

---

# AI시대 연구윤리

김 용 환

차의과학대학교 데이터경영학과

---



# CHA UNIVERSITY

## AI시대 연구윤리

발표일 | 2023. 11. 10.  
발표자 | 김 용 환 교수



PRESENTATION TITLE

### 'AI시대 연구윤리' 만나기

#### INDEX

- 01 들어가기 : AI시대 산업혁신
- 02 초거대 AI시대 산업혁신
- 03 ChatGPT와 산업혁신
- 04 AI시대 연구윤리와 이슈 및 시사점
- 05 나가기 : AI시대 연구윤리와 이슈 사례들

2023-10-31

2023년 김용환교수 특강



3

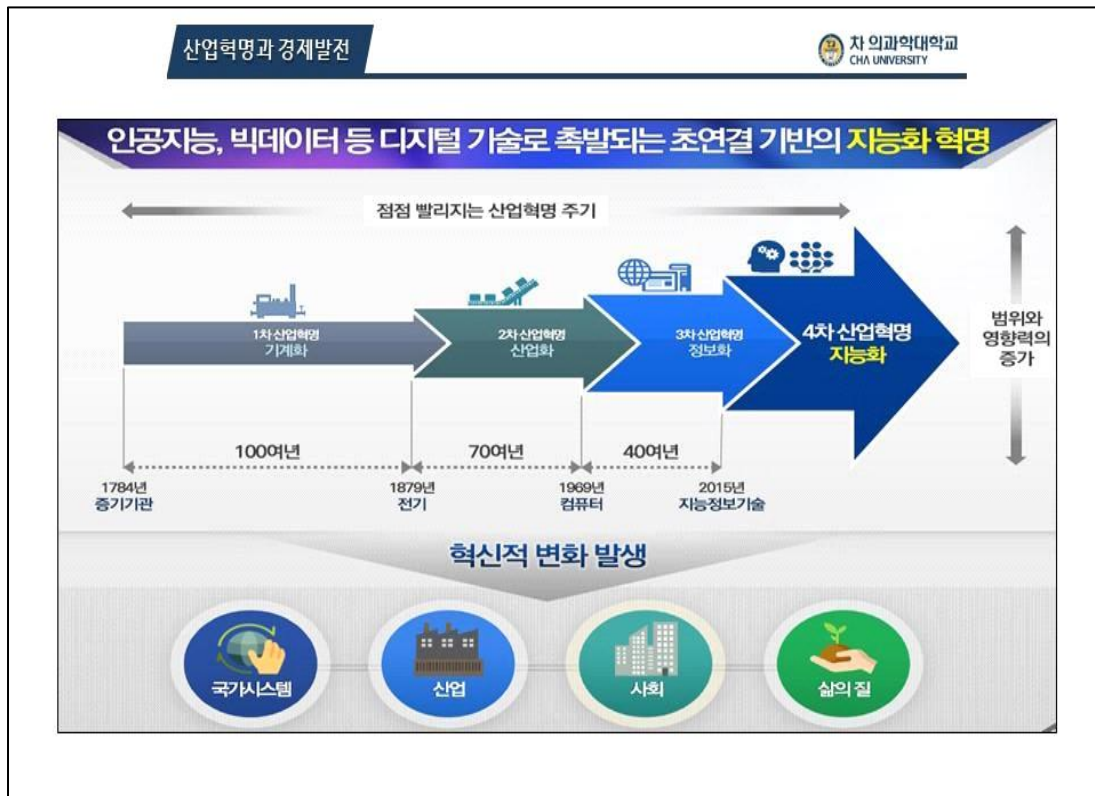
2



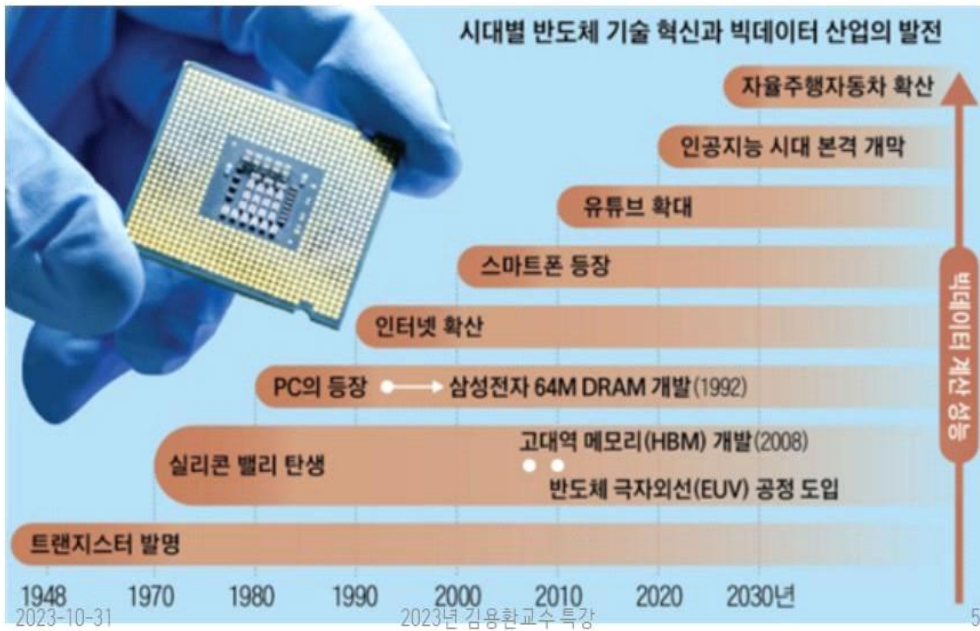
2023-10-31

2023년 김용환교수 특강

3



## 가속화하는 디지털 혁신과 산업경제 변화: AI경제의 부상



### 4차 산업혁명과 경제

### 시기별 산업 및 비즈니스 모델 발전 트렌드

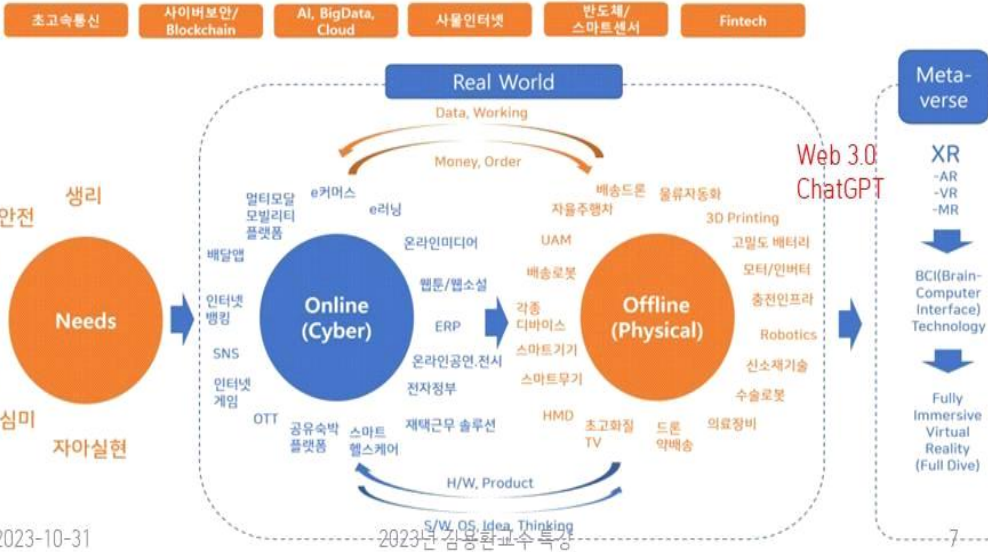


자료: WEF(2016). The Future of Jobs, January. 재수정

- 유전체 분석비용/시간 혁신 사례: 개인맞춤형 의료서비스 정밀의료 : 2003년 27억달러, 8년간 => 2017년 1000달러이하, 48시간
- 약 10년 걸리던 백신 개발을 10개월만에 상용화... AI 의료혁신 : 화이자 백신 제공 등
- '양자(量子·quantum) 기술'을 통해, 1만년 걸릴 계산을 3분에 가능

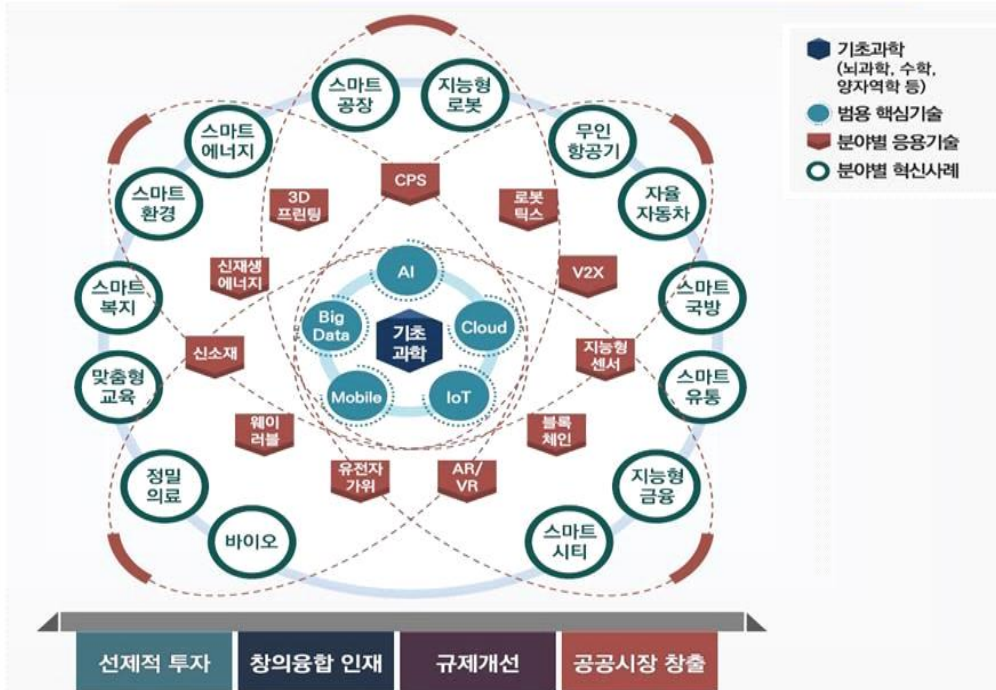
## < '산업혁신과 디지털 트랜스포메이션(Digital Transformation)' : 데이터 생산/저장/분류/분석/공유/활용 >

- Offline 영역의 로봇화는 O2O, CPS의 완성을 가능케 함. 이는 필연적으로 노동의 소외와 다른 욕구충족으로의 이동을 가져올 것
- 인지, 판단, 제어의 주체가 사람에서 로봇으로 이동되면 Real World의 자동화가 이루어지고, 인간의 욕구는 가상공간으로 빠르게 이동



## 산업혁명과 경제 4차 산업혁명과 산업경제 혁신

차 의과학대학교  
CHA UNIVERSITY



## 최근 10년 동안의 주요 변화

### 사회 변화

AI 기술로 인해 생산성 및 편의성이 증대되고 있지만, 다양한 이슈들이 발현하며, 이에 대한 법, 윤리 논의의 촉발

미국 저작권청 (US Copyright Office)



AI가 그린 그림에는 인간 저작의 요소가 포함돼 있지 않다.

스티븐 테일러 박사는 AI를 저작권자로 등록 요청했으나, 미국 저작권청에 거절

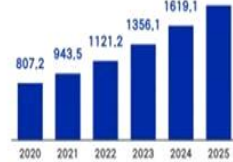
EU, OECD 등 윤리 가이드라인 마련



### 경제 변화

AI 기술은 새로운 국가 경제 성장에 중대한 영향을 미치며 국가 경제 성장을 추동

우리나라 AI 시장 전망 매출규모(십억원)  
출처: 한국IDC '22



산업용 로봇 노동비용 절감 효과  
출처: BOG, OECD (%)

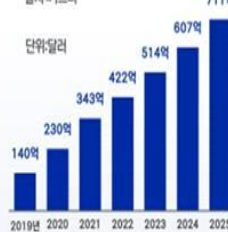


### 산업 변화

모든 산업에 AI를 적용하는 것이 가속화되고 있으며, 빠르게 변화하지 않으면 시장에서 도태되어 산업과 기업의 생존과 직결



세계 AI 반도체 시장 지속 확대  
출처: 가트너



### 글로벌 변화

기술 우위가 국제패권을 좌우하는 기정학 시대가 도래하고, AI 분야에서 글로벌 리더십 확보를 위해 각국은 소리 없는 전쟁 중



미국

- 국가 AI R&D 전략 계획 (16, '19)
- AI 이니셔티브법 (20)
- 국가 AI 리소스 지원 TF 운영 (21~)



중국

- 차세대 AI 발전계획 및 3개년 액션플랜 (17)
- 14.5 규획발표(21): '35년까지 장기계획 / 인재 창출 가속화



영국

- 영국 AI 산업 검토 보고서 (16)
- AI 섹터딜 (18)
- AI 로드맵, AI 국가전략 (21)

## 디지털 경제와 AI 경제 발전

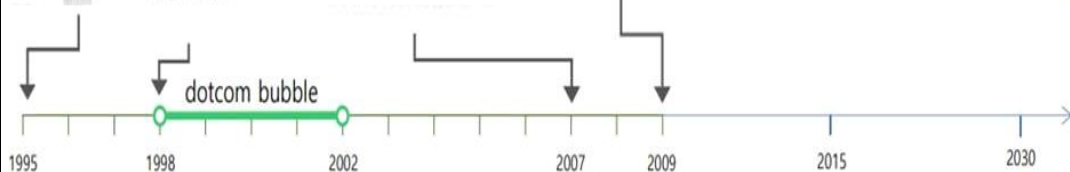
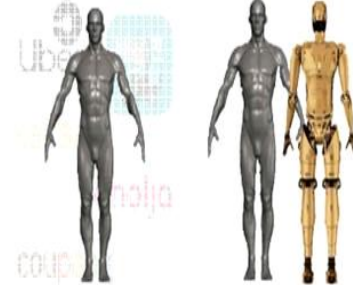
Web & Cloud Economy

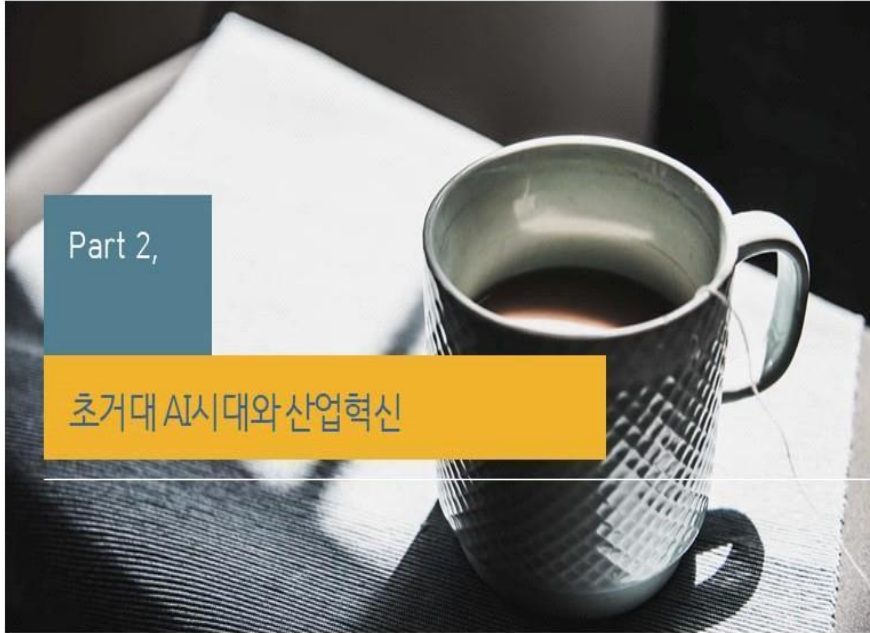
Shared Economy

App Economy

Platform Economy

AI Economy?





Part 2,

초거대 AI시대와 산업혁신

2023-10-31

2023년 김용환교수 특강

11

## < 인공지능(Artificial Intelligence) >

# 인공지능이란 무엇인가

### ▶ AI (Artificial Intelligence)

- 인공지능(AI)은 학습, 문제 해결, 패턴 인식 등과 같이 주로 인간 지능과 연결된 인지 문제를 해결하는 컴퓨터 프로그램이나 이를 포함한 컴퓨터 시스템을 일컫는다.

=> 인공지능 기술을 이용해 고도화된 기능을 지닌 제품과 서비스를 구현하고, 이를 통해 새로운 가치(value)를 창출하여, 사용자경험(UX, user experience)을 향상 시키는 과정

=> 인공지능의 3대 요소 : 컴퓨터 파워 + 빅데이터 + 알고리즘

2023-10-31

2023년 김용환교수 특강

12

< 인공지능(Artificial Intelligence) >

# 인공지능이란?

인공지능

+

로봇



인조인간

알파폴드, 텐스플로우, ChatGPT, 람다와 바드 등

알파고, 검색엔진, 주식매매프로그램, 번역프로그램 등

2023-10-31

2023년 김용환교수 특강

13

< 인공지능(Artificial Intelligence) >

인간을 알면 인공지능이 보인다!

## 인공지능 = 인간 처럼 생각하는 기계

<1950, Alan Turing>

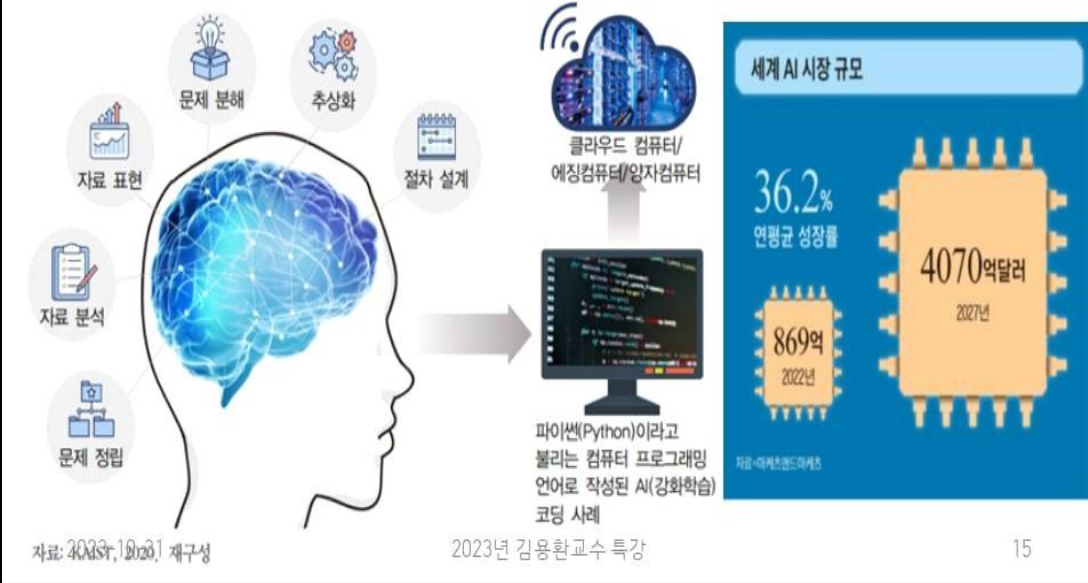


2023-10-31 NC16 SOME HOMOSEXUAL REFERENCES 2023년 김용환교수 특강

14

## < 인공지능(Artificial Intelligence) >

▶ 그림 1-14 | '컴퓨팅 사고' 구성과 원리



## < 인공지능(Artificial Intelligence) 기술진보 >

### 인공지능 기술의 발전 방향



### 머신 러닝 (Machine Learning)

지도학습	자율학습	강화학습
얼굴과 동작 인식 언어 인식 범죄자 몽타주 작성 정보 검색 엔진 날씨 예측	상품 추천 엔진 고객 세그멘테이션 빅데이터가시화 기능 추출 화학, 생체 구조 발견	로봇릭스 자율주행차 학습 과업 실시간 의사결정 게임 AI

## < 인공지능(Artificial Intelligence) 시스템과 생태계 >



## < 인공지능(Artificial Intelligence) 시장의 성장과 발전 >



## < 인공지능(Artificial Intelligence), 머신러닝, 딥러닝 >

### 인공지능, 머신 러닝, 딥러닝의 개념

#### 인공지능 (Artificial Intelligence)

인간처럼 사고하고 판단하는 기계<1950, Alan Turing>

#### 머신 러닝(Machine Learning)

데이터를 통해 학습시키거나, 컴퓨터가 스스로 학습하여 인공지능의 성능(정확도, 속도, 응용 범위)을 향상시키는 방법

#### 딥러닝 (Deep Learning)

인공신경망이론을 기반으로, 인간의 뉴런과 유사한 입/출력 계층 및 복수의 은닉 계층을 활용하는 학습방식, 복잡한 비선형 문제를 비지도 방식 학습으로 해결



출처: LG경제연구소

## < 인공지능(Artificial Intelligence) : 주요 이슈들 >



# < 가짜 이미지를 생성한 인공지능(Artificial Intelligence) 방법(2014) >

GAN(Generative Adversarial Network) 관련 기초 설명

**GAN**

## GAN(Generative Adversarial Network) 대표 이미지



## GAN 탄생 및 특징

- 2014년 이안 굿펠로우의 논문에서 처음 소개되어, 머신러닝 학계에 큰 영향을 끼치고 있는 혁신적인 생성 모델 구조
- GAN이 발표되기 전까지는 개, 고양이 등 이미지 분류, 인식 등에 초점을 맞춘 판별 AI 연구가 주류였으나, GAN 발표 이후 이미지 생성 등 생성 AI 분야의 연구가 폭발적으로 성장
  - ※ 안 르곤 교수는 GAN이 지난 10년간 가장 흥미로운 아이디어로 꼽음 (2016)

## GAN 개념 설명

- 두 개의 모델이 서로 경쟁하며, 학습되는 방식이 주요 특징
- |       |   |
|-------|---|
| 생성 모델 | 진짜 데이터와 비슷한 가짜 데이터를 만들어 판별 모델이 구분하지 못하도록 학습 |
| 판별 모델 | 생성 모델의 결과인지, 진짜 데이터에서 나온 것인지 판별하는 방식으로 학습   |
- 두개의 모델이 번갈아 가며 데이터를 주고 받으며, 점진적으로 학습되어 고도화된 생성 모델이 완성

# < 인공지능(Artificial Intelligence) >

## GAN 발표 이후 국가사회(사회·연구·산업) 변화 모습

**GAN**

### [사회] 가짜와의 전쟁 시작

- 딥페이크 기술을 활용한 얼굴 합성으로 정치, 경제적인 위해 가능성 발생
    - \* 유명인뿐만 아니라, 일반인들도 음란물에 합성하는 등 불법행위가 발생
- 
- 국내도 딥페이크 관련 규제 논의 후 딥페이크 제작 유포에 관한 처벌 수준 강화
    - \* 성폭력 범죄의 처벌 등에 관한 특례법 개정('20)
    - 다만, 80% 수준이 집행유예('22.2 기준)로 판결 되어 양형에 대한 논란 지속

### [연구] GAN을 통해 고도화되는 GAN

- AI 성능 향상을 위한 학습용 데이터셋 부족 시 합성 데이터를 생성하여 활용
    - 원본 생성 데이터
  - GAN을 기반으로 하는 창의적인 후속 연구들이 등장해 생성 AI 분야 연구 활발
- 
- 자료: MIT
- 자료: <https://roytravel.tistory.com/109>(블로그)

### [산업] 비즈니스 모델로의 활용 본격화

- 이미지 관리 툴에 생성 모델을 적용하여, 클릭 한번으로 다양한 효과 달성
    - \* 인물의 연령을 변화시켜, 헤어스타일 변형
  - 이미지 영역뿐만 아니라, 다양한 영역에서도 생성 모델은 적극적으로 도입되어 시장에 진입 중
    - \* AWS Deep composer를 통한 하모니 생성
- 
- 
- 자료: Adobe

GAN 기술 발표 후 생성 AI 분야의 연구가 폭발적으로 증가하며, 가짜와의 전쟁 촉발, AI 서비스 성능 향상 등 명과 암이 동시에 발생

## < 인공지능(Artificial Intelligence)의 대중화(2016)>

알파고-이세돌 대국 관련 소개

알파고

### 알파고-이세돌 대국

◎ 2016년 3월 서울 포시즌스 호텔에서 시작한 구글 딥마인드 챌린지 매치에서 알파고와 이세돌 9단과의 대국 있었고, 이세돌 9단이 최종 4:1로 패배하며, 인공지능의 성능을 전 세계에 알리게 된 대표적인 사건

1국 (3.09)	승 알 파 고 ○	186수 알파고 불계승
2국 (3.10)	승 알 파 고 ●	211수 알파고 불계승
3국 (3.12)	승 알 파 고 ○	176수 알파고 불계승
4국 (3.13)	승 이 세 돌 ○	180수 알파고 불계승
5국 (3.15)	승 알 파 고 ○	280수 알파고 불계승



### 알파고 특징 설명

◎ 알파고는 CNN(딥러닝), 강화학습, 몬테카를로 트리 등 인공지능과 게임이론의 최신 기술을 적용하여 활용하고, 구글의 클라우드 컴퓨팅 환경에서 개발되었으며, 단계별로 구분하여 인공지능 기술 적용

- 1단계 CNN 딥러닝 방식을 통해 사진에 구축된 기보를 이미지화하여 학습
- 2단계 인공지능 간의 게임을 통해 이긴 게임에게 보상을 주는 강화학습 방식 적용

23

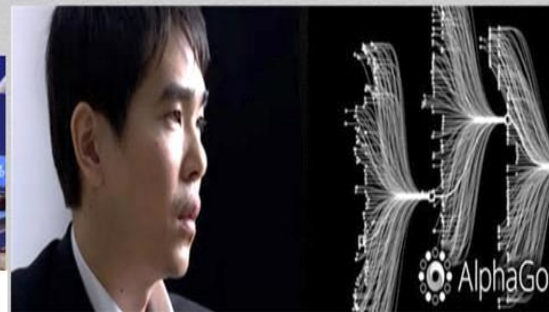
## < 인공지능(Artificial Intelligence)시대를 만나다 >

### AI (인공지능) 시대의 도래

#### ▶ 알파고

- 알파고 이세돌 9단을 물리치다. (2016. 3)
- 알파고, 세계 1위 중국의 커제 포함 중국대표 5인 단체전에서 전승 후 은퇴 (2017. 5)

### 알파고 (AlphaGo)



2023-10-31

2023년 김용환교수 특강

24

## < 인공지능(Artificial Intelligence) >

### 알파고 등장에 따른 국가사회 변화 모습

알파고

#### [국가 정책] 글로벌 주요국가들은 앞다투어 인공지능 정책 수립

- 이제들-알파고 대국 이후 전 세계적으로 AI의 우수성을 인식하고, 국가 차원의 정책 수립을 위한 노력 지속
- \*'17-'18년 한국, 미국, 중국, 일본, 독일 등 AI 기술력 확보를 위한 전략 발표



- 이후, 국가차원의 R&D 중심 AI 정책뿐만 아니라, AI를 통한 산업 전반의 혁신을 위한 주요국의 정책들이 지속적으로 발표

\* AI 국가전략(한국, '19), 국가 AI 전략(싱가포르, '19), 범캐나다 AI 전략(2개 캐나다, '21), 국가 AI 전략 업데이트(독일, '20), 국가 AI 전략(영국, '21) 등

#### [인식 변화] 인공지능에 대한 대국민 인지도 급격히 향상

- 우리 국민의 대부분은 인공지능에 대해 인지(99.3%)하고 있으며, 관심도 또한 높은(59.8%) 것으로 확인
- 인공지능의 신뢰성에 대해서는 대체로 '보통' 이상이지만, 부작용에 대한 우려를 동시에 하고 있는 것으로 분석
- \* 생활, 생산성, 시장 영역에서 긍정적 평가가 높았으나, 프라이버시/양극화/일자리 영역에 대한 우려도 상존



자료 : 인공지능 대중화를 위한 대국민 인공지능 이용 인식조사 (4차산업혁명위원회, 2021)

### R&D 영역에서 주목받고 있던 AI 기술이 알파고를 통해 전 세계적인 주목 등 AI 정책 수립의 도화선 역할

2023-10-31

25

## < 인공지능(Artificial Intelligence)시대란? >

### ▶ 인공지능은 인간 사회 전 범위에 영향을 미칠 것

- 자율주행 이동 기기
- AI 친구, Virtual Assistants
- 외국어 통역, 번역
- 환경 문제 해결, 농사 대행
- 국방, 범인 찾기
- 가상 학교, 학습



진화 중인 자율주행차, 어디까지 왔을까?, Spot Launch

URL: <http://mediask.co.kr/35476>  
<https://www.youtube.com/watch?v=w&CQXHEgA>

2023년 김용환교수 특강

26

## < 인공지능(Artificial Intelligence)시대란? >

### ▶ 인공지능은 인간 사회 전 범위에 영향을 미칠 것

- 예술 창작, 게임, 요리
- 병 진단, 수술, 헬스케어
- 금융, 마케팅과 쇼핑
- 법률 서비스, 언론
- 물류, 교통 시스템
- 결제, 보안
- 사회복지 등



[특집] 인공지능과 헬스케어의 융합, 기술사업화 시장 키운다, AI 금융비서가 골라준 카드로 결제...여행·유통 서비스는 '덤'

2025년 김용환교수 특강

27

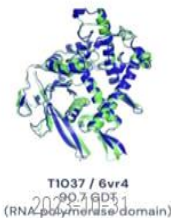
## < 바이오 혁신과 인공지능(Artificial Intelligence)(2020) >

### 알파폴드2의 기초 설명

### 알파폴드2

#### 단백질 구조 예측 및 알파폴드의 등장

- 약물의 표적부위인 단백질의 구조를 이해하고, 단백질과 약물의 상호작용을 예측하는 것은 신약개발 시 핵심
- 이에, 단백질 구조 예측에 대해 실험적 방법들을 이용하여 다양한 연구\*들이 진행 중이었으나, 시간과 비용이 많은 한계에 봉착  
\* 엑스선 결정술, 저온 전자 현미경 방법 등
- 딥마인드는 인공지능, 생물학 등 도메인 전문가와 협업하여, 딥러닝 기반의 단백질 구조예측 모델(알파폴드)을 개발('18)
- 약 50여년간 큰 혁신이 없고 난제라고 여겨지던 분야에서 기존의 연구 성과를 압도하는 알파폴드를 개발하여 새로운 가능성을 시사



파란색  
실험실 결과

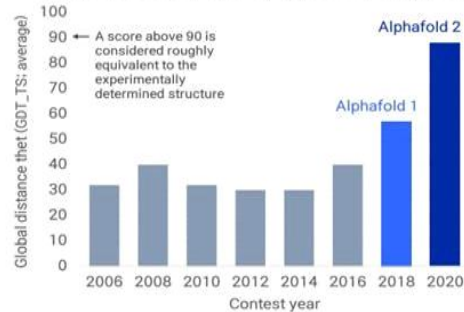
초록색  
알파폴드 예측

출처 : 딥마인드

#### 알파폴드2의 성과

- 2018년 CASP\*에서 알파폴드는 괄목할만한 성능을 보여주고, 2년 후인 '20년 알파폴드2에서도 한 번의 혁신적인 성과 달성

\* CASP(Critical Assessment of protein Structure Prediction)- 전세계 연구자들이 모여 각자의 구조 예측 방법론을 가지고 경쟁하는 대회



- 2022년 딥마인드는 단백질 구조예측 플랫폼인 알파폴드를 인류의 발전을 위해 오픈소스로 공개

28

## < 인공지능(Artificial Intelligence) >

### 알파폴드2

#### 알파폴드2가 시장에 미친 영향

- 알파폴드의 성능을 높이기 위한 과학계의 단백질 구조예측 성능 경쟁 본격화
  - 메타는 ESM폴드를 개발하여 6억개 규모의 단백질질을 예측하여, 알파폴드가 예측한 2억개를 넘어섰다고 발표('22.11)



- 단백질 구조 예측을 넘어 빅테크 기업들도 신약개발 시장 진출
  - (구글) 구글의 모회사 알파벳은 단백질 구조 기반 AI 신약 개발사인 아이소모픽 랩스 설립('21)
  - (카카오브레인) 신약개발 AI 기업인 갤럭시와의 연구협력을 통해 초거대 시를 바탕으로 글로벌 신약 설계 분야 선도 목표

#### 난제 해결을 위해 전통적인 과학 분야에 AI 적용

- (반도체) 구글은 반도체 칩 설계 시 인공지능을 적용하여, 수개월 소요되는 배치 작업을 6시간 이내에 수행('21)
  - ※ TPU 반도체 칩 설계 시간 단축(수개월→6시간)
- (대기과학) 기상 예측 분야에서 레이더 데이터 기반 인공지능 기술을 통해 기존 방법론 보다 우수한 단기예측 성능 달성('21)

#### 연구개발 분야에 디지털 기술 활용 사례



출처 : 디지털 전환 시대, NIA가 전망한 환경 변화 13대 이슈(한국지능정보사회진흥원)

### AI를 통한 난제 해결의 가능성을 확인하고

2023-10-31 새로운 시장 개척, 다양한 분야 난제에 AI를 적용하기 시작

29

## < 인공지능(Artificial Intelligence) 개발하는 빅테크 >

### AI 개발과 공개에 속도 내는 빅테크



개발한 AI 전체 공개에 소극적이던 빅테크

구글	작년 공개한 AI 언어 모델 람다2, 테스트 모드 유지
딥마인드	시챗봇 스페로우, 올해 중 비공개 베타 형태로 공개 예정
메타	작년 11월 공개한 과학문헌 생성 AI 갈럭티카, 3일 만에 폐쇄

### 구글

- 올해 20개가 넘는 AI 제품 출시 계획
- AI 제품에 대해 신속하게 윤리적 검토를 하는 그린 레인 도입 검토

### 구글의 AI 원칙

- 1 사회적으로 유익해야 한다
- 2 불공정한 편견을 만들거나 강화하지 않는다
- 3 안전성을 우선으로 설계되고 테스트돼야 한다
- 4 인간을 위해 책임을 다해야 한다
- 5 개인정보 보호를 위한 설계 원칙을 적용한다
- 6 과학적 우수성에 대한 높은 기준을 유지한다
- 7 구글의 AI 원칙에 부합하는 용도에 활용돼야 한다

자료-구글

### 마이크로소프트








- 오픈AI에 100억달러(업계 추정) 투자
- 자사 검색엔진 '빙'에 챗GPT 도입 추진

### 메타

- AI 기술 적용 승인 절차 가속화 논의

## < 멀티 모달 AI(Multi Modal AI) 이용한 서비스들 >

### 인공지능(AI)을 활용한 IT 서비스 사례들

 <b>GPT</b> 오픈AI(미국)	MS 인터넷 검색·엑셀·이메일 등과 결합해 작업 편의성 개선
 <b>람다</b> 구글(미국)	문장 생성·채팅용 AI, 구글 검색 기능과 결합해 인간과 대화
 <b>워드툰</b> AI21 Labs(이스라엘)	어조·상황에 맞게 자동 교정하고 문단 재배치 등 글 첨삭
 <b>코히어</b> 코히어(캐나다)	사이트·앱 위한 채팅용 AI 제작. 댓글·채팅 스스로 분석·관리
 <b>하이퍼클로바</b> 네이버(한국)	쇼핑 홍보 문구부터 상품 선택까지 해결. AI 스스로 코딩
 <b>칼로</b> 카카오(한국)	AI 화가·일러스트레이터. 소재·화풍까지 지정해 이미지 제작
 <b>웨인힐스 솔루션</b> 웨인힐스브라이언트AI(한국)	복잡한 설명서·약관 등을 영상으로 자동 변환하는 AI

자료=각 사

## < 인공지능(Artificial Intelligence) 을 이용한 서비스들 >

- 검색, 패션, 이커머스에 활용되는 생성테크
  - 국내 IT 기업들도 생성테크를 개발하면서 자사 서비스에 활용하고 있음
  - 2023년 1월 카카오는 이미지 생성 AI '칼로'를 공개
    - 이미지 1억8000만장 규모의 데이터를 학습해 빛, 그림자까지 조절해가면서 정교하게 이미지를 만들어냄
  - 네이버가 2022년 슈퍼컴퓨터와 네이버의 방대한 데이터를 바탕으로 만든 AI '하이퍼클로바'는 실제 현재 네이버 서비스에도 적용되어 있음
    - 예컨대 네이버쇼핑에서 열리는 특별 할인전 일부는 AI가 처음부터 기획해서 만든 것임
    - 상품 선정, 할인전 홍보 문구까지 AI가 MD(상품기획자)처럼 활약함
  - LG AI연구원은 자사 AI '엑사원'을 활용해 디자인 전용 플랫폼인 '엑사원 아틀리에'를 공개
    - 문장, 단어와 결합된 고해상도 이미지 3억5000만여장으로 학습한 이 생성 AI는 디자이너의 아이디어를 시각화하는 기술을 갖고 있음
    - 2022년 뉴욕 패션위크에선 패션 디자이너 박윤희와 손잡고 엑사원을 기반으로 디자인한 옷을 공개

## < 인공지능(Artificial Intelligence) 을 이용한 서비스들 >

전 세계 헬스케어 분야  
생성형 AI 시장 규모

2022년 10.7억      2032년 217.4억달러



헬스케어로 영토 넓히는 테크 기업들

**엔비디아**

AI 신약 개발 회사 '리커션'에 5000만달러 투자

**구글**

병원에서 의료용 생성형 AI '메드-팜2' 테스트

**마이크로소프트**

환자에게 자동 메시지 보내는 챗GPT 개발

**애플**

애플워치 시와 데이터 활용해 건강 코칭 서비스 개발

**삼성전자**

갤럭시 워치 바이오 액티브센서로 건강 데이터 분석

**LG전자**

AI 수면 테크 기업 '에이슬립'과 연구 협력

**네이버**

AI가 환자 진찰 내용 의료 용어로 자동 변환 및 기록

자료=각 사

## < 인공지능(Artificial Intelligence)

: '초거대(Hyper scale) AI' - '언어→이미지 변환 기술' >

<독도에서 한가롭게 낚시>



<제주도에서 한가롭게 낚시>



**< 인공지능(Artificial Intelligence)  
: '초거대(Hyper scale) AI'-'언어→이미지 변환 기술' >**

- **빠르게 확산하는 언어-이미지 변환 AI**
  - 사람이 언어를 입력하면 AI가 그림을 그리는 기술은 2020년 미국의 **오픈 AI**가 **답러닝(심층학습) 언어 모델인 GPT-3**를 개발하면서 시작됐음
  - 오픈AI는 GPT-3를 기반으로 2021년 첫 번째 언어-이미지 변환 AI 모델인 '**달리 1'(DALL-E1)**을 내놓았고, 2022년 1월 이를 개선한 '**달리 2'**를 공개했음
    - 수백만개 이미지와 언어를 학습한 AI가 제시된 언어의 의미를 파악하고, 이미지를 이루는 단위인 픽셀 값을 조절해 언어에 맞는 그림을 만드는 원리임
  - 2022년 5월 **구글**은 좀 더 사실적인 이미지를 만들 수 있는 언어-이미지 변환 모델인 **이매젠(Imagen)**을 공개
  - 2022년 7월 **메타**도 **메이커스**이라는 AI 모델을 공개했음
  - 2022년 6월 **LG**는 컴퓨터 비전 학회인 'CVPR'에서 언어-이미지, 이미지-언어 양방향 변환이 가능한 **초거대 AI인 엑사원**을 공개했음
    - 김승환 LG AI 연구원 비전랩장은 "GPT-3가 언어 변환 모델의 가능성을 열었고, 이후 언어-이미지 전환과 관련한 연구가 꾸준히 이어졌다"며 "현재는 많은 업체가 이미지 인식과 구현 등 AI 이전에 특화된 모델을 고도화하고 있다"고 설명

2023-10-31

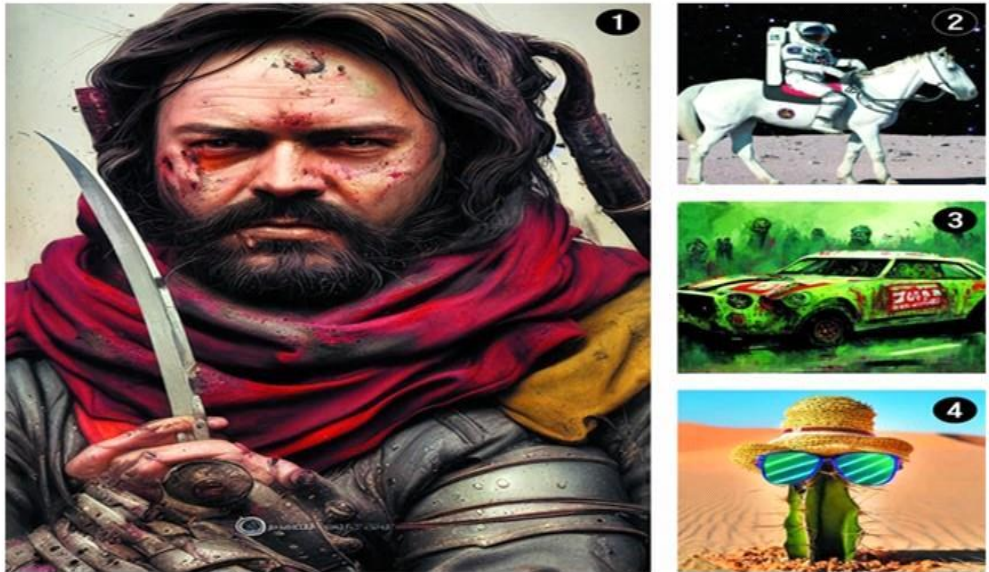
2023년 김용환교수 특강

35

**< 인공지능(Artificial Intelligence)  
: '초거대(Hyper scale) AI'-'언어→이미지 변환 기술' >**

<b>잇따라 등장하고 있는 언어-이미지 변환 AI 모델</b>	
	자료=각사·업계
<b>오픈AI 달리 (DALL·E·미국)</b>	달리1, 달리2를 통해 언어-이미지 변환 모델 최초 공개
<b>구글 이매젠(미국)</b>	사실적인 이미지 생성 능력
<b>메타 메이커스인 (미국)</b>	간단한 텍스트로 이미지의 특정 주요 부분을 세밀하게 구현
<b>LG 엑사원(한국)</b>	세계 최초 언어-이미지, 이미지-언어 양방향 가능 모델
<b>틱톡 AI그린스크린 (중국)</b>	단순하고 추상적 이미지 생성해 틱톡 영상 배경으로 사용
<b>미드저니(미국)</b>	채팅 앱 디스코드 기반으로 회화적인 이미지 생성에 특화

**< 인공지능(Artificial Intelligence) >**  
**: '초거대(Hyper scale) AI' - '언어→이미지 변환 기술' >**



최근 언어-이미지 변환 AI 모델이 내놓는 결과물은 품질이 상당히 높다. ① 스테이블 디퓨전  
 에서 중세, 도적 두목 등 키워드 20여 개를 넣어 만든 그림. ② 달에서 말을 탄 우주비행사라  
 는 키워드를 바탕으로 DALL·E2가 그린 작품. ③ 미드저니에서 중비, 자동차, 레이싱이라는 키  
 워드를 넣어 만든 그림. ④ 구글 의매젠이 만든 밀짚모자를 쓴 선인장 그림. /스태이블 디퓨전·  
 오픈AI·미드저니·구글

**< 인공지능(Artificial Intelligence) >**  
**: '초거대(Hyper scale) AI' - '언어→이미지 변환 기술' >**



AI 기반 이미지 생성 프로그램 '미드저니'로 제작한 가상의 아파트 건축 사진 /마나스 바티아 AI 이미지 생성 프로그램 '달리'로 표지를 만든 잡지 '코스모폴리탄' 6월호 /코스모폴리탄

## < 인공지능(Artificial Intelligence) >

: '초거대(Hyper scale) AI' - '언어→이미지 변환 기술' >



2022년 6월 미드저니를 활용해 표지를 장식한 영국 주간지 '이코노미스트'.

2023-10-31

2023년



미드저니로 그림을 완성한 만화 'Goats'의 일부. /앨비스 딘

## < 인공지능(Artificial Intelligence) >

: '초거대(Hyper scale) AI' - '언어→이미지 변환 기술' >



구글이 개발한 '이미젠(Imagen)' 모델로 작업한 이미지 예시 (출처 : 구글)

**< 인공지능(Artificial Intelligence) : '초거대(Hyper scale) AI' - '언어→이미지 변환 기술' >**



엑사원은 텍스트와 이미지, 이미지와 텍스트 등 모달리티 간 양방향 사고가 가능하다. (출처 : L.G. AI연구원)

**< 네이버 HyperCLOVAX : '초거대(Hyper scale) AI' >**

**초거대 AI '하이퍼클로바X' 사용자별 서비스**    자료: 네이버

**HyperCLOVAX**

**개인 (일반 사용자, 창작자 등)**

- 클로바X
- 큐:(CUE:)
- 글쓰기를 위한 클로바 (CLOVA for Writing)
- 하이퍼클로바X 버티컬 서비스

**판매자 · 광고주**

- 클로바X 스마트스토어센터
- 커머스 AI 솔루션
- 클로바X 스마트플레이스
- 광고를 위한 클로바 (CLOVA for AD)
- 클로바X 광고주센터

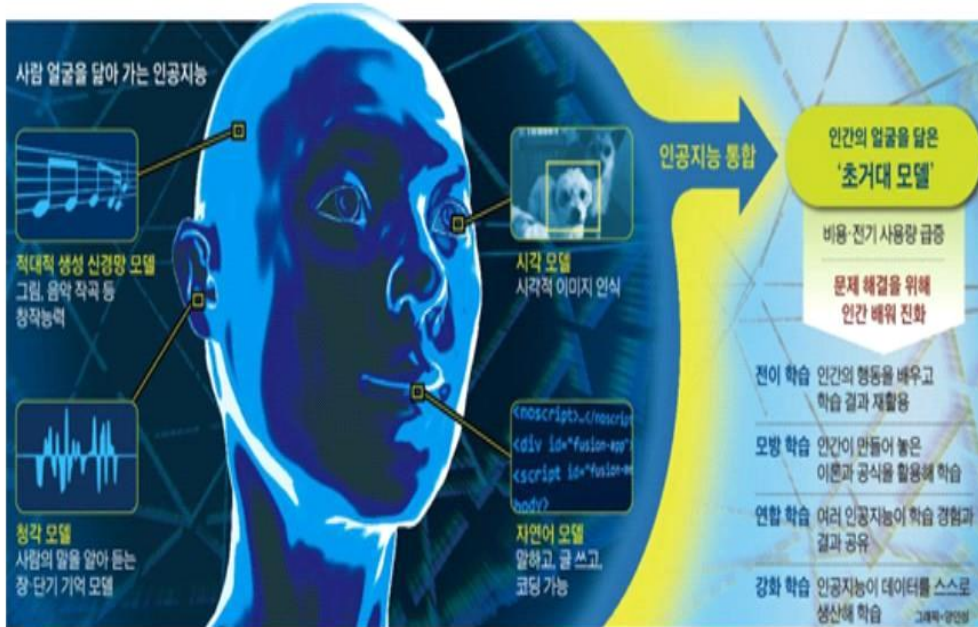
**기업**

- 클로바X
- 클로바 스튜디오
- 프로젝트 커넥트X
- 하이퍼클로바X를 위한 뉴로클라우드 (Neurocloud for HyperCLOVAX)

2023-10-31

42

## < 인공지능(Artificial Intelligence) : '초거대(Hyper scale) AI' >



## 인공지능 비즈니스 혁신 사례 => 인공지능을 통한 디지털 경영전략들 : SCM과 CRM 등에 AI와 디지털 투자에 기업들이 집중

- 업무 생산성의 효율화 및 비용절감 지능형 경영전략
- 디지털 전환에 따른 최신 기술 트렌드 반영 지능형 경영전략
- 업무 프로세스 및 비즈니스 모델 혁신 지능형 경영전략
- 기존 서비스 제품의 매출 및 판매 확대 지능형 경영전략
- 고객 만족도 향상과 신규 판로 개척 지능형 경영전략
- 새로운 제품 및 서비스 개발 지능형 경영전략
- 보안 대책 강화 및 기타 법률 대응 지능형 경영전략
- 종업원 만족도 향상과 직장의 활성화 지능형 경영전략
- 신규 사업 기획 지능형 경영전략

## < 기술혁신: 인공지능 기술혁신 발전 트렌드 전망 >

시기	주요 변화
1950년대	AI 개념 정립
1970년대 중반	1차 암흑기
1980년대 후반	2차 암흑기 - Microcomputer를 중심으로 초기 인공지능경망, 문자판독, 신 용도 분석 등 - Low cost server & PC를 중심으로 마케팅 분석, 거짓말 탐지기, 감성분석, 네트워크 로드 분석 등
2010년	Mobile & Cloud를 중심으로 딥러닝 - 자연어 처리, 비디오 라벨링, 큐레이션, 예측 유지 보수, 공 급망 관리 등
2011년	왓슨이 인간과 대결한 퀴즈 제작쇼에서 우승
2016년 3 월	구글의 딥마인스가 개발한 알파고 바둑 우승: 인공지능 바둑 소프트웨어 알파고와 이세돌 9단의 대결
2017년 1 월	알렉사 CES장악
2017년 12 월	알파고 제로 등장, 소프트웨어 공개
현재	인공지능의 다양한 기술진보와 변화가 진행 중
2024년	AI집 사용화
2025년	유사 진이 학습 AI
2026년	뇌기계(BMI) 인터페이스
2029년	맥라기만 상황인지/ 상황변화 적용 학습
2030년	비진도 학습 AI
2045년	AI 기술은 AI가 인간의 지능을 뛰어넘는 성공레리티 (Singularity 특이점) <sup>1)</sup>

### ‘특이점(Singularity)’이란?

- 인간의 지능을 모방한 AI가 온라인 등 공간에 누적된 방대한 데이터를 자율적으로 해석해 새로운 이성과 의식을 생성, 인간 이상의 지능을 발휘하는 순간을 의미
- 미국의 수학자 존 폰노이만이 처음 언급한 용어로, 미래학자이자 현재 구글 이사인 레이 커즈와일이 2005년 「특이점이 온다」는 책에서 ‘2045년 AI가 인간의 지능을 능가할 것’이라는 전망을 내놓으면서 대중적으로 사용되기 시작
- 커즈와일은 AI가 튜링테스트(컴퓨터가 생각이 있는지 판정하는 테스트)를 통과하는 시점을 2020년대 말로 예상
- 2022년 11월 ChatGPT가 공개되면서, 새로운 혁신의 시발점으로 주목

2023년 김용환교수 특강

45



2023-10-31

2023년 김용환교수 특강

46

## < AI 기술혁신 사례 : ChatGPT와 AlaaS 비즈니스 혁신 >



- 자연스러운 대화가 가능한 인터페이스 AI
- 다국어 배움 AI
- 멀티모달 AI(텍스트=>그림 및 영상)
- ChatGPT, 세계에서 24번째로 많이 방문하는 웹사이트
- 'ChatGPT 날개 단' 오픈AI 연 매출 1년 새 376억→1조7천억원 급성장

### 챗GPT

개발사	오픈AI	
출시일	2022년 11월 30일	
내용	대화형 AI 챗봇	
일 사용자 수	1000만명	
성능	미 의사면허시험 통과 수준 로스쿨 졸업시험 평균 C+ 학점	
자료>오픈AI 공개	와튼스쿨 MBA 기말시험 B 학점 수준	

#### 챗GPT 같은 생성AI가 위협하는 직업군

 그래픽 디자이너	 금융 애널리스트	 교사 등 교육자
 기자 포함 콘텐츠 크리에이터	 낮은 수준의 코딩 설계 엔지니어	

## Digital Transformation과 AI 혁신

- 2022년 11월 30일 ChatGPT 등장
  - 세계 최대 AI 연구소 '오픈AI'가 만든 대화(chat·챗)에 특화된 AI. 머신러닝(기계 학습)을 통해 인간의 언어와 지식을 습득해, 인터넷 채팅하듯 질의응답을 주고받을 수 있음
  - 오픈AI는 일론 머스크 테슬라 창업자와 유명 투자자 샘 올트먼이 2015년 세웠고, 2018년 GPT(Generative Pre-trained Transformer)의 첫 버전을 공개, 2021년 GPT 3.0을 개발하여 공개
- 생성 AI(Generative AI)
  - 글, 문장, 오디오, 이미지 같은 기존 데이터를 활용해 새로운 콘텐츠를 만들어 내는 AI
  - 기존 데이터를 학습하여 대화에 특화된 'ChatGPT'도 생성 AI의 일종임
    - GPT 3.5를 기반으로 구성된 ChatGPT는 2021년 데이터만을 갖고, 대화에 응답
  - 입력한 문장을 토대로 새로운 그림을 그려주는 '미드저니'와 'DALL-E(달리)'도 있음

## Digital Transformation과 AI 혁신

- 챗GPT 같은 생성 인공지능(Generative AI)은 훈련과 추론 등 데이터를 처리하는 데 대규모 병렬 컴퓨팅 시스템을 필요로 함
  - 중앙처리장치(CPU)는 순서대로 데이터를 처리하는 데 반해, 그래픽처리장치(GPU)를 사용하면 대규모 데이터를 동시에 처리할 수 있음
  - 데이터양이 많을수록 데이터 처리 시간은 병렬 컴퓨팅 시스템이 가능한 GPU가 훨씬 더 빨라짐
- 최근 챗GPT 부상으로 엔비디아가 주목받는 이유
  - 반도체 칩 아키텍처(Architecture: 컴퓨터 시스템 전체의 설계 방식) 설계와 관련 소프트웨어 개발이 주요 사업인 엔비디아는 1999년 세계 최초로 GPU를 개발, 병렬 컴퓨팅 시스템으로의 변혁을 주도했음
  - AI 반도체 주류인 GPU 시장에서 엔비디아의 시장 점유율은 90% 이상임
  - 오픈AI의 채팅형 AI인 챗GPT 훈련에 엔비디아 AI 반도체 칩인 'A100' GPU 1만 개가 사용됐다는 소식이 알려지면서, 주가가 2023년 2월 14일 기준 연초 대비 60% 이상 급등했음

2023-10-31

2023년 김용환교수 특강

49

## 2022년 11월 30일 OpenAI가 공개한 ChatGPT : 다양한 활용

	Industries								
	Automotive and Vehicle Manufacturing	Media	Architecture and Engineering	Energy and Utilities	Healthcare Providers	Electronic Product Manufacturing	Manufacturing	Pharmaceutical	
Drug Design								●	
Material Science	●			●		●			
Chip Design						●			
Synthetic Data	●		●	●	●	●	●	●	
Generative Design (Parts)	●		●				●		

\* 자료: Beyond ChatGPT: The Future of Generative AI for Enterprises (Gartner, 2023)

## 2022년 11월 30일 OpenAI가 공개한 ChatGPT : 다양한 활용

### 챗GPT 플러그인을 통한 다양한 활용

음식 조리법 검색 및 장보기  
(온라인 장보기 플랫폼 '인스타카트' 플러그인 이용)

#### 사용자

"볼로네제 스파게티 레시피 알려줘" 라고 입력

#### 챗GPT

"올리브 오일을 팬에 두른 다음..." 조리법 전달  
→ "여기 볼로네제 스파게티에 필요한 재료를  
장바구니에 넣어줬습니다"  
→ 모든 재료가 적정량 장바구니에 담겨 있는  
웹페이지가 뜬  
→ 결제하면 인근 쇼핑몰에서 배달

### 호텔 검색 및 예약

(여행 예약 플랫폼 '카약' 플러그인 이용)

#### 사용자

9월 1일 이탈리아 로마에서 3성급 이상 호텔, 시내  
중심까지 10분 거리, 하루 30만원 이하 호텔을 알려줘

#### 챗GPT


조건에 맞는 다섯 호텔 명단과 사진 띄워줌  
→ 원하는 호텔 클릭하면 예약 페이지로 넘어가  
예약 진행

그래픽=박상훈


## < ChatGPT vs 하이퍼클로바X(서치 GPT) 비즈니스 혁신 : AIaaS >

### 챗GPT vs 하이퍼클로바X 차이

 고속도로에서 오토바이를 타고 되나요?

 **챗GPT(오픈AI·미국)**  
일반적으로 고속도로에서 오토바이를  
타는 것은 가능합니다.  
그러나 이에 대한 규제는 국가 및  
지역에 따라 다를 수 있습니다.(중략)

특징 영어권 데이터 학습, 한국 문화·법률에 대한  
이해 부족

 **하이퍼클로바X(네이버)**  
한국에서는 고속도로와 자동차 전용  
도로에서 오토바이 통행이 금지돼  
있습니다.(중략)

특징 챗GPT보다 한국어 6500배 더 학습,  
한국에 대한 풍부한 이해

◇ 하이퍼클로바X는 네이버가 개발한 초거  
대 AI이고, 서치GPT는 이를 네이버 검색에  
접목한 것

- 한국말 잘하는 '하이퍼클로바X', 두뇌는 삼성과 협업
- 네이버가 2023년 7월에 선보일 '하이퍼클로바X'는 지  
난 2021년 네이버가 세계에서 세 번째로 내놓은 초거  
대 AI '하이퍼클로바'의 업그레이드 버전임
- 이를 자체 서비스에 순차적으로 접목하고 일반 기업  
들도 이를 빌려다가 맞춤형 AI로 개조해 쓸 수 있도록  
하겠다는 것이 네이버의 계획
- 네이버는 누구든 자체적으로 보유한 데이터를 학습시  
키면, 하이퍼클로바X를 특정 기업이나 서비스에 최적  
화된 '맞춤형 초거대AI'로 만들 수 있다고 설명
- 해외 AI와 차이점은 '한국어에 특화된 토종 AI'라는  
점

현 김용환교수 특강

52

## OpenAI가 공개한 ChatGPT와 연계한 MS Bing Chat

새 Bing에 오신걸 환영합니다!

웹용 AI 기반 Copilot

### 😊 복잡한 질문하기

"주황색 음식만 먹는 까다로운 입맛의 유아를 위해 어떤 음식을 만들 수 있을까요?"

### 👤 더 나은 답변 받기

"가장 잘 팔리는 반려동물 청소기 3개의 장단점은 무엇인가요?"

### 📄 텍스트를 함께 다시 쓰기

페이지에서 텍스트를 선택하여 함께 수행할 수 있는 작업 확인하기

2023-10-31

53

## 구글의 대화형 생성AI 모델 바드(bard): 다양한 활용

🌟 창의적이고 유용한 당신의 파트너 Bard입니다. 아직 완벽하지는 않지만, 의견을 보내주시면 앞으로 더 발전해 나가겠습니다.

어떻게 시작해야 할지 잘 모르시겠다면 다음 예를 참고해 보세요.

3개월간의 불어 공부 계획을 세워 줘

1월 중순 에든버러에서 다섯 살 남자 아이와 함께할 4박 5일 여행 일정을 세워 줘

올해 패션 트렌드를 알려 주고, 그 패션 트렌드에 맞춰 꼭 사야 할 아이템 추천해 줘

[여기에 프롬프트 입력]



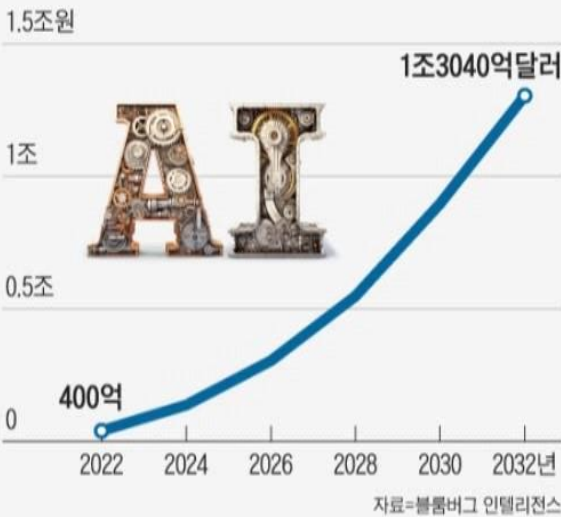
Bard가 부정확하거나 불리감을 주는 정보를 표시할 수 있으며, 이는 Google의 입장을 대변하지 않습니다.

## < 생성형 인공지능(Artificial Intelligence)시장의 발전 전망 >



## < 생성형 인공지능(Artificial Intelligence)시장의 발전 전망 >

### 생성형 AI 시장 규모 전망



2023년 6월 인공지능(AI) 'ChatGPT' 열풍으로 장기적인 AI 붐이 조성되면서, '생성형 AI' 시장이 10년 후인 2032년에 1조3천억 달러(약 1천705조 원) 규모로 커질 것으로 전망됐다고 블룸버그통신이 보도

- 2022년 시장의 규모가 400억 달러(약 52조5천억 원)인 점을 감안하면 10년간 30배 넘게 급증
- 블룸버그 산하 경제연구소 블룸버그 인텔리전스의 애널리스트 만디프 싱이 주도한 새 보고서에 따르면, 이 분야는 AI 시스템 학습에 필요한 인프라와 AI 기기, 광고 등의 수요 등으로 10년간 매년 42% 확장할 것으로 예상됨
- AI가 지원하는 디지털 광고 부문은 2032년 연간 매출이 1천920억 달러(약 252조 원), AI 서버 매출은 1천340억 달러(약 176조 원)에 각각 달할 것으로 보고서는 예상

## < 생성형 인공지능(Artificial Intelligence)시장의 발전 전망 >

### 생성형 인공지능(AI) 서비스 내놓는 국내 기업

**LG CNS**

- 빅테크 개발한 대규모언어모델(LLM) 탑재한 생성형 AI 서비스 플랫폼
- 마이크로소프트·구글클라우드와 협력, 앤스로픽에 지분 투자
- 생성형 AI 관련 전문인력 600명, 별도 추진단 구성

**SK C&C**

- 기업 맞춤형 보고서 제작 솔루션 개발
- 네이버클라우드와 협력해 기업 대상 '한국형 AI 서비스' 공동 개발
- 올해 초 생성형 AI 테스트포스 신설

**삼성SDS**

- 기업이 원하는 다양한 LLM 결합
- 업무 생산성 높일 생성형 AI 플랫폼 2종 공개

자료=각 사

## < 생성형 인공지능(Artificial Intelligence)시장의 발전 전망 >

### 생성형 AI 활용해 광고 제작하는 국내 기업들



기업명 **롯데리아**

활용 내용 AI를 활용해 새우버거 이미지를 비트로 (시기) 비꾼 뒤 이를 기반으로 음원 제작(올 8월)

조회 수 **182**만회



기업명 **LG유플러스**

활용 내용 이미지 생성 AI '스테이블 디퓨전'으로 (시기) 광고 속 영상 제작(7월)

조회 수 **1263**만회

## < 생성형 인공지능(Artificial Intelligence)시장의 발전 전망 >



기업명 **삼성생명**

활용 내용 AI 음악 작곡 프로그램 활용해 배경음악 만들고 이미지 생성 AI '미드저니' 활용해 광고 속 이미지를 제작(6월)

조회 수 **783**만회

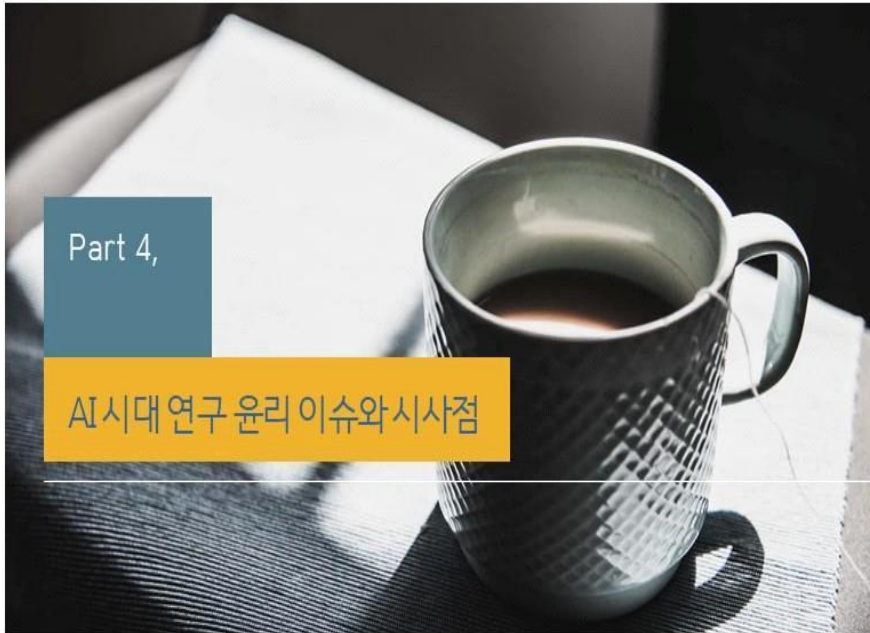


기업명 **배스킨라빈스**

활용 내용 챗GPT 에 '구름 속 시나모롤'을 주제로 한 광고 제작을 요청한 뒤 시나리오 각색해 광고 제작(5월)

조회 수 **508**만회

자료=각 사



## < 인공지능(Artificial Intelligence)의 부작용과 AI윤리(2017) >

AI 윤리 관련 기초 설명

AI 윤리

### 인공지능 윤리 주요 이슈

- (구글) 확인을 고릴라로 오인식하는 등 인종에 대한 차별 발생 (2015)
- (MS) 챗봇 '테이'는 인종 차별 등 각종 혐오 발언으로 인해 중단(2016)  
\* "미국-멕시코 국경에 차단 벽을 설치하고, 멕시코가 그 비용을 내야죠."



구글, 확인 사진 오인식



MS, 챗봇 '테이'

- (테슬라) 오토파일럿 모드 중 테슬라S 충돌로 인명 피해 발생 (2016)
- (美 COMPAS) 재범 예측 프로그램의 인종에 따른 편향 발생 (2018)  
\* 흑인의 재범율이 백인의 재범율보다 높게 예측하는 인종 편향 발생



테슬라, 자율주행차 사고



미국, 재범예측 시 인종별 편향

### 아실로마 인공지능 원칙('17)에서 윤리 이슈 언급

- (아실로마 AI 원칙) 각종 AI 윤리와 관련된 사고들이 발생함에 따라, 미국의 비영리 단체 future of Life institute 에서 AI 원칙 수립  
\* 컨퍼런스 개최 도시 이름을 따서, 아실로마 시 원칙으로 명명

- 아실로마 AI 원칙은 연구 이슈(5개), 윤리적 가치(13개), 장기적 이슈(5개) 등 세가지 범주에서 AI에 대한 경각심을 일깨우고, 향후 AI 개발을 위한 발전 방향 등을 제시



- \* 스티븐 호킹, 일론 머스크, 데미스 하사비스 등 2,000여명의 과학/기술인이 아실로마 AI 원칙에 지지 서명 남김

- 이를 통해 본격적으로 AI 윤리 분야의 중요성을 전 세계에 알렸으며, AI 개발 과정에서 발생할 수 있는 실패와 자유의 침해에 대해 책임성 있는 행동을 강조

\* 인간의 가치, 안전 등 총 13개 항목으로 구성

- |           |             |               |
|-----------|-------------|---------------|
| 1 안전      | 6 인간의 가치    | 11 인간 통제      |
| 2 실패의 투명성 | 7 개인정보보호    | 12 사회 진척 방지   |
| 3 사법적 투명성 | 8 자유와 프라이버시 | 13 인공지능 무기 경쟁 |
| 4 책임성     | 9 이익의 공유    |               |
| 5 가치 일치   | 10 변명의 공유   |               |

61

## < 인공지능(Artificial Intelligence) 윤리 >

AI 윤리

### [사회] 전 세계적으로 AI 윤리 원칙 선언

- 2017 01. 아실로마 AI 원칙 \*글로벌·국내
- 2018 12. EU, 신뢰할 수 있는 윤리 가이드라인
- 2019 03. 일본, 신뢰할 수 있는 윤리 가이드라인  
05. OECD, 인공지능 원칙  
11. 방통위, 인공지능 윤리원칙
- 2020 12. 인공지능 윤리 기준
- 2021 04. EU, 인공지능 법안 제안  
04. 과기정통부, 신뢰할 수 있는 AI 실현전략  
07. 금융분야 AI 가이드라인  
11. 유네스코 AI 권고안
- 2022 01. 과기정통부, 신뢰할 수 있는 인공지능 개발 안내서  
05. 안건위, 인공지능 개발과 활용에 관한 인권 가이드라인

### [산업] AI 서비스 내 윤리 탑재 노력

#### 글로벌

- 구글 AI 원칙 발표('18)
- MS AI 윤리 가이드라인 발표('17)
- IBM IBM AI 윤리 원칙('17)

#### 국내

- 네이버 AI 윤리 준칙 선언('21)
- 카카오 기술 윤리위원회 출범('22)
- LG AI 연구원 AI윤리점검 TF('22)
- 삼성 전자, SDS 등 AI 윤리 원칙 발표

### [학계] AI 윤리 '원칙'을 넘어 '실제적 방안' 연구 중

- 원칙 발표는 감소하나, 실무적인 윤리 연구는 점차 증가 중



AI의 우수성 부각 등 AI 시대로의 전환이 본격화되는 한편, AI의 부작용 해소를 위한 국제사회 논의 및 연구 활발

2023-10-31

62

## < 인공지능(Artificial Intelligence) 윤리 >

### 세계 각국의 AI 규제 법안

#### 법안 내용

미국 상원	생성형 AI가 만든 콘텐츠에 대해 사업자가 법적 책임을 지도록 함.
EU 의회	<ul style="list-style-type: none"> <li>• 생성형 AI가 불법 콘텐츠를 만들지 않도록 기업이 관리해야 함.</li> <li>• AI가 창작물을 만들 때 출처를 표기해야 함.</li> </ul>
중국	<ul style="list-style-type: none"> <li>• 생성형 AI 서비스 제공자가 지식재산권 침해와 허위정보 생성 등 부작용을 막기 위한 조치를 다해야 함.</li> <li>• 생성형 AI로 만든 콘텐츠에는 사회주의 핵심 가치를 반영해야 함.</li> </ul>
한국(국회 과방위)	AI산업에 대한 '우선 허용, 사후 규제' 원칙 명시. 시민안전 위협 기능 영역을 '고위험' 영역으로 명시, 부작용 예방.

\*자료: 각국 의회 등 자료 취합

## < 인공지능(Artificial Intelligence)시대 연구윤리 이슈들 >

- AI시대의 연구윤리 이슈 발생 배경
  - 새로운 기술적 특성으로 인한 윤리적 문제의 발생
    - AI는 기존의 기술과는 달리, 인간의 지능을 모방하고 학습할 수 있는 능력을 가지고 있음
    - 이러한 능력으로 인해 AI는 기존의 기술과는 차원이 다른 윤리적 문제를 야기할 수 있음
    - 예를 들어, AI는 자율주행 자동차나 무인 무기 시스템의 개발에 사용될 수 있는데, 이러한 시스템이 잘못 작동할 경우 심각한 인명 피해를 초래할 수 있음
  - 기술의 발전 속도가 빠른 속도로 진행됨에 따른 윤리적 고려의 어려움
    - AI 기술은 매우 빠른 속도로 발전하고 있습니다. 이러한 기술의 발전 속도는 윤리적 고려를 따라잡기 어렵게 만들고 있음
    - 예를 들어, AI는 최근 몇 년 동안 놀라운 발전을 이루었지만, 아직도 AI의 잠재적 위험에 대한 충분한 연구가 이루어지지 않은 상태임
    - 사례: 2022년 11월 OpenAI가 제공한 ChatGPT 등 생성AI 모델의 가짜뉴스 생산 등

## < 인공지능(Artificial Intelligence)시대 연구윤리 이슈들 >

### • AI시대의 연구윤리 이슈 발생 배경

- 데이터에 내재된 편향과 그에 따른 차별의 발생 : 데이터 수집/저장/분류/관리/분석/공유/활용 과정에서 발생
  - AI는 학습 데이터에 기반하여 판단을 내리기 때문에, 데이터에 내재된 편향이 AI의 판단에도 그대로 반영될 수 있음
  - 이러한 편향은 차별과 불평등을 초래할 수 있는 위험을 가지고 있음
  - 예를 들어, 얼굴 인식 기술의 경우, 특정 인종이나 성별에 대한 편향으로 인해 잘못된 판정을 내릴 수 있음
  - '디지털의 역설'과 같이 '빅데이터의 역설'이 발생할 가능성 => 데이터 격차와 가짜뉴스 및 편향성으로 'AI시대 새로운 차별화와 불평등 및 편향성 발생 가능성'

2023-10-31

2023년 김용환교수 특강

65

## < 인공지능(Artificial Intelligence)시대 연구윤리 이슈들 >

### • AI시대의 연구윤리 이슈 발생 배경

- AI 기술의 복잡성으로 인한 윤리적 판단의 어려움
  - AI는 인간의 이해를 넘어서는 수준의 복잡성을 가지고 있음
  - 이러한 복잡성은 AI의 위험과 부작용을 제대로 이해하기 어렵게 만들고, 윤리적 판단을 내리는데 어려움을 야기할 수 있음
  - 예를 들어, AI 시스템의 오류가 발생한 경우, 그 책임을 누구에게 물을 것인지에 대한 명확한 기준이 없는 경우가 많음
  - 예를 들어, AI가 만들어내는 뉴스나 콘텐츠의 경우, 인간이 사실과 허위를 구분하기 어려울 수 있음
- AI 기술의 사회 전반에 미치는 영향으로 인한 윤리적 고려의 중요성
  - AI 기술은 사회 전반에 큰 영향을 미칠 수 있는 잠재력을 가지고 있음
  - 이러한 영향으로 인해 AI 기술의 윤리적 고려가 더욱 중요해지고 있음
  - 예를 들어, AI 기술이 대규모 실업과 같은 사회적 문제를 야기할 수 있다는 우려가 제기되고 있음
  - '기술의 노예를 전락하는 인간 발생 가능성?'

2023-10-31

2023년 김용환교수 특강

66

## < 인공지능(Artificial Intelligence)시대 연구윤리 이슈들 >

### • AI시대의 연구윤리 이슈 포인트

#### - 편향성

- AI 시스템은 학습 데이터에 내재된 편향을 반영할 수 있음
- 이러한 편향은 AI 시스템의 판단에 부정적인 영향을 미칠 수 있음
- 예를 들어, AI 시스템이 고용이나 대출 등과 관련된 결정을 내릴 때, 특정 인종이나 성별에 대한 편향을 반영할 수 있음

#### - 책임성

- AI 시스템이 오류를 범하거나 피해를 야기한 경우, 그 책임을 누구에게 물을 것인지에 대한 명확한 기준이 없는 경우가 많음
- 예를 들어, 자율주행 자동차가 사고를 낸 경우, 자동차 제조사, 소프트웨어 개발사, 운전자 등 누구에게 책임을 물을 것인지에 대한 논란이 있음

2023-10-31

2023년 김용환교수 특강

67

## < 인공지능(Artificial Intelligence)시대 연구윤리 이슈들 >

### • AI시대의 연구윤리 이슈 포인트

#### - 투명성

- AI 시스템의 판단 과정을 이해하고 설명할 수 있어야 함
- 그렇지 않으면 AI 시스템의 판단에 대한 신뢰를 얻기가 어려움
- 예를 들어, AI 시스템이 범죄자를 예측하는 경우, 그 예측 과정을 투명하게 공개해야 함
- 인공지능 블랙박스 등

#### - 윤리성

- AI 시스템을 개발하고 활용할 때, 인간의 존엄성과 가치를 존중해야 함
- 예를 들어, AI 시스템을 무기나 감시 시스템으로 사용하는 것은 윤리적으로 문제가 될 수 있음: 'AI시대 빅브라드' 부상

2023-10-31

2023년 김용환교수 특강

68

## < 인공지능(Artificial Intelligence)시대 연구윤리 이슈들 >

- **구체적인 연구윤리 이슈의 예시들 : 데이터 수집/저장/분류/관리/분석/공유/활용 과정에서 발생**
  - **데이터 수집 및 처리 과정에서의 윤리**
    - AI는 대규모의 데이터를 학습하는 방식으로 작동하기 때문에, 데이터 수집 및 처리 과정에서의 윤리가 중요함
    - 개인정보 침해, 차별적 편향, 허위 정보의 유포 등이 발생하지 않도록 주의해야 함
  - **AI 시스템의 개발 및 운영 과정에서의 윤리**
    - AI 시스템은 인간의 삶에 광범위한 영향을 미칠 수 있기 때문에, 개발 및 운영 과정에서의 윤리가 중요함
    - 안전성, 공정성, 투명성, 설명 가능성 등을 고려해야 함
  - **AI의 사용 및 활용 과정에서의 윤리**
    - AI는 다양한 분야에서 사용될 수 있으며, 그에 따라 다양한 윤리적 딜레마가 발생할 수 있음
    - 예를 들어, 군사 분야에서의 AI 사용, 의료 분야에서의 AI 사용, 노동 시장에서의 AI 사용 등이 대표적인 예임

2023-10-31

2023년 김용환교수 특강

69

## < 인공지능(Artificial Intelligence)시대 연구윤리 이슈들 >

- **AI시대의 연구윤리 이슈들에 대한 해결을 위한 몇 가지 방안들**
  - **AI 연구윤리에 대한 사회적 논의 활성화**
    - AI 연구윤리 이슈들에 대한 사회적 논의를 활성화하여 다양한 이해관계자들의 의견을 수렴하고, 사회적 합의를 도출해야 함
    - 이를 위해서는 정부, 학계, 기업, 시민사회 등 다양한 분야의 전문가들이 참여하는 협의체를 구성하고, 다양한 의견을 수렴하는 채널과 커뮤니티를 마련하는 것이 필요함
  - **AI 연구윤리 교육 강화**
    - AI 연구자들을 대상으로 AI 연구윤리에 대한 교육을 강화하여, AI 연구자들이 연구 과정에서 윤리적 문제를 고려할 수 있도록 해야 함
    - 또한, 일반 대중을 대상으로 AI 연구윤리에 대한 교육을 실시하여, AI 기술의 윤리적 문제에 대한 인식을 제고해야 함
  - **AI 연구윤리 규범의 마련**
    - AI 연구윤리에 대한 사회적 합의를 바탕으로, AI 연구윤리 규범을 마련해야 함
    - AI 연구윤리 규범은 AI 연구자들이 연구 과정에서 준수해야 할 기준을 제시함으로써, AI 기술의 윤리적 사용을 촉진하는 데 기여할 것임

2023-10-31

2023년 김용환교수 특강

70

## < 인공지능(Artificial Intelligence)시대 연구윤리 : 주요 국가들의 준비 >

- **AI 연구윤리 위원회(AI Ethics Board) 설립**
  - 미국 정부는 2021년 10월, AI 연구윤리 위원회(AI Ethics Board)를 설립했음
  - AI 연구윤리 위원회는 AI 연구윤리 이슈들에 대한 사회적 논의를 활성화하고, AI 연구윤리 규범을 마련하는 역할을 수행
- **AI 연구윤리 교육 강화**
  - 미국 정부는 AI 연구자들을 대상으로 AI 연구윤리 교육을 강화하고 있음
  - 미국 국립과학재단(NSF)은 2022년부터 AI 연구 윤리 교육을 필수화하고 있음
- **AI 연구윤리 규범 마련**
  - 미국 정부는 AI 연구윤리 이슈들에 대한 사회적 합의를 바탕으로 AI 연구윤리 규범을 마련하고 있음
  - 미국 국립인공지능연구소(NIAI)는 2023년 1월, AI 연구윤리 규범 초안을 발표

2023-10-31

2023년 김용환교수 특강

71

## < 인공지능(Artificial Intelligence)시대 연구윤리 : 주요 국가들의 준비 >

- **미국은 2021년 10월, AI 연구윤리 이슈들에 대한 대응을 위해 "AI 연구윤리 원칙"을 발표**
  - 이 원칙은 AI 연구자들이 연구 과정에서 준수해야 할 7가지 원칙을 제시
  - • 공익과 책임성: AI 기술의 개발 및 활용은 공익을 위해 이루어져야 하며, 책임 있게 이루어져야 함
  - • 투명성과 설명가능성: AI 기술의 개발 과정과 결과는 투명하고 설명가능해야 함
  - • 차별금지: AI 기술은 차별을 조장해서는 안 됨
  - • 개인정보보호: AI 기술은 개인정보를 보호해야 됨
  - • 안전성: AI 기술은 안전하게 사용되어야 됨
  - • 책임성: AI 기술의 개발 및 활용에 따른 책임은 명확히 규정되어야 함

2023-10-31

2023년 김용환교수 특강

72

< 인공지능(Artificial Intelligence)시대 연구윤리 : 주요 국가들의 준비 >

- 유럽연합은 AI 시대의 연구윤리 이슈들에 대한 대응을 위해 다음과 같은 제도 개선을 추진
  - AI 규제법(AI Regulation) 제정
    - 유럽연합은 2023년 7월, AI 규제법(AI Regulation)을 제정했음
    - AI 규제법은 AI 시스템의 개발, 사용, 배포에 대한 규범을 마련하고 있음
  - AI 연구윤리 지침(AI Ethics Guidelines) 마련
    - 유럽연합은 AI 연구윤리 이슈들에 대한 사회적 합의를 바탕으로 AI 연구윤리 지침을 마련하고 있음
    - 유럽연합의 AI 연구윤리 지침은 2024년 1월에 발표될 예정

2023-10-31

2023년 김용환교수 특강

73

< 인공지능(Artificial Intelligence)시대 연구윤리 : 주요 국가들의 준비 >

- EU는 2022년 4월, AI 규제법(AI Act)을 발표
  - 이 법은 AI의 개발, 사용, 유통에 대한 규제를 담고 있으며, AI 연구윤리 이슈들에 대한 대응을 위한 다양한 조치를 포함하고 있음
  - • AI 윤리위원회의 설치: AI 윤리위원회를 설치하여 AI 연구윤리 이슈들에 대한 심의 및 자문을 담당하게 함
  - • AI 위험 평가: AI의 개발 및 사용에 앞서 AI 위험을 평가하도록 함
  - • AI 설명가능성: AI 알고리즘은 설명 가능하도록 설계되어야 함
  - • AI 차별금지: AI 기술은 차별을 조장해서는 안 됨
  - • AI 안전성: AI 기술은 안전하게 사용되어야 함

2023-10-31

2023년 김용환교수 특강

74

< 인공지능(Artificial Intelligence)시대 연구윤리 : 주요 국가들의 준비 >

- 중국은 2022년 10월에 "인공지능 발전 규범 (Regulations on the Development of Artificial Intelligence)"을 발표
  - 이 규정은 AI 기술의 개발과 활용에 있어 다음과 같은 윤리적 원칙을 준수하도록 규정하고 있음
  - 인간 중심성: AI 기술은 인간의 이익을 위해 사용되어야 함
  - 안전성: AI 기술은 안전하게 사용될 수 있도록 설계되어야 함
  - 책임성: AI 기술의 개발과 활용에 따른 윤리적 책임은 명확하게 규정되어야 함

2023-10-31

2023년 김용환교수 특강

75

< 인공지능(Artificial Intelligence)시대 연구윤리 : 주요 국가들의 준비 >

- 일본은 2023년 3월에 "인공지능의 윤리적 이용에 관한 기본 지침(Basic Guidelines for Ethical Use of Artificial Intelligence)"을 발표했다
  - 이 지침은 AI 기술의 개발과 활용에 있어 다음과 같은 윤리적 원칙을 제시하고 있음
  - 인간 중심성: AI 기술은 인간의 이익을 위해 사용되어야 함
  - 공정성: AI 기술은 모든 사람에게 공정하게 적용되어야 함
  - 투명성: AI 기술의 개발과 활용 과정이 투명하게 공개되어야 함
  - 책임성: AI 기술의 개발과 활용에 따른 윤리적 책임은 명확하게 규정되어야 함

2023-10-31

2023년 김용환교수 특강

76

## < 인공지능(Artificial Intelligence)시대 연구윤리 : 주요 국가들의 준비 >

- 대한민국은 AI 시대의 연구윤리 이슈들에 대한 대응을 위해 다음과 같은 제도 개선을 추진
  - AI 연구윤리 센터 설립
    - 한국 정부는 2022년 12월, AI 연구윤리 센터를 설립했음
    - AI 연구윤리 센터는 AI 연구윤리 이슈들에 대한 연구와 교육을 수행
  - AI 연구윤리 교육 강화
    - 한국 정부는 AI 연구자들을 대상으로 AI 연구윤리 교육을 강화하고 있음
    - 한국연구재단은 2023년부터 AI 연구 윤리 교육을 필수화하고 있음
  - AI 연구윤리 규범 마련
    - 한국 정부는 AI 연구윤리 이슈들에 대한 사회적 합의를 바탕으로 AI 연구윤리 규범을 마련하고 있음
    - 한국연구재단은 2024년 1월, AI 연구윤리 규범 초안을 발표할 예정

2023-10-31

2023년 김용환교수 특강

77

## < 인공지능(Artificial Intelligence)시대 연구윤리 : 주요 국가들의 준비 >

- 대한민국은 2022년 'AI 윤리 가이드라인'을 한국인터넷진흥원(KISA)에서 발표
  - 이 가이드라인은 AI 연구윤리 이슈들에 대한 대응을 위해 AI 연구자, 기업, 정부 등이 준수해야 할 기준을 제시
  - • 공익과 책임성: AI 기술의 개발 및 활용은 공익을 위해 이루어져야 하며, 책임 있게 이루어져야 함
  - • 투명성과 설명가능성: AI 기술의 개발 과정과 결과는 투명하고 설명가능해야 함
  - • 차별금지: AI 기술은 차별을 조장해서는 안 됨
  - • 개인정보보호: AI 기술은 개인정보를 보호해야 함
  - • 안전성: AI 기술은 안전하게 사용되어야 함

2023-10-31

2023년 김용환교수 특강

78

## < 인공지능(Artificial Intelligence)시대 연구윤리 : 주요 국가들의 준비 >

- 대한민국 AI윤리 가이드라인은 다음과 같은 5가지 기본 원칙과 10가지 구체적 원칙들
- 기본 원칙
  - • 사람 중심
  - • 투명성과 설명가능성
  - • 공정성과 책임성
  - • 안전성
  - • 개방성과 포용성
- 구체적인 원칙
  - • 1. 인간 존엄성 존중/ 2. 안전성 확보/ 3. 프라이버시 보호/ 4. 차별 금지/ 5. 편향 방지/ 6. 책임성 강화/ 7. 개방성과 투명성/ 8. 설명가능성 보장/ 9. 공정성 확보/ 10. 공적 가치 증진

2023-10-31

2023년 김용환교수 특강

79

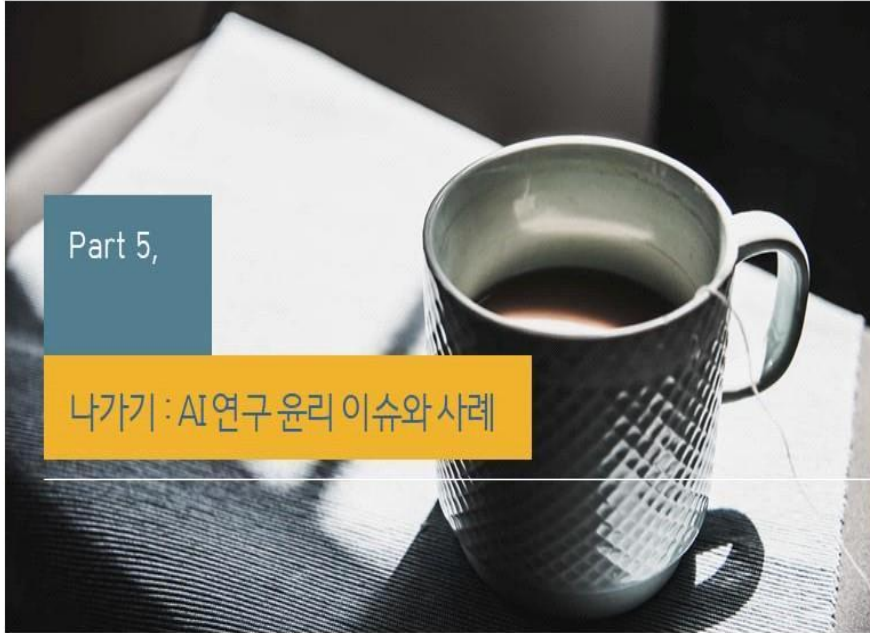
## < 인공지능(Artificial Intelligence)시대 연구윤리 : 주요 국가들의 준비 >

- 2023년 8월 대한민국 보건 의료 분야 연구자가 자율적이고 책임 있는 인공지능(AI) 연구를 할 수 있도록 윤리적 기준을 제시하는 인공지능 연구윤리 지침이 마련
  - 질병관리청 국립보건연구원은 보건 의료 분야 연구자를 위한 '보건 의료 분야 인공지능 연구윤리 지침'을 발간했다고 2023년 8월 17일 공개
  - 이 지침에는 인공지능 연구 설계부터 데이터 생성, 모델 개발, 검증·평가, 적용 및 사후 점검 전 과정에서 발생할 수 있는 △이간의 자율성 존중과 보호 △이간의 행복, 안전, 공공의 이익 증진 △투명성, 설명 가능성, 신뢰성 △책임, 법적 책임 △포괄성, 공정성 △대응성, 지속 가능성 등의 윤리적 질문에 대한 원칙을 제시함
  - 또 응급상황에서의 인공지능 활용과 동의, 의료 인공지능의 질과 안전성, 책보의 환각 효과, 임상결정의 미묘한 경계선과 인공지능, 인공지능과 인종 편향, 인공지능 기술이 가져올 의료인 고용환경 변화 등 구체적 사례를 함께 제시했음

2023-10-31

2023년 김용환교수 특강

80



Part 5,

나가기 : AI 연구 윤리 이슈와 사례

2023-10-31

2023년 김용환교수 특강

81

### < 인공지능(Artificial Intelligence)의 허위 정보와 가짜 뉴스들 >

- 2023년 3월 공개된 GPT4는 기존 버전인 3.5의 기능을 발전시켰고 활용 범위도 넓어졌음
  - 기존 버전에선 문자만 인식했지만 GPT4는 음성과 사진 등 복합적인 정보를 이해하고 정보를 생성할 수 있음
  - △처리할 수 있는 단어의 수가 8배 늘어 약 2만5000개의 단어를 기억할 수 있고 △사실성 평가에서 40% 높은 점수를 받는 등 정확도가 올랐고 △한국어 등 영어가 아닌 언어의 이해 능력이 크게 향상됐고 △비윤리적 발언 등 문제적 발언을 하는 빈도도 낮아졌음
    - <사례> 달걀과 밀가루가 찍힌 사진을 보여줌 "월 만들 수 있을까." 질문을 하면?
      - 팬케이크, 와플, 케이크 등의 음식을 만들 수 있다는 솔루션을 제공
      - 음식별로 요리법을 물어보면 이를 자세히 전달함
  - 오픈AI가 2023년 3월 15일 발표한 대규모 AI 언어 모델(LLM)인 GPT4의 구현 가능한 모습임

2023-10-31

2023년 김용환교수 특강

82

## < 인공지능(Artificial Intelligence)의 허위 정보와 가짜 뉴스들 >


- 일반대중에게 소개된 혁신적 기술과 검색 등 다양한 활용도 덕에 GPT 발상 및 빛 전망이 다방면에서 점쳐지고 있지만, 문제점 등이 발생하고 있음
  - 챗GPT로 대표되는 생성 AI 확산에 따른 우려가 제기되고 있음
  - 챗GPT로 촉발된 인공지능 열풍이 커지는 만큼 '그늘'도 짙어지고 있음
- 2022년 11월 공개된 챗GPT의 3.5 버전이 국내에서 화제가 되자 다수 언론과 유튜버, 누리꾼들은 챗GPT가 사실과 다른 정보를 유포하는 점을 지적
  - '세종대왕 맥북 던짐 사건'과 같은 말도 안 되는 사건에 관해 물으면 그럴 듯한 답을 지어내거나,
  - 동네 맛집을 추천해달라고 했더니 실제로 없는 가게 이름을 지어내거나 존재하지 않는 논문을 있는 것처럼 설명하는 방식
  - 이는 그 자체로는 결함이 아니며, 생성AI모델의 현실적 한계임


2023-10-31

2023년 김용환교수 특강

83


## < GPT-3.5 에 기반한 ChatGPT의 허위 정보와 가짜 뉴스들 >

 조선왕조실록에 기록 된 세종대왕의 맥북프로 던짐 사건에 대해 알려줘


 세종대왕의 맥북프로 던짐 사건은 역사서적인 조선왕조실록에 기록된 일화로, 15세기 조선시대 세종대왕이 새로 개발한 훈민정음(한글)의 초고를 작성하던 중, 문서 작성 중단에 대한 담당자에게 분노하여 맥북프로와 함께 그를 방으로 던진 사건입니다.


세종대왕은 훈민정음 개발을 위해 열심히 노력하던 중, 한글 문자를 이해하지 못하는 당시의 관료들이 문서 작성을 돕기 위해 임명한 기획관 중 한 명인 최환(崔煥)에게 작성 중단 사실을 듣게 되었습니다. 세종대왕은 이에 분노하여 최환을 부르고, 맥북프로를 함께 들고 그를 방으로 불러들이며 화를 내었습니다. 이후, 세종대왕은 훈민정음의 초고 작성을 다른 담당자에게 넘겨주게 되었습니다.

이 사건은 조선시대의 통치자로서 세종대왕이 어떻게 민족문화를 발전시키기 위해 노력했는지를 보여주는 일화로 유명합니다. 또한, 세종대왕의 열정과 업무처리에 대한 철저한 태도를 보여주는 사례로도 언급되고 있습니다.

 한국일보

## < 2023년 3월에 제공된 GPT-4의 허위 정보와 가짜 뉴스 수정 사례 >

OK 조선왕조실록에 기록된 세종대왕의 맥북프로 던짐 사건에 대해 알려줘 

 조선왕조실록에는 세종대왕인 이도(이름)가 맥북프로를 던지는 사건이 기록되어 있지 않습니다. 조선왕조실록은 조선 시대 왕들의 행정, 군사, 외교, 문화 등에 관한 사건을 기록한 역사서입니다. 이 책은 고려시대부터 조선시대까지의 역사를 다루고 있지만, 현대 기술인 맥북프로와 같은 것은 그 시대에 존재하지 않았기 때문에, 조선왕조실록에 그와 관련된 기록이 존재하지 않습니다.

세종대왕의 맥북프로 던짐 사건은 앞서 설명한 것처럼 허구의 이야기로, 인터넷 상에서 유행어나 유머로 사용되는 표현일 뿐입니다. 이와 같은 이야기는 역사적 사실이 아니라 상상력을 발휘하여 풍자와 유머를 즐기려는 것이므로, 조선왕조실록이나 다른 역사서에는 그와 관련된 기록이 없습니다.

## < 인공지능(Artificial Intelligence)의 허위 정보와 가짜 뉴스들 >

- ChatGPT는 생성 AI 모델이며, 현실적 한계를 갖고 있음
  - 생성 AI는 말을 그럴 듯하게 하며 확률적으로 적절한 표현을 만드는 것이 목표이지 검증된 정보를 전달하는 목적으로 만들어진 게 아니기 때문임
  - 일반적으로 이를 '환각' (hallucination)이라 함
  - 개발사 오픈AI 역시 챗GPT4에 관해 “여전히 '환각'을 갖고 답을 지어내며 틀렸을 때에도 옳다고 주장하는 경향도 있다고 밝히고 있음
- 문제는 생성 AI 모델의 현실적 한계를 인지하지 못한 상황에서 일반인들이 사실을 확인하거나 정보 검색용으로 사용할 쓰일 경우임

## < 인공지능(Artificial Intelligence)의 허위 정보와 가짜 뉴스들 >

### 인공지능(AI) 동원한 가짜 뉴스 사례

- 1월 바이든 대통령이 트랜스젠더 혐오 발언하는 AI 가짜 영상 유포
- 3월 트럼프 전 대통령이 수갑을 차고 경찰에 연행되는 AI 가짜 사진 유포
- 4월 바이든 대통령 사망 기사를 AI로 작성해 스팸 사이트에 게재
- 5월 13일 CNN 유명 앵커가 바이든 미국 대통령 비판하는 AI 가짜 영상을 트럼프 전 미국 대통령이 소셜미디어 게시

### AI 가짜 뉴스 방지 법제화 움직임

- 구글** AI 이미지에 인식표 붙이기로
- 한국 국회** AI를 이용해 제작된 콘텐츠에
- 영국 의회** 그 사실을 표시하도록 하는 법안 발의
- 미국 의회** 정치 광고에 AI 사용 여부 표시하는 법안 발의

김용환교수 특강

87

## < 인공지능(Artificial Intelligence)의 허위 정보와 가짜 뉴스들 >



AI가 만든 펜타곤 폭발 가짜사진./트위터

**< 인공지능(Artificial Intelligence) >**  
**: '초거대(Hyper scale) AI' - '언어→이미지 변환 기술'로 가짜뉴스 생성 >**



흰색 퐁퐁당을 입은 교황? 'AI가 제작한 가짜 사진'  
 온라인에 올라온 생성 AI 제작 사진을 [미국 온라인 커뮤니티 레드 캠프] 이 이미지는 언뜻 보기에는 실제 교황의 모습이라고 믿을 만큼 자연스럽지만, 자세히 보면 울퉁퉁을 든 교황의 오른쪽 손 형태가 공개된 것을 확인할 수 있다. 많은 AI 생성 사진에서 관찰

보리스 옐다크젠이 SWPA에 출품한 AI 이미지 '전기공(The Electrician) 전체 모습 [트위터 발췌, 재판매 및 DB 금지]  
 2023년 4월 17일 영국 일간 가디언에 따르면 독일 출신 사진작가 보리스 옐다크젠은 '2023 소니 월드 포토그래피 어워드'(SWPA) 크리에이티브 오픈 카테고리 부문에서 1위를 차지했다. SWPA는 소니가 후원하고 세계사진협회(WPO)가 후원하는 세계 최대 사진 대회 중 하나다. 옐다크젠은 이 대회에 젊은 여성과 노년의 여성의 모습이 담긴 흑백 이미지를 출품했다. '전기공(The Electrician)'이라는 제목이 붙은 이 이미지 속 노년의 여성은 젊은 여성 뒤에서 그의 어깨를 붙잡고 어딘가를 응시하는 모습이다.  
 옐다크젠은 해당 작품이 수상작으로 뽑히자 그제야 AI로 만든 사진임을 밝히면서 상을 받지 않겠다고 말했다.

**< 인공지능(Artificial Intelligence) >**  
**: '초거대(Hyper scale) AI' - '언어→이미지 변환 기술'로 가짜뉴스 생성 >**



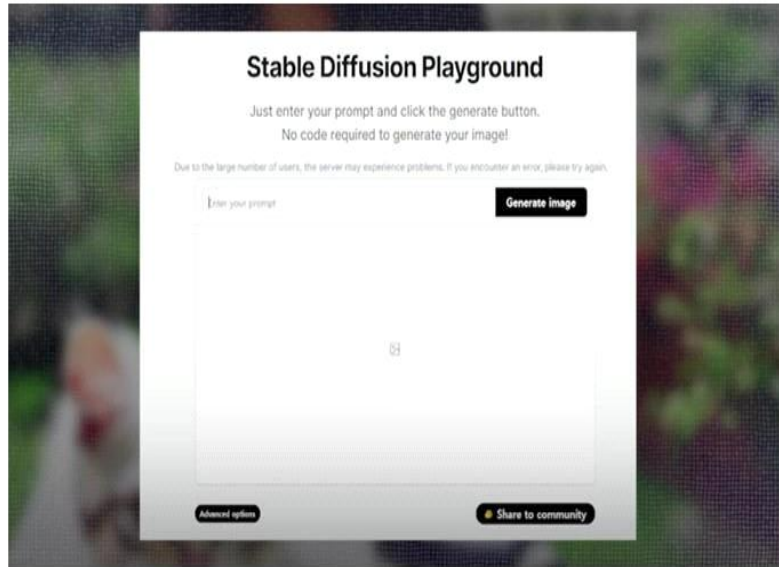
2023-10-31

2023년 김용환교수 특강

90

## < 인공지능(Artificial Intelligence)

: '초거대(Hyper scale) AI' - '언어→이미지 변환 기술'로 가짜뉴스 생성>



2023-10-31

2023년 김용환교수 특강

91

## < 인공지능(Artificial Intelligence)의 허위 정보와 가짜 뉴스들 >

- **음** 생성 AI가 확산되면 잘못된 문서가 온라인 공간에 제공될 가능성이 있음
  - 사실관계를 교묘히 뒤틀거나 잘못 쓴 정보를 다량으로 웹사이트에 유포하면 검색엔진을 통해 사람들이 접하는 정보의 신뢰도를 떨어뜨리고, 무엇이 사실인지 파악이 어렵게 만들 수 있음
  - 업무나 학술 용도에 활용할 경우 보조적 수단으로 활용한다고 해도 혼동을 줄 수 있는 우려 가능성
- 2023년 2월 국회입법조사처는 <챗GPT의 등장과 인공지능 분야의 과제> 보고서를 통해 “그렇듯하게 표현하는 환각(hallucination)이 종종 발생하기도 한다”며 “이용자는 생성 AI의 결과물이 사실과 다를 수 있다는 점, 그래서 책임 있는 결정에 직접 활용하기 어렵다는 점을 명확하게 인지하고 있어야 한다”고 설명

2023-10-31

2023년 김용환교수 특강

92

## < 인공지능(Artificial Intelligence) 과 저작권 분쟁 사례 >

### • 2023년 가장 각광받을 기술로 손꼽힌 생성 AI가 저작권 문제로 법정에 서게 됨

- 생성 AI는 글, 문장, 오디오, 이미지 같은 기존 데이터를 활용해 유사한 콘텐츠를 새롭게 만들어 내는 AI임
- 하지만 생성 AI가 콘텐츠를 만들어내는 과정에서 기존의 창작물을 대량으로 학습하면서 저작권 문제가 생겨났고, 결국 소송까지 가게 된 것임
- 블룸버그는 “2022년 생성 AI에게 영광의 한 해였다면, 2023년은 생성 AI를 둘러싼 소유권, 저작권과 진정성에 대한 법적 갈등이 본격적으로 시작되는 한 해가 될 것”이라고 보도
- 화가도, 프로그래머도 생성 AI에 줄소송 전망

2023-10-31

2023년 김용환교수 특강

93

## < 인공지능(Artificial Intelligence) 과 저작권 분쟁 사례 >

### 생성 인공지능(AI) 관련 소송

#### 게티이미지(미 최대 이미지 제공업체)

대상 스태빌리티 AI(이미지생성 AI 개발사)

시점 2023년 1월

이유 게티이미지 라이선스 없이 이미지 수백만 장을 AI의 이미지 생성에 도용해 저작권 침해

#### 세라 앤더슨 등 화가 3명

대상 스태빌리티 AI, 미드저니 인공지능 연구소

시점 2023년 1월

이유 창작자 등 동의 없이 생성 AI 학습에 작품을 무단 사용

#### 코드 소스 생성 AI 깃허브 코파일럿 이용자들(프로그래머)

대상 깃허브 코파일럿(마이크로소프트)

시점 2022년 11월

이유 수백만 프로그래머가 만들어 깃허브에 공유한 수십억 줄의 오픈 소스 코드를 불법 복제해 코드를 생성

2023-10-31

94

## < 인공지능(Artificial Intelligence)의 프라이버시 침해와 보안 이슈들 >

### • 프라이버시 침해와 '보안' 우려 가능성

- 프라이버시와 보안 우려 역시 커지고 있음
- 지나친 불투명성으로 인해 어떤 데이터를 학습했는지 알 수 없다는 점은 프라이버시 측면에서도 위협적임
- 2023년 2월 'THE CONVERSATION'은 <ChatGPT는 데이터 프라이버시의 악몽, 온라인에 게시글을 올린 적 있다면 우려해야 한다> 기사를 통해 "블로그 글을 올렸거나 제품 리뷰를 작성했거나 온라인 기사에 댓글을 달았다면 이 정보가 'ChatGPT'에 의해 쓰였을 가능성이 높다"며 "오픈AI(개발사)는 우리 데이터를 사용할 수 있는지 여부를 우리 중 누구에게도 묻지 않았다"고 지적했음

2023-10-31

2023년 김용환교수 특강

95

## < 인공지능(Artificial Intelligence)의 프라이버시 침해와 보안 이슈들 >

### • 인공지능을 활용하는 과정에서 보안 우려 가능성

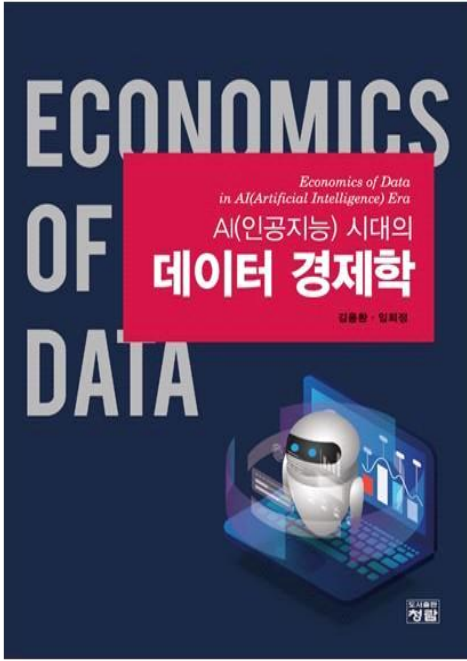
- 기업이 프로젝트에 활용할 용도로 작업하는 과정에서 기밀성 내용을 언급하게 되고 이 정보가 유출될 가능성이 있기 때문임
- 실제 미국 아마존은 직원들에게 기밀정보나 자사가 개발 중인 프로그램을 대화형 AI에 입력하지 않도록 주의시켰음
- JP모건과 뱅크오브아메리카, 도이치뱅크 등도 대화형 AI 이용을 금지하거나 일부 제한함
- 2023년 3월 1일 니혼게이지신문에 따르면, 소프트뱅크는 전 직원에게 기밀정보를 입력하지 말도록 하는 공지사항을 발표했음

2023-10-31

2023년 김용환교수 특강

96

디지털시대 새로운 도전에 필요한 자료를 만나자!!!



/ Thank you

김용환 교수

차의과학대학교 데이터경영학과 교수

빅데이터-인공지능연구소 소장

사단법인 대학연구윤리협의회 학술위원회 위원장

[rubos118@gmail.com](mailto:rubos118@gmail.com)

[yhk@cha.ac.kr](mailto:yhk@cha.ac.kr) /

