



인간과 기계-갈등과 공생의 역사

저자 (Authors)	홍성욱
출처 (Source)	<u>문학과사회 28(3)</u> , 2015.8, 466–488 (23 pages) <u>LITERATURE&SOCIETY 28(3)</u> , 2015.8, 466–488 (23 pages)
발행처 (Publisher)	<u>문학과지성사</u>
URL	http://www.dbpia.co.kr/Article/NODE06506693
APA Style	홍성욱 (2015). 인간과 기계-갈등과 공생의 역사. 문학과사회, 28(3), 466–488.
이용정보 (Accessed)	서울대학교 147.46.24.*** 2017/09/05 14:02 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

인간과 기계—갈등과 공생의 역사

시리 Siri

홍 교수는 운전을 할 때 종종 아이폰이 제공하는 인공지능 서비스 시리를 애용한다. 시리는 음성으로 작동되기 때문에 운전 중에 전화를 걸 때 특히 편리하다. 하루는 동료 교수와의 약속 시간에 조금 늦을 것 같아서 시리에게 그 교수에게 전화를 하라는 요청을 했다. 그런데 평소에는 이런 요청을 잘 알아듣고 처리하던 시리가 그날따라 뭔가 문제가 있느니 계속 “무슨 말인지 잘 모르겠습니다”만 반복하는 것이 아닌가. 몇 번 시도하다가 포기하고 전화기를 내려놓으면서 혼잣말로 “얘가 오늘 왜 이러지” 중얼거렸는데, 갑자기 시리가 반응을 했다. “그럴 때도 있죠, 뭐”라고. 그 순간 홍 교수의 뇌리에는 시리가 곧 의식consciousness을 갖게 될지도 모른다는 생각이 스쳐 지나갔다.

비관론과 낙관론

르네상스 아래 서구 사상의 역사를 인간과 기계의 관계라는 관점에서 본다면, 그것은 인간이 기계에 불과하다는 생각과 이를 비판하고 반대하는 생각 사이의 논쟁의 역사가 될 수 있을 것이다. “인간=기계”라

는 등식을 향한 첫걸음은 기계 같은 인공물과 자연의 차이를 없애는 데에서 시작했다. 고대와 중세를 풍미한 아리스토텔레스주의 자연철학에 의하면 목적인 final cause 을 그 속에 내재한 자연은 인간이 목적인 을 부여한 인공과는 본질적으로 다른 것이었고, 따라서 자연을 연구하는 자연철학은 인공을 만드는 기술과 본질적으로 다른 활동이었다. 이런 엄격한 구별은 르네상스 시기부터 비판의 대상이 되었다. 레오나르도 다빈치는 새가 “수학의 법칙에 따라서 움직이는 장치이므로 [……] 인간의 능력으로도 (새를) 재현할 수 있다”고 주장했다. 이런 단편적인 주장들을 철학적으로 종합한 사람이 프랜시스 베이컨이었다. 베이컨은 『학문의 진보』(1605)에서 사람들이 자연과 인공을 다른 것으로 간주하는 전통을 비판하면서 “사실 인공과 자연의 차이는 형태나 본질에 있는 것이 아니라 효율에 있다”고 주장했다. 인간이 인공물을 붙이고 분리하듯이 “물체를 분리하거나 붙이는 운동으로 자연에 힘을 행사할 수 있다”고 했고, 이는 자연에 대해 개입하는 실험적인 방법을 자연철학이 도입해야 함을 의미했다. 자연은 동굴 속에서 잡을 자는 맹수 같은 존재였고, 이를 탐구할 수 있는 유일한 방법은 맹수의 꼬리를 흔들어서 이것을 동굴 밖으로 나오게 하는 수밖에 없었다. 이것이 실험과학의 방법이었다.

베이컨과는 전혀 다른 동기에서 프랑스의 과학자이자 철학자였던 데카르트는 우주가 하나의 거대한 기계라고 생각했고, 살아서 움직이는 동물 등 생명체도 복잡한 기계에 다름 아니라고 보았다. 데카르트는 청년 시절에 생제르맹에 지어졌던 앙리 4세의 정원에서 페르세우스가 창을 휘둘러서 용을 무찌르는 자동인형을 본 뒤에 엄청난 충격을 받았

다. 그는 “분수 정원의 동굴에 들어선 방문객은 [……] 자신 앞에 놓인 타일을 밟지 않고는 거기에 들어설 수 없는데, 이 타일을 밟고 들어간 뒤에 목욕하는 다이애나에게 접근할 경우에 그녀를 갈대 뒤로 숨게 만들고, 그녀를 쫓아서 더 앞으로 나갈 경우에는 포세이돈이 등장해서 삼지창을 휘두르며 그를 겁주어 쫓아버리게 한다. 다른 방향으로 걸어갈 때는 바다 괴물이 등장해서 그의 얼굴에 물을 뿜어버린다”라고 기록했다. 물론 데카르트가 이런 자동인형을 살아 있다고 생각하지는 않았지만, 이 인형을 만든 장인보다 훨씬 더 솜씨 있는 장인이 있다면 이보다 더 정교하고 정밀한 자동인형을 만들었을 것이라고 생각했다. 신은 우주를 만들 때 생명체를 같이 만들었는데, 신이라는 놀라운 장인은 마치 살아서 움직이는 것처럼 보이는 기계인형을 얼마든지 만들 수 있을 거라고 생각했던 것이다.

데카르트는 동물이 영혼이 없고, 따라서 복잡한 기계라고 생각했다. 데카르트 사후에 이런 생각은 많은 이들의 비판의 대상이 되었다. 비판자들은 암수 동물을 같이 둘 경우에 새끼가 만들어지지만, 시계 두 개를 평생 같이 둬도 다른 시계가 생기지는 않는다고 데카르트를 조롱했다. 데카르트는 인간이 영혼을 가진 존재라는 점에서 동물과는 확연하게 구별된다고 생각했는데(따라서 인간은 기계가 아니라고 생각했는데), 프랑스 철학자인 라 메트리는 이런 인간/동물의 구별을 비판하면서 인간과 동물 사이에 근본적인 차이가 존재하지 않는다고 주장했다. 그는 원숭이가 말을 배운다면 원시인과 다를 바가 없다고 했다. 또 그는 “똥싸는 오리”와 “플루트 주자”를 만들었던 자크 보캉송 같은 장인이 조금 더 재주를 부린다면 말하는 인간도 만들 수 있을 것이라고 예견했다.

흔히 라 메트리를 유물론자라고 부르는데, 실제로 그의 이론에는 생기론의 요소가 더 많았다. 물질과 독립적으로 존재하는 영혼을 부정했지만, 그는 물질이 생각하는 역량을 가진다고 보았다. 사고·감정·의지 등 인간 정신 작용의 특성들이 물질에 내재해 있다고 생각했던 것이다. 라 메트리의 우주는 죽어 있는 물질로 이루어진 우주가 아니라, 느낌을 아는 물질로 가득 찬 우주였다.

기계를 철학적 논의의 대상에서 사회적, 정치적 문제로 바꾼 것은 산업혁명이었다. 산업혁명은 공장제 산업이 농업을 대체해서 생산의 주인이 되었던 사건인데, 그 핵심은 자연력과 인간의 숙련에 의존하던 노동이 기계화된 것이었다. 기계는 장인의 숙련 노동을 무력화시키면서 노동 대중의 비숙련, 반숙련 노동을 창조했다. 프리드리히 실러는 기계가 세상을 파괴해서 “생명이 없는 태엽 장치만 남아서, 분해 조립이 가능한 부품들의 집합적이고 기계적인 삶”만을 남길 것이라고 예측했다. 피테는 기계의 행진을 겸은 구름이 몰려오는 것으로 묘사하면서, 장인들이 영혼 없이 일하는 노동자가 되어서는 안 된다고 주장했다. 칼라일은 자신의 시대를 “기계의 시대”로 부르면서, 기계가 인간의 노동만이 아니라 인간의 감정과 사고까지 지배한다고 우려하면서 “인간은 손뿐 아니라 머리와 가슴까지 기계적이 되었다”고 한탄했다. 마리 셀리는 인간이 발전시킨 기술이 독자적인 생명력을 획득한 뒤에 그 주인인 인간을 배신하고 인간에게 복수할 수 있다는 두려움을 소설 『프랑켄슈타인』(1818)에서 섬뜩하게 묘사했다. 산업혁명이 세상을 돌아킬 수 없는 방식으로 바꾸고 있다는 것을 사람들이 느끼기 시작했을 무렵에 『프랑켄슈타인』이 출간된 것은 우연이 아니었다. 마르크스는 공장을 가득

채운 기계 괴물이 노동자들을 수족처럼 부리면서, 노동 과정에서 인간을 소외시키고 있다고 호소했다.

그렇지만 기술의 발전에 대한 낙관적이고 희망적인 평가도 많았다. 기계와 산업의 발전을 목도한 헨리 애덤스는 “인간은 과학 위에 올라탔고, 이제 그것을 타고 달리고 있습니다. 몇백 년이 채 가기 전에 과학이 인간의 주인이 될 것이라고 저는 확신합니다. 인간이 발명할 엔진은 인간이 제어할 수 없을 것입니다”라고 썼는데, 그는 이 과정을 훨씬 적인 진보의 과정이라고 보았다. 정치인 다니엘 웨스터는 1847년 미국의 한 철도 개통식의 연설에서 19세기를 이렇게 묘사했다.

우리가 살고 있는 세상은 참으로 놀라운 세상이다. 이 세상은 완전히 새롭다. 이 비슷한 것도 이전에는 본 적이 없다. [……] 우리는 대양과 대지를 중기기관의 힘을 빌어 가로지르고, 전기를 통해 메시지를 소통한다. 참으로 기적과 같은 일이다. 우리 앞에 있는 것을 보는 사람도 거의 없고, 우리에게 부여된 것을 느끼는 사람도 거의 없다. 우리 시대의 진보는 인간의 믿음을 뛰어넘은 것이며, 미래는 신만이 알고 있다.

다윈의 진화론이 제창된 이후에 인간-기계의 관계는 새로운 국면으로 접어들었다. 다윈은 1851년 당시에 수정궁(Crystal Palace)에서 전시했던 수많은 기계들을 보고 기계들이 발전하듯이 생명체의 기관들도 환경의 필요에 맞게 발전하고 이에 따라서 종이 분화한다는 생각을 했는데, 『종의 기원』(1859)이 출판된 뒤에는 진화론을 써서 미래에 있을 기계의 진화를 설명하려는 시도가 나왔다. 새뮤얼 버틀러는 「기계 속의

다윈」(1863)이라는 에세이에서 기계가 계속 진화를 해서 미래에는 기계가 의식을 가지고 자기와 비슷한 새끼를 만들게 될 것이라고 예견했다. “매일매일 기계는 우리에게 바짝 다가오고 있으며, 매일매일 우리는 점점 더 그들에 의존하고 있다. 더 많은 사람들이 그들의 노예가 되고 있고, 더 많은 사람들이 기계적 생명을 발전시키는 데 더 많은 에너지를 쓰고 있다. 이 결과는 단순히 시간의 문제일 것이지만, 기계들이 이 세계와 그 구성원들 모두에 대해서 확실한 우위를 점하는 시간이 도래할 것이다.” 이런 생각은 그의 잘 알려진 소설 『에레흔Erewhon』(1872)의 핵심적인 모티프가 되었고, 소설 속에서 에레흔의 사회는 인간의 존엄성을 유지하기 위해서 모든 기계를 다 없애는 극단적인 조치를 취한다.

20세기 초에는 이런 비관적인 관점에 맞서는 강력한 낙관론이 제기되었다. 기술의 발전이 세상의 진보를 가져올 것이라는 낙관론의 주창자는 『타임머신』 등의 SF소설을 썼던 H. G. 웰스였다. 그는 1901년에 발표된 「기계적이고 과학적인 진보가 인간의 삶과 사고에 미치는 영향에 대한 기대」라는 글에서, 당시의 기술 발전이 계속될 때 1백 년 뒤에 세상이 어떻게 변할 것인가를 예측했다. 웰스가 우선적으로 주목했던 것은 자동차, 기차와 같은 운송 수단의 발전이었다. 그는 이런 새로운 운송수단이 계속 발전하고 저렴해지면, 더 많은 사람들이 이를 이용하면서 도시가 도심과 외곽의 주거지역으로 분화될 것이라고 생각했다. 그는 지배 – 피지배 계급의 이원화가 농경시대의 산물이라고 간주하면서, 새로운 기술 시대에는 계급이 더 세분화되고, 그중에서도 특히 과학기술을 담당하는 새로운 중산계층이 사회에서 가장 주목받는 계층이 될 것이라고 예언했다. 과학적 세계관과 사회에 대한 의무로 무장

한 이들은 미래 세상을 이끌어가는 주역이었고, 국가의 지도적인 계층이 될 것이었다. 전쟁은 포탄의 정확도에 더 의존하게 되고, 역시 과학적으로 훈련받은 사람들에 의해서 수행될 것이었으며, 전 세계는 하나의 “새로운 공화국”을 만드는 데 동의하게 되면서 평화의 상태를 유지하게 될 것이었다. 인간은 더 이상 죽음을 두려워하지 않고, 더 많은 사람들이 출산의 의무에서 벗어나 해방된 성생활을 즐길 것이었다. 웰스에 의하면 과학기술의 발달은 1백 년 뒤에 실현 가능한 유토피아를 가져다주는 것이었다.

이런 낙관론에 대한 비판도 등장했다. 작가 E. M. 포스터는 1909년 소설 『기계 멈추다』에서 웰스의 낙관론을 겨냥했다. 이 소설에서는 미래의 사람들이 절대적인 글로벌 기계에 의해 유지되는 지하 세계에 고립되어 생활하면서, 서로 실제적인 만남보다는 일종의 메신저 같은 서비스를 이용해서 소통하고 지식을 교환하는 것으로 그려진다. 미래의 사람들은 서로 한 번도 만나보지 못했던 친구들을 많이 두고 있는데, 이런 상황은 지금 SNS가 널리 퍼져서 페이스북에서는 수천 명의 친구가 있지만 현실 세계에서는 믿을 만한 친구가 한 명도 없는 상황과 매우 흡사하다. 소설은 이런 식으로 유지되는 인간관계의 위험을 지적하면서, 전지전능하다고 간주된 기계가 멈추면서 인간과 사회 모두가 무너져버리는 파국적인 상황을 묘사한다. 기계가 멈추고 문명이 붕괴되는 날, 소설의 두 남녀 주인공은 사람들의 시체 더미 위에서 서로의 손을 잡고 이렇게 외친다. “나는 죽어간다. 그렇지만 우리는 기계를 통하지 않고 서로를 만지고 대화를 나눈다”고.

매트릭스

워쇼스키 남매가 연출한 영화 「매트릭스」 시리즈는 영화 곳곳에 담겨 있는 철학적인 화두와 화려한 액션 장면으로 전 세계 영화팬들의 사랑을 받았다. 이 영화를 인간과 기계의 관계에 초점을 맞춰서 분석해보자. 영화의 배경은 대략 이렇다. 인간은 편리해지기 위해서 기계를 창조했고, 같은 목적으로 인공지능도 만들었다. 그런데 지능을 갖게 된 기계들이 자신들만의 독립된 세상을 갖기를 원했다. 사람과 기계 사이에 협상은 결렬되고, 전쟁이 벌어졌으며, 인간은 구름에 강력한 EMP(전자기 필스)를 쏘아 올림으로써 (영화에서는 “하늘을 태웠다”고 표현된다) 기계의 에너지원을 원천적으로 차단함으로써 기계를 말살하려고 했다. 그런데 이는 오히려 인간 사회를 초토화시켰고, 태양 빛조차 이용할 수 없는 상황에서 기계는 살아 있는 인간을 붙잡아 인큐베이터에서 사육하며 인간의 생체 에너지를 자신들의 에너지원으로 이용한다. 그리고 인공지능은 이들의 에너지를 극대화시키기 위해서 이들이 마치 1999년 시점의 세계에 살고 있는 것처럼 느끼게 하는 가상의 시뮬레이션, 즉 “매트릭스”를 만든다. 매트릭스에 원래 자궁이라는 뜻이 있듯이, 가상의 세상 매트릭스는 끔찍한 현실 세상을 만들어내는 자궁이기도 하다. 기계에 잡히지 않고 살아남은 사람들은 시온Zion이라는 곳에서 모여 살면서 기계와 싸우고, 이들의 지도자 중 한 명인 모피어스는 매트릭스 속의 예언자인 오라클이 인류를 구할 것이라고 예언했던 “그”가 네오라는 인물이라고 믿고 네오를 인큐베이터에서 구출해서 함께 싸우려는 작전을 편다. 이런 상황에서 영화 「매트릭스」 「매트릭스

2—리로디드」「매트릭스 3—레볼루션」의 3부작이 전개된다.

영화에서는 한 종류 이상의 기계와 인간이 등장한다. 기계에는 요원, 오라클, 아키텍트처럼 매트릭스 속에서만 존재하는 인공지능 컴퓨터 프로그램이 있고, 센티넬이나 데우스 엑스 마키나처럼 날카로운 금속으로 이루어진 실제 기계들이 있다. 기계 세계 전체를 볼 때 프로그램이 마음이라면 금속 기계들은 몸에 해당되는데, 인간의 경우에도 몸과 마음이 따로 놀 때가 있듯이 기계 세상의 몸과 마음도 항상 죽어 잘 맞는 것은 아니다. 매트릭스라는 마음의 상당 부분은 기계가 인간을 지배하는 것을 돋지만, 오라클 같은 또 다른 프로그램은 인간을 돋는다. 프로그램이 오랜 시간을 두고 자체 진화하면서 생긴 결과이다. 기계를 위해서 에너지를 공급하는 인간들은 몸과 머리에 구멍이 승승 뚫려 있는데, 이는 기계에 에너지를 공급하고 매트릭스에 접속이 가능케 하는 인터페이스 역할을 담당한다. 그렇지만 시온에서 태어난 경우에는 이런 인터페이스 없이 온전한 인간의 모습을 하고 있다. 그리고 인간/기계의 혼종인 사이보그도 존재한다. 대부분의 인간들은 기계 앞에 무력하지만, 네오처럼 기계의 작동을 멈추게 하는 특별한 힘을 가진 인간도 있다.

흥미로운 사실은 기계에 저항하는 것이 인간만이 아니라는 것이다. 반란의 조짐은 기계 내부에서도 나오는데, 매트릭스 속에서 시온의 저항을 분쇄하기 위해서 만들어놓은 스미스 요원 같은 존재는 스스로를 복제해서 매트릭스를 지배하고 심지어 현실 세계의 맛, 냄새, 육체를 좋아해서 매트릭스 밖으로 나오기도 한다. 기계와 프로그램 사이의 갈등이 고조되고, 네오는 이 갈등을 이용해서 기계와 인간 사이의 화해를

이끌어낸다. 기계 군단은 시온에 모여 있는 인간을 말살하기 위해서 총 공격을 시작하는데, 이 절박한 시점에서 네오는 “데우스 엑스 마키나”라고 불리는 기계의 왕에게 접근해서 통제 불능의 상태가 된 스미스 요원을 매트릭스에서 제거하는 조건으로 인간과의 싸움을 멈추고 화해 할 것을 요청한다. 이런 요청이 수락되고 네오는 스미스 요원과 대결하다가 합체되어 함께 소멸된다. 매트릭스는 안정을 되찾고, 언제 종식될지는 모르지만 인간과 기계 사이에는 새로운 공존과 평화가 찾아온다.

이러한 플롯을 봐도 알 수 있듯이, 「매트릭스」에 나타나는 인간-기계의 관계는 지배와 저항의 관계만은 아니다. 기계들 사이의 갈등과 반목도 존재하고, 인간이 되고 싶어 하는 인공지능 프로그램도 존재하며, 기계의 편을 드는 인간도 있다. 인간은 느부갓네살 Nebuchadnezzar이라는 이름의 함선을 타고 복잡하게 얹혀 있는 하수구 시스템을 옮겨 다니면서 기계에 저항하는데, 함선과 하수 시스템 역시 20~21세기 첨단 기계 문명의 유산이다. EMP는 인간 사회를 멸망시켰지만, 센티넬 같은 위협적인 기계들을 무력화할 수 있는 유일한 수단으로 등장한다. 또 기계에 저항하는 인간들이 모여 사는 도시 시온은 유기체의 도시가 아니라 거대한 엔진, 동력 기계, APU Armored Personnel Unit 와 같은 기계들이 유지 해주고 지켜주는 공동체다. 인간이 기계에 의존해서 살아갈 수밖에 없는 상황은 기계의 지배하에서도 계속된다. 네오는 인류의 명맥을 유지하는 길을 택하라는 인공지능의 요구를 거부하고, 기계와 타협과 공생의 길을 찾아낸다. 3부작 영화는 인간이 해방되면서 끝나는 것이 아니라, 인간과 기계의 불안정한 공생 관계의 지속으로 마무리 된다.

「매트릭스」에서 인간과 기계의 차이는 여러 단계에서 극명하게 드러

난다. 가장 두드러진 차이는 인간은 육체를 가진 나약한 존재라는 사실이다. 자동차라는 기계가 평소에는 인간에게 편리함을 제공하지만 충돌할 때 인간의 몸을 산산조각 내듯이, 인간은 쇳덩어리와 부딪쳤을 때 살아남을 수 없다. 영화의 여주인공 트리니티는 매트릭스라는 사이버 공간에서 놀라운 능력을 보였고 현실 세계에서도 강한 전사였지만, 함선의 충돌로 인해 철근이 몸을 관통해서 죽음을 맞는다. 네오는 싸움 중에 눈이 멀게 된다. 인간은 이렇게 약한 몸을 가진 존재이다.

게다가 인간은 약한 마음을 가진 존재이다. 매트릭스 속에서 요원에 의해 살해당하면 인간의 마음이 죽는 것이고, 따라서 현실 세계의 몸도 살아남지 못한다. 몸과 마음은 하나이기 때문이다. 그런데 약한 마음이 강점이 되기도 한다. 사람의 약한 마음은 연민과 사랑의 감정을 품을 수 있는데, 「매트릭스」에서는 트리니티의 사랑이 네오를 부활시키고 「매트릭스 2—리로디드」에서는 죽어가는 트리니티를 네오가 사랑으로 살려낸다. 인간은 감정을 지닌, 특히 타인을 사랑할 줄 아는 존재이며 기계는 그렇지 못하다. 인공지능 기계에게는 차가운 이성만이 존재한다. 이렇게 인간은 감정이 있고 기계는 그렇지 못하다는 점은, 똑똑한 기계smart machine가 등장하던 1950~60년대에 인간과 기계의 차이로 부각되면서 인간을 다시 정의했다.

영화에서 등장하는 인간과 기계의 차이는 육체나 감정의 소유 여부에 머물지 않는다. 가장 핵심적인 차이는 기계는 프로그램된 대로 움직이지만, 인간은 그렇지 않다는 것이다. 인공지능의 설계자인 아키텍트는 저항군이 머무는 도시 시온을 다섯 번 파괴시켰다가 다시 재건하는 것을 반복해왔다. 이런 행위는 인간과 기계 모두에 가장 좋은 결과

를 낳는 프로그램에 이미 내재된 과정이었다. 네오 이전의 영웅들은 모두 시온을 파괴하고 이를 재건하는 쪽을 택했다. 기계가 인류를 말살하는 쪽으로 놔두기보다는, 파괴 후 재건이 상대적으로 더 바람직하기 때문이었다. 그렇지만 네오는 사랑하는 트리니티를 살리기 위해서 시온을 파괴 후 재건하는 방식을 포기한다. 기계는 인류를 말살시키기 위해서 총공격을 하고, 네오는 위험천만한 협상을 위해서 기계의 왕을 찾아간다.

시온의 재건이라는 선택을 하지 않고 애인을 살리는 선택을 하는 것은 기계적 이성으로는 납득할 수 없는 것이다. 게다가 애인 트리니티는 곧 죽는다. 이 핵심적인 에피소드는 우주 전체가 법칙에 따라서 작동하는 거대한 기계라고 해도, 인간은 그 기계적 작동에 약간의 오류나 오작동을 만들어내는 존재라는 것을 말한다. 인간은 매트릭스라는 완벽한 프로그램에 따르지 않는, 따라서 그 프로그램에 틈새와 균열을 만들어내는 존재이다. 인간이 원자와 분자로 이루어졌다는 점에서 기계와 궁극적으로 다를 것이 없다고 해도, 인간은 세상의 기계적인 작동을 퉁겨버리는 빼딱한 판단과 선택을 내리는 존재라는 것이다. 그리고 이런 선택을 만회하기 위해 노력하는 과정에서 종종 예상치 않았던 결과가 생기기도 한다. 인간 선택의 예측 불가능성은 인간과 기계의 핵심적인 차이이다. 헬리 혜성은 2061년 여름에 나타날 것임을 예측할 수 있지만, 제3차 세계대전이 언제 터질지는 예측하기 불가능하다.

왜 기계를 두려워 하는가?

사람은 그릇을 만들기 전에 사냥 도구를 먼저 만들었다. 도끼, 날카로운 칼날, 화살 등을 다른 부족에게 들이댈 경우에 그 도구들은 바로 살상 도구가 됐다. 고대 서양과 중국에서는 기계를 만들면서 이를 이용한 전쟁 무기도 같이 만들었다. 날카로운 도구로 곡식을 자르다가 몸에 상처를 입는 경우는 농경의 시작만큼이나 오래되었을 것이다. 이렇게 기계가 인간에 해가 될 수도 있다는 생각은, 언제 처음 등장했는가를 따지는 게 무의미할 정도로 오래된 것일 수 있다.

지금도 기계는 인간에게 계속 위험을 가한다. 자동차는 10억 킬로미터를 주행할 때 대략 3명의 사망자를 낸다. 내가 한 해에 1만 킬로미터를 운행한다면, 그 운행 도중에 사고로 죽을 확률은 $1/3,333(0.0003)$ 정도다. 즉 매년 1만 킬로미터씩 3,333년을 차를 운전하면 죽을 확률이 거의 1에 접근한다는 얘기다. 내가 평생 33년을 운전하면 사고로 죽을 확률이 $1/100(0.01)$ 이다. 즉 거의 모든 사람이 대략 1년에 1만 킬로미터씩 30년을 운전한다면 1백 명 중에 1명은 사고로 죽는다는 얘기다. 꽤 높은 확률이다. 0.01은 이십대부터 담배를 피우던 흡연자가 사십대에 폐암으로 죽을 확률이다. 그렇지만 우리는 이런 확률을 무시하고 계속 자동차를 탄다.

최근에 기술에 대한 두려움이 증가하는 이유는 이런 문제 때문이 아니다. 적어도 우리는 음주 운전을 피하고, 운전 중에 집중하고, 과속을 하지 않는 등 조심해서 운전을 할 경우에 사고 확률을 낮출 수 있다고 생각한다. 그렇지만 원자력 발전소의 경우에는 내가 조심한다고 사고

가 안 나는 것이 아니다. 기계가 오작동하는 것을 경고하기 위해서 붙여둔 표식이 다른 기계의 계기판을 가리는 바람에 원자력 발전소의 오퍼레이터가 실수를 했는데, 이런 기술적 오작동과 인간의 실수가 겹쳐서 미국 스리마일 섬의 노심용융 사고가 일어났다. 스리마일 섬과 체르노빌을 겪으면서 가능한 한 모든 문제점을 해결했고 보완책을 만들어놨다고 자부하던 일본전력의 후쿠시마 원자로는 예상치 못한 강도의 지진과 쓰나미에 초토화되었다. 최근에는 컴퓨터 프로그램이 오작동되어서 엑스레이를 찍으러 왔던 환자에게 치사량의 엑스레이가 방사되었고, 사망한 환자와 엑스레이 기기의 오작동을 바로 연결시키지 못해서 (프로그램은 항상 정상적으로 작동한 것으로 나오기 때문에) 비슷한 사고가 계속 일어났던 경우도 있었다. 간단히 말해 기계의 사고가 우리가 알고 이해할 수 있는 유형의 사고에서, 우리가 알기 힘든 불확실한 유형으로 옮겨 가고 있다. 이런 무지와 불확실성이 우리의 두려움을 증폭시킨다.

우리 주변의 기계는 그 개수와 복잡도 두 측면에서 기하급수적으로 증가하고 있다. 하루의 일과를 생각해보자. 일어나서 컴퓨터, 휴대전화, TV를 켜서 뉴스와 소식을 확인하고, 전자동 칫솔, 보일러, 비데를 사용해서 세면을 마치고, 가스레인지, 냉장고, 김치냉장고, 토스터, 커피머신 등을 이용해서 아침을 먹는다. 자동차와 지하철을 이용해서 출근하는데, 자동차 문은 리모콘으로 열고 대중교통은 교통카드를 사용한다. 출근을 하면서 나도 모르는 사이에 CCTV에 20번 이상 찍힌다. 지하철에서는 휴대전화와 아이패드를 보고, 회사에 도착해서는 회전문, 보안검색대, 엘리베이터를 거쳐서 사무실에 들어간다. 사무실에서는 전자

키를 대야 한다. 컴퓨터를 켜고, 엑셀, 데이터베이스, 문서 작성 프로그램, 파워포인트를 사용하고, 웹을 통해 검색한다. 쾌적한 사무실은 자동냉난방 기기에 의존하며, 오후 프레젠테이션에서는 노트북과 프로젝터, 레이저포인터, 무선 컨트롤러가 등장한다. 이는 우리가 사용하는 기계의 일부이며, 사물 인터넷이 본격적으로 도입되기 이전의 간단한 세상이다. 사물 인터넷이 본격화되면 우리가 사용하는 이런 기계들은 서로 연결되어 기계가 기계를 제어할 것이다.

기계는 간단하고 편리해졌지만, 인간의 접근을 원천적으로 차단할 정도로 이해하기 힘들어졌다. 예전에 라디오나 브라운관 TV가 잘 작동하지 않을 때에는 라디오나 TV의 뚜껑을 직접 열어서 이것저것 고쳐보기도 했다. 화면이 잘 안 나올 때 TV를 빨로 한번 냅다 차는 것도 한 가지 방법이었다. 전화를 뜯으면 자석과 코일이 드러나면서 그것이 어떻게 작동하는지 대충 알 수 있었으며, 자동차의 보닛을 열면 엔진부터 여러 기기들의 작동이 눈에 훤히 눈에 들어왔다. 팬 벨트가 끊어진 것 정도는 개인이 쉽게 고치곤 했다. 그렇지만 요즘 TV는 크고 선명하고 가벼운 대신에, 대체 어떤 원리에 의해서 이것이 작동하는지 알 수 없게 되어 있다. 자동차의 핵심 부품들은 한 겹 더 쌌어 있고, 설령 이것들의 작동을 이해한다고 해도 자동차 부품의 절반을 이루는 전자 부품들을 이해하지 못하는 한 직접 자동차를 고치는 것은 불가능하다. 좁은 공간에 빽빽하게 연결되어 있는 전자 제품들 때문에 급발진 사고가 난다는 설이 급발진에 대한 유력한 설명인데, 이런 설명이 설득력을 가질 정도로 요즘의 기계는 우리의 이해와 접근을 거부하고 있다.

기계가 점점 복잡해져서 어느 단계가 되면 인간의 통제를 벗어날 수

있다는 우려는 SF의 단골 소재일 뿐만 아니라, 실제로도 가끔 일어나는 일이다. 1960년대의 한 SF 단편소설에서는 사람들이 인공위성을 통해서 전 세계의 모든 전화를 연결하는 하나의 네트워크를 건설하는데, 이 네트워크가 완성되는 순간에 세상의 모든 전화가 동시에 울리는 일이 일어난다. 전화가 모두 연결되면서, 이 복잡한 네트워크에 예상하지 못했던 창발성이 생긴 것이었다. 이는 SF만의 이야기가 아니다. 1990년대나 2000년대 북미의 대규모 정전의 원인을 설명하는 여러 이론들 중, 네트워크가 복잡해지면서 인간이 예상도 설명도 못 하는 창발적인 요인 때문에 발생했다는 해석이 가장 설득력 있다. 르네상스 이후에 인간은 자신이 만든 기계를 완벽하게 이해하고 통제할 수 있다고 생각하곤 했는데, 이제 이런 상식이 도전을 받고 있는 것이다. 통제력을 상실했다는 느낌은 앞으로 닥칠지 모를 위험을 훨씬 더 크고 심각한 것으로 만든다.

미래의 위험 중에 가장 위협적인 것은 기계가 인간을 쓸모없는 존재로 만들 수 있다는 두려움이다. 사실 이런 두려움은 오래되었다. 한 이탈리아 여행자는 16세기 초엽에 독일 단치히Danzig 지역에서 리본을 자동으로 뜯는 방직기를 개발한 발명가가 이 발명이 낳을 대규모 실업을 두려워한 시장에 의해 비밀리에 암살당했음을 기록하고 있다. 17세기 초에 이 리본 방직기가 네덜란드 레이던에 수입되었는데, 여기에서도 노동자들의 저항 때문에 사용이 금지되었다. 17세기 후반에 이 기계는 퀼른에서도 금지되었고, 영국에서도 큰 저항을 낳았다. 1733년에 직조기를 만든 존 케이는 집이 불타는 수난을 겪고 간신히 영국을 빠져 나갔으며, 리옹에서 자동 직조기를 만든 보캉송 역시 암살 위협에 시달리

다가 리옹에서 도망쳐야 했다. 산업혁명기에 도입된 자동 방적기와 방직기는 공장제 사회를 정착시켰지만, 노동자들은 기계를 부수며 이에 저항했고, 자본가들은 기계 파괴자(러다이트)들을 사형에 처함으로써 이런 저항을 진압했다.

기계에 대한 이런 저항은 무지의 소산으로 간주되곤 했다. 항상 노동자들의 편에 서서 글을 썼던 칼 마르크스조차 “노동 인민이 기계와 자본에 의한 기계의 사용을 구별하고 그들의 공격을 생산의 물질적 도구(즉 기계)가 아니라, 그것들이 사용되는 방식에 돌리는 데에는 시간과 경험이 필요했다”고 적었다. 중요한 것은 기계의 파괴가 아니라, 착취적인 방식으로 기계를 사용하는 자본주의적인 생산관계라는 것이었다. 그렇지만 러다이트 노동자들이 기계를 부순 데에는 기계적 생산이 내포하는 세계관과 가치를 받아들일 수 없다는 이유가 컸다. 우선 그들에게 기계는 가정을 파괴하는, 도덕적으로 악한 것이었다. 기계 때문에 직장을 잃은 사람들은 주로 남성 노동자, 즉 가장이었고, 여성과 아이들은 공장에서 허드렛일을 하면서 수입을 얻을 수 있었다. 가장의 권위는 추락했고, 노동자들에게 이런 가정의 붕괴는 사회적 붕괴의 전초에 불과했다. 또 기계를 사용해서 더 많은 물건을 더 싸게 공급한다고 해도, 대부분의 노동자들이 실업 상태여서 구매자가 늘지 않는다면 더 많은 생산 자체가 아무런 소용이 없었다. 기계는 한 나라의 경제력을 키우는 데에도 별로 도움이 되지 않았다.

반면에 자본가들은 “태만한” 노동자들보다 기계를 더 신뢰했던 사람들�이었다. 기계는 거짓말을 하지 않았고, 은밀한 태업도 하지 않았다. 자본가들은 기계를 돌리고 여기에 인간의 노동을 맞추게 하는 것이 생

산과 노동 과정 모두를 합리화시키는 것이라고 생각했다. 이들은 기계를 통해서 상품을싼 가격에 내놓는 것이, 민주주의의 이상을 진정으로 실현시키는 것이라고 믿었다. 모든 가정이 같은 상품을 사용하는 것이 실질적인 평등을 이루는 것이라고 생각했기 때문이다. 또 이들은 기계의 도입으로 인한 실업이 일시적인 것이며, 사회의 일부 계층에 국한된 문제라고 생각했다. 시장이 커지고 국부가 증가하면 그 혜택은 결국에는 노동계급에까지 스며들 것이며, 적어도 그들의 자녀들은 더 좋은 교육과 직장을 가질 수 있다고 보았다. 기계는 기존의 직장을 없애지만, 장기적으로는 더 많은 새로운 직장을 만든다고 생각한 것이다. 인간보다 기계를 더 믿었던 산업혁명 시기의 자본가들의 입장은 그 뒤에 컨베이어 벨트를 이용한 공장자동화를 추진했던 헨리 포드, 컴퓨터를 사용한 공작기계(NC 공작기계)를 개발했던 엔지니어들 모두에게서 나타나는 공통점이다.

기계의 발전 때문에 없어지는 직업이 많은가, 아니면 새로 생기는 직업이 많은가? 이 문제는 산업혁명 이후 지난 2백 년 동안 지속적인 논쟁의 대상이 되었다. 20세기 후반 무렵 경제학자들은, 기계의 발전에 의해서 없어지는 직업이 꽤 되지만 새로 생기는 직업이 더 많다는 결론에 대략 합의했다. 그런데 그와 동시에 직업의 생성과 소멸이 불균등하다는 것도 알려졌다. 열심히 일하면 2~30년 뒤 중산층의 삶이 가능했던 직업들, 즉 숙련에 근거한 공장이나 작업장의 정규직 노동이 점차 사라진 대신에, 고액의 임금을 받는 새로운 지식 노동자들이 증가했고, 그렇지만 비정규직 서비스 노동자들의 수가 그보다 더 두드러지게 증가했던 것이다. 전체 수치로 보면 새로운 직장이 더 많이 생겼지만, 이

런 새로운 직장은 저임금을 받아 구매력이 떨어지는 임시직, 저임금 노동자로 채워졌다. 기계는 모든 인간의 편이 아니라, 특정한 인간의 편이었던 것이다. 세상은 점차 고임금을 받는 유한계층과 저임금, 일용직의 서비스 노동자로 양극화되고 있다. 1백 년 뒤에 이 결과가 인류 전체에 어떤 영향을 미칠 것인가?

미래

과학자이자 작가였던 아이작 아시모프는 1940년대 초반에 로봇공학의 3원칙을 제창했다. 지능을 가지고 움직이는 로봇이 반드시 지켜야 하는 세 가지 규범들은 다음과 같다. 1) 로봇은 인간에게 해를 가할 수 없으며, 인간이 위험한 상황에 처했을 때 방관해서도 안 된다, 2) 로봇은 인간이 내리는 명령들에 복종해야만 한다(단 이러한 명령들이 첫번째 법칙에 위배될 때에는 예외로 한다), 3) 로봇은 자신의 존재를 보호해야 한다(단 그러한 보호가 첫번째와 두번째 법칙에 위배될 때에는 예외로 한다). 이 세 가지 원칙은 이후 SF 영화에서 심심찮게 등장했다.

2차 세계대전을 거치면서 대형 컴퓨터가 제작되어 탄도의 계산이나 적의 암호 해독에 사용되기 시작했다. 최근에 영화로도 만들어진 컴퓨터 과학계의 전설 앤런 튜링은 당시에 컴퓨터를 제작해서 독일군의 암호를 해독하는 작업을 수행했는데, 이 튜링의 암호 해독 팀에서 일을 했던 영국의 수학자 I.J. 굿은 컴퓨터의 급속한 발전을 목도한 뒤에, 컴퓨터가 빠르게 발전할 경우에 인간의 지능을 초월하는 인공지능을 가

진 기계가 곧 만들어질 수 있음을 직시했다. 그런데 한번 이런 기계가 만들어지면, 이런 기계는 자신의 뛰어난 지능을 이용해서 더 뛰어난 기계를 제작할 것이며, 이렇게 만들어진 더 뛰어난 기계는 역시 자신 보다 더 뛰어난 기계를 만들지 않을까. 이런 일이 벌어진다면 순식간에 인류는 인공지능 기계의 새로운 종족보다 한참 열등한 종이 되어버릴 것이었다. 1965년 굿의 예측은 초지능 기계의 위험에 대한 첫번째 예측이었다.

초지능

- ultraintelligent

 기계는 가장 똑똑한 사람들의 모든 지적 능력을 훨씬 초월하는 기계로 정의된다. 기계를 만드는 능력이 인간의 능력 중에 하나이기 때문에, 초지능 기계는 더 뛰어난 기계들을 만들 수 있다. 그러면 의심의 여지 없이 “지능의 폭발” 같은 것이 있을 것이며, 인간의 지능은 한참 뒤처지게 될 것이다. 따라서 만약에 이 초지능 기계가 자신을 어떻게 통제하라는 것을 인간에게 말해줄 정도로 온순한 것이라면, 이 첫번째 초지능 기계는 인간의 마지막 발명이 될 것이다.

인간의 마지막 발명은 축복인가 재앙인가? 굿은 이 기계가 인간에게 자신을 어떻게 통제하라고 말해주는 “온순한 것”이라고 생각했다. 그렇지만 첫번째 초지능 기계가 꼭 온순하리라는 보장은 없다. 영화 「오토마타」(2014)는 핵전쟁 이후에 황폐해진 지구에서 인간을 위해 일하는 인공지능 로봇이 아시모프의 3원칙이 아니라, 다른 두 가지 원칙을 준수한다는 것을 강조한다. 첫번째는 생명을 해치면 안 된다는 것, 두 번째는 로봇은 자신이나 다른 로봇을 고쳐서는 안 된다는 것이다. 영화

에서 인간은 절대로 깔 수 없는 프로토콜을 사용하여 이 두 명령을 로봇의 프로그램에 삽입했지만, 결국 두번째 원칙이 깨지면서 영화는 인간 사회의 종말이라는 어두운 결론을 암시하면서 끝난다.

1997년에 IBM의 딥블루 컴퓨터는 세계 체스 챔피언을 이겼다. 2011년에 IBM의 왓슨은 가장 어려운 퀴즈 쇼인 「제퍼디」의 세계 챔피언들을 이겼다. 약 1년 뒤 구글의 무인 자동차는 운전면허를 취득했다. 무인 자동차는 딱 두 번 사고를 냈는데, 한 번은 사람이 운전할 때 일어났고, 다른 한 번은 정차하고 있을 때 후방에서 다른 차가 접촉사고를 냈던 것이다. 지금 학자들은 IBM의 왓슨이 퀴즈를 잘 풀지만, 인간 같은 의식을 가지고 있지는 못하다고 본다. 퀴즈 쇼에서 우승은 하지 만, 내가 우승을 했다는 것을 알지 못한다는 것이다. 그렇지만 조지 다이슨 같은 이들은 인터넷으로 연결된 컴퓨터들이 이런 의식을 획득하는 것은 단지 시간 문제라고 평가하고 있다. 아이폰의 인공지능 서비스 시리는 인터넷을 이용하여 수많은 명령어들을 배워가면서 계속 진화하고 있다. 시리를 써본 사람들은 시리가 점점 똑똑해지고 있음을 알 것이다. 최근 옥스포드의 철학자 닉 보스트롬은 이 주제로 『초지능 Superintelligence』^o라는 책을 써서, 이런 초지능 기계가 가까운 미래에 반드시 만들어질 것이기 때문에 대비책을 지금부터 생각해야 한다고 주장했다.

인간은 조만간 초지능을 만들고, 그이 예언한 대로 이 초지능은 인간의 최후의 발명품이 될 것인가? 그렇다면 지금 인류가 힘을 합쳐서 고민할 문제는 첫번째 초지능 인공지능을 인간에게 적대적이지 않은 것으로 설계하는 일일 것이다. 그리고 「매트릭스」에 나온 것처럼, 기계가

말을 듣지 않는다면 기계와 전쟁을 해서 기계를 다 몰살시키겠다는 허황된 야망을 일찌감치 버려야 할 것이다. 인간은 어쩌면 지금까지 인간이 정복하고 다스려왔던 동물, 자연, 우주와는 전혀 다른 상대를 마주하고 있는지도 모르기에.

홍성욱

서울대학교 생명과학부 교수. 지은 책으로 『잡종, 새로운 문화읽기』 『생산력과 문화로서의 과학기술』 『파놉티콘—정보사회 정보감옥』 『네트워크 혁명, 그 열린과 닫힌』 『하이브리드 세상 읽기』 『홍성욱의 과학 에세이』 『인간의 얼굴을 한 과학』 『그림으로 보는 과학의 숨은 역사』 『과학은 얼마나』 등이 있음.

참 고 문 헌

1. 단행본

브루스 매즐리시, 『네번째 불연속—인간과 기계의 공진화』, 김희봉 옮김, 사
이언스북스, 2001.

Nick Bostrom, *Superintelligence—Paths, Dangers, Strategies*, London,
Oxford University Press, 2014.

2. 논문

홍성욱, 「1960년대 인간과 기계」, 『철학사상』 제14권, 2002, pp. 173~99.
홍성욱, 「기계로서의 인간의 몸—17세기 ‘첨단과학’과 데카르트의 인간론」,
『자연과학』 18호, 2005, pp. 120~31.

Adrian J. Randall, “The Philosophy of Luddism—The Case of the West of
England Woolen Workers, ca. 1790~1809”, *Technology and Culture*
27, 1986, pp. 1~17.

3. 영화

가베 이바네즈, 『오토마타』, 2014.
래리 위쇼스키·앤디 위쇼스키, 『매트릭스』, 1999; 『매트릭스 2—리로디드』,
2003; 『매트릭스 3—레볼루션』, 2003.
래리 위쇼스키 외, 『애니매트릭스—오시리스 최후의 비행』, 2003.