DSU Dongseo University 동서대학교

인공지능의 미래와 윤리

박 수 현

학습 목표

인공지능을 수준별로 살펴보고 지능의 폭발과 윤리

- 약한 인공지능과 강한 인공지능
- 미래의 인공지능의 기술적 특이점과 인간 종속
- 슈퍼 인공지능이 창의성에도 도전할 수 있는지 그 가능성을 예상해본다.
- 인공지능 시대에 요구되는 윤리성과 윤리 강령

4.1 인공지능의 수준별 분류

- 인공지능의 수준에 따른 분류
  - 인공지능 관련 이론과 기술 수준은 지속적으로 발전 중
  - 초기의 인공지능은 인간의 두뇌를 흉내 내는 정도의 수준
  - 인간의 지능과 같거나 그 이상의 능력을 목표로 추구
  - 자아의식까지 가진 강력한 인공지능을 꿈꾸고 있음
  - 인공지능은 수준에 따라 2가지 방법으로 분류
    - 약한 인공지능과 강한 인공지능
    - 좁은 인공지능, 일반 인공지능, 슈퍼 인공지능

4.1 인공지능의 수준별 분류

- 약한 인공지능과 강한 인공지능
  - 1980년 미국의 존 설(Searle) 교수가 중국어 방 논증을 제안하면서 최초로 사용한 개념
  - **약한 인공지능**: 유용한 도구로써 설계된 인공지능
  - **강한 인공지능**: 인간을 완벽하게 모방한 인공지능

[그림 4.1] 약한 인공지능과 강한 인공지능

4.1 인공지능의 수준별 분류

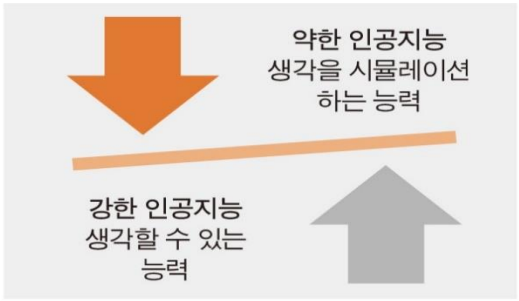
- 약한 인공지능 (weak AI)
  - 특정 분야 내에서 인간의 지능을 흉내 내는 지능적인 활동
  - "인공지능이란 단어를 검색하여 결과를 보여라." 등 업무
  - 고양이 그림을 찾거나 문서를 다른 언어로 번역하는 일
  - 현재 수준의 인공지능 기술 대부분은 약한 인공지능에 속함

[그림 4.2] 약한 인공지능에 속하는 언어 번역

4.1 인공지능의 수준별 분류

- 약한 인공지능 (weak AI)
  - 특정 분야 내에서 인간의 지능을 흉내 내는 지능적인 활동 즉, 인간 능력의 일부를 대체하는 수준
  - 인공지능이란 단어를 검색하여 결과를 보여라, 고양이 그림을 찾거나 문서를 다른 언어로 번역하는 일 등
  - 알파고는 바둑에서 이길 확률로만 계산, 다음 수 결정, 바둑 외 능력은 없고 특정 영역에 국한된 약한 인공지능
  - 현재 수준의 인공지능 기술 대부분은 약한 인공지능에 속함

[그림 4.2] 약한 인공지능에 속하는 언어 번역



[그림 4.3] 약한 인공지능과 강한 인공지능

4.1 인공지능의 수준별 분류

- IBM 왓슨(Watson) 컴퓨터
  - 미국의 인기 퀴즈대회에서 인간 챔피언들을 이겼음
  - 의학 진단용으로도 훌륭한 성과
  - 최근 인공지능 기술의 자율주행 셔틀버스, 수집된 데이터를 활용하여 인공지능이 자율적으로 운행
  - 승객과의 일상 대화도 가능, 승객 목적지의 최적 경로로 스스로 운행
  - 다양한 활용 분야 측면에서는 강한 인공지능에 접근
  - 약한 인공지능에 속함

[그림 4.4] 퀴즈대회에서 우승한 IBM의 왓슨

[그림 4.5] 자율주행 셔틀버스

4.1 인공지능의 수준별 분류

- 강한 인공지능 (strong AI)
  - 인간 수준으로 다양한 일을 할 수 있는 인공지능
  - 인간과 같이 생각하고, 판단하며, 상황을 이해
  - SF 영화에서의 미래지향적 인공지능 수준
  - 인공지능 자체가 대부분 법적 책임을 짐
  - 인간의 의식 수준으로 생각하는 힘과 감정도 가짐

[그림 4.6] 강한 인공지능 수준의 영화 〈어벤저스〉

4.1 인공지능의 수준별 분류

- 약한 인공지능과 강한 인공지능 비교

〈표 4.1〉 약한 인공지능과 강한 인공지능과의 차이

약한 인공지능	강한 인공지능
특정 분야에서만 활용 가능	다양한 분야에서 활용 가능
인간의 지능을 흉내 내는 수준	인간과 유사 또는 뛰어넘는 지능 수준
인간 두뇌의 제한된 일부 기능	인간 두뇌의 일반 지능
현재의 인공지능 수준	미래지향적 인공지능 수준
제작자나 소유자가 책임	인공지능 자체가 대부분 책임
지능적인 것같이 행동	실제로 지능적인 행동
시리, 알파고, 전문가 시스템 등	공상 소설이나 SF 영화에 등장
느낌이나 감정이 없음	자아의식과 감정도 가짐
특정 분야(바둑)에서 인간 능가	아직도 요원하며 예측 어려움

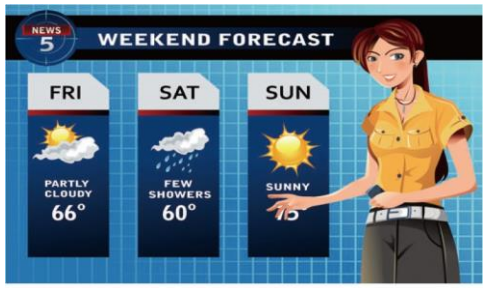
4.1 인공지능의 수준별 분류

- 좁은 인공지능, 좁은 인공지능, 슈퍼 인공지능
  - 좁은 인공지능은 한 가지 업무에 특화된 인공지능
  - 일반 인공지능은 인간 수준의 인공지능
  - 슈퍼 인공지능은 인간의 지능보다 뛰어난 인공지능

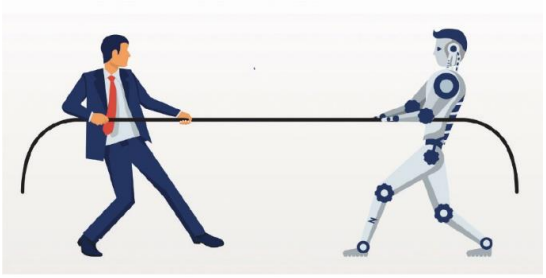
[그림 4.7] 인공지능의 3가지 발전 단계

4.1 인공지능의 수준별 분류

- 좁은 인공지능 (Narrow AI)
  - 한 가지 또는 특정한 영역에 국한된 인공지능
  - 체스, 바둑, 또는 일기예보 등 특정 분야에 국한된 인공지능
  - 체스의 딥 블루, 퀴즈의 왓슨, 얼굴인식, 알파고, 자율 자동차, 애플의 시리 등
- 일반 인공지능 (General AI)
  - 인간 수준의 능력을 가진 인공지능
  - 생각하는 능력, 사회적인 능력, 창의적인 능력도 가능
  - 인간의 학습 수준 또는 그 이상으로 학습 가능
  - 단순 응용의 수준을 넘어 일반화에 초점을 맞춤



[그림 4.8] 일기예보 인공지능



[그림 4.9] 인간 지능에 필적하는 일반 인공지능

4.1 인공지능의 수준별 분류

- 슈퍼 인공지능 (Super AI)
  - 모든 면에서 인간보다 훨씬 뛰어난 지능을 가진 인공지능
  - 과학적 창의력, 일반적인 지혜, 사회적 능력 등을 가짐
  - 어벤저스, 메트릭스 등의 SF 영화 속에서 존재하는 로봇 (복제인간)
  - 뛰어난 지능과 능력으로 인간을 지배하려 할 가능성(?)

[그림 4.10] 슈퍼 인공지능 수준의 영화 〈외계인〉

[그림 4.11] 슈퍼 인공지능 스카이넷

4.1 인공지능의 수준별 분류

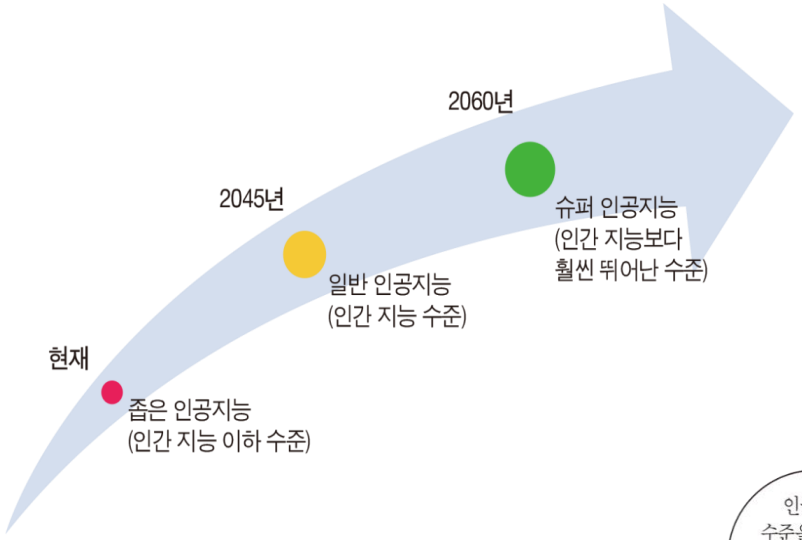
- 좁은 인공지능, 일반 인공지능, 슈퍼 인공지능 비교

〈표 4.2〉 좁은 인공지능, 일반 인공지능, 슈퍼 인공지능

	좁은 인공지능	일반 인공지능	슈퍼 인공지능
다른 이름	전용 인공지능	범용 인공지능	초인공지능
주요 특징	한 가지 또는 특정한 영역에 국한된 인공지능	인간 두뇌와 대등한 수준의 인공지능	모든 면에서 인간보다 뛰어난 인공지능
구현 시기	현재	2045년 무렵	2060년 이후
응용 분야	체스, 바둑 등	다방면에 적용 가능	현재의 SF 영화 수준
지능 수준	인간 지능의 흉내 수준	인간과 유사한 지능 수준	인간을 뛰어넘는 수준
대응 분류	약한 인공지능에 대응	강한 인공지능에 대응	강한 인공지능에 대응

4.1 인공지능의 수준별 분류

- 인공지능 구현 시기



[그림 4.12] 3가지 형태의 인공지능의 구현 시기

4.2 약한 인공지능 (알파고)

- 알파고 개발
  - 361!(팩토리얼), 즉 10의 179승이란 엄청난 경우의 수
  - 하사비스(Hassabis)가 인공지능 바둑 프로그램을 개발
  - 2,900만 기보를 딥러닝 알고리즘으로 학습
  - 2014년 구글에 약 5천억 원에 인수됨
  - 2016년 KAIST '인공지능과 미래'란 하사비스 초청 강연
    - https://www.youtube.com/watch?v=lcZ1T9v22oc
    - https://www.youtube.com/watch?v=cqaLuDCyit0

[그림 4.14] KAIST에서 강연하는 하사비스

4.2 약한 인공지능 (알파고)

- 알파고와 이세돌 9단 대국
  - 2016년 3월 이세돌 9단과 알파고와의 세기적인 대국
  - 구글에서 개최, 알파고 대결에 세계적 관심 집중
  - 알파고가 이세돌 9단에 4대 1로 승리
  - 그 후 인공지능에 대한 관심 급증

[그림 4.15] 알파고와 대국하는 이세돌 9단

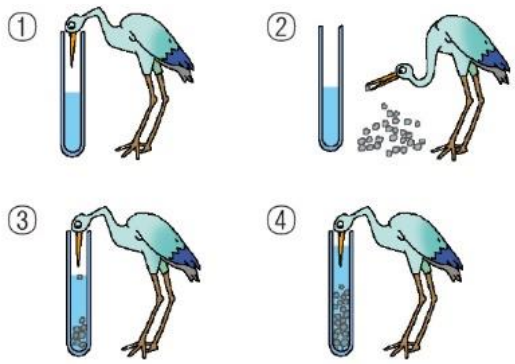
4.2 약한 인공지능 (알파고)

- 알파고 발전
  - '알파고 마스터(AlphaGo Master)'로 개량
  - 스스로 학습하는 '알파고 제로', '알파 제로'로 발전
  - 얼마 후 알파고는 바둑계에서 완전 은퇴 선언
  - 실시간 전략 게임인 '알파 스타' 개발
  - 세계 상위 0.2% 내의 '그랜드 마스터' 레벨에 오름

[그림 4.16] 스타크래프트 2

4.2 약한 인공지능 (알파고)

- 이솝 이야기의 두루미 물 마시기 장면
  - 두루미가 좁고 긴 유리관에 있는 물을 마시려 함
  - 물이 안 닿자 조약돌 몇 개 유리관에 넣기
  - 조약돌을 더 넣으니 수면이 좀 더 위로 올라옴
  - 조약돌을 더 넣고 나서 물을 마심



[그림 4.18] 두루미의 물 마시기 스토리텔링

4.2 약한 인공지능 (알파고)

- 강아지 구조하는 스토리텔링
  - 아이들이 강아지를 구하러 둑 밑으로 내려감
  - 책가방 끈에 의지하여 강아지를 물 밖으로 끌어냄
  - 책가방 끈에 의지하여 강아지와 둑 위로 올라옴
  - 강아지를 구출하고 난 후 그곳을 떠남

[그림 4.19] 스토리텔링의 장면들

- 보다 많은 그림이 순서 없이 나열된 경우에도 스토리텔링 가능?
- 가능하면 강한 인공지능의 영역에 다가선 수준으로 평가 가능

4.2 약한 인공지능 (알파고)

- 강한 인공지능 (중국어방 논증)
  - '강한 인공지능'은 인공지능이 인간 의식 수준 도달 의미
  - 인공지능이 '생각'을 가짐에 대한 철학적 문제로 파급
  - 1980년 중국어 방 논증(The Chinese Room Argument)
  - 미국의 언어 철학자 존 설(John Searle)이 고안한 사고 실험
  - 논증 사유 실험을 통해 튜링 테스트를 회의적으로 반박
  - 컴퓨터 프로그램의 지능적인 행동과 이해는 별개라고 주장

[그림 4.20] 존 설

4.2 약한 인공지능 (알파고)

- 강한 인공지능 (중국어방 논증 과정)
  - 중국어를 모르고 영어만 아는 사람이 방 안에 있다고 가정
  - 방에 필기도구와 중국어 표현의 영어 대응지침서 목록 비치
  - 외부에서 중국어 표현의 질문을 방안으로 넣음
  - 방 안 사람은 대응지침서에 따라 중국어로 답변을 씀
  - 바깥의 심사관에게 전달
  - 심사관은 방 안 사람이 중국어를 안다고 믿게 된다는 것

[그림 4.21] 중국어 방 논증의 상황도

4.2 약한 인공지능 (알파고)

- 강한 인공지능 (중국어방 논증 실험 결과)
  - 방 안 사람은 중국어를 모른 채 대응지침서에 따라 대답
  - 결론적으로 방 안 사람의 중국어 이해 여부 판정 불가
  - 튜링 테스트 통과가 지능에 대한 보장이 없다고 주장
  - 앨런 튜링의 튜링 테스트를 반박하는 논증
  - 언어와 마음에 대한 철학적 논쟁을 불러일으키는 계기

〈표 4.3〉 중국어 방 논증과 튜링 테스트 비교

	중국어 방 논증	튜링 테스트
주어진 상황	중국어를 모르는 사람이 답변	컴퓨터가 대화에 참가
실험의 대상	영어만 할 줄 아는 사람	지능을 가진 컴퓨터 프로그램
언어 이해도	문맥과 관련 없이도 대응	어느 정도 문장을 이해함
문법적 수준	구문론적	구문론적 + 의미론
지능의 정도	영어를 아는 사람	인공지능

4.2 약한 인공지능 (알파고)

- 인공지능에 대한 철학자들의 비판
  - 많은 철학자들이 인공지능 실현 가능성 강력히 부정
  - 철학자 루카스(John Lucas)는 인공지능 가능성 부정
  - 데넷(Dennett)은 마법의 영혼 없는 인간은 기계에 불과
  - 드레이퍼스(Dreyfus)는 초기 인공지능의 한계성 비판
  - 유치원생 수준의 동화를 이해하는 컴퓨터는 왜 없느냐고 지적
  - 와이젠바움은 인공지능의 오용이 인간의 삶 평가 절하 주장


[그림 4.22] 존 루카스와 대니얼 데넷

[그림 4.23] 허버트 드레이퍼스

4.3 인공지능의 미래와 기술적 특이점

- 특이점이란 무엇인가?
  - 인공지능이 새로운 문명을 만드는 가설적 미래 시점
  - 인공지능 기술이 인간 능력을 뛰어넘어 새 문명을 만드는 시점
  - 지능은 달리 통제할 수 없으며 스스로 발전
  - 미래학자들은 인공지능이 통제 불가능한 수준 발달 예상
  - 특이점에 지능의 폭발(intelligence explosion)이 일어남

- 인공지능이 발전하여 인간의 지능을 뛰어넘는 기점
  - 강한 인공지능 이상의 슈퍼 인공지능이 출현하는 시기
  - 어빙 굿(Irving Good)이 지능의 폭발을 처음 주장
  - 1965년 사람을 훨씬 능가하는 기계를 '초지능 기계'라 함
  - 초지능 기계를 통해 지능 폭발이 일어날 것' 으로 전망

4.3 인공지능의 미래와 기술적 특이점

- 인류는 인공지능에 종속될 것인가?
  - 미국의 인공지능 연구가 유드코우스키(Yudkowsky)
  - 특이점이 1996년부터 시작되었다고 주장
  - 컴퓨터의 속도는 2년마다 2배씩 증가하며 빨라짐
  - '무어의 법칙'을 인공지능에 적용
  - 특이점 도달 이후 인공지능은 엄청난 속도로 발달할 것 주장

[그림 4.29] 유드코우스키

4.3 인공지능의 미래와 기술적 특이점

- 유드코우스키의 예측
  - 1만 년 전에 인류 문명이 시작됨
  - 인쇄술, 컴퓨터의 발명, 인공지능연구
  - 30년 후에는 지능의 폭발이 일어남

〈표 4.4〉 인류의 과거와 인공지능 이후의 미래 예측

연도	사건
5만 년 전	현재의 인간인 호모사피엔스가 시작
1만 년 전	인류 문명이 시작
600년 전	인쇄술의 발명
70년 전	컴퓨터의 발명
60년 전	인공지능의 시작
20년 전	딥러닝의 시작
현재	인공지능 연구가 활발히 진행 중
앞으로 30년 후	지능의 폭발이 일어나고 인류는 인공지능에 종속될 것으로 주장

4.3 인공지능의 미래와 기술적 특이점

- 특이점 개념에 대한 지지와 비판
  - 특이점 주의자들은 강한 인공지능을 확신하며 지지
  - 인간의 대뇌 분석을 통해 인공신경망 구현을 믿음
  - 미국 시사주간지 「TIME」에서 커버 페이지에 소개
  - 특이점이 2045년이라고 특이점 주의자들의 주장

[그림 4.30] 2045년의 특이점 주장

4.3 인공지능의 미래와 기술적 특이점

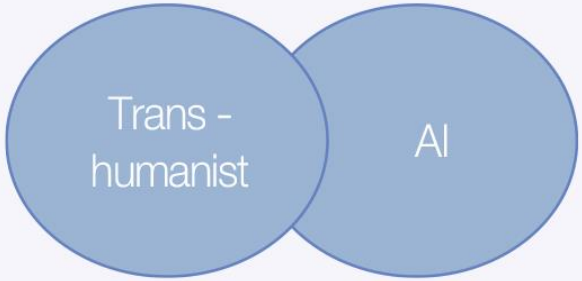
- 미래학자들의 특이점 관점
  - 특이점이 실현될 수 있을 수 있는지에 대한 비판
  - 강한 인공지능 구현에 대해 회의적으로 비판
  - 핀커, 캐플런 등 미래학자들이 특이점 예측에 대해 비판
  - 신학자들도 종교적 차원에서 '지능 폭발'에 부정적 견해
  - 트랜스 휴머니즘과 같은 인공 창조물 비판

미래학자 제리 캐플런

4.3 인공지능의 미래와 기술적 특이점

여기서 잠깐! 트랜스 휴머니즘이란?

트랜스 휴머니즘은 1957년 영국에서 처음으로 등장한 개념인데, 인공지능 등의 과학 기술을 이용하여 인간의 신체적, 정신적 능력을 개선할 수 있다고 믿는 신념이나 운동이다. 트랜스 휴머니즘을 믿는 사람들은 인류가 2050년경 특이점에 도달할 것이며, 인간 이후의 존재인 '포스트 휴먼(post-human)' 시대가 올 것이라 믿고 있다.



4.4 슈퍼 인공지능의 가능성과 대비책

- 슈퍼 인공지능의 가능성과 대비책
  - 슈퍼 인공지능(Super AI)은 미래의 가상적인 인공지능
  - 시기는 기술적 특이점과 관련이 깊음
  - 인간의 지능과 같은 한계가 없음
  - 영국의 철학자 닉 보스트롬(Nick Bostrom)의 주장
  - '슈퍼 인공지능으로 인해 인간이 멸종할 수도 있다'

[그림 4.31] 닉 보스트롬

4.4 슈퍼 인공지능의 시대의 도래와 지능폭발

- 슈퍼 인공지능의 구현 시기
  - 세계 인공지능 전문가들 대상 여론조사: 2040년 ~ 2050년
  - 인공지능학회 참석자들 대상: 30년 ~ 60년 후
  - 인공지능 연구원들 대상: 평균 답변은 2045년
  - 일부는 거의 불가능할 것이라는 견해
  - 일부 연구원들은 앞으로 수백 년 이상 걸릴 것 답변
  - 소프트뱅크 손정의 회장은 슈퍼 인공지능이 30년 전후에 실현될 것 예상, IQ 3,000인 인공지능이 출현할 것으로 기대

4.4 슈퍼 인공지능의 시대의 도래와 지능폭발

- 슈퍼 인공지능 시대에 대한 대비책
  - 슈퍼 인공지능은 통제할 수 없을 만큼 강력해질 것
  - 인간에게 큰 위협이 될 우려, 미리 준비해야 할 것
  - 영화 <2001 스페이스 오디세이>에서 슈퍼 인공지능 등장
  - 인간에게 적대감을 가진 HAL이라는 인공지능 컴퓨터 등장
  - HAL은 입술의 움직임만으로도 말을 알아들을 수 있음
  - 인간과 같이 감정을 느끼거나 추론할 수도 있음

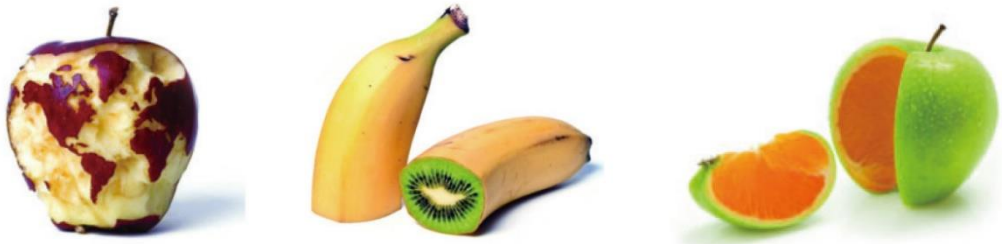
[그림 4.34] 슈퍼 인공지능에 가까운 〈2001 스페이스 오디세이〉

4.4 슈퍼 인공지능의 시대의 도래와 지능폭발

- 슈퍼 인공지능 시대에 대한 대비책
  - <터미네이터>는 기계 군단이 인간 말살 목적 이야기
  - <매트릭스>는 인간과 지능 기계 사이의 전쟁을 소재
  - <2001 스페이스 오디세이>에서 슈퍼 인공지능HAL은 인간에게 적대감을 가진 인공지능 컴퓨터 등장
  - 슈퍼 인공지능은 통제할 수 없을 만큼 강력해질 것
  - 인간에게 큰 위협이 될 우려, 미리 준비해야 할 것
  - 슈퍼 인공지능의 위협은 통제하기 어려울 것
  - 인간에 대한 사랑을 가진 인공지능 설계가 필요
  - 범국민적인 인공지능 윤리 교육 필요

4.4 슈퍼 인공지능의 가능성과 대비책

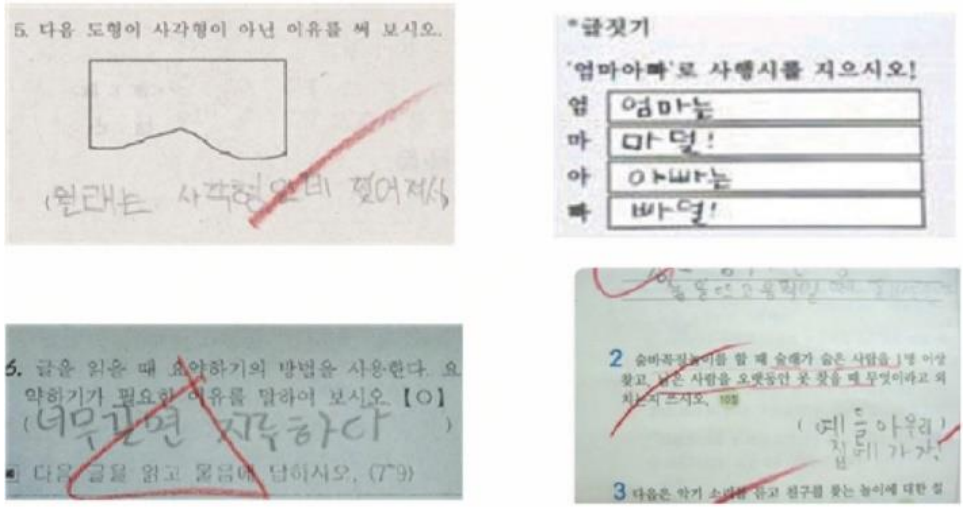
- 슈퍼 인공지능의 창의성 도전?
  - 창의성(creativity)이란 독창적이되 의미 있고 유용한 능력
  - 남과는 달리 새롭고 적절한 것을 만들어내는 능력
  - 레오나르도 다빈치(da Vinci)와 월트 디즈니(Disney) 등
  - 슈퍼 인공지능이라도 인간의 창의성 도전은 어려울 것



[그림 4.36] 창의적 발상

4.4 슈퍼 인공지능의 가능성과 대비책

- 창의적인 것
  - 어린이들의 순수한 발상
  - 창의성의 기반이 됨
  - 인공지능도 인간 또는 그 이상 수준으로 창의적일 수 있을까?



[그림 4.37] 어린이들의 발상

4.5 인공지능 윤리 강령

- 인공지능의 위험성과 윤리 강령
  - 인공지능 윤리 기준의 제정과 실천 대책 필요
  - 나쁜 의도로 사용되지 않도록 경계하고 대비
  - 자율 살상 무기는 암시장에서 거래될 가능성 큼
  - 인공지능 로봇을 통해 살상 무기로 공격할 가능성
  - 얼굴인식 기능을 갖춘 자율 드론 개발중
  - 구글은 무기 시스템용 AI 기술 지원 배제 지침

[그림 4.38] 인공지능 로봇을 통한 공격

[그림 4.39] 공격용 자율 드론

4.5 인공지능 윤리 강령

- 인공지능 윤리 제정
  - 인공지능 윤리 기준의 제정과 실천 대책 필요
  - 나쁜 의도로 사용되지 않도록 경계하고 대비
  - 자율 살상 무기는 암시장에서 거래될 가능성 큼
  - 인공지능 로봇을 통해 살상 무기로 공격할 가능성
  - 얼굴인식 기능을 갖춘 자율 드론 개발중
  - 구글은 무기 시스템용 AI 기술 지원 배제 지침
  - 인공지능 윤리 제정
  - 인공지능 윤리나 윤리 규범의 실천이 매우 중요

[그림 4.40] 인공지능 윤리와 규정

4.5 인공지능 윤리 강령

- 인공지능 윤리
  - 인공지능 개발자들을 제어하는 규칙들과 기준들
  - 연구 대상자들이 지켜야 할 기본적인 윤리
  - 연구 과정이나 내용을 조작하지 않을 윤리
  - 사회적 문제의 가능성을 고려하며 연구할 윤리
  - 예측되는 결과들의 윤리적 문제 여부 판단
  - 혹시라도 모르는 재난에 대한 책임 의식
  - 예방적 차원에서의 윤리 의식

4.5 인공지능 윤리 강령

- 챗봇과 인공지능 윤리
  - 챗봇(Chatbot)은 채팅 로봇, 챗로봇 등으로 불림
  - 챗봇은 인공지능 커뮤니케이션 소프트웨어
  - 문자 대화를 통해 질문에 대한 관련 정보 제공
  - 사용자의 과거 대화 내용 분석한 후 대화
  - 다음 질문을 예측할 수 있는 기능들이 보강
- Microsoft 챗봇 테이의 윤리적 문제
  - 2016년 MS가 개발한 인공지능 챗봇 '테이(Tay)'
  - 인종 및 성차별 발언을 내보내는 큰 소동 일으킴
  - MS는 출시 하루 만에 공식 사과, 운영 중단
  - 그 후 MS는 인공지능 윤리 문제에 적극적 대처

[그림 4.42] 챗봇 테이의 윤리적 문제

4.5 인공지능 윤리 강령

- 인공지능의 아실로마 원칙
  - 2017년 미국 아실로마(Asilomar)에서 열린 인공지능 콘퍼런스
  - 인공지능과 로봇 연구자 등이 '인공지능 기술 23원칙' 발표
  - 이른바 '아실로마 인공지능 원칙(Asilomar AI Principles)'
  - 물리학자 스티븐 호킹, 미래학자 레이 커즈와일 등 서명

[그림 4.43] 아실로마 콘퍼런스의 장면

4.5 인공지능 윤리 강령

- 인공지능의 아실로마 원칙
  - 2017년 미국 아실로마(Asilomar)에서 열린 인공지능 콘퍼런스
  - 인공지능과 로봇 연구자 등이 '인공지능 기술 23원칙' 발표
  - 이른바 '아실로마 인공지능 원칙(Asilomar AI Principles)'
  - 물리학자 스티븐 호킹, 미래학자 레이 커즈와일 등 서명

[그림 4.44] 인공지능의 23가지 아실로마 원칙

- 인류에게 유익한 지능을 만드는 것이어야 함
- 초지능은 인류의 이익을 위해서만 개발되어야 함
- 인공지능 기반 무기 경쟁을 피해야 함 등

4.5 인공지능 윤리 강령

- 인공지능의 아실로마 원칙 (계속)

  연구 문제 (Research Issues : 5가지)
  - 연구목적은 유익한 지능을 만드는 것이어야 함
  - 유익한 사용을 보장하기 위한 연구 자금 동반
  - AI 연구자 및 정책 입안자 간 교류 필요
  - 협력, 신뢰, 투명성의 연구 문화 육성
  - AI 시스템 개발팀은 적극적으로 협력해야 함

  윤리와 가치(Ethics and Values : 13가지)
  - AI 시스템 운영 과정에서 안전과 보안의 확보
  - AI 설계자는 도덕적 의미에서의 이해 관계자
  - AI 시스템은 인간의 존엄성 등과 공존하도록 설계
  - 개인정보 보호
  - 프로세스를 존중하고 개선해야 함
  - 치명적인 자율 무기 경쟁을 방지해야 함 등

4.5 인공지능 윤리 강령

- 인공지능의 아실로마 원칙 (계속)

  장기적인 문제(Longer-term Issues : 5가지)
  - AI 역량에 대한 주의
  - AI의 계획과 관리
  - AI의 위험 관리
  - AI 시스템은 엄격한 안전 관리 및 통제의 대상
  - 초지능의 공공성

4.5 인공지능 윤리 강령

- 구글의 우리의 원칙

  2018년 선언한 7개 항목

  - 사회에 이익이 되는 AI 이용
  - AI 관련 불공정한 편견 만들지 않을 것
  - AI는 인간의 지시와 통제를 받음
  - AI의 프라이버시 보호
  - 원칙에 부합하는 용도로 사용하기 등

[그림 4.45] 구글의 우리의 원칙 7개 항목

4.5 인공지능 윤리 강령

- 한국의 인공지능 윤리 단체
  - 우리나라에도 인공지능 윤리에 관한 관심이 커지고 있음
  - 2019년 비영리 인공지능 윤리 단체 설립됨
  - 추구하는 목표
    - 인류의 행복과 발전에 기여할 수 있도록 지원
    - 인공지능의 부작용과 위험성 정의
    - 인공지능의 윤리와 안전에 대한 연구 등

가까운 곳에서 인공지능 경험하기

인공지능 지도 앱 이용하기

인공지능 기술을 이용한 지도 앱은 현재 여러 개 있다. 우리는 그중에서 마음에 드는 앱을 골라 스마트폰에 설치하여 사용하면 된다. 출발할 지역과 도착할 지역을 지정하면 걸리는 시간과 도착 시각까지 알려주며, 빠르게 도착할 수 있는 교통수단까지 알려주니 매우 편리하다.

인공지능 실습하기

텐서플로로 구현된 신경망 특정화자 목소리 시연

[Deep Voice 2: Multi-Speaker Neural Text-to-Speech]
특정 화자의 목소리를 흉내 내어 문장을 읽어줌

https://carpedm20.github.io/tacotron/

(클릭)

DSU Dongseo University 동서대학교

수고하셨습니다.

인공지능 입문