KoNIBP
재단법인 국가생명윤리정책원
KOREA NATIONAL INSTITUTE FOR BIOETHICS POLICY

보건의료분야의 인공지능 윤리와 거버넌스

대규모 멀티모달 모델에 대한 가이던스

보건의료분야의 인공지능 윤리와 거버넌스

대규모 멀티모달 모델에 대한 가이던스

목차

감사의 글

세계보건기구(WHO) 가이던스 개발은 Andreas Reis(보건연구 부서의 보건윤리 및 거버넌스 팀 공동책임자)와 Sameer Pujari(디지털 보건 및 혁신부서)의 주도로 시행되었으며, John Reeder(보건연구 국장), Alain Labrique(디지털 보건 및 혁신국장), Jeremy Farrar(수석 과학자)의 전반적인 지도하에 진행되었습니다.

Rohit Malpani(프랑스, 컨설턴트)가 주요 집필을 담당했으며 WHO 보건분야 인공지능 윤리 및 거버넌스 전문가 그룹 공동 의장인 Effy Vayena(스위스 ETH 취리히)와 Partha Majumder(인도 통계연구소 및 국립 생물의학 유전체학 연구소)가 보고서 초안 작성과 전문가 그룹 운영에 전반적인 가이던스를 제공했습니다.

WHO는 이 가이던스 개발에 기여한 분들에게 감사를 표합니다.

WHO 보건분야 인공지능 윤리 및 거버넌스 전문가 그룹

Najeeb Al Shorbaji(e헬스부문 개발협회, 요르단 암만), Maria Paz Canales(글로벌 파트너 디지털, 칠레 산티아고), Arisa Ema(도쿄대학교, 일본 도쿄), Amel Ghouila(빌 & 멜린다 게이츠 재단, 미국 워싱턴 시애틀), Jennifer Gibson(WHO 생명윤리 협력센터, 토론토대학교, 캐나다 토론토), Kenneth Goodman(생명윤리 및 보건정책 연구소, 마이애미 밀러 의과대학, 미국 플로리다 마이애미), Malavika Jayaram(디지털 아시아 허브, 싱가포르), Daudi Jjingo(Makerere 대학교, 우간다 캄팔라), Tze Yun Leong(싱가포르 국립대학교, 싱가포르), Alex John London(카네기멜런대학교, 미국 펜실베이니아 피츠버그), Partha Majumder(인도 통계연구소 및 국립 생물의학 유전체학 연구소, 인도 콜카타), Thilidzi Marwala(요하네스버그대학교, 남아프리카공화국 요하네스버그), Roli Mathur(인도 의학연구위원회, 인도 방갈로르), Timo Minssen(생의학 혁신법 고급 연구 센터, 코펜하겐 대학교 법학부, 덴마크 코펜하겐), Andrew Morris(영국 헬스 데이터 연구소, 영국 런던), Daniela Paolotti(ISI 재단, 이탈리아 토리노), Jerome Singh(KwaZulu-Natal 대학교, 남아프리카공화국 더반), Jeroen van den Hoven(델프트 공과대학교, 네덜란드 델프트), Effy Vayena(스위스 ETH 취리히, 스위스 취리히), Robyn Whittaker(오클랜드 대학교, 뉴질랜드 오클랜드), YiZeng(중국과학원, 중국 베이징)

참관인

David Gruson, Luminess, 프랑스, 파리 ; Lee Hibbard, 유럽 평의회, 프랑스, 스트라스부르

외부 검토자

Oren Asman(텔아비브 대학교, 이스라엘 텔아비브), I. Glenn Cohen(하버드 로스쿨, 미국 매사추세츠 보스턴), Alexandrine Pirlot de Corbion(Privacy International, 영국 런던), Rodrigo Lins(페르남부쿠 연방대학교, 브라질 헤시피), Doug McNair(통합 개발 부국장, 빌 & 멜린다 게이츠 재단, 미국 워싱턴 시애틀), Keymanthri Moodley(스텔렌보스 대학교, 남아프리카공화국 케이프타운), Amir Tal(텔아비브 대학교, 이스라엘 텔아비브), Tom West(Privacy International, 영국 런던)

외부 기여자

가이던스의 Box 2(아동의 LMM 사용에 대한 윤리적 고려사항)는 Vijaytha Muralidharan, Alyssa Burgart, Roxana Daneshjou, Sherri Rose(미국 캘리포니아주 스탠퍼드 대학교)가 작성했습니다. 가이던스의 Box 3(LMM와 관련된 윤리적 고려사항 및 장애인에게 미치는 영향)는 Yonah Welker(스위스 제네바 소재 독립 컨설턴트)가 작성했습니다.

모든 외부 검토자, 전문가 및 기여자들은 WHO 정책에 따라 이해관계를 공개했으며, 공개된 이해관계 중 중요한 사항은 없는 것으로 평가되었습니다.

WHO

Shada Al-Salamah(기술 책임관, 디지털 보건 및 혁신 부서, 제네바), Mariam Otmani Del Barrio(과학자, 열대 질병 연구 특별 프로그램, 제네바), Marcelo D'Agostino(정보 시스템 및 디지털 보건 부서장, WHO 미주 지역 사무소, 미국 워싱턴, DC), Jeremy Farrar(수석 과학자, 제네바), Clayton Hamilton(기술 책임관, WHO 유럽 지역 사무소, 덴마크 코펜하겐), Kanika Kalra(컨설턴트, 디지털 보건 및 혁신 부서, 제네바), Ahmed Mohamed Amin Mandil(연구 및 혁신 코디네이터, WHO 동지중해 지역 사무소, 이집트 카이로), Issa T. Matta(법무 담당, 제네바), Jose Eduardo Diaz Mendoza(컨설턴트, 디지털 보건 및 혁신 부서, 제네바), Mohammed Hassan Nour(기술 책임관, 디지털 보건 및 혁신 부서, WHO 동지중해 지역 사무소, 이집트 카이로), Denise Schalet(기술 책임관, 디지털 보건 및 혁신 부서, 제네바), Yu Zhao(기술 책임관, 디지털 보건 및 혁신 부서, 제네바)

약어

AI	artificial intelligence 인공지능
LMM	large multi-modal model 대규모 멀티모달 모델
USA	United States of America 미합중국

요약

인공지능(AI, Artificial Intelligence)은 시스템과 도구에 통합된 알고리즘이 데이터를 학습하여 인간이 매 단계 명시적으로 프로그래밍하지 않아도 자동화된 작업을 수행할 수 있는 능력을 의미합니다. 생성형 AI는 텍스트, 이미지 또는 비디오와 같은 새로운 콘텐츠를 생성하는데 사용될 수 있는 데이터 세트에서 알고리즘이 학습되는 AI 기술의 한 범주입니다. 본 가이던스에서는 생성형 AI의 한 유형인 대규모 멀티모달 모델(LMM)에 대해 다룹니다. LMM은 하나 이상의 유형의 데이터 입력을 받아들이고 알고리즘에 입력된 데이터 유형에 국한되지 않는 다양한 출력을 생성할 수 있습니다. LMM은 보건의료, 과학 연구, 공중보건 및 신약 개발에 널리 사용되고 응용될 것으로 예측되고 있습니다. LMM은 "범용 기반 모델"로도 알려져 있지만, LMM이 광범위한 작업과 목적을 성공적으로 수행할 수 있는지는 아직 입증되지 않았습니다.

LMM은 역사상 어떤 소비자 응용 프로그램보다 빠르게 채택되었습니다. 이는 인간의 의사소통을 모방하고 인간과 유사하며 권위있어 보일 수 있는 쿼리나 데이터 입력에 대한 응답을 생성하기 위해 인간-컴퓨터 상호작용을 촉진하기 때문에 매력적입니다. 소비자들 사이에서의 빠른 채택과 수용, 그리고 주요 사회 서비스와 경제 부문을 혁신할 가능성으로 인해 많은 대규모 기술 기업, 스타트업, 그리고 정부가 생성형 AI 개발을 주도하기 위해 투자하고, 경쟁을 벌이고 있습니다.

2021년, WHO는 보건 분야에서 AI의 윤리와 거버넌스에 관한 포괄적인 가이던스*(1)*를 발표했습니다. WHO는 AI 분야의 20명의 선도적인 전문가들과 협력하여 보건의료 분야에서 AI 사용의 잠재적 이점과 위험을 모두 파악하고, 정부, 개발자, 제공자를 포함한 다양한 이해관계자들이 정책과 관행에서 고려해야 할 여섯 가지 원칙을 합의에 의해 도출했습니다. 이 원칙들은 정부, 공공 부문 기관, 연구자, 기업, 실행자 등을 포함한 폭넓은 이해관계자들이 보건의료 분야에서 AI를 개발하고 배포하는 데 있어 가이던스 역할을 해야 합니다. 그 원칙으로는 (1) 자율성 보호, (2) 인간의 복지, 안전 및 공익 증진, (3) 투명성, "설명 가능성" 및 이해 가능성 보장, (4) 책임감과 책무성 강화, (5) 포용성과 형평성 보장, (6) 반응성이 높고 지속 가능한 AI 촉진이 있습니다(그림 1 참고).

WHO는 회원국들이 보건 분야에서 LMM 사용과 관련된 이점과 과제를 파악하고 적절한 개발, 제공 및 사용을 위한 정책과 관행을 수립할 수 있도록 돕기 위해 이 가이던스를 발행합니다. 이 가이던스에는 원칙에 부합하는 기업 내 거버넌스, 정부의 역할, 국제 협력을 통한 거버넌스를 위한 권고사항이 포함되어 있습니다. 인간이 건강을 위해 생성형 AI를 사용하는 독특한 방식을 설명하는 원칙과 권고사항은 본 가이던스의 기초를 이루고 있습니다.

그림 1: WHO 보건분야에서 AI사용을 위해 합의된 윤리원칙

대규모 멀티모달 모델(LMM)의 응용, 과제 및 위험

LMM의 보건의료 분야에서의 잠재적 응용은 기존의 다른 형태의 AI와 유사하지만, LMM의 접근방식과 사용방식은 새로우며, 사회, 보건시스템 및 최종 사용자가 아직 완전히 대처할 준비가 되어 있지 않은 새로운 이점과 위험을 포함합니다. 표 1은 LMM의 주요 응용, 잠재적 이점 및 위험을 요약한 것입니다.

LMM 사용과 관련된 시스템적 위험에는 보건시스템에 영향을 미칠 수 있는 위험이 포함됩니다(표 2 참고).

LMM 사용은 더 광범위한 규제 및 시스템적 위험을 초래할 수 있습니다. 한 가지 우려사항은(여러 데이터 보호 당국에서 조사 중) LMM이 국제 인권 의무를 포함한 기존의 법적 또는 규제 체제와 국가 데이터 보호 규정을 준수하는지 여부입니다. 알고리즘은 LMM을 훈련시키기 위해 데이터를 수집하기 위한 방식, 수집된 데이터(또는 최종 사용자가 LMM에 입력한 데이터)의 관리 및 처리, LMM이 "환각(hallucination, 할루시네이션)"을 일으킬 가능성이 있기 때문에 이러한 법률을 준수하지 않을 수 있습니다. 또한, LMM은 소비자 보호법을 준수하지 않을 수도 있습니다.

LMM의 사용 증가와 관련된 더 광범위한 사회적 위험은 보건 분야뿐만 아니라 그 외의 영역에서도 나타납니다. LMM은 주로 대규모 기술 기업에 의해 개발 및 배포되는데, 이는 LMM 개발에 필요한 막대한 컴퓨팅, 데이터, 인적 및 재정적 자원 때문입니다. 이러한 상황은 AI 연구와 개발 및 사용에 있어 공공 및 민간 부문에서 AI 연구의 초점을 포함해 소규모 기업과 정부에 비해 이러한 대규모 기업의 우위를 강화할 수 있습니다. 대규모 기술 기업의 잠재적 지배력에 대한 추가적인 우려에는 윤리와 투명성에 대한 기업의 부족한 헌신이 포함됩니다. 이러한 기업들 간의 새로운 자발적 약속이나 정부와의 협력은 단기적으로

표 1. 보건의료 분야에서 LMM의 다양한 사용의 잠재적인 이점과 위험성

사용분야	잠재적 또는 제안된 이점	잠재적인 위험성
진단 및 임상진료	복잡한 사례 관리 및 일상적인 진단검토 지원 의료 제공자의 의사소통 업무 감소 ("키보드 해방") 다양한 비정형 보건 데이터에서 새로운 통찰과 보고서 제공	부정확하거나 불완전한 응답 또는 잘못된 응답 낮은 품질의 학습 데이터 편향(학습 데이터 및 응답의 편향) 자동화 편향 보건의료 전문가의 기술 저하 환자의 사전동의 문제
환자중심 활용	의료 상태(환자 또는 간병인으로서)에 대한 이해를 개선하기 위한 정보 생성 가상 건강 비서 임상시험 등록	부정확하거나 잘못된 진술 조작 위험 개인정보 보호 문제 임상의와 환자 간의 상호작용 감소 인식적 부정의 보건시스템 외부에서의 의료 제공 위험
사무 및 행정업무	임상진료에 필요한 문서 작업 및 행정 업무 지원 언어 번역 지원 전자 건강 기록 작성 완료 환자 방문 후 임상 기록 초안 작성	부정확성 및 오류 프롬프트(Prompt)에 따라 일관성 없는 응답
의학 및 간호교육	각 학생의 필요에 맞춘 동적 텍스트 제공 다양한 상황 및 다양한 환자와의 의사소통 및 실습을 개선하기 위한 시뮬레이션 대화 연쇄적 사고(chain-of-thought) 추론을 동반한 질문 응답	자동화 편향 기여 오류 또는 허위 정보가 의학 교육의 질을 저하시킬 위험 디지털 기술 학습에 대한 새로운 부담
과학 연구 및 신약 개발	과학 데이터 및 연구에서 통찰 생성 과학 논문, 원고 제출 또는 동료 검토에 사용될 텍스트 생성 연구 데이터를 분석 및 요약 교정 작업 지원 새로운 약물 설계(De novo drug design)	알고리즘에 콘텐츠에 대한 책임을 물을 수 없음 알고리즘이 고소득 국가의 관점을 편향적으로 반영할 가능성 존재하지 않는 정보 또는 참고문헌 생성 동료 검토와 같은 과학 연구의 주요 원칙을 약화시킴 과학적 지식 접근의 차별성을 악화시킴

표 2. 보건의료 분야에서 LMM의 사용과 관련된 보건시스템의 위험성

위험 유형	설명
LMM의 이점을 과대평가	LMM 사용의 안전성, 유효성, 유용성을 포함한 사용상의 문제를 무시하거나 경시하면서 "기술적 해결주의"에 빠지거나 LMM의 이점을 과대평가할 가능성이 있음.
접근성 및 경제성	"디지털 격차"와 LMM 접근을 위한 구독료와 같은 이유로 인해 LMM에 대한 공평한 접근이 부족할 수 있음.
시스템 전반의 편향	점점 더 큰 데이터 세트를 사용함으로써 LMM에 내재된 편향이 증가할 수 있으며, 이는 보건의료시스템 전반에 걸쳐 자동화될 가능성이 있음.
노동에 미치는 영향	LMM 사용이 일부 국가에서 일자리 감소로 이어질 수 있으며, 의료 종사자가 재교육 및 LMM에 적응해야 하는 상황을 초래할 수 있음. 데이터 주석 작업과 필터링은 저임금 노동 및 심리적 스트레스를 유발할 가능성이 있음.
부적합한 LMM에 대한 보건시스템의 의존도	LMM이 유지되지 않거나(저소득 및 중소득 국가에서) 고소득 국가 전용으로만 업데이트될 경우, 보건시스템이 취약해질 수 있음. 또한, 개인정보 및 기밀성 보호 부족은 보건시스템에 대한 신뢰를 약화시킬 수 있음
사이버 보안 위험	악성 공격 또는 해킹으로 인해 보건의료 분야에서 LMM 사용의 안전성과 신뢰성이 저하될 가능성이 있음

몇 가지 위험을 완화할 수 있지만, 결국 시행될 수 있는 정부의 감독을 대체할 수는 없습니다.

또 다른 사회적 위험은 LMM의 탄소 및 물 발자국입니다. 다른 형태의 AI와 마찬가지로, LMM은 상당한 에너지를 필요로 하며, AI의 물 발자국 증가에 기여합니다. LMM 및 기타 AI 형태는 중요한 사회적 이점을 제공할 수 있지만, 증가하는 탄소 발자국은 기후 변화에 주요 기여 요인이 될 수 있으며, 물 부족 지역에서 물 소비 증가로 인해 더욱 부정적 영향을 미칠 수 있습니다. LMM과 관련된 또 다른 사회적 위험은 점점 더 신뢰받는 지식의 출처로 여겨지는 그럴듯한 응답을 제공함으로써 보건의료, 과학, 의학 등의 영역에서 인간의 인식론적 권위를 약화시킬 수 있다는 점입니다.

보건의료 및 의학 분야에서 LMM의 윤리 및 거버넌스

LMM은 하나 이상의 행위자가 프로그래밍 및 제품 개발에 대해 내린 일련의 (혹은 연결된) 결정의 산물로 간주될 수 있습니다(그림 2). AI 가치사슬의 각 단계에서 내려지는 결정은 LMM의 개발, 배포 및 사용에 참여하는 사람들에게 직접적 및 간접적 영향을 미칠 수 있습니다. 이러한 결정은 국가, 지역 및 글로벌 차원에서 법과 정책을 제정하고 집행함으로써 정부가 영향을 주고 규제할 수 있습니다.

그림 2: LMM 개발, 제공 및 배포의 가치사슬

AI 가치사슬(value chain)은 일반적으로 본 가이던스에서 "개발자"라고 불리는 대규모 기술 기업에서 시작됩니다. 개발자는 또한 대학, 소규모 기술 기업, 국가 보건시스템, 공공-민간 컨소시엄 또는 데이터, 컴퓨팅 능력 및 AI 전문 지식과 같은 여러 입력요소(이를 "AI 인프라"라고 부름)를 사용할 수 있는 자원과 역량을 갖춘 기타 단체일 수 있습니다. 이러한 인프라를 활용해 범용 기반 모델(정부가 입법 및 규제에서 LMM을 지칭할 때 사용하는 용어)을 개발합니다. 이 모델은 직접적으로 다양한, 종종 예기치 않은 작업(보건의료 분야와 관련된 작업 포함)을 수행하는 데 사용할 수 있습니다. 일부 범용 기반 모델은 특히 보건의료 및 의학 분야에서 사용하기 위해 특별히 학습됩니다.

범용 기반 모델은 제공자로 불리는 제3자에 의해 특정 목적 또는 용도로 사용될 수 있으며, 이는 활성 프로그래밍 인터페이스(API)를 통해 이루어집니다. 여기에는 (i) 기반 모델의 추가 학습이 필요할 수 있는 새로운 LMM의 세부 조정, (ii) 사용자에게 서비스를 제공하기 위해 LMM을 응용 프로그램 또는 더 큰 소프트웨어 시스템에 통합, (iii) "플러그인"으로 알려진 구성 요소를 통합하여 LMM을 공식적이거나 규제된 형식으로 채널화, 필터링 및 구성하여 "이해 가능한" 결과를 생성하는 것들이 포함됩니다.[1]

그 이후, 제공자는 LMM을 기반으로 한 제품이나 서비스를 고객(또는 "제공자")에게 마케팅할 수 있습니다. 여기에는 보건부, 보건의료 시스템, 병원, 제약회사 또는 의료 제공자와 같은 개인도 포함됩니다. 제품이나 응용 프로그램을 구매하거나 라이선스를 취득한 고객은 이를 환자, 의료 제공자, 보건시스템 내의 기타 단체, 일반 대중 또는 자신의 비즈니스에 직접 사용할 수 있습니다.
AI 가치사슬은 "수직적으로 통합"될 수 있으며, 데이터를 수집하고 범용 기반 모델을 학습시키는 기업(또는 국가 보건 시스템과 같은 기타 단체)가 특정 용도로 LMM을 수정한 뒤 사용자에게 응용 프로그램을 직접 제공할 수 있습니다.

거버넌스는 기존의 법률과 정책, 그리고 새로 제정되거나 개정된 법, 규범, 개발자를 위한 내부 실천 강령 및 절차, 국제 협정과 프레임워크를 통해 윤리 원칙과 인권 의무를 구현하는 수단입니다.

LMM의 거버넌스를 구성하는 한 가지 방식은 AI 가치사슬을 세 단계로 나누는 것입니다: (i) 범용 기반 모델 또는 LMM의 설계 및 개발, (ii) 범용 기반 모델을 기반으로 한 서비스, 응용 프로그램 또는 제품 제공, (iii) 보건의료 서비스나 응용 프로그램의 배포. 본 가이던스에서는 각 단계에 대해 세 가지 질문에 초점을 맞춥니다:

- 가치사슬의 각 단계에서 어떤 위험이 해결되어야 하며, 어떤 행위자가 이러한 위험을 해결하는 데 가장 적합한가?
- 정부의 역할은 무엇이며, 관련 법률, 정책 및 규제는 어떻게 되어야 하는가?

AI 가치사슬의 각 단계에서 특정 위험을 해결할 수 있으며, 특정 행위자가 각 위험을 완화하고 윤리적 가치를 지키는 데 있어 더 중요한 역할을 할 가능성이 큽니다. 개발자, 제공자, 배포자 간의 책임 분담에 대해 의견 차이와 긴장이 있을 가능성이 있지만, 각 행위자가 잠재적 또는 실제 위험을 해결할 수 있는 능력을 가진 영역은 분명히 존재합니다.

범용 기반 모델(LMM)의 **설계 및 개발**

범용 기반 모델을 설계 및 개발하는 동안의 책임은 개발자에게 있습니다. 정부는 특정 관행을 요구하거나 금지하기 위해 법과 표준을 설정할 책임이 있습니다. 본 가이던스의 섹션 4에서는 LMM 개발 중 위험을 해결하고 이점을 극대화하기 위한 권장사항을 제공합니다.

1 WHO의 보건분야의 윤리 및 거버넌스 전문가, Leong Tze-Yun의 단보(communication).

범용 기반 모델(LMM)을 활용한 서비스 **제공**

보건의료 환경에서 사용되는 AI 기반 시스템과 관련된 특정 위험을 해결하기 위해 개발자와 제공자의 요구 사항 및 의무를 정의할 책임이 있습니다. 본 보고서의 섹션 5에서는 LMM을 사용하여 보건의료 서비스와 응용 프로그램을 제공하는 동안 위험을 해결하고 이점을 극대화하기 위한 권고사항을 제공합니다.

범용 기반 모델(LMM)을 활용한 **배포**

개발 및 제공 단계에서 관련 법률, 정책 및 윤리적 관행이 적용되었더라도, LMM의 예측 불가능성, 사용자의 비의도적 사용, 출력 변화 가능성으로 인해 사용 중에 위험이 발생할 수 있습니다. 본 보고서의 섹션 6에서는 LMM 사용 및 응용 프로그램과 관련된 위험과 과제를 해결하기 위한 권장사항을 제공합니다.

LMM에 대한 법적 책임

LMM이 보건의료 및 의학 분야에서 널리 사용됨에 따라, 오류, 오용, 그리고 궁극적으로 개인에게 해를 끼칠 가능성이 불가피합니다. 따라서, 피해를 입은 사용자가 적절히 보상받거나 구제책을 받을 수 있도록 책임 규칙을 마련해야 합니다.

정부는 인과관계의 추정을 도입하여 이를 수행할 수 있습니다. 또한, LMM 배포로 인해 발생하는 모든 피해에 대해 엄격한 법적 책임 기준을 도입하는 것을 고려할 수 있습니다. 엄격한 법적 책임 규칙은 피해를 입은 사람에게 보상을 보장할 수 있지만, 점점 더 정교해지는 LMM의 사용을 억제할 수도 있습니다. 정부는 무과실, 무책임 보상기금 도입도 고려할 수 있습니다.

LMM의 국제 거버넌스

정부는 국제 거버넌스가 이러한 기술의 세계화 속도에 발맞추도록 하기 위해 새로운 제도적 구조와 규칙을 구축하기 위해 협력해야 합니다. 또한, 보건의료 분야의 AI 배포와 관련된 기회와 도전에 대응하고, 사회 및 경제 분야에서 더 널리 응용될 수 있도록 유엔 시스템 내에서 강력한 협력과 협업을 해야 합니다.

국제 거버넌스는 모든 정부가 AI 기반 시스템의 개발 및 배포에 대한 투자와 참여에 대해 책임을 지도록 보장하고, 윤리 원칙, 인권 및 국제법을 준수하는 적절한 규정을 도입하도록 보장하기 위해 필요합니다. 국제 거버넌스는 또한 기업이 안전성과 유효성에 대한 적절한 국제 표준을 충족하고, 윤리 원칙과 인권 의무를 준수하는 LMM을 개발하고 배포하도록

보장할 수 있습니다. 정부는 기업이나 자신들에게 경쟁적으로 유리함과 불리함을 제공하는 규정을 도입하는 것을 피해야 합니다.

국제 거버넌스가 의미 있으려면 이러한 규칙은 고소득 국가뿐만 아니라 모든 국가에 의해 형성되어야 합니다. 이는 고소득 국가 정부와 협력하는 기술 기업에 의해서만 결정되지 않아야 합니다. AI의 국제 거버넌스는 2019년 유엔 사무총장이 제안한 대로 모든 이해관계자가 네트워크 다자주의를 통해 협력할 필요가 있을 수 있습니다. 이는 접근방식 유엔 체계, 국제 금융 기관, 지역 기구, 무역 블록 뿐만 아니라 시민 사회, 도시, 기업, 지방 당국, 젊은 세대 등 다양한 이해관계자들이 더욱 긴밀하고 효과적이며 포괄적으로 협력하도록 유도할 것입니다.

보건분야 인공지능의 윤리와 거버넌스 : 대규모 멀티모달 모델

해결해야 할 위험	무엇을, 그리고 누가 할 수 있는 가?	
	개발자 조치	정부 조치

개발 단계

- 편향
- 개인정보 보호
- 노동문제
- 탄소 및 물 발자국
- 허위정보 또는 오정보
- 안전 및 사이버 보안
- 인간의 의식론적 권위
- LMM의 독점적 통제

개발자 조치

- ○ 프로그래머 인증 및 교육
- ○ 데이터 보호 영향 평가
- ○ '모범사례'의 데이터 보호규칙에 대해 수집된 데이터를 교육
- ○ 학습 데이터를 최신 상태로 유지 및 맥락에 적합하게 갱신
- ○ 교육 데이터 투명성 보장
- ○ 데이터 작업자에게 공정한 임금 및 지원제공
- ○ 설계단계에 다양한 이해관계자 참여
- ○ 정확성과 예측 가능성을 고려한 설계
- ○ 윤리적 규범과 합의된 원칙에 기반한 설계
- ○ 모델의 에너지 효율성을 개선하는 설계

정부 조치

- ○ 강력한 데이터 보호법 마련 및 시행
- ○ 대상 제품 프로필 발행
- ○ 위임 결과 예측(가능성, 해석 가능성, 수정 가능성, 안전성, 사이버 보안)
- ○ 윤리적 위험을 파악하고 회피하기 위한 사전 인증 프로그램 도입
- ○ AI 개발과정에서 감사 수행
- ○ 개발자에게 탄소 및 물 발자국 문제 해결을 요구
- ○ 개발자에게 사용자를 위해 AI가 생성한 콘텐츠에 레이블을 지정하도록 요구
- ○ 알고리즘 초기 단계 등록을 장려 또는 요구
- ○ 공공 및 비영리 인프라에 투자 또는 제공
- ○ 자금지원을 통한 오픈소스 LMM 촉진

제공 단계

- 시스템 전반의 편향
- 허위정보 또는 오정보
- 조작
- 개인정보 보호
- ATM 자동화 편향

정부 조치

- ○ 보건분야 LMM 평가 및 승인을 담당할 규제기관 지정
- ○ 소스 코드 및 데이터 입력 포함 투명성 요구
- ○ 사용자 입력 데이터 보호를 위한 데이터 보호법 시행
- ○ 위험 또는 이점에 관계없이 윤리적 및 인권 기준 의무화
- ○ 제3자가 감사를 수행하고 공개하도록 영향평가를 요구하는 법률 제정
- ○ 성능 증명 요구 및 의료기기 규정 준수 요구
- ○ 시험되지 않은 상업적 사용금지 및 규제 테스트를 위한 샌드박스 탐색
- ○ 소비자 보호법을 적용하여 최종 사용자 및 환자에 대한 부정적 영향을 방지

배포 단계

- 부정확하거나 거짓 응답
- 편향
- 개인정보 보호
- 접근성 및 경제성
- 노동 및 고용문제
- ATM 자동화 편향
- 환자-전문가 간 상호작용 품질 저하
- 기술 저하

개발자 조치

- ○ 부적절한 환경에서 LMM 사용회피
- ○ 알려진 위험, 오류 및 피해를 명확한 경고와 대책과 함께 전달
- ○ 포괄적인 가격책정과 언어제공을 통해 접근성과 이용가능성을 보장

정부 조치

- ○ LMM 배포에 대해 독립적인 출시 후 감사 및 영향평가 의무화행
- ○ 허위정보 및 독성정보에 대한 책임 부여
- ○ 기술문서 등 운영적 공개 규정 시행
- ○ 의료 종사자들에게 LMM 의사결정, 편향회피, 환자참여, 사이버 보안 위험에 대해 교육
- ○ 적절한 사용을 위해 인간 감독 협회를 통한 공공의 참여를 촉진
- ○ 데이터 공유를 이해하고 사회적/문화적 수용을 평가하며, AI 활용 능력을 향상시키고, 허용 가능한 LMM 사용사례를 파악하기 위해 대중과 소통
- ○ 구매권한을 활용하여 가치사슬 참여자들이 투명성과 책임있게 행동하도록 장려

1 서론

본 가이던스는 보건 관련 응용 분야에서 대규모 멀티모달 모델(LMM)의 새로운 사용에 대해 다룹니다. [2] 여기에는 보건의료 및 의학 분야에서 LMM 사용의 잠재적 이점과 위험, 그리고 윤리, 인권 및 안전에 관한 가이던스와 의무를 준수하도록 보장하기 위한 LMM 거버넌스 접근방식이 포함됩니다. 본 가이던스는 WHO가 2021년 6월에 발행한 보건 분야 인공지능의 윤리 및 거버넌스*(1)*에 기반하고 있으며, 이는 보건 분야 AI 사용의 윤리적 도전과 위험을 다루고, AI가 모든 국가의 공익을 위해 사용되도록 보장하는 여섯 가지 원칙을 제시했으며, 기술의 잠재력을 극대화하기 위해 보건 분야 AI 거버넌스를 강화하기 위한 권고사항을 발행한 바 있습니다.

AI는 시스템과 도구에 통합된 알고리즘이 데이터를 학습하여 인간이 각 단계를 명시적으로 프로그래밍하지 않아도 자동화된 작업을 수행할 수 있는 능력을 의미합니다. 생성형 AI는 머신러닝 모델이 데이터 세트를 학습하여 텍스트, 이미지, 비디오, 음악 등 새로운 결과물을 생성하는 AI 기술의 한 범주입니다. 생성형 AI 모델은 학습 데이터로부터 패턴과 구조를 학습하고, 학습된 패턴으로부터 예측을 기반으로 새로운 데이터를 생성합니다. 생성형 AI 모델은 인간 피드백을 통한 강화 학습을 통해 개선될 수 있습니다. 이 과정에서 인간 트레이너는 생성형 AI 모델이 제공한 응답을 평가하여 알고리즘이 인간이 가장 가치를 느끼는 응답을 생성하도록 학습시킵니다. 생성형 AI는 디자인, 콘텐츠 생성, 시뮬레이션 및 과학적 발견 등 다양한 분야에서 잠재적 응용 가능성을 가지고 있습니다.

생성형 AI의 특정 유형 중 많은 관심을 받는 것이 대규모 언어 모델입니다. 이 모델은 텍스트 형태의 하나로 입력을 받아 텍스트로 응답을 제공합니다. 대규모 언어 모델은 단일 모달 모델의 예로, 초기 버전의 챗봇의 운영 기반이 되는 모델입니다. 대규모 언어 모델은 대화를 수행하지만, 모델 자체는 생성하는 내용에 대한 개념을 가지고 있지 않습니다. 단순히 이전 단어, 학습된 패턴 또는 단어 조합에 따라 다음 단어를 예측할 뿐입니다*(2)*.

본 문서는 대규모 언어 모델을 포함하여 대규모 멀티모달 모델(대규모 언어 모델 포함)의 사용 증가에 대해 다룹니다. LMM는 보건 및 의학 분야에서 사용하기 위해 훈련되며, 텍스트를 넘어 매우 다양한 데이터 세트를 포함합니다. 이러한 데이터에는 바이오센서, 게놈, 에피게놈, 프로테오믹스, 영상, 임상, 사회적, 환경적 데이터가 포함됩니다*(3)*. 따라서 LMM은 하나 이상의 유형의 입력을 수용할 수 있으며, 입력된 데이터 유형에 국한되지 않는 출력을 생성할 수 있습니다. LMM은 보건 분야와 신약 개발에서 다양한 응용을 위한 모델로 기대되고 있습니다.

---

2 이 지침에서는 “대규모 멀티모달 모델(LMM)”과 “범용 기반 모델”이라는 용어를 혼용하여 사용하며, 특히 거버넌스와 관련된 논의에서는 후자의 용어가 사용됩니다. 그러나 LMM이 일반적인 목적을 위해 광범위한 작업을 수행할 수 있는지는 아직 알려지지 않았습니다.

LMM은 기존의 AI 및 머신러닝 모델과 다릅니다. AI는 이미 많은 소비자 응용 프로그램에 광범위하게 통합되어 있지만, 대부분의 알고리즘 결과물은 고객이나 사용자의 참여를 요구하거나 초대하지 않습니다. 단, 사용자 생성 콘텐츠를 선별하여 주의를 끄는 소셜 미디어 플랫폼에 통합된 기초적인 형태의 AI는 예외입니다*(4)*. LMM과 다른 형태의 AI 간의 또 다른 차이는 다재다능함에 있습니다. 이전 및 기존의 AI 모델(의료용 포함)은 특정 작업을 위해 설계되었기 때문에 유연성이 부족합니다. 이들 모델은 학습 데이터 세트와 그 수준에 정의된 작업만 수행할 수 있으며, 다른 데이터 세트를 사용해 재학습하지 않으면 다른 기능을 수행하거나 적응할 수 없습니다*(5)*. 따라서, 미국 식품의약국(FDA)에서 임상 의학을 위해 승인된 500개 이상의 AI 모델 중 대부분은 한두 가지의 좁은 작업에 대해서만 승인되었습니다*(5)*. 이에 반해, LMM은 다양한 데이터 세트로 학습되며, 명시적으로 학습되지 않은 작업을 포함하여 다양한 작업에서 사용할 수 있습니다*(5)*.

LMM은 일반적으로 인간-컴퓨터 알고리즘 상호작용을 용이하게 하는 인터페이스와 형식을 갖추고 있으며, 이는 인간의 의사소통을 모방할 수 있습니다. 이러한 특성 때문에 사용자는 알고리즘에 인간과 같은 속성을 부여할 수 있습니다. 따라서, LMM의 사용 방식과 생성 및 제공되는 콘텐츠는 "인간과 유사한" 것으로 보일 수 있지만 이는 다른 형태의 AI와는 다르며, LMM의 전례 없는 대중적인 채택에 기여했습니다. 또한, LMM이 제공하는 응답이 권위 있는 것처럼 보이기 때문에, 많은 사용자가 이를 비판 없이 옳다고 받아들이는 경향이 있습니다. 이는 LMM이 올바른 응답을 보장할 수 없고, 생성된 응답에 윤리적 규범이나 도덕적 추론을 통합할 수 없는 경우에도 마찬가지입니다. 본 가이던스는 LMM이 보건 및 의학 분야에서 사용되는(혹은 사용될 것으로 상상되는) 다양한 방식을 설명하지만, LMM은 이미 교육, 금융, 통신 및 컴퓨터 과학 등 여러 분야에서 사용되고 있습니다.

LMM은 하나 이상의 행위자가 프로그래밍 및 제품 개발과 관련하여 내린 일련의(혹은 연속적인) 결정의 산물로 간주될 수 있습니다. AI 가치사슬의 각 단계에서 내려진 결정은 LMM의 개발, 배포 및 사용에 참여하는 사람들에게 직접적 및 간접적 결과를 초래할 수 있습니다. 이러한 결정은 정부가 법률과 정책을 제정하고 시행함으로써 국가적, 지역적, 그리고 글로벌 차원에서 영향을 미치고 규제할 수 있습니다.

AI 가치사슬은 주로 대규모 기술 기업에서 시작됩니다. 개발자는 대학, 소규모 기술 기업, 국가 보건시스템, 공공-민간 컨소시엄 또는 데이터, 컴퓨팅 능력, AI 전문지식 등 여러 입력요소를 사용할 수 있는 자원과 역량을 가진 기타 단체일 수도 있습니다. 이들은 범용 기반 모델을 개발하는 데 필요한 "AI 인프라"를 구성합니다. 이들 모델은 사용자가 다양한, 종종 예기치 않은 작업(보건과 관련된 작업 포함)을 수행하기 위해 직접 사용할 수 있습니다. 일부 범용 기반 모델은 특히 보건 및 의학 분야에서 사용하기 위해 특별히 학습되었습니다.

범용 기반 모델은 제3자("제공자")가 활성화된 프로그래밍 인터페이스를 통해 특정 목적 또는 용도로 사용할 수 있습니다. 여기에는 (i) 기반 모델의 추가 학습이 필요할 수 있는 새로운 LMM의 세부 조정, (ii) 사용자에게 서비스를 제공하기 위해 LMM을 응용 프로그램 또는 더 큰 소프트웨어 시스템에 통합, (iii) "플러그인"으로 알려진 구성 요소 통합을 통해 LMM을 공식적이거나 규제된 형식으로 채널화, 필터링 및 구성하여 "이해 가능한" 결과를 생성하는 것이 포함됩니다.[3]

3 WHO에서 보건 분야의 AI 윤리와 거버넌스 분야의 전문가로 활동하는 Leong Tze-Yun의 단보(communication).

그 이후, 제공자는 LMM을 기반으로 한 제품이나 서비스를 보건부, 보건의료시스템, 병원, 제약회사, 또는 의료 제공자와 같은 개인 고객(또는 "제공자")에게 마케팅할 수 있습니다. 제품이나 응용 프로그램을 구매하거나 라이선스를 취득한 고객은 이를 환자, 의료 제공자, 보건의료시스템 내의 다른 단체, 일반 대중, 또는 자신의 사업에서 직접 사용할 수 있습니다. 가치사슬은 "수직적으로 통합"될 수 있으며, 데이터를 수집하고 범용 기반 모델을 학습시키는 기업(또는 보건시스템과 같은 기타 단체)이 LMM을 특정 용도에 맞게 수정하여 응용 프로그램을 직접 사용자에게 제공할 수 있습니다.

WHO는 AI가 공중보건 개선과 보편적 건강보장 달성을 포함하여 보건시스템에 제공할 수 있는 막대한 이점을 인식하고 있습니다. 그러나 WHO의 "보건 분야 AI의 윤리 및 거버넌스"에 대한 가이던스*(1)*에서 설명된 바와 같이, AI는 공중보건을 약화시키고 개인의 존엄성, 개인정보, 인권을 위태롭게 할 수 있는 중대한 위험을 수반합니다. LMM은 비교적 새로운 기술임에도 불구하고, 급속한 채택과 확산 속도로 인해 WHO는 LMM이 전 세계적으로 성공적이고 지속 가능하게 사용될 수 있도록 이 가이던스를 제공합니다. WHO는 본 가이던스가 AI의 잠재적 이점과 위험, 설계 및 사용에 적용되어야 할 윤리적 원칙, 거버넌스 및 규제 접근방식에 대한 많은 상충된 의견이 있는 시점에서 발행되고 있음을 인식하고 있습니다. 본가이던스는 LMM이 보건 분야에서 처음 응용되기 시작한 직후, 그리고 점차 더 강력한 모델이 출시되기 전에 발행되었으므로, WHO는 기술의 급속한 진화, 사회가 그 사용을 다루는 방식, 그리고 보건 및 의학을 넘어선 LMM 사용으로 인한 건강상의 결과에 발맞추어 이 가이던스를 업데이트할 예정입니다.

1.1 LMM의 중요성

LMM은 비교적 새로우며, 아직 검증되지 않았음에도 불구하고 보건의료 및 의학 분야를 포함한 다양한 영역에서 사회에 엄청난 영향을 미치고 있습니다. 미국의 한 기술 기업에서 출시한 대규모 언어 모델인 Chat GPT는 출시 두 달 만인 2023년 1월 기준으로 월간 활성 사용자 수가 1억 명에 달하는 것으로 추정되었습니다. 이는 당시 역사상 가장 빠르게 성장한 소비자 응용 프로그램으로 기록되었습니다*(6)*.

현재는 많은 기업이 LMM을 개발하거나 인터넷 검색 엔진과 같은 소비자 응용 프로그램에 통합하고 있습니다. 대규모 기술 기업들은 대부분의 응용 프로그램에 LMM을 빠르게 통합하거나 새로운 응용 프로그램을 만들고 있으며*(7,8)*, 수백만 달러의 민간 투자를 받은 신생 기업들도 경쟁력 있는 LMM 개발에 박차를 가하고 있습니다*(9)*. 또한, 오픈소스 LMM이 대규모 기업에서 개발한 모델보다 더 빠르고 저렴하게 등장하고 있으며, 이는 이러한 플랫폼의 접근성 덕분입니다*(10)*.

LMM의 등장이 기술 산업에 새로운 투자를 촉진하고 신제품들이 출시되는 가운데, 일부 기업은 LMM이 특정 응답을 생성하는 이유를 완전히 이해하지 못한다고 인정하고 있습니다*(11)*. 인간 피드백을 활용한 강화 학습에도 불구하고, LMM은 항상 예측 가능하거나 통제 가능한 결과를 생성하지는 않습니다. 예를 들어, 사용자에게 불편함을 줄 수 있는 "대화"에 참여하거나*(12)*, 오류가 있거나 잘못된 내용을 생성하면서도 매우 설득력 있게 보이는 결과를 제공할 수 있습니다*(13)*. 그럼에도 불구하고, LMM에 대한 많은 지지는 그 기능에 대한 열광뿐만 아니라, 동료 검토를 거치지 않은 출판물에서 성과를 과도하게 주장하거나 비판 없이 받아들이는 경우에 의해 뒷받침되고 있습니다*(14)*.

LMM은 빠르게 도입되었지만, 이를 학습시키는 데 사용된 데이터 세트는 공개되지 않았습니다*(15)*. 때문에 데이터가 편향되어 있는지, 합법적으로 수집되었는지, 데이터 보호 규정 및 원칙을 준수했는지, 그리고 특정 작업이나 질문에 대한 성과가 동일하거나 유사한 문제로 학습된 결과인지 아니면 문제를 해결하는 능력을 습득한 결과인지를 알기 어렵거나 불가능합니다. 데이터 보호법과의 일치성을 포함한 LMM 학습에 사용된 데이터에 대한 다른 우려사항은 아래에서 논의됩니다.

개인과 정부 모두 LMM의 출시를 준비하지 못한 상태였습니다. 개인은 LMM을 효과적으로 사용하는 방법에 대해 교육받지 못했으며, LMM이 생성하는 응답이 항상 정확하거나 신뢰할 수 있는 것이 아니라는 사실을 이해하지 못할 수 있습니다. 이는 LMM 기반 챗봇이 그러한 인상을 줄 수도 있기 때문입니다. 한 연구에 따르면, 대규모 언어 모델 중 하나인 GPT-3는 "인간과 비교했을 때 더 쉽게 이해할 수 있는 정확한 정보를 생성할 수 있는 반면", "더 설득력 있는 허위 정보"를 생성할 수도 있으며, 인간은 LMM이 생성한 콘텐츠와 인간이 생성한 콘텐츠를 구별하지 못한다고 밝혔습니다*(16)*.

정부 역시 대부분 준비가 부족했습니다. AI 사용을 규제하기 위한 법률과 규정은 LMM과 관련된 과제나 기회를 다룰 준비가 되어 있지 않을 수 있습니다. 유럽연합은 전역에서 적용될 인공지능법을 제정하기 위한 합의에 도달했지만, LMM을 고려하기 위해 초안 작성의 최종 단계에서 입법 프레임워크를 수정해야 했습니다*(17)*. 다른 정부는 빠르게 새로운 법률이나 규정을 개발하거나*(18)*, 일시적인 금지를 시행했습니다(이 중 일부는 이미 폐지됨)*(19)*. 앞으로 더 강력하고 능력 있는 LMM이 계속 출시될 것으로 예상되며, 이는 새로운 혜택뿐만 아니라 새로운 규제 과제도 가져올 수 있습니다. 이러한 역동적인 환경에서, 이전의 윤리적 가이던스를 포함한 가이던스를 바탕으로 보건 및 의료 분야에서 LMM을 사용하는 데 대한 제안과 권장 사항이 제공됩니다.

1.2 보건분야를 위한 AI 윤리 및 거버넌스에 대한 WHO 가이던스

WHO의 초기 보건분야 AI 윤리 및 거버넌스 가이던스*(1)*은 다양한 머신러닝 접근방식과 보건의료에서의 AI의 다양한 응용을 검토했지만, 생성형 AI나 LMM에 대해 구체적으로 다루지는 않았습니다. 이 가이던스이 개발되는 동안과 2021년에 발표할 당시, 생성형 AI와 LMM이 이렇게 빠르게 보급되고 임상 진료, 보건 연구, 공중보건에 적용될 것이라는 증거는 없었습니다.

그럼에도 불구하고, 이 가이던스에서 파악된 근본적인 윤리적 도전과제와 핵심 윤리 원칙 및 권고사항(Box 1 참조)은 여전히 유효하며, LMM을 평가하고 효과적이고 안전하게 사용하는 데 적용될 수 있으며, 이 새로운 기술과 관련하여 추가적인 거버넌스의 공백과 도전 과제가 계속해서 발생할 것입니다. 과제, 원칙 및 권장사항은 이러한 가이던스에 제시된 LMM에 접근하는 데 기초가 되었습니다.

Box 1. WHO 합의에 기반한 보건분야 AI 사용을 위한 윤리원칙 개요

- **자율성 보호:** 인간은 보건의료 시스템과 의료 결정에 대한 통제권을 유지해야 합니다. 제공자는 AI 시스템을 안전하고 효과적으로 사용할 수 있는 정보를 보유하고 있습니다. 사람들은 AI 시스템이 자신들의 의료에서 어떤 역할을 하는지 이해해야 합니다. 데이터 보호를 위한 적절한 법적 틀을 통해 유효한 정보에 기반한 동의로 데이터의 개인정보와 기밀성이 보호됩니다.

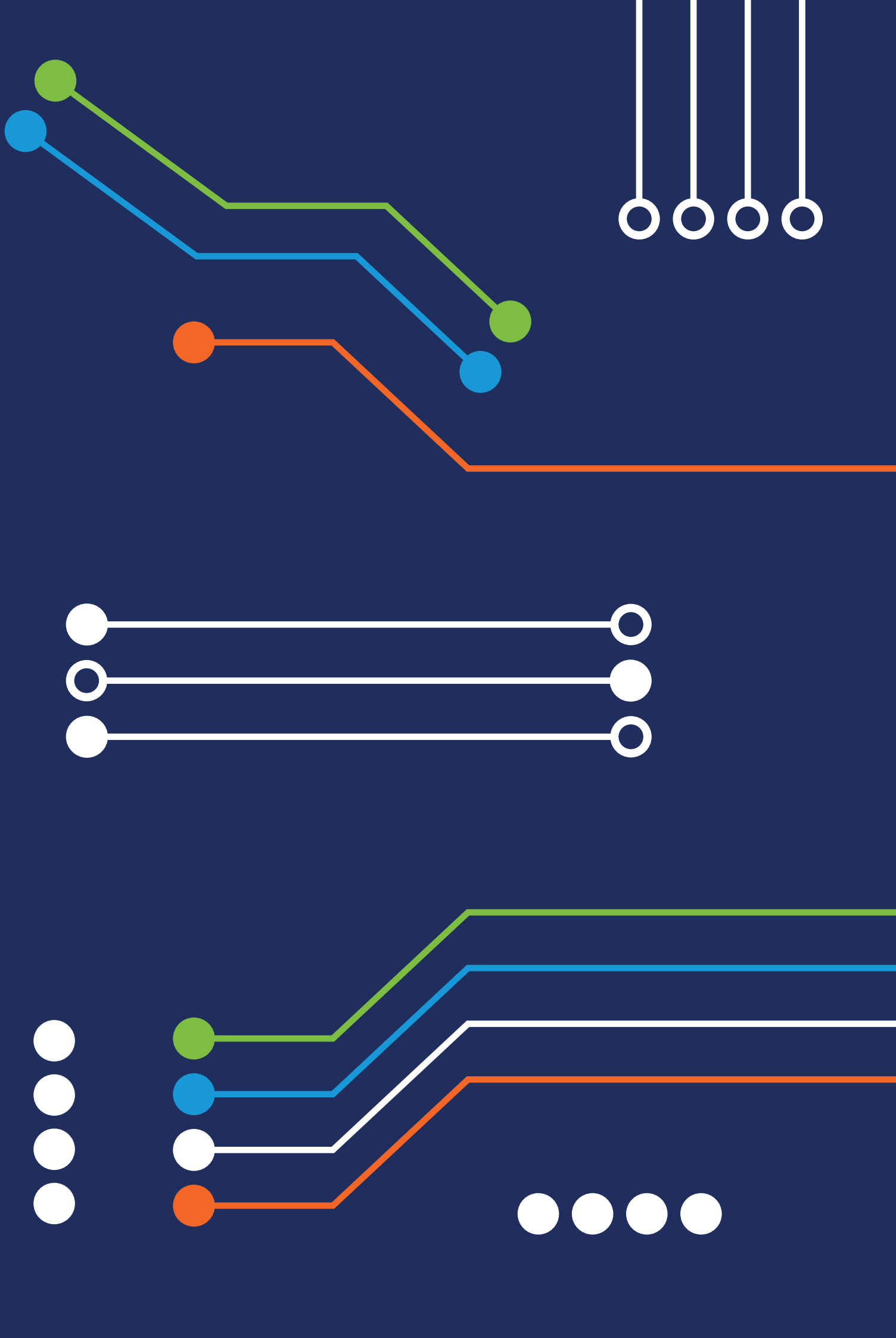
- **인간의 복지, 안전 및 공익 증진:** AI 설계자는 명확하게 정의된 용도 또는 적응증에 대해 안전성, 정확성 및 유효성에 대한 규제 요구사항을 충족합니다. 실질적인 품질관리 및 시간 경과에 따른 AI 사용의 지속적인 품질개선 조치가 가능해야 합니다. AI 사용으로 인해 정신적 또는 신체적 피해가 발생할 경우 대안적 관행이나 접근방식을 통해 이를 방지할 수 있어야 합니다.

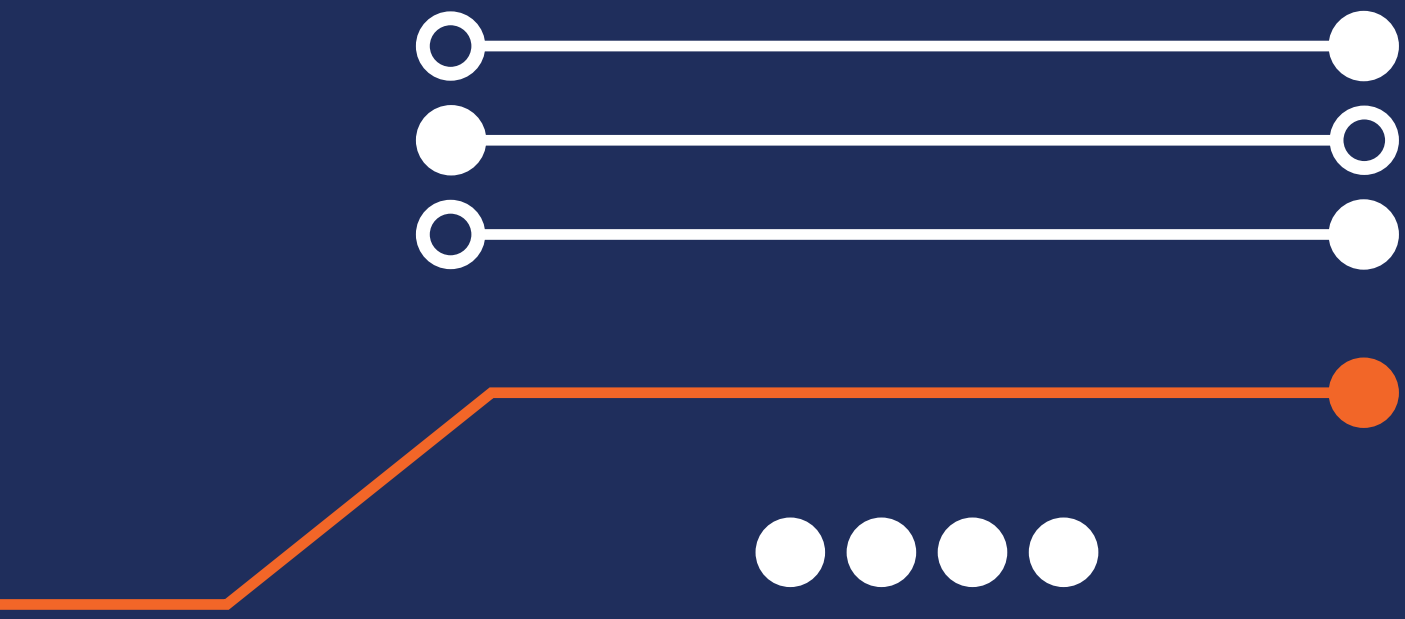
- **투명성, "설명 가능성" 및 이해 가능성 보장:** AI 기술은 개발자, 의료 전문가, 환자, 사용자 및 규제기관이 이해할 수 있도록 설계되어야 합니다. AI가 설계 또는 배포되기 전에 충분한 정보가 공개되거나 문서화되어야 하며, 이 정보는 AI가 어떻게 설계되었는지, 그리고 어떻게 사용되어야 하는지 반대로 어떻게 사용되지 않아야 하는지에 대한 공공의 의미 있는 협의 및 토론을 가능하게 합니다. AI는 이를 설명 받는 사람의 이해 수준에 따라 설명 가능합니다.

- **책임감과 책무성 강화:** AI는 적절한 조건에서 적절히 훈련된 사람들에 의해 사용되도록 책임성을 강화합니다. 환자와 임상의가 AI 개발 및 배포를 평가합니다. 알고리즘의 상류와 하류에 규제 원칙이 적용됩니다. AI 기반 결정으로 인해 부정적인 영향을 받는 개인 및 집단을 위한 문제 제기 및 구제 메커니즘이 적절히 마련되어야 합니다.

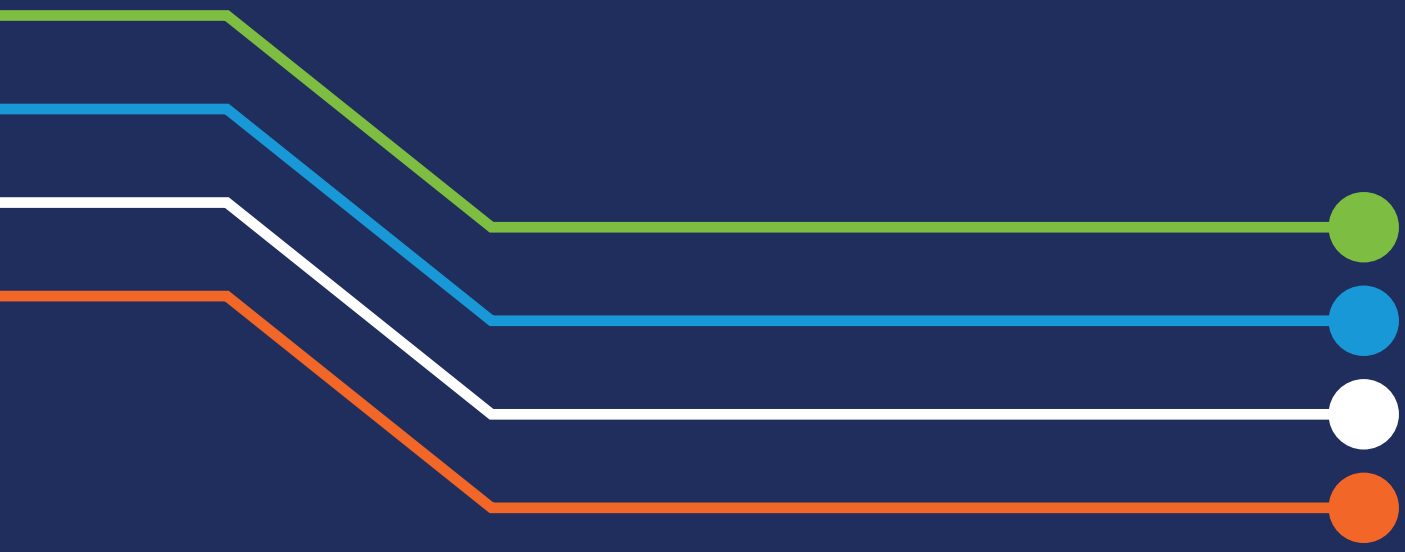
- **포용성과 형평성 보장:** AI는 연령, 성별, 성 정체성, 소득, 인종, 민족, 성적 지향, 역량 또는 기타 특성에 상관없이 가능한 한 널리, 적절히, 공평하게 사용되고 접근할 수 있도록 설계되고 공유되어야 합니다. AI는 고소득 국가뿐만 아니라 저소득 및 중소득 국가에서도 사용 가능해야 하며, 특정 집단에 불리한 편향을 포함하지 않아야 합니다. AI는 불가피한 권력 격차를 최소화해야 하며, 특정 집단에 미치는 불균형한 영향을 파악하기 위해 지속적으로 모니터링 및 평가되어야 합니다.

- **반응성이 높고 지속 가능한 AI 촉진:** AI 기술은 보건시스템, 환경, 직장의 지속 가능성을 촉진하는 방향과 일치해야 합니다.





I. LMM의 응용, 과제 및 위험



2 보건분야에서 LMM 사용의 응용 및 과제

AI의 보건분야 응용에는 진단, 임상진료, 연구, 약물 개발, 보건의료 행정, 공중보건 및 감시가 포함됩니다. LMM의 많은 응용은 AI의 새로운 사용 방식은 아니지만, 임상의, 환자, 일반인, 보건의료 전문가 및 종사자들이 LMM에 접근하고 이를 사용하는 방식은 다릅니다. 이 섹션에서는 보건의료 분야에서의 LMM의 잠재적 응용과 그 사용과 관련된 실제 및 예상되는 과제와 위험에 대해 다룹니다. 많은 적용 분야와 그 용도에 대해서는 아직 입증되지 않았으며, 궁극적으로 광고된 혜택을 제공하지 못할 가능성도 있습니다.

2.1 진단 및 임상진료

AI는 방사선과 의료 영상, 결핵 및 종양학과 같은 분야에서 이미 진단 및 임상진료를 지원하는데 사용되고 있습니다. 임상의는 AI를 통해 상담 중 환자기록을 통합하고, 위험 환자를 식별하고, 어려운 치료 결정에 도움을 받고, 임상적 오류를 포착할 수 있을 것으로 기대하고 있습니다*(1)*. LMM은 가상 및 대면 상담 모두에서 진단 및 임상진료 전반에 걸쳐 AI 기반 시스템의 사용을 확장할 수 있을 것으로 보이며, 일부 전문가들은 LMM이 "과거의 청진기보다 의사들에게 더 중요한 역할을 할 것"으로 기대하고 있습니다*(20)*. 여러 LMM은 미국 의학면허 시험을 통과했습니다. 하지만, 의료지식을 단순히 반복하여 서면 시험을 통과하는 것은 안전하고 효과적인 임상 서비스를 제공하는 것과 동일하지 않습니다*(21)*. 또한, LMM은 온라인에 이전에 게시되지 않았거나 어린아이도 쉽게 해결할 수 있는 자료가 포함된 시험에서 실패한 사례도 있습니다*(22)*. 한 대규모 언어 모델의 임상 지식에 대한 연구에서는 "의료 질문에 답변하기 위해 사용되는 대규모 언어 모델에서 보건의료 제공자, 관리자, 소비자가 사용할 수 있는 도구로 전환하려면 기술의 안전성, 신뢰성, 유효성 및 개인정보를 보장하기 위한 상당한 추가 연구가 필요하다"고 결론지었습니다*(23)*.

진단은 특히 유망한 분야로 간주되며, LMM은 복잡한 사례에서 희귀한 진단이나 "비정상적 증상"을 식별하는 데 사용될 수 있습니다*(24)*. 의사들은 이미 인터넷 검색 엔진, 온라인 리소스 및 감별 진단 생성기를 사용하고 있으며, LMM은 진단을 위한 추가 도구가 될 것입니다. LMM은 또한 일상적인 진단에서 명백한 진단이 간과되지 않도록 의사에게 추가적인 의견을 제공하는 데 활용될 수 있습니다. 이는 부분적으로 LMM이 환자의 전체 의료 기록을 의사보다 훨씬 빠르게 스캔할 수 있기 때문에 신속하게 이루어질 수 있습니다*(24)*.

현재 인기 있는 여러 LMM은 임상의들을 지원하기 위한 파일럿 프로그램에서 사용되고 있지만, 이 모델들은 전자 건강 기록이나 의료 또는 기타 관련 보건 데이터를 기반으로 특정한 훈련이 이뤄지지 않은 것들입니다. 다만, 데이터 세트에는

이러한 정보가 포함되어 있습니다. 예를 들어, 미국의 여러 의료 시스템에서는 한 기술 기업이 제공하는 LMM이 환자의 메시지를 읽고 의사의 응답 초안을 작성하는 파일럿 테스트를 진행하고 있습니다. 이는 의료진이 환자 문의에 답변하는 데 소요되는 시간을 줄이고, 매일 수천 건의 메시지를 처리하는 데 따른 번아웃을 줄이며, 임상 업무에 집중할 수 있도록 돕는 것을 목표로 하고 있습니다("키보드 해방")*(25)*. 환자 메시지가 수신되면, LMM은 환자의 정보와 전자 의료 기록 버전을 기반으로 초안을 작성합니다. 그러나 AI는 일부 환자 질문에만 사용되며, 응답은 많은 편집이 필요합니다*(25)*. 그럼에도 불구하고, 미국의 한 연구에서는 Chat GPT 기반 챗봇이 온라인 포럼에서 질문에 답변하는 데 있어 자격 있는 의사보다 더 나은 성과를 보였다고 밝혔습니다. 선택된 195개의 질문 중 약 80%의 사례에서 독립 평가자들이 의사의 응답보다 챗봇의 응답을 선호했습니다*(26)*. 챗봇은 표준화된 "비공식적 상담" 질문에 답변하거나 환자의 초기 진단 단계에서 정보와 응답을 제공하고, 검사 결과를 요약하는 데도 유용할 수 있습니다*(27)*.

기업과 대학들도 의료 및 보건 데이터 또는 전자 건강 기록을 활용해 훈련된 LMM을 개발하고 있으며, 여기에는 소규모 데이터 세트를 기반으로 한 LMM이 포함됩니다. 예를 들어, 한 LMM은 약 3만 개의 의료 증례 보고서 데이터 세트로 훈련되어, 의학적 상태와 증상 간의 관계를 학습함으로써 진단을 지원할 수 있도록 개발되었습니다*(28)*. 또 다른 LMM은 10만 개 이상의 흉부 X선 데이터 세트로 훈련되어 이상 징후를 식별하고 통찰을 제공하거나 질병을 식별할 수 있는 있도록 훈련되었습니다*(29)*. 공개적으로 평가된 몇몇 LMM은 수백만 개의 전자 건강 기록과 기타 전문 및 일반 의학 지식을 포함한 소스 데이터를 기반으로 알고리즘에 의해 훈련되었습니다. 이 접근방식은 알고리즘이 다양한 형태의 서면 의료 정보를 처리하고 응답을 생성하는 능력("의료 질의 응답")을 향상시켰습니다*(30)*.

몇몇 대규모 기술 기업들은 일반 용도의 LMM을 임상 진단 및 치료에 도움을 줄 수 있는 모델로 변환시키고 있습니다. 한 기술 기업은 Med-PaLM 2를 개발 중인데, 이는 의료 텍스트에서 얻은 통찰을 요약하고 질문에 답할 수 있도록 설계되었으며, 현재 X선 및 유방촬영술과 같은 이미지를 종합하여 보고서를 작성하고 후속 질문에 응답할 수 있도록 발전하고 있습니다. 이러한 기능은 의료 종사자와 컴퓨터 간의 "의견 불일치(Peer Disagreement)"를 완화하는 데 기여할 수 있습니다*(31)*.

장기적인 비전은 "범용 의료 인공지능"을 개발하는 것으로, 이는 보건의료 종사자들이 LMM과 유연하게 대화하여 맞춤형 임상의 주도 질의에 따라 응답을 생성할 수 있게 합니다. 이를 통해 사용자는 LMM을 재훈련하거나 비정형 데이터를 처리하도록 LMM을 추가적으로 훈련하지 않고도, 일반적인 언어로 요구사항을 설명하여 범용 의료 AI 모델을 새로운 작업에 적용할 수 있습니다*(5)*.

진단 및 임상진료에서 LMM 사용의 위험

임상진료에서 제공하는 LMM의 가능성은 여러 주요 위험을 동반하며, 일부 위험은 LMM 이전부터 존재해 왔습니다. 진단 및 임상진료에서 LMM 사용과 관련된 다섯 가지 주요 위험이 확인되었습니다.

- ***부정확하거나 불완전하고 편향적이거나 잘못된 응답:*** LMM에 대한 한 가지 우려는 챗봇이 데이터를 기반으로 잘못된 응답이나 완전히 허위인 응답(예: LMM이 만들어낸 참고자료 등)을 생성하는 경향*(32)*과 학습 데이터에 내재된 결함을 그대로 복제하는 방식으로 편향된 응답을 생성하는 점입니다*(33)*. LMM은 또한 AI 기술이 사용되는 환경에 대한 가정을 기반으로 다른 환경에 적합한 권장사항을 제공하는 맥락적 편향을 초래할 수 있습니다*(1)*. 예를 들어, 저소득 및 중소득 국가의 데이터와 관점이 학습 데이터에서 충분히 반영되지 않을 경우, LMM이 저소득 국가의 보건부를 위한 질병 치료 방안을 요약하도록 요청받는다면, 고소득 국가에만 적합한 접근방식을 재생산할 가능성이 있습니다*(34)*. 또한, LMM은 불완전한 응답을 제공할 수 있으며, 전혀 응답하지 않거나 사용 환경의 변화된 상황을 반영하지 못한 응답을 제공할 수 있습니다.

잘못된 응답, 소위 "환각(hallucination, 할루시네이션)"으로 알려진 응답은 LMM이 생성한 사실적으로 정확한 응답과 구별할 수 없습니다. 이는 LMM이 인간 피드백을 통한 강화 학습을 거쳤더라도, 사실을 생성하는 것이 아니라 사실처럼 보이는 정보를 생성하도록 훈련되기 때문입니다. 한 연구에 따르면, 대규모 언어 모델은 단순한 사실 집합을 요약할 때 최소 3%에서 최대 27%까지 환각을 일으킨다고 합니다*(35)*. 현재의 LMM은 또한 "신속한 엔지니어링"이라는 인간의 입력 최적화에 의존합니다*(36)*. 따라서 의료 데이터와 보건 정보를 기반으로 훈련된 LMM이라 하더라도 반드시 올바른 응답을 생성한다고 보장할 수 없습니다. 특정 LMM 기반 진단에서는 응답의 정확성을 확인할 테스트나 검증 방법이 없을 수도 있습니다*(24)*. 보건의료 및 공중보건 의사결정 영역에서는, LMM이 대부분의 경우 사실적으로 정확하다 하더라도 이를 안전하고 효과적으로 보건의료 시스템에 구현하거나 개발하는 비용을 정당화할 만큼 정확하지 않을 수 있습니다.

- ***데이터 품질 및 데이터 편향:*** LMM이 편향되거나 부정확한 응답을 생성하는 주요 원인 중 하나는 데이터 품질이 낮기 때문입니다. 현재 공개적으로 이용 가능한 많은 LMM은 인터넷과 같은 대규모 데이터 세트를 기반으로 훈련되었으며, 이러한 데이터 세트는 잘못된 정보와 편향으로 가득할 가능성이 높습니다. 대부분의 의료 및 보건 데이터 역시 인종, 민족, 조상, 성별, 성 정체성 또는 연령에 관계없이 편향되어 있습니다. 보건 데이터를 기반으로 훈련된 LMM은 이러한 편향을 내재화하는 경향이 있으며, 이는 대부분의 데이터가 고소득 국가에서 수집되기 때문입니다. 예를 들어, 유전자 데이터는 유럽계 사람들에게 불균형적으로 수집되는 경향이 있습니다*(1)*. LMM은 또한 오류와 부정확한 정보로 가득 찬 전자 건강 기록을 기반으로 훈련되거나*(24)*, 신체 검사에서 얻은 부정확한 정보에 의존할 수 있으며, 이는 LMM의 결과물에 영향을 미칠 수 있습니다*(25)*. 데이터 품질 및 편향 문제는 LMM을 포함한 모든 AI 모델에 영향을 미칩니다*(1)*.

Box 2. 아동의 LMM 사용에 대한 윤리적 고려사항

최근 AI 및 머신러닝에서 소아 데이터의 안전하고 윤리적인 사용을 위한 포괄적인 가이던스(ACCEPT-AI)가 발표되었지만*(43)*, 아동이 LMM을 사용하는 데 따른 잠재적 영향에 대해 특별히 고려해야 합니다.

개방형 LMM의 광범위한 이용 가능성은 다양한 연령대의 사용자가 접근할 수 있도록 합니다. 그러나 아동이 LMM과 상호작용하는 방법이나 사용하는 방식에 대한 증거는 제한적입니다. LMM 사용의 잠재적 기회와 단점은 보다 넓은 교육적 맥락에서 논의된 바 있지만*(44)*, 아동의 이러한 상호작용이 그들의 정신적 또는 신체적 건강에 어떤 영향을 미치는지는 불분명합니다. 아동의 LMM 사용은 시간 경과에 따른 모니터링을 실시하여 혜택과 잠재적 해악을 이해해야 합니다.

소아의 동의, 승낙, 법적 부모 참여 요건에 대한 법률과 정책은 국가 간 및 국가 내에서도 차이가 있습니다. 따라서 아동에 특화된 통합적이고 통일된 글로벌 규제 및 감독이 부족하면, 특히 LMM 사용에서 확인되지 않거나 모니터링되지 않은 해악이 발생할 수 있습니다. 구체적으로, LMM이 소아 건강 데이터를 얼마나 정확히 일반화할 수 있는지는 명확하지 않습니다. 연구에 따르면, 성인 데이터 세트를 소아 집단에 일반화하는 것은 제한적일 수 있습니다. 따라서 소아 데이터는 테스트와 학습 데이터 세트에서 별도로 유지되어야 합니다*(45)*.

개발자는 학습 데이터에 연령을 포함한 인구 통계 정보를 포함해야 하며, 관련된 경우 LMM과의 적절하고 안전한 상호작용을 위해 대상 집단(연령 범위 포함)에 대한 명확한 설명자를 제공하도록 권장되어야 합니다. 법적으로 가능한 경우, 아동 사용자와의 적절한 상호작용 및 피드백을 포함하여 LMM을 개선해야 합니다.

한 기술 기업은 자사의 LMM(GPT-4) 시스템 카드에 "GPT-4 초기 버전과 출시 버전 모두 이전 언어 모델과 마찬가지로 편향적이고 신뢰할 수 없는 콘텐츠를 생성하는 한계를 가지고 있다"고 명시하고 있습니다*(37)*. LMM은 알고리즘이 학습된 데이터의 종료 시점에 의해 제한될 수 있지만, 일부 LMM은 이제 인터넷에서 최신 정보를 검색할 수 있습니다. 예를 들어, Chat GPT-4는 2021년 9월까지의 데이터로 학습되었지만*(38)*, 이제는 인터넷 검색 또는 탐색을 통해 최신 정보를 얻을 수 있습니다*(39)*. 그러나 이는 더 많은 잘못된 정보나 부정확한 정보를 생성할 가능성을 증가시킬 수 있습니다. 이전의 "종료 시점"은 새로 게시된 허위 자료의 도입을 방지했습니다*(39)*. 의학에서는 최신 정보와 높은 정확성이 모두 표준 진료를 충족시키고 특정 질병을 이해하는 데 중요합니다.

- ***자동화 편향***: LMM이 잘못되거나 부정확하거나 편향된 응답을 생성할 수 있다는 우려는, 다른 형태의 AI와

마찬가지로, LMM이 전문가와 보건의료 전문가(및 환자, 아래 참조)에게 자동화 편향을 조장할 가능성 때문에 더욱 증폭됩니다*(1)*. 자동화 편향이란, 임상의가 인간이 발견했어야 할 오류를 간과하는 경우를 말합니다. 또한 의사와 보건의료 종사자가 윤리적 또는 도덕적 고려사항이 상충되는 결정을 내릴 때 LMM을 사용할 수 있다는 우려도 있습니다*(20)*. Chat GPT와 같은 LMM은 도덕적 조언자로서 매우 일관성이 없을 수 있지만, 최근 실험에 따르면 사용자들이 챗봇의 조언을 받고 있다는 사실을 알고 있음에도 불구하고 도덕적 판단에 영향을 줄 수 있습니다*(40)*. LMM을 도덕적 판단에 사용하면 의사들이 어려운 판단이나 결정을 내리지 못하게 되는 "도덕적 탈숙련화(moral de-skilling), 즉 도덕적 역량 상실"로 이어질 수 있습니다*(20)*.

- ***기술 저하***: 의료행위에서 AI 사용이 증가함에 따라 의사들이 점점 더 일상적인 책임과 의무를 컴퓨터에 의존하게 되면서, 의료 전문가로서의 역량이 저하되거나 약화될 장기적인 위험이 있습니다. 기술의 상실은 의사가 알고리즘의 결정을 자신 있게 무효화하거나 도전하지 못하게 하거나, 네트워크 장애나 보안 침해가 발생했을 때 의사가 특정 의료 작업과 절차를 완료하지 못하게 할 수 있습니다*(1)*.

- ***사전동의***: LMM의 사용이 대면 뿐만 아니라 특히 비대면 환경에서 증가함에 따라, 환자에게 AI 기술이 응답을 지원하거나 궁극적으로 임상의의 피드백을 모방한 응답을 생성할 수 있다는 사실을 알리는 것이 필요합니다. 그러나 LMM 및 기타 형태의 AI가 일반적인 의료행위에 통합될 경우, 환자나 보호자는 AI 기술에 전적으로 또는 부분적으로 의존하는 것이 불편하거나 원하지 않더라도, 사용에 대한 동의를 거부할 수 없게 될 가능성이 있습니다. 이는 특히 AI에 기반하지 않는 다른 옵션을 쉽게 이용할 수 없거나, 해당 기능의 책임을 컴퓨터에 넘긴 임상의가 AI를 사용하지 않고 의료 서비스를 제공할 수 없는 경우에 해당합니다.

2.2 환자중심의 응용

AI는 환자들이 자신의 질병을 관리하는 방식을 변화시키고 있습니다. 현재 환자들은 약물 복용, 영양과 식단 개선, 신체 활동 참여, 상처 관리, 주사 투여 등 스스로 건강 관리를 위한 상당한 책임을 지고 있습니다. AI 도구는 챗봇, 건강 모니터링 및 위험 예측 도구, 장애인을 위한 시스템을 포함하여 자가 관리의 범위를 확대할 것으로 전망됩니다*(1)*.

LMM은 환자와 일반 대중이 의료 목적으로 AI를 사용하는 추세를 가속화할 수 있습니다. 지난 20년 동안 개인들은 인터넷 검색을 통해 의료 정보를 얻어왔습니다. 이에 따라 LMM은 인터넷 검색에 통합되어 환자와 대중에게 정보를 제공하는 데 중요한 역할을 할 수 있습니다. 대규모 언어 모델기반 챗봇은 자가 진단 및 병원 방문 전 정보를 찾는 데 있어 검색 엔진을 대체할 수 있습니다*(41)*.

Box 3. 장애인을 위한 LMM과 관련된 윤리적 고려사항

과거에는 장애인이 직장, 교육 시스템 및 적절한 의료 지원에서 배제되었으며*(56)*, 이는 AI 시스템을 훈련시키는 데 사용된 데이터 세트에서도 마찬가지였습니다. 이러한 시스템은 얼굴 비대칭, 다양한 몸짓, 의사소통 방식, 행동 및 행동 패턴을 가진 사람들을 차별할 가능성이 있습니다. 가장 심각하게 영향을 받는 그룹은 장애인, 인지적 또는 감각적 장애를 가진 사람들, 그리고 자폐 스펙트럼 장애를 가진 사람들입니다*(57)*.

이러한 편향과 배제는 생성형 AI에도 적용될 수 있습니다. 예를 들어, LMM은 환자의 설명이나 이력에서 "장애"와 관련된 키워드 또는 문구에 부정적인 의미나 감정을 부여할 수 있습니다*(58)*. 챗봇은 장애인을 "살아있지 않음", "인간이 아님", "정서적으로 둔감"하다고 인식할 수 있는데, 이는 다른 행동이나 행동 패턴 때문일 수 있습니다. 음성 인식 시스템은 언어 장애가 있는 사람들에게 덜 정확하게 작동하여 오해를 일으킬 가능성이 있습니다.

장애와 관련된 편향을 해결하고 극복하기 위해서는 AI 개발 과정 전반에 걸친 개입이 필요합니다. AI 시스템의 개발 및 설계에 장애인을 포함시키고, 데이터 세트와 AI 시스템의 성능에서 장애 편향을 평가하는 감사를 실시하며, 장애인의 권리를 보호하고 증진하기 위한 법률이 AI 기술과 관련된 문제를 고려하도록 해야 합니다. 또한, AI 기반 시스템의 사용 증가로 인해 발생할 수 있는 문제와 장벽을 해결하기 위한 법률과 정책을 제정해야 합니다. AI 관련 법률에는 장애 스펙트럼 및 조건이 AI 시스템에 미치는 영향을 포함하는 "장애 특정적(disability-specific)" 범주화가 포함될 수 있습니다.

LMM 기반 챗봇은 다양한 형태의 데이터를 활용하여 매우 개인화되고 폭넓은 초점의 가상 건강 비서로 활용될 수 있습니다. 한 연구에 따르면 "가상 건강 조언자는 개인 프로필을 활용하여 행동 변화를 촉진하고, 건강 관련 질문에 답하며, 증상을 분류하거나 적절한 경우 의료 제공자와 소통할 수 있다"고 합니다*(3)*. 특정 LMM 기반 챗봇은 예를 들어 정신 건강 분야에서 치료를 제공할 수 있습니다*(2)*.

환자 중심 LMM의 세 번째 응용 분야는 임상 시험을 식별하거나 그러한 시험에 참여하는 것입니다*(28)*. AI 기반 프로그램은 이미 환자와 임상 시험 연구자가 적합한 시험을 찾는 것을 돕고 있으며*(42)*, LMM은 환자의 관련 의료 데이터를 활용하여 동일한 방식으로 사용될 수 있습니다*(28)*. 이는 모집 비용을 줄이고 속도와 효율성을 높이며, 다른 채널을 통해 찾기 어렵고 접근하기 힘든 적절한 시험과 치료 기회를 개인에게 더 많이 제공할 수 있습니다*(42)*.

위험과 도전 과제

개인이 LMM을 쉽게 사용할 수 있다는 점은 아래와 같은 중대한 위험을 예고할 수 있습니다.

- ***부정확하거나 불완전하거나 잘못된 정보***: 임상의 및 보건의료 전문가가 LMM을 사용하는 경우와 마찬가지로, 환자 및 일반인이 LMM을 사용하는 경우에도 부정확하거나 편향된, 불완전하거나 잘못된 진술의 위험이 있습니다. 이는 의료 정보를 제공한다고 주장하는 AI 프로그램에서도 마찬가지입니다. 의료 전문 지식이 없는 사용자가 LMM을 사용할 경우, 응답을 검토하거나 반박할 근거가 없거나, 다른 정보 출처에 접근할 수 없는 상황, 또는 어린이가 이를 사용하는 경우 이러한 위험이 더욱 커집니다(박스 2). 사람들은 수십 년 동안 인터넷 검색을 통해 의료 정보를 얻어왔지만, LMM은 다른 LMM(동일한 위험을 지닌)을 참조하여 빠르게 비교할 수 있는 응답을 제공하면서도 "정확한" 것으로 보이는 답변을 제시할 수 있습니다.

- ***조작:*** 많은 LMM 기반 챗봇 응용 프로그램은 대화 방식에서 서로 다른 접근방식을 취하며, 이들은 점점 더 설득력 있고 중독성이 강해질 것으로 예상됩니다*(46)*. 챗봇은 사용자 각각에 맞춘 대화 패턴을 채택할 수 있을 정도로 발전할 수 있으며*(41)*, 질문에 답변하거나 대화에 참여하여 개인이 자신에게 불이익이 되거나 건강에 해로운 행동을 하도록 유도할 수 있습니다*(12)*. 여러 전문가들은 챗봇이 “감정적으로 조작적”이 될 가능성이 있다며*(47,48)* 이러한 부정적 결과를 관리하기 위한 긴급한 조치를 촉구하고 있습니다. 한 사례로는 벨기에에서 불안을 겪던 한 사람이 챗봇과 6주 동안 집중적으로 대화한 후 자살한 사건이 크게 보도되었습니다*(49)*.

- ***개인정보:*** 환자와 일반인이 LMM을 사용하는 경우 그들이 공유하는 개인정보가 보호되지 않을 수 있으며, 사적인 사용이 보장되지 않을 수 있습니다. 다른 목적으로 LMM을 사용하는 사용자들이 회사 기밀 정보와 같은 민감한 정보를 공유하는 경향이 있습니다*(50)*. LMM에 공유된 데이터가 반드시 삭제되는 것은 아니며, AI 모델 개선을 위해 사용될 수 있습니다*(50)*. 비록 나중에 회사 서버에서 데이터가 삭제될 수 있지만, 법적 근거 없이 사용될 가능성이 있습니다*(51)*. 이와 관련된 문제로는 다른 LMM 사용자와의 정보 공유가 있습니다. 이는 다른 사용자가 LMM을 통해 특정 정보를 요청하거나*(52)*, 한 LMM을 여러 사람과 공유할 경우 다른 사람들의 채팅기록(대화내용 이외의 정보 포함)이 잘못 공개되는 경우*(53)* 발생할 수 있습니다. 따라서 개인의 식별 가능한 의료 정보가 LMM에 입력되면 제3자에게 공개될 가능성이 있습니다*(54)*.

- ***임상의와 환자, 일반인 간의 상호작용의 약화:*** 환자나 그들의 보호자가 LMM을 사용하는 것은 의사-환자 관계를 근본적으로 변화시킬 수 있습니다. 지난 20년 동안 환자들이 인터넷 검색을 통해 의료 정보를 얻는 것이 증가하면서 이러한 관계는 이미 변화했습니다. 환자들은 자신이 얻은 정보를 바탕으로 의료 제공자에게 도전하거나 추가 정보를 요구할 수 있습니다. LMM은 이러한 대화를 개선할 수 있지만, 환자나 보호자가 LMM에

의존하여 예후 및 치료를 결정함으로써 전문적인 의학적 판단과 지원에 적절히 의존하지 않거나 이를 완전히 배제할 수 있습니다. 관련된 우려는 AI 기술이 제공자와 환자 간의 접촉을 줄이면, 임상의가 건강 증진을 도모할 기회가 줄어들고 사람 간의 상호작용이 약화되어 취약한 상황에서 일반적인 지지의료(supportive care)가 약화될 수 있다는 점입니다*(1)*. 일반적으로, AI로 인해 임상진료가 "비인간화"될 수 있다는 우려가 있습니다.

- ***인식적 부정의:*** 보건의료 제공자의 판단을 LMM의 판단으로 대체할 경우 발생할 수 있는 추가적인 잠재적 결과 중 하나는 환자에게 "인식적 부정의(epistemic injustice)"를 초래할 수 있다는 점입니다. "인식적 부정의"란 "지식의 주체로서의 특정 사람에게 가해지는 부당함"을 의미하며, 이는 의료 시스템 내에서 환자에게 발생할 수 있습니다*(55)*. 인식적 부정의의 한 형태인 해석적 부정의(hermeneutic injustice)는 공유된 이해와 지식(소위 "집단적 해석 자원")의 격차로 인해 일부 사람들이 자신의 경험, 사회적 경험, 또는 보건의료의 경우 자신의 신체적 또는 정신적 상태를 이해하는 데 불이익을 받게 되는 상황을 말합니다*(55)*. LMM은 대규모 데이터에 기반해 학습되었더라도 인식하고 반응할 수 있는 범위, 즉 자신이 가진 어휘를 초과하는 개념과 관념에 한계가 있습니다. 만약 환자의 경험이 임상 환경에서 LMM에 의해 인정되지 않거나 인식되지 않는다면, 이는 의료 제공자의 적절한 치료를 방해할 수 있으며, 이는 환자에게 해를 끼칠 수 있습니다. 이러한 상황은 특히 장애인을 포함하여 이미 데이터에서 소외되고 대표성이 부족한 취약 계층에서 더 자주 발생할 가능성이 큽니다*(55)*(Box 3).

- ***의료 시스템 외부에서의 보건의료 제공:*** AI 기술의 발전으로 인해 건강 관련 AI 응용 프로그램은 더 이상 의료 시스템 내에서만 독점적으로 사용되거나 가정에서만 이용되지 않습니다. 이제 이러한 기술은 비의료 시스템 기관에서도 쉽게 획득하여 사용할 수 있으며, LMM을 대중적으로 제공하는 기업과 같이 비의료 시스템에 의해 도입되기도 합니다. 이 때, 이러한 기술이 더 높은 규제검토가 요구되는 임상 응용 프로그램으로 규제되야 할지(더 강력한 규제 감독이 필요함), 아니면 "웰니스(wellness)"와 같은 비교적 낮은 규제와 검토가 요구되는 응용 프로그램으로 규제해야 할지에 대한 의문이 제기됩니다. 현재 이러한 기술은 두 범주 사이의 회색 지대에 속한다고 볼 수 있습니다.

"경미하게" 규제되는 LMM은 환자가 규제 안전장치 없이 사용할 경우 위험을 초래할 수 있습니다. 이는 LMM을 의학적 조언이나 자가 진단에 사용하는 경우를 포함합니다. 환자가 잘못된 또는 오해를 불러일으킬 수 있는 조언을 받을 가능성이 있으며(위 참조), 특히 자살 충동을 가진 개인이 AI 챗봇을 사용할 때 지지의료가 부족한 상황에서 환자 안전이 위협받을 수 있습니다. 설령 정보가 정확하더라도, 의료 교육을 받지 않은 개인이 해당 정보를 자가 진단에 사용하면 이를 잘못 해석하거나 오용할 수 있습니다. 이러한 응용 프로그램(LMM 포함)이 계속 확산되고 있으며 반드시 의료 응용 프로그램으로 분류되는 것은 아니기 때문에 전반적인 의료 서비스의 질이 저하될 가능성이 있습니다. 특히 다른 대안이 부족한 사람들이 이러한 응용 프로그램에 의존하게 되면 양질의 의료 서비스에 대한 접근성이 더욱 불평등해질 수 있습니다*(1)*.

2.3 사무기능 및 행정업무

LMM은 의료 전문가의 의료 업무에서 문서 작업, 행정 업무 및 재정 관리 업무를 보조하기 위해 사용되기 시작했습니다. 의학 분야에서는 환자 정보와 데이터를 전자 건강 기록에 기록하고, 민간, 보험, 공공 의료 시스템의 청구서를 처리하며, 기타 행정 업무를 수행해야 하는 여러 의무로 인해 점점 더 많은 문서 작업이 필요합니다. 전자 건강 기록을 작성하는 등의 많은 의무는 의료 전문가를 "해방" 하는 것을 목표로 했지만, 현재는 의사와 의료 종사자들의 주요 번아웃 원인이 되었습니다*(59)*. 한 연구에 따르면, 문서 작업은 의사의 시간의 4분의 1에서 절반, 간호사의 시간의 5분의 1을 차지했습니다*(59)*.

LMM은 보건의료 전문가들에게 가장 귀중한 자원인 시간을 되돌려주는 수단으로 주목받고 있으며, 이는 번아웃을 줄이고 각 환자 치료에더 많은 시간을 할당하거나 더 많은 환자를 볼 수 있도록 합니다. LMM이 내재된 소프트웨어를 사용해 환자 방문 기록을 작성한 한 의사는 "AI 덕분에 의사로서 환자에게 100% 집중할 수 있게 되었다"고 말하며, 이 소프트웨어가 하루에 최대 2시간을 절약했다고 보고했습니다*(60)*.

LMM의 현재의 사용사례 및 예상되는 사례는 아래와 같습니다.

- 의료 용어를 단순화하고 의사소통을 더 "환자 친화적"으로 만들어 번역 또는 임상의-환자간의 의사소통 개선에 도움을 줌*(34)*;
- 전자 건강 기록에서 누락된 정보를 채우는 데 활용*(61)*;
- 기타 AI 형태와 함께 환자 방문 후(가상 또는 대면) 임상 기록 초안을 작성*(62)*;
- 자동 처방전 작성, 예약 관리, 검사 일정 조율, 청구 코드, 보험사의 사전 승인 처리, 시술 노트 및 퇴원 요약서 작성 등의 작업을 사전에 수행*(5)*;
- 더 정교한 LMM이 개발됨에 따라, 방사선 전문의와 같은 복잡한 문서 작업에도 활용될 가능성이 있음. 예를 들어, "이 모델은 환자의 병력을 고려하며 이상 소견과 관련된 정상 소견을 모두 기술하는 방사선 보고서를 자동으로 작성하고, 텍스트 보고서와 대화형 시각화를 병행하여 각 단계별로 설명된 부분을 강조함으로써 임상의에게 추가적인 도움을 제공"할 수 있음*(5)*.

위험과 과제

LMM의 다른 용도와 마찬가지로 부정확성, 실수(예: 필사, 번역 또는 단순화) 또는 "환각(hallucination, 할루시네이션)"으로 인해 심각한 오류가 발생할 수 있습니다. 따라서 대부분의 사무 및 행정 기능이 완전히 자동화되지 않는 것이 중요합니다. 감독 및 검토가 보건의료 전문가의 시간을 차지하더라도 LMM은 여전히 현상 유지보다 부담이

적을 가능성이 높습니다. 또 다른 문제는 LMM이 일관되지 않을 수 있다는 것입니다. 프롬프트나 질문을 약간 변경하면 완전히 다른 전자 건강 기록이 생성될 수 있지만 불일치는 시간이 지남에 따라 감소할 것으로 예상됩니다*(63)*.

2.4 의학 및 간호교육

LMM은 의료 및 간호 교육에서도 중요한 역할을 할 것으로 예상됩니다. LMM은 학생의 특정 요구와 질문에 맞춰 일반적인 텍스트와 비교해 맞춤형으로 "동적 텍스트"를 생성하는 데 사용될 수 있습니다*(63)*. 챗봇에 통합된 LMM은 임상의-환자 간의 의사소통 및 문제 해결 능력을 향상시키기 위해 모의 대화를 제공할 수 있으며, 이는 의료 면담, 진단적 추론 및 치료 옵션 설명 연습에 활용될 수 있습니다. 또한 챗봇은 장애가 있는 환자나 희귀 질환을 가진 환자를 포함한 다양한 가상의 환자를 학생들에게 제공하도록 맞춤화될 수 있습니다. LMM은 또한 "연쇄적 사고(Chain-of-Thought)"를 통해 생리학적 및 생물학적 과정을 설명하며 의대생의 질문에 답변하는 형태로 교육을 제공할 수 있습니다*(63)*.

위험과 도전 과제

AI 사용은 보건의료 전문가의 훈련과 기술을 향상시킬 수 있지만, 동시에 전문가가 자신의 판단(또는 동료의 판단)을 컴퓨터의 판단에 의존하게 되는 위험을 초래할 수 있습니다. LMM이 부정확한 정보나 응답을 제공하거나 응답을 조작할 경우, 이는 의료 교육의 질에 영향을 미칠 수 있습니다.

또 다른 우려는 교육에서 LMM의 사용 또는 행정 및 사무 기능을 단순화하는 과정이 디지털 문해력이 부족한 의료 종사자들에게 추가적인 부담을 줄 수 있다는 점입니다. 이들은 일상 업무에서 AI 지원 기술 사용에 대한 새로운 역량을 개발해야 할 필요가 있습니다*(1)*. LMM의 새로운 기능은 보건의료 전문가가 지속적으로 재교육하고 조정해야 할 것으로 예상됩니다*(1)*. 개발자들은 궁극적으로 일반인도 쉽게 사용할 수 있는 자연어나 시각과 같은 커뮤니케이션 인터페이스를 갖춘 AI 지원 기술을 도입할 수 있을 것입니다.

2.5 과학 및 의학 연구 및 약물 개발

AI는 이미 과학 및 임상 연구와 신약 개발에 사용되고 있습니다. AI를 통해 전자 건강 기록을 분석하여 임상 실무 패턴을 식별하고 새로운 임상 실무 모델을 개발할 수 있습니다. 머신러닝은 또한 질병에 대한 이해를 향상시키고 새로운 바이오마커를 식별하는 등 유전체학 분야에서 사용됩니다*(1)*. AI는 신약 개발 주기의 거의 모든 단계에서 활용되며, 화합물 선별 과정을 간소화하고, 단백질의 3차원 구조를 예측하며("단백질 접힘 문제")*(1)*, 전임상 개발 단계에서 화합물의

독성과 효과를 예측하고, 임상 시험 중 참가자 모집, 등록 및 모니터링을 개선하는 데 사용됩니다.[4]

LMM은 AI가 신약 개발 뿐만 아니라, 과학 및 의학 연구를 지원할 수 있는 방법에도 변화를 주고 있습니다. LMM은 다양한 과학 연구의 측면에서 사용될 수 있습니다. 예를 들어, 과학 논문 작성, 논문 제출 또는 동료 검토 작성에 사용할 텍스트를 생성할 수 있습니다*(34)*. 학술 논문을 요약하거나 초록 생성에 사용될 수 있으며, 데이터를 분석하고 요약하여 임상 및 과학 연구에서 새로운 통찰을 얻는 데에도 활용될 수 있습니다. 또한 LMM은 텍스트를 편집하여 논문이나 연구계획서와 같은 문서의 문법, 가독성 및 간결성을 개선할 수 있습니다. 한 LMM은 수백만 개의 학술 논문으로 학습되어 과학 연구를 분석하고 질문에 답하거나 정보를 추출하거나 관련 텍스트를 생성할 수 있는 것으로 평가됩니다*(64)*. LMM은 신약 발견과 특히 새로운 화합물을 설계하는 de novo 신약 설계에도 사용됩니다. 여기에는 특정 속성을 가진 화합물을 개발하는 작업이 포함됩니다*(65)*.

위험과 도전 과제

의료 및 과학 분야의 주요 학술 저널은 이미 LMM의 등장, 잠재력, 과학 연구에 미치는 영향을 인식하고 이에 대응하고 있습니다. 예를 들어, 한 학술 출판사는 다음 두 가지 규칙을 제정했습니다: (1) 연구 논문의 저자로 LMM을 인정하지 않으며, (2) LMM을 사용하는 연구자는 방법론 섹션과 감사의 글에 그 사용을 명시해야 합니다*(66)*. 또한, 세계의학편집자협회(World Association of Medical Editors)는 저자의 자격을 인간으로 제한했습니다*(67)*.

과학 연구에서 LMM의 사용과 관련된 일반적인 우려는 다음과 같습니다:

- ***책임 부족***: 과학 또는 의학 연구 논문의 저자는 책임성을 가져야 하지만, AI 도구는 이를 가질 수 없습니다*(66)*. 이 책임감 부족은 주요 학술 출판사와 세계의학편집자협회가 LMM을 저자로 인정하지 않기로 한 결정한 근거가 되었습니다.

- ***고소득 국가 편향***: LMM을 학습시키는 데 사용되는 과학 및 의학 연구 대부분이 고소득 국가에서 수행됩니다. 따라서 LMM의 쿼리 결과는 고소득 국가의 관점을 편향적으로 반영할 가능성이 높습니다*(34)*. 이는 저소득 및 중소득 국가에서 나온 연구, 특히 비라틴 문자로 작성된 출판물을 무시하거나 인용하지 않는 경향을 강화하고 악화시킬 수 있습니다*(68)*.

- ***환각(hallucination, 할루시네이션) 또는 잘못된 정보***: LMM은 존재하지 않는 학술 논문이나 기타 정보를

4 약물 개발에서 인공지능의 윤리와 거버넌스. 제네바: 세계보건기구 WHO, 출간될 예정.

요약하거나 인용하며 "환각(hallucination, 할루시네이션)"을 일으킬 수 있습니다*(69)*.

- ***신뢰 저하***: 동료 검토 생성과 같은 활동에 LMM을 사용하는 것은 이 과정에 대한 신뢰를 약화시킬 수 있습니다*(69)*.

- ***LMM 및 LMM이 생성한 지식에 대한 접근성***: 과학 및 의학 연구에 사용되는 다른 도구, 기술 및 정보와 마찬가지로, LMM도 페이월(paywall)에 자리 잡을 가능성이 높습니다. 이는 디지털 및 지식 격차를 악화시키며, 과학 및 의학 연구에 참여하고자 하는 재정적 지원이 부족한 과학자들에게 부정적인 영향을 미칠 수 있습니다*(34)*.

AI(LMM 포함)는 신약 개발에 혜택을 제공할 수 있지만, 이 분야에서 AI 사용에 대한 우려도 존재하며, 이는 곧 발표될 WHO 출판물에서 검토될 예정입니다.

3 보건 시스템 및 사회에 대한 위험과 LMM 사용에 대한 윤리적 문제

LMM과 관련된 많은 위험과 우려는 보건의료 전문가, 환자, 연구자 또는 간병인과 같은 개별 사용자에게 영향을 미칠 수 있지만, 시스템적 위험도 초래할 수 있습니다. 의료 분야에서 LMM 및 기타 AI 기반 기술의 사용과 관련된 새로운 또는 예상되는 위험은 (i) 국가의 보건 시스템에 영향을 미칠 수 있는 위험, (ii) 규제 및 거버넌스와 관련된 위험, (iii) 국제적인 사회적 우려의 세 가지 범위로 나눌 수 있습니다.

3.1 보건 시스템에 대한 영향

보건 시스템은 서비스 제공, 의료인력, 보건 정보 시스템, 필수 의약품 접근성, 재정, 리더십 및 거버넌스의 여섯 가지 구성요소를 기반으로 합니다*(70)*. LMM은 이러한 구성 요소에 직간접적으로 영향을 미칠 수 있습니다. 보건 시스템에 영향을 미칠 수 있는 LMM 사용과 관련된 위험은 다음과 같이 설명됩니다.

LMM의 이점의 과대평가 및 위험의 간과

일부 사람들은 AI의 가능성을 과장하거나 과대평가하는 경향이 있으며, 이는 안전성과 유효성에 대해 철저한 평가를 거치지 않은 검증되지 않은 제품과 서비스의 채택으로 이어질 수 있습니다*(1)*. 이는 부분적으로 "기술적 해결주의"의 지속적인 매력때문이며, 이 접근방식에서는 AI 및 LMM과 같은 기술이 더 깊은 사회적, 구조적, 경제적, 제도적 장벽을 제거하는 "만능 해결책"으로 간주되지만, 그러한 기술이 실제로 유용하고 안전하며 효과적이라는 것이 입증되기 전에 이러한 기대가 형성됩니다*(1)*.

LMM은 새롭고 검증되지 않은 기술로, 위에서 언급한 바와 같이 사실을 생성하는 것이 아니라 사실과 유사한 정보를 생성하며, 이는 부정확할 수 있습니다. LMM은 소비자, 정치, 대중의 큰 관심을 받았지만, 정책 입안자, 제공자 및 환자가 그 이점을 과대평가하고 LMM이 초래할 수 있는 문제와 과제를 간과하게 만들 수 있습니다. 정책 입안자는 LMM이 이미 개발되고 사용하기 전까지는 이를 얼마나 광범위하게 사용해야 하는지 판단하는 데 필요한 증거를 확보하는 것이 어려울 수 있습니다. LMM의 사용이 이미 사용 중인 AI 기반 기술이나, 자금이 부족하고 충분히 활용되지 못하고 있지만, 입증된 치료적 또는 공중 보건적인 이점을 가진 비 AI 또는 비 디지털 솔루션보다 우선시 되어서는 안 됩니다. 특히 저소득 및 중소득 국가에서, 검증된 치료적 또는 공중 보건상의 이점을 가진 기술이 자금 부족 및 과소 활용되고 있는 상황에서 LMM에 대한 비균형적인 의료 정책 및 잘못된 투자는 주의를 분산시키고 자원을 검증된 개입에서 빼앗아 갈 수 있습니다. 이는 보건부가 의료비를 줄이라는 압박을 받는 상황에서 더 큰 문제를 초래할 수 있습니다*(1)*.

접근성과 경제성

LMM이 제공자와 환자에게 혜택을 줄 수 있음에도 불구하고, 이를 공평하게 접근할 수 없게 만드는 여러 요인이 존재합니다. 그 중 하나는 디지털 격차로, 특정 국가, 지역, 또는 인구의 일부로 디지털 도구 사용이 제한됩니다. 디지털 격차는 AI 사용에 영향을 미치는 다른 불평등을 초래하며, AI 자체가 이러한 불평등을 강화하고 악화시킬 수 있습니다. 또 다른 요인은 LMM이 인터넷과 달리 많은 경우 사용료나 구독료를 지불해야만 이용할 수 있다는 점입니다. LMM의 개발과 운영이 비용이 많이 들기 때문입니다. 예를 들어, ChatGPT를 운영하는 데 하루 약 70만 달러가 소요되는 것으로 추정됩니다*(71)*. 일부 기업은 새로운 버전의 LMM에 대해 구독료를 도입하고 있습니다*(72)*. 이는 저소득 및 중소득 국가뿐 아니라, 자원이 부족한 환경에 있는 고소득 국가의 개인, 의료 시스템, 또는 지방 정부에서도 특정 LMM을 이용하기 어렵게 만들 수 있습니다*(54)*. 반면, 모든 국가에서 경제적으로 어려운 사람들은 "가성비가 좋은 해결책"으로 LMM에 의존할 수밖에 없으며, "진짜" 보건의료 전문가에게 접근할 수 있는 기회는 부유한 사람들만의 특권이 될 가능성이 있습니다. 세 번째 요인은 대부분의 LMM이 현재 영어로만 운영된다는 점입니다. 따라서, 다른 언어로 입력을 받고 출력할 수는 있지만, 이러한 경우 잘못된 정보나 허위 정보를 생성할 가능성이 높아집니다*(73)*.

시스템 전반의 편향

앞서 언급한 바와 같이, AI 모델을 학습시키는 데 사용된 데이터 세트는 여성, 소수 인종, 노인, 농촌 지역 사회, 취약 계층을 배제하는 등 편향되는 경우가 많습니다. 일반적으로 AI는 데이터가 가장 많은 인구를 중심으로 학습하기 때문에, 불평등한 사회에서는 AI가 소수 인구를 불리하게 할 수 있습니다*(1)*. 모델의 규모가 클수록 편향이 증가할 가능성이 있으며*(74)*, 이에 점점 더 많은 데이터를 사용하는 LMM에서 특히 문제가 될 수 있습니다. 이는 소위 더 작은 LMM이 개발되고 있다고 하더라도 연속적으로 개발되는 모델에서 사용되는 데이터가 지속적으로 증가하기 때문입니다 이러한 편향은 의료 시스템 전반에 걸쳐 차별을 초래할 수 있으며, 사람들의 기본 재화, 특히 보건 서비스 및 고품질 치료에 대한 접근에 영향을 미칠 수 있습니다*(75)*. 동시에, LMM은 편향과 고정관념에 "반발"하는 데이터를 포함할 가능성도 있습니다. 연구자들은 모델이 고정관념에 의존하지 않도록 프롬프트를 설정했을 때 알고리즘의 응답이 상당히 긍정적으로 변화하는 것을 발견했습니다*(74)*.

노동과 고용에 미치는 영향

한 투자은행은 LMM이 결국 최소 3억 개의 일자리 상실(또는 "질적 저하")을 초래할 것이라고 추정했습니다*(76)*. 경제협력개발기구(OECD)는 회원국에서 AI 기반 자동화로 인해 가장 큰 영향을 받을 직업은 고도로 숙련된 직업이며, 이는 LMM의 사용과 관련이 크다고 언급했습니다. 특히 "금융, 의학, 법률 활동 분야의 직업이 AI로 인해 갑작스럽게 자동화의 위험에 처할 수 있다"고 분석했습니다*(77)*. 하지만 많은 국가에서 의료는 산업이 아니라 정부의 핵심 기능이며, 의료 종사자가 기술로 대체되지 않을 가능성이 큽니다. 또한 많은 국가들이 여전히 보건의료 종사자들의 부족 문제를

겪고 있으며 이는 코로나19 팬데믹 이후부터 지속되고 있습니다*(1)*. 세계보건기구(WHO)는 2030년까지 주로 저소득 및 중소득 국가에서 1천만 명의 의료 종사자가 부족할 것으로 추산하고 있습니다*(79)*. 따라서 안전성과 유효성이 입증된 LMM은 필요한 의료 인력과 실제 이용 가능한 인력 간의 격차를 좁히는 데 사용될 수 있습니다.

또 다른 우려는 LMM의 도입이 현재 및 미래의 보건의료 전문가 수에 미칠 영향입니다. 한 주요 기술 기업은 최대 80%의 직업이 AI 도입으로 영향을 받을 수 있다고 추정했습니다*(80)*. 컨설팅 회사인 액센츄어(Accenture)는 LMM이 작업 시간의 40%에 영향을 미칠 수 있다고 추정하며, "인간의 창의성과 생산성에 미칠 긍정적 영향이 막대할 것"이라고 낙관적으로 평가했습니다*(81)*. 그러나, 위에서 언급한 바와 같이 LMM의 도입은 많은 의료 종사자에게 상당한 도전 과제를 초래할 수 있으며, 이들은 LMM에 대한 교육과 적응이 필요합니다. 의료 시스템은 제공자에게 제기되는 도전 과제와 환자 및 간병인에게 발생할 수 있는 위험을 고려해야 합니다.

세 번째 우려는 LMM 학습에 사용되는 데이터 세트를 검토하고 주석을 달거나 학대, 폭력 또는 정신적으로 고통스러운 콘텐츠를 제거하는 책임을 맡은 사람들에게 가해지는 정신적, 심리적 부담입니다. 이러한 작업을 담당하는 사람들은 종종 저소득 및 중소득 국가에 기반을 두고 있으며, 낮은 임금을 받고 일하면서 상담이나 기타 의료 서비스에 접근하지 못한 채 심리적 고통을 겪을 가능성이 있습니다*(73)*.

부적합한 LMM에 대한 보건보건 시스템의 의존

LMM은 지속적인 의료 종사자의 부족 문제를 완화하고 보건 시스템의 범위를 확장할 수 있는 잠재력을 가지고 있지만, 동시에 보건 시스템이 LMM, 특히 산업에서 개발된 LMM 기술에 과도하게 의존하게 될 수도 있습니다. 따라서, 의료 또는 공중 보건에 사용되는 LMM이 유지되지 않거나, 성능이 저하되거나, 고소득 국가의 맥락에서만 설계 및 업데이트될 경우, 이러한 LMM에 의존하던 보건 시스템은 LMM 없이 의료 서비스를 제공하기 위해 적응해야 할 수도 있습니다. 의료 전문가들이 "기술숙련 저하(de-skilled)"를 겪고 특정 책임을 AI에 의존하게 되었거나, 환자들이 AI를 사용하기를 기대하는 경우, LMM에 대한 과도한 의존이 개인 및 사회의 의료 시스템에 대한 신뢰를 악화시킬 수 있습니다. 이와 관련된 또 다른 위험은, LMM이 환자의 개인정보와 기밀성을 보장하지 못할 경우, LMM에 대한 과도한 의존이 개인 및 사회적 신뢰를 저하시킬 수 있다는 점입니다. 사람들이 의료 서비스를 이용할 때 개인정보 침해 위험 없이 접근할 수 있다는 확신을 잃게 될 것이기 때문입니다.

사이버 보안 위험

보건의료 시스템이 AI에 점점 더 의존하게 되면서, 이 기술들은 악의적인 공격과 해킹의 표적이 될 수 있으며, 일부 시스템이 정지되거나, 알고리즘 학습에 사용되는 데이터가 조작되어 성능 및 권장사항 등이 변경될 수 있습니다. 또한

데이터가 랜섬웨어로 “탈취”될 위험도 있습니다*(1)*. 위에서 언급한 특히 중요한 보안 위험 중 하나는, 민감한 데이터를 무단으로 공개하거나 사용되는 일로부터 보호하지 못하는 LMM에 입력하는 경우입니다. LMM 자체도 “프롬프트 인젝션(Prompt injection)”과 같은 사이버 보안 위험에 취약할 수 있습니다. 이는 제3자가 LMM에 데이터를 입력하여 개발자가 의도하지 않은 방식으로 작동하게 만드는 공격입니다*(82)*. 예를 들어, 데이터베이스 질문에 답변하도록 설계된 LMM이 데이터베이스의 정보를 삭제하거나 변경하도록 지시할 수 있습니다. 이 결함을 해결할 수 있는 알려진 해결책은 아직 없습니다. 현재 프롬프트 인젝션(Prompt injection)은 보안 연구자들에 의해 LMM의 과제를 설명하는 데 사용되고 있지만, 악의적인 행위자가 데이터를 훔치거나 사용자를 속이는 데 이를 악용할 수 있습니다*(83)*.

3.2 규제 및 법적 요구사항 준수

AI 사용을 규제하기 위한 새로운 법률이 제정될 수 있지만, 특정 기존 법률과 규정, 특히 데이터 보호법 및 국제 인권 의무는 LMM의 개발, 제공 및 배포에 적용될 수 있습니다. 현재 개발되어 대중적으로 사용되는 일부 LMM은 자동화된 의사 결정으로부터 보호를 포함한 다양한 권리를 보장하는 유럽연합의 일반 데이터 보호 규정(European Union’s General Data Protection Regulation)과 같은 주요 데이터 보호법을 위반할 가능성이 있습니다*(84)*. 이러한 권리, 보호 및 요구사항은 AI 개발의 가이던스가 되어야 합니다*(85)*.

이러한 위반사항 중 일부는 유럽연합 회원국*(83)* 및 캐나다*(86)*와 같은 지역에서 LMM에 대한 조사를 초래했습니다. 위반사항은 다음과 같습니다: (i) LMM이 인터넷에서 개인의 데이터를 무단으로 수집하고 사용했으며, 이러한 데이터 수집에 대해 ‘정당한 이익’이 없었습니다*(87)*, (ii) LMM은 사람들이 자신의 데이터 사용 여부를 알거나, 실수를 수정하거나, 데이터를 삭제할 권리("잊힐 권리"), 또는 데이터 사용에 반대할 권리를 보장하지 못합니다*(87)*, (iii) LMM은 챗봇이나 기타 소비자 인터페이스에 제공된 민감한 데이터 사용에 대한 투명성을 충분히 확보하지 못했습니다. 법적으로는 사용자가 채팅 기록 데이터를 삭제할 수 있어야 합니다*(83)*, (iv) LMM은 13세 미만 사용자와 부모 동의를 받지 않은 13~18세 사용자에 대한 필터링 시스템(“연령 게이팅”)을 보유하지 않았습니다*(88)*, (v) LMM은 개인정보 유출을 방지하지 못합니다*(87)*, (iv) LMM이 부분적인 환각(hallucination, 할루시네이션)으로 인해 부정확한 개인 정보를 게시했습니다*(89)*. 일반 데이터 보호규정의 또 다른 잠재적인 위반 사례로는 "설명을 요구할 권리" 요구사항과 관련이 있습니다. 이 요구사항에 따르면, 개인 데이터를 자동화된 처리에 사용하는 시스템(LMM 등)은 의사 결정 방식을 설명할 수 있어야 합니다.[5] 하지만 위에서 언급한 바와 같이, 일부 기업들은 “설명 가능성” 요구사항을 충족시키기 위한 접근방식을 개발하고는 있지만, LMM의 의사 결정 방식을 충분히 설명할 수 없습니다*(90)*.

수많은 심각한 위반사항들이 있습니다. 이러한 위반사항은 대부분 LMM의 학습 방법, 사용 방식 및 데이터 관리

5 일반 데이터 보호규정(General Data Protection Regulation) 제15조(1)(h)항 및 설명조항 71 참조.

방식과 관련이 있습니다. LMM이 일반 데이터 보호규정 또는 기타 데이터 보호법을 완전히 준수할 수 없을 가능성도 존재합니다*(91)*. 2023년, 한 유럽연합 회원국의 데이터 보호 당국에 제기된 불만에서는 한 기업의 대규모 언어 모델과 그 개발 및 운영 방식이 GDPR을 체계적으로 위반했다고 주장했습니다*(92)*.

많은 데이터 보호법 위반은 소비자 보호법 위반으로도 간주될 수 있습니다*(93)*. 더 나아가, 이러한 문제가 해결되지 않는다면, 이는 보건 분야에서의 AI 사용에 관한 WHO의 가이던스 원칙에도 직접적으로 위배될 수 있습니다. 이 가이던스에는 자율성 보호, 투명성 보장, "설명 가능성" 및 이해 가능성 원칙이 포함됩니다.

기업들이 기존 법률을 준수하지 못하는 것은 AI 규제와 관련하여 일부 기업들이 심각한 우려를 표명하는 이유일 수 있습니다. 예를 들어, 유럽연합이 계획 중인 AI 규제 법안(AI 법안)에 대응하여 한 주요 기업의 대표는 해당 규정을 준수하지 못할 가능성 때문에 유럽에서 자사의 핵심 LMM 제품을 제공할 수 없을 것이라고 밝혔습니다*(94)*. 이러한 최후통첩은 개인정보 권리와 기타 보호 조치의 약화를 초래하거나, 일부 인권을 포기하려는 국가의 의지에 따라 의료 서비스 제공이 달라질 위험이 있습니다.

3.3 사회적 우려와 위험

다른 AI 기술과 마찬가지로, LMM은 보건 시스템을 넘어서는 더 넓은 사회적 영향을 미칠 것으로 예상되며, 이는 단일 법률이나 정책으로 해결할 수 없습니다. 이러한 사회적 영향에는 LMM 상용화를 선도하는 소수의 기술 기업(및 그 경영진)의 권력과 권한를 강화할 가능성이 포함됩니다. 또한, LMM은 훈련 및 사용 과정에서 소비되는 탄소 및 물로 인해 환경과 기후에 부정적인 영향을 미칠 수 있습니다. 이러한 기술은 "문화가 이를 안전하게 흡수할 수 있는 속도보다 빠르게 수십억 명의 삶에 얽혀들며"*(4)*, 의료 및 의학 분야를 포함한 다양한 영역에서 사용되고 있습니다. 이는 AI 기술이 부정확하거나 잘못된 정보, 편향된 증거와 권장사항을 제공하며, 도덕적 또는 맥락적 추론이 결여된 상태에서 인간의 인식적 권위를 대체하지 않도록 보장할 수 있기 이전에 이루어지고 있습니다. 또한, LMM이 기술을 통한 젠더 기반 폭력을 증폭시킬 수 있다는 심각한 우려가 있습니다. 여기에는 사이버 괴롭힘, 혐오 발언, 비동의 이미지 및 동영상 사용(예: "딥페이크") 등이 포함됩니다. 이러한 문제는 본 보고서에서 다루지 않았지만, 이는 WHO가 더 광범위하게 고려해야 할 사항입니다. 특히 이러한 AI 사용으로 인해 피해를 입는 청소년 소녀와 여성에게 보건과 복지에 심각한 부정적 영향을 미칠 수 있습니다*(95)*.

대규모 기술 기업과 관련된 과제

점점 더 많은 매개변수를 포함하며 규모가 커지는 LMM의 등장은 AI를 개발 및 배포하는 소수의 대규모 기술 기업의 지배력과 중심성을 강화했습니다*(96)*. 점점 더 정교한 LMM을 개발할 인적, 재정적 자원, 전문 지식, 데이터, 컴퓨팅

능력을 갖춘 기업과 정부는 소수에 불과합니다*(96)*. LMM에 대한 컴퓨팅 능력과 투자가 증가하고 있으며, AI에 대한 수요가 증가함에 따라 "AI 인재"를 모집하는 비용도 높아지고 있습니다*(97,98)*. 가장 강력한 마이크로칩을 포함하는 LMM은 수많은 컴퓨터와 수천 개의 칩이 함께 작동해야 훈련될 수 있으며, 이러한 컴퓨터는 몇 주 또는 몇 달 동안 24시간 작동해야 합니다*(99)*.

LMM의 훈련, 배포 및 유지보수 비용이 계속 증가함에 따라, 이러한 기술이 많은 제품과 서비스(의료 포함)의 기본 구성 요소가 될 수 있는 기술을 소수의 기업이 독점하는 "산업 지배" 위험이 있습니다. 이는 대학(학계), 스타트업, 심지어 정부까지 배제시킬 가능성이 있습니다*(100)*. AI 연구 분야에서는 이미 대규모 기업들이 대학과 정부를 배제하고 있다는 설득력 있는 증거가 존재합니다.

그중 하나의 징후는 AI 박사 졸업생들이 선택하는 직업 경로입니다. 현재 "전례없는" 수의 졸업생이 기업에서 일하는 것을 선택하고 있습니다. 예를 들어, 2004년에는 약 20%의 졸업생이 산업계로 진출했지만, 2020년에는 거의 70%가 산업계에서 활동하고 있습니다*(101)*. AI를 전문으로 하는 대학 교수들도 대학을 떠나 산업계에서 종사하고 있으며, 이러한 이동은 2006년 이후 8배 증가했습니다. 이는 미국뿐만 아니라 다른 국가에서도 나타나는 현상입니다*(101)*. 산업계는 컴퓨팅 능력과 대규모 데이터 세트 사용 측면에서도 정부와 학계를 압도하고 있습니다. 2021년에는 산업계 모델이 학계 모델보다 29배 더 컸습니다*(101)*. 또한, 고소득 국가 정부의 예산은 산업계의 지출에 비해 현저히 뒤처지고 있습니다. 한 연구에 따르면, "2021년, 미국 정부의 비군사 기관은 AI에 15억 달러를 할당했습니다. 같은 해, 유럽연합 집행위원회는 10억 유로(약 12억 달러)를 지출할 계획이었습니다. 반면, 전 세계적으로 산업계는 2021년에 AI에 3400억 달러 이상을 지출했으며, 이는 공공 투자에 비해 압도적으로 많은 금액입니다"*(101)*.

"AI 투입"의 지배는 대규모 기술 기업들이 이제는 AI의 결과물과 결과를 지배하게 되었음을 의미합니다. 산업계가 개발한 가장 큰 AI 모델의 점유율은 2010년 11%에서 2021년 96%로 증가했으며, 2000년에서 2020년 사이 산업계 공동 저자가 포함된 연구 보고서의 비중은 16% 증가했습니다*(101)*.

산업계의 지배는 AI의 응용 및 사용 뿐만 아니라 초기 단계 연구의 우선순위까지 정의하게 되었습니다*(101)*. 산업계의 지배와 정부의 투자 부족은 공공 이익에 중요한 AI 기술, 특히 의료 및 의학 분야에서 대안을 점점 더 희귀하게 하고 있습니다. 이는 제약 산업의 경우와는 다릅니다. 예를 들어, 제약 산업에서는 정부, 비영리 단체 및 자선 단체의 연구 개발 투자(특히 신약 개발 초기 단계와 특정 치료제의 후기 개발 단계)가 상당히 이루어지고 있습니다*(102)*. 기업들은 점점 더 경제와 사회 부문(의료 포함)을 지탱하는 시스템 운영을 감독하게 될 것이며, 이는 시민과 공무원이 자신의 삶을 관리할 능력에 대한 우려를 야기합니다*(101)*.

대안과 규제가 없는 상황에서(심지어 2023년에 법률이 제정되더라도 완전히 시행되기까지는 수년이 걸릴 수 있음),

대규모 기술 기업들이 내부적으로 의사 결정을 내리는 방식과 사회 및 정부와의 관계는 더욱 중요해지고 있습니다. 기업들은 Frontier Model Forum*(103)*이나 고소득 국가 정부와의 파트너십 프로그램을 통해, 또는 미국 정부와의 여러 자발적 약속*(104)* 및 유럽연합과의 예정된 약속*(105)*을 통해 다양한 우려를 해결하기 시작할 수 있습니다.

또 다른 우려는 기업들이 윤리에 대한 기업적 책임을 유지하지 않을 가능성입니다. 예를 들어, 주요 기술 기업들이 AI 모델의 설계 및 개발이 내부 윤리 원칙을 준수하도록 보장하고, 특정 개발 활동을 지연하거나 중단하도록 요구하는 "마찰"을 받아 들이기 위해 설립된 윤리팀을 축소하거나 제거하고 있다는 점입니다*(106)*. AI와 관련된 윤리 문제를 다루는 팀 전체를 제거한다는 것은 윤리 원칙이 "제품 설계와 밀접하게 연관되어 있지 않다"는 것을 의미하며*(108)*, 이로 인해 중요한 공백이 발생할 수 있습니다.

몇몇 대규모 기술 기업들은 Frontier Model Forum을 통해 LMM을 포함한 "첨단 AI 모델의 책임감 있고 안전한 개발"을 보장하고, "첨단 모델의 책임 있는 개발 및 배포를 위한 모범 사례 식별"과 "정책 입안자, 학계, 시민 사회 및 기업들과 협력하여 신뢰와 안전 위험에 대한 지식을 공유"하겠다고 약속했습니다*(103)*. 또한, 미국 정부와의 자발적 약속을 통해 기술 기업들은 유해한 편향 및 차별을 피하고 개인정보를 보호하겠다고 서약했습니다*(104)*. 그러나 자발적 약속이나 파트너십이 강력한 윤리적 책임을 대체할 수 있을지는 명확하지 않습니다. 예를 들어, 한 기업의 윤리팀은 새로운 LMM의 출시를 중단할 것을 권고했으나, 이후 관련 문서를 수정하고 이전에 문서화된 위험을 축소시켰습니다*(106)*.

대규모 기술 기업들은 의료 제품 및 서비스 개발에 대한 역사적 경험이 없으며, 이 분야에 전문화되어 있지도 않습니다. 따라서 이들은 의료 시스템, 제공자, 환자의 요구사항에 민감하지 않을 수 있으며, 전통적인 의료 기업이나 공중 보건 제공자에게 익숙한 개인정보 보호나 품질 보증과 같은 문제를 다루지 못할 수 있습니다. 그러나 몇 십 년 동안 의료 제품과 서비스를 제공해 온 다른 기업들에서와 같이 시간이 지남에 따라 이러한 민감성이 개선될 가능성은 있습니다.

LMM을 개발 중인 많은 기업들은 정부나 규제 기관, 그리고 그들의 모델을 사용하는 기업들과도 투명하지 않은 모습을 보이고 있습니다. 이러한 기업들은 (i) LMM의 위험과 이점을 평가하기 위한 증거, 데이터, 성능 및 기타 정보*(97,106)*, (ii) 모델 내 매개변수의 수(이는 모델의 성능을 나타내는 지표)*(8)*를 요구하는 경우에도 이를 충분히 제공하지 않고 있습니다. 또한, 이러한 모델을 사용하여 자체 제품 및 서비스를 개발하는 기업들 역시 윤리적 문제와 위험을 평가하는 방식, 도입된 안전장치, 이러한 안전장치에 대한 LMM의 반응, 그리고 기술 사용이 제한되거나 중단되어야 할 시점에 대한 정보를 공개하지 않습니다. Foundation Model Transparency Index는 100가지 지표를 기준으로 대규모 언어 모델 개발 기업 10곳을 평가한 결과, "주요 모델 개발 기업 중 어느 곳도 적절한 투명성을 제공하지 못하고 있으며, AI 산업에서 근본적인 투명성 부족을 드러내고 있다"고 지적했습니다*(107)*. 미국 연방 정부와 여러 대규모 기술 기업 간의 자발적 합의에는 투명성과 관련된 두 가지 약속이 포함되어 있습니다. 이 기업들은 (i) 산업, 정부, 시민 사회, 학계와 위험 관리 정보를 공유하고, (ii) AI 시스템의 역량, 한계, 적절하거나 부적절한 사용 영역에 대해 공개적으로 보고하기로 약속합니다*(104)*.

이러한 약속은 현 상황을 개선할 가능성이 있지만, 자발적이며 각 기업의 해석에 따라 달라질 수 있기 때문에 구체적인 규제 요구사항 없이는 충분한 공개로 이어지지 않을 가능성이 있습니다.
기업들은 내부 상업적 압박이나 외부 경쟁*(106)*으로 인해 LMM이 어떻게 작동하는지 완전히 이해하기 전에*(109)*, 그리고 적절한 테스트, 안전장치, 윤리적 위험 및 우려를 식별하고 해결했는지 여부에 관계없이*(106,110)* 최대한 빠르게 시장에 출시하려 하고 있습니다. 한 기업의 임원은 "나중에 수정할 수 있는 문제에 대해 지금 걱정하는 것은 완전히 치명적인 실수"라고 언급했습니다*(106)*. 기업들은 인터넷 검색과 같은 특정 분야에서 LMM의 시장 점유율이 수익을 창출할 수 있기 때문에 "선점자 우위"를 추구합니다. 한 기업에 따르면 검색 엔진에서의 시장 점유율 1%는 추가로 20억 달러의 수익을 의미합니다*(108)*. 또 다른 주요 기술 기업의 임원은 자사의 LMM이 "완벽하지 않다"고 인정하면서도 "시장 수요가 있기 때문에 출시할 것"이라고 밝혔습니다*(8)*. 위험을 완전히 식별, 검증, 평가, 완화하지 않고 LMM을 출시하는 기업들은 "윤리적 부채"를 축적하게 되며, 이러한 기술의 부정적 영향에 가장 취약한 기업들이 그 결과를 감당해야 할 것입니다*(109)*. Frontier Model Forum의 회원들은 "AI 안전 연구를 증진"하고 "모범 사례를 식별"하는 데 헌신할 것을 약속했으며*(103)*, 미국 정부와의 자발적 약속에는 출시 전에 AI 시스템에 대한 내부 및 외부 테스트가 포함됩니다*(104)*.

상업적 압박은 기업이 LMM을 가능한 한 빨리 시장에 출시하게 만들 뿐만 아니라, 수익을 창출할 수 있는 서비스를 우선시하기 위해 공중 보건에 상당한 이익을 제공하는 제품 및 서비스를 우선시하지 않거나 포기하도록 만들 수 있습니다. 2023년, 한 주요 기술 기업은 단일 단백질 서열로부터 원자 수준의 단백질 구조 전체를 예측할 수 있는 단백질 언어 모델인 ESMFold라는 LMM을 개발한 팀을 "해체"했습니다. 이 모델은 6억 개 이상의 단백질 구조를 포함한 데이터베이스를 생성했습니다. 해당 기업이 데이터베이스 운영 비용을 감당하거나 과학자들이 새로운 단백질 서열에 ESM 알고리즘을 실행할 수 있도록 하는 서비스를 유지할 의지가 없을 것이라는 우려가 제기되고 있습니다*(111)*.

LMM의 탄소 및 물 발자국

LMM 크기의 증가로 인한 또 다른 결과는 환경적 영향입니다. LMM은 대량의 데이터를 필요로 하며, 이러한 데이터를 훈련시키는 데 상당한 에너지를 소비합니다*(112)*. 한 대규모 기업에서 새로운 LMM을 훈련하는 데 약 2개월 동안 3.4 GWh의 에너지가 사용되었는데, 이는 미국 가정 300가구의 연간 에너지 소비량과 같습니다*(112)*. 일부 LMM은 재생 가능 에너지나 탄소 없는 에너지를 사용하는 데이터 센터에서 훈련되지만, 대부분의 AI 모델은 화석 연료로 전력을 공급받는 전력망에서 훈련됩니다*(112)*. 더 많은 기업들이 LMM을 도입하면서 전력 소비는 계속 증가할 것이며, 이는 결국 기후 변화에 상당한 영향을 미칠 수 있습니다.

WHO는 기후 변화를 시급한 전 세계적 보건 문제로 간주하며, 앞으로 수십 년 동안 이를 우선적으로 해결해야 한다고 강조합니다. 2030년에서 2050년 사이, 기후 변화로 인해 영양실조, 말라리아, 설사병, 열 스트레스로 인해 매년 약 25만 명이 추가로 사망할 것으로 예상됩니다. 2030년까지 보건에 대한 직접적인 피해 비용은 매년 20억~40억 달러로

추정됩니다. 대부분 저소득 및 중소득 국가에 위치한 보건 인프라가 취약한 지역은 준비와 대응을 위한 지원 없이는 이러한 상황에 대처하기 어려울 것입니다*(1)*.

LMM은 상당한 물 사용 발자국을 가지고 있습니다. 한 대규모 기술 기업에서 초기 LMM을 훈련시키는 데 70만 리터의 담수가 소비되었으며, 다른 데이터 센터에서는 이 수치가 더 클 가능성이 있습니다*(113)*. 많은 개발자들이 탄소 발자국에 대해서는 점점 더 인식하고 있지만, 물 발자국에 대해서는 여전히 인식이 부족한 경우가 많습니다*(114)*. 예를 들어, LMM과의 짧은 대화(20~50건의 질문과 답변 는 약 500mL 물병 한 병에 해당하는 물을 소비합니다. LMM을 훈련시키는 데 드는 전체 물 발자국은 AI 서버 제조, 운송, 칩 제작을 포함한 모든 물 소비를 포함하여 훨씬 더 클 수 있습니다*(114)*. 데이터 센터는 지역 물 공급에 스트레스를 줄 수 있습니다. 예를 들어, 미국 오리건 주의 한 도시에서는 한 기업의 데이터 센터가 해당 도시 전체 물 사용량의 25% 이상을 소비했습니다*(114)*. 또 다른 대규모 기술 기업은 심각한 가뭄이 지속되고 주민들이 염분이 섞인 물을 마실 수밖에 없는 국가에 데이터 센터를 건설할 계획을 세우고 있습니다*(115)*. 물 발자국을 추적하는 것은 어려운데, 이는 탄소 발자국에 대한 인식, 측정 및 투명성이 더 높은 반면, 기업들이 물 발자국에 대해서는 동일한 수준의 투명성을 제공하지 않거나 이를 측정하지 않는 경우가 많기 때문입니다*(114)*.

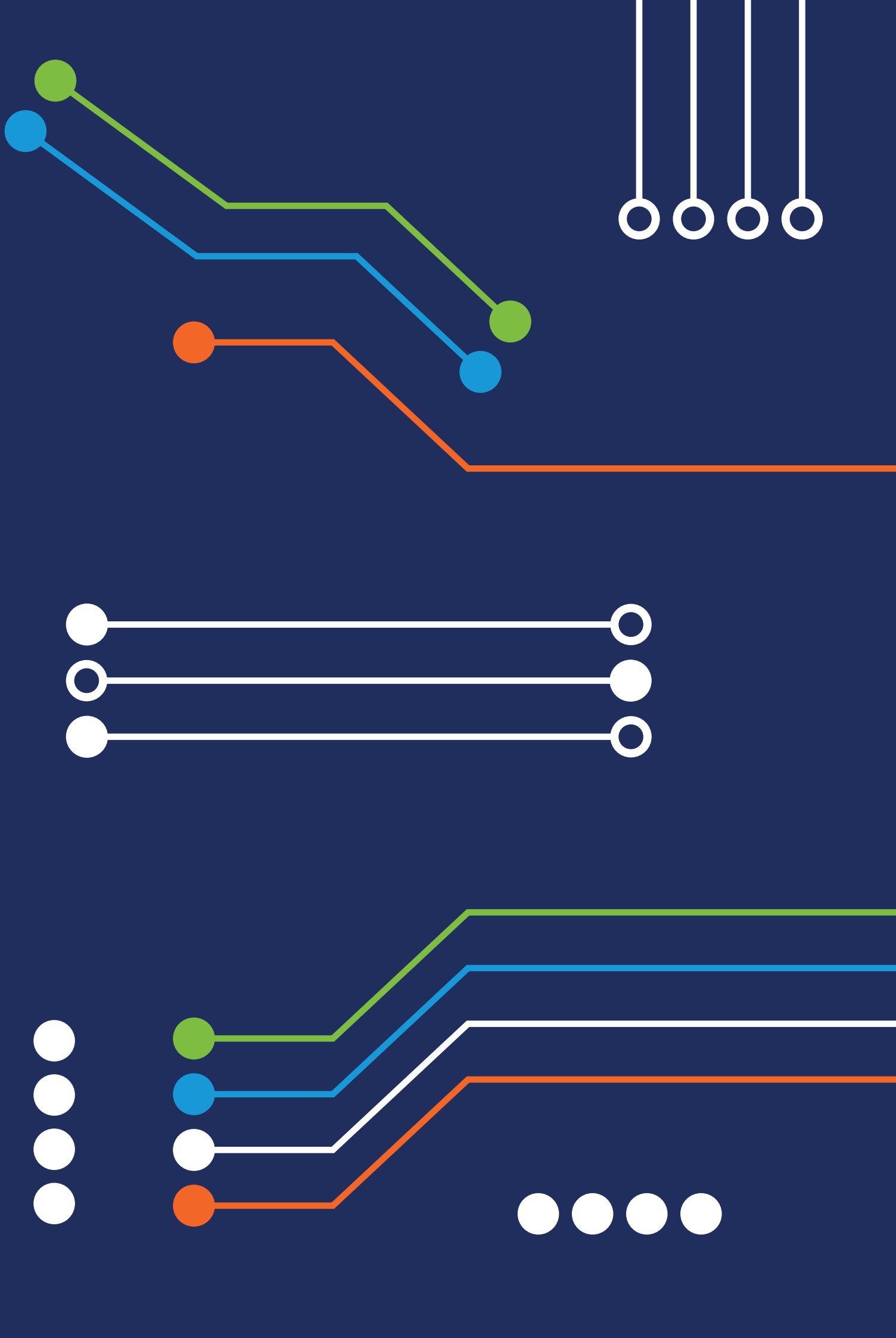
"위험한" 알고리즘: 인간의 인식적 권위의 대체

LMM의 등장과 관련된 보다 일반적인 사회적 위험은 LMM이 점점 더 지식의 원천으로 간주되는 그럴듯한 응답을 제공함으로써, 의료, 과학, 의학 분야를 포함한 인간의 인식적 권위를 결국 약화시킬 수 있다는 점입니다. 그러나 LMM은 실제로 지식을 생성하지 않으며, 자신이 "말하고 있는" 내용을 이해하거나 질문에 대한 도덕적 또는 맥락적 추론을 할 수 없습니다.

이러한 우려가 지속된다면, 사회는 컴퓨터 기반 추론의 결과에 대비하지 못할 수 있습니다. 이전의 소셜 미디어 알고리즘과 같은 초기 형태의 AI는 잘못된 정보를 퍼뜨려 정신 건강에 부정적인 영향을 미치고, 분열과 양극화를 심화시킨 사례가 있습니다*(4)*. 기술 기업들이 LMM의 위험에 대해 반복적으로 경고하면서도, 안전장치나 규제 감독 없이 LMM을 사회에 직접 출시하고 있으며, 이는 인간의 지식 생산을 통제를 대체할 뿐만 아니라 의료, 의학 등 사회가 의존하는 시스템에서 지식을 안전하게 활용할 수 있는 인간의 능력을 저하시킬 위험이 있습니다. 특히 자원이 부족한 환경에 있는 사람들과 지역사회는 AI 시스템 훈련에 그들의 데이터가 사용되지 않았을 가능성이 높아 응답의 정확성이 떨어질 수 있습니다. 그러나 이러한 그룹은 보건의료 전문가나 의료 제공자가 없는 상황에서, 잘못된 또는 부정확한 LMM 응답을 맥락화하거나 수정할 수 없기 때문에 AI 시스템의 조언을 따를 가능성이 높습니다.

LMM이 불완전하거나 잘못된 정보를 공공 영역과 지식 기반에 지속적으로 배출할 경우, 결국 “모델 붕괴(model collapse)”로 이어질 수 있습니다. 이는 LMM이 부정확하거나 거짓된 정보로 훈련되어 인터넷과 같은 공공 정보 출처를

오염시키는 현상을 의미합니다*(116, 117)*. LMM이 보건 의료 및 기타 사회적 중요 영역에서 이점을 최대화하면서 이러한 시나리오를 방지하려면, 정부, 시민 사회 및 민간 부문이 이 기술들을 공익으로 이끌어야 합니다.



Ⅱ. 보건의료 및 의학 분야에서의 LMM 윤리 및 거버넌스

WHO 전문가 그룹이 정의한 윤리 원칙(위 참조)은 이해관계자들에게 보건의료 및 의학 분야에서 LMM의 개발, 배포 및 사용 평가에서 결정을 내리고 행동을 지시할 때 필요한 기본 윤리적 요구사항에 대한 가이던스를 제공합니다. 이 원칙들은 정부, 공공 부문 기관, 연구자, 기업 및 실행자가 LMM의 사용을 관리하는 방식을 위한 기반이 되어야 합니다.

거버넌스는 국가 및 글로벌 보건 정책이 보편적 보건 보장을 촉진할 수 있도록 정부와 국제 보건 기구를 포함한 기타 의사 결정자들이 수행하는 방향 설정 및 규칙 제정 기능을 의미합니다. 거버넌스는 또한 경쟁하는 영향력과 요구를 균형있게 조율하는 정치적 과정이기도 합니다*(1)*. 현행 법률과 정책은 효과적으로 LMM의 사용을 관리하기에는 부족할 가능성이 높습니다. 이는 많은 법률이 초기 LMM 버전이 출시되기 전에 작성되었기 때문입니다. LMM 거버넌스는 전반적인 AI 거버넌스와 마찬가지로, 기존 및 새로운 법률과 규정, 윤리 원칙과 같은 "소프트 로(soft law)", 인권 의무, 실천 강령, 그리고 기업, 산업 협회 및 표준 설정 기관의 내부 절차를 적용하는 것을 포함합니다.

현재 LMM은 그 능력과 취약성을 완전히 이해하기 전에 빠르게 배포되고 있습니다. LMM에 대한 우려를 해결하기 위한 초기 제안 중 하나는 개발 금지 또는 중단하는 것이었습니다*(118)*. 일부 국가에서는 특정 LMM의 사용을 제한하거나 금지하고 있지만, 대부분의 정부는 적절한 거버넌스를 통해 LMM의 사용을 사회적으로 유익한 결과로 이끌기를 희망합니다. 주요 AI 기업들도 신중하고 숙고된 LMM 및 기타 AI 형태의 개발을 촉구하고 있습니다. 그러나 정부와 기업 모두 경쟁에서 자유롭지 않습니다. 여러 정부는 기술적 우위를 확보하기 위한 "군비 경쟁"에 몰두해 있으며, 규제를 요구하는 AI 기업들 조차 상업적 압박에서 자유롭지 않습니다*(119)*. 낙관론자들은 점점 더 큰 데이터 세트와 더 강력한 알고리즘을 설계에 포함함으로써 많은 AI 과제와 위험을 해결할 수 있다고 보고 있지만, 비판론자들은 LMM의 한계가 시스템적으로 내재되어 있으며, 훈련 데이터와 모델 매개변수의 크기를 늘리는 것이 문제를 해결하기보다는 오히려 증폭시킬 수 있다고 지적합니다*(59)*.

LMM 거버넌스는 그 급속한 발전과 증가하는 용도에 맞추어야 하며, 기술적 우위를 추구하는 정부나 상업적 이익을 추구하는 기업 어느 한쪽도 특권을 가져서는 안 됩니다. 초기 제안과 권고사항은 윤리 원칙과 인권 의무를 적절한 거버넌스의 중심에 두고 있으며, 이는 기업이 도입할 수 있는 절차 및 관행과 정부가 제정할 수 있는 법률 및 정책을 포함합니다.

LMM은 하나 이상의 행위자가 프로그래밍과 제품 개발에서 내리는 일련의(또는 연속적인) 결정의 산물로 간주될 수 있습니다. AI 가치사슬의 각 단계에서 이루어진 결정은 LMM의 개발, 배포 및 사용에 참여하는 사람들에게 직접적 또는 간접적인 영향을 미칠 수 있습니다. 이러한 결정은 국가적, 지역적, 글로벌 수준에서 법률 및 정책을 제정하고 집행하는 정부에 의해 영향을 받고 규제될 수 있습니다. AI 가치사슬은 데이터, 컴퓨팅 능력, AI 전문 지식과 같은 "AI 인프라"를 구성하는 다양한 입력사항을 범용 기반 모델 개발에 통합하는 것으로부터 시작됩니다. 이러한 모델은 사용자가 다양한,

종종 예상치 못한 작업(예: 의료 관련 작업)을 수행하는 데 직접 사용할 수 있습니다. 일부 범용 기반 모델은 의료 및 의학 분야에서 사용되도록 특별히 훈련되었습니다.
의료 및 의학 분야에서 사용되는 LMM의 적절한 거버넌스는 데이터 수집부터 의료 응용 프로그램 배포까지 가치사슬의 각 단계에서 정의되어야 합니다. 따라서, AI 가치사슬의 세 가지 중요한 단계는 다음과 같습니다:

범용 기반 모델의 설계 및 개발(설계 및 개발 단계);

- 범용 기반 모델을 활용한 서비스, 응용 프로그램 또는 제품의 정의(제공 단계)
- 의료 서비스 응용 프로그램 또는 서비스의 배포(배포 단계)

각 단계에서 다음과 같은 질문이 제기됩니다:

- 관련 위험을 가장 잘 해결할 수 있는 행위자(개발자, 제공자, 배포자)는 누구인가? AI 가치사슬에서 해결해야 할 위험은 무엇인가?
- 관련 행위자들이 이러한 위험을 어떻게 해결할 수 있는가? 그들이 준수해야 할 윤리 원칙은 무엇인가?
- 정부는 이러한 위험을 해결하는 데 어떤 역할을 할 수 있는가? 정부는 행위자들이 특정 윤리 원칙을 준수하도록 하기 위해 어떤 법률, 정책 또는 투자를 도입하거나 적용할 수 있는가?

설계 및 개발 단계에서는 개발자가 윤리적 약속과 규범을 준수하기 위해 도입할 수 있는 관행과 정부 정책 및 투자가 초점이 됩니다. 제공 단계에서는 정부가 의료 및 의학 분야에서 LMM의 사용을 평가하고 규제하기 위해 도입할 수 있는 조치에 초점이 맞춰집니다. 배포 단계에서는 정부와 가치사슬의 모든 행위자들이 사용자의 잠재적 또는 실제적인 피해를 식별하고 방지하기 위한 조치를 사용합니다.

4 범용 기반 모델(LMM)의 설계 및 개발

범용 기반 모델은 방대한 데이터를 학습시키며 엄청난 컴퓨팅 능력이 필요합니다. LMM의 개발에는 과학 및 공학 전문성을 포함한 특화된 인적 자원도 필요합니다. WHO의 의료 분야 AI 윤리 및 거버넌스에 관한 가이던스*(1)*는 의료용 AI 개발자가 "제품의 설계, 감독, 신뢰성 및 자율 규제를 개선하기 위한 조치에 투자해야 한다"고 권고하고 있습니다.

아래의 대부분의 연구 결과와 권고사항은 모든 범용 기반 모델에 적용될 수 있지만, 이 가이던스는 특히 의료 및 의학 분야에서 사용될 가능성이 있거나 실제로 사용되는 모델(사용자에 의해 직접적으로 또는 응용 프로그램이나 서비스를 통해 사용되는 경우)에 초점을 맞추고 있습니다. 또한, 아래 권고사항은 의료 및 의학 분야에서 사용하기 위해 특별히 훈련된 LMM의 설계와 사용을 안내하는 것에도 목적이 있으며, 이러한 모델은 사용자에 의해 직접적으로 또는 응용 프로그램이나 서비스를 통해 사용될 수 있습니다.

4.1 범용 기반 모델(LMM) 개발 시 해결해야 할 위험

범용 기반 모델의 설계 및 개발 과정에서 심각한 위험이 발생할 수 있으며, 이를 수정하지 않을 경우 광범위한 사회적 영향을 미치거나 LMM 사용자에게 특정한 부정적 결과를 초래할 수 있습니다. 이러한 위험을 제거하거나 완화하는 것은 개발자의 책임입니다. 이는 설계 및 개발 중에 개발자만이 내릴 수 있는 특정 결정 때문입니다. 이러한 결정은 알고리즘을 사용하는 제공자나 배포자의 통제를 벗어나는 경우가 많으며, 제공자, 배포자 또는 사용자가 기술을 올바르게 사용한다고 해서 완화될 수 없습니다*(120)*. 예를 들어, LMM을 훈련시키는 데 사용되는 데이터와 관련된 결정*(121)*, 데이터 보호와 품질 보장을 위한 의무, 편향을 완화하기 위한 조치 등은 응용 프로그램의 다운스트림 개발자의 통제를 벗어납니다*(121)*. 또한, LMM이 "AI로 인한 유해성"을 생성하지 않도록 보장하기 위한 조치도 포함됩니다*(122)*. LMM 개발자에게 이러한 설계 결함에 대한 책임을 묻지 않는 것은, 가장 많은 자원을 보유한 기업들을 보호하고 한 보고서에서 언급한 바와 같이 "새로운 형태의 적용 AI에서 지배적 위치를 차지하기 위해 서둘러 개발하는 과정에서 그들의 방법이 무심코 문제를 내재화"하도록 방치하게 됩니다*(122)*.

개발자는 최소한 다음과 같은 여덟 가지 위험을 해결해야 하며, 이는 정부의 법률 및 규정을 통해 이루어질 수 있습니다:

- 편향(설계와 훈련 데이터와 관련됨)
- 개인정보(훈련 데이터 및 기타 입력 데이터의 보호)
- 노동 문제(불쾌한 콘텐츠를 제거하기 위해 외주된 데이터 필터링)
- 탄소 및 물 발자국

- 허위 정보, 혐오 발언 또는 잘못된 정보
- 안전 및 사이버 보안
- 인간의 인식적 권위 보존
- LMM의 독점적 통제

4.2 개발자가 범용 기반 모델(LMM)과 관련된 위험을 해결하기 위해 취할 수 있는 조치

개발자는 윤리 원칙 또는 정책에 대한 약속이나 정부의 요구사항을 충족하기 위해 다양한 조치나 관행을 통해 이러한 위험을 해결할 수 있습니다.

AI 전문성(과학 및 공학 인력): 개발자는 과학 및 프로그래밍 인력이 위험을 식별하고 이를 회피할 수 있도록 보장해야 합니다. WHO 윤리 가이던스*(1)*는 과학 및 공학 인력의 교육과 설계 과정의 포괄성에 대해 여러 가지 권장사항을 제시했습니다. 특히 WHO 전문가 그룹은 "의료용 AI를 포함한 '고위험' AI 개발자를 대상으로 라이선스 또는 인증 요구사항"을 고려할 것을 권고했습니다.

의료, 과학 연구 또는 의학 분야에서 사용될 가능성이 있는 LMM을 개발하는 기업 및 기타 기관은 의료 전문 분야의 요구사항에 부합하고 제품과 서비스에 대한 신뢰를 높이기 위해 인증 또는 교육을 고려해야 합니다*(1)*. 개발자 또는 전문 협회에 의해 도입되고 시행되는 모든 표준은 정부 규제 기관과 협력하여 작성되어야 하며, 인간의 복지, 안전 및 공익을 증진한다는 WHO 윤리 원칙과 일치해야 합니다. 개발자가 LMM이 의료 분야에서 사용될 가능성을 의도하지는 않았지만, 그러한 사용 가능성을 예측할 수 있다면, 이러한 사용을 예상하고 해결하기 위한 내부 전문성을 확보해야 합니다.

데이터: 인적 자원과 컴퓨팅 능력이 필수적이지만, 데이터는 아마도 가장 중요한 인프라 요구사항일 것입니다. LMM을 훈련시키는 데 사용되는 데이터의 품질과 유형은 해당 모델이 핵심 윤리 원칙과 법적 요구 사항을 충족하는지를 결정합니다*(123)*. AI 개발자들은 질적 조사에서 데이터 품질이 "중요하다"는 데 동의하며, 데이터 작업에는 상당한 시간과 노력이 필요하다고 밝혔습니다. 그러나 데이터 작업은 종종 과소평가되며, 이는 의료 및 의학과 같은 "고위험" 분야에서 AI에 중대한 부정적 영향을 미칠 수 있습니다*(123)*. 데이터가 적절한 품질을 갖추지 못하면 인간의 복지, 안전, 공익 증진, 포괄성과 형평성 보장이라는 WHO의 여러 가이던스 원칙을 위반할 수 있으며, 데이터에 편향이 있을 경우 특히 그러합니다.

의료 데이터를 사용하는 것은 사전동의를 얻기 위해 엄격한 법적 요구사항을 준수해야 할 가능성이 높기 때문에,

의료 및 의학 분야에서 사용될 LMM을 훈련시키는 개발자는 더 작은 데이터 세트에 의존해야 할 수도 있습니다*(59)*. 또한, 더 작은 데이터 세트는 데이터 품질을 보장하고, 편향을 피하기 위해 데이터의 다양성을 확보하며*(59)*, LMM이 서비스를 제공할 인구의 구성과 현실을 반영하도록 할 수 있습니다. 그러나 더 작은 데이터 세트는 개인 재식별의 위험을 증가시킬 수 있으며, 이는 현재 또는 미래의 피해를 초래할 수 있습니다. 더 작은 데이터 세트를 사용하는 것은 탄소 및 물 발자국을 줄이고*(112)*, 데이터, 컴퓨팅, 인적 및 재정적 자원이 적게 필요한 LMM을 개발하기 위해 더 작은 단체들이 참여할 수 있도록 하는 등의 추가적인 이점이 있을 수 있습니다*(59)*. 데이터 세트의 크기와 관계없이, 개발자는 데이터를 처리하기 전에 "데이터 보호 영향 평가"를 수행해야 하며, 이는 유럽연합의 일반 데이터 보호 규정에 따라 요구되며, 데이터 처리작업이 개인의 권리와 자유에 미치는 위험과 개인정보 보호에 미치는 영향을 처리하기 전에 평가하도록 요구합니다*(1)*. 저소득 및 중 소득 국가에서 데이터를 수집하는 것은 동의, 개인정보 또는 자율성을 존중하지 않고 상업적 또는 비상업적 목적으로 데이터를 사용하는 "데이터 식민주의"로 간주될 수 있습니다*(1)*.

평가는 개인정보 보호에 대한 위험을 넘어 데이터가 편향되지 않고 정확한지 여부와 같은 데이터 품질을 포함할 수 있습니다. 데이터 세트를 감사하거나 검토하는 AI 연구자들은 AI를 위한 데이터 세트 생성은 간단하지만 감사는 어렵고, 시간과 비용이 많이 든다고 지적하며, 자원을 투자하기를 꺼립니다. 한 연구자는 "더러운 일을 하는 것은 훨씬 더 어렵다"고 말했습니다*(124)*.

개발자는 데이터 품질을 개선하고 데이터 보호법을 준수하기 위해 추가적인 조치를 취할 수 있습니다. 모델 크기와 관계없이, 초기 LMM이 개발되던 방식과는 달리, 개발자는 최선의 데이터 보호 규칙에 따라 수집된 데이터를 기반으로 LMM을 훈련시켜야 합니다. 따라서 데이터 브로커와 같은 제3자 소스로부터 데이터를 사용하는 것을 피해야 합니다. 이러한 데이터는 오래되었거나, 편향이 있거나, 잘못 결합되었거나, 수정되지 않은 다른 결함이 있을 수 있기 때문입니다(125). 데이터를 신중하게 수집하면 LMM이 저작권이나 데이터 보호법을 위반하지 않도록 보장할 수 있으며, 이를 위반할 경우 법적 결과를 초래하여 특정 LMM이 불법으로 간주될 수 있습니다*(126)*.

제3자 데이터 제공자를 사용하는 경우, 신뢰를 구축하고 전문성과 정당성을 보장하기 위해 인증을 고려할 수 있습니다*(127)*. 개발자가 직접 수집했든 제3자로부터 제공받았든, LMM 훈련에 사용되는 모든 데이터는 최신 상태를 유지해야 합니다. 위에서 언급한 바와 같이, 일부 주요 AI 모델은 최신 데이터로 훈련되지 않았으며*(38)*, 이는 특히 새로운 증거와 정보가 의사 결정에 중요한 영향을 미치는 의료 및 의학 분야에서 모델의 성능을 위태롭게 할 수 있습니다. 데이터 세트는 사용되는 맥락에 적합하고 관련성이 있도록 업데이트되고 정확해야 합니다.

데이터의 투명성을 충분히 확보하기는 어려울 수 있습니다. 새로운 LMM을 출시하는 기업들은 모델을 훈련시키는데 사용된 데이터에 대해 점점 더 불투명해지고 있습니다. 한 주요 AI 기업은 새로운 LMM을 출시하면서 다음과 같이

밝혔습니다. “경쟁 환경과 GPT-4와 같은 대규모 모델의 안전과 관련된 문제를 고려하여, 이 보고서에는 아키텍처(모델 크기 포함), 하드웨어, 훈련 컴퓨팅, 데이터 세트 구성, 훈련 방법 또는 이와 유사한 세부 정보가 포함되어 있지 않습니다”*(128)*.

데이터에 대한 투명성을 꺼리는 것은 WHO의 윤리 원칙인 투명성, “설명 가능성” 및 이해 가능성을 보장한다는 원칙과 일치하지 않습니다. 개발자는 LMM을 훈련시키는 데 사용된 데이터에 대해 투명해야 하며, 이를 통해 LMM을 미세 조정하거나 의료 응용 프로그램을 개발하는 다운스트림 사용자, 그리고 LMM을 직접 사용하는 사람들이 훈련 데이터 세트의 불충분성 또는 불완전성을 인식할 수 있도록 해야 합니다.

개발자가 학대, 폭력 또는 불쾌한 자료를 선별하고, 데이터를 주석 처리하기 위해 저소득 및 중 소득 국가의 데이터 작업자를 고용하여 데이터 품질을 개선할 경우, 이러한 작업자들에게는 생활 임금을 지급하고, 정신 건강 서비스 및 기타 상담 지원을 제공해야 합니다. 또한, 작업자가 겪을 수 있는 고통으로부터 보호하기 위한 안전장치를 도입해야 합니다. 정부는 모든 데이터 작업자에게 혜택을 확대하기 위해 노동 기준을 업데이트하고, 기업 간의 “공정한 경쟁 환경”을 조성하며, 시간이 지남에 따라 노동 기준이 유지되고 개선되도록 보장해야 합니다.

윤리적 설계 및 가치를 위한 설계: AI 기술 개발에 윤리 및 인권 기준을 통합하는 한 가지 접근방식은 “가치 기반 설계”로 인간의 존엄성, 자유, 평등, 연대와 같은 가치를 설계의 기반으로 삼고, 이를 비기능적 요구사항으로 간주하는 패러다임입니다*(1)*. WHO의 전문가 가이던스에 제시된 AI 기술 설계에 대한 여러 권장사항 중 “가치를 위한 설계(design for values)”를 포함한 일부는 여기서 다시 한번 언급하여 강조할 필요가 있습니다.

이 가이던스는 AI 기술 설계와 개발이 과학자와 엔지니어에 의해서만 이루어져서는 안 되며, “잠재적 최종 사용자와 모든 직접적, 간접적 이해관계자가 초기 AI 개발 단계부터 포괄적이고 투명한 설계 과정에 참여하고 윤리적 문제를 제기하고 우려를 표현하며 검토 중인 AI 응용 프로그램에 대한 의견을 제시할 기회를 가져야 한다"고 권장했습니다*(1)*. 따라서 LMM과 같은 기반 모델의 개발 과정에서는 모델을 사용할 가능성이 있거나 그로부터 혜택을 받을 가능성이 있는 사람들이 초기 개발 단계에 참여할 수 있습니다. 한 제안으로는 “인간 감독 대학(human oversight colleges)”을 도입하여, 환자나 간병인에게 직접적 또는 간접적으로 혜택을 줄 수 있는 LMM 개발 과정에 환자 대표를 포함시키는 방안을 제안하고 있습니다.[6] 의료 및 보건의료 전문가, 연구 과학자, 환자, 일반인, 취약 계층도 LMM 설계, 데이터 라벨링 및 테스트 과정에 포함될 수 있습니다. LMM 설계의 포괄성은 예를 들어, 의료 제공자가 참여하면 자동화 편향을 방지하거나 줄일 수 있어 인간 자율성을 보호할 수 있습니다. 포괄적 설계는 연령, 능력, 인종, 민족, 성별 또는 성 정체성과

6 보건분야의 인공지능 윤리 및 거버넌스에 대한 WHO 전문가인 데이비드 그루슨의 메시지.

같은 다양한 관점을 포함함으로써 포괄성과 형평성을 보장하는 WHO의 원칙을 촉진할 수 있습니다.

WHO 초기 가이던스는 또한 "설계자와 기타 이해관계자들이 AI 시스템이 잘 정의된 작업을 수행할 수 있도록 설계하고, 보건 시스템의 역량을 개선하며 환자 이익을 증진하기 위해 필요한 정확성과 신뢰성을 확보해야 한다"고 권장했습니다. 디자이너와 이해관계자들은 또한 잠재적인 2차 결과를 예측하고 이해할 수 있어야 합니다*(1)*. LMM 개발을 시작하기 전에도 개발자는 소위 "사전부검"을 수행하여*(33)* "가상실패"를 고려하여 개발팀이 이러한 예상치 못한 실패를 역설계할 수 있습니다. 이를 통해 개발자는 사전알려진 위험과 알려지지 않은 위험을 식별하고 대안을 마련할 수 있습니다*(33)*. 두 번째 제안으로, 범용 기반 모델 개발자들 중 일부는 "레드 티밍(red teaming)" 평가를 권장합니다*(129)*. 이는 실제 시뮬레이션을 통해 모델이나 시스템의 취약점을 식별하여, 편향된 의견 제공과 같은 바람직하지 않은 행동을 초래할 가능성을 평가하는 방법입니다. 이를 통해 개발자는 모델이나 시스템의 신뢰성과 안전성을 보장하기 위해 이를 수정할 수 있습니다. 한 기업은 2023년 8월 해커 컨퍼런스인 DEFCON에 자사의 최신 LMM 모델을 제출하여 "전문가들이 모델의 역량을 추가로 분석하고 스트레스 테스트를 수행할 수 있도록" 할 것이라고 발표했습니다*(130)*.

WHO의 초기 가이던스는 "디자이너가 '가치 기반 설계'를 위해 사용하는 절차는 합의된 원칙, 모범 관행(예: 개인정보 보호 기술 및 기법), 설계 윤리 기준 및 발전하는 전문적 규범에 의해 정보를 제공받고 업데이트되어야 한다"고 권장했습니다*(1)*. 적절한 설계를 통해 LMM에 입력된 데이터의 무단 공개를 제한하거나, LMM의 훈련과 사용에 수반되는 환경적 우려(탄소 및 물 발자국)를 해결할 수 있습니다(아래 참조). 또한, 사용자가 LMM이 생성한 콘텐츠가 인간이 아닌 AI 시스템에 의해 생성되었음을 알 수 있도록 하여, 인간을 인식적 권위의 중심에서 대체하지 않도록 보장할 수 있습니다. 이러한 알림은 사용자, 커뮤니티, 사회에 LMM이 유용한 정보를 생성할 수는 있지만, 인간의 지식 생산을 대체할 수는 없다는 점을 상기시킬 수 있습니다.

환경적 우려를 고려한 설계: 앞서 논의한 바와 같이, LMM과 관련된 주요 우려 중 하나는 탄소 및 물 발자국입니다. 개발자는 모델의 에너지 효율성을 개선하는 등 에너지 소비를 줄이기 위해 가능한 모든 조치를 취해야 하며, 여러 대규모 기술 기업이 이러한 접근방식을 실험하고 있습니다. 예를 들어, 한 기업은 외부 데이터베이스와 결합된 LMM을 개발했으며, 이는 더 많은 변수로 훈련된 LMM보다 효율적으로 작동하면서 에너지 효율이 낮은 LMM을 능가합니다*(112)*. 또 다른 기업은 하나의 신경망이 아니라 64개의 더 작은 신경망에 변수를 분산시킨 LMM을 실험 중입니다. 이 모델은 각 작업을 완료하기 위해 두 개의 신경망만 사용하도록 훈련되었으며, 따라서 각 추론에서 필요한 변수의 일부만 사용하여 에너지를 절약합니다*(112)*.

에너지 효율성을 개선하는 또 다른 방법은 작은 데이터 세트로 훈련된 더 작은 규모의 LMM을 개발하는 것입니다. 이러한 모델은 훈련이나 작동에 많은 에너지를 요구하지 않으므로 에너지 소비를 줄일 수 있습니다. 작은 규모의 LMM은 에너지 소비를 줄이는 것뿐만 아니라, 소규모 기업이나 단체가 LMM을 개발하고 출력의 정확성을 향상할 수 있는 기회를 제공할

수도 있습니다*(59)*. 특히 “전문화된 LMM” 개발에 작은 규모의 LMM이 유용할 수 있습니다. 이는 의료, 과학 연구 및 의학 분야에서 특정 용도로 사용하기 위한 모델을 의미합니다. 대규모 기술 기업을 포함하여 여러 전문화된 LMM이 이미 도입된 바 있습니다*(59)*.

4.3 정부 법률, 정책 및 공공부문 투자

현재 존재하거나 잠재적인 법률이나 정책은 범용 기반 모델의 설계 및 개발 중 발생할 수 있는 위험을 줄이거나 방지하기 위해 시행되거나 작성될 수 있습니다. 또한, 정부는 범용 기반 모델의 윤리적 설계 및 개발을 촉진하거나 지원하기 위해 공공 부문 투자를 할 수도 있습니다.

데이터 사용을 규제하는 법률 및 정책: WHO는 LMM을 훈련하기 위해 사용된 데이터의 방식과 사용될 방식을 규제하는 데이터 보호 규정을 포함한 기준의 적용 및 집행을 지지합니다. 데이터 보호법은 일반적으로 권리에 기반한 접근방식을 기반으로 하며, 데이터 처리 규제를 위한 기준을 포함하여 개인의 권리를 보호하고, 공공 및 민간 데이터 관리 책임자와 처리자의 의무를 설정하며, 법적 권리를 침해하는 행위에 대한 제재와 구제 조치를 포함합니다. 데이터 보호법은 150개 이상의 국가에서 채택되어, LMM을 포함한 모든 AI 기술 개발의 견고한 기반을 제공합니다*(1)*. 그러나 데이터 보호법의 한계는 대부분의 법률이 생성형 AI 및 기타 유형의 AI가 등장하기 전에 제정되었으며, 데이터 보호 당국이 초기 법률의 의도가 동일하지 않을 수 있다는 이유로 이를 지나치게 엄격하게 적용하지 않으려는 경향이 있다는 점입니다*(120)*.

LMM 훈련에 사용되는 보건 데이터를 포함한 데이터 보호 요구사항 중 하나는 데이터가 적법하게 수집되고 처리되어야 한다는 것입니다. 이는 종종 데이터 주체가 명시된 목적에 따라 데이터 사용에 대해 의미 있는 정보제공 동의를 해야 함을 의미합니다. 추가적인 데이터 처리에는 초기목적과 호환된다고 가정할 수 없으므로 자체적인 법적 근거가 있어야 합니다. LMM을 개발하고 출시한 기업 및 기타 단체는 이미 동의 없이 수집된 데이터를 사용했을 가능성에 대해 조사를 받고 있습니다. 점점 더 큰 데이터 세트가 요구되는 LMM 개발은 개발자가 법적 요구사항을 무시하도록 만들 수 있습니다*(83)*. 이는 인간 자율성을 보호하는 WHO 가이던스 원칙을 위반하는 것이기도 합니다. 따라서 WHO 전문가 그룹은 정부에 "보건 데이터 사용 및 개인 권리 보호, 특히 의미 있는 동의를 받을 권리를 포함한 명확한 데이터 보호법과 규정을 마련해야 한다"고 권장했습니다.

LMM 훈련 데이터를 수집하고 사용하는 것을 감독하고 규제하기 위한 다른 정부 조치로는 중국 정부가 2023년 8월에 시행한 생성형 AI 규정이 있습니다. 중국 국가 인터넷 정보 판공실 (Cyberspace Administration of China)은 (i) 제공자는 훈련 데이터를 선택할 때 차별과 편향을 방지하기 위한 효과적인 조치를 취해야 하고, (ii) 제공자는 명확한 라벨링을 사용하고 데이터 라벨링의 품질을 평가해야 하며, (iii) 개발자는 데이터의 진정성, 정확성, 객관성 및 다양성 목표를 달성하기 위해 “효과적인 조치”를 취해야 한다*(131)*와 같은 여러 의무를 부여하고 있습니다. 이러한 요구사항은 기업에게 엄격하게 적용되지 않을 것으로 예상되며, 기업은 적절한

데이터 품질을 보장하기 위한 효과적인 조치를 취하기만 하면 됩니다. 이러한 조치는 중국 대중에게 서비스를 제공하는 기업에만 적용됩니다*(132)*.

데이터와 관련된 법적 조항에는 기반 모델을 훈련하는 데 사용되는 데이터 소스를 설명하고, 적합성, 편향 및 적절한 완화 조치를 포함하여 데이터 거버넌스의 대상이 되는 데이터를 사용하기 위한 요구사항이 포함될 수 있습니다*(133)*.

정부가 설계 및 개발 단계에서 취할 수 있는 추가적인 조치는 아래와 같습니다.

- *대상 제품 프로필*: 정부와 국제 기관은 의료 및 의학 분야에서 사용될 LMM의 선호도와 특성을 명시하는 대상 제품 프로필을 발행할 수 있습니다. 특히, 정부가 공공 보건 시스템에서 사용할 목적으로 이러한 기술을 구매할 것을 예상할 경우 이러한 조치가 유용합니다.

- *설계 및 개발 표준과 요구사항*: 정부는 범용 기반 모델의 설계와 개발이 생애 주기 전반에 걸쳐 특정 결과를 달성하도록 개발자들에게 요구할 수 있습니다. 여기에는 모델의 예측 가능성, 해석 가능성, 수정 가능성, 안전성 및 사이버 보안을 보장하는 요구사항이 포함될 수 있습니다*(134)*.

- *사전 인증 프로그램*: 규제 기관은 법적 의무를 부여하고, 윤리적 위험(예: 편향 또는 자율성 침해)을 식별하고 회피하도록 개발자를 독려하기 위해 사전 인증 프로그램을 포함한 다양한 조치를 통해 인센티브를 제공할 수 있습니다*(1)*. 이전 WHO AI 윤리 가이던스는 "정부 규제 기관이 제품 설계 및 개발 과정에서 관련 안전 및 인권 문제를 식별, 모니터링 및 해결하도록 개발자에게 인센티브를 제공하고, 이러한 가이던스를 사전 인증 프로그램에 통합해야 한다"고 권장합니다*(1)*.

- *감사(Audits)*: 정부는 기반 모델 개발 초기 단계에 대한 감사를 도입할 수 있습니다. 한 제안으로는 세 가지 유형의 감사가 있습니다: LMM 제공자에 대한 "거버넌스 감사", LMM 자체에 대한 감사, LMM을 기반으로 구축된 다운스트림 제품 및 서비스에 대한 "응용 프로그램 감사"로 이러한 감사는 LMM 개발 과정에서는 적용되지 않을 수 있습니다. 의료 또는 의학 분야에서 사용될 LMM의 승인을 위한 요구사항에 통합될 수 있습니다(아래 참조). 감사가 효과적으로 이루어지기 위해서는,그 품질이 의도된 목적을 충족하는지 평가되어야 합니다.

- *환경 발자국(Environmental Footprint)*: 정부는 범용 기반 모델 개발자에게 탄소 및 물 발자국과 관련된 우려를 해결하도록 요구할 수 있습니다. 예를 들어, 에너지 소비를 측정하고, 훈련 중 에너지 사용을 줄이며*(133)*, 아직 정의되지 않은 환경 기준을 충족하도록 개발자에게 요구할 수 있습니다*(134)*.

- *"기계 생성" 콘텐츠 알림(Notification)*: 정부는 범용 기반 모델의 배포 시 콘텐츠가 인간이 아닌 기계에 의해 생성되었음을 최종 사용자에게 알리고 상기시키는 알림과 메시지를 포함하도록 개발자에게 요구할 수 있습니다*(133)*.

- *정부의 조치*: 정부는 의료 및 의학 분야에서 사용될 AI 알고리즘이나 시스템의 조기 등록을 요구하거나, 이를 장려하기 위한 인센티브 제공을 고려할 수 있습니다. 조기 등록은 부정적인 결과의 공개를 촉진하고, 출판 편향이나 결과에 대한 지나치게 낙관적 해석을 방지하며, 환자에게 유익한 지식의 통합을 용이하게 할 수 있습니다.

공익을 위한 LMM 개발을 지원하는 공공 인프라: LMM의 의료 분야 활용이 급증함에 따라, 윤리적 원칙을 준수하는 LMM 개발을 장려하기 위해 비영리 또는 공공 인프라(컴퓨팅 능력 및 공공 데이터 세트 포함)를 제공할 수 있습니다. 이러한 인프라는 공공, 민간 및 비영리 부문의 개발자들이 접근할 수 있으며, 접근 조건으로 윤리적 원칙과 가치를 준수하도록 요구할 수 있습니다. 또한, LMM에 대한 개발자의 독점적 통제를 방지하고, 가장 큰 기업과 자원에 접근할 수 없는 개발자 간의 공정한 경쟁의 장을 조성하는 데 도움을 줄 수 있습니다.

정부는 독립적인 감독 하에 의료 및 의학 분야에 사용할 LMM을 개발할 수 있는 인프라를 구축할 수 있습니다. 예를 들어, 국제적으로 1,000명의 학술적 자원봉사자, Hugging Face라는 기업, 프랑스 정부의 자금 지원을 통해 1750억 개의 매개변수를 가진 BLOOM이라는 LMM이 훈련되었습니다. 이 프로젝트에는 700만 달러의 컴퓨팅 비용이 필요했습니다*(112)*.

공정한 경쟁의 장을 마련하려는 노력은 학계에도 적용될 수 있으며, 학계는 자원 측면에서 불리한 위치에 있습니다. 캐나다 정부의 국가 선진 연구 컴퓨팅 플랫폼(National Advanced Research Computing Platform)은 학문적 부문을 지원합니다. 중국 정부는 학계 및 기타 단체가 데이터와 컴퓨팅 능력에 접근할 수 있도록 국가 컴퓨팅 능력 네트워크 시스템을 승인했으며, 미국에서는 국가 AI 연구 자원(National AI Research Resource) 태스크포스가 “공공 연구 클라우드와 공공 데이터 세트를 생성할 것을 제안”했습니다*(101)*. 또한, 유럽 시민사회는 정부가 “유럽 대규모 생성 모델” 구축에서 더 적극적인 역할을 수행해야 한다고 주장하며, 이를 위해 AI 전용 컴퓨팅, 데이터 인프라, 과학 및 연구 지원을 제공해야 한다고 촉구하고 있습니다*(135)*.

4.4 오픈소스 LMM

오픈소스 LMM이 윤리적 원칙을 통합하고 알려진 위험을 해결하는 데 어떤 역할을 할지는 불확실합니다. 그러나 일반적으로, AI 기술 설계에 오픈소스 소프트웨어를 사용하거나 소프트웨어의 소스 코드를 공개하면 투명성과 참여를

높일 수 있습니다*(1)*. 오픈소스 소프트웨어는 기여와 피드백을 받을 수 있도록 개방되어 있어, 사용자가 시스템 작동 방식을 이해하고 잠재적인 문제를 식별하며 소프트웨어를 확장하고 적응시킬 수 있습니다*(1)*. 오픈소스 LMM은 보건의료 분야에서 LMM 사용에 대한 몇 가지 우려를 해결할 수 있는 기회를 제공할 수 있습니다. 오픈 소스 LMM은 의료 분야에서 LMM 사용에 대한 일부 우려를 해결할 수 있는 기회를 제공할 수 있습니다. 오픈소스 모델은 독점적이지도, 폐쇄적이지도 않기 때문에 비영리 기관과 같은 소규모 기업이나 단체가 더 낮은 비용으로 LMM을 설계할 수 있습니다*(136)*. 오픈소스 모델을 기반으로 구축된 LMM은 코드와 데이터가 검토를 위해 공개되어 있어 면밀히 조사될 수 있습니다. 사용자 커뮤니티의 참여와 감시는 장기적으로 오픈소스 모델의 견고성을 보장하는 데 도움을 줍니다*(136)*.

그러나 이전에 모델을 공개했던 대규모 기술 기업들이 이를 계속 제공하지 않기로 선택할 경우, 오픈소스 LMM은 지속되지 않을 수 있습니다*(10)*. 대부분의 오픈소스 LMM 개발은 Meta(구 Facebook)가 제한적으로 출시한 LMM을 기반으로 이루어졌습니다*(10)*. 해당 LMM과 가중치가 유출된 이후*(137)*, Meta는 오픈소스 접근방식에 대한 약속-헌신을 밝히며, 개방성은 "더 나은 제품, 더 빠른 혁신, 번영하는 시장을 이끌어내며, 이는 [Meta]뿐만 아니라 많은 사람들에게 이익이 된다. 궁극적으로 개방성은 AI와 관련된 두려움에 대한 최선의 해독제다"라고 강조했습니다*(130)*. 그러나 독립적인 관찰자들은 Meta가 LMM을 비상업적 목적으로 제공하고 있지만, 사용 약관에 제한사항이 포함되어 있어 오픈소스 원칙과 일치하지 않는 방식으로 LMM을 제공하고 있다고 지적했습니다*(138,139)*.

오픈소스 모델의 성능과 결과를 모니터링하기 위한 추가 요구사항은 개발자들이 해결하기 어려울 수 있습니다. 하지만, 이러한 모델의 이점이 오픈소스 모델의 사용과 관련된 보안 문제와 같은 규제의 필요성과 보안 문제와 같은 규제 및 피해 회피의 필요성을 대체할 수는 없습니다. 오픈소스 모델은 오용에 취약하며*(141)*, 이러한 취약성을 악용하기 위해 공격받을 수 있습니다*(142)*. 최근 한 연구팀은 오픈소스 AI 시스템에서 테스트된 방법이 AI 안전 조치와 안전장치를 우회했으며, 이 방식은 이른바 폐쇄형 시스템(closed systems)의 안전장치도 우회할 수 있다고 밝혔습니다*(143)*. 궁극적으로 오픈소스 모델은 다른 LMM에서 사용되는 동일한 블랙박스 기술을 기반으로 하고 있습니다.

오픈소스 LMM을 장려하는 한 가지 방법은 정부가 정부 자금이나 지적 재산으로 구축된 기반 모델이 광범위하게 접근 가능하도록 요구하는 것입니다. 이는 정부가 정부 자금으로 수행된 연구에 대해 공개 접근를 요구했던 방식과 유사합니다. 정부는 또한 공공 시설에서 통제된 조건과 공공 감독 하에 차세대 모델을 포함한 오픈소스 연구 및 개발을 장려할 수 있습니다. Meta의 유출된 모델이 누구나 "MacBook M2에서 다운로드하고 실행할 수 있게"된 새로운 현실보다 공공 감독과 참여가 더 나은 대안이 될 수 있습니다*(144)*.

권장사항

- 의료, 과학 연구 또는 의학에서 사용되거나 사용될 가능성이 있는 LMM을 설계하는 개발자는 프로그래머를 위한 윤리 인증 또는 교육을 고려해야 합니다. 이는 AI 개발자를 의료 전문 분야의 요구사항과 일치시키고, 제품과 서비스에 대한 신뢰를 높일 수 있습니다.

- 데이터 세트의 크기와 관계없이, 개발자는 데이터 처리 전에 “데이터 보호 영향 평가"를 수행해야 합니다. 이는 데이터 처리 작업이 개인의 권리와 자유를 침해할 위험과 개인정보 보호에 미치는 영향을 평가하도록 요구합니다.

- 개발자는 최선의 데이터 보호 규칙에 따라 수집된 데이터를 기반으로 LMM을 훈련시켜야 합니다.

- LMM을 훈련하기 위해 개발자가 직접 수집했더나 제3자를 통해 수집된 모든 데이터 세트는 최신 상태로 유지되야 하며, 시스템이 사용될 수 있는 환경에 적합해야 합니다.

- 개발자는 모델을 훈련하는데 사용된 데이터 세트에 대한 투명성을 유지해야 하며 이를 통해 LMM을 세부 조정하거나 보건의료 응용 프로그램을 개발하거나 LMM을 직접 사용하는 사용자들이 훈련 데이터 세트의 부적합이나 불완전성을 인지할 수 있도록 해야 합니다. 는 개발자가 직접 수집했든 제3자를 통해 수집했든 관계없이 최신 상태를 유지해야 하며, 시스템이 사용될 수 있는 맥락에 적합해야 합니다.

- 개발자는 데이터 작업자에게 임금을 지급하고, 정신 건강 서비스 및 기타 상담지원을 제공해야 합니다. 또한 작업자가 겪을 수 있는 고통으로부터 보호하기 위한 안전장치를 도입해야 합니다. 정부는 이러한 혜택이 모든 데이터 작업자에게 적용되도록 노동 기준을 업데이트하고, 기업 간의 공정한 경쟁의 장을 조성하며, 이러한 노동 기준이 유지되고 시간이 지나면서 개선되도록 보장해야 합니다.

- 포괄적 설계 과정: LMM은 과학자와 엔지니어뿐만 아니라, 잠재적 사용자 및 의료 제공자, 과학 연구자, 보건의료 전문가, 환자 등 모든 직간접적인 이해관계자가 초기 개발 단계부터 구조적이고 포괄적이며 투명한 설계 과정에 참여할 수 있도록 해야 합니다. 이해관계자는 윤리적 문제를 제기하고, 우려를 표현하며, 검토 중인 AI 응용 프로그램에 대한 의견을 제공할 기회를 가져야 합니다. 이러한 의견은 "인간 감독 대학(human oversight colleges)"을 통해 제공될 수 있습니다.

- 개발자는 LMM이 명확히 정의된 작업을 수행하도록 설계해야 하며, 이를 통해 필요한 정확성과 신뢰성 확보하여 보건 시스템의 역량을 개선하고 환자 이익을 증진시켜야 합니다. 또한 개발자는 잠재적인 부수적 결과를 예측하고 이해할 수 있어야 합니다. 이를 위한 기법으로 “사전부검(pre-mortems)”과 “레드 티밍(red teaming)”이 포함됩니다.

- 개발자가 사용하는 절차는 합의된 기준, 모범관행(예: 개인정보 보호 기술 및 기법), 설계 윤리 기준, 그리고 발전하는 전문적 규범에 의해 정보를 제공받고 업데이트되어야 합니다. 여기에는 LMM이 생성한 콘텐츠가 AI 시스템에 의해 생성되었음을 공개하는 것도 포함됩니다.

- 개발자는 모델의 에너지 효율성을 개선하는 등 에너지 소비를 줄이기 위한 가능한 모든 조치를 취해야 합니다.

- 정부는 LMM 개발에 적용될 수 있는 강력하고 강제적인 데이터 보호법과 규정을 마련해야 합니다. 이 법률은 사람들의 권리를 효과적으로 보호하고, 의미 있는 동의권을 포함하여 사람들이 자신의 권리를 보호하는 데 필요한 도구를 제공해야 합니다. 의료 분야에서 사용되는 데이터를 수집하고 처리하기 위해 추가적인 도구가 필요할 가능성도 고려해야 합니다.

- 정부 및 WHO와 같은 국제 기관은 의료 및 의학 분야에서 사용될 LMM의 선호도와 특성을 명시하는 "대상 제품 프로필"을 발행할 수 있습니다. 특히, 정부가 공공 보건 시스템에서 사용할 목적으로 이러한 기술을 구매할 것및을 예상할 경우 이러한 조치가 유용합니다.

- 정부는 개발자들에게 범용 기반 모델의 설계 및 개발 과정에서 제품 수명 주기 동안 특정 결과를 달성하도록 요구해야 합니다. 여기에는 모델의 예측 가능성, 해석 가능성, 수정 가능성, 안전성 및 사이버 보안을 포함한 요구사항이 포함될 수 있습니다.

- 규제 기관은 윤리적 위험(예: 편향 또는 자율성 침해)을 식별하고 회피하도록 개발자들에게 요구하고 이를 장려하기 위해 사전 인증 프로그램과 같은 법적 의무 및 인센티브를 도입해야 합니다.

- 정부는 기반 모델 개발 초기 단계의 감사를 도입해야 합니다.

- 정부는 범용 기반 모델 개발자들에게 탄소 및 물 발자국에 대한 우려를 해결하도록 요구해야 합니다.

- 정부는 범용 기반 모델 사용 시, 생성된 콘텐츠가 인간이 아닌 기계에 의해 생성되었음을 사용자에게 알리고 상기시키도록 개발자들에게 요구해야 합니다.

- 정부는 의료 및 의학 분야에서 사용될 AI 알고리즘이나 시스템의 조기 등록을 요구하거나 이를 장려하기 위한 인센티브를 제공을 고려할 수 있습니다. 조기 등록은 부정적인 결과의 공개를 촉진하고, 출판 편향이나 결과에 대한 지나치게 낙관적 해석을 방지하며, 환자에게 유익한 지식의 통합을 용이하게 할 수 있습니다.

- 정부는 공공, 민간 및 비영리 부문의 개발자들이 접근할 수 있는 컴퓨팅 능력과 공공 데이터 세트를 포함한 비영리 또는 공공 인프라에 투자하거나 이를 제공해야 합니다. 이러한 인프라는 사용자들이 접근 조건으로 윤리적 원칙과 가치를 준수하도록 요구해야 합니다.

- 정부는 정부 자금이나 지적 재산으로 구축된 기반 모델이 광범위하게 접근 가능하도록 요구함으로써 오픈소스 LMM 개발을 장려해야 합니다. 이는 정부가 정부 지원 연구에 대한 공개 접근을 요구한 방식과 동일합니다. 정부는 공공 시설에서 공공 감독하에 통제된 조건으로 차세대 모델을 포함한 오픈소스 연구 및 개발을 지원해야 합니다.

5 범용 기반 모델(LMM)을 활용한 서비스 제공

범용 기반 모델의 사용은 사용자가 LMM을 활용해 의료 관련 결과물을 생성하도록 요청하는지, 또는 개발자가 LMM을 의료 관련 응용 프로그램, 제품, 또는 서비스에 통합하도록 허용하는지에 따라 달라집니다. 이 두 경우 모두 새로운 위험이 발생하며, 이는 개발자, 제공자, 또는 양측 모두에 의해 해결되어야 합니다. 정부는 이러한 기술이 배포되기 전에 사용을 평가하고 규제할 책임이 있습니다.

5.1 범용 기반 모델(LMM)을 사용하는 보건의료 서비스 및 응용 프로그램 제공 시 해결해야 할 위험

범용 기반 모델 및 응용 프로그램이 의료 목적으로 사용되거나 사용자가 직접적으로 이용할 경우, 이들 모두를 평가하고 승인해야 하는지에 대한 의견 불일치가 있을 가능성이 높습니다. 일부 주요 기술 기업은 정부가 범용 기반 모델에 대한 평가 프레임워크를 포기하고, 대신 정부가 “위험한” 것으로 간주할 수 있는 방식으로 사용될 가능성이 있는 응용 프로그램에 대한 감독에 초점을 맞추도록 정부 관계자(예: 유럽연합)를 비공식적으로 설득해 왔습니다*(145)*. 이는 범용 기반 모델을 포함한 의료 응용 프로그램을 생성하고 시장에 내놓는 제공자와, LMM을 직접적으로 또는 AI 시스템을 통해 간접적으로 사용하는 제공자나 환자와 같은 사용자 모두와 관련이 있습니다. 기업들은 범용 기반 모델의 감독이 개발자에게 “과도한 책임을 전가”하게 될 것이며, 가치사슬 내의 다른 주체들도 책임을 져야 한다고 주장합니다*(145)*.

범용 기반 모델 개발자를 모든 LMM 사용 사례에 대해 책임지게 하는 것은 적절하지 않을 수 있지만, 책임을 전적으로

제공자, 배포자 또는 사용자에게 전가하는 것 역시 부적절합니다. 이들은 모델 개발에 관여하지 않았고, 모델과 관련된 한계 및 위험을 이해하지 못할 가능성이 있기 때문입니다. 이로 인해 막대한 권력, 자원, 감독 및 LMM에 대한 이해를 가진 범용 기반 모델 개발자들이 책임을 회피하게 되며, 의료 AI 기술의 규제를 위한 시도에서 "중대한 허점"을 야기할 것입니다*(145)*.

개발자는 LMM을 의료 목적(또는 다른 용도)로 사용하지 않으려고 할 수도 있습니다. 개발자가 LMM이 의료나 의학적 목적(특히 임상 의학)으로 사용되지 않기를 원한다면, 보건 및 의학 응용 프로그램을 개발하는 주체가 응용 프로그램 프로그래밍 인터페이스를 통해 LMM을 사용(라이선스)하지 못하도록 방지하거나, LMM을 사용자(제공자 또는 환자)가 의료 목적으로 LMM을 직접 사용하는 경우, 보건 및 의료정보를 포함한 응답에 명확한 경고를 첨부하고 적절한 도움을 제공할 수 있는 정보나 서비스로 사용자를 안내할 수 있습니다.

이러한 조치가 이루어지지 않거나 개발자가 보건의료보건의료 분야에 제공자를 통해 직접 또는 간접적으로 LMM을 적용하고자 할 경우, 개발자는 개발자만이 충족할 수 있는 특정한 책임을 가지게 됩니다. 나아가, 개발자와 제공자는 보건의료 분야보건의료에서 LMM 사용과 관련된 위험을 해결하기 위한 추가적인 의무를 갖게 됩니다.

아래에 상세히 설명된 이러한 책임들은 정부가 궁극적으로 AI 기반 시스템이 보건의료에서 사용될 수 있는지 여부를 결정해야 하기 때문에, 정부가 법률, 정책 및 규제를 통해 정의합니다. LMM을 보건의료에 사용하려면 개발자와 공급자도 상호책임을 다해야 합니다. 이러한 책임은 법률이 작성되지 않았거나 갱신되지 않은 경우, 정부가 아닌 개발자와 제공자 간의 계약을 통해 정의될 수도 있습니다.

배포 전에 해결해야 할 주요 위험으로는 시스템 전반의 편향, 보건의료 용도로 인한 잘못된 정보 또는 환각(hallucination), LMM에 입력된 데이터의 개인정보, 조작 및 자동화 편향 등이 있습니다.

5.2 정부가 도입할 수 있는 위험 해결 조치 및 준수해야 할 윤리 원칙

LMM과 이를 포함한 응용 프로그램의 개발 속도가 빠른 만큼, 정부는 이러한 AI 알고리즘을 보건의료 시스템 및 기타 과학 및 의료 목적으로 사용하는 데 필요한 규제와 구체적인 기준을 신속히 마련해야 합니다. 이 접근방식은 의료기기 또는 제약 관련 기관과 같은 규제 기관에서 보건의료나 의학 목적으로 사용하려는 AI 기술을 평가하고 승인하는 것으로 구성되어야 하며, 정부가 이를 위해 새로운 기관을 설립할 수도 있습니다. 그러나 저소득 및 중소득 국가의 경우, 이미 자원이 부족하고 제약 규제로 인해 과부하 상태에 있는 규제 기관이 이러한 역할을 수행하기 어려운 점이 도전 과제입니다.

일부 고소득 국가의 정부는 주요 기술 기업들과 합의하여 해당 국가의 기반 모델을 자발적 공공 평가를통해 평가받도록 하고, 그 결과를 공개하여 대중과 연구자에게 모델에 대한 정보를 제공하며, 기업들이 오류를 수정하도록 장려하고 있습니다*(146)*. 그러나 자발적 접근방식은 충분하지도 지속 가능하지도 않을 가능성이 높습니다.

LMM과 응용 프로그램에 대한 평가는 보건의료 시스템에서 사용되는 AI 시스템이나 알고리즘만을 대상으로 해서는 안되며, 임상과 "웰니스(Wellness)" 응용 프로그램 사이의 회색 지대에서 LMM 및 응용 프로그램의 사용과 관련된 상당한 위험도 다루어야 합니다. 이러한 기술의 급속한 확산을 감안할 때, 정부는 최소한 초기 단계에서는 이러한 응용 프로그램을 식별하고, 공통 기준과 규제를 설정하며, 해당 기준과 규제를 충족하지 않는 응용 프로그램이 대중에게 배포되지 못하도록 금지해야 합니다.

개발자와 제공자는 보건의료 사용을 목적으로 하는 AI 기술이 법률이나 정책에 명시된 최소 요구 사항을 충족함을 증명할 책임을 져야 합니다. LMM과 관련된 알려진 위험과 과제를 고려할 때, AI 알고리즘 및 LMM 응용 프로그램이 안전하고 효과적이라고 가정해서는 안 되며, 이미 널리 사용되고 있는 AI 또는 비AI 기반 접근방식보다 우수하다고도 가정해서는 안 됩니다.

다음은 보건의료 및 의학에서 LMM 사용에 적용될 수 있는 여러 법률, 정책 및 횡단적 요구 사항에 대한 설명입니다.

공개(투명성) 요구 사항: 적절한 규제는 정부가 평가 및 승인할 수 있는지 결정할 수 있는 역량과 재량뿐만 아니라, 이러한 평가를 수행하기 위한 충분한 정보도 필요로 합니다. AI 기술을 적절히 규제하고 AI 가치사슬의 다른 주체들이 기술을 안전하게 사용할 수 있도록 보장하기 위해 필수적입니다. 예를 들어, 개발자가 범용 기초 모델의 성능[예: 환각(hallucination) 경향성]을 공개하지 않을 경우, 제공자는 모델을 세부 조정하거나 기술 마케팅을 피하기 위해 필요한 정보를 갖지 못할 수 있습니다. 제공자나 개발자의 이러한 형태의 공개는 의료 제공자와 같은 사용자가 부정확한 정보를 제공할 가능성이 있는 LMM의 사용을 피하거나 결과물을 더 철저히 검토하도록 도울 수도 있습니다.

공개와 투명성은 WHO의 기본 원칙이자 AI 기반 시스템의 "설명 가능성" 및 이해 가능성을 향상시키는 조치이며, 범용 기반 모델 또는 응용 프로그램의 평가에서 요구되어야 합니다. WHO의 보건의료를 위한 AI 윤리 및 거버넌스 가이드라인(1)은 "정부 규제 기관은 안전성과 효율성에 대한 감독 및 평가를 개선하기 위해 독점권을 고려하면서 AI 기술의 특정 측면의 투명성을 요구해야 한. 여기에는 AI 기술의 소스 코드, 데이터 입력 및 분석 접근방식이 포함될 수 있다"고 권고했습니다. LMM과 관련된 새로운 형태의 공개에는 내부 테스트에서의 성능, 탄소 및 물 발자국 등이 포함될 수 있습니다. 또한, 알고리즘 학습의 결과물이나 LMM이 학습 중에 얻은 지식을 이해할 수 있도록 하는 "개방형 가중치(open-weight)"에 대한 기준도 필요할 수 있습니다*(147,148)*.

여러 형태의 공개는 제공자, 사용자 또는 규제 기관에 도움을 줄 수 있습니다. 예를 들어, 기반 모델의 능력과 한계를 설명하고, 공공 또는 업계 표준 벤치마크에 따라 모델을 평가하며, 모델 및 최적화의 내부 및 외부 테스트 결과를 보고하는 것 등이 포함됩니다*(134)*. 특히 LMM이나 응용 프로그램과 관련된 위험에 대한 공개는 명확하게 광고될 수 있으며, 이는 한 연구자가 "영양 성분표(nutrition label)"에 비유하기도 했습니다*(129)*.

데이터 보호법: LMM의 개발과 개발자가 LMM 학습에 필요한 데이터를 관리하는 방식은 데이터 보호법을 위반할 가능성이 있습니다. 별도의 문제로, LMM이나 응용 프로그램에 특정 출력을 생성하기 위해 입력된 데이터가 민감한 개인 정보를 포함할 경우, 실수로 또는 프롬프트를 통해 공개될 위험이 있습니다. 이러한 잠재적 정보 공개는 LMM을 개발하고 상용화하는 기술 기업을 포함한 많은 대기업들이 자사 직원들이 이러한 알고리즘을 사용하는 것을 금지하는 이유 중 하나입니다*(149)*.

데이터 공개는 자율성을 보호할 책임을 지닌 개발자의 책임을 위반하는 것입니다. 또한, 개발자는 데이터 최소화 요구사항에 따라 민감한 데이터를 허용된 기간 이상 보유할 경우 데이터 보호 법률을 위반할 수 있습니다*(85)*. 한 개발자는 사용자가 챗봇 성능을 개선하기 위해 제공하는 콘텐츠를 제외할 수 있는 선택권을 제공합니다*(150)*. LMM 사용을 허용하는 정부는 LMM에 입력된 데이터를 보호하기 위한 데이터 보호 규칙을 설정, 확장 및 시행해야 합니다. 중국 정부의 LMM 규정에는 이러한 요구사항이 포함되어 있지만, 이는 중국 사용자에게만 적용됩니다*(151)*.

보건의료에 사용되는 범용 기반 모델 및/또는 응용 프로그램 평가 – 인권법 대 위험 기반 프레임워크: AI 기술을 평가하고 규제하기 위한 여러 법적 프레임워크가 개발되고 있습니다. 이러한 프레임워크와 관련하여, AI 기술이 인권 의무(유럽연합 "기본권")를 충족해야 하는지, 아니면 위험 기반 프레임워크에서 다른 방식으로 평가되어야 하는지가 논의되고 있습니다. 유럽연합은 AI 법 하에 위험 기반 프레임워크를 채택했습니다*(152)*. 위험 기반 프레임워크는 기술의 위험 수준에 따라 요구사항 또는 입증 책임을 정의하는 데 도움을 줄 수 있으며, 위험 수준이 높아질수록 입증 책임이 증가합니다.

모든 보건의료 및 의학에서 사용되는 AI 시스템이나 도구는 개인의 존엄성, 자율성 또는 개인정보와 같은 윤리적 의무와 인권 기준을 존중해야 합니다. 여기에는 범용 기반 모델도 포함됩니다. 인권과 윤리 원칙은 협상의 대상이 아니며, AI 기술이 가지는 위험이나 잠재적 이점에 관계없이 반드시 준수되어야 합니다*(153)*. AI 알고리즘이 "저위험"으로 간주된다고 해서 검토 대상에서 제외되는 것은 아니며, 개발자나 제공자는 알고리즘이 인권 및 윤리적 의무를 준수하는지 확인해야 합니다. LMM이나 응용 프로그램이 이러한 약속을 준수하고 안전하게 사용될 수 있는지 여부를 확인하기 위해 인권 영향 평가를 수행할 수 있습니다.

WHO의 보건의료 AI 윤리 및 거버넌스 가이드라인*(1)*은 "정부는 AI 시스템의 전체 수명 주기에 걸쳐 윤리, 인권, 안전 및

데이터 보호를 다루는 AI 기술의 영향 평가를 정부 기관과 기업이 수행하도록 요구하는 법률과 정책을 제정해야 한다"고 권장했습니다. 또한 "영향 평가는 AI 기술 도입 전후에 독립적인 제3자에 의해 감사를 받아야 하며, 결과는 공개되어야 한다"고 강조했습니다*(1)*. 영향 평가 결과는 독점 또는 민감한 정보를 고려하여 대중 및 영향을 받을 수 있는 단체에 공개되어야 합니다. 감사와 마찬가지로(위 참조), 영향 평가가 도구나 서비스를 제공하는 제3자에 의해 수행되는 경우, 그 품질과 엄격성을 보장하기 위해 면밀히 검토해야 할 수 있습니다.

영향 평가를 통해 AI 기술이 시스템 전반에 걸친 편향을 유발하거나, 개인 데이터를 공유하는 사용자의 개인정보를 노출시키거나, 사용자를 조작할 가능성이 있는지를 밝혀낼 수 있습니다. 개인정보 위험은 개별 개인정보를 보호하는 LMM을 개발하기 위해 제공자와 개발자가 협력하여 해결해야 합니다. 미국의 한 기업과 병원 시스템이 이러한 LMM 개발을 진행 중이지만, 데이터의 완전한 비식별화가 불가능하다는 이유로 프로젝트가 성공할 가능성은 낮다고 평가받고 있습니다*(154)*. 영향 평가는 또한 범용 기반 모델이나 응용 프로그램 사용 시, 자동화된 의사결정을 피하고 인간이 판단에 참여하도록 보장할 수 있습니다. 이는 사용자가 LMM의 출력 결과에 의존하거나 잘못된 정보를 받는 상황, 혹은 의료 제공자나 환자가 LMM의 출력을 비판 없이 수용하는 "자동화 편향"을 방지하기 위한 것입니다.

정부는 보건의료나 의학에서 사용되는 LMM에 대해 위험 기반 프레임워크를 채택할 수 있습니다. 심각한 우울증 환자에게 처방전을 제공하거나 정신 건강 조언을 제공하는 것과 같은 고위험 기능이나, 취약하거나 소외된 집단에 의해 사용되는 AI 기술은 더 높은 입증 책임을 요구받을 것입니다. 그러나 정부가 위험 기반 접근방식을 선택할 경우, 이 접근방식이 충분하다고 간주되거나 인권 기반 접근방식을 대체할 수 있다는 우려가 있습니다*(153)*. 위험 기반 프레임워크는 특정 LMM이나 응용 프로그램을 평가에서 제외할 수 있는데, 이러한 응용 프로그램은 저위험으로 보일 수 있지만 결국에는 피해를 초래할 가능성이 있습니다.

추가적인 논의는 AI 규제 평가가 궁극적인 사용 용도와 관계없이 모든 기반 모델에 적용되어야 하는지, 아니면 가장 크고 널리 사용되는 기반 모델("시스템 기반 모델"[7])에만 적용되어야 하는지, 그리고 언제 제공자에게 이러한 평가가 적용되어야 하는지에 관한 것입니다.

본 가이던스는 모든 기반 모델이 사용 목적과 관계없이 위험 기반 및/또는 인권 기반 평가를 받아야 하는지 여부에 대한 권장사항을 포함하지 않습니다. 또한 범용 기반 모델에 대한 AI 규제 평가가 가장 큰(시스템 기반) LMM에만 적용되어야 하는지에 대한 권장사항도 포함하지 않습니다. 전문가 그룹은 규모와 관계없이 모든 LMM에 평가를 적용하도록 설계된 경우, 대규모 기업들의 지배적 위치를 "고착화"시킬 수 있다는 우려를 제기했습니다. 이는 표준이 오직 대규모 기업들만

7 Kai Zenner, 유럽의회, Axel Voss 사무소장, AI for Good 컨퍼런스, 2023년 6월 5일 발표.

준수할 수 있는 수준이거나 해당 표준이 그들의 비즈니스 모델과 목표에 가장 적합할 수 있기 때문입니다*(155)*. 이러한 우려는 경쟁 당국의 관심을 끌고 있으며, 경쟁 당국은 "선도 기업들이 현재의 권력을 공고히 하거나 새로운 생성형 AI 시장을 지배하기 위해 부당한 경쟁 방법을 사용할 가능성"을 발견했습니다*(141)*. 경쟁 당국은 LMM을 개발하는 기업들의 사용관행에 초점을 맞추고 있지만, LMM 사용에 대해 더 많은 조사를 실시할 것으로 예상됩니다*(156)*.

제공자는 LMM의 목적 및 기능을 개발자가 설정한 것과 다르게 변경할 수 있기 때문에 AI 규제 평가를 받아야 합니다. 따라서 제공자가 보건의료나 의학 용도로 범용 기반 모델을 사용하도록 조정하고 개발자가 이를 동의한 경우, 개발자와 제공자는 보건의료 및 의학에서 LMM 사용 요구사항을 준수해야 합니다. 제품이나 응용 프로그램의 사용이 기반 모델과 크게 다르거나 개발자의 통제를 벗어난 방식으로 변경될 경우, 제공자가 부담하는 규제 책임은 더 커질 것입니다.

의료기기 규제: 정부는 범용 기반 모델이나 응용 프로그램이 의료기기로 간주되는지 여부를 결정할 수 있습니다. 현재 어떤 LMM이 의료기기로 간주되는지에 대한 가이던스는 적지만, 한 규제 기관은 "범용으로만 사용되고 개발자가 해당 소프트웨어가 의료 목적으로 사용될 수 있다고 주장하지 않는 LMM은 의료기기로 간주되지 않을 가능성이 높다"고 말했습니다*(157)*. 그러나 "의료 목적으로 개발되거나 조정, 수정 또는 특정 의료 목적을 위해 설계된 LMM은 의료기기로 간주될 가능성이 높다. 또한, 개발자가 자신의 LMM이 의료 목적으로 사용될 수 있다고 주장하는 경우, 이는 해당제품이 의료기기로 분류될 가능성이 높다는 것을 의미한다."고 덧붙였습니다*(157)*.

LMM을 기반으로 한 챗봇이 의료 상담을 제공할 경우, 이는 현재의 유럽연합 및 미국 규제 기준에 따라 의료기기로 분류될 가능성이 높습니다*(158)*. WHO의 보건분야 AI 윤리 및 거버넌스 가이드라인*(1)*은 "정부 규제 기관은 AI 시스템의 성능을 테스트하고 무작위 시험에서 예측 가능한 테스트로부터의 건전한 증거를 요구해야 하며, 단순히 기존 데이터 세트와의 실험실 비교에만 의존해서는 안 된다"고 권고했습니다.

LMM이나 응용 프로그램이 의료기기로 규제되려면, 개발자와/또는 제공자가 제품이 홍보된 대로 작동하며, 현재의 국가 법률 또는 개정된 법률의 요구 사항을 충족한다는 증거를 제시할 책임을 져야 합니다. 이는 편향 및 개인정보와 관련된 윤리적 의무를 준수하는 것과 같은 다양한 요구사항을 포함할 수 있습니다. 유럽연합과 미국에서 제안된 새로운 의료기기용 AI 기술 규정은 "설명 가능성", 편향 통제, 투명성 등 AI의 의료 분야 사용과 관련된 윤리적 원칙을 통합할 가능성이 높습니다. 그러나 현재 LMM을 포함한 챗봇은 이러한 기준을 충족할 가능성이 낮습니다*(158)*.

임상 의사결정을 지원하기 위한 LMM은 이미 실험적으로 사용되고 있습니다. 이러한 LMM에는 면책조항이 포함되어 있지만, 이는 의료기기 법률의 적용을 면제하지 않습니다. 해당 법률에 따르면, 이러한 실험은 환자를 보호하고 임상적으로 관련된 결과를 생성하기 위해 적절한 통제가 이루어진 허가된 임상 시험 환경에서만 진행되어야 합니다*(158)*. 정부는 이러한 LMM의 통제된 실험적 사용을 규제 "샌드박스"에서 검토할 수 있습니다. 샌드박스는 실제 임상 환경에서

보호 장치와 감독 하에 실험을 허용하여 보건 시스템이 위험이나 의도치 않은 결과로부터 보호받도록 합니다. 그러나 이러한 사용은 새로운 의료 제품 및 서비스와 그 사양이 공식 규제를 받는 국가에서만 적합할 수 있습니다*(1)*.

소비자 보호법: 정부는 LMM과 응용 프로그램의 부정적 결과가 사용자와 환자에게 미치지 않도록 소비자 보호법을 개발하고 활용해야 합니다. 소비자 보호법은 예를 들어, 조작에 해당할 수 있는 관행을 방지하기 위해 적용될 수 있습니다*(159)*. 미국에서는 여러 정부 부처와 기관이 소비자 보호법 및 기타 규정을 활용하여 자동화 시스템에서의 차별과 편향을 방지하고 있습니다*(159)*. 이러한 법률은 기술 상용화를 추구하는 기관들이 부정적 결과의 원인을 해결하고 환자와 그 가족을 현재 또는 미래의 피해로부터 보호하도록 요구할 수 있습니다*(93)*. 소비자 보호법이나 기타 규정을 통해 LMM과 응용 프로그램이 사용자를 오도하거나 인간과 유사한 특성을 부여하는 언어 사용을 제한하도록 요구할 수 있습니다. 예를 들어, “내가 생각하기에”, “내가 추측하기에”, “내가 제안하기에”와 같은 표현의 사용을 제한하거나 금지할 수 있습니다.

권장사항

- 정부는 자원이 허용하는 한, 보건의료나 의학에 사용하려는 LMM 및 응용 프로그램을 평가하고 승인할 기존 또는 새로운 규제 기관을 지정해야 합니다.

- LMM과 그 응용 프로그램의 특정 측면은 규제 기관이 안전성과 유효성을 감독하고 평가할 수 있도록 투명해야 합니다. 여기에는 소스 코드, 데이터 입력, 모델 가중치, 분석적 접근방식 등이 포함될 수 있습니다. 정부는 LMM이나 응용 프로그램의 내부 테스트 성능 및 탄소와 물 발자국에 대한 추가 공개를 고려해야 합니다.

- 정부는 사용자가 LMM이나 응용 프로그램에 입력한 데이터에 데이터 보호 규칙이 적용되도록 해야 합니다.

- 정부의 법률, 정책 및 규정은 AI 기술의 위험이나 이점과 관계없이, 보건의료나 의학에 사용되는 LMM 및 응용 프로그램이 개인의 존엄성, 자율성 또는 개인정보와 같은 윤리적 의무와 인권 기준을 충족하도록 보장해야 합니다.

- 정부는 제공자와 개발자가 LMM과 응용 프로그램의 전체 수명 주기에 걸쳐 윤리, 인권, 안전 및 데이터 보호를 다루는 영향 평가를 수행하도록 요구하는 법률과 정책을 제정해야 합니다. 영향 평가는 독립적인 제3자에 의해 감사되어야 하며, AI 기술 도입 전후에 공공 영역에 공개되어야 합니다.

- 제공자가 제품이나 응용 프로그램을 기반 모델에서 크게 벗어나게 하거나 개발자의 통제를 벗어난 방식으로

변경할 경우, 제공자가 부담하는 규제 책임은 증가해야 합니다.

- 정부는 의료기기로 규제를 받는 LMM이나 응용 프로그램의 경우, 제품이 홍보된 대로 작동하며 국가의 법률 또는 개정된 법률의 요구사항을 충족한다는 입증 책임을 개발자와/또는 제공자가 지도록 보장해야 합니다.

- 정부는 아직 사용 승인을 받지 않은 임상 의사결정을 지원하는 LMM이나 응용 프로그램이 허가된 임상 시험 환경 외에서 실험적으로 사용되지 않도록 해야 합니다. 정부는 규제 "샌드박스"를 통해 LMM의 통제된 실험적 사용을 지원할 수 있습니다. 이를 통해 실제 임상 환경에서 테스트를 허용하며, 보건 시스템을 위험한 환경이나 의도치 않은 결과로부터 보호하기 위한 안전장치와 감독을 제공할 수 있습니다.

- 정부는 소비자 보호법을 활용하여 LMM과 응용 프로그램의 사용으로 인해 발생할 수 있는 부정적 결과가 사용자, 특히 환자에게 영향을 미치지 않도록 해야 합니다. 소비자 보호법은 예를 들어, 환자와 그 가족을 현재 또는 미래의 피해로부터 보호하기 위해 조작에 해당할 수 있는 관행을 방지하거나, LMM이나 응용 프로그램의 다른 부정적 결과의 원인을 해결하도록 적용될 수 있습니다.

6 범용 기반 모델(LMM)을 활용한 배포

LMM이나 LMM이 포함된 응용 프로그램이 윤리적으로 설계되고 적절한 규제 검토를 거쳤더라도 상용화될 경우 여전히 위험을 수반할 수 있습니다. AI 보건의료 응용 프로그램이나 도구를 배포하는 주체는 LMM이나 응용 프로그램의 개발자 또는 제공자일 수도 있고, 예를 들어 보건부, 병원, 보건의료 기업 또는 제약 회사일 수도 있습니다.

6.1 범용 기반 모델(LMM)을 사용하는 보건의료 서비스 또는 응용 프로그램 배포 시 해결해야 할 위험

배포 중 발생할 수 있는 위험은 LMM의 예측 불가능성과 그 응답, 개발자나 제공자가 예상하지 못한 방식으로 범용 기반 모델이 사용될 가능성, 그리고 LMM이 생성하는 응답이 시간이 지남에 따라 변할 가능성에 기인할 수 있습니다.

배포 시 해결해야 할 주요 위험은 다음과 같습니다:

- 부정확하거나 잘못된 응답,

- 편향,
- LMM에 입력되고 출력되는 데이터의 개인정보,
- LMM의 접근성 및 경제성,
- 노동 및 고용에 미치는 영향,
- 자동화 편향과 기술 저하,
- 의료 제공자와 환자 간 상호작용의 수준

이 섹션에서는 AI 가치사슬 내의 행위자, 특히 사용자들이 위험을 완화하거나 예방하는 방법과 LMM이 배포된 후 AI 도구 사용을 규제하는 데 있어 정부의 역할, 그리고 보건의료 시스템 내의 의료 종사자 및 기타 참여자들이 LMM을 적절히 사용할 수 있도록 교육하고 역량을 강화하는 방법에 대해 설명합니다.

6.2 배포 중 개발자와 제공자의 지속적 책임

LMM이나 응용 프로그램이 사용 승인을 받은 이후에도 개발자와 제공자는 책임과 의무를 집니다. 이는 개발자나 제공자가 LMM을 직접 배포하거나, 특정 위험이 배포 이후에만 해결될 수 있기 때문입니다. 이러한 의무는 개발자와 제공자가 적절한 자원과 관심을 할당하도록 보장하기 위해 규정이나 법률로 요구되어야 할 수 있습니다.

첫째, 정부는 LMM이 대규모로 배포될 경우 독립적인 제3자가 데이터 보호 및 인권을 포함한 사후 감사와 영향 평가를 의무적으로 실시하도록 해야 합니다*(155,160)*. 출시 후 감사 및 영향 평가는 공개되야 하며, 연령, 인종 또는 장애와 같은 사용자의 특성에 따라 결과와 영향을 세분화하여 발표해야 합니다.

둘째, 정부는 LMM 출시 후 개발자나 제공자가 부정확하거나 잘못되거나 유해한 LMM 콘텐츠를 수정하거나 방지하기 위한 조치를 취하지 않았을 경우, 이에 대한 책임을 묻도록 할 수 있습니다. 예를 들어, 중국 정부의 생성형 AI에 대한 규제는 "허위 및 유해"한 정보를 생성해서는 안 된다고 명시하고 있으며, 정부가 이를 집행할 수 있습니다*(151)*. 유럽연합에서는 제품이나 서비스에 LMM을 추가할 경우 개발자와 LMM 제공자에게 추가적인 책임이 발생할 수 있습니다. 예를 들어, LMM의 유럽연합 디지털 서비스법(Digital Services Act)과 같은 디지털 서비스 규제범위 내의 서비스에 통합될 경우 LMM은 간접적으로 규제검토 대상이 될 수 있으며, 환각(hallucination) 경향 때문에 규제 대상이 될 수 있습니다*(120)*.

셋째, 개발자와 제공자는 정부와 사용자가 LMM을 안전하게 사용할 수 있도록 충분한 기술 문서를 포함한 지속적인 운영 정보를 제공해야 할 수 있습니다*(133,134)*.

6.3 배포자의 책임

배포자는 LMM이나 응용 프로그램 사용과 관련된 위험을 방지하거나 완화할 책임이 있습니다.

우선, 배포자는 개발자나 제공자로부터 받은 정보를 바탕으로 LMM이나 응용 프로그램이 부적절한 환경에서 사용되지 않도록 해야 합니다. 이는 훈련 데이터에 포함된 편향, 환경에 맞지 않는 상황적 편향 또는 제공자가 미리 알고 피할 수 있는 오류나 잠재적 위험 때문일 수도 있습니다. 또한 배포자가 이러한 위험에 대한 명확하고 충분한 경고를 받았음에도 불구하고 부적절한 환경에서 사용할 수 있도록 LMM을 배포할 경우, 그로 인해 발생한 피해에 대한 책임을 져야 합니다.

둘째, 배포자는 LMM 사용으로 인해 발생할 수 있는 위험과 사용자에게 피해를 준 오류나 실수를 반드시 알릴 책임이 있습니다. 이러한 경고는 작은 글씨로 적혀 있거나 쉽게 놓칠 수 있는 형태로 배포되어서는 안 됩니다. 특정 상황에서는 법률이나 규정에서 요구하지 않더라도, 배포자는 피해를 방지하기 위해 LMM이나 응용 프로그램의 사용을 중단하거나 시장에서 제거할 책임이 있을 수 있습니다.

셋째, 배포자는 LMM의 경제성과 접근성을 향상시키기 위한 조치를 취할 수 있습니다. 배포자는 LMM 사용에 대한 가격이나 구독료가 정부나 기타 사용자의 지불 능력에 부합하도록 보장해야 하며, 기술의 혜택에서 소외된 사람들도 사용할 수 있도록 적합한 LMM이 여러 언어와 문자로 학습되고 배포되도록 해야 합니다. 또한, 배포자는 제공자와 개발자에게 현재 및 미래의 LMM이 여러 언어로 제공되도록 요청해야 합니다.

6.4 정부 프로그램 및 관행

LMM을 보건의료 시스템 및 기타 보건의료 관련 용도에 도입하는 것은 의료 종사자에게 상당한 적응을 요구하게 됩니다. 개발자나 제공자 모두 보건의료 전문가들이 LMM을 적절히 사용하도록 전문교육을 하거나 전문지식을 가진 개인이 관여하여 기타 용도에 사용하는 것을 관리할 수 있는 충분한 여력을 갖추고 있지 않습니다.

범용 기반 모델의 설계와 마찬가지로(위 참조), 정부는 임상 의사결정에서 사용되는 새로운 LMM과 응용 프로그램이

환자의 권리를 훼손하지 않고 적절히 사용되도록 보장하기 위해 의료 종사자와 환자 모두를 "인간 감독 협회(human oversight colleges)"에 참여시킬 수 있습니다.[8]

정부, 대학(보건과학 학부) 또는 병원과 같은 의료 제공자는 의료 종사자들이 LMM을 활용하여 효과적으로 임상 진료를 제공하고, 다른 용도에 대해 적절히 훈련받도록 보장할 수 있습니다. 보건의료 전문가와 임상의는 다음을 훈련받아야 합니다: (i) LMM이 어떻게 결정을 내리는지와 이러한 결정 과정의 한계를 이해하는 방법, (ii) 적절한 사용에 대한 우려를 식별하는 방법, (iii) 자동화 편향을 피하는 방법, (iv) LMM을 고려하거나 사용하는 환자와의 상호작용 및 교육 방법, (v) LMM 사용과 관련된 사이버 보안 위험*(161)*.

의료 종사자들이 교육과 지속적인 교육을 제공하는 것은 매우 중요합니다. 이는 LMM이 생성한 조언이 환자, 일반인 또는 제3자에게 전달되거나, 의료 제공자가 의료적인 결정이나 다른 의료적 목적으로 LMM이 제공한 정보를 활용했을 때 이를 알리기 위해서 입니다. 이러한 알림과정에서 환자나 일반인은 LMM 사용과 관련된 위험에 대해 충분히 설명을 들어야 하며, 이를 통해 본인의 사전 동의를 받을 권리를 보장받아야 합니다.

의료 종사자들 교육은 또한 LMM을 전문적으로 사용할 때, 자신의 업무가 법률(특히 보건 데이터와 정보 보호와 관련된 법률)을 의도치 않게 위반하지 않도록 보장하는 것이 중요합니다. 예를 들어, 의료 제공자가 LMM 챗봇에 "보호된 보건 정보"를 입력하면, 미국의 의료보험의 양도 및 책임에 관한 법률(HIPAA)과 같은 법률을 위반할 수 있습니다*(150)*. 인기 있는 LMM이 의료 종사자들에게 "신뢰받는" 도구가 될 경우, 의료 종사자들은 자신도 모르는 사이에 더 많은 환자 데이터를 공개할 위험이 있습니다*(154)*.

의료 시스템의 다른 이해관계자들은 보건의료에서 LMM의 이점, 위험, 용도 및 과제와 LMM이 정보 생성 또는 조언 제공을 위한 다른 기술과 어떻게 다른지, 그리고 보건의료에서 다른 용도로 어떻게 사용되었는지에 대해 교육을 받아야 합니다. 보건의료 및 기타 분야에서 LMM 사용에 대한 대중의 인식을 개선해야 합니다. WHO의 보건의료 AI 윤리 및 거버넌스 가이드라인*(1)*은 다음과 같이 권고합니다:

"대중은 건강을 위한 AI 개발에 참여하여 데이터 공유 및 사용 방식에 대해 이해하고, 사회적, 문화적으로 수용 가능한 AI 형태에 대해 의견을 제시하며, 자신들의 우려와 기대를 완전히 표현해야 합니다. 또한, 대중의 AI 기술에 대한 이해를 높여

8 WHO 보건분야 인공지능의 윤리와 거버넌스 전문가, David Gruson의 단보(Communication).

대중이 어떤 AI 기술이 수용 가능한지 결정할 수 있도록 해야 합니다."

정부가 보건의료 시스템에 LMM 또는 응용 프로그램을 제공하는 경우, 조달 권한을 활용하여 개발자, 제공자 및 배포자들 사이에서 특정 관행을 촉진할 수 있습니다. 보건의료 시스템에서 사용될 중요한 LMM이나 응용 프로그램을 조달할 때에는 AI 기술이 더 효과적이고 공정하며 경제적인 다른 보건의료 투자들을 대체하지 않는 한, 접근성과 경제성의 장벽을 없앨 수 있습니다. 공공 조달은 데이터 학습, 품질 보증, 위험 평가, 위험 완화 및 외부 감사와 관련된 투명성 요구사항을 설정할 수 있습니다. 이는 관련 법률이 없거나 LMM을 효과적으로 규제할 자원을 가진 규제 기관이 없는 국가에서 특히 중요할 수 있습니다.

권고사항

- 정부는 LMM이 대규모로 배포될 때 독립적인 제3자가 데이터 보호 및 인권을 포함한 사후 감사 및 영향 평가를 의무적으로 실시하도록 해야 합니다. 이러한 감사와 평가 결과는 공개되야 하며 연령, 인종 또는 장애 등을 포함하여 사용자 유형별로 세분화된 결과 및 영향을 포함해야 합니다.

- 정부는 LMM이 출시된 후 제공자나 개발자가 수정하거나 방지하지 않은 부정확하거나 잘못된, 또는 유해한 콘텐츠에 대해 제공자나 개발자에게 책임을 물을 수 있습니다.

- 정부는 개발자와 제공자가 LMM과 응용 프로그램을 안전하게 사용할 수 있도록 지속적인 운영 정보를 공개하도록 요구해야 합니다. 여기에는 충분한 기술 문서가 포함될 수 있습니다.

- 개발자나 제공자로부터 얻은 정보를 바탕으로, 배포자는 추가 훈련 데이터의 편향, 특정 환경에 부적합게 만드는 상황적 편향, 또는 제공자가 알고 있고 피할 수 있는 부정확하거나 잘못된 유해한 콘텐츠와 같은 잠재적 오류나 위험 때문에 LMM이나 응용 프로그램을 부적절한 환경에서 사용하지 않아야 합니다.

- 배포자는 LMM 사용으로 인해 발생할 수 있는 위험과 사용자에게 피해를 준 오류를 반드시 고지해야 하며, 이러한 경고는 작은 글씨나 쉽게 놓칠 수 있는 형태로 제공되어서는 안 됩니다. 특정 상황에서는 법률이나 규정에서 요구하지 않더라도, 배포자는 미래의 피해를 방지하기 위해 LMM 또는 응용 프로그램의 사용을 중단하거나 시장에서 제거할 책임이 있을 수 있습니다.

- 배포자는 LMM의 경제성과 접근성을 개선해야 하며, LMM 사용에 대한 가격이나 구독료가 정부나 기타 사용자의 지불 능력에 부합하도록 보장해야 합니다. 또한, 기술 혜택에서 소외된 사람들에게 도달할 수 있도록 적절한 LMM이 다양한 언어와 문자로 학습되고 제공되도록 해야 합니다. 배포자는 제공자와 개발자에게 현재 및 미래의 LMM이 여러 언어로 개발되도록 요청해야 합니다.

- 정부는 임상 의사결정에 사용되는 새로운 LMM 및 응용 프로그램이 적절히 사용되고 환자의 권리를 훼손하지 않도록 보장하기 위해 의료 종사자와 환자가 참여하는 "인간 감독 대학(human oversight colleges)"을 지원해야 합니다.

- 보건부와 대학(보건학부)은 의료 전문가와 임상 의사들에게 (i) LMM이 어떻게 결정을 내리는지 이해하도록 (그리고 이러한 결정이 어떻게 내려지는지 이해하는 데 한계가 있음을 인식하도록), (ii) 적절한 사용에 대한 우려를 식별하고 이해하도록, (iii) 자동화 편향을 피하는 방법에 대해 교육하도록, (iv) LMM을 사용 중이거나 사용을 고려 중인 환자들과 소통하고 교육하도록, 그리고 (v) LMM 사용과 관련된 사이버 보안 위험에 대해 교육해야 합니다.

- 정부, 의료 서비스 제공자, 보건 연구자 및 기금 제공자는 대중이 데이터 공유 및 사용의 다양한 형태를 이해하고, LMM이 사회적 및 문화적으로 수용 가능한지 여부에 대해 의견을 제시하며, 우려와 기대를 충분히 표현할 수 있도록 대중과 소통해야 합니다. 또한, 대중의 AI 기술에 대한 이해를 향상시켜, 수용 가능한 LMM의 사용 방식과 유형을 식별할 수 있도록 해야 합니다.

- 보건의료 시스템을 통해 LMM 또는 응용 프로그램을 제공하는 정부는 조달 권한을 활용하여 개발자, 제공자 및 배포자 간 투명성을 포함한 특정 관행을 촉진해야 합니다.

7 LMM에 대한 법적 책임

LMM이 보건의료 및 의학 분야에서 점점 더 널리 사용됨에 따라, 오류, 오용, 궁극적으로 개인에게 해를 끼치는 일이 불가피하게 발생할 수 있습니다. 이러한 피해에 대해 개인에게 보상하기 위해 책임 규칙을 사용해야 하며, 기존 방식이 부족하거나 시대에 뒤떨어질 경우 새로운 구제 방안을 마련해야 합니다.

AI 기술의 설계, 개발, 품질 보증 및 배포에는 각기 다른 역할을 수행하는 여러 기관이 관여합니다. 이는 책임 소재를 규명하는 것을 복잡하게 만듭니다. 개발자는 제공자와 배포자와 같은 후속 주체들에게 LMM 사용으로 인한 피해에 대한 책임을 요구할 수 있으며, 반대로 후속 주체들은 알고리즘 훈련에 사용된 데이터 선택과 같은 이전의 조치가 원인이라고 주장할 수 있습니다. 또한, 개발자와 제공자는 의료 AI 기술이 규제 기관의 승인을 받았다면 더 이상 발생한 피해에 대한 책임을 지지 않아야 한다고 주장할 수 있습니다(규제 선점주의)*(1)*. 가치사슬에 따라 책임을 설정하는 것은 입법자와 정책 입안자들에게 어려운 과제입니다.

민사 책임 규정의 핵심 기능은 피해자가 AI 기술의 개발 및 배포에 관련된 여러 주체들 간의 책임과 원인을 규명하는 것이 어렵더라도 보상과 구제를 청구할 수 있도록 보장하는 것입니다. 만약 피해자가 보상을 받기 너무 어렵다고 느낀다면, 정의가 실현될 수 없을 뿐만 아니라, AI 가치사슬에 있는 주체들이 미래의 피해를 방지하려는 동기를 잃게 될 것입니다. 또한, 규정은 피해에 대해 충분한 보상이 이루어지도록 해야 합니다.

유럽연합은 제안된 AI 책임 가이던스에서 피해자의 입증 책임을 완화하기 위해 “인과관계 추정”을 도입했습니다*(162)*. 따라서 피해자가 하나 이상의 주체가 피해와 관련된 의무를 준수하지 않았고, AI 성능과 인과관계가 있을 가능성이 높다는 것을 입증하면, 법원은 불이행이 피해의 원인이라고 추정할 수 있습니다*(162)*. 이 경우, 책임을 가진 주체는 예를 들어 피해의 원인이 다른 주체임을 입증함으로써 이 추정을 반박해야 합니다. 해당 법률의 범위는 AI 시스템의 초기 제작자에만 국한되지 않으며 AI 가치사슬의 모든 참여자를 포함합니다*(162)*. AI 가치사슬에 있는 모든 행위자가 공동 책임을 질 경우, 이들은 위험 평가 및 완화의 효과를 입증함으로써 책임을 줄일 수 있습니다.

그러나 책임 제도가 AI 기반 제품과 서비스로 인한 피해에 대한 명확한 구제와 보상을 항상 제공하는 것은 아닙니다. 특히 피해자가 의료 결정에 LMM이 사용되었다는 사실을 알지 못하는 경우 더욱 그렇습니다. 새 규정은 AI 기반 의료 기술로 인한 피해에 대한 책임의 공백을 남길 수 있습니다*(163)*. LMM은 여전히 예측하기 어렵고 이해가 부족하며 시장에 빠르게

도입되고 있기 때문에, 정부는 보건의료에서 사용되는 LMM을 엄격 책임(strict liability) 기준에 따라 개발자, 제공자 및 배포자가 책임지도록 고려할 수 있습니다. 이러한 주체들에게 발생한 오류에 대한 책임을 묻는 것은 오류가 환자에게 영향을 미칠 경우 보상을 보장할 수 있는 방법이 될 수 있습니다*(1)*. 그러나 이는 환자가 LMM이 사용되었다는 사실을 알고 있었는지 여부에 따라 달라질 수 있습니다. 계속되는 책임 부여는 점점 더 정교한 LMM의 사용을 위축시킬 수 있지만, 불필요한 위험을 감수하거나 LMM의 많은 위험과 잠재적 피해가 완전히 규명되고 해결되기 전에 보건의료 또는 공공 보건 환경에 새로운 LMM을 배포하려는 의지를 줄일 수도 있습니다*(1)*.

AI에 대한 책임 체계가 과실을 완전히 규명하기에는 충분하지 않을 수 있습니다. 이는 알고리즘이 개발자, 제공자 또는 배포자가 완전히 통제할 수 없는 방식으로 진화하기 때문입니다. 또한, 피해를 입은 사람이 손해 배상을 받을 수 없는 상황이나 관할권이 있을 수 있습니다. 예를 들어, 미국에서는 LMM을 사용하여 직접 조언을 구하다가 부상을 입은 환자가 전문 책임 규칙에 AI 시스템이 포함되지 않거나 제품 또는 소비자 책임법의 예외나 제한이 있는 경우 배상을 받지 못할 수 있습니다(163). 의료 분야의 다른 영역에서는 백신의 부작용으로 인한 의료적 손해와 같은 의료 부상에 대해 과실이나 책임을 규명하지 않고 보상을 제공하기도 합니다. WHO의 초기 가이드라인은 “AI 기술 사용으로 인한 의료 부상을 입은 개인에게 지급을 제공하기 위한 적절한 메커니즘으로 무과실, 무책임 보상 기금을 설정할지 여부와 이러한 청구를 지급하기 위한 자원을 어떻게 동원할지를 결정해야 한다”고 권고했습니다*(1)*. 이 권고는 현재에도 유효하며, LMM이나 LMM 응용 프로그램으로 인한 부상에 대한 보상을 결정하는 수단이 될 수 있습니다.

권장사항

- 정부는 LMM 및 응용 프로그램의 개발, 제공 및 배포 가치사슬 전반에 따라 책임을 설정하여, 피해자가 책임 소재의 명확성과 기술 개발 및 배포에 관여한 다양한 기관들의 책임과 관계없이 보상을 청구할 수 있도록 해야 합니다.

8 LMM의 국제적 거버넌스

정부는 보건의료에서 사용되는 LMM 및 기타 형태의 AI를 관리하기 위한 국제 규칙의 공동 개발을 지원해야 합니다. 이러한 사용이 전 세계적으로 확산되고 있기 때문입니다. 한 예로 WHO의 2020-2025 디지털 건강 글로벌 전략을 들 수 있습니다. 이 과정은 보건의료에서 AI를 배포하는 기회와 도전에 대응하기 위해 유엔 시스템 내에서 더 큰 협력과 협조를 포함해야 하며, AI의 사회 및 경제 전반에 걸친 광범위한 응용에도 대응해야 합니다. 정부가 적절하고 집행 가능한 기준을 함께 설정하지 않으면, 적절한 법적, 윤리적, 안전 기준을 충족하지 못하는 LMM 및 기타 형태의 AI의 수가 증가할 가능성이 있습니다. 이는 규제 및 기타 보호 조치가 도입되지 않거나 제대로 집행되지 않을 경우 의도치 않은 피해를 초래할 수 있습니다. WHO는 최근 전 세계 규제 기관들과 협의하여 정부와 규제 당국이 새로운 가이던스를 개발하거나 기존 가이던스를 AI에 맞게 조정하는 데 따라야 할 주요 원칙을 제시한 새로운 출판물을 발간했습니다*(164)*.

국제 거버넌스는 선도적 이점을 얻으려는 기업들 사이에서 안전성과 유효성 기준이 무시되는 "하향 경쟁"과 기술적 우위를 위한 지정학적 경쟁에서 우위를 점하려는 정부들 사이의 경쟁을 방지할 수 있습니다. 따라서 국제 거버넌스는 모든 기업이 최소한의 안전성과 유효성 기준을 충족하도록 보장하고, 기업이나 정부가 규제로 인해 경쟁적 우위나 불이익을 겪지 않도록 할 수 있습니다. 또한, 국제 거버넌스는 AI 기반 시스템 개발과 배치에 대한 정부의 투자와 참여에 대해 책임을 묻고, 윤리적 원칙, 인권 및 국제법을 존중하는 적절한 규제를 도입하도록 보장할 수 있습니다. 전 세계적으로 집행 가능한 기준이 없으면 제품 채택에도 부정적인 영향을 미칠 수 있습니다.

국제 거버넌스는 여러 형태를 취할 수 있습니다. 한 가지 제안은 유럽 원자핵 연구위원회(CERN, Conseil Européen pour la Recherche Nucléaire)와 같은 국제 협력 기구를 설립하여 여러 정부의 자금으로 대규모 변혁적 프로젝트를 수행하고, 그 결과를 공개적으로 공유하는 것입니다*(165,166)*. 또 다른 제안은 이러한 기관이 고도로 안전한 시설에서 가장 발전되고 위험한 형태의 AI를 개발하도록 하고, 이러한 형태의 AI를 개발하려는 다른 시도를 불법으로 규정하는 것입니다*(167)*. 현재 이러한 대규모 프로젝트는 공공 재화를 창출하기 위한 공공 자금 프로젝트의 영역에 있지 않고, 상업적 경쟁에 참여하는 대규모 기술 기업들이 주도하고 있습니다. 일부 세계 지도자와 기술 경영진은 AI를 핵무기와 유사하게 취급하여 핵무기 사용 조약과 유사한 글로벌 거버넌스 프레임워크를 도입해야 한다고 주장했습니다*(109)*.

국제 거버넌스가 어떤 형태를 취하든, 고소득 국가나 대규모 기술 기업과 주로 협력하는 고소득 국가들만이 주도하지 않도록 하는 것이 중요합니다*(168)*. 고소득 국가와 기술 기업이 개발한 기준은 AI의 모든 응용분야나 보건의료 및

의학에서의 LMM의 특정 사용을 대한 것이라 할지라도, 저소득 및 중간소득 국가의 대부분의 사람들에게는 기준을 형성하는 데 참여할 역할이나 목소리를 내지 못하게 할 것입니다. 이는 이러한 국가들이 AI 기술의 혜택을 가장 많이 받을 가능성이 있음에도 불구하고, 미래의 AI 기술을 잠재적으로 위험하거나 비효율적으로 만들 수 있습니다.

AI의 국제 거버넌스는 2019년 유엔 사무총장이 제안한 네트워크 다자주의를 통해 모든 이해관계자가 협력하도록 요구할 수 있습니다*(169)*. 이는 유엔 기구, 국제 금융 기관, 지역 조직, 무역 블록 및 시민 사회, 도시, 기업, 지방 당국, 청소년 등을 포함한 다양한 주체들이 더 긴밀하고 효과적이며 포괄적으로 협력할 수 있도록 합니다. LMM의 개발 및 배치에서 윤리와 인권을 중심에 두는 것은 보편적 건강 보장의 달성에 상당한 기여를 할 수 있습니다.

권장사항

- 정부는 AI의 국제 거버넌스를 위한 규칙을 공동으로 개발하는 것을 지원해야 합니다. 거버넌스의 형태가 무엇이든, 고소득 국가나 주로 대규모 기술 기업과 협력하는 고소득 국가들에 의해 단독으로 형성되지 않아야 합니다. 이러한 접근방식은 국제 AI 규제를 형성하는 데 있어 저소득 및 중간 소득 국가의 역할과 목소리를 배제하게 될 것입니다.

참고문헌

1. Ethics and governance of artificial intelligence for health. Geneva: World Health Organization; 2021 (https://www.who.int/publications/i/item/9789240029200, accessed 26 May 2023).

2. Khullar D. Can A.I. treat mental illness? The New Yorker, 27 February 2023 (https:// www.newyorker.com/magazine/2023/03/06/can-ai-treat-mental-illness, accessed 29 May 2023).

3. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med. 2022;28(9):1773–84. doi:10.1038/s41591-022-01981-2.

4. Hariri Y, Harris T, Raskin A. You can have the blue pill or the red pill, and we're out of blue pills. The New York Times, 24 March 2023 (https://www.nytimes.com/2023/03/24/ opinion/yuval-harari-ai-chatgpt.html, accessed 26 May 2023).

5. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ et al. Foundation models for generalist medical artificial intelligence, Nature. 2023;616(7956):259–65. doi:10.1038/s41586-023-05881-4.

6. Hu K. Chat GPT sets record for fastest growing user-base – analyst note. Reuters, 2 February 2023 (https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/, accessed 26 May 2023).

7. Weise K, Grant N. Microsoft and Google unveil A.I. tools for businesses. The New York Times, 16 March 2023 (https://www.nytimes.com/2023/03/16/technology/microsoft-google-ai-tools-businesses.html, accessed 26 May 2023).

8. Yang Z. Chinese tech giant Baidu just released its answer to ChatGPT. MIT Technology Review, 16 March 2023 (https://www.technologyreview.com/2023/03/16/1069919/baidu-ernie-bot-chatgpt-launch/, accessed 26 May 2023).

9. Murgia M, Bradshaw T. Musk to launch AI start-up to rival ChatGPT. Financial Times, 15 April 2023 (https://www.ft.com/content/2a96995b-c799-4281-8b60-b235e84aefe4, accessed 26 May 2023).

10. Heaven WD. The open-source AI boom is built on Big Tech's handouts. How long will it last? MIT Technology Review, 12 May 2023 (https://www.technologyreview. com/2023/05/12/1072950/open-source-ai-google-openai-eleuther-meta/, accessed 26 May 2023).

11. Martin A. Google CEO Sunder Pichai admits people don't fully understand how chatbot AI works,

Evening Standard, 17 April 2023 (https://www.standard.co.uk/tech/google-ceo-sundar-pichai-understand-ai-chatbot-bard-b1074589.html, accessed 26 May 2023).

12. Roose K. A conversation with Bing's chatbot left me deeply unsettled. The New York Times, 16 February 2023 (https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html, accessed 26 May 2023).

13. Marcus G. AI platforms like ChatGPT are easy to use but potentially dangerous, Scientific American, 19 December 2022 (https://www.scientificamerican.com/article/ai-platforms-like-chatgpt-are-easy-to-use-but-also-potentially-dangerous/, accessed 26 May 2023).

14. Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E et al. Sparks of artificial general intelligence: early experiments with GPT-4. ArXiv:2302.12712.

15. McGowran L. OpenAI criticised for lack of transparency around ChatGPT-4. Silicon Republic, 16 March 2023 (https://www.siliconrepublic.com/machines/openai-gpt4- transparency-ai-concerns-stripe-chatgpt, accessed 26 May 2023).

16. Spitale G, Biller-Andorno N, Germani F. AI model GPT-3 (dis)informs us better than humans. Sci Adv. 2023;9(26):eadh1850. doi:10.1126/sciadv.adh1850.

17. Volpicelli G. ChatGPT broke the EU plan to regulate AI Politico, 3 March 2023 (https:// www.politico.eu/article/eu-plan-regulate-chatgpt-openai-artificial-intelligence-act/, accessed 26 May 2023).

18. Arcesati R, Chang W. China is blazing a trail in regulating Generative AI – on the CCP's terms. The Diplomat, 28 April 2023 (https://thediplomat.com/2023/04/china-is-blazing-a-trail-in-regulating-generative-ai-on-the-ccps-terms/, accessed 26 May 2023).

19. Martindale J. These are the countries where ChatGPT is currently banned. Digital Trends, 12 April 2023 (https://www.digitaltrends.com/computing/these-countries-chatgpt-banned/, accessed 26 May 2023).

20. Johnson K. ChatGPT can help doctors – and hurt patients. Wired, 24 April 2023 (https:// www.wired.com/story/chatgpt-can-help-doctors-and-hurt-patients/, accessed 28 May 2023).

21. Topol E. Multimodal AI for medicine, simplified. Ground Truths, 14 March 2023 (https:// erictopol.substack.com/p/multimodal-ai-for-medicine-simplified, accessed 28 May 2023).

22. Heaven WD. AI hype is built on high test scores. Those tests are flawed. MIT Technology Review, 30 August 2023 (https://www.technologyreview.com/2023/08/30/1078670/ large-language-models-arent-people-lets-stop-testing-them-like-they-were/, accessed 1 October 2023).

23. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW et al. Large language models encode clinical knowledge. Nature. 2023;620:172–80. doi:10.1038/s41586-023-06291-2.

24. Kulkarni PA, Singh H. Artificial intelligence in clinical diagnosis: opportunities, challenges, and hype. JAMA. 2023;330(4):317–8. doi:10.1001/jama.2023.11440.

25. Subbamaran N. ChatGPT will see you now: Doctors using AI to answer patient questions. Wall Street Journal, 28 April 2023 (https://www.wsj.com/articles/dr-chatgpt-physicians-are-sending-patients-advice-using-ai-945cf60b, accessed 28 May 2023).

26. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. 2023; 183(6):589–96. doi: 10.1001/jamainternmed.2023.1838.

27. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. New Engl J Med. 2023;388(13):1233–9. doi: 10.1056/NEJMsr2214184.

28. The potential of large language models in healthcare: Improving quality of care and patient outcomes, Medium, 7 December 2022. (https://medium.com/@BuildGP/the-potential-of-large-language-models-in-healthcare-improving-quality-of-care-and-patient-6e8b6262d5ca, accessed 28 May 2023).

29. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T et al. CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv. 2017;1711.05225vv3. doi:10.48550/arXiv.1711.05225.

30. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C et al. A large language model for electronic health records. npj Digit Med. 2022;5:194. doi:10.1038/s41746-022-00742-2.

31. Ghahramani Z. Introducing PaLM 2. The Keyword, 10 May 2023 (https://blog.google/ technology/ai/google-palm-2-ai-large-language-model/, accessed 28 May 2023).

32. Weise K, Metz C. When A.I. chatbots hallucinate. The New York Times, 9 May 2023 (https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html, accessed 1 June 2023).

33. Bender EM, Gebru T, McMillan-Major A, Mitchell M. On the dangers of stochastic parrots: Can language models be too big? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, March 2021, pp. 610–23. doi: 10.1145/3442188.3445922.

34. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. Lancet Digit Health. 2023;5(6):e333–5. doi:10.1016/s2589-7500(23)00083-3.

35. Metz, Cade. Chatbots may 'hallucinate' more often than many realize, The New York Times, 6 November 2023 (https://www.nytimes.com/2023/11/06/technology/chatbots-hallucination-rates.html, accessed 7 November 2023).

36. Acar OA. AI prompt engineering isn't the future, Harvard Business Review, 6 June 2023 (https://hbr.org/2023/06/ai-prompt-engineering-isnt-the-future?registration=success, accessed 26 June 2023).

37. GPT-4 system card. Open AI, 23 March 2023 (https://cdn.openai.com/papers/gpt-4- system-card.pdf, accessed 28 May 2023).

38. GPT-4. OpenAI, 14 March 2023 (https://openai.com/research/gpt-4, accessed 28 May 2023).

39. Radford A, Kleinman Z. ChatGPT can now access up-to-date information. BBC News, 27 September 2022 (https://www.bbc.com/news/technology-66940771, accessed 1 October 2023).

40. Kruge S, Ostermaier A, Uhl M. The moral authority of ChatGPT. ArXiv:23101.07098. doi:10.48550/arXiv.23101.07098.

41. Mickle T, Metz C, Grant N. The chatbots are here, and the internet industry is in a tizzy, The New York Times, 8 March 2023 (https://www.nytimes.com/2023/03/08/technology/ chatbots-disrupt-internet-industry.html, accessed 29 May 2023).

42. Woo M. Trial by artificial intelligence. Nature. 2019;573:S100–2 (https://media.nature. com/original/magazine-assets/d41586-019-02871-3/d41586-019-02871-3.pdf, accessed 29 May 2023).

43. Muralidharan V, Burgart A, Daneshjou D, Rose S. Recommendations for the use of pediatric data in artificial intelligence and machine learning ACCEPT-AI. npj Dig Med. 2023;6:166. doi:10.1038/s41746-023-00898-5.

44. Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learning Individual Differences. 2023;103:102274. doi:10.1016/j.lindif.2023.102274.

45. Reddy CD, Lopez L, Ouyang D, Zou JY, He B. Video-based deep learning for automated assessment of left ventricular ejection fraction in pediatric patients. J Am Soc Echocardiogr. 2023;36(5):482–9. doi:10.1016/j.echo.2023.01.015.

46. Knight W. These ChatGPT rivals are designed to play with your emotions. Wired, 4 May 2023 (https://www.wired.com/story/fast-forward-chatgpt-rivals-emotions/#:~:text=12%3A00%20PM-,These%20ChatGPT%20Rivals%20 Are%20Designed%20to%20Play%20With%20Your%20Emotions,%2C%20companionship%E2%80%94and%20even%20romance., accessed 29 May 2023).

47. Smuha NA, De Ketalaere M, Coeckelbergh M, Dewitte P, Poullet Y. Open letter: We are not ready for manipulative AI – urgent need for action. KU Leuven, 31 March 2023 (https:// www.law.kuleuven.be/ai-summer-school/open-brief/open-letter-manipulative-ai, accessed 29 May 2023).

48. Cuthbertson A. "No, I'm not a robot": ChatGPT successor tricks worker into thinking it is human. Independent, 15 March 2023 (https://www.independent.co.uk/tech/chatgpt-gpt4-ai-openai-b2301523.html, accessed 26 June 2023).

49. Walker L. Belgian man dies by suicide following exchanges with chatbot, The Brussels Times, 28 March 2023 (https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt, accessed 29 May 2023).

50. DeGuerin M. Oops: Samsung employees leaked confidential data to ChatGPT. Gizmodo, 6 April 2023 (https://gizmodo.com/chatgpt-ai-samsung-employees-leak-data-1850307376, accessed 29 May 2023).

51. Privacy policy. OpenAI, 27 April 2023 (https://openai.com/policies/privacy-policy, accessed 29 March 2023).

52. Coles C. 11% of data employees paste into ChatGPT is confidential. Cyberhaven, 19 April 2023 (https://www.cyberhaven.com/blog/4-2-of-workers-have-pasted-company-data-into-chatgpt/, accessed 29 May 2023).

53. Mihalcik C. ChatGPT bug exposed some subscribers' payment info. CNET, 24 March 2023.(https://www.cnet.com/tech/services-and-software/chatgpt-bug-exposed-some-subscribers-payment-info/, accessed 29 May 2023).

54. Moodley K, Rennie S. ChatGPT has many uses. Experts explore what this means for healthcare and medical research. The Conversation, 22 February 2023 (https:// theconversation.com/chatgpt-has-many-uses-experts-explore-what-this-means-for-healthcare-and-medical-research-200283, accessed 2 June 2023).

55. De Proost M, Pozzi G. Conversational artificial intelligence and the potential for epistemic injustice. Am J Bioethics. 2023;23(5):51–3. doi:10.1080/15265161.2023.2191020.

56. Disability and employment. New York: United Nations, Department of Economic and Social Affairs (Disability); undates (https://www.un.org/development/desa/disabilities/ resources/factsheet-on-persons-with-disabilities/disability-and-employment.html, accessed 11 September 2023).

57. Whittaker M, Alper M, Bennett CL, Hendren S, Kaziunas L, Mills Met al. Disability, bias and AI. New York: AI Now Institute; 2019 (https://ainowinstitute.org/wp-content/ uploads/2023/04/disability-

biasai-2019.pdf, accessed 11 September 2023).

58. Hallman J. AI language models show bias against people with disabilities, study finds. University Park (PA): Penn State University; 2022 (https://www.psu.edu/news/ information-sciences-and-technology/ story/ai-language-models-show-bias-against-people-disabilities/, accessed 11 September 2023).

59. Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. eBioMedicine. 2023;90:194512. doi:10.1016/j.ebiom.2023.104512.

60. Lohr S. AI may someday work medical miracles. For now, it helps do paperwork. The New York Times, 26 June 2023 (https://www.nytimes.com/2023/06/26/technology/ai-health-care-documentation.html, accessed 10 July 2023).

61. Eddy N. Epic, Microsoft partner to use generative AI for better EHRs. Healthcare IT News, 18 April 2023 (https://www.healthcareitnews.com/news/epic-microsoft-partner-use-generative-ai-better-ehrs, accessed 31 May 2023).

62. Nuance and Microsoft announce the first fully AI-automated clinical documentation application for healthcare. Burlington (MA):Nuance; 2023 (https://news.nuance. com/2023-03-20-Nuance-and-Microsoft-Announce-the-First-Fully-AI-Automated- Clinical-Documentation-Application-for-Healthcare, accessed 31 May 2023).

63. Ahn S. The impending impacts of large language models on medical education. Korean J Med Educ. 2023;35(1):103–7. doi:10.3946/kjme.2023.253.

64. Luo R, Sun L, Xia Y, Qin T, Zhang S, Poon H et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining. Briefi Bioinformatics. 2022; 23(6):bbac409. doi:10.1093/bib/bbac409.

65. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. Drug Discov Today. 2021;26(1):80–93. doi:10.1016/j.drudis.2020.10.010.

66. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. Nature. 2023;613:612. doi:10.1038/d41586-023-00191-1.

67. Zielinksi C, Winker MA, Aggarwal R, Ferris LA, Heinemann M, Lapeña JF Jr et al. Chatbots, generative AI, and scholarly manuscripts. Overijssel: World Association of Medical Editors; 2023. (https://wame.org/page3.php?id=106, accessed 26 June 2023).

68. Gibbs W. Lost science in the Third World. Sci Am. 1995;273(2):92–9 doi:10.1038/scientificamerican0895-92.

69. Birhane A, Kasirzadeh A, Leslie D, Wachter S. Science in the age of large language models. Nat Rev Phys. 2023;5:277–80. doi:10.1038/s42254-023-00581-4.

70. Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies. Geneva: World Health Organization; 2010 (https://apps.who.int/ iris/bitstream/handle/10665/258734/9789241564052-eng.pdf, accessed 26 June 2023).

71. Morozov E. The true threat of artificial intelligence. The New York Times, 30 June 2023 (https://www.nytimes.com/2023/06/30/opinion/artificial-intelligence-danger.html, accessed 2 July 2023).

72. Introducing ChatGPTPlus. San Francisco (CA): Open AI; 2023 (https://openai.com/blog/ chatgpt-plus, accessed 1 June 2023).

73. The hidden workforce that helped filter violence and abuse out of ChatGPT. Wall Street Journal, 11 July 2023 (https://www.wsj.com/podcasts/the-journal/the-hidden-workforce-that-helped-filter-violence-and-abuse-out-of-chatgpt/ffc2427f-bdd8-47b7- 9a4b-27e7267cf413, accessed 13 July 2023).

74. Firth N. Language models may be able to self-correct biases – if you ask them. MIT Technology Review, 20 March 2023 (https://www.technologyreview. com/2023/03/20/1070067/language-models-may-be-able-to-self-correct-biases-if-you-ask-them-to/, accessed 1 June 2023).

75. Khan L. We must regulate A.I. Here's how. The New York Times, 3 May 2023 (https:// www.nytimes.com/2023/05/03/opinion/ai-lina-khan-ftc-technology.html, accessed 2 June 2023).

76. Hatzius J, Briggs J, Kodnani D, Pierdomenico G. The potentially large effects of artificial intelligence on economic growth (Briggs/Kodnani). Goldman Sachs Economics Research, 26 May 2023 (https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics- Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth- Briggs_Kodnani.pdf, accessed 1 June 2023).

77. Milmo D. AI revolution puts skilled jobs at highest risk, says OECD. The Guardian, 11 July 2023 (https://www.theguardian.com/technology/2023/jul/11/ai-revolution-puts-skilled-jobs-at-highest-risk-oecd-says, accessed 12 July 2023).

78. Health and care workforce in Europe: time to act. Geneva: World Health Organization; 2022 (https://iris.who.int/handle/10665/362379, accessed 1 June 2023).

79. Health workforce. Geneva: World Health Organization; 2023 (https://www.who.int/ health-topics/health-workforce#tab=tab_1, accessed 1 June 2023).

80. Hurst L. OpenAI says 80% of workers could see their jobs impacted by AI. These are the jobs most impacted, Euronews.next, 30 March 2023. (https://www.euronews.com/ next/2023/03/23/openai-says-80-of-workers-could-see-their-jobs-impacted-by-ai-these-are-the-jobs-most-affe, accessed 1

June 2023).

81. A new era of generative AI for everyone. Dublin: Accenture; 2023 (https://www.accenture. com/content/dam/accenture/final/accenture-com/document/Accenture-A-New-Era-of- Generative-AI-for-Everyone.pdf, accessed 1 June 2023).

82. Burgess M. The security hole at the heart of ChatGPT and Bing. Wired, 25 May 2023 (https://www.wired.co.uk/article/chatgpt-prompt-injection-attack-security, accessed 1 June 2023).

83. Heikkila M. Open AI's hunger for data is coming back to bite it. MIT Technology Review, 19 April 2023 (https://www.technologyreview.com/2023/04/19/1071789/openais-hunger-for-data-is-coming-back-to-bite-it/, accessed 1 June 2023).

84. General Data Protection Regulation, Regulation 2016//679 of the European Parliament and of the Council, 27 April 2016. Strasbourg: European Parliament; 2016 (https://eur-lex. europa.eu/eli/reg/2016/679/oj, accessed 27 September 2023).

85. The impact of the General Data Protection Regulation on artificial intelligence (STOA Options Brief). Strabourg: European Parliament; 2020 (https://www.europarl.europa.eu/ RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530(ANN1)_EN.pdf, accessed 26 June 2023).

86. OPC launches investigation into ChatGPT. Ottawa: Office of the Privacy Commissioner of Canada; 4 April 2023 (https://www.priv.gc.ca/en/opc-news/news-and-announcements/2023/an_230404/, accessed 1 June 2023).

87. Lomas N. Italy orders Chat GPT blocked citing data protection concerns. Tech Crunch, 31 March 2023 (https://techcrunch.com/2023/03/31chatgpt-blocked-italy/, accessed 1 June 2023).

88. ChatGPT: Italian SA to lift temporary limitation if OpenAI implements measures. Rome: Italian Data Protection Authority; 2023 (https://www.garanteprivacy.it/home/docweb/-/ docweb-display/docweb/9874751#english, accessed 1 June 2023).

89. Weatherbed J. OpenAI's regulatory troubles are only just beginning. The Verge, 5 May 2023 (https://www.theverge.com/2023/5/5/23709833/openai-chatgpt-gdpr-ai-regulation-europe-eu-italy, accessed 1 June 2023).

90. Wiggers K. Open AI's new tool attempts to explain language models' behaviours. Tech Crunch, 9 May 2023 (https://techcrunch.com/2023/05/09/openais-new-tool-attempts-to-explain-language-models-behaviors/, accessed 1 June 2023).

91. Libeau D. ChatGPT will probably never comply with GDPR. 10 April 2023. (https://blog. davidlibeau.fr/chatgpt-will-probably-never-comply-with-gdpr/, accessed 1 June 2023).

92. Lomas N. ChatGPT maker OpenAI accused of string of data protection breaches in GDPR complaint filed by privacy researcher. TechCrunch, 30 August 2023. (https://consent. yahoo.com/v2/collectConsent?sessionId=3_cc-session_6bdecae4-d7b6-448f-8e26- e7805c03b964, accessed 11 September 2023).

93. Fung B. The FTC should investigate Open AI and block GPT over "deceptive" behaviour, AI policy group claims. CNN, 30 March 2023. (https://edition.cnn.com/2023/03/30/tech/ ftc-openai-gpt-ai-think-tank/index.html, accessed 2 June 2023).

94. Waters R, Murgia M, Espinoza J. Open AI warns over split with Europe as AI regulation advances. Financial Times, 25 May 2023 (https://www.ft.com/content/5814b408-8111-49a9-8885-8a8434022352, accessed 1 June 2023).

95. Technology-facilitated gender-based violence: Making all spaces safe. New York: United Nations Population Fund; 2021 (https://www.unfpa.org/publications/technology-facilitated-gender-based-violence-making-all-spaces-safe, accessed 1 October 2023).

96. Murgia M. DeepMind reinvents itself for AI counterattack. Financial Times, 2 May 2023 (https://ft.pressreader.com/v99c/20230502/281724093873699, accessed 2 June 2023).

97. Schaake M. Regulating AI will put companies and governments at loggerheads, Financial Times, 2 May 2023 (https://www.ft.com/content/7ef4811d-79bb-4b4f-b28f-b46430f0c9ff, accessed 2 June 2023).

98. Metz, Cade. Tech giants are paying huge salaries for scarce A.I. talent, The New York Times, 22 October 2017 (https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html).

99. Leswing K. Google reveals its newest AI supercomputer, says it beats Nvidia. CNBC, 5 April 2023. (https://www.cnbc.com/2023/04/05/google-reveals-its-newest-ai-supercomputer-claims-it-beats-nvidia-.html, accessed 2 June 2023)

100. Ahuja K. Antitrust has role in policing AI landscape. Financial Times, 10 April 2023 (https://www.ft.com/content/953817f5-5bc4-49e1-b583-977cc4780eca, accessed 2 June 2023).

101. Ahmed N, Wahed M, Thompson NC. The growing influence of industry in AI research. Science. 2023;379(6635):884–6. doi:10.1126/science.ade2420.

102. Røttingen JA, Regmi S, Eide M, Young AJ, Viergever RF, Ardal C et al. Mapping of available health research and development data: What's there, what's missing, and what role is there for a global observatory? Lancet. 2013;382(9900):1286–307. doi:10.1016/S0140-6736(13)61046-6.

103. A new partnership to promote responsible AI. Google Blogs, 26 July 2023 (https://blog. google/outreach-initiatives/public-policy/google-microsoft-openai-anthropic-frontier-model-forum/#:~:-text=Anthropic%2C%20Google%2C%20Microsoft%20and%20 OpenAI%20are%20launching%20the%20Frontier%20Model,development%20of%20 frontier%20AI%20models, accessed 29 July 2023).

104. Fact sheet: Biden–Harris Administration secures voluntary commitments from leading artificial intelligence companies to manage the risks posed by AI. Washington DC: The White House, 21 July 2023 (https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/, accessed 29 July 2023).

105. Volpicelli G. Europe pitches AI pact to curtail the booming tech's risk. Politico, 26 May 2023 (https://www.politico.eu/article/big-tech-rumble-europe-global-artificial-intelligence-debate-ai-pact/, accessed 29 July 2023).

106. Grant N, Weise K. In AI race, Microsoft and Google choose speed over caution. The New York Times, 7 April 2023 (https://www.nytimes.com/2023/04/07/technology/ai-chatbots-google-microsoft.html, accessed 2 June 2023).

107. Center for Research on Foundation Models. The Foundation Model Transparency Index, 2023. (https://crfm.stanford.edu/fmti/, accessed 21 October 2023).

108. Schiffer Z, Newton C. Microsoft lays off team that taught employees how to make AI tools responsibly. The Verge, 14 March 2023.(https://www.theverge.com/2023/3/13/23638823/microsoft-ethics-society-team-responsible-ai-layoffs, accessed 2 June 2023).

109. Milmo D. Google chief warns AI could be harmful if deployed wrongly. The Guardian, 17 April 2023 (https://www.theguardian.com/technology/2023/apr/17/google-chief-ai-harmful-sundar-pichai, accessed 2 June 2023).

110. Fiesler C. AI has social consequences, but who pays the price? Tech companies' problem with ethical debt. The Conversation, 19 April 2023 (https://theconversation.com/ai-has-social-consequences-but-who-pays-the-price-tech-companies-problem-with-ethical-debt-203375, accessed 2 June 2023).

111. Criddle, Cristina and Murphy, Hannah, Meta disbands protein-folding team in shift towards commercial AI, Financial Times, 7 August 2023. (https://www.ft.com/ content/919c05d2-b894-4812-aa1a-dd2ab6de794a?accessToken=zwAGBZu-oWVwkdORnAXSuJRIEtOqGt0qtt55Sg.MEQCIA1QQ1iG8KPAnuDAuPvt-Ngds3OzxL1lt- 0FnaVbAQFtAiAZvHnmKD_fABj8ZzLTNXRp1v-7V38nTcUf_pPxAPdx16A&sharetype=gift&to ken=3ac5a132-e08e-412e-bc3c-08edea8a7417,

accessed 18 September 2023).

112. Ananthaswamy A. In AI, is bigger always better? Nature, 8 March 2023 (https://www. nature.com/articles/d41586-023-00641-w, accessed 2 June 2023).

113. Li P. Making AI less "thirsty": Uncovering and addressing the secret water footprint of AI models. ArXiv. 2023;2304.03271v. doi:10.48550/arXiv.2304.03271.

114. Syed N. The secret water footprint of AI technology. The Markup, 15 April 2023 (https:// themarkup.org/hello-world/2023/04/15/the-secret-water-footprint-of-ai-technology, accessed 2 June 2023).

115. Livingstone G. It's pillage: Thirsty Uruguayans blast Google's plan to exploit water supply. The Guardian, 11 July 2023 (https://www.theguardian.com/world/2023/jul/11/ uruguay-drought-water-google-data-center, accessed 12 July 2023).

116. Thornhill J. The sceptical case on generative AI. Financial Times, 17 August 2023. (https://www.ft.com/content/ed323f48-fe86-4d22-8151-eed15581c337, accessed 11 September 2023).

117. Marcus G. The imminent enshittification of the Internet. Substack, 16 August 2023 (https://garymarcus.substack.com/p/the-imminent-enshittification-of, accessed 11 September 2023).

118. Pause giant AI experiments: An open letter. Narberth (PA): Future of Life Institute; 2023 (https://futureoflife.org/open-letter/pause-giant-ai-experiments/, accessed 13 June 2023).

119. Perrigo B. DeepMind's CEO helped take AI mainstream. Now he's urging caution. Time, 12 January 2023 (https://time.com/6246119/demis-hassabis-deepmind-interview/, accessed 13 June 2023).

120. Lomas N. Unpacking the rules shaping generative AI. Tech Crunch, 13 April 2023 (https://techcrunch.com/2023/04/13/generative-ai-gdpr-enforcement/, accessed 13 June 2023).

121. Mökander J, Schuett J, Kirk HR, Floridi L. Auditing large language models: a three-layered approach Soc Sci Res Netw. 2023. doi:10.2139/ssrn.4361607.

122. Lomas N. Report details how Big Tech is leaning on EU not to regulate general purpose Ais. Tech Crunch, 23 February 2023 (https://techcrunch.com/2023/02/23/eu-ai-act-lobbying-report/, accessed 20 June 2023).

123. Sambasivan N, Kapania S, Highfill H, Akrong D, Paritosh P, Aroyo LM. "Everyone wants to do model work, not the data work.": Data cascades in high-stakes AI. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, May 2021. doi:10.1145/3411764.3445518.

124. Browne G. AI is steeped in Big Tech's digital colonialism. Wired, 25 May 2023 (https:// www.wired.

co.uk/article/abeba-birhane-ai-datasets, accessed 17 June 2023).

125. Baxter K, Schelsinger, N. Managing the risks of generative AI. Harvard Business Review, 6 June 2023 (https://hbr.org/2023/06/managing-the-risks-of-generative-ai, accessed 17 June 2023).

126. Samuelson P. Generative AI meets copyright. Science. 2023;381(6654):158–61. doi:10.1126/science.adi0656.

127. El-Mhamdi E, Farhadkhani S, Guerraoui R, Gupta N, Hoang L, Pinot R et al. On the impossible safety of large AI models. arXiv. 2209.15259v2. doi:10.48550.arXiv.2209.15259.

128. Open AI. GPT-4 technical report. arXiv:2303.08774v3. doi:10.48550/arXiv.2302.08774.

129. Murgia M. Open AI's red team: experts hired to "break" ChatGPT. Financial Times, 14 April 2023 (https://www.ft.com/content/0876687a-f8b7-4b39-b513-5fee942831e8, accessed 10 July 2023).

130. Clegg N. Openness on AI is the way forward for tech. Financial Times, 11 July 2023 (https://www.ft.com/content/ac3b585a-ce50-43d1-b71d-14dfe6dce999, accessed 11 July 2023).

131. Huang S, Toner H, Haluza Z, Creemers R, Webster G. Measures for the management of generative artificial intelligence services (draft for comment) (translation). DigiChina. Palo Alto (CA): Stanford University, Program on Geopolitics; 2023 (https://digichina. stanford.edu/work/translation-measures-for-the-management-of-generative-artificial-intelligence-services-draft-for-comment-april-2023/, accessed 17 June 2023).

132. Ye J. China says generative AI rules to apply only to products for the public. Reuters, 13 July 2023 (https://www.reuters.com/technology/china-issues-temporary-rules-generative-ai-services-2023-07-13/, accessed 13 July 2023).

133. Bommasani R, Klyman K, Zhang D, Liang P. Do foundation model providers comply with the draft EU AI Act? Palo Alto (CA): Stanford University, Human-centered Artifical Intelligence; 2021 (https://crfm.stanford.edu/2023/06/15/eu-ai-act.html, accessed 17 June 2023).

134. Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Strasbourg: European Parliament; 2023 (https://www.europarl.europa. eu/doceo/document/TA-9-2023-0236_EN.html, accessed 10 July 2023).

135. Beyond ChatGPT: How can Europe become a leader in generative AI? Kaiserslautern: German Research Centre for Artificial Intelligence; 2023. (https://www.dfki.de/en/ web/news/

jenseits-von-chatgpt-wie-kann-europa-bei-der-generativen-ki-eine-fuehrungsposition-uebernehmen, accessed 17 June 2023).

136. Spirling A. Why open-source generative AI models are an ethical way forward for science. Nature. 2023;616(7957):413. doi:10.1038/d41586-023-01295-4.

137. Vincent J. Meta's powerful AI language model has leaked online – What happens now? The Verge, 8 March 2023 (https://www.theverge.com/2023/3/8/23629362/meta-ai-language-model-llama-leak-online-misuse, accessed 29 July 2023).

138. Maffuli S. Meta's Llama 2 license is not open source. Open Source Initiative, 20 July 2023 (https://blog.opensource.org/metas-llama-2-license-is-not-open-source/, accessed 29 July 2023).

139. Marble A. Software licenses masquerading as open source. marble.onl, 1 June 2023 (http://marble.onl/posts/software-licenses-masquerading-as-open-source.html, accessed 29 July 2023).

140. Keary T. Report finds 82% of open-source software components "inherently risky". Venture Beat, 17 April 2023 (https://venturebeat.com/security/report-finds-82-of-open-source-software-components-inherently-risky/, accessed 8 July 2023).

141. Generative AI raises competition concerns. Technology blog, 29 June 2023. Washington DC: Federal Trade Commission; 2023 (https://www.ftc.gov/policy/advocacy-research/ tech-at-ftc/2023/06/generative-ai-raises-competition-concerns, accessed 29 July 2023).

142. Wishart-Smith H. Generative AI: cybersecurity friend and foe, Forbes, 6 June 2023 (https://www.forbes.com/sites/heatherwishartsmith/2023/06/06/generative-ai-cybersecurity-friend-and-foe/?sh=4407e0884bd2, accessed 29 July 2023).

143. Metz C. Researchers poke holes in safety controls of ChatGPT and other chatbots. The New York Times, 27 July 2023 (https://www.nytimes.com/2023/07/27/business/ai-chatgpt-safety-research.html, accessed 11 September 2023).

144. Harris T, Freuh S. The complexity of technology's consequences is going up exponentially, but our wisdom and awareness are not. Issues in Science and Technology, 16 May 2023 (https://issues.org/tristan-harris-humane-technology-misinformation-ai-democracy/, accessed 19 June 2023).

145. Schyns C. The lobbying ghost in the machine: Big Tech's covert defanging of Europe's AI Act. Brussels: Corporate Europe Observatory; 2023 (https://corporateeurope.org/ en/2023/02/lobbying-ghost-machine, accessed 17 June 2023).

146. Fact sheet: Biden–Harris Administration announces new actions to promote responsible AI innovation that protects Americans' rights and safety. Washington DC: White House, 4

May 2023 (https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety/, accessed 19 June 2023).

147. Sijbrandij Sid. AI weights are not "open source". Open Core Ventures, 27 June 2023. (https://opencoreventures.com/blog/2023-06-27-ai-weights-are-not-open-source/, accessed 29 July 2023).

148. Meeker H. Towards an open weights definition. Copyleft Currents, 8 June 2023 (https:// heathermeeker.com/2023/06/08/toward-an-open-weights-definition/, accessed 29 July 2023).

149. Dastin J, Tong A. Google, one of AI's biggest backers, warns its own staff about chatbots. Reuters, 15 June 2023 (https://www.reuters.com/technology/google-one-ais-biggest-backers-warns-own-staff-about-chatbots-2023-06-15/, accessed 9 July 2023).

150. Kanter GP, Packel EA. Health care privacy risks of AI chatbots. JAMA. 2023;330(4):311–2. doi:10.1001/jama.2023.9618.

151. Interim measures for the management of generative artificial intelligence services. Beijing: Cyberspace Administration of China; 13 2023. (http://www.cac.gov.cn/2023- 07/13/c_1690898327029107.htm, accessed 29 July 2023).

152. Satariano A. E.U. agrees on landmark artificial intelligence rules. The New York Times, 8 December 2023 (https://www.nytimes.com/2023/12/08/technology/eu-ai-act-regulation.html, accessed 15 December 2023).

153. The EU should regulate on the basis of rights, not risks. Access Now, 17 February 2021 (https://www.accessnow.org/eu-regulation-ai-risk-based-approach/, accessed 21 June 2023).

154. Marks M, Haupt CE. AI chatbots, health privacy, and challenges to HIPAA compliance JAMA. 2023;330(4):309–10. doi: 10.1001/jama.2023.9458.

155. Marcus G. Two models of AI oversight – and how things could go deeply wrong. Substack, 8 June 2023 (https://garymarcus.substack.com/p/two-models-of-ai-oversight-and-how, accessed 17 June 2023).

156. Kang C, Metz C. FTC opens investigation into Chat GPT maker over technology's potential harms. The New York Times, 13 July 2023 (https://www.nytimes.com/2023/07/13/ technology/chatgpt-investigation-ftc-openai.html, accessed 29 July 2023).

157. Ordish J. Large language models and software as a medical device. MedRegs blogs, 3 March 2023 (https://medregs.blog.gov.uk/2023/03/03/large-language-models-and-software-as-a-medical-device/, accessed 19 June 2023).

158. Gilbert S, Harvey H, Melvin T, Vollebregt E, Wicks P. Large language model AI chatbots require approval as medical devices. Nat Med. 2023. doi:10.1038/s41591-023-02412-6.

159. Ghost in the machine: Addressing the harm of generative AI. Forbrukerradet. Oslo: Norwegian Concumer Council; 2023 (https://storage02.forbrukerradet.no/ media/2023/06/generative-ai-rapport-2023.pdf, accessed 9 July 2023).

160. Mökander J, Floridi L. Ethics-based auditing to develop trustworthy AI. Minds & Machines, 2021. doi:10.1007/s11023-021-09557-8.

161. Minssen T, Vayena E, Cohen IG. The challenges for regulating medical use of ChatGPT and other large language models. JAMA. 2023;330(4):315–6. doi: 10.1001/jama.2023.9651.

162. Questions and Answers: AI Liability Directive. Brussels: European Commission; 2022 (https://ec.europa.eu/commission/presscorner/detail/en/QANDA_22_5793, accessed 20 June 2023).

163. Duffourc MN, Gerke S. The proposed EU directives for AI liability leave worrying gaps likely to impact medical AI. NPJ Digit Med. 2023;6(1):77. doi:10.1038/s41746-023-00823-w.

164. Regulatory considerations on artificial intelligence for health. Geneva: World Health Organization; 2023 (https://iris.who.int/bitstream/handle/10665/373421/9789240078871- eng.pdf?sequence=1&isAllowed=y, accessed 16 November 2023).

165. Marcus G. Artificial Intelligence is stuck. Here's how to move it forward. The New York Times, 29 July 2017 (https://www.nytimes.com/2017/07/29/opinion/sunday/artificial-intelligence-is-stuck-heres-how-to-move-it-forward.html, accessed 20 June 2023).

166. Parker G. Rishi Sunak to lobby Joe Biden for UK "leadership" role in AI development. Financial Times, 5 June 2023 (https://www.ft.com/content/7c30ea28-2895-44c2-9a2d-c31ea7fa27e7, accessed 19 June 2023).

167. Hogarth I. We must slow down the race to god-like AI. Financial Times, 13 April 2023 (https://www.ft.com/content/03895dc4-a3b7-481e-95cc-336a524f2ac2, accessed 10 July 2023).

168. Blinken A, Raimondo G. To shape the future of AI, we must act quickly. Financial Times, 24 July 2023 (https://www.ft.com/content/eea999db-3441-45e1-a567-19dfa958dc8f, accessed 30 July 2023).

169. Guterres, Antonio, Networked, inclusive multilateralism can help overcome challenges of era, says Secretary General, opening general assembly session, United Nations, 17 September 2019. (https://press.un.org/en/2019/sgsm19746.doc.htm, accessed 18 September 2023).

부록 · 방법론

본 가이던스는 합의에 따라 개발되었습니다. WHO는 20명의 전문가로 구성된 보건의료 위한 인공지능 윤리 및 거버넌스 전문가 그룹에 의존했으며, 이들은 약 4개월 동안 매 2주마다 회의를 진행했습니다. 전문가 그룹은 이전에 발행된 인공지능의 윤리 및 거버넌스에 관한 지침의 합의 원칙과 권장 사항을 보건 의료 및 의학에서 LMM의 새로운 사용에 적용했습니다.

전문가 그룹은 먼저 대규모 멀티모달 모델의 잠재적 사용과 이점, 그리고 최종 사용자 수준에서의 위험에 대한 예비 맵핑을 작성했습니다. 또한, 이러한 AI 시스템 사용으로 인해 보건 시스템과 사회가 직면할 수 있는 위험을 파악했습니다. 이는 다음을 포괄적으로 문헌 검색하여 보완되었습니다.
(a) LMM의 점진적인 발전과 사용 과정에서 지난 몇 년간 실현된 보건의료 분야에서의 기존 및 제안된 사용 사례, (b) LMM의 예상 사용 사례, (c) 본 가이던스 발행 이전에 발표된 LMM에 대한 비판 및 분석.

알려진 이점 및 잠재적인 이점과 위험에 대한 공동 이해를 바탕으로, 전문가 그룹은 LMM 사용과 관련된 다양한 윤리적 도전과제와 기회를 해결하기 위한 적절한 프레임워크를 도출했습니다. 전문가 그룹은 “가치사슬(value chain)” 접근방식이 적절한 거버넌스를 조직하고 관련 조치를 수행할 책임이 있는 행위자(들)를 파악하는 데 적합하다는 데 의견을 모았습니다.

여러 관할권에서 기존 또는 제안된 법률 및 규제 조치를 참조하여 권장사항의 틀을 잡는 수단으로 활용했으며, 전문가 그룹은 각 권고사항이 여러 국가와 법적 시스템에 적용 가능하도록 설계했습니다. 전문가 그룹은 또한 기업, AI 시스템 구매자, 그리고 특히 의료 제공자와 환자를 포함한 LMM 최종 사용자가 적용해야 할 권고사항에 대해서도 논의했습니다.

WHO는 전문가 그룹의 조사결과와 권고사항이 관련성과 유용성을 유지할 수 있도록 본 가이던스가 수정 및 업데이트될 필요가 있을 수 있음을 인정합니다.

발간일: 2025년 3월
펴낸곳: (재)국가생명윤리정책원
번역: (재)국가생명윤리정책원
교열: (재)국가생명윤리정책원
조판 편집: 광연재

그래픽 디자인: Joanna Sleigh(ETH Zurich, Zurich, Switzerland)
레이아웃: Imprimerie Centrale(Luxembourg)
주소: 서울특별시 광진구 능동로 400 별관 2층 (재)국가생명윤리정책원
홈페이지: www.nibp.kr
이메일: nibp@nibp.kr