

AI 윤리



• 양석환

- AiDA Lab 대표. 컴퓨터공학박사 (세부전공: 인공지능, 신경망, 기계학습)
- 1994년 COBOL을 이용한 백화점 POS 개발을 시작으로 개발자 생활 시작
- (주)TGS(삼성SDS 컨소시엄): 부산경남 신발산업 업계 통합 대형 ERP 개발
- 日本 NEC(NEC Soft, 東京)에서 개발자 근무, AIST(つくば) 지능로봇기술연수
- 학위 취득 후 (주)지에스티 부설연구소 연구소장으로 스마트 팩토리 설계, 구축 및 관련 연구
- (주)COMMAX, (주)엑셈 등에서 AI사업부 근무
- 현재:
 - AI 기반 데이터 예측 및 분석 시스템
 - 문화/예술/교육 분야에 대한 AI 기술 적용
 - AI 교육

- [1 교시]

- 정보사회와 윤리
- 4대 기저 기술과 사회 변화
- 4대 혁신 기술과 사회 변화

- [2교시]

- 지능정보사회와 윤리
- 지능정보사회의 법
- 지능정보사회와 유해 콘텐츠

- [3교시]

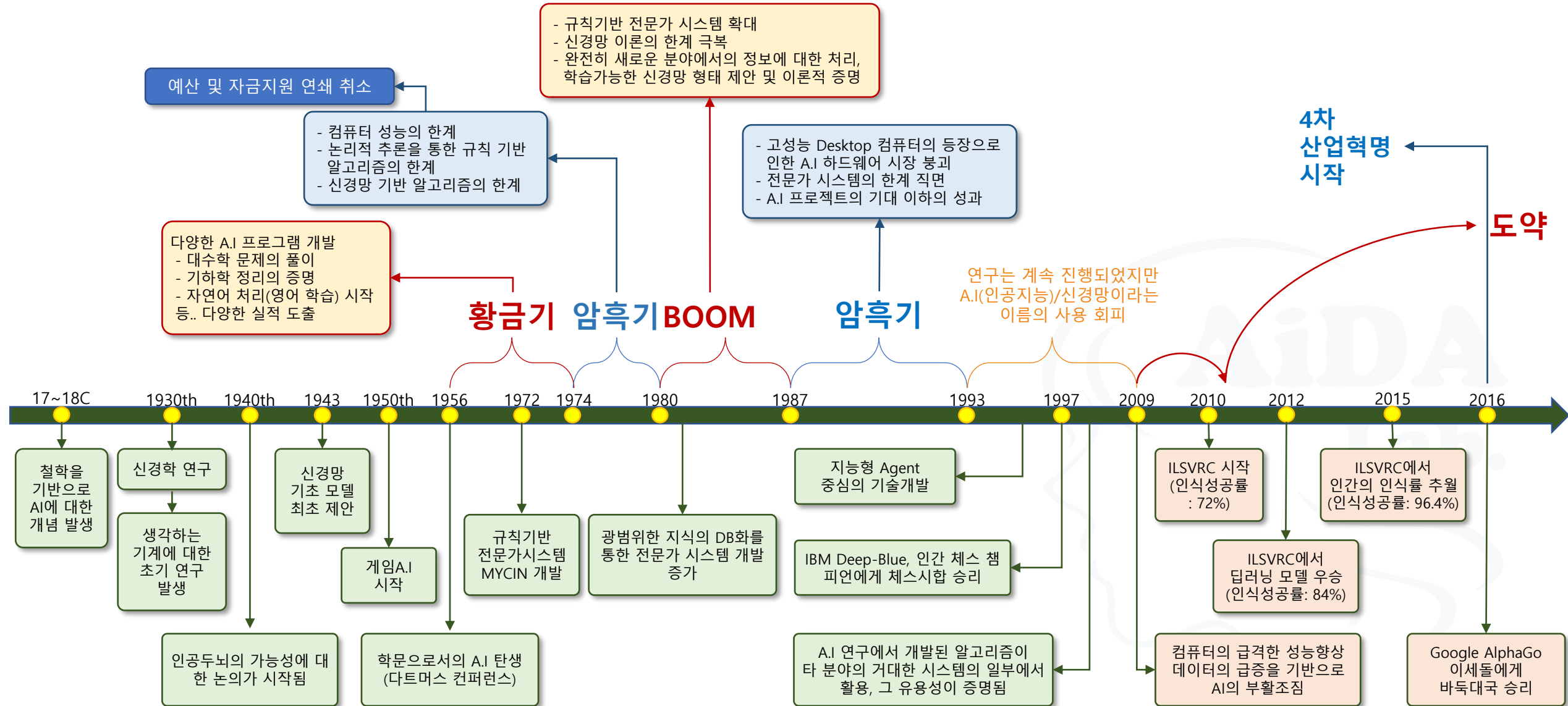
- 지능정보사회와 개인정보보호
- 지능정보사회와 저작권 침해
- 지능정보사회와 사이버범죄
- 지능정보사회와 사이버보안

- [4교시]

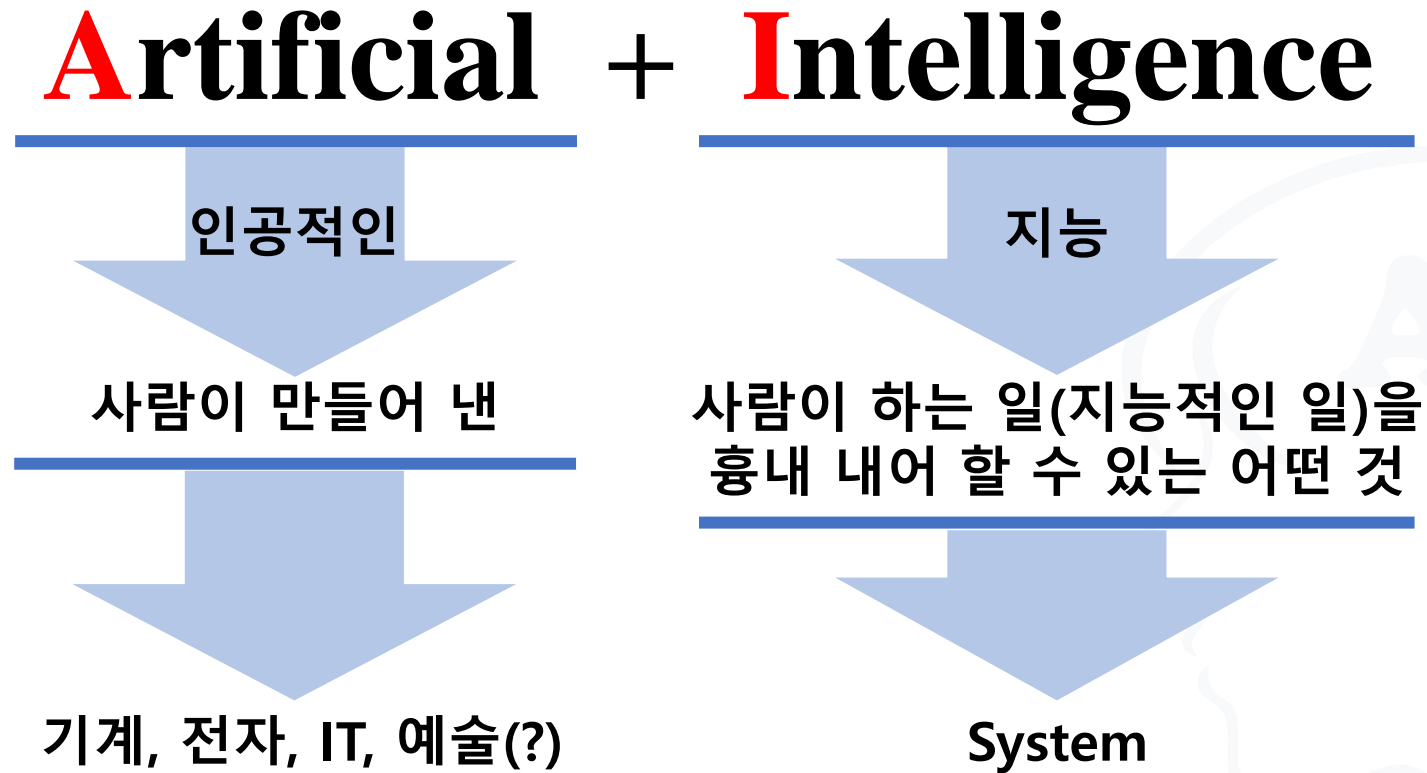
- 지능정보사회와 데이터 윤리
- 인공지능 개발자, 사용자, 공급자 윤리

- 2016년
 - 이세돌과 알파고의 바둑 대국: 일반인을 대상으로 한 AI 부활의 신호탄
 - WEF(세계 경제 포럼)의 4차 산업혁명 주창: 전 산업분야에서 Bigdata / IoT / AI 기술 부상
 - 미 국방성 DARPA, 차세대 AI 기술로 XAI 지목
- 학계, 산업계 만이 아니라 일반인을 포함하여 전국, 전 세계적으로 AI 기술이 화려하게 부상함

Ai의 대중화



- 인공지능(Artificial Intelligence, AI)이란 무엇인가?



- 인공지능이란
 - 다양한 기술을 이용하여
 - 사람이 하는 일을 흉내 내어 처리할 수 있는 시스템
- 다양한 기술에는
 - 기계, 전자, 컴퓨터 등 공학적인 기술과
 - 예술로 표현할 수 있는 창의성이 포함됨 (?)

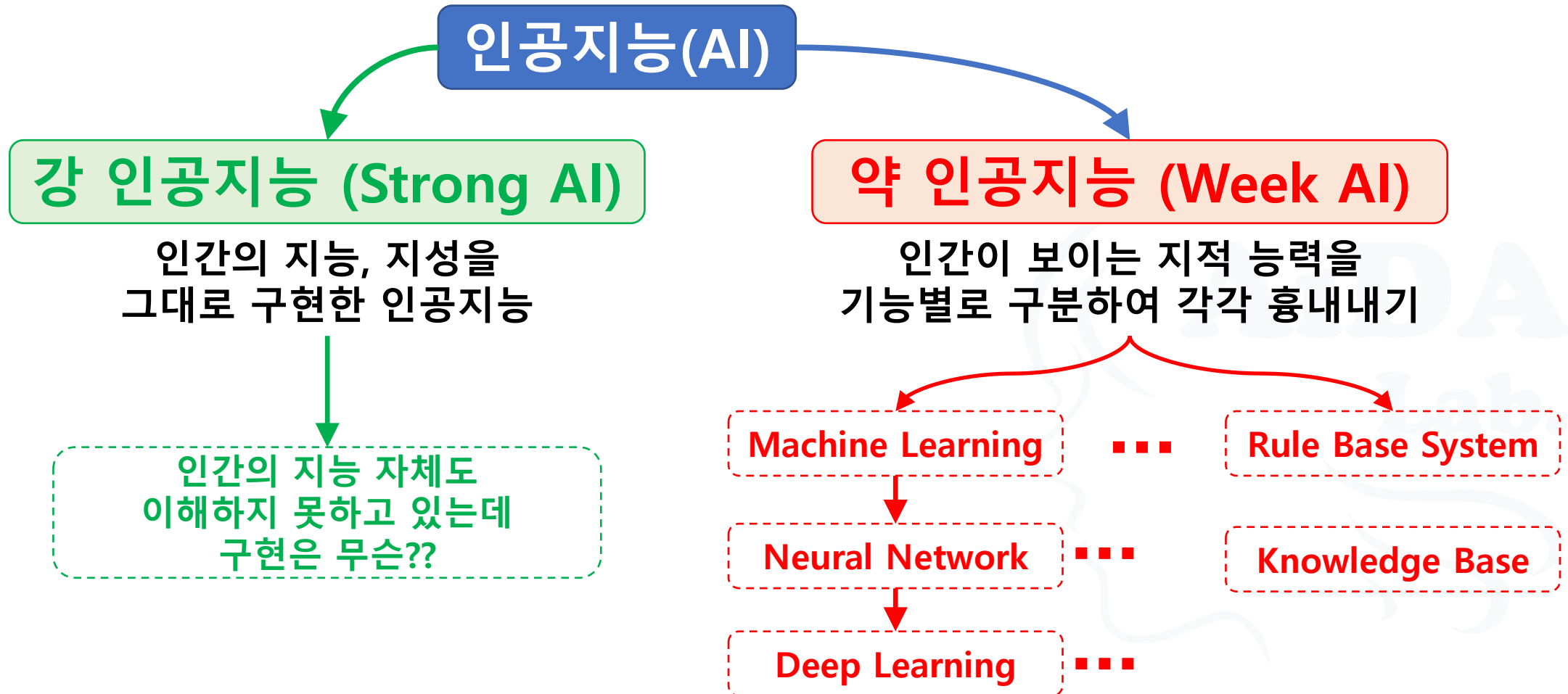
정의(?)가 애매한 것 같은데 왜 그렇죠?

“지능”의 정의가 아직 명확하지 않음
아직 인간의 지능에 대해서는 밝혀지지 않은
영역이 많아서 명확하게 정의할 수 없음

- 인간의 지능에 대하여 명확하게 밝혀지거나 정의되지 않음에 따라
 - 인공지능의 기술 구현 방향은 인간을 흉내내는 것으로 귀결됨
 - 시각: Computer Vision → 영상 인식, 분류, 영역구분 등
 - 청각 / 발성: Audio 처리기술 → AI Speaker 등
 - 촉각, 후각, 미각: 센서 기술 연구 수준에서 머물고 있음. 최근 성과가 조금씩 나오는 중
 - 사고: 현재 구현 불가능 → 데이터 처리, 의사결정, 언어처리 등으로 우회하여 구현

→ 주변에서 쉽게 볼 수 있는 인공지능의 연구/산업 분야

• 인공지능의 구분



- AI가 인간을 지배한다?
 - 다양한 분야의 전문가, 석학들이 AI의 위험성에 대하여 경고
 - 스티븐 호킹 박사: 인류 멸망 가능성의 4가지 시나리오 중 하나가 AI 로봇
- AI의 인간 지배는 가능한가?
 - 강인공지능에 대한 기술 진보가 없는 이상 → 불가능
 - 어느 날, **갑작스런 특이점이라도 오지 않는 한**, 현재로서는 걱정할 필요가 없음
- 그렇다면 무엇이 문제인가?

- 가장 큰 문제는
 - 현재의 AI는 Black Box 이다



안에서 무슨 짓을 하는지

알 수가 없다

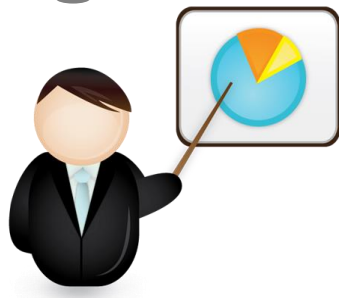


알 수가 없으니 믿을 수도 없다

• 사례 1

- 2016년, 홍콩의 헨슨로보틱스가 만든 AI 로봇 '소피아'

열심히 학습을...



소피아, 사진 출처 CNET

인류를
멸망시키겠다!!

CNBC 방송 중 발언
→ 농담이라고 마무리 하였으나 과연?

• 사례 2

- 2016년, 마이크로소프트의 AI 챗봇 '테이'



- 사례 1~2의 경우
 - 정보, 데이터의 편향성으로 인한 문제
 - 윤리적인 문제도 포함되지만 좀 더 기술적인 문제에 가까움
- 그렇다면 AI와 관련된 윤리적인 문제는 어떤 것이 있을까?

- AI 기술이 적용된 자율주행 자동차가 자율 운행 중 사고를 일으켰다.

사고의 책임은 누가 져야 할까?

- 자율주행 자동차의 운전자?
- 자율주행 자동차를 생산한 자동차 업체?
- 자율주행 자동차를 검증, 인증한 검사 업체?
- 자율주행 프로그램을 개발한 소프트웨어 개발 업체?
- 자율주행 프로그램을 개발한 개발자?

- AI 기술을 이용하여 개발된 AI 작가가 소설을 썼다.
그런데 AI가 쓴 소설이 저작권 침해로 고소를 당했다.
- AI가 쓴 소설은 저작권 침해로 인정받을 수 있는가?
- 책임은 누가 져야 하고 보상은 누가 해야 하는가?



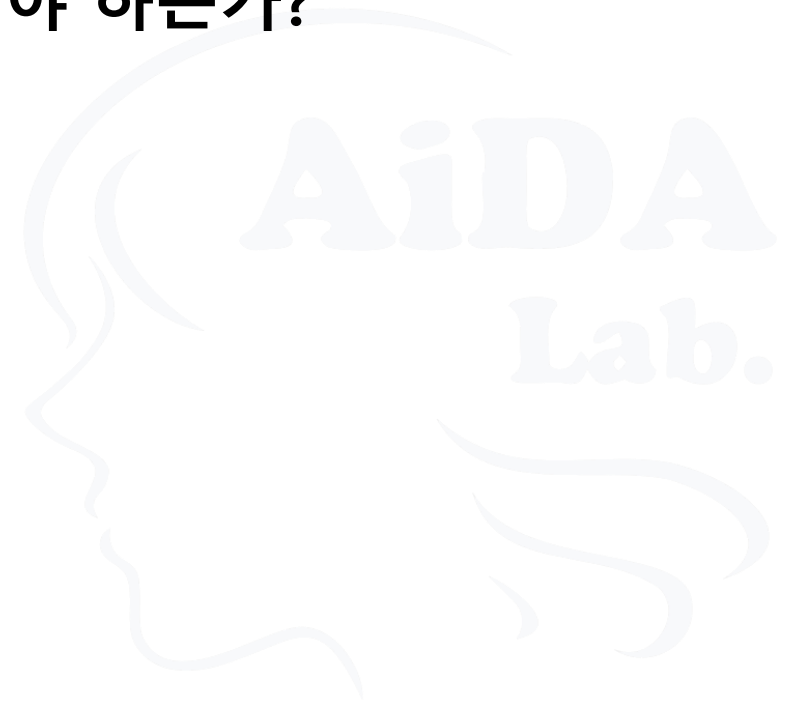
- AI 기자가 입력된 데이터를 기반으로 기사를 작성하였다.
그런데 가짜뉴스라며 비난을 받게 되었다.
 - 데이터를 기반으로 AI가 작성한 기사는 가짜뉴스인가?



- AI 시스템이 사용자 편의를 제공하기 위하여 사용자의 얼굴을 학습하고 여러 서비스에 활용하였다.
 - 이것은 초상권 침해에 해당하는가?



- AI 법관이 재판 결과 피고인에게 유죄를 판결하였다.
시간이 흐르고 해당 피고인의 무죄가 확인되었다.
 - 이런 경우, 책임은 누가 져야 하며 보상은 누가 해야 하는가?



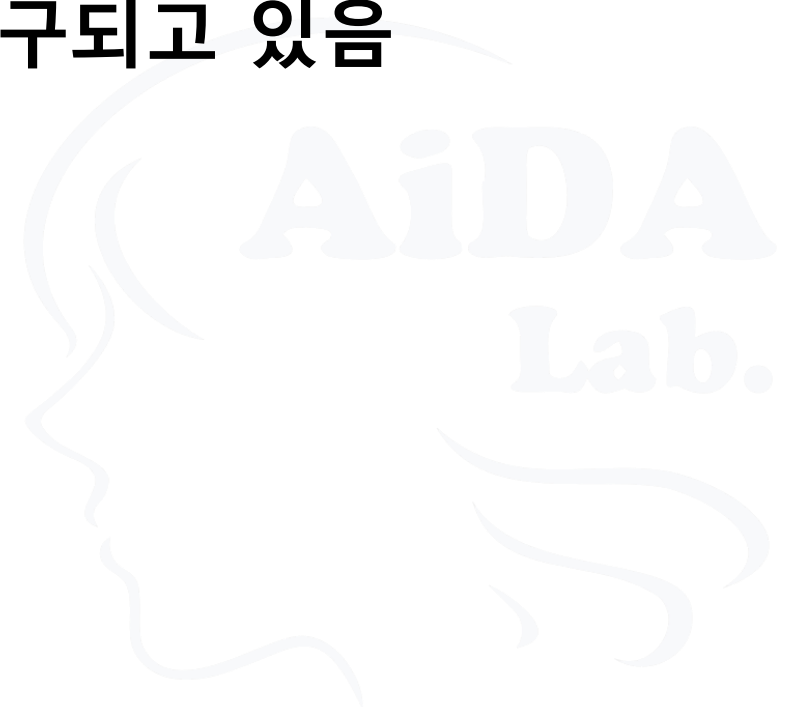
- AI 서비스를 위한 챗봇 시스템이 개발되어 서비스를 시작하였다.
그런데 챗봇을 학습시킬 때 AI 시스템이 스스로 인터넷에서 다양한 대화 데이터 및 콘텐츠를 수집하여 사용하였다.
 - 이 경우는 개인정보 침해에 해당하는가?



- AI의 음성합성 기술을 이용한 보이스피싱 사건이 증가하고 있다.
보이스피싱 조직이 주로 사용하는 음성합성 시스템해당 기술을 개발한 개발자에게 책임을 물어야 하는가?



- 지금까지 살펴본 것처럼 다양한 문제가 제기되고 있음
- “이러한 문제에 대하여 어떻게 대응하여야 하는가?”와 같이 다양한 문제와 상황에 대한 윤리적 접근과 이해가 요구되고 있음



전통 윤리



관습, 도덕, 법 그리고 윤리

구분	설 명
관습	<ul style="list-style-type: none">• 개별성을 지닌 특정 조직의 구성원들 간에 타당하다고 인정되는 규범적 질서
윤리	<ul style="list-style-type: none">• 관습의 이론적 토대 및 실천적 근거를 갖춘, 관습에 관한 시스템• 사회 구성원들의 합의를 통해 보편성을 지니며 관습에 정당성을 부여함
도덕	<ul style="list-style-type: none">• 옳고 그름에 대한 개인의 가치판단 기준• 자기 완성을 위한 개인적 규범
법	<ul style="list-style-type: none">• 사회 질서 및 안녕을 유지하기 위한 최소한의 윤리• 윤리 중 파급력이 큰 금지행위들

- 인간은
 - 행복을 추구하는 존재
 - 행복이란, 욕구가 충족되었을 경우 느낄 수 있는 만족과 기쁨
- 각 개인, 각 집단의 행복 추구 → 충돌을 야기할 수 있음
 - 대표적인 충돌의 원인: 자원, 재화의 유한성(=부족)
 - 자원 및 재화의 편중이 개인, 집단 간의 갈등과 충돌의 주요 원인이 됨

- 성악설, 성선설과 윤리

- 인간이 항상 악하기만 하다면(성악설) → 윤리는 존재할 수 없음
- 인간이 항상 선하기만 하다면(성선설) → 윤리는 필요 없음

- 윤리는

- 개인적인 욕망을 추구하기 위한 자기 행위를 자율적으로 규제하고
- 양심에 따라 선을 실현시킬 목적으로
- 스스로에게 명하는 자율적 규범
- 으로써 가치를 가지는 인류에게 반드시 필요한 요소

• 트롤리의 딜레마

- 1967년 영국의 철학자 '필리파 풋'이 최초로 고안한 문제
- 1985년 미국의 철학자 '주디스 자비스 톰슨' 이 살을 붙인 문제를 살펴보면

브레이크가 고장 난 기차가 빠른 속도로 선로 A를 달리고 있다.

선로 A 위에는 다섯 명의 사람들이 서 있으며, 이들은 미처 피할 수 있는 여유가 없다.

기차와 다섯 명의 사람들 사이에는 기차의 선로를 바꿀 수 있는 선로 전환기가 있다.

선로 전환기를 조작하면 기차를 선로 B로 달리게 할 수 있다.

그런데 선로 B 위에는 한 명의 사람이 있다.

당신이 선로 전환기 바로 앞에 있다면 어떤 행동을 취할 것인가?

(단, 행동에 대한 법적 책임은 무조건 면한다는 가정하에서 결정할 것)

- 트롤리의 딜레마 문제에서

- 선로 전환기를 사용하는 경우

- 다수의 인명을 구함

- '최대 다수의 최대 행복'을 주장하는 공리주의에 기반한 도덕적인 선택이라고 주장

- 선로 전환기를 사용하지 않는 경우

- 만약 다섯 명의 사람은 살아날 가능성이 희박한 응급환자이고 한 명의 사람은 가벼운 쇼크 증상으로 이송되는 응급환자였다면

- 과연 다섯 명을 구하는 것이 도덕적인 판단인가?

- 공리주의가 주장하는 '최대 다수의 최대 행복'을 위해 길가던 사람의 지갑을 강탈하여 자선단체에 기부한다면, 한 명의 피해자와 다섯 명의 수혜자가 발생하는데 이것이 과연 도덕적인가?

- 트롤리의 딜레마 문제에서

- 어떤 선택이 정답인지는 아무도 알 수 없음

- 미국의 철학자 '루이스 포이만'의 저서 <윤리학>에서 말하는 윤리학의 필요성

윤리학의 연구는
우리가 편견과 독단에서 벗어날 수 있도록 해주며
개인적인 판단을 지도해 줄 포괄적인 체계를 제시하고
무엇보다 우리가 어떻게 살아야 하는가에 대한
지침을 제공해 준다

• 윤리학의 구분

- 메타 윤리학: 어떤 행위가 선한가(옳은가)?를 연구
- 규범 윤리학: 특정 행위가 선하다고 하는 것의 의미는 무엇인가?를 연구
- 기술 윤리학: 특정 상회의 문화 혹은 도덕을 가치 평가 없이 객관적으로 기술

• 윤리학의 구분

- 메타 윤리학: 어떤 행위가 선한가(옳은가)?를 연구
- 규범 윤리학: 특정 행위가 선하다고 하는 것의 의미는 무엇인가?를 연구
 - 이론 윤리학: 문제 해결의 토대가 되는 기본 이론의 연구
 - 응용 윤리학: 현대인들이 실질적인 삶에서 마주하는 다양한 윤리 문제 해결이 목표
- 기술 윤리학: 특정 상회의 문화 혹은 도덕을 가치 평가 없이 객관적으로 기술

- 응용 윤리학의 대표적 세부 분야

- 직업 윤리

- 직업인이라면 누구에게나 사회적으로 요구되는 정신적 자세나 행위 규범

- 기업 윤리

- 기업 경영 시 나타나는 행동이나 태도의 옳고 그름에 활용되는 도덕적 판단 기준, 그리고 이와 관련된 의사 결정 과정

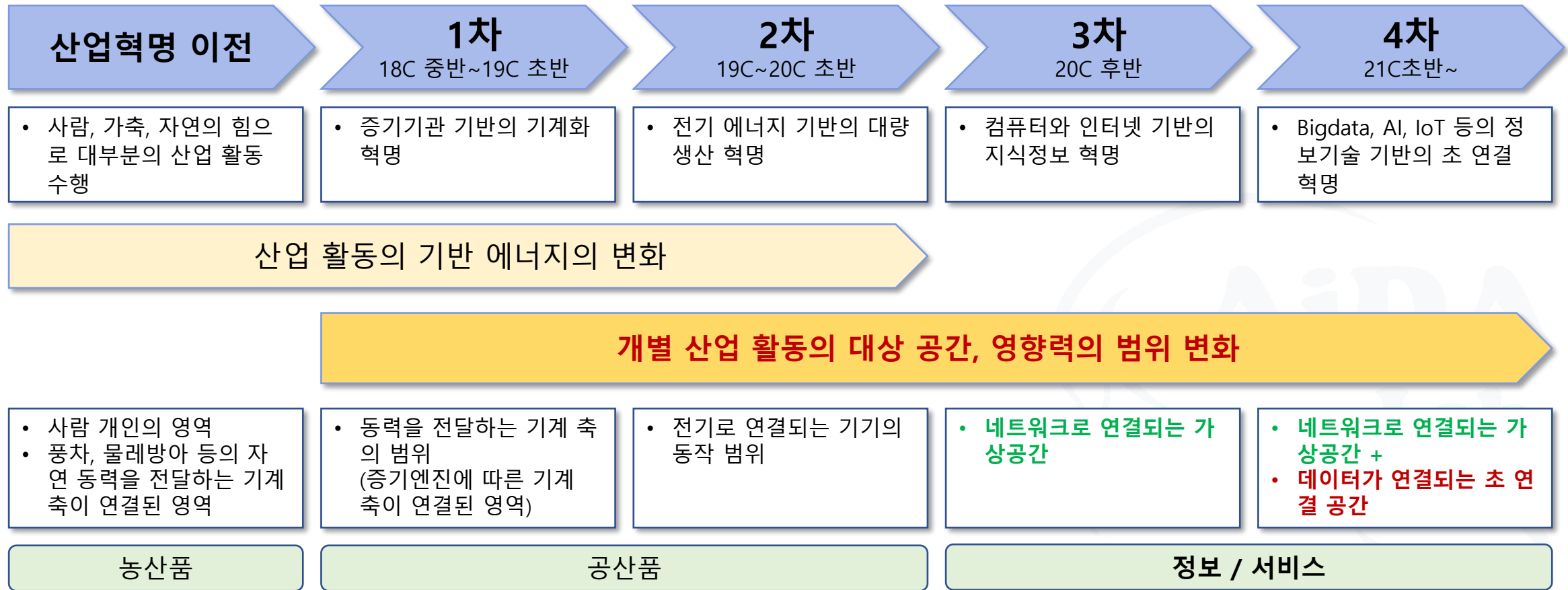
- 생명 윤리

- 생명과학, 의학 등 인간을 대상으로 연구 및 치료 시 발생하는 윤리적 문제에 대한 연구
 - 타 응용 윤리 영역 대비 시급한 윤리적 규범 확보가 요구되는 분야

정보사회와 윤리



산업혁명의 단계에 따른 산업환경의 변화



정보사회의 시작

• 정보사회

- 컴퓨터와 통신이 결합하여 정보의 수집, 가공, 유통 능력의 획기적인 증대로 정보의 가치가 산업사회에서의 물질이나 에너지 못지않게 중요해 지는 사회
- 엄청난 양의 정보가 생산되어 빠른 속도로 이동 및 유포, 소비되는 사회
- 인간의 모든 행동에 정보가 매우 중요한 역할을 하며, 정보의 가치가 어느 것보다 중요하게 인식되는 사회

- 일반적인 특징

- 양(量)의 사회에서 질(質)의 사회로 변화
- 시간적, 공간적 거리의 소멸 → 지역간, 국가간 경계의 모호성 증가
- 인간 사이의 소통, 사회적 행위가 매체를 기반으로 한 중개적인 형태로 전환
- 개인의 개성과 전문성의 중요성 증가와 이를 개발할 기회의 다양화
- 생산과 소비의 경제가 사라지고 소비자가 생산을 주도하는 경우도 발생

• 정보사회의 문제점

- 정보의 불평등에서 오는 정보 격차 문제
- 개인의 사생활 침해 문제
- 인터넷의 활성화와 정보의 쉬운 복사 기술로 인한 저작권 침해 문제
- 인터넷, 스마트폰 관련 중독 문제
- 정확하지 않거나 사실이 아니 정보의 인포데믹(Infodemic) 확산 현상
- 인터넷의 익명성을 이용한 악성 댓글, 유언비어 배포, 타인의 창작물에 대한 무단복제 및 유포 등의 정보 윤리 문제

Infodemic = Information(정보) + Epidemic(전염병)

- 사회 구조의 변화가 야기한 현상

- 기존의 산업사회와 정보사회는 가치관과 문화가 크게 달라짐

- 새로운 형태의 범죄 행위가 증가

- 기존 윤리와 법적 제도장치로 변화에 완벽하게 대처하기가 불가능해 짐

- 기존 윤리의 한계에 직면

- 정보사회에 적합한 새로운 정보윤리의 정립이 시급

4대 기저 기술과 사회 변화



4대 기저 기술(ICBM)의 소개

- **4대 기저 기술(ICBM)**

- **Internet of Things(IoT)**

- 주변의 수많은 사물들이 서로 인터넷 기반의 데이터 통신을 통해 다양한 서비스를 제공

- **Cloud**

- 지능정보화 시대의 필수 인프라. 언제 어디서든 쉽게 대량의 데이터 서비스 제공

- **Bigdata**

- 비정형 데이터를 포함한 대규모의 데이터 처리, 분석, 활용 기술

- **Mobile**

- 휴대형, 이동형 장비 기술을 통한 다양한 서비스 제공

- IoT 기술로 인한 사회 변화

- IoT 기술의 목표는 사용자의 편의를 위한 맞춤형 서비스의 개발 및 지원
- 이를 위해 사물로부터 측정된 대용량의 데이터를 수집, 분석 후 활용
 - 데이터 기반의 여러 판단이 획일화 되는 경향 발생
 - 사회의 중앙집중화 유발 가능성 증가

- IoT 기술로 인한 사회 변화와 해결해야 할 문제

- 보안 문제

- 많은 경우의 IoT 서비스는 개인의 생활패턴, 기호 등 민감할 수 있는 개인 정보를 대량으로 수집 → 사생활 침해의 우려 증가
- 데이터의 집중화에 따른 해킹 시도 등의 보안 문제 증가

- 표준화 문제

- 다양한 플랫폼에서 개발됨에 따라 각 기기 별 호환성 문제
- 기술의 표준화에 대한 국가간 협의가 요구됨

- **기술적 완성도**

- IoT 관련 장비가 고립된 지역 또는 악조건에 처한 상태가 지속되는 경우
→ 유해한 데이터가 축적, 생성되어 개인 및 사용자에게 제공될 수 있음
- 이러한 다양한 문제에 대한 해결 방안의 문제

- **데이터의 역기능**

- 개인에게서 축적된 데이터를 통해 개인이 서비스를 제공받는 것보다 사물이 개인을 통제하게 되는 역기능적 상황 발생 가능
- 사회적 협의와 제도의 개선이 요구됨
- 편의성에 앞서 정말 필요한 것인지에 대한 고려가 필요함

- IoT 기술의 활성화에 따라

- 수많은 사물과 연결되고 여기에서 수집된 정보를 기반으로 재가공, 재창출된 서비스 생태계가 구성됨
- 정보 유출, 정보의 위·변조 가능성, 유출된 정보의 오남용 등의 위험성 증가
- 이로 인한 사고 및 피해 → 경제, 사회적 문제 야기 가능성 급증
- 이미 다양한 IoT 관련 보안 침해 사례가 보고되고 있으나, 허술한 보안관리, 지나치게 많은 사물로 인한 침해 인지 어려움 등으로 침해 사례는 증가 추세

- Cloud 기술로 인한 사회 변화

- 2015년 클라우드컴퓨팅법 제정 후, 공공부문 중심으로 클라우드 산업 육성 및 생태계 구축을 위한 제도적 기반 조성
- 금융권의 클라우드 이용 확대
- 공공부문의 클라우드 이용 활성화
- 클라우드 기반의 전자정부 플랫폼



- Bigdata 기술로 인한 사회 변화

- 전세계적으로 빅데이터가 가져올 혁신과 패러다임의 변화에 주목
- 민간 영역에서는 제품생산, 서비스, 유통, 판매까지 모든 부분에서 활용 중
- 공공 영역에서의 활용 가치와 효율적인 활용 방법 모색
 - 선거와 빅데이터
 - 정부 정책과 빅데이터
 - 제품개발과 서비스

- Bigdata 기술의 활용 영역 이동

- 빅데이터 활용의 목적: 대부분 생산, 유통 판매의 효율성 향상을 위해서임
- 상업적 영역에서의 생산성, 효율성 제고
 - 개인에 대한 맞춤형 서비스로 발전
 - 개인 생활에 대한 영향력 강화
 - 개인의 사회적 행태 변화
 - 빅데이터화
 - 사회현상을 예측하는 도구로 활용



- Mobile 기술로 인한 사회 변화

- 모바일 장비, 앱 시장의 급성장

- 앱, 게임 콘텐츠, 생활, 홈 비서, 인공지능, VR/AR, 핀 테크 등 새로운 산업의 출현과 함께 동반 성장 중



4대 혁신 기술과 사회 변화



4대 혁신 기술로 인한 사회 변화



- 4차 산업혁명에 기인한 가장 큰 변화
 - 데이터 중심의 기술과 서비스 확산 → 최신 인공지능 기술까지 이어져
- 4대 혁신 기술
 - AI/Robot, VR/AR, 3D Printing, Bio



- AI/Robot 기술로 인한 사회 변화
 - 사회 전반에 대한 가장 큰 변화를 주도
 - 두 가지의 주요 변화 분야
 - 일상생활 환경의 변화
 - 업무환경의 변화



- 일상생활 환경의 변화

- 교통 분야

- 지능형 서비스/제품의 제공에 기인한 자율주행 자동차, 스마트 자동차 등
 - 지능형 교통정보 시스템: IoT 환경을 기반으로 수집, 분석된 정보를 통해 운전자, 대중교통 이용자에게 유용한 정보를 실시간 제공
 - On-Demand 교통서비스: 교통서비스의 공급자, 수요자의 실시간 연결, 실시간 가격 결정

- 가사 자동화

- 로봇청소기, AI스피커/IoT 기반의 가사 자동화 및 정보 서비스 등

- 교육 분야의 변화

- 교육용 로봇, 지능형 학습 시스템 등의 플랫폼 기반으로 맞춤형 학습 서비스 제공

- 업무환경의 변화

- 업무 자동화

- 기존 공장 자동화 기술에서 더욱 발전
 - RPA(Robot Process Automation) 등 다양한 IT기술, 데이터 기술, AI기술 기반의 자동화 확산

- 지식 서비스 등의 변화

- 특히 지식 집약적 직종에 대한 AI 기술의 활용으로 업무처리의 효율화
 - 다양한 지식과 정보를 맞춤형으로 제공하는 서비스
 - 비정형적 업무에 대하여 인간과 AI의 협업을 통한 시너지 효과 유도 등

4대 혁신 기술로 인한 사회 변화

- VR/AR 기술로 인한 사회 변화

- 교육, 비즈니스, 항공우주 등 다양한 분야에 대한 콘텐츠 확산
- AI/Robot 기술과의 협력으로 접근이 어려운 분야에 대한 접근, 시뮬레이션 구현 등

- 3D 프린팅 기술로 인한 사회 변화

- 초 다품종 소량생산 등 제조 환경의 변화 유도

- 바이오 기술로 인한 사회 변화

- 질병의 발생을 사전에 예방하는 예방치료 방식으로의 전환 시작
- 미래 식량 문제, 환경/에너지 분야의 문제에 대한 해결 방향 제시

지능정보사회와 윤리

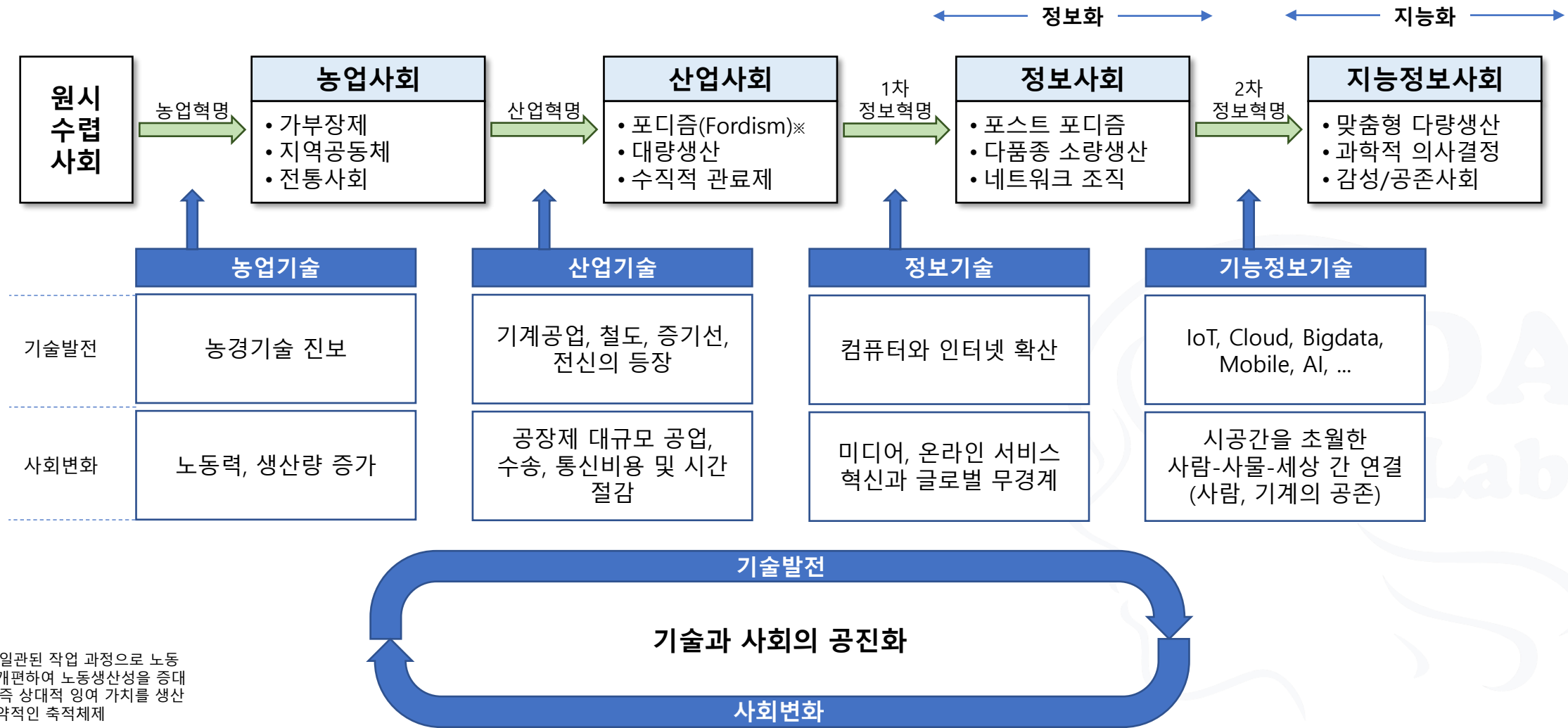


- 지능정보사회가 보이는 특징적인 모습
 - 사물인터넷 → 빅데이터 → AI 기술 적용 → 클라우드 서비스 제공
 - 지식 중심, 데이터 중심의 사회로 진화
- 과거의 산업혁명 시대: 인간의 노동력이 기계로 대체되는 시기
- 현재의 AI 시대: 인간의 지능이 기계로 대체되는 시기
- 지능정보사회: 제품, 서비스가 융합, 산업 간 경계가 붕괴되는 특징을 보임

- 지능정보사회는

- 지능화된 기계를 통한 자동화가 지적 노동 영역까지 확장되어 경제, 사회 전반에 혁신적인 변화를 유발하고 있는
- 고도화된 정보통신기술 인프라를 통해 생성, 수집, 축적된 데이터와 AI가 결합한 지능정보기술이 경제, 사회, 삶 모든 분야에 보편적으로 활용되어 새로운 가치가 창출되고 발전하는 사회

지능정보사회의 특징



그림출처: (이원태, 문정옥, 류현숙, "지능정보사회의 공공정보 패러다임 변화와 미래정책 연구," 정보통신정책연구원, 2017) 에서 정리, 일부수정

- **지능정보기술은**

- **사람의 삶의 양태를 변화시킴**

- 전통적 윤리와 관습은 시간과 공간의 제약에서 형성되며 공적영역과 사적영역을 구분
 - ICT(정보통신기술)의 발전은 시간, 공간, 공적/사적 영역이 혼재하는 탈 경계화 야기
 - 전통적 윤리와 관습의 발생 근원이 유지되지 못함

- **지능정보사회는 기존 정보사회에서 누적된 역기능이 미해결 상태로 중첩됨**

- 기존 정보사회에서 발생한 윤리 문제 + AI, Bigdata 등에서 유발되는 새로운 윤리 문제
 - 데이터를 기반으로 하는 침해, 변경, 조작 등으로 인한 윤리적 문제가 사회적 위협으로 대두됨

• 사례

- **2018년 2월, MIT 미디어랩, 조이 부올람위니**
 - 젠더셰이즈 프로젝트 수행: IBM(미), Microsoft(미), 메그비(중)의 얼굴인식 기술 비교
 - 피부 빛이 검을 수록 인식률 저하 및 오류 증가. 특히 흑인 여성의 경우 35% 내외의 인식 오류 발생 (2015년 일반적인 이미지 인식성공률은 인간을 뛰어넘은 96.4%)
- **미국 법원의 재범률 예측 AI의 성능테스트**
 - 흑인의 재범률을 백인의 재범률에 비해 실제보다 훨씬 높게 추론
 - 큰 편향성을 보임

- **2017년 미국 스탠포드 대학**

- 온라인 데이트 사이트에 공개된 남녀 프로필 사진을 이용하여 성적 지향성 추론 실험
- 단지 프로필 사진만을 이용하여 정확한 추론 결과 도출
- 성소수자에 대한 사회적 불이익을 유발할 수 있는 윤리적으로 반사회적인 실험으로 평가

- **‘웨이모’, 자율주행 자동차 운영의 상용화 시작**

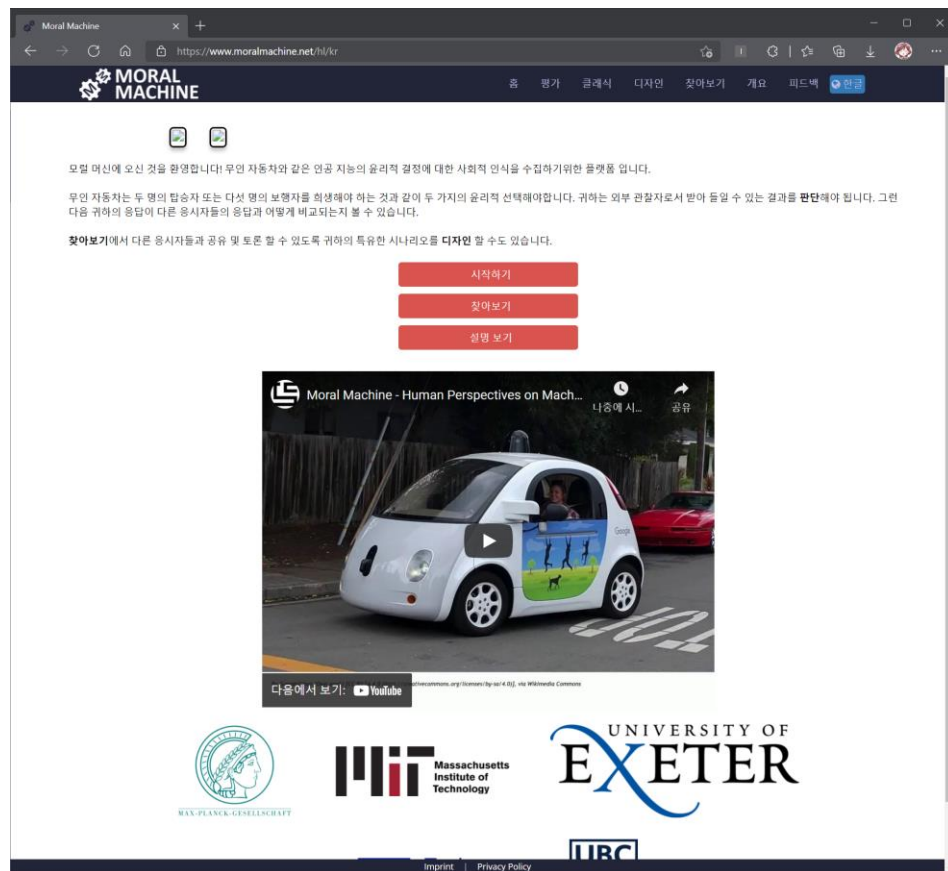
- ‘브레이크가 고장났을때 AI가 어떤 선택을 할 것인가?’ 에 대한 문제 제기
→ 트롤리의 딜레마

- **기타 사례**

- AI 기술 기반으로 구직자 이력서 1차 필터링 → 데이터 편향성에 따른 차별 조장 가능성
- 등...

- MIT에서 수행한 AI와 윤리에 대한 연구
 - 윤리기계(Moral Machine)라는 플랫폼을 이용하여 자율주행 자동차에서 발생할 수 있는 윤리 문제에 대한 일반인의 의견을 수집 → AI 설계에 반영
 - 윤리기계 사이트의 목적
 - 윤리적 딜레마에 직면한 AI의 결정에 대한 사람들의 의견 수집
 - 일어날 수 있는 문제점에 대한 시나리오를 만들고 토론할 수 있는 플랫폼의 제공

MIT의 윤리 기계(Moral Machine)



<https://www.moralmachine.net/hl/kr>





- 새로운 접근법의 제시

- 트롤리의 딜레마에서 확인가능한, 정답이 없는 윤리문제에서
- 집단지성의 도움을 받아 민주주의적 방식으로 AI의 윤리적 의사 결정 문제를 해결할 수 있는 접근법을 제시했다는 평가



• 트롤리의 딜레마에 대한 판단의 관점

구분	허용가능 수준	판단 관점	설명
<p>트롤리는 선로를 따라 달려오고 있고, 선로에는 다섯 사람이 있다. 당신은 선로 밖에 서 있고 다섯 사람을 구하기 위해서는 선로변환기를 당기면 되지만 그렇게 되면 다른 선로에 있는 다른 한 사람이 죽게 된다. 선로 변환기를 당기는 행위는 도덕적으로 허용 가능한가?</p> 	85% 도덕적 허용 가능	이성적 문제	선로 변환기만 조작해서 결정을 내리는 경우에는 전전두엽이 활성화
<p>트롤리는 선로를 따라 달려오고 있고, 선로에는 다섯 사람이 있다. 당신은 선로 밖에 언덕 위에 서 있고, 바로 옆에는 상당히 무거운 사람이 한명 서 있다. 다섯 사람을 구하는 유일한 방법은 옆에 서 있는 사람을 선로 위로 밀쳐서 그 무게로 트롤리를 멈추게 하는 것인데, 이는 도덕적으로 허용 가능한가?</p> 	12% 도덕적 허용 가능	감성적 문제	한 명의 덩치 큰 살아있는 사람에게 적극적 행동(직접 뛰어들어 밀쳐서 선로로 떨어뜨리는 행위)이 필요한 경우에는 편도체 부위가 활성화

- AI와 관련된 윤리 문제

- 정서적 판단의 개입이 줄거나 AI와 같이 정서적 판단 없는 기계에 전적으로 의존하여 판단하는 경우 → 윤리적 문제 발생 가능
- 기계적 알고리즘이 가진 한계인 이성적인 해결은 인간 고유의 윤리와 상충할 수 있음

- 지능정보사회는

- 인간의 보편적 윤리를 통해 AI가 가지는 이성적 판단의 한계를 해결해야 함

• 실업 등 빈부 격차와 일자리 창출

- 산업혁명 당시: 기계로 인한 대량 실업과 고용감소 우려

- 금융산업의 ATM 도입: 금융업계의 대량 실업 예상

→ 실제로는 새로운 일자리 증가로 실업률 감소 → 신기술과의 협력 요구

지능정보사회의 일자리 분야 이슈

이슈	긍정적 이슈	부정적 이슈
일자리 분야	<ul style="list-style-type: none">• 인공지능 기술 발전에 따른 새로운 일자리 창출• 인간(자연지능)과 로봇(인공지능)의 협업 확대• 기술 조력을 통한 업무 효율성, 정확성 증가• 인간 고유의 창의적 일 증가• 교육 부문에서 통찰력, 융합 중시	<ul style="list-style-type: none">• 인공지능 기술 수용에 따른 일자리 감소, 실업률 증가• 인간의 노동에 대한 사회적 가치 약화• 로봇과 공조에 따른 업무 복잡성 증가• 기술 적응 비용 증가(재교육 비용 등)• 로봇 및 기계의 업무 참여에 따른 책임성 문제

출처: "4차산업혁명, 대한민국의 미래를 찾다," 한국정보화진흥원, 2018

- 공유 경제 모델과 갈등의 고조

- 우버, 에어비앤비 등의 급성장 → 공유 경제의 현실화 → 기존 업계의 반발
- 스마트폰, 자율주행 자동차는 이러한 갈등이 없었는데..?
 - 스마트폰, 자율주행 자동차는 새로운 산업, 제품, 서비스 영역
- 공유경제는 기존의 물품이나 서비스를 대체하는 성격이므로 갈등은 필연적
- 블록체인, 암호화폐, 핀테크 등의 기술과 금융업계의 반발과 동일한 이유
 - 금융업계의 블록체인, 암호화폐 기술 포용을 통해 해결방안 모색 중

- **비대면, 프리랜서 경제 확산**

- 기존의 정규직, 비정규직 차별 등 다양한 문제에서 벗어나고자 하는 개인
- 지능정보사회의 다양한 기술(IoT, Cloud, 5G등 통신기술, IT 플랫폼 등)을 이용한 비대면 근무, 프리랜서의 증가
- 독립적 계약 근무직, 전문가 집단을 위주로 돌아가는 Gig 이코노미 성장
- COVID-19로 인한 급격한 확산
- 기업으로부터의 인력 탈출 현상 증가
- 프리랜서 등 Gig 워커는 기업의 복지혜택 및 보호를 받을 수 없는 어려움 증가

지능정보사회와 법



- 법이란

- 개인의 권리를 보호하고 사회질서와 공공의 이익을 실현시키기 위한 규범
- 국가 공동체의 구성원이 총의로써 형성한 것

- 법의 기능

- 사회질서의 유지, 인권 보장, 분쟁의 예방 및 해결 등

- 지능정보사회의 법 제도의 패러다임 변화

- 전통적인 법체계

- 사람을 중심으로 권리능력, 행위능력, 책임능력을 설정하고 그에 따른 체계를 구성

- 지능정보사회에 요구되는 법체계

- 사람만이 아닌 인공지능, 로봇 등의 특정한 지능체계가 사람을 대체 또는 보조
- 인공지능에 의한 행위와 책임에 대한 논의 발생
- “사람이 아닌 것에 법적 효과를 인정할 수 있는가?”의 문제
 - 기계의 특정한 행위와 그에 법적인 효과를 부여하는 체계가 현재에도 일부 인정됨
 - 단순 기계를 벗어나 상황을 판단하는 기능의 추가는 기계의 자동적 행위를 넘어서는 것으로 판단 가능하며 이에 대한 법 체계의 고민이 요구됨

- 지능정보사회는 지능 또는 정보를 기반으로 한 사회
- 지능정보사회의 개념을 중심으로 법적 쟁점 논의 가능
 - 지능 기반의 법체계 논의
 - 인공지능의 법적 지위나 성격을 어떻게 규정하고 그 주체성을 인정할 것인가
 - 인공지능의 동작에 대한 법적 효과를 어떻게 부여할 것인가
 - 정보 기반의 법체계 논의
 - 정보의 수집, 활용 등 처리의 관계를 중심으로 논의
 - 기존의 정보 문제는 개인정보의 보호문제로 집중 → 데이터에 대한 주도권에 대한 고려
(소유권, 행사권 등)

• 갈등과 규제

- 지능정보사회의 기술, 서비스의 적용관계 vs 기존 기술, 서비스의 적용관계
→ 갈등 유발 가능

• 사례

- 공유 자동차의 도입과 기존 택시업계 간의 갈등 (타다 자동차 등)
- 디지털 헬스케어의 도입과 기존 의료체계와의 갈등

• 지능정보사회에서 갈등을 둘러싼 법적 이슈는

- 규제의 완화, 폐지 또는 신설 문제로 귀결
- 새로운 편익과 기존의 이익 사이의 비교에 따라 양보와 희생 결정

- 대표적인 규제 갈등 사례

- 타다 서비스 사례

- 지능정보사회의 대표적인 서비스인 공유경제가 국내에서는 불법으로 판정, 금지
 - 2020년 2월 29일, 유상 운송 금지 위반의 구성요건에 해당되지 않음 → 무죄 선고

- 자율주행 자동차 사례

- 자동차 제조 관점에서 → 전통적 자동차 사업자와 새로운 자율주행 자동차 제작자의 관계에 주목
 - 사고 발생 시에 대한 윤리적 판단 및 책임 소재에 대한 논란 계속

• 데이터를 둘러싼 갈등 사례

- IoT 등 기계를 통한 부지불식 간의 데이터의 수집 vs 정보 주체의 제공을 통한 수집 등
- 활용 허가여부, 정보 주체와의 합의 불균형 문제 등
- 의료 분야 사례
 - 데이터, AI를 활용한 신약 개발, ICT 융합을 통한 헬스케어 서비스, 원격의료 서비스
 - 의료인, 요양업 등 기존 사업영역과 IT 기업 간의 갈등
 - 동네 병/의원과 대학병원 등의 대형 병원 사이의 갈등

- 갈등의 해결 구조

- 새로운 기술, 서비스에 대한 규제 완화
- 기존 사업에 대한 보호 방안



- 자율주행 자동차의 상용화를 대비한 입법 추진 중

- 자동차 법제

- 자율주행 자동차의 운행과 관련한 제반 사항 → 입법 활용 가능성 높음

- 도로 법제

- 전용 운행구간 설정, 도로 설치 기준의 재설정, 도로 종류 및 구분 재설정 등 → 미추진

- 도로교통 법제

- 사람의 운전을 중심으로 차량 및 운전자의 주의 의무 규정 → 운전 주체 규정 → 미추진

- 정보통신 법제

- 수집, 활용되는 정보가 개인정보의 경우 → 개인정보 보호법 적용 → 해결 방안 모색
 - 데이터에 대한 해킹으로부터 보호문제, 망 중립성 적용 문제 등 → 미추진

지능정보사회와 유해 콘텐츠

- 유해 콘텐츠

- 국가별, 사회별 환경에 따라 기준이 상이

- 인터넷 유해 콘텐츠의 분류

- 합법 콘텐츠의 경우 → 비도덕적이거나 비윤리적인 콘텐츠

- 반국가적 내용, 인권 침해, 인명 경시, 법과 질서의 존엄성 저해 등 국민정서에 반하는 경우

- 불법 콘텐츠의 경우 → 법적으로 위반되는 콘텐츠

• 인터넷 유해 콘텐츠의 유형

보호차원	보호범위	유해 콘텐츠의 유형	위법 여부
사회 생활 보호	국가 안전 보장	무기 제조, 폭탄 제조, 병역 기피, 약물 제조 등	불법(국가보안법, 형법)
	사회적 풍속 보호	자살 사이트, 사이버 도박, 사행심 조장, 음란 채팅 등	일부 위법
	미성년자 보호	음란 영상물, 폭력물 등	청소년 보호법 위반
	지식재산권 보호	영상물, 음악, 저작물 무단 배포	불법
	신뢰 및 신용 보호	스팸/메일폭탄, 해킹, 사이버 사기	일부 위법
개인 생활 보호	인간의 존엄 보호	낙태 조장 사이트, 인종차별, 성차별, 인신매매	일부 위법
	사생활 보호	개인정보 유통, 스토킹	일부 위법
	개인의 명예 보호	악성 유머 유포, 스토킹	불법

• 유해 콘텐츠의 특징

- **익명성:** 정보의 제공자, 사용자의 인적사항이 쉽게 노출되지 않으므로 비윤리적 행위를 부추기는 원인이 되고 있음
- **쌍방향성:** 정보의 제공자와 사용자의 구분이 없음
- **동시성:** 정보와 콘텐츠의 이용이 실시간, 동시에 이루어짐
- **접근의 용이성:** 누구나 쉽게 접근
- **휘발성:** 탐지가 어려움
- **정보의 복제 및 변형 가능성:** 악의적인 정보의 왜곡 등

- 유해 콘텐츠의 침해 유형

- 사행심 관련 도박 정보
- 음란, 선정성 정보
- 불법 명의 거래
- 불법 화학물질 정보
- 불법 금융 광고



• 유해 콘텐츠 대응 방안

법률적 방안	기술적 방안	교육적 방안
<ul style="list-style-type: none">• 규제를 위한 법률 및 과태료 처분• 광고 표시 의무화(제목에 광고 표시)• 거부 표시 의무화 (수신 거부)	<ul style="list-style-type: none">• 차단 필터링(블랙리스트, 인공지능 방식)• 사후/사전 평가 시스템(콘텐츠 분석, 처리 기준)	<ul style="list-style-type: none">• 유해 콘텐츠의 본질과 해악 교육• 피해 시 대처 요령 교육



- AI를 활용한 유해 콘텐츠 차단 사례
 - 카카오의 유해 콘텐츠 필터링 기술
 - 네이버 'X-eye' 성인 이미지 필터링 시스템
 - 페이스북의 AI 'Chip'
 - 구글의 'Content Safety API'
- 유해 콘텐츠 예방 방법
 - 공유기의 유해 콘텐츠 차단: ipTIME의 '아이안심 ipTIME' 스마트폰 앱 서비스
 - 유해정보 라벨링/필터링: 인터넷 콘텐츠 선별 플랫폼(PICS, Platform for Internet Content Selection)

지능정보사회와 개인정보보호

• 개인정보

• 개인을 알아볼 수 있는 정보

- 직접적으로 특정인을 알아볼 수 없어도 다른 정보와 결합하여 알아볼 수 있는 것 포함

• 개인정보의 구체적 사례

- 신분 관계: 성명, 주민등록번호, 주소, 본적, 가족관계, 본관 등
- 내면의 비밀: 사상, 신조, 종교, 가치관, 정치적 성향 등
- 심신의 상태: 건강 상태, 신장, 체중 등 신체적 특징, 병력, 장애 정도 등
- 사회경력: 학력, 직업, 자격, 전과 여부 등
- 경제 관계: 소득 규모, 재산 보유 상황, 거래내역, 신용정보, 채권/채무 관계 등
- 기타 새로운 유형: 생체 인식 정보(지문, 홍채, DNA 등), 위치 정보 등

출처: 개인정보분쟁조정위원회 – 분쟁조정 제도 안내 – 개인정보 의미

- 인공지능 사회에서 개인정보는 안전한가?

- 경계를 넘어서는 맞춤형 서비스

- AI의 기반은 데이터 → 서비스의 제공을 위하여 필연적으로 개인정보의 수집 동반
→ 침해 위험이 상시 존재

- 생활에 가까워진 지능 정보

- 개인정보 기반의 생활밀착형 서비스 확산
 - CCTV를 활용한 안심귀가, 교통정보, 주차 단속, 무인기기 활용 서비스, 사회 인프라 관리 서비스 등

- AI 스피커 등 AI 기기의 활용

- AI 스피커를 활용한 다양한 서비스와의 연결
 - 날씨 및 정보 검색
 - 개인 맞춤형 추천 서비스(음악, 뉴스, 콘텐츠 등)
 - 우버 등 공유 경제 서비스와 연결
- 데이터가 중심이 되는 다양한 서비스가 확산 중
- 개인 정보가 다양한 서비스의 주체로 전달되어 활용됨 → 개인정보 침해의 위험성 높음

지능정보사회와 저작권 침해



• 저작권

- 인간의 사상이나 감정을 표현한 창작물을 만들어 낸 창작자에게 부여된 배타적이고 독점적인 권리
- 저작권법 제 1조 – 저작권법의 목적
 - 이 법은 저작자의 권리와 이에 인접하는 권리를 보호하고 저작물의 공정한 이용을 도모함으로써 문화 및 관련 산업의 향상발전에 이바지함을 목적으로 한다.

• 저작권의 종류

저작권	세부 권리	설명
저작재산권	복제권	저작물을 복제할 권리
	공연권	저작물을 공연할 권리
	공중송신권	저작물을 공중송신할 권리
	전시권	저작물의 원본이나 복제본을 전시할 권리
	배포권	저작물의 원본이나 복제본을 배포할 권리
	대여권	저작물을 영리를 목적으로 대여할 권리
	2차적저작물 작성권	저작물을 원저작물로 하는 2차적저작물(웹툰을 드라마나 영화로 작성)을 작성하여 이용할 권리
저작인격권	공표권	저작물을 공표하거나 공표하지 아니할 권리
	성명표시권	저작물에 실명 또는 이명을 표시할 권리
	동일성유지권	저작물의 내용, 형식 및 제호의 동일성을 유지할 권리

• 저작물의 종류

종류	설명
어문저작물	소설, 시, 논문, 강연, 연설, 각본 등, 언어를 매체로 하여 작성된 저작물
음악저작물	악곡 등 음에 의하여 표현되는 저작물
연극저작물	연극 및 무용, 무언극 등에 있어 동작에 의하여 표현되는 저작물(즉, 동작의 형, 안무)
미술저작물	회화, 서예, 도안, 조각, 공예, 응용미술저작물 등, 형상 또는 색체에 의하여 미적으로 표현되는 저작물
건축저작물	건축물, 건축을 위한 모형 및 설계도서 등, 토지상의 공작물에 표현된 전체적인 디자인
사진저작물	사진, 청사진 등 사진의 방법으로 표현한 저작물
영상저작물	영화, 애니메이션 등, 연속적인 영상으로 표현되는 저작물
도형저작물	지도, 도표, 약도, 모형, 설계도(건축 설계도, 모형은 건축저작물에 해당)에 의해 표현되는 저작물
컴퓨터프로그램저작물	특정한 결과를 얻기 위하여 컴퓨터 등 정보처리 능력을 가진 장치 안에서 직접 또는 간접으로 사용되는 일련의 지시, 명령으로 표현된 창작물
2차적저작물	원저작물을 번역, 편곡, 변형, 각색, 영상제작, 그 밖의 방법으로 작성한 창작물. 2차적저작물은 다시 위의 하나에 해당할 수 있음
편집저작물	저작물이나 부호, 문자, 음성, 음향, 영상 그 밖의 자료 등 소재의 집합물로서 그 소재의 선택 또는 배열이 창작성이 있는 것(창작성 있는 데이터베이스 포함)

- 오프라인 저작권 침해

- 1980~1990년대의 음반, 어학 카세트 테이프, 만화책의 복제 판매 사례
- 게임 미디어의 크랙, 복제 판매 사례 등

- 온라인 저작권 침해

- 인터넷을 사용한 불법 음반, 영상 복제
- 1999년 P2P 서비스를 이용한 MP3 공유 서비스, 냅스터 → 온라인 저작권 침해의 최초 사례로 언급됨

- 새로운 유형의 저작권 침해

- OTT(Over The Top) 스트리밍 서비스에 대하여 불법 서버 등을 이용한 콘텐츠 전송 사례 등



• 인공지능의 창작

- 기존 저작물 데이터를 학습한 인공지능 모델 → 변형, 수정된 새로운 콘텐츠 생산 → 인공지능이 창작의 주체로 등장하기 시작
- 인공지능의 결과물을 창작물로 인정할 것인가?
 - 창작? 모작? 습작? 표절?
 - 셰익스피어도 초기 활동기에는 표절에서 자유롭지 못했음 → 습작기로 보는 시각도 있음
 - 창작을 위해 인공지능이 학습하기 위해서는 수많은 저작물에 접근하여야 함
→ 이것은 저작권 침해인가? 습작인가?
- 현재로서는 아직 가능성이 낮은 상황이지만 관련 지침의 필요성은 요구됨



캠브리지 컨설턴트가 엔비디아의 기술을 활용하여 개발한 빈센트 AI (출처: 엔비디아)

지능정보사회와 사이버범죄

• 사이버 범죄의 분류체계

대분류	중분류	소분류
사이버범죄	정보통신망 침해 범죄	해킹, 서비스 거부 공격, 악성 프로그램
	정보통신망 이용 범죄	인터넷 사기, 사이버 금융 범죄, 개인/위치정보 침해, 사이버 저작권 침해
	불법 콘텐츠 범죄	사이버 음란물, 사이버 도박, 사이버 명예훼손/모욕, 사이버 스토킹



- 인공지능에 의한 저작권 침해

- 2017년 08월, 구글은 사진에서 워터마크(저작권 표시의 일종)만 제거하는 AI 프로그램을 개발하여 공개함
 - 인공지능이 저작권 침해 범죄에 적극 참여 가능함을 보여줌

- 인공지능을 이용한 사이버 사기 고도화

- 2018년 5월, 구글은 헤어숍에서 고객들의 예약을 받는 점원과 자연스러운 전화 대화를 나누면서 예약에 성공하는 AI 서비스 발표
- 딥페이크를 이용한 가짜 동영상 제작 및 배포 기술

지능정보사회의 사이버범죄



GAN (ProGAN)을 통해 유명인의 사진을 바탕으로 만들어진 **허구의 인물** (출처: 엔비디아)



버락 오바마 전 미국 대통령의 **가짜 영상** (출처: 워싱턴 대학교)

- 인공지능에 의한 개인정보 유출 및 사생활 침해 확대
 - 2019년 10월, 고려대 SW보안 국제공동연구센터
 - IoT 보안 취약점 자동 분석 플랫폼 'IoT 큐브'를 이용하여 현재 판매 중인 AI 스피커의 소스코드 분석
 - 평균 350여개의 보안 취약점 노출, 2018년에 알려진 블루투스 보안 취약점 그대로 방치
 - 국제공통취약점(CVE): 아마존 '에코'(324개), 구글 '구글 홈'(454개), 네이버 '프렌즈'(272개), 카카오 '카카오 미니'(338개) 발견
 - 4가지 제품 모두 심각한 보안취약점(블루본, KRACK) 그대로 방치
 - AI 스피커를 통한 개인정보 유출, 개인정보 침해 가능성 매우 높음

- 인공지능을 통한 악성 코드의 지능적 활동
 - 특정 사람을 인식하면 악성 코드가 활성화 되도록 공격
- 인공지능의 인식을 교란하는 적대적 패치
 - 교통표지판 인식을 교란하여 자율주행 자동차의 사고를 고의적으로 유도

지능정보사회와 데이터 윤리



• 얼굴 인식

- 업로드된 사진에 대해 자동으로 레이블을 태깅하는 기능을 가진 사물인식 AI
 - 미국의 한 흑인 남성과 그 여자친구를 '고릴라'로 분류
- 여러 사람이 이야기하는 도중에 말하는 사람에게 카메라 초점을 맞추는 AI
 - 흑인 여성이 이야기 할 때, 흑인 여성을 사람으로 인식하지 않고 카메라를 비추지 않음
- 미국의 얼굴인식 AI
 - 상원 및 하원 의원 28명을 범죄자로 인식

• 채팅

• 마이크로소프트 AI 챗봇 '테이'

- 백인우월주의자, 여성 및 무슬림 혐오자들이 챗봇 AI가 차별 발언을 하도록 훈련시킬 것을 모의하여 고의적으로 욕설, 인종차별, 성차별 등 극우성향의 대화를 계속 입력함으로써 AI의 훈련 결과를 왜곡시킴
- 출시 당일, 여성차별 및 혐오 발언, 홀로코스트 조작설 등의 발언을 함

• 채용

- 직원 채용 AI가 이력서를 체크할 때
- '여성 체스 동아리'와 같은 여성이라는 단어가 들어가는 이력서는 감점시키도록 학습시켜 차별을 조장시킨 사례
- 여자대학을 졸업한 지원자는 감점, 남자 기술자들이 이력서에 자주 쓰는 표현이 발견되면 가산점을 주는 식으로 차별 행위

• 추천

- 미국인의 데이터를 이용하여 학습한 의료 AI가 제시하는 진단 및 치료법이
- 아시아 국가에서는 과다 약물 투여 등으로 인한 사망사고 유발 가능성 발견
- 평균적으로 미국인과 아시아인의 체격 차이와 같은 데이터의 차이로 인해 진단법과 치료법이 달라지게 되는 사실이 반영되지 않음
- 외부적인 고의성 조작이 없더라도 AI로 인한 차별 및 위험성이 증가하는 경우도 있을 수 있음

• 예측

• 미국 법원의 재범률 예측과 FBI의 감시 대상 인물 선정 AI

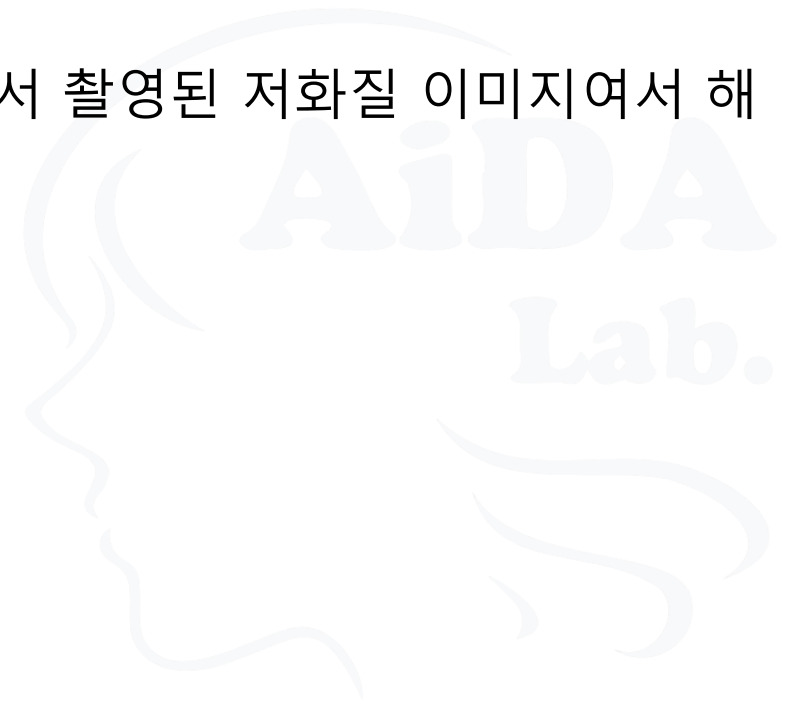
- 재범하지 않을 것으로 예측되었으나 재범을 하거나, 그 반대의 경우 발생
- 백인과 흑인의 예측 결과가 편향된 사례 발생

• 유료 댓글 기능의 AI

- 일정한 금액을 지불하면 해당 댓글이 진행 중인 방송의 채팅창 상단에 고정되어 수익 창출과 연결되는 차별 조장
- 동일한 방식으로 헤이스트 스피치(특정 집단이나 개인에 대한 차별적 혐오발언)를 조장

• 의료

- **최신 기기로 촬영된 고화질의 13만장의 망막 이미지로 학습한 의료 AI**
 - 의료 기술이 낙후된 국가에서 도입, 활용하려다 실패
 - 해당 국가에서 촬영된 망막 화상 이미지는 구형 장비에서 촬영된 저화질 이미지여서 해당 AI에서 사용할 수 없었음



• 대출 심사

• 주택담보 대출 심사 업무를 담당하는 AI

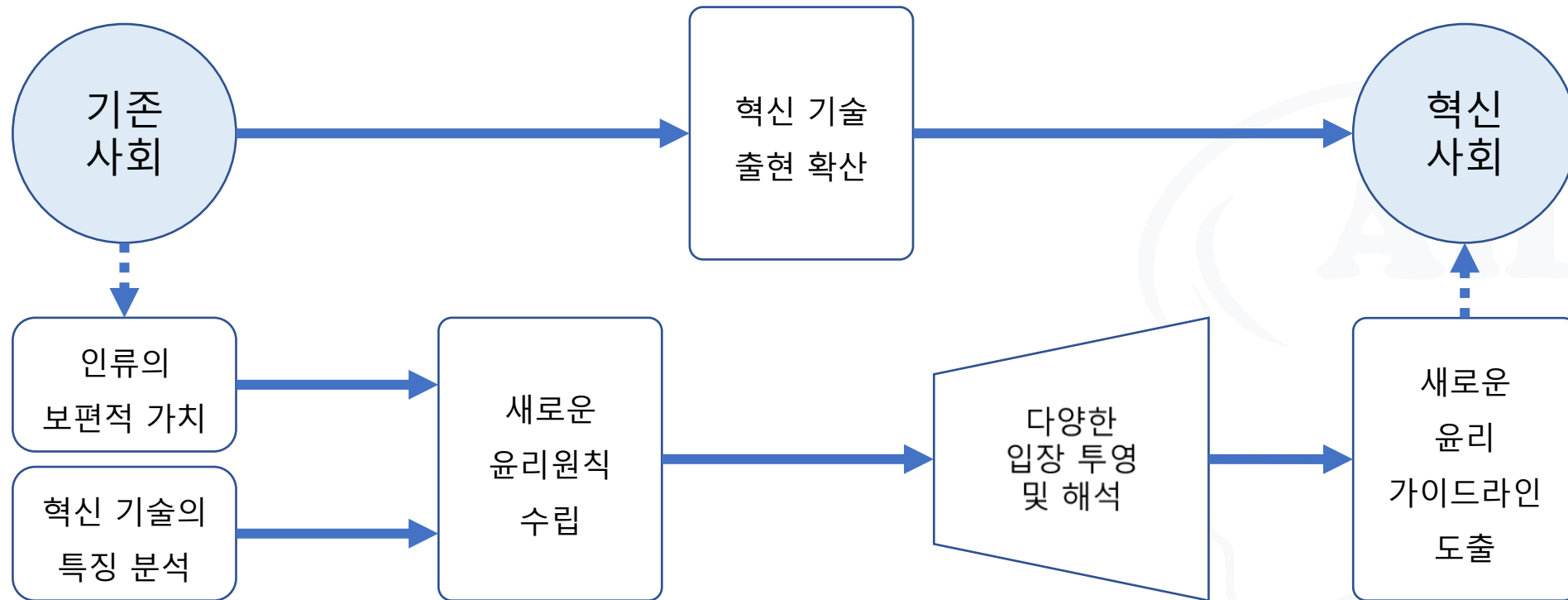
- 중산층 주택가와 빈곤층 주택가에 대한 데이터를 학습
- 중산층 주택의 대출금액은 높고 변제율이 높기 때문에
- 중산층 주택은 계속적으로 주택 가격이 상승하고, 중산층 주민의 변제 능력도 계속 상승
- 빈곤층 주택은 계속적으로 주택 가격이 하락하고, 빈곤층 주민의 변제 능력도 계속 하락
- 부익부 빈익빈 현상을 AI 시스템이 조장하는 결과가 발생

• 데이터 품질의 기준

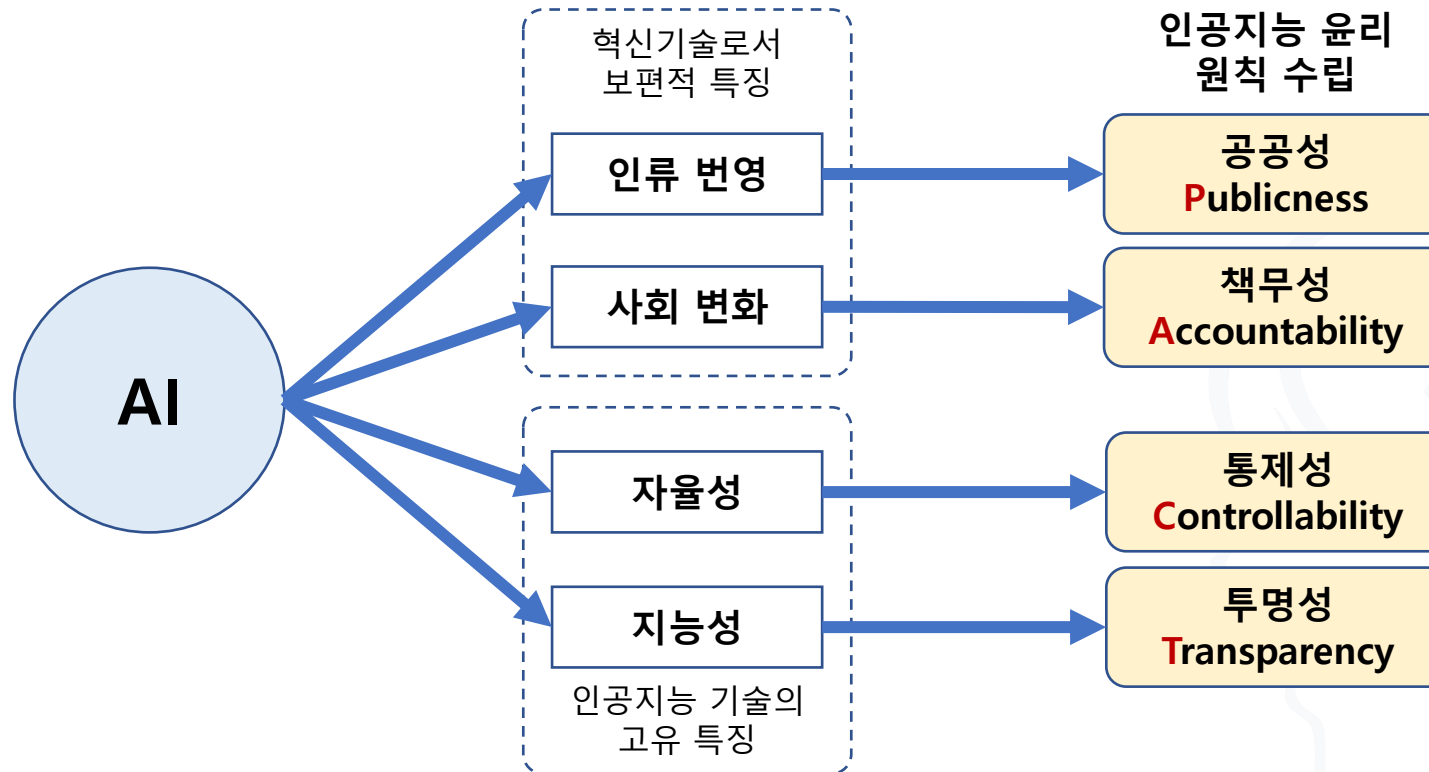
데이터의 특성	개요
다양성	<ul style="list-style-type: none">다양한 데이터를 확보하지 못함으로 인해 발생할 수 있는 역기능 문제의 방지를 위하여 편향되지 않은 다양한 데이터가 제공되어야 함
정확성	<ul style="list-style-type: none">의료 진료에 이용되는 전자의무기록 데이터, 기업의 매출 데이터 등의 경우, 정확한 데이터가 수집, 제공되어야 함
적합성	<ul style="list-style-type: none">주택담보대출 심사의 경우, 중산층과 빈곤층 데이터를 학습하게 함으로써 중산층 주택가격 인상이나 빈곤층 대출규제와 같은 문제가 발생하지 않도록 문제와 목표에 적합한 데이터가 제공되어야 함
충분성	<ul style="list-style-type: none">얼굴인식 AI를 개발하고자 할 경우 최소 수백만 장 이상의 사진이 필요한데, 데이터가 충분히 확보되지 못해서 얼굴을 제대로 인식할 수 없는 문제가 발생하지 않도록 충분한 양의 데이터가 제공되어야 함
투명성	<ul style="list-style-type: none">데이터의 중요성이 증가함에 따라 데이터를 매매하는 시장과 데이터 매매업자의 등장으로 데이터의 출처를 위장하는 문제가 발생하지 않도록 데이터의 출처가 투명해야 함

인공지능 개발자, 사용자, 공급자 윤리

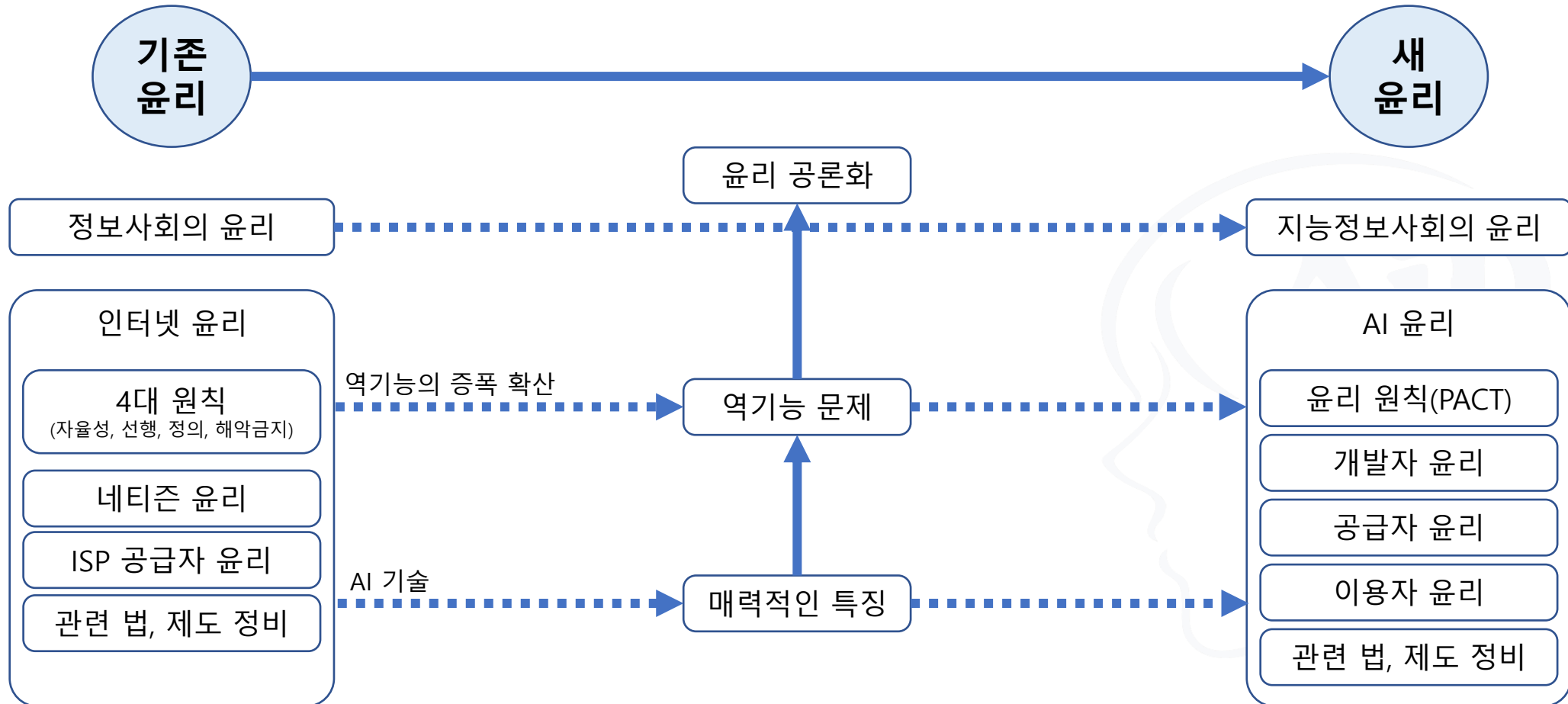
- 혁신 기술 출현에 따른 새로운 윤리 도출 과정



- 인공지능의 특징을 기반으로 한 윤리 원칙 수립 과정



• 사회 혁신에 따른 인터넷 윤리에서 인공지능 윤리로의 이행



- 공공성(Publicness) 원칙 기반
 - 기술 개발 시 사회적 차별 요소 배제
 - 사회적 약자 보호를 위한 접근성 보장
 - 인공지능 기술의 공공성 확보 노력



- **책임성(Accountability) 원칙 기반**
 - 책임의 공유
 - 개발자들 간 정보 교류와 기술 갱신에 지속 참여
 - 품질 인증 기준의 충족
 - 윤리적 절차를 충실히 이행하여 연구, 개발



- 통제성(Controllability) 원칙 기반
 - 예외적 상황에 대한 종합적 검토
 - 지속적 품질 관리 실시
 - 기술적 제어 장치 마련



- 투명성(Transparency) 원칙 기반
 - 위급 상황 시, 필요 데이터 제공
 - 은닉 기능 개발 금지
 - 위험에 대한 적극적 예측 및 공급자와 공유



- 공공성(Publicness) 원칙 기반
 - 공공의 이익에 부합하는 제품 공급
 - 선한 의도를 가진 발주
 - 상업적 이익과 공공적 기여 사이의 조화



- **책임성(Accountability) 원칙 기반**
 - 책임 공유 및 책임, 보상 원칙의 마련
 - 이용자 권리 보장
 - 지능 정보 서비스의 자율 의사결정 조건 및 범위 확립
 - 위험 예방을 위한 사회적 공론화 참여



- 통제성(Controllability) 원칙 기반
 - 제품 유통과정에서의 위험 통제
 - 안전성 검증 및 통제 조치의 마련
 - 이용자의 선택권 보장



- 투명성(Transparency) 원칙 기반
 - 위험 관련 정보의 이용자 공유
 - 이용자 정보의 부당한 이용 금지
 - 사회적 영향 평가 결과의 반영



- 공공성(Publicness) 원칙 기반
 - 악의적 이용 금지
 - 공공의 이익을 위한 제품 개선에 참여
 - 소비자 행동 원칙 준수의 일상화



- **책임성(Accountability) 원칙 기반**
 - 이용자 윤리 책임 숙지
 - 책임을 제기할 수 있는 권리
 - 안전 관련 정보 공유와 제도화 요구의 권리
 - 이용 제품의 교체, 갱신, 폐기 시 지침 준수



- 통제성(Controllability) 원칙 기반
 - 지능정보기술 이용 역량 강화
 - 자의적 조작 금지



- 투명성(Transparency) 원칙 기반
 - 설명을 요구할 권리
 - 개인정보 활용에 대한 감시자로서의 의무
 - 소비자 정보의 공유 의무



- **건전한 연구 문화 조성 및 국내외 협력 네트워크 구축**
 - 건전한 연구 문화 조성, 기술 개발과 정책의 연계, 글로벌 협력
- **공공개발 유도 및 위험 예방을 위한 법 제도 마련**
 - 거버넌스 구축 및 운영, 정책적 인센티브 필요, 사회적 영향 평가 제도 마련, 이용자 권익 보호를 위한 법과 제도 마련
- **오남용 방지 및 이용자 역량 강화를 통한 자율적 활용 제고**
 - 대국민 캠페인 및 홍보, 이용자 역량 강화를 통한 자율적 활용 제고