碩士學位論文

AI가 인간을 해치는 날

(『AIが人間を殺す日』 飜譯論文)

濟州大學校 通譯飜譯大學院

韓 日 科

文 辛 化

2018 年 7 月

AI가 인간을 해치는 날

(『AIが人間を殺す日』飜譯論文)

指導敎授 坂野愼治

文 辛 化

이 論文을 通譯飜譯學 碩士學位 論文으로 提出함

2018年 7月

文辛化의 通譯飜譯學 碩士學位 論文을 認准함

審査委員長 _____㊞

委 員 _____㊞

委 員 _____㊞

濟州大學校 通譯大學院

2018年 7月

역자 서문

2016년 알파고와 바둑기사 이세돌의 대결 이후 국내에서의 AI에 대한 관심도도 매우 높아졌다. 실제로 그 이후 AI 기술도 비약적인 발전을 이루었으며 다양한 산업 분야에서 도입을 서두르고 있다.

이에 따라 AI는 청소기, TV, 냉장고, 에어컨, 스피커 등 가정에서 일상적으로 사용하고 있는 제품부터 스포츠, 채용심사, 콘텐츠 제작, 대학 행정 등 매우 광범위한 분야로까지 도입이 확대되고 있다.

AI가 더 많은 분야에서 활용되게 된다면 우리의 생활은 지금보다 더 편리하고 윤택해질 것이다. SF 영화에서나 볼 수 있었던 날아다니는 자동차, 집안일을 도맡아 하는 휴머노이드 등이 머지않아 현실에 등장할지도 모른다. 어쩌면 전혀 상상도 하지 못했던 방향으로 '스마트폰'과 같은 새로운 혁명이 일어날 수도 있다.

그러나 AI의 도입으로 기대되는 긍정적인 효과와 더불어 우려의 목소리도 적지 않다. AI 스피커는 도청에 대한 위험성이 제기되었으며, 금융권의 AI 도입은 해킹으로 인한 개인정보 유출 문제가 가장 우려되고 있다.

이런 가운데 최근 테슬라의 자율주행차 '오토파일럿'으로 몇 차례의 교통사고가 나면서 사상자가 발생하였고, 이 사고가 과연 운전자의 과실인지 오토파일럿의 책임인지에 대한 공방이 끊임없이 이어지고 있다.

이렇게 사회 전반적으로 AI가 관심을 끌면서 관련 서적도 다양하게 출판되고 있다. 역자도 여러 서적을 찾아보았으나 국내에 발간된 서적 중 AI의 실태를 쉽게 풀어 설명하면서도 위험성에 대하여 제대로 경고하는 책은 생각보다 많지 않았다.

앞으로 AI의 확산은 불가피한 흐름이 되었으며, 도입으로 기대되는 효과도 적지 않다. 그러므로 우리가 더욱 현명하게 AI를 받아들이기 위해서는 AI 기술의 실태와 위험성에 대해 현실적으로 인지할 필요가 있다. 이 책을 통해 좀 더 많은

사람이 AI를 제대로 알고 경각심을 가질 수 있게 되었으면 한다.

마지막으로 도움을 주신 많은 분께 감사의 말씀을 전한다.

국문 초록

이번에 번역한 고바야시 마사카즈(小林雅一)의 『AI가 인간을 해치는 날(AIが人間を殺す日)』(集英社, 2017)은 현시점의 AI 기술이 어느 정도 위치에 와있으며, 1차·2차·3차 산업혁명으로 이루어낸 자동화와 AI의 도입으로 이루어질 4차 산업혁명의 자동화가 근본적으로 어떤 차이점이 있는지를 보여주는 책이다. 더불어 단순히 일상을 윤택하게 해주는 범위를 넘어 생사를 좌우하는 중대한 분야까지 확대되고 있는 AI 기술에 대해 우리가 어떤 것을 미리 알고 경계해야 할지 짚어보고 있다.

이 책은 '들어가는 말', '제1장 AI 위협론의 허와 실', '제2장 자율주행차의 사각지대', '제3장 로봇 닥터의 오진', '제4장 자율무기의 조준', '제5장 수퍼오토메이션의 함정', '나오는 말'로 이루어져 있다.

본 논문에서는 '들어가는 말', '제1장', '제2장'의 일부를 번역하였다.

'들어가는 말'에서는 전체적인 책의 줄거리와 이 책을 쓰게 된 배경을 소개한다. '제1장 AI 위협론의 허와 실'은 현재의 AI 기술 수준과 AI에 대한 잘못된 위협론, 자율주행차와 의료, 무기 분야에 도입된 AI에 대해 개략적으로 설명하고 있다. '제2장 자율주행차의 사각지대'는 자율주행차에 대한 본격적인 설명으로 테슬라에서 개발한 자율주행차(오토파일럿)가 일으킨 사망사고를 자세히 분석하고 있다.

목 차

들어가는 말

최근 몇 년 동안 세계적으로 가열되어 온 AI(Artificial Intelligence: 인공지능) 개발 경쟁이 이제는 새로운 국면에 접어들고 있다.

지금까지의 주된 AI 제품은 스마트폰의 음성 조작 기능, 청소 로봇, 대화형 인공지능 스피커 등 전적으로 IT · 가전제품에 탑재되어 편이성을 높여주는 '가벼운 용도의 인공지능'이었다.

그러나 앞으로는 인간과 사회에 더욱 크고 심각한 영향을 미치는 '중요한 용도의 인공지능'이 활발하게 개발될 전망이다.

이 책은 그중에서도 자동차와 의료, 그리고 무기에 탑재되는 인공지능을 다루고 있다. 이들 분야 전부 AI의 판단이 인간의 생사를 좌우하는 중대한 분야이다.

가장 선두에 있는 것이 바로 자동차에 AI를 탑재한 '자율주행 자동차'이다.

2009년쯤부터 미국의 구글이 가장 먼저 자율주행차 개발에 착수하기 시작하였고, 이후 미국과 유럽, 일본 등 각국의 주요업체에서도 이 경쟁에 뛰어들면서 지금은 차세대 자동차 비즈니스의 핵심 사업으로 주목받고 있다. 이미 미국의 테슬라, 일본의 닛산자동차 등 일부 업체에서 부분적 자율주행 기능을 제공하는 등 상용화도 시작되었다.

2020년 이후에는 운전자가 필요 없는 (완전) 자율주행차도 제품화될 전망이다. 이런 꿈만 같은 자동차는 지금까지 신체적 핸디캡과 고령 등 여러 이유로 운전을 포기할 수밖에 없었던 사람들에게까지 이용의 폭이 넓어질 것이다. 나아가 승차시간을 유익하게 활용하는 등 사회 생산성 향상 면에서도 큰 기대를 모으고 있다.

이에 필적하는 효과를 불러오는 것이 바로 의료 분야의 AI 도입이다.

이미 미국 IBM의 AI인 '왓슨(Watson)'이 대량의 의학 논문 검색을 통해 의사

가 알아내지 못한 병명을 제시하여 환자의 생명을 구한 사례가 보고되기도 하였다.

그리고 앞으로는 최첨단 AI인 딥러닝(Deep Learning) 기술로 MRI나 CT 등의 단층 화상을 해석함으로써 질병의 조기 발견 및 의료비 절감을 실현할 수 있게 된다. 나아가 여러가지 질병 예측 및 예방까지 가능할 것으로 보인다. 이 분야는 이미 임상연구 단계에 이르렀으며 향후 수년 이내에 전 세계에 상용화될 전망이다.

최종적으로는 무기 산업 분야에 탑재되는 AI가 있다.

스스로 표적을 지정해 날아가는 미사일, 상공에서 지상의 테러리스트를 감시하는 자율 드론(무인기) 등 인공지능을 탑재한 여러 무기가 세계 각국에서 개발되어 실증 실험 단계까지 이르렀다. 특히 미 국방부에서는 AI 무기가 단순한 임시방편이 아닌, 통칭 '제3의 상쇄(3rd Offset)'라는 근본적 군사 개혁의 핵심사항으로 자리 잡고 있다. 언젠가는 다른 국가들도 이러한 절차를 밟을 것이다.

위의 3가지 케이스로 알 수 있듯 AI는 앞으로 우리 생활 및 사회 나아가 국가 시스템의 중추까지 파고들어 (좋든 나쁘든) 극적인 변화를 불러일으킬 가능성이 크다.

그런 만큼 만약 AI가 오작동이나 폭주를 일으킨다면, 그 피해도 헤아릴 수 없을 정도로 심각하다. 최악의 경우 인간의 죽음으로 이어진다.

여기서 문제점은 우리가 과연 이 강력한 AI를 제어할 수 있냐는 것이다. 결론부터 말하자면 그것이 불가능할 수도 있다.

다음은 자율주행차의 사례이다. 이미 테슬라의 (부분적) 자율주행 시스템인 '오토파일럿(Autopilot)'으로 인해 미국에서 사망사고가 발생한 바 있다. 운전자가 자율주행이라는 일종의 AI에 대한 원리 및 구조를 전혀 이해하지 못하고, 그 성능을 과대평가한 채 운전 자체를 AI에게 맡겨버린 것이 이 사고의 주요 원인이었다.

그러나 잘 생각해보면 내부의 구조 및 기술을 이해하지 못하면서 우리의 목숨

을 기계에 맡기는 일은 부지기수이다. 일반 자동차, 고속철도, 제트 여객기만 하더라도 (일부의 기계 마니아를 제외한) 절대다수의 사람들은 내부 구조를 거의 이해하지 못하면서 이용하고 있다.

그런데도 기본적으로 우리 사회가 기계를 받아들이고, 자리 잡을 수 있었던 이유는 일반 사용자가 이해할 수 없더라도 이를 개발한 과학자나 기술자, 즉 전문가들이 그 원리와 시스템을 정확하게 파악하고 있었기 때문이다. 사고나 고장 등 트러블이 생기면 전문가들이 대처하여 결과적으로는 기계가 인간의 통제하에 있었다.

그러나 앞으로 사회 곳곳에 도입되어 갈 AI 기술은 이런 부분이 불투명해지고 있다. 우리 같은 일반 사용자를 넘어 AI 연구개발에 종사하는 과학자와 기술자들조차 내부의 메커니즘 및 사고회로의 파악이 어렵게 될지도 모른다.

한 가지 예를 살펴보자.

고도의 의료기술로 세계적으로 알려진 미국의 마운트 사이나이 병원에서는 2015년 (임상 연구의 일환으로) 질병 예측 시스템을 개발하였다. 이 AI 시스템은 (앞서 거론된 딥러닝이 기초가 되어) '딥페이션트(Deep Patient)'라 불리게 되었는데, 개발에 참여한 의료학자들조차 경악할 정도로 예측이 잘 맞았다.

딥페이션트는 이 병원에서 관리하는 7만 6000여 명의 전자 진료기록 카드를 읽고, 신장·체중·혈액 및 소변검사 결과 등 헬스케어 데이터를 분석하여 환자들이 각각 언제 어떤 병세가 나타나는지 보란 듯 맞췄다.

질병의 종류는 각종 암, 당뇨병뿐만 아니라 조현병 같은 정신 질환까지 포함하여 78종에 이른다. 딥페이션트는 다른 모든 기술 방식을 압도적으로 뛰어넘는 수준으로 정확하게 발병 확률을 산출해냈다. 그러나 과학자들은 이런 뛰어난 예측 능력에 감탄하면서도 한편으로 불안감을 떨쳐낼 수 없었다.

이 AI 시스템은 어떤 사람이 어떤 병에 걸리는지는 예측할 수 있지만, 결코 그 근거가 무엇인지 알려주지 않았기 때문이다. 내부의 사고회로가 (전문가를 포함한) 우리 인간에게는 전혀 보이지 않는 미지의 블랙박스 상태였다.

현대의 AI는 놀라울 만큼 정확하게 정답을 산출할 수 있다. 그러나 그렇다고 블랙박스화된 AI를 무조건 수용하고 생사가 걸린 중대한 판단을 맡기는 것이 과연 현명한 선택일까? 인간이 어떠한 형태로든 제어에 개입해야 하지 않을까? 만약 그렇다면 AI와 인간은 앞으로 어떤 식으로 관계를 만들어 가야 할까?

이 점을 살펴보는 것이 이 책의 역할이다. 이러한 이유로 인공지능의 기본 원리도 가볍게 다루고는 있으나 독자의 이해를 최우선으로 집필하였으므로 전문 예비지식이 없어도 이해하는데 어려움은 없을 것이다.

이 책이 위험수위에 가까워진 AI의 실태를 적절히 보여주고, AI와 어떻게 마주해야 할지에 대한 힌트가 되었으면 한다.

제1장 AI 위협론의 허와 실

1950년대 미국에서 탄생한 AI는 지금에 이르러 역사상 세 번째 전성기를 맞이하고 있다. 항간에 넘쳐나는 많은 자료에는 AI가 불러올 편이성과 생산성 향상 등 장밋빛 미래를 꿈꾸는 것들이 많다.

그러나 이런 밝은 전망과는 정반대로 AI가 파멸적인 미래를 불러오리라 예측하는 자료도 적지 않다. 가히 'AI 위협론'이라 할 수 있는 내용이다.

예를 들어 인간이 진화한 AI와 로봇에게 일자리를 빼앗긴다는 견해가 있다.

2013년 영국의 옥스퍼드대학 연구자들이 발표한 「고용의 미래」라는 논문과 함께 일본 노무라 종합연구소에서도 유사한 2015년도 조사결과를 발표하였고, 이를 바탕으로 예전부터 가까운 미래에 AI와 로봇이 대규모 고용파괴를 불러올 것이라고 했다. 이들 모두 국내외 600~700종에 달하는 직종을 구체적으로 열거하며, 앞으로 10~20년 이내에 이 직종의 약 50%가 컴퓨터와 로봇·자율주행차 등에 빼앗길 것으로 예측했다.

우리에게는 매우 절실한 '고용'이라는 주제가 이런 충격적인 전망과 맞물리면서 세간의 주목도는 한층 더 높아졌다. 고명한 연구자들에게 트집을 잡고 싶지는 않지만, 이런 조사결과 내용에는 고개를 갸우뚱하게 만드는 부분도 있다.

한 예로 '이발사', '목수', '패션모델' 등은 가까운 미래에는 사라진다고 한다. 대체 무슨 근거로 이런 결과가 나온 걸까? 실제로 앞의 두 직종은 (손재주와 같이) 섬세한 운동 능력과 고도의 소통 능력이 필요하므로 로봇으로 대체하기는 매우 어려울 것으로 보이는 분야이다.

패션모델의 경우는 아예 논할 가치도 없다. 대체 누가 로봇이 입은 옷을 입고 싶어 할까?

실제로 2015년 미국에서 개최된 'DARPA 로보틱스 챌린지(DARPA Robotics

Challenge)'에서는 건물의 건설현장과 배송센터에서 인간 노동자를 대신하는 작업용 휴머노이드의 실현이 매우 어렵다는 사실을 재확인하기도 했다.

미래에 이런 로봇이 상용화되더라도 그 과정은 절대로 순탄치 않을 것이다. 그리고 실현되었을 때 인간의 일자리를 빼앗기보단 오히려 각종 서비스, 건축 및 유통업계의 인력 부족을 보완해주는 긍정적 효과가 더 크지 않을까?

패턴인식 직종이 위험하다

한편, 컴퓨터와 AI의 발달로 사라지는 직종도 존재한다. 장기적으로 상승할 주식을 골라서 투자하는 펀드매니저가 이에 속한다.

지난 2017년 3월, 총액 5조 달러(한화 5000~6000조 원)를 운용하는 세계 최대 자산 운용회사인 미국의 블랙록(BlackRock)은 53명이던 펀드매니저를 17명 이하로 줄인다는 방침을 발표했다. 비용도 많이 들지 않고 운용 실적도 양호한 알고리즘 거래, 나아가서는 AI를 활용한 보다 효율적 시스템으로 전환하겠다는 것이다("At BlackRock, Machines Are Rising Over Managers to Pick Stocks" Landon Thomas Jr. (2017), *The New York Times*, March 28, 2017).

그 외에도 (택시나 트럭) 운전사, (금융기관의) 대출 심사원, 방사선과 전문의 등 가까운 미래에 AI와 자율주행이 보급됨에 따라 사라질 위기에 처한 직종은 적지 않다.

언뜻 위 직업들은 분야 및 수입이 전부 제각각으로 보이지만, 하나의 공통점을 가지고 있다. 업무에서 차지하는 '패턴인식'의 비율이 크다는 점이다.

대부분의 미디어에서 과장 보도되고 있는 정보와는 다르게, 현재의 인공지능이 활약하고 있는 분야는 '패턴인식'이라는 극히 한정된 분야뿐이다. 이는 컴퓨터와 로봇이 화상이나 음성을 인식하거나 빅데이터에서 하나의 규칙성(패턴)을 찾아내는 기술이다.

패턴인식 분야에서만큼은 이미 인공지능이 인간을 앞지른 것으로 보인다. 그 말은 결국 패턴인식에 대한 의존 비율이 높은 직종일수록 컴퓨터와 AI에게 빼앗기기 쉽다는 것을 의미한다. 위와 같은 직업이 바로 그에 해당한다.

펀드매니저는 거시적, 미시적 경제 지표와 중앙은행의 동향, 시장 심리 등을 통해 상향 신호가 있는 투자종목을 선별한다.

대출심사 담당자는 신용 카드 결제 이력과 주택·자동차 담보 대출 여부, 일상적인 소비 형태 등을 보고 상환이 불가능해질 리스크를 예측한다.

또, 방사선과 전문의는 MRI나 CT 촬영 단층 화상을 보고 악성종양 등으로 의심되는 부분은 없나 찾아낸다.

이들은 전부 일종의 패턴인식이다. 하지만 인간이 가지고 있는 능력은 이것이 전부가 아니다. 사람들은 평소 자신의 업무가 '단조롭고 따분하다'라고 비하하지만, 무의식중에 통찰력, 관찰력, 커뮤니케이션 능력, 타인에 대한 공감·배려, 감수성 등을 활용하고 있다.

앞으로 AI가 인간 고유의 능력을 갖출지도 모른다. 그러나 만약 그렇다 하더라도 그때까지는 앞으로 몇십 년, 혹은 그 이상의 오랜 세월이 필요할 것이다.

그러므로 현시점에서 AI로 인한 고용 상실은 패턴인식이 중심인 특정 분야로 한정된다. 그 외의 직종은 AI에게 빼앗기기보다 오히려 상호 부족한 능력을 보충하는 형태로 서서히 인간과 기계가 역할을 분담하여 재구성될 확률이 높다.

2045년 문제와 화성의 인구 폭발

한편 이와 별개로 한층 심각한 AI 위협론도 들려온다. 이른바 '2045년 문제'라는 '싱귤래리티(Singularity: 기술적 특이점)'의 도래이다. 이것은 미국의 저명한 발명가 레이 커즈와일 박사가 꽤 오래전부터 주장해온 미래에 대한 예측이다.

이에 따르면 2045년쯤에는 AI의 바탕이 되는 컴퓨터 프로세서 처리능력(AI의

기본이 되는 기술)이 인간의 지력을 뛰어넘고, 언젠가는 AI가 의식과 감정까지 갖추게 된다고 한다. 머나먼 미래에는 AI와 로봇이 인류를 지배하게 되어 우리의 생존 자체를 위협받을 수도 있다고 한다(커즈와일 박사가 처음부터 이렇게 말한 것은 아니지만 점점 이야기에 살이 붙으면서 과장되었다).

커즈와일 박사는 약간 괴짜 같은 인물로 통하기 때문에, 만약 그가 혼자 이런 주장을 했다면 이 정도로 진지하게 받아들여 지지 않았을지 모른다. 그러나 실제로는 그 외에도 세계적인 물리학자 스티븐 호킹 박사와 우주여행 사업 등 큰 스케일을 자랑하는 기업가 일론 머스크 등 각계의 저명인사들 또한 같은 경고를 했다. 이로 인해 싱귤래리티처럼 놀랍도록 진화한 AI가 인류의 생존을 위협할 수 있다는 AI 위협론에 대한 관심이 높아지고 있다.

세계적 유명인사에게 이의를 제기하는 것이 주제넘을지는 모르지만, 호킹 박사와 머스크는 AI 전문가가 아니다. 이들이 인공지능 구현을 위한 구체적 기술이나 그 내부 메커니즘에 대한 일정 수준 이상의 전문 지식을 갖고 있다고 보기는 어렵다. 그런데도 어떻게 AI의 발전 방향성과 잠재적 위험성을 점칠 수 있었을까? 오히려 일종의 흥미와 충격적인 예측으로 세간의 관심을 끌고자 했던 것이 아닐까?

물론 '싱귤래리티와 같은 사태가 미래에 절대 일어나지 않는다'고는 단언할 수 없다. 먼 미래에는 실제로 이런 시대가 올지도 모르지만, 그것은 'AI로 인한 고용 상실'보다 더 많은 시간이 필요할 것이다. 이에 대해서는 AI 전문 연구자이자 딥러닝의 일인자로 알려진 미국의 스탠퍼드 대학 앤드루 응(Andrew Ng) 준교수가 다음과 같은 비유로 풍자하고 있다.

"(현대 사회의) 우리가 AI로 인류가 파멸될지도 모른다고 우려하는 것은 화성의 인구 폭발을 지금부터 걱정하는 것과 마찬가지이다."

먼 미래에는 우주 개발이 비약적으로 발전하여 인류가 화성으로 이주하게 될지도 모른다. 그렇다 한들 화성의 인구가 폭발하려면 앞으로 몇백 년이 더 걸릴 것이다. 마찬가지로 AI와 로봇이 초월적 진화에 성공하여 인간에게 해를 끼치거나 인류를 지배하게 되더라도 아주 먼 미래의 일이라는 것이다. 싱귤래리티와 같

은 AI 위협론도 현재를 살아가는 우리에게는 비현실적인 문제라고 할 수 있다.

Human out of the Loop - 제어권을 상실한 인간

그러면 AI의 진정한 위협은 아직 발견되지 않은 것일까? 아쉽게도 그렇지는 않다. 실제로 앞에서 언급한 고용 파괴, 싱귤래리티와는 별도의 중대한 위험성이 존재한다. 그것이야말로 더 현실적이고 눈앞에 닥쳐오고 있는 AI의 '진정한 위협'이다.

기계 및 AI와 인간의 관계성을 규정하는 이 문제를 전문가들은 'Human out of the Loop(인간이 제어권을 상실했다)'고 하고 있다. 더 간단하게는 '수퍼오토메이션(Super Automation)'이라고 한다.

역사적인 관점에서 이 문제를 살펴보자.

18세기 영국의 '증기기관 발명'으로 시작된 근대 과학 문명의 발달사는 한 마디로 '자동화의 역사'라고 해도 과언이 아니다. 1차 산업혁명 시대에는 '증기기관차' 등의 교통 및 수송수단 자동화와 함께 '방직기' 등의 제품 제조 공정까지 어느 정도 자동화를 이루었다.

이후 19세기의 전자기학(電磁氣學) 발달을 거쳐 20세기 전반에는 전기 모터를 사용한 '컨베이어 벨트의 제품 조립라인'이 등장하는 등 자동화도 다음 단계에 이르렀다. 이것이 2차 산업혁명이다.

그리고 20세기 초반부터 '양자역학(量子力學)'을 공학에 응용하기 시작하며, 20세기 후반에는 '집적 회로(반도체 제품)'가 발명되었고 이를 바탕으로 한 전자 산업이 꽃 피우게 되었다. 이 연장 선상에서 컴퓨터와 IT, 나아가 산업용 로봇으로 '두뇌 노동', '공장의 생산공정' 자동화가 가속화되었다. 이것이 3차 산업혁명이다.

지금은 21세기 초반 하나의 커다란 붐을 일으킨 'AI'와 'IoT(Internet of Things: 사물인터넷)'라는 기폭제로 인해 4차 산업혁명이 시작되고 있다. 하지만

그로 인한 오토메이션은 지금까지와는 본질적으로 다르다.

1차부터 3차 산업혁명까지의 오토메이션에서 교통과 수송수단 및 공장은 일정 수준의 자동화를 이루었지만, 최종적으로 그것을 제어하는 것은 인간이었다.

1차 산업혁명에서 탄생한 증기기관차, 2차 산업혁명의 상징인 자동화는 각각 '증기기관', '가솔린엔진' 등 동력(구동 시스템)을 자동화한 것이었다. 3차 산업혁명의 주역인 산업용 로봇도 특정 부분의 절단, 용접 및 복수 부품 조립 등 공장 노동자와 기술자들이 사전에 프로그램한 명령에 따라 정형화된 작업을 수행하는 정도였다.

즉, 동력 등의 '구동 시스템'이 자동화되더라도 기계를 자유자재로 조종하기 위한 핸들 등의 '제어 시스템'은 인간을 위한 영역으로 남아 있었다.

그러나 현재 진행형인 4차 산업혁명에서는 인간의 마지막 보루인 '제어 시스템' 즉, 기계를 컨트롤하는 권리마저 기계에 넘어가려 하고 있다.

이것이 앞에서 언급된 '수퍼오토메이션' 혹은 'Human out of the Loop'이며, 근대 과학 문명의 발달사 사상 최후의 자동화 프로세스이다. 이 점이 기존 현상과 가장 결정적인 차이임을 유념해야 한다.

수퍼오토메이션은 과거와는 다른 쾌적함, 편리함과 더불어 SF같은 미래사회를 우리에게 약속한다. 그러나 만약 수퍼오토메이션이 폭주하거나 오작동 및 제어 불능 상태에 빠진다면 그 공포와 피해도 차원이 다르고 가히 파멸적이다.

핸들도 브레이크도 없는 자동차의 폭주

그 선봉은 아마도 상용화를 앞둔 자율주행일 것이다. 자세한 내용은 제2장에서 다룰 예정이지만, 이미 테슬라의 '오토파일럿(Autopilot)' 등 부분적 자율주행(반자율주행)이 상용화 중이다. 이를 넘어선 완전 자율주행 혹은 이와 근접한 기능의 제품화는 애초 목표가 2020년이었으나 실제로는 이보다 더 걸릴 전망이다.

그렇다 하더라도 그렇게까지 먼 미래의 이야기는 아니다. 실제 상용화가 되면 얼마나 편리하고 쾌적할지 쉽게 상상할 수 있다. 먼저, 신체적인 핸디캡을 안고 있는 사람들이나 질병 때문에 시력, 체력 등이 쇠약해진 사람들처럼 지금까지 자동차 보급(모터리제이션)의 혜택을 누리지 못한 많은 사람이 자율주행차로 원할 때 원하는 곳으로 이동할 수 있게 된다.

일반 회사원들도 이동 중 서류를 읽거나 거래처에 메일을 보낼 수 있어 자율주행차가 움직이는 사무실이 된다. 또, 목적지에 도착하면 스스로 주차장으로 이동하고, 회식 후에도 마음 편히 귀가할 수 있다.

자동차에서 다 함께 파티를 하거나 게임을 하고 DVD 영화를 볼 수 있다. 그리고 운전은 AI에 맡겨둔 채, 풍경을 감상하면서 애정행각에 몰두하는 커플도 분명히 나올 것이라고 업계 관계자들은 입을 모은다. 한마디로 즐겁고 쾌적한 일이라면 '뭐든지 할 수 있다'는 것이다.

그러나 자율주행차가 폭주하거나 제어 불능에 빠진다면 그 공포는 상상을 초월한다. 이를 짐작하게 하는 것은 구글 자회사인 알파벳이 2015년 캘리포니아주에서 공개한 무당벌레 모양의 자율주행차 시제품이다.

이 자동차에는 핸들과 브레이크가 없었다. 이것을 본 DMV(Department of Motor Vehicles: 캘리포니아주 차량국)에서 '이대로는 너무 위험하니 만약을 위해 핸들과 브레이크, 엑셀을 만들라'고 행정명령을 내렸을 정도였다.

2017년 6월 구글 산하의 웨이모(Waymo)는 자사에서 설계한 자율주행차의 공공도로 시범운행을 중단하고 앞으로는 협약을 맺은 FCA(Fiat Chrysler Automobiles)에서 생산한 차량을 사용한다고 발표했으나, 인간이 운전에 관여하지 않는 완전 자율주행차를 목표로 하는 것에는 변함이 없었다.

만약 이런 자동차가 정말 제품화되었을 때, 운전자를 태운 채 제어불능에 빠졌다고 가정해보자. 고속도로를 달리던 자율주행차가 갑자기 폭주한다면, 그 안에 타고 있던 운전자는 아무 대책 없이 어딘가 부딪힐 때까지 기다릴 수밖에 없을지 모른다. 그 상황의 공포는 말로 다 헤아릴 수 없다.

자율주행의 산업적 임팩트

이러한 잠재적 위협을 인지하고도 왜 구글과 세계의 자동차업체들은 굳이 자율주행 시스템을 개발에 착수하였을까?

그것은 앞으로 파생될 매우 커다란 산업적 임팩트 때문이다. 많은 주요 선진국의 기간산업이기도 한 '자동차' 분야가 자율주행의 실현으로 새롭게 탈바꿈된다면 막대한 신규수요 및 고용 창출의 경제적 효과를 기대할 수 있다.

특히나 이 부분에 주력하고 있는 것이 미국이다. 2016년 1월 오바마 정권은 자동차업체의 자율주행 연구개발에 향후 10년간 40억 달러(한화 약 4조 3천억 원)의 개발 지원금을 출연할 계획을 발표하였다(물론, 2017년 1월에 출범한 트럼프 정권이 오바마 정권의 정책을 뒤집을 가능성도 있지만, '오바마 케어' 같은 사회보장 정책과 다르게 자율주행과 같은 산업 진흥 정책은 트럼프 정권에서도 지지를 이어갈 가능성이 크다).

이런 지원책을 발표한 미국 운수부 장관 곁에는 미국의 제너럴 모터스(GM), 포드 모터와 같은 주요 자동차업체 및 구글의 간부 등의 조력자들이 포진하여 정부와 민간이 함께 자율주행 시스템을 육성하려 하고 있다.

또, 주 정부와 각 주에서 선출된 연방 의원들도 이런 자율주행 기류에 쌍수를 들고 환영하는 분위기이다. 버지니아주에서는 실제 약 112km에 달하는 산과 계곡 등의 변화무쌍한 공공도로를 테스트 코스로 지정하고, 각 자동차업체가 자율주행차 시험주행을 하기를 바라고 있다.

자동차 산업의 중심지인 디트로이트가 있는 미시간주에서는 약 4만 평의 거대 부지에 유사 시가지를 조성하여 자율주행 운행 테스트를 할 수 있는 환경을 조성했다. 이들 모두 GM 등의 주요업체가 자신들의 지역에 자율주행 개발의 거점 및 공장을 건설하고, 이것이 새로운 고용 및 세수로 이어지기를 고대하고 있다.

미국이 이렇게까지 자율주행에 돈을 쏟아붓는 이유는 무엇일까? 그것은 현재 자동차 산업이 사상 초유의 전환기를 맞이하고 있고, 이를 잘만 살린다면 미국이

다시금 이 분야에서 세계를 선도할 수 있을지도 모른다고 보고 있기 때문이다.

세계의 자동차 시장에서 존재감을 드러내고 있는 업체는 유럽의 폴크스바겐, 다임러, BMW, 르노 등과 함께 동아시아의 도요타, 르노 산하의 닛산, 현대 등이 있다. 이들의 효율적인 생산체제와 우수한 품질, 높은 브랜드 파워와 판매망 앞에 미국 업체들은 세계 시장에서 고전을 면치 못하고 있었다.

그러나 이제는 자동차가 근본적인 변화를 맞이하고 있다. 구동 시스템에서는 기존의 가솔린 엔진이 전기 모터로, 제어 시스템에서는 수동 운전이 자동 운전으로 전환될 것이다. 그렇게 되면 독일, 일본 등의 업체가 지금까지 축적해온 고도의 기술력 대부분은 무효가 되고 자동차 기술 개발도 거의 처음부터 다시 시작해야만 한다.

미국의 업체가 거대한 트렌드를 선점할 수 있다면, 유럽과 동아시아를 누르고 세계 자동차 산업의 맹주로 복귀할 수 있게 된다. 이것은 미국 내 제조업 전체에 활기를 다시 불어넣고, 엄청난 고용을 창출하는 트리거가 될지 모른다. 자율주행은 미국에 있어 제조업의 권위회복을 위한 천재일우의 기회인 것이다.

그러나 만약 이런 미국의 노림수가 들어맞는다면 반대로 유럽과 동아시아의 자동차 산업은 커다란 타격을 입게 된다. 특히 일본이 받는 영향은 심각하다. 일본의 주력산업이던 전자 산업은 왕년의 기세를 잃었고 현재는 자동차 산업에 의지하고 있다고 해도 과언이 아니기 때문이다.

이 분야마저 새로운 기술 개발의 흐름에서 도태된다면, 일본의 경제와 고용에는 괴멸적인 피해를 불러올 것이다. 그렇기에 도요타와 닛산 등 일본의 주요업체도 구글과 같은 미국이나 유럽 제조업체에 뒤지지 않도록 필사적으로 자율주행 연구개발에 힘을 쏟고 있다.

기계는 사람보다 믿을 수 있나?

이런 산업적 임팩트와 더불어 세계의 자동차업체가 자율주행에 주력하는 또 한 가지 커다란 이유는 '안전성 향상'이다. 즉, 인간보다는 (자동차와 같은) 기계 자체에 제어를 맡기는 것이 사고를 줄일 수 있다고 생각하는 것이다.

이것은 앞서 언급한 '기계가 제어 불능에 빠졌을 때의 공포'와는 정반대의 관점이지만 일리 있는 말이기는 하다. 우리 인간은 자동차를 운전하면서 항상 운전에만 집중하고 있다고 할 수 없기 때문이다. 운전자가 한눈을 팔거나 운전 외의 다른 생각에 빠지거나, 졸음운전을 하는 등의 일은 심심치 않게 일어난다.

특히 최근에는 운전 중 통화나 앱을 사용하는 운전자도 있다. (법적인 처벌이 아무리 강화되어도) 음주운전은 여전히 일어나고 있고, 조수석 베이비시트에 앉힌 아이를 수시로 달래가며 운전하는 사람도 있다. 고령 운전자나 병을 앓고 있는 사람이 운전 중 갑자기 발작을 일으키거나 의식불명에 빠지는 일도 있다.

이런 휴먼 에러(인위적 실수)가 자동차로 보행자를 치거나 다른 자동차와 충돌하는 각종 사고의 주요 원인이 되고 있다. 미국 내의 조사에 따르면 자동차 사고 전체의 약 94%가 이 휴먼 에러로 인해 발생한다고 한다.

그렇다면 이렇게 '사고를 일으키기 쉬운 인간'이 아닌 자동차라는 기계에 운전 자체를 맡기면 어떻게 될까? 기계는 운전하면서 졸거나 한눈을 팔지 않고, 음주운전도 하지 않는다. 스마트폰을 만지거나 의식을 잃지도 않는다. 이런 기계에 운전을 맡기는 것이 더 믿을 수 있고 안전하지 않을까 하는 생각에서 자율주행 연구개발이 시작되었다.

그러나 여기에서 우려되는 점은 '기계는 정말로 인간보다 믿을 수 있는가'하는 문제이다. 우리가 안심하고 제어권을 넘겨줄 정도로 자율주행차라는 기계가 높은 신뢰성을 가지고 있을까? 이 점이 가장 중요한 문제이다.

이를 확실히 알기 위해서는 자율주행차 시스템, 즉 실현을 위한 기술까지 파고

들어 검증해 볼 필요가 있다. 여기서는 자동차라는 기계를 제어하는 AI가 중요한 역할을 맡는다. 다음은 요점만 추려서 살펴보고자 한다.

3종류의 AI

자동차 선진국인 미국과 독일, 일본, 이탈리아 등에서는 대학 및 주요업체를 중심으로 자율주행 연구개발이 1960~70년대에 시작되었다. 그러나 자율주행용 시제품 차량에 탑재된 당시 하드웨어 처리능력에는 한계가 있어 초기 연구의 성과가 상용화까지 이르지 못했다.

이후 1980~90년대에 걸쳐 미국의 카네기멜론 대학의 가나데 다케오 교수 등을 중심으로 본격적인 자율주행 연구개발이 진행되었다. 그리고 이것이 2004~6년에 걸쳐 미 국방부 산하의 DARPA(국방고등연구계획국)에서 개최한 그랜드 챌린지(Grand Challenge: 자율주행차 경주) 등을 거치면서 결실을 보았고, 자율주행 시스템이 상용화에 가까운 단계까지 성장했다.

이러한 학술적 연구성과를 구글이 헤드헌팅 하였고 시판 차량을 자율주행용으로 개조하였다. 그리고 2009년쯤부터 공공도로에서 테스트 주행을 시작하여 동영상 사이트 '유튜브' 등으로 대대적으로 홍보하였다.

현재는 이에 자극을 받은 미국과 유럽, 동아시아의 자동차업체 및 IT 기업들이 다시금 자율주행 개발에 본격적으로 나서고 있다.

지금까지 자율주행의 연구개발을 선도해 온 구글은 2016년 12월 새롭게 설립한 자회사 '웨이모'에 자율주행사업을 이관하고 드디어 상용화 단계에 접어들었다. 이것이 자율주행의 현주소이다.

지금의 자율주행차에는 '60년 이상 쌓아온 세계적 AI 연구'의 집대성이라 할 수 있는 기술이 탑재되어 있다. 이들은 AI 개발사 초기에 번성했던 '규칙 기반 AI', 1990년대부터 성행한 '통계·확률형 AI', 최신예의 '뉴럴 네트워크' 3종류의 기술로 크게 구분된다.

먼저, '규칙 기반 AI'란 인간이 '만약 ~라면, ~를 해라'라는 규칙을 여러 개 설정하고 이를 자율주행차와 같은 기계가 이해할 수 있는 프로그래밍언어로 기술하여 자동차에 이식하는 방식이다.

예를 들어 '빨간 신호에는 정지해라', '(일본의 경우) 자동차는 좌측으로 통행해라', '교차로에서 우회전할 때는 맞은편 차선에서 직진하는 차량을 우선으로 해라' 등의 규칙을 다수 정해놓는다. 이러한 규칙을 자동차에 이식해서 이를 따르도록 하면 자율주행을 어느 정도까지 실현할 수 있다.

한편, 통계·확률형 AI에서는 탑재된 각종 센서로 취득한 외부데이터를 기계가 스스로 확률적으로 처리하여 자율주행을 할 수 있게 한다. 이 기술은 '은닉 마르코프 모델'이라고 불리는 수학적 모델을 기초로 한다(물론 최종적으로는 '파이썬(Python)', 'C' 등 각종 프로그래밍언어가 사용된 소프트웨어로 변환된다).

은닉 마르코프 모델에서는 '자동차'와 '보행자' 등 여러 이동체의 위치를 직전의 상태로 추정한다. 이를 각종 센서로 측정하는 작업과 '베이즈의 정리'라는 확률 논리로 보강하여, 이동체의 현재 위치를 최대한 정확하게 파악한다.

이 정도의 설명으로는 조금 이해가 어려울 수도 있으나, 은닉 마르코프 모델을 바탕으로 한 자율주행 시스템은 제2장에서 다시 설명할 예정이므로 그 내용을 읽으면 충분히 이해할 수 있을 것이다.

마지막으로 뉴럴 네트워크란 인간(혹은 동물)의 뇌를 (극히 한정된 범위이지만) 참고하여 개발한 인공지능이다. 뉴럴 네트워크 연구 자체는 1950년대에 시작된 전통적인 기술이었으나 상용화할 만큼의 수준이 된 것은 21세기 들어선 후이다.

이것은 '딥 뉴럴 네트워크(DNN)' 혹은 '딥러닝' 등으로 불린다. 특히 화상 및 음성 인식 등의 '패턴인식' 작업에 적합하다. 마찬가지로 최신 자율주행차에 활용되는 기술이다.

이 뉴럴 네트워크와 통계·확률형 인공지능은 기술자가 자동차에게 일일이 규칙을 가르치지 않아도 기계 자체가 각종 센서로 측정된 외부데이터를 학습하며 발전하기 때문에 일반적으로는 AI 안에서도 '머신러닝'이라고 불리는 분야에 속한

다.

이러한 AI 기술에는 저마다의 장단점이 존재하므로 자율주행차에서는 적절히 혼용되고 있다.

자율주행차가 주위의 이동체를 파악하려 하는 경우에는 센서로 취득한 외부데이터를 '통계·확률형 AI(은닉 마르코프 모델)'로 처리하여 이동체의 현재 위치를 산출해내고 있다. 센서로 측정한 외부데이터에는 반드시 오차가 나오므로 이를 처리하기 위해서는 통계·확률적 방식이 가장 적절하다.

한편, 이 이동체가 보행자인지 강아지나 고양이 같은 동물인지 혹은 바람에 굴러다니는 도로 위 비닐봉지인지 등을 구체적으로 정확하게 파악하기 위해서는 패턴인식에 특화된 '뉴럴 네트워크(딥러닝)'가 사용된다.

하지만, '빨간 불에는 정지'한다는 약속은 자동차 스스로 각종 센서로 머신러닝하기 보다는 기술자가 규칙을 입력하는 편이 손쉽고 간단하다(물론 빨간 불을 인식하기 위해서는 센서가 사용된다). 이처럼 '정확히 규정할 수 있는 행동원칙'을 자율주행차에 입력할 때는 규칙 기반 AI가 지금도 사용되고 있다.

상용화를 저해하는 문제

위와 같은 요소기술은 DARPA 그랜드 챌린지가 개최된 2005년쯤에는 거의 기본적으로 등장했다. 그리고 이 대회에서 우승한 스탠포드 대학의 서배스천 스런 교수의 개발팀을 구글이 그대로 영입하여 자율주행 개발을 맡겼다.

이 구글 팀은 2008년 2월 자율주행용으로 개조한 (도요타의 하이브리드차) '프리우스'를 사용하여 샌프란시스코에서 베이 브릿지를 건너 근처 섬까지 피자를 배달하는 주행실험을 했다. 이 모습이 미국의 케이블 TV '디스커버리 채널'을 통해 방영되어 주목을 모으기도 하였다(다만 이 실험은 대규모의 경찰이 동원되어 주위에 바리게이트를 치고 다른 차량 및 보행자가 통행할 수 없게 엄중히 제한된 환경에서 실시되었다).

이때부터 구글을 필두로 미국과 유럽, 일본 등 각국의 업체들이 본격적인 자율주행 개발 경쟁에 뛰어들었다.

이후 약 10년의 세월이 흐른 지금, 자율주행 시스템은 어느 수준까지 이르렀을까?

가끔 신문 및 TV 방송 보도에서 기자들이 자율주행차에 시승하는 장면이 나오기도 한다. 대부분의 체험담은 교통량이 비교적 적은 시간대에 고속도로를 주행하는 것이다. 이때, 자율주행차는 앞차와 적절한 간격을 두고 핸들과 브레이크, 엑셀 등의 조작도 큰 무리 없이 해낸다.

또한, 옆에서 달리던 자동차가 차선을 변경해 들어오려고 하면 매너 있게 앞 공간을 양보하고, 필요에 따라서는 옆 차선으로 부드럽게 차선을 변경한다.

이것만 보면 당장이라도 자율주행차를 상용화할 수 있는 완벽한 기술 수준에 이른 것처럼 생각할 수 있지만 실제로는 그렇지 않다. 그 사실은 각 자동차업체가 최근 미국 규제 당국에 제출한 '공공도로 테스트 주행' 데이터를 보면 더욱 명확하다.

자세한 내용은 제2장에서 소개하겠지만, 이 데이터에 따르면 '다임러(메르세데스-벤츠)' 같은 세계적인 업체가 개발 중인 자율주행차조차 몇km를 달리는 동안 적어도 한 번 이상의 문제로 차에 탑승한 오퍼레이터가 직접 운전을 했다. 이것이 자율주행 시스템의 현주소이다.

그렇다면 왜 TV에서는 문제없이 자율주행을 하다가도 공공도로 주행 테스트에 나서면 문제가 드러나는 것일까? 그것은 미리 철저한 준비로 이루어지는 TV 방송용 데모 주행과 달리 현실의 공공도로 환경에서는 무슨 일이 일어날지 알 수 없기 때문이다.

도로 공사를 예로 들 수 있다. 도로 공사 현장에서는 몇 개의 라바콘을 설치해서 비상 라인을 만든다. 자동차는 정식 차선을 무시하고 이 차선을 이용해서 공사 현장을 우회한다. 일반적으로는 자동차를 유도하는 작업자가 깃발을 흔들며 '이쪽으로 가주세요'라고 비상 라인 쪽으로 유도한다. 이때, 자동차는 앞의 신호가 빨간불이어도 무시한 채 전진할 수 있다.

인간이 운전하는 평범한 자동차라면, 이런 공사 현장은 어려움 없이 지나갈 수 있다. 그러나 적어도 지금의 자율주행 시스템으로는 그것이 불가능하다. 자율주행차에 탑재된 인공지능이 내부에서 분열을 일으키기 때문이다.

먼저, 각종 센서로 측정된 외부데이터를 은닉 마르코프 모델로 처리하는 '통계·확률형 AI'는 라바콘이 늘어서 있는 비상 라인을 인식하고 '도로에 구멍이 파인 위험한 공사 현장을 우회하는 방법은 이 차선을 지나가는 수밖에 없다'고 판단한다.

이 경우 중앙선을 넘어 맞은편 차선을 침범하게 된다. 게다가 신호는 빨간불이다. 둘 다 '규칙 기반 AI'로 엔지니어가 자율주행차에 입력한 명령에 위반된다.

즉, 자동차가 센서와 머신러닝으로 스스로 도출해낸 결론과 미리 인간이 입력해놓은 명령이 상반됨에 따라 자율주행차가 무엇을 따라야 할지 혼란스러워하는 것이다. 그리고 기계는 전방에서 작업자가 흔드는 깃발의 의미를 이해할 수 없다. 이것들이 상충하며 자율주행차가 운전을 포기할 수밖에 없게 된다.

이런 사례들은 그 밖에도 더 존재한다. 예를 들어, 자율주행차 앞에서 달리던 자동차가 갑자기 사고를 일으키면, 연쇄 사고를 피하기 위해서라도 급하게 중앙선을 넘어 맞은편 차선으로 피해야 한다. 최악의 경우는 인도를 침범해야 할지도 모른다. 그러나 이것도 사전에 기술자가 입력한 교통 법규를 위반한다. 이런 상황에서 자율주행차는 어떻게 행동해야 좋을지 판단할 수 없다.

현재 시제품 단계까지 간신히 도달한 자율주행차가 앞으로 상용화 단계로 발전하기 위해서는 가끔 일어나는 비상사태 혹은 전혀 예상치 못한 사태 등 어떤 상황에도 대응할 수 있게 되어야 한다.

하지만 기술자가 사전에 모든 상황을 상정하고, 이를 규칙 기반 AI 같은 형태로 일일이 프로그래밍하는 것은 사실상 불가능한 일이다. 무엇이 일어날지 모르는 현실 세계에서 이러한 상황을 헤아리다 보면 끝이 없기 때문이다.

그러면 어떻게 해야 할까? 여러 업체에서 선택지 중 하나로 검토하고 있는 방법은 비상사태 시 오퍼레이터가 원격으로 자율주행차를 조종하는 방식이다. 2017년 1월 닛산자동차는 미국의 라스베이거스에서 개최된 국제 전자제품 박람회(CES)에서 이를 위해 개발한 시스템을 발표했다.

이 시스템에서는 원격지에 있는 관제 센터가 자율주행차를 무선 인터넷으로 상시 감시한다. 만에 하나 앞의 사례처럼 도로 공사 등 비상사태가 발생하면 관제 센터의 오퍼레이터가 자동차 제어권을 이어받아 원격으로 조종한다. 이 기술은 '끊김 없는 자율주행(SAM: Seamless Autonomous Mobility)'으로 불리는 기술로 미 항공우주국 NASA의 기술을 바탕으로 개발되었다고 한다.

이와 기본적으로 같은 방법을 '웨이모', 스마트폰 앱으로 배차 서비스 사업을 제공하는 미국의 '우버', 일본의 '도요타' 등 경쟁사에서도 검토 중이다.

이처럼 같은 자율주행이라도 자동차의 제어권을 AI에게 완전히 넘기는 것이 아니라 인간이 어떠한 형태로든 제어에 관여하는 방식은 앞에 나온 'Huma out of the Loop'와 대비되는 방식으로 'Human in the Loop(인간이 제어권을 갖는다)'이라고 한다.

이것은 이미 '반자율주행(Semi Autonomous)'이라는 기술로 상용화되었다. 예를 들면 미국의 전기자동차업체인 '테슬라'가 2015년 10월에 공개한 '오토파일럿' 시스템 등이 해당한다. '브레이크 및 엑셀 조작', '전방 차량의 추적', '차선 변경' 등의 일부 운전 기능은 자동화되었으나, 자율주행 중에도 기후나 도로 및 교통상황의 변화 등 필요에 따라 일반 수동 운전으로 변환할 수 있다.

이러한 반자율주행은 제조업체들의 최종 목표인 완전 자율주행으로 가기 위한 중간 단계로 보인다. 이것은 현실적인 접근방식이기는 하지만, 나름의 위험성을 동반한다. 자율주행차 같은 기계와 인간 사이에서 제어권 전환이 잘 안 되는 경우가 발생하고 있다. 실제로 이와 같은 이유로 2016년 이후 테슬라의 오토파일럿

이 미국과 중국 등에서 몇 건의 교통사고를 일으켜 사상자가 발생했다.

이를 토대로 지금 심각하게 고려해야 하는 것은 제어권을 인간이 가지고 있어야 하는지의 여부이다. 필요하다면, 어느 정도까지 제어해야 할까? 이는 자율주행차의 제어권을 둘러싼 기계와 인간의 주도권 다툼이다. 그 향방과 이로 인한 중대한 위협은 제2장에서 살펴보고자 한다.

의료에 진출하는 AI

자율주행차와 더불어 인공지능에 의해 우리의 생사가 좌우되는 분야가 있다. 바로 의료이다. 평소 우리의 건강과 목숨을 맡고 있는 이 중요한 분야에 고성능이기는 하지만 수수께끼로 둘러싸인 AI가 진출하려 한다.

일반적으로 두뇌가 명석한 의사라고 하더라도 결국은 살아있는 인간이다. 그리고 인간의 최대 단점은 시야(지식의 범위)가 한정되어 있다는 사실이다. 아무리 지식을 갈고닦은 우수한 의사라고 해도 이 세상에 존재하는 모든 질병과 원인, 치료법 등에 정통할 수는 없다.

이에 반해 고속 프로세서와 대용량 기억장치를 갖춘 AI는 세계에서 매일 발표되어 축적되는 대량의 의학 논문을 눈 깜빡할 사이에 독파하고 의사가 알지 못한 병명과 치료법을 제시한다. 이러한 생각이 바탕이 되어 인공지능이 첨단 의료 현장에 도입되기 시작하였다.

이 분야를 개척한 상징적인 존재가 미국의 IT 대기업 IBM에서 개발한 '왓슨'이다. 왓슨은 원래 미국에서 국민적인 인기를 자랑하는 퀴즈 프로그램 '제퍼디!(Jeopardy!)'에 도전하기 위해 개발된 이색적인 컴퓨터였다.

왓슨은 인간처럼 자연어를 이해한다고 한다. 이 능력으로 동서고금의 역사, 문화, 정치, 경제, 예술 등의 문서를 모조리 독파하여 익혔다(물론 이것은 비유일 뿐으로 실제로는 엔지니어가 문서를 디지털화하여 왓슨에 입력한 것이다).

2011년 왓슨은 염원하던 제퍼디에 출연하여 전설적인 역대 챔피언 2명과 대전을 펼쳤다. 그리고는 잇달아 출제되는 고난도 문제에 정답을 맞추며 2명의 챔피언을 제압했다. 이 모습을 지켜본 미국의 시청자들은 왓슨과 현대 AI가 도달한 기술 수준에 경악했다.

이에 확신을 얻은 IBM은 단순히 자사 PR용 기계였던 왓슨을 본격적인 비즈니스용 컴퓨터로 개조하기로 했다. 약 3년의 준비 기간을 거쳐 2014년 왓슨 사업부를 정식으로 출범시켰다. 여기에는 당초 10억 달러(1조 원 이상)의 예산이 투입되었고, 이후 만 명의 스태프가 배치되는 등 향후 IBM의 기간 사업으로 왓슨을 육성하기 위해 전력을 다하고 있다.

이 책을 집필하고 있는 지금도 왓슨은 미국과 일본을 포함하여 세계 49개국에서 도입되어 활용되고 있으며, 그 분야도 25 업종에 달한다. 구체적인 용도로는 기업의 경영지원, 콜센터 고객지원업무, 세무 서비스, 금융 컨설팅 등 다방면에 걸쳐있다.

그중에서도 IBM이 초반부터 주력해왔고, 현재 사업부의 약 3분의 2 인원이 투입된 것이 '왓슨 헬스'라는 의료 비즈니스 분야이다. 이 분야는 신약 개발, 암 진단 지원, 게놈 해석 조언(환자의 DNA 등 유전정보를 해석하여 환자 개개인에게 최적의 치료법을 제공하는 신형 의료) 등 여러 의료 목적으로 왓슨을 응용하려 하고 있다.

IBM은 이를 위해 미국의 선진 의료 기관 등과 연계하여 그들이 축적한 암과 관련된 대량의 연구논문 및 수많은 환자의 '게놈(Genome: 전체 유전 데이터)'을 왓슨에 주입하여 입력했다. 이렇게 '퀴즈왕'을 '의료 전문가'로 탈바꿈시켜 의사의 어시스턴트로 활용하려 한 것이다.

이러한 시도는 눈부신 성과를 불러일으켰다. 미국에서는 왓슨이 암 치료에 활용된 약 1,000건의 사례 중 30%에서 의사는 생각지 못한 치료법을 제안하여 의학 관계자들을 충격에 빠뜨리기도 했다.

일본에서도 도쿄대학 의료학 연구소가 처음으로 왓슨을 도입하였고, 2015년 7월에는 급성 골수성 백혈병을 앓는 60대 여성의 진단에 사용해보았다. 그녀는 그

해 1월 도쿄대학병원에 입원한 후, 반년 동안 2종류의 항암제 치료를 받았으나 회복의 조짐은 보이지 않았고 패혈증까지 우려되고 있었다.

그러나 왓슨이 여성 환자의 유전정보를 해석하여 급성 골수성 백혈병 중에서도 진단이 어려운 '2차성 백혈병'이라는 특수한 질병이라는 것을 밝혀냈다. 이 정보를 바탕으로 의사가 항암제 처방을 변경하자 효험을 보이면서 병증이 완화되었고, 환자는 2개월 만에 퇴원할 수 있었다. 이 사례가 아마도 AI가 인간의 생명을 구한 최초의 케이스로 TV와 신문 등 일본의 각종 매체에 크게 보도되었다.

왓슨은 인도·태국·싱가포르 등의 병원에서도 도입되어 미국과 일본처럼 환자에 대한 진단과 치료법 등을 의사에게 적절하게 조언하고 있다. 또한, 일본의 공익단체법인 '암 연구회'와 AI 개발 회사 'FRONTEO 헬스케어' (둘 다 근거지를 도쿄로 두고 있음)가 공동 개발 중인 '인공지능의 암 프레시전 의료 시스템' 등 왓슨 이외의 다른 시스템도 점차 생겨나고 있다.

AI로 인한 새로운 의료 과실

이렇게 의료 분야에 진출한 인공지능으로 인해 밝은 미래가 펼쳐질듯 하지만 우려되는 심각한 요소도 있다. 새로운 의료 과실의 위험성이다.

물론, 이 책을 집필하고 있는 지금은 왓슨 등의 AI 사용으로 환자가 사망하거나 증상이 악화된 경우는 보고되지 않았다. 그러나 이것은 왓슨이 아직 테스트 운용 단계에 있기 때문일 것이다. 앞으로 왓슨 같은 AI가 본격적으로 병원과 클리닉 등 의료현장에 보급되면, 인공지능과 의사 사이에도 의견 차이가 발생할 것이다.

현재의 왓슨은 '의사의 어시스턴트' 위치에 있다. 대량의 의학 논문을 축적하여 의사가 알지 못한 병명 및 치료법을 제안할 수는 있지만, 이 조언을 참고로 최종 진단을 내리고 치료법을 결정하는 것은 결국 인간인 의사의 역할이다.

이것은 앞서 자율주행에서 소개한 'Human in the Loop'에 해당한다. 의료처럼

인간의 생사와 관련된 분야에서 인공지능에게 환자의 진단 및 치료를 일임하는 것은 아무래도 저항감이 크다. 그러므로 인간이 최종적인 결정권을 가지고 책임을 져야 한다. 이러한 방식은 당분간 유지될 것이다.

그러나 우려되는 것은 의사와 왓슨 같은 AI 사이에서 의견이 다를 때의 상황이다. 왓슨에는 '왓슨 패스'라는 의사 지원 기능이 갖춰져 있다. 의사는 이 기능으로 왓슨의 사고 경로를 되짚어 볼 수 있다. 즉, 왓슨이 어떤 논문을 참고하여 어떤 판단 기준과 사고방식으로 진단 결과와 치료법을 제시하게 되었는지 의사도 구체적인 경위를 살펴볼 수 있다.

그러나 아무리 왓슨의 사고 경로가 판명되었다 해도 의사와 의견이 다를 가능성은 충분히 있다. 실제로 인도의 병원에서 이런 사례가 있었다(제3장에서 소개).

왓슨이 제공하는 진단 및 치료법은 실제로 절대적인 정답이 아니며, 어디까지나 '정답일 확률이 높은 의료 정보'에 불과하다. 이 정보를 참고하여 최종 결정을 내리는 것은 의사이지만, 환자의 질환에 대해 자신의 견해와 왓슨의 조언이 어긋나는 경우 판단을 내리기가 상당히 어려울 것이다.

의사에게는 항상 '마이너리티 리포트(소수파의 의견)'라는 문제가 따라다닌다. 만약 한 질병에 대해 적힌 의학 논문이 100편 있다고 하자. 그중 90편은 'A'라는 치료법, 남은 10편은 A와 상충하는 'B' 치료법을 권장한다. 통계적으로 보면 A 치료법이 정답이겠지만, 의료 분야에서는 반드시 다수의 뜻이 정답이라 할 수 없다. 소수인 B가 올바른 치료법인 경우도 간혹 있는 것이다.

이런 와중에 왓슨이 다수의 의견인 A를 제안하고 의사가 소수 의견 B를 지지하거나 혹은 그 반대의 경우도 있을 수 있다. 만약 의사가 왓슨의 조언을 따르지 않고 소신대로 치료했지만 안타깝게도 환자의 증상이 악화하거나 사망했을 때, '왜 정답률이 높은 AI의 의견을 구태여 부정하고 자신의 의견을 고집했는지' 주위의 비판에 대한 걱정은 없을까?

반대로 앞으로 왓슨과 같은 의료 AI의 성능이 점점 향상됨에 따라 의사가 AI에 점점 의존하게 되어 최종적으로는 진단 및 치료를 사실상 전부 맡겨버릴 수도 있다.

이와 유사한 사태가 이미 '장기(將棋)'에서 발생하고 있다. 장기 소프트의 성능이 해마다 성장함에 따라 장기기사가 AI 소프트의 수를 정답으로 간주하는 경향이 늘어났고, 기사와 기사 간 대전에서 부정 의혹 문제가 일으키는 결과를 낳았다.

그러나 장기 소프트나 의료 AI의 답이 절대적인 정답은 아니다. 인공지능이 틀릴 가능성도 남아 있다. 장기는 승패가 나뉘는 것으로 끝나지만, 의료에 응용되는 AI에는 환자의 목숨이 달려 있다. 가까운 미래에 의사가 AI에 크게 의존하게 된다면, 그 AI로 인한 새로운 의료 과실은 현재와 비교할 수 없을 정도로 복잡한 결과를 초래할 것이다.

의료에 응용되는 딥러닝

이런 경향을 조장하는 것이 의료 현장의 딥러닝 도입이다. 제3장에서 자세히 설명하겠지만, 왓슨은 전통적인 '규칙 기반 AI'을 따르는 인공지능임에 반해 딥러닝은 금세기 들어 급격하게 발달한 뉴럴 네트워크 기술의 최신 모델이다(다만, IBM이 2015년에 '알케미 API'라는 벤처기업을 인수하면서 화상 해석 등 왓슨의 일부 기능에도 딥러닝을 탑재했다).

이런 딥러닝을 의료에 응용하려는 노력이 세계적으로 퍼지고 있다. 이를 선도하는 것은 구글(알파벳) 산하의 AI 개발기업인 영국의 딥마인드이다.

딥마인드는 2016년 3월 독자적으로 개발한 '알파고'라는 바둑 소프트는 세계 정상급인 한국인 바둑기사 이세돌 9단에게 승리하면서 단번에 세계적으로 유명해졌다. 알파고는 2017년 5월 중국 저장성에서 세계 최강 기사인 커제 9단도 3전 전승으로 승리했다. 이처럼 놀라운 바둑 소프트에 사용되는 AI 기술이 딥러닝이다.

딥러닝의 기술과 구조는 역시 제3장에서 자세히 다루겠지만, 기본적으로는 '머신러닝' 분야의 기술이라고 할 수 있다. 즉, 컴퓨터와 같은 기계가 '빅데이터'를

교재로 학습하여 발전하는 기술이다.

앞에서 딥러닝을 뉴럴 네트워크라는 기술의 최신 모델이라고 설명했는데, '머신러닝' 분야의 기술이라고 다시 설명하면 혼동할 수도 있다. 정확히 말하면, 머신러닝이라는 큰 분야 안에 뉴럴 네트워크가 포함되는 개념이라고 생각하면 거의 문제없을 것이다.

딥러닝은 뇌 후두부에 있는 시각 영역의 연구성과(이론)를 적용하고 있어, 화상 인식에 가장 최적화되어 있다. 또, 뇌의 지각 영역에는 범용성이 있으므로 시각 영역의 구조를 응용한 딥러닝은 청각 등과 같은 음성 인식에도 특화되어 있다. 이들은 일반적으로 총칭 '패턴인식'(패턴인식도 머신러닝의 일종)이라고 한다.

알파고에 탑재된 딥러닝이 바둑에서 강점을 발휘하는 이유는 바둑판 위의 흑과 백의 바둑돌이 만들어내는 패턴을 인간보다 신속 정확하게 인식할 수 있기 때문이다. 알파고는 사전에 입력된 수백만에 달하는 대국 기보 데이터와 알파고 끼리의 대국 데이터 등을 머신러닝으로 '어떤 패턴일 때 우세이고, 열세인지' 판단할 수 있다. 그리고 실전에서 바둑판 위의 흑돌·백돌의 패턴을 조금이라도 자신에게 우세한 방향으로 만들기 위한 수를 선택한다.

아마 인간 바둑기사도 실전에서는 거의 이와 유사하게 즉각적인 판단을 내리면서 대전을 펼칠 것이다. 알파고가 세계 수준의 기사에게 이길 정도가 되었다는 것은 뇌과학을 바탕으로 하는 AI의 패턴인식 및 형세판단력이 세계 톱클래스 기사도 뛰어넘을만한 수준에 이르렀다는 것을 의미한다.

블랙박스화되는 의료

딥마인드는 딥러닝의 우수한 패턴인식 능력을 바둑 같은 보드게임뿐 아니라 생명을 구하는 의학 분야에도 응용하려 하고 있다. 그들이 영국이나 인도의 병원과 공동으로 벌인 임상시험은 매우 좋은 결과를 얻었고, 향후 딥러닝으로 여러 질병의 진단 및 예측, 예방 등이 혁명적인 진전을 이룰 것이라는 사실도 거의 확

실시되고 있다.

그러나 한편으로는 심각한 문제에 대한 지적도 있다. 딥러닝과 같은 뉴럴 네트워크 기술에 공통으로 나타나는 '블랙박스화' 현상이다(제3장에서 설명).

블랙박스화란 문자 그대로 딥러닝이라는 '상자' 내부에서 무슨 일이 일어나는지 외부에서는 짐작할 수 없다는 것이다. 이는 앞의 IBM '왓슨'과는 대조적이다. 의사들은 '왓슨 패스'라는 지원 툴을 사용하여 왓슨이 무슨 경위로 어떤 병명과 치료법을 제시하게 되었는지 사고 경로를 나중에 추적할 수 있다.

이에 반해 내부가 블랙박스화된 딥러닝은 어떻게 결론에 이르게 되었는지 의사는 알 수 없다. 만약, 딥러닝으로 '이 환자는 이러한 희소질환을 앓고 있습니다. 이 질병에는 최근 개발된 이 신약이 효과적입니다'라고 조언을 들더라도 왜 그런 결론이 났는지 알 수 없는 것이다.

의사는 이런 상황에서 매우 고민스러운 결단을 내려야 한다. 딥러닝은 언제나 매우 높은 확률로 정답을 제시해준다. 즉, 이 조언에 따르면 환자를 구할 가능성도 커진다.

그렇다 해도 합리적인 이유를 알 수 없다면, 인명을 좌우하는 결단을 내리기가 쉽지 않다. 진찰실에서 의사가 환자와 그 가족에게 '이유는 잘 모르겠지만 딥러닝이 말하는 대로 이 치료법을 시도해 봅시다'라고 한다면, 받아들일 사람이 있을까?

그러나 놀랍게도 첨단 의료 분야는 실제로 이런 방향으로 나아가고 있다. 그래서 딥러닝의 블랙박스 문제를 해결하기 위하여 '이유를 설명할 수 있는 인공지능(Explainable AI)' 연구개발도 시작되고 있으나, 아직은 착수 단계이다. 이런 위험한 전개에 대해 사회적 논의도 이루어지지 않은 채, 의료 현장에서 AI 도입이 성급하게 진행되고 있다. 이 현상과 향방에 대해 제3장에서 다시 자세히 설명하고자 한다.

자율무기의 등장

위와 같은 의료는 생명을 '구하기' 위한 노력이지만, 반대로 생명을 '빼앗기' 위한 활동에도 AI가 도입되고 있다.

세계적 군사 강국인 미국은 최근 인공지능을 탑재한 차세대 무기 개발을 조용히 추진하고 있다. 예를 들어 스텔스 무인 전투기, 목표물 조준을 스스로 정하는 미사일, 상공에서 지상의 테러리스트를 감시하는 자율 드론 등 살인 로봇이라고도 불리는 AI 무기들이 '터미네이터'와 같은 SF 영화를 빠져나와 현실 세계의 새로운 위협이 되고 있다.

미 국방부에서는 AI가 탑재된 자율무기가 극히 일부의 예외적 존재가 아닌 광범위한 차세대 전력의 핵심으로 자리 잡았다. 현재는 AI를 바탕으로 한 근본적 군사 개혁을 한창 진행하는 중이다.

주요 동기는 중국과 러시아 등과의 군사력 차별화이다. 20세기 중반부터 지금까지 미국은 항상 (냉전 시의) 소련 등 다른 군사 강국과 차이를 두기 위하여 그 시대의 첨단 기술을 적용하여 강력한 신무기를 개발해왔다.

그 예로는 제2차 세계대전 중 맨해튼계획에서 나온 원자폭탄과 그것으로 인한 핵무기, 또한 1970년대 이후 개발된 레이저 정밀 유도 무기 등을 들 수 있다. 그러나 현재에 와서는 러시아·중국 등 5대 강국과 인도, 파키스탄, (아마도) 이스라엘, 북한까지 핵무기를 손에 넣고 정밀 유도 무기도 흔해지면서 미군은 기술적 우위성을 잃었다.

때문에 미 국방부는 이제 21세기의 혁명적 기술인 인공지능을 이용해 다시 중·러 등 다른 군사 강국과의 격차를 벌리려 하고 있다. 미국의 국방 관계자는 이것을 '3차 군사 상쇄 전략'이라고 부른다.

세계적인 군수업체인 미국의 록히드 마틴에서 개발하고 있는 'LRASM(Long Range Anti-Ship Missile)'이 그 시초로 상징적인 의미를 가지고 있다. LRASM에는 인공지능의 일종인 고도의 패턴인식 기술이 탑재되어 있어 미사일이 적의 전

함 등 공격대상을 발견하고 파괴할 수 있다.

이런 AI 무기는 미국 외에서도 개발·도입되고 있다. 영국에서 개발 중인 스텔라 무인 전투기 '타라니스(Taranis)'와 이스라엘이 이미 배치한 대레이더 미사일 '하피(Harpy)', 방공 시스템인 '아이언 돔', 나아가 한국이 비무장지대에 배치한 감시경계 로봇 'SGR-1' 등은 전부 무기가 스스로 판단하여 적을 공격하는 능력을 갖추고 있다. 그 외에도 프랑스, 노르웨이 등 자율무기 개발을 진행하고 있는 국가는 무수히 많다.

이런 사례들을 볼 때, 'AI 도입에 의한 군사력 상쇄'는 아마 미군뿐만 아니라 세계적인 경향이라고 할 수 있다.

기존의 무기와 결정적인 차이는?

AI와 같이 시대를 이끌어가는 첨단 기술로 전쟁과 전장의 양상이 일변하는 것은 어제오늘 일이 아니다. 예로부터 통치자들은 적을 물리치기 위해 새로운 기술에 눈을 돌려 무기에 적용해왔다.

투석기, 전투용 마차, 중세의 장궁, 석궁, 근세의 화기(대포, 총기)와 군함, 근세 이후의 기관총, 전차, 독가스, 잠수함, 군용항공기, 미사일, 레이저, 핵무기 등등……. 전쟁의 역사는 군사기술의 역사이기도 하다.

그러나 이런 과거의 혁신적 무기와 앞으로 전쟁에 투입될 AI 무기에는 결정적인 차이가 있다.

지금까지의 무기에는 주로 파괴력과 공격 범위를 확대할 수 있는 신기술이 도입되었다.

이에 반해, 현재 개발되고 있는 AI 무기에는 공격대상이 되는 적을 설정하거나 상대를 공격할지 판단하는 능력을 갖추고자 한다. 기존에 이런 판단은 군대의 지휘관과 전장의 군인들이 맡는 역할이었다.

이제는 무기가 '인간이 사용하는 도구'가 아닌 '인간을 대신하는 전투 주체'로 질적인 변화를 시작한 것이다.

현재 '3차 군사 상쇄 전략'을 추진하고 있는 미 국방부는 공식문서로 이러한 점을 언급하였다(제4장에서 설명). 이에 따르면 AI 무기가 아무리 전장에 투입된다고 해도 공격의 최종 판단을 내리고 책임을 지는 것은 지휘관과 병사라고 한다.

그러나 실제로는 추상적으로 기술되어 있고, 풀이에 따라 다르게 해석될 수 있다. 앞으로는 무기의 자율성이 커진다는 의미를 내포하는 것이다.

지금까지 기술한 것처럼 AI 위협론의 본질은 '그 제어에 인간이 관여하지 않는 것'에 있다. 하지만 적어도 현 단계에서 그런 진화의 방향성을 정하는 것은 연구자와 우리 일반인을 포함한 인류의 책임이다. 이어서 각 장에서는 '자율주행', '의료', '무기' 순으로 우리의 목숨과 직결되기 시작한 AI 개발의 현재와 그 향방에 대해 자세히 살펴보고자 한다.

제2장 자율주행차의 사각지대

2016년 5월 미 테슬라의 전기자동차 '모델S'가 일으킨 충돌사고는 AI의 잘못된 판단이 한순간 귀중한 목숨을 앗아간 최초의 사고로 역사에 기록될 것이다.

자세한 사고의 경위는 뒤에 설명하겠지만, 요컨대 고속도로를 주행 중이던 모델S는 맞은편 차선에서 왼쪽 커브를 틀던 대형 트레일러 차량과 충돌했다. 이 사고로 모델S는 크게 파손되고 운전자가 사망했다. 이 차종에는 AI 기술을 기본으로 하는 '오토파일럿'이라는 (한정적 혹은 부분적) 자율주행 시스템이 탑재되어 있다. 사고 발생 당시 모델S는 이 자율주행 모드로 주행하고 있었다.

최근 몇 년간 구글과 세계 각국의 자동차업체는 자율주행차의 개발을 가속하였고, 시가지를 포함한 공공도로에서 장거리의 시험주행을 반복해왔다. 이 과정에서 '접촉사고'와 같은 가끔 가벼운 사고들은 있었지만, 운전자와 동승자가 사망 혹은 중상을 입는 중대한 사고는 한 번도 발생한 적이 없었다.

그래서 순조롭게 개발이 진행되면 앞으로 점차 부분적 자율주행 시스템이 제품에 도입되고, 2020년쯤에는 완전 자율주행 내지는 이와 근접한 기능이 상용화될 것이라는 예측이 많았다. 그때 마침 사망사고가 벌어지면서 그때까지의 낙관적인 관측에 심각한 악영향을 끼쳤고, 자율주행 상용화에 대한 각 제조업체의 장래계획에도 적잖은 영향을 주었다.

미국의 GM은 2016년 가을에 제품화할 예정이던 (한정적인) 자율주행차의 발매를 2018년까지 연기했다. 테슬라의 자동차 사망사고로 인해 자율주행 시스템의 안정성을 한층 강화하고 시장에 투입하려는 의도로 보인다.

또, 미국의 포드는 '(오토파일럿 같은 한정적인 자율주행 시스템이 아니라) 운전자가 필요 없는 완전 자율주행 시스템을 2021년까지 상용화할 것이며, 처음에는 일반 소비자용이 아닌 (우버와 같이) 스마트폰을 사용한 라이드 셰어(차량 공유) 사업 등에 제공할 것'이라는 계획을 밝혔다.

한편, 독일의 BMW는 테슬라의 자동차 사고 후에도 '자율주행 시스템을 개발한다는 방침에는 변함이 없지만, 제품화는 2021년 이후가 될 것이며 이 기술은 테슬라 모델S에 탑재된 자율주행 기술(오토파일럿)과는 크게 다를 것'이라고 발표하였다.

스웨덴의 볼보도 이 사고 이후 '한정적 자율주행이 아니라 (구글이 개발하고 있는) 완전 자율주행 시스템을 목표로 하겠다'는 방침을 내놓았다(유럽의 업체들은 이미 기초적 자율주행 시스템을 상용화했다).

이러한 사례로 각 업체에서 자율주행 시스템과 관련된 방침 및 계획을 상당 부분 수정하고 있는 것을 알 수 있다.

사망사고의 현장 검증

테슬라 자동차 모델S의 충돌사고는 왜 그렇게 큰 충격을 불러온 것일까? 그 전에 사고의 구체적인 상황을 알아둘 필요가 있다.

이 사고는 2016년 5월 7일 미국의 플로리다주를 횡단하는 고속도로 'US-27A'에서 발생했다. 참고로 US-27A와 같은 미국의 '간선도로(Highway)'는 기본적으로 무료로 이용할 수 있어 제도상으로는 국도와 같다.

그러나 실제로는 제한속도가 '55~57마일(시속 약 89~121km)' 정도의 고속이며, 본 차선에 합류하기 위한 램프(입·출구)까지 설치되어 있어 '사실상의 고속도로(Freeway)'이다(반대로 시가지나 주택가 등을 달리는 일반도로는 'Surface road' 혹은 'Residential street' 등으로 부른다).

사고는 이 고속도로 US-27A를 남동 방향으로 (오토파일럿으로) 자동 주행하던 모델S(그림 1: V02)가 맞은편 차선에서 교차로 'NE 140th Court'로 진입하기 위해 급하게 좌회전한 대형 트레일러 차량(그림1: V01)의 우측면으로 돌진하면서 발생했다.

그림 1 테슬라 자율주행차로 인한 충돌(사망) 사고 현장

출처: Florida Highway Patrol/일부 변경

대형 트레일러는 차체가 상당히 높으므로 차체의 바닥 면과 도로 사이에 상당한 틈새가 생긴다. 이 때문에 모델S는 트레일러의 진행 방향 우측면에 충돌했다기보다는 그 틈새에 파고든 형태가 되었다. 그때까지 시속 65마일(105km)로 주행하고 있던 모델S는 그 기세로 트레일러의 하부 틈새를 통과하여 트레일러의 좌측면으로 빠져나왔다.

이로 인해 모델S의 천장은 트레일러의 아랫면과 격하게 마찰하여 분리되었고 그 충격으로 경로도 우측으로 크게 이탈했다. 그리고 고속도로 펜스를 뚫고 직진하여 전방의 전신주를 들이받았다. 결국, 차체는 크게 파손되고 운전자도 사망했다.

위와 같은 경위로 알 수 있듯, 모델S에 탑재된 오토파일럿은 급커브로 눈앞을 가로막은 트레일러를 인식하지 못하고 그대로 차량에 돌진했다.

테슬라는 사고 직후 오토파일럿이 트레일러의 하얀 차체와 파란 하늘 배경을 구별하지 못하여 장애물로 인식하지 못한 것이 사고의 원인이 되지 않았을까 추측했다.

그러나 그 후, 사고 당시 오토파일럿이 정상 동작하고 있었으나 자동 브레이크가 작동하지 않아 사고가 났다는 새로운 견해를 내보였다. 물론 일반 운전자 중에는 자동 브레이크도 오토파일럿의 기능임에도 둘을 굳이 구별하는 테슬라의 입장에 위화감을 표명하는 의견도 있었다.

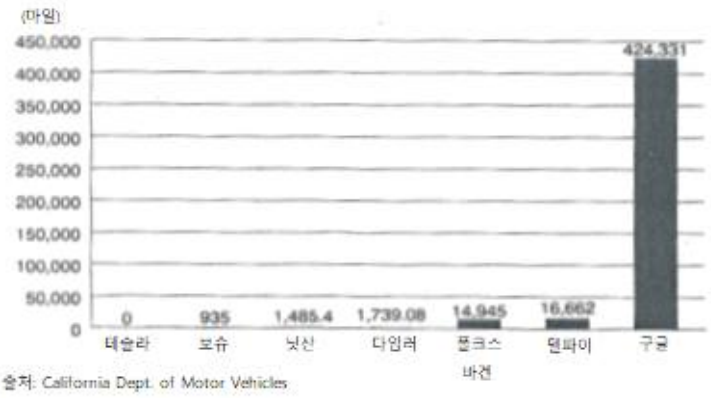
이듬해 발표된 미국 NTSB(연방교통안전위원회)의 조사결과에 따르면, 사망한 운전자는 (일정한 시간 동안 핸들을 잡지 않을 때 울리는) 경고음도 무시한 채 손을 놓고 있었다고 한다.

공공도로 테스트 주행이 부족했다

이 사고를 미연에 방지하는 것은 불가능했을까? 공교롭게도 실제로 오토파일럿의 안전성에 경종을 울리는 데이터가 사고 발생 수개월 전에 보고된 바가 있다. 바로 2016년 1월 미국 DMV가 발표한 조사결과이다.

이 조사에서 DMV는 자율주행 시스템을 개발 중인 여러 기업을 대상으로 주행 테스트의 데이터 제출을 요청했다. 대상은 2016년 1월까지의 과거 14개월 동안 캘리포니아주 공공도로에서 실시된 자율주행차 주행 테스트 결과였다.

그림 2 미 캘리포니아주의 각 업체별 자율주행차 테스트 주행거리



출처: California Dept. of Motor Vehicles

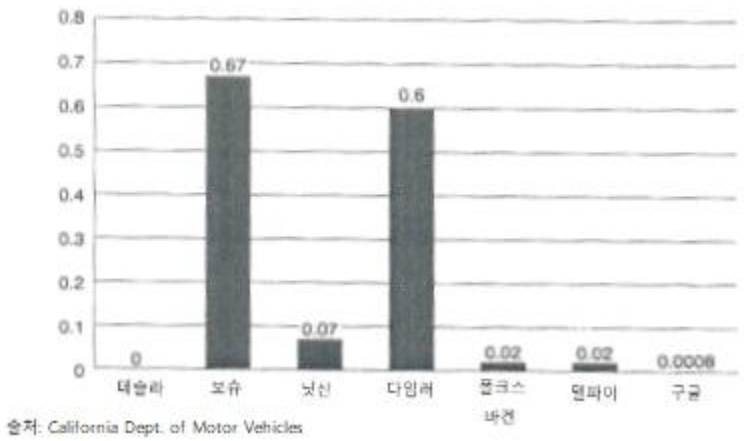
요청에 응한 것은 구글, 다임러(메르세데스-벤츠), 폴크스바겐, 닛산자동차, 보슈, 델파이, 테슬라였다. 물론 다른 곳에서도 주행 테스트가 이루어지고 있었을 테지만 공공도로 자율주행 테스트를 위한 환경이 가장 잘 정비된 곳이 캘리포니아주였다. 따라서, 이 2016년 1월에 제출된 데이터가 업체들의 자율주행 개발 및 테스트 상황을 상당 부분 반영하고 있다고 해도 과언이 아닐 것이다.

이 데이터를 보면, 먼저 공공도로의 테스트주행 총거리는 구글이 압도적으로 타의 추종을 불허한다(그림 2). 물론 구글이 제일 처음 본격적인 자율주행 개발을 시작했으므로 가장 우위에 있는 것이 당연하다. 하지만 최근 수년 동안 미국, 독일 및 일본을 비롯한 전 세계 각국의 업체도 자율주행 분야에 뛰어들었고 이를 어필해왔다. 그런데도 최근까지 이렇게 차이가 벌어지는 것은 다소 놀랍기도 하다.

2016년에 발표된 이 DMV의 조사결과에는 자율주행 테스트 중 트러블이나 비상사태 혹은 복잡한 도로상황으로 인해 승차 중이던 인간이 자율주행에 개입하여 직접 운전을 한 횟수도 포함되어 있다.

하지만 자율주행 테스트의 거리가 멀어질수록 사람이 개입한 횟수가 많아지는 것은 당연하다. 좀 더 공정하게 평가하기 위해서는 개입횟수를 주행거리로 나누어 산출할 필요가 있다. 그림 3이 바로 그 데이터이다.

그림 3 단위거리(1마일)당 개입 횟수



출처: California Dept. of Motor Vehicles

인간의 개입횟수가 적을수록 자율주행 시스템의 완성도가 높다고 볼 수 있다. 이 그래프에서도 구글이 압도적으로 좋은 성적을 보여주고 있다.

그리고 눈에 띄는 것이 테슬라의 데이터이다. 공공도로에서 테슬라의 주행거리와 인간의 개입횟수는 모두 “0”이다. 데이터로 보면 테슬라는 (적어도 캘리포니아주에서는) 공공도로 테스트주행을 전혀 하지 않았다.

그러나 테슬라는 (이 조사가 이루어진 시점에) 이미 다른 회사들을 제치고 부

분적 자율주행 시스템인 오토파일럿을 상용화했다. 테슬라가 공공도로 테스트를 전혀 하지 않는 이유는 무엇일까?

미국의 워싱턴포스트지는 '테슬라가 (오토파일럿을 공개한 후) 사용자들을 실험대상으로 하여 공공도로 테스트를 하고 있다'라고 하고 있다("These charts show who's lapping whom in the race to perfect the driverless car" Brian Fung and Matt McFarland, *The Washington Post*, Jan. 15, 2016).

실제로 2016년 5월에 사망사고가 발생하기 전부터 오토파일럿의 사용자 중에는 매우 위험한 순간을 경험한 사람도 적지 않다. 테슬라는 이런 데이터를 인터넷으로 수집 및 해석해서 오토파일럿의 완성도를 높여 가고 있다.

한 예로 오토파일럿이 공개된 지 얼마 안 되었을 즈음, 자율주행 중이던 테슬라 자동차가 고속도로 출구의 램프와 가까워지면 사용자가 지시하지 않았는데도 자동으로 램프에 다가가는 현상이 몇 차례 보고되었다. 이후, 이 문제를 일으킨 버그는 다음 버전에서 수정되었다.

이런 방식은 구글과 페이스북 등 IT 기업이 각종 소프트를 베타 버전(시제품) 단계에서 공개하고, 사용자의 반응을 보면서 서서히 버그 등의 문제를 수정하여 완성도를 높여가는 방식을 떠올리게 한다. 컴퓨터·스마트폰을 위한 소프트웨어라면 그래도 문제가 없을지 모르지만, 오토파일럿처럼 인명을 좌우하는 자율주행 시스템의 경우는 크게 잘못되었다고 말할 수밖에 없다.

미 정부는 소비자 보호보다 산업육성을 우선

미국의 소비자 단체들은 테슬라를 향해 비난을 쏟아냈다.

또, 미국의 정부 기관인 'NHTSA(고속도로교통안전국)'와 'NTSB(연방교통안전위원회)'도 사고 직후부터 이 원인에 대한 조사에 착수했다.

그러던 중 2017년 1월에 NHTSA는 '오토파일럿이라는 제품 자체에는 결함이

없었다'는 조사 결과를 발표하였다.

그에 따르면, '오토파일럿은 어디까지나 전방 차량과의 추돌사고 등을 방지하기 위한 목적으로 설계되었으며, (이번 사고처럼) 맞은편 차선을 달리던 차량이 이쪽 차선에 침입하여 눈앞을 가로지르는 사태는 이 시스템의 능력 범위 밖의 일'(NHTSA의 공식 견해)이라고 한다.

여기에서 주의할 점은 테슬라가 미리 운전자 측에 '오토파일럿은 (자율주행이 아니라) 운전 지원을 위한 기능이며 할 수 있는 기능에는 한계가 있다고 미리 양해를 구했다는 것'이다. NHTSA도 '테슬라가 오토파일럿의 한계를 사전에 운전자에게 알린 이상, 그 한계를 벗어난 사태로 인해 사고가 발생했다면 그 책임을 오토파일럿(테슬라)에 넘기는 것은 불가능하다'고 판단했다.

하지만 이것은 '오토파일럿이 완전히 안전한 시스템이고, 운전자가 안심해서 사용할 수 있다'는 이야기가 아니다. 오히려 '극히 한정된 기능밖에 없으므로 운전자가 이 사실을 인식하고 신중하게 사용해달라'는 의미이다. '오토파일럿에 결함이 없다'는 것은 어디까지나 '테슬라가 미리 운전자에게 공지한 제품 스펙과 다름이 없다'는 의미에 지나지 않는다.

그리고 사고로 사망한 운전자 측의 책임에 대해서도 언급하고 있다. NHTSA의 공식 견해에 따르면 사고를 피하지 못한 것은 운전자가 손을 놓고 있는 등 오토파일럿의 예상 방식으로 운전하지 않았기 때문에 그 책임이 운전자 본인에게 있다고 한다.

이런 내용에 대해 테슬라는 당연히 환영하고 있으며, 테슬라뿐만 아니라 세계의 주요 자동차업체에서도 긍정적인 견해을 보이고 있다. 만약 정반대의 판단이 내려져 오토파일럿에 책임을 묻게 되었다면, 가까운 미래에 새롭게 거대한 시장을 형성할 '자율주행 시스템' 개발 및 제품화에 저해요인이 되었을 것이다. 이를 미연에 방지할 수 있었으므로 관계 업계의 모두가 가슴을 쓸어내렸다.

반대로 운전자의 입장에서는 이번 NHTSA의 조사 및 판단에 몇 가지 의구심이 들 수밖에 없다. 먼저, 테슬라 측의 과장 광고를 NHTSA가 용인했다는 점이다.

테슬라는 공식적으로 확실하게 '오토파일럿은 자율주행차가 아니다'라고 사전 공지를 하였지만, 실제로 미디어 광고에서는 반자율주행에 가까운 기능인 것처럼 소개했다. 이로 인해 (이번 사고로 사망한 운전자처럼) 오토파일럿의 능력을 실력 이상으로 과신하는 운전자가 다수 나왔다는 점은 부정할 수 없다. NHTSA는 조사결과에서 이점에 대해 언급하면서도 그 책임을 강하게 추궁하지는 않았다.

어중간한 자율주행은 운전자를 혼란스럽게 한다

또 한가지는 오토파일럿과 같은 (사실상) 반자율주행 시스템의 위험성이 확인된 것이다. 사실 이번 사망사고 외에도 오토파일럿과 관련된 사고, 그리고 거의 사고로 이어질 뻔한 사례는 미국 안에서만 총 10건이 발생했다. 미국보다 앞서 2016년 1월에는 중국의 허베이성에서도 오토파일럿을 사용하다가 발생한 것으로 보이는 사망사고가 있었다(테슬라는 '오토파일럿의 작동 여부는 불명확'하다는 입장이다).

이번에 NHTSA가 미국 내의 오토파일럿 관련 사고를 조사하던 중 밝혀진 '운전자의 혼란(Mode Confusion)'이라는 현상이 있다.

반자율주행은 자동차의 제어권이 '운전자'와 '자율주행 시스템'을 왔다 갔다 한다. 그러는 동안 운전자는 '지금 운전을 자신이 하고 있는지, 자율주행 시스템이 하고 있는지' 혼란을 느끼게 된다. 이것이 '운전자의 혼란'이며, 오토파일럿과 관련된 사고의 주요 원인이 되고 있다. 이런 현상에 대해 예전부터 우려의 목소리는 있었지만, 이번 NHTSA의 조사를 계기로 실체가 확인되었다.

구글과 같이 한 번에 완전 자율주행 시스템의 상용화를 노리는 것이 아니라, 반자율주행 시스템부터 착수하여 서서히 스펙을 쌓아 가고 최종적으로 완전한 자율주행을 이루는 것이 세계 주요 자동차업체들이 적용해온 방식이다. 그러나 여기에는 '운전자의 혼란'이라는 과제가 반드시 따라온다.

한편, NHTSA는 '(오토파일럿의 일부인) 자동 스티어링(Autosteer) 기능을 도

입한 이후, 테슬라 자동차 사고율이 (도입 전보다) 40%나 낮아졌다'는 조사결과도 발표했다. 이 시스템의 장단점을 저울질하면서 이런 판단까지 하게 된 것이라 할 수 있다.

참고 문헌

간자키 요지(2017), 『최신 인공지능 쉽게 이해하고 넓게 활용하기』, 위키북스

미야케 요이치로, 모리카와 유키히토(2017), 『인공지능 70 재미있게 알아보는 AI 키워드』, 제이펍

열린책들 편집부(2017), 『열린책들 편집 매뉴얼 2018』, 열린책들

이은용 외 4명(2017), 『최신 ICT 시사상식』, 전자신문사

정용찬(2013), 『빅데이터』, 커뮤니케이션북스

pmg 지식엔진연구소(2017), 『시사상식사전』, 박문각

大野治(2017), 『俯瞰図から見える日本型"AI(人工知能)"ビジネスモデル』, 日刊工業新聞社

小林雅一(2015), 『AIの衝撃』, 講談社

국립국어원 http://www.korean.go.kr

국립국어원 표준국어대사전 http://stdweb2.korean.go.kr

네이버 일본어 사전 http://jpdic.naver.com

두산백과 http://www.doopedia.co.kr

goo辭書 http://dictionary.goo.ne.jp

weblio辭書 http://www.weblio.jp

Yahoo Japan https://www.yahoo.co.jp

日本語抄錄

今回翻譯した小林雅一の『AIが人間を殺す日』(集英社, 2017)は、現時点のAI技術の位置と、第1次・第2次・第3次産業革命を通じて成し遂げた自動化とAIの導入による第4次産業革命の自動化が、根本的にどのような違いがあるかを示している。また、單純に日常の豊かさを与える範囲を超えて、生と死を左右する重大な分野へまで擴大されているAI技術に對して、私たちが知るべきことは、そして警戒すべきことは何なのかを指摘している。

本書は「はじめに」、「第1章　AI脅威論の虛實」、「第2章　自動運轉車の四角」、「第3章　ロボ・ドクターの誤診」、「第4章　自律的兵器の照準」、「第5章　スーパー・オートメーションの罠」、「おわりに」で構成されている。

本論文では 「はじめに」、「第1章」、「第2章」の一部を翻譯している。

「はじめに」では、全体的な話に對する概略と本書が書かれた背景を紹介している。「第1章　AI脅威論の虛實」では、現時点のAI技術で眞の脅威になっていることは何なのか、また自動運轉技術や医療、そして兵器に導入されたAIを事例を通して具体的に說明している。「第2章　自動運轉車の四角」では自動運轉技術に對する本格的な說明と共に、米テスラが開發した自動運轉車(オートパイロット)による死亡事故を詳しく分析している。