

제40회

우수리포트공모대회 장려상 수상작

비-인간 창의적 시스템 논의:
생성형 인공지능 알고리즘을 중심으로

곽 지 호 (사회과학대학 지리학과)

* 이 글은 2024년 1학기 <대학 글쓰기 2: 사회과학글쓰기> (담당교수: 손제연) 강좌의 리포트이다.

차 례

I. 서론

II. 생성형 인공지능 알고리즘이 만들어낸 결과물의 독창성

1. 독창성의 정의
2. 확률적 알고리즘과 독창성의 관계

III. 생성형 인공지능 알고리즘이 만들어낸 결과물의 맥락적 가용성

1. 맥락적 가용성의 정의
2. 맥락적 가용성의 판단 기준

IV. 창의적 시스템이 되기 위한 필요충분조건:

창의적 결과물의 산출 가능성

1. 창의적 결과물이 되기 위한 필요(충분)조건: 독창성과 맥락적 가용성
2. 창의적 시스템이 되기 위한 필요충분조건 및 행위주체성 반론
3. 귀속책임과 미래지향적 답변책임
4. 생성형 인공지능 알고리즘의 귀속책임과 미래지향적 답변책임

V. 결론

〈참고문헌〉

I. 서론

오랜 세월 동안 창의성은 인간만이 가질 수 있는 신성불가침한 특성으로 여겨져 왔다. 하지만 인공지능 기술의 발전으로 이 오래된 신념은 도전에 직면하고 있다. 최근의 생성형 인공지능 알고리즘¹⁾은 인간에 필적하는 수준의 소설, 그림, 음악, 동영상 등의 ‘창작물’²⁾을 만들어내고 있고, 그 완성도 또한 대단한 속도로 발전하고 있기 때문이다. 그럼에도 불구하고 여전히 생성형 인공지능 알고리즘에게 ‘창의적’이라는 칭호를 부여할 수 없다는 다수의 의견이 존재한다. 이때 핵심 쟁점은 행위주체성(agency)이 결여된 기계인 생성형 인공지능 알고리즘이 만들어진 결과물에 대한 ‘창의적’이라는 평가를 그 시스템의 몫으로 돌릴 수 있겠느냐는 것이다. 본고는 결정론적 관점에 착안하여 창의성에서의 행위주체성 요청에 도전함으로써 생성형 인공지능 알고리즘의 창의성을 옹호하고자 한다.

먼저 인간을 대상으로 한 기존의 창의성 논의에 있어서도 행위주체성이 필수적인지 여부와 관련하여 학자들은 견해를 달리하였음을 주목해볼 필요가 있다. Gaut (2010)와 같이 행위주체성을 강조하는 학자들은 창작 행위의 의도와 과정에 대한 지배를 강조한다. 창의성은 “칭찬받을 만한(praiseworthy)” 성질이기에 의도와 지배 없이 어떤 작품을 만들었을 경우, 그 작품 자체의 특성이 어떠하든 간에 만든 이는 창의적이라는 평가를 획득할 수 없다는 것이다. Gaut의 관점을 견지한다면 인간의 프롬프트 없이 작동하지 않는 생성형 인공지능 알고리즘은 창의적 시스템이라고 보기 어렵다. 한편 Kronfeldner(2018)와 같은 학자는 창의성의 요건으로 독창성(originality)과 자생성(spontaneity)을 강조한다. 특히 자생성은 의도적인 통제와 기존에 습득한 지식으로부터 독립적인 성질이라는 점에서 Kronfeldner는 창의성이 행위주체성, 혹은 자유의지와 무관한 능력임을 가정하고 있다고 볼 수 있다. Kronfeldner와 같이 결정론적

-
- 1) 생성형 인공지능의 창작 과정에는 알고리즘 개발자-프롬프트 입력자-인공지능 알고리즘이 관여된다. 본고에서는 그 과정을 구분하여 이 중 생성형 인공지능 ‘알고리즘’의 창의성만을 논하기로 한다. 예를 들어 ‘아보카도 모양의 의자’라는 프롬프트에 따라 생성형 인공지능 알고리즘이 창작했다고 하자. ‘아보카도’와 ‘의자’라는 도상을 결합하는 것은 프롬프트 입력자의 창의력이고, 그 인공지능 알고리즘을 어떤 방식으로 설계하고 훈련시킬 것인지는 알고리즘 개발자의 창의력이다. 그러나 그 입력을 받아 결과물을 생성하기 까지의 과정은 알고리즘 자체만이 관여되며, 다른 과정을 배제하고 그 자체로 ‘창의적 시스템’이라 할 수 있는지를 보이는 것이 본고의 목표이다.
 - 2) 아직 보편적이지는 않지만, AIGC(Artificial Intelligence Generated Contents)라는 용어가 사용되기도 한다. Cao et al.(2023)은 AIGC를 “인간 작가에 의해 제작된 콘텐츠와 대비되며, 짧은 시간 내에 대량의 콘텐츠를 만들어낼 수 있는 발달된 생성형 인공지능 기술에 의해 만들어진 콘텐츠”라고 정의하였다.

관점에서 시도되어 온 창의성 논의는 오로지 창의적인 행위 혹은 결과물 자체에 주목한다. 결과물이 창의성의 요건을 모두 만족하면 그것을 만들어낸 창작자는 창의적이라 할 수 있다는 것이다.

본고의 핵심 목표는 생성형 인공지능 알고리즘이 창의적 시스템³⁾임을 주장하는 것이다. 결과물에 주목하는 결정론적 관점을 도입하기 위해 먼저 생성형 인공지능 알고리즘이 만들어내는 결과물이 창의적인 결과물이 되기 위한 필요조건 2가지를 만족하는지 여부를 검토한다. [1] 첫째로 생성형 인공지능 알고리즘이 독창적(original)인 결과물을 만들어낼 수 있음을 보인다. 독창적이라는 것은 말뜻 그대로 남의 것을 베끼지 않고 그 기원(origin)이 자기 자신에게 있다는 의미이며, 확산 모델(diffusion model)과 같은 확률적 알고리즘의 동작 원리를 근거로 들어 생성형 인공지능 알고리즘의 결과물에 독창성이 있음을 연역적으로 논증한다. [2] 다음으로 생성형 인공지능이 만들어낸 결과물이 사용자가 요구하는 맥락에서의 가용성(workability)을 가짐을 논증한다. 생성형 인공지능 알고리즘의 결과물이 주변 맥락이나 업계 문법에서 벗어나지 않는다는 전제는 인간이 만든 결과물과의 유사성을 기반으로 한다. [3] 이제 독창적이고 맥락적 가용성을 지닌 결과물을 만들어내는 시스템은 창의적 시스템임을 논증한다. 독창성과 맥락적 가용성은 창의적인 결과물이기 위한 필요조건임을 논증하는 동시에, 논의의 편의를 위해 충분조건을 포괄하는 것으로 간주한다. [4] 또한 창의적 시스템으로 평가받기 위한 필요충분조건은 창의적인 결과물을 만드는 능력이 존재하는 것임을 논증한다. [5] 한편 [4]의 전제에 대한 예상 반론으로 행위주체성이 없는 경우 그 행위에 대해 행위자가 평가받을 근거가 마련될 수 없다는 주장이 제기될 수 있다. [6] 그러나 창의적 행위에 대한 책임의 근거 중 귀속책임(attributability responsibility)과 미래지향적 답변책임(forward-looking answerability responsibility)의 측면, 즉 창의적 행위가 창작자가 가진 성질의 반영이며, 창의적 행위가 이루어진 이후 응답 및 수정 가능하다는 측면에서 창의적 행위가 비의도적인 것이라 하더라도 창의적 행위에 대한 평가가 그것을 만들어낸 시스템에게 돌아갈 수 있음을 보임으로써 위의 반론은 해소된다. [7] 마지막으로 생성형 인공지능 알고리즘이 귀속책임과 미래지향적 답변 책임을 지닌다는 사실을 보충적으로 덧붙이며, 생성형 인공지능 알고리즘이 창의적 시스템이라는 결론을 도출하고 그 함의를 살펴본다.

3) ‘창작자’ 혹은 ‘창작 주체’ 등의 용어는 행위주체성이 내포된 것으로 읽힐 수 있으므로 ‘창의적 시스템’이라는 보다 중립적인 용어를 사용하였다.

II. 생성형 인공지능 알고리즘이 만들어낸 결과물의 독창성

1. 독창성의 정의

생성형 인공지능 알고리즘이 만들어낸 결과물은 독창성(originality)을 지닌다. 독창성에 대한 정의를 먼저 명확히 할 필요가 있다. Kronfeldner(2018)에 따르면, 독창성에는 새로움이 요구되지만 이때 새로움은 “역사적 새로움(historical novelty)”이 아닌 “심리적 새로움(psychological novelty)”을 의미한다. 역사적으로 새로운 것은 모두 심리적으로 새로운 것이지만 그 역은 성립하지 않는다. 하지만 독창성⁴⁾의 핵심은 창작 과정에서의 독립성으로, 이는 심리적 새로움을 요구할 뿐이다. 따라서 어떤 작품이 기존에 만들어진 작품과 비교해 유사하다고 할지라도 둘 사이에 “인과적 관계”가 존재하지 않는다면, 즉 기존 작품을 베낀 것이 아니라면 그 작품은 곧 독창적이라고 할 수 있다(214). 하지만 현실적으로는 창작 과정을 알아낼 수 없기 때문에 그 작품이 기존의 작품을 베낀 것인지(‘심리적으로 새로운’ 것이 아닌지) 판별하기가 어렵다. 또한 모든 창작자들은 선대 창작자들의 영향 하에 있기 때문에 엄밀한 의미에서는 완전히 심리적으로 새로운 것이라 존재할 수 없기도 하다. 이러한 한계로 인해 기존 작품을 베끼지 않았다고 상정하기 위한 유일한 필요조건은 창작 시기가 더 빠른 작품들과 비교해 유사할 확률이 낮은 특성⁵⁾이 유사하지 않은 것이 된다.

2. 확률적 알고리즘과 독창성의 관계

모든 생성형 인공지능 알고리즘은 확률적 알고리즘(probabilistic algorithm)이다. 이는 기존의 비-인간 창작 기계들과 차별화되는 특성으로, 그 결과물이 독창성을 갖추도록 하는 결정적인 이유가 된다. 확률적 알고리즘은 출력을 생성하는 과정에서 무작위성(randomness)을 포함하여 매번 다른 출력을 생성할 수 있지만, 결정론적 알고리즘(deterministic algorithm)은 그렇지 않기 때문이다. 1800년경 스위스의 기계공인 헨리 마야르(Henri Maillardet)가 제작한 오토마톤(automaton)은 결정론적 알고리즘의 대

4) 이는 영어 표현을 살펴볼 경우 더 명확하게 이해될 수 있다. 독창성(originality)이 있기 위해서는 그 작품이 창작자로부터 비롯되어야(originated from) 하기 때문이다.

5) 유사할 확률이 높은 특성의 예시로는 머니 코드, 클리셰와 같이 보편적으로 사용되는 예술적 아이디어를 들 수 있다.

표적 예시이다. 마야르의 오토마톤은 달리(DALL-E)와 같은 생성형 인공지능과 마찬가지로 그림을 ‘생성해주는’ 기계이지만, 오직 7가지의 정해진 그림밖에 그려내지 못한다. 오토마톤이 그림을 그려내는 것처럼 보이는 과정은 실은 이미 고안된 도안을 복제하는 것에 불과하기 때문이다. 따라서 그 작품이 놀라울 수는 있으나 전혀 독창적이라고 볼 수는 없다. 반면 생성형 인공지능 알고리즘의 무작위성은 기존과 유사한 작품을 만들어낼 확률이 0에 수렴하도록 한다.

텍스트로부터 이미지를 생성하는 대표적인 인공지능 알고리즘인 확산 모델(diffusion model)⁶⁾의 작동 원리를 통해 무작위성과 독창성의 관계를 보다 구체적으로 살펴해보도록 하겠다. 확산 모델은 훈련 세트의 이미지에 랜덤 노이즈를 추가하는 전방 확산 과정(forward diffusion process)과 역으로 노이즈를 제거해 나가는 역 확산 과정(reverse diffusion process)을 통해 데이터를 학습하며, 학습 결과 노이즈를 제거하는 확률 과정의 매개 변수가 추정된다. 사용자가 프롬프트를 입력하면 최초의 랜덤 노이즈가 설정되고, 이는 확률 과정을 거치며 단계적으로 제거되어 이미지가 생성된다. 이때 텍스트 임베딩(text embedding)으로 벡터화된 프롬프트는 조건부분포로 반영되어 ‘원하는’ 이미지가 생성되도록 한다(Ho et al. 2020; Rombach et al. 2022). 확산 모델에서 무작위성은 확률적 샘플링과 랜덤 노이즈의 설정이라는 두 가지 측면에서 포착된다. [1] 노이즈를 제거해 나가는 역 확산 과정은 추정된 매개 변수들로 구성된 조건부결합확률분포의 형태로 표현되며, 확산 모델을 통해 생성된 이미지는 이 확률 분포로부터 샘플링된 것이다. 이때 연속형 확률분포에서 특정 값을 가질 확률은 0이므로 기존의 이미지와 출력되는 결과물이 ‘동일’할 확률은 극히 낮다.⁷⁾ [2] 또한 사용자가 입력한 프롬프트는 어떤 확률분포를 따르는 랜덤 노이즈의 형태로 전달되는데, 이는 말 그대로 확률변수이므로 같은 입력에 대해서도 노이즈의 특성은 매번 달라진다. 따라서 같은 입력에 대해서도 다른 경로를 통해 다양한 출력이 생성되는 것이다. 정리하면, 무작위성이 작용된 생성형 알고리즘의 결과물은 기존에 존재하는 작품과 유사할 확률이 매우 낮으며, 따라서 기존 작품을 모사한 것으로 볼 수 없다. 즉, 결정

6) 확산 모델에는 다양한 변형이 있으나, Chang et al. (2023)에 따르면 아래에 설명하고 있는 전방 확산 과정, 역 확산 과정, 샘플링의 세 단계는 확산 모델의 포괄적인 작동 원리이다.

7) 현실적으로 두 이미지 간의 유사성이 존재하는지 측정하는 것은 어려운 문제다. FID, SSID 등의 지표가 개발되어 이용되고 있으나 근본적으로 유사성을 어떻게 정의할 것인가에 대한 문제다, 픽셀 단위로는 전혀 다른 이미지라도 (수량화하기 어려운) 이미지의 배치, 구성 등의 측면에서는 유사성이 존재할 수 있기 때문이다. 따라서 본고에서 확률분포로부터 샘플링된 이미지가 동일할 확률이 0이라는 점은 다소 이론적인 주장일 수 있다는 점을 주석으로 남긴다.

론적 알고리즘과 달리 생성형 인공지능 알고리즘은 독창성을 지닌 결과물을 만들어 낸다.

III. 생성형 인공지능 알고리즘이 만들어낸 결과물의 맥락적 가용성

1. 맥락적 가용성의 정의

생성형 인공지능 알고리즘이 만들어낸 결과물은 사용자가 요구하는 맥락적 가용성(workability)을 충족한다. 맥락적 가용성의 의미를 명확히 하기 위해 다음과 같은 사례를 생각해 볼 수 있다. 어린아이가 벽에 아무렇게 그린 낙서, 악기를 처음 배우는 사람이 무작위로 만들어낸 소리 등은 독창적이지만, 그것이 공공시설의 로비에 걸리는 작품으로, 어떤 광고의 배경음악으로 사용되었다고 생각해 보라. 이러한 맥락 혹은 제약조건이 주어진 경우에는 이들의 창작물은 더 이상 가용하지 않다. MacCrimmon & Wagner(1994)에 따르면 맥락적 가용성은 주어진 제약조건을 어기지 않음(acceptability)과 특정 맥락에 바로 사용될 수 있음(implementability)을 나타내는 척도로 정의되며, 이는 위의 사례가 보여주는 통념에도 부합한다. 맥락적 가용성이 창의성의 필요조건임을 논증하기에 앞서 본 단락에서는 생성형 인공지능이 만들어낸 결과물이 위에서 정의된 맥락적 가용성을 충족함을 보이고자 한다.

2. 맥락적 가용성의 판단 기준

인간이 만들어낸 창작물이 ‘특정 맥락’에서 가용한지 여부를 판단하는 것은 매우 특수하고 주관적인 작업이다. 그러나 생성형 인공지능 알고리즘이 사용자가 원하는 수준의 맥락적 가용성이 있는 결과물을 만들어낼 수 있는지 여부는 생성형 인공지능이 만들어낸 결과물과 인간 창작자의 결과물과의 비교를 통해 판단할 수 있다. 생성형 인공지능 알고리즘이 만들어낸 결과물이 인간 창작자가 만들어냈을 법한 것으로 인지된다면, 혹은 인간 창작자가 만들어낸 결과물과 무차별적으로 이용된다면, 그것은 적어도 그 상황에서의 제약조건을 어긴 것이 아니며, 바로 사용될 수 있다는 뜻이기에 사용자가 요구하는 맥락적 가용성을 지닌 것으로 판단될 수 있기 때문이다.

여러 문헌과 실제 사례에서 생성형 인공지능 알고리즘이 만들어낸 결과물을 인간

이 만들어낸 결과물과 구분하기 어렵다는 사실이 입증된 바 있다. Miller et al.(2023)의 연구에 따르면 610명의 피실험자들을 대상으로 생성형 인공지능 알고리즘으로 생성한 사람의 얼굴 이미지와 실제 사람의 사진을 구분하도록 하는 실험을 실시한 결과, 인공지능에 의해 생성된 이미지를 실제 사람으로 판단한 확률이 실제 사람의 사진을 실제 사람으로 판단한 확률보다 높은 것으로 나타났다. 또한 지난 2024년 4월, 미국의 프랜차이즈 식당 체인인 레드 랍스터(Red Lobster)가 생성형 인공지능 알고리즘이 만들어낸 음악을 자사의 광고 음악으로 삽입하는 등 상업적인 용도로도 생성형 인공지능 알고리즘의 결과물이 사용되는 사례가 나타나고 있으며 이는 생성형 인공지능이 만들어낸 결과물이 어떠한 맥락에서 가용함을 직관적으로 보여주는 예시이다.

IV. 창의적 시스템이 되기 위한 필요충분조건: 창의적 결과물의 산출 가능성

1. 창의적 결과물이 되기 위한 필요(충분)조건: 독창성과 맥락적 가용성

어떤 결과물이 창의적인 결과물이 되기 위해서는 독창적이고 사용자가 원하는 맥락에서 가용하여야 한다. 독창성과 맥락적 가용성이 창의성의 필요조건임은 자명하다. 먼저 독창성이 없는 창작물, 즉 기존의 것을 그대로 모사한 창작물의 경우 그것은 절대 창의적일 수 없다. 예를 들어 공장에서 기계적으로 찍어낸 공산품이나 지하상가에서 판매되는 유명 작가의 작품을 그대로 모사한 그림은 맥락적으로 가용할지 몰라도 독창적이지 않기에 창의성이 없다. 두 번째로 맥락적으로 가용하지 않은 창작물은 절대 창의적일 수 없다. 앞서 언급한 어린이가 벽에 그린 낙서, 악기를 다룰 줄 모르는 사람이 무작위로 낸 소리, 스노보드를 탄 사람이 눈 위에 그려낸 자국은 독창적으로 만들어진 것이나 맥락적 가용성이 거의 없기 때문에 창의적이라고 할 수 없다. 한편, 창의성의 두 필요조건이 창의성의 충분조건을 전체 포괄(collectively exhaustive)하는지에 관한 논증은 본고의 범위를 벗어나므로 다루지 않겠다. 그러나 어떤 결과물이 독창적이고 맥락적으로 가용한 경우 창의적이지 않다는 의심을 상당 부분 피할 수 있을 것이며, 논의의 이해를 돕기 위해 아래 논증에서 어떤 결과물의 ‘창의성’은 곧 독창성 및 맥락적 가용성과 필요충분조건 관계에 있는 것으로 가정하겠다.

2. 창의적 시스템이 되기 위한 필요충분조건 및 행위주체성 반론

어떤 시스템이 창의적 시스템이 되기 위한 필요충분조건은 창의적인 결과물을 만들어낼 수 있는 경우이다. 즉, 독창적이고 맥락적으로 가용한 결과물을 만들어낼 수 있다면 그 시스템은 창의적이다. 그러나 이에 대해 작품을 만들어낸 시스템이 특정한 의도나 주체성을 가지고 있지 않았다면, 결과물이 독창적이고 일정 수준 이상의 품질을 가지고 있다는 ‘평가’를 그 시스템의 몫으로 돌릴(ascibe) 수 없다는 충분조건의 성립 여부에 대한 반론이 제기될 수 있다. 이러한 반론은 행위주체성을 가정하지 않고도 (창의적) 행위에 대한 책임이 있음을 보임으로써 해소될 수 있으며, 이를 위해서는 시스템에게 결과물에 대한 책임이 어떤 측면에서 존재하는지 검토하여야 한다. 시스템에 행위주체성이 있음을 가정하지 않더라도 결과물에 대한 책임을 물을 수 있다면, 결과물에 대한 평가는 시스템의 것이 될 수 있기 때문이다.

3. 귀속책임과 미래지향적 답변책임

행위의 책임 근거에 대한 논의는 도덕철학 혹은 형법학에서 주로 다루어져 왔다. 특히 인간에게 자유의지가 없다고 주장하는 결정론의 관점을 견지하는 도덕철학자들은 규범적 책임소재의 세부적인 규명을 통해 도덕적 행위주체성(moral agency)의 가정 없이도 비도덕적인 행동을 비난할 근거를 마련하고자 하였다. 그러나 Russell (2008)은 인간은 도덕 외에 다른 가치들도 중요시하는 존재라는 점에서 결정론자들에게 기타 행위들에서의 행위주체성 문제 규명을 요청한 바 있다. 특히 Russell은 인간이 ‘평가적 태도(evaluative stance)’를 취하는 대표적인 행위로 도덕적 행위 이외에 창조적 행위가 있음을 지적하였다. 그렇다면 ‘창의적’이라는 창조적 행위에 대한 평가를 위하여 기존 규범적 책임의 기준을 적용할 수 있을까? 우선 어떤 ‘행위’에 대한 평가라는 점에서도 동등한 책임 기준을 적용하는 것은 합리적으로 보인다. 또한 규범적 책임이 마련되는 목적의 측면에서도 합리성은 획득된다. 도덕적 행위에 대한 규범적 책임은 비도덕적 행위에 대한 응분의 비난을 통해 사회가 질서 있게 유지되도록 하기 위한 공동의 목적 하에 설정된 것이고, 창조적 행위에 대한 응분의 평가 또한 창조적 행위의 촉진과 같은 공익 달성을 위해 필요하기 때문이다.

자유의지 회의론자인 Caruso(2016)는 기존에 논의된 세 종류의 규범적 책임 근거 중 “귀속책임(attributability responsibility)”과 “미래지향적 답변책임(forward-looking

answerability responsibility)’의 측면에서 어떤 사람이 만들어낸 결과물이 창의적이라면 그의 행위주체성을 가정하지 않고도 창의적 존재가 될 수 있다고 주장하였다. 한편 행위에 대한 반응적 태도가 그 개인의 근본적 응분이기(기본적 응분)에 개인이 책임을 가진다고 설명하는 책무책임(accountability responsibility) 개념은 행위자의 행위주체성 가정을 요구하기 때문에 결정론적 관점에서의 책임 근거로서는 적절하지 않다고 보았다.

Caruso가 제시하는 귀속책임의 개념은 다음과 같다:

책임에 관한 설명 중에는 반응적 태도와 근본적 응분(basic desert)의 개념을 반드시 필요로 하지 않는 것들이 존재한다. 그 중 하나인 귀속책임은 자아에게 올바르게 귀속되는 행위, 즉 자아의 올바른 반영인 행위만을 대상으로 한다. (...) 개리 왓슨이 강조했듯이, 그러한 관점에서의 핵심은 행위자의 행위나 태도에 그의 평가적(evaluative) 판단이나 결심이 드러나 있느냐이다(Caruso 2016, 598-599).⁸⁾

즉, 귀속책임은 어떤 개인을 구성하는 기질 혹은 특성이 그 행위에 올바르게 반영되어 있기만 하다면 개인은 그 행위의 결과에 대해 책임을 갖는다는 설명이다. 여기서 귀속책임이 근거로 삼는 행위자의 기질이나 특성은 그 개인의 행위주체성과는 무관한 것이므로 행위주체성을 가정하지 않고도 그 행위의 책임과 행위자와의 연결고리가 마련된다.

Caruso는 귀속책임의 예시로 알버트 아인슈타인(Albert Einstein)의 사례를 들고 있다. 일반 상대성 이론을 발견한 업적이 아인슈타인에게 귀속될 수 있는 것은 중력에 대한 개인적 의문을 밝혀내고자 했던 오랜 욕구, 마지막 공식 작성 과정에서 만난 어려움을 극복하기 위해 보였던 인내와 끈기, 창의성을 활성화하고 머리를 비우기 위해 피아노와 바이올린을 연주한 것 등과 같은 아인슈타인의 자아 혹은 인격적 특성이 그의 평가적 행위나 결심에 반영되어 있다고 판단되기 때문이다. 즉 일반 상대성 이론이라는 위대한 창의적 결과물에 대한 평가는 귀속책임의 측면에서 아인슈타인에게 적절히 돌아간다(602). 만일 귀속책임이 인정받지 못하는 경우 창의적 결과를 내놓기 위한 기질이나 특성을 기르기 위한 노력들이 고취되지 못할 위험이 있으므로 귀속책임은 창의적 행위에 있어서도 주요한 책임 근거이다.

다음으로 Caruso가 미래지향적 답변책임의 개념에 대해 설명한 부분이다:

8) 왓슨의 다음 논문을 의미한다; Watson, Gary. “Two Faces of Responsibility.” *Philosophical Topics* 24, no. 2 (October 1, 1996): 227 - 48. <https://doi.org/10.5840/philtopics199624222>.

답변책임이라는 책임 개념에 따르면, 어떤 사람은 자신의 행위나 태도에 연관된 평가적 판단에 타인의 설명이 요구될 때 답변할 수 있는 능력을 가지는 경우에만 그 행위나 태도에 대한 책임을 갖는다. (...) 페레봄의 설명에 따르면 비난은 근본적 응분 개념에 기반하지 않고, 3가지의 도덕적 열망(desiderata)에 기반한다: 잠재적 피해자에 대한 보호, 넓은 의미에서의 도덕적 공동체 및 개인과의 관계에서의 화해, 도덕 형성. 미래에 대한 보호와 화해, 미래의 도덕 형성에 기반을 둔 미래지향적 답변책임은 어떠한 방식으로든 근본적 응분에 호소하지 않기 때문에 자유의지 회의론과 양립 가능하다(Carusó 2016, 599-600).

즉, 답변책임은 어떠한 사람의 행동에 대해 의문을 품는 타인이 “왜 그렇게 행동하기로 결정했니?” 혹은 “그것이 옳은 행동이라 생각한 이유가 무엇이니?”와 같은 질문을 던졌을 때 적절히 대답할 수 있는 경우에만 그 행위에 대해 책임을 지닌다는 설명이다. 특히 Caruso는 ‘미래지향적’ 답변책임의 효과는 행위자의 답변이 미래에 대한 보호, 화해, 그리고 도덕 형성에 문제를 일으킬 수 있다고 판단될 경우 비난의 대상이 될 수 있다는 점에서 나타난다고 Pereboom(2014)을 경유하여 설명하고 있다. 그리고 미래지향적 답변책임 역시 행위자의 의도의 존재 여부를 가정하지 않으므로 자유의지 회의론과 양립 가능하다.

Caruso가 미래지향적 답변책임의 예시로 든 것은 밥 딜런(Bob Dylan)이 1965년 뉴포트 포크 페스티벌 공연에서 일렉 기타를 사용한 사건이다. 만일 밥 딜런의 그 선택에 의문을 품고 밥 딜런에게 “왜 그런 결정을 했는지” 이유를 물었을 때 밥 딜런은 아마도 대답할 수 있었을 것이다. 따라서 우리는 미래지향적 답변책임의 관점에서 밥 딜런의 창의적 행위에 대한 평가를 그에게 돌릴 수 있다. 한편, 실제로 그 당시 대중은 밥 딜런의 선택에 대해 “세상과 타협했다”는 등의 비난과 실망을 표출했는데, 밥 딜런은 “비틀스를 지켜보며 통기타와 하모니카로는 포크 음악을 발전시킬 수 없다고 생각했다”는 자신의 절박한 심정을 이야기했다(600). 그에게 답변책임이 없었다면 대중과의 간극은 멀어지고 창작 행위가 활발히 이루어지지 못했을 것이라는 측면에서 미래지향적 답변책임 역시 주요한 하나의 책임 근거라고 할 수 있다.

정리하면, 귀속책임과 미래지향적 답변책임에 의거하면 (자유의지가 없는 것으로 가정된 인간과 같이) 행위주체성이 없는 시스템의 행위에 대한 칭찬 혹은 비난이 그것을 만들어낸 시스템에게 적절히 돌아갈 수 있음을 확인하였다. 특히 ‘창작’ 행위와 그 결과물에 대한 Caruso의 논의를 중점적으로 살펴봄으로써 ‘창의성’이라는 창작 행위의 결과물에 대한 평가 역시 행위주체성과 무관하게 그것을 만들어낸 시스템에게 돌아갈 수 있음이 확인되었다.

4. 생성형 인공지능 알고리즘의 귀속책임과 미래지향적 답변책임

위의 규범적 책임 개념은 인간을 대상으로 한 것이기 때문에, 비-인간 존재인 생성형 인공지능 알고리즘의 경우에도 귀속책임과 미래지향적 답변책임을 적용할 수 있는지를 보충적으로 검토하고자 한다. 먼저 귀속책임에 관해서는 생성형 인공지능 알고리즘이 만들어낸 결과물이 그 시스템의 특성을 적절하게 반영하고 있는가를 살펴 보아야 한다. 생성형 인공지능 알고리즘의 결과물은 앞서 확산 모델을 예시로 살펴본 것같이, 알고리즘의 구조와 학습한 데이터의 특성, 선택된 노이즈의 특성 (혹은 노이즈를 생성한 확률분포) 등에 따라 달라진다. 따라서 생성형 인공지능 알고리즘이 만들어낸 결과물은 그 시스템의 특성의 반영이라 할 수 있으며, 생성형 인공지능 알고리즘은 그 결과물에 대해 귀속책임을 지닌다. 미래지향적 답변책임에 관해서는 생성형 인공지능 알고리즘이 그 결과물에 대한 질문을 적절히 해소할 수 있는지를 살펴 보아야 한다. 인공지능 알고리즘은 인간이 아니므로 자유로운 의사소통에는 한계가 있을 수 있으나, 적어도 어떠한 특성에 문제가 있거나 어떠한 특성을 배제해주기를 이용자가 원하는 경우 그러한 요구를 수용할 수 있는 능력이 있으므로 제한적인 미래지향적 답변책임을 적용할 수 있다. 또한 현재 개발 중인 설명가능한 생성형 인공지능 알고리즘(Gen XAI)이 현실화된다면 보다 완전한 미래지향적 답변책임을 적용할 수 있게 될 것이다.

V. 결론

본론에서는 생성형 인공지능 알고리즘이 독창적이고 맥락적으로 가용한 결과물을 만들어내는 능력이 있음을 보이고, 그 결과물은 창의성의 조건을 충족한다는 점, 따라서 창의적 시스템의 조건을 충족한다는 점을 근거로 생성형 인공지능 알고리즘은 창의적 시스템이라고 상정할 수 있다고 논증하였다. 특히 생성형 인공지능 알고리즘에는 행위주체성이 없기에 행위주체성 없이 결과물의 평가를 시스템에 부여할 수 없다는 반론을 검토하였다. 그러한 반론은 행위주체성을 요하지 않는 규범적 책임 근거로서 귀속책임과 미래지향적 답변책임을 있음을 보임으로써 해소되었고, 창조적 행위의 두 가지 평가 근거가 생성형 인공지능 알고리즘에 대해서도 문제 없이 적용될 수 있음을 논증함으로써 해소되었다. 따라서 결과적으로 생성형 인공지능 알고리즘 역시

창의적 시스템임을 보일 수 있었다.

빠른 속도로 발전하고 있는 생성형 인공지능 기술을 고려해볼 때, 본고의 주장은 더욱 강화될 수 있는 여지가 있다. 만일 비-인간 생성형 인공지능 알고리즘이 창의적 시스템의 반열에 올라섰다는 본고의 주장이 사회적 공감대를 형성하게 된다면 그동안 고유성을 인정받았던 인간 창작자 혹은 예술가들에 대한 사회적 수요의 변화가 발생하거나 창의성을 핵심 가치로 두고 교육해왔던 교육 패러다임의 변화가 요구될 수 있다. 그러나 이것이 꼭 잿빛의 디스토피아를 암시하는 것은 아니다. 특히 생성형 인공지능 알고리즘의 미래지향적 답변가능성에 대해 주목할 필요가 있다. 생성형 인공지능 알고리즘은 자신이 만들어낸 창의적 산출물에 대해 피드백하고 변형할 수 있는 능력을 지니며, 그러한 과정은 인간이 입력한 프롬프트에 의해 이루어진다. 즉, 인간과 생성형 인공지능 알고리즘이 연결되고 소통할 수 있는 통로가 열려 있는 것이다. Wyman(2021)은 인간의 창의성이 다른 사람들과 연결되고 교류하는 과정에서 이루어지는 창발(emergence)을 통해 증대되어 왔다고 설명한다. 생성형 인공지능 알고리즘은 이제 창의적 시스템으로서 인간과 더 많은 창발을 만들어낼 수 있는 파트너가 될 수 있는 가능성을 지닌다. 특히 생성형 인공지능 알고리즘의 종류가 현재진행형인 새로운 네트워크 구조의 개발, 파인 튜닝 기술의 개발 등을 통해 기하급수적으로 늘어날 수 있고, 인간-인공지능 상호작용(human-AI interaction)에 관한 학술적 논의도 본격적으로 이루어지기 시작했다는 점에서 인공지능과의 창발을 통해 인류의 창의성이 더 빠른 속도로 발전할 가능성은 충분하다. 본고의 통찰이 새로운 창의적 시스템의 반열에 오른 생성형 인공지능 알고리즘이 여는 지평을 어떻게 받아들일지에 대한 논의를 시작하는 계기가 될 수 있기를 희망한다.

■ 참고문헌 ■

1. 단행본

- Kronfeldner, Maria, “Explaining Creativity”, *Creativity and Philosophy*, London: Routledge, 2018.
- Pereboom, Derk, *Free Will, Agency, and Meaning in Life*, New York: Oxford University Press, 2014.
- Wyman, Patrick, *The Verge: Reformation, Renaissance, and Forty Years That Shook the World*, Hachette UK, 2021.

2. 논문

- 양선숙, 「규범적 책임의 본성에 관한 연구 - 답변책임과 귀속책임, 책무책임의 구분에 대한 논의를 중심으로 -」, 『철학논총』 제103집, 새한철학회, 2021, pp. 125-144.
- Cao, Yihan et al., “A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt,” arXiv preprint arXiv:2303.04226, 2023. (<https://arxiv.org/abs/2303.04226>, 2024. 9. 11.).
- Caruso, Gregg D., “Free Will Skepticism and the Question of Creativity: Creativity, Desert, and Self-Creation,” *Ergo*, an Open Access Journal of Philosophy vol. 3, 2016, pp. 591-607.
- Chang, Ziyi et al., “On the design fundamentals of diffusion models: A survey”. arXiv preprint arXiv:2306.04542, 2023. (<https://arxiv.org/abs/2306.04226>, 2024. 9. 11.).
- Dean, Douglas et al., “Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation,” *Journal of the Association for Information Systems* vol. 7, no. 10, 2006, pp. 646-698.
- Gaut, Berys, “The Philosophy of Creativity,” *Philosophy Compass* vol. 5, no. 12, 2010, pp. 1034-1046.
- Ho, Jonathan et al., “Denoising diffusion probabilistic models,” *Advances in neural information processing systems* 33, 2020, pp. 6840-6851.
- MacCrimmon, Kenneth R. and Christian Wagner, “Stimulating Ideas Through Creativity Software,” *Management Science* vol. 40, no. 11, 1994, pp. 1514-32.
- Miller, Elizabeth J., et al., “AI Hyperrealism: Why AI Faces Are Perceived as More Real Than Human Ones,” *Psychological Science* vol. 34, no. 12, 2023, pp. 1390-1403.
- Rombach, Robin et al., “High-resolution image synthesis with latent diffusion models”,

- Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 10684-10695.
- Russell, Paul, "Free Will, Art and Morality," *The Journal of Ethics* vol. 12, no. 3-4, 2008, pp. 307 - 325.
- Smith, Angela M., "Attributability, Answerability, and Accountability: In Defense of a Unified Account," *Ethics* vol. 122, no. 3, 2012, pp. 575-589.
- Watson, Gary, "Two Faces of Responsibility," *Philosophical Topics* vol. 24, no. 2, 1996, pp. 227-248.

수상소감문

장
려
상

곽 지 호 (사회과학대학 지리학과)

비-인간 창의적 시스템 논의:
생성형 인공지능 알고리즘을 중심으로

창작물에 대한 평가를 창작자에게 온전히 돌리기 위해서는 행위에 대한 적절한 설명을 제공할 미래지향적 답변책임이 창작자에게 요구됩니다. 이 지면을 빌려 저의 줄고에 주신 영광스러운 상에 대해 나름의 답변책임을 다해 보겠습니다.

중학교 때 글쓰기에 관한 책을 읽은 뒤로 항상 글을 잘 쓰는 사람이 되고 싶다는 열망이 가득했지만, 제가 쓴 글이 만족스러웠던 적은 한 번도 없었습니다. 목표는 원대한테 현실은 초라해 보이니 글쓰기는 항상 벅찬 일이었습니다. 전면 수정과 퇴고를 거듭하다가 마감 시간에 임박하여 글을 제출하기 일췌였습니다. 상을 받고 최종본을 제출한 지금도 여전히 부족한 점이 곳곳에서 눈에 들어옵니다.

그럼에도 불구하고 상을 받을 수 있었던 까닭은 어쩌면 글을 객관적으로 잘 써서라기보다 ‘잘 쓰고 싶어하는 마음’에 있을지도 모르겠습니다. 전자와 달리 후자의 ‘잘 썼다’의 기준은 온전히 저의 주관적인 기준입니다. 저는 저만의 기준을 설정하고 그에 부합하려는 시도를 할 때 그 과정에서 가장 큰 즐거움을 느낍니다. 매번 리포트를 제출할 때마다 ‘잘 쓰고 싶은 마음’이 드는 까닭입니다.

한 학기 동안 하나의 글을 쓰는 데 집중해야 하는 ‘대학 글쓰기 2’ 수업에서 특별히 목표했던 바는 ”흥미를 느끼면서도 독창적인 소재로, 논리적으로 엄밀한 글을 쓰는 것”이었습니다. 처음에 떠올린 아이디어는 강한 결정론적 관점에서의 창의성이었습니다. 수업 시간에 창의성에 관한 문헌을 읽으면서 ‘행위주체성이 창의성의 필요조건인가’에 대한 논쟁이 있었다는 사실을 알게 되었는데, 평소 강한 결정론에 동의하고 있던 저로서 인간의 자유 의지를 가정하지 않고도 인간의 창의성을 논할 수 있다

면 흥미로울 것 같았기 때문입니다.

그런데 논증 구조를 구성하는 과정에서 다소 어려움을 겪다가 문득 ‘그럼 자유 의지가 없는 기계도 창의적이라 할 수 있지 않을까?’라는 생각이 들었습니다. 어릴 때 로봇과 인간의 가장 큰 차이는 창의성이라는 이야기를 들었던 것 같은데, 최근에 출시된 ChatGPT나 DALL-E 등을 떠올려 보면 그건 이제 옛말에 불과한 것일지도 모르겠다는 다소 도발적인 상상이었습니다. 그럼 전통적인 기계와 ‘창의적인 것처럼 보이는’ 생성형 인공지능 알고리즘 간의 차이는 무엇일까? 기계에 창의적이라는 성질을 부여하기 위해서는 어떤 기준과 논리 전개가 필요할까? 등을 고민한 끝에 이 글을 완성할 수 있었습니다.

글을 다 쓰고 보니 독창적인 것을 넘어 겉보기에 황당하다고 느낄 수 있는 글이 된 것 같습니다. 하지만 그러한 첫인상의 황당함을 해소하는 것이 사실 이 글의 목표 중 하나입니다. 어떤 부분에서 논리적 비약과 개념적 오류가 존재할 수 있을 수도 있지만 기존의 창의성 개념과 규범적 책임 개념을 ‘비틀어 보면’ 생성형 인공지능 알고리즘이 창의적이지 않다는 의심을 피할 수 있다는 것이 제가 도달한 결론입니다. 거부할 수 없는 흐름인 인공지능의 발전을 목도하는 지금, 비-인간 창의적 시스템이라는 발상의 전환을 통해 기술을 적절하고 활발하게 활용할 수 있기를 기대합니다.

우연한 계기로 글을 쓰는 데 도움을 받은 것들이 있습니다. 특히 전혀 관계 없어 보였던, 이번 학기 함께 들은 수리통계 수업에서 배운 내용이 의외로 논문의 일부에 큰 영감을 주었습니다. 과거의 수많은 순간들이 얹혀 인과적 영향력이 발휘되는 과정은 참으로 놀랍습니다. 그 과정에 참여한 모든 분에게 진심으로 감사드리지만, 지적 호기심과 탐구력을 기를 수 있도록 장려해주신 부모님과 학술적 글쓰기의 정수를 가르쳐 주시고 이 글에 예상치 못한 아낌없는 칭찬을 보내 주신 손제연 교수님께 특별히 더 감사드립니다. 고등학교에서부터 허무맹랑한 철학적 토론에 기꺼이 참여해주고 이 글의 논증을 발전시키는 데에도 많은 도움을 준 경제학부 정다엘 군에게도 감사합니다. 다듬어지지 않은 문장들을 꼼꼼히 읽어 주신 심사위원들께도 감사드리며, 이상은 먼 타지에서의 새로운 삶에 원동력이자 더 좋은 글을 쓰는 연구자가 되기 위한 디딤돌로 삼겠습니다.

심사평

장 려 상

곽 지 호 (사회과학대학 지리학과)

**비-인간 창의적 시스템 논의:
생성형 인공지능 알고리즘을 중심으로**

오늘날을 살아가는 사람이라면 누구나 고민하지 않을 수 없는 문제를 다루었다는 점에서 시의성 높은 글이었습니다. 덕분에 저도 미처 몰랐던 여러 측면을 생각해보면서 즐겁게 읽었습니다.

독창성과 맥락적 가용성이라는 두 가지 필요조건을 충족함으로써 생성형 인공지능 알고리즘이 창의적이라는 주장에 충분히 공감했습니다. 하지만 행위주체성과 관련한 예상 반론을 해소하는 과정은 이해하기가 쉽지 않았습니다.

그 이유를 찾아보면 첫째, 이 글에서 채택한 귀속책임과 대조되는 책무책임과 근본적 응분 개념이 선행연구(카루소 2016)의 인용과 관련되어 소략하게 다루어졌기 때문입니다. 행위주체성에 대한 주장이 분명히 제시되고 이를 대체하는 개념이 소개되는 과정이 있었다면 독자에게 한층 친절한 글이 되었을 것입니다. 둘째, ‘미래지향적 답변 책임’과 관련해 의문이 남습니다. 밥 딜런이 포크 축제에 전자 기타를 사용했던 이유를 설명하는 것과 인공지능이 결과물 수정 요구에 응하는 것은 답변의 범위와 성격이 다른 것이 아닐지요? 인공지능은 ‘왜 그렇게 행동하기로 결정했는지, 혹은 그것이 옳은 행동이라 생각한 이유는 무엇인지’에 답변하지 못하는 것으로 보입니다. 이와 관련해 글쓴이가 답변책임과 미래지향적 답변책임을 구분하고 있는지도 궁금합니다. 혹시 구분하고 있다면 답변책임 중에서 특히 미래지향적 답변책임을 통해 인공지능의 창의성이 주장되는 것이라 볼 수 있을까요?

제목은 어떻게 결정된 것일까요? ‘생성형 인공지능 알고리즘의 창의성 옹호’와 같이 보다 직관적인 제목을 택하지 않은 이유가 궁금해집니다.

이 상 원 (기초교육원 강의교수)