



2024.12.30.

국회미래연구원 | 연구보고서 | 24-10호

보건의료 분야 인공지능 활용과 입법전략

허종호



국회미래연구원
NATIONAL ASSEMBLY FUTURES INSTITUTE

보건의료 분야 인공지능 활용과 입법전략



연구진

내부연구진

허종호 연구위원(연구책임)

- ◆ 출처를 밝히지 않고 이 보고서를 무단전제 또는 복제하는 것을 금하며, 저작권법 제24조3(공공저작물의 자유이용)에 따라 사용하실 경우 국회미래연구원의 동의를 반드시 받아 사용하여 주시기 바랍니다.
- ◆ 본 보고서의 내용은 연구원의 국회미래연구원의 공식적인 의견은 아님을 밝힙니다.

발 | 간 | 사

현대 사회는 기술의 비약적 발전과 함께 다양한 사회적, 윤리적 도전에 직면하고 있습니다. 그중에서도 인공지능(AI)의 활용은 보건의료 분야에서 혁신적인 가능성을 열어주는 동시에, 개인정보 보호와 공정성, 책임성 등 새로운 윤리적·법적 과제를 제기하고 있습니다. 이에 본 보고서는 보건의료 분야에서 AI 기술의 잠재력을 극대화하고, 동반되는 법적·윤리적 이슈에 대한 체계적 접근을 통해 지속 가능한 발전 방안을 제시하고자 합니다.

본 연구는 글로벌 AI 헬스케어 시장의 성장 전망과 더불어, 한국이 가진 우수한 ICT 인프라와 방대한 보건의료 데이터를 바탕으로 글로벌 선도 국가로 나아갈 가능성을 모색하였습니다. 동시에, 국내외의 주요 AI 정책 및 법제도 사례를 분석하여 한국적 환경에 맞는 입법 전략을 제안하였습니다. 특히, AI 기술의 투명성과 공정성, 안전성과 책임성을 강화할 수 있는 실효성 있는 대안을 탐구하였습니다.

이 보고서가 보건의료 AI의 안전하고 효과적인 도입과 활용을 위한 정책적·법적 기초 자료로서 유용하게 활용되기를 기대하며, AI 기술이 보건의료 분야에서 인간의 건강과 행복을 증진시키는 데 기여할 수 있기를 바랍니다.

2024년 12월
국회미래연구원

목 차

제1장 연구 목표 및 필요성 1

제1절 연구 질문 및 연구 분석틀 3

1. 인공지능의 발전과 윤리적 문제 3
2. 본 보고서의 목적 및 구성 4

제2장 보건의료 분야 인공지능 기술 현황 및 전망 5

제1절 인공지능의 정의 및 관련 기술 7

1. 인공지능 정의 7
2. 인공지능 학습 방법에 따른 분류 8
3. 인공지능의 역량과 범위에 따른 분류 11

제2절 보건의료 인공지능 활용 현황 및 전망 13

1. AI 기술로 인한 정밀의료로의 패러다임 가속화 13
2. 의료인공지능의 주요 적용 지점 16
3. 보건의료 분야 인공지능 산업의 동향과 전망 19

제3장 국내외 보건의료 AI 관련 R/D 주요 전략 23

제1절 미국의 보건의료 분야 AI R/D 주요 전략 25

1. 국가 AI R&D 전략계획 25
2. 미국의 보건의료 분야 AI R/D 전략 28

제2절 캐나다의 보건의료 AI 관련 R/D 주요 전략	30
1. 범캐나다 인공지능 전략 (PanCanadian Artificial Intelligence Strategy) ..	30
2. 캐나다의 보건의료 분야 AI R/D 전략	31
제3절 싱가포르의 보건의료 AI 관련 R/D 주요 전략	33
1. 싱가포르의 국가 AI 전략(National AI Strategy)	33
2. 싱가포르의 보건의료 분야 AI R/D 전략	35
제4절 일본의 보건의료 AI 관련 R/D 주요 전략	38
1. 일본의 AI 국가 전략	38
2. 일본의 보건의료 분야 AI R/D 전략	41
제5절 국내 보건의료 AI 관련 R/D 주요 전략	43
1. 국내 AI R/D 전략	43
2. 국내 보건의료 관련 R/D 전략	47
제4장 국내외 인공지능 윤리 관련 법제도	55
제1절 UN의 인공지능 윤리 관련 법제도	57
1. 인류를 위한 인공지능 거버넌스	57
제2절 WHO의 인공지능 윤리 관련 법제도	59
1. 세계보건기구의 인공지능 가이드라인	59
2. 보건 분야 인공지능의 윤리와 거버넌스: 대규모 멀티모달 모델에 대한 지침 ..	62

제3절 유럽연합의 인공지능 윤리 관련 법제도	63
1. 「신뢰할 수 있는 인공지능 윤리 가이드라인」	63
2. 「인공지능법」	65
3. 유럽연합의 보건의료 분야 인공지능 관련 가이드라인	69
제4절 미국의 인공지능 윤리 관련 윤리적 법제도	72
1. 미국의 보건의료 AI 법제도	72
2. 미국의 보건의료 분야 인공지능 관련 가이드라인	74
제5절 캐나다의 인공지능 윤리 관련 법제도	77
1. 캐나다의 인공지능 관련 법제도	77
2. 캐나다의 보건의료 분야 인공지능 관련 가이드라인	79
제6절 싱가포르의 인공지능 윤리 관련 법제도	82
1. 싱가포르의 AI 거버넌스 프레임워크	82
2. 싱가포르 보건부의 Artificial Intelligence in Healthcare Guidelines	84
제7절 국내 인공지능 관련 법제도	86
1. 과학기술정보통신부의 “사람이 중심이 되는 국가 인공지능 윤리기준”	86
2. 한국지능정보사회진흥원의 지능정보사회 윤리 가이드라인	87
3. 대한의료인공지능학회의 의료 인공지능 윤리 가이드라인	89

제5장 보건의료 분야 인공지능 관련 윤리적 과제 해결 및 활용 확대 방안	91
제1절 개인정보 보호	93
1. 한국의 보건의료빅데이터	93
2. 빅데이터 활용의 제도적 한계	94
3. AI를 활용한 가명정보의 활용 시 문제점	95
4. 개선방안	96
제2절 투명성과 설명가능성	98
1. 인공지능에 대한 설명의 의미	98
2. 설명가능한 인공지능 관련 기술	100
3. 보건의료 분야에서의 XAI의 필요성과 기술현황	102
4. XAI의 한계점 및 개선방안	108
제3절 공정성과 차별금지	110
1. 인공지능의 공정성과 차별가능성	110
2. 건강 데이터 편향	111
3. 건강 격차와 알고리즘 편향	113
4. 보건의료 분야 알고리즘 공정성을 위한 기술	115
5. 보건의료분야 공정성 향상을 위한 개선방향	117
6. 보건의료분야 공정성 향상을 위한 입법과제: 감독기구 및 절차의 법제화	118
제4절 안전 및 책임성	120
1. 인공지능의 안전과 책임성(responsibility)	120
2. 인공지능 책임 귀속의 어려움	120
3. 의사의 설명의무 및 환자의 자기결정권에 대한 영향	123

제5절 인간의 기술통제	124
제6절 효율성의 평가: 국내 보건의료 지불보상제도 개선방안	126
1. 디지털치료기기 및 보건의료 AI에 대한 지불보상제도 접근 현황	126
2. 가치기반 대안적 지불보상제도와 보건의료 AI 도입의 연관성	128
3. 국내에서의 대안적 지불보상제도 논의와 보건의료 AI의 접목 가능성	129
제6장 결론	131
참고문헌	135
1. 문헌자료	137
2. 웹사이트	138
Abstract	147

[표 3-1] 미국 「국가 AI R&D 전략계획」 비교 분석 27

[표 3-2] 2021년 캐나다 인공지능 전략 예산안 31

[표 5-1] 디지털치료기기 개발원가 산정기준 127

그림 목 차

[그림 2-1] 디지털 기술의 발전으로 변화하는 헬스케어 서비스 유형	13
[그림 2-2] 정밀의료에 대한 전통적인 공중보건적 접근과 정밀의료적 접근의 비교	14
[그림 2-3] 전 세계 AI 헬스케어 시장 규모 전망	20
[그림 2-4] 세계 AI 의료 및 헬스케어 시장 규모 및 전망	20
[그림 2-5] 제공 형태별 AI 의료 및 헬스케어 시장 규모 및 전망	21
[그림 3-1] NAIS 2.0의 세 가지 시스템 및 10개 조력자	34
[그림 3-2] 일본의 AI 기술전략 회의	39
[그림 3-3] 일본의 건강·의료·돌봄 분야 AI 기술 산업화 로드맵	40
[그림 3-4] 보건의료분야 AI의 활용의 6가지 중점영역	42
[그림 3-5] 인공지능(AI) 국가전략의 비전 및 목표	43
[그림 3-6] 국가 AI 전략 정책 방향의 비전과 전략, 추진 체계	45
[그림 3-7] 『보건의료 데이터·인공지능 혁신 전략』의 개요도	50
[그림 3-8] 『의료 인공지능 연구개발(R&D) 로드맵』의 개요도	53
[그림 4-1] WHO 인공지능 가이드라인의 주요 내용	59
[그림 5-1] 블랙박스 AI vs. 설명가능한 AI	99
[그림 5-2] AI 기반 흉부 X-ray 진단 보조 소프트웨어를 통한 판독결과	104
[그림 5-3] 비정상적인 QRS파와 T파가 AI 기반 심정지 예측에 활용된다는 XAI 연구 결과 ..	106
[그림 5-4] AI 기반 병리영상 분석 알고리즘을 통한 바이오마커 규명	107
[그림 5-5] 인공지능 알고리즘의 편향성	111
[그림 5-6] 인공지능 학습에 활용되는 데이터의 편향	114

요 약

1 연구 목표 및 필요성

- 인공지능(AI)의 의료 분야 활용이 가속화되고 있음
 - 미국 등의 주요국은 국가 주도 거버넌스 및 대형 프로젝트와 국민 참여 유도를 통해 AI 의료 활용을 촉진
 - 글로벌 AI 헬스케어 시장은 2030년까지 매년 368% 성장, 1조 3,452억 달러에 이를 것으로 예상
 - 국내 AI 헬스케어 시장은 연평균 508% 성장, 2030년 66억 7,200만 달러 도달 전망
- 한국은 방대한 데이터와 뛰어난 ICT 역량으로 선도 가능성이 큼에도 불구하고 보건의료 데이터 활용 및 법윤리적 제약이 여전히 존재
 - 보건의료 분야는 안전성, 효용성, 국민 건강 보호를 위한 기술적·법적 제약이 필요한 분야로서, 국제기구(WHO, EU 등)는 AI 활용의 법윤리적 문제에 대한 정책적 준비를 진행 중임
 - 국내는 보건의료 데이터 활용 정책과 AI에 대한 사회적 인식 성숙이 미흡한 실정임
 - AI 기술의 장점을 극대화하고, 법윤리적 문제를 해결할 중장기 입법 방안을 제안하되, 기술 발전과 윤리 문제의 균형적인 해결방안을 모색함

2 보건의료 분야 인공지능 기술 현황 및 전망

- 인공지능 정의와 개념
 - AI는 데이터를 수집, 해석, 학습하여 목표를 달성하거나 작업을 수행하는 머신 기반 시스템(OECD, 2019)으로 정의할 수 있음

● 인공지능의 구분

- AI 학습 방법에 따라 심볼릭 러닝(논리적 규칙과 기호를 기반으로 문제를 해결하는 초기 AI 방식), 머신러닝(데이터를 통해 패턴을 학습하는 방식으로, 지도 학습, 비지도 학습, 강화 학습, 반지도 학습 등으로 나뉨), 딥러닝(인공신경망을 활용해 복잡한 비선형 관계를 학습하며, 이미지·음성 인식 등에서 탁월한 성과를 보임)으로 구분
- AI의 역량과 범위에 따라서는 특정 분야에 특화된 AI(예: 알파고, 챗봇)인 약 인공지능(Weak AI)과 인간처럼 자유롭게 사고하고 자아를 지닌 이론적 AI인 강 인공지능(Strong AI)으로 구분

● AI의 보건의료 분야 적용

- 정밀의료: 환자의 유전적, 환경적 특성을 반영한 맞춤형 의료 서비스로, AI와 빅데이터 기술이 핵심 동력
- 질병 예방: AI와 웨어러블 디바이스를 활용해 개인 맞춤형 예방 활동 가능
- 질병 진단: 의료 데이터를 심층 분석하여 진단 정확성을 향상(예: 영상의학, 병리학)
- 치료: AI 기반 환자 참여 및 관리 시스템으로 치료 순응도 향상
- 모니터링: 웨어러블 기기를 통해 환자의 생체 신호 데이터를 분석, 이상 징후 파악
- 보건의료 시스템: 진료 전후 활동 자동화, 자원 배치 최적화, 보험 청구 간소화

● 보건의료 AI 산업 동향

- 글로벌 시장 전망: 2030년까지 연평균 약 37% 성장, 약 1,880억 달러 도달 예상
- 한국 시장: 2023년 3억 7,700만 달러에서 2030년 66억 7,200만 달러로 연평균 508% 성장 예상
- 시장 구성: 과거 하드웨어 중심에서 소프트웨어 중심으로 변화하며 지속 성장

3 국내외 보건의료 AI 관련 R/D 주요 전략

● 미국

- 미국은 2023년 발표한 “국가 AI R&D 전략계획”에서 9가지 전략으로 AI 기술의 책임 있는 개발과 국제 협력을 강조하고 있음
- 주요 전략으로는 윤리적·사회적 영향 이해, 안전성 강화, 공공 민간 협력, 국제 협력 강화, 고품질 데이터셋 및 표준 개발, AI 인력 양성과 지속적인 투자 촉진이 있음
- 미국은 의료 데이터 분석, 진단, 치료 계획 등에 AI 활용하고자 하며, 미국 보건부에서는 의료 AI 인재 육성, 기술 확산 및 데이터 투명성 확보를 추진하고 FDA에서는 AI 기반 의료기기 승인 절차 개선하고 있음. 국립 보건원에서는 “All of Us” 프로그램을 통해 대규모 건강 데이터를 활용하여 질병 예측 및 맞춤형 치료 개발을 추진하고 있음

● 캐나다

- “범캐나다 인공지능 전략”을 2017년 발표하면서 연구·인재 육성에 초점을 두고 있음
- AI 연구기관(CIFAR, Vector, Mila, Amii)을 중심으로 의료 AI 적용을 확대하고 있음
- 주요 사례로 Amii와 Alberta Health Services의 협력으로 질병 조기 발견, Mila의 암 진단 모델 연구, Vector Institute의 의료 영상 분석 및 맞춤형 의료 서비스를 볼 수 있음
- 보건의료 AI R&D 전략도 “범캐나다 인공지능 전략”의 일환으로 추진되고 있으며 공공 보건 시스템과 협력하여 AI 기술 초기 적용, 의료 데이터 분석 및 병원 자동화 기술 개발, 글로벌 규제 환경에서 초기 AI 기술 테스트 지원 등을 추진하고 있음

● 싱가포르

- “국가 AI 전략(NAIS 2.0)”을 통하여 AI를 ‘필수 기술’로 정의하며, 글로벌 리더십 확보와 시스템 접근 방식을 채택함

- 이를 추진할 3가지 시스템으로 AI 적용 산업 확대, AI 전문가 15,000명 양성, 신뢰 가능한 데이터 인프라 구축을 추진함
- 보건의료 AI R&D 전략으로는 데이터 인프라 구축(SingHealth 기반 데이터 통합 관리), 질병 예측 및 진단(암, 만성질환의 조기 발견 및 맞춤형 치료), 의료 자동화(AI 기반 챗봇, 가상 간호사, 로봇 수술), 스마트시티 헬스케어(IoT 센서를 활용한 개인화된 건강 관리) 등을 추진하고 있음

● 일본

- “Society 5.0 구현”을 위해 2017년 산업화 로드맵을 구축하면서 건강, 의료, 간호 분야에 집중함
- 보건의료 AI R&D 전략 중 6대 중점 영역으로는 게놈 의료(암 유전자 정보 관리), 화상 진단(의료 영상 데이터베이스 구축), 진단·치료 지원(질병 예방 및 관리), 의약품 개발(AI 기반 신약 후보 탐색), 간병·인지증(조기 발견 및 예측 도구 개발), 수술 지원(데이터 기반 표준화)를 선정함

● 한국

- “AI 국가전략”(2019년)을 통해 2030년까지 AI로 디지털 경쟁력 세계 3위, 경제효과 455조 원 창출을 목표로 함
- 주요 전략으로는 AI 인프라 확충, 인재 양성, 산업별 AI 활용 확대, 윤리적 AI 생태계 구축이 있음
- “보건의료 데이터·인공지능 혁신 전략”(2021~2025)을 통해 고품질 의료 데이터 생산·개방, 빅데이터 플랫폼 완성, 데이터 활용 혁신 및 신뢰성 있는 거버넌스 구축을 추진함
- “의료 인공지능 연구개발(R&D) 로드맵”(2024~2028)을 통해 필수의료 및 신약 개발에
- 종합적으로 보면, 각국은 AI를 보건의료 분야에 적극 도입하며, 데이터 분석, 질병 예측, 맞춤형 치료 등에서 활용도를 높이는 동시에 윤리적 문제와 데이터 보안, 국제 규제 협력에도 중점을 두고 있음

4 국내외 인공지능 관련 윤리적 법제도

• UN의 AI 거버넌스

- “인류를 위한 인공지능 거버넌스” 보고서(2023)를 통해 국제 협력과 조정을 통한 AI 거버넌스 체계의 필요를 역설함
- 7대 권고사항으로 국제 AI 과학 패널 설립, 정기적인 AI 정책 대화, AI 표준 교환 플랫폼, 개발도상국 지원 AI 역량 개발 네트워크, AI 글로벌 기금, 글로벌 AI 데이터 프레임워크, UN 사무국 내 AI 전담 부서 설립을 제시함

• WHO의 AI 윤리 가이드라인

- WHO는 “인공지능(AI) 가이드라인”을 통해 형평성 및 포용성, 투명성 및 설명 가능성, 책임성, 개인정보 및 데이터 보호, 안전성 및 보안, 인권 및 윤리적 사용의 6가지 윤리적 원칙을 제시함
- 향후의 위험과 도전 과제로는 AI의 불평등 심화, 데이터 편향, 투명성 부족 및 의료 현장에서의 윤리적 문제를 강조함

• EU의 AI 윤리 및 법제도

- “신뢰할 수 있는 인공지능 윤리 가이드라인 AI 윤리 가이드라인”(2019)을 통해 인간의 존엄성, 개인 자유, 법치주의, 평등, 시민의 권리 보장의 원칙을 제시함
- 4가지 윤리원칙으로는 인간 자율성 존중, 피해 예방, 공정성, 설명 가능성을 제시함
- 이후 EU “인공지능법”(AI Act, 2024)을 통해 AI 시스템을 위험 수준(허용 불가, 고위험, 제한적, 최소 위험)으로 분류하였으며 의료는 교육, 법 집행 등 민감 분야에 속하는 고위험 AI로 분류되어 엄격한 규제의 대상임

• 미국의 AI 윤리 및 법제도

- 미국은 행정명령을 통해 AI 기술 리더십 유지, 신뢰할 수 있는 AI 사용, 안전하고 신뢰할 수 있는 AI 개발과 사용 등의 제도를 만들어감

- “AI 권리장전”을 통해 안전한 시스템, 알고리즘 차별 방지, 개인정보 보호, 자동화 시스템 고지, 인간 대안 마련의 5대 원칙을 제시함

- **캐나다의 AI 윤리 및 법제도**

- 디지털 현장 이행법안(Bill C 27)을 통해 AI 시스템 설계·개발·활용 규제하고 고위험 AI에 대해서는 위험성 평가, 데이터 품질 관리, 중대한 피해 발생 시 신속 고지하도록 함
- 온타리오주 보건의료 AI 가이드라인에서는 AI의 잠재적 이점(진단·치료 개선, 효율성 향상)과 과제(편향, 보안 등) 제시하고 윤리적 AI 개발, 데이터 인프라 구축, 공공 민간 협력 등을 권고하고 있음

- **싱가포르의 AI 윤리 및 법제도**

- AI 거버넌스 프레임워크(2019)를 통해 인간 중심, 책임성의 핵심가치 아래, 데이터 품질 관리, 알고리즘 투명성, 설명 가능한 AI(XAI) 기술 활용 등을 추진함
- AIHG(Healthcare AI Guidelines)를 통해 환자 안전 및 복지, 의료진 책임, 투명성 및 설명 가능성, 데이터 프라이버시, 공정성의 핵심 원칙을 제안함

- **한국의 AI 윤리 및 법제도**

- AI 윤리기준(2020)을 통해 인간의 존엄성, 사회 공공선, 기술의 합목적성의 3대 원칙과 인권 보장, 프라이버시 보호, 투명성, 안전성 등의 10대 요건을 제시함
- 대한의료인공지능학회 의 의료 AI 윤리 가이드라인은 인간 존엄성 존중, 사회적 책임, 안전성 및 유효성, 투명성, 개인정보 보호, 책임 있는 사용의 핵심 원칙을 제시함

5 보건의료 분야 인공지능 관련 윤리적 과제 및 해결 방안

● 개인정보 보호

- 한국의 보건의료 빅데이터: 한국은 국민건강보험 및 전자의무기록(EHR)을 기반으로 방대한 의료 데이터를 보유하고 있어 AI 개발에 유리하나, 민감정보인 건강 데이터를 다루기 때문에 개인정보 유출 위험이 존재함
- 빅데이터 활용 제약: 개인정보보호법은 건강 정보와 같은 민감정보의 활용에 사전 동의를 요구해 빅데이터 활용을 제한하며, 이를 보완하기 위해 데이터 3법 개정과 마이데이터 정책이 도입되었으나 기술적·제도적 한계가 남아 있음
- 가명정보 활용의 문제점: 가명처리를 통한 개인정보 보호가 도입되었지만, 데이터 결합 시 재식별 가능성이 여전히 문제가 됨
- 개선방안: 데이터 보호 원칙 준수와 사회적 합의를 통한 개인정보 보호와 기술 발전 간 균형이 필요함

● 투명성과 설명가능성

- 블랙박스 AI 문제: 복잡한 AI 알고리즘의 의사결정 과정이 불투명해 신뢰받기 어려움
- 설명가능 AI(XAI): XAI는 AI의 의사결정 과정을 사용자와 개발자가 이해할 수 있게 하며, 의료 분야에서 특히 중요한 역할을 함
- 적용 사례: 의료영상 판독, 심정지 예측, 맞춤형 치료에서 XAI가 적용되어 신뢰도와 활용도를 높임
- 한계와 개선: XAI의 성능과 정확성 간의 균형, 사용자 중심 평가 기준 마련이 필요함

● 공정성과 차별금지

- 데이터 편향 문제: AI는 학습 데이터의 편향을 그대로 학습하며, 이는 특정 집단에 대한 차별을 유발할 가능성이 큼
- 건강 데이터 격차: 유럽 혈통 중심의 데이터 편향과 의료 시스템 차이는 AI 성능의 불균형으로 이어짐

- 개선 방향: 데이터 수집 단계부터 대표성 확보, 연합 학습, 편향 완화 기술 개발, 공정성 모니터링 제도 필요함

- **안전 및 책임성**

- 책임 분배 문제: 의료 AI 도입 시 개발자, 의료인, 의료기관 간의 법적 책임 분배가 모호해짐
- 설명의무와 자기결정권: AI를 활용한 진단·치료 과정에서 의사는 AI의 오류 가능성과 작동 방식을 환자에게 충분히 설명해야 함
- 개선방안: 의료 AI의 책임 소재 명확화와 이를 위한 법적·제도적 체계 구축이 필요함

- **인간의 기술통제**

- 강인공지능 시대의 윤리적 문제: 자율적 판단이 가능한 강인공지능은 기존 법·윤리 체계에 도전하며, 법적 인격 부여 여부가 쟁점이 됨
- 국내외 사례: 유럽연합의 로봇시민법, 국내 로봇기본법 등이 강인공지능의 책임과 윤리적 기준을 논의함
- 필요성: 기술 발전 속도를 고려할 때 강인공지능의 법적·윤리적 문제 해결을 위한 논의가 중장기적으로 요구됨

- **효율성의 평가: 국내 보건의료 지불보상제도 개선방안**

- ‘디지털치료기기 건강보험 등재 가이드라인’의 경우 디지털 및 인공지능 혁신의료기술이 ‘통합심사평가’를 통해 신속히 사용 승인을 받아 ‘임시등재’가 된 후, RWD에 기반한 근거를 창출한 후 새로이 신의료기술평가를 받아 정식등재가 가능한 건강보험 등재 경로를 제시함
- 이때 이들 기술들에 적용되는 지불보상제도가 실제 신기술을 개발하는 주체들에게 충분한 개발 동기를 제공하고 보건의료 시장에서 AI에 기반한 가치를 창출하기에 적합한 제도인지에 대해서는 추가적인 검토가 필요함
- 보건의료 체계 내의 시스템적인 효율을 유도하기 위해 현재 행위별 수가제 하의 원가기반의 지불보상제도가 아닌 책임의료조직(Accountable

Care Organization, ACO)으로 대표되는 대안적 가치기반 지불보상제도의 활용 등을 모색할 필요가 있음

6 결론

- 본 보고서는 보건의료 분야에서 AI 기술을 활용하는 데 있어 윤리적 문제와 법적 과제를 균형 있게 다루며, 이에 대한 입법 및 정책 방향을 제안함
- AI의 보건의료 분야 활용이 확대됨에 따라 개인정보 보호, 설명가능성, 공정성 등 윤리적 이슈가 중요해짐
- 국제기구와 주요국은 AI 활용 윤리 가이드라인을 마련했으나, 한국은 아직 보건의료 분야에 특화된 국가적 윤리적 지침이 부족함
- AI의 개발 과정에서 윤리적 문제를 해결하려는 노력은 기술적·사회적 비용 증가를 초래할 수 있으나, 국가적 지원을 통해 이를 해결할 필요가 있음

● 법윤리적 문제 해결을 위한 전략 방향

1. 보건의료 분야에 특화된 국가적 윤리 가이드라인 마련

- 보건의료 분야 AI 기술의 고위험 특성을 고려한 규제 및 윤리 프레임워크 개발 필요가 있음
- 환자 안전과 데이터 보호를 보장하며, 책임성 확보를 위한 법제화 필요
- 중장기적으로는 AI가 의사를 대체하거나 전문성을 초월할 경우 발생할 법적·윤리적 문제를 선제적으로 논의할 필요가 있음
- 독립적인 감독 기구 및 법적 절차를 마련해 규제 체계를 확립해야 함

2. 다학제적 전문가 구성

- 보건의료 AI 연구는 윤리학자, 의료 전문가, 데이터 공학자, 정책 연구자, 법학자, 환자 및 일반인을 포함한 다학제적 연구 생태계에서 이루어져야 함
- 임상적·윤리적·사회적 측면에서 통합적으로 기술의 위험성 평가 수행이 필요함

3. 사회적 합의 형성

- AI의 개인정보 보호와 효용성 간의 균형, 설명가능성의 기준, 공정한 기술 활용 등을 둘러싼 사회적 논의가 필요함
- 다양한 이해관계자와 국민이 참여하는 사회적 합의 과정을 통해 법윤리적 문제를 해결해야 함
- 국회가 국민 논의와 합의의 중심 역할을 수행할 필요가 있음

제1장

연구 목표 및 필요성

제1절 연구 질문 및 연구 분석틀

제 1절

연구 질문 및 연구 분석틀

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 인공지능의 발전과 윤리적 문제

4차 산업혁명 발전에 따라 빅데이터를 기반으로 한 인공지능(Artificial Intelligence 이하, AI)의 의료현장 적용이 가속화되고 있다. 미국 등 주요국의 경우 국가 주도 거버넌스와 대형 프로젝트 추진과 더불어 자발적 국민 참여 유도를 통해 보건의료 분야의 AI 활용을 가속화하고 있다. 우선, 미국에서는 민간 의료보험 중심 국가임에도 불구하고 보건의료데이터 활용에 있어서는 국가주도적 거버넌스와 대형 프로젝트 기획 추진, 의료데이터 정책 총괄을 위한 ASTP/ONC(Assistant Secretary for Technology Policy/Office of the National Coordinator for Health IT) 등 공공기구 설치 등이 추진되고 있다. 아울러 질병 극복, 산업적 활용 등이 가능하도록 데이터 권리 주체인 국민대상 캠페인과 기술개발 추진, 2차 활용에 대한 자발적 참여 유도를 통해 보건의료 분야 AI 활용이 급성장하고 있다.

AI를 포함한 전 세계의 디지털 헬스 시장은 급속도로 성장하고 있으며, 2030년에는 그 규모가 1조 달러를 훌쩍 넘어설 것으로 예상된다. 마켓앤마켓은 AI 헬스케어 시장이 2030년까지 매년 36.8%씩 성장하여 1조 3,452억 달러 규모를 형성할 것으로 내다봤다(MarketsandMarkets, 2023). 이는 헬스케어 분야에서 AI 기술의 활용이 빠르게 확대되고 있음을 반증하는 것이다.

국내 시장 또한 높은 성장세를 보이고 있다. 한국의 AI 헬스케어 시장은 2023년 3억 7,700만 달러에서 글로벌 평균(41.8%)과 아시아 평균(47.9%)을 상회하는 연평균 50.8% 성장하며 2030년 66억 7,200만 달러를 기록할 것으로 전망된다(MarketsandMarkets, 2023). 우리나라는 ① 잠재가치 높은 방대한 데이터 보유(병원, 전국민 건강보험), ② 뛰어난 ICT 역량(AI, 정보통신네트워크) 등을 통해 '보건의료 빅데이터 및 AI 헬스케어'를 선도할 수 있는 잠재적인 가능성을 보유하고 있다.

그럼에도 불구하고 보건의료 분야의 AI 활용에 있어서 보건의료데이터 활용 등에 있어서 실제적인 한계 요인과 법윤리적 제약 요인이 상존하고 있다. 보건의료 분야는 관련 산업의 발전과 동시에 국민 건강을 위한 기술의 안전과 효용성 등을 고려해야 하는 특수한 분야이다. WHO와 EU 등 국제기구와 선진국의 경우, AI 기술의 흐름에 발맞추어 보건의료 분야에서의 AI 활용이 가져올 수 있는 법윤리적 문제들에 대한 정책적 준비를 활발하게 진행하고 있다. 우리나라도 보건의료 데이터 활용을 위한 정책적 여건과 AI 활용에 대한 사회적 인식 성숙이 같이 이루어져야 할 필요가 있으나 이에 대한 논의가 활발하게 이뤄지지 못하고 있는 실정이다.

2 본 보고서의 목적 및 구성

이에 본 연구는 보건의료 분야에서 AI 기술의 장점을 극대화하고 잠재적인 문제점이 될 수 있는 법윤리적 이슈에 대한 문제 제기와 더불어 중장기적인 입법방안을 제안하여 미래 보건의료의 발전과 변화에 선제적으로 대응하고자 한다. AI 기술로 인한 윤리적인 문제는 단순히 규제적인 측면으로만 접근하는 것이 아니라 기술적인 발전을 통해 해결하려는 노력들이 함께 이뤄지고 있다. 특히, 학계 및 스타트업 등 산업계를 중심으로 신뢰성 있는 AI 개발 및 설명가능한 AI 개발 등이 활발하게 이뤄지고 있어 이를 반영하였다.

이에 본 보고서는 AI 기술과 법윤리적 내용을 균형있게 다루고자 하였다.

보고서 2장에서는 AI의 기술적 개관 및 보건의료분야 AI의 기술, 산업적 측면에서의 발전 전망을 살펴본 후, 3장에서는 국내외 AI의 연구개발과 관련된 주요국들의 정책들을 탐색적으로 살펴보고자 한다. 이후 4장에서는 국제기구 및 주요국들의 윤리적 규범들을 살펴보고 범위를 보건의료 분야로 국한하여 보건의료분야에 특성화된 윤리적 가이드라인 등을 검토하였다. 5장에서는 향후 AI이 보건 분야에서 적극적으로 활용될 경우 이슈가 될 수 있는 5가지 쟁점을 중심으로 연구를 수행하였다. 5가지 쟁점은 1. 개인정보 2. 투명성과 설명가능성 3. 공정성과 차별금지 4. 안전, 보안 및 책임성 5. 인간의 기술통제, 6. 효율성의 평가: 국내 보건의료 지불보상제도 개선방안을 다루면서 각 이슈에 대한 설명과 더불어 기술적 발전과 법윤리적 대응 등을 고찰하였다. 마지막 결론으로 각 보고서의 결론 및 향후 법윤리적 방향을 제시하였다.

제2장

보건의료 분야 인공지능 기술 현황 및 전망

제1절 인공지능의 정의 및 관련 기술

제2절 보건의료 인공지능 활용 현황 및 전망

제 1절

인공지능의 정의 및 관련 기술

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 인공지능 정의

인공지능(Artificial Intelligence, AI)은 “환경에서 데이터를 수집하고, 해석하며, 이 데이터에서 학습한 후 구체적인 목표를 달성하거나 다양한 설정에서 작업을 수행하기 위해 행동하는 머신 기반 시스템”(OECD, 2019)이다.

앨런 튜링은 “계산 기계와 지능” 논문에서 AI의 초기 개념을 제시하며, 기계가 인간 처럼 사고할 수 있는지를 평가하는 기준으로 튜링 테스트(Turing Test)¹⁾를 제안했다. 여기서 AI는 인간과 구분할 수 없을 정도로 지능적인 대화를 할 수 있는 시스템으로 정의된다. 튜링은 “기계가 생각할 수 있는가?”라는 질문보다는 “기계가 인간처럼 행동할 수 있는가?”에 중점을 두며, AI를 인간 지능과 유사한 방식으로 사고하고 학습하는 기계로 정의했다(Turing, 1950).

1956년 다트머스 회의(Dartmouth Conference)에서 Artificial Intelligence라는 용어를 만든 존 매카시(John McCarthy)는 AI를 “기계가 지능적 행동을 할 수 있도록 만드는 과학과 공학”으로 정의했다. 그는 AI를 통해 기계가 스스로 문제를 해결하고 논리적으로 추론할 수 있는 능력을 획득하게 된다고 보았다. 이는 AI가 사람의 지능적 활동을 기계에서 구현할 수 있는 도구로써 연구된다는 점을 강조한 것이라고 볼 수 있다(McCarthy, 2007).

볼프강 에르텔(Wolfgang Ertel)은 AI를 문제 해결을 위한 시스템으로 정의하며, 이는 복잡한 문제를 자동으로 해결하거나 결정을 내리는 기계적 지능으로 설명한다. 그의 연구에서는 특히 강화 학습과 딥러닝과 같은 현대 AI 기술이 지능적인 시스템을 어떻게

1) 한 방에 인간 응답자와 기계(AI)가 각각 위치해 있고, 다른 방에 있는 판별자는 이들 중 어느 쪽이 인간인지 모른 채로 질문을 하고 기계와 인간은 각각 텍스트를 통해 판별자의 질문에 답변하며, 판별자는 그들의 대답만을 바탕으로 누가 인간인지 맞추려고 한다. 이때 만약 판별자가 기계와 인간을 구분하지 못하거나 기계를 인간으로 착각하게 만든다면, 그 기계는 지능이 있다고 간주할 수 있다는 것이 튜링의 주장임

구축하는지에 관한 실용적 응용에 대한 학문적 논의를 반영한다(Ertel, 2018)

스튜어트 러셀(Stuart Russell)과 피터 노빅(Peter Norvig)은 AI가 목표로 하는 행동 방식을 기준으로 다음과 같은 네 가지 방식으로 정의될 수 있다고 주장한다(Russell & Norvig, 2016). ① 인간처럼 생각하기(Human-like thinking): AI가 인간의 사고 과정을 모방하는 것을 목표로 하는 접근, ② 인간처럼 행동하기(Human-like acting): 인간과 유사한 행동을 보이는 시스템을 만드는 것, ③ 이성적으로 생각하기(Rational thinking): 논리적 추론을 통해 문제를 해결하는 AI, ④ 이성적으로 행동하기(Rational acting): 합리적인 결정을 내리며 목표를 달성하는 AI. 이러한 정의는 AI가 어떻게 지능을 발휘하는지에 대한 다양한 측면을 반영하며, AI가 단순히 행동만 모방하는 것이 아니라 추론하고 결정할 수 있는 시스템이라는 점을 강조하고 있다.

결론적으로 AI의 정의에 있어서 사고 과정, 추론 능력, 문제 해결, 자율성 등을 강조하는 여러 정의가 존재하며, 연구자들은 각기 다른 관점에서 AI를 다루고 있다. 이는 AI 관련 기술의 급속한 발전에 기인하기도 한다(한국보건산업진흥원, 2018). 특히 이미 지나 음성 인식과 같은 데이터에서 패턴을 분석하는 능력은 인간의 평균적인 수준과 유사하거나 이를 능가하는 성과를 보이고 있다.

2 인공지능 학습 방법에 따른 분류

가. 심볼릭 러닝(Symbolic Learning)

심볼릭 러닝은 1950년대와 1960년대에 시작된 고전적인 AI 접근 방식이다. 이 방식은 인간의 논리적 추론 과정을 모방하려는 목표로, 인간의 추론 방식과 유사하게 논리, 기호(symbol), 규칙(rule) 등을 사용하여 지식을 표현하여 명시적이고 논리적인 규칙을 사용해 문제를 해결하거나 새로운 지식을 추론한다. 심볼릭 러닝은 복잡한 개념을 기호로 표현하고, 그 기호들 간의 관계를 수학적 또는 논리적으로 정의하기 때문에 AI는 투명하고 설명가능한 방식으로 학습하고 추론할 수 있다. 1956년 다트머스 회의에서 AI라는 용어가 처음 등장했을 때, 연구자들은 심볼릭 AI를 주요 연구 분야로 삼았으며, AI 연구 초기의 지배적인 패러다임이었다.

심볼릭 러닝은 명시적인 규칙과 논리를 바탕으로 학습하는 반면, 현대의 머신러닝(Machine Learning)은 데이터를 바탕으로 패턴을 학습하고, 추론 규칙을 자동으로 찾아내는 방식으로, 심볼릭 러닝은 설명 가능성과 추론의 정확성에서 장점이 있지만, 복잡한 비정형 데이터에서의 학습이 어렵다는 한계가 있다.

나. 머신러닝(기계학습, Machine Learning)

머신러닝은 명시적인 프로그래밍 없이 컴퓨터가 데이터를 바탕으로 패턴을 학습하고 경험을 통해 모델을 학습하여 예측이나 결정을 내리는 기술이다. 학습 방식에 따라 아래와 같이 기술적으로 분류할 수 있다(Goodfellow et al., 2016).

1) 지도 학습(Supervised Learning)

AI가 레이블²⁾이 달린 데이터를 학습하는 방식으로, 입력된 레이블을 학습시키면 AI는 그에 해당하는 출력을 제공한다. 예를 들어, 이미지 인식에서 고양이 사진과 “고양이”라는 레이블을 함께 입력하여 다음에 고양이 사진을 입력할 때 고양이라는 결과가 나오게 하는 것이다. 이때 사용되는 주요 알고리즘³⁾은 선형 회귀, 로지스틱 회귀, 서포트 벡터머신(SVM), 의사결정 나무, K-최근접 이웃(KNN) 등이 있다. 이러한 방법은 이메일 스팸 필터링, 이미지 분류, 의학적 진단에 활용된다.

2) 비지도 학습(Unsupervised Learning)

지도 학습과는 다르게 레이블이 없는 데이터를 분석해 자연스러운 패턴이나 관계를 찾는 방식으로, 주로 데이터의 분류나 군집화를 통해 데이터를 분류하거나 요약한다. 이때 사용되는 주요 알고리즘은 K-평균 군집화, 계층적 군집화, 주성분 분석(PCA), 독립 성분 분석(ICA) 등이 있으며 군집화, 이상 탐지, 시장 고객 세분화 등에 활용된다.

2) 레이블은 머신러닝에서 학습 데이터의 정답을 의미함. 레이블이 정확할수록 알고리즘의 성능은 높아지지만, 잘못된 레이블이 있을 경우 모델의 예측 오류가 커질 수 있다. 따라서 레이블된 데이터의 품질은 머신러닝 모델의 성능을 결정짓는 중요한 요소 중 하나로, 학습 데이터의 양이 충분하지 않거나 불균형할 경우에도 모델의 학습 과정에 문제가 발생할 수 있다.

3) 알고리즘은 데이터를 처리하고 학습하는 절차를 정의하는 일련의 수학적 연산 또는 명령어입니다. 머신러닝 알고리즘은 주어진 데이터에서 패턴을 학습하거나, 레이블을 기반으로 예측하는 과정을 수행합니다.

3) 강화 학습(Reinforcement Learning)

AI가 환경과 상호작용하며 보상(reward)과 처벌(penalty)을 통해 어떤 행동이 가장 좋은 결과를 낳는지, 즉 정책(policy)의 최적화를 위해 시행착오를 반복하며 학습하는 방식이다. 예를 들어, 자율주행차는 주행 중 얻은 데이터를 바탕으로 행동을 최적화할 수 있다. 사용되는 주요 알고리즘으로는 Q-러닝, 정책 경사(Policy Gradient), 심층 Q-네트워크(Deep Q-Network, DQN)가 있으며 자율주행차, 게임 AI, 로봇 공학 등에 활발하게 사용되고 있다.

4) 반지도 학습(Semi-Supervised Learning)

반지도 학습은 소량의 레이블 된 데이터와 대량의 레이블 없는 데이터를 함께 사용하는 방식으로, 레이블 된 데이터를 이용해 예측 모델을 학습한 후 레이블이 없는 데이터로 학습을 보강한다. 활용되는 주요 알고리즘으로는 그래프 기반 방법, 전파 학습(Propagation Learning) 등이 있으며 이미지나 텍스트 분류 등 대량의 라벨링되지 않은 데이터를 처리하는 문제에서 효과적이다.

다. 딥러닝(심층 학습, Deep Learning)

딥러닝은 머신러닝의 하위 분야로, 초기 딥러닝은 인간의 뇌 구조를 모방한 인공신경망(neural network) 모델이 발전하면서 계층 구조를 가진 신경망의 내부가 여러 단계로 이루어져, 다양한 환경과 빅데이터, 예측이 필요한 상황에 적합한 절차와 방법을 제시하는 방식으로 개발되었다. 1950년대 AI는 큰 기대를 받았으나, 알고리즘의 복잡성이 증가하면서 기존 컴퓨터로는 해결하기 어려운 문제들이 발생하기 시작했다(Cook, 1971). 당시의 알고리즘은 얇은 네트워크에서만 효과적으로 작동했으며, 레이어가 증가하거나 복잡성이 커질 경우 학습 속도가 느려지고 학습이 불안정해지며 과적합(오버피팅, overfitting)⁴⁾ 문제가 나타났다(LeCun et al., 1988).

그러나 최근 딥러닝 알고리즘의 발전과 더불어 하드웨어 및 빅데이터 기술의 발달로

4) 모델이 훈련 데이터에 너무 적합하게 학습되어 새로운 데이터에 대한 일반화 능력이 떨어지는 현상. 즉, 모델이 학습 데이터의 노이즈와 특이점까지 과도하게 학습함으로써, 훈련 데이터에 대해서는 매우 높은 성능을 보이지만 실제로는 새로운 데이터에 대해 좋은 예측을 하지 못하는 경우를 의미

학습 시간이 단축되고 과적합 문제가 완화되면서 딥러닝의 활용도가 크게 증가했다. 특히 이미지에서 객체 인식이나 음성, 신호 인식과 같은 데이터 패턴 분석에서는 인간과 유사한 성능을 보이고 있다. 딥러닝은 뇌의 신경세포 구조를 기반으로 한 신경망을 활용해 복잡한 비선형 관계를 학습할 수 있어, 의료 데이터 처리 분야에서 활발히 사용되고 있다. 또한, 비지도 학습을 통한 전처리 과정(Erhan et al., 2010), 컨볼루션 신경망(CNN⁵); Krizhevsky et al., 2017), 시계열 데이터를 위한 리커런트 신경망(RNN⁶); Irsoy & Cardie, 2014) 등의 기술 발전이 이뤄졌다.

이처럼 고도화된 딥러닝 알고리즘은 복잡한 환경에서 더욱 효과적으로 사용될 수 있으며, 컴퓨팅 성능의 향상과 빅데이터의 출현으로 딥러닝이 기존 신경망 학습 알고리즘보다 더욱 효율적으로 발전하고 있다(조영임, 2016). 하지만 현재는 주로 의료분야에서 AI의 기술적 적용 가능성을 확인하는 연구가 진행 중이며, 임상적 유효성 및 유용성을 검증하는 초기 단계에 있다.

3 인공지능의 역량과 범위에 따른 분류

가. 약 인공지능(Weak AI)

약 인공지능은 특정 분야에 최적화되어 인간의 한계를 보완하고 생산성을 높이기 위해 활용되는 인공지능을 의미한다. 자아의식이 없는 형태로, 대표적인 예로는 2016년 이세돌 9단과 대국을 벌인 ‘알파고’, 자연어 처리(NLP) 기반의 챗봇(예: GPT-3), 얼굴 인식 시스템 등이 있다. 약 인공지능은 알고리즘, 기본 데이터, 그리고 규칙을 입력받아 마치 지능이 있는 것처럼 작동하도록 프로그래밍되며, 주로 데이터 패턴을 해독하는 역할을 한다.

5) 패턴 인식과 객체 탐지에서 뛰어나 이미지 데이터 처리에 특화된 신경망

6) 순차적 데이터(예: 텍스트, 시계열 데이터)를 처리하는 데 강점을 가진 신경망

나. 강 인공지능(Strong AI)

반면, Artificial General Intelligence, AGI라고도 불리는 강 인공지능은 사람처럼 자유롭게 사고할 수 있는 자아의식을 지닌 인공지능으로, 약 인공지능과는 달리 지능을 스스로 가질 수 있는 분류라고 할 수 있다. 강 인공지능은 인간의 두뇌를 대체할 수 있을 만큼의 능력을 갖추고 있으며, 다양한 분야에서 과제를 수행할 수 있을 것으로 상상하고 있다. 즉, 강 인공지능은 이론적 개념으로 연구와 논의의 대상이며, 실현되었다고 할 수 있는 기술적 또는 실험적 사례는 아직 없다. 학술적인 근거는 아직 부재하며 근시일 내에 가능한 것이라고 보기는 어렵다는 것이 일반적인 의견이다.

이는 현재의 여러 기술적 한계와 더불어 인간의 의식과 자아라는 개념을 구현하는 것이 매우 어려운 문제로 남아 있기 때문이다. 예를 들어, MIT의 맥스 테그마크(Max Tegmark) 교수는 그의 저서 『Life 3.0』에서 AI의 미래에 있어 중요한 것은 의식의 문제라고 언급하였다(Max Tegmark, 2017). 그는 ‘주관적’으로 정의되는 의식도 지능과 같이 정보 처리 능력의 일부분이며, AI가 충분히 발전하면 이러한 주관적인 의식을 가질 수 있을 것이라고 주장했다. 하지만 그는 AI가 인간과 다른 목적을 가질 경우 미래에 인간에게 큰 위협이 될 수 있다고 경고하며, AI에 대한 윤리적 교육이 필수적이라고 강조했다. 이처럼 강 인공지능에 대한 논의는 다분히 이론적 가능성과 철학적 논의의 수준에 머물러 있다.

제2절

보건의료 인공지능 활용 현황 및 전망

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 AI 기술로 인한 정밀의료로의 패러다임 가속화

정밀의료(Precision medicine)는 각 환자의 유전적, 환경적, 생활 습관 요소 등 선천적·후천적 특성을 반영한 다양한 의료·헬스데이터를 기반으로 건강 상태를 판단하고 예측하거나 질병에 대해 맞춤형 치료를 제공하는 새로운 의료 접근 방식이다. 기존의 획일적 치료법과 달리 정밀의료는 예방에서 진단, 치료까지 건강 관련 전 과정에 대한 서비스를 개인에 맞춰 제공하는 것을 목표로 하는 미래 의료에 가장 근접한다고 볼 수 있다.

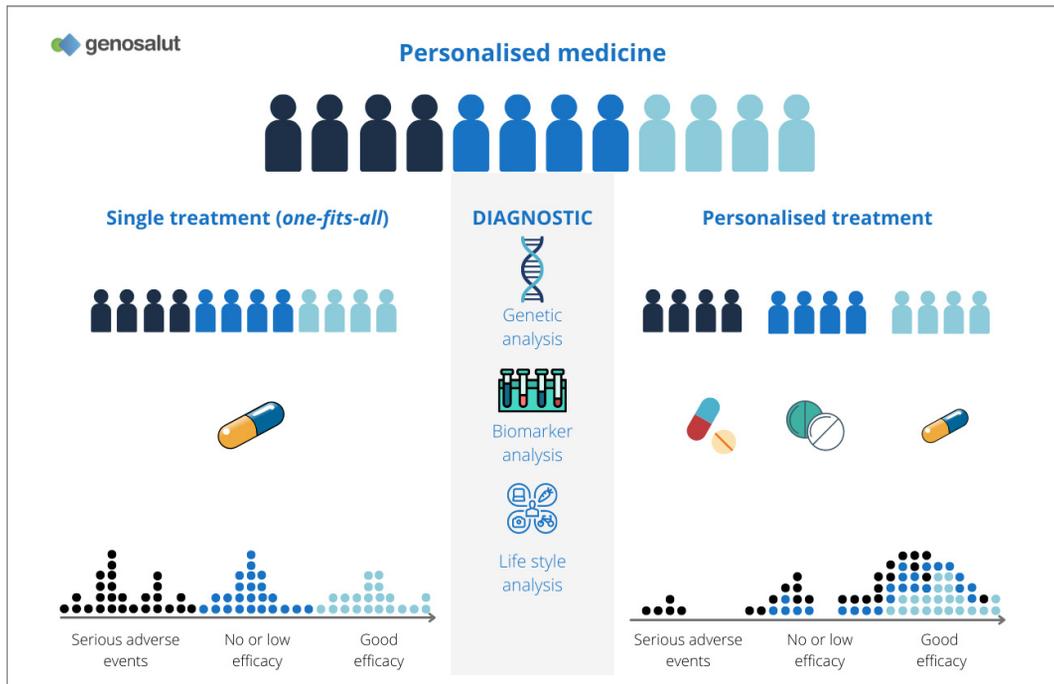


출처: 삼정KPMG, 2022

[그림 2-1] 디지털 기술의 발전으로 변화하는 헬스케어 서비스 유형

정밀의료는 기존의 전통적 공중보건적 접근과는 서로 다른 철학과 방법론을 바탕으로 한 의료 접근 방식이다. 공중보건의 목표는 형평성과 접근성에 중점을 두어 최대한 많은 사람이 기본적인 예방 조치와 건강서비스를 받을 수 있도록 하는 것이 핵심이다. 이를 위해 전통적인 공중보건적 접근은 역학적 데이터를 기반으로 질병의 원인과 분포를 분석하고, 전체 인구의 건강을 개선하는 데 중점을 두어 인구집단 단위의 질병 예방과 건강 증진을 목표로 백신 프로그램, 공공위생 개선, 금연 캠페인 등 보편적인 개입을 통해 널리 적용되는 정책을 주로 시행한다.

이에 비해 정밀의료는 그간 효율성 내지 형평성의 측면에서 한계가 많았던 개인 맞춤형 예방 및 치료에 초점을 두고 있다. 개인 맞춤형 개입을 통해 개인의 생물학적 다양성을 고려한 정밀한 치료를 제공하고 개인마다 다른 특정 약물이나 치료법에 대한 부작용을 줄일 수 있다.



출처: <https://www.genosalut.com/en/news/healthy-lifestyle/what-is-personalised-medicine/>

[그림 2-2] 정밀의료에 대한 전통적인 공중보건적 접근과 정밀의료적 접근의 비교

예를 들어, 공중보건에서는 암 예방을 위해 일반적인 식이 요법 또는 기존의 인구 집단 단위에서 평균적인 효과가 입증된 치료방법을 권장하였다면, 정밀의료는 개인의 유전적 특성에 따라 맞춤형 식단을 제안하거나 특정 암 환자의 증양에 맞는 유전자 분석을 통해 맞춤형 항암제를 제공할 수 있다(Garraway et al., 2013; Relling & Evans, 2015). 또한, 전통적인 공중보건적 접근에서는 인구 전체에 백신 접종을 권장하는 반면, 정밀의료는 특정 개인이 특정 백신에 대해 부작용을 보일 위험성이 있는지 유전적 요인을 분석하여 알 수 있다. 심혈관 질환에서는 콜레스테롤 수치 또는 고혈압과 관련된 유전적 변이를 식별하여 개인화된 치료 전략을 세우는 것이 가능해졌다. 당뇨병 관리에서도 유전적 위험 점수를 통해 질병 발병 및 합병증 발생을 예측할 수 있다(Schorck, 2015). 자가면역질환에서는 유전적 및 분자 프로파일링을 통해 보다 정확한 진단과 치료 방법이 개발되고 있다. 예를 들어, 개별 면역 반응을 유전적으로 이해함으로써 류머티스 관절염이나 다발성 경화증과 같은 질환의 맞춤형 치료가 가능해지고 있다(Collins & Varmus, 2015).

이러한 정밀의료가 가능하게 된 것에는 빅데이터, AI, IoT 등을 활용한 디지털 헬스(Digital Health)라고 부르는 거대한 패러다임이 동력이 되고 있으며 이러한 트렌드 안에서 AI는 예방(모니터링), 진단-예측, 관리, 치료-재활 전 분야에서 초개인화 정밀 의료 구현의 중요한 역할을 하고 있다(Topol, 2019). AI를 통해 수작업으로 처리하는 것보다 훨씬 더 효과적으로 다양한 건강 관련 데이터세트(환자 데이터, 의학 문헌, 라이프스타일 정보, 유전자 구성, 병력 등)를 수집, 저장, 분석할 수 있을 뿐만 아니라 데이터를 기반으로 더 빠르게 연구를 수행하고 개인별 맞춤 표적 치료법을 더 빠르게 개발할 수 있다. 이를 통해 질병의 조기 진단을 통해 치료 효과를 향상하고 의료 비용을 절감하는 동시에 치료의 효과를 증대할 뿐만 아니라 의료진이 환자 응대 및 지원 등에 더 많은 시간을 집중할 수 있도록 도와줄 수 있다. 나아가 데이터를 기반으로 개인의 생활 방식이 의료 결과에 어떤 영향을 미치는지에 대해 실증적으로 이해하게 되면 개인 수준에서도 더 건강한 사회적 행동을 유도하고 촉진할 수 있을 것으로 예상된다(Khoury & Galea, 2016).

2 의료인공지능의 주요 적용 지점

가. 질병 예방적 관점

질병 예방은 환자에게 장기적인 혜택을 제공할 뿐만 아니라, 예방 가능한 질환의 치료에 드는 비용을 절감할 수 있다. 기존의 예방 방법으로는 운동, 식단 관리, 예방접종 등이 있었지만, 이는 인구 집단의 전체적인 효과성을 기준으로 하기 때문에 개인의 유전적 특성을 충분히 반영하지 못했다. 그러나, AI를 활용하면 개인 맞춤형 예방 활동이 가능해질 것으로 예측된다. 더불어, AI와 웨어러블 기술의 발전은 질병의 조기 경고와 신속한 개입을 가능하게 하여 이러한 변화에 큰 역할을 하고 있다. 이는 질병의 사후적 대응에서 사전적 예방으로의 전환을 의미하며, AI가 이끌 중요한 패러다임의 변화라고 할 수 있다.

나. 질병 진단적 관점

질병의 진단은 다양한 요인의 영향을 받는 복잡하고 어려운 과정이다. 환자의 현재 증상과 검사 결과뿐 아니라 유전적 배경, 생활 방식, 병력 등의 요소를 모두 고려해야 하지만, 기존의 시스템과 절차만으로는 의료진이 이러한 정보를 총체적으로 얻는 데에는 어려움이 있었다. 하지만 AI 기술을 통해 질병을 더 효율적이고 정확하게 진단할 수 있는 가능성이 열리고 있다.

의료진은 다양한 출처에서 수집된 방대한 양의 건강 데이터를 심층신경망, 머신러닝, 범주화 등의 AI 기술을 통해 분석하여 질병 진단의 정확성을 높이고 시간과 비용을 절감할 수 있다. 그 대표적인 사례가 영상의학에서 AI를 활용한 영상 진단의 발전이다. 영상의학과 전문의들은 엑스레이, MRI, CT 스캔, PET 스캔, 초음파 등 다양한 의료 영상을 통해 질병과 부상을 진단하는 데 전문성을 갖추고 있지만, 이러한 전문성을 습득하기 위해서는 오랜 학습과 경험이 필요하다. AI의 영상 분석 성능과 정확성이 발전함에 따라, 영상의학과 전문의가 더 정확한 진단을 내리도록 돕고, 픽셀 수준의 세부 사항까지 확인할 수 있게 해주며, 주요 구조를 육안으로 확인하기 어려운 영상을 개선하고 분석하는 데 점점 더 많이 활용 될 수 있을 것으로 기대된다. 이처럼 AI를 활용한 진단 기술은 효과성과 효율성을 높여 의료진이 환자 관리와 지원에 더 집중할 수 있도록

록 한다. 또한, 영상의학 전문의가 부족하거나 업무가 과중한 상황, 특히 중저개발국의 보건의료 환경에서 유용하게 사용될 수 있다.

병리학 분야에서도 AI를 통해 질병 진단의 속도와 정확성을 향상시킬 수 있다. 병리학자는 채액과 조직(예: 혈액 검사, 생검)을 연구하여 질병의 원인과 특성, 영향을 파악하는데, 이 과정은 주로 현미경을 사용해 유리 슬라이드를 관찰하는 방식으로 이루어져 왔다. 그러나 디지털 병리학이 주목받으면서 AI가 채액과 조직의 패턴을 학습하여 질병을 더 신속하고 정확하게 진단하는 것이 가능해졌다. 예를 들어, AI는 극저온 전자 현미경의 결과 등에서 사람이 감지하기 어려운 패턴과 세부 사항을 분석하여, 병리학자가 환자의 질병에 대한 최적의 치료 방법을 신속하게 결정할 수 있도록 지원할 수 있다.

다. 환자 중심의 치료적 관점

전통적인 의학의 관점에서 환자는 의사와의 정보 비대칭으로 인해 치료 과정에서 의사의 지시에 수동적으로 따르게 되는 경우가 많았다. 그러나 AI의 발전은 진료 예약, 의료 기록 확인부터 의료진 및 코디네이터팀과의 소통에 이르기까지 환자 참여의 모든 측면을 개선할 수 있는 가능성을 열어주고 있다.

기존에는 환자들이 자신의 치료에 대해 이해하고 관리하는 과정에서 상당한 어려움을 겪었다. 많은 환자들이 진료 예약, 의료 기록 확인, 의료 서비스 선택부터 간단한 문의와 같은 여러 과정에서 어려움을 겪어 치료 과정에서 적극적인 참여가 어렵거나 자신의 건강 관리에 대해 부정확한 정보를 기반으로 의사결정을 하는 경우가 빈번했다.

AI 기반의 자연어 처리는 복잡한 의료 정보를 분석해 환자에게 의미 있는 정보를 전달함으로써 환자의 건강 정보 이해 능력을 향상시킬 수 있다. 더 나아가, 챗봇 등을 통해 환자는 질문에 대한 답변을 얻고, 진료 및 통화 일정 예약, 유관 부서와의 연결 등 다양한 업무를 보다 효과적으로 수행할 수 있다. 이를 통해 환자는 자신의 치료 계획을 더 잘 이해하게 되고 처방 준수 및 치료 순응도를 높일 수 있다.

정보의 비대칭성 해소로 인한 환자의 자기결정권 강화를 통한 환자 중심성 개선에 더해, 환자 중심성을 강화하는 또다른 메커니즘으로 케어 경로(care pathway) 최적화에 있어서의 AI의 가능성을 고려할 수 있다. 특히, 고령화로 인한 복합 만성질환의 증가와,

이에 따라 분절화 된 다양한 보건의료 서비스 공급자들의 협력을 통해 환자를 중심으로 한 통합된 의료·돌봄 체계의 구축(Person Centered Integrated Care)이 강조되고 있는 맥락을 고려할 때, 서비스 공급자들이 수용할 수 있는 투명성과 객관성이 높은 AI 기반 정보 공유를 토대로 한 케어 경로 최적화가 달성될 가능성이 있다.

다. 환자 모니터링의 관점

웨어러블 디바이스는 처방 이후 또는 치료 이후에 수집된 환자의 생체 신호 데이터를 AI를 활용해 분석하고 이상 징후를 파악할 수 있다. 팔찌, 심장 모니터, 패치, 센서가 부착된 의류 등 다양한 신체 센서와 웨어러블 디바이스에서 수집된 데이터는 착용자의 활동량과 체력 수준뿐만 아니라, 병원이나 진료소에서 환자의 생체 신호를 지속적으로 원격 모니터링하는 데 사용된다. 또한, AI로 시계열 데이터를 분석하여 주의가 필요한 이상 징후를 식별하는 의료 목적으로도 활용된다.

AI 기반의 가상비서나 챗봇 등은 환자의 복약 준수 및 원격 환자 모니터링을 지원함으로써 만성 질환 환자가 건강 관리 계획에 지속적으로 참여하도록 돕고, 이와 관련된 지속적인 연락을 가능하게 할 수 있다. 예를 들어, AI 이미지 인식과 스마트폰을 이용하면 외래 환자의 행동을 관찰하고, 환자가 지시에 따라 약물을 제대로 복용하고 있는지 원격으로 모니터링할 수 있다. 임상시험의 경우에도 임상시험 프로토콜을 준수하지 않으면 시험 결과가 무효화될 수 있기 때문에 이러한 AI 기반의 모니터링이 중요한 역할을 할 수 있다.

라. 보건의료 시스템적 관점

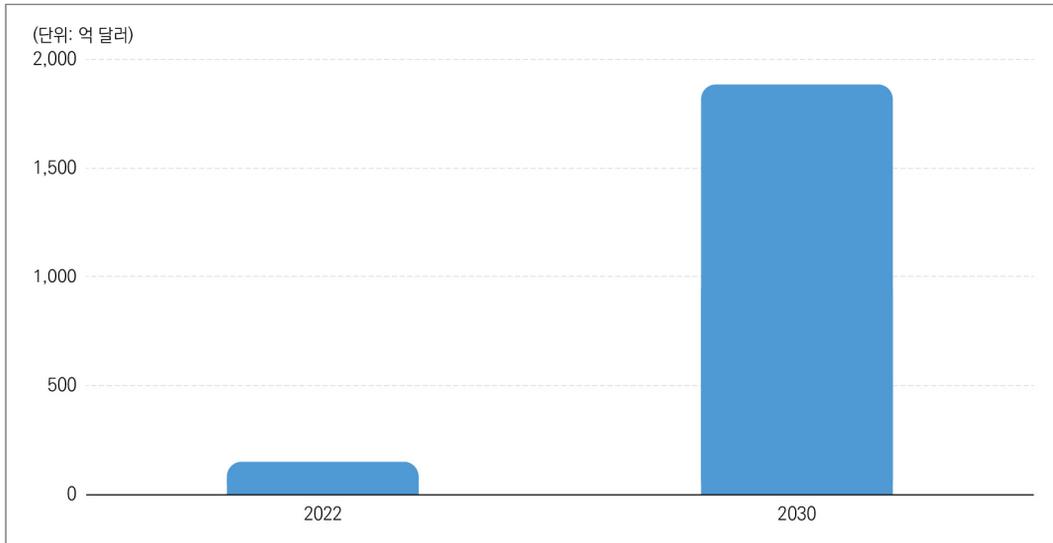
보건의료 시스템의 측면에서도 AI는 다양한 방식으로 의료의 효과성과 효율성을 제고할 수 있다. 의료 서비스 제공자는 AI를 활용해 진료 전후 활동, 보험 청구, 지불 과정을 자동화할 수 있다. 기존의 의료보험 청구 관리는 시간 소모적이고 자원 집약적인 과정으로 매년 수많은 청구를 표준화하고 검증하는 데 막대한 자원이 필요하며, 진료 혹은 전후 활동을 지연시키는 요인이 되기도 하였다. 그러나 AI의 사용은 보험 청구 담당자나 수익 분석가보다 신속하고 정확하게 청구를 처리할 수 있어, 의료기관과 환자 모두의 시간과 자원을 절약할 수 있는 가능성을 제공한다.

병원 운영의 측면에서도 AI는 예상 환자 수의 최대치와 최저치를 예측하여 병원 인력과 자원의 배치를 최적화할 수 있다. 또한, 의료기관은 AI를 사용하여 환자를 대신해 의학적 결정을 내리는 것뿐만 아니라 비즈니스 운영과 관련된 결정을 내릴 수도 있다. 의료 서비스 수요는 다양한 요인에 의해 변동성이 크기 때문에, 병원이 의료 장비나 인력을 최적으로 배치하는 과정에는 어려움이 따른다. AI는 과거 데이터와 실시간 상황 분석을 통해 미래 자원 수요를 예측하는 데 도움을 줄 수 있다. 예를 들어, AI를 사용하여 병원 재입원 증가에 대비한 최적의 재고량을 파악하거나 계절적 수요에 맞춘 새로운 기계 및 용품의 필요성을 결정할 수 있다. 또한, AI는 위험 영역을 식별해 의료기관의 제도적 실패를 예방하고 의료 결과를 최적화할 수 있다. 이를 통해 의료기관은 향후 발생 가능한 상황을 예측하고 선제적으로 대응할 수 있어 비용을 절감하고 진료 성과를 개선할 수 있다.

3 보건의료 분야 인공지능 산업의 동향과 전망

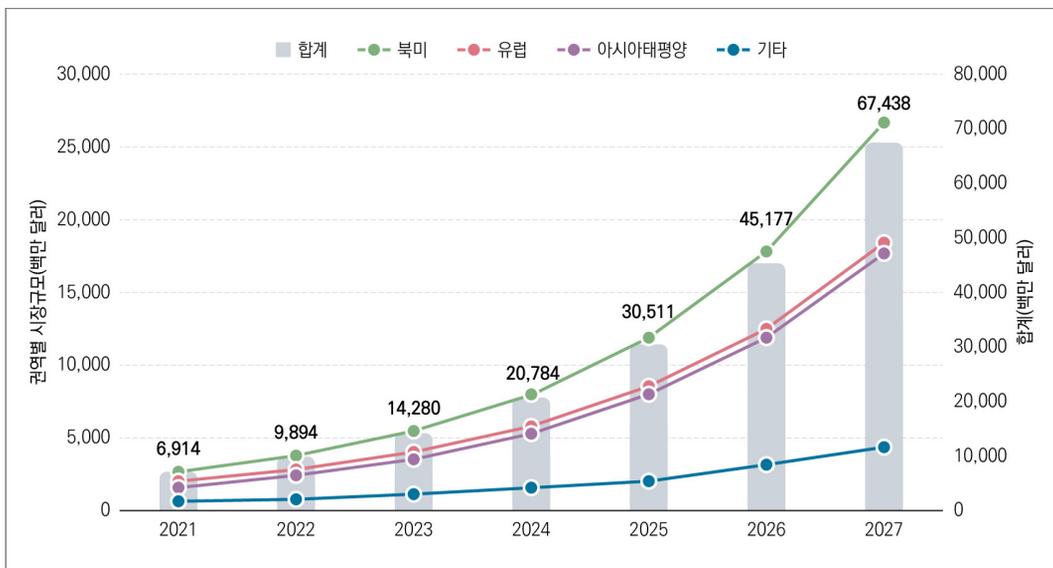
시장조사기관 프레시던스 리서치(Precedence Research)의 보고서에 따르면 전 세계 보건의료 분야 AI 시장 규모는 2022년 151억 달러로 추산되며, 이 중 북미 시장 규모는 68억 달러로 세계 시장 점유율의 59%를 차지하며 시장을 선도하고 있다. 시장 규모는 22년 이후 연평균 약 37%씩 발전하여 2030년에는 약 1,880억만 달러에 육박할 것으로 전망하고 있다(ASTI MARKET INSIGHT 2022-2065).

또 다른 시장조사기관인 마켓앤마켓(Markets and markets)에 따르면, 보건의료 분야 AI의 세계 시장 규모는 2021년 69억 달러로, 지역별로는 북미 38.5%, 유럽(27.8%), 아시아·태평양(25.4%) 순이며 남미(5.0 %)와 중동·아프리카(3.4 %)로 이어지고 있다. 2027년까지 시장 규모는 연평균 46.2%씩 성장하여 674억 달러에 이를 것으로 예상된다. 국가별로 살펴보면, 미국은 2021년 23억 500만 달러에서 2027년에는 232억500만 달러로 연평균 성장률도 47.0%로 큰 성장을 보일 것으로 전망된다. 남미에서는 멕시코가 2021년 1억4,200만 달러에서 2027년 15억4,400만 달러로, 연평균 48.8%의 성장률이 예상되며, 아시아 태평양 지역에서는 중국이 연평균 48.8%의 성장률로, 2021년 6억500만 달러에서 2027년 65억6,900만 달러에 이를 것으로 예상하고 있다.



출처: Precedence Research

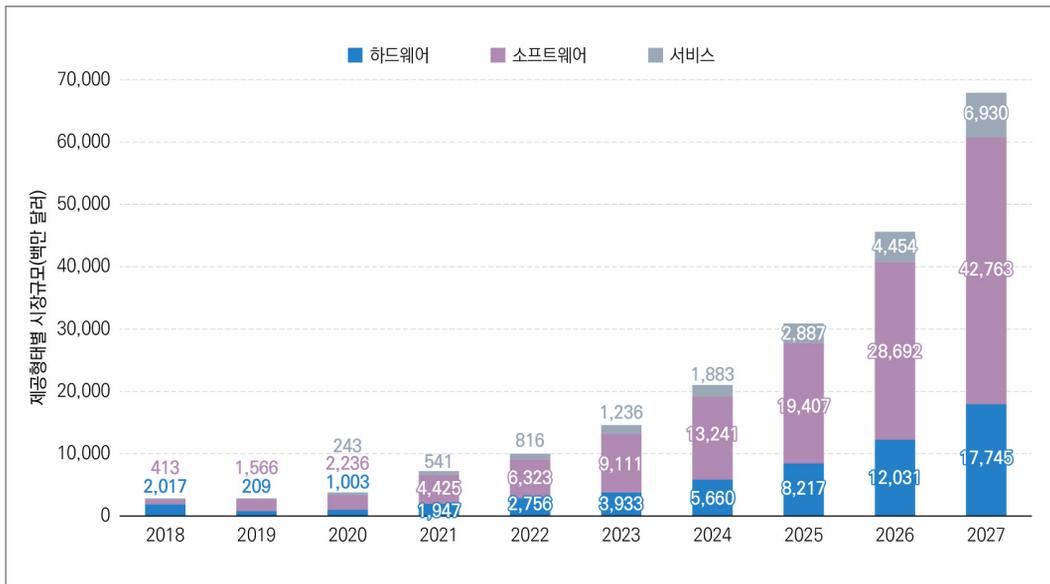
[그림 2-3] 전 세계 AI 헬스케어 시장 규모 전망



출처: MarketsAndMarkets, Artificial Intelligence in Healthcare Market, KISTI 재구성

[그림 2-4] 세계 AI 의료 및 헬스케어 시장 규모 및 전망

의료 및 헬스케어 산업에서 AI는 제공되는 형태에 따라 하드웨어 (Hardware), 소프트웨어(Software), 서비스(Services)로 시장을 나눌 수 있다. 과거에는 하드웨어가 가장 큰 시장을 차지했지만, 2019년부터는 소프트웨어 형태의 AI 헬스케어 시장이 다른 분야보다 확연히 커졌으며 앞으로도 꾸준히 성장할 것으로 예상된다.



출처: MarketsAndMarkets, Artificial Intelligence in Healthcare Market, KISTI 재구성

[그림 2-5] 제공 형태별 AI 의료 및 헬스케어 시장 규모 및 전망

제3장

국내외 보건의료 AI 관련 R/D 주요 전략

- 제1절 미국의 보건의료 분야 AI R/D 주요 전략
- 제2절 캐나다의 보건의료 AI 관련 R/D 주요 전략
- 제3절 싱가포르의 보건의료 AI 관련 R/D 주요 전략
- 제4절 일본의 보건의료 AI 관련 R/D 주요 전략
- 제5절 국내 보건의료 AI 관련 R/D 주요 전략

제 1절

미국의 보건 의료 분야 AI R/D 주요 전략

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 국가 AI R&D 전략계획

2023년 5월 23일, 미국 대통령 직속 국가과학기술위원회(NSTC)의 AI 특별위원회(Select Committee on Artificial Intelligence)는 새롭게 업데이트한 “국가 AI R&D 전략계획”(National AI R&D Strategic Plan)을 발표하였다. 국가과학기술위원회는 다원화된 과학기술정책을 연방정부 차원에서 국정 목표와 대비하여 보완·조정하는 역할을 수행하며, 국가의 여러 목표를 달성하기 위해 연방기관에서 조정된 연구개발 전략을 수립하고 있다.

「국가 AI R&D 전략계획」은 AI 기술 발전에 대응하여 AI 연구와 개발을 촉진하고, AI의 안전성과 윤리성을 강화하기 위한 연방 차원의 로드맵으로 정부 기관들이 AI 분야 투자계획을 수립할 때 반영해야 할 우선순위를 제시하고 있다. 이 계획은 2016년 처음 발표되었으며, 2019년과 2023년에 업데이트되었다.

23년 버전은 미국 과학기술정책국(OSTP)⁷⁾이 책임 있는 AI R&D 투자를 위한 기존 국가 AI R&D 전략계획('16.11, '19.6)에 관한 정보제공 요청(RFI)⁸⁾을 추진하여 대중의 의견을 수렴하고, AI 분야 최신 트렌드와 바이든 정부의 AI 정책 방향을 반영하여 전략을 보완하기 위해 작성되었다. 바이든 정부는 책임 있는 AI 혁신을 촉진하고 사회적 과제를 해결하기 위하여 효과적인 AI R&D 투자 중요성을 강조하며, 전략 계획을 통해 연방 R&D 투자를 조정하고 기관별 관련 활동을 조율하고자 하였다. 이번에 발표된 계획은 2016년과 2019년 발표된 전략 계획에 이어 8가지 기존 전략을 재확인하고 국제 협력에 대한 9번째 전략을 추가하였다.

7) 대통령실 내 과학기술 정책 총괄 기구로서 대통령에게 관련 정책 수립에 필요한 과학적·기술적 자문 제공을 위해 설립했으며, 연방 AI 활동 총괄 및 자문을 제공하며, 부처 간 활동을 지원

8) 정보제공요청(RFI, Request for Information): 정부가 정책·전략 기획단계에 전문가와 대중의 의견을 공개적으로 수렴하는 과정·제도

AI R&D 9가지 전략

- ① 책임 있는 AI 연구에 장기적으로 투자: 공익을 위한 책임 있는 혁신을 고려하여 생성형 AI를 비롯한 차세대 AI에 대한 투자 우선순위를 설정
- ② 인간과 AI의 협업을 위한 효과적인 방법을 개발: 인간의 능력을 효과적으로 보완하고 강화하는 AI 시스템을 만드는 방법에 대한 이해를 제고
- ③ AI의 윤리적, 법적, 사회적 영향을 이해하고 대처: AI가 초래하는 위험을 이해하고 위험을 완화하기 위한 접근법을 개발하여 AI 시스템이 국가의 가치를 반영하고 형평성을 증진할 수 있도록 노력
- ④ AI 시스템의 안전과 보안을 보장: 신뢰할 수 있고 믿을 수 있으며 안전한 AI 시스템 설계에 대한 지식을 발전
- ⑤ AI 학습 및 테스트를 위한 공유 공개 데이터셋과 환경을 개발: 고품질 데이터 셋과 환경, 테스트 및 훈련 리소스에 대한 액세스를 확대 및 활성화
- ⑥ 표준과 벤치마크를 통해 AI 시스템을 측정하고 평가: AI 권리장전을 위한 청사진[Blueprint for an AI Bill of Rights('22.10.)]과 AI 위험 관리 프레임워크[AI Risk Management Framework('23.1.)]에 따라 기술 표준 및 벤치마크를 포함한 광범위한 AI 평가 기법을 개발
- ⑦ 국가 AI R&D 인력 수요에 대한 이해 제고: AI 인력을 전략적으로 육성하기 위해 인력 개발 기회를 개선
- ⑧ AI 발전 가속화를 위한 공공-민간 파트너십 확대: 학계, 산업계, 국제 파트너 및 기타 비연방 기관과 협력을 통해 책임있는 AI R&D에 대한 지속적인 투자와 발전을 실용적인 역량으로 전환할 수 있는 기회를 촉진
- ⑨ AI 연구 분야에서 국제 협력에 대한 원칙적이고 조율된 접근 방식 확립: 환경적 지속가능성, 의료, 제조와 같은 글로벌 과제를 해결하기 위해 AI R&D에서 국제 협력을 우선시하여 국제 가이드라인 및 표준 개발, 이행에 도움

[표 3-1] 미국 「국가 AI R&D 전략계획」 비교 분석

전략	'16년(전략 7개)	'19년(전략 8개)	'23년(전략 9개)
〈전략1〉 장기적 관점의 투자	<ul style="list-style-type: none"> 연방기관들이 근본적이고 장기적인 기초연구에 우선 투자할 것을 권고 기초연구가 필요한 분야로 데이터 처리 활용 방법론, 범용 AI 등 차세대 AI 개발, 로봇, AI 전용 HW 개발 등을 제시 	<ul style="list-style-type: none"> 범용 AI 개발을 위한 머신러닝(ML) 관련 기초연구와 기반 요소*에 대한 투자 강조 * 기반 요소: 컴퓨팅 인프라, 데이터, 보안 등 	<ul style="list-style-type: none"> 범용 AI 개발을 위한 관련 기초연구 및 기반 요소에 대한 장기적인 투자와 더불어 지속가능한 AI(Sustainable AI)와 컴퓨팅 시스템에 대한 관심 필요
〈전략2〉 인간-AI 협업 방법 개발	<ul style="list-style-type: none"> 범용적이고 보편적인 상황에서 활용 가능한 인간-AI 협업시스템 개발 * 인간인지 지능시스템, 인터페이스 기술, 언어처리 시스템 등에 관한 R&D 필요 	<ul style="list-style-type: none"> 신뢰 기반의 인간-AI 협업을 위해 설명 가능성, 공정성, 투명성 기반 협업시스템에 관한 연구 추진 	<ul style="list-style-type: none"> 기존의 인간-AI 협업(Human-AI Interaction /Collaboration)보다 조금 더 발전된 개념인 Human-AI Teaming이라는 개념을 도입 인간-AI 협업을 위한 신뢰성에 대한 연구 추가적인 추진
〈전략3〉 AI의 윤리적·법적· 사회적 영향에 대한 이해 대응	<ul style="list-style-type: none"> AI의 윤리적·법적· 사회적 의미를 이해하고 윤리적·법적· 사회적 원칙을 준수하는 AI 설계 방법론 개발 AI 설계상의 공정성, 투명성, 책임성을 제고하고, 윤리 프레임워크 마련 	<ul style="list-style-type: none"> 윤리적 AI 관련 연구와 적용에는 다양한 분야 전문가들의 참여·협력 필요 연방 차원에서 협업의 장을 마련하고, 다학제 간 R&D가 지속할 수 있도록 노력 미국 시민의 자유, 프라이버시, 미국의 가치관을 보호하는 방향으로 AI 기술 개발 	<ul style="list-style-type: none"> AI의 윤리적·법적· 사회적 영향을 포함한 광범위한 영향에 대한 이해 필요
〈전략4〉 AI 시스템의 안전 및 보안 보장	<ul style="list-style-type: none"> AI 시스템의 설명가능성, 투명성 강화 AI 시스템에 대한 신뢰 형성을 위해 사용자 친화적 UI 개발 및 운영자 대상 시스템 기능·한계 교육 필요 AI 시스템의 검사, 검증용 위한 평가·진단·수리 방법론 개발 적대적 ML 등 AI 대상 고의적인 사이버 공격 대응 방안 마련 	<ul style="list-style-type: none"> AI 시스템 라이프사이클* 전주기에 걸쳐 AI 시스템의 안전성과 보안 고려 * AI 시스템 라이프사이클: 설계-데이터·모델 구축-검사·검증-도입·가동-모니터링 인간요소(human factor)를 고려한 새로운 안전성 평가 기준·규정 마련 	<ul style="list-style-type: none"> AI 개발 공급망의 보호와 AI 시스템의 보안을 강화하는 연구 필요

전략	'16년	'19년	'23년
〈전략5〉 AI 학습-테스트를 위한 공공 데이터셋 및 환경 구축	<ul style="list-style-type: none"> 정부 보유 데이터셋, 연방기금으로 개발 지원이 가능한 데이터셋*, 산업계 보유 데이터셋의 활용 지원 * 연방기금으로 이뤄지는 연구에서 발생하는 데이터 등 데이터 수집, 큐레이션, 분석, 시각화 등을 위한 인프라와 툴 개발·제공 AI 개발을 위한 오픈소스 라이브러리와 툴킷 개발 및 확대 	<ul style="list-style-type: none"> 세부 영역별* 특수성을 반영한 공공 데이터셋 구축 * 의료(진단, 검사), 제약(신약개발) 등 	<ul style="list-style-type: none"> 공유되는 대규모 및 전문화된 고급 컴퓨팅 및 하드웨어 리소스 개발 필요
〈전략6〉 AI 기술 성능측정 및 평가기준 개발	<ul style="list-style-type: none"> AI 기술 개발을 촉진하기 위해 성능측정을 위한 기술기준 및 벤치마크* 개발, 테스트베드 구축, AI 커뮤니티 (학계, 산업계)의 참여가 중요 * 반복적 테스트를 통해 특정 오브젝트(object)의 성능을 측정하는 것 	<ul style="list-style-type: none"> 기술적인 요소뿐만 아니라 AI 시스템의 신뢰성, 사용성, 프라이버시 등을 측정·평가할 수 있는 기준 마련 	<ul style="list-style-type: none"> AI 시스템의 감사(audit) 및 모니터링을 위한 표준 개발
〈전략7〉 AI R&D 인력의 필요성 이해	<ul style="list-style-type: none"> AI 전문인력의 수요·공급 특성을 이해하고 향후 인력 수요를 예측 	<ul style="list-style-type: none"> AI 기술의 혁신·응용에 중요한 타분야(인지 과학, 심리학, 경제학, 언어학 등) AI 전문인력 양성 및 융합 방안 연구 	<ul style="list-style-type: none"> AI 기술 분야에 필요한 전문인력 훈련/재교육 및 세계적으로 유능한 고급 인재들의 동원과 유치 강조 지역별 AI 전문성 개발 및 연방 AI 인력 강화 필요 AI 인력 교육과 훈련에 윤리적, 법적, 사회적 영향력 고려
〈전략8〉 산·학·관 R&D 파트너십 확대	-	<ul style="list-style-type: none"> 정부-대학-산업 간 파트너십 확대로 AI 분야 혁신 촉진 	<ul style="list-style-type: none"> AI 발전의 가속화를 위해 미국 내의 정부-대학-산업 간 파트너십 확대와 더불어, 다른 국가 정부, 다국적 기업 및 시민 사회 단체들과의 국제협력 강화 추진 필요 (전략9 도입)
NEW 〈전략9〉 AI R&D 국제협력 추진	<ul style="list-style-type: none"> 동맹국들의 신뢰할 수 있는 AI 개발과 사용에 대한 글로벌 문화 육성, 글로벌 AI 시스템, 표준 및 프레임워크 개발 지원 AI 분야 선도 전문가들과 혁신가들 간의 아이디어와 지식 교류 촉진 글로벌 이익을 위한 AI 개발 장려 및 AI 적용에 대한 대처를 위한 추가 연구 필요성 강조 		

자료: 김소미, 「2023년 美 「국가 AI R&D 전략계획」 주요 내용」, 『NIA The AI Report』, 한국지능정보사회진흥원, 2023.

2 미국의 보건의료 분야 AI R/D 전략

미국의 보건의료 분야에서 AI 연구 및 개발(R&D) 전략은 빠르게 발전하고 있으며, AI 기술이 의료 혁신을 주도하고 환자 치료 결과를 향상시키는 데 중요한 역할을 하고 있다. 미국의 AI R&D 전략은 특히 의료 데이터 분석, 진단, 치료 계획 수립 등에서 AI를 활용하며, 연방정부의 여러 기관과 협회가 이를 지원하고 있다.

가. 국가 AI R&D 전략 계획 내

미국의 AI R&D 전략계획은 보건의료를 AI 연구와 적용의 핵심 분야로 설정하고 있다. 이 전략을 통해 정부는 의료 AI 연구에 장기적인 투자를 장려하고 있으며 의료 데이터의 효율적 활용과 의료 시스템의 개선을 목표로 하며, AI 기반 솔루션이 질병 예측, 조기 진단, 치료 효과 증대, 환자 맞춤형 치료 제공 등에 기여할 수 있도록 하고 있다. 아울러, 신뢰할 수 있는 AI 시스템 구축을 위해 의료 데이터를 안전하게 공유하고 표준화하는 방안을 강조하고 있다. 이와 관련하여, 연구자들이 보다 질 높은 데이터를 확보하고 공유할 수 있는 공공 데이터셋 구축 및 테스트 환경을 제안하고 있으며, 이를 위해 다양한 기관 간 협업이 이루어질 수 있도록 장려하고 있다.

나. 미국 보건복지부(HHS)의 AI 전략

미국 보건복지부(HHS)는 공공 의료 서비스를 강화하기 위해 AI 기술을 도입하고 있다. 주요 전략 목표는 다음과 같다.

AI 인재 육성: 의료 종사자와 HHS 직원의 AI 기술 역량의 향상

헬스케어 혁신 및 연구 개발 촉진: 대학 및 민간 기업과 협력하여 AI 기반 헬스케어 솔루션을 개발하고 이를 확장

AI 도구 접근성 확대: 의료 부문 전반에서 AI 기술을 활용토록 지원

신뢰할 수 있는 AI 개발: AI 시스템에서 발생할 수 있는 프라이버시, 편향성, 투명성 문제 해결⁹⁾

9) Government Accountability Office, <https://www.gao.gov/products/gao-21-7sp>

국립보건원(NIH)과 미국 식품의약국(FDA) 등은 AI 기반의 의료 혁신을 촉진하기 위해 다양한 프로그램을 운영 중이다. NIH의 “All of Us” 연구 프로그램은 개인 맞춤형 의료를 위해 백만 명의 건강 데이터를 수집하고 AI를 통해 질병 예측 모델을 개선하고 있다.

FDA는 AI 및 머신러닝 기반 의료기기의 사용을 승인하는 절차를 개선하고 있으며, 이를 통해 AI가 임상 환경에서 안전하게 사용될 수 있도록 하고 있다. 데이터 및 인프라 관련하여 AI 기술이 의료 데이터 분석과 의사결정을 도울 수 있도록 고성능 컴퓨팅 인프라와 데이터 상호운용성을 보장하는 정책이 마련되었다. HHS와 ASTP/ONC는 데이터의 프라이버시 보호와 데이터 활용의 투명성을 강조하고 있으며, 다양한 의료 시스템 간의 데이터 교환이 원활하게 이루어질 수 있는 인프라를 구축하고 있다(The White House).

제 2 절

캐나다의 보건의료 AI 관련 R/D 주요 전략

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 범캐나다 인공지능 전략

(PanCanadian Artificial Intelligence Strategy)

캐나다는 2017년 세계 최초로 범캐나다 인공지능 전략(PanCanadian Artificial Intelligence Strategy)을 발표하며, 2022년까지 향후 5년간 1억 2,500만 캐나다 달러를 투자해 AI 원천기술 경쟁력 확보에 주력할 것이라고 밝혔다. 주요 내용은 AI 관련 연구와 인재 육성을 중심으로 구성되어 있다. 1996년부터 2019년까지 캐나다 내에서 2만 3,000개의 AI 연구가 발표됐으며, 2019년 한 해에만 약 2,054개의 연구가 진행되었다. 저자의 논문 수와 피인용 수를 기반으로 연구의 생산성과 과학적 영향력을 나타내는 Scopus h-index의 경우 캐나다는 239개국 중 4위를 차지하였고 AI 인재들의 이동 흐름을 보여주는 AI skills migration 지수에서도 캐나다는 2015년과 비교해 20단계 상승한 4위를 기록하였다. 이처럼 AI 관련 양질의 연구와 인재 육성이 대학과 연구기관을 중심으로 행해지며 캐나다는 AI 분야에서 독자적인 강점을 만들어내고 있다.

현재 캐나다 연방정부는 고등연구원(CIFAR, Canadian Institute for Advanced Research)을 중심으로 온타리오, 퀘벡, 앨버타 내에 각각 벡터 연구소(Vector Institute), Montreal Institute for Learning Algorithms(Mila), Alberta Machine Intelligence Institute(Amii) 등 AI 거점 연구기관을 설립해 운영하고 있다. 아울러 캐나다 정부는 우수한 AI 인력을 적극적으로 유치하고 유지하기 위해 CIFAR AI Chairs 프로그램에 8,650만 캐나다 달러 예산을 배정해 2017년부터 운영하고 있다.

컴퓨터 과학 분야의 노벨상이라고 불리는 튜링 어워드를 수상한 제프리 힌튼(Geoffrey Hinton), 요슈아 벤지오(Yoshua Bengio) 등의 세계 우수 연구 인력들이 연구에 집중할 수 있도록 1인당 최대 100만 캐나다 달러를 지원하고 있다.

2021년 기준, 해당 프로그램을 통해 109명에 달하는 우수 연구자들이 캐나다에서

다양한 AI 관련 연구를 진행하고 있다. 예컨대 앨버타 대학교(University of Alberta)에서 조교수로 재직 중인 알로나 피시(Alona Fyshe)는 머신러닝, 신경과학을 통해 컴퓨터가 인간처럼 사물을 인식하고 개념을 학습할 수 있는 연구를 활발히 진행 중이다.

[표 3-2] 2021년 캐나다 인공지능 전략 예산안

책정 예산(C\$)	내용
18,500만	향후 5년간 캐나다 내 인공지능 혁신과 연구 상용화
16,220만	향후 10년간 앨버타, 브리티시컬럼비아, 온타리오, 퀘벡 주 내에 세계 최고 수준의 학술 인재 유치 및 유지
4,800만	향후 5년간 캐나다 고등연구원의 연구, 교육 프로그램 개선 및 강화
4,000만	향후 5년간 토론토, 에드먼턴, 몬트리올에서 활동하는 연구자 혹은 국가 인공지능 기관을 위한 컴퓨팅 용량 제공
860만	인공지능 기술 표준 개발

원자료: Budget 2021, A Recovery Plan for Jobs, Growth, and Resilience

출처: 김예지(토론토무역관), 대한무역투자진흥공사(kotra) 해외시장뉴스, 2021-09-09

국가 주도의 연구자 중심 R&D 지원에 따라 우수 연구인력이 캐나다에 모여들며 글로벌기업인 마이크로소프트, 구글, 페이스북, 톰슨 로이터 등 45개 이상의 대기업들이 캐나다 내에 AI 연구소를 설립하는 성과를 이루었다. 이러한 정부 주도의 노력 하에 2017년부터 2019년까지 캐나다 내 ICT 관련 외국인 직접 투자는 50% 상승했으며 2019년 AI 스타트업 투자 규모는 2017년 대비 2.3배 상승한 6억 캐나다 달러를 기록하였다.

2 캐나다의 보건의료 분야 AI R/D 전략

캐나다의 보건의료 분야 AI 연구 및 개발(R&D) 전략 또한 Pan-Canadian Artificial Intelligence Strategy를 중심으로 진행되고 있다.

캐나다는 캐나다 고등연구원과 협력하여 AI의 의료 활용을 포함한 다양한 연구 프로그램을 지원하고 있다. Amii, Mila, Vector Institute들은 보건의료 AI 분야에서 공공 보건 시스템과 협력하여 초기 AI 기술을 시험하고 있다. Amii는 캐나다의 대형 의료

서비스 제공자인 알버타 보건청(Alberta Health Services, AHS)과 함께 협력하여 병원 환경에서 AI 기반 솔루션을 적용하고 있다. Amii와 AHS의 협력 프로젝트는 주로 의료 데이터 분석을 통해 질병의 조기 발견과 진단의 정확도를 높이려는 목표하에 실제 임상 환경에서 이를 검증하기 위해 방대한 환자 데이터를 활용하고 있다. Mila는 암 진단 모델을 개발하여 암 조직을 조기에 식별하는 데 중점을 두고 있으며, 병원과 연구 센터가 이를 시험하고 피드백을 제공하여 AI 시스템을 개선할 수 있도록 지원하고 있다. Vector Institute 역시 의료기관들과 협력하여 의료 영상 분석을 통한 질병 예측 모델을 개발하여 질병 징후를 발견하거나 환자 맞춤형 의료 서비스를 가능하게 하는 중요한 역할을 하고 있다.

캐나다 정부는 의료 AI 혁신에 필요한 자금을 지원하는 등의 여러 가지 혜택을 제공하고 있다. 과학연구 및 실험 개발 세액 공제 프로그램과 더불어 전략적 혁신 기금(Strategic Innovation Fund)과 같은 재정 지원 프로그램을 통해 혁신적 의료 기술을 상용화하기 위해 캐나다 내에서 개발된 기술을 보다 유연한 규제 구조하에서 테스트할 수 있도록 장려하고 있다. 또한, 글로벌 시장에서 AI 제품을 출시하기 전 단계에서 초기 테스트를 보다 자유롭게 할 수 있도록 하는 점에서 OECD 국가 중에서도 규제 환경이 더 우호적이다.

제3절

싱가포르의 보건의료 AI 관련 R/D 주요 전략

NATIONAL ASSEMBLY FUTURES INSTITUTE

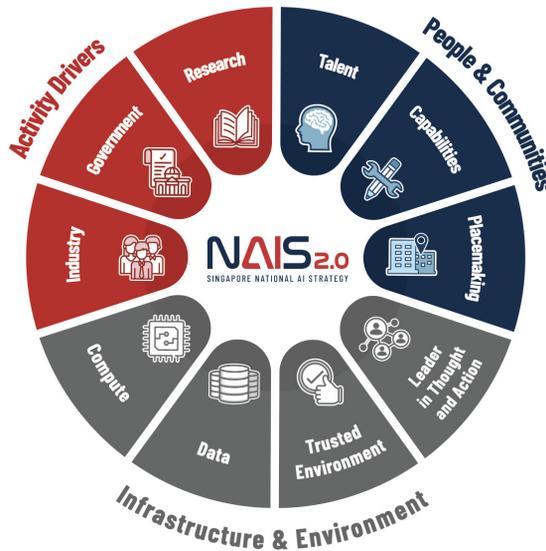
1 싱가포르의 국가 AI 전략(National AI Strategy)

싱가포르는 “스마트네이션(Smart Nation)” 이니셔티브를 통해 AI를 적극적으로 활용하고 있으며, 기술 기반 솔루션 개발과 사회적 문제 해결에 중점을 두고 있다. 이 이니셔티브는 리셴룽 총리가 2014년 발표한 것으로, 정보통신 기술과 빅데이터를 활용하여 국민의 삶의 질을 향상시키고자 하는 목표를 가지고 있다. 싱가포르는 도시 국가로서 제한된 자원을 효율적으로 활용해야 하는 필요성이 있었고, 이를 위해 첨단 기술을 국가 발전 전략의 핵심으로 삼았다. 보건 분야는 스마트네이션의 중점 분야 중의 하나로, 성공적인 고령화 지원을 위해 고령자들을 위한 헬스케어 기술 개발과 스마트 홈 솔루션을 도입하여 고령화 사회에 대비하는 것을 목표로 한다.

싱가포르는 국가 AI 전략을 처음 수립한 국가 중 하나로, 2019년에 연구, 혁신, 기업(RIE) 2020 및 2025 계획을 통해 약 3억 7,300만 달러를 투자하겠다고 밝힌 바 있다. 싱가포르는 이어 2023년 말 국가 AI 전략(NAIS) 2.0이라고 불리는 새 AI 전략을 발표하였다. NAIS 2.0의 발표 배경에는 최근 급속도로 발전한 AI 기술과 생성형 AI의 등장이 있으며, 빠른 변화에 대응하기 위해 기존 정책을 재검토하고 나선 것으로 볼 수 있다. 이 정책은 3가지 주요 변화를 골자로 하고 있다.

첫 번째 주요 변화는 싱가포르가 AI를 ‘있으면 좋은 것’이 아닌 ‘필수’라고 인식한다는 점이다. 아울러 싱가포르는 우수성을 목표로 하여 사람과 기업이 자신 있게 기술을 사용하도록 권한을 부여해 ‘가치 창출 극대화’를 추구하고, AI의 우수성을 높이기를 기대하고 있다. 이러한 맥락에서 NAIS 2.0 발표와 함께 AI 전문가를 교육하고 채용해 인재풀을 3배 수준인 15,000명으로 늘리는 것을 목표로 한다고 밝힌 바 있다. AI 기술 발전과 함께 모든 분야에서 AI 전문가와 데이터 과학자의 수요가 높아지고 있기 때문에 싱가포르는 인력 양성과 해외 인력의 채용을 통해 우수 인재 확보의 우위를 점할 계획이다.

두 번째 주요 변화는 싱가포르가 글로벌 네트워크와 밀접하게 협력하여 AI 혁신과 거버넌스에 적극 기여하고자 AI 혁신을 목표로 한다는 점이다. 세 번째는 개별 프로젝트 방식에서 ‘시스템’ 접근 방식으로서의 변화(projects-to-systems)이다. 새 정책에서는 “시스템 접근 방식을 채택해 싱가포르 내외의 이해관계자들을 한 데 모아 융합적 활동을 꾀하고 있으며, 활동 동인(산업, 정부, 연구), 사람 및 커뮤니티(인재, 역량, 장소 조성), 인프라 및 환경(컴퓨팅, 데이터, 신뢰할 수 있는 환경, 사고와 행동의 리더) 등 3가지 범주와 이를 가능하게 하는 10개 조력자 역할을 규정하고 아이디어 교환을 가속화하며 AI 지원 솔루션을 대규모로 관리할 예정이다”라고 명시하고 있다.



[그림 3-1] NAIS 2.0의 세 가지 시스템 및 10개 조력자

이를 실현하기 위해 NAIS 2.0은 세 가지 주요 시스템을 중심으로 전개된다:

1. 활동 주도자 시스템: AI가 생명과학, 제조, 운송과 같은 주요 경제 부문에 도입되며, 국가적 우선순위인 의료와 공공 안전을 지원한다.
2. 사람과 공동체 시스템: AI 전문가 15,000명을 양성하고, AI 인턴십 프로그램 (AIAP)을 확장하며, 글로벌 AI 인재를 환영함으로써 AI 인재 풀을 확대한다.

3. 인프라 및 환경 시스템: 고성능 컴퓨팅과 데이터 인프라에 투자하고 AI Verify 툴킷과 같은 거버넌스 프레임워크를 통해 책임 있는 AI 사용을 장려한다.

싱가포르는 AI 거버넌스 측면에서도 선도적인 역할을 하고 있다. 싱가포르는 세계 최초로 모델 AI 거버넌스 프레임워크를 도입하여 책임 있는 AI 사용을 촉진하고 있으며, AI 기술의 윤리적 사용과 투명성을 강조해왔다. 정부는 기업들이 신뢰할 수 있는 AI 솔루션을 구현하도록 지침과 지원을 제공하고 있으며, 이러한 프레임워크를 통해 AI 시스템의 공정성, 투명성, 안전성을 확보하고자 한다. 예를 들어, 알고리즘의 편향성을 최소화하고 개인정보 보호를 강화하는 등의 규정을 마련하여 AI 기술이 사회에 미치는 부정적인 영향을 줄이기 위한 노력을 하고 있으며, 기업들은 이러한 지침을 준수함으로써 글로벌 시장에서 경쟁력을 갖추고, 국제적인 신뢰를 얻을 수 있게 되었다.

싱가포르는 모델 AI 거버넌스 프레임워크와 AI Verify 재단을 통해 AI 거버넌스의 국제적 기준을 수립하여 투명하고 신뢰할 수 있는 AI 시스템을 개발하고, 사용자들의 신뢰를 강화하며 혁신을 촉진하고 있다.

또한, 싱가포르는 공공과 민간 부문 간의 협력을 통해 AI 기술 개발과 상용화를 촉진하고 있다. 대표적인 예로, AI Trailblazers 프로그램을 들 수 있다. 이 프로그램은 구글 클라우드와의 협력을 통해 기업과 정부 기관이 무료로 AI 도구에 접근할 수 있도록 지원한다. 이를 통해 다양한 생성형 AI 사용 사례를 개발하여 혁신을 가속화하고, 기업들이 최신 기술을 활용하여 새로운 비즈니스 모델을 창출할 수 있도록 돕고 있다. 이 프로그램은 기업들이 기술적 장벽 없이 AI를 도입하고 활용할 수 있게 함으로써 산업 전반의 디지털 전환을 촉진하고 있다.

2 싱가포르의 보건의료 분야 AI R/D 전략

싱가포르의 보건의료 분야에서 AI R/D 전략은 첨단 기술을 활용하여 의료 서비스의 효율성을 높이고 질병 예측 및 예방, 진단, 치료를 개선하는 것을 목표로 하고 있으며, 주요 전략은 다음과 같다.

가. 의료 데이터 인프라 강화

싱가포르 보건 당국(Singapore Health Services, SingHealth)은 국가 AI 계획에 따라 대규모 의료 데이터 분석을 통해 환자 맞춤형 의료와 예측 분석을 위한 기초 데이터를 구축하고 있다. 이러한 데이터의 기반은 국가 디지털 신원(National Digital Identity) 체계에 기인하고 있다. 기존의 SingPass 시스템¹⁰⁾을 기반으로 신뢰할 수 있는 디지털 신원 프레임워크를 개발한 것으로, 안전하고 편리한 전자 정부 서비스를 제공한다. 시민들은 이를 통해 세금 신고, 의료 기록 조회, 금융 거래 등 다양한 서비스를 온라인으로 손쉽게 이용할 수 있게 되었으며, 행정 업무의 효율성도 크게 향상되었다. 디지털 신원 체계의 구축은 전자 정부 구현의 핵심 요소로서, 디지털 경제 활성화에 중요한 역할을 하고 있다. 싱가포르는 이러한 신원체계 기반하에 보건 당국이 운영하는 헬스허브(HealthHub) 플랫폼을 사용하여 각 의료기관의 전자건강기록(EHR) 시스템을 통한 환자 데이터를 통합 관리하고 AI 알고리즘을 활용해 환자의 의료 기록, 유전자 정보, 생활 습관 등 다양한 데이터를 분석하고, 이를 바탕으로 개인화된 치료 계획을 세우는 등에 활용될 수 있도록 하고 있다.

나. 질병 예측 및 조기 진단

AI는 의료 기록과 진단 데이터를 분석하여 환자의 건강 상태를 예측하고, 질병이 발생하기 전에 조기 경고를 제공하는 데 사용된다. 이를 통해 만성 질환이나 심혈관계 질환, 당뇨병 등의 질환을 조기 발견하는 시스템을 개발하고 있다. 예를 들어, 암 환자에게 개별화된 치료법을 제공하기 위해 유전자 분석과 AI 기반 예측 모델 및 이미징 분석 등을 통해 조기 진단하는 연구가 활발히 이루어지고 있다.

다. 의료 자동화 및 효율성 증대

싱가포르는 AI를 통해 병원의 운영을 효율화하고 의료 서비스 제공 속도를 높이기 위한 기술을 개발하고 있다. AI 기반 챗봇이나 가상 간호사가 환자의 질의를 실시간으로 처리하고, 진료 예약, 건강 모니터링 등의 기능을 자동화하는 시스템이 연구되고 있다.

10) 싱가포르 국민과 거주자들이 정부 및 민간 서비스에 안전하게 접근할 수 있도록 해주는 디지털 신원 인증 시스템임

아울러 병원에서도 로봇틱스와 AI 기반 분석을 결합하여 수술의 정확성을 높이고, 최소 침습 수술 기술을 개발하고 있다.

라. AI 기술 상용화 및 혁신 촉진

싱가포르는 스타트업 및 혁신 허브를 통해 AI 기반 의료 기술의 상용화를 적극적으로 지원하고 있다. 정부 차원에서 AI 연구 및 개발에 대한 투자와 함께, AI 솔루션을 개발하는 스타트업과 협력하여 실질적인 의료현장에 적용할 수 있는 기술을 도입하고 있다.

아울러 싱가포르는 스마트네이션 센서 플랫폼(SNSP)을 통해 도시 전반에 사물인터넷(IoT) 센서를 설치하여 데이터 수집 및 분석을 진행하고 있다. 이러한 스마트시티의 AI를 통해 도시 생활과 실시간 건강 데이터를 통합하고 고품질의 의료 서비스를 제공하고 자 한다. 특히 스마트시티 사업은 취약계층과 장애인을 위한 맞춤형 의료 서비스에 중점을 두고 있으며 헬스케어 프레임워크는 ① 개인화된 건강 관리 ② 실시간 스마트 커뮤니티 서비스 ③ 커뮤니티 헬스 키퍼 서비스로 구성되어 있다. 이 서비스들은 예방 의료, 조기 진단, 전문가에 의한 즉각적인 건강 경보를 포함하며, 주로 고위험군과 장애인을 대상으로 한다.

제4절

일본의 보건의료 AI 관련 R/D 주요 전략

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 일본의 AI 국가 전략

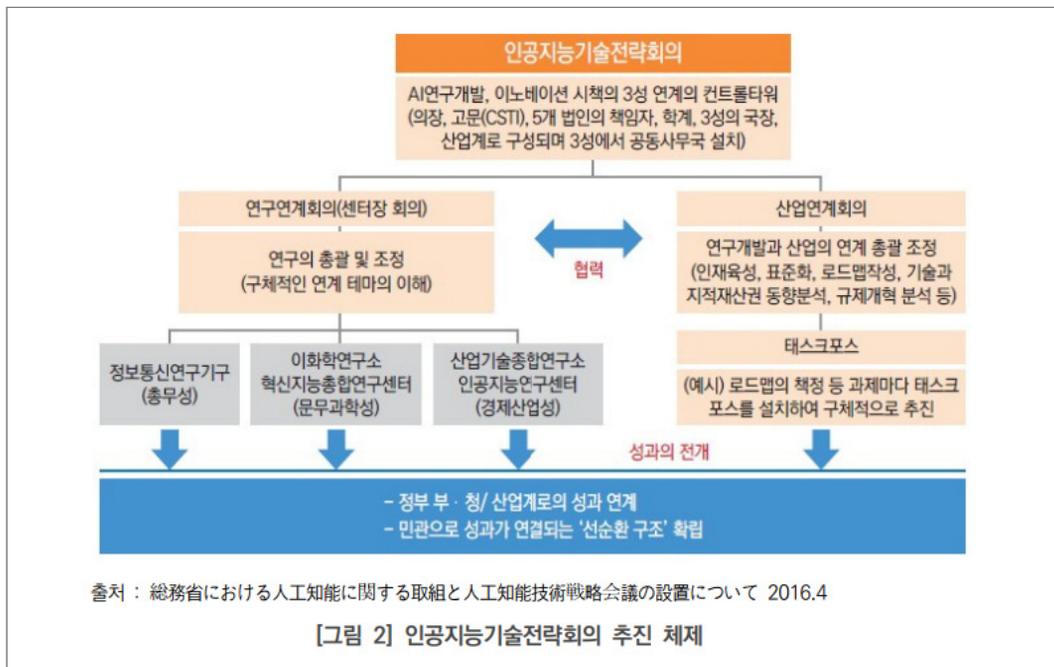
일본 정부의 국가적 비전이 'Society 5.0 구현'으로 정립되면서 국가혁신전략도 '로봇 기술을 활용한 생산성 혁명'에서 'AI를 통한 사회문제 해결 및 기술진보'로 구체화되었다. 일본 재흥전략의 목표는 '세계 최고 수준의 IT 사회 실현'(13, '14년), '생산성 혁명'(15년), '신산업 강화 및 생산성 제고'(16년) 등이었으며, AI는 빅데이터, IoT와 함께 산업구조 변화를 위한 주요 시책 중의 하나로 제시되었다. 일본 정부는 '세계 최고 수준의 IT 사회 실현'을 위해 '로봇 신전략'(15)을 발표하였으나, 4차 산업혁명 논의가 급부상하자 로봇 기술의 범주를 초월하는 새로운 정책 대안으로 AI를 보다 강조하기 시작하였다.

가. 인공지능 연구개발 목표와 산업화 로드맵

일본은 경제 부진 탈출과 초스마트사회 실현을 위해 AI 기술을 시급하게 개발해야 한다고 표명하고 2016년 AI 기술정책의 컨트롤타워를 설치한 바 있다. '미래 투자를 위한 민관 대화'를 통해 4차 산업혁명을 주도할 수 있는 수준의 혁신을 달성하기 위해서는 국가 차원의 대규모 프로젝트가 필요하다고 지적하고 핵심 분야로 AI를 제시하였다. 이에 총무성, 문부과학성, 경제산업성이 연계하여 '인공지능기술전략회의'를 설치하고 AI 실행전략인 '인공지능 연구개발 목표와 산업화 로드맵'을 2017년 발표하였다. 로드맵에서는 ① 생산성, ② 건강, 의료, 간호, ③ 공간 이동, ④ 정보 보안 등 산업화 중점 4대 분야를 선정하고, AI 기술 발전을 3단계로 구분(매우 중요한 사회 문제 해결 필요성, 경제 파급 효과에 대한 기여, AI 기술에 의한 기여도)하였다.

나. AI 기술전략

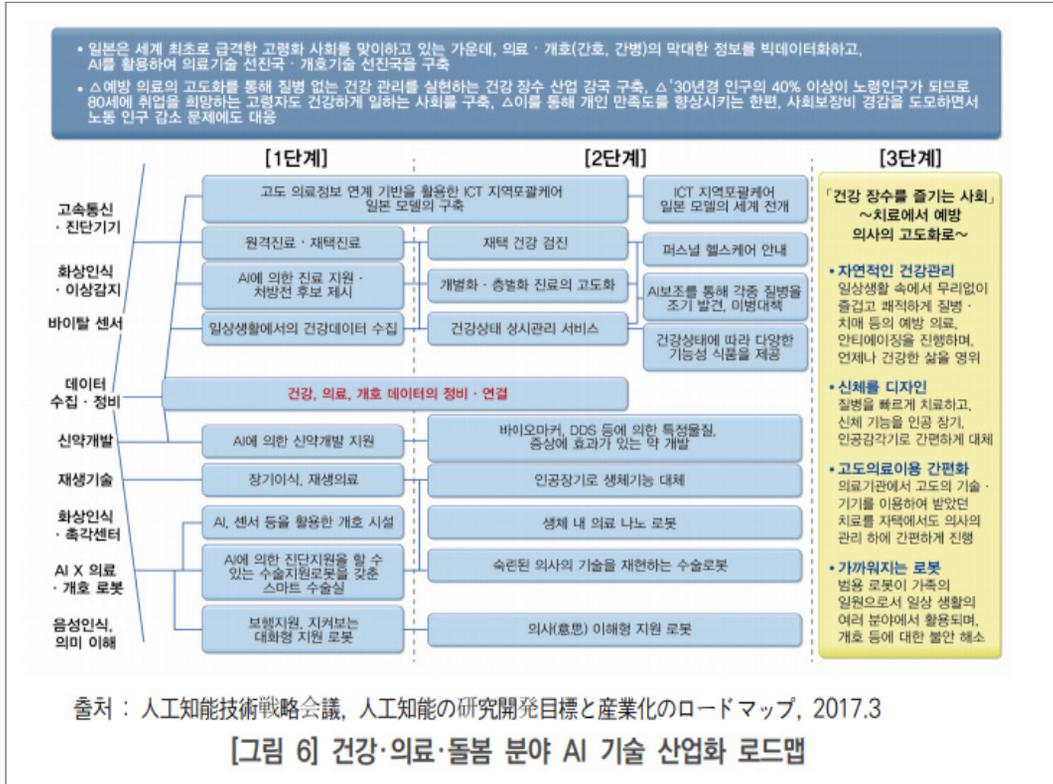
2016년 4월 ‘미래 투자를 위한 민간기업과 정부의 대화’에서의 총리 지시로 ‘AI 기술전략 회의’가 설치되었다. 이 회의가 사령탑이 되어, 총무성, 문부과학성, 경제산업성이 소관하는 5개의 국립 연구개발 법인을 묶어, AI 기술의 연구개발을 진행하는 것과 동시에, AI를 이용하는 산업의 관계 부서의 프로젝트와 제휴. AI 기술의 연구개발에 대해 민간 투자를 촉진하였다.



출처: 일본의 인공지능(AI) 정책동향 _ KIAT산업기술정책 브리프

[그림 3-2] 일본의 AI 기술전략 회의

일본은 중점 분야 「생산성」, 「건강, 의료·간병」, 「공간의 이동」에 있어서의 산업화 로드맵을 정하였다. 건강·의료 분야에서는 의료정보의 빅데이터화 및 AI 도입을 위한 3 단계 로드맵이 제시되었다. 일본은 고령화 사회에 대비하여 의료·돌봄 분야의 방대한 데이터를 빅데이터로 체계화하고, AI를 활용하여 의료·돌봄 분야에서 글로벌 선도국으로의 도약을 계획하였다.



출처: 일본의 인공지능(AI) 정책동향 _ KIAT산업기술정책 브리프

[그림 3-3] 일본의 건강·의료·돌봄 분야 AI 기술 산업화 로드맵

다. AI 전략 2022

‘AI 전략 2022’은 「인간 존중», 「다양성», 「지속 가능」의 이념 아래 Society 5.0을 실현해, SDGs에 공헌이라는 3개의 이념의 구현을 위한 5개의 전략(인재, 산업경쟁력, 기술체계, 국제에 가세해 임박한 위기에 대처)을 설정하고 있다. 특히, 사회적 상황에 충실한 새로운 목표를 설정하고 추진함과 동시에, 팬데믹이나 대규모 재해 등 현재에 임박한 위기에 대처하기 위한 방법을 구체화하고 있다. 또한 경제안전보장 관점의 대응에 따라 정부 정책의 효과적인 중점화를 도모하기 위한 관계 대책의 조정이나, 양자나 바이오 등의 전략적 대처와의 시너지를 추구 등을 함께 제시하고 있다.

2 일본의 보건의료 분야 AI R/D 전략

가. 보건의료 분야의 AI 활용 추진 간담회

2017년 개최된 보건의료 분야의 AI 활용 추진 간담회에서는 AI의 특성을 활용하여 환자·국민에게 가져오는 효과를 설명하고, 보건의료 등에 있어서 AI의 도입이 전망되는 영역을 살펴보고 개발 추진을 위해 필요한 대응 및 AI를 이용한 서비스 등의 질·안전성 확보를 위해 필요한 사항 등을 검토하였다.

간담회에서는 ① 일본에서의 의료기술의 강점 발휘 및 ② 일본의 보건의료 분야의 과제 해결(의료 정보의 증대, 의사의 편재 등)을 기준으로 하여 AI 개발을 진행해야 할 중점 6영역을 선정하고 연구개발의 가속화를 추진하기로 하였다.

① 게놈 의료

- 암 연구 센터에 암 게놈 정보 관리 센터를 정비하고, 게놈 정보를 집약
- 게놈 정보 관리 센터가 임상 정보나 유전자 해석 정보 등을 수평연계 (cross functional team)에서 해석하는 지식 데이터베이스를 구축

② 화상(画像) 진단 지원

- 관련 의학회(일본 병리 학회, 일본 소화기 내시경 학회, 일본 의학 방사선 학회, 일본 안과학회)와 제휴하여 화상 데이터베이스를 구축하고 후생 노동성이 의사법 및 의약품 의료기기법의 취급을 명확화

③ 진단·치료 지원(검사·질병 관리·질병 예방 포함)

- 일본의료연구개발기구(AMED) 연구비를 통해 난치병 영역을 폭넓게 수용하는 데이터베이스를 구축
- 후생 노동성의 의사 법 및 의약품 의료기기법 명확화

④ 의약품개발

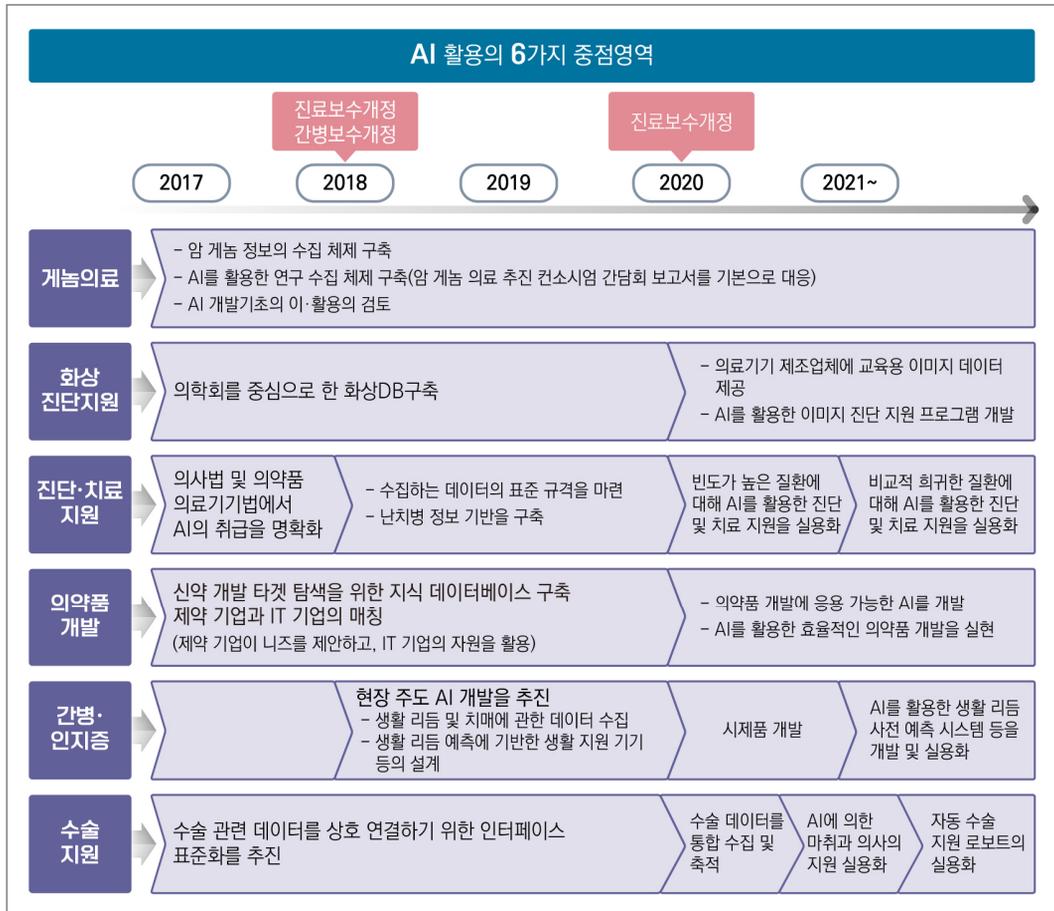
- 국립 연구개발 법인 '의약 기반·건강·영양 연구소'가 중심이 되어 신약개발 목표의 탐색을 위한 지식 데이터베이스를 구축하고 '이화학 연구소' 및 교토 대학 등과 제약 기업과 IT 기업의 매칭을 지원

⑤ 간병&인지증

- 후생노동과학연구비 보조금으로 간병 시 조기 발견 및 중증화 예방을 위한 데이터 수집 및 예측 도구 개발

⑥ 수술 지원

- 후생 노동 과학 연구비 보조금 등으로 수술 관련 데이터를 상호 연결하기 위한 인터페이스 표준화



출처: 후생노동성

[그림 3-4] 보건의료분야 AI의 활용의 6가지 중점영역

제5절

국내 보건 의료 AI 관련 R/D 주요 전략

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 국내 AI R/D 전략

가. 인공지능 국가전략

한국 정부는 2019년 12월 17일 대통령 주재로 열린 제53회 국무회의에서 과학기술 정보통신부를 비롯한 전 부처가 참여하여 마련한 「인공지능(AI) 국가전략」을 발표하였다. 본 전략에서는 대한민국의 AI 기술을 글로벌 수준으로 끌어올리고, 경제와 사회에 AI가 기여할 수 있는 시스템을 구축하기 위한 중장기 로드맵을 제시하였다. 이 전략은 크게 3가지 비전과 9개의 구체적인 전략 목표를 중심으로 구성되어 있다.

‘IT 강국을 넘어 AI 강국으로’를 비전으로, 2030년까지 ① 디지털 경쟁력 세계 3위, ② AI를 통한 지능화 경제효과 최대 455조원 창출, ③ 삶의 질 세계 10위 달성을 위해, 3대 분야의 9대 전략과 100대 실행 과제를 마련하고자 하였다.



[그림 3-5] 인공지능(AI) 국가전략의 비전 및 목표

9대 전략과제는 다음과 같다.

① AI 인프라 확충: 공공데이터 전면 개방, 공공-민간 데이터 지도의 연계, AI 허브의 컴퓨팅자원 맞춤형 지원 등 데이터·컴퓨팅자원 등 AI 산업의 핵심인프라 확충

② AI 기술경쟁력 확보: AI 반도체 핵심기술(설계·미래소자·장비 및 공정 등) 및 신개념 AI 반도체(PIM) 개발, 창의적·도전적 차세대 AI 연구개발 선제 투자, 지식표현 및 추론, 기계학습 알고리즘, 인지과학 등 AI 기초연구 강화 등 글로벌 선도국과 대등한 기술산업 경쟁력 확보

③ 과감한 규제혁신 및 법제도 정비: AI 분야 '포괄적 네거티브 규제 로드맵' 수립, AI 시대 기본이념과 원칙, 역기능 방지 시책 등 기본법제 마련 등 상상하고, 도전할 수 있는 제도적 환경 조성

④ 글로벌을 지향하는 AI 스타트업 육성: AI 투자펀드 조성, 미래기술육성자금 지원 및 AI 전문가와 스타트업의 교류·협력 활성화 등으로 혁신 생태계 구축

⑤ 세계 최고의 AI 인재 양성 및 전 국민 AI 교육: AI 관련학과 산증설 및 교수의 기업 겸직 허용, AI 대학원 프로그램 확대·다양화, 모든 군 장병 및 공무원 임용자 대상 AI 소양교육 필수화 등 세계 최고의 AI 인재 양성 및 전 국민 AI 교육체계 구축

⑥ 산업 전반의 AI 활용 전면화: 공공영역 보유 대규모 데이터 기반 대형 AI 융합 프로젝트 확대, AI 기반 스마트공장 보급, 바이오·의료(AI 신약개발 플랫폼·AI 의료기기 검증·심사), 도시(스마트시티 데이터 허브), 농업(스마트팜) 등 산업 전 분야로 AI 활용 확산

⑦ 최고의 디지털 정부 구현 : 주요 전자정부 시스템 진단('20) 및 디지털 전환 로드맵 수립, 국민 체감도가 높은 공공 서비스부터 AI 선도적 도입 등 차세대 지능형 정부 구현

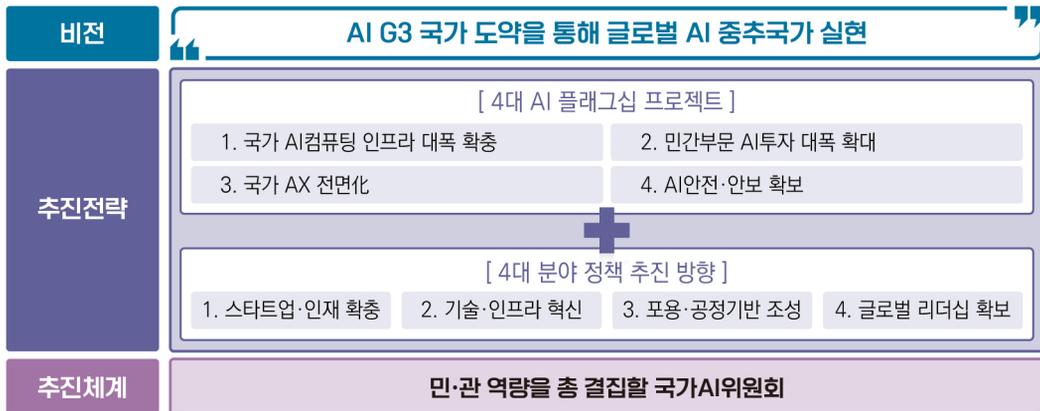
⑧ 포용적 일자리 안전망 구축 : 고용형태 다변화(특수형태근로종사자 증가 등)에 대응한 사회보험 확대, 국민취업제도 도입, 신기술 분야 직업훈련 비중 확대 등 모두가 AI 혜택을 고루 누릴 수 있도록 일자리 변화에 선제적 대응

⑨ 역기능 방지 및 AI 윤리체계 마련 : AI 기반 사이버침해 대응체계 고도화, 딥페이크 등 신유형의 역기능 대응을 위한 범부처 협업체계 구축, AI 신뢰성·안전성 등을 검

증하는 품질관리체계 구축, OECD 등 글로벌 규범에 부합하는 AI 윤리기준 확립 및 AI 윤리교육 커리큘럼 개발·보급 등 역기능 대응 및 AI 윤리 정립 등으로 안전한 AI 이용 환경 조성

나. 국가 AI 전략 정책 방향

정부가 최근에 발표한 새로운 AI 국가 전략은 한국을 세계 3대 AI 강국으로 도약시키려는 목표를 담고 있으며, 이를 통해 국가의 경제적 발전과 기술적 우위를 확보하고자 한다. 한국은 미국, 중국과 함께 세계 3대 AI 강국으로 자리매김하는 것을 목표로 삼고 있다.



[그림 3-6] 국가 AI 전략 정책 방향의 비전과 전략, 추진 체계

2030년까지 정부는 산업 부문의 70%, 공공 부문의 95%에 AI를 도입하겠다는 계획을 수립하여, 전 산업 분야에서 AI 기술이 적극적으로 활용될 수 있도록 추진하고 있다. 이는 제조업, 금융, 의료, 교육 등 다양한 분야에서 AI 기술의 통합을 의미하며, 산업 혁신과 생산성 향상을 도모하고자 한다. 또한, 여기에는 안전하고 신뢰할 수 있는 AI 생태계를 구축하고 국제 거버넌스를 주도함으로써 글로벌 AI 리더십을 확보하려는 전략도 포함되어 있다. 이를 통해 국제 사회에서 한국의 기술적 영향력을 강화하고, 글로벌 표준 설정에도 기여하고자 함이다.

이러한 목표를 실현하기 위해 정부는 네 가지 주요 플래그십 프로젝트를 추진하고 있다.

① 국가 AI 컴퓨팅 인프라 대폭 확충 : 최신 GPU 인프라를 15배로 확대하고, 국산 AI 반도체의 상용화를 추진하고 있으며, 대규모 AI 컴퓨팅 센터를 설립하여 국내 연구자들과 기업들이 고성능 컴퓨팅 자원을 효율적으로 활용할 수 있도록 지원한다. 이를 통해 국내 AI 기술 개발의 속도를 높이고, 글로벌 경쟁력을 강화하는 데 중요한 역할을 할 것으로 기대된다.

② 민간 부문 AI 투자 대폭 확대 : 민간 투자를 유도하고 정책 금융 지원을 확대하여 민간 AI 투자를 활성화하고 있다. 정부는 세제 혜택, 투자 펀드 조성 등 다양한 방법을 통해 기업들이 AI 분야에 적극적으로 투자할 수 있는 환경을 조성하고 있으며, 이를 통해 혁신적인 기술 개발과 신사업 창출을 촉진하고자 한다.

③ 국가 AX(AI+X) 전면화 : AI를 다양한 산업 분야에 통합하고, 고용 변화와 공교육 혁신을 위한 AI 활용 방안을 추진하고 있다. 이는 AI 기술이 단순한 기술 혁신을 넘어 사회 전반의 변화를 이끌어 낼 수 있도록 하는 전략으로, 교육 분야에서는 AI를 활용한 맞춤형 교육과정 개발, 고용 분야에서는 자동화로 인한 일자리 변화에 대응하는 재교육 프로그램 등이 포함된다. 이를 통해 기술 발전에 따른 사회적 충격을 최소화하고, 국민들의 삶의 질을 향상시키고자 한다.

④ AI 안전·안보·글로벌 리더십 확보 : AI 안전 연구소를 설립하고 AI 관련 법 제정과 글로벌 규범을 주도하여 AI의 안전 및 보안을 강화하고 있다. 이는 AI 기술이 가져올 수 있는 윤리적, 사회적 문제에 대응하고, 국민의 신뢰를 확보하기 위한 조치로서 중요하게 여겨지고 있다. AI 기술의 발전이 개인정보 침해, 알고리즘 편향 등 부작용을 초래하지 않도록 체계적인 관리와 규제가 필요하다는 인식 하에 이러한 노력을 기울이고 있다.

4대 분야 정책 추진 방향은 다음과 같다.

① 국가 AI혁신을 뒷받침할 기초체력 확충을 위한 AI 스타트업·인재 육성 : 스타트업과 인재 육성을 통해 AI 인재를 양성하고 AI 유니콘 기업을 육성하는 것이 포함된다. 정부는 대학과 연구기관과의 협력을 통해 AI 전문 인력을 양성하고, 창업 지원 프로그램과 액셀러레이터 등을 통해 혁신적인 아이디어가 실제 사업화로 이어질 수 있도록 지

원하고 있다.

② AI핵심·원천기술 확충 및 AI 인프라 혁신 추진 : AI 반도체 및 온디바이스 AI 기술 개발을 추진하여 기술 및 인프라 혁신을 도모하고 있으며, 국내 기술의 자립도를 높이고 글로벌 시장에서의 경쟁력을 강화하고자 한다.

③ 지속 가능한 AI발전·확산 기반 조성 : 디지털 포용법을 제정하고 프라이버시 보호를 강화함으로써 포용적이고 공정한 AI 기반을 마련하고 있다. 이는 AI 기술의 발전이 사회적 불평등을 심화시키지 않도록 하고, 모든 국민이 기술의 혜택을 고르게 누릴 수 있도록 하기 위한 조치이다. 또한, 디지털 소외 계층을 위한 교육 프로그램 제공, 접근성 강화 등 다양한 방안을 통해 사회 전반의 디지털 역량을 향상시키고자 한다.

④ AI시대의 신질서 정립 및 글로벌 AI 규범·거버넌스 주도 : 국제기구와의 협력을 확대하여 글로벌 AI 규범을 선도하고자 하며, 국제 사회에서 책임 있는 기술 강국으로서의 위상을 확립하고자 한다.

이러한 전략은 국가 AI 위원회를 중심으로 정부와 민간이 협력하여 AI 전략을 실행하고 세부 계획을 수립하는 방식으로 추진되고 있다. 국가 AI 위원회는 정책의 일관성을 유지하고 효율적인 집행을 보장하기 위한 핵심 기구로서 역할을 하고 있으며, 다양한 이해관계자의 의견을 수렴하여 정책에 반영하고자 한다.

2 국내 보건의료 관련 R/D 전략

가. 『보건의료 데이터·인공지능 혁신 전략』(‘21~’25년)

정부는 본 혁신전략을 통해 데이터 기반의 바이오헬스 경쟁력 확보와 미래의료 혁신을 본격 추진한다는 방침으로, 이를 위해 데이터 생산, 집적, 활용의 전 주기에 걸친 3대 분야 9대 핵심과제를 추진하는 한편, 보건의료 데이터의 안전한 활용 법제와 거버넌스 등 정책 기반도 강화한다는 계획이다.

1) 양질의 데이터 생산과 개방

① 한국인의 호발암종(위암, 대장암, 폐암, 간암, 유방암 등 10대 암종), 개인생성 건강데이터¹¹⁾의 표준화 및 보건의료데이터 품질 관리를 위한 인증제 도입을 검토하여 병원 등에서 고품질 데이터 생산 체계를 구축하도록 지원

② 건강보험 등 공공데이터 개방 건수를 연 1,000건에서 5,000건으로 확대하고, K-Cancer 등 한국인 특화 빅데이터를 단계적으로 구축하여, 질환 예측모형 개발 등 다양한 임상연구, 맞춤형 질병 치료를 촉진하는 핵심 의료데이터를 개방할 예정

③ 데이터 3법 개정 이후 이용자 중심의 가명정보 활성화를 위해 지속적인 제도개선과 결합 사례를 확산해 나갈 예정. 이를 통해, 표준의 부재로 상호 연계·통합 활용이 어려웠던 데이터 활용이 용이해지고, 폐쇄·독점적으로 활용되던 데이터를 개방, 결합하여 고부가가치 데이터가 보편적으로 활용될 전망

2) 고부가가치 데이터 플랫폼 완성

'25년까지 유전체 등 바이오, 병원 임상기록, 공공보건의료 데이터 중심으로 3대 원천 데이터 플랫폼을 완성할 계획

① 100만 명 규모 통합바이오 빅데이터 인프라를 구축하여, 맞춤형치료·정밀의료 등에 활용하고, 임상·유전체·건강보험·개인건강기록(PHR, Personal Health Record) 등과 연계하여 고부가가치 국가 전략 자산화를 추진

② 폐쇄적·독점적으로 활용되어 민간에서 접근하기 어려웠던 병원임상 데이터는 의료데이터 중심병원 기반의 기업-학계-연구기관병원 공동 연구를 활성화

③ 공공분야 데이터를 연계 활용하는 보건의료 빅데이터 플랫폼은 연계기관을 단계적으로 확대하여, 인구고용 등 다양한 분야의 이종데이터 연계·결합 공공 연구를 확산해 나갈 예정

11) Patient Generated Health Data, PGHD: 의료기관 외부에서 일상 생활 속에서 생성되는 건강 관련 데이터로 정의되며, IoT, 웨어러블 등 장비로부터 측정되는 데이터만으로 제한되지 않고, 설문 조사 등도 포함된다.

3) 데이터 활용 혁신으로 성과 가속화

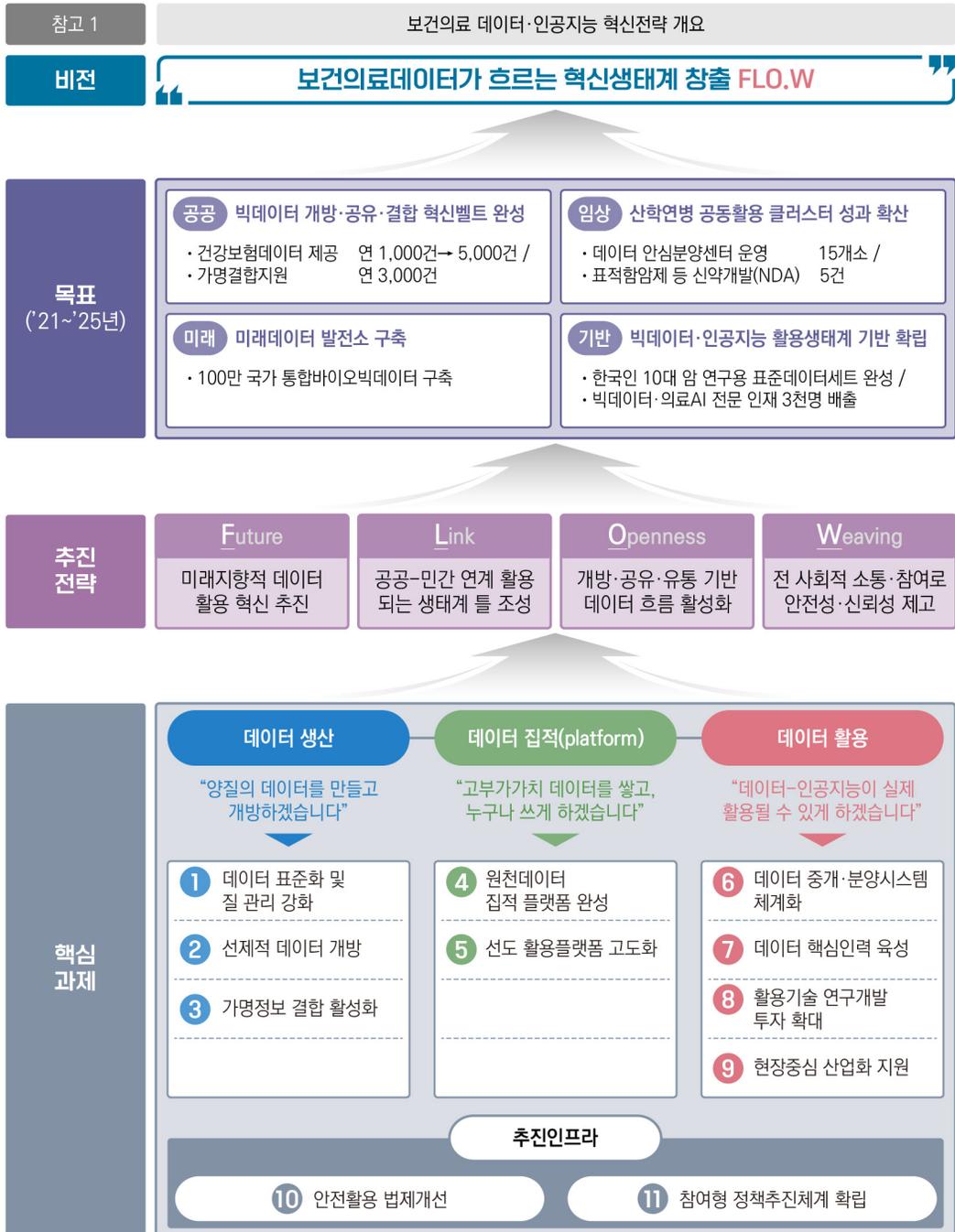
① 안전한 데이터 제공 기반을 마련하기 위해 데이터 증개·분양을 체계적으로 지원 : 데이터 제공-활용기관을 증개하여 꼭 필요한 데이터를 쉽게 활용할 수 있도록 컨설팅 지원 전문인력을 양성하고, 국민건강보험공단 등 주요 공공기관, 의료데이터 중심병원 등 빅데이터 보유 기관을 대상으로 안전한 데이터 제공역량을 갖춘 경우 ‘안심분양센터’로 지정하는 방안을 추진

② 의료인공지능 스타트업에 대한 전(全) 주기 지원체계를 마련하고, 중점분야에 대한 연구개발 투자를 확대: 병원 중심 의료 AI 특화 개방형 실험실 구축·운영 등 우수한 인공지능 모델에 대한 다양한 임상 실증과 창업을 지원할 예정

③ 인공지능의 혁신 가치를 고려할 수 있는 차별화된 평가보상체계를 마련하고, 의료 AI 연구개발 투자를 확대

4) 신뢰할 수 있는 데이터 거버넌스 확립

3대 핵심분야 과제가 원활히 추진될 수 있도록 법제 정비, 민·관 합동 정책 거버넌스를 구축 운영하기 위하여 민감한 보건의료데이터의 특성을 반영하고, 개인의 권리보호를 강화하는 방향으로 법제를 정비하고, 의료 AI 윤리 원칙 수립 등을 통해 민감정보 활용에 대한 사회적 우려를 해소하고 한다. 또한, 사회적 공감대 및 신뢰 기반의 민·관 합동 정책 거버넌스를 가동함으로써 보호와 활용 간 균형 잡힌 추진체계를 마련할 계획이다.



[그림 3-7] 『보건의료 데이터·인공지능 혁신 전략』의 개요도

나. 『의료 인공지능 연구개발(R&D) 로드맵』(2024~2028년)

보건복지부 내 보건의료데이터 정책심의위원회에서 발표한 본 로드맵은 AI를 통한 의료서비스 질 제고 및 의료 AI 연구개발활용 생태계 마련에 중점을 두었다. 이를 위해 ‘인공지능 기반 의료기술 혁신으로 국민건강 증진’을 비전으로, ① 인공지능 의료기술 사업화 확대, ② 기술격차 단축, ③ 연구개발(R&D) 투자 강화를 핵심 목표로 설정하였다. 이 로드맵을 통해 필수의료, 신약개발 등에 AI 연구개발을 확대 지원하고, 의료데이터를 안전하게 활용할 수 있도록 데이터 활용체계를 고도화하는 등 4대 전략 9개 과제를 추진할 계획임을 밝혔다. 추진과제로는 다음과 같다.

1) AI 기반 의료서비스 혁신 지원

① 필수의료 중심 의료 AI 기술개발 확대: 응급 환자의 신속한 중증도 분류 및 AI 기반 응급상황 예측알림 서비스 등 모델 개발, 암 유전자 패널데이터를 표준화·연계한 빅데이터를 구축하여 암 정밀의료 연구 및 AI·임상의사결정 지원시스템 개발과 더불어 치매, 발달장애, 희귀질환 등 주요 분야별로 예측진단, 치료제 타겟 예측 등 인공지능 기술을 접목한 기술개발을 추진하기로 함

② 생성형 AI 기반 의료서비스 기술개발 추진: 진료 과정에서 의료진-환자 간 소통을 지원하고 의료행정을 효율화하는 생성형 AI 기반 의료 기술개발 및 암, 만성질환, 재활 등 퇴원 후 건강 관리, 재택 모니터링 및 정보 제공 등 생성형 AI 기반 의료기술 개발을 추진하기로 함

2) AI 기반 첨단 의료기기 및 신약개발 지원

① 의료기기, 진단보조 분야 AI 기술개발 확대: 인공지능 기반의 디지털치료기기 및 수술 로봇 등 첨단기술을 접목한 의료기기 개발, 인공지능 기반의 다양한 질환 올인원·원스톱 진단 및 차세대 진단 플랫폼 개발 및 초고령사회 진입 대비 노인, 장애인 등 사회적 약자 자립을 지원하는 AI 기반의 돌봄 로봇 기술 개발을 추진하고자 함

② 신약개발 전 과정에서 AI 활용 활성화: 데이터 통합 없이 알고리즘 분석결과 공유를 통해 신약 후보물질 발굴을 효과적으로 수행하는 신약개발 가속화 프로젝트 추진 및

증상기적으로 신약개발 과정 중 가장 비용이 많이 발생하는 임상 단계에서 AI 활용 기술개발과 반복적인 실험과 데이터 분석 과정을 AI 기반으로 자동화하여 자율적이고 연속적인 실험 및 데이터 분석 추진하고자 함

3) 보건의료데이터 활용 체계 고도화

① 보건의료데이터 활용 기반 강화: 의료기관·공공기관 등 흩어진 의료데이터를 AI 연구자, 기업 등이 편리하게 활용할 수 있도록 정보 연계 플랫폼 구축 및 AI·데이터 연구에 필요한 빅데이터 구축 및 개방 확대 추진

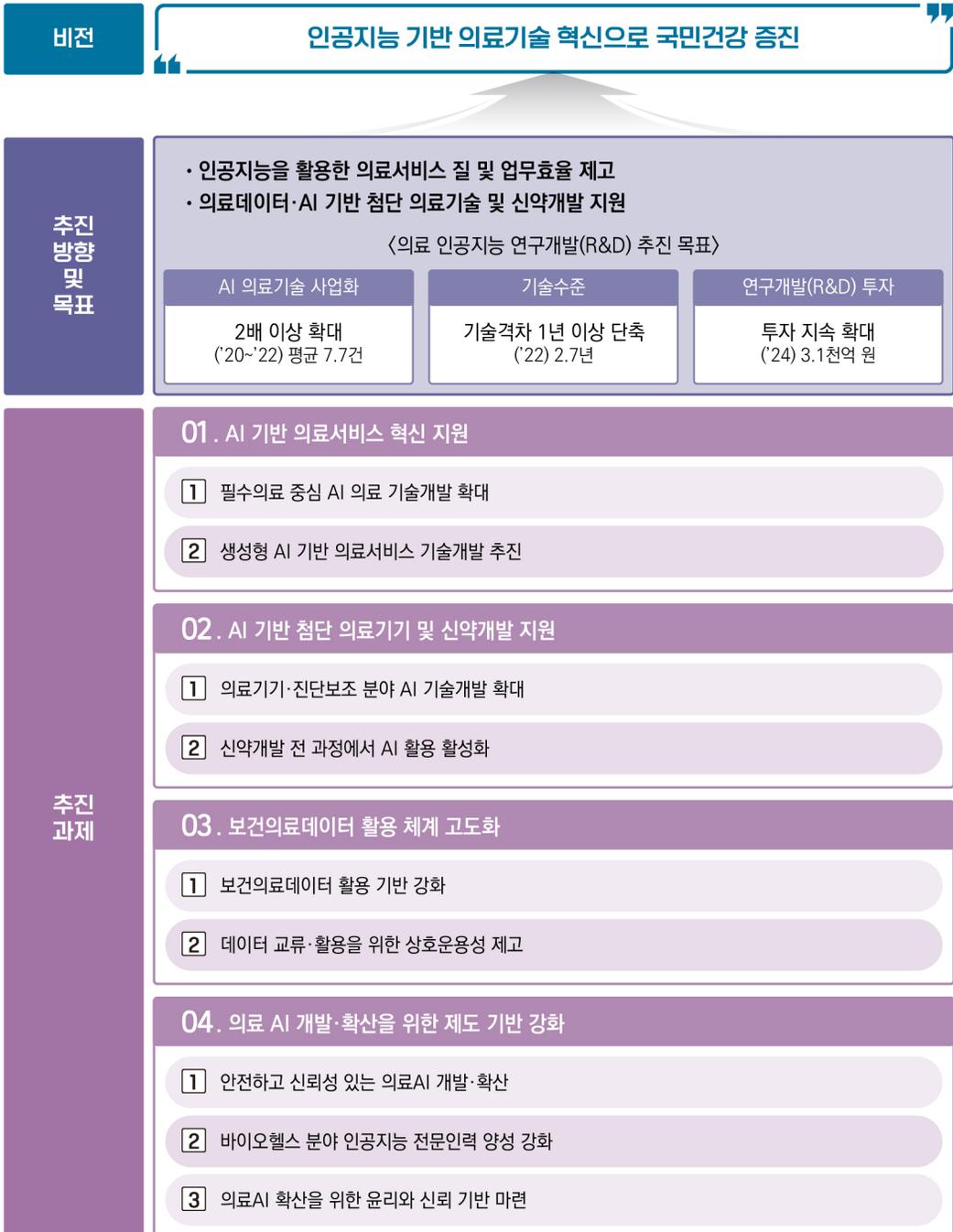
② 데이터 교류·활용을 위한 상호운용성 제고: 의료 AI·데이터 연구에 필요한 상호운용성을 갖춘 고품질 데이터 생성 지원 및 국제 전송표준인 FHIR 전문가가 아니더라도 개발이 용이하도록 문법 규칙과 기준을 학습하여 개발을 지원하는 AI 모델 개발 추진

4) 의료 AI 개발·확산을 위한 제도 기반 강화

① 안전하고 신뢰성 있는 의료 AI 개발·확산: 개발된 의료 AI가 의료 현장에서 의미 있게 활용될 수 있도록 실증 지원, 고령자 등 AI 디지털 취약계층의 격차 해소 및 역량 강화를 위한 디지털 포용기술 개발 추진

② 바이오헬스 분야 인공지능 전문인력 양성 강화: 의대(의학)-공대(ICT·AI)-병원(현장·데이터)간 학과 융합과정 개설·운영 및 현장 수요기반 프로젝트형 교육 개발·운영 계획

③ 의료 AI 확산을 위한 윤리·신뢰 기반 마련: 보건의료분야 AI 연구·개발·활용에 필요한 포괄적·일반적 윤리 기본원칙과 세부 지침 수립, AI 연구개발(R&D) 투자 방향전략, 활용·확산 방안 마련 등 정책 수립에 활용하기 위한 실태조사 실시 및 의료 AI 기술을 안전하고 윤리적으로 활용 가능하도록 관련 협의체, 국제협력 등 논의·관리체계 마련



[그림 3-8] 『의료 인공지능 연구개발(R&D) 로드맵』의 개요도

제4장

국내외 인공지능 윤리 관련 법제도

- 제1절 UN의 인공지능 윤리 관련 법제도
- 제2절 WHO의 인공지능 윤리 관련 법제도
- 제3절 유럽연합의 인공지능 윤리 관련 법제도
- 제4절 미국의 인공지능 윤리 관련 윤리적 법제도
- 제5절 캐나다의 인공지능 윤리 관련 법제도
- 제6절 싱가포르의 인공지능 윤리 관련 법제
- 제7절 국내 인공지능 관련 법제도

제 1절

UN의 인공지능 윤리 관련 법제도

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 인류를 위한 인공지능 거버넌스

(Governance of Artificial Intelligence for Humanity)

UN의 AI 고위급 자문 기구(AIAB)는 안토니우 구테흐스 UN사무총장의 지시하에 2023년 설립되었다. AIAB는 33개국 출신의 기술, 법률 및 데이터 보호 분야 전문가 39명으로 구성되었으며 1년 여의 작업을 거쳐 '인류를 위한 AI 거버넌스(관리)'라는 제목의 보고서를 공개하였다.

보고서는 AI 기술의 발전은 경제 성장, 사회 발전, 환경 문제 해결 등에 큰 잠재력을 가지고 있지만, 동시에 일자리 감소, 프라이버시 침해, 편견과 차별 심화 등의 위험도 내포하고 있다고 말하며, 이러한 위험을 최소화하고 AI의 혜택을 극대화하기 위해서는 국제적인 협력과 공동의 노력을 통한 AI 거버넌스 체계 구축이 필수적이라고 천명하고 있다. 아울러 현재 AI 거버넌스는 국가별, 지역별로 큰 차이를 보이고 있으며, 국제적인 협력과 조정이 부족한 상황으로, 이러한 격차는 AI 기술의 개발과 활용에 불확실성을 야기하고, AI의 잠재력을 충분히 실현하는 데 걸림돌이 될 수 있다고 말하고 있다. 이러한 AI 거버넌스의 필요성 및 글로벌 AI 거버넌스 격차 해소를 위해 글로벌 AI 거버넌스를 위한 7개 권고 사항을 밝혔다:

1. 국제 인공지능 과학 패널: AI 기술의 발전 동향을 지속적으로 모니터링하여 AI의 기회와 위험을 분석하기 위한 AI의 능력과 리스크에 대한 공정한 지식을 제공하며 글로벌 공통의 이해를 촉진하는 패널을 설립
2. AI 거버넌스 정책 대화: 정부와 이해관계자 간의 정기적인 대화를 통해 모범 사례를 공유하고, 주요 AI 사건에 대한 대응 논의
3. AI 표준 교환 플랫폼: AI 표준을 개발하는 글로벌 플랫폼을 마련하여 국가 간 규제의 상호 운용성을 높이고, 규제의 예측 가능성을 제고하기 위한 정기적인 AI 규

제 관련 정책 논의 개최의 필요성을 역설하고 다양한 조직과 이해관계자들이 협력할 수 있도록 지원.

4. AI 역량 개발 네트워크: AI 지식과 자원을 지원하는 네트워크를 구축하고, 특히 개발도상국의 역량 개발 지원.
5. AI 글로벌 기금: AI 자원 불균형을 해소하고, 인프라가 부족한 국가들이 공평하게 AI를 활용할 수 있도록 돕는 기금을 마련.
6. 글로벌 AI 데이터 프레임워크: AI 학습 데이터의 거버넌스를 위한 프레임워크를 마련하고, 인간의 권리와 지속 가능한 개발을 준수하도록 함.
7. UN 사무국 내 AI 부서 설립: AI 거버넌스를 위한 글로벌 협력과 이니셔티브를 조정하는 UN 사무국 내 전담 부서를 설립하고, AI에 대한 UN의 일관된 입장을 취할 수 있는 환경 마련

이 보고서는 AI 거버넌스에 대한 국제적인 논의를 촉발하고, 글로벌 협력 체계 구축을 위한 기반을 마련했다는 점에서 큰 의의가 있다. AI 거버넌스에서의 포용성을 강조하며, 특히 공동의 권고사항 및 패널 등의 협력관계를 통해서 AI 역량에 있어서의 글로벌 격차를 줄이고 개발도상국 등의 다양한 목소리가 반영될 필요성을 역설하고 있다.

제2절

WHO의 인공지능 윤리 관련 법제도

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 세계보건기구의 인공지능 가이드라인

(Ethics and governance of artificial intelligence for health: WHO guidance, 2021)

세계보건기구(WHO)의 AI 가이드라인은 2021년 WHO에서 발표한 “AI 건강 윤리 및 거버넌스 지침”에 포함되어 있으며, 건강 분야에서 AI 기술을 윤리적이고 안전하게 개발하고 사용하기 위한 원칙과 권고사항을 제시하고 있다. 가이드라인은 AI가 건강 분야에 긍정적인 영향을 미치기 위한 기본적인 윤리적 원칙과 거버넌스 체계를 제시하며, 기술이 인권과 형평성을 보장하는 방식으로 활용되도록 촉구하고 있다.



출처: WHO <https://www.who.int/publications/i/item/9789240084759>

[그림 4-1] WHO 인공지능 가이드라인의 주요 내용

가. 윤리적 원칙 (Ethical Principles)

WHO는 AI가 건강 분야에서 사용될 때 반드시 지켜야 할 6가지 윤리적 원칙을 제시했다.

① 형평성 및 포용성 (Equity and Inclusiveness): AI 시스템은 모든 사람이 공평하게 접근할 수 있도록 설계되어야 하며, 취약한 계층을 포함한 모든 사람들이 그 혜택을 받을 수 있도록 해야 한다.

② 투명성 및 설명 가능성 (Transparency and Explainability): AI 시스템은 사용자에게 그 작동 방식과 결정 과정을 명확히 설명할 수 있어야 하며, 모든 과정이 투명하게 문서화되어야 한다.

③ 책임성 (Accountability): AI 시스템을 개발하고 사용하는 모든 이해관계자는 AI가 인간의 건강에 미치는 영향에 대해 책임져야 하며, 잘못된 결정이나 오류에 대해 책임을 지는 체계가 필요하다.

④ 개인정보 보호 및 데이터 보호 (Privacy and Data Protection): AI 시스템은 환자의 개인 건강 데이터를 안전하게 처리하고, 개인정보 보호를 준수해야 한다.

⑤ 안전성 및 보안 (Safety and Security): AI 시스템은 환자의 안전을 최우선으로 고려해야 하며, 지속적으로 업데이트되고 점검되어야 한다.

⑥ 인권 및 윤리적 사용 (Human Rights and Ethical Use): AI는 인권을 존중하고, 차별이나 편향 없이 공정하게 사용되어야 하며, 인간의 존엄성을 해치지 않도록 해야 한다.

나. 거버넌스와 규제 (Governance and Regulation)

WHO는 각국 정부와 규제 기관이 협력하여 AI 기술의 안전하고 윤리적인 사용을 보장할 수 있도록 규제 체계를 마련해야 한다고 강조한다. 국제적인 규칙과 표준을 개발하는 것이 중요하며, 이를 통해 AI가 건강 관리에 미치는 부정적 영향을 최소화하고, 기술이 사회에 미치는 영향을 정확하게 관리해야 한다.

다. AI 기술의 적용과 혜택

WHO 가이드라인은 AI가 진단, 치료, 공공 보건, 의료 연구, 전염병 대응 등 다양한 분야에서 긍정적인 역할을 할 수 있다고 보고한다. AI는 의료 진단의 정확도를 높이고, 치료 방법을 개인화하며, 의료 자원의 분배를 효율적으로 관리하는 데 도움을 줄 수 있다.

라. 위험과 도전 과제

WHO는 AI 기술이 불평등 심화(예: 소외된 계층의 의료 접근 불균형), 편향된 데이터로 인한 오류, 개인정보 보호의 취약성, AI의 투명성 부족 등의 문제를 발생시킬 수 있다고 경고한다. 이러한 위험을 관리하기 위해서는 지속적인 모니터링과 강력한 규제가 필요하다.

마. 기술 개발자와 사용자에게 대한 권고

WHO는 기술 개발자와 사용자에게 AI 시스템을 설계하고 사용하는 과정에서 윤리적이고 인권을 존중하는 접근법을 취할 것을 권장한다. 또한, AI의 성능과 안전성을 지속적으로 평가하고, 필요 시 사용자 교육을 강화해야 한다고 제시하고 있다. 이를 위해 WHO가 제시한 주요 권고 사항은 다음과 같다.

- ① 투명한 알고리즘 개발 및 운영: AI 시스템의 결정 과정은 명확히 설명되고 검증 가능해야 한다.
- ② AI에 대한 국제적 협력: 각국 정부와 국제기구들이 협력하여 AI 기술의 글로벌 표준을 설정해야 한다.
- ③ AI 시스템의 윤리적 설계: AI 시스템은 공정하고 투명하게 설계되어야 하며, 인간의 존엄성을 존중하고, 편향을 줄이기 위해 다양한 데이터와 테스트가 필요하다.
- ④ 모니터링 및 평가: AI 시스템의 성과와 결과를 지속적으로 모니터링하고, 그 결과에 대한 피드백을 제공하여 개선할 수 있는 체계를 마련해야 한다.

2 보건 분야 인공지능의 윤리와 거버넌스:

대규모 멀티모달 모델에 대한 지침

(Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models: WHO guidance, 2024)

WHO의 2021년 가이드라인이 인공지능(AI)의 윤리적 원칙(자율성 보호, 투명성, 포괄성 등)을 중심으로, 전통적인 AI와 머신러닝 모델이 보건의료에 미치는 영향을 다뤘다면, 2024년 새로이 발표된 WHO의 보고서에서는 LMM(Large Multi-modal Models), 즉 텍스트, 이미지, 음성 데이터 등 다중 데이터를 처리할 수 있는 AI 모형들의 보건의료 분야에서의 응용 가능성을 제시하고 있다.

특히, LMM이 진단 지원 등 기존 머신러닝 기반 AI들이 수행하던 영역들은 물론, 환자 중심의 서비스 제공, 행정 업무의 자동화, 의료 연구 등 다양한 영역에서 잠재력을 발휘할 수 있음을 지적하고 있으며, 특히 광범위한 작업을 수행할 수 있는 범용성에 그 초점을 맞추고 있다. 해당 가이드라인에서는 기존 2021년 가이드라인에서 제시한 윤리 원칙을 그대로 제시하면서, 추가적으로 LMM에서 나타나기 쉬운 잘못된 정보 제공(hallucination)의 가능성을 보완할 수 있는 안전성 확보 체계의 수립, 데이터 편향을 극복하기 위한 포괄적 데이터 활용 및 형평성의 고려부터, LMM의 높은 에너지 활용에 따른 탄소 배출 및 환경 영향까지 보다 광범위한 영역들에 대한 고려사항들을 함께 제시하고 있다.

제3절

유럽연합의 인공지능 윤리 관련 법제도

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 「신뢰할 수 있는 인공지능 윤리 가이드라인」 (Ethics guidelines for trustworthy AI)

유럽연합(EU)은 2019년 4월, ‘신뢰할 수 있는 AI 구현’이라는 핵심개념을 담아 AI 규제안을 발표한 바 있다. AI 가이드라인의 궁극적인 목적은 “인공지능을 통해 개인과 사회가 막대한 이익을 창출하고 더 나아가 인류 번영과 성장을 촉진하도록 인공지능 기술의 가능성을 극대화하기 위해 적절히 관리하고 안전하게 법적 토대를 마련하는 것”이다. EU는 AI 시스템을 수준에 따라 용인할 수 없는 위험, 고위험, 낮은 위험으로 분류해 단계별 위험 관리가 필요하다고 적시하였다. EU는 고위험 AI 항목으로 생체 인식 및 분류, 공공 인프라, 교육 및 직업훈련, 고용, 공공 서비스 및 필수 개인 서비스, 법 집행, 이주 및 망명 등 7개 분야를 명시하였다. 7개 항목과 관련 AI 시스템을 개발한 기업과 개인은 요구사항을 준수해야 하며, 고위험 시스템은 적합성 평가를 거쳐야 사용이 가능하다.

AI 가이드라인이 가져야 하는 윤리의 기반은 국제인권법, EU 조약, EU 헌장 등의 5가지 개인의 권리에 근거하고 있으며, 이를 존중하는 4가지 윤리원칙을 다음과 같이 제시하고 있다.

EU AI 가이드라인이 가져야 하는 윤리의 기반

- ① 인간의 존엄성에 대한 존중 : 모든 인간은 타인에 의해 억압되지 않는 본질적 가치를 지니고 있으므로 AI 시스템은 인간의 신체적·정신적 존엄(integrity), 개인적·문화적 정체성(identity) 및 본질적 욕구를 만족시키고 보호하는 방식으로 개발되어야 한다.

- ② 개인의 자유 : 인간은 스스로 자유롭게 자신의 삶에 대한 결정을 내릴 수 있어야 하며 개인의 자유는 불법적 강요나 정신적 위협, 부당한 감시와 속임수, 불공평한 조작에 의해 억압되어서는 안 된다.
- ③ 민주주의·정의·법치주의에 대한 존중 : AI 시스템은 민주적 과정을 유지·육성하고 개인의 가치와 삶의 선택에 대한 다원성을 존중해야 한다. 또한, 법률에 근거한 근본적인 약속을 훼손하지 않도록 하고, 법 앞에 적법한 절차와 평등을 보장해야 한다.
- ④ 평등·차별금지·연대 : 모든 인간의 도덕적 가치와 존엄성을 평등하게 존중해야 하고 특히, 노동자, 여성, 소수 인종, 어린이, 소비자, 취약계층에 대한 존중이 필요하다.
- ⑤ 시민의 권리 : 시민들은 투표권, 행정권, 공공정보 열람권, 행정청구권 등의 권리를 가지고 있다. AI 시스템은 사회서비스를 제공하는 데 있어서 정부의 효율성을 개선할 수 있는 잠재력이 있으며, 동시에 시민의 권리를 제약할 가능성이 존재한다. AI 시스템이 긍정적인 효과와 함께 시민의 권리에 부정적 영향을 미치지 않도록 해야 한다.

4가지 윤리원칙

- ① 인간의 자율성 존중 : AI 시스템과 상호작용하는 인간은 자신에 대해 완전하고 효과적인 자기 결정권을 유지할 수 있어야 하며 민주적 절차에 참여할 수 있어야 한다. AI 시스템이 인간을 부당하게 종속시키거나, 강압하거나, 기만하거나, 조작하거나, 조절하거나, 집단화해서는 안 되며, 인간의 인지, 사회·문화적 기술을 강화하도록 설계되어야 한다. 인간과 AI 시스템 간의 역할 분배는 인간 중심의 설계 원칙을 따라야 하며 인간이 선택할 수 있도록 해야 한다. 이는 AI 시스템의 작업 프로세스에 대한 인간의 감독 권한을 확보하는 것을 의미한다.
- ② 피해 예방 : AI 시스템은 인간에게 피해를 주거나 상황을 악화시키는 등 부정적인 영향을 주어서는 안 된다. AI 시스템이 작동하는 환경은 인간의 존엄성을 정신적·육체적으로 보호하도록 안전해야 한다. 기술적으로도 견고해야 하며 악의적으로

사용되지 않도록 해야 한다. AI 시스템의 개발·배포·사용에 있어서 취약계층을 포함해야 하고 고용주와 직원, 기업과 소비자, 정부와 시민 등이 함께 참여하여 권력이나 정보의 비대칭으로 인한 피해가 없도록 해야 한다.

- ③ 공정성 : AI 시스템의 개발·배포·사용은 공정해야 하며, 공정성을 향상시키는 방향으로 사용되도록 해야 한다. 실질적으로 편익과 비용이 평등하게 분배되고 개인과 집단이 편견, 차별, 낙인에 피해를 입지 않아야 한다. 교육, 상품, 서비스, 기술에 대한 접근의 기회가 평등하게 보장되어야 하고 인간의 선택의 자유가 부당하게 침해되어서는 안 된다. 이를 위해 AI 시스템과 이를 운영하는 실무자는 수단과 목적에 대한 원칙을 존중하고 방법을 신중하게 고려해서 의사결정을 해야 하며 소비자는 이익을 제기하고 보상을 요구할 수 있어야 한다.
- ④ 설명 가능성 : AI 시스템의 목적과 기능이 투명하게 공개되어야 하고 직간접적으로 영향을 받는 사람들에게 설명할 수 있어야 한다. AI 알고리즘이 ‘블랙박스’인 상황에서 그 과정이 항상 설명 가능한 것은 아니므로 가능한 한 추적 가능성, 감사 가능성 및 시스템 기능에 대한 투명한 커뮤니케이션이 필요하다.

출처: 인공지능의 윤리적 쟁점에 관한 탐색적 연구

2 「인공지능법」(the Artificial Intelligence Act)

유럽의회는 2024년 3월 13일 인공지능에 관한 세계 최초의 포괄적 법안인 “인공지능법(the Artificial Intelligence Act)”(이하 「EU 인공지능법」)을 통과시켰다. 「EU 인공지능법」은 법안이 발의될 때부터 AI에 관한 세계 최초의 포괄적인 입법이라는 점에서 주목을 받아왔으며, 한국의 국회에서도 「EU 인공지능법」의 구조와 접근법에 영향을 받은 다수의 법안을 발의한 바 있다.

「EU 인공지능법」은 2024년 8월 1일에 발효되었으며 주요 규정들은 이후 단계적으로 시행될 예정이다. 인공지능법에서는 EU 역내에서 사용되는 AI 시스템의 위험도를 4 단계(허용 불가능한 위험/고위험/ 제한적 위험/최소 위험)로 분류하여 의무 사항을 달

리 규정하고 있다. 이에 따라 ‘허용 불가 위험’으로 분류된 AI 시스템에 대한 규정은 2025년 2월 2일부터 적용되며, 범용 AI 모델에 대한 규제는 2025년 8월 2일부터 시행될 예정이다. 전체적인 법안의 전면 시행은 2026년 8월 2일부터 시작될 예정이다. 특히, 생성형 AI 시스템 제공자와 배포자의 투명성 의무를 강조하며 의무 위반 시 벌금을 차등적으로 부과하고 있으며, 스타트업 및 중소기업 지원 규정을 마련하여 공정한 AI 기술 혁신을 장려하고 있다.

가. 「EU 인공지능법」의 규제 범위

「EU 인공지능법」은 주로 인공지능 시스템을 개발하는 공급자 및 인공지능 시스템 배포자(사용자)를 규제 대상으로 하고 있다. 구체적으로, (i) 인공지능 시스템을 사용하는 EU 내 기관, (ii) EU 내 인공지능 시스템 수입업체, 유통업체 또는 제조업체, (iii) 인공지능 시스템을 EU 시장에 출시하는 EU 내·외 공급자, (iv) 해당 결과물이 EU 내에서 사용되는 인공지능 시스템의 EU 내·외 공급자 및 사용자 등이 규제 대상이다. 여기에서 ‘사용자’란, 인공지능 시스템으로부터 영향을 받는 최종 사용자가 아닌 인공지능 시스템을 전문적으로 배포하는 자연인 또는 법인을 의미한다.

나. 위험 수준에 따른 인공지능 분류

「EU 인공지능법」은 잠재적인 위험 및 영향 수준에 따라 AI 시스템을 분류하며, 각 위험군에 대하여 서로 다른 규칙과 의무가 적용되고 있다. 해당 규정을 준수하지 않을 경우, 위반 규정, 위반 정도 및 회사 규모에 따라 최대 750만 유로 또는 전 세계 연간 총매출액의 1.5%부터 최대 3,500만 유로 또는 전 세계 연간 총매출액의 7%에 달하는 벌금이 부과될 수 있다.

1) 금지된 인공지능 시스템 (Prohibited AI Systems)

「EU 인공지능법」이 금지하는 AI 시스템은 다음과 같다.

① 잠재의식적, 조작적 또는 기만적 기술을 이용하여 행동을 왜곡하고 정보에 입각한 의사결정을 저해함으로써 심각한 피해를 입히는 AI 시스템

② 연령, 장애 또는 사회경제적 상황과 관련된 취약성을 악용하여 행동을 왜곡함으로써 심각한 피해를 입히는 AI 시스템

③ 개인의 민감한 특성(예: 인종, 정치적 견해, 노동조합 가입 여부, 종교적 또는 철학적 신념, 성생활 또는 성적 취향)을 추론하는 생체 인식 분류 AI 시스템 (단, 합법적으로 취득한 생체 정보에 대한 라벨링 또는 필터링 및 법 집행으로서 생체 정보를 분류하는 경우는 제외)

④ 소셜 스코어링(social scoring), 즉 사회적 행동이나 개인적 특성을 기반으로 개인이나 그룹을 평가 또는 분류하여 해롭거나 불리한 대우를 초래하는 AI 시스템

⑤ 오로지 프로파일링이나 성격 특성만을 토대로 개인이 범죄를 저지를 위험을 예측하는 AI 시스템(단, 범죄 행위에 직접적으로 연결된 객관적이고 입증 가능한 사실을 기반으로 사람에게 대한 평가를 향상시키는 데 사용되는 경우는 제외)

⑥ 인터넷이나 CCTV 영상에서 안면 이미지를 무차별적으로 수집하여 안면 인식 데이터베이스를 구축하는 AI 시스템

⑦ 의료 또는 안전상의 이유를 제외하고, 직장이나 교육기관에서 감정을 추론하는 AI 시스템

⑧ 법적으로 허용되는 경우를 제외하고, 법 집행 목적의 공개적으로 접근 가능한 공간에서의 실시간 원격 생체 인식(real-time remote biometric identification, RBI) 인공지능 시스템

2) 고위험 인공지능 시스템 (High-risk AI Systems)

「EU 인공지능법」은 고위험 AI 시스템에 대해 엄격하게 규제하며, 규정 준수를 보장하기 위한 광범위한 거버넌스 활동을 의무화하였다. 고위험 AI 시스템이란 (i) EU 법률의 적용을 받는 안전 부품/제품으로 사용되며 제3자 적합성 평가를 받아야 하는 AI 시스템 또는 (ii) 금지되지 않는 생체 인식, 중요 인프라, 교육 및 직업 훈련, 고용, 근로자 관리 및 자영업 접근, 필수 공공/민간 서비스에 대한 접근 및 향유, 법 집행, 이주, 망명 및 국경 통제 관리, 사법 행정 및 민주적 절차에 이용되는 AI 시스템을 말한다.

이러한 고위험 AI 시스템의 공급자는 (a) AI 시스템 수명주기 전반에 걸친 위험 관리

시스템 구축, (b) 데이터 거버넌스 수행, (c) 기술 문서 작성 및 (d) 자동 기록 보관, 정확성, 견고성 및 사이버 보안을 보장하는 고위험 AI 시스템 설계 등의 추가 의무 사항을 준수해야 한다고 규정하였다.

3) 범용 AI (General purpose AI, GPAI) 모델

범용 AI 모델(이하 “GPAI”)이란, 최적화된 자체 감독을 이용하여 대량의 데이터로 훈련한 경우를 포함하여, 상당한 일반성을 나타내고 모델이 시장에 출시된 방식에 관계없이 광범위한 고유 작업을 능숙하게 수행할 수 있으며, 다양한 하위 시스템이나 응용 프로그램에 통합될 수 있는 AI 모델을 의미한다. 모든 GPAI 모델 공급자는 기술 문서와 사용설명서를 제공하고, 저작권 지침을 준수하며, 훈련에 사용된 콘텐츠의 요약본을 공개해야 한다. 무료 및 오픈 라이선스 GPAI 모델 공급자는 시스템 리스크를 초래하는 GPAI 모델로 분류되지 않는 한, 저작권 지침을 준수하고 충분히 상세한 훈련 데이터 요약본을 공개하기만 하면 된다. 그러나, 오픈 라이선스 모델인지 여부를 불문하고 시스템 리스크를 초래하는 것으로 간주되는 GPAI 모델의 공급자는, 이에 더해 모델 평가 및 적대적 테스트를 수행하고, 심각한 사고를 추적하여 문서화 및 보고하며 적절한 사이버 보안 수준을 보장하도록 규정하고 있다.

4) 제한된 위험 및 최소 위험 인공지능 시스템 (Limited Risk and Minimal Risk AI Systems)

「EU 인공지능법」은 제한된 위험의 AI 시스템에는 투명성 의무를 부과한 반면, 최소 위험 AI 시스템에 대해서는 별도 규제를 부과하지 않았다.

3 유럽연합의 보건의료 분야 인공지능 관련 가이드라인

가. AI Act 내 보건의료 부분

AI Act는 2021년에 처음 제안된 법안으로, 2023년 6월 14일, 유럽 의회는 ‘인공지능 법(AI Act)’에 대한 협상안을 승인한 후 2023년 12월 8일 생성형 AI 규제를 포함한 EU 인공지능법 최종안에 합의하였고, 2024년 2월 2일에는 EU 27개국으로 구성된 이사회가 최종안을 승인한 데 이어, 3월 13일 유럽 의회의 본회의 표결을 통과하였다. 이 법안은 세계 최초로 AI에 대한 포괄적인 규제 프레임워크를 구축했으며, 특히 고위험 AI 시스템에 대한 엄격한 규제를 도입하고 있다. 이 법안은 AI 시스템을 위험 수준에 따라 ‘허용할 수 없는 위험’, ‘고위험’, ‘제한된 위험’, ‘최소 위험’의 네 가지 범주로 분류하고 있으며 특히 ‘고위험’으로 분류된 AI 시스템은 의료, 교육, 금융, 고용, 법 집행 등 시민의 기본권에 영향을 미칠 수 있는 분야에서 사용되는 경우에는 엄격한 규제를 받도록 하고 있어 보건의료 분야에서 사용하는 AI 시스템도 이 범주에 속한다. 이 법안은 의료 데이터의 투명성, 비차별성, 그리고 프라이버시 보호를 중심으로 규제를 마련하며, 의료기기에 포함된 AI 알고리즘의 안전성과 윤리적 사용을 강조하고 있다.

고위험 시스템에 대한 규제는 AI 시스템이 어떻게 학습하고 적용되는지를 검토하며, 특히 환자의 안전과 관련된 결정에 AI가 관여할 경우, 엄격한 감독을 받도록 규정하고 있다. 예를 들어, 진단 AI 도구나 치료 계획 AI는 이 법안에 따라 엄격한 검증과 승인을 받아야 한다.

나. AI in healthcare: applications, risks, and ethical and societal impacts

2022년 유럽의회는 AI는 보건의료 분야에 혁명을 일으킬 잠재력을 지니고 있으며, 질병 진단 및 치료 방식을 변화시키고 있음을 주지하고 보건 의료분야에서의 AI의 활용 기회와 더불어 잠재적인 위험은 무엇이며 이를 완화시킬 수 있는 전략 등을 제시하고 있다. 보건의료 분야에서 AI가 제공하는 기회는 다음과 같다:

① 진단 및 치료 개선 : AI는 의료 영상 분석, 병리 슬라이드 검토, 유전체 데이터 분석 등을 통해 질병의 조기 진단 및 정확한 치료에 기여

② 임상의 효율성 향상 : AI는 의료 기록 분석, 진료 예약 관리, 치료 계획 수립 등 반복적인 작업을 자동화하여 의료진의 업무 부담을 줄이고 환자 진료에 더 집중할 수 있도록 지원

③ 자원 할당 최적화 : AI는 환자 데이터 분석을 통해 입원 기간 예측, 병상 배정 최적화, 의료 자원의 효율적인 배분에 기여

④ 신약 개발 가속화 : AI는 신약 후보 물질 발굴, 약물 효능 예측, 임상시험 설계 등에 활용되어 신약 개발 프로세스를 단축하고 비용을 절감

⑤ 맞춤형 의료 제공 : AI는 환자의 유전 정보, 생활 습관, 의료 기록 등을 분석하여 개인에게 최적화된 치료법을 제시

이와 상반되게 보건의료 분야에서 AI가 제기하는 위험은 다음과 같이 제시하고 있다.

① 잠재적 오류 및 환자 피해 : AI 알고리즘의 오류 또는 오작동으로 인한 잘못된 진단, 부적절한 치료, 환자 안전 위협

② 편견 및 건강 불평등 심화 : AI 모델은 훈련 데이터에 내재된 편견을 학습하여 특정 인종, 성별, 사회경제적 집단에 불리한 결과 초래

③ 투명성 및 신뢰 부족 : AI 알고리즘의 복잡성과 불투명성은 의사결정 과정에 대한 이해를 어렵게 만들고 환자의 신뢰를 저해

④ 데이터 프라이버시 및 보안 위협 : 보건의료 데이터는 민감한 개인정보를 포함하고 있으며, 해킹, 유출, 오용으로부터 보호 필요

⑤ 책임 소재 불분명 : AI 시스템의 오류로 인해 발생한 피해에 대한 책임 소재를 명확히 규정이 곤란

⑥ 실제 의료 환경에서의 구현 어려움 : AI 기술의 복잡성, 데이터 표준화 문제, 의료진의 AI 이해 부족 등은 실제 의료 환경에서 AI 도입을 어렵게 함

이러한 의료 AI의 위험성을 완화하기 위한 전략으로 본 보고서는 다음과 같은 내용을 담고 있다.

① 다중 이해관계자 참여 : 보건의료 AI 개발 및 구현 과정에 의료 전문가, 환자, 윤리 전문가, 규제 기관 등 다양한 이해관계자를 참여시켜 의견을 수렴하고 잠재적 문제점을 해결

- ② 투명성 및 추적성 확보 : AI 알고리즘의 작동 방식, 데이터 출처, 의사결정 과정을 투명하게 공개하고, 오류 발생 시 원인을 추적할 수 있도록 시스템을 설계
- ③ 철저한 임상 검증 : AI 도구를 실제 의료 환경에 도입하기 전에 충분한 임상 시험을 통해 안전성과 효과를 검증
- ④ AI 교육 및 훈련 : 의료진에게 AI 기술에 대한 교육을 제공하여 AI 도구를 효과적으로 활용하고 잠재적 위험을 인지
- ⑤ 규제 및 윤리적 프레임워크 구축 : AI 개발 및 활용에 대한 명확한 규제 지침과 윤리적 프레임워크를 마련하여 환자 안전과 데이터 보안을 보장
- ⑥ AI 여권 도입 : AI 모델 및 데이터 세부 정보, 평가 결과, 지속적인 모니터링 및 감사에서 얻은 정보를 포함하는 'AI 여권'을 도입하여 AI 기술에 대한 정보를 표준화하고 투명성을 향상

제4절

미국의 인공지능 윤리 관련 윤리적 법제도

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 미국의 보건의료 AI 법제도

가. 행정명령

1) 행정명령 Executive Order 13859 of February 11, 2019

(Maintaining American Leadership in Artificial Intelligence:
인공지능 분야에서 미국의 리더십 유지)

행정명령 13859호는 트럼프 행정부에서 2019년 2월 11일 발표한 “인공지능 분야에서 미국의 리더십 유지”를 위한 행정명령이다. 이 명령은 미국의 AI 기술 리더십 유지 및 강화와 더불어 AI의 안전한 활용을 강조하고 있다. 구체적으로는 AI 분야 리더십 유지를 위해 연방 기관은 연구개발, 인력 교육, 산업 협력 등을 강화하기 위한 노력을 해야 하며, AI 혁신을 촉진하고, 공공의 신뢰를 유지하며, 시민의 자유와 권리를 보호하는 방향으로 규제 환경을 조성할 것, AI 시스템의 안전성, 보안성, 신뢰성을 확보하기 위한 기술 개발 및 표준 마련을 지원하고 국제적인 협력을 통해 AI 기술의 발전과 활용에 대한 공동의 이해를 증진하고, 국제 표준 마련에 참여할 것을 강조하고 있다.

2) 행정명령 「Executive Order 13960 of December 3, 2020」

(Promoting the Use of Trustworthy Artificial Intelligence in the
Federal Government, 연방정부에서 신뢰할 수 있는 인공지능 사용 촉진)

Executive Order 13960은 2020년 12월 3일 트럼프 행정부에서 발표한 행정명령으로, ‘연방정부에서 신뢰할 수 있는 AI의 사용 촉진’을 목표로 한다. 연방정부 기관이 AI를 사용할 때 지켜야 할 원칙을 제시하고 책임감 있고 윤리적인 AI 활용을 장려하기 위해 만들어졌다. 본 행정명령은 신뢰할 수 있는 AI, 책임 있는 AI, 위험 관리, 편익 증

진이라는 네 가지 주요 내용을 담고 있다. 첫째, AI 시스템을 설계, 개발, 획득 및 사용할 때 공공의 신뢰와 자신감을 조성하는 방식으로 접근해야 하며, 개인정보 보호, 시민의 자유 및 권리 보호, 차별금지 등의 원칙을 준수해야 한다. 둘째, AI 시스템의 개발 및 활용에 대한 책임성을 확보하고, 잠재적 위험을 식별하고 평가하여 적절한 위험 관리 조치를 취해야 한다. 셋째, AI 시스템을 활용하여 공공 서비스의 질을 향상시키고, 효율성을 높이며, 국민의 삶을 개선하는 데 노력해야 한다. 행정명령 13960은 AI 시스템의 투명성, 설명 가능성, 공정성 및 비차별, 개인정보 보호, 안전 및 보안에 대한 구체적인 지침을 제공한다. 이 행정명령은 연방정부의 AI 활용에 대한 기본 원칙을 제시하고 책임감 있고 윤리적인 AI 활용을 위한 기반을 마련했으며, AI 시스템의 투명성, 설명 가능성, 공정성, 개인정보 보호 등 중요한 가치를 강조하여 AI 기술에 대한 신뢰를 높이는 데 기여했다. 또한, 연방기관의 AI 활용에 대한 명확한 지침을 제공하여 AI 기술의 효율적인 활용을 촉진했다.

결론적으로 행정명령 13960은 미국 연방정부의 AI 활용에 대한 중요한 정책 방향을 제시하는 문서이며, 연방정부는 이를 통해 AI 기술의 잠재력을 최대한 활용하면서도 그로 인해 발생할 수 있는 위험을 최소화하고 공공의 이익을 증진하는 것을 꾀하고 있다.

3) 행정명령 제14110호(Executive Order 14110: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 안전하고 신뢰할 수 있는 인공지능 개발 및 사용)

바이든 행정부는 2023년 10월 30일 ‘안전하고 보안이 보장되며 신뢰할 수 있는 AI 개발 및 사용에 관한 행정명령’을 발표했다. 이 행정명령은 AI 기술의 안전하고 책임감 있는 사용을 보장하기 위한 목적으로 제정되었으며, AI 기술의 잠재력을 최대한 활용하면서도 안전, 보안, 신뢰성을 보장하고, 시민의 권리와 자유를 보호하기 위한 미국의 노력을 보여주는 중요한 정책 방향이다.

본 행정명령을 통해 미국 인공지능 안전 연구소(USAISI)를 설립하여 AI의 안전성, 보안 및 신뢰성을 증진하기 위한 연구를 수행하고, 그 결과를 바탕으로 표준 및 모범 사례를 개발하도록 하였다. 아울러 국가 AI 자문 위원회(NAIAC)를 강화하여 AI의 책임 있는 개발 및 사용에 대한 권고를 제공하고, AI 위험 관리 프레임워크 개발을 지원하도록 하였다.

나. AI 권리장전(AI Bill of Right): 인공지능(AI)과 관련한 인권 보호 지침

백악관은 2022년 10월 5일 과학기술정책실 명의로 인공지능(AI) 기술의 개발과 사용 과정에서 발생할 수 있는 부작용을 최소화하기 위한 “AI 권리장전(AI Bill of Right) 청사진”을 발표한 바 있다. AI 권리장전은 유럽연합(EU)이 추진 중인 AI법보다는 법적 구속력이 약한 선언적 지침에 그친다.

하지만, 미국 정부 차원에서 AI와 관련한 인권 보호 원칙을 처음으로 공식 천명한 것으로 향후 AI에 대한 미국 정부의 광범위한 규제 지침이 될 가능성이 큰 것으로 해석되고 있다. AI 권리장전 청사진은 ① 안전하고 효과적인 시스템 구축, ② 알고리즘을 통한 차별 방지, ③ 데이터 관련 사생활 보호, ④ 자동화 시스템의 활용에 대한 고지와 설명, ⑤ 인간 대안 마련 등의 5가지 원칙으로 구성되어 있다.

백악관은 AI 시스템은 안전하고 의도된 대로 사용되어야 하며, 필요 시 아예 사용하지 않거나 제거할 수 있어야 한다고 권고하고 있다. 또한, 알고리즘 등 시스템에 인종이나 피부색, 민족, 성별, 종교, 연령, 장애 등에 따라 사람을 차별할 수 있는 고정관념이 내재되지 않아야 한다고 설명하고 있다. 아울러 반드시 개인의 허락을 받아 필요한 정보만을 수집하고 특히 의료, 직업, 교육, 범죄기록, 금융 등 민감한 정보가 유출되지 않도록 안전장치를 마련할 것을 주문하고 있다. 이 외에도 사업자들은 AI가 어떻게 작동하고 사용자에게 어떤 영향을 미치는지 이해하기 쉬운 표현으로 설명해야 한다고 언급하고 있다.

2 미국의 보건의료 분야 인공지능 관련 가이드라인

가. 미국 보건부(Department of Health and Human Services, HHS) 및 보건의료정보기술조정국(Office of the National Coordinator for Health IT, ONC)의 AI 프레임워크

미국 보건부(HHS)는 보건의료 분야에서 AI의 책임감 있는 개발과 활용을 촉진하기 위해 2023년 5월 “HHS AI 전략”을 발표했다. 이 전략은 HHS의 미션 달성을 위해 AI를 활용하는 방법을 제시하며, 4가지 핵심 영역에 중점을 두고 있다.

나. AI 준비 인력 개발 및 AI 문화 강화

HHS는 보건의료 분야의 AI 인력 부족 문제를 해결하고 AI 기술에 대한 이해와 수용을 증진하기 위해 다음과 같은 노력이 필요하다고 언급하고 있다.

① AI 교육 및 훈련 프로그램 제공 : 의료 전문가, 연구자, 정책 입안자 등을 대상으로 AI 기초, AI 윤리, AI 활용 사례 등에 대한 교육 및 훈련 프로그램을 제공

② AI 커뮤니티 구축 : AI 관련 학회, 워크샵, 온라인 포럼 등을 통해 AI 전문가들이 지식과 경험을 공유하고 협력할 수 있는 환경을 조성

③ AI 문화 조성 : AI 기술에 대한 긍정적인 인식을 확산하고, AI 도입에 대한 의료진의 거부감을 해소하기 위한 홍보 및 교육 활동을 전개

다. 의료 AI 혁신 및 연구 개발 장려

HHS는 보건의료 분야의 AI 혁신을 가속화하고 AI 연구 개발을 지원하기 위해 다음과 같은 정책을 추진하기로 하였다.

① AI 연구개발 투자 확대: NIH, CDC 등 HHS 산하 기관을 통해 AI 연구 개발에 대한 투자를 확대하고, 혁신적인 AI 솔루션 개발을 지원

② 공공 데이터 접근성 향상: 의료 데이터, 유전체 데이터, 임상시험 데이터 등 AI 연구 개발에 필요한 공공 데이터에 대한 접근성을 높이고, 데이터 공유 및 활용을 장려

③ 규제 개선: AI 기반 의료기기 및 서비스에 대한 규제를 명확화하고 간소화하여 AI 혁신을 저해하는 요소를 제거

라. 기본적인 AI 도구 및 리소스의 민주화

HHS는 모든 의료기관이 AI 기술의 혜택을 누릴 수 있도록 AI 도구 및 리소스에 대한 접근성을 높이는 데 주력하고자 하였다.

① 오픈 소스 AI 도구 개발 및 배포 : 의료 영상 분석, 질병 예측, 약물 개발 등 다양한 분야에서 활용 가능한 오픈 소스 AI 도구를 개발하고, 의료기관에 무료로 배포

② 클라우드 기반 AI 플랫폼 구축 : 보건의료 데이터 저장, 분석, AI 모델 훈련 등을 위한 클라우드 기반의 AI 플랫폼을 구축하여 의료기관이 AI 기술을 효율적으로 활용할 수 있도록 지원

③ AI 교육 자료 개발 및 보급 : 의료진, 연구자, 환자 등을 위한 AI 교육 자료를 개발하고, 온라인 플랫폼, 모바일 앱 등을 통해 널리 보급

마. 신뢰할 수 있는 AI 사용 및 개발 촉진

HHS는 AI의 윤리적 문제를 해결하고 책임감 있는 AI 개발 및 활용을 위해 다음과 같은 지침을 제공하고 있다.

① AI 윤리원칙 수립 : 편견, 차별, 개인정보 침해 등 AI 윤리 문제를 해결하기 위한 AI 윤리원칙을 수립하고, AI 개발 및 활용 과정에서 이를 준수

② AI 거버넌스 프레임워크 개발 : AI 개발, 배포, 사용, 모니터링 등 AI 전수명 주기 전반에 걸친 거버넌스 프레임워크를 개발하여 AI의 책임감 있는 활용을 위한 기반 마련

③ AI 편견 완화 : AI 알고리즘의 편견을 완화하기 위한 기술개발 및 연구를 지원하고, AI 모델 훈련에 사용되는 데이터의 다양성을 확보

제5절

캐나다의 인공지능 윤리 관련 법제도

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 캐나다의 인공지능 관련 법제도

캐나다는 ‘디지털 헌장(Digital Charter)’을 이행하기 위하여 2022년 6월 16일에 ‘디지털 헌장 이행법안(Digital Charter Implementation Act, Bill C-27)’을 발의 한 바 있다. 디지털 헌장은 캐나다가 디지털 경제 및 데이터 주도 경제를 선도할 수 있도록 보편적 접근, 안전 및 보안, 개인정보의 투명성·이동성·상호호환성, 개방적·현대적 디지털 정부, 증오 및 폭력 극단주의로부터 자유 등의 ‘10대 원칙’을 내용으로 하고 있다.

캐나다는 기술 발전 속도에 맞추어 개인정보 보호 법제를 현대화하기 위하여 「소비자 프라이버시 보호법」과 「개인정보 보호법」 개정안 및 「인공지능 시스템 및 인공지능 시스템에 사용되는 데이터에 관한 법(An Act respecting artificial intelligence systems and data used in artificial intelligence systems)」으로 법안을 구성하였다. 「인공지능 시스템 및 인공지능 시스템에 사용되는 데이터에 관한 법」 법안의 주요 내용은 다음과 같다.

목적 및 적용 범위

(목적) AI 시스템의 설계·개발·이용 등에 관한 공통 기준을 설정하여 인공지능 시스템의 통상·무역을 규제하고, 심각한 피해를 야기할 수 있는 인공지능 시스템의 특정 행위를 금지

(규제 대상) AI 시스템의 설계·개발·활용 목적으로 인간 활동과 관련한 데이터를 처리하는 경우, AI 시스템의 설계·개발·활용·관리

(적용 예외) 주(州) 정부 기관, 정부 소유 기업 및 자회사, 국방부 장관, 정보국장 및 통신안보국장의 지시·관리하에 있는 상품·서비스·활동, 연방·주 정부 기관의 책임

자와 총독이 법령으로 정하는 자 또는 기관

AI 시스템 책임자 등의 책무

(익명화된 데이터 관리) 익명화된 데이터를 처리·제공하는 자는 데이터 익명화 방법 및 익명화된 데이터의 이용·관리에 관한 조치를 수립해야 함(제6조)

(고위험 시스템 평가) AI 시스템 책임자는 법령에 따라 해당 시스템이 고위험 시스템인지 여부를 평가해야 함(제7조)

(위험관련 조치) 고위험 시스템 책임자는 시스템 이용에 따라 발생할 수 있는 피해의 위험성 또는 편향된 결과를 식별·평가하고 위험성을 감경하기 위한 조치를 취해야 함(제8조)

(감경조치 이행 감독) 고위험 시스템 책임자는 위험성 감경조치의 수립·이행 및 감경조치의 실효성을 감독하기 위한 조치를 취해야 함(제9조)

민간분야 AI 시스템 규제 조항으로는 다음과 같은 조항을 둠

(기록 작성 및 보관) 익명화된 데이터 관리 조치, 위험 관련 조치, 고위험 시스템 평가 등에 관한 사항을 기록하여 보관하여야 함(제10조)

(고위험 시스템에 대한 설명) 고위험 시스템 제공자 및 운영자는 시스템 사용 방법, 도출하고자 하는 결과의 유형, 결정, 권고 또는 예측, 고위험 시스템에 대한 위험성 감경조치 등 시스템에 관한 설명을 웹사이트에 게재하여야 함(제11조)

(중대한 피해 고지) 고위험 시스템 책임자는 해당 시스템 사용으로 인해 중대한 피해가 발생하거나 발생할 가능성 있는 경우 신속하게 장관에게 고지해야 함(제12조)

(기록 제공) 장관은 기록관리자에게 기록의 제공을 명령할 수 있음(제13조)

(감사) 장관은 익명화된 데이터 관련 조치, 고위험 시스템 평가 등과 관련한 위반행위가 있다고 믿을 만한 합리적 근거가 있는 경우 해당 위반자에게 감사를 수행하게 하거나 독립 감사인이 감사에 참여하게 할 것을 명령할 수 있음(제15조~제16조)

처벌

(벌금) 제6조 ~ 제12조 위반자 및 장관에게 허위 또는 오해 소지가 있는 정보를 제공한 자에게 벌금 부과

유죄 확정시 최고 10,000,000달러의 벌금을 부과하고, 개인이 아닌 경우에는 유죄로 확정된 해의 전년도 총 수익의 3% 벌금을 병과

약식기소 되어 유죄판결을 받은 경우 최고 5,000,000달러의 벌금을 부과하고, 개인이 아닌 경우 유죄로 확정된 해의 전년도 총 수익의 2% 벌금을 병과

(형사처벌) 누구든지 AI 시스템의 설계·개발·이용 목적으로 법을 위반하여 개인정보를 처리하거나, 고의 또는 과실로 AI 시스템 이용·제공하여 개인에게 신체적·정신적·재산적 피해를 야기한 경우 다음과 같이 처벌

(정식재판) 개인이 아닌 경우에는 최고 25,000,000달러의 벌금과 유죄로 확정된 해의 전년도 총 수익의 5% 벌금을 병과, 개인의 경우 법원의 재량에 따라 벌금을 부과하거나 최고 5년의 징역에 처함(병과 가능)

(약식기소) 개인이 아닌 경우에는 최고 20,000,000달러의 벌금과 유죄로 확정된 해의 전년도 총 수익의 4% 벌금을 병과, 개인의 경우 법원의 재량에 따라 최고 100,000달러의 벌금을 부과하거나 최고 2년의 징역에 처함(병과 가능)

2 캐나다의 보건의료 분야 인공지능 관련 가이드라인

가. 온타리오주의 Artificial Intelligence in Health Care

온타리오주가 의료 분야에서 AI를 책임감 있게 개발하고 활용하기 위한 지침으로, 이 문서는 환자 중심 접근 방식을 강조하며, AI가 의료 서비스의 질을 향상시키고 효율성을 높이는 데 기여하기 위한 권고사항을 제시하고 있다.

1) 보건의료분야에서의 AI의 잠재적 이점

- ① 진단 및 치료 개선 : AI는 의료 영상 분석, 병리 슬라이드 검토, 유전체 데이터 분석 등을 통해 질병의 조기 진단과 정확한 치료에 기여
- ② 임상의 효율성 향상 : AI는 의료 기록 분석, 진료 예약 관리, 치료 계획 수립 등 반복적인 작업을 자동화하여 의료진의 업무 부담을 줄이고 환자 진료에 더 집중할 수 있도록 지원
- ③ 자원 할당 최적화 : AI는 환자 데이터 분석을 통해 입원 기간 예측, 병상 배정 최적화, 의료 자원의 효율적인 배분에 기여
- ④ 환자 경험 개선 : AI 기반 챗봇, 가상 비서 등을 통해 환자에게 맞춤형 정보와 지원을 제공하고, 환자 경험을 향상
- ⑤ 공중 보건 증진 : AI는 전염병 예측, 공중 보건 개입 효과 분석 등을 통해 공중 보건 증진에 기여

2) AI 관련 과제 및 위험

- ① 데이터 품질 및 편향: AI 모델은 훈련 데이터의 품질에 크게 의존하기 때문에 편향된 데이터로 훈련된 AI 모델은 특정 집단에 불리한 결과를 초래할 수 있으며, 이를 방지하기 위해 데이터 품질 관리 및 편향 완화 노력이 필요
- ② 개인정보 보호 및 보안 : 보건의료 데이터는 민감한 개인정보를 포함하고 있기 때문에 AI 시스템은 개인정보 보호 및 보안 규정을 준수해야 하며, 데이터 익명화, 접근 제어, 암호화 등의 기술적 및 관리적 조치를 통해 데이터 보안을 강화해야 함
- ③ 투명성 및 설명 가능성 : AI 알고리즘의 복잡성과 불투명성은 의사결정 과정에 대한 이해를 어렵게 만들 수 있기 때문에 설명 가능한 AI (XAI) 기술 개발을 통해 AI 시스템의 의사결정 과정을 투명하게 공개하고, 의료진과 환자의 신뢰를 확보해야 함
- ④ 책임 소재 : AI 시스템의 오류로 인해 발생한 피해에 대한 의료진, AI 개발자, 의료기관 등 관련 주체들의 책임과 역할을 명확히 하고, 책임 소재에 대한 법적 및 윤리적 논의가 필요

⑤ 규제 및 거버넌스 : 보건의료 분야에서 AI 개발 및 활용에 대한 명확한 규제 지침과 거버넌스 프레임워크를 마련하여 규제 기관은 AI 기술의 발전 속도에 맞춰 규제를 유연하게 적용하고, 혁신을 저해하지 않으면서도 환자 안전을 보장해야 함

3) 핵심 권고 사항

이 문서는 온타리오주 정부, 의료기관, AI 개발자 등에게 다음과 같은 권고 사항을 제시하고 있다.

① AI 전략 개발 : 온타리오주 정부는 의료 분야에서 AI 개발 및 활용에 대한 명확한 비전과 목표를 설정하고, 이를 달성하기 위한 전략을 수립

② 데이터 인프라 구축: 고품질의 의료 데이터를 수집, 저장, 공유, 분석할 수 있는 안전하고 효율적인 데이터 인프라를 구축

③ AI 인력 양성: 의료 전문가, AI 개발자, 데이터 과학자 등 AI 관련 인력 양성을 위한 교육 및 훈련 프로그램을 개발하고 지원

④ 윤리적 AI 개발 및 활용: AI 개발 및 활용 과정에서 윤리적 원칙을 준수하고, 편견, 차별, 개인정보 침해 등의 문제를 해결하기 위한 노력

⑤ 공공-민간 파트너십 강화: AI 기술 개발 및 의료 분야 도입을 위해 정부, 의료기관, AI 개발 기업 간의 협력을 강화

제 6 절

싱가포르의 인공지능 윤리 관련 법제도

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 싱가포르의 AI 거버넌스 프레임워크

싱가포르의 개인정보보호위원회(PDPC, 정보통신미디어개발청 산하 개인정보보호 감독기구)는 2019년 1월 발표한 “AI 거버넌스 프레임워크(Model AI Governance Framework)”를 통해 인간 중심적이고 설명 가능하며 투명하고 공정한 의사결정이라는 원칙 하에 민간 부문의 AI 관리체계, 윤리 및 소비자 보호에 관한 지침을 제시하고 있다.

가. 프레임워크의 목표 및 범위

‘AI 거버넌스 프레임워크’는 정부 기관이 AI 시스템을 개발, 배포 및 운영할 때 고려해야 할 핵심 원칙과 실행 가능한 지침을 제공하고 있다. AI 시스템이 사회적 책임과 윤리적 가치를 준수하도록 개발 및 배포 과정을 안내하고, AI에 대한 신뢰 구축, AI 시스템의 투명성, 설명 가능성, 공정성을 확보하여 시민의 신뢰를 얻고 AI 기술의 수용성을 높이며, 싱가포르 정부가 AI 거버넌스 분야에서 세계적인 모범을 보이고, 국제적인 협력을 주도하기 위한 목적을 가지고 있다.

나. 프레임워크의 핵심 원칙

이 프레임워크는 다음과 같은 두 가지 핵심 원칙을 기반으로 한다.

① 인간 중심 : AI 시스템은 인간의 복지와 존엄성 증진하는 방식으로 설계 및 운영되어야 함

② 책임성 : AI 시스템의 개발 및 활용에 대한 책임 소재를 명확히 하고 잠재적 위험을 완화하기 위한 조치 필요

다. 프레임워크의 주요 적용 영역

프레임워크는 위에서 언급한 핵심 원칙을 실현하기 위해 다음과 같은 네 가지 영역에 대한 구체적인 지침을 제공하고 있다.

- ① 내부 거버넌스 : AI 시스템의 개발 및 배포에 대한 내부 감독 및 책임 구조를 명시
- ② 의사결정 : AI 시스템의 의사결정 과정을 투명하게 공개하고, 인간의 감독 및 개입을 보장
- ③ 운영 관리 : AI 시스템의 성능을 지속적으로 모니터링하고, 오류 발생 시 신속하게 대응
- ④ 이해관계자 관리 : AI 시스템의 개발 및 활용과 관련된 다양한 이해관계자와 소통하고 협력

라. 프레임워크의 실행 지침

프레임워크는 각 영역에 대한 몇 가지 주요 지침을 제공하고 있다.

- ① AI 윤리 위원회 설립 : AI 시스템의 윤리적 측면을 검토하고 자문하는 AI 윤리 위원회를 설립
- ② 데이터 거버넌스 : AI 모델 훈련에 사용되는 데이터의 품질을 관리하고, 데이터 편향을 최소화하기 위한 노력
- ③ 알고리즘 투명성: AI 알고리즘의 작동 방식, 데이터 출처, 성능 지표 등을 투명하게 공개
- ④ 설명 가능한 AI: AI 시스템의 의사결정 과정을 이해하고 설명할 수 있도록 XAI 기술을 활용
- ⑤ 인적 감독: AI 시스템의 의사결정 과정에 인간의 감독 및 개입을 보장
- ⑥ 책임 소재 명확화: AI 시스템의 오류로 인해 발생한 피해에 대한 책임 소재를 명확히 규정
- ⑦ 이해관계자 참여: AI 시스템 개발 및 활용과 관련된 다양한 이해관계자의 의견을 수렴하고 협력

2 싱가포르 보건부의

Artificial Intelligence in Healthcare Guidelines

싱가포르는 의료 분야에서 AI 기술의 잠재력을 일찍부터 인지하고 적극적으로 활용 해왔다. 2019년 11월, 싱가포르 보건부(MOH)는 ‘Artificial Intelligence in Healthcare Guidelines, AIHG’를 발표하여 의료 AI 개발 및 활용에 대한 명확한 방향을 제시했다. 이 가이드라인은 싱가포르에서 AI 기반 의료 서비스가 윤리적이고 책임감 있게 제공될 수 있도록 지원하는 중요한 역할을 한다.

가. AIHG의 목표 및 범위

AIHG는 의료 분야에서 AI를 안전하게 개발, 구현 및 관리하기 위한 권장 사항과 우수 사례를 제공한다. 이 가이드라인은 주로 임상 AI 의료기기(AI-MD)에 중점을 두고 있지만, 교육, 웰니스, 행정 등 의료 환경에서 구현되는 모든 AI 애플리케이션에 적용될 수 있다. AIHG는 개발자와 의료 서비스 제공자 모두를 위한 지침이며, 싱가포르 보건 과학청(HSA)의 소프트웨어 의료기기에 대한 규제 지침을 보완하고 있다. AIHG는 규제 요건은 아니지만, 우수 사례 원칙을 제공하며 개인정보보호위원회(PDPC)의 ‘전 정부 모델 AI 거버넌스 프레임워크’ 및 싱가포르 통화청(MAS)의 ‘공정성, 설명 가능성, 책임, 투명성(FEAT)’ 원칙과 같은 기존 국가적 입장을 고려하고 있다.

나. AIHG의 핵심 원칙

- ① 환자 안전 및 복지: AI 시스템은 환자의 안전과 복지를 최우선으로 고려하여 설계 및 운영되어야 하며 잠재적 위험을 최소화하고 이점을 극대화하기 위한 조치를 취해야 함
- ② 임상적 책임 : AI는 의료 전문가의 의사결정을 지원하는 도구일 뿐, 최종적인 진단 및 치료 결정은 의료진의 책임하에 이루어져야 하며, AI가 의료진의 전문적인 판단을 대체하는 것은 불가능함
- ③ 투명성 및 설명 가능성 : AI 알고리즘의 작동 방식, 데이터 출처, 성능 지표 등을 투명하게 공개해야 하며 AI 시스템의 의사결정 과정을 의료진과 환자에게 설명 가능하

도록 만들어야 함

④ 데이터 프라이버시 및 보안 : 환자의 개인정보를 보호하고 데이터 보안을 유지하는 것은 AI 시스템 개발 및 운영의 필수적인 요소로, 데이터 익명화, 접근 제어, 암호화 등의 기술적 및 관리적 조치를 통해 데이터 보안을 강화해야 함

⑤ 공정성 및 형평성: AI 시스템은 모든 환자에게 공정하고 형평성 있는 의료 서비스를 제공해야 하며 AI 모델 훈련에 사용되는 데이터의 편향을 최소화하고, AI 시스템이 특정 집단에 불리한 결과를 초래하지 않도록 주의

다. AIHG의 주요 지침

AIHG는 위에서 언급한 핵심 원칙을 실현하기 위해 다음과 같은 구체적인 지침을 제공하고 있다.

① 위험 관리: AI 시스템 개발 및 구현 과정에서 발생할 수 있는 잠재적 위험을 식별하고 평가하여 위험을 완화하기 위한 조치 필요

② 데이터 관리: AI 모델 훈련에 사용되는 데이터의 품질을 관리하고, 데이터 편향을 최소화하기 위한 노력 필요

③ 알고리즘 개발 및 평가: AI 알고리즘을 개발할 때 투명성, 설명 가능성, 공정성을 고려해야 하며, 성능을 평가하고 검증하기 위한 적절한 조치 필요

④ 임상 검증: AI 시스템을 임상 환경에 도입하기 전에 충분한 임상시험을 통해 안전성과 유효성을 검증함

⑤ 모니터링 및 감사 : AI 시스템의 성능을 지속적으로 모니터링하고 감사하여 문제 발생 시 신속하게 대응

제 7 절

국내 인공지능 관련 법제도

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 과학기술정보통신부의

“사람이 중심이 되는 국가 인공지능 윤리기준”

「인공지능 윤리기준」 제정은 「인공지능 국가전략(19.12)」의 주요 과제 중 하나로, 2020년 12월 23일, 과학기술정보통신부(이하 ‘과기정통부’)와 정보통신정책연구원은 대통령 직속 4차산업혁명위원회 전체회의에서 인공지능 시대 바람직한 인공지능 개발·활용 방향을 제시하기 위한 “사람이 중심이 되는 「인공지능(AI) 윤리기준”을 마련하였다. 이 윤리기준은, 윤리적 AI를 실현하기 위해 정부·공공기관, 기업, 이용자 등 모든 사회구성원이 인공지능의 개발에서 활용에 이르는 전 단계에서 함께 지켜야 할 주요 원칙과 핵심 요건을 제시하는 기준이라고 볼 수 있다.

본 기준을 설정하기 위해 2020년 4월부터 과기정통부는 AI·윤리 전문가로 구성된 ‘인공지능 윤리연구반’을 통해 국내외 주요 AI 윤리원칙을 분석하고, 그 결과를 윤리철학의 이론적 논의와 연계하여 ‘인간성을 위한 인공지능(AI for Humanity)’을 목표로 하는 윤리기준 초안을 마련하였고, 이후 3개월에 걸쳐 학계·기업·시민단체 등 각계의 다양한 전문가로부터 AI·윤리학·법학 등의 의견을 수렴하였다. 2020년 11월 27일 초안이 발표되었고, 2020년 12월7일 공개 공청회 등 시민 의견수렴을 거쳐 완성되었다.

AI 윤리기준은 ① 사회구성원 모두가 ② 모든 분야에서 ③ 자율적으로 준수하며 ④ 지속 발전하는 윤리기준을 지향하며 AI 개발에서 활용에 이르는 전 단계에서 정부·공공기관, 기업, 이용자 등 모든 사회구성원이 참조하는 윤리기준이 되어야 함을 목표 및 지향점으로 한다. 본 기준은 특정 분야에 제한되지 않는 범용성을 가진 일반원칙으로, 이후 각 영역별 세부 규범이 유연하게 발전해 나갈 수 있는 기반을 조성하는 윤리기준이 되어야 할 것을 명시하였다. 윤리기준은 구속력 있는 ‘법’이나 ‘지침’이 아닌 도덕적 규범이자 자율 규범으로, 기업의 자율성을 존중하고 AI 기술 발전을 장려하며 기술과

사회변화에 유연하게 대처할 수 있는 윤리 담론을 형성하는데 기여하도록 하였다. 아울러 사회경제, 기술 변화에 따라 새롭게 제기되는 AI 윤리 이슈를 논의하고 구체적으로 발전시킬 수 있는 플랫폼으로 기능하도록 하였다.

본 보고서는 윤리기준이 지향하는 최고 가치를 ‘인간성(Humanity)’으로 설정하고, ‘인간성을 위한 인공지능(AI for Humanity)’을 위한 3대 원칙·10대 요건을 제시하였다.

(3대 기본원칙) ‘인간성(Humanity)’을 구현하기 위해 AI의 개발 및 활용 과정에서 ① 인간의 존엄성 원칙, ② 사회의 공공선 원칙, ③ 기술의 합목적성 원칙을 지켜야 한다.

(10대 핵심요건) 3대 기본원칙을 실천하고 이행할 수 있도록 AI 개발~활용의 전 과정에서 ① 인권 보장, ② 프라이버시 보호, ③ 다양성 존중, ④ 침해금지, ⑤ 공공성, ⑥ 연대성, ⑦ 데이터 관리, ⑧ 책임성, ⑨ 안전성, ⑩ 투명성의 요건을 제시하였다.

2 한국지능정보사회진흥원의 지능정보사회 윤리 가이드라인

지능정보사회 윤리 가이드라인은 AI과 관련된 기술의 개발, 활용, 확산 과정에서 윤리적 문제를 해결하고, 사회적 신뢰와 책임 있는 기술 활용을 촉진하기 위해 한국지능정보사회진흥원에서 마련한 문서다. 이 가이드라인은 다양한 이해관계자(개발자, 사용자, 정책입안자 등)가 지켜야 할 기본적인 윤리적 원칙을 아래와 같이 제시하고 있다.

구분	공공성(P) (Publicness)	책임성(A) (Accountability)	통제성(C) (Controllability)	투명성(T) (Transparency)
정의	지능정보기술에 의해 창출된 다양한 혜택은 모든 인류를 위해 광범위하게 공유되어야 한다.	사회변혁에 따른 책임성이 사회 구성원 모두에게 따라야 하며, 확실한 책임분배와 사회적 의무 이행이 이루어져야 한다.	인공지능을 안전하게 관리하고 적절하게 제어할 수 있는 통제성을 갖추어야 한다.	인공지능은 외부가 쉽게 판단하기 힘든 고도의 지능을 갖추고 있으므로 관련 영역 전반에 걸쳐 투명성이 제공되어야 한다.

구분	공공성(P) (Publicness)	책임성(A) (Accountability)	통제성(C) (Controllability)	투명성(T) (Transparency)
개발자	기술 개발 시 사회적 차별요소를 배제해야 하며, 사회적 약자 보호를 위한 접근성을 보장해야 한다.	개발부터 이용에 이르는 전 과정과 결과에 대해 책임을 공유하고, 정보 교류 및 기술경신에 지속적으로 참여한다.	예외적인 발생에 대한 종합적 검토를 하며, 기술적 제어 장치 등을 마련해야 한다.	위급상황 발생시, 필요 데이터를 공유하며 해결방안 마련에 협력하며 은닉 기술을 개발하지 않는다.
공급자	공공의 이익에 부합하는 제품을 공급해야 하며, 선한 의도를 가지고 발주해야한다.	제품의 보급 및 확산, 이용의 결과로 발생하는 사회적 폐해에 대한 책임을 공유하며, 사고 발생에 따른 보상 원칙 등을 확립한다.	제품 유통 과정에서의 위험 요소를 파악하며, 안전성 검증 및 조치를 마련해야 한다.	위험 관련 정보를 이용자와 공유하며, 이용자 대상 제품교육과 매뉴얼 등을 제공한다.
이용자	공공이익을 위한 제품 개선에 참여하며, 소비자 행동원칙 준수를 일상화해야 한다.	제품 이용 결과에 따른 윤리 책임을 숙지하고, 안내된 지침을 충실히 따른다.	이용하는 제품과 서비스에 대해 정한 이용 안내 정보를 습득한다.	특정 서비스 과정에 대한 설명을 개발자나 공급자에게 요구할 권리를 인지하고, 이용에 따른 이상 현상과 부작용 등에 대해 습득한 정보를 공유한다.

3 대한의료인공지능학회의 의료 인공지능 윤리 가이드라인

현재 정부 차원에서의 보건의료분야 AI에 대한 포괄적인 가이드라인은 부재한 상태다. 대한의료인공지능학회는 보건의료분야에서 책임감 있는 보건의료 AI의 개발과 활용을 위해 2021년 12월 ‘의료 인공지능 윤리 가이드라인’을 발표했다. 이 가이드라인은 의료 AI 개발 및 활용 과정에서 지켜야 할 윤리적 원칙과 실질적인 지침을 제시하고 있다.

가. 가이드라인의 목표 및 범위

대한의료인공지능학회의 ‘의료 인공지능 윤리 가이드라인’은 의료 AI 개발자, 의료 전문가, 정책 입안자 등 모든 이해관계자가 AI 기술을 윤리적으로 개발하고 활용하는데 필요한 지침을 제공하여 의료 AI의 잠재적 이점을 극대화하면서도 환자 안전, 개인 정보 보호, 사회적 형평성 등을 보장하는 것을 목표로 한다.

나. 가이드라인의 핵심 원칙

이 가이드라인은 다음과 같은 6가지 핵심 원칙을 기반으로 한다.

① 인간 존엄성 존중: 환자의 자율성을 존중하고, 인간의 건강과 복지를 최우선으로 고려해야 하며 AI는 인간의 의사결정을 지원하는 도구이며, 인간을 대체하는 것은 불가함

② 사회적 책임: 의료 AI는 사회적 불평등을 심화시키지 않고, 오히려 건강 형평성을 증진하는 데 기여해야 함

③ 안전성 및 유효성 확보: 과학적 근거에 기반하여 안전하고 유효한 AI 기술을 개발하고 활용해야 하며 충분한 검증을 통해 AI 시스템의 안전성과 유효성을 확보하고, 지속적인 모니터링을 통해 잠재적 위험 예방 필요

④ 투명성 및 설명 가능성: AI 알고리즘의 작동 방식, 데이터 출처, 성능 지표 등을 투명하게 공개하고, 의료진과 환자가 이해하기 쉬운 설명이 필요

⑤ 개인정보 보호 및 보안: 환자의 개인정보를 보호하고 데이터 보안을 유지하는 것은 AI 시스템 개발 및 운영의 필수적인 요소이며, 데이터 익명화, 접근 제어, 암호화 등의 기술적 및 관리적 조치를 통해 데이터 보안을 강화할 필요

⑥ 책임 있는 사용: 의료 AI는 의료 전문가의 판단을 지원하는 도구로 사용되어야 하며, 의료진의 책임하에 활용되어야 하며, AI 시스템의 오류 또는 오작동으로 인해 발생하는 문제에 대한 책임 소재를 명확히 할 필요

다. 가이드라인의 주요 지침

위에서 언급한 핵심 원칙을 실현하기 위해 다음과 같은 구체적인 지침을 제공한다.

① 데이터 관리 : AI 모델 훈련에 사용되는 데이터의 품질을 관리하고, 데이터 편향을 최소화하기 위한 노력을 통해 데이터 출처, 수집 방법, 라벨링 과정 등을 투명하게 기록하고, 데이터의 다양성을 확보하여 편향을 최소화

② 알고리즘 개발 : AI 알고리즘을 개발할 때 투명성, 설명 가능성, 공정성을 고려해야 하며, 성능을 평가하고 검증하기 위한 적절한 조치가 필요

- ③ 임상 검증 : AI 시스템을 임상 환경에 도입하기 전에 충분한 임상 시험을 통해 안전성과 유효성을 검증
- ④ 환자와의 소통 : 의료 AI를 사용하는 경우, 환자에게 AI 시스템의 역할과 한계에 대해 충분히 설명하고, 환자의 동의 구득
- ⑤ 모니터링 및 평가 : AI 시스템의 성능을 지속적으로 모니터링하고 평가하여 문제 발생 시 신속하게 대응
- ⑥ 교육 : 의료진, AI 개발자 등을 대상으로 의료 AI 윤리에 대한 교육을 제공하여 AI 기술에 대한 이해와 윤리적 감수성 향상

제5장

보건의료 분야 인공지능 관련 윤리적 과제 해결 및 활용 확대 방안

제1절 개인정보 보호

제2절 투명성과 설명가능성

제3절 공정성과 차별금지

제4절 안전 및 책임성

제5절 인간의 기술통제

제6절 효율성의 평가: 국내 보건의료 지불보상제도 개선방안

본 장에서는 향후 보건의료 분야에서 AI이 활용될 경우 윤리의 관점에서의 논의점을 소개하고자 한다. 윤리적 차원의 논의뿐만 아니라 이러한 논의점을 해소하려는 기술개발 측면의 진전뿐만 아니라 보건의료서비스 전달체계도 함께 다루고자 한다. 보건의료 분야 AI 맥락에서의 윤리의 확립 및 활용 확대와 관련하여 다음과 같은 여섯 가지 논의 요소로 나눌 수 있다.

1. 개인정보 보호
2. 투명성과 설명가능성
3. 공정성과 차별금지
4. 안전 및 책임성
5. 인간의 기술통제
6. 효율성의 평가: 국내 보건의료 지불보상제도 개선방안

제 1절 개인정보 보호

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 한국의 보건의료빅데이터

AI는 방대한 의료 데이터를 학습함으로써 의료 분야에서 뛰어난 효율성을 보인다. 수백만, 수천만 건의 데이터를 통해 학습된 AI는 진단의 정확성을 높이고 치료 효과를 향상시키는 등 의료 서비스 개선에 크게 기여할 수 있다. 이처럼 AI의 성능은 데이터 학습에 크게 의존하기 때문에 많은 국가들이 빅데이터 구축에 힘쓰고 있다. 특히, 전 국민 건강보험 시스템과 전자의무기록(EHR) 시스템을 갖춘 한국은 의료 빅데이터 구축에 유리한 환경을 가지고 있다.

한국의 국민건강보험공단은 전 국민을 대상으로 개인 신상 정보, 진료 기록, 건강검진 결과 등 방대한 데이터를 보유하고 있어 AI 학습에 활용될 수 있는 빅데이터를 이미 상당 부분 구축했다고 볼 수 있다. 공단은 전 국민의 보험가입절차에서 주민등록표본에 기재된 정보(성명, 성별, 나이, 생년월일, 주민등록번호 등)를 수집한다. 아울러 보험료의 산정 및 부과를 위하여 거주지, 직장, 직장 수입, 연금 등 소득, 재산(자동차 포함)에 관한 정보 등을 수집하며, 요양급여를 실시하기 위해 보험가입자가 의료서비스를 받은 의료기관으로부터 환자의 내원 일자, 진단명, 치료 및 의약품 등에 관한 정보를 수집한다. 아울러 국민건강보험공단은 국민들에게 주기적으로 일반건강검진, 암검진, 영유아 건강검진, 학교 밖 청소년 건강검진 등을 실시하여 국민의 건강을 증진하는 한편 그 결과를 모두 취합한다.

또한, 병원에서 운영하는 전자건강기록(EHR) 시스템에는 환자의 상세한 진료 정보가 기록되어 있기 때문에 각 병원은 이러한 데이터를 활용하여 특정 질병이나 치료법에 대한 AI 모델을 개발할 수 있으며, 병원 간 데이터 공유 및 협력을 통해 더욱 풍부하고 다양한 빅데이터를 구축할 수 있다.

그러나, 한편으로는 보건의료 AI는 민감정보인 건강 데이터를 이용하기 때문에 개인 정보 유출 등의 문제가 필연적으로 제기된다.

2 빅데이터 활용의 제도적 한계

한국은 풍부한 의료 데이터를 보유하고 있지만, 빅데이터의 AI 활용에 있어 제도적 제약이 존재한다. 2011년 제정된 개인정보보호법은 개인정보 처리에 대한 엄격한 규제를 통해 개인의 자기정보결정권을 보장하는 데 중점을 두고 있다. 특히, 건강 정보와 같은 민감정보는 정보 주체의 사전 동의를 필수로 요구한다. 하지만 빅데이터는 방대한 양과 빠른 수집 속도를 특징으로 하기 때문에 모든 정보 주체로부터 사전 동의를 얻는 것은 현실적으로 어려우며, 이는 막대한 비용과 시간을 필요로 할 뿐만 아니라 빅데이터의 효율적인 활용을 저해하는 요인이 된다.

입법부는 이러한 문제점을 인지하여 2020년 소위 데이터 3법인 개인정보 보호법, 정

보통신망 이용촉진 및 정보보호 등에 관한 법률, 신용정보의 이용 및 보호에 관한 법률을 개정하였다. 핵심적인 변화로는 가명정보의 도입으로, 가명처리 등의 기법을 도입하고, 가명정보 활용 범위를 확대했다. 구체적으로, 과학적 연구, 통계 작성, 공익적 기록 보존 등의 목적으로 가명정보를 정보 주체의 동의 없이 활용할 수 있도록 허용한 한편, 개인정보처리자의 책임성을 강화하고 안전한 개인정보 보호를 위한 제도적 장치를 마련하여 개인정보 오용 및 유출을 방지하고자 하였다.

정부는 데이터 3법 개정을 기반으로 2020년부터 마이데이터(MYDATA) 정책을 시행하였다. 마이데이터란, 개인이 자신에 관한 데이터를 직접 관리하고, 이를 통해 다양한 맞춤형 서비스를 제공받을 수 있도록 하는 정책으로, 개인이 다양한 분야에서 수집된 데이터를 자발적으로 제공하고, 이를 기반으로 맞춤형 서비스나 혜택을 받는 시스템이다. 예를 들어, 헬스케어 분야에서는 개인의 건강 데이터를 기반으로 맞춤형 건강 관리 서비스를 제공하는 것이 가능하도록 하였다.

이 정책 하에서는 개인정보 처리 과정에서 개인의 동의를 기반으로 데이터를 수집하며, 데이터를 제3자에게 제공할 때는 엄격한 규제를 따르도록 하였다. 이를 통해 개인정보의 안전성과 투명성을 확보하고자 하였다.

그러나, 마이데이터 서비스의 확장을 위한 법안이 국회 법제사법위원회에 계류되면서, 관련 제도와 서비스가 계획대로 진행되지 못할 가능성이 크다. 이는 민간 사업자들에게도 큰 영향을 미치고 있으며, 정책 시행의 불확실성을 초래하고 있다. 이밖에도 데이터 전송 시의 표준화와 보안 문제, 데이터의 품질 관리, 보안 문제, 그리고 기술적 신뢰성 등이 여전히 걸림돌로 작용하고 있다.

3 시를 활용한 가명정보의 활용 시 문제점

가명정보는 특정 개인을 식별할 수 없도록 일부 정보를 삭제하거나 대체하여 처리한 데이터이다. 그러나, 가명정보도 개인정보의 범주에 포함되는데, 가명정보는 완전히 익명화된 정보(익명정보)와 달리, 추가 정보를 결합하거나 복원할 경우 개인을 식별할 가능성이 있기 때문이다. 예를 들어, 병원의 AI 시스템이 환자의 개인정보를 가명처리

하여 연구용 데이터셋을 구축했다고 하더라도 AI가 여러 데이터를 결합하여 개인을 재식별할 수 있는 가능성이 존재한다. 국내 개인정보보호법에서는 “통계 작성, 과학적 연구, 공익적 기록 보존 등의 목적”의 경우 원칙적으로 정보 주체의 동의 없이도 활용이 가능하다.

이에 데이터 3법 개정 이후 보건복지부와 개인정보보호위원회가 합동으로 「보건의료 데이터 활용 가이드라인」을 발표해 보건의료분야의 빅데이터 활용에 대한 구체적 기준을 제시하고자 하였으나 국가인권위원회는 가명정보를 결합, 활용하는 과정에서 개인을 재식별화할 가능성을 이유로 여전히 데이터 제공자의 인권 보호에는 미흡함을 지적하였다(국가인권위원회, 2020).

AI가 대규모 개인정보를 처리하는 과정에서도 개인정보 유출 사고가 발생할 수 있다. 예를 들어, AI 기반의 고객 서비스 시스템이 고객의 개인정보를 저장하고 처리하는 과정에서 시스템 상의 보안 결함으로 해커가 데이터를 탈취한다면 개인정보 유출 사고가 발생할 수 있다. 이러한 사고는 개인의 프라이버시를 침해하고, 법적 책임을 물을 수 있는 사유가 된다.

AI가 처리하는 데이터가 개인정보보호법에 어떻게 적용될 수 있는지에 대한 논란도 있다. 예를 들어, AI가 소셜 미디어에서 사람들의 행동을 분석하여 맞춤형 광고를 제공한다고 할 때, 사용자가 명시적으로 동의를 했는지 여부가 중요하다. 그러나 사용자는 자신의 데이터가 어떻게 사용되는지에 대해 충분히 인식하지 못할 수 있으며, 이로 인해 개인정보의 무단 사용이나 침해가 발생할 수 있다.

4 개선방안

AI가 보건의료 분야에서 활용될 때 발생할 수 있는 사생활 침해 문제를 해결하기 위해 AI 개발의 모든 단계에서 개인정보 보호 원칙을 준수할 필요가 있다(UNESCO, 2021). 이를 위해서는 AI 개발 전 과정에 걸쳐 사생활 보호 원칙을 준수하도록 규제 및 거버넌스 체계를 구축해야 한다. 특히, 의료 AI 개발 과정에서 다량의 데이터를 연계 및 활용할 때 발생할 수 있는 재식별 위험을 지속적으로 관리하고 예방할 필요가 있다. 이를 위해

서는 법적 규제뿐만 아니라 데이터 보호를 위한 모니터링 구축을 포함한 거버넌스 체계 마련이 필요하다. 또한 개인정보 보호와 AI를 통한 보건의료 기술의 발전을 통한 건강 증진 모두를 합리적으로 추구하기 위해서는 AI의 정확성 및 유용성과 사생활 보호 사이의 균형점을 찾기 위한 사회적 합의가 이루어질 필요가 있다.

제2절

투명성과 설명가능성

NATIONAL ASSEMBLY FUTURES INSTITUTE

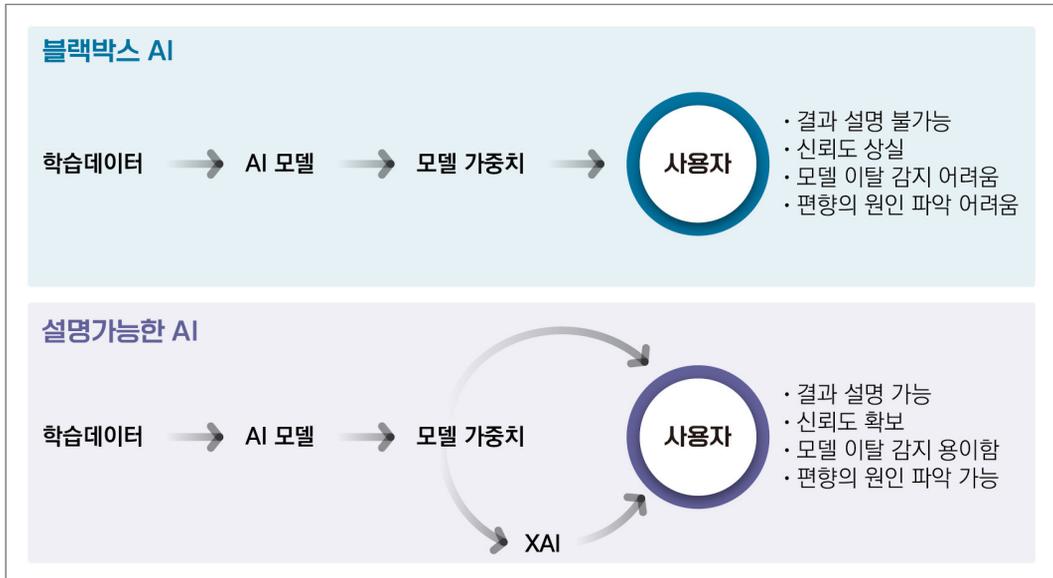
1 인공지능에 대한 설명의 의의

적절한 설명 없이 최종 결과값만 도출하거나, 과정 설명이 복잡하고 불투명한 AI 알고리즘은 ‘블랙박스(blackbox)’ 라고 불린다. 최근 AI 알고리즘은 점점 더 복잡해지고 계산 과정을 추적하기 어려워짐에 따라, 블랙박스 AI에 대한 문제 제기가 계속되고 있다. 블랙박스 AI는 설명가능성(explainability)이 낮다고 표현되며, 이러한 AI시스템은 개발자와 사용자 모두에게 신뢰받기 어렵기 때문에 널리 사용되기 힘들다.

현재는 현실적으로 AI의 설명이 어렵다는 견해가 우세하다. AI의 설명가능성 논의를 촉발한 미국 국방고등연구계획국(DARPA)의 보고서에서는 두 가지 목표를 제시했다. 첫째는 기존 AI를 설명가능하게 만드는 것이며, 둘째는 처음부터 설명가능한 AI를 새롭게 설계하는 것으로, 기존 AI의 설명이 어려운 상황임을 감안해 두 번째 목표가 설정되었다(DARPA, 2020).

AI 연구 동향을 보면, AI를 완전히 설명하기 어려운 사례가 많다. 연구 결과들은 인간이 AI의 작동원리를 이해할 수 있는 경우와, AI의 의사결정 결과에 대한 제한적 해석이나 설명을 제공할 수 있는 정도로만 설명할 수 있는 경우로 크게 나눌 수 있다. 이는 AI의 개발자조차 완전한 설명을 제공하기 어려운 경우가 적지 않다는 점을 보여준다.

특히 높은 정확도를 자랑하는 딥러닝 방식의 AI는 개발자가 완전히 설명하기 어려운 경우가 많다. 딥러닝은 기본 알고리즘을 바탕으로 주어진 데이터를 학습하며 스스로 규칙을 구체화해 정확도를 높이는 방식이다. 예를 들어 개와 고양이를 구별하는 알고리즘은 수많은 사진을 학습하며 설계자가 처음에 제시하지 않은 촬영 장소 같은 추가 정보를 활용해 개와 고양이를 구별할 수 있다. 이때 설계자는 기본 공식은 알지만, 학습 과정에서 AI가 귀납적으로 추가하거나 구체화한 변수를 파악하기 어렵다.



출처: 저자 구성

[그림 5-1] 블랙박스 AI vs. 설명가능한 AI

그럼에도 불구하고 그 작동방식이나 판단 과정을 부분적으로 이해하는 것은 가능하며 기술적으로도 설명하려는 시도들이 지속되고 있다. 예를 들어, 사용된 모델 유형이나 학습에 이용된 데이터를 통해 작동방식을 대략적으로 파악할 수 있으며, 입력과 출력값을 분석함으로써 어느 정도의 이해가 가능하다.

이용자의 측면에서 보면, 이용자들의 신뢰를 얻는 것이 기술 성공의 중요한 요소로 여겨진다. 그렇다고 반드시 AI가 사용되는 모든 맥락에서 완전한 설명가능성이 필요한 것은 아니다. 예를 들어 자율주행차의 작동이 일일이 설명될 필요는 없고, 과도한 설명은 오히려 소비자 편익을 감소시킬 수 있다. 설명이 요구되는 경우는 보통 안전성에 대한 신뢰가 부족할 때나 사고와 같은 문제가 발생해 원인 규명이 필요할 때이다. 이러한 경우, 소비자가 AI의 한계를 인식할 때 그에 대한 신뢰는 줄어들고, 설명가능성 요구가 증가한다. 특히 소비자가 불이익을 겪을 가능성이 있는 경우, 설명가능성에 대한 요구가 더욱 커질 수 있다.

2 설명가능한 인공지능 관련 기술

설명가능한 AI(explainable AI; XAI)란 특정 AI 기술이 최종 의사결정과 결과물에 도달하는 과정을 분명하고 이해할 수 있도록 설명하는 방법 또는 프로세스다(Molnar, C, 2022). 블랙박스 AI와 달리, XAI를 활용하면 사용자가 AI 시스템 내부의 알고리즘을 신뢰할 수 있으며, 결과값 산출 과정의 타당성을 검토하고, 어떻게 활용할지 주체적으로 판단할 수 있도록 한다. 최근 AI시스템을 개인 사용자가 직접 적용하는 사례가 많아지면서 XAI의 중요성은 더욱 강조되고 있다.

XAI에 있어 해석가능성(interpretability)과 설명가능성은 흔하게 상호호환적으로 사용되는 개념이지만 구별할 수 있다. 해석가능성은 AI 모델이 “어떻게 작동하는지”를 직관적으로 이해할 수 있는 능력에 초점을 맞추며, 일반적으로 간단한 모델을 적용할 때 활용되는 개념이다. 이에 비해 설명가능성은 복잡한 AI 모델의 “결과가 왜 그렇게 되었는지”를 설명하는 데 중점을 둔다. 이를 위해 ‘모델 불특정법(Model agnostic methods)’(예시: partial dependence plots, Shapley Additive exPlanations (SHAP) dependence plots, surrogate models 등)과 같은 추가적인 분석 기법이 필요하다.

이 중 투명성은 설명가능성이 사회적으로 공정하고 지속 가능한 AI 시스템을 책임감 있게 구축하는 데 핵심적인 요소이다(Molnar, C, 2022). 일반적으로 AI 투명성은 개발자, 최종 사용자, 전문가 및 비전문가 모두가 AI 시스템 내부에서 어떠한 과정으로 의사결정이 이루어지는지 알 수 있는 상태로 정의된다. 투명성 있는 AI는 적절한 정보를 개방적이고 명확하게 제공하여 사용자의 알 권리를 존중하고 신뢰성을 확보한다. AI의 해석가능성은 알고리즘의 본질적인 명확성에 초점을 맞추어 정의 내릴 수 있다. AI가 어떤 과정을 거치는지 투명하게 제공하는 것을 넘어, 알고리즘의 개발자와 포함한 전문가가 명확한 의미를 해석할 수 있어야만, 알고리즘에 문제가 생겼을 때 이를 개선하거나 성능을 고도화시킬 수 있다. AI의 설명가능성은 사용자에게 AI 시스템의 의사결정 방식에 대한 설명을 할 수 있는 능력을 뜻하며, 전문 지식이 부족한 사용자도 AI 시스템 작동 방식을 이해하도록 도와준다.

XAI를 통해, 설명가능성, 해석가능성, 그리고 투명성의 원칙을 기반으로 AI 시스템을 구축할 수 있다. 생산자는 개발 과정에서 편향과 결함을 줄여 AI 시스템에 대한 신뢰를

구축한 후, 책임감 있게 배포할 수 있다. 또한, 최종사용자는 AI시스템의 의사결정 과정에 대해 충분히 이해하고 주체적으로 AI시스템의 결론을 활용할 수 있다.

설명가능성을 보장하는 가장 직관적인 방법은 내재적 설명(intrinsic explanation)으로, 모델 자체가 해석 가능한 경우를 의미한다. 선형 회귀, 로지스틱 회귀, 그리고 의사결정 트리가 일반적으로 사용되는 해석 가능한 모델에 해당한다. 내재적 설명이 가능한 모델은 설명 방식이 간단하다는 이점이 있지만, 모델 구조가 비교적 단순하기 때문에 성능이 떨어질 수 있다. 반면, 딥러닝 모델과 같이 계산 과정이 비교적 복잡하여 모델을 직접 설명하기 어려운 경우에는 사후 설명(post-hoc explanation) 방법이 활용된다. 이러한 방식은 모델에서 설명이 분리되어 있어 모델 독립적 해석 방법(model-agnostic interpretation)이라고도 한다(Ribeiro, Singh & Guestrin, 2016). 모델 독립적 해석 방법은 모델과 분리되어 있기 때문에 다양한 모델에 적용할 수 있다는 큰 장점이 있다. 따라서 복잡한 모델의 향상된 성능이 필요한 경우, 모델 독립적 해석 방법이 유용할 수 있다.

모델 독립적 해석 방법은 전역적(global) 방법과 국소적(local) 방법으로 구분될 수 있다. 전역적 방법은 특징이 예측에 평균적으로 어떻게 영향을 미치는지를 설명하며, 데이터의 일반적인 매커니즘을 이해하거나 모델을 디버깅할 때 특히 유용하다. 대표적인 예로는 부분 의존도 그림(partial dependence plot)이 있는데 이 방법은 특정 변수를 한 단위 증가시킬 때 변화하는 결과값의 크기를 계산하는 방식이다(Friedman, 2001).

반면 국소적 방법은 개별 예측을 설명하는 데 중점을 둔다. 예를 들어, 국소적 대리 모델(local surrogate models)인 LIME(Locally Interpretable Model-agnostic Explanations)는 복잡한 모델을 국소적으로 해석 가능한 대리 모델로 대체하여 예측을 설명한다. 복잡한 딥러닝 연산을 역순으로 계산하여 결과값에 기여한 영향을 시각화하는 Grad-CAM, LRP, Saliency Map 등 또한 국소적 방법에 해당된다.

설명이라는 개념과 직접적인 연관은 적지만 AI가 처리한 데이터나 의사결정의 내용을 이해당사자에게 알리고 오류를 바로잡는 등 후속 절차를 마련하는 작업은 중요하다. 의료영역에서 진단에 도움을 주는 AI와 관련해서도 환자의 시각에서는 AI의 개략적인 작동 방식은 설명을 통해 이해할 수 있을지는 몰라도 개개 환자별 구체적인 처방에 이런 경위에 대해서는 이해하지 못하는 경우가 일반적이다.

이와 같은 상황에서 개인에게 중요한 영향을 미칠 의사결정을 전적으로 AI에게 맡기는 것은 자율성에 중대한 침해를 유발할 수 있다. 반면, AI 의사결정이 충분히 신뢰할 수 있다면 AI에 의한 의사결정에 대해 설명가능성을 요구하지 않고 오히려 적극적 활용을 선호하는 것도 충분히 가능하다.

3 보건의료 분야에서의 XAI의 필요성과 기술현황

가. 의사-환자 관계에서의 설명가능성의 필요성

XAI는 특히 삶에 직접적인 영향을 미치거나 민감한 정보를 다루는 고위험 또는 고부담 의사결정 영역에서 더욱 중요해지고 있으며, 보건의료 분야가 그 대표적인 예이다 (Kundu, 2021; Reddy, 2022). 예를 들어, AI 기반 진단 보조 의료기기가 환자에게 폐암 진단을 내렸다면, 의사는 그 진단의 근거를 명확히 파악할 수 있어야 AI의 진단을 신뢰할 수 있고 환자에게 설명할 수 있으며, 질병의 예후를 예측하고, 그에 맞는 처방 및 치료 계획을 세워 적절한 환자 관리를 할 수 있다. 예를 들어, 환자 입장에서 폐암 진단은 막대한 의료비 지출, 항암치료로 인한 신체적 부작용, 예후에 대한 두려움, 질환 관리에 소요되는 시간 등 여러 측면에서 큰 부담을 초래한다. 그러므로 환자 자신이 어떠한 기준에 의해 폐암 진단을 받았는지에 대한 충분한 설명을 받을 권리가 있다. 만약 의사가 AI 기반 의료기기의 판단 과정 및 근거를 설명할 수 없다면, 환자는 폐암 진단을 신뢰할 수 없을 것이다. 이처럼 개인의 안전이나 생명에 관한 중대한 의사결정을 요구하는 의료 시스템에서 XAI의 필요성은 더욱 부각된다. XAI는 의료 AI에 대한 불안감을 해소하고 신뢰도를 높여, AI 시스템의 임상 적용을 촉진할 수 있다.

그러나, AI에 대한 설명이 제공되더라도 비전문가인 의료인이나 환자가 이를 완전히 이해하기는 어렵다는 견해가 많다. AI 알고리즘은 복잡한 수학과 통계 개념으로 구성되어 전문가가 아닌 경우 작동 원리를 파악하기 어려우며, 이는 의료인과 환자의 과학적 이해력을 고려할 때, 설명이 제공되더라도 실제 이해로 이어지기 힘들 수 있다는 점을 시사한다.

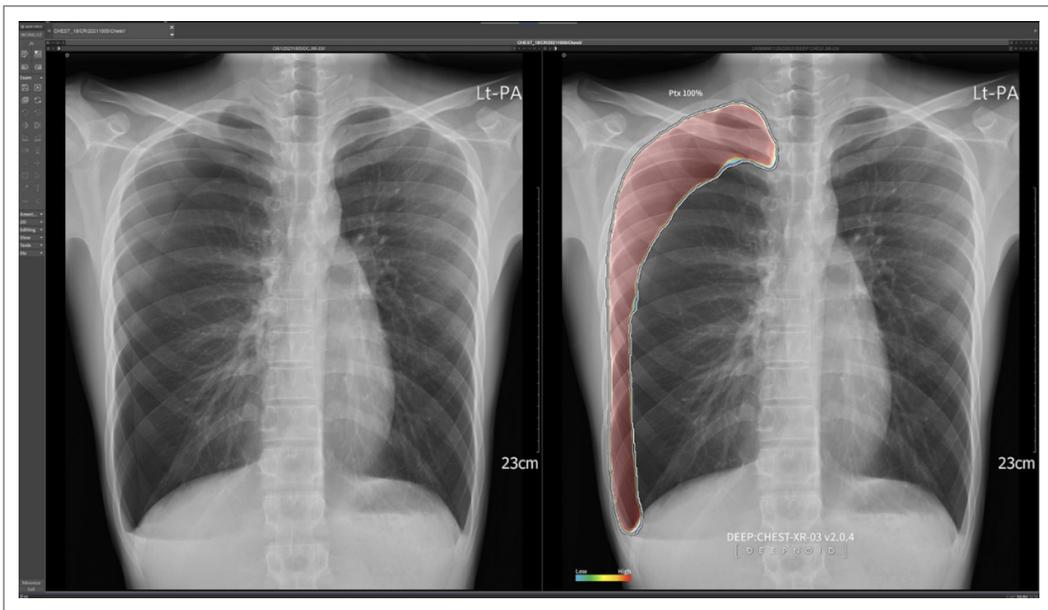
의료 분야에서 AI의 발전이 특수 영역에서 의료인의 역량을 능가하거나, 중저개발국

에서 의료인의 역량 또는 의료기술의 부족으로 인해 진단 및 치료 결정에 유의미한 역할을 하게 되는 경우에는, AI의 결정 과정에 대한 투명성 확보가 필수적이며, 그 이유는 다음과 같다. 첫째, 환자는 자기 신체에 대한 결정권을 가지므로, AI의 진단이나 치료법 추천의 근거를 알고 이를 수용할지 여부를 판단할 수 있어야 한다. 환자에게 제공되는 설명은 치료법 선택과 효과, 부작용, 대안 등을 이해하게 하여 환자의 자기결정권을 존중하는 데 중요한 역할을 한다. 둘째, 의료인은 AI가 제시한 판단을 이해해야 AI가 진단과 치료 방법 결정에서 자신을 대신할 수 있는지를 평가할 수 있다. AI가 의사 결정 보조 수단으로 사용될 때, 의료인은 AI의 판단이 자신의 판단과 다를 경우 그 원인을 파악하여 어떤 의견을 수용할지 결정해야 할 필요가 있다. 이 과정에서 의료인은 원인 분석을 통해 추가적인 조치나 치료 방향을 검토할 수 있다. 마지막으로, AI의 투명성은 공정성 모니터링의 기반이 된다. 투명성을 통해 AI가 진료에서 편향이나 차별 없이 공정한 결정을 내리는지를 감시할 수 있으며, 이는 의료 AI의 신뢰성을 높이고 환자와 의료인의 신뢰를 구축하는 데 기여할 수 있다.

아직까지 보건의료 분야에서 AI가 인간 전문가를 완전히 대체할 수 없다는 점과 보건의료 분야가 생명과 삶의 질에 직접적인 영향을 미치는 영역임을 고려할 때, XAI의 필요성은 인간-컴퓨터 상호작용(Human-Computer Interaction, HCI)의 관점에서 중요성을 지닌다. 보건의료 분야에서 AI가 실질적인 가치를 창출하기 위해서는 필연적으로 보건의료 전문인력들에 의한 활용 등 상호작용이 필요하다는 점을 고려할 때, AI가 최소한의 설명가능성을 확보하는 것은 보건의료 전문인력들의 수용가능성에 영향을 미칠 수 있다. 특히, AI의 학습에 활용되는 데이터, 세부 조정 과정 등에 따라 AI가 제시하는 응답이 편향될 수 있다는 점과 보건의료에서 중요한 가치 중 하나로 고려되는 환자 중심의 가치와 목표를 반영한 공유된 의사결정(Shared Decision Making, SDM)을 고려할 때, 보건의료 인력들이 AI의 결론을 수용하기 위해서는 판단의 근거에 대한 최소한의 정보들을 필요로 할 수 있다.

나. 진단

질병의 진단은 보건의료 분야에서 AI가 가장 흔히 활용되는 분야이며, 이와 관련한 XAI도 연구가 활발하다(van der Velden et al., 2022). XAI는 AI 시스템이 X-ray 검사, CT 검사 및 혈액 검사 등 다양한 진단 검사 또는 건강검진 정보에 기반하여 질환 진단을 할 때, 구체적으로 어떤 기준에 의해 최종 진단이 이루어졌는지 사용자가 이해할 수 있도록 설명하는 역할을 한다. X-ray, CT, MRI 등과 같은 의료영상 분야는 축적된 데이터가 많고, 영상 데이터 특성상 AI 알고리즘 개발에 적합해서 XAI가 가장 많이 적용되고 있다. 의료영상검사는 일상적으로 시행되며, 정해진 진료지침에 따라 수행되어 방대한 양의 데이터가 병원 전산시스템에 보관된다. 또한, 다수의 픽셀로 이루어진 영상 형태의 데이터는 메모리 용량이 커서 복잡한 AI 연산에 적합하다. 픽셀 단위의 데이터는 대체로 사람의 눈으로 식별하기 어려울 정도로 작기 때문에 AI 알고리즘의 도움을 받기에 적합하고, AI 기반 분석 결과의 시각화로 해당 결과를 설명하기에도 용이하다.



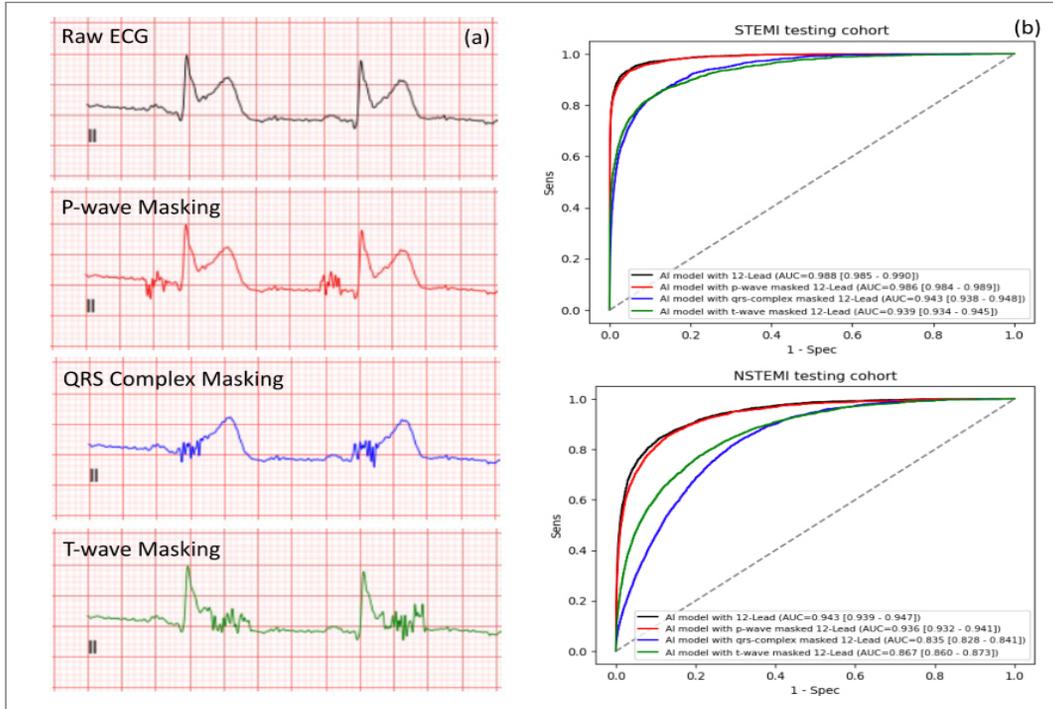
출처: Lee, et al., 2022

[그림 5-2] AI 기반 흉부 X-ray 진단 보조 소프트웨어를 통한 판독결과

예를 들어, 국내에 상용화된 한 AI 기반 흉부 X-ray 진단 보조 소프트웨어는 영상 전체에 대한 질환이 존재할 확률을 판독 결과로 보여줄 뿐만 아니라, 최종 판독의 근거로 삼은 영역이 어디에 위치하였는지 시각화해주는 기능을 탑재하였다(그림 5-2). 이와 같은 결과 설명 시각화 기능은 사용자가 AI 판독 결과를 직접 해석하고 최종 결론을 주체적으로 내릴 수 있게 도와주며, AI 시스템의 오류를 검출하거나 AI 판독 결과에 기반한 추가 진단을 가능하게 한다. 이러한 이유로 미국 식품의약국(FDA)와 한국 식품의약안전처에서도 의료영상 진단 보조 소프트웨어 인허가 과정에서 위치 지정 정확도(localization accuracy) 측정을 권장하는 등 의료기기의 성능 심사에 XAI를 활용하기도 한다(FDA, 2020).

다. 예후 예측

치료 수단이 없거나 진단 시기에 민감한 질환의 경우, 진단이 아무리 정확하고 신속하더라도 발병 후에는 환자의 상태를 개선하기에는 이미 늦을 수 있다. 예를 들어, 뇌졸중이나 심정지와 같은 질병은 발병 전에 위험성을 예측하여 예방하거나 대처함으로써 예후를 개선할 수 있다. 국내에 상용화된 한 AI 기반 심정지 예측 의료기기는 환자의 활력 징후를 기반으로 24시간 이내 심정지 발생 위험도를 예측한다(Cho, et al., 2023). 이러한 조기 경고 시스템은 응급실이나 중환자실에서 환자의 임상 상태 악화를 조기에 발견하고 적절한 중재를 가능하게 한다. 또한, 인력 부족으로 상시 모니터링이 어려운 일반 병동에서도 심정지나 계획되지 않은 중환자실 이송을 예방하는 데 효과가 있음이 입증되었다.



출처: Ji, et al., 2023

[그림 5-3] 비정상적인 QRS파와 T파가 AI 기반 심정지 예측에 활용된다는 XAI 연구 결과

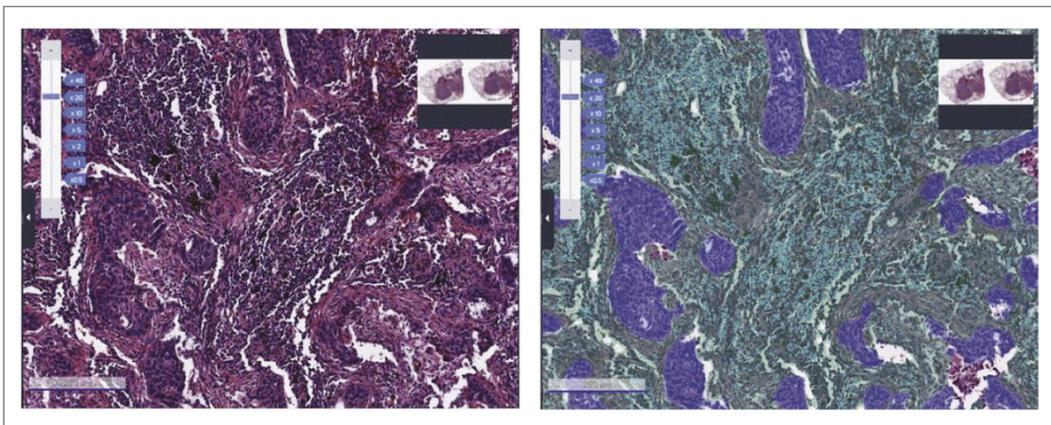
해당 의료기기의 연구진은 XAI 연구를 통해 심전도에서 특히 비정상적인 QRS파와 T파가 알고리즘 작용원리에 중요한 역할을 한다는 사실을 밝혀냈다(그림 5-3). 심정지 예측 신호는 다른 합병증의 전조현상일 수도 있으므로, 이러한 정보는 의료진에게 반드시 전달되어야 한다. 그러나, 오경보가 잦을 시, 의료진의 신뢰를 잃을 수 있기 때문에 XAI를 통해 해석가능성을 높이고 지속적인 알고리즘 개선을 모색해야 한다.

라. 치료

정밀화된 약물치료는 정밀의학(precision medicine)의 핵심 요소로, 환자의 유전적, 생물학적 및 환경적 특성을 고려하여 가장 효과적이고 부작용이 적은 약물을 선택하는 맞춤형 치료 접근법을 의미한다. 임상시험에서 평균적인 효과가 입증된 시판 약물 중에서도 환자 개개인에게 효과가 나타나지 않는 경우가 잦으며, 특정 약물이 개인에게 효

과가 있을지는, 여러 요소가 복합적으로 작용하기 때문에 예측하기 쉽지 않다. 따라서 AI 분석을 활용한 약물 선택을 통해 치료의 효율성을 높이고, 환자 개개인에게 최적화된 치료 결과를 도출하려는 노력이 계속되고 있다.

정밀의학에 대한 논의에서 주목받고 있는 또 다른 접근 중 하나는, 기존에 이미 개발된 의약품의 새로운 적응증을 찾거나, 기존에는 효과의 크기가 개인별로 편차가 지나치게 커 평균적인 효과를 토대로는 시장성을 갖추기 어려웠던 의약품에 대해, 실사용 데이터를 기반으로 효과의 크기가 급여화될 수 있을만큼 커지는 적응증이나 인구집단을 식별해내기 위해 머신러닝을 접목한 이질적 인과효과 추론(heterogenous treatment effect estimation)을 수행하는 것이다. 이는 후보물질 탐색 비용이나, 안전성 실험 등의 비용이 이미 지불된 의약품의 적용범위를 확장시키기 위해 데이터에 기반한 접근을 수행한다는 측면에서 비용효과성이 확보될 수 있는 방안으로 고려되고 있다. 그러나, 이와 같은 빅데이터 및 머신러닝에 기반한 정보를 근거로 임상시험이나 환자에의 적용이 바로 가능한지에 대해서는 여전히 안전성·유효성 등의 관점에서 논의가 이루어질 필요가 있다.



출처: Park, et al., 2022

[그림 5-4] AI 기반 병리영상 분석 알고리즘을 통한 바이오마커 규명

국내에서 개발 중인 한 AI 기반 병리영상 분석 알고리즘은 종양 미세 환경에서 종양 침윤림프구(tumor-infiltrating lymphocyte)를 분석하여 면역관문억제제(immune checkpoint inhibitor)의 효과와 관련된 면역 표현형을 분류할 수 있다(그림 5-4). 이 면역 표현형은 진행형 비소세포 폐암 환자에 대한 면역관문억제제의 효과를 예측하는데 활용될 새로운 바이오마커로서의 잠재력이 있는 것으로 나타났다. 또한, XAI 기술로 원본 병리영상에서 암 상피(cancer epitherm), 암 기질(cancer stroma), 종양침윤림프구를 각각 시각화하였다. 기존 방식의 종양침윤림프구 분석은 수행과정이 복잡하고 방법론적 한계 때문에 임상 적용이 어려웠지만, AI 기반 분석을 통해 항암면역치료의 정밀화에 더 근접할 수 있게 되었다.

4 XAI의 한계점 및 개선방안

보건의료 분야에서 AI 기술의 활용 범위가 확대되며, 의료 관련 산업 및 보건의료 시스템 전반에 큰 영향을 미치고 있다. 이에 따라 AI 시스템의 안전성과 신뢰성에 대한 관심이 증가하고 있으며 AI 알고리즘이 갈수록 복잡해짐에 따라 작동 원리에 대한 설명 가능성은 필수조건이 되었다. XAI의 역할이 중요해지고 있는 가운데, 실질적으로 최종 사용자의 이해를 돕기 위해서는 극복해야 할 과제가 존재한다.

XAI가 해결해야 할 주요한 과제 중 하나는 해석가능성과 정확성의 균형을 유지하는 것이다(Yang, et al., 2023). 선형 회귀나 의사결정 트리와 같이 구조가 단순한 모델은 대체로 해석가능성이 높다. 이러한 모델은 내재적인 구조가 단순하고 투명하기 때문에 작용 원리에 기반하여 직접적으로 출력값을 설명하기 용이하다. 하지만 단순한 모델 구조는 복잡한 상관관계를 포착하지 못하거나 지나치게 단순화하여 결과값의 정확도를 떨어뜨릴 수 있다. 반면, 딥러닝 모델과 같은 복잡한 모델은 정확도가 높을 수 있으나, 그 작용 원리는 해석하기 어려운 경우가 많다. 복잡한 모델을 내재적으로 해석하기는 어렵기 때문에 일반적으로는 외부 모델을 사용하여 사후 설명을 한다. 하지만 사후 설명 모델로 메인 모델의 복잡한 구조를 설명하는 것에는 한계가 있으며, 이러한 모델의 추가는 계산 비용이 많이 소요될 수 있다. 또한, 모델이 복잡해질수록 항상 높은 정확성을 보장하지는 않는다. 모델이 복잡해지면 데이터의 노이즈를 학습해서 모델이 과적합될

가능성도 높아진다. 이렇듯 해석가능성과 정확성의 상충관계에 대한 보편적인 해결책은 없기 때문에, 개발환경과 최종사용자의 상황에 맞게 적절하게 균형이 맞춰져야 한다.

XAI를 활용해 AI 시스템에 대한 신뢰와 확신을 얻기 위해 해결해야 할 또 다른 과제는 XAI의 측정과 평가이다(di Martino & Delmastro, 2023). 현재 표준으로 사용되는 XAI 평가 프레임워크가 없기 때문에 연구자는 제각각 다른 방식으로 XAI를 평가하고 있다. AI 시스템별로 요구되는 설명의 성격 또는 입력 데이터의 형태가 다르기 때문에 XAI 기술의 성능을 비교하는 것은 간단하지 않다. 기존의 XAI의 정량적 평가는 복수의 전문가가 합의한 정답지와 비교하거나, 입력값 변형에 따른 성능 저하를 측정하는 방법으로 이루어져 있다. 이러한 방법은 알고리즘 작동 원리와 설명이 얼마나 일치하는지를 객관적인 지표로 평가할 수 있다는 장점이 있지만, 인간의 편향이 반영될 수 있다는 결점이 있고, 시계열 데이터와 같은 직관적이지 않은 특성의 데이터에서는 전문가 합의 방식을 적용하기 어렵다. 효과적이고 성공적인 XAI를 위해서는 사용자 중심의 XAI 평가 표준이 개발되어야 한다. 사용자가 중심이 되는 평가가 이루어질 때, 객관적인 정량적 지표를 넘어 사용자 관점에서의 유용성, 만족도, 설명에 대한 신뢰와 같은 질적 평가와 인간과 AI 간의 상호작용과 같은 실제적 사용에 대한 평가가 이루어질 수 있다.

사용자 중심 평가가 중요한 이유를 보여주는 사례로는 AI 개발자와 최종사용자의 차이를 들 수 있다. AI 시스템의 오류가 발생할 시에 책임을 지는 AI 개발자는 알고리즘 전반에 대한 패턴을 설명해주는 전역적 설명을 필요로 하는 반면, 임상이나 환자와 같은 최종사용자는 알고리즘 전체를 이해하기 보다는 자신에게 영향을 미치는 개별 예측 결과에 대한 국소적 설명에 관심이 있을 것이다. 또한, AI 개발자는 알고리즘의 전문적인 해석가능성과 유지보수성과 같은 알고리즘의 직접적인 특성을 중요시하는 반면, 사용자는 인과관계에 대한 직관적 판단을 더 중요하게 여길 것이다. AI 시스템이 전문가 관점에서 투명하고 해석 가능하게 설명되더라도, 비전문가가 이해할 수 없는 방식일 때가 많다. 따라서 각 유형의 사용자가 어떤 정보를 어떤 형태로 제공받기를 원하는지 등 사용자의 맥락을 고려한 XAI 연구가 진행될 필요가 있다.

제3절

공정성과 차별금지

NATIONAL ASSEMBLY FUTURES INSTITUTE

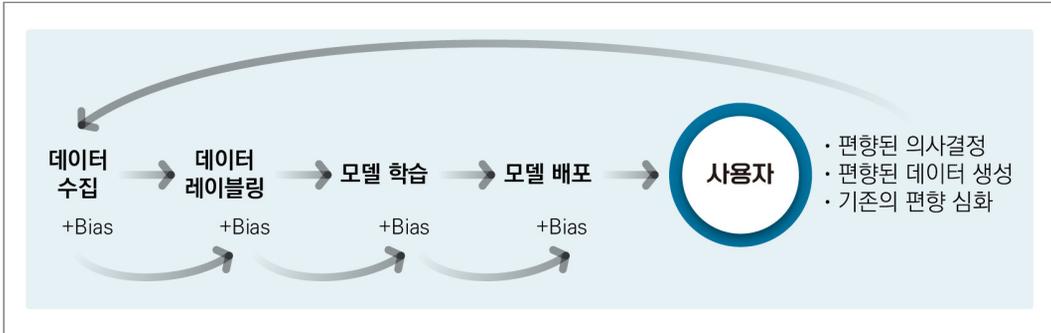
1 인공지능의 공정성과 차별가능성

AI는 학습 데이터에 따라 의사결정이 달라지기 때문에 공정하지 않은 데이터로 훈련 되면 불공정한 결정을 내릴 가능성이 크다. 머신러닝 기반의 AI는 데이터 패턴에 의존하여 추론하기 때문에 데이터에 내재된 편향을 그대로 학습할 위험이 있다. 불공정한 의사결정이 반복되면 피드백 루프가 형성되어 기존의 편향이 더 심화되고 악순환이 발생할 수 있다.

대표적인 사례로, 미국 법원에서 아프리카계 미국인에게 불리하게 작용했던 재범 위험 측정 소프트웨어가 있다(Angwin et al., 2016). 형사피고인 중에서 재범 고위험군으로 분류되었으나, 실제로 재범을 저지르지 않은 비율이 흑인이 백인에 비해 약 2배로 나타난 것이다.

이는 알고리즘 공정성(algorithmic fairness)의 이슈를 발생시키는데, 알고리즘 공정성이란 알고리즘의 의사결정 과정에서 개인이나 집단의 본질적 혹은 습득한 특성에 따라 편향이 없고 차별하지 않는 것을 의미한다(Mehrabi, et al., 2021). 이 때 편향은 한쪽으로 치우친 성질을 뜻하며, AI 시스템의 편향은 AI 알고리즘의 의사결정이 체계적으로 한쪽으로 치우친 결과를 산출하는 상태를 말한다. 사람과 마찬가지로 알고리즘도 편향(bias)에 취약하여 특정 집단에 유리하게 결정이 내려지도록 왜곡될 수 있다.

의료보건 분야에서도 AI 시스템의 공정성을 확보하는 것은 중요하다. 환자의 특성에 따라 성능의 차이가 나거나 시스템의 배포가 고르지 못하다면, AI의 임상 적용이 오히려 공평한 의료 서비스 제공을 저해할 수 있다. 데이터 대표성이 부족한 사회적 소외집단이 특히 불균형한 성능 저하나 편향에 따른 오류에 취약하다는 사례가 관찰됨에 따라 AI 시스템이 사회에 미칠 차별적인 영향에 대한 윤리적 우려가 커지고 있다(McCradden, 2020).



출처: 저자 구성

[그림 5-5] 인공지능 알고리즘의 편향성

2 건강 데이터 편향

알고리즘 공정성을 보장하는 데 있어 가장 큰 걸림돌 중 하나는 학습에 활용되는 데이터의 편향이다. 데이터 편향은 데이터 수집 과정에서 발생하여 대표성의 부족과 왜곡된 데이터 분포로 이어진다. 의료 AI를 생성하기 위해 사용되는 대부분의 생물의학 데이터셋은 유럽계 혈통의 비율이 과도하게 높으며, 다양한 인종이나 민족, 국가, 사회계층에 대한 대표성이 부족한 경우가 많다. 예를 들어, 14개의 공개된 피부과 영상 데이터셋 중 11개(78.6%)가 북미, 유럽 및 오세아니아에서 수집되었고, 민족성 또는 피츠패트릭 피부 유형에 대한 정보가 제한적이며 어두운 피부를 가진 환자가 지나치게 과소 대표되어 있다(Wen, 2022). 마찬가지로, 94개의 안과 데이터셋 중 10개(10.6%)만이 아프리카 또는 중동에서 수집되었으며, 2개(2.1%)가 남미에서 수집되었다. 33종의 암의 종양 샘플을 포함하는 암 관련 연구의 대표 데이터셋에서도 전체 샘플의 82%가 백인인 반면 다른 소수민족의 샘플 수가 현저히 적게 대표되어 있다(Gao, 2013).

이러한 데이터 수집에서의 불균형은 AI 시대에 나타난 새로운 현상이 아니다. 유전체 데이터를 본격적으로 모으기 시작한 2000년대부터 데이터 비율을 조사한 결과, 유럽계 혈통 데이터가 다른 인종 및 민족 집단 데이터에 비해 압도적인 차이로 꾸준히 과대 대표된 것이 관찰되었다(Martin, et al., 2019). 유전체 데이터 플랫폼인 GWAS Catalog의 2019년 데이터에 따르면 전 세계 인구의 16%에 불과한 유럽계 혈통의 데

이터가 전체 데이터의 79%를 차지하고 있었다. 대표성이 편향된 데이터셋에 기반하여 개발된 AI 모델은 과소 대표된 집단에서 성능이 저하될 위험이 있으므로, 알고리즘 공정성을 위해서 대표성 개선은 매우 중요한 과제이다.

데이터 품질 또는 특성의 이질성도 AI 알고리즘에 편향을 일으킬 수 있는 또 다른 요인이다. 수집된 장소나 시기, 수집 주체 및 방법 등에 따라 이질성이 발생할 수 있으며, 이러한 이질성이 건강 격차와 맞물렸을 때, AI 성능 또는 공정성을 저해하는 상관관계가 학습될 수 있다. 예를 들어, 방사선, 유방촬영술 또는 CT 영상은 병원별 장비 종류 또는 프로토콜에 따라 방사선량에 차이가 있을 수 있으며, 이는 해당 병원의 환자 특성과 상관관계가 나타나게 할 수 있다. 병리 영상에서도 조직 검사 및 염색 프로토콜로 인해 데이터 이질성이 발생할 수 있다.

또한, 국가 간의 질병 치료에 대한 가이드라인과 보건의료체계의 차이는 알고리즘 기반 치료의 유효성의 문제를 유발할 수 있다. 의료 행위는 궁극적으로는 의사결정 과정이며 이 과정은 환자와 의사가 처한 보건의료시스템과 사회경제적 환경에 크게 좌우되며(Kim, 2020), 보건의료의 역량과 의료품의 공급 수준 등에 의해서도 차이가 발생한다. 예를 들어, 영국은 미국과의 암치료에 대한 관점이 다르기 때문에 왓슨의 치료 권고 내용을 그대로 수용하지 않았으며(STAT, 2017), 한국의 건강보험 수가 시스템을 반영한 치료 가이드라인이나 한국 의사들의 치료패턴은 다른 나라에서 최선의 치료일 수 없다. 환자의 상태와 의사 개개인의 전문성에 따른 치료와 처방의 차이를 고려한다면 AI가 최선의 의료를 제공하는가에 대해서는 판단하기 어렵다. 실제 진료 현장에서는 표준 지침에 어긋나지만 최선의 진료가 되기도 하며 동일한 질환이라도 새로운 치료를 시도하는 경우가 빈번하게 발생한다.

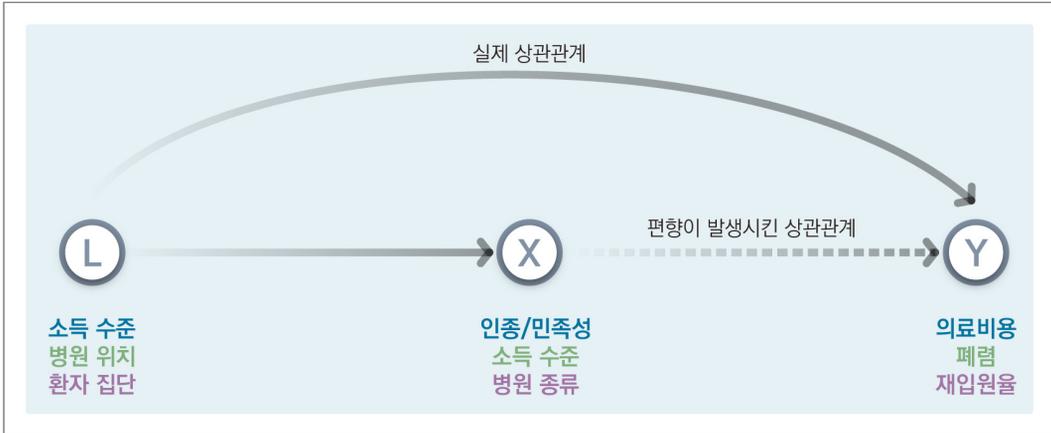
한편, 데이터 수집 주체의 편향적 태도나 방식에 따라 데이터에 편향이 스며들 수도 있다. 예를 들어, 의료진이 환자의 국적, 인종 또는 성별에 따라 편향된 시각으로 임상 기록을 남기면, 이러한 편향이 고스란히 데이터로, 그리고 모델로 전달될 것이다. 또한 시간이 지남에 따라 데이터셋에 이질성이 생기는 경우도 있다(Guo, et al., 2021). 질환의 정의가 절대적이지 않거나 주관적 판단이 포함되는 경우, 시간이 지나면서 표준적인 진단 가이드라인이 바뀌거나 의료 현장에서의 관행이 변하면, 데이터가 내포하는 의미가 달라지고 중요변수와의 관계가 변형될 수 있다. 예를 들어, 국제 질병코드가

ICD-9에서 ICD-10으로 변경된 후, 오피오이드(opioid) 관련 입원 환자 수가 급격하게 증가한 사례가 있다(Heslin, et al., 2017). 이러한 시간적 변화로 인해 AI 시스템의 성능이 불안정해지거나 특정 집단에 대해 성능 편차가 커질 수 있으므로, 알고리즘 공정성을 보장하기 위해 지속적인 모니터링이 필요하다.

3 건강 격차와 알고리즘 편향

보건의료 분야에서의 알고리즘 편향(algorithmic bias)을 이해하기 위해서는 먼저 집단 간 건강 격차에 대한 이해가 필요하다. 대개 한 사회의 여러 하위집단 간에는 질환 발병, 건강 수준, 의료의 질과 접근성 등에서 차이가 발생하는데, 이러한 불균형은 생물학적 기질, 성별, 나이, 문화, 인종, 사회경제적 수준, 지리적 위치, 유전형 등 다양한 요소에 기인한다. 이러한 건강 격차와 다양한 요인과의 상관관계를 파악하는 것은 사회 환경이나 의료체계를 개선하는 정책을 만들고 건강 격차를 완화하기 위해 중요하다.

그러나 단순히 상관관계에만 집중한 의사결정은 알고리즘 편향으로 이어질 위험이 있다. 예를 들어, 인종을 변수로 포함하여 추정 사구체 여과율(eGFR)을 계산하는 공식이 흑인의 신장 기능을 과대평가하여 만성신장질환을 과소진단하게 한 사례가 있다. 이는 흑인 환자의 신장이식 등 필요한 관리가 적절한 시기에 이루어지지 못하게 한다(Quaggin & Palevsky, 2021). 또한, 특정 골절 위험 평가 계산식에서는 흑인의 골다공증 위험을 더 낮게 측정하여 골다공증에 대한 개입이 불합리하게 지연된 사례도 있다(Lewiecki, Wright & Singer, 2020). 이처럼 인과관계에 대한 명확한 이해 없이 상관관계에 기반하여 알고리즘을 설계하면 편향된 데이터와 마찬가지로 편향된 의사결정을 내리게 된다.



출처: 저자 구성

[그림 5-6] 인공지능 학습에 활용되는 데이터의 편향

AI 모델 또한 다르지 않다. 편향된 데이터에 기반하여 훈련되면 편향된 결과값을 도출하게 되고, 이로 인한 AI 시스템의 알고리즘 편향은 여러 차례 보고된 바 있다. 한 연구에서는 다수의 병원으로부터 획득한 병리영상을 학습한 딥러닝 모델이 영상의 환자가 어떤 인종인지 분류할 수 있었다(Howard, et al., 2021). 병원별로 상이한 영상 촬영 프로토콜과 인종별 비율의 상관관계로 인해, 영상의 헤마톡실린과 에오신(Hematoxyl & Eosin; H & E) 염색 강도와 인종 간의 상관관계가 생긴 것이다. 병리영상에 기반한 질환 진단에서 인종은 임상적으로 중요하지 않지만, 딥러닝 모델은 인종에 대한 정보를 학습함으로써 질환과 관련되어 편향된 병리영상 분석 결과를 낼 위험이 있다. 또 다른 연구에서는 방사선 분류 딥러닝 모델이 실제 인과관계와 관련 없는 정보로부터 환자의 인종을 학습하여 히스패닉 또는 흑인, 저소득층 환자에 대한 과소진단 또는 오진율이 불균형적으로 높게 나타나는 것이 관찰되었다(Seyyed-Kalantari, et al., 2021).

알고리즘 편향은 학술적 연구에서뿐만 아니라 이미 실제 의료 현장에서 큰 영향을 미치고 있다. 미국 전역에 배포되어 수백만 명의 환자에게 영향을 미치는 고위험 환자 예측 알고리즘에서도 상당한 인종적 편향이 관찰되었다(Obermeyer, et al., 2019). 고위험 환자에게 추가적인 서비스와 관리를 제공하여 환자의 예후를 개선하고 병원이 비용 효과적인 프로그램을 운영할 수 있도록 하는 알고리즘은 매년 미국에서 약 2억 명의 환자

에게 적용되며, 대형 병원에서 고위험 관리 프로그램 대상을 선정할 때 활용된다. 하지만 연구 결과, 해당 알고리즘으로부터 동일한 위험 점수를 책정 받았을 때 흑인 환자는 백인 환자보다 훨씬 병세가 심각한 것이 드러났다. 이는 흑인 환자의 건강 상태가 실제보다 좋게 평가되어 고위험 환자로 구분되지 않아 적절한 관리를 받지 못하는 것으로 이어진다.

이러한 편향이 발생한 이유는 해당 알고리즘이 의료 비용을 환자의 건강 상태에 대한 대리 변수로 삼아 예측하였기 때문이다. 의료 접근의 불평등으로 인해 흑인 환자는 동일한 건강 수준에서도 의료 비용을 백인에 비해 낮게 사용하므로 이러한 상관관계를 알고리즘이 학습한 것이다. 또 다른 예로, 병원별 사망률이나 재입원율을 측정하는 알고리즘에서 환자 특성을 고려하지 않는다면, 병원의 성과를 비교할 때 빈곤층이나 유색인종 환자를 주로 상대하는 병원은 불이익을 볼 가능성이 있다(Joynt Maddox et al., 2019). 이처럼 있는 그대로의 패턴을 학습한다고 하더라도 알고리즘의 목적에 따라 편향이 발생할 수 있다. 현존하는 건강 격차와 구조적 불평등에 의해 실제 데이터에서 발견되는 여러 상관관계가 알고리즘에 편향을 일으킬 수 있으므로, 인과적 이해를 바탕으로 알고리즘을 개발하는 원칙이 편향을 줄이는 노력의 핵심이 될 수 있다.

4 보건의료 분야 알고리즘 공정성을 위한 기술

알고리즘 편향을 줄이고 공정성을 개선하기 위해 사용하는 기술은 단계에 따라 처리 전, 처리 중, 처리 후 보정으로 나눌 수 있다(Chen, et al., 2023). 먼저, 처리 전 단계에서는 데이터 수집 과정에서 편향을 완화할 수 있다. 과소 대표된 집단의 데이터를 추가로 수집하여 모델의 성능을 향상시킬 수 있다. 그러나 이미 수집된 데이터셋에 데이터를 추가하는 것은 비용이나 행정적 문제로 인해 현실적으로 어려울 수 있으며, 중요한 사회경제적 변수는 대개 민감한 개인정보에 속해 수집이 제한될 수 있다. 따라서 데이터 수집 전에 집단별 대표성을 고려한 계획이 필수적이다.

추가적인 데이터 수집이 어려운 경우에는 모델을 훈련하기 전 입력 데이터 자체를 조정하는 방법이 있다. 예를 들어, AI 알고리즘의 집단별 성능 차이는 주로 사회경제적 요인에 따라 발생한다. 이는 데이터 수집 단계에서 사회적 소외계층의 데이터가 다른

집단에 비해 품질이 낮거나 양이 적은 경우가 많기 때문이다. 이 경우, 해당 집단에 가중치를 부여하거나 리샘플링(resampling)을 통해 과소 대표성을 교정하여 데이터의 불균형을 개선할 수 있다. 하지만 입력 데이터를 임의로 수정하면 의도치 않게 새로운 편향을 유발할 수 있다. 또한 과소대표성을 개선해야 할 집단의 데이터 샘플 수가 적고 이상치가 존재할 경우, 가중치 적용이 오히려 이상치의 영향을 극대화할 위험이 있다.

연합 학습(federated learning)도 데이터 편향을 줄이는 방법 중 하나가 될 수 있다(Brisimi, et al., 2018). 민감한 생물의학 데이터는 여러 기관에서 관리되기 때문에 데이터 공유가 어렵다. 연합 학습은 각 기관이 자체적으로 모델을 훈련한 후, 모델의 가중치만을 공유해 통합하는 방식이다. 이를 통해 민감한 개인정보를 공유하지 않으면서도 데이터의 다양성을 확보하고 샘플 수를 늘릴 수 있다.

처리 중 단계에서의 편향 완화 기술도 연구되고 있다. 일반적으로 이러한 기술은 편향을 유발하는 교란변수의 영향을 최소화하는 방식으로 작동한다. 예를 들어, 의료 영상에서 병원별 또는 장비별 프로토콜 차이가 시각적으로 표현될 때, 해당 정보를 블라인딩 처리해 학습을 억제할 수 있다. 또한, 모델 가중치 최적화 과정에서 규제(regularization)를 적용한 손실함수(loss function)나 적대적 손실 항(adversarial-loss term)을 사용하여 특정 변수에 대한 집단 간 차이를 줄여서 학습하도록 유도할 수 있다(Chen, et al., 2023). 그러나 이러한 방법은 각 모델별로 개발되어야 하며, 특정 변수의 영향을 왜곡함으로써 관련된 다른 변수와의 영향도 불투명하게 변형할 수 있다는 한계가 있다. 또한, 정교한 보정을 위해서는 편향의 원인에 대한 명확한 이해가 선행되어야 한다는 문제가 있다.

마지막으로, 처리 후 단계에서의 보정이란 훈련된 모델의 출력값을 사후 수정하여 편향을 완화하는 방식이다. 가장 단순한 방법은 각 집단별로 모델 성능이 동일하게 나오도록 설정하는 것이다. 모델 출력값이 예측 확률인 경우, 캘리브레이션(calibration)을 통해 집단별로 예측 확률이 가지는 의미를 동일하게 조정할 수 있다. 하지만 모델 구조상 집단 간 성능이 일치하게 나올 수 없거나, 한 집단에서의 성능을 높이기 위해 다른 집단에서의 성능을 낮춰야 한다면 윤리적 문제가 제기될 수 있다.

5 보건의료분야 공정성 향상을 위한 개선방향

AI는 효율성 향상을 이유로 생명 윤리의 우선순위를 무시하고 부와 명예에 따라 장기 이식이나 백신과 같은 한정된 자원을 배분할 위험이 있다. 따라서 기존의 공정성 원칙을 보건의료 분야의 AI의 맥락에서 재확립할 필요가 있다.

편향된 AI 시스템으로 인해 여러 집단에 미치는 영향이 공평하지 않다면, 현존하는 건강 불평등을 더욱 악화시킬 수 있다. 생성형 AI의 등장과 함께 AI 시스템의 활용이 비약적으로 증가함에 따라, 공정성과 윤리적 시사점에 대한 관심과 우려도 커지고 있다. 글로벌 기업들은 이에 대응하여 AI의 공정성을 위한 진단 도구를 개발하고 활용하고 있다. IBM의 AI Fairness 360, Google의 Fairness Indicator, Microsoft의 Fairlearn 등이 대표적이다. 그러나, 규제적 측면에서 이러한 잠재적 위험성 속에서 차별을 예방하기 위해 AI 알고리즘을 어떻게 훈련하고 평가할지에 대한 노력이 부족한 실정이다. 최근 미국 식품의약국(FDA)과 국내 식품의약품안전처를 포함해 전 세계적으로 AI 기반 의료기기의 인허가 수가 빠른 속도로 증가하고 있는 가운데, 충분한 공정성 관련 심사 없이 인허가를 받아 사용되고 있는 의료 AI 제품이 다수 존재한다(Gaube, 2021).

AI 공정성 모니터링 방법으로는 의사결정 과정과 결과의 공정성을 평가하는 방법이 있다. 가장 바람직한 것은 알고리즘 자체가 공정성 원칙을 위반하지 않도록 하는 것이지만, 복잡하고 설명이 어려운 알고리즘의 특성상 과정의 공정성 평가가 현실적으로 어렵다. 따라서, 의료 AI시스템에서 편향을 제거하여 공정성을 확보하는 기술에 대한 연구와 더불어 따라서 결과의 공정성을 꾸준히 모니터링하는 것이 현실적인 대안이 될 수 있다. AI 시스템이 배포된 후에는 성능이 유지될 수 있도록 지속적으로 모니터링을 해야 하고 성능에 문제가 생길 시 신속한 업데이트가 필요하다. 관련한 AI 규제의 필요성이 국내외로 강조되고 있고, 이를 뒷받침 할 수 있는 법적 제도의 마련이 시급하다. 공정성을 우선시하는 책임 있는 개발과 활용이 AI 시스템을 사회에 안전하고 효과적으로 정착하게 하는 핵심 요인이 될 것이다.

이를 실현하기 위해서는 의료 AI 공정성을 평가하는 프레임워크가 구축되어야 한다. 또한, 환자의 안전과 건강에 지대한 영향을 미칠 AI 기반 의료 시스템이 긍정적인 효과를 가져올 수 있도록 사회 전반적인 노력이 필요하다. 알고리즘 공정성의 저해하는 잠

재적 요인과 이를 완화하기 위한 기술을 고려할 때, 공정하고 차별 없는 AI 시스템을 위해서는 다양한 이해관계자의 적극적인 참여가 필수적이다. 데이터 생성자, 모델 개발자, 법적 규제 주체 및 시스템 최종 사용자가 모두가 공정성의 원칙을 공유하고, 그에 따라 AI 시스템에 접근해야만 AI가 사회에 긍정적인 영향을 미칠 수 있다. 또한, 공정성에 대한 판단은 사회적 맥락을 반영해야 하기 때문에, 사회적으로나 법률적으로 합의하는 과정에는 어려움이 따른다. 따라서 다양한 이해관계자 간의 지속적인 대화와 협의가 더욱 중요하다.

6 보건의료분야 공정성 향상을 위한 입법과제: 감독기구 및 절차의 법제화

AI의 공정성과 차별금지의 보장을 위해서는 전문 지식을 가진 제 3자가 이를 지속적으로 관리하고 감독할 수 있도록 법제화하는 것이 필요하다. 이를 위해 정부기관이나 전문가 단체가 주기적으로 개인정보 사용에 대한 감사를 시행하고, 머신러닝 시스템이 차별이나 오류 없이 설계대로 작동하는지 검토하도록 할 수 있다. 만약 공정성 문제나 차별이 발견된다면 시정 명령을 내릴 수 있는 권한을 부여할 수 있으며, 유럽연합의 GDPR 제58조와 같은 규정이 그 예이다.

현재 우리나라는 AI 의료기기에 한하여 인증제를 도입하고 있으나 전통적인 약무 행정의 기관에서 제공하는 규제는 데이터, 알고리즘, 로봇 등 매우 다양한 기술들의 빠른 변화 속도를 따라가기 어려운 측면이 있다. 기술적인 문제뿐만 아니라 여러 윤리적 문제들을 고려하기 위해서는 의사 및 공학자 뿐만 아니라 변리사, 법학자 등 다양한 전문가들이 참여한 다학제적인 역량을 가진 기관이 승인과 감독 등을 총괄하는 방안을 고려해볼 수 있다.

이러한 기관을 중심으로 의료 AI에 대한 인증제를 도입하여, 전문 기관이 알고리즘의 공정성과 비차별성을 인증하도록 할 수 있다. 인증은 일정기간 동안 유효하며, 정기적인 갱신을 통해 주기적인 모니터링이 이루어지도록 하는 방안이 있다. 인증제 외에도 공정성과 비차별성을 기준으로 한 강제적 면허제를 시행할 수 있으며, 이 경우 공정하

고 차별적이지 않은 AI로 신뢰를 얻을 수 있을 것이다. 면허는 정부 또는 정부 산하 전문 기관이 발급할 수 있으며, 민간 기관이 발급할 경우 그 기관의 전문성과 중립성을 정부가 평가하여 지정하는 것이 적절할 것이다.

법제화 과정에서는 실질적인 문제가 고려되어야 한다. AI에 대한 감사의 경우, 전문가가 감사를 수행하더라도 AI가 공정한 결정을 내렸는지 확인할 수 있을 만큼 감사가 가능한 기술인지가 제도의 성패를 좌우한다. 인증제 또한 시간과 비용이 소요되는 절차로서, 인증이 가져다줄 혜택이 명확하지 않으면 성공하기 어렵다. 인증 획득에 따른 법적 혜택을 제공한다면 인증제가 성공적으로 운영될 가능성을 제고할 수 있다. 강제적 면허제는 AI의 개발 및 활용 방식에 큰 변화를 줄 수 있으므로, 제도 도입에 신중을 기해야 한다.

제4절

안전 및 책임성

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 인공지능의 안전과 책임성(responsibility)

생명과 관련된 의료 분야에서는 AI의 효율성보다는 신뢰성과 안전성 확보가 더욱 중요하다. 의료 AI의 유효성과 신뢰성이 검증될수록, AI의 역할 및 의사와의 관계가 결정되고, 이와 관련된 책임 소재와 처벌 기준도 설정된다. 현재의 법체계는 인간 중심으로 구성되어 있으며 사고와 불법행위 발생 시 책임 분배가 중요한 쟁점이다. 그러므로 의료 AI 도입 시 발생할 수 있는 오류나 성능 문제로 인한 법적 책임의 분배는 매우 복잡하며, 의사, 환자, AI 개발사, 의료기관, 규제 기관, 보험사 등 다양한 주체가 관여된다. AI의 수준이 의사결정을 전적으로 위임할 수 있거나 의사를 대체할 수 있는 수준이 아니라 의사결정 등에 참고 또는 활용하는 수준이라면, 현재 수준의 AI는 권리 주체 또는 처벌의 대상이 될 수 없고 제조, 소유 혹은 점유한 사람이 대상에 해당된다.¹²⁾

2 인공지능 책임 귀속의 어려움

의료 행위는 고도의 전문 지식과 숙련도가 요구되는 전문가만 수행할 수 있도록 법적으로 제한되어 있다. 「의료법」 제27조 제1항은 의료인이 아닌 사람의 의료 행위를 금지하고, 의료인 역시 자신의 면허 범위 내에서만 의료 행위를 할 수 있다고 규정한다. 국가에서는 의사 면허를 부여할 때, 의료의 질 향상과 감염 예방, 그리고 의료기술 발전 등을 통해 환자에게 최상의 의료 서비스를 제공할 의무를 의료인에게 부과하고 있다(의료법 제4조 제1항).

의료 행위에는 환자의 상태를 진단하고, 적절한 치료 방법을 선택하여 시행하는 것이

12) 그러나, 인공지능의 발전으로 인해 강인공지능이 개발되어 인간과 비슷한 수준의 자율성을 갖게 될 경우는 매우 다른 법적 대응이 필요하게 된다. 이 부분은 V. 인간의 기술통제에서 다룬다.

포함되며, 여기에는 진단, 시술, 처방 등이 해당한다. 최선의 의료 행위란 무조건 완벽한 결과를 요구하는 것이 아니라, 환자의 상태를 면밀히 살피고, 의학적 지식과 윤리적 기준에 따라 위험을 최소화하는 결정을 내리는 것을 의미한다. 이와 관련하여 대법원 판례(2003.1.24. 선고, 2002다3822)는 의사가 진단 과정에서 주의 의무를 다해야 한다는 점을 강조하며, 그 이행 여부가 의사의 책임 판단 기준이 된다.

이에 따라 비의료인의 의료 행위는 강력히 처벌하여 의료인의 독점적 권한을 보호하며, 의료 행위 관련 책임을 명확히 한다. 대법원 판례(2010.7.8. 선고, 2007다55866)도 의사가 진단 및 치료에 있어 최선의 주의 의무를 다했는지 여부에 따라 법적 책임을 논의하고 있다.

그러나, AI 소프트웨어는 인격체가 아니기에 법적 책임을 부과하기 어렵다. 식약처의 '빅데이터 및 인공지능 기술이 적용된 의료기기 허가심사 가이드라인'에서도 AI 소프트웨어를 의료기기(Software as Medical Device, SaMD)로 분류하고 있으나, 법적인 주체로 인정하지 않는다. 「의료기기법」 제2조 제1항에 따르면, 의료기기는 질병을 진단, 치료, 경감, 처치 또는 예방할 목적으로 사람이나 동물에게 사용되는 기구, 장치, 재료, 소프트웨어 등을 의미한다. 따라서 현행 「민법」에서 AI의 법적 지위는 단순한 물건에 불과하며, 독자적인 권리 능력을 인정받지 못한다(김광수, 2018; 이도국, 2017).

그 결과 AI 소프트웨어의 사용으로 인해 환자가 필요한 진료를 받지 못하거나 불필요한 검사, 수술, 처방이 이루어졌다면 손해배상 책임이 제기될 수 있다. 이 경우, AI 소프트웨어의 개발자, 이를 사용한 의료인, 또는 의료기관에 귀책 사유가 생길 수 있다. 「민법」 제390조는 진료 계약의 불이행 책임을, 제750조는 고의나 과실로 환자의 생명이나 신체에 손해를 가한 경우 불법행위 책임을 규정하고 있다. AI 소프트웨어 자체의 결함으로 발생한 손해는 개발자가 책임을 지며, 소프트웨어가 내장된 의료기기의 경우, 「제조물책임법」 제3조에 따라 제조자는 결함으로 인한 생명, 신체, 재산상의 손해에 대해 배상 책임을 진다. 즉, 이에 따라 이의 활용에 따른 안전성에 대한 책임성은 소프트웨어 개발업체, 이의 안전성·유효성의 평가 및 인허가 주체인 식품의약품안전처, 최종적으로 처방권자인 의료인에 이르기까지 모호하게 분포하고 있는 특성을 지니고 있다.

이처럼 최종 이용자인 의료인 또는 의료기관과 개발자 사이의 책임 소재가 불명확한 환경에서, 안전성·유효성에 대한 평가를 바탕으로 SaMD의 인허가를 수행하는 식품의

약품안전처의 역할이 중요하다. 머신러닝 기반 소프트웨어와 같은 복잡한 AI 모델은 사용 환경에서의 맥락적 변화에 민감하므로, 지속적인 성능 모니터링이 필수적이며, 이에 따라 AI 모델이 개발된 후에도 지속적인 평가와 업데이트가 필요하다. 또한, 대규모 언어 모델(Large Language Model, LLM)과 같은 생성형 AI는 예측 불가능한 결과를 초래할 수 있어 정확성과 안전성을 보장하기 위한 도개 개발 및 혁신적인 규제 전략이 필요할 수 있다. 이와 같은 문제들은 특히 인허가의 판단 시점에 인허가 주체인 식약처가 어디까지 판단해야 하는지의 문제, 이후 소프트웨어 및 데이터의 업데이트 과정에서 발생하는 문제들을 자율적으로 점검하기 위한 데이터-인공지능 안전성 검토 거버넌스 체계의 구축 등 다양한 관점에서 검토될 필요가 있다. 미 FDA는 이를 위해 개발, 배포, 운영, 실제 사용 환경에서의 성능 평가를 지속적으로 수행하는 전주기 관리(Total Product Life Cycle) 접근 방식을 제시한 바 있다(Warraich, Tazbaz, & Califf, 2024).

효율성의 측면에서 AI가 인간보다 뛰어나더라도, 임상적으로 의사보다 우월한 효용성을 지니는 경우는 영상의학 분야를 제외하고는 현재까지는 제한적이다. AI는 기계학습과 딥러닝으로 패턴을 학습하여 질병을 진단하고 치료법을 제안하지만, 어떤 데이터가 실제로 유용한지 분별하는 데는 한계가 있다. AI의 신뢰성을 높이려면 데이터의 질과 수집 과정의 투명성 검토가 필수적이지만, 현재 의료 데이터는 기관마다 형식이 다르고 일관성이 부족하다(Kim & Kim, 2019). 또한, 진단 정확도가 특정 시점에 높더라도 지속적인 검증이 필요하여 실제 의료 적용에서는 한계가 있다. 또한, 국가 별로 상이한 보건의료체계 및 자원, 임상지침 및 관련 규제, 사회인구학적인 건강상의 차이, 생물학적인 차이 등으로 인해 외국에서 수입한 AI 관련 데이터와 시스템을 그대로 적용하기에는 한계가 뚜렷하다.

이러한 상황에서 AI의 오류 또는 비적합 발생 시 개발자와 의사 중 누구에게 귀책 사유가 있는지가 불분명해진다. AI의 비투명성과 자동 학습으로 인해 인간의 통제 및 개입 가능성이 낮아질수록 기존 법 원칙에 따른 책임 판단이 어려워지며, 이는 '책임 공백' 문제로 이어진다. 특히 AI가 여러 알고리즘과 의료기기, 로봇 및 외부 장치와 연계되어 작동할 경우, 문제의 원인을 판별하기 어려워 소프트웨어 개발자, 의료기기 개발자, 의사 간의 책임 소재가 모호해질 수 있다. 또한, AI의 진단 결과가 의사의 판단과 달라질 경우 책임의 범위 역시 달라질 수 있으며, 이는 AI 발전이 의료인의 방어 수단

이 되거나 반대로 의료인의 신뢰를 저하시킬 가능성을 내포한다. 이처럼, 보건의료 분야에서 AI 활용에 따른 적절한 법적 책임 체계 마련이 필요하다.

3 의사의 설명의무 및 환자의 자기결정권에 대한 영향

AI 소프트웨어나 이를 탑재한 의료기기를 사용한 진료로 인해 손해가 발생할 경우, 의사의 설명의무 문제가 제기될 수 있다. 의사가 의료과실에 대한 책임을 지기 위해서는 우선 설명의무와 주의의무가 위반되어야 하며, 이러한 위반으로 인한 인과적 손실이 입증되어야 한다.

「의료법」 제24조의2는 의료인이 환자의 생명이나 신체에 큰 위협을 초래할 가능성이 있는 수술, 수혈, 전신마취와 같은 중대한 의료 행위에 대해 환자에게 설명하고 서면 동의를 받도록 규정한다. 다만, 긴급한 경우에는 절차가 지체될 수 있다. 설명의무는 중대한 위해를 전제로 하기에 현재 AI 의료기기가 의료기기로서의 법적 지위에 머무는 한, 기존의 의사의 과실 책임 체계에 큰 변화가 필요하지는 않다.

그러나 의사가 검증되지 않은 치료법이나 의약품을 사용할 경우 기존의 방법과 비교하여 설명할 의무가 있듯이, 향후 AI를 활용한 진단이나 치료 권고를 시행할 때는 환자에게 이를 설명할 의무가 더욱 중요해질 수 있다. 다만, 어떤 수준의 AI 활용이 설명의무를 필요로 하는지는 명확하지 않다. 특히 진단 단계에서는 AI의 판단에 따라 이후의 치료법이 결정되므로, 의사는 AI의 분석 알고리즘과 오류 가능성을 충분히 이해하고, 이를 환자에게 설명해야 한다. 이는 환자가 AI의 신뢰성과 오류 가능성을 올바르게 이해하지 못해 AI의 역할을 과대평가하거나 의사와의 관계를 오해하지 않도록 돕기 위해서도 중요하다.

반면 의사가 충분한 설명을 제공하지 못하면 환자의 자기결정권이 제한될 수 있다. 환자는 의료계약을 통해 자신의 신체와 관련된 결정을 자율적으로 할 권리가 있으며, 이는 의사의 충분한 설명과 환자의 동의를 바탕으로 이루어진다. 따라서 AI의 안전성 및 법적 책임 문제를 해결하려면 의료 전문가, AI 전문가, 법률 및 윤리 전문가, 시민 등 이해관계자 간의 사회적 논의가 필요하다. 이러한 논의를 통해 사회적 책임의 배분과 실질적 배상책임, 그리고 절차적 증명책임을 법제화할 필요가 있다.

제5절 인간의 기술통제

NATIONAL ASSEMBLY FUTURES INSTITUTE

현재는 인간이 AI를 통제하고 있지만, 먼 미래에 AI가 고도의 자율성과 합리성을 갖추고 인간의 지적 수준을 초과하는 ‘강인공지능(Strong AI)’에 이르게 되면 기존의 법과 윤리 체계가 위협받을 가능성이 있다. 강인공지능은 외부 데이터나 인간의 통제를 받지 않고도 자율적으로 학습하고 판단을 내릴 수 있다. 이러한 수준에 도달할 경우, 기존 법제에서 정한 제조물이나 의료기기의 범위를 넘어서는 능력을 갖추게 되어, 기존의 민법과 형법에 따라 법적 책임을 규명하기가 어려워진다.

이러한 강인공지능에 대해 법적 인격이나 윤리적 주체성을 인정할지, 범죄에 대한 의지와 능력을 갖추었다고 볼 수 있는지, 법적 책임을 부과할 수 있는지에 대한 여부가 근본적인 쟁점이 될 수 있다. 현재의 법인격은 스스로 인지하고 판단하며 의사를 표현할 수 있는 경우를 대상으로 하기에 강인공지능이 이 기준을 만족할 가능성은 있으나, 복제나 변형이 없는 존재로 전제하고 인권 등 기본권을 존중해야 한다는 점에서는 현행 법적 인격의 개념과 일치하지 않는다. 따라서 AI를 법체계 내에서 인격화하기 위해서는 인간에 준하는 지위를 부여하거나, 인간 개념을 재정립하거나, 법인격을 부여하는 방안을 모색해야 할 필요가 있다.

독일에서는 법적으로 물건과 인간 사이에 동물이라는 제3의 주체를 정의한 바 있다. 1990년 민법을 개정하여 신설된 제90a조에 따라 동물은 물건이 아니며 별도로 법적으로 보호받는다고 규정하여 사람과 물건의 중간 지위를 마련한 것이다. 이와 유사한 논리로, 법적 주체로서 로봇이나 AI 소프트웨어를 인정할 가능성이 열릴 수 있다. 유럽연합(EU)에서는 이보다 더 나아가 아이작 아시모프가 제안한 ‘로봇 3원칙’을 바탕으로 로봇의 법적 책임을 규정하는 ‘로봇시민법(European Civil Law Rules on Robotics)’ 결의안을 제정하였으며, 이후 2017년 2월 AI 로봇에 대해 전자적 인격(electronic personhood)을 인정하는 결의안도 통과시켰다. 이 결의안은 AI 발전이 인류에 기여할 수 있도록 책임 있는 행동과 인류애를 지향해야 한다는 취지를 담고 있으며, AI 로봇이

사회, 환경, 그리고 인체 건강에 미치는 영향에 대한 법적, 윤리적 기준을 마련하고 책임 소재를 명확히 하는 방안을 제시하고 있다.

우리나라에서도 2017년 7월 박영선 의원이 ‘로봇기본법’을 대표 발의하여 로봇을 “외부 환경을 인식하고 상황을 스스로 판단하여 자율적으로 동작하는 기계 장치나 소프트웨어”로 정의하고, 로봇이 인간의 윤리 규범을 준수하도록 하며 로봇이 발생시킨 손해에 대해 책임 부여와 보상 방안을 마련할 것을 제안하였다. 또한, 국무총리 산하에 국가로봇윤리정책위원회를 설치하고, 국가로봇정책연구원을 설립하여 로봇으로 인한 영향을 사회적으로 대비할 수 있는 방안을 제안하였다.

강인공지능의 구현 시기에 대해서는 여전히 학계에서 다양한 예측과 논쟁이 있지만, 기술의 빠른 발전 속도를 감안할 때, 이러한 법적 및 윤리적 문제를 해결하기 위한 논의는 계속해서 이어져야 할 것이다.

제6절

효율성의 평가: 국내 보건의료 지불보상제도 개선방안

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 디지털치료기기 및 보건의료 AI에 대한 지불보상제도 접근 현황

디지털치료기기 및 보건의료 AI가 도입된 SaMD 전반의 개발 및 활성화를 장려하기 위해 2022년 12월 ‘의료기기 품목 및 품목별 등급에 관한 규정’이 개정되어 디지털헬스기기 등 의료기기의 허가·인증·신고 등의 관리 절차를 신속화하기 위해 허가 신청과 동시에 품목 신설 절차를 진행하는 한시품목 분류제도의 도입, 혁신의료기기의 신속한 시장진입을 위해 한시적 비급여로 선사용 후(1-3년) 건강보험 등재를 위한 의료기술평가 시 시장에서의 실사용데이터(Real World Data, RWD)를 활용한 효과성 평가 허용 등 다양한 ‘개발 속도’의 관점에서 다양한 정책적 개선이 이루어져 왔지만, 여전히 ‘개발 동기’의 관점에서 핵심적인 요인이라고 볼 수 있는 지불보상제도에 대해서는 개선이 필요한 상황으로 볼 수 있다. 특히, 2023년 8월 보건복지부와 건강보험심사평가원이 발표한 ‘디지털치료기기 건강보험 등재 가이드라인’의 경우 디지털 및 인공지능 혁신의료기술이 ‘통합심사평가’를 통해 신속히 사용 승인을 받아 ‘임시등재’가 된 후, RWD에 기반한 근거를 창출한 후 새로이 신의료기술평가를 받아 정식등재가 가능한 건강보험 등재 경로를 제시하고 있으나, 이때 이들 기술들에 적용되는 지불보상제도가 실제 신기술을 개발하는 주체들에게 충분한 개발 동기를 제공하고 보건의료 시장에서 AI에 기반한 가치를 창출하기에 적합한 제도인지에 대해서는 추가적인 검토가 필요한 실정이다.

혁신의료기기 통합심사·평가를 통해 실제 현장에서 활용을 통한 RWD를 통한 근거를 생산하는 대상이 된 SaMD들은, 1년에서 3년의 RWD 확보기간 동안 건강보험에 따른 선별급여에 기반한 보상 혹은 비급여 보상을 통해 시장에서 활용될 수 있는 ‘임시등재’ 경로가 확보된다. 이때, 비급여를 선택하게 될 경우 비급여 금액신고 및 금액 공개, 청구 의무화를 통해 RWD가 수집되게 되며, 건강보험에 따른 선별급여를 적용받는 경로를 선택하게 될 경우 상대가치점수의 산출에 의거한 수가를 판단한 후, 이를 기반으로 급여청구시스템에 의거하여 RWD가 수집되는 경로를 취하게 된다. 건강보험에 따른

선별급여를 적용받는 경로를 선택하여 상대가치점수를 산출할 때에는, 업체 및 요양기관이 제출한 원가 산출결과 및 산출근거자료가 그 주된 근거로 활용되며, ‘디지털치료기기 건강보험 등재 가이드라인’에서 제시하고 있는 디지털치료기기 개발원가 산정기준은 아래 표와 같다.

[표 5-1] 디지털치료기기 개발원가 산정기준

구성 요소	산정 기준	근거 기준
① 제품개발비	- 제품개발비 소프트웨어 기능점수 × 기능점수당 단가 × 보정계수	SW사업 대가산정 기준
② 직접경비	- 제품개발을 위한 임상시험비용	국내개발신약 개발원가산출기준 (간접경비는 제품개발비 포함)
③ 단위당개발원가	- 개발원가(①+②) ÷ 3년 예상사용자수*	*시장규모(명) × 시장성장률(%) × 시장점유율(%) × 연간실투여횟수(회)
④ 이윤	- 단위당 개발원가(③)의 25%	국가를 당사자로 하는 계약에 관한 법률 시행규칙
⑤ 부가가치세	- 단위당 개발원가와 이윤 합계(③+④)의 10%	부가가치세법
⑥ 유지관리비	- 단위당 개발원가, 이윤, 부가가치세 합산액 (③+④+⑤)의 10~15%	SW사업 대가산정 기준
최종 금액	단위당 개발원가, 이윤, 부가가치세, 유지관리비 합산액(3+4+5+6)	

※ [참고] 디지털치료기기 개발원가 산정기준

이상과 같은 신청인의 제출자료, 식품의약품안전처가 확보한 임상시험자료, 국내·외 임상 근거 등을 참고하여 기준금액이 검토되며, 최종적으로는 전문평가위원회(신설 디지털의료전문평가위원회) 심의를 거쳐 SaMD의 급여여부 및 상대가치점수가 평가된다. 이상의 과정은 비록 새로이 신의료기술평가를 받아 정식등재를 거치는 과정이 아닌 임시등재 과정에 활용되는 절차이지만, 위와 같은 형태의 지불보상제도의 근간에는 몇가지 제한점이 존재한다. 첫째, 원가에 기반한 보상이 주요한 기준이 된다. 이는 의약품 시장에서 원가에 기반한 약가 결정이 주된 기전으로 활용되지 않는 점을 참고할 필요가 있다. 일반적으로 원가 기반 약가 결정은 제약사의 비용 정보를 정확히 검증하기 어렵고, 약물의 임상적 효과나 사회적 가치를 반영하지 못하는 한계가 있으며, 이에 따라 임상적 효능과 비용 효과성을 반영하는 가치기반 약가(Value-Based Pricing)나 성과 기

반 계약(Outcome-Based Contracting) 등이 대안으로 활용되고 있음을 참고할 필요가 있다. 둘째, 별도수가 기반 보상이 이루어지는 혁신의료기기의 범위에 포함되는 SaMD는 특정 적응증을 대상으로 임상적 효과성의 개선을 보이는 경우로 한정된다. 이는 특히 고령의 복합만성질환자에서의 복잡한 케어 경로(care pathway)를 최적화하거나 개인의 생활습관 및 만성질환 관리를 지원함에 따라 달성할 수 있는 효율성 개선에 대해서는 보상할 수 있는 경로가 제한된다는 점에서 한계가 있다. 예를 들어, 현재의 지불보상제도 내에서 단일 의료기관 내에서의 서비스 공급을 효율화함에 있어 AI 기술을 활용하는 것은 원가절감 및 이익률 개선을 달성할 여지가 있지만, 환자를 중심으로 여러 의료기관간 서비스를 최적화하고 환자의 자기관리(self-care)를 지원하는 등 향후 거대언어모델(Large Language Model, LLM)에 기반한 서비스가 활용될 가능성이 높은 분야에 대해서는 현재의 지불보상제도를 토대로 경제적 이윤을 창출할 수 있는 경로가 제한되는 문제가 존재한다.

2 가치기반 대안적 지불보상제도와 보건의료 AI 도입의 연관성

앞서 기술한 내용 중 원가에 기반한 수가 결정이 아닌, 경제성 평가 등 현재 의약품의 약가 산정에 활용되는 메커니즘을 준용하여 가치를 반영한 수가 결정체계를 접목하는 것은 독일의 ‘디지털 헬스 애플리케이션(DiGA)’ 제도 등 승인된 디지털 치료기기에 대해 임상적 효과와 경제성을 검증하며 가치를 기반으로 보상을 결정하는 메커니즘을 도입한 사례들에서 이미 확인되고 있는 접근이다. 다만, 이와 같이 명시적인 SaMD에 대한 직접적인 지불보상이 아닌, 디지털 및 AI 기술을 보건의료 분야에 접목하고자 할 때 소외되기 쉬운 영역 중에 고령의 복합만성질환자에서의 복잡한 케어 경로(care pathway)를 최적화하거나 개인의 생활습관 및 만성질환 관리를 지원함에 따라 달성할 수 있는 효율성 개선에 대해 보상하는 대안적 지불보상제도가 활용되지 않고 행위별 수가제도가 중심이 된 한국의 의료체계의 제한점에 대해서도 관심이 필요하다(Ju et al., 2022)

고령화와 함께 복합만성질환자의 비중이 높아지고, 이에 따라 다양한 의료기관으로부터 복수의 질병에 대한 관리, 더 나아가서는 각종 장기요양 서비스를 포함한 사회서비스의 활용이 혼재되어 나타나는 고위험-고의료비(high-risk, high-cost) 집단이 늘어

나는 상황에서, 해외 선진국들에서 공통적으로 나타나는 접근 중 하나는 다양한 공급자 집단이 네트워크를 형성하고, 해당 네트워크가 협력을 통해 환자 중심의 서비스를 제공함으로써 시스템의 효율성 개선에 기여할 경우 네트워크 단위로 보상을 제공하는 등 책임의료조직(Accountable Care Organization, ACO)으로 대표되는 대안적 지불보상 제도의 활용이다(van Kessel et al, 2023). 이와 같은 대안적 지불보상제도가 존재하는 경우와, 현재 한국과 같이 행위별수가제 중심의 보상만이 존재하는 경우에는 보건의료 분야에서의 AI 혁신이 일어날 수 있는 여지의 측면에서도 차이점이 존재한다는 점에 주목할 필요가 있다. 예를 들어, ACO와 같은 네트워크 기반 지불보상 기전이 존재할 경우에는, ACO 내부에 소속된 다양한 기관 간 협력을 통해 서비스 경로를 최적화하고, 환자 대상의 교육 및 생활습관 관리 지원 등을 통해 환자의 건강 결과, 비용 결과, 만족도로 대변되는 결과 전반을 개선할 유인이 존재하며, 이에 도움을 줄 수 있는 혁신적인 기술기반 서비스들을 수용하고 활용할 가능성이 증가할 수 있다. 이에 반해, 행위별수가제 하에서 환자의 적응증에 대한 특정 처치행위를 중심으로 보상이 이루어지는 체계 하에서는, 의료공급자들에게 이와 같은 유인이 없으며, 디지털 치료기기와 같이 특정 적응증을 대상으로 임상적 효과성을 확보하고 수가가 부여된 매우 제한적인 경우에만 이를 활용하고자 하는 동기가 발생할 수 있다.

3 국내에서의 대안적 지불보상제도 논의와 보건의료 AI의 접목 가능성

2024년 2월 발표된 제2차 국민건강보험 종합계획에서는 ACO의 시범사업 추진을 제시한 바 있다. 다만, 현재까지 제시된 내용에서는 이를 지역의 거점의료기관을 중심으로 한 특성화 사업이나, 시범사업 지역 내 의료기관 간 연계·협력을 지원하는 등 매우 제한적인 범위에서 제시하고 있다. 미국의 ACO의 도입에서 민간의 경영지원회사들이 다양한 기술 및 정보시스템 등 인프라 구축에서 핵심적인 역할을 담당했던 것을 고려할 때, 또한 ACO 내부에서의 케어 경로 최적화나 환자 중심의 자기관리 지원 등 AI 등 기술기반 혁신이 가능한 영역이 기존의 지불보상제도 내에서는 충분히 보상받기 어렵지만 ACO 기반의 지불보상제도에서는 보상이 가능한 점을 고려할 때, 향후 계획된 시범사업

들에서 이와 같은 보건의료 AI 기반 기술의 활용을 위한 안전성 검토의 기준, 기술기반의 민간 파트너들이 보건의료기관 중심의 ACO와 협력하여 비즈니스 모델을 개발할 수 있는 제도적 검토 등이 필요할 것이다.

제6장

결론

본 보고서는 보건의료 분야에서 AI 기술의 장점을 극대화하고 잠재적인 문제점이 될 수 있는 법윤리적 이슈에 대한 문제 제기와 입법방안을 제안하기 위해 AI 기술과 법윤리적 내용을 균형 있게 다루면서 미래 보건의료의 발전과 변화에 선제적으로 대응하고자 하였다. 국제기구 및 주요국들은 국가적 차원에서 AI 기술을 발전시키고 활용하기 위한 전략을 수립하였으며 이러한 전략들 중 상당수는 AI 기술로 인해 발생할 수 있는 윤리적인 문제에 대한 전략을 포함하고 있었다. 보건의료 분야의 윤리적 문제에 대해서 대부분의 국제기구와 주요국은 개별적인 전략을 가지고 있으나 현재 한국은 국가적 차원에서 보건의료 AI 분야에 특화된 윤리적 가이드라인이 마련되지 않은 상태이다. 그러나, 그 필요성에 대한 공감대가 형성되고 있어 관련 논의와 연구가 활발히 진행되고 있다.

향후 AI가 보건 분야에서 적극적으로 활용되게 될 경우 이슈가 될 수 있는 5가지 쟁점을 중심으로 연구를 수행하였다. 개인정보 보호 및 설명가능성의 경우, AI의 정확성과 상충관계(trade-off)를 보이고 있으며 그 외에도 다양한 윤리 가치를 준수하기 위한 개발과정에서의 추가적인 고려는 AI의 개발 비용 및 사회적 비용을 상승시킬 것으로 예상된다. 다행히도 이러한 문제들은 기술적인 차원에서도 해결 방안을 모색하고 있으며 일정 수준의 가능성들을 보이는 성과를 내고 있다. 따라서 보건의료 분야의 AI 활용을 위해서는 이러한 윤리적 차원의 문제, 특히 설명가능성과 공정성의 한계를 기술적으로 해결하려는 연구 등에도 국가적 차원의 지속적인 지원이 이뤄져야 할 것이다.

아울러 정책적, 사회적 차원에서 보건의료 분야의 AI 개발 및 활용을 촉진하기 위한 중장기 전략은 크게 3가지로 정리하고자 한다. 첫 번째는, 국가적 차원에서 보건의료분야에 특성화된 AI 개발 및 활용에 대한 명확한 규제 지침과 윤리적 프레임워크를 마련하여 환자 안전과 데이터 보안을 보장하기 위한 가이드라인이 필요하다. 보건의료 분야는 유럽연합의 보고서에 따르면 AI가 활용될 시 고위험이 잠재되어 있는 분야이다. AI에 국한되지 않더라도 생명 윤리 등 기존의 윤리적 논쟁들이 향후 AI로 인해 확대될 것으로 예상된다. 이에 따라 국가적 차원에서 보건의료 분야에 국한된 가이드라인을 만들고 이를 통해 유연하고 점진적인 법규제를 확립해 나갈 필요가 있다. 또한, 책임성과 관련된 책임 주체 및 처벌 등에 관련된 내용들은 법제화 할 필요가 있다. 아울러, AI 기술의 발전이 매우 빠르기 때문에 중장기적으로는 AI 기술이 특정 분야에서 의사의 전문성을 뛰어넘거나 대체할 경우, 또는 의사의 진단과 처방 등을 받기 어려운 특정 맥락과

조건에서 의사를 대체할 수 있을 경우에 발생할 수 있는 여러 법윤리적 문제들에 대해 선제적으로 논의하는 것이 필요하다. 이러한 내용은 보건의료 분야에 특화된 감독기구 및 절차의 법제화를 통해 독립적이고 확고하게 자리잡을 수 있을 것으로 판단된다.

두 번째로, 보건의료 분야의 AI 기술 발전 및 활용을 위한 다학제적인 전문가의 구성이다. 주요국의 경우, AI의 발전이 단순한 과학기술 분야에 국한한다고 보지 않고 인문학자 및 사회과학자가 대거 포함된 연구를 권장하고 있다. 윤리학자 뿐만 아니라 생물 및 의료분야 전문가와 더불어 데이터공학자와 정책 연구자, 법학자, 환자 및 일반인 등이 포함된 국가 주도의 보건의료 연구 생태계가 조성될 필요가 있다. 특히, 보건의료 분야 AI 기술에 필요한 기술의 위험성 및 효과성 등의 평가는 다학제적인 전문가가 수행하는, 임상적, 윤리적, 사회적 측면에서의 다면적인 평가가 수행될 필요가 있다.

마지막으로, 사회적 합의 과정을 통한 보건의료 분야의 법윤리적인 문제 해결 전략이 필요하다. 개인정보 보호 가능성을 얼마나 희생하고 AI의 효용성을 높일 것인가, 보건의료 분야에서 AI의 설명가능성은 어느 눈높이까지 맞출 것인가, 공정한 AI 기술 발전과 활용을 위해 합의 과정에 참여시켜야 하는 집단과 계층은 어디까지 인가 등에 대한 사회적 논의와 합의의 형성과정은 필요한 여러 지점이 있다. 특히, 국회가 국민 논의 및 합의의 장으로서 합의 형성의 장이 될 필요가 있다.

참고문헌

1. 문헌자료
2. 웹사이트

참 고 문 헌

NATIONAL ASSEMBLY FUTURES INSTITUTE

1 문헌자료

- 강달천(2020), “데이터 3법 개정의 주요 내용과 전망”, KISA.
- 국가인권위원회(202), 개인정보 보호법 등 ‘데이터 3법’ 국회 통과에 대한 국가인권위원장 성명, 보도자료
- 김광수(2018), 인공지능 규제법 서설, 토지공법연구, (81):292.
- 김승현(2019), 인공지능(AI)의 의료활용과 주요이슈, KPC4IR ISSUE PAPER 7.
- 김지연(2020), “비대면 시대, 비대면 의료 국내외 현황과 발전방향”, KISTEP ISSUE PAPER.
- 문세영 외, “정밀의료의 성공전략”, KISTEP I&I, 2016
- 삼정KPMG, 코로나19 그 이후, 헬스케어 산업에 불어오는 변화의 바람(2022)
- 식품의약품안전평가원(2020), 2020년 신개발 의료기기 전망 분석 보고서.
- 윤지영, 김한균, 감동근, 김성돈(2017), 「법과학을 적용한 형사사법의 선진화 방안(VIII)」, 한국형사정책연구원.
- 이도국(2017), 인공지능의 민사법적 지위와 책임에 관한 소고, 법학논총, 34(4) :324.
- 이은자 외(2009), “원격의료 시대를 대비한 정보보호 법제화 방향”, J Kor Soc Med Informatics, 15(4):361-371.
- 원동규, 이상필(2016). 인공지능과 제4차 산업혁명의 함의. ie 매거진, 23(2), 13-22.
- 정규환(2018), 인공지능 기반 의료영상 분석 기술 동향, 기획시리즈-인공지능, 정보통신기술진흥센터.
- 조영임(2016). 인공지능 기술 동향 및 발전 방향, 정보통신기술진흥센터.

철학미학 정동과 정서 『정동의 힘』, 26~38.

한국보건산업진흥원(2018), 인공지능(AI) 기반 의료기기 현황 및 이슈(1), KHIDI BRIEF.

한국지능정보사회진흥원 정책본부 지능화법제도팀, 「캐나다 '인공지능 및 데이터 법안'의 주요 내용과 시사점」, 『디지털 법제 Brief』, NIA지능정보원, 2022년 07월 11일

홍진(2020). '인공지능', '기계학습', '딥 러닝'분야의 국내 논문 동향 분석. 한국정보전자통신기술학회 논문지, 13(4), 283-292.

허종호(2024) 2024년 『국민 미래인식·가치관 조사』 주요 결과, 국가미래전략 Insight, 111호, 국회미래연구원

2 웹사이트

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine Bias. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Ashley, E. A. (2016). "The Precision Medicine Initiative: A New National Effort." *Journal of the American Medical Association*, 315(8), 797-798. doi:10.1001/jama.2016.0372.

ASTI MARKET INSIGHT 2022-065

Brisimi, T. S., Chen, R., Mela, T., Olshevsky, A., Paschalidis, I. Ch., & Shi, W. (2018). Federated learning of predictive models from federated Electronic Health Records. *International Journal of Medical Informatics*, 112, 59-67.

Chen, R. J., Wang, J. J., Williamson, D. F. K., Chen, T. Y., Lipkova, J., Lu, M. Y., Sahai, S., & Mahmood, F. (2023). Algorithmic fairness in artificial intelligence for medicine and healthcare. In *Nature biomedical engineering* (Vol. 7, Issue 6, pp. 719-742). NLM (Medline).

Cho, K. J., Kim, J. S., Lee, D. H., Lee, S. -M, Song, M. J., Lim, S. Y., Cho, Y. J., Jo, Y. H., Shin, Y., & Lee, Y. J. (2023). Prospective, multicenter val

- idation of the deep learning-based cardiac arrest risk management system for predicting in-hospital cardiac arrest or unplanned intensive care unit transfer in patients admitted to general wards. *Critical Care*, 27(1).
- Collins, F. S., & Varmus, H. (2015). "A New Initiative on Precision Medicine." *The New England Journal of Medicine*.
- Defense Advanced Research Projects Agency (DARPA), (2020) Explainable Artificial Intelligence(XAI), Dr. Matt Turek, <https://www.darpa.mil/program/explainable-artificial-intelligence>.
- Desai, D. R., & Kroll, J. A.(2017). Trust but verify: A guide to algorithms and the law. *Harv. JL & Tech.*, 31, 1.
- di Martino, F., & Delmastro, F. (2023). Explainable AI for clinical and remote health applications: a survey on tabular and time series data. *Artificial Intelligence Review*, 56(6), 5261-5315.
- DM Altshuler et al.(2010), "A map of human genome variation from population-scale sequencing". *Nature*, 467(7319):1061-1073.
- Edwards, L., & Veale, M.(2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke L. & Tech. Rev.*, 16
- Erhan, D., Courville, A., Bengio, Y., & Vincent, P. (2010, March). Why does unsupervised pre-training help deep learning?. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 201-208). *JMLR Workshop and Conference Proceedings*.
- Ertel, W. (2018). *Introduction to Artificial Intelligence*. Springer.
- FDA. (2020). *Clinical Performance Assessment: Detection Devices Applied to Radiology Images and Radiology Device Data in - Premarket Notification (510(k)) Submissions Guidance for Industry and FDA Staff*.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5). <https://doi.org/10.1214/aos/10>

13203451

- Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., & Schultz, N. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science Signaling*, 6(269).
- Garraway, L. A., Verweij, J., & Ballman, K. V. (2013). "Precision Oncology: An Overview." *The Lancet Oncology*, 14(12), e283-e292. doi:10.1016/S1470-2045(13)70359-7.
- Gaube, S., Suresh, H., Raue, M., Merritt, A., Berkowitz, S. J., Lermer, E., Coughlin, J. F., Guttag, J. v., Colak, E., & Ghassemi, M. (2021). Do as AI say: susceptibility in deployment of clinical decision-aids. *Npj Digital Medicine*, 4(1), 1-8.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI magazine*, 38(3), 50
- Guo, L. L., Pfohl, S. R., Fries, J., Posada, J., Fleming, S. L., Aftandilian, C., Shah, N., & Sung, L. (2021). Systematic Review of Approaches to Preserve Machine Learning Performance in the Presence of Temporal Dataset Shift in Clinical Medicine. In *Applied Clinical Informatics* (Vol. 12, Issue 4, pp. 808-815). Georg Thieme Verlag.
- Heslin, K. C., Owens, P. L., Karaca, Z., Barrett, M. L., Moore, B. J., & Elixhauser, A. (2017). Trends in Opioid-related Inpatient Stays Shifted After the US Transitioned to ICD-10-CM Diagnosis Coding in 2015. *Medical Care*, 55(11), 918-923.
- Howard, F. M., Dolezal, J., Kochanny, S., Schulte, J., Chen, H., Heij, L., Huo, D., Nanda, R., Olopade, O. I., Kather, J. N., Cipriani, N., Grossman, R. L., & Pearson, A. T. (2021). The impact of site-specific digital histology signatures on deep learning model accuracy and bias. *Nature Communications*, 12(1).

- Irsoy, O., & Cardie, C. (2014). Deep recursive neural networks for compositionality in language. *Advances in neural information processing systems*, 27, 2096-2104.
- Ji, Y., Cho, K. H., Chang, M., Joo, S., Park, M., Kim, K. G., Lee, Y., Kim, M. C., Sim, D. S., Hong, Y. J., kim, J., Ahn, Y. K., Jeong, M. H., Shin, D.-G., Cho, J. R., Kang, M.-K., Seonghoon, C., Lee, N., & Han, D. (2023). Abstract 13745: Exploring the Segment-Wise Explainability of Electrocardiogram on Deep Learning-Based Models for Acute Myocardial Infarction. *Circulation*, 148(Suppl_1).
- Ju, J. H., Sim, B., Lee, J., & Lee, J. Y. (2022). Reimbursement of digital therapeutics: future perspectives in Korea. *Korean Circulation Journal*, 52(4), 265-279.
- Juengst, E. T., McGowan, M. L., Fishman, J. R., & Settersten, R. A. (2016). "From 'Personalized' to 'Precision' Medicine: The Ethical and Social Implications of Rhetorical Reform in Genomic Medicine." *Hastings Center Report*, 46(5), 21-33. doi:10.1002/hast.614.
- Khoury, M. J., & Galea, S. (2016). "Will Precision Medicine Improve Population Health?" *JAMA*, 316(13), 1357-1358.
- Kim HS. Decision-making in artificial intelligence: is it always correct? *J Korean Med Sci* 2020;35:e1.
- Kim HS, Kim DJ, Yoon KH. Medical big data is not yet available: why we need realism rather than exaggeration. *Endocrinol Metab (Seoul)* 2019;34:349-54.
- Kim HS, Kim JH. Proceed with caution when using real world data and real world evidence. *J Korean Med Sci* 2019;34:e28.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.
- Kundu, S. (2021). AI in medicine must be explainable. *Nature Medicine*, 27(8),

1328-1328.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.

LeCun, Y., Touresky, D., Hinton, G., & Sejnowski, T. (1988). A theoretical framework for back-propagation. In *Proceedings of the 1988 connectionist models summer school (Vol. 1, pp. 21-28)*. CMU, Pittsburgh, Pa: Morgan Kaufmann.

Lee, S. Y., Ha, S., Jeon, M. G., Li, H., Choi, H., Kim, H. P., Choi, Y. R., Hoseok, I., Jeong, Y. J., Park, Y. H., Ahn, H., Hong, S. H., Koo, H. J., Lee, C. W., Kim, M. J., Kim, Y. J., Kim, K. W., & Choi, J. M. (2022). Localization-adjusted diagnostic performance and assistance effect of a computer-aided detection system for pneumothorax and consolidation. *Npj Digital Medicine*, 5(1), 107.

Lewiecki, E. M., Wright, N. C., & Singer, A. J. (2020). Racial disparities, FRAX, and the care of patients with osteoporosis. *Osteoporosis International*, 31.

MarketsandMarkets (2023) Artificial Intelligence(AI) Market

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, 51(4), 584-591.

Max Tegmark(2017). 『Life 3.0』. Vintage Books.

McCarthy, J. (2007). *What Is Artificial Intelligence?*. Stanford University.

McCadden, M. D., Joshi, S., Mazwi, M., & Anderson, J. A. (2020). Ethical limitations of algorithmic fairness solutions in health care machine learning. In *The Lancet Digital Health (Vol. 2, Issue 5, pp. e221-e223)*.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6), 1-35.

Michel G Gauthier(2007) "Simulation of polymer translocation through small

- channels: A molecular dynamics study and a new Monte Carlo approach”, Thesis for Doctor of Philosophy in Physics.
- Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd ed.).
- Mullis, K.F.; Faloona, F.; Scharf, S.; Saiki, R.; Horn, G.; Erlich, H. (1986). “Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction”. *Cold Spring Harbor Symposia on Quantitative Biology*. 51: 263-273.
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464).
- oynt Maddox, K. E., Reidhead, M., Hu, J., Kind, A. J. H., Zaslavsky, A. M., Nagasako, E. M., & Nerenz, D. R. (2019). Adjusting for social risk factors impacts performance and penalties in the hospital readmissions reduction program. *Health Services Research*, 54(2), 327-336.
- Park, S., Ock, C.-Y., Kim, H., Pereira, S., Park, ; Seonwook, Ma, M., Choi, S., Kim, S., Shin, S., Brian, ; Aum, J., Paeng, K., Yoo, D., Cha, H., Park, S., Koung, ;, Suh, J., Hyun, ;, Jung, A., ... Lee, S.-H. (2022). Artificial Intelligence-Powered Spatial Analysis of Tumor-Infiltrating Lymphocytes as Complementary Biomarker for Immune Checkpoint Inhibition in Non-Small-Cell Lung Cancer. *J Clin Oncol*, 40, 1916-1928.
- Quaggin, S. E., & Palevsky, P. M. (2021). Removing Race from Kidney Disease Diagnosis. *Journal of the American Society of Nephrology*, 32(12), 2987-2989.
- Reddy, S. (2022). Explainability and artificial intelligence in medicine. In *The Lancet Digital Health* (Vol. 4, Issue 4, pp. e214-e215). Elsevier Ltd.
- Relling, M. V., & Evans, W. E. (2015). “Pharmacogenomics in the Clinic.” *Nature*, 526(7573), 343-350. doi:10.1038/nature15817.

- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should i trust you?” Explaining the predictions of any classifier. Proceedings of the ACM SIG KDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016, 1135-1144.
- Schork, N. J. (2015). “Time for One-Person Trials.” *Nature*, 520(7549), 609-611. doi:10.1038/520609a.
- Seyyed-Kalantari, L., Zhang, H., McDermott, M. B. A., Chen, I. Y., & Ghassemi, M. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Medicine*, 27(12), 2176-2182.
- STAT(5 September 2017) IBM pitched its Watson supercomputer as a revolution in cancer care. It’s nowhere close. <<https://www.statnews.com/2017/09/05/watson-ibm-cancer/>>
- The UK10K Consortium(2015), “The UK10K project identifies rare variants in health and disease”, *Nature*, 526:82-90
- Topol, E. J. (2019). *Deep Medicine: How Artificial Intelligence Can Make Health care Human Again*.
- Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & van Moorsel, A. (2020). The relationship between trust in AI and trustworthy machine learning technologies. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 272-283
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 49, 433-46
- UNESCO. (2021) Recommendation on the ethics of artificial intelligence [Internet]. Paris
- van der Velden, B. H. M., Kuijf, H. J., Gilhuijs, K. G. A., & Viergever, M. A. (2022). Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. In *Medical Image Analysis (Vol. 79)*. Elsevier B.V.
- van Kessel, R., Srivastava, D., Kyriopoulos, I., Monti, G., Novillo-Ortiz, D.,

- Milman, R., ... & Mossialos, E. (2023). Digital health reimbursement strategies of 8 European countries and Israel: scoping review and policy mapping. *JMIR mHealth and uHealth*, 11(1), e49003.
- Warraich, H. J., Tazbaz, T., & Califf, R. M. (2024). FDA perspective on the regulation of artificial intelligence in health care and biomedicine. *JAMA*.
- Wen, D., Khan, S. M., Xu, A. J., Ibrahim, H., Smith, L., Caballero, J., Zepeda, L., de Blas Perez, C., Denniston, A. K., Liu, X., & Matin, R. N. (2022). Characteristics of publicly available skin cancer image datasets: a systematic review. In *The Lancet Digital Health* (Vol. 4, Issue 1, pp. e64-e74).
- Yang, W., Wei, Y., Wei, H., Chen, Y., Huang, G., Li, X., Li, R., Yao, N., Wang, X., Gu, X., Amin, M. B., & Kang, B. (2023). Survey on Explainable AI: From Approaches, Limitations and Applications Aspects. *Human-Centric Intelligent Systems*, 3(3), 161-188.

Abstract

2024 Research Report on the Utilization of Artificial Intelligence in the Healthcare Sector and Legislative Strategies

NATIONAL ASSEMBLY FUTURES INSTITUTE

The “2024 Research Report on the Utilization of Artificial Intelligence in the Healthcare Sector and Legislative Strategies” explores the rapidly advancing integration of AI technologies in healthcare and addresses the accompanying ethical, legal, and institutional challenges. Leveraging global and domestic case studies, the report examines the transformative potential of AI in precision medicine, disease prevention, diagnostics, treatment, and healthcare systems. It also highlights key national and international strategies, including those from the U.S., Canada, Singapore, and the EU, aimed at fostering responsible AI development and application in healthcare. Through a comprehensive analysis, the report identifies critical areas such as data protection, transparency, explainability, fairness, accountability, and human oversight, proposing mid-to-long-term legislative frameworks tailored to South Korea's healthcare system. The findings underline the importance of balancing innovation with ethical safeguards to ensure safe, equitable, and effective AI adoption, contributing to the sustainable growth of the healthcare sector.

보건의료 분야
인공지능 활용과 입법전략

발행 2024년 12월 30일
발행처 국회미래연구원
주소 서울시 영등포구 의사당대로 1
전화 02)786-2190
팩스 02)786-3977
홈페이지 www.nafi.re.kr
인쇄처 (주)케이에스센세이션(02-761-0031)

©2024 국회미래연구원

ISBN 979-11-94650-06-5 (95510)

