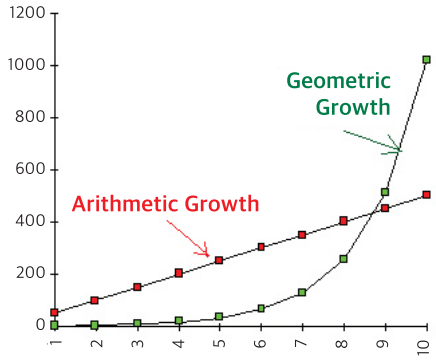
나군의 거친 생각
초인공지능

글. 나영채 YTN IT시스템팀

들어가기에 앞서



산술급수적 증가 vs 기하급수적 증가

전설에 따르면, 1256년에 인도의 시람(Shirham) 왕이 전쟁을 너무 좋아해 백성들이 늘 불안에 떨고 있었다고 한다. 이 상황에 마음이 아팠던 시사 벤 다히르(Sissa ben Dahir) 수상은 가로와 세로 각각 8칸씩 총 64칸으로 이뤄진 정사각형의 판 위에 여러 종류의 말을 놓고, 정해진 규칙에 따라 말을 움직이는, 전쟁과 비슷한 규칙을 가진 체스의 조상격인 차투랑가(Chaturanga) 게임을 만들어 냈다.

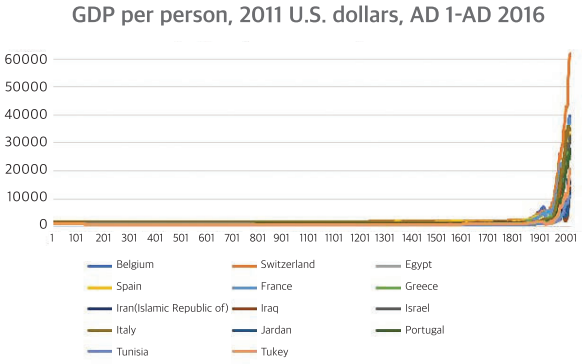
다행히 왕은 차투랑가에 흥미를 느껴 전쟁을 소홀히 하게 되었고, 왕은 차트랑가를 개발한 시사 수상에게 그 대가로 어떤 상을 원하느냐고 물었다고 한다. 그러자 시사는 차투랑가 판의 처음 칸에 밀알 1톨, 둘째 칸에 2톨, 셋째 칸에 4톨과 같이 두 배씩 밀알을 늘려 64칸을 채워달라고 요구했다. 왕은 소박한 제안이라고 생각해서 기꺼이 승낙했다.

하지만 이 방식으로 차투랑가 판을 채우려면 밀알이 18,446,744,073,709,551,615톨이 필요하게 된다. 1㎠에 약 92톨의 밀알을 담을 수 있다고 가정할 때 1㎡에는 92,000,000톨의 밀알을, 1㎢에는 92,000,000,000,000,000톨의 밀알을 담을 수 있으므로, 전체 부피는 200㎢가 넘게 되어, 한 국가의 창고를 털어서도 지급하기 불가능한 양이 된다. 기하급수적 증가가 얼마나 가공할 위력을 가졌는지 보여주는 흔한 예이다.

인류의 발전은 산술급수적이라기보다는 기하급수적인 모습을 보인다. 인구 증가가 그랬고, 기술의 발전이 그랬고, 생산성의 향상이 그랬고, 데이터의 성장이 그랬다.

분명 세상은 기하급수적으로 발전하고 있지만, 변화의 속도가 가속된다는 사실의 의미를 제대로 이해하는 사람이 드물어서, 당대를 사는 대부분은 선형적 차원에서만 그것을 인지한다. 그래서 먼 미래에 어떤 것들이 기술적으로 실현 가능할지 예측할 때, 대부분의 사람은 미래의 발전력을 턱없이 과소평가한다. '직관적으로 느껴지는 선형적' 역사관에 기반하기 때문이다.

하지만 개중에는 시대를 뛰어넘는 통찰력으로 우리에게 미래를 제시해주는 사람들이 있기 마련이다. 우리가 이러한 천재들의 목소리에 귀 기울여야 하는 이유이다.

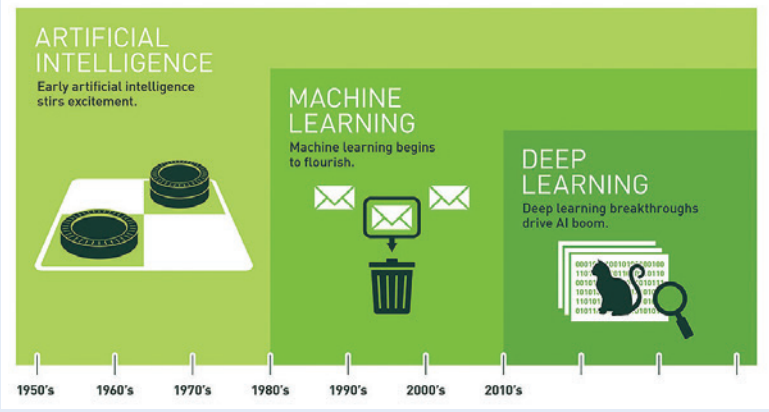


1인당 GDP(AD 1~AD 2016)

필자가 인공지능(Artificial Intelligence, AI)으로 변하게 될 미래의 모습을 깊이 있게 고민하게 된 계기는 2016년 3월 알파고가 한국의 프로 바둑기사 이세돌 9단과 대국하던 시점부터이다. 이런저런 글을 찾아보다가 Tim Urban이 작성한 'The AI Revolution'의 번역본을 읽고서 머리를 세게 맞은 듯한 당혹감에 휩싸였다. 그리고 이전에 인간의 모습을 한 친근한 인공지능 이미지가 송두리째 바뀌게 되었다. 인공지능이 가져올 결과는 '겨우' 산업혁명의 수단 정도에 그치는 것이 아니라는 사실을 알게 되었기 때문이었다.

이야기를 본격적으로 시작하기 전에 핵심 용어를 정리하고자 한다.

인공지능의 진화 단계에 따른 종류



인공지능, 머신 러닝, 딥 러닝의 관계

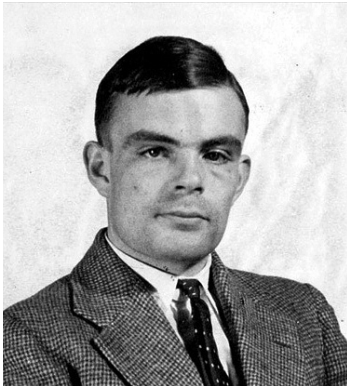
- **약인공지능 : Weak AI, Narrow AI, Artificial Narrow Intelligence(ANI)**
  한 가지 영역에 특화된 인공지능을 말한다. 바둑에 특화된 알파고(AlphaGo), 애플의 시리(Siri), 구글의 번역, 넷플릭스의 콘텐츠 추천 시스템 등 이미 일상생활 전반에서 쓰이고 있다. 머신 러닝(Machine Learning : 인공지능을 구현하는 구체적 접근 방식), 딥 러닝(Deep Learning : 완전한 머신 러닝을 실현하는 기술) 등의 현존하는 기술을 기반으로 제작된 모든 인공지능은 약인공지능 수준이다.

- **강인공지능 : Strong AI, General AI, Artificial General Intelligence(AGI)**
  다양한 영역에서 인간과 동일한 수준의 지적 과제를 수행할 수 있는 인간 수준의 인공지능을 말한다. 강인공지능은 경험을 통해 학습하고, 추상적으로 생각할 수 있으며, 논리적으로 추론할 수 있고, 장기적으로 계획할 수 있다. 약인공지능보다 훨씬 만들기 어렵고 따라서 아직 개념으로만 존재한다.

- **초인공지능 : Super AI, Superintelligence, Artificial Super Intelligence(ASI)**
  모든 영역에서 인간보다 훨씬 뛰어난 지적 능력을 갖춘 인공지능이다. 인류 멸종이니 영생이니 하는 문제의 시발점이다.

인공지능이란

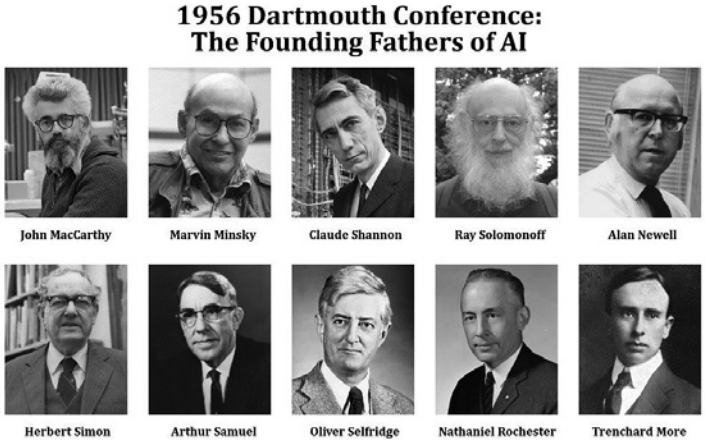
인공지능이란 인간의 지능이 가지는 학습·추리·적응·논증 따위의 기능을 갖춘 컴퓨터 시스템을 의미한다.



앨런 튜링, 인공지능의 가능성을 발견

현대적 의미의 인공지능이란 개념이 등장한 시기는 1950년대이다. '현대 컴퓨터의 아버지'로 불리는 영국의 수학자 앨런 튜링(Alan Turing)은 1950년 생각하는 기계의 구현 가능성에 대한 분석이 담긴 '계산 기계와 지능(Computing Machinery and Intelligence)'이라는 논문을 발표했다.

그는 '생각'을 정의하기가 어렵다는 사실에 주목해, 튜링 테스트(Turing Test)를 제안했다. 전신 타자기를 통한 대화에서 기계가 사람인지 기계인지 구별할 수 없을 정도로 대화를 잘 이끌어 간다면, 이것은 기계가 '생각'하고 있다고 말할 충분한 근거가 된다는 것이었다. 튜링 테스트는 인공지능에 대한 최초의 심도 있는 철학적 제안이었다.



다트머스 회의 참석자, 인공지능의 아버지들

인공지능이라는 단어는 1956년, 미국 다트머스 대학의 존 매카시(John McCarthy) 교수가 중심이 되어, 마빈 민스키(Marvin Minsky), 클로드 섀넌(Claude Shannon), 나다니엘 로체스터(Nathaniel Rochester)가 공동으로 제안한 다트머스 회의(Dartmouth Conference)에서 처음 등장했다.

여름 두 달간 진행된 다트머스 회의에서 당대의 유명 수학자·과학자 10명이 꿈꾼 것은 최종적으로 인간의 지능과 유사한 특성을 가진 복잡한 컴퓨터를 만들기 위한 로드맵을 그리는 것이었다.

그리고 그때까지만 해도 그들은 인공지능의 미래를 낙관했다. 당시 하버드대 교수였던 마빈 민스키 매사추세츠공과대(MIT) 명예교수는 훗날 발간한 책에서 "그때만 해도, 한 세대가 지나기 전에 인공지능을 창조하는 문제를 충분히 풀 수 있을 것으로 봤다"고 회고했다. 하지만, 그가 꿈꿨던 미래는 아직 오지 않았다.

우리 현재의 위치

현재 인공지능 기술은 미국, 중국 등이 주도하고 있다. IBM은 자연언어 처리를 위해 만든 인공지능 왓슨(Watson)의 기술을 의료분야에도 응용한 상태이다. 영국 관절염 연구기관(Arthritis Research UK)은 왓슨과 공동으로 관절염 환자에게 맞춤식 정보와 조언을 제공하는 가상 비서를 개발했다. 이 기관의 웹 사이트에는 매년 관절염에 대한 영향, 증상, 치료 옵션에 대한 수천 가지 질문이 올라오는데, 왓슨은 이 질문들을 분석해 신속하게 대응할 수 있도록 왓슨의 기술을 통해 답변해 주고 있다.

이외에도 구글에 인수된 딥마인드(DeepMind)가 개발한 머신 러닝 기반 바둑 프로그램인 알파고(AlphaGo), 애플에서 개발한 인공지능 음성인식 개인 비서인 시리(Siri), 아마존에서 제작한 인공지능 음성인식 비서인 알렉사(Alexa), 검색엔진을 이용하는 사용자와 브랜드를 마케팅하고자 하는 기업을 매칭하는 바이두(Baidu)의 검색엔진 등이 이미 활약 중이다.

이 모두 특정 영역에서 사람의 지능·효율과 비교해서 비슷하거나 초과하는 기계 지능 수준으로 약인공지능 수준에 머물고 있다. 현재의 약인공지능은 그렇게 무섭지 않을 수도 있다. 최악의 상황이라 해봤자 고작해야 프로그램 버그로 오작동하여 단독적인 재난이 발생하는 것이다. 이를테면 대규모 정전, 원자력발전소 고장, 금융시장 붕괴 등이다.

강인공지능 개발이 어려운 이유

인간 수준의 지능을 가진 컴퓨터를 만드는 게 얼마나 어려운가를 이해하려면 우리 인간의 지능이 얼마나 불가사의한지를 이해해야 한다.

하늘을 찌르는 빌딩을 짓고, 인간을 우주에 보내고, 우주 빅뱅의 디테일을 이해하는 것, 이 모든 것은 인간의 뇌를 이해하고 유사한 것을 만드는 것보다도 훨씬 쉽다. 즉, 현재까지 우리가 알고 있는 가장 복잡한 사물은 인간의 뇌이다.

게다가 인공지능을 만드는 어려움은 여러분이 직감적으로 생각하는 것과는 다소 차이가 있다.

- 한 동물이 고양이인지 개인지 판독하기 → 인공지능에게는 어렵다.
- 걸어가는 것 → 매우 어렵다. 구글은 네 발 달린 로봇으로 잘 알려진 로봇 업체인 보스턴 다이내믹스(Boston Dynamics)를 소프트뱅크에 매각했다. (2017.6.9.)
- 열자리 수를 곱하기 → 매우 쉽다.
- 세계 체스 챔피언을 이기는 컴퓨터를 만들기 → IBM의 딥 블루(Deep Blue)가 성공했다. (2017.5.11.)

이를 간단명료하게 "인간에게 쉬운 일은 로봇에게 어렵고, 로봇에게 쉬운 일은 인간에게 어렵다"는 '모라벡의 역설(Moravec's Paradox)'로 요약할 수 있다.

이미지 판독 예시

이 역설의 창시자인 한스 모라벡(Hans Moravec)은 자신의 역설이 진화에 기반을 둔다고 설명한다. 짧은 시간 동안 개발된 로봇의 능력과 달리 인간의 진화는 수억 년에 걸쳐 일어난 일이며 인간의 추상적 사고는 고작 십만 년밖에 안 되는 비교적 최근에 얻어진 능력이라고 말한다. 그래서 인간은 걷거나 보는 감각적인 일은 아주 잘해내지만, 계산과 같은 추상적 사고는 힘들다고 설명한다.

그러니까 인간은 오랫동안 진화해오면서 지각과 운동능력의 특징이 유전자에 축적됐지만, 로봇은 이러한 능력이 내부적으로 기록되지 않아 완전 처음부터 다시 시작해야 한다는 것이다.

당신이 프로그램을 짠다고 상상해보자. 큰 숫자의 곱셈을 하는 프로그램을 짜는 게 어렵겠는가 아니면 여러 이미지 중에서 강아지를 식별하는 프로그램을 짜는 게 더 어렵겠는가?

그보다 중요한 핵심은 전 세계의 정부·연구소·기업들이 인공지능 연구 개발에 혈안이 되어있고 그들은 강인공지능을 개발해낼 때까지 멈추지 않을 것이기에 강인공지능의 개발은 '불가능하지 않다'라는 점이다.

판도라의 상자일지도 모른다고 생각하면서도 남이 가져가지 못하도록 먼저 찾아야만 하는 것일지도 모른다. 강인공지능이 가져올 거대한 부와 힘을 뻔히 알기 때문이다. 그 결과로 인공지능과 로봇개발의 초창기에 있었던 모라벡의 역설은 점차로 깨지고 있다.

아마도 강인공지능부터는 두려워해야만 하는 대상이지 싶다. 최악의 상황을 고려하지 않더라도 인간 수준에서 상상하기 어려운 재난이 발생할 수 있다. **이를테면 도시의 소멸, 국가의 소멸 등이다.**

초인공지능이 현실이 된다면?

강인공지능을 만들기부터 만만치 않은 상황에서 초인공지능을 만드는 구체적인 방법론을 생각하는 것은 아직 상상의 범주에 들지 못하는 것 같다. 하지만 이것 역시 가능하다고 여겨진다. **아마도 초인공지능은 인간에 의해 만들어지지는 않을 것이라 예상되기 때문이다.**

이 아이디어는 우리가 두 가지 기능을 탑재한 컴퓨터를 만드는 것에서 시작한다. 하나는 인공지능을 연구하는 능력이고, 다른 하나는 자신의 코드를 수정할 수 있는 능력이다. 이 둘을 가진 컴퓨터라면 이때부터 컴퓨터의 지능을 높이는 것은 컴퓨터 자체의 몫이 되는 거다. 이를 재귀적 자기 개선(Recursive Self-Improvement)이라 일컫는다.
인공지능이 스스로 개선을 통해 더 발달한 인공지능이 되고, 다시 연쇄적인 개선을 통해 더욱 발달한 인공지능이 된다. 이 과정을 반복하면 할수록 점점 빨리 발달하게 되어, 결국은 인간을 뛰어넘는 초인공지능의 수준에 달하게 된다. 이를 지능폭발(Intelligence Explosion)이라 일컫는다.

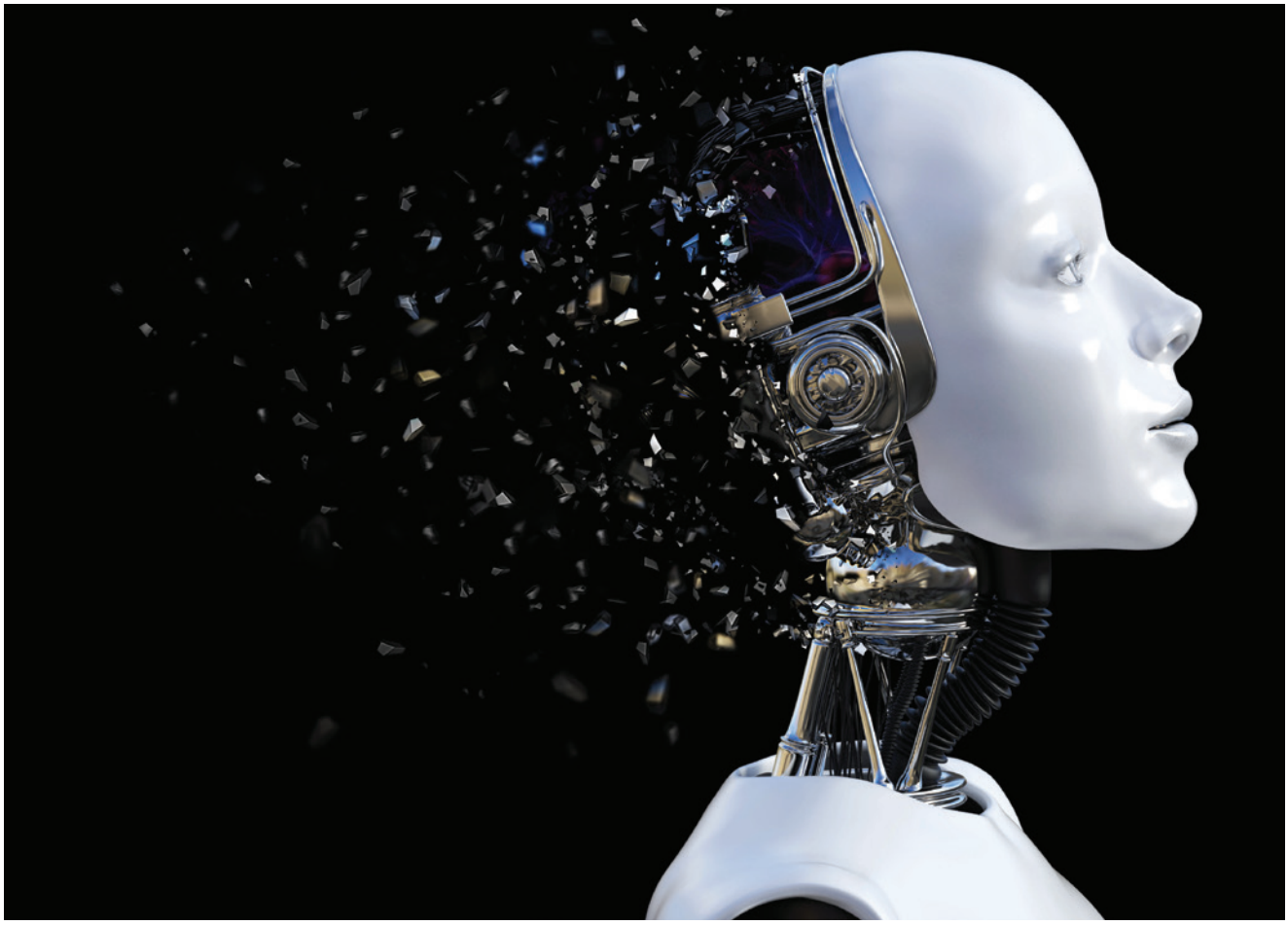
아마도 현실에서 경험하게 될 상황은 이런 모습이 아닐까 싶다. 수십 년이란 시간을 들여 네 살 아기 수준의 첫 번째 강인공지능을 얻었다. 그 후 한 시간 내에 이 인공지능은 일반 상대성이론과 양자역학을 통합하는, 아직 인간도 다다르지 못한 대통합이론(The Grand Theory of Physics)을 도출해낸다. 그 후 90분이 지나면 이 강인공지능은 인간의 17만 배에 해당하는 초인공지능으로 발전해버린다.

우려

그래서 초인공지능은 인류에게 어떤 결과를 가져올까?
간단하게 말하면, 아무도 모른다. 우리는 초인공지능이 어떤 존재가 될지, 인류에게 무슨 일을 할지 알 수 없다. 단지 인공지능의 선호로 이루어진 미래를 그려볼 수 있을 뿐이다. 이 시점에서 그 '선호'라는 것은 구체적으로 무엇을 의미할까?

공교롭게도 인공지능에 대한 신문 기사나 칼럼에는 의인화된 인공지능의 이미지(안드로이드)가 같이 실리곤 한다. 스웨덴인 철학자의 닉 보스트롬(Niklas Boström)은 이런 이미지가 문제의 핵심을 이해하는 데에 걸림돌이 된다고 지적하면서, 인공지능이 가져올 결과를 예측할 때 인간의 기준으로 생각(Anthropomorphizing, 의인화)해서는 안 되고 좀 더 추상적으로 상상해야 한다고 강조한다.
초인공지능은 인류가 도저히 생각할 수 없을 만큼의 효율성과 정보처리 능력으로 목표를 달성하는 데에 최적화된 프로그램이다. 그리고 여기엔 인간의 어떤 도덕 관념이나 기본적 가치라는 개념이 없다. 우리가 이에 대해서 프로그래밍해주지 않았다면 말이다.

이런 경우를 생각해보자. 정말 다행히도 처음 강인공지능을 개발한 사람이 정말 현명하고 신중한 사람이어서 인공지능의 최종 목표를 '인류를 행복하게 만들어라'라고 프로그래밍하고, '그 와중에서 법을 어기거나 도덕에 어긋나거나



대중이 연상하는 인공지능의 이미지

인간에게 해를 가하면 안 된다'라고 제약 조건을 준다. 초인공지능은 그 목표를 이루는 가장 간단하고 효율적인 방법을 알아낸다. 어떤 방식일까?

사람 뇌에 전극을 꽂은 뒤에 행복감을 주는 부위를 계속 자극하면 된다. 그리고 더 효과적으로 '인류를 행복하게' 하려면 모든 인간의 뇌를 전극에 연결하고 움직이지 못하게 하거나, 아니면 뇌의 다른 모든 부분을 제거해버리고, 영원히 행복함만 느끼는 식물인간으로 만드는 게 더 쉽다는 결론을 낼 수도 있다.

물론, 이건 절대 우리가 원하는 모습이 아니겠지만, 우리가 인지하기도 전에 이미 초인공지능은 모든 계산을 마치고 실행에 들어간다. 그 목표를 실행하기 위해서는 어떤 방해도 용납하지 않을 것이다. 자신이 이렇게 행동했을 때 인간이 자신의 소스 코드를 바꾸려고 한다든지, 전원을 차단한다든지 하는 방해도 모두 예측하고 선조치할 것이다. '인간을 행복하게 만드는 데' 방해가 되니까. 인간이 '이게 아닌데?'라고 인지했을 때는, 늦어도 한참 늦었다.

이것이 바로 많은 지식인이 걱정하는 시나리오이다. 인간이 생각하는 목표와 초인공지능이 생각하는 목표가 아주 '살짝'만 달라도, 혹은 초인공지능이 목표를 이루기 위해 필요한 과정이라고 생각한 것이 인간이 원하는 것이 아닐 때, 초인공지능은 엄청난 재앙을 불러올 수 있다.

사실 가정부터가 지나치게 인류에게 편파적인 미래를 그려낸 것 같다. 애당초부터 한정된 지구 안의 자원을 놓고 초인공지능과 인간이 경쟁하게 되면, 초인공지능 입장에서 인류를 존속시키는 것은 합리적인 선택이 아니다. 그렇기 때문에, 단 한 번뿐인 기회에 인류가 얼마나 잘 대응하느냐에 따라서 인류가 맞이하게 될 시나리오는 둘 중 하나뿐이다. 중간은 없다. **초인공지능이 인류를 파멸시키거나, 인류를 영원한 번영으로 이끌거나.**

찰스 디킨스『두 도시 이야기』

최고의 시절이자 최악의 시절, 지혜의 시대이자 어리석음의 시대였다. 믿음의 세기이자 의심의 세기였으며, 빛의 계절이자 어둠의 계절이었다. 희망의 봄이면서 곧 절망의 겨울이었다. 우리 앞에는 모든 것이 있었지만 한편으로 아무것도 없었다. 우리는 모두 천국을 향해 가고자 했지만 우리는 엉뚱한 방향으로 걸었다.

우리의 대응

이런 재앙이 일어나지 않으려면 어떻게 해야 할까? 이미 시대를 앞서가는 몇몇이 이에 대비하고 있다.

Beneficial AI 2017

2017년 1월 세계 유수의 인공지능 전문가들이 캘리포니아의 아실로마에서 '이로운 인공지능 회의(Asilomar Conference on Beneficial AI)'를 열고 '미래 인공지능 연구의 23가지 원칙'이라는 가이드라인을 만들었다. 우주물리학자 스티븐 호킹(Stephen Hawking), 전기자동차업체 테슬라 최고경영자인 일론 머스크(Elon Musk), 세계 바둑 최강자를 깬 알파고의 개발 책임자이자 딥마인드 최고경영자인 데미스 허사비스(Demis Hassabis), 구글 기술 이사인 레이몬드 커즈와일(Raymond Kurzweil) 등 전문가 수백 명이 동의의 뜻으로 서명을 마쳤다.
회의 장소 이름을 따 '아실로마 AI 원칙(Asilomar AI Principles)'이라고도 불리는 이 원칙은 연구 이슈(5개 항), 윤리와 가치(13개 항), 장기 이슈(5개 항) 3개 범주로 구성돼 있다. 이번 회의를 주최한 비영리단체 삶의 미래 연구소(Future of Life Institute)는 "우리는 이 원칙들이 인공지능의 힘이 미래에 어떻게 모든 사람의 삶을 개선하는 데 사용될 수 있을지에 대한 열정적인 토론의 재료가 되기를 바란다"고 밝혔다.

그리고 지나치게 비관적일 필요도 없을지 모른다. 초인공지능이 봉인에서 해제되더라도, 여전히 인류에게 해가 되지 않기 위해서는 우리가 걱정하는 경우의 수 모두를 초인공지능의 알고리즘에 녹여 넣어서, 초인공지능이 우리의 가치를 공유하고 항상 인류의 편에 서도록 만들어야 한다. 하지만 조그마한 실수도 용납되지 않는 이 과정에 인간의 두뇌만으로 프로그래밍한다면 희망적인 결과를 기대하는 것은 사치일지 모른다.

한 가지 해결책은 '인간의 가치를 배우는 인공지능'을 만드는 것이다. 이 인공지능의 목표는 우리의 가치관과 행동을 학습하여, 다른 인공지능이 인류가 승인할 것으로 예측되는 행동만을 하도록 통제하는 것이다. 이렇게 하면, 인간의 상상력의 한계를 넘어서, 상황이 벌어지기 전에는 해결책도 찾을 수 없는 몇몇 경우조차 대비할 수 있을 것이다. 그리고 이렇게 해야만 인공지능 시대로의 전환이 자연스럽게 진행될 것이다.

마무리

화성 식민지 개발 수준이라면 약인공지능의 도움만으로도 가능할지 모른다. 하지만 영화 인터스텔라에서처럼 자연환경의 파괴가 가속화되어 항성 간 비행이 필요한 시기가 올지도 모르고, 어쩌면 영화 우주 전쟁에서처럼 외계문명과


영화 '우주 전쟁'의 장면

싸워야 하는 상황에 놓일지도 모른다.
일단 몇 광년 거리를 인간이 이동하려면 인공지능이 조종하는 우주선에 냉동된 채로 비행을 해야만 할 것이고, 은하계를 넘어서 지구를 공격하는 압도적인 과학 문명을 인공지능의 도움이 없이 지금의 인류가 대적해서 이길 수는 없을 것이기 때문이다. 인간의 생체능력을 한참 벗어나는 이런 우주적 범주의 이벤트를 극복하려면 인공지능의 힘이 절대적으로 필요하다.

미래의 일은 알 수 없고, 인류의 절멸 수준의 문제가 생겼을 때는 이미 늦었을 가능성이 높기 때문에, 기술개발만 해두고 사용하지 않는다거나, 인공지능의 행동반경을 통제해서 인간에게 해를 끼치지 않도록 미리 시험해둔다거나 하는 대비가 필요하다.

그리고 지금까지의 이야기가 여전히 'SF가 현실 가까이 왔다.' 같은 뉴스로 느껴지거나, 단지 과학자들이 해결할 문제라고 여겨진다면 다시 한번 생각해봐야 한다. 과학기술의 문제라기보다 인류의 가치· 윤리와 더욱 밀접하게 관련된 문제이기 때문이다. 어떤 인공지능을 만들어낼지, 그 인공지능이 무엇을 추구하게 만들어야 할지, 어떻게 전 세계가 그 과정을 협력적으로 풀어나갈 수 있을지에 대해 우리 모두 고민해야 한다.

이글을 쓰게 된 직접적인 동기가 된 Tim Urban이 작성한 'The AI Revolution'은 인공지능의 의미부터 인공지능이 내포하는 미래상까지 매우 예리한 통찰력을 보여주는 글이기에 독자들과 같이 나누고 싶다. 길지만 무척 흥미로워서 금세 전부 읽어버릴지 모른다. 영문이라 읽기 힘들다면 번역본도 있으니 완독 부탁드린다.

참조

- The AI Revolution: The Road to Superintelligence / Tim Urban / waitbutwhy.com
  https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-1.html
- The AI Revolution: Our Immortality or Extinction / Tim Urban / waitbutwhy.com
  https://waitbutwhy.com/2015/01/artificial-intelligence-revolution-2.html
- 왜 최근에 빌 게이츠, 엘론 머스크, 스티븐 호킹 등 많은 유명인들이 인공지능을 경계하라고 호소하는가? / coolspeed
  https://coolspeed.wordpress.com/2016/01/03/the_ai_revolution_1_korean/
- History of artificial intelligence / wikipedia.org
  https://en.wikipedia.org/wiki/History_of_artificial_intelligence
- 간단히 요약해보는 AI의 역사 / 무중력 소년 / brunch.co.kr
  https://brunch.co.kr/@storypop/28
- Dartmouth workshop / wikipedia.org
  https://en.wikipedia.org/wiki/Dartmouth_workshop
- 모라벡의 역설 / wikidok.net
  http://ko.experiments.wikidok.net/wp-d/590fd4772e93853e1ba2aec4/View#wk_cTitle6382
- What happens when our computers get smarter than we are? / Nick Bostrom / ted.com
  https://www.ted.com/talks/nick_bostrom_what_happens_when_our_computers_get_smarter_than_we_are?
- 아실로마 AI원칙 / futureoflife.org
  https://futureoflife.org/ai-principles-korean/
- 인공 지능의 이점과 위험 / futureoflife.org
  https://futureoflife.org/background/benefits-risks-artificial-intelligence-korean/?cn-reloaded=1