파악하는 데 있다. 이 관점에서 기술교양의 결여는 단순한 전문지식의 부족이 아니라 사회가 기술적 대상과 관계 맺는 방식의 빈곤을 의미한다. 기술적 개체화와 사회적 가치가 상호 이해될 수 있는 조건, 즉 '기술 교양'의 결여는 인식론적 손실이기 이전에 윤리적 손실이다. 기술 교양이 결여된 사회에서 기술은 점점 더 이해 불가능한 블랙박스가 되고, 시민은 기술의 운용에 참여하지 못한 채 결과만 받아들이게 된다. 사회가 기술의 개체화 과정에 능동적으로 참여할 수 없게 될 때, 기술 발전은 내재적 방향성을 잃고 외부의 경제적·도구적 논리에 일방적으로 종속될 수 있다.

생성형 AI는 거대한 데이터, 연산 인프라, 서버, 네트워크, 전력, 냉각 시스템, 모델과 상호작용하는 인간들의 집합적 실천, 이 모두를 자신의 '연합 환경'으로 갖추지 않고서는 존재할 수 없다. 이 복합적 환경은 단지 외부의 조건이 아니라 AI의 존재 자체를 가능하게 하는 내적 조건이다. AI는 독립적으로 존재하는 지적 주체가 아니라, 인간과 비인간의 요소가 함께 구성하는 '기술적 앙상블'인 셈이다. AI를 무조건적인 위협으로만 간

16) 질베르 시몽동 (2011), 213쪽.

www.kci.go.kr

주하면 기술의 잠재적 공공성을 놓치게 되고, 반대로 지나친 효율과 편의성에 기대면 기술이 인간의 개체화와 사회적 성숙을 가로막는 지점을 놓치게 된다. 시몽동의 사유는 이 둘을 넘어서, 기술을 이해하고 관계 맺는 집단적 역량 자체를 윤리의 핵심으로 삼도록 요구한다. 기술적 대상의 고유한 존재를 이해하는 기술적 정신과 기술 교양 없이는 어떤 규제 원칙도 기술의 내재적 발전과 실질적으로 소통할 수 없다. 기술 교양을 통해 기술적 정신이 확립될 때, 비로소 기술적 규범성과 사회적 규범성이 서로를 이해하고 조율하는 대화가 가능해진다.

따라서 시몽동의 기술윤리는 기술적 실천과 동시대의 규범적 실천 사이의 관계와 거리를 지속적으로 성찰할 것을 요구한다. AI 시스템 자체에 윤리 규범을 공학적으로 내장하는 문제와, AI라는 기술적 실천이 법, 정치, 교육, 노동, 인간의 권리와 같은 동시대의 규범적 실천과 맺는 관계를 성찰하는 문제는 동일하지 않다. 전자는 설계의 문제이고, 후자는 철학적·정치적·문화적 성찰의 문제이다. AI 윤리가 후자의 차원까지 나아가지 않는다면, 기술적 구현의 성공을 윤리적 성공으로 오인할 위험이 있다. 시몽동적 관점에서 윤리는 기술을 외부에서 감시하는 장치가 아니라, 기술의 개체화 과정과 사회적 의미가 서로 단절되지 않도록 매개하는 조건이어야 한다. 기술이 어떤 연합된 환경 속에서 개체화되고 있는지, 그리고 그 개체화가 인간의 심리적·집단적 개체화와 어떻게 맞물리는지를 살피는 데 윤리의 핵심이 있다.

Ⅲ. 라투르의 매개적 기술윤리: 기술의 블랙박스화에 저항하는 도덕적 반복

1. 기술의 우회, 위임, 그리고 블랙박스화

「도덕과 기술: 수단의 종말」(2002)에서 라투르가 비판하는 것은 전통적인 '목적-수단 도식'이다. 이 도식에 따르면, 도덕은 목적을 규정하고 기술은 그 목적을 실현하는 수단을 제공한다. 라투르는 이 도식이 기술과

www.kci.go.kr

도덕 모두를 오해한다고 주장한다. 기술의 측면에서 보면, 기술적 수단은 목적을 '중립적으로' 실현하지 않는다. 기술적 우회는 목적을 변형시키고, 수단은 새로운 목적을 창출하기 때문이다. 도덕의 측면에서 보면, 도덕적 목적은 기술적 수단을 거치지 않고는 현실에서 실현될 수 없다. 도덕은 항상 이미 기술적으로 매개되어 있기 때문이다.[17]

『존재양식의 탐구』에서도 역시 기술 도구주의를 비판한다. "목적과 수단의 관계보다 기술에 덜 적합한 것은 없다. 목적과 수단은 동시에 발명되는 것이기 때문이다. 기술을 단순한 '과학의 적용'이나 단순한 '자연에 대한 지배'로 보는 것은 심각한 오해이다"[18]. 라투르에게 기술은 의도의 단순 실행이 아니라, 의도를 번역하고 굴절시키며 재배치하면서 행위 자체를 변형시키는 매개자다. 중개(intermédiaire)가 입력된 것을 거의 그대로 전달하는 통로라면, 매개(médiateur)는 자신을 통과하는 것을 변형하고 새로운 결과를 산출한다.

『존재양식의 탐구』에 따르면, 기술, 정치, 법, 종교, 도덕 등의 여러 존재양식들은 각각의 고유한 궤적(각 양식의 연속성)과 적정성의 조건(잘 작동할 때와 잘못 작동할 때의 구분 기준)을 갖는다. 여기서 '기술'의 고유한 특성은 '우회(détour)'와 '위임(délégation)'을 통해 존재한다는 데 있다.[19] 우선, '우회'는 기술적 행위가 성립하기 위해 반드시 이종적 요소들의 연합을 거쳐야 함을 뜻한다. 못을 박으려면 망치가 필요하고, 망치를 생산하려면 철광석, 채굴 기술, 제련 공정, 운송 체계, 시장 교환이 필요하다. 이 연쇄는 원래의 의도(못을 박는다)를 달성하기 위해 이질적인 실재들을 연합시키는 우회이다. 라투르의 논점은, 이 우회가 원래의 의도를 '중립적으로' 전달하지 않는다는 것이다. 목표를 달성하려면 이질적인 재료·지식·도구·규칙·조직·인프라를 연합시켜야 하며, 이 과정에서 원래의 의도는 기술적 매개물에 의해 번역·굴절·재배치된다. 이질적 행위소들을 연

17) Bruno Latour (2002), "Morality and Technology: The End of the Means," *Theory, Culture & Society*, 19(5-6), pp. 247~260.
18) 브뤼노 라투르 (2023), 『존재 양식의 탐구』, 황장진(역), 사월의책, 327쪽.
19) 브뤼노 라투르 (2023), 314쪽.

www.kci.go.kr

결하여 하나의 집합체를 구축하는 기술적 작업은 구성 요소들 간의 저항과 이질성을 교묘하게 조정하며 적합한 연결을 찾아내는 '번역'이다. "장인, 건축가, 엔지니어가 매일 실천하는 것이 바로 이러한 매번 완전히 고유한 이동, 번역이다."[20] 우회는 행위의 형태 자체를 변형시킨다. 망치를 든 인간은 단순히 손의 힘이 증가한 것이 아니라, 망치와 결합하여 폭력의 새로운 가능성을 가진 다른 존재가 된다.

그리고 '위임'은 우회를 통해 연합된 행위자들의 역할이 기술적 대상 안에 '접혀(plié)' 안정화되는 과정이다. 기술적 대상은 단일한 설계자의 의도가 그대로 구현된 산물이라기보다, 서로 다른 요구와 제약이 조정되고 안정화된 결과라고 할 수 있다. 기술적 대상은 '지금-여기'에서 쓰이는 단일한 물건이 아니다. 그것이 가능해지기까지 동원된 노동·물류·표준화·법과 정책·자원 채굴·과학적 지식·정치적 결정 등이 압축된 '시공간의 접힘'이 거기에 들어가 있다. 스마트폰의 경우, 그 안에는 수십 년에 걸친 반도체 기술의 역사, 희토류를 채굴한 노동자들의 땀, 특허 분쟁의 법정, 통신 표준을 둘러싼 국제 협상 등이 함께 '접혀' 있다. 기술을 단순 사용 행위로만 이해하면 기술을 가능케 한 이 다층적 조건들이 지워진다. 시몽동에게 기술적 개체가 연합된 환경과의 앙상블로 존재하듯이, 라투르에게도 기술적 실재는 단일한 '물건'이라기보다 '시공간을 가로질러 응축된 네트워크의 형태'로 존재한다.

그런데 기술적 대상이 성공적으로 작동할 때, 그것을 가능하게 만든 이 우회와 위임의 연쇄는 사용자에게 보이지 않게 된다. 문제는 기술이 성공적으로 작동할수록, 이 복잡한 네트워크가 '블랙박스화(blackboxing)'되고, 사용자는 기술을 투명한 도구로 오인하게 된다는데 있다. 라투르는 바로 이 블랙박스화가 기술의 힘과 윤리적 위험이 함께 잠복하는 지점이라고 본다. 블랙박스가 닫힐수록, 누가 무엇을 통해 무엇을 가능하게 했는지에 대한 책임의 경로가 흐려지고, 어떤 행위자들이 연합되고 배제되었는지에 대한 가치의 경로가 가려지며, 기술적 결정을 되돌릴 수 있는

20) 브뤼노 라투르 (2023), 337쪽.

www.kci.go.kr

수정 가능성도 약화되기 때문이다. 기술적 대상이 블랙박스화될 때, 그 안에는 특정 인간 집단의 이해관계, 가정, 편견이 위임되어 있다. 속도방지턱은 '모든 차는 이 구간에서 서행해야 한다'는 규범을 수행하지만, 그 규범의 수행은 구급차와 일반 승용차를 구별하지 않는다. 총기는 '상대를 무력화할 수 있다'는 행위 가능성을 창출하지만, 그 가능성은 총기 제조사, 로비 집단, 법률 체계의 정치적 선택들이 안정화된 결과이다. 기술적 대상 안에는 언제나 이런 종류의 위임, 즉 도덕적, 정치적, 경제적 선택들의 결정화가 들어 있다.

라투르의 주장은, 기술적 대상들도 '도덕적 행위자(moral agent)'라는 것이 아니라, 도덕적 판단은 인간 주체에게만 귀속될 수 없다는 것이다. 왜냐하면 행위(agency)는 인간과 비인간이 결합된 연합 속에서 분산적으로 수행되기 때문이다. 따라서 도덕적 검토는 개인의 의도만이 아니라, 그 행위를 가능하게 한 기술적·제도적 연합 전체를 대상으로 해야 한다. 이는 기술적 비인간 요소에 인간과 같은 도덕적 책임을 부과한다는 뜻이 아니며, 책임 귀속의 물음을 폐기한다는 뜻도 아니다. 비인간 요소가 행위의 성립에 인과적으로 기여한다는 사실과, 책임을 질 수 있는 존재가 누구인가는 별개의 문제이기 때문이다. '누가 책임이 있는가'라는 물음을 '이 행위는 어떤 연합을 통해 수행되었는가, 그 연합 속에서 어떤 선택들이 은폐되었는가'라는 물음으로 보충하는 것이다. 이 보충을 통해 책임 귀속은 개별 행위자에게 고립적으로 묻는 방식에서 벗어나, 연합을 구성하고 유지해 온 설계자·기업·제도 등 인간 행위자들에게 그 연합의 구성에 대한 책임을 묻는 방식으로 정교해진다. 책임 귀속의 단위가 개인의 의도에서 연합의 구성과 유지로 확장되는 것이지, 책임이 사라지거나 비인간에게 이전되는 것은 아니다.

2. 양심의 거리낌과 다시 시작함

『존재양식의 탐구』에서 도덕은 기술과 마찬가지로 고유한 존재양식을 갖는다. 도덕은 보편적 규범 체계도 아니고 주관적 양심의 내면성도 아니

www.kci.go.kr

다. 도덕의 적정성 조건은 "양심의 거리낌(scruple)"이다.[21] 라투르는 이 단어의 어원인 라틴어 'scrupulus: 신발 속의 작은 돌멩이'를 참조한다. 도덕은 매끄럽게 진행되는 행위 연쇄 속에 던져져 행보를 머뭇거리게 만드는 일종의 '신발 속의 작은 돌멩이'와 같은 것, 즉 양심의 거리낌과 같은 것이다. 도덕적 경험의 핵심은 '내가 지금 하고 있는 것이 정말로 내가 하고자 했던 것인가?'라는 물음이 불현듯 제기되는 순간이다. 이 머뭇거림은 행위를 중단시키거나 방해하는 것이 아니라, 행위자로 하여금 자신이 수행하고 있는 것의 의미를 재검토하게 만든다. 즉 도덕은 행위를 정당화하거나 단호한 결단을 내려주는 장치가 아니라 오히려 '확신을 중단시키는 마찰이자 방해'의 역할을 한다는 것이다.

이 양심의 거리낌과 불가분한 개념이 "반복(reprise)"이다.[22] 이 반복, 즉 '다시 시작함'은 단순히 중단된 행위의 재시작이 아니다. 그것은 양심의 거리낌이 제기한 물음, 즉 '지금 내가 하고 있는 것이 무엇인가'에 응답하는 방식으로 행위 연쇄를 다시 펼쳐 보는 것이다. "만약 내가 타자를 통해서만 존재한다면, 우리 중 어느 것이 목적이고 어느 것이 수단인가? 그것을 거쳐 통과해야하는 나는 그것의 수단인가, 아니면 그것이 나의 수단인가? 내가 목적인가, 아니면 그것이 나의 목적인가?"[23] 물으면서, 목적과 수단의 최적 분배, 즉 "어떤 계산으로도 최적을 최적화할 방법이 전혀 없지만, 그럼에도 불구하고 최적에 도달해야 한다."[24]는 양심의 거리낌을 반복하는 것이다. 이 '다시 시작함'의 작업 속에서 행위자는 자신이 어떤 연합을 통해 행위하고 있는지, 그 연합 속에서 어떤 존재들이 동원되었고 어떤 존재들이 배제되었는지, 그 결과로 발생한 것이 원래 의도했던 것과 일치하는지 등을 다시 검토한다.

따라서 도덕이 잘못 작동하는 경우는 두 가지 방향에서 나타난다. 첫째, 양심의 거리낌이 전혀 작동하지 않는 경우. 이것은 도덕적 주체가 없

21) 브뤼노 라투르 (2023), 662쪽.
22) 브뤼노 라투르 (2023), 664쪽.
23) 브뤼노 라투르 (2023), 656쪽.
24) 브뤼노 라투르 (2023), 666쪽.

www.kci.go.kr

는 경우가 아니라, 모든 행위가 자동화되어 주저함의 여지 자체가 없어지는 경우이다. 가령, 알고리즘이 모든 결정을 내리고, 그 결정의 근거가 블랙박스 안에 감추어져 있을 때, 양심의 거리낌이 발생할 자리가 없어진다. 둘째, 양심의 거리낌이 있더라도 '다시 시작함'이 발생하지 않는 경우. 머뭇거리지만 결국 기존의 연쇄를 그대로 지속하는 것이다. 진정한 '반복/다시 시작함'은 연쇄의 실질적 재검토와 수정 가능성을 포함해야 한다.

기술과 도덕이 교차할 때의 핵심 위험으로 라투르가 주목하는 것은 "최적의 계산을 자동화하는 전복"[25]을 일으키는 자본주의 경제다. 기술적 수단이 성공적으로 작동할수록, 즉 블랙박스화가 진전될수록, 수단이 산출하는 지표들(효율성, 정확도, 참여율 등)이 원래의 도덕적 목적을 대체하기 시작한다. 가치를 사실의 문제로 바꾸어 최적을 계산할 수 있다고 믿는 물신화된 '경제'가 양심의 거리낌에 대한 탐구를 중단하고 목적과 수단의 분배에 대한 개입을 중단시킨다. 목적은 더 이상 독립적으로 검토되지 않고, 수단이 최적화하는 방향으로 재정의된다. '우리가 무엇을 위해 이것을 하는가'라는 물음이 '이것이 어떻게 최적화되었는가'라는 물음으로 교체된다. 이것이 수단의 목적화이며, '양심의 거리낌'이 차단되는 전형적인 방식이다.

그렇다면 기술과 도덕의 교차에서 도덕적 반복이 작동하기 위한 조건은 무엇인가? 라투르의 관점에서 가능한 답변은, 기술적 우회가 블랙박스화될 때 도덕은 그 블랙박스를 열어 내부의 연합, 즉 어떤 존재들이 동원되었고, 어떤 선택들이 이루어졌으며, 어떤 존재들이 배제되었는가를 다시 가시화하는 것이다. 이것은 기술을 파괴하거나 블랙박스를 영구히 열어두는 것이 아니다. 그것은 블랙박스가 닫히는 속도보다 다시 시작함의 능력이 더 빠르게 유지될 수 있도록 구조적 조건을 설계하는 것이다. 도덕은 규칙을 행위에 적용하는 것이 아니라, "목적과 수단의 최적 분배에 대한 양심의 거리낌이라는 반복"[26]을 수행하며, 우리가 무엇을 하고 있는가,

25) 브뤼노 라투르 (2023), 679쪽.

www.kci.go.kr

이 행위에 누가 연루되어 있는가, 무엇이 누락되었는가 계속해서 되묻는 실천이다. 기술이 효율을 위해 우회하고 연결하며 블랙박스화한다면, 도덕은 그 블랙박스를 다시 열어 무엇이 접혀 있는지 확인하고, 그 연결이 정당한지, 배제된 행위자는 없는지 물으며 매개 연쇄를 다시 펼쳐 조정하는 '다시 시작함'의 운동이다.

오늘날 대규모 AI 시스템은 방대한 '우회'와 '위임'의 연쇄다. 특정 콘텐츠를 차단하거나 특정 대상을 우선적으로 추천하는 판단은 데이터 라벨, 모델 구조, 손실 함수, 임계값, 필터, 인터페이스, 운영 지침, 인간 검토 절차 등으로 분산되어 구현된다. 학습 데이터를 수집하고 정제한 노동자들의 결정, 아키텍처를 설계한 엔지니어들의 선택, 정렬 과정에서 내부화된 가치 판단, 운영 정책이 정한 사용 한계, 이 모든 것이 하나의 시스템 안에 위임되고 응축되어 있다. 그러나 시스템이 성공적으로 작동할수록, 이 위임 연쇄는 사용자에게 보이지 않게 된다. 사용자는 이 응축된 결정들의 총체와 매일 상호작용하면서도, 자신이 무엇과 관계 맺고 있는지를 알 수 없다. 이 상황에서 라투르가 말하는 '양심의 거리낌'의 조건은 무엇인가. 개인의 비판적 태도만으로는 충분하지 않다. 마찰이 연쇄 밖에서 오는 규범의 형태로 부과될 때, 그것은 도덕적 양심이 아니라 외재적 제약에 불과하다. AI 시스템 안에 도덕이 작동하게 하려면, 위임 연쇄의 일부를 가시화하는 인터페이스, 자동화된 결정에 이의를 제기할 수 있는 절차적 채널, 학습 데이터와 정렬 과정에 대한 감사 가능한 기록 등이 설계 안에 내장될 필요가 있다. 이것들은 기술적 투명성의 요구가 아니라, 도덕의 마찰이 기술 연쇄 안에서 발생할 수 있게 하는 구조적 조건이다.

AI 시대의 핵심 위험은 기술적 우회가 지나치게 성공하면서 도덕적 가책과 반복이 개입할 여지가 사라지는 현상이다. 효율성·정확도·참여율·수익성 같은 기술적 지표가 원래의 사회적 목적을 대표해 버리며, 목적이 수단에 의해 잠식되는 상황이 발생한다. 추천 알고리즘의 경우, 참여율이나 체류시간은 원래 목적을 위한 수단적 지표였으나 조직의 핵심성과지표

26) 브뤼노 라투르 (2023), 657쪽.

www.kci.go.kr

(KPI: Key Performance Indicator)와 결합하면서 목적 자체로 승격된다. 이때 알고리즘이 어떤 인간 관계와 공론장을 만들어내는지는 블랙박스에 가려진다. 도덕적 '양심의 거리낌'은 '참여율이 높아졌으니 성공'이라는 자동성을 멈추게 하고, '무엇을 위해 최적화하고 있었는가'를 다시 묻게 하는 '다시 시작함'의 요구로 작동해야 한다. 특히 0.001초 만에 결정을 내리는 AI의 초고속 환경에서, '신발 속의 작은 돌멩이'는 단순한 심리적 주저함이 아니라 설계 단계에 의도적으로 삽입된 지연 전략이어야 한다. 추적 가능한 경로를 유지·설계하는 실천, 즉 감사·기록·변경이력·의사결정 절차의 가시화가 윤리의 중심에 놓여야 한다. AI 윤리의 실질적 과제는 원칙의 선언이 아니라, 기술적 연쇄 안에 '도덕'이 개입할 수 있는 구조적 계기를 어떻게 설계할 것인가여야 한다.

IV. 결론: 발생과 매개의 교차, 역동적 AI 윤리 거버넌스를 향하여

시몽동의 기술적 대상은 완성된 물건이 아니라 발명과 구체화의 역사를 통해 자신의 고유한 존재양식을 획득한 실재이며, 그 존재양식은 내재적 규범성(구체화의 방향성, 연합 환경과의 정합성)을 포함한다. 따라서 기술을 단지 수단으로 보는 것은 기술 존재의 존재론적 두께를 무시하는 것이다. 시몽동의 윤리적 실천은 기술적 대상의 발생 역사와 내재적 논리에 대한 집단적 이해를 기술 교양으로 쌓아가는 문화적 성숙의 형태를 취한다. 이 실천은 교육적이고 장기적이다. 그것은 즉각적인 규제나 통제가 아니라, 기술을 둘러싼 문화적 생태계의 변화를 통해 기술적 실천과 사회적 가치의 관계를 조율하는 과정이다.

라투르는 기술을 투명한 중개자가 아니라 행위 목적을 변형하는 매개자로 분석한다. 우회는 의도를 굴절시키고, 위임은 새로운 행위 가능성을 창출하며, 블랙박스화는 정치적 선택을 자연화한다. 따라서 기술을 단지 수단으로 보는 것은 기술이 실제로 행위에 미치는 효과를 무시하는 것이

www.kci.go.kr

다. 라투르의 윤리적 실천은 정치적 저항의 형태를 띤다. 기술이 행위를 변형시킨다면, 그 변형이 누구에게 유리하고 누구를 배제하는지를 지속적으로 물어야 한다. 즉 블랙박스화된 네트워크를 열어 그 안의 선택들과 배제된 행위자들을 가시화하며 연합의 조건을 재협상하고 목적-수단 분배를 다시 하는 것이다. 이 실천은 즉각적이고 반복적이다. 그것은 블랙박스가 닫힐 때마다 다시 시작되어야 하는 지속적인 감시이자 협상이다.

요컨대, 시몽동은 기술-윤리 관계에서 '무엇을 알아야 하는가'의 차원을 강조하고, 라투르는 '누가 결정하는가'의 차원을 강조한다. 이 두 차원은 상호 보완적이다. 기술적 결정에 민주적으로 참여하기 위해서는 기술적 정신이 필요하고, 기술적 정신의 형성은 그것을 가능하게 하는 정치적 조건, 즉 기술적 지식에 대한 공정한 접근, 기술 설계 과정의 투명성에 의존한다. 기술적 정신의 확립은 기술교육의 혁신을 요구하지만, 그 혁신은 기술적 지식과 설계 과정에 대한 접근을 민주화하는 정치적 조건 없이는 이루어질 수 없다. '이 기술적 대상이 어떤 발생적 계통과 환경적 조건 속에서 현재의 형태에 이르렀는가'를 묻는 발생 분석이 없으면 현재 시스템의 형성 조건을 놓치기 쉽고, '이 기술적 배치에서 인간과 비인간 행위소들은 어떤 방식으로 결합하여 서로의 목적과 능력을 변형하는가'를 묻는 매개 분석이 없으면 형성된 시스템이 운영 과정에서 감추고 있는 목적-수단 배치의 변형을 놓치기 쉽다.

가령, 이러한 발생과 매개의 이중적 관점은 최근 생성형 대규모 언어모델(LLM)의 '인간 가치 정렬(alignment)' 실무를 분석할 때 유용할 수 있다. 모델이 유해 답변을 거부하고 윤리적 규범을 따르도록 조정하는 RLHF(인간 피드백 기반 강화학습) 과정은 단순히 추상적 윤리 원칙을 알고리즘에 직접 새겨넣는 작업이 아니다. 시몽동의 발생적 관점에서 볼 때, 이 과정은 대규모 사전 학습 단계의 데이터 선별 역사, 신경망 아키텍처의 구체화, 그리고 미세조정이라는 기술적 계통 속에서 특정 시스템이 형성되어 온 통시적 조건의 일부이다. 동시에 라투르의 매개적 관점에서 보면, '유해한 답변을 지양하라'는 규범적 요구는 외주 라벨러의 평가

www.kci.go.kr

가이드라인, 보상 모델의 손실 함수, 프롬프트 템플릿, 그리고 배포의 안전 필터링 시스템이라는 이질적 행위소들의 연결망을 거치며 '우회'되고 '위임'된다. 따라서 LLM의 윤리적 문제는 사후적인 설명가능성 검증이나 일회성 체크리스트에 그칠 것이 아니라, 사전 학습 데이터의 생성 계통부터 피드백 보상 체계의 번역 경로, 그리고 실시간 배포 환경의 상호작용에 이르는 생애주기 전체를 가시화하고 지속적으로 재검토하는 제도적 거버넌스로 구현될 필요가 있다.

오늘날 AI 윤리는 공정성·투명성·책임성 같은 원칙을 선언적으로 제시하거나 규제와 법규 준수의 문제로 축소되는 경향이 있다. 그러나 시몽동과 라투르의 관점에서 보면, 이러한 접근은 윤리적 문제가 발생하는 존재론적 조건을 충분히 포착하지 못한다. AI 시스템은 단일한 기술 객체가 아니라 데이터 생산 관행·알고리즘 설계·조직의 목표·경제적 계산·사용자 행태가 결합된 기술적 앙상블이자 매개의 연쇄다. 윤리적 실패는 특정 알고리즘의 결함이나 개인의 악의에서 비롯되기보다, 기술적 구성 속에서 목적과 수단, 의미와 계산의 관계가 왜곡될 때 발생한다. 그리고 그 왜곡은 기술이 완성된 이후가 아니라 설계되고 배치되는 과정에서 이미 시작된다. 따라서 이들의 관점에 따르자면, 기술적 대상의 발생 조건과 계통을 이해하고, 목적·가치·행위·책임의 번역과 위임 경로를 추적하며, 이러한 경로를 지속적으로 다시 판단할 수 있는 제도적 공간 확보가 AI 윤리에 요구된다. 이는 공정성, 투명성, 설명가능성, 책임성이라는 기존 원칙들이 실제 기술적 맥락에서 작동하기 위한 적용 조건을 확장하는 것이기도 하다. 기술윤리는 기술을 인간의 외부에 두고 통제할 것인가의 문제가 아니라, 기술과 함께 형성되는 인간과 사회를 어떻게 사유하고 조직할 것인가의 문제다. AI 윤리 역시 '기술을 통제·관리하는 윤리'로부터 '기술과 함께 어떤 공동체를 구성할 것인가'를 묻는 윤리로의 전환 위에서 실질적인 의미를 가질 수 있을 것이다.

www.kci.go.kr

참고문헌

김재희 (2017), 『시몽동의 기술철학: 포스트휴먼 사회의 청사진』. 아카넷.

브뤼노 라투르 (2023), 『존재 양식의 탐구』, 황장진(역), 사월의책.

질베르 시몽동 (2011), 『기술적 대상들의 존재양식에 대하여』, 김재희(역), 그린비.

Gianni, Robert (2026), "AI ethics after generative AI: from consequences to effects," *AI & SOCIETY*, Advance online publication.

Jobin, Anna, Marcello Ienca, Effy Vayena (2019), "The Global Landscape of AI Ethics Guidelines," *Nature Machine Intelligence*, 1, pp. 389~399.

Latour, Bruno (2002), "Morality and Technology: The End of the Means," Theory, *Culture & Society*, 19(5-6).

Mittelstadt, Brent (2019), "Principles Alone Cannot Guarantee Ethical AI," *Nature Machine Intelligence*, 1, pp. 501~507.

Simondon, Gilbert (1961/2014), "La mentalité technique," *Sur la technique(1953-1983)*, PUF.

Spiegel, Irina (2024), "Criticizing Ethics According to Artificial Intelligence," arXiv, arXiv:2408.04609.

White, Joel (2026), "The Questions Concerning the Ethics of AI," *Triple Ampersand Journal (&&&)*. (https://tripleampersand.org/the-questions-concerning-the-ethics-of-ai/)

과학기술정보통신부, 「대한민국 인공지능 윤리원칙」, 인공지능 윤리 소통채널, 2026. (https://ai.kisdi.re.kr/aieth/main/contents.do?menuNo=400054, 2026. 9. 2. 접속)

European Parliament and Council of the European Union (2024), Regulation (EU) 2024/1689 (Artificial Intelligence Act), Official Journal of the European Union, 12 July. (https://eur-lex.europa.eu/eli/reg/2024/1689/oj, 2026. 8. 31. 접속)

Tabassi, Elham (2023), "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," National Institute of Standards and Technology(NIST) AI 100-1. (https://doi.org/10.6028/NIST.AI.100-1, 2026. 8. 31. 접속)

UNESCO (2021), Recommendation on the Ethics of Artificial Intelligence, UNESCO. (https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence, 2026. 8. 31. 접속)

www.kci.go.kr

Abstract

Beyond Instrumentalism -
A Paradigm Shift in AI Ethics through Simondon and Latour

Kim, Jae-Hee (Eulji University, Faculty of Liberal Arts, Assistant Professor)

This paper critiques the instrumentalist limitations of contemporary AI ethics—which currently relies heavily on declarative and ex-post regulatory models—and reconstructs a dynamic model of AI ethics by integrating the technological philosophies of Gilbert Simondon and Bruno Latour. Simondon's theory of ontogenesis conceptualizes AI as a processual individual undergoing concretization within an associated milieu, calling for the cultural establishment of a technical culture that mediates between technical practices and social values. Furthermore, Latour's theory of mediation views AI as a network of detours and delegations that translates goals and actions, proposing moral repetition —which introduces friction into black-boxed chains of action—as a concrete ethical practice. The convergence of these two perspectives shifts AI ethics away from the static application of norms. Instead, it transforms AI ethics into a dynamic practice of governance that visualizes, continuously monitors, and revises the genetic lineages and delegation pathways throughout the entire lifecycle of the technology.

Keywords: AI Ethics, Instrumentalism, Concretization, Associated Milieu, Technical culture, Detours, Delegations, Technical Mediation, Moral Repetition, Simondon, Latour

논문 투고일: 2026년 09월 07일
심사 완료일: 2026년 10월 05일
게재 확정일: 2026년 10월 15일

www.kci.go.kr