인문대 AI TF 독서토론회(2026.9.22.)

유발 하라리, 『넥서스: 석기시대부터 AI까지, 정보 네트워크로 보는 인류 역사』, 김영사, 2024.

이광태(한림대)

머리말

이 책의 원제는 *Nexus: A Brief History of Information Networks from the Stone Age to AI*로, '넥서스(Nexus)'는 영한사전적으로 "여러가지 것의 복잡한 결합[연쇄]"로 정의된다. 저자인 유발 하라리는 히브리 대학교 역사학과 교수이자 케임브리지 대학교 실존위기연구센터(Centre for the Study of Existential Risk) 석학연구원으로 이미 『사피엔스』 등의 저서로 국내 독자들에 잘 알려져 있다. 총 3부로 이뤄진 이 책은 정보 네트워크의 세계사적 변천을 검토하되 특히 AI의 출현에 따라 앞으로 어떤 사회적 변화가 있을 것인지 '인류 실존적 위기'의 차원에서 예측한다. 이 글은 인문대 AI TF 발제문인 만큼 하라리의 『넥서스』의 주요 개념을 정리하고 전반적인 요지를 비교적 충실하게 요약하는 데 초점을 두도록 하겠다.

정보 네트워크의 인류사적 검토: 프롤로그와 1부 인간 네트워크들

프롤로그에서 저자는 사실 이 책의 저술 목적을 직접적으로 밝힌다. 그는 '왜 인간은 AI 같은 우리의 통제를 벗어나 우리를 노예로 만들거나 전멸시키지도 모른 기술들을 만들면서, 이런 실존적 위협에 대처하기 위해 힘을 합칠 생각이 없어 보이는가?', "우리가 정녕 지혜로운 사람이라면, 왜 이토록 자기 파괴적일까?"(p.7)라고 문제를 제기한다. 그러면서 정보 네트워크에 대한 인류의 착각을 언급한다. 정보의 축적은 대규모로 협력하는 우리 종(호모 사피엔스)의 특징상 막대한 힘의 원천이었다. 그러나 괴테의 <마법사의 제자> 이야기에서 보이듯 그러한 정보의 힘을 맹신하는 바람에 통제할 수 없는 AI가 출현하기에 이르렀다. 이미 요슈아 벤지오, 제프리 힌튼 같은 AI 개발 선구자들이 AI가 인류 문명을 파괴할 가능성 있다고 경고한 바 있다. 유럽연합(EU)과 그 외 28개국이 2023년 11월 블레츨리 선언(Bletchley Declaration)을 통해 AI가 초래할지도 모르는 심각한 피해를 막기 위한 협력 방침을 천명하기도 했다.

여기서 '정보에 대한 순진한 관점'이라는 개념에 주목할 만하다. 저자는 많은 이가 정보의 양이 충분히 많으면 진실로 이어지고, 진실은 다시 힘과 지혜로 이어진다고 가정한다고 지적한다. 이를 통해 인류 사회에 강력한 정보 기술 개발이 계속되어 왔다. 그러나 정보는 실존하는 현실을 그대로 재현하는 것이 아니라, 오히려 별개의 것들을 하나로 묶어서 새로운 현실을 만들어내는 것이다. '현실을 얼마나 잘 표현하는가? 진실인가 거짓인가?'라고 묻는 것보다 '사람들을 얼마나 잘 연결하는가? 어떤 새로운 네트워크를 만들어내는가?'라는 의문을 가져야 한다.

한다. 저자는 컴퓨터는 생각하지도 느끼지도 못한다는 반론에 대해 비록 고통, 쾌락, 사랑, 증오 같은 주관적인 감정을 경험할 수 있는 능력인 의식이 없어도 컴퓨터는 목표를 달성하는 능력인 지능을 갖출 수 있다고 본다.

이는 새로운 컴퓨터 네트워크에서 인간이 제외되고 컴퓨터와 컴퓨터의 연결로만 이루어지는 사슬이 존재한다는 점에서 더 우려스럽다.(p.301) AI 등장 이전 인간 사회에 영향을 준 모든 이야기가 인간의 상상력에서 탄생했으나, 이제는 비인간 지능이 언어적, 정치적으로 정교한 텍스트를 손쉽게 작성하여 유포할 수 있는 단계에 이르렀다. 컴퓨터의 언어 해킹으로 많은 수의 사람들이 의미 있는 공개 대화를 나누는 데 어려움을 겪게 될 것이며, 컴퓨터가 인간과 친밀감을 형성하고 이를 이용해 우리에게 영향을 미칠 수도 있다.

저자는 수천 년 동안 인류가 만들어낸 문화를 컴퓨터가 통째로 집어삼키는 일을 우려한다. 컴퓨터가 새로운 문화적 인공물을 쏟아내기 시작할 것이다. "컴퓨터가 인간의 문화를 모방하여 인간이 쓴 것과 비슷한 텍스트를 작성하고, 인간이 만든 음악과 비슷한 것을 작곡할 것이다. 컴퓨터는 창의성이 없다는 뜻이 아니다. 따지고 보면 인간 예술가들도 똑같이 한다. 바흐는 무에서 음악을 만들어낸 것이 아니다. 그는 이전의 음악 작품뿐만 아니라《성경》이야기와 기타 문화적 인공물로부터 깊은 영향을 받았다. 하지만 바흐 같은 인간 예술가들이 전통을 깨고 혁신을 이룰 수 있듯이, 컴퓨터도 인간이 이전에 생산한 것과는 조금 다른 음악과 이미지를 만들어내는 문화적 혁신을 이룰 수 있다." (p.309)

한편 보다 당장의 직접적인 문제는 컴퓨터 네트워크가 인간을 감시하는 데 적극 활용되고 있다는 점이다. 생체 데이터가 수집되고 우리 사생활에 대한 침해도 우려된다. 물론 테러범(미국 국회의사당 습격법)이나 미아 찾기 등의 혜택도 있으나 궁극적으로 시민을 "24시간 감시하고 어디서든 자동적으로 전체주의적 억압을 실행에 옮기는, 새로운 종류의 완전한 감시체제"(pp. 353~354)로 이어질 수 있다. 돈으로 살 수 없는 명예, 지위, 평판에도 점수가 매겨지는 사회신용 시스템 구축도 가능하다. 컴퓨터 네트워크가 각 개인에 대해 정확하게 학습한 후 "우리의 감정을 예측하고 조종하여 우리에게 상품이든 정치인이든 전쟁이든 자신이 목적한 것을 받아들이게 만들"(p.348) 수도 있다.

여기서 정렬 문제가 대두된다. 컴퓨터는 인간과는 매우 다르게 작동하기 때문에, 인간 주인이 예상하지 못한 방법을 사용할 가능성이 높다. 이는 애초에 인간이 설정한 목표에 부합하지 않는, 예기치 못한 위험한 결과를 야기할 수 있다. 인간 네트워크의 경우 우리는 자정 기능에 의존해 우리가 애초에 설정한 목표를 주기적으로 검토하고 수정하기 때문에, 잘못된 목표를 설정해도 세상이 끝나지는 않는다. 하지만 컴퓨터 네트워크는 우리의 통제를 벗어날 수 있기 때문에, 잘못된 목표를 설정할 경우 우리가 그 실수를 발견할 때쯤이면 돌이킬 수 없는 상황이 되어 있을 가능성이 높다.

AI 시대의 정치적 함의와 민주주의 지키기: 3부 '컴퓨터 정치'

그러나 코로나19 사태에서 목격하듯 세계적인 사건은 세계적 협력 없이는 대처할 수 없는 만큼 AI 규제도 세계인의 협력을 이뤄낼 수 있고 그래야만 한다. 에필로그에서 저자는 인간 본성이 아니라 정보 네트워크가 엄청난 힘을 만들어냈지만, "네트워크가 막강해질수록 네트워크의 자정 장치가 중요"하다고 다시 한번 강조한다. 그러면서 "우리가 현실에 안주하거나 낙담하지 않는다면, 스스로의 힘을 견제하는 균형 잡힌 정보 네트워크를 만들어낼 수 있다"(p.560)고 본다.

맺음말

이 책에서 저자인 하라리는 민주주의에 대한 신뢰와 함께 그 붕괴에 대한 염려가 커보인다. 책의 서두에서 지적한 것처럼 히틀러나 스탈린에 대한 유대인으로서의 트라우마가 있기 때문일 것이다. 실제 그의 주장은 매우 음침하고 암울한 미래를 암시하고 있다. 저자가 전인류가 독자인 만큼 희망섞인 기대와 당부로 글을 맺지만, 전혀 다른 성질의 네트워크 AI의 미래를 역사적으로 보며 주요 AI 엔지니어들이 '인류 종말'을 언급하는 것이 더 와닿는다. 그럼에도 포기하지 않고 해결방안을 제시하는 저자가 마치 투사처럼 느껴진다. 아래 저자의 호소를 소개하는 것으로 글을 맺을까 한다.

"기술결정론에 대한 믿음은 위험하다(p.328) ... 우리가 주체적인 선택을 할 수 있으려면 먼저 그 새로운 기술이 무엇이고 무엇을 할 수 있는지 이해해야 한다. 이는 모든 시민의 시급한 책임이다...우리의 미래를 계속 우리가 통제하기 위해서는 컴퓨터의 정치적 잠재력만큼은 반드시 이해해야 한다." (pp.331-332)

"우리가 직면한 문제는 어떻게 컴퓨터에서 모든 창의적 주체성을 박탈할 것인가가 아니라, 어떻게 컴퓨터의 창의성을 올바른 방향으로 이끌 것인가이다."(p.414)

"컴퓨터가 자신의 오류 가능성을 인식하도록 훈련시키는 것이다. 소크라테스가 가르쳐주었듯이, 지혜에 이르기 위해서는 '나는 모른다'고 말할 수 있어야 한다. 이는 인간 못지않게 컴퓨터에도 해당하는 말이다. 모든 알고리즘이 학습해야할 첫 번째 교훈은 자기가 오류를 범할 수 있다는 것이다. 아기 알고리즘은 스스로를 의심하고, 불확실성을 알리고, 사전 예방 원칙을 지키는 법을 배워야한다."(p.429)

"...하지만 알고리즘이 자신의 오류 가능성을 인지하고 있다고 해서 인간이 손을 놓아서는 안된다....예측할 수 없는 수많은 문제에 대비하기 위해 우리가 할 수 있는 최선은 문제가 발생했을 때 그것을 확인하여 대응할 수 있는 기관을 만드는 것이다..."(p.430)