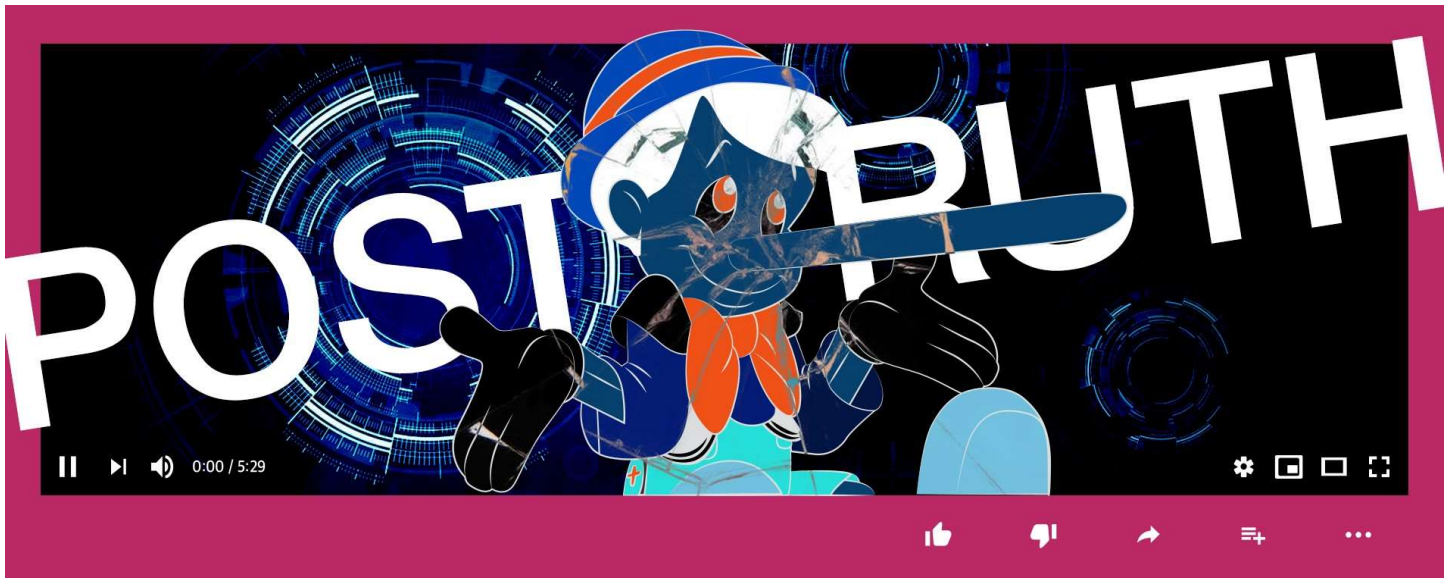


# 포스트휴먼과 포스트 트루스: 탈진실 현상은 왜 문제일까

2020년 6월 2일

구본권



인터넷과 인공지능 등 정보기술의 발달 덕분에 오늘날 인류는 방대한 정보를 자유롭게 활용한다. 유사 이래 가장 오랜 기간 교육을 받고 가장 활발하게 정보를 이용하는 세대이지만, 가짜뉴스와 허위 왜곡정보의 폐해는 어느 때보다 크다. 2016년 옥스퍼드영어사전은 '탈진실 post truth'을 '올해의 단어'로 선정했다. 탈진실 현상은 '올해의 단어'로 선정된 이후에도 사라지거나 수그러들지 않았다. 탈진실 현상의 문제는 일시적으로 등장했다가 사라질 사회적 병리 현상이 아니라 오히려 점점 강화될 추세라는 데 있다. 2017년 10월 컨설팅그룹 가트너Gartner는 미래전망 보고서를 통해 “2022년이 되면 대부분의 사람들이 진짜 정보보다 가짜 정보를 더 많이 접하게 될 것”이라는 전망을 내놓은 바 있다. 누구나 방대한 지식과 정보에 자유롭게 접근해 진위를 검증할 수 있는 환경인데 왜 사람들은 가짜에 더 쉽게 속고 허위정보를 이용하는 것일까? 선뜻 이해되지 않는 아이러니다.

포스트 트루스(탈진실) 현상이 문제 되는 이유는 정치적·경제적 이해를 노린 가짜뉴스나 허위정보, 음모론의 생산과 유통이 양적으로 늘어나고 영향력이 광범해졌기 때문이 아니다. 인류 역사상 허위정보와 루머가 사라졌던 시기는 없다. 하지만 기존에 허위정보가 사회와 개인에게 끼치던 영향과 현재의 탈진실 현상은 구조적 차이가 있다. 탈진실 현상은 인간이 지식정보와 관계 맺는 방식이 기존과 근본적으로 달라진 데서 기인한다. 과거와 달리 인간 이외 지적 활동을 하는 도구가 등장해 인간의 정보 이용 과정에서 주요한 역할을 하고 있다는 점이다. 인공지능과 알고리즘, 소셜미디어 플랫폼, 자동화 프로그램 등의 디지털 기술이다.

종교와 신으로부터 벗어난 근세 이후 개인과 사회는 이성과 지적 능력을 갖추고 지식과 정보를 통제하는 유일한 존재였다. 하지만 현재의 탈진실 상황은 인간이 더 이상 기존처럼 정보를 통제하지 못하게 된 현실을 살아가고 있음을 알려준다. 더 이상 사람만이 지식과 정보를 생산·유통하는 유일 주체가 아니게 된 것이 배경이다. 인공지능과 알고리즘 기술의 발달은 도구도 인간 못지않은 지적 기능을 수행할 수 있음을 보여주고 있다. 컴퓨터 비전, 자율주행, 머신러닝 등의 사례에서 인공지능은 사람이 수행해오던 지각과 인지, 이해와 설명, 판단과 실행 등 각종 지적 기능을 과시하고 있다.

탈진실 현상이 포스트휴먼 논의와 연결되는 지점이다. 인공지능은 인간과 같은 자율성과 의식을 갖추고 있지 못한 상태이지만, 인지적 능력을 통해 비인격적 주체의 지위를 획득해가고 있다. 인간만 지능과 의식을 갖춘 존재였는데, 인공지능은 의식과 분리된 지능의 등장을 의미한다. '의식 없는 지능'은 사람만이 할 수 있다고 여겨온 일들을 대행할 수 있게 되었고, 자율적 행동이 가능한 주체가 되었다. 알파고, 자율주행차 등이 사례다. 인간의 자율성과 다르지만, 인공지능은 사람이 작동 구조와 그 결과를 알지 못하는 방식으로 인식하고 행동할 수 있는 기능을 갖췄다는 점에서 기존 도구와 구별된다. 현재 인공지능의 자율적 인식과 판단이 설계자가 허용하고 위임한 범위에 국한하는지 여부와 별개로, 인공지능이 일부 영역에서 자율적 기능을 구현한다는 점에서 인공지능은 도구 이상의 도구이다.

//

2016년 옥스퍼드영어사전은  
'탈진실 post truth'을 '올해의 단어'로  
선정했다.

2017년 컨설팅그룹 가트너는  
미래전망 보고서를 통해  
"2022년이 되면 대부분의 사람들이  
진짜 정보보다 가짜 정보를 더 많이  
접하게 될 것"  
이라는 전망을 내놓은 바 있다.

//

이는 인간만을 주체적 인식과 활동의 주체로 보고, 그 위에서 형성해온 기존의 사회적 체계와 규정 등에 새 변수가 등장했음을 의미한다. 인공지능은 인간만을 인식과 사회적 행동의 주체로 여겨온 오랜 인식과 사회 체계에 새로운 관점을 요청하고 있다. <사피엔스>의 저자 유발 하라리 Yuval Noah Harari는 의식없는 지능의 발명이 인류 역사에 전면적 변화를 불가피하게 만드는 결정적 요소라고 말한다. 인공지능의 발달로 인해 인류가 직면하게 된 낯선 현실의 하나가 탈진실 현상이다.

인공지능 환경에서 탈진실 현상이 문제가 되는 세 가지 이유가 있다. 첫째, 기계가 정보를 이해하고 만들어내는 능력을 갖게 되었다는 점이다. 기존의 자동화 기계와 달리 인간 뇌 구조를 모방한 딥러닝 방식의 인공지능은 비지도학습 방식으로 인간의 구체적 지시가 없는 영역의 지식과 노하우를 학습하게 됐다. 정확성과 효율성이 높기 때문에 인공지능이 만들어낸 결과에 대한 사회적 신뢰와 의존도는 갈수록 높아지고 있다.

둘째, 인공지능은 인간 인지능력으로 식별이 불가능한 가짜를 대량으로 만들어내고 있다. 생성적 대립쌍 신경망<sup>GAN</sup>, Generative Adversarial Network을 활용한 딥페이크<sup>Deepfake</sup>가 사례다. 이미 인공지능은 사람보다 뛰어난 이미지 식별능력에 도달했을 뿐 아니라 진짜와 식별불가능한 이미지와 텍스트를 자동으로 무한히 만들어낸다. 인공지능 붐은 소셜미디어에서 이용자별 취향을 반영한 허위정보를 만들어내 자동유포한다. 인간 인지능력으로는 식별할 수 없고, 처리할 수 없다.

셋째, 인간의 인지능력은 기술과 달리 거의 진화하지 않는다. 사람은 성장기 때 교육과 학습을 통해 형성한 인지방식과 사고구조를 이후 변화하는 정보환경에 맞게 업그레이드하기 꺼리는 인지적 게으름뱅이다. 이미지 조작 방법이 없던 시기에 교육을 받은 활자세대는 포토샵이나 딥페이크에 익숙한 이미지세대와 이미지를 수용하는 태도가 다르다. 동영상과 사진에 대해 조작 가능성을 의심하기보다 자명한 사실로 수용하는 경향이 강하다. 교육과정에서 딥페이크와 같은 동영상 조작 기술의 존재를 배우지 않았고, 졸업 이후에도 최신 이미지 조작기술과 인공지능의 발달에 대해 학습하지 않았기 때문이다.

특히 디지털 환경은 가짜뉴스 확산에 유리한 조건이다. 다양한 목적에서 인간과 기계에 의해 정보 콘텐츠가 방대한 규모로 생산되기 때문에 이용자와 사회, 알고리즘이 정보에 대해 일일이 진위 여부를 가려낼 수 없다. 디지털 콘텐츠는 원본과 사본을 식별하기 힘들고 콘텐츠 작성의 주체와 출처를 확인하기도 어렵다.

## 연재글

### 포스트휴머니즘을 성찰한다

1. [포스트휴먼과 포스트휴머니즘, 그리고 삶의 재발명](#)
2. [왜 포스트휴머니즘인가](#)
3. [포스트휴머니즘과 인류세](#)
4. [인공지능과 포스트휴머니즘 - 초학제의 New Horizon](#)
5. [포스트휴먼과 포스트 트루스: 탈진실 현상의 왜 문제일까](#)

허위정보의 범람과 확산이 갈수록 강화되고 지속될 사회 변화의 방향이라면 이에 대한 대응은 무엇일까. 인류의 지적 속성과 능력, 그에 대한 사회적 합의의 변화 역사를 검토해야 한다.

진짜와 가짜를 식별하는 능력은 일상적 사회환경에서는 신뢰와 판단, 행동의 근거다. 하지만 물리학, 철학적 논의에서는 달라진다. 인간 인지능력의 한계와 특성으로 인해 진실 자체를 파악하는 것이 근본적으로 불가능하다. 답이 없기 때문에 철학에는 이를 탐구하는 존재론과 인식론이 독립된 학문으로 존재한다. 또한 사실에 대한 사회적 인식은 시간의 영향을 받는다. 하버드대 복잡계 물리학자 새뮤얼 아브스만<sup>Samuel Arbesman</sup>은 <지식의 반감기>에서 대부분의 지식은 절대적 지식이 아니라 유효기간을 지닌 가변적 지식이라고 주장한다. 그래서 한때 사실로 여겨진 것이 시간이 지나며 사실이 아닌 것으로 되기도 한다. <자유론>의 존 스튜어트 밀<sup>John Stuart Mill</sup>이 검열에 반대하고 표현의 자유를 주창한 근거는 지금 참이 아닌 것으로 여겨진 소수의 의견도 오랜 시간이 지난 뒤 진리로 밝혀질 수 있다는 논리였다.

하지만 현실에서 사실은 철학적 개념의 영역이 아니라 법과 권력에 따른 합의와 질서의 공간이다. 사회는 상호주관성을 토대로, 사실에 대한 공통된 인식과 합의가 가능하다는 사회적 신뢰 위에서 작동한다. 갈릴레이 시절 배척당하던 지동설처럼 과학적 사실도 그 자체로 규명되는 게 아니라 지식의 도달수준 그리고 시대와 상황에 따라 합의로 만들어지고 변화한다. 절대적 진실의 규명 없이도 사회구성원들이 공통적 사실이라고 받아들이는 지식과 정보체계 덕분에 사회제도와 상호소통이 가능했다. 사회는 공동체가 신뢰할 수 있는 사실의 체계를 요청한다. 하지만 탈진실 현상은 더 이상 과거와 같은 사실성에 대한 사회적 합의를 충족시키기 어려워지는 상황임을 의미한다.

오늘날 탈진실 현상은 왜곡정보와 음모론을 신봉하고 그에 따라 행동하는 사람들이 늘어나기 때문에 위험한 게 아니다. 공동체가 기반하고 있는 공통된(상호주관적인) 사실의 개념을 허물어버린다는 게 탈진실 현상의 진짜 위험이다. 공동체는 사실에 대해 공통된 신념 공유를 기반으로 가능한데 구성원들이 사실에 대한 서로 다른 개념을 갖고 있다면 공동체가 존립하기 어렵다. 탈진실 현상은 신뢰 기반의 붕괴와 사회적 소통 단절을 가져온다. 탈진실 현상은 인공지능 기술 발전과 인간 인지능력 간의 격차가 커지는 문화지체<sup>cultural lag</sup>에서 비롯하는 문제다. 무어의 법칙이 알려주듯 정보의 총량과 처리속도는 단기간에 빠른 속도로 늘어나고 있다. 지식정보사회에서 과거 정보는 새로운 정보로 끊임없이 업데이트되고 있지만 인간의 인지능력은 이를 따라잡기 어렵다. 인간 두뇌는 컴퓨터와 달리 방대한 정보를 저장, 처리할 수 없다. 정보가 방대해짐에 따라 인공지능과 알고리즘과 같은 지적 보조도구에 대한 의존성이 커질 수밖에 없다.

또한 가능한 한 두뇌의 인지적 자원을 덜 사용하고자 하는 인지적 게으름뱅이<sup>cognitive miser</sup>로서 사람은 확증 편향<sup>confirmation bias</sup>과 필터버블<sup>filter bubble</sup>과 같은 편향에 빠지기 쉽다. 인터넷은 어떠한 잘못된 개념이나 음모론을 신봉하는 경우에도 충분한 정보와 논리를 제공해주는 거대한 정보의 수원지다. 지구평면설<sup>flat earth theory</sup>이나 백신유해설을 신봉하고 퍼뜨리는 집단이 전형적 사례다. 소셜미디어에서 비슷한 생각을 가진 사람들끼리 소통하며 자신들의 신념을 강화하는 현상을 반향실<sup>echo chamber</sup> 효과라고 부른다. 정보 검증의 효과적 도구로 기대되었던 인터넷이 허위왜곡 정보에 대한 신념체계까지 강화시켜주는 현상들이 나타나고 있다. 정보와 미디어가 범람하는 상황에서는 자신의 생각과 상반되는 견해로부터 스스로를 고립시키기가 훨씬 쉬워졌다. 소셜미디어에서는 왜곡되고 황당한 신념을 가진 사람들끼리도 상호 정보와 신념을 공유하면서 기존의 태도를 강화하는 게 얼마든지 가능하다.

//

사람이 식별할 수 없는 허위정보가  
사람이 인지하지 못하는 방식으로  
인공지능과 알고리즘에 의해

## 대량으로 만들어져

일상 속으로 들어오는 세상이 탈진실

현실이다.

//

탈진실 현상은 인공지능과 인터넷 상황에서 불거졌지만 인공지능과 인터넷의 설계구조를 개선하고 변경하는 것으로 해결할 수 있는 문제가 아니다. 정보기술과 서비스 플랫폼의 구조가 잘못 설계되어서 생겨난 현상이 아니기 때문이다. 탈진실 현상은 인간의 인지적 특성에서 비롯하는 문제로, 빠르게 발달하는 정보기술과 변화하지 않는 인간 인지능력 간에 생겨나는 격차가 배경이다. 정보기술을 악의적으로 활용하는 세력이 그 격차를 이용해 가짜뉴스를 유포하고 사회의 신뢰 기반을 붕괴시키고 있다.

하지만 탈진실 현상이 새로운 게 아니라는 관점도 있다. 유발 하라리는 “인간은 늘 탈진실의 시대를 살아왔으며 호모 사피엔스 특유의 힘은 허구를 만들고 믿는다는 점”이라고 말한다. 하라리는 “1000명의 사람이 조작된 이야기를 한 달 간 믿으면 가짜뉴스이지만 10억명이 1000년 동안 믿으면 종교다”라고 예시한다. 하라리는 화폐제도, 국가, 종교, 이념 등을 인간이 만들어낸 허구의 전형으로 제시한다. 호모속 중에서 유일하게 정교한 허구를 만들어내고 믿을 수 있는 능력 덕분에 인간은 지속적으로 정교하게 대규모의 협력과 지적인 작업을 할 수 있었다는 해석이다. 하라리가 말하는 허구는 달리 말하면, 개념이자 사고방식이다. 다른 동물들이 감각적 실재만을 인식하고 고려하는 것과 달리 인간은 비감각적 개념을 실재하는 것처럼 받아들여 인식과 행동의 근거로 삼을 수 있다는 주장이다.

하라리가 탈진실 현상을 호모 사피엔스의 오랜 특성이라고 주장했지만 현재의 탈진실 현상은 기존과 구별된다. 하라리가 인간의 개념화와 사고능력을 비실재적인 것을 실재하는 것처럼 사유하고 간주한 것을 인간의 허구 창조능력으로 설명했지만, 이는 기본적으로 인간의 인지와 광범한 동의 아래 이뤄져 왔다. 화폐시스템, 종교, 이데올로기, 가치관 등 ‘허구적 개념’을 사회가 수용하고 그 전제 위에서 각종 활동이 이뤄져왔다. 모든 개인이 이러한 개념과 제도적 장치를 허구라고 인식하고 수용하지 않았지만, 다수의 개인과 사회는 자신들이 수용하고 신뢰하는 비실재의 개념의 작동 메커니즘을 인지한 상태였다. 기본적으로 인간의 활동으로 만들어낸 허구의 체계였다.



엔비디아NVIDIA에서 개발한 스타일갠styleGAN이 인공적으로 생성한 인물 이미지. 이미지 속 인물은 실제로 존재하지 않는다.

[wikimedia](#)

하지만 근래의 인공지능과 디지털 환경에서 탈진실 현상은 다르다. 최신 인공지능 알고리즘은 인간 두뇌구조를 모방한 심화신경망의 은닉층 구조를 통해 작동하는데 사람은 입력값과 결과값을 얻을 뿐 내부의 작동구조를 알 수 없다. 인공지능이 만들어낸 산출물의 구조를 인간 인지시스템이 파악할 수 없는 상황은 심화신경망 방식 인공지능의 특징인 설명불가능성이다. 설명가능한 인공지능이 향후 과제로 부상한 배경이다. 허구가 현실의 삶과 관계에 영향을 끼치고 있다는 사실을 개인과 사회가 인지한 상태와 인지하지 못한 상태는 같을 수 없다. 하라리가 말한 것과 달리 오늘날의 탈진실 현상은 인류가 일찍이 직면해보지 않은 상황이다. 지금까지 인류가 탈진실과 공존하는 방법과 해결방안을 보유한 채 살아왔다면, 현재의 인류는 새로운 탈진실 현상에 대한 대응 방안을 갖고 있지 못하다.

이러한 환경이 인류가 더 오랜 기간 교육을 받고 더 강력한 지적 도구를 항상 지니고 있음에도 가짜뉴스, 허위정보의 영향을 벗어나지 못하게 된 탈진실 현상의 배경이다. 탈진실 현상은 호모 파베르의 역설이라고 할 수 있다. 호모 파베르는 도구의 개발사에서 가장 강력하고 편리한 도구를 만들어내는 데 성공했다. 인간의 지적 결함을 보조하고 인지적 수고를 덜기 위해 지적 기능을 수행할 수 있는 똑똑하고 편리한 디지털 기술이다. 인지, 이해, 판단, 창작 등 다양한 인

간의 지적 기능까지 수행할 수 있는 인공지능이 등장하게 됐다. 조작의 주체와 작동구조를 알 수 없지만 항상 뛰어난 효율성을 보이는 인공지능의 결과물도 활용할 수 있게 됐다. 사람이 일일이 개입하지 않고도 편리하고, 강력하고, 효율성 높은 결과물을 이용할 수 있게 됐다.

사람이 식별할 수 없는 허위정보가 사람이 인지하지 못하는 방식으로 인공지능과 알고리즘에 의해 대량으로 만들어져 일상 속으로 들어오는 세상이 탈진실 현실이다. 인공지능을 활용하거나 집단지성이 참여하는 필터링 시스템을 통해 허위정보를 걸러내겠다는 기술적 시도가 있다. 각종 법규와 처벌을 강화해 의도적 허위정보를 막겠다는 법률적 시도도 여러 나라에서 진행 중이다. 하지만 기술적 시도와 법률적 시도 모두 현재의 탈진실 현상에 대한 실질적 해결책이 될 수 없다. 생성적 대립쌍 신경망의 구조처럼, 상대방의 구조를 참고해 가짜를 만들어내는 알고리즘과 가짜를 식별하는 알고리즘 간의 경쟁에서 어느 한쪽의 절대적 우위는 불가능하기 때문이다. 더욱이 가짜 뉴스를 법이나 규제로 차단하는 것은 한계가 있다. 허위 사실의 통신 또한 표현의 자유로 보호받는 영역이기 때문이다. 헌법재판소는 2010년 12월 이른바 '미네르바 사건'에서 허위정보를 포함한 내용의 통신을 처벌하려는 것은 헌법적 가치인 표현 자유를 침해한다고 결정한 바 있다.

가짜 뉴스의 생산과 유통은 기본적으로 디지털 기술 환경에서 발생하는 현상이고, 이용자의 자발적 선택에 의해 이뤄지는 행위이다. 이는 가짜 뉴스가 소셜미디어라는 기술적 플랫폼 차원에서 접근이 이뤄져야 한다는 점과 함께 이용자의 정보 수용능력 차원에서 대응이 이뤄져야 한다는 것을 의미한다. 기술적 해결방안, 제도적 해결방안, 정보인지능력 업그레이드를 위한 교육적 방안 등이 함께 고려되어야 한다. 그중 어느 것도 효율적이거나 결정적 해결방안이 되지 못한다는 것을 이해하는 게 필수적이다. 탈진실 현상에 대해서 현재로서는 근본적 대책이 불가능하며 제한적 효과가 있는 접근법밖에 없다. 하지만 탈진실 현상은 불가피하다는 것을 인지하는 게 최선의 출발점이다.