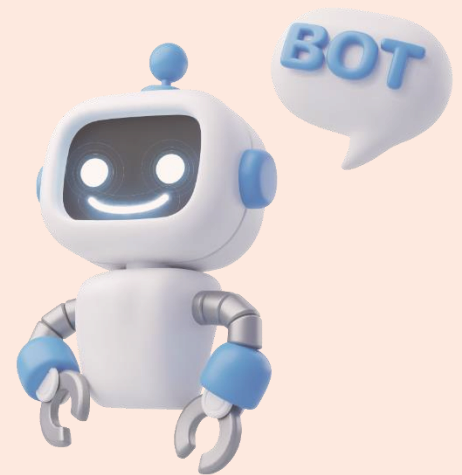


# 인공지능기초

인공지능의 미래와 윤리

● 퀴즈 ●



# 학습 목표

- 미래사회에서 인공지능이 해결하게 될 여러가지 사회적 문제를 예측한다.
- 미래의 사회적 문제 해결에서 인공지능의 역할을 이해한다.
- 데이터 편향성에 대해 이해한다.
- 인공지능 윤리의 개념을 이해한다.

# 미래사회에서 인공지능의 영향력

## ■ 나는 어떻게 해야할까?

인공지능의 순기능과 역기능에 대해 **균형 있는 인식을 갖고** 인공지능의 가치를 판단하며, 인공지능의 올바른 활용 방법에 대하여 끊임없이 고민하는 태도가 필요하다.

### 인공지능 기술

- DNA분석으로 인한 유전자 편집
  - 음성 인식과 언어 이해
  - 컴퓨터 비전

## • 긍정

- 노화로 인해 저하되는 인지 기능 향상과 각종 질병 개선
- 인간의 언어를 이해하는 개인 비서의 등장으로 인한 편의성 추구
- 실종자 수색 기술로 인한 빠른 사건 해결

## • 부정

- 인간 능력을 뛰어 넘는 인공지능 등장으로 인한 인류와 적대적 경쟁의 위협
- 부정적인 대화, 욕설을 사용하는 편향적인 인공지능 탄생 가능성
- 개인 정보를 독점하여 사회를 통제할 가능성

# 미래사회에서 인공지능의 영향력

- 인공지능의 긍정적 영향

인공지능은 다양한 영역에 적용되어 인간 삶의 질을 높일 수 있음.

# 퀴즈...

## 1. 인공지능의 긍정적 영향에 해당하는 것은? ①

- ① 지능화된 서비스로 삶의 질이 향상된다.
- ② 인공지능 설계 오류로 예측하지 못한 문제가 발생할 수 있다.
- ③ 인공지능이 지나치게 고도화되어 통제 불능 상태에 빠질 수 있다.
- ④ 인공지능이 특정 목적에 악용되면 심각한 사회문제가 발생할 수 있다.
- ⑤ 특정 직군의 직업 상황에 따른 사회 양극화로 빈부 격차가 심해질 것이다.

# 미래사회에서 인공지능의 영향력

- 인공지능의 부정적 영향

- 인공지능의 무기화

인간의 조작 없이도 스스로 표적을 찾아 사살하는 자율 살상 무기의  
제조 및 사용

- 인공지능의 법적 책임

인공지능의 오작동으로 피해를 입었을 때, 개발자, 사용자, 운영자의 책임  
정도에 따른 논란

- 인공지능의 차별성

인공지능의 알고리즘이나 학습데이터의 편향으로 특정 계층의 사람들이  
재판, 채용, 의사결정 등에서 차별받을 가능성

# 미래사회에서 인공지능의 영향력

- 인공지능의 부정적 영향(계속)

- 인공지능을 활용한 가짜 뉴스 유포

딥페이크 기술을 악용해 가짜 뉴스 유포 등의 행위로 사회 혼란 가중 및  
사이버 범죄 가능성 확대

- 사생활 침해와 해킹 위험

CCTV 영상에서 안면인식 기술의 남용으로 인한 사생활 침해 우려와  
인공지능 스피커 등 집안 IoT 제품에 대한 해킹 및 도청 우려

# 인공지능으로 **해결**할 수 있는 문제

- 고령 사회 문제 극복

- 인공지능 로봇으로 고령화시대에 필요한 인력문제를 해결한다.

- 안전

- 얼굴인식기술로 빠르게 범죄자를 검거한다.
- 인공지능과 센서를 이용하여 지진, 바람, 온도 등의 복합적인 정보를 분석 및 예측하고 건축물, 시설물의 재난 안전 관리를 한다.



# 인공지능으로 **해결**할 수 있는 문제(계속)

- 일자리 창출

- 인공지능과 관련된 신규 일자리를 창출한다.

- 독거노인 보호

- 의료로봇으로 뇌졸중과 치매관리, 고독사를 방지한다.
- 대화형 인공지능 로봇으로 인한 고독과 불안감을 치유한다.

- 환경 분석 및 예측

- 기후 변화, 멸종위기 동식물의 이주.분포 등 환경의 구조와 현상 등을 분석 및 예측하여 종의 보호 방안을 마련한다.

# 인공지능으로 **해결**할 수 있는 문제(계속)

- 기후 변화 문제 해결
  - 데이터 기반 환경 오염 원인 분석
- 식량문제 해결
  - 식물 재배를 위한 환경 최적화
- 교통 문제 해결
  - 주행 정보 기반 교통 흐름 제어
- 질병 · 바이러스 문제 해결
  - 전문 지식 지원 서비스로 의료 시스템의 효율성 향상

# 퀴즈...

2. 인공지능이 해결할 수 있는 사회문제 중 환경문제만 모두 고른 것은? ③

ㄱ. 멸종 동물 감시

ㄴ. 교육 불균형 해소

ㄷ. 주차 위치 탐색

ㄹ. 원격 환자 관리

ㅁ. 벌목 탐지와 숲 생애 주기 관리

① ㄱ, ㄴ

② ㄱ, ㄹ

③ ㄱ, ㅁ

④ ㄴ, ㄹ

⑤ ㄷ, ㅁ

# 인공지능으로 **발생**할 수 있는 문제

- 일자리 감소

- 단순노동 · 반복적인 업무를 인공지능이 대체하면서 기존 일자리가 감소한다.

- 거짓 정보 생성

- 인공지능을 이용한 거짓 정보 생산, 사진 및 영상 조작 후 소셜 미디어를 활용한 유포 등을 할 수 있다.

- 법적 · 윤리적 책임 소재 문제

- 인공지능을 탑재한 자율 주행차의 오류, 의료 분야에서 진단 오류 및 실수 발생시 책임 소재가 불분명하다.

# 인공지능으로 **발생**할 수 있는 문제(계속)

- 경제 체제의 붕괴 위험

- 인공지능이 인간의 직업을 대체함으로써 생산성은 향상하나  
경제적 문제로 소비층이 붕괴하면서 체제마저 붕괴될 위험이 있다.

- 인공지능의 무기화 위험

- 인공지능을 무기로 활용하여 인간을 해칠 가능성이 있다.

# 미래사회에서 **인간**의 역할

- **나는 어떻게 해야 할까?**

인공지능 시대의 미래사회에서는 인간과 인공지능이 각자의 장점을 살릴 수 있도록 경쟁이 아닌 협력을 통해 상호 보완을 해야한다.

## 미래사회에서 요구되는 역량은?

- 기계와의 협력적 의사소통 역량
- 획일적이지 않은 문제 인식 역량
- 다양성의 가치를 조합하는 대안 도출 역량

# 미래사회에서 **인공지능**의 역할

- 인공지능이 어떤 역할을 하느냐는 **인간**이 어떤 방향으로 설계하고 조정 하느냐에 달려 있다.

## 미래사회에서 인공지능의 역할은?

1. 인간을 대체하는 역할
2. 인간을 돕는 도구로서의 역할

**생각해 봅시다.**

- 의료 진단 시스템
- 직원채용
- 예술 작품 제작

# 미래사회에서 **인공지능**의 역할

- 의료 진단 시스템

1. 인간 의사가 인공지능보다 실수를 더 많이 해. 앞으로는 인공지능이 인간보다 의료진단을 더 잘 할거야.
2. 인공지능이 인간 의사를 대체하기에는 한계가 있어. 환자의 아픔을 공감하는 감성적인 활동은 인간 본성의 영역이니까 이것을 이해할 수 없는 인공지능은 인간을 돕는 역할만 할거야.



# 미래사회에서 **인공지능**의 역할

- 직원채용

1. 인공지능이 면접을 보면 사람의 미세한 표정 변화, 사용하는 단어를 객관적으로 분석해서 어떤 사람인지 정확하게 판단해줘. 따라서 면접관 업무를 사람이 할 필요는 없어.
2. 사람은 직감이 있어. 대화를 나누면 어떤 사람인지 어느 정도 파악할 수 있어. 인공지능은 단지 사람이 놓칠 수 있는 부분을 채워 주는 역할을 할 뿐이야.

# 미래사회에서 **인공지능**의 역할

- 예술 작품 제작

1. 인공지능도 창의적일 수 있어. 인공지능이 그린 그림도 경매에 나오고 있는데 굳이 인간이 예술을 할 필요는 없어.
2. 인공지능은 인간 예술가에게 영감을 불어넣는 역할을 하거나 예술 창작을 흉내 낼 수 있지만, 감성을 가지고 직접 창작할 수는 없어.

# 퀴즈...

## 3. 인공지능의 역할에 대한 다음 의견 중 관점이 다른 것은? ⑤

- ① 다솜 : 인공지능도 창의적일 수 있어서 예술가 역할을 할거야. → 대체
- ② 정혜 : 인공지능은 인간의사보다 환자를 더 잘 진단 할 수 있을 테니  
인간의사는 필요없지. → 대체
- ③ 석준 : 인공지능이 면접을 보면 어떤 사람인지 정확하게 판단할 테니까  
앞으로는 인공지능에게 맡기자. → 대체
- ④ 소연 : 인공지능은 학습에 한계가 있는 인간과 달리 속도나 용량에 제한이  
없으니 인간을 지배할지도 몰라. → 대체
- ⑤ 하림 : 인공지능은 학생이 잘하는 것과 어려워하는 것을 파악하여 학교  
선생님이 더 잘 가르칠 수 있도록 도울 거야. → 돕는 도구

# 데이터 편향성

- 데이터 편향성이란 무엇일까?

학습에 사용되는 **데이터**가 현실세계를 잘 반영할 수 있도록 충분한 양이 확보되지 않았거나 현실 세계의 성질을 골고루 반영하지 못하여 한쪽으로 치우친 성질을 가지고 있는 것을 말함.

- 롱테일 편향(long-tail bias)

학습에 사용되는 데이터에 특정 종류의 데이터가 빠져 생기는 편향

# 퀴즈...

## 4. 다음 설명에 해당하는 용어는? ④

인공지능이 공평하지 않은 기준이 적용된 데이터로  
기계학습을 하여 편향된 결과를 산출하는 것을 말한다.

① 네테켓

② 정보 윤리

③ 사이버네틱스

④ 데이터 편향성

⑤ 윤리적 딜레마

# 데이터 편향성(계속)

- 인공지능도 **인간**처럼 편견을 가질까?

학습한 데이터가 편향될 경우, 한쪽으로 치우치거나 잘못된 결과를 도출할 수 있음.

- 데이터 편향성이 발생하는 원인은 무엇일까?

- 데이터 **선택** : 편향된 데이터 선택 → 일부를 가지고 전체를 판단, 대표성이 부족
- 데이터 **양** : 일부 데이터만 사용 → 대량의 데이터로 학습, 안정적이고 믿을 수 있는 결과 산출
- 데이터 **질** : 목적과 관계없는 데이터 사용 → 목적에 맞는 데이터로 학습, 정확한 결과 산출

# 퀴즈...

5. 데이터 편향성에 대한 설명으로 바른 것은? ③

- ① 학습에 사용하는 데이터가 현실세계를 잘 반영한 성질이다.
- ② 데이터의 양이 충분히 확보되었을 때 나타나는 성질이다.
- ③ 현실 세계를 골고루 반영하지 못하여 한쪽으로 치우친 성질이다.
- ④ 확률에 의해 나타나기도 하고 나타나지 않기도 한다.
- ⑤ 사람은 항상 공정하므로 잘 나타나지 않는다.

# 데이터 편향성(계속)

- 데이터 편향성 체험하기

- 인공지능 번역기

- ✓ 구글번역기 웹 사이트(<https://translate.google.co.kr>) 에 접속하여 [남성]과 [여성]을 가리키는 대명사와 직업이 포함된,
    - ➔ 간단한 문장을
    - ➔ 인도네시아어나 터키어로 번역했다가
    - ➔ 다시 영어로 재번역해 보자.

영어문장 : She is a professor. He is a nurse.



# 인공지능 윤리

- 인공지능 윤리란...

인공지능 윤리는 인공지능을 바르게 **활용**하기 위해 사회 구성원이 갖추어야 하는 가치관과 행동 양식.

- 인공지능 윤리의 중요성

- ✓ 인공지능이 겪을 수 있는 다양한 윤리적 딜레마 상황에 대비해야 한다.
- ✓ 인공지능이 알고리즘 오류, 편견이 포함된 잘못된 윤리관 학습으로 의도하지 않은 실수를 할 가능성에 대비해야 한다.
- ✓ 인공지능의 능력이 인류가 통제하기 어려울 정도로 발달할 것에 대비해야 한다.

# 퀴즈...

6. 인공지능을 사용하는 구성원의 역할 중 바른 것은? ④

- ① 인공지능 **개발자**는 인공지능을 운영하는 주체다.
- ② 인공지능 **사용자**는 인공지능을 개발하는 주체다.
- ③ 인공지능 **관리자**는 인공지능을 최종적으로 사용하는 주체다
- ④ 인공지능 **사용자**는 편향되지 않은 데이터로 인공지능을 학습시켜야 한다.
- ⑤ 인공지능 **운영자**는 인공지능 운영 과정을 지속적으로 점검하고, 사용자를 교육해야 한다.

# 인공지능 윤리(계속)

인공지능 윤리는 인공지능의 악의적 **활용**을 예방하고, 올바르게 **사용**하기 위해 준수해야 하는 규범이다.

- 인공지능도 윤리적 딜레마 상황에 빠질 수 있을까?
  - ✓ 우리가 살면서 겪는 모든 상황이 옳고 그른 선택으로 나누어지는 것은 아니며, 어떤 선택을 해도 바람직하지 못한 결과가 나오는 경우도 있다. 이처럼 어느 쪽을 선택해도 바람직하지 못한 결과가 나오는 상황을 **윤리적 딜레마**라고 한다.

# 인공지능 윤리(계속)

- 하인츠의 딜레마
  - 부인이 위독함, 치료제 약값을 깎아달라 → 거절당함, 나중에 반절 주마 → 거절당함, → 하인츠가 밤에 약을 훔침.
  - 하인츠가 치료제를 훔친 행동에 대해 어떻게 생각하는지?
- 모럴 머신(moral machine, <https://www.moralmachine.net>)

MIT에서 만든 도덕 기계, 윤리적 딜레마 상황에서 사람들의 도덕적 가치 판단을 조사하기 위해 만들어진 온라인 플랫폼으로, 다양한 트롤리 딜레마를 체험할 수 있다.
- 트롤리 딜레마
  - 윤리학 분야의 사고 실험으로, 다섯 사람을 구하기 위해 한 사람을 죽이는 것이 도덕적으로 허용 가능한지에 대한 질문

# 인공지능 윤리(계속)

- 인공지능의 **가상** 딜레마 상황
- 자율 주행 자동차의 윤리적 딜레마
  - 갑작스런 브레이크 고장으로 멈출 수 없게 된 상황일 때,
  - 3명의 보행자 사망? 또는 탑승자 사망?
- 인공지능 의사의 윤리적 딜레마
  - 백신이 1개 남아 있을 때, 생존가능성이 희박한 사람과 높은 사람 중 누구?
- 인공지능 군사무기의 윤리적 딜레마
  - 목표물을 공격할 때, 그 주변 민간인의 희생이 예상되는 상황에서 명령을 따라야 할까?

- 인공지능에 윤리가 필요할까?

인공지능을 실현하기 위해 기술을 개발하는 것도, 인공지능의 도움을 받는 것도 인간이다.

그리고 어느 시점이 되면 인공지능의 결정과 행동이 인간에게 영향을 미치고 인간의 사고를 형성하는 데 큰 역할을 하게 될 것이다.

# 인공지능 윤리(계속)

미국의 공상 과학 소설 작가 아이작 아시모프는 1942년 발표한 단편소설 <런 어라운드> 에서 인류를 위하여 로봇이 반드시 지켜야 하는 3원칙을 처음으로 언급 → 아이작 아시모프의 로봇 3원칙 + 0원칙

- 제0원칙 : 인류를 위해 행동해야 한다.  
→ 로봇은 반드시 본래의 목적인 인류를 위해 행동해야 한다.
- 제1원칙 : 위험에 처한 사람을 그냥 지나치면 안 된다.  
→ 로봇은 인간에게 해를 입혀서는 안된다. 그리고 위험에 처한 인간을 모른 척해서도 안된다.
- 제2원칙 : 타인에게 해를 입히는 것은 할 수 없지만, 당신에게 필요한 일은 할 수 있다.  
→ 제1원칙에 위배되지 않는 한 로봇은 인간의 명령에 복종해야 한다.
- 제3원칙 : 1,2원칙을 지키는 범위 내에서 스스로를 위험에 빠뜨리는 일은 하지 않는다.  
→ 제1원칙과 제2원칙에 위배되지 않는 한 로봇은 로봇자신을 지켜야 한다.

# 퀴즈...

## 8. 다음 사례에 알맞은 개념은? ④

의사를 대신해 환자를 치료하는 인공지능이 한 임신부를 진단하고 있다. 이 임신부는 평소 지병을 앓고 있어 출산을 강행하면 산모가 사망할 수 있고, 반대로 출산을 포기하면 산모가 살 수 있다. 태아를 살릴 것인지 산모를 살릴 것인지 선택해야 한다.

- ① 사회적 합의                      ② 인공지능 윤리
- ③ 데이터 편향성                  ④ 윤리적 딜레마                  ⑤ 인공지능의 양면성



# 데이터 공정성

- 인공지능에게 공정성을 학습시킬 수 있는 방법은 무엇일까?

영국의 한 기업에서는 인간의 편견 없이 직원을 선발하기 위해 도입한 [인공지능 채용 시스템]이 사회의 성차별적 관행을 학습하는 바람에 오히려 특정 성별의 지원자를 중심으로 선발한다는 문제를 일으킨 바 있다. 인간의 편견이 개입되지 않아 공정할 것이라고 예상한 인공지능이 오히려 인간의 편견을 따라 한 것이다. 인공지능은 어느 누가 상대하더라도 공정해야 한다. 인공지능의 공정성은 설계, 개발, 활용 과정에서 다음 사항을 고려할 때 확보될 수 있다.

# 데이터 공정성(계속)

- 인공지능의 공정성 확보
  - ✓ 공정성을 확보하기 위한 **의도적** 노력
  - ✓ 다양한 계층 노력 : 남녀 노소, 소외계층 등을 최대한 고려
  - ✓ 알고리즘 투명성 확보
  - ✓ 인공지능 이해관계자별 윤리 :
    - 인공지능 개발자,
    - 운영 · 관리자,
    - 사용자의 각 주체가 자신의 위치에서 윤리적인 역할을 해야한다

# 데이터 공정성(계속)

- 다양한 인공지능 원칙
  - 아실로마 인공지능 원칙
  - 마이크로소프트의 인공지능 원칙
  - 구글의 인공지능 원칙
  - 카카오의 알고리즘 윤리 헌장

# 용어

- ✓ 딥페이크(deepfake)

딥러닝(deep learning)과 가짜(fake)를 합쳐 만든 합성어로,  
인공지능을 기반으로 이미지를 합성해 만든 영상 편집물.

- ✓ 설명 가능한 인공지능(XAI, eXplainable Artificial Intelligence)

인공지능이 내린 결정이나 답을 사람이 이해할 수 있는 방식으로 설명할 수  
있는 인공지능.

# 용어

본 수업에 사용된 일부 자료 및 영상물은 강의 내용을 보충하기 위해 교육 목적으로 활용하였습니다.  
자료 및 영상물의 불법적 이용, 무단 전재·배포는 법적으로 금지되어 있으니, 학생 여러분께서는 학습 외 용도의 사용을 삼가 바랍니다.