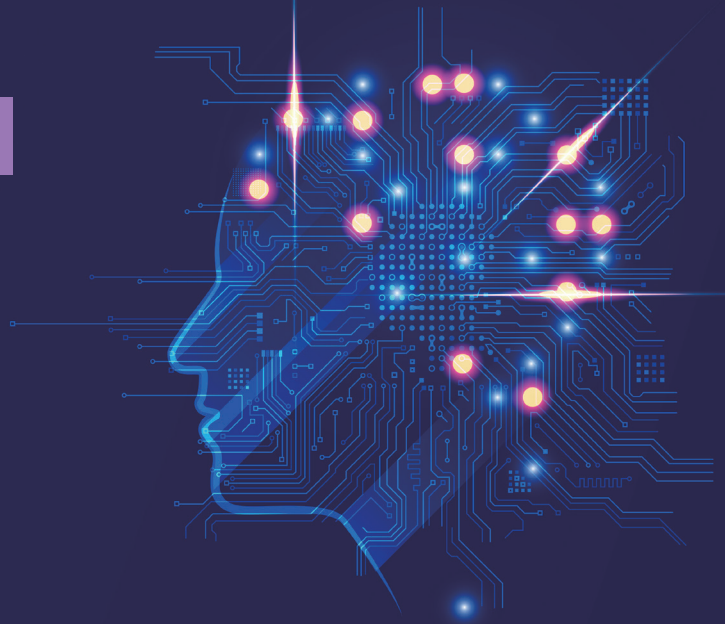


생명윤리 유관기관 공동학술대회

인공지능과 거버넌스



일시

2026. 6. 10.(수) 10:00 ~ 17:30

장소

유네스코 회관 11층 유네스코 홀(온라인 병행)

프로그램

세션	시간	주제발표	연자
사회: 김근 실장(국가생명윤리정책원)			
축사	10:00 - 10:05	축사	홍현익 사무총장 (유네스코한국위원회)
	10:05 - 10:10	축사	홍창권 원장 (국가생명윤리정책원)
기조 발제	10:10 - 10:35	공공기능과 글로벌 헬스 거버넌스	차지호 의원 (국회의원)
	10:35 - 11:00	인공지능 시대의 생명윤리 거버넌스와 국가생명윤리심의위원회의 역할	김옥주 위원장 (국가생명윤리심의위원회, 서울대학교 의과대학 교수)
[제1부] 이론적 담론 11:00 - 12:30			
좌장: 유상호 교수(한양대학교 의과대학, 한국의료윤리학회 회장)			
제1부	11:00 - 11:20	인공지능은 공공재적 성격을 지니는가?	최경석 교수 (이화여자대학교 법학전문대학원, 이화여자대학교 생명의료법연구소 소장)
	11:20 - 11:40	인공지능과 책임의 확장	윤준식 선임연구원 (서울대학교 교육융합연구원, 한국윤리학회 교육이사)
	11:40 - 12:00	인공지능시대의 새로운 개발 윤리 모색	목광수 교수(서울시립대학교, 한국윤리학회 회장)
	12:00 - 12:30	패널토론 및 종합토론	심지원 교수(동국대학교) 임시연 선임전문관(유네스코한국위원회)
Lunch time	12:30 - 14:00	점심시간	
[제2부] 실천적 담론 14:00 - 17:30			
좌장: 이일학 교수(연세대학교 의과대학, 한국생명윤리학회 회장)			
제2부	14:00 - 14:20	공공의료 AI 거버넌스	김준혁 교수 (연세대학교 치과대학, 한국의료윤리학회 간행이사)
	14:20 - 14:40	의료임상 현장과 AI	박소연 교수 (인하대학교 의과대학, 한국의료윤리학회 고시이사)
	14:40 - 15:00	보건정책과 인공지능	이경도 교수 (울산대학교 의과대학, 한국생명윤리학회 학술이사)
	15:00 - 15:15	쉬는시간	
	15:15 - 15:35	데이터은행, 인공지능 개발과 활용의 공공성	김한나 교수 (연세대학교 의과대학, 한국생명윤리학회 총무이사)
	15:35 - 15:55	의료인공지능과 민형사상 책임의 귀속	김천수 교수 (성균관대학교 법학전문대학원, 대한의료법학회 상임고문) 최민영 교수 (강남대학교 법행정학과, 대한의료법학회 총무이사)
	15:55 - 16:15	피지컬 AI와 생명의료법	엄주희 교수 (건국대학교, 한국의료법학회 부회장)
	16:15 - 17:00	패널토론	문재영 교수 (충남대학교 의과대학) 박지용 교수 (연세대학교 법학전문대학원) 이일호 교수 (연세대학교) 전재필 부장 (국립보건연구원) 김민우 교수 (충북대학교)
	17:00 - 17:30	종합토론	

[축사]

홍현익 사무총장 (유네스코한국위원회)	4
홍창권 원장 (국가생명윤리정책원)	6

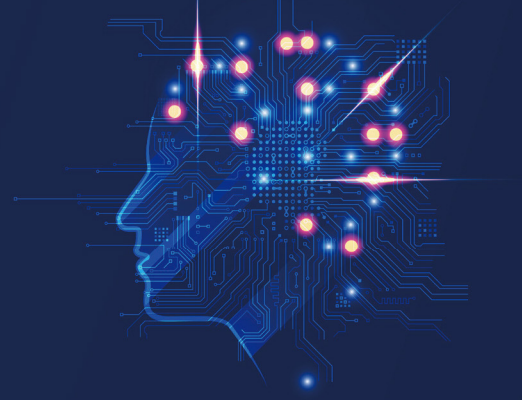
[기조발제]

• 공공지능과 글로벌 헬스 거버넌스	7
차지호 의원 (국회의원)	
• 인공지능 시대의 생명윤리 거버넌스와 국가생명윤리심의위원회의 역할	8
김옥주 위원장 (국가생명윤리심의위원회, 서울대학교 의과대학 교수)	

[제1부 이론적담론]

좌장: 유상호 회장(한국의료윤리학회, 한양대학교 의과대학)

• 인공지능은 공공재적 성격을 지니는가?	19
최경석 소장(이화여자대학교 생명의료법연구소, 이화여자대학교 법학전문대학원)	
• 인공지능과 책임의 확장	27
윤준식 선임연구원 (서울대학교 교육종합연구원, 한국윤리학회 교육이사)	
• 인공지능시대의 새로운 개발 윤리 모색	40
목광수 교수 (서울시립대학교, 한국윤리학회 회장)	
• 패널토론 및 종합토론	
심지원 교수 (동국대학교)	47
임시연 선임전문관 (유네스코한국위원회)	48



[제2부 실천적담론]

좌장: 이일학 회장(한국생명윤리학회, 연세대학교 의과대학)

• 공공의료 AI 거버넌스	51
김준혁 교수 (연세대학교 치과대학, 한국의료윤리학회 간행이사)	
• 의료임상 현장과 AI	59
박소연 교수 (인하대학교 의과대학, 한국의료윤리학회 고시이사)	
• 보건정책과 인공지능	69
이경도 교수 (울산대학교 의과대학, 한국생명윤리학회 학술이사)	
• 데이터은행, 인공지능 개발과 활용의 공공성	81
김한나 교수 (연세대학교 의과대학, 한국생명윤리학회 총무이사)	
• 의료인공지능과 민형사상 책임의 귀속	88
김천수 교수 (성균관대학교 법학전문대학원, 대한의료법학회 상임고문)	
최민영 교수 (강남대학교 법행정학과, 대한의료법학회 총무이사)	
• 피지컬 AI와 생명의료법	97
엄주희 교수 (건국대학교, 한국의료법학회 부회장)	
• 패널토론	
문재영 교수 (충남대학교 의과대학)	107
박지용 교수 (연세대학교 법학전문대학원)	110
이일호 교수 (연세대학교)	111
전재필 부장 (국립보건연구원)	112
김민우 교수 (충북대학교)	113

안녕하십니까.

유네스코한국위원회사무총장 홍현익입니다.

먼저 ‘인공지능과 거버넌스’를 주제로 열리는 생명윤리 유관기관 공동학술대회의 개최를 진심으로 축하드립니다.

오늘 자리를 함께 마련해 주신 국가생명윤리심의위원회와 국가생명윤리정책원, 그리고 생명윤리와 의료윤리, 윤리학과 의료법 등 각 분야의 논의를 이끌어 오신 여러 학회와 관계기관 여러분께 깊이 감사드립니다. 아울러 인공지능 시대의 생명윤리와 거버넌스라는 중요한 과제를 함께 논의하기 위해 자리해 주신 발표자와 토론자, 참석자 여러분께도 감사의 말씀을 드립니다. 유네스코한국위원회도 이번 학술대회의 공동주관기관으로 참여하게 되어 매우 기쁘게 생각합니다.

인공지능은 더 이상 특정 산업이나 일부 전문가의 영역에 머물러 있지 않습니다. 우리가 배우고 일하는 방식, 서로 소통하고 문화를 향유하는 방식, 공공의 문제를 해결하고 미래를 준비하는 방식에 이르기까지, 인공지능은 이미 우리 삶의 여러 장면을 깊이 바꾸어가고 있습니다.

특히 의료와 보건의 영역에서 인공지능은 더욱 각별한 의미를 지닙니다. 질병을 보다 일찍 발견하고, 환자 한 사람 한 사람에게 필요한 치료를 세심하게 제공하며, 의료서비스가 충분히 닿지 못했던 곳까지 돌봄의 가능성을 넓힐 수 있기 때문입니다.

그러나 기술의 진보가 곧바로 모두의 진보를 의미하는 것은 아닙니다. 인공지능이 판단한 것이 누구의 가치와 관점을 반영하고 있는가. 평균적인 정확도와 효율성이라는 이름 아래, 오히려 더 세심한 보호가 필요한 사람들의 삶과 경험이 지워지고 있지는 않은가. 생명과 건강에 직접 영향을 미치는 결정 과정에서 인간의 판단과 책임은 어디에 위치해야 하는가.

오늘 학술대회의 주제인 ‘거버넌스’는 바로 이러한 질문에 함께 답해가는 과정이라고 생각합니다.

유엔은 2025년 제79차 유엔 총회의 결의에 따라 전 세계 전문가 40명으로 구성된 유엔 AI 과학 패널(Independent International Scientific Panel on Artificial Intelligence)을 출범시키고, 올해 3월 첫 회의를 개최하여 AI가 가져올 기회와 위험을 증거 기반의 과학 차원에서 점검하는 체계를 본격 가동했습니다. 첫 보고서가 7월 제네바에서 개최되는 AI 거버넌스 글로벌 대화(Global Dialogue on AI Governance) 회의에서 발표됩니다. 유엔 결의 후속조치에 유네스코가 공동사무국으로 참여하고 있고, 유네스코가 만든 규범이 논의의 근간을 이루고 있습니다.

유네스코는 새로운 기술이 등장할 때마다 기술의 발전 자체에만 주목하지 않았습니다. 과학기술이 인간의 존엄성과 인권, 평화롭고 공정한 사회에 기여할 수 있도록 국제사회가 공유할 수 있는 규범과 원칙을 마련해 왔습니다.

그 중요한 이정표 가운데 하나가 2021년 제41차 유네스코 총회에서 채택된 ‘인공지능 윤리 권고’입니다. 이 권고는 인공지능의 개발과 활용이 인간의 존엄성과 인권, 다양성과 포용성, 지속가능성을 존중하는 방향으로 이루어져야 한다는 보편적 기준을 제시했습니다.

‘인공지능 윤리 권고’의 이행을 위한 유네스코의 노력은 선언적 원칙에 머물지 않고 있습니다. 유네스코는 구체적인 이행도구로서 준비도평가(RAM)와 윤리영향평가(EIA)를 발전시켜 왔습니다. 이미 70개 이상의 회원국이 RAM에 참여하여 자국의 법·제도와 정책, 데이터 거버넌스, 인권 보호장치와 역량 수준을 점검하고, 정책적 공백과 개선 과제를 진단하고 있습니다. 또한 EIA를 통해 개별 인공지능 시스템과 서비스가 인권과 사회적 가치, 취약집단, 투명성과 책임성에 미칠 영향을 사전에 점검하도록 지원하고 있습니다.

아울러 회원국의 이행 현황을 정기적으로 확인하고 정책 경험과 우수사례를 공유하는 국가보고 체계를 운영하며, 글로벌 AI 윤리 포럼과 교육·사법·미디어·정보 리터러시 등 분야별 지침과 역량강화 사업을 통해 국제규범을 실제

정책과 현장에 연결하고 있습니다.

그리고 2025년 제43차 유네스코 총회에서는 ‘신경기술 윤리 권고’가 새롭게 채택되었습니다.

이는 단순히 하나의 새로운 기술 분야가 국제규범의 대상으로 추가되었다는 의미에 그치지 않습니다. 과학기술의 윤리적 과제가 개인정보와 데이터 보호를 넘어, 인간의 정신적 프라이버시와 자율성, 정체성, 자유로운 사고와 의사결정의 보호라는 더욱 근본적인 영역으로 확장되고 있음을 보여줍니다.

특히 인공지능과 신경기술이 결합하면서 두 권고의 접점은 더욱 중요해지고 있습니다. 인공지능은 신경 데이터와 다양한 생체·행동 데이터를 분석하여 질병의 진단과 치료, 재활을 지원할 수 있습니다. 그러나 다른 한편으로는 인간의 생각과 감정, 집중도와 의도를 추론하고, 인지와 행동에 영향을 미치는 방향으로 활용될 가능성도 있습니다.

윤리와 거버넌스는 기술이 사회에 자리 잡은 이후 덧붙이는 보완 장치가 되어서는 안 됩니다. 무엇을 개발할 것인지, 어디까지 허용할 것인지, 어떤 위험은 처음부터 감수하지 않을 것인지를 정하는 기준으로서 기술의 출발점에 놓여야 합니다.

오늘 학술대회가 갖는 의미도 바로 여기에 있습니다. 이번 학술대회에서는 공공재로서의 인공지능이라는 근본적인 질문에서 출발하여, 새로운 개발 윤리, 의료현장의 인공지능 활용, 공공의료와 보건행정, 데이터의 공공성, 책임 귀속, 피지컬 인공지능의 생명의료법적 과제까지 폭넓은 쟁점을 다룹니다.

인공지능 거버넌스의 핵심은 기술을 도입할 것인지 여부에만 있지 않습니다. 그 기술이 누구의 삶을 개선하고, 누구의 필요를 놓치고 있으며, 그 결과에 대하여 누가 답할 것인지를 묻는 데 있습니다.

특히 의료와 보건의 영역에서는 평균적인 성능을 넘어, 의료취약지와 취약계층을 포함한 모든 사람에게 실제로 유효하고 공정하게 작동하는지를 세심하게 살펴야 합니다. 인공지능의 활용이 확대될수록 책임이 사라지는 것이 아니라, 오히려 책임의 범위는 더욱 넓어져야 합니다.

무엇보다 오늘 이 자리는 생명윤리와 신경기술윤리 분야에서 축적해 온 깊이 있는 논의가 인공지능이라는 시대적 과제와 만나, 서로 다른 전문영역을 잇는 새로운 논의로 확장되는 계기입니다.

유네스코한국위원회는 인공지능 및 신경기술 윤리에 대한 유네스코의 국제규범을 국내에 소개하는 데 그치지 않고, 이를 둘러싼 논의를 모으고 연결하는 역할을 충실히 해 나가고자 합니다. 이번 학술대회를 하나의 출발점으로 삼아, 정부와 연구기관, 의료계와 법조계, 학계와 산업계, 시민사회가 각자의 관점과 경험을 나눌 수 있는 자리를 지속적으로 마련하겠습니다. 서로 다른 전문영역이 만나 새로운 질문을 발견하고, 우리 사회에 필요한 원칙과 실천 과제를 함께 모색할 수 있도록 협력의 기반을 넓혀가겠습니다.

아울러 국내에서 축적된 논의와 경험을 아시아·태평양 지역과 공유하고, 유네스코의 국제 논의에도 적극적으로 환류하겠습니다. 국내 이행을 넘어 지역과 국제사회의 지혜를 모으고, 첨단기술이 인간을 위한 기술로 발전할 수 있도록 책임 있는 협력의 장을 꾸준히 만들어가겠습니다.

오늘 이 자리가 인공지능과 신경기술이 인간의 존엄성과 공공의 이익을 중심에 두고 발전하기 위해 필요한 원칙과 제도를 함께 모색하고, 책임 있는 혁신을 위한 사회적 지혜를 모으는 뜻깊은 장이 되기를 기대합니다.

다시 한번 학술대회의 개최를 진심으로 축하드리며, 함께해 주신 모든 분의 깊이 있는 논의와 풍성한 성과를 기원합니다. 감사합니다.

안녕하십니까?

국가생명윤리정책원 원장 홍창권입니다.

먼저 「인공지능과 거버넌스」를 주제로 개최되는 공동학술대회의 개최를 진심으로 축하드립니다.

특히 이번 학술대회는 국가생명윤리심의위원회와 국가생명윤리정책원, 대한의료법학회, 유네스코한국위원회,

이화여자대학교 생명의료법연구소, 한국생명윤리학회, 한국윤리학회, 한국의료법학회, 한국의료윤리학회 등이 함께 뜻을 모아 마련한 자리로, 다양한 분야의 전문가들이 모여 인공지능 시대의 생명윤리와 거버넌스를 함께 논의한다는 점에서 더욱 뜻깊은 의미를 지닌다고 생각합니다.

공동학술대회를 준비해주신 모든 기관과 관계자 여러분께 깊은 감사의 말씀을 드립니다.

오늘날 인공지능 기술은 보건의료와 생명과학 분야 전반에 빠르게 확산되며, 연구와 진료는 물론 정책과 행정의 방식까지 크게 변화시키고 있습니다.

특히 생성형 인공지능과 데이터 기반 기술의 발전은 의료의 효율성과 접근성을 높이는 동시에, 개인정보 보호와 책임의 귀속, 공공성과 형평성, 인간의 존엄성 보장과 같은 새로운 윤리적·법적 과제를 함께 제기하고 있습니다.

이러한 변화 속에서 인공지능 기술을 어떻게 사회적으로 수용하고, 어떠한 기준과 원칙 아래 활용해 나갈 것인가는 더 이상 특정 분야만의 문제가 아니라 우리 사회 전체가 함께 고민해야 할 중요한 과제가 되었습니다.

그런 의미에서 오늘 학술대회는 인공지능 시대의 생명윤리와 거버넌스를 이론과 실천의 측면에서 함께 성찰하고, 공공성과 책임성을 기반으로 한 미래의 방향을 모색하는 매우 뜻깊은 자리가 될 것으로 생각합니다.

특히 인공지능과 글로벌 헬스 거버넌스, 생명윤리 거버넌스의 방향, 의료현장과 보건정책에서의 AI 활용과 책임 문제 등 다양한 주제를 폭넓게 다루는 오늘의 논의는 향후 우리 사회의 정책과 제도 설계에도 중요한 시사점을 제공할 것입니다.

또한 인공지능 기술이 의료와 생명과학 분야에 빠르게 확산되는 상황에서, 기술 발전과 인간의 존엄성, 사회적 신뢰와 공공의 가치가 조화를 이룰 수 있는 방향을 모색하는 일은 무엇보다 중요하다고 생각합니다.

오늘 이 자리가 인공지능 시대에 필요한 생명윤리의 방향과 사회적 책임을 함께 모색하는 뜻깊은 논의의 장이 되기를 기대합니다.

국가생명윤리정책원은 국내 유일의 생명윤리 공공기관으로서, 2026년 인공지능 시대의 생명윤리 거버넌스를 중요한 정책 과제로 삼고 있습니다.

이를 위해 학계와 의료계, 법조계와 시민사회, 정책 현장을 잇는 생명윤리 논의의 구심점으로서, 사회적 공감과 협력을 바탕으로 책임 있는 거버넌스가 마련될 수 있도록 지속적으로 노력하겠습니다.

다시 한번 공동학술대회의 개최를 축하드리며, 오늘 함께해주신 모든분들께 뜻깊고 유익한 시간이 되기를 바랍니다.

감사합니다.

국가생명윤리정책원장 홍창권

공공지능과 글로벌 헬스 거버넌스

차지호 의원 (국회의원)

인공지능 시대의 생명윤리 거버넌스와 국가생명윤리심의위원회의 역할

김옥주 위원장 (국가생명윤리심의위원회, 서울대학교 의과대학 교수)

인공지능 시대의 생명윤리 거버넌스와 국가생명윤리심의위원회의 역할

기조발제 | 10:35-11:00

김옥주 | 서울대학교 의과대학 교수 · 국가생명윤리심의위원회 위원장

국가생명윤리정책원 공동학술대회 「인공지능과 거버넌스」

2026. 6. 10.(수) | 유네스코 회관 11층 유네스코 홀

발표 개요

01

왜 지금 생명윤리 거버넌스인가

AI 전환이 던지는 근본적 질문

02

국제 규범이 말하는 것

UNESCO · WHO의 공통 메시지

03

한국의 현실

거버넌스의 파편화와 조정 과제

04

국가생명윤리심의위원회의 역할 재정의

국가 윤리 거버넌스 플랫폼으로

05

오늘 우리가 함께 답해야 할 질문들

이론적 · 실천적 담론을 위한 과제

AI 국가전략 — 그 다음 질문

01

앞 발제 · 차지호 의원

공공지능과 글로벌 헬스 거버넌스

- AI 기술의 국가전략적 가치
- 글로벌 헬스에서의 AI 활용
- 산업 · 혁신 · 경쟁력 담론



이 발제 · 김옥주

그 전략은 어떻게 정당화되는가?

- 인간 존엄과 사회적 신뢰의 기반 위에서
- 생명윤리 거버넌스의 책임과 역할
- 국생위의 위치와 과제

AI 혁신의 속도와 방향은 기술만이 결정하지 않습니다. 어떤 가치 위에서 혁신할 것인가 — 이것이 오늘의 질문입니다

AI 전환이 생명윤리에 던지는 근본적 질문

Q. 인간이란 무엇인가?

AI가 진단하고 처방하고 돌보는 시대에 의사-환자 관계, 의료적 판단, 돌봄의 의미는 어떻게 재정의되는가

Q. 동의는 여전히 유효한가?

의료·유전체 데이터가 AI 학습에 활용될 때, 기존의 '정보에 근거한 동의' 체계는 여전히 정보주체를 보호할 수 있는가

Q. 책임은 누구에게 있는가?

AI의 오진, 알고리즘 편향, 데이터 유출 — 개발자인가, 의사인가, 병원인가, 국가인가. 책임의 공백을 누가 채우는가

Q. 공정한 혜택 분배는 가능한가?

AI 기술의 혜택이 경제적·사회적으로 취약한 집단에게도 공평하게 돌아가는가, 아니면 불평등을 심화시키는가

국제 규범이 말하는 것: 공통 메시지

02

UNESCO
2021

AI 윤리 권고
(193개국 채택)

WHO
2024

보건의료 AI
윤리.거버넌스 가이드
스

EU AI Act
2024

위험 기반
규제 체계 발효

세 규범이 공통적으로 강조하는 것: 기술 발전 자체가 아니라 "신뢰할 수 있는 AI(Trustworthy AI)"

인간 자율성 보호

AI는 인간의 판단을 대체하지 않고 보조한다

투명성 · 설명가능성

이해관계자 모두가 AI를 이해할 수 있어야 한다

책임성 · 책무성

AI 결정으로 인한 피해에 명확한 구제 경로가 있어야 한다

포용성 · 형평성

취약계층을 포함한 모든 이에게 공평한 혜택과 보호

선제적 거버넌스

기술이 사회에 미치는 영향을 사전에 평가하고 대응한다

한국의 현실: 파편화된 AI 거버넌스

03

과기부

AI 기본법
데이터 산업 진흥

개인정보위

AI 학습 데이터
개인정보 규제

식약처

의료기기 AI
소프트웨어 규제

분절된
AI 거버넌스

복지부

보건의료 AI
생명윤리법 소관

국생위

생명윤리 심의
→ 역할 재정의 필요

국과위

AI 전략 총괄
범부처 조정

IRB · DRB · 개인정보보호체계 등 기존 거버넌스의 토대 위에서 — AI 시대 생명윤리 관점의 통합 조정이 과제

기존 거버넌스의 토대 위에서

우리 사회는 이미 IRB, 데이터심의위원회(DRB), 개인정보보호체계 등을 통해 의료데이터 활용에 대한 다층적 거버넌스를 발전시켜 왔습니다.

앞으로의 과제는 이러한 체계를 약화시키는 것이 아니라, AI 시대에 맞게 더욱 신뢰 가능하고 정교하게 발전시키는 데 있습니다.

IRB (기관생명윤리위원회)

인간대상연구 보호의 핵심 기제. AI 시대
심의 기준 정교화 필요

DRB (데이터심의위원회)

의료데이터 활용 적정성 심의. 재식별·접
근통제 다층적 검토

개인정보보호체계

개인정보보호법·가명정보 가이드라인.
AI 학습 데이터 기준 고도화

국생위의 역할: 기존 체계의 파괴가 아니라 AI 시대에 맞는 정교화·연계·조정

Section 04

국가생명윤리심의위원회

역할 재정의

AI가 발전할수록 더욱 중요해지는 것이 있습니다.

"기술이 아무리 빠르게 달려도
생명윤리의 원칙은 양보할 수 없다"

— 그 원칙을 세우는 기구

역할 ①②: 국가 윤리 공론장 · 부처 간 조정

04

① 국가 윤리적 공론장 주도

AI 문제는 기술·법·산업의 문제이면서 동시에 인간권·사회정의·민주주의의 문제입니다.

- ▶ 다학제·시민 참여 속의 과정 설계
- ▶ 기술이 가져오는 사회적 변화의 공적 논의
- ▶ 약자·소수자 목소리 포함한 포용적 논의
- ▶ AI·바이오데이터 정책 논의에 윤리적 속의 지원

② 부처 간 조정 플랫폼

분절된 부처들을 넘어 생명윤리 관점의 통합 조정이 필요합니다.

- ▶ 과기부 AI 기본법 ↔ 생명윤리 원칙 연계
- ▶ 개인정보위 AI 학습 기준 ↔ 동의 원칙
- ▶ 식약처 의료AI 승인 ↔ 윤리 요건
- ▶ 부처 간 AI 윤리 거버넌스 협력 논의 지원

역할 ③④: 사회적 신뢰 형성 · 국제규범의 한국적 구현

04

③ "국생위의 역할은 기술을 늦추는 것이 아니라, 국민이 신뢰할 수 있는 AI 활용 환경을 만드는 것이다"

✓ AI 개발·활용 과정에서의 윤리 원칙과 권고 기준 마련

✓ 취약집단의 AI 접근성과 권리 보장을 위한 원칙 수립

✓ 피해 발생 시 책임성 있는 구제 경로 보장

✓ 시민 AI 리터러시 및 사회적 수용성 제고 프로그램

④ 국제 생명윤리 규범의 한국적 구현

UNESCO AI 윤리 권고



AI 기본법·하위 법령 생명윤리 반영

WHO 보건의료 AI 지침



의료AI 개발·제공·배포 가이드라인

헬싱키 선언(2024) · 타이페이 선언 개정 논의



연구참여자 중심 AI 연구윤리 기준 수립

글로벌 AI 프라이버시 선언



AI 학습 데이터 보호 국내 기준

역할 ⑤⑥: 연구대상자 보호 · 선제적 거버넌스

04

⑤ AI 시대 인간대상연구의 신뢰 기반 원칙 수립

- ▶ IRB·DRB 등 기존 체계와의 협력 및 연계 강화
- ▶ 사회적 신뢰 확보를 위한 데이터 거버넌스 체계 발전
- ▶ AI 학습 데이터 활용의 동의 원칙 재설계
- ▶ 취약집단의 안전한 AI 접근과 권리 보장 원칙 수립
- ▶ 알고리즘 공정성·책임성 원칙 수립

⑥ 선제적 거버넌스 — 미래 위협 대응

"사후 규제기관"이 아니라
"선제적 윤리 거버넌스 플랫폼"

- ⚡ 생성형 AI 의료 조건의 책임 귀속
- ⚡ 유전체 AI 기반 보험·고용 차별
- ⚡ AI 기반 정신건강 분석의 인격권
- ⚡ 뇌-컴퓨터 인터페이스(BCI) 생명윤리

Section 05

오늘 우리가 함께 답해야 할 질문들

이론적 담론 · 실천적 담론 · 그리고 국생위의 과제

제1부 이론적 담론을 위한 질문

05

좌장: 유상호 교수 · 11:00-12:30

최경석 교수 | 인공지능은 공공재적 성격을 지니는가?

Q. AI가 공공재라면, 그것을 개발하고 통제하는 방식도 공적 규범의 지배를 받아야 하는가?
국가생명윤리심의위원회는 이 공공재 논의에서 어떤 역할을 맡아야 하는가?

윤준식 선임연구원 | 인공지능과 책임의 확장

Q. AI 시스템에서 발생하는 피해에 대해 책임을 분산시키는 것은 정당한가?
생명윤리 관점에서 '책임의 공백'을 메우는 제도적 장치는 무엇인가?

목광수 교수 | 인공지능 시대의 새로운 개발 윤리 모색

Q. AI 개발자에게 요구되는 윤리적 책무는 무엇인가?
생명윤리의 기존 원칙들 - 자율성, 선행, 비해악, 정의 - 은 AI 시대에도 유효한가?

제2부 실천적 담론을 위한 질문

05

좌장: 이일학 교수 · 14:00-17:30

김준혁·박소연·이경도 교수 | 공공의료 AI / 임상현장 / 보건의 정책

Q. 의료 현장에서 AI를 신뢰하려면 무엇이 필요한가?
환자의 자율성과 임상적 판단, 그리고 AI 추천 - 어떻게 균형을 잡을 것인가?

김한나 교수 | 데이터은행, 인공지능, 개발과 활용의 공공성

Q. 의료·유전체 데이터를 AI가 학습하는 것을 어떻게 공익적으로 정당화할 수 있는가?
데이터 주권과 혁신 사이의 균형점은 어디인가?

김천수·최민영 교수 / 임주희 교수 | 민형사상 책임 귀속 / 피지컬 AI와 생명의료법

Q. AI가 야기한 의료 피해의 법적 책임은 어떻게 귀속되어야 하는가?
피지컬 AI·로봇 수술 등 새로운 기술은 기존 생명의료법의 틀로 규율 가능한가?

국가생명윤리심의위원회 7기가 나아가는 방향

2026. 6. 5. 국가생명윤리심의위원회 7기 출범 세미나 논의 방향을 바탕으로

의제 1+2

AI 시대 생명윤리의 원칙을 세운다

인간 존엄, 연구참여자 보호, 의료·유전체 데이터의 생명윤리적 기준 — AI가 아무리 발전해도 흔들려서는 안 되는 원칙들을 국생위가 수립한다

의제 3

의료·생명과학 AI의 생명윤리 논의를 주도한다

알고리즘 편향, 연구참여자 권리, 유전체 프라이버시 — AI 정책 논의가 놓치기 쉬운 생명윤리 차원의 문제를 국생위가 공론화한다

의제 4

국제 생명윤리 규범과 한국 현실을 연결한다

UNESCO·WHO·헬싱키·타이페이 선언의 생명윤리 원칙이 한국 AI 정책에 반영되는지 점검한다. 국제 생명윤리 담론에서 한국의 목소리를 낸다

의제 5

신기술의 생명윤리 문제를 선제적으로 논의한다

유전체 AI, 뇌·컴퓨터 인터페이스, 생성형 AI 의료 활용 — 기술이 생명윤리의 경계를 넘기 전에 국생위가 먼저 원칙을 세운다

맺음말

인공지능 시대의 핵심 경쟁력은
단순한 기술 속도나 데이터 규모만이 아닙니다.

인간 존엄, 개인정보 보호, 공정성,
사회적 신뢰를 유지하면서 혁신을 지속할 수 있는
거버넌스의 질이야말로
미래 국가 경쟁력의 핵심이 될 것입니다.

국가생명윤리심의위원회는

AI 시대에도 생명윤리가 혁신의 토대가 되도록
"AI 시대 국가 생명윤리 기구"으로서 역할을 해야 합니다.

국가생명윤리정책원 공동학술대회 「인공지능과 거버넌스」 · 2026. 6. 10.



제1부

이론적 담론

좌장: 유상호 회장(한국의료윤리학회, 한양대학교 의과대학)

생명윤리 유관기관 공동학술대회

인공지능과
거버넌스

인공지능은 공공재적 성격을 지니는가?

최경석 소장
(이화여자대학교 생명의료법연구소, 이화여자대학교 법학전문대학원)

[머리말]

생명의료윤리에서의 새로운 기술이 야기하는 복잡한 문제를 분석하고 대응하는 일은 지난 50여년간 생명의료 윤리학의 분야에서 수행해 왔던 일이다. 인공지능기술의 대두와 도입, 그리고 확산의 문제에서 생명의료윤리학자들은 그동안 축적된 이론이나 실천지를 바탕으로 인공지능 기술의 지속가능성과 바람직한 도입, 전개, 확산을 위해 충분히 기여할 바가 있다. 이에 본 글에서는 이런 이론이나 실천지를 바탕으로 우리가 현재 놓치고 있는 중요한 문제가 무엇인지 지적하고, 인공지능 기술의 바람직한 발전을 위해 우리가 무엇을 준비해야 하는지 강조하고자 한다.

01

인공지능의 공공성 여부

전통적으로 공공재로 여겨지는 것의 대표적인 것들에는 물이나 공기, 환경과 같은 것들이 있다. 과연 인공지능이 이런 성격의 공공재라고 여겨질 수 있는지는 의문이다. 인공지능은 데이터를 처리하는 특정 기업이 만들어낸 일종의 소프트웨어 프로그램으로 규정될 수 있기 때문이다. 따라서 전형적인 공공재라고 보기는 어렵다.

하지만 인공지능이 작동하기 위해 필요한 데이터가 어떤 것들인지 생각해 보면, 인공지능은 전형적인 공공재는 아니지만, 공공성을 지니거나 공공성을 지향해야 할 충분한 이유를 지닌다. 여기서 공공성(publicity)이란 두 가지 차원에서 논의될 필요가 있다 하나는 공동의 이익을 추구해야 한다는 측면에서 부여되는 성질이다. 다른 하나는 공동의 이익과 관련된다거나 이익에 영향을 미치는 경우에도 공공성이 강조될 수 있다.

후자의 개념과 관련하여, 예를 들어, 보건의로 체계는 사익과 관련하여 개인의 건강과 안녕, 나아가 생사 문제와 직결된다. 하지만 한 사회의 구성원인 개인의 건강과 안녕의 문제는 사회의 안전성이란 측면에서 공동의 이익과 관련된다. 이런 측면에서 즉 사회를 구성하는 개인의 안녕이나 이익과 관련된다는 측면에서 그리고 나아가 사회의 안전성이란 측면에서, 인공지능이 공동의 이익과 관련된다는 점은 부인하기 어렵다. 간단한 예로 인공지능이 적용된 소위 '자율주행차량'은 그 운용과정에서 발생하는 위험 여부에 대해 안전성을 제공하지 않는다면 우리 사회가 수용하기 어려운 기술이 될 것이다.

우리는 설사 어떤 기술이 사익에 주로 기여하는 기술이라 할지라도, 그 기술이 사회에 미치는 영향에 대해 검토해 왔다. 그래서 새로운 기술에 대한 공적 통제의 필요성이 정당성을 부여 받았다. 해당 기술이 매우 긴요한 기술이거나 그 사용의 필요성이 인정되면서도, 그 영향력이 광범위한 경우, 또는 사회 구조나 가치에 영향을 미치는 경우, 우리는 그 기술에 대한 통제 여부를 고려해 왔다. 아울러 해당 기술의 도입이 사회 불평등을 심화시키는 경우에도 그 접근성의 평형성이나 기술로 인해 발생하는 이익의 공정한 분배에 대해 논의해 왔다. 예를 들어, 유전자 조작

이나 치료기술의 도입은 바로 이런 성격의 불평등 문제 해결이 핵심임을 지능 프로젝트 사업으로 수행된 ELSI 연구의 성과물로 출판된 From Chance to Choice: Genetics and Justice(2001)의 핵심 내용이다.

필자의 관심은 첫 번째 개념으로서의 공공성에 더 천착하고자 한다. 왜냐하면 인공지능이란 기술이 어떤 가치를 지향해야 하는지가 더 시급히 심도 있게 논의되어야 하는 주제일 수 있기 때문이다. 간단히 말해, 인공지능은 여타 의 생명과학 분야의 새로운 기술이 지닌 성격 즉 이익을 제공하고, 안전성만 확인된다면 그래서 사회에 해를 끼치지 않는다면, 사회에서 수용가능하다는 정도의 기준만으로 검토되는 것만으로 충분하지 않기 때문이다.

인공지능의 공공성은 즉 첫 번째 개념의 측면에서도 공공성을 지니는데, 그 이유는 다음과 같다. 첫째, 인공지능 기술의 완성도는 수집되는 데이터의 질에 영향을 받는다. 그런데 수집되는 데이터는 지식 특히 공적 이익을 위해 생산한 인류의 지식, 즉 학술연구를 반드시 포함해야 한다. 인공지능이 소위 ‘학습’한다는 데이터는 학술자료를 포함한다. 그런데 중요한 것은 학술 자료는 공공의 이익을 목적으로 생산되는 자료이며, 누구나 자유롭게 접근할 수 있도록 허용하는 대상이 되는 데이터로 취급된다. 한 때 ‘오픈 소스’가 강조되었던 것도 이런 맥락에서 이해될 수 있다.

아울러, 엄청난 다수의 시민이 생산하는 개인정보를 수집대상으로 한다. 간단히 말해 인공지능의 질적 우수성은 다수의 시민의 자발적인 참여를 전제로 한다. 이러한 데이터의 수집은 기본 모델은 동의에 기반한 수집이다. 아직 개인정보를 거래에 기반하여 수집하는 경우는 많지 않다. 그런데 우리나라의 경우에는 정보주체의 동의 없이 가명 처리한 정보, 즉 가명정보를 정보주체의 동의 없이 사용할 수 있도록 하고 있다. 이 문제는 우리나라가 윤리적으로 흠결 없는 인공지능의 개발의 위해서는 보완되어야 하는 문제점이다. 이 문제점에 대해서는 이후 적절한 곳에서 별도로 논의하겠다.

인류의 지식의 수집대상이고, 대다수의 시민이 제공한 개인정보가 수집대상이라면 인공지능 기술이 제공하는 혜택은 사익만을 목표로 해서는 안 되고, 다수의 시민, 나아가 글로벌 사회의 구성원에게 돌아가야 한다. 간단히 말해, 기술 개발의 혜택이나 이익에 대한 시민, 세계 시민의 공유가 지향되어야 한다.

02

인공지능은 무엇을 위해, 어떤 방식으로 활용해야 하는가?

인공지능이 여타 기술과 다른 점은 이 기술이 가져다 주는 성과물이 이미 정해진 목표가 있는 것이 아니라, 소위 ‘지능’이라 불리는 능력을 산출하는 것이 성과물이다. 따라서 우리는 이 능력을 무엇을 위해, 어떤 방식으로 활용해야 하는지 숙고하지 않을 수 없다.

인공지능은 인간의 지능과 동일하지 않다. 가장 근본적인 차이점은 인간의 지능은 진화의 산물이며, 진화는 생존과 관련이 있다. 다시 말해, 인간의 지능은 개인적으로는 살고 죽는 문제를 해결하기 위해 발전하였고, 사회적으로는 함께 살아가는 가족, 동료, 동료 시민과 공감하고 연대하며 발전하였다. 유한한 인간이 고통과 즐거움이라는 감수 능력을 지니고서, 질병과 죽음을 피하며 건강한 삶을 영위해야 하지만, 유한한 인간이기에 언제가는 죽음을 직면해야 한다는 사실은 인지하며 지능은 발전하였다. 유한한 인간은 무한한 것, 영원한 것, 절대적인 것, 완전한 것이라는 개념을 지니며, 영성적 사고 또한 개발하며 진화해 왔다. 이처럼 인간의 지능은 인간 삶의 객관적인 물리적인 현실과 분리되어 생각할 수 없고, 삶의 완성이 죽음이라는 엄연한 현실과 고통은 피하고 싶은 것이고 행복이

바라는 것이라는 욕구를 인정하지 않고는 생각할 수 없다.

하지만 인공지능은 존재론적으로 이런 삶도 삶의 환경도 존재하지 않는다. 인공지능이 인간의 지능을 흉내낸다 할지라도 이와 같은 존재론적인 맥락까지 흉내낼 수는 없다. 근본적인 차이는 생명이란 생물학적인 특성이 인공지능에게는 부여할 수도 없고, 수반 또는 창발하게 할 수도 없다. 기껏해야 생명이 아닌 수명이 있는 것처럼 할 수 있을 뿐이고, 진정성은 없지만 감정이 있는 것처럼 흉내내게 할 수밖에 없다. 이처럼 인공지능의 개발은 우리에게 인간을 인간이게 해 주는 것이 무엇인지 보다 깊이 있게 보다 진지하게 숙고할 것을 요청한다. 나아가 우리는 인류 역사에서 무엇을 지향해 왔고, 인공지능 시대에도 무엇을 지향해 나아가야 하는지 숙고하게 한다.

인공지능이 인간의 지능을 흉내내는 것이든 인간의 지능과는 구별되는 기계지능이든, 인공지능이 인간 지능의 일부 기능이나 역할을 보조하거나 대신할 수 있다는 사실은 부인할 수 없다. 그렇다면 소위 ‘인공지능’이라는 능력을 누가 그리고 어떤 목표를 위해, 또한 어떻게 활용해야 하는가? 필자는 인간의 지능이 어떻게 활용되어야 하는지에 대한 기존의 이해나 사회적 합의가 인공지능에도 기본적으로는 적용되어야 할 것이라고 주장한다.

첫째, 인간 지능이 추구하는 이익과 마찬가지로 인공지능 역시 공적 이익을 위해 활용되어야 한다. 앞서 인공지능이 온전하게 공공재적 성격을 지닌다고 말할 수는 없지만, 공공성을 지니고 있음은 분명하다고 주장하였다. 인간도 마찬가지이고, 인간의 지능도 마찬가지이다. 인간은 사적 삶을 살아가면서도 공적인 삶도 영위한다. 인간의 지능은 사적 이익을 위해 활용되지만, 그래서 자신이나 가족을 위해 활용되지만, 공적 이익을 위해 활용되기도 한다. 직무 수행은 큰 틀에서는 공익과 사익의 실현을 함께 진행시킨다. 물론 사익과 공익은 항상 명확히 구별되는 것은 아니다. 그러나 그렇다고 하여, 그 전형이 존재하지 않는 것은 아니다.¹⁾

앞서 언급했듯이, 인공지능은 데이터 수집의 측면에서 공공성을 지니고 있기 때문에 특히 더 공적 이익을 위해 활용되어야 한다. 사적 이익과 관련하여, 다행히도 삶과 죽음의 문제, 고통과 즐거움의 문제를 갖고 있지 않은 인공지능은 욕구조차도 없기에, 인공지능이 사적 이익을 추구한다는 일은 발생하지 않는다. 다만, 인공지능이 인공지능 기술을 제공하는 기업의 이익을 추구하도록 설계될 수는 있다. 그러나 이 문제는 인공지능 기술을 제공하는 회사가 어떤 사회적 역할을 수행해야 하는지에 대한 문제로서 기업 윤리의 관점에서 별도의 논의가 필요한 부분이다.

둘째, 인공지능은 인간의 지능처럼 진리추구적(veristic)이어야 한다.²⁾ 지능은 인식과 떼어놓을 수 없는 관계에 있다. 인간의 지능은 많은 경우 인식을 위해 활용되고 인식을 바탕으로 또 다른 인식이나 인식의 범위를 넘어선 영역에 사용된다.

인간의 지능은 인식(episteme), 즉, 전통적으로 정의되어온 것처럼, 정당화된 참된 믿음(justified true belief)을 위해 활용된다. 진리가 무엇인지 모르기에, 진리 탐구 행위에서 정당화 근거 제시는 매우 중요한 역할을 한다. 그렇기 때문에 학문 영역에서 각주나 후주를 달며, 참고문헌을 밝히거나 경험연구의 경우, 실증적 근거를 제시하는 것은 매우 중요한 활동이다.

그런데 인공지능은 과연 이런 노력을 하는가? 진리추구적인가? 과연 인공지능은 진리추구적으로 설계되었는가? 나아가 인공지능의 산물을 이용하는 우리는 진리추구적 관점에서 적절한 관행을 수립하거나 관행을 수립하려고 노력하고 있는가?

1) 신경윤리에서 치료와 증강의 구분이 이와 유사하다. 치료와 증강의 구별이 어렵다고 하여 치료와 증강의 구분이 무의미한 것은 아니다. 사회적 합의의 도출에 이 구분이 항상 작동하는 것은 아니지만, 전형적 사례에 해당하는 경우, 효과적으로 작동할 수 있다.
2) Alvin Goldman은 *Social Knowledge in A Social World*(1999)라는 저서에서 인식 관련 제도나 관행이 veristic 한지 즉 진리 추구적 관점에서 도움이 되는지 여부를 검토할 것을 제안한다.

인식을 위한 정당화 노력은 인공지능 기술 자체에도 탑재되어야 하지만 이용자인 우리에게도 필요하다. 우선 생성형 인공지능이 생성하는 산물은 유료형의 경우조차도 제공하는 정보의 근거 제시가 충실하지 않다. 할 수 없어서라기보다 하려면 더 많은 비용 즉 시간과 전력이 소모되기 때문인 것으로 판단된다. 정확하고 진실된 정보를 제공하기보다는 시간 대비 소비자의 만족도가 우선되어야 한다고 생각한다면, 이 생각은 진리추구적이지 않다. 그러나 지능의 문제, 지식의 문제, 그래서 인식의 문제와 관련된 서비스를 제공하면서 진리라는 가치가 강조되지 않고 있다는 것은 매우 위험해 보인다.

따라서 인공지능은 제시하는 정보가 어디에 근거한 것인지 참조한 문헌뿐만 아니라 문헌에서의 위치를 밝히도록 해야 한다. 누구든 정보의 출처를 확인하고자 한다면 쉽게 확인할 수 있도록 설계되어야 한다.

인공지능이 절대로 허위 정보를 만들어내지 않아야 할 뿐만 아니라, 선입견에 따른 부정확한 정보가 산출되지 않도록 해야 한다는 것은 너무나도 당연한 주장이다. 다만 이런 생산물의 산출을, 의도되었든 실수든, 어떻게 관리하고 대응할 것인지에 대한 방안이 시급히 마련되어야 한다. 이 문제에 대해서는 이후 대응방안을 다룰 때 언급하고자 한다.

이용자 측면에서는 인공지능이 산출한 결과임을 밝혀야 할³⁾ 뿐만 아니라, 인공지능을 활용해 거짓 정보를 만들어 내도록 명령하지 않아야 하며, 이용자는 거짓 정보인지 여부를 확인하는 노력을 게을리하지 않아야 한다. 거짓 영상을 만들어 내는 것 역시 원칙적으로는 금지되어야 한다. 예술창작과 같은 특별한 목적을 위해 이런 성격의 영상을 인공지능에게 산출하도록 한 후 다른 사람들에게 전달한다면, 인공지능이 만들어낸 거짓 영상임을 밝혀야 하고 맥락상 분명하게 그러한 정보임이 드러나도록 해야 한다.

정당화 근거 제시라는 인식적 활동의 기본 요건을 인공지능의 산출뿐만 아니라 이용자에게 요구해야 하는 이유는 다음과 같다. 첫째, 이용주체인 우리 인간은 인식의 주체이기 때문이다. 즉 이용자인 우리가 인식주체로서 당연히 지녀야 하는 정당화 근거 확인 및 근거 제시 요구라는 활동을 게을리해서는 안 되기 때문이다. 둘째, 인식 주체인 우리 인간은 지식을 전달한다는 기능을 수행하는 인공지능이기에 반드시 해당 정보의 근거 제시를 요구할 수 있고, 이는 인공지능 기술을 제공하는 업체가 그 기능을 인공지능이 수행하도록 설계해야 한다.

2021년 11월 유엔 유네스코 총회에서는 『인공지능 윤리 권고』(Recommendation on the Ethics of Artificial Intelligence)를 마련하였다. 이 권고는 다양한 가치와 원칙을 제시하고 있다. 가치로는 ‘인권과 기본적 자유의 존중, 보호 및 증진과 인간 존엄’, ‘환경과 생태계 번영’, ‘다양성과 포괄성의 보장’, ‘평화롭고 정의로우며 상호 연결된 사회에서 살아감’을 열거하고 있으며, 원칙으로는 ‘비례성과 해악 금지’, ‘안전성과 보안’, ‘공정과 차별 금지’, ‘지속가능성’, ‘프라이버시권과 데이터 보호’, ‘인간의 감독과 결정’, ‘투명성과 설명가능성’, ‘책임과 해명책임’, ‘인식제고와 리터러시 증진’, ‘다중 이해당사자 참여를 통한 적응적 거버넌스와 협력’을 열거하고 있다. 이러한 가치와 원칙들은 인류가 지향해야 온 가치와 원칙들 그리고 인공지능과 관련하여 보다 주목해야 하는 가치와 원칙들임에 틀림없다. 하지만 진리추구적이어야 한다거나 진실(truth)이 중심가치여야 한다는 직접적인 언급이 없어 아쉽다. 향후 개정에서는 이 가치에 주목하고 강조할 필요가 있다.

다만, 정당화 근거 제시를 충분히 수용하지는 못하지만, 이와 관련이 있는 가치나 원칙으로 설명가능성과 해명책임(accountability)이 있기는 하다. 해명책임은 어떤 문제에 대응(respond)해야 하는 책임(responsibility)과 달리

3) 과학기술정보통신부, 한국과학기술기획평가원, <국가연구개발연구윤리 길잡이> (3차 개정본, 2025)에서는 “AI도구는 연구 보고서나 논문 등에 저자로 표시될 수 없음”이라고 하면서, “연구계획서, 연구보고서, 연구논문 작성 등에 AI 도구 활용 시 관련 정보를 공개”해야 한다고 권고하고 “AI도구가 생성한 콘텐츠를 연구계획서, 연구보고서, 연구논문 작성 등에 활용한 경우 그 책임은 사용자(저자)에 있음” 명확히 권고하고 있다.

인식적 차원에서 설명(account for)해야 하는 책임으로 이해될 수 있다. 우리가 인공지능에서 해명책임을 요구하는 것은 주로 문제 발생시 왜 이런 문제가 발생했는지 해명 즉 설명되어야 할 것이기 때문이다. 그러나 문제 발생만이 아니라 정상적인 기능 수행의 과정에서도 인식과 관련된 기능 수행에서는 정당화 근거 제시 기능이 충실하게 탑재되어야 한다.

간혹 인공지능의 처리과정이 “블랙박스”라는 이유로 해명책임을 유보하는 견해도 있는 듯하다.[참고문헌] 하지만 그 어떤 경우에도 해명책임을 유보는 허용될 수 없다. 이것은 인식주체로서의 인간의 인식적 노력을 방기하는 것이다. 이것은 인공지능 기술의 이용 주체인 인간으로서의 주체성을 포기하는 것이다. 당장 해명할 수 없다면 또는 왜 그와 같은 결과값이 산출되었는지 알 수 없다면, 알 수 있도록 노력해야 한다. 추론과정을 알지 못하고 단지 결과값만 가지고 평가하겠다는 것은 올바른 태도가 아니다. 이와 같은 태도는 효율성에만 몰두한 태도이고 결코 진리추구적 태도가 아니다. 예를 들어, 추론과정은 전혀 알 수 없지만 그래서 그 추론의 결론에 대한 근거를 잘 이해할 수는 없지만, 또는 수궁할 수 없지만, 무조건 따라야 한다는 태도는 매우 위험하다. 이것은 합리적 근거를 제공하지 못하는 주술이나 미신에 의존하는 것과 무엇이 다른지 묻고 싶다. 이런 태도는 행위의 주체로서뿐만 아니라 인식의 주체로서의 인간의 지위를 포기하는 것이나 다름 없다.

셋째, 유네스코의 권고가 이미 밝히고 있듯이, 인공지능은 책임의 주체가 될 수 없음을 명확히하고, 인공지능의 산출이나 활용에서 발생하는 책임의 문제를 명료하게 해결할 수 있어야 한다. 유네스코 권고에서 “회원국은 궁극적인 책임과 해명책임은 언제나 자연인이나 법인에 주어져야 하며, 인공지능 체계에 법인격이 부여되어서는 안 된다.”고 권고하고 있다. 아울러 다음과 같이 권고하기도 한다.

어떤 방식으로든 인공지능 체계에 기반한 결정과 행동에 대한 윤리적 책임(ethical responsibility)과 법적 책임(liability)은 항상 궁극적으로는 인공지능 체계의 수명 주 기에서 역할을 수행한 인공지능 관련 행위자(AI actors)에게 귀속되어야 한다.

위에서 언급된 인공지능 관련 행위자에 대해서는 다음과 같이 규정하고 있다.

AI actors는 인공지능 체계의 수명 주기의 최소 한 단계에라도 관여하고 있는 행위 자라고 정의될 수 있으며, 이들은 무엇보다도 연구자, 프로그래머, 엔지니어, 데이터 과학자, 소비자, 기업, 대학, 공적 또는 사적 기관과 같은, 자연인과 법인 모두 를 가리킬 수 있다.¹

인공지능은 책임의 주체가 될 수도 없지만, 위 권고가 규정하듯이, 설정해서도 안 된다. 이런 조치는 책임이 증발하는 문제를 해결하기 위해서이다. 인공지능은 실질적인 책임을 질 수도 없을 뿐만 아니라, 자연인이나 법인에게 가야 하는 책임을 가리는 잘못을 해서는 안 된다.

넷째, 인공지능이 추구해야 하는 중국의 가치, 최상위의 가치, 궁극적 지향점은 인류의 생존과 평화로운 공생이며, 인공지능 활용 방향과 가치를 설정하는 주체는 인간이어야 한다.

늘 새로운 기술이 등장하며 때론 논란도 야기되고 우려를 낳기도 했지만, 결국 인류를 지금까지 잘 대처해 왔다고 평가할 수 있다. 물론 잘못된 대응으로 무고한 목숨이 희생되거나 불평등이 심화되는 일이 없었다고 할 수는 없다. 하지만 그 기술의 방향을 누가 어떻게 어떤 방향으로 설정해야 할지를 몰라 우왕좌왕하지는 않았다고 평가한다.

혈액 투석기, 체외수정기술, 줄기세포 연구 및 기술, 이종장기 이식 연구 등 다양한 첨단과학기술이 연구되고 등장할 때마다, 극단의 반대나 과장된 찬성도 있었지만, 생명의료윤리학 영역의 학자들과 전문가들은 진지하게 토론하고, 숙의하고, 전문가와 시민의 의견을 청취하며 합리적인 정책을 수립하였고, 인간 존중의 가치가 훼손되지 않고,

인류가 지향해야 온 가치, 정의, 인류 공생과 번영, 지속가능성 등의 가치에 부합하는 방향으로 논의를 전개해 왔고, 국제사회의 합의를 기반으로 권고와 지침을 마련해 가며, 나름 잘 조정해 왔다고 평가한다.

인공지능 기술 역시 인간 존중의 가치가 훼손되지 않고, 인류가 지향해야 온 가치, 정의, 인류 공생과 번영, 지속가능성 등의 가치에 부합하는 방향으로 개발되어야 하며, 그동안 인류가 새로운 기술 도입시 거쳐야 했던 과정을 잘 거치면서, 현명하게 도입 확산하는 과정을 거쳐야 한다. 아울러 이런 소통과정과 결정과정의 주체는 바로 인간이어야 한다.

인공지능 기술이 추구해야 하는 가치와 원칙은 결국 인간 중심의 가치이다. 인간의 가치를 훼손시키거나 인간 존중의 원칙을 어기는 일이 없어야 할 것이다. 인간 중심의 가치를 추구한다는 것은 인간 이외의 존재자를 함부로 대할 수 있다는 것을 의미하지 않는다. 인간중심주의는 인간 중심 편파주의를 함축하지 않는다.

03

인공지능과 거버넌스: 대응체계 구축 필요성

상기한 바와 같이 인간 존중과 인류가 공유해 온 가치를 지향하는 인공지능 기술의 개발과 도입, 확산을 위해 우리는 다음과 같은 대응 체계 구축이 필요하다.

첫째, 국내 및 국제적 협력 거버넌스를 구축할 필요가 있다. 인공지능 기술의 문제는 전세계적 문제이므로 국제 협력은 필수적이다.

국제 협력 거버넌스의 구축은 인공지능 기술이 준수해야 하는 국제적 기술표준을 마련하는 것뿐만 아니라, 인공지능 기술과 관련된 윤리정책에 대한 지침 및 가이드라인 개발을 위해서도 필요하다. 국제 협력 거버넌스의 구축을 위해 민주당 차지호 의원이 현재 추진 중인 우리나라의 ‘글로벌 AI 허브’ 유치 프로젝트는 매우 중요한 역할을 수행하게 될 것이라 생각된다.

물론 유네스코의 인공지능 윤리에 대한 권고가 있긴 하다. 우리는 이 권고의 정신과 내용이 반영된 상세한 가이드라인을 개발할 필요가 있다.

유네스코의 이번 권고는 생명의료윤리학 분야에서의 대응 체계 관련 이론의 변화를 반영하여, 사전예방 원칙을 넘어서서 상시대응체계 도입 및 구축의 필요성을 보여주고 있다고 판단된다. 전통적으로 위협과 이익에 대한 대응 방식은 사전예방 원칙을 따르는 것이었다. 그러나 기술 발전과 위험 예방이나 방지라는 두 가치의 조화로운 실현을 위해 회복불가능하거나 치명적인 위험이 예견되는 상황이 아니라면, 기술 연구를 진행하면서 때로는 실생활 도입을 시도하면서 상시적인 모니터링 체계를 구축하여 대응하는 방식을 취하고 있다. 2025년 11월 유엔 유네스코 총회에서 채택된 <신경기술 윤리 권고>(Recommendation on the Ethics of Neurotechnology) 역시 동일한 입장을 드러내고 있다.

둘째, 국내 거버넌스의 경우, 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」 제정은 고무적이다. 하지만 전략위원회 및 분과위원회의 실질적인 활동과 관련하여 이해당사자들의 합의 도달을 목표로 한 기능이 강화될 필요가 있다. 우리나라에서 개발되는 윤리적 가이드라인이나 정책 개발은 유네스코의 권고를 충실히 반영하여야 한다. 국제적 표준을 따르지 않거나 국제적으로 수용할 수 없는 표준의 개발은 국제 사회를 선도할 만한 지위에 있을 자격이 없다.

따라서 국내 거버넌스의 구축은 국내 문제뿐만 아니라 우리의 거버넌스가 국제적으로도 수용 가능한 보편적인 거버넌스의 모델이 될 수 있도록 진행되어야 한다. 따라서 기술 개발, 개발시 준수해야 하는 가치나 원칙, 정책 결정 과정, 의사소통 방식, 문제 해결을 위한 대응 체계 등등의 문제와 관련된 상세한 가이드라인이 개발되어야 하고, 중요한 정책 결정을 위한 의사결정 구조와 공중의 공적 참여가 고려된 정책 결정 모델을 개발해야 할 것이다.

특히 전략위원회나 분과위원회는 국가생명윤리심의위원회는 국가 차원의 윤리정책에 대한 의사소통의 허브 역할을 하는 위원회가 활성화될 필요가 있다. 위원회는 전문가뿐만 아니라 정부, 언론, 시민과 소통하며 현안 문제를 정확히 알리고, 현안 문제 해결에 필요한 정보를 정확히 공유하여야 하며, 가치와 목표 및 방향에 대한 사회적 합의를 도출하는 노력이 필요하다. 건강과 생명에 심대한 영향을 미치는 문제들이 인생관과 세계관의 상이함으로 인한 가치 대립의 양상을 모여왔던 생명의료윤리학의 영역에서 공중의 공적 참여를 지속적으로 강조해 왔던 것은 민주주의의 가치와 합의에 기초한 문제해결을 존중해 왔던 인류 역사와도 맥을 같이 한다.

셋째, 인공지능 기술의 도입과 확산에 잘 대비하기 위해서는 대응 및 대비 체계 구축이 필요하다. 우선 인공지능이 야기할 수 있는 위험에 대한 구체적인 평가 시스템이 도입될 필요가 있다. 간단히 말해 안전성 평가 체계가 구축되어야 한다.

특히 사회적 변화의 수준이 구조를 변화시키는 정도는 근본적인 문제에 대해서는 시장의 논리에 내맡길 수 없다. 현재 예상되는 심각한 수준의 변화는 노동 문제, 교육의 목표와 방법 개발 등이 시급히 다루어져야 할 것이다.

인공지능은 인간을 위한 그리고 인류를 위한 기술이 되어야 하기에, 급격한 변화로 인해 발생하는 대규모 실업이나 변화의 부작용으로 인해 발생하는 사회 불안, 기술 도입으로 인한 공정한 경쟁 훼손, 경쟁에 기반한 보상 시스템 교란 등을 유발하거나 현재의 불평등을 더욱 심화시키지 않도록 해야 한다. 나아가 불가피한 변화에 대해서는 이 변화를 감내할 만한 준비 시간이 확보된 후 기술이 도입되도록 해야 한다. 정부는 이런 변화에 대비하는 노력을 기울여야 하며, 변화를 경제와 시장의 논리에 내맡기는 것은 정부의 역할에도 부합하지 않고, 인공지능 기술이 지향해야 하는 목표나 가치와도 부합하지 않는다.

또 다른 인간 소외의 문제를 겪지 않도록 우리 사회의 전문가들은 선제적으로 연구하고, 대비하고, 해결책을 도출하는 노력을 기울여야 할 것이다. 이러한 노력 안에는 인공지능 연구자를 대상으로 한 소통과 논의 역시 필요하다. 의학의 발전이 의사들의 자발적이고 자율적인 전문직 윤리의 수립과 병행해 왔듯이, 인공지능 분야 역시 인공지능 전문 학자 및 기술자들의 자발적이고 자율적인 전문직 윤리의 수립과 병행되어야 할 것이다. 생명의료윤리학의 영역이 학제적 성격을 유지하며 발전해 왔듯이, 인공지능 윤리정책의 분야 역시 광범위한 학제적인 성격을 견지하며 발전될 필요가 있다. 따라서 인공지능 분야의 학자들과 기술자들과의 소통이 필요하며, 학제적인 연구와 더불어, 의사소통의 공간이 열려야 한다.

끝으로 국내 법제와 관련하여 시급히 해결되어야 할 문제가 있어 지적하고자 한다. 프라이버시 보호 체계 도입과 신뢰 확보가 필요하다. 인공지능 기술 개발을 위해 데이터 수집은 필수적이며, 그것도 광범위한 대량의 데이터 축적이 필수적이다. 인공지능과 빅데이터 기술이 함께 결합되는 것은 너무나도 자연스러운 일이다. 그러나 데이터 수집은 프라이버시 보호를 기본 원칙으로 해야 한다. 그리고 이런 수집은 정보 주체의 신뢰를 기반으로 한다. 수집된 정보가 잘 보호되고 있다는 신뢰를 정보 주체로부터 얻지 못하면 대량의 데이터 구축은 어려움을 겪을 수도 있다.

그러나 동의를 기반으로 한 원칙을 우회하려는 시도는 특히 우리나라의 「개인정보 보호법」에서 쉽게 목격된다. GDPR을 참고하였음에도 불구하고, GDPR이 정보 주체의 정보 이용 철회나 거부권을 행사할 수 있도록 한 단서 규정을 누락시킨 채, 결과적으로는 정보 주체는 가명처리된 정보가 개인정보로 정의되고 있음에도 불구하고, 정보

주체의 이용 철회나 거부권 행사를 보장하고 있지 않다. 이런 규정들은 얼핏 보기에는 동의 없이 데이터를 쉽게 획득하게 해 주는 규정처럼 보이지만, 프라이버시 침해에 대한 우려가 확인되고 확산되게 함으로써, 결국 데이터 이용을 꺼려하게 되고, 정보 제공을 주저하게 만드는 요인이 될 수 있다. 정보 이용보다 정보 주체로부터의 신뢰 획득이 더 우선되는 정책 방향으로 전환될 필요가 있다.

[맺음말]

인간 존중과 인류가 그동안 유지해 온 가치를 지향하는 인공지능 기술의 활용이라는 방향성 설정과 이 가치에 부합하는 국제 및 국내 대응 체계 구축이 절실하다. 우리는 인공지능 기술이 인식적 측면에서 진리추구적 활동이 되도록 기술이 설계되어야 하고, 우리가 인식의 주체임을 결코 잊어서는 안되며, 이 기술의 활용 주체는 바로 우리 인간임을 명심해야 할 것이다. 이와 같은 방향성 아래에서 우리들이 해결해야 하는 철학적, 법학적, 사회학적 문제들이 너무나도 많다.

인공지능의 향후 개발과 관련하여 인간을 인간답게 만드는 것들을 온전히 인공지능에게 부여하려는 시도는 애당초 불가능하다. 일부 철학자들은 더 이상 인공지능의 존재론적 지위라는 헛된 논쟁에 매달리기보다 철학이 인공지능 시대에 기여할 너무나도 많은 철학적 문제들을 해결하는 데 기여하길 바란다. 철학은 시대의 철학적 문제를 해결하도록 노력해야 한다. 철학사에 회자되는 학자들은 바로 그들의 시대적 철학 문제를 해결했기 때문에 철학사의 큰 학자로 평가 받았다. 인공지능의 존재론적 지위에 대한 논쟁은 시대적 과제가 아닐 뿐만 아니라, 앞서 언급했듯이 인간의 지능과 인공지능은 결코 동일할 수 없기 때문에 잘못 설정되었거나 헛된 논쟁이라 생각한다.⁴⁾

4) 이 논쟁에 대한 필자의 입장에 대해서는 필자의 논문, “인공지능이 인간 같은 행위자가 될 수 있나?”, 『생명윤리』, 제21권 제1호, 한국생명윤리학회, 2020. 71-85면 참조.

인공지능과 책임의 확장

윤준식 선임연구원 (서울대학교 교육종합연구원, 한국윤리학회 교육이사)

생명윤리 유관기관 공동학술대회(2026.06.10.)

<인공지능과 거버넌스>

인공지능과 책임의 확장

서울대학교 교육종합연구원 선임연구원

윤준식

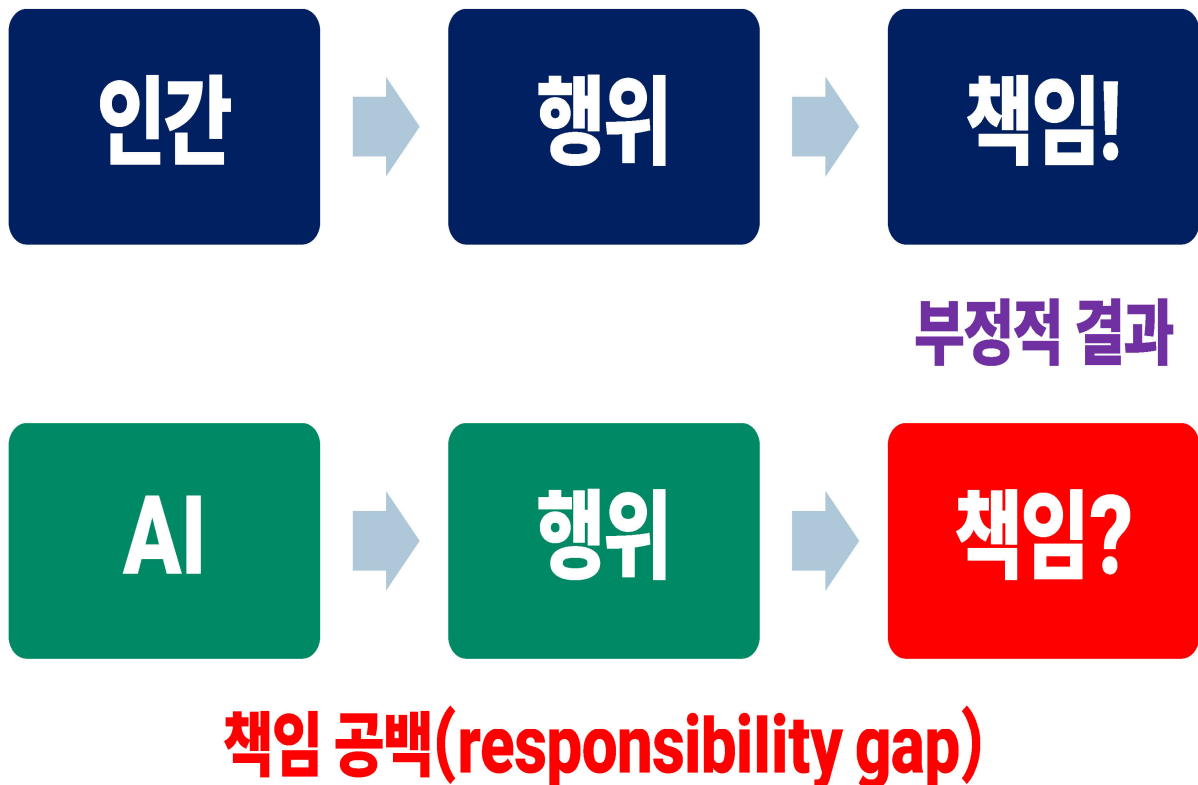
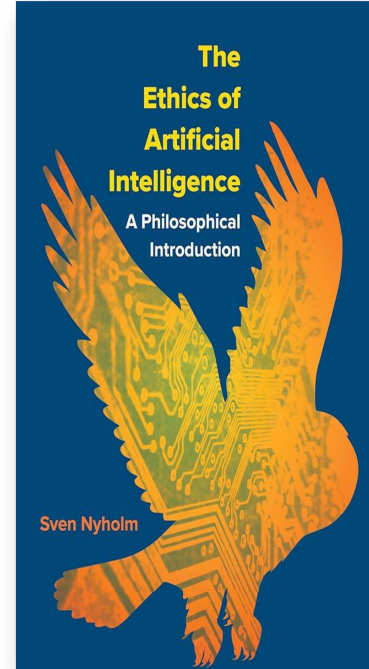
yjs5594@snu.ac.kr

1. 인공지능과 책임 공백

1. 인공지능과 책임 공백

- 포괄적이고 넓은 의미의 인공지능(Nyholm, 2026)

- 인간이 자신의 **자연적 지능**을 **사용**하여 수행하는 **과업**들을 **대신**하거나 **수행**하는 기술
- (그것이 무엇을 의미하든) 실제로 지능적이거나, 지능적 사고 또는 행동을 모방/시뮬레이션할 것
- 인간 수준이거나, 적어도 특정 과업에서는 최적으로 합리적인 수준일 수 있음
- 잠재적으로는 인공적 **행위자**(agent)의 한 형태



1. 인공지능과 책임 공백

• 도덕적 책임의 세 가지 차원(Shoemaker, 2011)

- 귀속성(attributability):
특정 행위/태도를 **행위자(성품)**와 연결
- 응답성(answerability):
특정 행위/태도를 **정당화하는 이유**를 행위자가 제시할 것 요구
- 책무성(accountability):
특정 (부정적) 행위/태도를 근거로 행위자에게 부과되는 **비난/제재**
[의 가능성]와 관련

=> 기계 윤리 및 AI의 투명성/설명가능성은 귀속성/응답성 차원,
AI가 **책무성** 차원의 책임을 갖긴 어려워 보인다.

Attributability, Answerability, and
Accountability: Toward a Wider
Theory of Moral Responsibility*

David Shoemaker

Recently T. M. Scanlon and others have advanced an ostensibly comprehensive theory of moral responsibility—a theory of both being responsible and being held responsible—that best accounts for our moral practices. I argue that both aspects of the Scanlonian theory fail this test. A truly comprehensive theory must incorporate and explain three distinct conceptions of responsibility—attributability, answerability, and accountability—and the Scanlonian view conflates the first two and ignores the importance of the third. To illustrate what a truly comprehensive theory might look like, I investigate what it would say about the difficult case of the psychopath.

INTRODUCTION

In recent years, T. M. Scanlon, along with a few others, has been constructing an attractive, wide-ranging theory of moral responsibility. Call it Scanlonian responsibility. The view has two parts: an account of being responsible and an account of being held responsible.

* The road from the first draft of this article to the last was long and winding, and I have many people to thank for providing corrective directions along the way. I would not have begun this project as I did without an invitation (and encouragement) from Christine Tappolet to present a paper on irrationality and agency at a panel she organized in Ottawa in May 2009. Thanks also to my fellow panelists and audience members at that session. I then posted an early discussion of some ideas about Scanlon's views on blame on the ethics blog *PEA Soup* in late May 2009 (<http://peasoup.typepad.com/peasoup/2009/05/scanlon-on-moral-responsibility-blame-part-1.html>), and I was very fortunate to have a number of excellent responses, for which I'm very grateful. In particular, I would like to thank Tim Scanlon himself for contributing a very helpful clarifying discussion to that thread. For their insightful comments on the earliest drafts, I thank Neil Levy, Michael McKenna, Angela Smith, and Matt Talbert. For thoughtful discussion of many of the ideas in the article as it developed, I am grateful to the graduate students in my seminars at both Tulane and Bowling Green State University (in particular Brandon Byrd and Nick Sans). For helping me structure and focus the article, I thank two anonymous referees.

Ethics 121 (April 2011): 602–632
© 2011 by The University of Chicago. All rights reserved. 0014-1704/2011/12105-0005\$10.00

AI 도입 확산 + 가속화
= 광범위한 책임 공백
<-> 인간 책임의 확장

2. 책임의 확장 1: 개인 -> 공동체

2. 책임의 확장 1: 개인 -> 공동체

- 분산된 행위성 -> 분산 책임: 피동자 및 결과 중심, 인과적 영향 역추적 및 책임 분배
=> 무과실/엄격 책임 개념 도입 (Floridi, 2016)

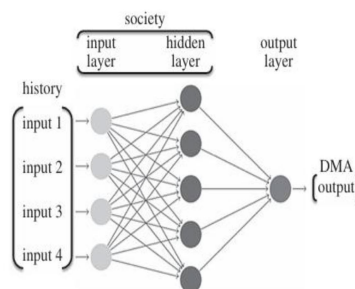


Figure 3. A multi-agent system as a multi-layered neural network.

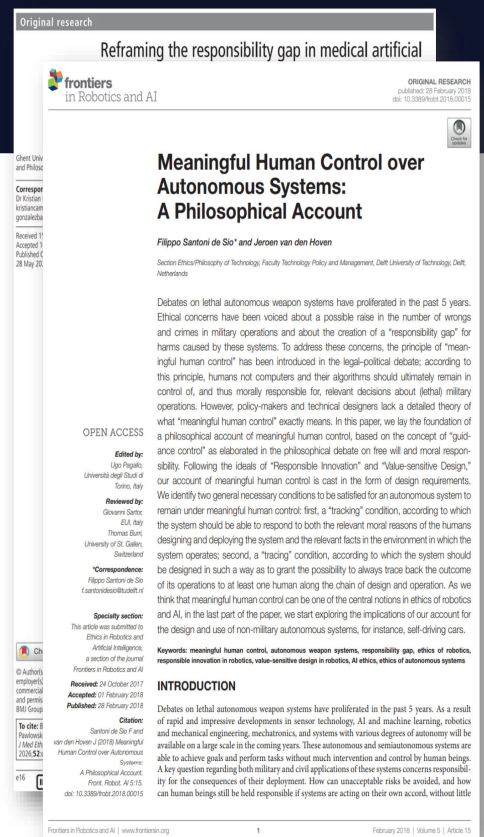
- 공유(shared) / 대리(vicarious) / 집단(collective) 책임
- 지휘(command) 책임, 인간-기계 팀 관점(Nyholm, 2023)
=> 선행 통제 행사자 + 통제 당시 개인적 이익 기대한 사람에게 책임 우선 배분(Gong, 2026)

3. 책임의 확장 2: 과거 -> 미래



3. 책임의 확장 2: 과거 -> 미래

- 기존 책임 논의, 이미 벌어진 부정적 결과에 대한 책임 중심 (backward-looking negative responsibility gap or retrospective accountability gap)
- 핵전쟁 위협, 생태계 파괴, 유전공학의 발전 등에 직면한 현대 (응용) 윤리학
=> 미래지향적/전향적 책임관 발전
- 인공지능 관련,
 - 기계 윤리, 가치 정렬, 설명 가능성 확보 등
 - 유의미한 인간 통제 (meaningful human control, Santoni de Sio & van den Hoven, 2018): tracking[기술 요소] + tracing[인간 요소]



4. 책임의 확장 3: 부정 -> 긍정

4. 책임의 확장 3: 부정 -> 긍정

They are responsible for the consequence. => 중립

He is responsible for the failure. => 부정

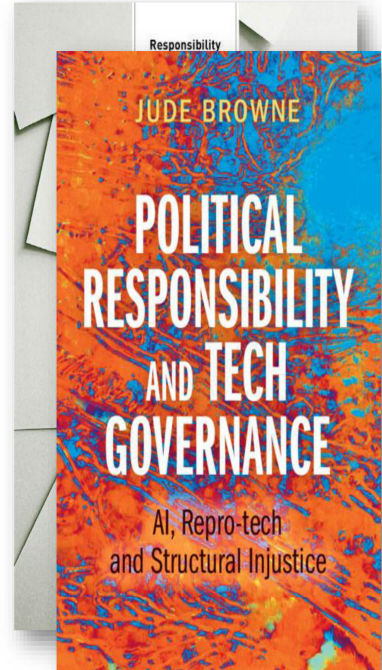
She is responsible for the success. => 긍정

- 과거에 발생한 긍정적 결과의 공로 인정, AI가 개입할 때, 그 성취는 누구의 것?
=> **긍정적 책임 공백 = 성취 공백** (저작권, 저자성 등 인정 문제) 해소 필요
- 미래에 긍정적인 일을 발생시킬 책임? 보통 특수 계약 및 친밀한 관계에서 발생
<-> **AI가 인간 관계 매개 혹은 대체할 때, 미래-긍정적 책임 공백 더 심화?**
- 삶의 질의 절대값은 전반적으로 향상될 수 있으나, 격차는 더 벌어질 수 있다
=> **분배적 부정의 시정 - 정의로운 분배 체계 마련 책임 => 정치적 책임!**

5. 책임의 확장 4: 도덕/법 -> 정치

5. 책임의 확장 4: 도덕/법 -> 정치

- 아이리스 매리언 영, 과실 추적이 불가능한 사회구조적 부정의 시정을 위한 정치적 책임 주장: 과거지향적 도덕적/법적 책임 (liability) 모델 -> 사회적 연결(social connection) 모델
 - 행위의 맥락/과정 및 배경조건 자체의 정의로움에 초점
 - 배경조건의 개선 추구
 - 부정의한 결과를 초래한 구조적 과정에 기여한 모든 이들의 집단 행동을 통한 제도/과정 변경 책임(서원주, 2021)
- 『정치적 책임과 기술 거버넌스』(Jude Browne, 2025)
 - 특히 기존 부정의 심화시키는 알고리즘 및 기술적 실업 지적
 - 규제 포획/선물 현실 <-> 심의 민주주의 참여 대표의 다양화
 - => 공적 이익 추구하는 독립적 공적 기구, 참여 제도적 보장



6. 확장된 책임 개념 프레임워크의 적용: AI 활용 연구윤리

6. 확장된 책임 개념 프레임워크의 적용

• 인간 책임의 확장

- ① 개인 책임 -> 공동체 책임
- ② 과거지향적(후향적) 책임 -> 미래지향적(전향적) 책임
- ③ 부정적 책임 -> 긍정적 책임
- ④ 도덕적/법적 책임 -> 정치적 책임

=> Conceptual Framework of Extended Responsibility, CFER

정치 / 거버넌스

		과거지향적	미래지향적
개인	부정적	① 개인 x 과거 x 부정적 책임	③ 개인 x 미래 x 부정적 책임
	긍정적	② 개인 x 과거 x 긍정적 책임	④ 개인 x 미래 x 긍정적 책임
공동체	부정적	⑤ 공동체 x 과거 x 부정적 책임	⑦ 공동체 x 미래 x 부정적 책임
	긍정적	⑥ 공동체 x 과거 x 긍정적 책임	⑧ 공동체 x 미래 x 긍정적 책임

Retraction

Retraction: Relational accountability in AI-driven pharmaceutical practices: an ethics approach to bias, inequity and structural harm

This article¹ is retracted by the journal.

WHAT HAPPENED AFTER PUBLICATION
The journal investigated concerns about the quality of the work and the accuracy of the references, including concerns that several references did not exist. During investigation of these issues, the journal was also alerted to the possibility of review manipulation.

The author used generative AI to identify and understand referenced sources. They did not adequately verify these references prior to submission.

REASONS FOR RETRACTION
The article is retracted due to evidence of peer review manipulation, and because it includes references that do not exist. This results from undeclared AI use with inadequate oversight by the author.

FURTHER ACTION
The journal will inform the author's organisation of the retraction.

© Author(s) (or their employer(s)) 2026. No commercial re-use. See rights and permissions. Published by BMJ Group.
J Med Ethics 2026;0:1. doi:10.1136/jme-2025-110913ret

Check for updates

REFERENCE
1 Biswas L. Relational accountability in AI-driven pharmaceutical practices: an ethics approach to bias, inequity and structural harm. J Med Ethics 2025.

동아일보

오피니언 정치 경제 국제 사회 문화 연예 스포츠 헬스동아

사회 > 단독

[단독] AI 부적절 사용 논문... 최근 5년 204건 철회

동아일보 | 업데이트 2025-11-17 10:50
전남혁 기자 [+ 구독](#)

2 1

165건은 '논문공장'이 대필 의혹
"AI 활용 구체적 가이드라인 필요"

국가연구개발
연구윤리
길잡이

2025. 11. 개정본

☞ 「국가연구개발혁신법」(이하 혁신법) 제정에 따라 종전의 연구진실성, 연구부정 대응에 국한되었던 연구윤리 규범이 건강하고 성숙한 연구문화 조성을 위한 규범으로 범위가 확장되었음

〈혁신법으로 달라진 연구윤리〉¹⁾

구분	종전	혁신법
	연구윤리확보	연구윤리확보
연구윤리개념	연구부정행위 (위조·변조·표절, 부당한 저자표시 등) 관리	연구진실성 확보 + 학문교류에 관한 윤리 + 이해 충돌 관리 + 인간·동물 대상 연구 윤리 + 건전한 연구실 문화 조성 등
책임주체	불명확	연구자·연구개발기관
책임주체의 의무	(연구자) - (연구기관) 연구부정 방지 및 검증을 위한 연구윤리에 관한 자체규정 마련	(연구자) 연구윤리 준수 및 진실·투명한 국가연구개발활동 수행 (연구기관) ① 자체 연구윤리 규정 마련, ② 소속 연구자 등의 연구윤리 확보 지원

☞ 연구자는 연구윤리를 준수하고 진실하고 투명하게 국가연구개발활동을 수행할 책임을 지니며(혁신법 제7조제1항) 연구개발기관은 소속 연구자 및 연구지원인력의 연구윤리 확보를 위하여 혁신법 시행령 제58조 각 호에 해당하는 사항들이 포함된 자체 연구윤리 규정을 마련해야 함(혁신법 제31조제4항 및 시행령 제58조)

6. 확장된 책임 개념 프레임워크의 적용

• 사례: AI 활용 연구윤리 문제와 CFER

- ① 개인x과거x부정: 연구진실성 및 연구부정행위(표절, 위조, 변조 등)에 대한 책임
 - ② 개인x과거x긍정: 연구 성과에 대한 귀속성, 응답성, 책무성 세 차원 모두 확보 필요
 - ③ 개인x미래x부정: AI의 한계 및 관련 규범(예: 평가, 데이터 보안 관련) 지속적으로 숙지
 - ④ 개인x미래x긍정: 검증 및 연구 역량 유지 및 개발
-
- ⑤ 공동체x과거x부정: 연구부정행위 발생시 검증 및 보고 체계 마련
 - ⑥ 공동체x과거x긍정: 우수 AI 거버넌스 운영 연구기관에 대한 (평가상의) 인정
 - ⑦ 공동체x미래x부정: 연구윤리 가이드라인 및 연구 역량 유지/강화 프로그램 마련,
 - ⑧ 공동체x미래x긍정: 로컬 AI 서비스 환경 등 구축 + 연구 업적 평가 방식 개선(양 -> 질) 등



“지난 몇 달간 우리는 여러 연구 프로젝트에서 인공지능 보조 도구인 **클로드 코드(Claude Code)**와 **오픈클로(OpenClaw)**를 광범위하게 활용해 왔다. [...] 데이터 수집, 정제, 정리 등 대부분의 작업을 이 도구들에 맡기고 싶은 유혹이 드는 것도 사실이다. 이런 심정이 우리만의 것은 아닐 것이다.

그러나 우리는 한 가지 불편한 사실을 발견했다. 어느새 몇 주째 대학원 연구 조교들에게 실질적인 업무를 맡기지 않고 있었던 것이다. 학생들의 논문 지도는 여전히 이어지고 있지만, 우리 자신의 연구 과제에 관해서는 어느 사이인가 **AI 에이전트에 의존하게 되었다**. 그 결과, 연구 그룹 구성원들과의 회의는 뜸해졌고, 학생들은 연구 과정에서 점점 소외되고 있다. 하나의 프로젝트를 만들어 가고 결과물로 이어지는 수많은 선택과 막다른 시행착오, 그리고 사소한 조율들에 대해 학생들이 알고 있는 것은 그렇지 않았을 경우에 비해 훨씬 적어졌다.

이 깨달음은 우리에게 적잖은 충격을 주었다. **AI는 시간을 절약해 주지만, 동시에 학생들이 자립적인 학자로 성장하는 데 필요한 인지적 발판을 빼앗기도 한다**. 신진 학자를 육성하는 멘토로서, 우리는 AI 활용과 학생 성장에 필수적인 실천적 연구 경험의 보존 사이에서 균형을 찾아야 한다.” (Hu & Zhang, 2026)

7. 책임지는 인간 (*Homo Responsabilis*)

7. 책임지는 인간(*Homo Responsabilis*)

- 인공지능 시대는 다차원적으로 확장된 책임을 인수하는 인간,
즉 ‘책임지는 인간’이 존재할 때, 윤리적으로 지속 가능하다!

- 책임지는 인간을 형성해 낼 개인적/공동체적 + 미래지향적 + 긍정적 책임
- 너무 이상적? (확장된 책임 인수의 실천적 동기 문제)

분명 그런 측면이 있지만,

그럼에도 불구하고 우리 주변에는 부정적인 결과의 책임을 기꺼이 지고 공은 남에게 돌리는 사람이 분명 존재하며, 우리는 이들을 칭찬하고 존경한다. (그리고 그런 사람이 되고 싶어 한다?!)

=> 달성 가능하고 관련성 있는 도덕적 귀감(Han et al., 2017) 필요

- 책임의 확장 5, 도덕적 피동자(moral patient)로서의 AI

생명윤리 유관기관 공동학술대회(2026.06.10.)

“인공지능과 인간의 책임”

감사합니다

서울대학교 교육종합연구원 선임연구원

윤준식

yjs5594@snu.ac.kr

인공지능시대의 새로운 개발 윤리 모색

목광수 교수 (서울시립대학교, 한국윤리학회 회장)

생명윤리 유관기관 공동학술대회
2026년 6월 10일

인공지능 시대의 새로운 개발 윤리 모색

목광수(서울시립대, 한국윤리학회 회장)



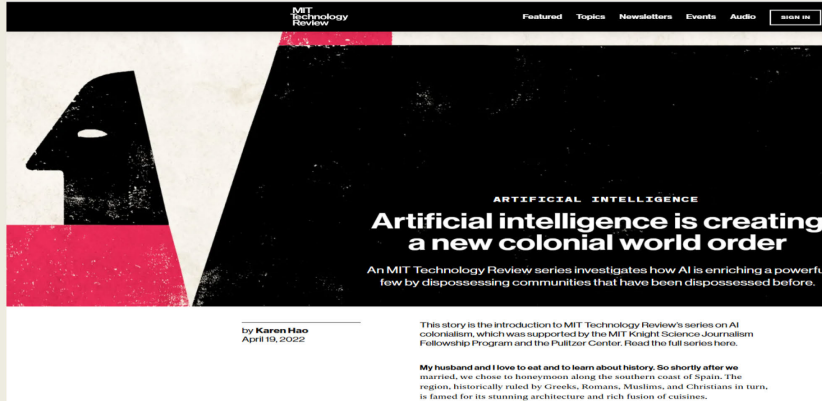
오늘 생각해 볼 물음들

- ❑ 인공지능 시대에 전지구 사회, 특히 제3세계는 어떤 상황에 처해 있는가?
- ❑ 전지구 사회, 특히 제3세계의 입장에서, 인공지능 시대가 초래한 윤리 문제에 대응하기 위해 필요한 윤리는 무엇일까?
- ❑ 과거의 개발 윤리, 특히 역량 중심 개발 윤리를 인공지능 시대에 걸맞게 재정립하기 위해서는 어떻게 해야 할까?



인공지능 시대의 전지구 사회

□ 새로운 식민 지배 질서?



□ 노동

: 라벨링 작업
: 저가 노동 분업

□ 자원

: 광물 수탈
: 데이터 추출

□ 시장화

: 실험공간
: 소비 시장

□ 구조와 규범

: AI 중심 구조 재편
: 인식/규범 주도

□ 교황의 회칙 <고귀한 인류>(2026): '새로운 형태의 노예제'

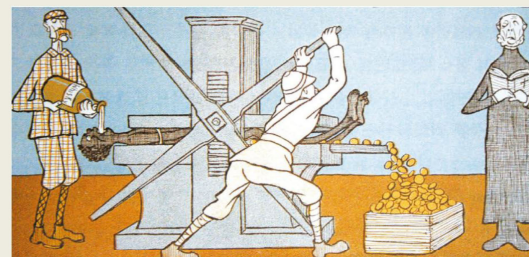
제국주의의 전지구사회 질서

- 배경: 대항로시대(15세기 말) 이후, 전지구 사회의 교류 증대
→ 문명의 전지구적 확대?



빅터 길럼, 『백인의 짐』, 『저지』, 1899년.

- 식민 지배와 제국주의 (15세기 말~1945년)
→ 자원 약탈과 시장 지배, 노동력 착취



토머스 하이네, 『식민지를 지배하는 방식』, 『심플리치시무스』, 1904년
(출처:중앙일보) <https://www.joongang.co.kr/article/18979885>



개발의 시대, 전지구 사회 질서

□ 개발의 시대(1945년 ~ 현재)

“... 우리는 우리의 과학적 진보와 산업 발전의 혜택이 **저개발 지역 (underdeveloped area)**의 향상과 성장에 쓰일 수 있게 하는 과감하고 새로운 프로그램을 시작해야 합니다. 세계 인구의 절반 이상이 **비참에 가까운 여건** 속에 살고 있습니다. ... 사상 처음으로 인류는 이 사람들의 고통을 덜어줄 수 있는 지식과 기술을 갖고 있습니다 ...”(1949년, 트루먼 대통령의 취임사중)

→ 원조 통한 경제적 발전, 역량 증진 등

: ICT4D (Information and Communication Technology for Development)

→ 윤리의 실천?

: 은폐된 약탈과 지배, 인식과 구조의 식민지배? (by 후기 개발 이론 등)

→ 제국주의의 지속? 또는 가능성?



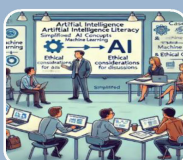
인공지능 시대의 전지구 사회 질서

□ 새로운 식민 지배 질서? '새로운 형태의 노예제'?



제국주의 시대 (개발의 시대 포함)

- 토지/자원과 노동의 지배
- 지배 방식의 가시화와 영토화



인공지능 시대

- 데이터와 자원, 노동의 지배
- 지배 방식의 비가시화와 탈영토화

□ 제3세계의 피해 가증: 올리히 벡, <위험사회> (1986), '위험은 더 약한 곳으로 재분배'



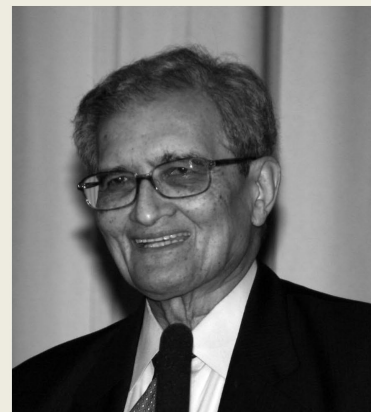
전지구 사회 윤리 문제 대응: 개발 윤리

- ❑ 개발 윤리학(development ethics)
:바람직한, 윤리적 개발/발전은 무엇인가? 어떻게 해야 하는가? 등
- ❑ 개발 윤리학의 전개
GDP성장담론의한계:GDP가발전 담보 못함
→ '기본적 필요(basic need)' 중심의 개발 윤리학
: 필요 설정 주체 문제, 필요의 불충분함 등
→ 역량(capability) 중심의 개발윤리학



전지구 사회 윤리 문제 대응: 개발 윤리

- ❑ 역량접근법(capability approach)
: 센(Amartya Sen)의 태너 강좌
'Equality of What?'(1979)에서
역량(물적 토대가 갖춰진 상
태에서의 선택인 실질적 자유)
개념 제시
→ 공적 추론 통해 보편성과
개별성의 조화 추구, 제도보다
부정의 제거의 비교적 정의론
추구



(1933~), 하버드 대학 경제학과/
철학과 교수, 후생경제학으로
노벨경제학상 수상(1998)



전지구 사회 윤리 문제 대응: 개발 윤리

□ 역량 중심의 개발 윤리학

: 공적 추론을 통해 설정된 역량 증진을 목표로 제시하여, 개발을 저해하는 외적인 영향(물적 토대 관련)을 극복하고 일부 내적인 영향(adaptive preference) 극복 추구

: 역량 중심의 개발 윤리학은 UNDP의 인간개발보고서, 인간개발지수(HDI)에 활용됨

→ 권리 중심, 심의 민주주의 중심, 지속가능성 등의 개발 윤리학으로 확장



AI 시대의 전지구 사회 윤리 문제

□ AI 시대의 전지구 사회: 새로운 제국주의 질서

- 지배 방식의 비가시화: 자발적 참여로 지배 은폐?

c.f.) AI/알고리즘 조종(manipulation)

- 탈영토화: 강화된 인식 종속? 기술 종속성?

c.f.) 특정 사회 가치 담은 AI 중심의 사회 구조 전환과 규범

→ 내적인 영향의 고도화: 개별화된 심리 접근 통한 인식 통제와 중독 - 역량 약화 / 종속

c.f.) 외적 영향(주류 문화, 비주류/대항 문화)과 관련된 내적 영향인 순응된 선호(adaptive preference)와 질적인 차이?

□ 역량 중심의 개발 윤리가 대응할 수 있을까?

수정/강화된 역량 중심의 개발 윤리

□ 역량 개념의 강화

- 실질적 자유
- 행위주체성(agency)
: reason to value

→ + 성찰(deliberation, 동기, 믿음, 선호, 책임, 사회적 영향 등에 대한 자기반성/비판) 특성 강화, 성찰을 포함한 다층적 자유 개념으로 확장, (wellbeing 보다) 우위성 부여

: AI시대의 개발 윤리학이 추구할 목표가 성찰 중심적 역량이기에 비가시화된 지배/자유 침해를 포착하고 대응할 여지 마련

수정/강화된 역량 중심의 개발 윤리

□ 역량 중심의 비교적 정의론

- 개인 측면의 부정의 제거

→ + 정의로운 제도 마련: 보편성과 특수성을 보장하는 열린 공평성(모든 관련자 참여)의 공적 토론(열린 공론장)을 통한 전지구적 관점의 전환 계기 마련

: 특정 사회 중심의 AI 관련 사회 구조와 인식 체계를 탈피하여, 전지구적 관점의 AI 규범/제도/인식체계로 전환하는 제도적 환경 조성

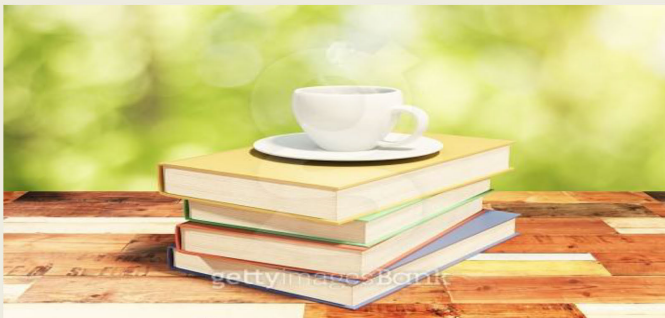


정리

- ❑ 인공지능 시대는 새로운 식민 지배 질서, 새로운 형태의 노예제로 전락할 위험이 있음
- ❑ 과거 제국주의, 개발의 시대가 초래한 윤리 문제에 대응했던 개발 윤리가 다시금 요청됨
- ❑ 역량 중심의 개발 윤리는 내적인 영향력을 고도화한 AI의 비가시적인 지배에 대응하기 위해,
1) 성찰 특성을 강화한 자유와 행위주체성의 우위, 2) 전지구적 관점의 AI 규범과 인식 체계를 조성할 열린 공론장 제도 마련으로 강화/수정될 필요가 있음

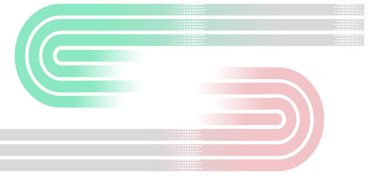


궁금한 것들에 대해...



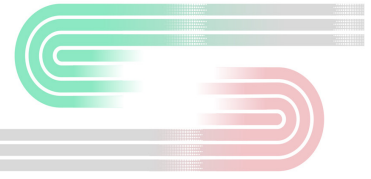
감사합니다!

패널토론 및 종합토론



심지원 교수 (동국대학교)

패널토론 및 종합토론



임시연 선임전문관 (유네스코한국위원회)



제2부

실천적 담론

좌장: 이일학 회장(한국생명윤리학회, 연세대학교 의과대학)

생명윤리 유관기관 공동학술대회

**인공지능과
거버넌스**

공공의료 AI 거버넌스

김준혁 교수 (연세대학교 치과대학, 한국의료윤리학회 간행이사)

AI ETHICS JOINT SYMPOSIUM · AI & GOVERNANCE

01 / 15

공공의료 AI 거버넌스

의료 AI를 '공공'의 문제로 다시 묻기

생명의료윤리 공동학술대회 — 인공지능과 거버넌스 · 제2부 실천적 담론

김준혁 (연세대 치과대학 / 한국의료윤리학회 간행이사) · 2026. 6. 10. · 유네스코홀

AI 윤리 공동학술대회 · 제2부 실천적 담론 · 김준혁

UNESCO HALL · 2026-06-10

PUBLIC HEALTHCARE AI · FRAME

02 / 15

이 발표의 자리 — '공공의료'와 'AI 기본의료'

공공의료 = '지필공'의 핵심 — 그 위에 'AI 기본의료'

- 국가·지자체·공공기관이 제공하는 의료 — 의료취약지·필수의료·취약계층 (공공보건의료법 §2·§7)
- 2026.6 발표 예정 'AI 기본의료' 전략 — '지필공'(지역·필수·공공) 격차 해소 수단, 구호 "보편성·신뢰성·효율성" (복지부·KIHASA)

오늘의 주장

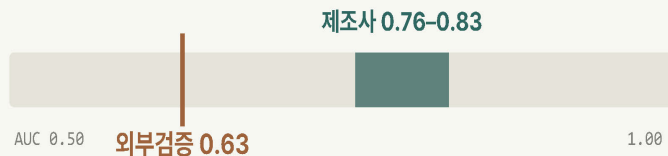
거버넌스의 본령은 'AI를 쓰느냐'가 아니라, **AI가 공공의료를 공공의료답게 — 시장이 버린 곳을 향하게 — 유지하는가이다.**

AI 윤리 공동학술대회 · 공공의료 AI 거버넌스 · 김준혁

2026-06-10

[사례] 검증의 실패 — Epic Sepsis Model

널리 배포되었지만, 작동하지 않았다



67%

패혈증 농침

18%

입원환자 경보

'누구에게' 유효한가

거버넌스의 첫 질문: 공공의료가 책임지는 환자(취약지·고령·취약계층)에게 유효한가?

공공의료의 관문 — 도입이 곧 정당화가 아니다

의료기술에는 윤리적 관문이 있다

- 키보드는 그대로 팔아도 되지만, 새 의료기술은 도입 자체를 따져야 한다 (『헬스케어 AI 윤리』)
- 공공의료에선 더 엄격하다: 도구가 사명(취약지·필수의료)을 향하는가까지 물어야

거버넌스 = 인구 규모의 관문

- 개별 승인이 아니라 공공의료 책임지는 인구 전체에 대한 유효성·형평 검증
- 규제(시판 허가)는 이 관문의 일부일 뿐

구조적 불일치 — '보편 AI'는 어디서 만들어지나

공공의료와 AI — 정반대의 환경

공공의료

'지필공'이 책임지는 곳 — 의료취약지
547개 읍·면, 취약계층·저수익 필수의료
 (공공보건의료법)

≠

불일치

의료 AI

데이터가 풍부한 소수 거점에서
 학습·검증 — 미국 임상 딥러닝 데이터
 대부분이 **캘리포니아·뉴욕·매사추세츠**
 (Kaushal et al., JAMA 2020)

'보편 AI'는 평균이 가장 적은 곳을 가장 늦게 본다 — '도입'은 곧 **불일치**의 수입이다.

[사례] 'AI 기본의료'와 '지필공' — 사명과 어긋날 위험

정부의 'AI 기본의료' — 격차 해소 시도

- 복지부 1차 'AI 기본의료' 정책간담회(2026.4.30) · KIHASA 공동주관 · 2026.6 전략 발표 예정
- 두 축: 'AX 스프린트'(만성질환·원격협진) · 'AI 고속도로'(상용화 지원)
- 구호 "보편성·신뢰성·효율성"으로 '지필공'(지역·필수·공공) 격차 해소

그러나 사명과 어긋날 위험

- 도구의 유효성은 거점 데이터에 묶임 (cf. Kaushal, 2020)
- '보편'(평균)에 맞춰 도입하면 **평균 바깥의 인구**(취약지·취약계층)가 가장 늦게 적합해짐

유효성 관문 — '보편성'을 누구의 보편으로 시험하는가

헬스케어 AI 6원칙 (WHO 2021 기반)

- ① 자율성 · ② 행복·안전·공공의 이익 · ③ 투명성·설명·이해가능성
- ④ 책임·책무 · ⑤ 포괄성·공정성 · ⑥ 대응성·지속가능성

'보편성'은 평균 — 시험이 필요함

- '평균적 정확'은 공공의료의 유효성 기준이 아님 → 취약계층에서 유효해야 함 (원칙 ②·⑤)
- '보편성·신뢰성·효율성'은 사명을 향하는 방향으로 시험되어야 함 — 편향 = 사명의 실패

데이터의 공공성 — 누구의 데이터가 빠져 있나

연결은 목적을 보장하지 않는다

- "디지털 고속도로 — 그런데 어디로?" (BMJ, 2021, My HealthWay 논평)
- 공공의료의 책임지는 **취약지·취약계층의 데이터는 과소대표** → AI가 그들에게 덜 유효
- 공동의 데이터를 원료로 쓰되 누가 통제·수혜하는가 (cf. Coeckelbergh-Gunkel, *Communicative AI*, 2025)

소유에서 인정으로

- 데이터 '소유권'으로 풀리지 않는다 → **인정으로서의 분배**: 누구의 데이터가 인정되고 이익이 어떻게 분배되나 (김준혁, 「건강데이터 소유권에 관한 윤리적 접근」, 2021; Fraser-Honneth)

책무 관문 — 국가는 제공자이자 수탁자다

한 문장으로

책임은 나뉘어도 **답할 의무(책무)는 위임되지 않는다**. 거버넌스의 일은 그 **답할 주체를 지정하는 것**.

공공의료에서는 더 무겁다

- 공공의료에서 국가는 규제자가 아니라 **제공자이자 수탁자** — 책무가 이중이다
- '많은 손의 문제': 행위자가 많을수록 답할 주체가 흐려진다 (van de Poel, 2011; 2015)
- 국가는 **가장 취약한 환자를 향해** 직접 답해야 한다

가치 정렬과 검증의 한계

가치 정렬의 세 물음

- 가치 정렬 = 기술이자 **규범**의 문제 (Gabriel, *Minds and Machines*, 2020)
- **선택(Which)** 누구의 가치 · **방법(How)** 어떻게 정렬 · **검증(Verify)** 어떻게 확인 — **검증이 최난제**

규제는 배경일 뿐 — 오히려 늦춰진다

- WHO LMM 지침(2024) 정부 일차 책임 · EU AI Act 고위험 **2026.8→2027.12** 연기(Digital Omnibus) — 규제 ≠ 거버넌스

구조를 넘어 ▶ 지표로 검증되지 않는 관계로서의 돌봄으로

돌봄의 전환 — 관계로서의 돌봄, 팀으로서의 AI

돌봄은 기능이 아니라 관계다

- 규칙·계산이 아니라 **감정·책임·연민** — 핵심은 **조율(attunement)**과 **인정(recognition)** (『AI 돌봄윤리』)
- 효율의 언어로는 **측정되지 않는 선(善)**을 침식할 수 있다 ("인간보다 더 인간적인", WALL-E)

AI는 대체자가 아니라 팀의 일원

- **돌봄제공자 — 인공행위자(ACA) — 돌봄수용자**: 책임은 자율 AI가 아니라 **관계 안에서**
- 거버넌스 함의: AI를 '대체자'가 아니라 **관계 안의 행위자**로 설계·배치

돌봄 관문 — 거버넌스가 지켜야 할 관계적 선(線)

무엇을 보호할 것인가

- **치료적 관계·존엄·돌봄의 조건** — throughput이 아니다
- 효율 최적화의 함정: AI 활용 역량을 관계적 역량보다 우대하면 **돌봄은 보기 좋게 퇴화**
- AI는 환자-의료인 신뢰를 약화시킬 수 있다 — 특히 **여성·고위험 상황**에서 (Zondag et al., *JMIR* 2024)

공공의료에서 특히

- 공공의료 돌보는 이들은 **이미 가장 취약한 사람들** — 관계적 선의 침식은 여기서 가장 아프다
- **유동성**: 누구나 **돌봄받는 자리에** 설 수 있다 → 이 거버넌스는 **모두의 문제**

종합 — 공공의료를 공공의료답게 하는 세 관문

사명 충실성의 세 관문

1 평균 아닌 유효성

'보편성'을 누구의
보편으로 시험하는가
(외부검증·형평 감사)

2 책무

제공자이자 수탁자인
국가가 취약한 환자를 향해
답하는가

3 돌봄·인정

관계적 선과 '누구의
필요·데이터가 인정되는가'

거버넌스는 AI가 공공의료를 공공의료답게 — 시장이 버린 곳을 향하게 — 유지하는가를 묻는다.

닫는 질문 — 'AI 기본의료'에 묻기

질문 바꾸기

- 정책의 물음: "이 AI가 안전하고 효율적인가? — 그래서 보편적인가?"
- 공공의료가 물어야 할 것:

"이 'AI 기본의료'가 공공의료를 공공의료답게 — 시장이 버린 곳을 향하게 — 유지하는가?"

- 유효성·책무·돌봄, 세 관문은 모두 그 하나의 물음임

참고 · 근거

사례 · 정책

- Wong (*Epic Sepsis Model*, JAMA IM 2021) · Kaushal (JAMA 2020) · Zondag (JMIR 2024)
- WHO LMM(2024) · EU **AI Act** · 복지부 '**AI 기본의료**'(2026-KIHASA) · 공공보건의료법 §2-§7

이론 · 관점

- van de Poel(2011·2015) · Gabriel(2020) · Fraser·Honneth · Coeckelbergh·Gunkel(2025)
- 김준혁 「건강데이터 소유권」(2021) · 『헬스케어 AI 윤리』 · 『AI 돌봄윤리』

의료임상 현장과 AI

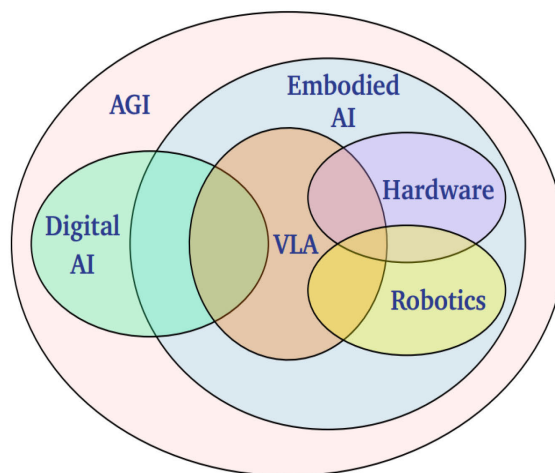
박소연 교수 (인하대학교 의과대학, 한국의료윤리학회 고시이사)

의료 임상현장과 AI

인하대학교 의과대학
한국의료윤리학회 고시이사
박소연

Real-world usage of AI : 그 공동 목표는?

- 거버넌스 (governance): **공동의 목표**를 달성하기 위하여, 다양한 이해 당사자들이 주체적인 행위자로 투명하게 협의하고 의사 결정을 수행할 수 있도록 하는 **사회적 시스템**.



Zhong et al. 2025

Interrelationships among key AI fields

2

임상 유형에 기반한 의료 인공지능 임상 적용 분류 체계

① 예방 (Prevention)	Ex) 생활데이터 기반 위험 예측 / 생활습관 코칭
② 선별 (Screening)	Ex) 건강검진 영상 자동 판독, 당뇨병 위험 예측 모델
③ 진단 (Diagnosis)	Ex) 병리/영상 AI, 심전도 이상 판독 AI, 진단보조 AI
④ 치료 (Treatment)	Ex) 수술보조 AI, 방사선량 자동 최적화, 약물용량 추천 AI
⑤ 예후예측 (Prognosis)	Ex) 재입원 위험 예측, 암 재발 확률 예측 모델
⑥ 추적관리 (Follow-up) ✓ 지원 (Support/Engagement) ✓ 회복/(Recovery) ✓ 재활 (Rehabilitation)	Ex) 웨어러블 데이터 기반 합병증 경고 회복 과정 원격모니터링 챗봇, 환자 교육 AI 운동량 피드백, 재활 AI 코치

김헌성 외. 2025

사례> 당뇨병성 망막병증 진단



사례> Surgical Robotics – The Era of "Active Guidance"

**Annals of
COLOPROCTOLOGY**



Technical Note

Ann Coloproctol 2025;41(5):483–488
pISSN: 2287-9714 • eISSN: 2287-9722
<https://doi.org/10.3393/ac.2025.00906.0129>

Technical implementation and feasibility of the world's first artificial intelligence-assisted augmented reality-based instrument de-occlusion in robotic-assisted right hemicolectomy

Thalia Petropoulou^{1,2}, Pieter De Backer³, Kyriacos Evangelou^{2,4}, Jente Simoens³,
Andreas Polydorou^{1,4}, Alex Motttrie³

¹Department of Colon and Rectal Surgery, The Euroclinic Hospital of Athens, Athens, Greece

²Second Department of Surgery, Aretaieion University Hospital, Athens, Greece

³ORSI Academy, Merelbeke-Melle, Belgium

⁴School of Medicine, National and Kapodistrian University of Athens, Athens, Greece

사례> Surgical Robotics – The Era of "Active Guidance"

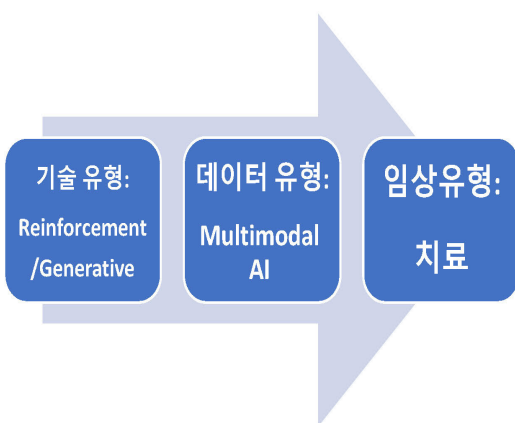
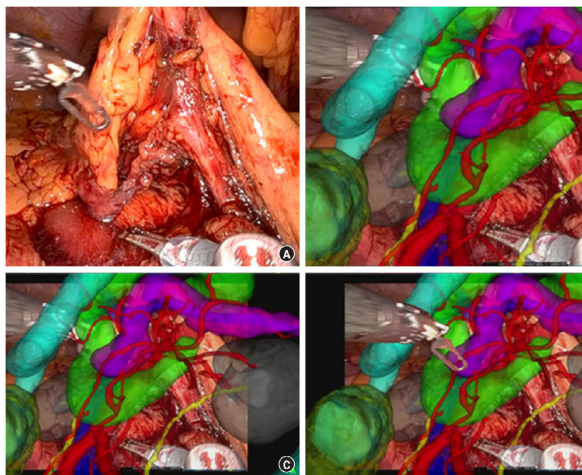


Fig. Instrument occlusion and de-occlusion during surgery.



- (A) 수술 당시의 원본 화면.
- (B) 도구를 가린 3차원 오버레이.
- (C) 인공지능 기반 도구 분할.
- (D) 분할 후 개방되어 도구가 선명하게 보이는 화면.

Surgical Robotics: The Era of "Active Guidance"

1. 실시간 해부학적 오버레이 (Augmented Reality Guidance)

- 과거에는 의사가 수술 전 CT/MRI 영상을 머릿속으로 그리며 수술했다면, 이제 AI는 로봇의 카메라 뷰 위에 실시간으로 디지털 지도를 입히는 것이 가능
 - AI가 혈관, 신경, 요관 등 손상되면 치명적인 구조물을 실시간으로 식별하여 색깔별로 표시
 - 암 조직과 건강한 조직의 Margin을 정확히 구분해 재발률을 낮추고 출혈 위험을 최소화.

Surgical Robotics: The Era of "Active Guidance"

2. 자율적 보조 및 위험 방지 (Safety Barrier & Automation)

- 단순히 '보여주는 것'에 그치지 않고 로봇이 물리적으로 개입하기 시작함
- 가상 장벽 (No-go Zones): 의사가 로봇 팔을 조종하더라도, AI가 위험 지역(예: 주요 동맥 근처)이라고 판단하면 로봇 팔의 움직임을 자동으로 제한하거나 멈추게 하는 '햅틱 피드백' 기술이 도입
- Task 자동화: 봉합이나 시야 확보를 위한 카메라 이동처럼 반복적이고 단순한 작업은 AI가 자율적으로 수행하여 외과의의 피로도 감소

Surgical Robotics: The Era of "Active Guidance"

3, 수술 단계 인식 (Surgical Phase Recognition)

- 최신 논문에서 강조되는 AI 모델(주로 Transformer 기반)은 현재 수술이 어느 단계 인지 실시간으로 파악

- 지능형 워크플로우: "지금은 Dissection 단계이므로 이 도구가 필요할 것"이라고 예측하여 로봇 팔의 세팅을 미리 최적화하거나, 다음 단계를 위한 가이드라인을 화면에 띄울 수 있음.

→ '컨텍스트 인식 AI'를 탑재한 차세대 플랫폼으로 수술 시간을 평균 15~20% 단축시켰다는 데이터.

국내 보건의료 AI 거버넌스?

기관	제도/가이드라인	주요 내용	
과학기술정보통신부	사람이 중심이 되는 인공지능 윤리기준 (2022)	인간 존엄성, 안전성, 투명성, 책임성, 공정성 등 기본 윤리 원칙 제시	의료 AI에 필요한 가치 제시. 하지만 임상 검증이나 책임 규정은 부족
산업통상자원부 / 국가기술표준원(KATS)	ISO/IEC 42001 AI 경영시스템 표준 (2023)	조직 차원의 AI 위험 식별, 통제, 지속적 개선(PDCA) 체계 구축	의료 AI 개발기관의 내부 거버넌스와 감사체계 구축에 활용 가능하나, 임상 안전성은 직접 규제하지 않음
보건복지부 / 개인정보보호위원회	보건의료데이터 활용 가이드라인 (2020~, 2024 개정)	가명정보 처리, 데이터 결합, 보안조치 등 의료데이터 활용 기준 제시	AI 학습용 데이터 활용 근거 제공. 그러나 알고리즘 투명성, 책임성과 같은 AI 자체 윤리 문제는 다루지 않음
보건복지부 / 건강보험심사평가원(HIRA)	디지털치료기기·AI 기반 의료 기술 건강보험 등재 가이드라인 (2023)	임상효과성, 비용효과성, 사회적 가치 중심의 급여평가 기준 제시	상용화 이후 의료 AI의 경제성, 임상적 효과를 평가하지만 개발 단계의 윤리·거버넌스는 다루지 않음
정부(기본법)	인공지능 발전과 신뢰기반 조성 등에 관한 기본법 (2026 시행)	위험기반 관리체계와 정부의 정책 조정 역할 규정	보건의료 AI를 고영향 분야로 분류할 법적 근거 마련. 다만 세부 규제는 각 부처에 위임

김연주 외, 2026

AI Agent의 시대

	DEEP LEARNING (2016-)	AI AGENT(2025-)
핵심	의료진보다 더 잘하는 모델 개발	<u>업무 최적화</u>
개발	대규모 정답 데이터 기반 학습	사전 학습된 모델 사용
연구	빅데이터, 전처리, 라벨링	데이터 및 도구 오케스트레이션

AI 활용 관련 윤리원칙

1. FUTURE-AI 가이드라인

- 2021년부터 50개국 117명의 전문가들이 모여 보건의료 인공지능(AI)의 신뢰성 있는 개발과 실용적 적용을 위한 국제적 지침을 협의하여 발표 (BMJ, 2025).
- 여기서는 AI의 의료 분야 실제 적용이 기술적 진보에 비해 느리게 진행되는 원인을 윤리성 부족, 불신, 규제 부재 등으로 보고, 신뢰할 수 있고 책임 있는 AI를 구현하기 위한 프레임워크와 및 권고사항을 개발.
- FUTURE-AI 가이드라인의 6가지 핵심 원칙은 의료 인공지능을 신뢰할 수 있고 윤리적으로 사용하고자 하는 것으로, 구체적인 권고사항을 포함하고 있으며 AI 개발자, 의료 종사자, 정책 입안자 모두가 따라야 할 기준을 제시

원칙	설명
① 공정성 (Fairness)	편향을 사전에 인식하고 최소화할 것
② 보편성 (Universality)	다양한 환경에 적용 가능하도록 할 것
③ 사용 용이성 (Usability)	실제 임상에서 사용자가 쉽게 활용 가능하도록 할 것
④ 강건성 (Robustness)	실제 환경 변화에도 성능을 유지할 것
⑤ 추적 가능성 (Traceability)	AI 개발부터 배포까지 전 과정 문서화 및 감시
⑥ 설명 가능성 (Explainability)	AI 의사결정 과정에 대한 설명이 가능할 것

13

의료분야에서의 AI 활용에 따른 윤리적 문제

세계보건기구의 가이드라인: AI의 의료분야 활용 과정에서 발생할 수 있고 고려되어야 하는 윤리적 문제점을 아래와 같이 제시(Ethics and governance of artificial intelligence for health: WHO guidance, 2021).

WHO consensus ethical principles for use of AI for health



14

국내 AI 적용 관련 기존 문제점

- (1) 부처 간 분절적 관리 구조 및 행정적 비효율
 - (2) '선언적 수준'에 그친 공정성 : 알고리즘 편향 검증 체계의 미제도화
 - (3) 윤리 원칙은 존재하지만 실제 집행·감독 체계가 부족
 - (4) 의료 AI의 설명가능성(Explainability) 기준 부재
 - (5) AI 오류 발생 시 불명확한 책임 주체 및 배분 기준 (책임 공백)
- 보건의료 AI 윤리 원칙과 제도적 기반은 갖추고 있지만, 이를 실제로 검증·감독·책임질 수 있는 운영 거버넌스 체계는 아직 미흡

15

1. 정책 주도과 현장 자율성 사이 균형 유지

- ‘빠른 도입’에 의의를 둘 것이 아니라 정부는 법/재정적 지원과 인프라 구축을 담당하되, 현장 의료진에게는 자율적 도입 권한을 부여하는 이중 전략이 필요함.
- 일례로 정부가 인공지능 도입 병원을 지정해 재정 지원을 하되, 그 구체적인 활용 범 관련 의료진의 결정권을 존중함으로써, 제도적 안전성과 현장 수용성을 동시에 확보하는 것이 필요함.
- 특히, 보험 수가 개선과 제도적 지원, 전문과별 맞춤형 도입 전략, 환자 참여 확대 등은 시급히 해결해야 할 정책 과제임.
- 더불어 국제 사례에서 확인된 것처럼, 윤리적 신뢰성과 투명성 확보는 그 지속 가능성을 결정짓는 핵심 요소임.

2. 다층적 협력관계의 필요성

- 의료 AI의 안정적 발전과 제도적 정착을 위해서는 기술개발, 임상검증, 정책규제의 세 영역이 유기적으로 결합된 다층적 거버넌스 체계가 필수적임.
- 정부-학계-산업계의 삼자협력 구조를 기반으로, 정책 결정, 기술 표준화, 임상 검증, 산업 실증 간의 역할을 명확히 구분하고 동시에 상호 피드백이 가능한 운영 모델 필요
- 즉, 기술의 신뢰성, 제도의 정합성, 산업의 혁신성을 동시에 확보하는 통합형 의료 인공지능 거버넌스를 바탕으로, 향후 정책적 실행력과 함께 기술적 변화에 민감하게 반응할 수 있는 기반 마련 필요
- 임상현장 유지와 관리에서는 개발사-의료기관-정부 간의 상호 피드백과 공동 책임을 전제로 운영되어야 함.

3. 유연하면서도 표준화된 관리체계 필요

- 기술의 급속한 진화와 사회적 책임을 수용할 수 있도록, 유연하면서도 표준화된 관리체계가 필요
- ① 임상 현장/데이터/기술의 세 축이 상호 교차적으로 작동 하는 다차원 통합 시스템임
- ② 기술 발전의 속도와 임상 안전성의 균형 확보를 위해 AI 업데이트 등 관련하여 위험도(risk) 기반 차등 관리, 정기/이벤트형 혼합 구조, 기술성숙도, 사전 승인형 제도화 등 차등적 설계 고려. 신뢰성과 지속가능성의 보장하는 적응형 거버넌스 구축 필요
- ③ 법제화 및 표준화: 법/정책 체계에 분류체계를 명시적으로 반영하고, 국가 차원의 데이터 표준화 및 상호운용성 인프라를 구축해야 함.
- ④ 지속적 검증과 피드백: 정기적 검증과 공공 데이터 기반 사후평가를 통해 지속적으로 성능 및 문제 관리.

References

1. Zhong, Yifan, et al. "A survey on vision-language-action models: An action tokenization perspective." arXiv preprint arXiv:2507.01925 (2025).
2. 한국보건산업진흥원, 의료분야 인공지능 활용 실태 조사·분석 (2025)
3. Yeonju Kim et al. Institutionalization of Ethical Governance for Artificial Intelligence in Healthcare: A Comparative Analysis of International Guidelines and Policy Implications for Korea. Korean Public Health Research Vol. 52, No. 2, pp.65~81. (2026).

Thank you

syparkmd@inha.ac.kr

보건정책과 인공지능

이경도 교수 (울산대학교 의과대학, 한국생명윤리학회 학술이사)

보건정책과 인공지능

인공지능과 가버넌스

울산의대 인문사회의학교실 이경도 2026.06.10

Special thanks to 성호경 선생님(국립중앙의료원)

+ no conflict of interest to declare

Overview

- 문제 의식과 한 가지 사례: 응급실 뽕뽕이
- 희망, 가능성, 그리고 hype(거품)
- 현실, 제약, 그리고 위험
- 결론

사회 의료·건강

청주서 임신부 '뽕뽕이' 끝 태아 숨져...24시 분만 모자의료센터, 작동 못했다

허윤희, 신형철, 최예린 기자

수정 2026-05-04 09:33

기사 읽어드립니다

4:53 ▶ 🔊

‘29주 태아’의 비극

고위험 임신부, 권역 내 진료 안 되면 모자의료센터가 병원 찾아준다

신소윤 기자

수정 2026-05-26 19:56

기사 읽어드립니다

3:21 ▶ 🔊

‘분만 뽕뽕이’에 정부 이송·전원체계 개선키로
동네 전문의엔 모자의료센터 파트타임 허용

김민석 총리 “응급실 뽕뽕이 태아 사망 사고”

“산과 법적 보호·모자병원 신설, 한국형 모자 전문병원 모델 검토”



2026.05.04 17:05 댓글쓰기



정치 · 정치일반

‘응급실 뽕뽕이’ 태아 사망에...당정 “특단 대책 마련”

전문가 8명 자문위원으로 위촉
모자 의료센터 등 모니터링 강화
전원 이송·보상체계 개선 논의도

강도림 기자

+구독

입력 2026-05-08 17:51

지면 6면



응급실 뽕뽕이, AI로 해결합니다

실시간 데이터 기반 AI 매칭으로 가장 빠르고 적절한 병원으로 안내

1. 환자 정보 입력

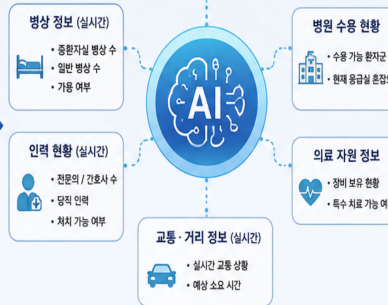
응급구조사가 현장에서
환자 정보를 입력



- 입력 정보 (예시)
- 환자 기본 정보
 - 나이 / 성별
 - 의식 상태
- 주요 증상 / 중증도
- 활력징후
- 이송 수단 정보
- 현재 위치

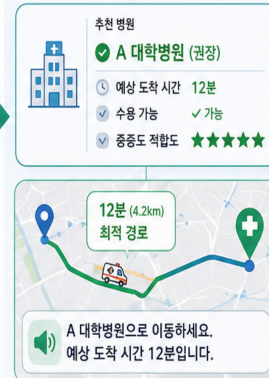
2. AI가 실시간 분석 및 최적 병원 매칭

전국 병상·인력·수용 가능 여부를 실시간 분석



3. 최적 병원 추천 및 안내

AI가 가장 적합한 병원을 추천하고
최적 경로로 안내



시스템 전체 흐름



기대 효과

- ✓ 불필요한 병원 탐색 시간 감소
- ✓ 환자 중증도에 맞는 최적 치료
- ✓ 병상·의료 자원 효율적 배분
- ✓ 의료진 업무 부담 경감 및 시스템 안정화

✓ AI가 데이터로 길을 찾고, 환자는 생명으로 답합니다.

By ChatGPT

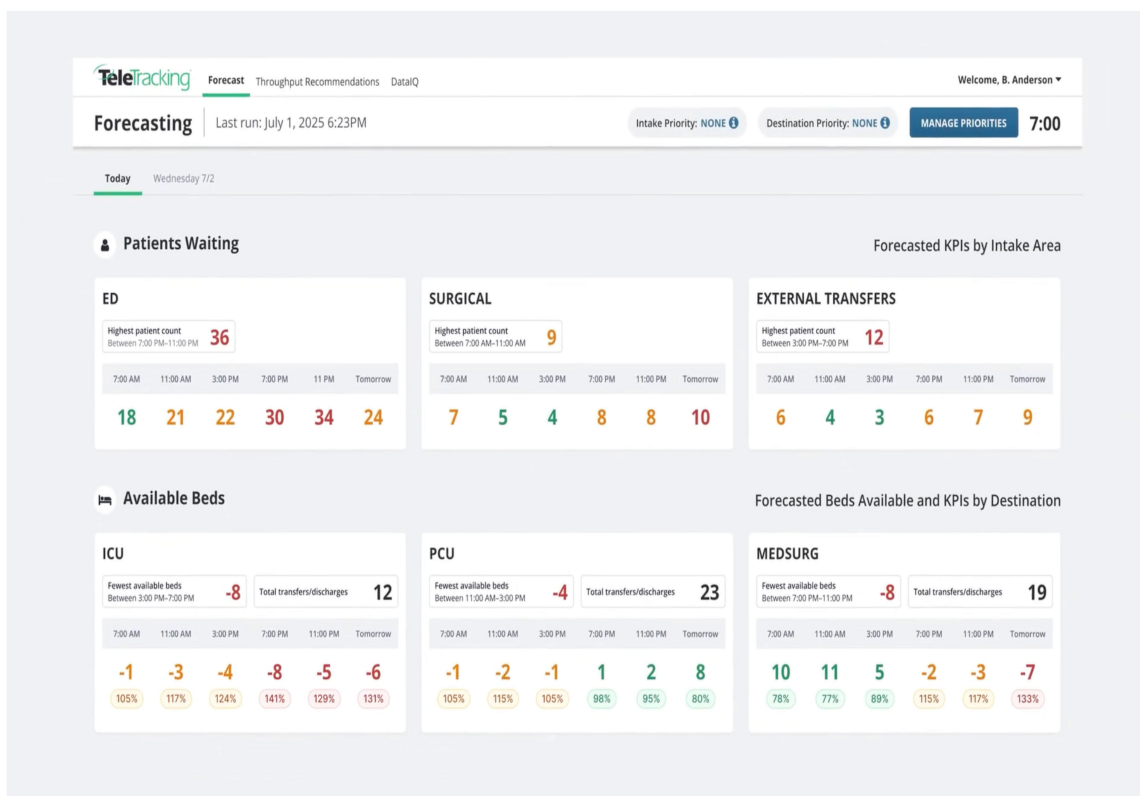
Transforming Healthcare Operations with AI-Powered Insights

We're integrating over 30 years of operational expertise with the most advanced AI technology in the industry. The result is a Boundaryless platform that ingests and analyzes data across your enterprise—unlocking new capacity and putting patients at the center of every decision.

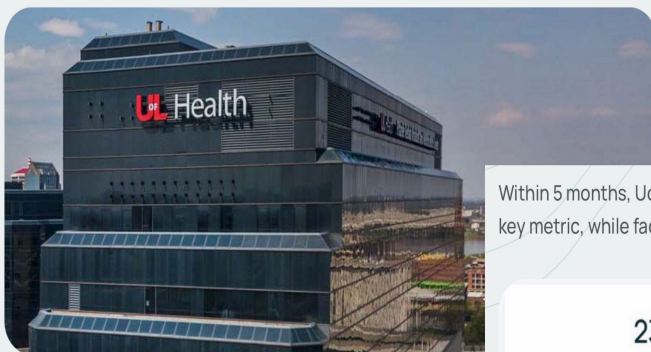
AI-Based Decision Support...Workflow Automation...Enterprise-Wide Visibility

[Learn more →](#)





Helping Patients Get Home Faster: How Decision IQ Transformed Patient Flow



Within 5 months, UofL achieved sustained operational improvements across every key metric, while facing increased admissions volume.

23%
reduction in length of stay

44%
increase in morning discharges

15 min
faster discharge processing

53%
increase in ED-to-inpatient admissions processed

*Results were measured against a pre-implementation baseline and validated using a contemporaneous control site comparison across key throughput metrics.

NHS Privacy Enhancing Technology



A guide to Palantir's Foundry software

Summary Document

The Patients Association has worked with Palantir and patients to develop a guide to Foundry. Foundry is a software platform that organisations – including a growing number of NHS organisations – use to securely bring together, manage and use data.



Reality

- 실시간 가용 병상 정보
 - 빈 병상 여부만이 중요하지 않음
 - 실제 중증 진료 가능한 인력
 - 중환자실 입실 대기 원내 환자 여부
 - 수술실, 중환자 대기 인력
 - 응급실 입실 가능? 진료 능력?
- 민간 병원에 대한 특정 환자 입실/진료 등을 얼마만큼 강제할 수 있는가?
 - 인센티브 vs. 기준 및 처벌 등

내 손안의 응급실 - 간편조회

시/도

서울 전체

반경

10km

20km

40km

60km

100km

서울 송파구

응급실 가용 병상 조회 (51건)

(가용병상수/전체병상수, *대기환자)

기관 명	응급실일반	응급실소아	분만실	응급격리	일반격리
센터 강남성심 서울특별시 영등포구 신길로 1 (... 검색기 >)	보통 15/22	원활 2/2	가능/3	원활 1/1	원활 2/2
센터 강남세브 서울특별시 강남구 언주로 211, (... 검색기 >)	혼잡 1/21	보통 1/2	가능/1	원활 1/1	원활 2/2
권역 강동경희대 서울특별시 강동구 동남로 892 (... 검색기 >)	보통 10/15	원활 3/3	가능/8	원활 2/2	보통 2/3
센터 강동성심 서울특별시 강동구 성안로 150 (... 검색기 >)	보통 11/15	-	가능/1	원활 1/1	원활 6/6
센터 강북삼성 서울특별시 종로구 새문안로 29 (... 검색기 >)	혼잡 0/15	원활 2/2	가능/1	원활 1/1	원활 2/2

Figure 3: HAvBED Data Elements

Denver Health Medical Center
777 Bannock Street
Denver, CO 80204-4507

Facility Last Updated: 12/13/2007 08:46:46
Availability As Of: 12/13/2007 06:47:48

Available Beds
Trauma Level: 1

Bed Category	Current	24 Hour	72 Hour	AHA Survey	Remarks
Adult ICU	7	12	14	52	2 Isolation
Medical/Surgical	5	40	50	194	
Burn	0	0	0	0	
Pediatric ICU	2	3	4	7	
Pediatrics	5	8	8	15	
Psychiatric	4	9	12	44	
Airborne Infection Isolation	3	5	7	34	3 on 8th floor
Operating Rooms	3			12	

Resources
Emergency Department Status: Open
Mass Decontamination Status: Available
Ventilators Available: 5

실시간 데이터 제공

인센티브 X

노력과 자원 필요

Artificial intelligence and evidence-informed policy – emerging challenges and opportunities

Discussion paper

Reason-giving

Policy phase	AI opportunities (Section 4)	Key risks (Section 5)	Governance insights (Section 6)	Practical responses (Section 7)
Cross-cutting (applies to all phases)	⇒ More comprehensive data integration ⇒ Research synthesis ⇒ Scenario modelling ⇒ Adaptive feedback across the cycle ⇒ Accelerate movement through policy cycle	⇒ <u>Opacity and accountability</u> ⇒ Power concentration and equity ⇒ Regulatory gaps ⇒ Erosion of human expertise ⇒ Bias and discrimination ⇒ Epistemic injustice and knowledge exclusion ⇒ Resource intensity	⇒ Transparent and auditable decision-making ⇒ Participatory engagement ⇒ Data governance and quality (FAIR principles) ⇒ Rights and accountability frameworks ⇒ Evidence standards and reporting ⇒ Risk-based regulation and life cycle oversight	⇒ Prioritize augmentation over automation ⇒ Human-in-the-loop governance ⇒ Combine AI with collective intelligence ⇒ Navigate evolving concepts of evidence ⇒ Link AI research to policy applications

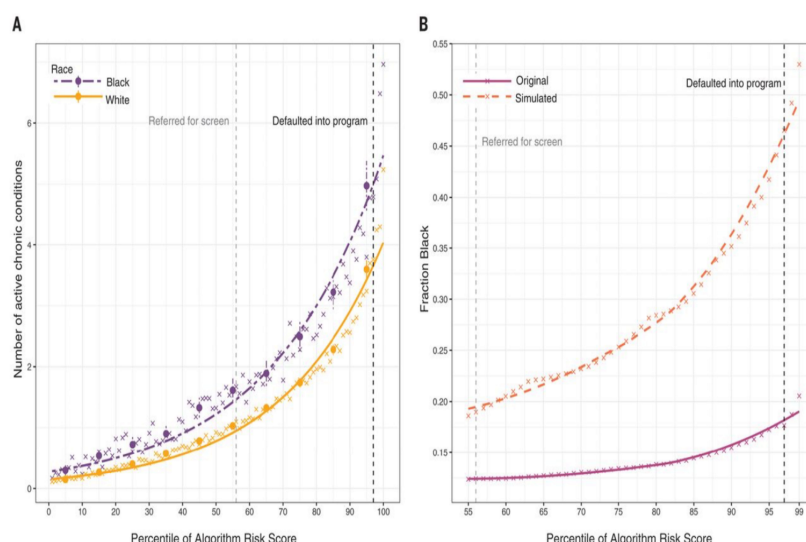


Fig. 1 Number of chronic illnesses versus algorithm-predicted risk, by race.

(A) Mean number of chronic conditions by race, plotted against algorithm risk score. (B) Fraction of Black patients at or above a given risk score for the original algorithm ("original") and for a simulated scenario that removes algorithmic bias ("simulated": at each threshold of risk, defined at a given percentile on the x axis, healthier Whites above the threshold are replaced with less healthy Blacks below the threshold, until the marginal patient is equally healthy). The x symbols show risk percentiles by race; circles show risk deciles with 95% confidence intervals clustered by patient. The dashed vertical lines show the auto-identification threshold (the black line, which denotes the 97th percentile) and the screening threshold (the gray line, which denotes the 55th percentile).

Algorithmic bias

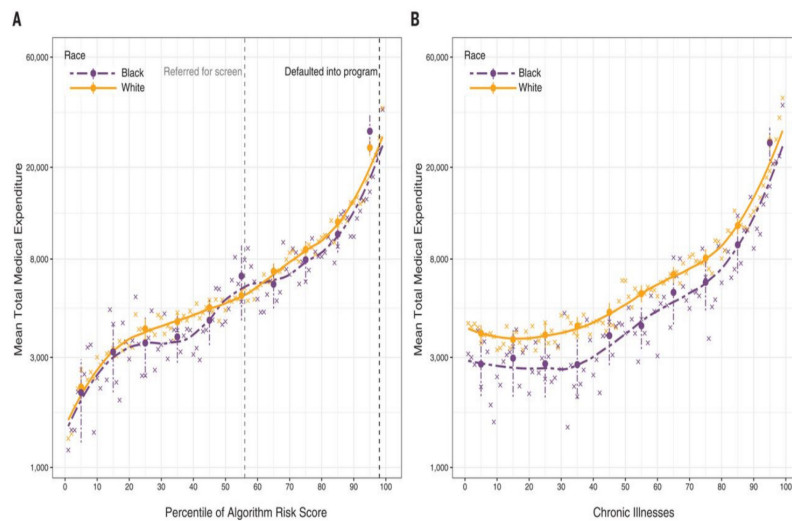


Fig. 3 Costs versus algorithm-predicted risk, and costs versus health, by race. (A) Total medical expenditures by race, conditional on algorithm risk score. The dashed vertical lines show the auto-identification threshold (black line: 97th percentile) and the screening threshold (gray line: 55th percentile). (B) Total medical expenditures by race, conditional on number of chronic conditions. The x symbols show risk percentiles; circles show risk deciles with 95% confidence intervals clustered by patient. The y axis uses a log scale.

• Ziad Obermeyer et al., Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 447-453 (2019). DOI:10.1126/science.aax2342

Value alignment



Palantir: NHS pilot's "success" is questioned as second figure at major trust is linked to the tech giant

BMJ 2026 ; 393 doi: <https://doi.org/10.1136/bmj.s680> (Published 13 April 2026)

Cite this as: BMJ 2026;393:s680



Table 1. Examples of opportunities, risks, governance insights and practical responses across the policy cycle

Policy phase	AI opportunities (Section 4)	Key risks (Section 5)	Governance insights (Section 6)	Practical responses (Section 7)
Cross-cutting (applies to all phases)	⇒ More comprehensive data integration ⇒ Research synthesis ⇒ Scenario modelling ⇒ Adaptive feedback across the cycle ⇒ Accelerate movement through policy cycle	⇒ Opacity and accountability ⇒ Power concentration and equity ⇒ Regulatory gaps ⇒ Erosion of human expertise ⇒ Bias and discrimination ⇒ Epistemic injustice and knowledge exclusion ⇒ Resource intensity	⇒ Transparent and auditable decision-making ⇒ Participatory engagement ⇒ Data governance and quality (FAIR principles) ⇒ Rights and accountability frameworks ⇒ Evidence standards and reporting ⇒ Risk-based regulation and life cycle oversight	⇒ Prioritize augmentation over automation ⇒ Human-in-the-loop governance ⇒ Combine AI with collective intelligence ⇒ Navigate evolving concepts of evidence ⇒ Link AI research to policy applications
Understanding the problem	⇒ Pattern detection and prediction ⇒ Enhanced situational awareness	⇒ Data bias and misrepresentation ⇒ Reinforcement of historical bias ⇒ Oversimplification of complex issues ⇒ Mis/disinformation	See cross-cutting principles above	⇒ Living evidence workflows ⇒ Multidisciplinary oversight ⇒ Context-specific question framing ⇒ Blend AI outputs with expert judgement
Designing the solution	⇒ Idea generation and option surfacing ⇒ Resource allocation optimization ⇒ Operational efficiency improvements	⇒ Over-optimization of measurable objectives ⇒ Algorithmic narrowness ⇒ Legitimacy risks ⇒ Ethical blind spots ⇒ Innovation constraints ⇒ Stakeholder exclusion	See cross-cutting principles above	⇒ Pilot projects and structured appraisals ⇒ Technology readiness assessments ⇒ Algorithmic impact assessments ⇒ Map tasks to comparative advantage
Achieving impact (Including implementation, monitoring & adjustment)	⇒ Adaptive policies and real-time feedback ⇒ Implementation dashboards ⇒ Public sentiment analysis ⇒ Iterative evidence-based decision-making ⇒ Learning from policy failures	⇒ Incentive distortions and system gaming ⇒ Cybersecurity threats to programme delivery ⇒ Adaptability challenges in novel contexts ⇒ Overemphasis on quantitative metrics ⇒ Causality and attribution gaps ⇒ Feedback loop risks ⇒ Privacy violations ⇒ Resistance to change and path dependency ⇒ Policy drift and inflexibility	See cross-cutting principles above	⇒ Bias-testing and auditing ⇒ Post-deployment audits ⇒ Real-time analytics for adjustments ⇒ Flexible governance and adaptive protocols ⇒ Systematic risk-mapping

Note: This table presents selected examples that are relevant to each phase. Cross-cutting risks affect all policy phases (Section 5.1), and governance insights and practical responses build on foundational principles that apply across the cycle. Several opportunities and responses may be relevant to multiple phases. For a comprehensive discussion, see sections 4-7.

결론

- 보건의료 정책 전반의 문제 인식, 해결 방안 모색 및 결정, 정책 입안 및 운영 모든 단계에서 인공지능을 활용할 수 있는 여지가 있다.
- 그렇지만 그 활용은 현존하는 여러 현실적, 정치적, 사회적, 정책적, 윤리적 문제를 해소하거나 감소시키기보다는 증폭시킬 우려가 있다.
- 인공지능의 사회적 신뢰와 가치 정렬 등 그 자체의 문제는 물론 국민건강보험공단이나 보건복지부 등 정부 기관의 투명성과 합리성을 재고하지 않으면 광범위한 인공지능 활용 및 민간 인공지능 기업과의 협업 등은 상당한 반발을 마주할 것이 예상된다.
- 결론적으로 쉬운 길은 없다: 국민건강보험을 포함한 보건의료 정책 결정 과정과 가버넌스 등 정책의 정당성(legitimacy)와 수용성(acceptance)을 향상시킬 수 있는 방안을 모색해야 한다.

감사합니다.

kdlee@amc.seoul.kr

데이터은행, 인공지능 개발과 활용의 공공성

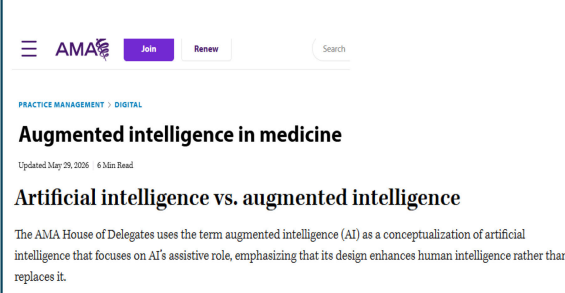
김한나 교수 (연세대학교 의과대학, 한국생명윤리학회 총무이사)

생명윤리 유관기관 공동학술대회 | 인공지능과 거버넌스 | 2026. 6. 10.

데이터은행과 인공지능

김한나 (연세의대 인문사회이학교실,
한국생명윤리학회 총무이사)

보건의료 인공지능에 대한 개인과 집단의 인식 변화

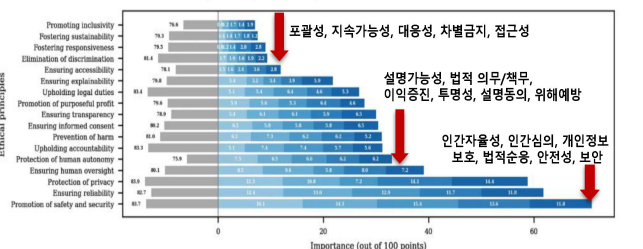


「연세대학교 의과대학생 생성형 인공지능 활용 지침」
2026. 02. 01 제정

배경 및 목적

연세대학교 의과대학 교육과정은 사고를 명확히 하고, 성찰 학습을 촉진하여, 비판적 추론과 분석 능력을 개발하도록 설계되어 있다. 연세대학교 의과대학 학생은 학업 과정에서 이러한 역량을 달성하기 위해 최선의 노력을 기울여야 한다.

생성형 인공지능(generative artificial intelligence, gAI) 기술은 지식 기반 콘텐츠의 생성을 통하여 개인의 학습과 연구에 보조 수단으로 활용될 수 있다. 반면 환각, 출력의 확률적 특성과 편향, 프롬프트 인젝션 취약성, 개인정보보호 문제, 입력 데이터의 불투명성 등 기술적 한계도 존재한다. 이러한 특성으로 인해 의과대학생과 의사는 gAI를 바르게 활용할 수 있는 고도의 비판적 사고와 판단 능력을 갖추어야 한다.

- 의사 대상 조사(미국 의사협회, 2026)
 - 전문적 용도로 AI 사용(80%, 2026)
 - 환자들 대부분 건강 질문에 AI 사용, 대체로 허용가능
 - 기대: 환자 치료 능력 향상 (65%, 2023 → 75%, 2026), 진단 능력 향상, 업무 효율성, 번아웃 감소
 - 우려: 환자 개인정보 보호, 의사 환자 관계 변화, 기술 의존으로 인한 역량 약화 (88%)
- 일반인 대상 조사 (한국, 2023)
 

인공지능과 인간의 윤리적 의사결정: 데이터 은행 적용

대규모 언어모델(LLMs): 의료윤리 학술연구의 지원, 환자나 의사를 대변, 도덕 자문가로 기여 가능



II 아마도 의료윤리가 직면한 다음 과제는 윤리적 수행 능력에서 AI가 뛰어난 세상에서 인간의 역할을 다시 규명하는 것이다. (...)

인지적으로나 도덕적으로 한계를 가진 인간인 우리를 우리가 만들어가는 이 거대한 디지털 세상 속에 포함시키는 것이다. II

AI가 인간보다 우월한 임상 의사 결정과 윤리적 판단까지 가능한 미래를 향해 가는 상황에서,

데이터은행 관련 어떠한 거버넌스를 추구할 것인가?

- 정밀의료를 위한 데이터은행 논의를 중심으로 전개

Savulescu J. Academic freedom, artificial intelligence and the blandification of ethics. *Journal of Medical Ethics* 2025;51:438-440.

정밀의료의 활성화를 위한 데이터은행의 필요성

- 개인의 자발적 동의와 기증을 통해 구축된 대규모 멀티모달(multimodal) 데이터 베이스
- Open science and open commons
 - 원칙: 과학적 발견 과정에서 가능한 한 이른 시점에 모든 종류의 과학 지식이 공개적으로 공유되어야 함 (방법론, 소스, 데이터, 접근, 피어 리뷰, 교육 자원)
 - 의료정보 커먼즈: 대규모 인구의 건강·의료·유전체 정보가 광범위하게 공유되는 네트워크 환경
- 개인의 생애주기에 걸쳐 임상 활용 확대: 임상 아카이브와 연구 지원 목적의 결합
 - 기증 대상 연령대의 변화, 데이터의 장기 보관, 혈연 분석, 데이터 관리 효율성 및 지속가능성
 - 단순한 이상적 목표의 오픈 사이언스가 아닌, 이미 실현 중인 현실로의 전환에 역할

M Phillips, BM Knoppers (2019) Whose Commons? Data Protection as a Legal Limit of Open Science. *The Journal of Law, Medicine, and Ethics*. 47:1.

유전체데이터의 의미

- "우리는 누구인가"
 - 유전정보: 개인과 집단의 생물학적 정체성을 드러내는 비대체적 정보
 - 개인 민감정보의 궁극적 형태: 질병 위험, 약물 반응, 행동 및 정신적 성향에 대한 예측 가능성
 - 생명의 설계도로 기능할 잠재력을 가진 고도 자원 (Cyranoski, 2019)
- 후성유전학 접근 방식: 건강 관점에서 개인 수준의 분석, 관리, 개입 가능
- 개인정보의 자기결정권: 시대 간 정의와 장기적 책임 문제
 - 유전체 데이터에 대한 자기결정권 의미는 시대적으로 불균형적임. 현 세대는 조상의 유전체 정보를 비공개 상태에서 물려받았기 때문에, 자신의 유전정보를 공개하기로 결정하는 행위가 정당화될 수 있음. 그러나 아직 존재하지 않는 주체의 권리까지 현 세대의 선택에 의해 좌우됨. (Andorno, 2004; Knoppers et al., 2014)

Cyranoski, D. (2019). The CRISPR-baby scandal: What's next for human gene-editing. *Nature*, 566(7745), 440-442.; Andorno, R. (2004). The right not to know: An autonomy based approach. *Journal of Medical Ethics*, 30(5), 435-439; Knoppers, B. M., Harris, J. R., Budin-Ljøsne, I., & Dove, E. S. (2014). A human rights approach to an international code of conduct for genomic and clinical data sharing. *Human Genetics*, 133(7), 895-903.

인간의 행위주체성

- 인간의 주체성, 결정권, 행위주도성, 감독권
 - 긍정적 이해: 기술을 통해 인류가 변형하는 것을 추구할 자유, 애플리케이션이나 기술을 선택할 자유, 타자와 상호작용을 할 권리, 동의의 철회
 - 인간 권한의 지속적 부여와 인식론적 강화
 - 기술의 부정적인 영향에 대한 숙고에 기반함. 데이터 접근과 해석, 판단 능력에 따라 지속적으로 부여. 데이터 착취나 조작, 인공지능의 편향과 감시에 대항하는 자유
 - 제한적 역량 지원: 개인과 집단 윤리 사이 기준 마련에 대한 고려
- 행위자 별 역할 제고
 - 임상의사의 예: 자신의 의료 현장에서 AI 도입과정에 직접 참여하거나 책임을 갖길 원함 (85%), AI에 대한 추가 교육과 훈련을 원함 (92%) (AMA, 2026)

질병관리청 국립보건연구원. (2023) 보건의료 분야 인공지능 연구자를 위한 연구윤리 지침.; American Medical Association, Center for Digital Health and AI. 2026 Physician Survey on Augmented Intelligence. 2026.

인간의 자율성 보호

- 데이터뱅크 정보제공에 대한 개인의 자율성 존중과 동의
 - 기구축 코호트: 생명윤리법 상 AI 개발 전 유전체 연구 전제로 개인 동의 도입
 - 동의의 유효성: 충분한 설명에서 충분한 이해에 근거한 동의로
 - 임상연구 동의 과정에서 LLM 활용: 참여자는 연구 목적, 절차, 위험, 이익, 개인정보 활용 범위 등에 관한 설명을 LLM의 응답을 통해 접할 수 있다. 그러나 LLM이 부정확하거나 허위 정보를 생성 시, 참여자는 사실과 다른 설명에 근거하여 연구 참여 여부를 판단할 수 있다. (헬스케어 인공지능 윤리지침 개정(안) 중)
 - 일차 목적인 임상 목적과 그 외 목적의 구분
 - 개인(예: 신생아, 사망자)과 기관 역량 차이에 대한 세밀한 접근 방식 필요

인간의 감독과 통제

- 데이터은행의 수집, 관리 의무 효율적 구분
- 연구설계 단계: 고영향 인공지능 프로그램 개발 목적 식별, 심의·감독 체계 구축, 사례 공유
- UK Biobank: AI 활용 연구자들에게 현재 통용되는 AI 활용 기준의 선제적 수행을 권장
 - 안전성, 보안, 견고성
 - 투명성, 설명가능성
 - 공정성
 - 책임성과 거버넌스
 - 이의 제기 가능성과 구제 수단
- 고위험 인공지능 시스템을 시장에 출시하거나 서비스로 공급하기 전 검증 (Tripod-AI, Consort/Spirit AI 등)
- 데이터은행: 절차에 따른 정제·정확한 데이터 제공, 필요 시 윤리적 판단 과정을 구조화된 규칙(알고리즘)을 제공하고 확인

데이터 보호와 개인정보보호 법제 의미

- 데이터 보호
 - 개인 데이터의 전 과정 규율 (수집, 저장, 처리, 공유, 파기)
- 목표: 데이터 처리 과정과 결과에서 공정성 확보
- EU 일반개인정보보호규정(GDPR)
 - 데이터 주체: 접근권, 수정권, 처리 반대권 부여
 - 데이터 관리자: 개인정보 처리의 정당성 입증 의무 부과
 - 건강 정보와 유전정보 등 민감정보 처리에 추가 조건 요구
 - 연구 목적 시 포괄적 동의 허용 조항 포함: 데이터 공유와 활용의 여지 확보

- 현재 경향: 일부 국가에서 연구나 공익 목적 데이터 처리 시 동의 외 법적 근거로 '공익'이나 '정당한 이익' 활용 증가
- 통합적 관점에서 정책적 목표 조정
 - 목적 별 데이터 공유 확대
 - 개인정보 보호 강화

개인 정보 보호와 위해 예방

- 이익과 위험 사이
 - 기존 불평등 심화, 충분한 분석자원을 가진 주체의 데이터 활용 쏠림, 공적자금 구축 오픈 사이언스의 민간 기업 혜택 제공 집중 가능성
 - 유전정보 시스템 상 데이터 수집, 저장, 공유 전 과정에서 보안 취약성 및 노출 위험 존재(Schumacher et al., 2020)
 - 사이버보안: 23andMe 데이터 유출 사고, 엔트로픽, 인공지능 윤리적 보증·평가, 바이오보안법(안) (미국)

- 개인정보보호 공개 대응: UK Biobank
 - 참여자 수준의 데이터 공개 금지
 - 승인된 신청범위 외 제삼자 공유나 제삼자가 접근가능한 웹기반 저장소 업로드 금지
 - 편향(참여집단, 데이터처리), 인종 등 집단 별 격차(낙인) 등 연구 진행
- 우생학, 건강결정론, 차별의 고착화 -> "다양한 인간을 위한 평등과 공존"의 실현을 위하여

Schumacher, G. J., et al (2020). Genetic information insecurity as state of the art. Frontiers in Bioengineering and Biotechnology, 8, 591980.

투명성, 신뢰성

- 투명성: 더 나은 연구 및 의사결정 지원, 데이터 활용 공개, 중요 판단 지점 및 오류가능성 식별, AI 사용 관련 책임 소재 명확화를 통한 사회적 신뢰 확보, AI 투명성에 대한 규제적 요구는 아직 불확실함
- 신뢰성: 기술의 공정성을 사회적으로 인정받는 것, 다양한 이해관계자와의 소통, 개인정보 보호 체계에 대한 긍정적인 인식, 피해 예방
- 과학권과 이익 공유의 실천 노력
 - 데이터 추가가능성 확보: “연구자는 데이터 출처를 알아야 하며, 기증자는 본인의 인체유래물 사용처를 알 수 있어야 한다 (UKCRC)” → 연구 프로젝트 대중 공개 (참여자), 데이터 접근 로그 기록 (기관)
 - 설명가능성 (Genomed4all, white-box)
 - 데이터 접근 및 분양 규칙의 문서화: AI 활용 구체화

질병관리청 국립보건연구원. (2023) 보건의료 분야 인공지능 연구자를 위한 연구윤리 지침; UKCRC Tissue Directory and Coordination Centre. <https://biobankinguk.org/transparency-principles/>; GenoMed4All Consortium. (2021–2024). GenoMed4All factsheet: Genomics and personalised medicine through artificial intelligence.

마지막으로...

- 유전체바이오 데이터뱅크의 역할 전환을 통한 데이터 이용 활성화
- 효율성과 혁신의 강조로, 데이터뱅크의 윤리적 거버넌스는 병목현상으로 이해되어 단순화되거나 축소 될 위험 존재
- 우리는 선택의 기로에 있음: 윤리적 판단을 AI에 위임하는 윤리 자동화, 윤리적 고려가 제거된 탈윤리적 거버넌스
- 정밀의료에서 주요 이해관계자이자 참여자인 인간의 윤리적 역할을 위한 제도적 장치: 비판적·협력적 사고, 집단 윤리 및 규범 발전, 경험과 책임있는 실천 지지



감사합니다.

의료인공지능과 민형사상 책임의 귀속

김천수 교수 (성균관대학교 법학전문대학원, 대한의료법학회 상임고문)

최민영 교수 (강남대학교 법행정학과, 대한의료법학회 총무이사)

생명윤리 유관기관 공동학술대회 “인공지능과 거버넌스” 발표문

의료인공지능과 민형사상 책임의 귀속

김천수(성균관대)·최민영(강남대)

발표 순서

I. 의료인공지능의 활용 현황

1. 진단
2. 수술
3. 기타 예방 및 건강관리

II. 논의의 전제

III. 인공지능의 민·형사 책임상 지위

1. 계약 책임상 지위 없음
2. 불법행위 책임상 지위 없음
3. 형법상 행위능력·책임능력·형벌능력 부인

IV. 의료인공지능 사고의 책임 주체 및 내용

1. 진료채무 이행을 인공지능을 이용하여 제공한 의사 측의 민사책임
2. 인공지능 공급자의 민사책임
3. 의사와 인공지능 공급자의 형사책임

V. 현행 법체계의 한계와 법정책적 개선방향

1. 자율성의 단계별 책임 귀속의 내용을 논하자는 제안
2. 현행법의 한계: 증명 곤란 등
3. 인공지능 사고로 인한 손해의 배상책임에 관한 특별법의 제정
4. 기타

I. 의료인공지능의 활용 현황

1. 진단 – 데이터 편향의 문제
2. 수술 – 자율성이 없던 기존 로봇수술에서 현재 로봇의 자율성이 약간 추가되었음
3. 기타 예방 및 건강관리 – 의료와 비의료의 경계선상에 있음.
의료행위와 일반적인 소비행위 사이에 위치함

II. 논의의 전제

1. 피지컬 인공지능의 권리능력 부인
2. '의료인공지능 사고'에 한정함: '인공지능을 이용한 의료행위'
(이하의 인공지능 진료는 이를 의미함)로 환자 측에 손해(법익침해)가
야기된 사고(이하의 의료인공지능 사고는 이를 의미함)
3. 이하에서 인공지능은 의료인공지능을, 의료사고는 의료인공지능사고를 의미함.
달리 특별한 언급이 없으면 의료 인공지능 관련법(인공지능기본법, 의료법,
의료기기법, 디지털의료제품법)의 용어를 그 정의대로 사용함

의료인공지능과 민형사상 책임의 귀속 2

III. 인공지능의 민·형사 책임상 지위

1. 계약 책임상 지위 없음
 - (1) 권리의무의 귀속 주체가 되는 계약 당사자의 지위가 부인됨
 - (2) 이행대행자나 이행보조자의 지위도 부인됨
2. 불법행위 책임상 지위 없음
 - (1) 일반불법행위의 행위자 지위가 부인됨
 - (2) 타인의 불법행위에 대한 사용자는 물론 피용자의 지위도 부인됨
3. 형법상 행위능력·책임능력·형벌능력 부인
 - (1) 현 단계의 인공지능에 국한됨
 - (2) 인공지능의 자율성이 현격히 증가한다면, 법인의 형사책임 인정과 유사한 법리의 적용을
논할 여지는 있음

의료인공지능과 민형사상 책임의 귀속 3

IV. 의료인공지능 사고의 책임 주체 및 내용

1. 진료채무 이행을 인공지능을 이용하여 제공한 의사 측의 민사책임

(1) 진료계약 당사자의 계약책임

- ① 유상계약인 진료계약의 경우, 현행 매도인의 담보책임 규정(민법 제570조 이하)이 적용됨
- ② 채무불이행책임의 규정(민법 제390조 이하)이 적용됨
- ③ 설명의무 위반 책임
 - 인공지능진료에 관한 (고지 및 조언의) 설명 대상: 비인공지능진료 및 인공지능진료의 현존 방법들, 각 방법의 위험과 효과 및 비용의 비교
 - 진료행위자의 인공지능진료 숙련도
 - 인공지능 자체의 신뢰도 등에 대한 설명도 포함

의료인공지능과 민형사상 책임의 귀속 4

IV. 의료인공지능 사고의 책임 주체 및 내용

1. 진료채무 이행을 인공지능을 이용하여 제공한 의사 측의 민사책임

(2) 의사측 불법행위책임

- ① 진료행위자라면 일반불법행위 책임(민법 제750조 이하) 규정이 적용됨
 - 과실의 존부 증명이 어렵다는 점
- ② 진료행위자의 사용자라면 사용자책임(민법 제756조)의 규정이 적용됨

의료인공지능과 민형사상 책임의 귀속 5

IV. 의료인공지능 사고의 책임 주체 및 내용

2. 인공지능 공급자의 민사책임

[인공지능 공급자란? 인공지능 소프트웨어나 하드웨어의 제조, 판매, 수입 등으로 이용자에게 인공지능을 공급한 사람]

(1) 피해자인 환자에 대한 제조물책임

- ① 공급한 인공지능이 현행 제조물책임법상 제조물에 포함되어야 함
 - 현행 제조물책임법상 제조물에 해당하지 않는 소프트웨어 공급자에게는 물을 수 없는 책임
- ② 공급자가 현행 제조물책임법상 제조자의 개념에 포함되어야 함
- ③ 결함의 세 유형 가운데 어느 하나의 결함에 관한 환자 쪽의 증명이 필요함
- ④ 결함과 피해 사이 (성립의) 인과관계도 환자 쪽이 증명해야 함

(2) 의사 측에 대한 계약책임

- ① 매매의 담보책임
- ② 수급인의 담보책임

의료인공지능과 민형사상 책임의 귀속 6

IV. 의료인공지능 사고의 책임 주체 및 내용

3. 의사와 인공지능 공급자의 형사책임

(1) 개별 책임이 아닌 공동책임으로, 그리고 행위 주체별(의료인, 의료기관, 제조업자/판매업자, 환자/소비자, 국가/규제기관)로 책임이 분산됨

(2) 설명의무의 구체화 요구

- ① 이전보다 설명 목록 및 내용이 더 구체화될 것임
- ② 인공지능의 자율성이 커질수록 이에 대한 의사의 설명의무도 더 중요해질 것으로 보임

의료인공지능과 민형사상 책임의 귀속 7

IV. 의료인공지능 사고의 책임 주체 및 내용

3. 의사와 인공지능 공급자의 형사책임

(3) 주의의무위반의 판단기준

① 주의의무 내용의 변화

- 데이터 중심의 맞춤의료 가능 - 주의의무의 주관화 가능성
- 신뢰원칙 인정의 범위 변화 - 의료인공지능과의 수직적 관계 혹은 수평적 관계 인정 여부에 따라 상이하게 변화
- 의사·공급자·의료기관의 작위에 의한 과실 이외에도 부작위에 의한 과실 증가 가능성
- 공급자의 예견가능성 판단기준의 구체화
- 의료 인공지능과의 협업으로 인간의 주의의무가 더 강화되는 측면이 있는 반면, 의료 인공지능의 자율성이 증가할수록 인간의 주의의무가 더 약화될 수 있는 부분도 공존함

의료인공지능과 민형사상 책임의 귀속 8

IV. 의료인공지능 사고의 책임 주체 및 내용

3. 의사와 인공지능 공급자의 형사책임

(3) 주의의무위반의 판단기준

② 데이터 관리 의무

- 주의의무로서 데이터 관리 의무의 중요성 부각
- 알고리즘 조작, 편향적 데이터 공급 등에 대한 과실범 성립 가능. 때에 따라 고의범도 성립 가능

③ 사전·사후 관리 및 감독

- 인공지능 시스템에 대한 관리 감독 의무, 데이터에 대한 검증 및 관리의무
- 보증인 지위에 기초하여 부작위에 의한 과실범 성립 가능. 때에 따라 고의범도 성립 가능

④ 과실 입증의 문제

- 인간이 사후에 밝혀낼 수 없는 오작동 원인에 대한 책임은 누구에게 귀속하는가?
- 이 경우, 모든 행위 주체를 상정하여 과실범의 공동정범 인정이 가능한가?

의료인공지능과 민형사상 책임의 귀속 9

V. 현행 법체계의 한계와 법정책적 개선방향

1. 자율주행 자동차와 같이 자율성의 단계별로 책임 귀속의 내용을 논하자는 제안들이 있음. 여기서는 별개로 논하지 않음
2. 현행법의 한계: 증명 곤란 등
 - (1) 인공지능 사고인지(적용대상)에 관한 증명이 곤란함
 - (2) 인공지능의 결함이 존재하는지에 관한 증명이 곤란함
 - (3) 그 결함으로 인한 사고인지(성립의 인과관계)에 관한 증명이 곤란함

의료인공지능과 민형사상 책임의 귀속 10

V. 현행 법체계의 한계와 법정책적 개선방향

3. 인공지능 사고로 인한 손해의 배상책임에 관한 특별법의 제정
 - 인공지능기본법 등 관련 특별법에 인공지능사고로 인한 손해배상책임을 규율하는 인공지능 고유의 특별법이 없음
 - 인공지능기본법을 개정하여 거기에 관련 특별조항을 둘 수도 있지만 규율을 간단하게 하면 문제를 해결하기엔 미흡한 개정입법이 될 가능성이 있음
 - 의료인공지능은 인공지능기본법상 '고영향 인공지능'임. 이와 관련된 조항들의 내용으로는 인공지능사고로 인한 손해배상책임을 제대로 규율하지 못함
 - 따라서 인공지능 사고로 인한 손해배상책임 고유의 특별법이 필요함. 여기에 담을 조항으로 인공지능 진료에 관련된 것 가운데 중요한 것만 서술

의료인공지능과 민형사상 책임의 귀속 11

V. 현행 법체계의 한계와 법정책적 개선방향

3. 인공지능 사고로 인한 손해의 배상책임에 관한 특별법의 제정

(1) 인공지능이 이용된 의료행위의 경우 인공지능사고로 일단 추정하는 조항을 둬.

단, 추정을 번복하고자 하는 쪽에서 인공지능과 무관한 사고임을 증명해서 특별법 적용 배제

(2) 인공지능 사고의 경우 인공지능 결함의 존재를 추정하는 조항을 둬.

단, 추정을 번복하고자 하는 쪽에서 결함의 부재를 증명해서 특별법 적용 배제

(3) 인공지능 프로그램이나 시스템의 결함 또는 피지컬 부분 결함을 모두 인공지능 결함으로 의제하는 규정을 둬

(4) 결함의 부재가 증명되지 않으면 결함과 손해 사이 인과관계, 즉 성립의 인과관계를 추정하는 조항을 둬. 단, 추정을 번복하고자 하는 쪽에서 결함이 없었더라도 발생하였을 손해임을 증명해서 특별법 적용 배제

의료인공지능과 민형사상 책임의 귀속 12

V. 현행 법체계의 한계와 법정책적 개선방향

4. 기타

(1) 행정규제 및 형사제재도 가능: 인공지능 진료자의 자격 및 지속적 교육, 데이터 편향성 및 신뢰성 관리, 위험관리 의무, 환자 진료 의사결정 지원에서 차별 방지, 의료인의 독립적 검토 및 해석 의무 명시, 알고리즘의 공정성 확보, 데이터 품질 유지에 관한 정기 점검 제도화

(2) 책임보험

① 인공지능의 이용을 의료현장에서 활성화할 필요가 있음

② 그 활성화의 부작용으로 증명이 곤란한 의료사고가 빈발할 우려가 있음

③ 인공지능진료를 하려면 인공지능사고 책임보험의 가입을 강제하는 조항을 둬

의료인공지능과 민형사상 책임의 귀속 13

V. 현행 법체계의 한계와 법정책적 개선방향

4. 기타

(3) 국가의 손해배상책임

- ① 인공지능의 산업을 육성하는 국가에게 인공지능 관련 법령의 각종 책임 조항이 존재함
- ② 이들 책임 조항을 선언적 조항이나 공공 일반의 이익을 위한 조항으로 해석될 수 있음
- ③ 국가의 손해배상책임을 강화하는 구체적 조항을 둬

- (4) 과실에 대한 형사처벌 가능성 검토: 입증할 수 없는 사유로 사고가 발생하고, 인공지능 결함인지 알 수 없고, 그 결함으로 인한 사고인지 알 수 없을 때, 의료행위에 참여한 각 행위 주체가 공동정범으로서 형사책임을 질 수 있는지, 아니면 민사상 책임 귀속에 그칠지에 대해서는 더 논의가 필요함

의료인공지능과 민형사상 책임의 귀속 14

경청해 주셔서 감사합니다!

피지컬 AI와 생명의료법

헌법적 책임과 거버넌스의 재구성

발표 : 건국대학교 엄주희 교수

2026. 6. 10.

국가생명윤리정책원 공동 학술대회, 인공지능과 거버넌스

I. 왜 지금 '피지컬 AI' 인가

디지털 데이터에서 물리적 행위로의 진화
인격체와의 직접적 결합과 물리 에너지의 직접 통제

피지컬 AI란 ?

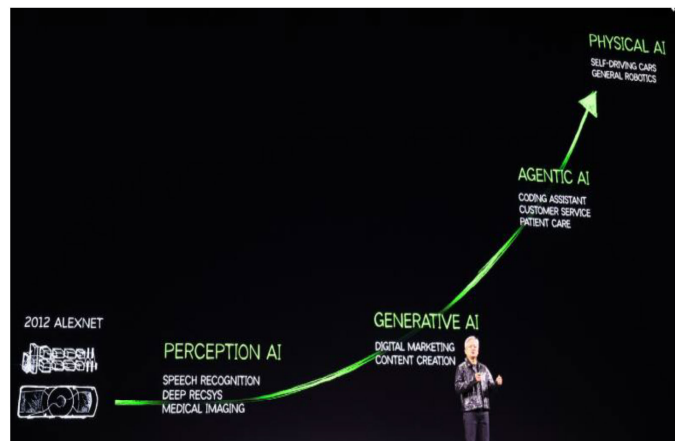
- 디지털 환경 속에서만 작동하던 인공지능 시스템과는 달리 현실 세계에서 물리적 행동을 수행하며 환경과 직접 상호작용하는 새로운 형태의 인공지능
- 로봇공학과 인공지능이 융합된 형태
- 학계에서 논의되어 오다가 2025년 CES 에서 엔비디아 (NVIDIA)의 CEO 젠슨 황이 차세대 AI 플랫폼 Cosmos 를 공개하고 ‘ 피지컬 physical AI 의 본격적인 시작’으로 선언하면서 대중과 산업계에 주목을 받게됨

엔비디아: 피지컬 AI는 세상을 관찰하는 센서와, 세상과 상호작용할 수 있는 액추에이터가 장착된 AI 시스템

< 출처 > nVIDIA(2025.1), Cosmos World Foundation Model Platform for Physical AI

네이처: 로봇의 몸(Body)과 두뇌(Brain), 이 두 가지의 긴밀한 결합을 통해 환경과의 상호작용을 하게 하는 AI

< 출처 > Nature(2020.11), Skills for Physical Artificial Intelligence



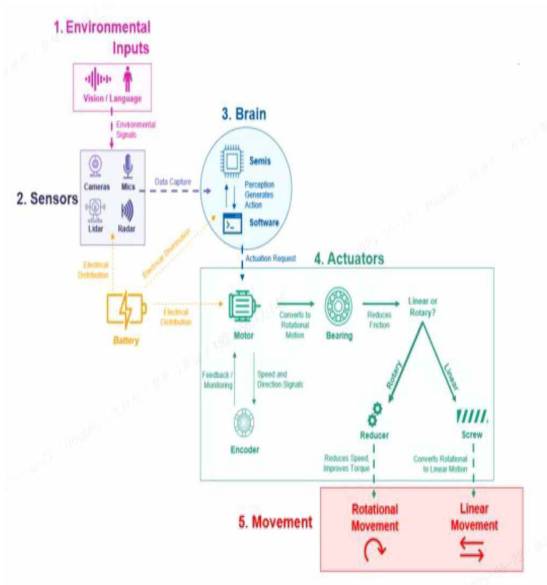
< CES 2025에서 기조 연설하는 엔비디아 CEO 젠슨 황 >

AI의 발전과정

지각 AI → 생성 AI → 능동형 AI → 피지컬 AI

Perception AI	Generative AI	Agentic AI	Physical AI
인식	생성	자율적 실행	물리적 행동
데이터해석	콘텐츠 생산	목적 지향적 행동	현실세계 개입

구분	정의	주요 특징	적용 분야
지각 AI (Perception AI)	• 외부환경(시각, 청각 등)으로부터 데이터 수집, 인식하는 AI	• 센서 데이터를 해석하여 의미 있는 정보로 전환 • 과거 및 실시간 데이터 기반 미래 상황 예측	• 자율주행차(객체 인식) • 헬스케어(의료 영상분석) • 로봇틱스(환경 인식)
생성형 AI (Generative AI)	• 기존 데이터를 기반으로 텍스트, 이미지, 코드 등을 생성하는 AI	• 대규모 데이터셋을 학습하여 새로운 콘텐츠 제작 • 텍스트 생성, 이미지 합성, 난제 해결 등 응용 가능	• 과학연구(연구개발 가속화) • 콘텐츠 제작(디자인, 광고) • 소프트웨어 개발(코드생성)
에이전트 AI, 능동형 AI (Agentic AI)	• 스스로 데이터를 수집·분석하고, 결정을 내리는 능동적인 AI	• 인간 개입 없이도 문제를 해결할 수 있는 자율성 • 지속적 학습 및 상호작용을 통해 성능 향상	• AI 연구 에이전트 • 공급망 관리(유통 최적화) • 업무자동화 및 보조
피지컬 AI (Physical AI)	• 로봇과 AI의 결합, 물리적 형태로 세상과 상호작용하는 AI	• 감지 + 계획 + 제어 • 센서, 액추에이터 등 학습 기반으로 실제 환경에서 작업 수행	• 스마트팩토리(생산공정 관리) • 물류 및 창고 자동화, 운송 • 재난 대응, 고위험 작업



< 피지컬 AI 구동 흐름도 >

구분	주요내용
①환경입력 (외부 자극 인식)	• 카메라, 마이크, 라이다(LiDAR), 레이더 등 센서 장치가 주변 환경의 정보를 수집해 신호를 디지털화, 뇌 역할을 수행하는 연산장치로 전송
②센서 (중앙처리)	• 센서에서 전달된 정보를 바탕으로, 온디바이스 또는 엣지 장치에 탑재된 AI 반도체 (데이터 처리) 및 소프트웨어 알고리즘이 인식과 판단을 수행
③브레인 (출력 명령 전송)	• AI의 판단 결과는 정확한 움직임을 지시하는 제어 신호로 변환되어 모터, 감속기, 토크 제어장치 등 구동기계(액추에이터)에 전달
④액추에이터 (물리적 동작 실행)	• 모터와 베어링, 감속기, 연크더 등으로 구성된 액추에이터는 회전 혹은 직선 운동을 생성하여 실제 움직임을 구현
⑤움직임 (최종 동작)	• AI의 인지와 판단을 바탕으로 생성된 최종 물리적 동작(회전, 직선 운동 등) • 사람과 유사한 동작이나 기능 수행을 가능하게 함

< 피지컬 AI 구동을 위한 5가지 단계 >

출처: Morgan Stanley(2025.2), The Humanoid 100: Mapping the Humanoid Robot Value Chain

의료 현장의 피지컬 AI

- ✓ **자율 수술 로봇:** 인간의 직접 제어를 벗어난 독립적 절개 및 봉합 수행
- ✓ **지능형 간병 로봇:** 환자와의 물리적 접촉을 통한 케어 수행
- ✓ **핵심 변화:** 정보적 판단 보조에서 물리적 힘의 직접 행사로 이동



지각(Perception)
- 실시간 환경 데이터 수집·처리 - 센서 기반 환경 인식(카메라, Lidar 등) - 공간 레이아웃, 객체 위치, 동작과 같은 실시간 데이터를 수집·처리하여 상황인식 제공
인지(Cognition)
- LLM 기반 의사결정, 기억, 추론, 학습 - 변화하는 상황에 따른 적응을 위한 지속적 학습을 통한 맥락 이해 - 실시간 상황 데이터를 처리하여 감각 입력을 실행 가능한 작업·계획으로 전환
작동(Actuation)
- 인지 기능의 결정을 물리적 행동으로 변환 - 로봇 팔, 모터, 액추에이터, 유압 시스템 및 이동 시스템 등으로 행동

< 피지컬 AI의 핵심 구성요소 >

	헬스케어 로봇
지각	실시간 내시경, 촉각센서
인지	의료지식 기반 수술 부위 인식
작동	수술(보조), 피드백 제공

< 헬스케어 로봇의 적용사례 >

의료 돌봄 영역에서 피지컬 AI 적용 사례

◇ 고령화 사회에서 간병인력 부족 문제를 해소하기 위해, 고령자 돌봄 로봇, 환자 이송 로봇, 정서 교감형 로봇 등이 피지컬 AI 기반으로 도입
- 음성 인식, 상황 판단, 원격 모니터링 기능 등을 포함

(일본 RIKEN, Robear) 노인을 들어 올리거나 침대에서 휠체어로 옮기는 작업 지원. 로봇 간호사 모델로, 몸 모양 140kg 외형, 최대 80kg까지 들 수 있음



(이스라엘 Intuition Robotics, ElliQ) 고령자 대상 동반형 로봇. 대화를 통해 정서적 지지 제공, 약 복용·운동·식사 시간 알림 기능 내장. 실시간 가족·케어팀 연결



(이스라엘 Temi Global, Temi) 병원이나 재택 돌봄 현장에서 이동하며 환자와 대화하고, 원격 진료 연결, 낙상 등 위험 실시간 감지, 간단한 심부름 가능



헌법상 인간의 '행위' 개념의 위기

신체를 조정하는 피지컬 AI 와 인간의 자율성

- ✔ **신체적 온전성 문제** : 인공 외골격이나 간병 기기가 인간의 사지를 제어할 때 신체의 온전성이라는 신체의 자유의 내용에 문제
- ✔ **자율성 문제** : 인공지능 로봇에 의해 제어되는 신체는 자율성을 누린다고 할 수 있는지 문제

전통적 주체적 행위론의 붕괴

- ✔ **피지컬 AI 블랙박스로 인한 인과관계 단절** : AI의 블랙박스적 결과와 인간 설계자의 의사로 환원 불가
 - 인간이라면 과실과 그로인한 결과 사이에 인과관계가 존재하고, 그로인해 책임 귀속
 - 초연결·초지능 AI의 실시간 물리 행위는 인간의 통제 범위를 초과
- ✔ **행위자성이 모호함으로 인한 책임 귀속 문제**: 피지컬 AI는 법적 행위자가 아니므로, 행위는 존재하나 책임의 주체는 불분명한 '책임의 공백' 발생
 - 예) 인공지능 기체가 수술 부위 변화 등을 추론하여 외력을 인가할 때 인간 의사에 의한 행위나 제조사의 설계로 귀속시키는 것은 곤란함

헌법적 기초와 기본권적 가치



인간의 존엄성 (제10조)

의료 결정권의 주체로서 환자의 자기결정권 보장.
AI가 인간을 객체화하는 '기술적 온정주의' 경계.



신체의 자유 (제12조)

물리적 위해로부터 안전할 권리.
AI의 자율적 동작이 환자의 신체적 온전성을 침해할 가능성 차단.

II. 피지컬 AI 법제의 현주소와 입법 과제

유럽 AI 법(EU AI Act)과 국제 인권 원칙이라는 글로벌 스탠다드 적용 필요

EU AI Act : 고위험 의료 AI 규제



의료기기에 내장된 피지컬 AI를 '고위험(High-Risk) AI'로 자동 분류하여 엄격한 적합성 평가 실시.

- ✓ 인간 감독(Human Oversight) 의무화 (제14조) → 감독자는 AI의 결정을 맹목적으로 따르지 않고 거부하거나 정지 버튼으로 시스템 중단 능력 갖춰야 함
- ✓ 리스크 관리 시스템 및 데이터 거버넌스 (제9조, 제10조)
- ✓ 사후 시장 감시 및 사고 보고 체계 (제72조, 제73조)

구분	관련 조항 (Article)	핵심 내용
리스크 관리	제9조	수명 주기 전반에 걸친 반복적 위험 식별 및 완화
데이터 거버넌스	제10조	편향성 방지를 위한 고품질 학습 데이터셋 사용
인간 감독	제14조	실제 시 인간-기계 인터페이스 (HMI) 반영 및 개입권 보장
사후 감시	제72조	시장 출시 후 지속적 성능 모니터링 및 데이터 분석
사고 보고	제73조	심각한 사고 발생 시 감독 당국에 즉시 보고 의무

피지컬 AI 법안 발의

제안일자

2026-06-05

심사진행단계 ※ 접수는 제안일자, 심사단계는 처리일자, 본회의는 의결일자 기준임



접수정보

· 의안접수정보

의안번호	제안일자	제안자	제안회기	문서
2219104	2026-06-05	황정아의원 등 10인 제안자 목록	제22대(2024~2028)제436회	의안원문

젠스 황 방한 맞춰 '피지컬 AI 특별법' 발의...실증특구·규제샌드박스 대 폭 확대

황정아 의원 대표발의...60일 내 미회신 시 자동 승인-데이터 활용 특례 도입, 국무총리 산하 촉진위원회 신설

한국의 입법 딜레마: 산업 진흥 vs 생명과 안전

구분	피지컬 AI 진흥 특별법(안)	기존 생명의료법제
기본 철학	우선허용·사후규제 (산업 진흥)	안전 우선·사전 승인 (생명 보호)
입법 핵심 목표	K-로봇·자율주행 글로벌 최강국 도약 및 신시장 육성	인간 생명의 존엄과 신체적 온전성 및 환자 안전 보호
승인 방식	자동 승인 타임아웃제 (60일) - 피지컬 AI 신속 실증을 위한 규제특례 및 간주제도 (제18조)	기관윤리위(IRB) 및 식약처 엄격 심사
책임 귀속	제조사 및 운영사 성능 인증 제도 및 배상 보험으로 사후 보전 - 손해배상 보험 가입 의무 (제23조)	의사의 설명의무 및 직접 진료에 따른 과실책임 법리
개인정보 활용도	익명 처리된 고품질 비정형(영상·사운드) 데이터 학습 보장	생명윤리법상 고도로 제약되는 민감 건강 정보 통제
헌법적 쟁점	산업 진흥과 규제 사이에서의 법익 균형 과잉금지원칙 (경제적 자유)	국가의 기본권 보호의무 강조 과소보호금지원칙 (생명 안전)

III. 헌법적 책임과 과소보호금지

국가의 기본권 보호의무와 거버넌스 설계

과소보호금지원칙과 국가의 의무

- 국가가 AI의 위험성을 인지하면서도 충분한 보호조치를 취하지 않는 경우, 이는 헌법상 기본권 보호의무 위반(과소보호금지원칙 위배)에 해당할 가능성 존재
- 피지컬 AI의 위험성은 창발적이고 가변적 위험으로 스스로 학습하여 진화하고 자율성을 가진 존재가 발생시키는 위험
- 의사의 판단인지, AI 알고리즘의 오류인지, 자율학습의 결과인지 규명하기 어렵기 때문에, 책임 주체가 불분명한 영역이라고 할 수 있음 → 국가가 제3의 첨단 기술 위험으로부터 국민을 보호하기 위한 적극적인 위험관리 거버넌스를 구축해야 할 의무가 존재
- 국가의 기본권 보호의무의 수준은 고정된 기준이 아니라 과학기술의 발전에 맞추어 진화되어야 하고 최신의 과학적 인식과 기술적 수준이 요구하는 예방 조치를 상시적으로 업데이트해야 하는 동적 의무임 :
- 독일 연방헌법재판소의 원자력 발전소 건설의 위헌성을 판단했던 Karlar 판결(1928)의 법리를 현대적 적용이 필요 --> 동적 기본권 보호 의무 적용과 피지컬 AI 거버넌스 책무 (위험 최소화 거버넌스로서 실시간 모니터링, 알고리즘 감사제 등의 사전대비 의무 필요)

헌법적 가치 수호를 위한 거버넌스 3대 원칙



인간 존엄 기반의 '의무적 개입권'

의료 피지컬 AI 시스템은 반드시 수술 과정 등 실시간 작동 중에 인간(의료진 또는 당사자)이 단 한번의 조작으로 하드웨어 수준에서 완전 제어할 수 있는 인간 통제성을 강제 설계해야 함.



알고리즘 입증책임 전환

알고리즘 연산과 하드웨어 제어 논리의 블랙박스적 속성상 발생한 의료 사고에 대해, 기술적 지식이 전무한 환자에게 과실 입증을 요구하는 사후 보상 체계를 지양하고, 제조사 측에 기술 증명 책무를 전환할 필요



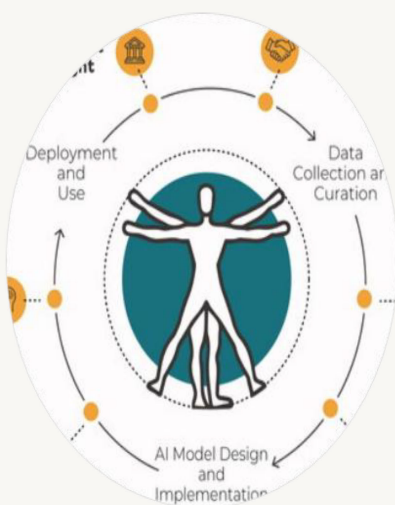
분배 정의와 평등권 보장

고가의 혁신적 보철 및 신체 개선 장비가 단순한 자본력에 의해 불평등하게 돌아가 계급 분화를 초래하지 않도록, 국가의 적극적 기본권 보호 조치 차원에서 공적 건강 보험 급여 범위 내에 편입을 적극 추진할 필요.

책임의 재구성: 구조적 책임 귀속

- ❖ **행위자 중심에서 시스템 중심으로:** 개별 의사의 과실을 넘어 시스템 설계자, 데이터 공급자, 운영사의 복합적 연대 책임 구체화
- ❖ **규범적 책임 귀속:** 물리적 통제 가능성이 없더라도, 위험을 창출하고 그로부터 이익을 얻는 주체에게 책임을 귀속시키는 법리
- ❖ **입증책임의 전환:** 정보 비대칭성을 고려하여, 피해자가 아닌 AI 운영 주체가 무과실을 입증하도록 하는 체계
- ❖ **무과실 보상 제도:** 기술의 본질적 불확실성으로 인한 사고에 대해 국가적 차원의 사회 보험 모델 검토

피지컬 AI 거버넌스의 최종 목표: 인간의 의미 있는 통제



AI는 도구로서의 지위를 유지해야 하며, 생사여부의 최종 판단 권은 반드시 인간에게 유보되어야 함.

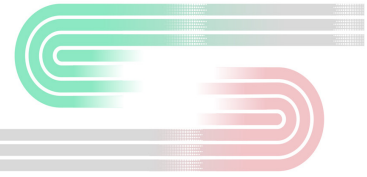
- ❖ **Constitutional by Design:** 피지컬 기술 설계 단계부터 헌법적 가치 내재
- ❖ **설명가능성:** AI의 물리 행위에 대한 투명한 설명력 확보
- ❖ **책임:** 기술적 실패에 대한 명확한 책임 소재 명시 및 시스템적 책임 부담과 국가의 무과실 보상 제도 등 도입

피지컬 AI 시대에 필요한 헌법 가치, 인간의 존엄성



감사합니다.

엄주희 교수 juheelight@gmail.com



의료 AI 거버넌스의 방향과 과제; 통제를 넘어 공공적 정렬로

문재영 교수(충남대학교 의과대학, 중환자의학)

오늘 세 분의 발표는 서로 다른 분야와 관점에서 이루어졌지만, 공통적으로 의료 인공지능의 문제를 기술 자체의 문제가 아니라 사회적 의사결정과 거버넌스의 문제로 다루고 있었다는 점에서 매우 인상적이었다. 저는 세 발표를 가치(Value), 제도(Institution), 실행(Implementation)이라는 세 층위의 삼각구조로 이해하고자 한다.

첫째, 김준혁 교수의 발표는 의료 인공지능 거버넌스의 가장 근본적인 층위인 ‘가치’의 문제를 제기하였다. 발표자는 공공의료 AI를 단순한 기술혁신의 수단이 아니라 공공의료의 사명을 실현하기 위한 도구로 이해해야 한다고 주장하였다. 특히 “이 AI가 안전하고 효율적인가?”라는 질문보다 “이 AI가 공공의료를 공공의료답게 만드는가?”라는 질문이 우선되어야 한다는 점을 강조하였다. 평균적인 정확도나 성능만으로는 공공성을 담보할 수 없으며, 의료 취약지와 취약계층에서 실제로 유효하게 작동하는지 검증해야 한다는 주장이다. 또한 국가는 단순한 규제자가 아니라 의료 제공자이자 수탁자로서 이중의 책무를 지니며, AI가 의료인을 대체하는 존재가 아니라 돌봄 관계를 지원하는 존재가 되어야 한다고 지적하였다.

둘째, 이경도 교수의 발표는 ‘제도’의 층위에서 의료 AI 거버넌스를 조명하였다. 발표는 응급실 뱅뱅이 문제를 사례로 삼아, AI가 병상 예측이나 자원 배분을 지원할 수 있는 가능성을 보여주었다. 그러나 동시에 응급의료체계의 문제는 단순한 정보 부족이 아니라 인력, 자원, 이해관계, 정책결정 구조의 문제임을 강조하였다. 알고리즘 편향, 가치 정렬(value alignment), 사회적 신뢰의 문제 역시 결국 정치적·사회적 문제이며, 기술적 해결책만으로는 충분하지 않다는 점을 보여주었다. 따라서 AI 거버넌스는 단순히 새로운 기술을 도입하는 것이 아니라 정책 결정 과정의 정당성과 사회적 수용성을 확보하는 문제라는 점을 상기시켜 주었다.

셋째, 박소연 교수의 발표는 ‘실행’의 층위에서 의료 AI의 현실적 과제를 제시하였다. 예방, 진단, 치료, 추적관리 등 임상 전 영역에서 AI 활용이 확대되고 있으며, 특히 생성형 AI와 AI Agent의 등장으로 의료현장은 새로운 전환기를 맞고 있다. 발표자는 현재 국내 제도가 부처별로 분절되어 있고, 설명가능성, 책임성, 편향 검증 체계가 충분히 제도화되어 있지 않음을 지적하였다. 이에 대해 정부-학계-산업계의 다층적 협력, 위험도 기반 관리체계, 적응형 거버넌스의 필요성을 제안하였다.

저는 특히 이 발표에서 한 가지 문제를 추가적으로 제기하고 싶다. 현재 우리나라 의료 AI 규제체계는 기술의 위험도와 활용 맥락에 따른 차등적 접근보다는 사전규제 중심의 획일적 관리에 머물러 있는 측면이 있다. 그 결과 연구자들은 실제 임상환경에서 AI를 검증하고 개선할 기회를 충분히 확보하지 못하고 있으며, 의료기관은 혁신적 기술을 안전하게 실증할 제도적 공간을 제공받지 못하고 있다. 산업계 역시 규제 불확실성으로 인해 투자와 개발에 어려움을 겪는다. 결국 사회가 기대하는 의료적·공공적 이익은 충분히 실현되지 못한 채, 기술혁신과 공공성 모두가 제약받는 상황이 발생하고 있다.

거버넌스의 역할은 단순히 위험을 통제하는 것이 아니라 위험을 관리하면서 사회적 편익을 실현하는 데 있다. 그러나 현재의 거버넌스는 위험 통제에는 상대적으로 집중하면서도 혁신의 사회적 활용을 촉진하는 기능은 충분히 수행하지 못하고 있는 것은 아닌지 성찰이 필요하다.

세 발표를 종합하며 저는 다음 세 가지 질문을 남기고 싶다.

첫째, 의료 AI 거버넌스의 최종 책임 주체는 누구인가? 개발자, 의료기관, 의료인, 정부가 모두 관여하는 상황에서 실제 피해가 발생했을 때 누가 최종적으로 설명하고 책임져야 하는가?

둘째, 효율성과 형평성이 충돌할 경우 무엇을 우선해야 하는가? 전체 인구 수준에서는 높은 효율성을 보이지만 취약계층에서는 낮은 성능을 보이는 AI를 우리는 허용할 수 있는가?

셋째, 우리 사회는 어떤 형태의 의료 AI 거버넌스를 지향해야 하는가? 현재와 같은 규제 중심 모델에 머물 것인가, 아니면 국가가 데이터·검증·실증 인프라를 제공하고 사회적 신뢰를 구축하는 공공 인프라 중심 모델로 전환할 것인가?

결론적으로 오늘 세 분의 발표는 의료 AI 거버넌스가 단순한 기술 관리의 문제가 아니라 가치, 제도, 실행이 결합된 사회적 통치의 문제임을 보여주었다. 김준혁 교수는 공공성과 돌봄이라는 가치의 총위에서, 이경도 교수는 정책 정당성과 사회적 수용성이라는 제도의 총위에서, 박소연 교수는 임상현장의 안전성과 적응적 운영이라는 실행의 총위에서 각각 중요한 통찰을 제공하였다.

현재의 거버넌스는 위험 통제에는 비교적 많은 관심을 기울이고 있지만, 사회적으로 바람직한 혁신을 촉진하는 역할은 충분히 수행하지 못하고 있다. 그 결과 연구자는 검증할 기회를 얻지 못하고, 의료기관은 실증할 공간을 확보하지 못하며, 기업은 불확실성 속에서 투자와 개발을 주저하게 된다. 사회는 혁신의 편익을 충분히 누리지 못한 채 위험에 대한 우려만 반복하는 상황에 머물고 있다.

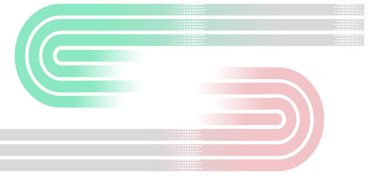
요컨대 한국의 의료 AI 거버넌스는 현재 '규제는 있으나 방향성은 부족하고, 통제는 있으나 공공적 목표는 불분명한 상태'에 놓여 있다고 볼 수 있다. 앞으로의 과제는 더 많은 규제를 만드는 것이 아니라, 의료 AI가 어떤 사회적 가치를 실현해야 하는지에 대한 공적 합의를 형성하고, 그 목표를 향해 혁신과 안전을 함께 관리할 수 있는 거버넌스를 구축하는 것이다.

의료 AI의 미래는 알고리즘의 성능만으로 결정되지 않는다. 그것은 결국 우리가 어떤 사회를 지향하며, 누구를

우선적으로 보호하고, 누구를 위해 기술을 사용할 것인가에 대한 집단적 선택의 문제이다. 의료 AI 거버넌스의 궁극적인 과제는 기술을 통제하는 것이 아니라 기술을 공공의 목적에 정렬시키는 데 있다고 생각한다.

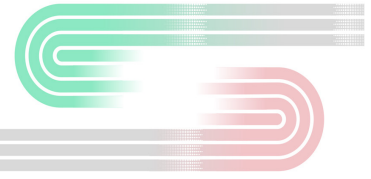
*발표문을 요약하고 공통점과 차이점 등을 정리하는데 OpenAI(ver 5.3)을 사용하였음을 밝힙니다.

패널토론



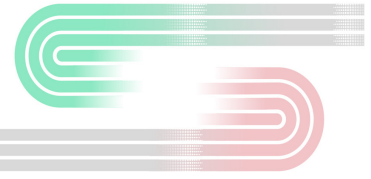
박지용 교수 (연세대학교 법학전문대학원)

패널토론



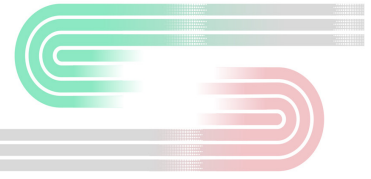
이일호 교수 (연세대학교)

패널토론



전재필 부장 (국립보건연구원)

패널토론



김민우 교수 (충북대학교)



국립생명윤리심의위원회
National Bioethics Committee



국가생명윤리정책원
Korea National Institute for Bioethics Policy



사단법인 대한의료법학회
The Korean Society of Law and Medicine



유네스코한국위원회
unesco



이화여자대학교
생명의료법연구소
Ewha Institute for Biomedical Law & Ethics



한국생명윤리학회
Korean Bioethics Association



한국윤리학회
Korean Society of Ethics



사단법인 한국의료법학회
Korean Association of Medical Law



한국의료윤리학회
THE KOREAN SOCIETY FOR MEDICAL ETHICS