5주차 과제

정명교 (글로벌바이오메디컬공학과 2021311097)

1. **튜링테스트란 무엇인가**

영국의 전산학자 앨런 튜링이 1950년 출판한 논문 <계산 기계와 지성>애서 튜링 테스트라는 게임을 제안했다. 기계가 스스로 사고할 수 있음을 증명하기 위해 튜링 테스트를 언급했는데 내용은 다음과 같다. 질의자 한 명과 응답자 둘이 있는데 응답자 중 하나는 [컴퓨터](https://namu.wiki/w/%EC%BB%B4%ED%93%A8%ED%84%B0)이고 나머지는 [인간](https://namu.wiki/w/%EC%9D%B8%EA%B0%84)이다. 질의자(심판)는 어느 쪽이 컴퓨터인지는 모르는 상태에서 키보드만을 이용한 응답을 통해 어느 쪽이 컴퓨터인지 판별할 수 있어야 한다. 이때, 심판이 컴퓨터의 반응과 인간의 반응을 구별할 수 없다면 해당 컴퓨터가 사고할 수 있는 것으로 보아야 한다고 튜링은 주장했다.

튜링 테스트를 통과했다는 것은 인공지능과 인간의 지능을 구별할 수 없음을 의미하는데, 이는 곧 컴퓨터가 사람처럼 생각할 수 있다는 결론으로 도출된다. 철학계에서는 튜링 테스트를 통과한 AI를 인간으로 취급할 것인지에 많은 논쟁이 있다. 튜링 테스트를 옹호하는 학자들은 인간과 기계를 구별할 수 없는 상황을 근거로 기계가 인간과 동등하다 혹은 인간의 지능을 가지고 있다고 주장하는 한편, 일각에서는 튜링 테스트는 특수한 상황에서만 적용되는 특별한 케이스로 범용성이 부족하고 이것만으로는 인공지능이 인간과 동일한 인지 과정과 사고 과정을 거치는 존재임을 증명할 수 없기에 인격을 부여할 수 없다는 주장을 제기한다. 아직까지도 튜링 테스트의 실효성에 대한 논란은 끊임없이 제기되고 있지만 그의 아이디어는 인공지능의 개념에 대한 중요한 기반을 제공했으며 인공지능을 판별하는 기준이 되었음은 확실하다고 할 수 있다.

인간은 생물로서의 직관과 경험을 가지고 있고 인공지능은 그것들을 데이터를 통한 훈련을 통해 습득한다. 앞으로 인공지능이 더욱 발전할 것임은 분명한데, 특이점을 뛰어 넘는 시대에 도달한다면, 그때는 진정으로 인간만이 가지는 고유한 가치에 대해 생각해볼 필요가 있을 것이다. 인공지능이 인간 특유의 직관에 접근할 수 있다면 인간과 인공지능의 차이점이 없어지는 것이 아닌지에 대한 생각도 든다.

1. **물체, 기계, AI, 동물, 인간의 존재방식은 어떻게 구별되는가**

인공지능의 등장으로 인하여 물체, 기계, AI, 동물이 인간과 어떤 면에서 다르지, 그들을 인간과 구별하는 기준에 대한 혼선이 빚어지고 있다. 존재방식에 대한 이해가 선행되어야 인간과 AI의 관계를 정의하는데 도움이 될 수 있기에 존재방식을 파악하고 그것들을 구별할 필요가 있다. 특히 인간에 기초하여 인공지능을 해석해야 하는지 혹은 인공지능에게 있어 인간은 어떠한 존재인가와 같은 상호간의 관계성을 규정하는 것이 큰 이슈로 대두되고 있기 때문에 그 중요성은 더욱 크다고 할 수 있다.

현재 ‘물리주의’라는 입장이 대세를 이루고 있는데, 이는 곧 만물이 궁극적으로 ‘물질’이라는 것으로 구성되어 있기에 이 세계에 존재하는 모든 것들을 물질이 갖는 성질을 이용해서 설명하려는 것이다. 물리주의에 따르면, 물체나 생물과 같은 모든 것은 물리적으로 구성되며 그렇기에 모든 현상을 물리적으로 설명할 수 있다고 본다. 모든 것을 통합하여 물리적으로 설명하는 것이 확실히 주류이긴 하나 이는 절대적인 진리로서는 작동할 수 없다. 물질만으로 기능을 구분하기에는 한계점이 분명히 존재하기 때문이다.

그렇기에 가장 큰 요점은 움직임에서 찾을 수 있다. 물체, 기계, 동물, 인간은 모두 움직인다. 다만 움직임의 원인과 배경 면에서 차이가 있다. 강의에서 언급된 바위와 도마뱀을 생각해보면 그 이유를 쉽게 알 수 있다. 물체, 즉 바위는 결코 스스로 운동하지 않는다. 운동을 당한다는 표현이 적합할 듯싶은데, 중력과 같은 물리 법칙의 영향으로 움직임이 발생하기 때문이다. 철학적으로는 타성 동이라고도 부르는데, 움직임의 원인을 내부가 아닌 외부에서 찾는다. 따라서 바위와 도마뱀이 같은 위치를 차지할 때, 바위는 운동의 원인이 외부에 있다. 하지만 도마뱀은 스스로 움직인다. 동물의 존재 목적은 생존에 있기에, 생존을 위한 움직임을 보인다. 예를 들어, 절벽에서 생활하는 산양은 낙사확률이 평균적으로 40%나 된다. 인간의 관점에서 보면 도저히 이해할 수 없지만 산양에게는 위험이라는 개념 자체가 없다고 볼 수 있다. 그들에게는 절벽이 존재 방식에 최적화된 환경이고 그곳이 삶의 터전인 곳으로 존재에 대한 충동이 발현되고 그것을 충족하는 공간이 바로 절벽인 것이다. 따라서 비록 도마뱀과 바위가 같은 위치에 있다고 할지라도, 그것들은 각기 다른 원인으로 치를 설명할 수 있기에 존재 방식이 서로 다르다고 볼 수 있다.

그렇다면 인간과 동물은 어떻게 구별할 수 있을까? 동물에게는 생존이 1순위이지만 인간은 아니다. 인간에게 절벽은 결코 생존에 유리한 환경이 아니다. 인간은 절벽을 생존을 위해 올라가지 않는다. 풍경의 장엄함을 몸으로 체험하기 위해서 절벽을 오를 뿐이다. 즉, 산양처럼 생존 환경이 아니라 다른 의미를 가지고 절벽에 오른다는 것입니다. 여기서 인간의 존재 방식을 알 수 있는데 인간은 생존을 위한 삶이 아닌, 실존하는 존재로서 그 가치를 지닌다고 볼 수 있다.

따라서 물체와 동물, 그리고 인간은 움직임을 바라는 관점의 차이가 수동적으로 발생하는 것인지, 주체적으로 행하는 것인지, 그리고 왜 하는 것인지에서 드러난다. 기계도 움직임이 있기는 하지만 보통 운동이라고 부르지는 않는다. 설계에 따라 공장에서 일어나는 부품들의 체계적인 작동이 더 맞는 표현이다. 이러한 관점에서 AI를 살펴보면, 인간과 가까우면서도 기계에 가까운, 둘 사이의 경계점에 놓여있는 애매한 상황이다. 그렇기에 더더욱 물체, 기계, 동물, 인간의 존재방식에 대한 명확한 이해와 구분이 필요하다. 관계를 구분짓지 못한다면 AI를 어떻게 디자인하고 설계할지를 결정하기 어려운 문제로 이어진다. 따라서 위와 같이 존재 방식에 대한 정확한 이해를 바탕으로 그들을 구별하고 그것들을 바탕으로 AI를 개발하도록 해야 한다.

1. **미래 인간과 AI의 관계는 어떻게 설계되어야 하는가**

AI 발전 속도는 엄청나다. 단순히 인간을 묘사하는 것이 아닌 모든 영역에서 인간을 초월하여 인간과 관련된 모든 곳에 탑재될 수 있는 것, 즉 특이점을 넘어서는 것을 개발 목표로 삼고 있다. 전문가들은 2045년이면 달성될 것으로 보고 있는데 우리와 삶과 AI가 밀접하게 연관 되어있는 미래가 그리 멀지 않음을 의미한다. 그렇기 때문에 인간과 AI의 관계를 어떻게 설계해서 그들을 개발해 나갈 것인가가 큰 화두이다.

크게 보면, 인간 미래에는 3가지 비전이 있는데 첫번째는 인간을 AI의 아바타로 보는 관점이다. 몇 년 전 화재가 되었던 알파고와 이세돌 9단과의 대전에서 이세돌의 맞은편에는 그저 알파고의 명령에 따라 바둑을 두는 사람이 있었다. 이 사람은 스스로 사유하여 바둑을 하는 것이 아니라 그저 알파고가 내린 결정에 따라 바둑알을 옮기는 역할만을 한다. 즉, 사고과정을 거쳐 주체적으로 의사 결정을 하는 인간으로서의 본질이 사라지고 단순히 AI의 결정에 맞게 행동하는 아바타에 불과하게 되는 것이다.

두번째는 죽지 않는 인간 이후의 존재자(Immortal Posthuman)가 되는 것이다. 1번의 대안 으로 제시된 것인데, AI 아래에 있는 존재가 아니라 AI를 이용해 더 나은 존재로 발전하고자 하는 관점이다. 인간의 신체와 지적 능력에는 분명한 한계가 존재한다. 정보를 저장하고 처리하는데 제한이 있고 죽음을 피할 수 없는 존재가 인간이다. 이러한 인간의 한계를 극복하고자 AI를 이용하자는 것이다. 이들은 미래에 인간이 경쟁력을 가지고자 한다면 인공 두뇌와 인공 피부 등 첨단 기술을 이용한 신체 개조가 이루어져야 한다고 하는데 일종의 증강 인류를 목표로 삼는다. 학자들은 모든 기술이 융합되어 급격한 기술적 비약이 일어나는 특이점이 온다면 인간 성능 개조가 가능해진다고 본다. 구글의 프로젝트 총괄자인 RAY KURZWEIL은 죽음은 정복 가능하며 자아의식을 복제해 빅데이터로 기록하는 Mind upload 기술이 가능해질 것이라고 말했다. 그는 죽음을 인간이 가진 최고의 결함으로 본 것이다.

마지막으로는 죽음을 인정하면서 본래적 자신을 향해 사는 존재이다. RAY KURZWEIL과 어깨를 나란히 한 엔지니어이자 CEO인 스티브 잡스는 정반대의 주장을 가졌다. 그는 죽음은 운명이며 죽음을 직면해야 남들과는 다른 주체적 존재로서의 자신을 찾을 수 있다고 말했다.죽음을 삶이 만든 최고의 발명이라 칭하며 이를 받아들임으로써 혁신이 가능하다는 주장을 펼쳤는데 이 관점은 죽음은 우리 모두의 숙명으로 받아들이고 나를 대신해서 죽을 자도 없으니 죽음을 통해 자기 자신으로 거듭나는 것이다. AI를 통해 영생을 꿈꾸는 것이 아니라, 인간의 한계성을 인정하고 되려 인간의 의미를 찾자는 것이다.

죽음을 어떻게 여기느냐에 따라 그것을 극복하고자 AI를 개발할 수도 있고, 죽음을 인정함으로써 본래의 자신을 되찾고자 활용할 수도 있다. 이는 AI를 어떻게 활용할 것인가라는 논점과도 연결된다. Techno Futurism는 과학기술의 발전을 인간이 이용함으로써 인간 존재를 질적으로 혁신할 수 있다는 과학 기술 결정론적 입장이다. 앞서 언급한 RAY KURZWEIL의 견해와 결을 같이 하는 입장으로 AI와 같은 첨단과학기술을 통해 죽음이나 정보 처리와 같은 한계를 극복함으로써 인간 성능을 증강하고 영생하는 존재로서의 post-human을 목표로 한다. Humane Futurism은 이와 달리 인간 존재의 의미를 다시 확립하고자 한다. 인간의 실존적-존재론적 허무화에 직면하여 인간 존재 의미를 재확보하려는 사상적 움직임이다. 이 관점에서는 죽음을 받아들임으로써 삶의 유한성에서 기원한 인간의 참된 가치를 되찾고 적절한 시간을 누려야 좋은 상태에 도달한다는 숙성의 미학을 강조한다.

현대인은 발전하는 AI 기술에 직면하여 몇 가지 문제점을 안고 있다. 인간의 위에 있는 존재로 개발을 할지, 인간의 한계를 극복하도록 돕는 도구로써 볼지, 아니면 인간이 지닌 한계를 받아들이고 삶의 가치를 찾는 것에 집중을 할지 등 결코 쉽지 않은 문제상황에 직면했다. 따라서 우리는 각자 자신이 생각하고 있는 AI의 미래 비전을 생각하여 AI와 인간의 관계에 대해 지속적으로 고민해보아야 한다.