

전 구

** 아시모프, 클라크, 하인라인의 작품들에 대한 스포일러가 여럿 포함되어 있습니다.

과학 기술을 넘어 사회 모든 분야에 걸쳐 인공지능에 대한 관심이 그 어느 때보다도 높아 보인다. 사람을 대신해 그림을 그리고, 글을 쓰고, 사람들의 궁금함을 풀어주는 역할도 자연스럽게 수행해 낸다. 나 자신도 대학 시절 신경회로망에 대한 연구를 처음 접한 후 항상 관심을 가져 왔으며 박사 과정에서도 기계 학습을 전공했지만, 그 이후 지난 15년간의 변화를 보면 그 발전 속도가 경이로울 지경이다. 한편 이런 빠른 발전에 대한 우려의 목소리 또한 심심치 않게 들려오곤 한다.

인류의 과학 기술이 발전하면서 사람처럼 생각하고 행동하는 기계에 대한 상상은 항상 끊이지 않아 왔다. 사람의 외형과 행동을 모방할 수 있는 자동 인형장치는 이미 수세기 전 등장하였으며, 이후 컴퓨터의 발전과 더불어 과학소설 (SF) 속에서 사람처럼 생각하는 로봇 혹은 컴퓨터의 존재는 빼놓을 수 없는 존재가 되었다. 그리고 요즘의 인공지능에 대한 우려처럼 SF 속 인공 지능의 묘사 또한 항상 긍정적인 것은 아니었다. 20세기 초, '로봇'이라는 단어가 최초로 사용된 체코의 작가 카렐 차펙의 희곡도 인류의 노동을 대신한 로봇들이 반란을 일으켜 인류와 싸우는 이야기를 다루고 있다. 그렇다면 인공지능과 기계 학습에 대한 기반 이론이 처음 확립되어 가던 20세기 중반, 당시의 SF 거장들이 묘사한 인공지능은 과연 어떤 모습이었을까?

인공지능의 위험성에 대한 우려에 대해 SF 팬으로 가장 먼저 떠오르는 것은 뭐니뭐니해도 아이작 아시모프가 만든 로봇 3원칙이다. 널리 알려져 있듯 로봇 3원칙은 다음과 같다. 첫째, 로봇은 사람을 다치게 하거나 위험에 방치해선 안된다. 둘째, 로봇은 사람의 명령을 들어야 한다. 셋째, 로봇은 자기 자신을 지켜야 한다. 여기서 하위 원칙은 상위 원칙을 벗어나지 않는 한에서만 적용 가능하며, 이 원칙들은 어길 경우에는 회로가 망가지도록 로봇의 인공지능을 구성하는 양전자 (positron) 두뇌에 절대적인 행동

원칙으로 각인되어 있다. 물론 양전자 회로라는 것이 대체 어떤 것인지, 혹은 로봇 자신의 행동을 스스로가 평가하여 발동되는 안전장치는 자가 모순이 아닌지 같은 질문들은 설명되지 않은 채 소설적 장치로서만 작동하지만, 타고난 이야기꾼이던 아시모프는 1940년대 이후 이 장치들을 활용하여 많은 명작을 써냈다.

그 중에서도 나의 기억에 가장 깊이 남아 있는 책은 역시 가장 먼저 읽었던 단편집 '로봇 머신 X'를 꼽고 싶다. 이 책은 원래 아이, 로봇 (I, Robot, 1950)이라는 단편집에 속한 이야기 중 네 편을 각색하여 아이디어 회관에서 아동용 SF 전집 중 한 권으로 펴낸 책으로, 초등학교 시절 학급 문고에서 빌려 읽은 뒤 수십년이 지난 지금까지도 그 안의 삽화까지 기억날 정도로 푹 빠져들었던 책이었다. 이 이야기들에서 로봇의 역할은 아이를 돌보고 수성에서 광물을 채취하는 물리적인 노동에서부터, 연애 상담과 세계 정부 운영과 같은 복잡한 일까지 이르게 된다. 모든 로봇은 3 원칙에 따라 작동하고 있지만 안전장치로 짜넣은 그 원칙들이 때로는 알 수 없는 행동을 하게 만드는 원인이 된다는 점이 이후 아시모프의 로봇 시리즈의 독특한 이야기거리가 된다. 가장 마지막에 실린 이야기이자 이 책의 제목이 된 머신 X에 대한 이야기가 특히 흥미로운데 (원제: 피할 수 있는 갈등, The Evitable Conflict, 1950), 이 로봇은 사람처럼 생긴 몸을 갖고 있지 않지만 사람의 질문에 대해 답을 제공하는, 로봇이라기 보다는 인공지능 컴퓨터라고 해야 마땅할 존재이기 때문이다. 또 이 단편에서 머신 X는 로봇 3 원칙을 문자 그대로 해석하기보다는 1 원칙을 인류 전체에 대한 이익으로 확장해 해석하며, 몇십년 후 아시모프가 그의 파운데이션 시리즈와 로봇 시리즈를 연결하기 위해 만들어낸 로봇 제 0 원칙을 암시하는 듯한 모습을 보여준다. 참고로 로봇 제 0 원칙은 다음과 같다: 로봇은 인류를 해쳐서는 안되며, 인류에 해가 될 일을 방관해서도 안된다. 즉 일부 사람에게 당장의 해가 되더라도 인류 전체에 걸친 이익이 더 큰 일이 있다면 로봇이 1 원칙을 무시하는 것이 가능해진다.

아시모프가 활발히 작품 활동을 하던 1950-1960년대는 세계대전이 끝나면서 현대 과학기술이 눈부신 발전을 보이던 시대이다. 국가와 대륙 사이의 물리적인 거리가 운송수단과 통신망의 획기적인 발전을 통해 좁혀지게 되었고 지구를 벗어나 우주로까지 인류의 진출이 시작되었다. 또한 컴퓨터 기술과 이론이 비약적으로 발전함과 동시에, 신경세포인 뉴런들이 어떻게 서로 연결되어 있고 어떻게 학습이 이루어지는지 대한

이론이 알려지고 이를 흉내낸 인공 신경망(artificial neural network)에 대한 연구가 처음 이뤄진 시기이기도 하다. 이 당시 등장한 인공신경망 구조는 동시에 자극되는 시냅스가 강화되는 생물학적 현상에 착안한 헤비안 학습(Hebbian learning)과, 주어진 예제의 정답에 가까워지도록 입력의 가중치(weight)를 조정하는 퍼셉트론 (perceptron) 알고리즘이 있었다. 이 두 방법은 모두 주어지는 입력과 출력에 따라 인공 신경망 내부의 연결 가중치가 선택적으로 조절되는 학습 방법을 이용한다는 공통점이 있다.

한편 같은 시기에 20 세기 초반에 장르를 확립한 과학 소설 (SF) 또한 황금기를 맞으며 수많은 작가들을 배출했다. 이 무렵부터, 스페이스 오페라로 불리며 낮은 평가를 받던, 미래나 우주가 배경이라는 것 외에는 여타 통속 소설과 차이가 없던 이전 시대의 소설들보다 훨씬 높은 수준의 과학적 기반과 현실적인 묘사들, 그리고 기발한 상상이 결합된 작품들이 등장하기 시작했다. 많은 SF 작가들 중에서도 특히 큰 인기를 얻은 세 명의 작가가 아이작 아시모프, 로버트 A. 하인라인, 그리고 아서 C. 클라크였다. 흔히 SF의 세 거장으로 불리는 이들은 과학에 대한 깊은 이해와 소설적 재미를 모두 갖춘 작품들을 다수 써냈다. 이들의 작품 속에는 인공지능 또한 다양한 관점에서 다루어지고 있는데, 전혀 다른 작품 속에서도 공통된 아이디어를 찾을 수 있다는 것이 또한 흥미롭다.

소설적 설정 속에서도 과학적 법칙에 의한 전개를 중요시하는 하드 SF 로 특히 유명했던 클라크는 자신의 단편을 기초로 하여 스탠리 큐브릭 감독과 함께 2001 년 스페이스 오디세이라는 명작을 만들어내게 된다. 이 영화는 우주 탐사에 대한 사실적이면서 자세한 묘사로 유명하지만, 이 이야기 속에서 가장 긴장을 불러 일으키는 존재는 목성 탐사 임무의 주도권을 놓고 주인공과 대결하는 인공지능 컴퓨터 HAL 9000 이다. <그림 2> 이야기의 시작과 끝이 인류의 진화와 그 너머로 확대되는 인식을 다루고 있다는 점에서 이야기의 클라이맥스에 인공지능과 인류의 갈등이 자리잡고 있다는 점이 의미심장하다. 2001 년 스페이스 오디세이는 두말할 것 없는 명작이지만, 사실 인공지능을 소재로 내게 가장 깊은 인상을 남긴 클라크의 작품은 따로 있는데, 대학 시절 SF 번역 모임을 통해 접하게 된 ‘프랑켄슈타인은 다이얼 F 로 (Dial F for Frankenstein, 1964)’라는 단편이다. 알프레드 히치콕의 영화 ‘다이얼 M 을 돌려라(Dial M for Murder)’와 메리 셸리의 프랑켄슈타인에서 따온듯한 제목의 이 이야기는 어느 날 전 세계의 전화기가

갑자기 울려대기 시작하면서 시작된다. 전 세계를 연결한 자동 전화교환망이 하나의 거대한 두뇌처럼 작동해 지능을 갖게 되었기 때문인데, 여기에 대해 클라크는 극중 인물의 대화를 통해 인간의 뇌와 전화 교환망의 유사성을 지적한다. 사람의 두뇌도 결국은 뉴런이라는 스위치가 서로 연결된 존재이고, 전화 교환망 역시 수많은 스위치들이 전선을 통해 연결되어 있기에, 전화 교환망의 스위치 개수가 뇌의 뉴런 수보다 많아지면서 지능이 깨어나는 일이 가능해졌다고 묘사된다. 물론 인공신경망에서처럼 뉴런의 가중치가 주어지는 입출력을 통해 강화되는 작용이 전화 교환망에서는 일어날 수 없기에 단순히 스위치의 수가 늘었다고 하여 지능이 생기는 것은 불가능한 일이며, 작가도 그런 사실을 충분히 알고 있었을 것으로 생각된다. 이 단편이 기억에 남는 이유는 현실에 존재하는, 사람이 만들어낸 복잡한 네트워크와 인간의 뇌 사이의 복잡도를 연결지어 짧으면서도 인상적인 이야기를 구상해냈기 때문이다.

역시 영화화된 스타쉽 트루퍼즈(Starship Troopers, 1999)로 가장 널리 알려진 하인라인은 자신의 군 경험과 정치 사회에 대한 독특한 관점으로 전쟁 및 다양한 사회상을 다루는 SF 소설들을 집필하였다. 그런 작품들 중 하나인 달은 무자비한 밤의 여왕(The Moon Is a Harsh Mistress)은 지구의 유형지로 달에 건설된 식민지를 배경으로 한다. 지구와는 전혀 다른 사회로 발전한 달 식민지의 사람들이 지구를 상대로 독립하기 위한 전쟁을 벌이는 이야기인데, 주인공이 속해 있는 소수의 비밀 결사가 거대한 지구를 상대로 전쟁을 벌이는 데 가장 큰 역할을 담당하는 것이 바로 달 전체의 모든 통신 및 제어 체계를 관리하는 슈퍼 컴퓨터 마이크(Mike)이다. 마이크는 본래 홈즈 IV 시스템(HOLMES IV, High-Optional, Logical, Multi-Evaluating Supervisor, Mark IV)이라 불리던 슈퍼컴퓨터인데, 이 컴퓨터를 구성하는 기본 소자인 뉴리스트어(neuristor)의 수가 인간의 뇌의 뉴런의 수보다 많아지면서 지능을 획득하게 된다. 그리고 주인공 및 소수의 인원만이 이 사실을 알게 된 후, 설록 홈즈의 형인 마이크로프트 홈즈의 이름을 딴 마이크라는 이름을 붙이고 동료가 되어 지구와의 싸움에 나서게 된다. 재미있는 점은 이렇게 고도로 자동화된 통신 시스템이 있는 달에서도 마이크와의 정보 교환은 오직 유선 전화를 통한 대화로만 가능하다는 점이다. 뉴리스트어라는 이름의 인공 신경망과 비슷한 소자를 가정한 것을 제외하면 위에 언급한 클라크의 단편과 여러 공통점이 엿보이는데, 단순한 소자로 구성된 네트워크의

복잡도가 증가하면서 지능이 생겨난다는 점, 그리고 전화 교환망이 이런 네트워크를 구성하는 데 핵심적인 역할을 한다는 점이다. 20세기 중반 당시에 전세계를 연결하는 자동 전화 교환망의 구성이 사람들에게 얼마나 크고 중요한 변화로 다가왔는지를 짐작할 수 있는 부분이다.

모든 작가들이 인공지능을 호의적으로 혹은 적극적으로 다루었던 것은 아니다. 비교적 최근에 재영상화로 화제를 모은 프랭크 허버트의 '들펀 (Dune, 1965)' 시리즈는 항성간 여행을 비롯해 과학기술이 마법처럼 보이는 지경에 이른 사회가 배경이지만 인공지능이나 발달된 컴퓨터는 전혀 등장하지 않는다. 그 이유는 이 소설의 배경이 되는 시대보다 아주 오래 전, 인류가 컴퓨터와 로봇에 지나치게 의존하게 되며 일어난 갈등과 싸움으로 모든 로봇과 컴퓨터가 파괴되었으며, 그 이후 고성능의 컴퓨터나 인공지능을 만드는 일 자체가 금기가 된 은하계를 배경으로 하고 있기 때문이다. 어떻게 보면 인공지능의 발전을 극단으로 가정한 소설이라고 할 수도 있겠지만, 소설의 귀족적, 중세적 인물상과 맞지 않는 장치를 작가의 선택으로 통제로 배제하기 위한 작가의 불가피한 선택은 아니었나 생각된다.

마지막으로 다시 아시모프의 다른 작품들에 대한 짧은 이야기. 아시모프의 단편들에는 사람과 어울리거나 싸우는 인격적인 자아를 가진 인공지능과 달리, 단지 질문에 대답하는 것이 역할의 거의 전부인 인공지능도 등장한다. 이 인공지능은 최초의 컴퓨터 시스템 중 하나였던 유니백(UNIVAC)을 패러디한 멀티백(Multivac)이라는 이름을 가진 컴퓨터인데, 이 컴퓨터는 로봇 시리즈와는 전혀 다르지만 역시 현재의 세계에서 연결되는 세계를 배경으로 하기에 로봇 3원칙의 제약을 받지 않는다. 장난으로 지어진 것만 같은 이름에서 상상되는 이미지와는 달리 멀티백은 이 세상 모든 농담의 근원을 탐구하기도 하고 (농담꾼, Jokester, 1956), 지금까지 묘사된 어떤 인공지능 컴퓨터들보다도 더 거대한 질문에 대한 답을 풀어내게 된다 (최후의 질문, The Last Question, 1956). 멀티백이 등장하는 작품들은 모두 단편이며 결말의 반전이 충격적인 작품들이기에 그 줄거리에 대한 언급은 최대한 삼가는 것이 좋겠다. 어떻게 보면 이 멀티백이야말로 우리가 익숙한 ChatGPT와 비슷한 존재라고 하겠다. 물론 현실의 인공지능 모델들은 아직 그렇게 거대한 질문들에 답을 제공하지는 못하지만, SF라면

얼마든지 생각해볼 수 있다. 우리가 지금 보는 인공지능 기술들이 더 발전해서 더 많은 일들에 대해 답을 줄 수 있게 된다면, 과연 인공지능은 우리의 어떤 질문에까지 답을 줄 수 있을까? 그리고 어떤 답이든 내줄 수 있는 인공지능이 존재한다면, 내가 가장 먼저 물어보고 싶은 질문은 어떤 질문일까? 인공지능이 과연 인류의 미래에 위협적인 존재가 될지 아닐지는 어쩌면 우리가 어떤 질문에 대한 답을 찾기 위해 인공지능을 개발하는지에 달려 있는 게 아닐까 생각해 본다.



<그림 1. 아이디어회관의 로봇 머신 X 표지>



<그림 2. HAL 9000, 2001 년 스페이스 오디세이. By Tom Cowap - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=103068276>>