



# 인공지능은 자아를 가질 수 있을까요?

「철학이란 무엇인가?」와 「인지심리학 및 실험」 강의페어링

심리학과 이준호 201323245, 정재영 교수님 지도

## 목적

알파고와 이세돌의 대국으로 많은 사람들이 인공지능에 대해 관심을 갖게 되었다. 그러나 아직은 SF 소설이나 영화같은 인공지능은 나타나기 어렵다. 현실과 상상 속 인공지능의 차이는 바로 '자아'다. 자아가 없는 인공지능을 약인공지능(weak A.I), 있는 인공지능을 강인공지능(strong A.I)이라 하는데, 현재까지 개발된 인공지능은 전자, 상상 속의 인공지능은 후자에 해당한다. 강인공지능에는 자아가 필요하고, 이 때문에 많은 학문들이 이 부분에 대해 연구하고 있다. 본 연구에서는 자아의 선행조건인 '의식'에 대해 서로 다른 학문들이 어떻게 이야기하고 있는지 알아보면서, 과연 인공지능이 자아를 가질 수 있는지 생각해본다.

## 철학: 의식이란?

- **데카르트의 인식론**  
'나'의 존재는 신의 뜻이라는 중세와 달리 근대 철학자 데카르트는 '나'의 존재 이유를 '내가 생각하기 때문'이라고 했다.
- **두 개의 코기토**  
원래 코기토(Cogito)명제는 중세 철학자 아우구스틴누스가 신의 존재를 증명하기 위해 사용했다. 그러나 데카르트는 이를 통해 의심할 수 없는 확실한 지식을 찾고자 했다.
- **방법론적 회의**  
코기토 명제를 통해 의심할 수 없는 진리를 찾고자 한 데카르트는 '의심하고 있는 나'의 존재는 의심할 수 없다는 결론에 이른다. 그 과정에서 '생각하는 나'는 신에게서 독립적으로 존재할 수 있는 주체가 되었다.
- **사유와 연장**  
데카르트에게 어떤 실체의 본질은 공간을 차지하는 '연장'과 생각하는 성질인 '사유'로 이루어져 있다. 이 관점에서 인간의 육체는 연장이며, 정신은 사유에 해당한다. 코기토 명제에 의해 인간은 '생각함으로써' 존재하므로, 사유가 연장에 앞선다. 따라서 데카르트는 인간의 육체와 정신이 분리되어 있다는 심신 이원론을 주장하였다.
- **현대 철학의 역할**  
현대의 철학은 다양한 학문이 모인 인지과학에서 의식과 자아같은 반드시 연구가 필요한, 그러나 아직 학문적으로 정의되지 않은 개념들을 명확히 정의하고자 노력하고 있다.



르네 데카르트(1596~1650)

## 심리학: 의식을 의식하기

- **초기의 의식 연구**  
최초의 심리학자 분트는 '심리학은 의식을 연구하는 과학'이어야 한다고 말하며, 마음을 구성하는 원자를 찾고자 했다.
- **의식의 신경생리적 관점**  
발달된 신경과학을 통해 뇌의 활동을 보다 자세하게 관찰할 수 있게 되었고, 현대 심리학은 사고과정에서 의식이 개입되는 순간 뇌의 활동을 관찰하여 의식을 측정할 수 있게 되었다. 이러한 바탕에서 의식은 뇌로부터 나오므로, 뇌가 곧 정신이라는 일원론적인 결론에 이르게 된다.
- **현대 심리학에서의 의식**  
현대 심리학에서의 의식은 주의, 기억 등 다양한 곳에서 연구되고 있다. 주의에서의 의식은 의식적/무의식적 처리로 나타난다. 의식적 처리는 낯설고 복잡한 문제를 심사숙고하여 해결할 때, 무의식적 처리는 익숙하고 반복적이고 단순한 문제를 처리할 때 사용된다.
- **의식의 기능적 가치**  
위와 같은 의식의 작용은 인간의 제한된 주의를 효율적으로 배분할 수 있게 한다. 따라서 이러한 정보처리적 관점에서의 의식은 한정된 자원을 가진 인간이 복잡한 세상을 안전하게 살아갈 수 있게 해주는 기능적 가치를 지닌다.

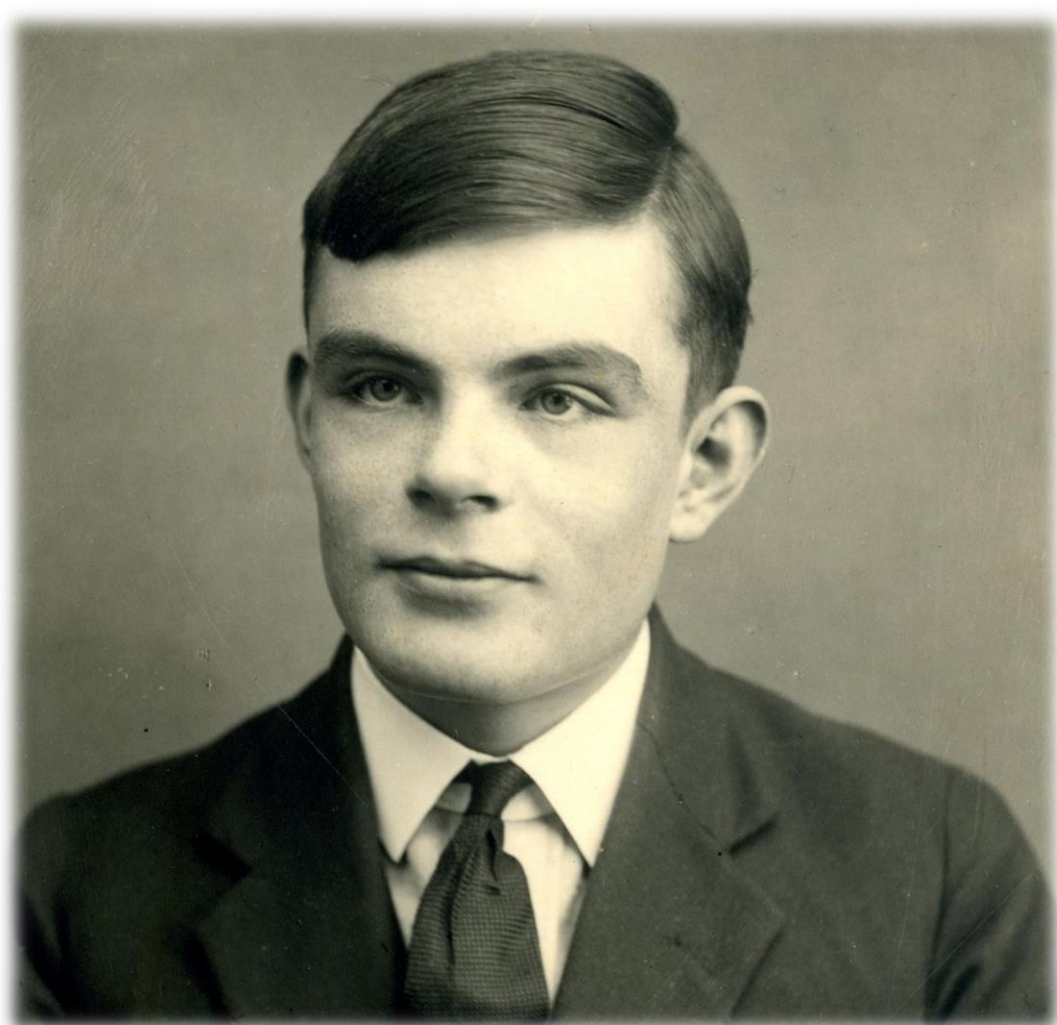


빌헬름 막시밀리안 분트 (1832~1920)

## 컴퓨터 과학: 의식을 구현하려는 사람들

- **커넥톰: 뇌의 설계도**  
뇌의 구조와 각 신경 간의 연결구조를 완벽하게 밝혀낸 것을 커넥톰Connectom이라 부른다. 인간의 커넥톰이 완성되면 정신의 생물학적 작용을 완전히 알 수 있게 되고, 이를 기계로 구현한다면 인간의 정신을 온전히 구현할 수 있다.
- **존 설의 중국어 방 논증: 기계는 의식을 가질 수 없다**  
중국어를 할 줄 모르는 사람이 방에 있고 그 안에는 중국어 문답이 적힌 카드가 있다. 이 때 밖에서 중국어로 적힌 질문을 보여주면, 안에 있는 사람은 중국어를 이해하지 못하지만 질문에 맞는 카드를 찾아 그에 맞게 답할 수 있다. 존 설은 겉으로 보서는 중국어로 대화하고 있지만, 정작 답을 하는 사람은 중국어를 이해하지 못하는 이 상황에 비유하여, 겉으로는 의식이 있는 것 처럼 보이는 기계는 실제로 의식을 가지고 있지 않을 수 있다고 주장하였다.
- **레이 커즈와일의 반박: 기계는 의식을 가질 수 있다**  
레이 커즈와일은 중국어방 논증에 대해 "중국어를 하는 것은 방(시스템)이지 안에 있는 사람이 아니다"라고 말했다. 인간의 신경세포가 의식을 갖는 것이 아니라 신경세포가 구성하는 인간이 의식을 갖듯, 컴퓨터도 그것이 구성하는 시스템이 의식을 갖게 될 것이며, 현재의 컴퓨터는 집적도와 성능이 낮지만 기술의 발전으로 극복할 수 있을 것이라 주장했다.

## 결론: 기계의 자아, 인간의 시대



앨런 튜링(1912~1954)

결국 인공지능은 자아를 갖게 될 것이다. 현재의 인공지능보다 강인공지능이 더 많은 일을 효율적으로 할 수 있기 때문이다. 인류는 늘 지적 욕구, 더 나은 세상, 경제적 이익 등으로 기술을 발전시켜왔고, 현대 기술의 정점인 인공지능 또한 그렇게 될 것이다. 인공지능의 의식은 데카르트의 의식처럼 스스로 생각하기에 존재 할 것이다. 인공지능의 의식은 프로그램 된 코드에 의해 구현되므로, 코드가 곧 의식이라는 일원론적인 결론을 내릴 수 있다. 반면 의식이 구현된 프로그램이 동작하는 하드웨어와 의식 그 자체인 소프트웨어는 별개이므로, 기계의 의식에 대한 이원론적 설명도 타당하다. 한편 기계에 의식이 필요한 이유는 그것이 더 효율적이기 때문이므로, 인공지능의 의식은 인간의 의식처럼 분명한 기능적 가치를 지닐 것이다. 그러나 그 의식은 인간의 의식과 동일한 기능을 가질 필요는 없다.

현재 인공지능은 인간을 모방해서 만들고 있다. 이는 지금까지 알려진 것 중 인간의 의식 시스템이 가장 지적이기 때문이다. 그러나 기계의 의식이 인간을 닮아야 할 이유는 없다. 일찍이 인공지능의 개념을 생각한 앨런 튜링은 "인공지능의 발명이란 자동차에서 바퀴를 떼어낸 뒤 그 자리에 발을 달기 위해 고심하는 것이다."라고 했다. 자동차는 달리기 위해 다리를 필요로 하지 않으며, 잠수함은 물 속을 다니기 위해 지느러미를 필요로 하지 않는다. 의식이 지적 존재에게 필요한 것이라 하더라도, 그것이 다르기 때문에, 기계에는 기계에 어울리는 의식이 있다. 그러나 우리는 아직 의식 없는 기계가 무엇을 필요로 하는지 알 수 없다. 기계가 의식을 가지고 스스로를 재생산하기 시작할 때, 그때부터 기계는 스스로에게 어울리는 자아를 만들어갈 것이다. 그리고 그 날부터 진정한 기계의 시대가 시작될 것이다.

## 참고문헌

『철학과 굴뚝 청소부』(이진경, 2005), 『마이어스의 심리학』(David G. Myers, 2008), 『인지심리학』(이정모 외 16명, 2009), 『심리학사(사상과 맥락)』(D. Brett King, Wayne Viney, William Douglas, 2009)