24년 추계학술대회

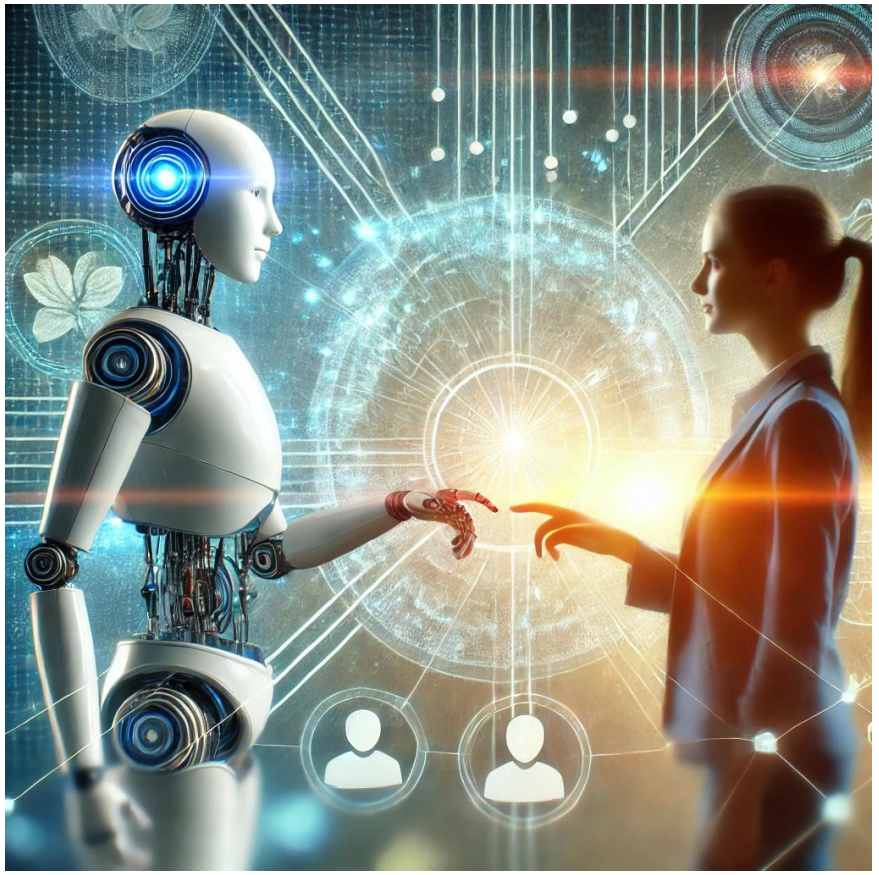
연구윤리 교육

한국보건간호학회

2024.11.08

목차

- 인공지능 시대의 전환
- 생성형 AI의 기술과 윤리
- 생성형 AI관련 윤리원칙
- 생성형 AI관련 윤리정책



인공지능 시대의 전환

지능을 가진 기계의 등장

- John McCarthy의 인공지능(Artificial Intelligence, AI)정의 및 활용
  - 기계가 지식을 가지고 스스로 학습하고 행동할 수 있어야 함
  - 대규모 데이터와 패턴을 학습하고 기존의 데이터를 활용함
  - 이차적인 텍스트, 이미지, 음악, 코딩 등 이용자의 요구에 맞춘 새로운 결과를 만들어 낼 수 있어야 함
  - AI 기술은 금융, 미디어, 교육, 의료, 법률, 행정 등 다양한 산업 및 업종 등에 도입되어 인간의 일상생활 속에 깊이 퍼져 있음
  - 학문탐구, 연구 및 창작 등에서 일하는 방식을 개선하여 생산성과 효율성 증진에 기여

AI 발달에 따른 활용기술의 변화

- 국가적 차원에서의 AI기술 활용 권장
  - 전 국민 일상화 목표로 산업현장 및 공공행정 분야에 적극적 도입 계획(AI 일상화 및 산업 고도화 계획_23.01_발표)
  - 금융, 미디어, 교육, 의료, 법률, 행정 등 다양한 산업 업종에 적용
  - 궁극적으로는 전 산업을 관통하는 AI기술 융합 기술 필요
  - 파운데이션 모델 구축을 통해 생성형 AI 서비스를 개발 하는 것으로 서비스 확대 진행 중

〈그림 2〉 생성형 AI를 활용한 서비스 개발 트렌드

출처) 삼정KPMG(2023b)

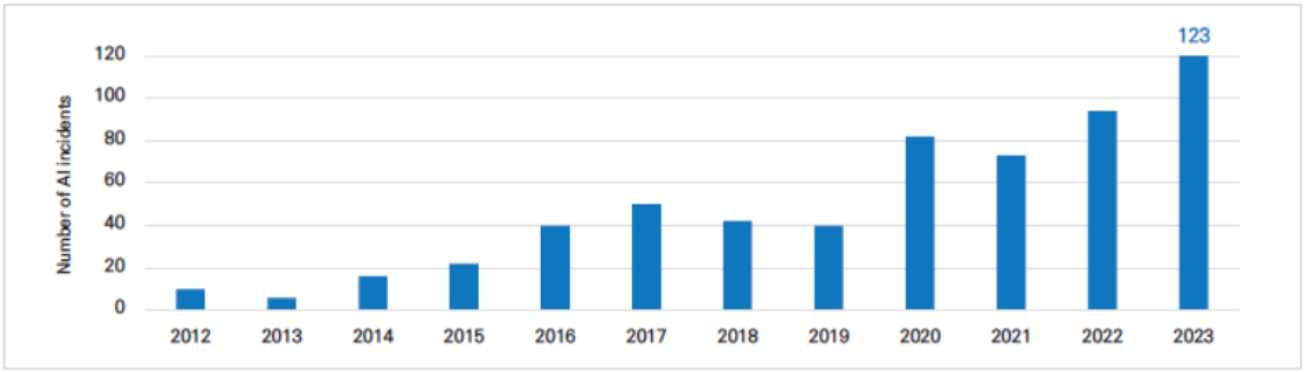
생성형 AI의 기술과 윤리

생성형 AI의 윤리적 이슈

- AI활용 시 개인정보 침해, 편향성 문제 등의 윤리 문제 발생
  - 사고과정 이해하지 못함, 비윤리적 목적이 있는 AI 등 AI위험요소에 대한 통제력 상실 발생
- 이로 인한 심각한 사회적 문제를 가져올 것에 대한 우려로 각국 정부는 생성형 AI를 포함한 인공지능의 역기능 대응을 위한 각종 제도적 보완책 마련에 집중하고 있음
- 생성형 AI의 경쟁이 본격화 되면서 무분별한 오남용과 이로 인한 역기능으로 이용자가 서비스를 회피하거나 직접적인 피해를 입는 것을 예방하기 위해 적법한 조치를 강구하고 있음
  - 저작권 등 분쟁, 소송 및 손해배상 대응

AI 활용의 위험성

〈그림 4〉 AI 사고 현황(2012~2023년)



출처) HAI(2024)

- AI의 무기화
  - 대량 사이버공격, 생화학 무기제조
  - 24년 3월 이스라엘이 하마스 전쟁에서 AI 자율살상무기를 활용
- AI시스템에 대한 통제력 상실
  - 우크라이나군의 AI 드론이 자체 판단으로 러시아군을 공격
- AI 시스템에 대한 공격
- 예측 불가능한 기술 발전
  - 딥페이크 기술로 인한 가짜뉴스, 프라이버시 침해 등 인간을 대상으로 한 범죄 행위 증가
- AI시스템의 오작동

생성형 AI관련 윤리원칙

개발과 규제 논의

- 개발과 규제의 충돌사례
  - AI의 산업분야의 활용은 인간에게 경제적 편익을 제공하지만 예기치 못한 위험의 발생도 야기하여 규제에 관한 논의가 지속적으로 이어짐
  - AI 기술 개발과 윤리 간의 의견대립으로 오픈AI CEO Sam Altman은 이사회 내부와 가치관의 차이(수익성 vs AI 안전성)로 해임(’23.11.)되었다 복귀함
- 이러한 문제로 EU는 세계 최초로 포괄적인 AI 규제법을 통과(’24.3.)시켜 기술 규제에 대한 보완책으로 생성형 AI 진흥안을 마련함
- AI 관련 각종 사고들이 발생함에 따라 AI 개발에 대한 공동 약속인 아실로마 인공지능 원칙(Asilomar AI Principles) 발표
  - 인공지능 원칙은 연구 이슈(5개), 윤리적 가치(13개), 장기적 이슈(5개) 등 세 가지 범주로 구성

Asilomar AI Principles_Research Issues

Research Issues	
Research Goal	The goal of AI research should be to create not undirected intelligence, but beneficial intelligence.
Research Funding	Investments in AI should be accompanied by funding for research on ensuring its beneficial use, including thorny questions in computer science, economics, law, ethics, and social studies.
Science-Policy Link	There should be constructive and healthy exchange between AI researchers and policy-makers.
Research Culture	A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI.
Race Avoidance	Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards.

https://futureoflife.org/open-letter/ai-principles/

Asilomar AI Principles_Ethics and Values

Ethics and Values	
Safety	AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible.
Failure Transparency	If an AI system causes harm, it should be possible to ascertain why.
Judicial Transparency	Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.
Responsibility	Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications.
Value Alignment	Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.
Human Values	AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.

https://futureoflife.org/open-letter/ai-principles/

Asilomar AI Principles_Ethics and Values

Ethics and Values	
Personal Privacy	People should have the right to access, manage and control the data they generate, given AI systems' power to analyze and utilize that data.
Liberty and Privacy	The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.
Shared Benefit	AI technologies should benefit and empower as many people as possible.
Shared Prosperity	The economic prosperity created by AI should be shared broadly, to benefit all of humanity.
Human Control	Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.
Non-subversion	The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.
AI Arms Race	An arms race in lethal autonomous weapons should be avoided.

https://futureoflife.org/open-letter/ai-principles/

생성형 AI관련 윤리정책

해외 인공지능 윤리정책

- UN과 OECD, G20/G7 등 국제기구와 협의체는 AI 윤리의 중요성을 인식하여 AI 윤리 및 신뢰성 관련 원칙 제정 및 파트너십을 통한 글로벌 협력을 추진함
- EU는 AI에 관한 세계 최초의 포괄적인 규제로 평가되는 AI 법안을 제정하고 있으며, AI에 따른 위험 통제 및 인권 보호, 투명성·책임성 보장 등을 위한 조치 및 국제협력을 추진 중임
- 미국은 바이든 정부 출범 이래 AI 기술 개발 및 활용을 촉진하는 동시에 AI에 의한 위험 관리와 시민 권리 보호를 위한 정책을 발표함
- 중국은 AI 사용자의 보호 및 콘텐츠 관리를 강화하고, 불량 정보의 확산을 방지하기 위해 엄격한 규제와 공급자의 의무를 부가함
    - 2023년 7월 정부는 ‘생성형 AI 서비스 관리 방법’을 통해 자국 내 콘텐츠를 생성하는 AI 서비스의 관리지침을 발표함

해외 인공지능 윤리정책

- 영국은 AI 주요 발전 국가로 AI의 혁신 촉진과 함께 위험에 대응하고 공공 신뢰를 높일 수 있도록 유연하고 반복적이며 협력적인 규제를 지향함
- 일본은 혁신을 방해하지 않도록 AI 규제를 최소화하는 방향을 견지해 왔으나, 최근 EU와 미국 등 AI 규제 강화에 대한 논의가 확산함에 따라 대응 방안을 검토 중임
- 세계 각국은 AI 윤리와 신뢰성 확보를 위해 자국 내 정책적 대응을 위한 원칙을 세워 규제 및 행정적 조치를 적극적으로 실행하고 있음
    - G7 회원국은 첨단 AI 시스템의 위험을 완화하기 위한 종합 AI 정책 프레임워크를 발표하고 AI 개발자가 준수해야 할 행동 강령과 모든 AI 사용자가 지켜야 할 지침을 제시함

국내 인공지능 윤리정책

- 과기정통부는 2023년 9월 ‘대한민국 인공지능 도약방안’을 발표하고, 민간 주도의 AI 윤리·신뢰성 확보 방안과 신뢰성 R&D 추진을 밝힘
  - 민간이 윤리 원칙을 준수하기 위해 자율적으로 운영하는 윤리위원회를 구성·운영하는 표준지침을 수립함
- 2023년 10월「인공지능 윤리·신뢰성 확보 추진계획」을 발표하고 중점 추진 방향을 제시함
  - AI 산업 발전의 전제 조건인 AI 윤리·신뢰성 확보 지원을 위해 분야별 가이드라인 확대 및 민간 자율 신뢰성 검·인증을 추진함
  - AI 위험성에 대응하기 위한 기술 개발 지원을 목적으로 AI 신뢰성 확보 기술 개발 및 AI 평가 데이터 신뢰성 확보 계획을 밝힘

인공지능과 인간의 공존

- 기술 발전으로 인한 문제에 유연하게 대응하기 위해서는 인공지능 개발자, 사용자, 정책입안자 각자의 역할에 따라 윤리적 책임을 지도록 해야 함
- AI 활용성 제고를 위한 AI 자원 접근성을 높이려고 하는 노력과 동시에 AI의 잠재적 위험성 감소를 위한 인간과 인공지능의 상호 성장을 위한 균형이 필요함
- 또한 인간과 AI가 조화롭게 공존하기 위해서는 AI 활용 목적이 인간의 가치와 일치해야 하므로, AI 기술을 활용하는 데 있어 인간의 윤리 의식과 도덕적 판단력이 중요함
  - 다양한 기구에서 발표되고 있는 AI 설계를 위한 지침도 공통적으로 인간의 가치관에 부합하게 인공지능이 작동하기 위해 필요한 사항을 포함하고 있음(Jacob Turner, 2023)
- AI 확산에 따라 발생할 수 있는 이슈를 도출하고, 각각에 대한 결과를 예측하고 준비할 수 있는 정책 설계가 필요함

감사합니다

참고문헌: 한국지능정보사회진흥원(2023). 생성형AI윤리 가이드북
한국과학기술정보연구원(2024). 인공지능 윤리:
인간과 인공지능의 조화로운 공존방안