AI에 대한 미국의 사회·윤리 연구 동향 : 정부와 학계

Social and ethical research trends on AI in the US : Government and academia

안성원 Ahn, SungWon • **선임연구원** Senior Researcher, SPRi • swahn@spri.kr

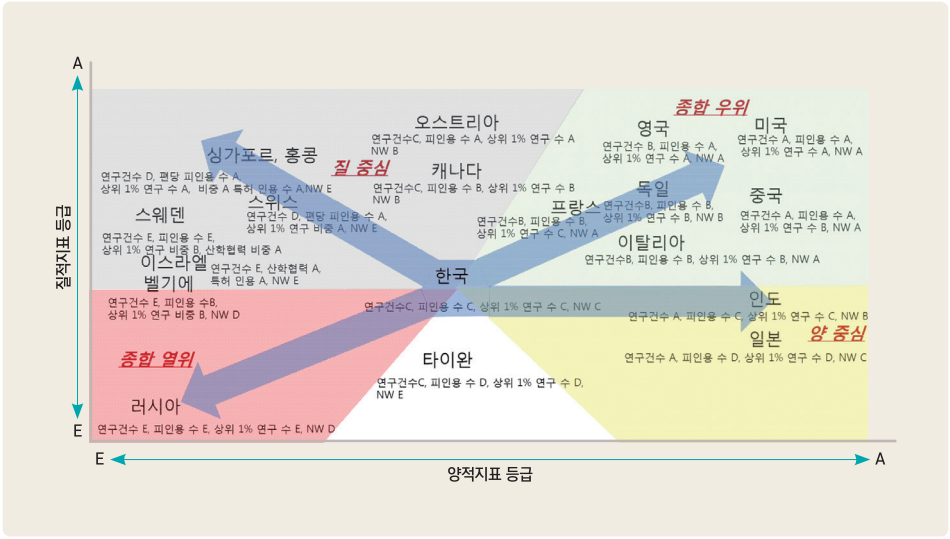
AI 분야 최선도국인 미국은 현재 AI 기술뿐만 아니라 윤리·사회·법률적 이슈들도 관심을 갖고 활발히 연구를 수행하고 있다. AI 기술의 확산과 그에 따른 사회파급 효과에 대한 충분한 고려가 있어야 부작용 없는 AI 시대를 맞이할 수 있다. 우리도 이처럼 AI 기술과 그 외적인 파급효과까지 고려하는 연구를 수행해야 한다.

The United States, the world leader in AI field, is currently investigating not only technology aspects of AI, but also ethical, social and legal issues in relation to AI. In order to prepare for the era of AI free of unexpected side effects, there must be sufficient consideration for the dissemination of AI technology and its social impact. We should also conduct research that considers not only the technical side of AI but comprehensively takes into account other consequent effects.

오래전부터 미국은 AI에 대한 적극적인 투자와 개발을 진행해 왔고, 현재 AI 분야에서 최선도국이 되어 있다. 따라서 미국은 AI 기술 수준과 격차를 판가름하는 기준이 되고 있으며, 수많은 연구 결과를 통해 전 세계 연구자와 기관들에게 참조를 제공한다.[1] 미국의 AI는 한국에 비해서 기술력은 1.8년 앞서 있으며, 기술수준은 약 28% 이상 앞서고 있다.[2] [그림 1]은 세계 주요국들의 AI 연구역량을 나타낸다. 미국은 종합 우위에 있는 6개국 중 연구의 양과 질적인 면에서 최상위에 있다.

지난 2월 트럼프 행정부는 '미국 AI 이니셔티브(American AI Initiative)' 행정명령을 발표하고 국가차원에서 AI에 대해 집중투자를 추진하고 있다. 미국 AI 이니셔티브는 연구개발, 인프라, 거버넌스, 인력, 국제적 참여 등 5대 영역을 대상으로 하며, 'AI 분야에서 미국의 리더 지위 유지'를 목표로 하고 있다. 미국 내에서는 다양한 형태의 이니셔티브(Initiative)[3]가 진행 중인데, 이들은 정부, 학교, 기업 또는 영리/비영리 단체 등이 주도하고 있다.

그림 1 인공지능 연구역량 국제비교



※ 자료 : 소프트웨어정책연구소, 인공지능 연구역량 국제비교 및 시사점, 2018.

미국은 다양한 AI 원천·응용 기술 개발과 투자만큼이나, AI로 인해 발생하는 사회적 파급효과에 대한 고려와 그에 대한 정책 연구도 활발하게 진행하고 있다.

1 미국은 2017년 기준 지난 5년간 AI분야 학술연구건수 2위(30,966건), 피인용수 1위(128,653건) – 소프트웨정책연구소, 인공지능 연구역량 국제비교 및 시사점, 2018. 참조
2 'ICT기술수준조사보고서(IITP 2017)' 인공지능 부문 기술격차를 기준으로 재산출
3 어떤 문제나 상황을 해결하거나 대응하기 위해 수행하는 프로젝트나 프로그램 등의 다양한 시도와 행위들, 정책 및 수행활동 등을 의미

정부차원의 AI 정책연구 및 지원

미국의 정부기관들은 AI로 인한 미래상을 국가 차원에서 기획하는 다양한 정책 및 기술개발 연구를 진행하고 있다. 국가 차원 전략은 AI 분야에 대한 우선적 장기투자, 인간과 AI의 협력체제 구축, 윤리·사회적 영향력을 고려한 기술개발과 제도 확충 등이다. 이 목표를 바탕으로 AI의 정책적 연구방향 제시, 관련 연구지원 사업 추진, 기술개발 등을 진행하고 있다.

1. 국립과학재단

기초과학 연구를 지원하는 국립과학재단(National Science Foundation, NSF)[4]은 오래전부터 AI에 대한 연구를 지원해 왔다. 지난 2018년 8월부터는 지원사업에 대한 10가지 핵심분야[5]를 공표하였고 그중 AI 분야의 지원 사업으로는 데이터 혁신활용(HDR)과 인간-기술 프론티어 미래직업(FW-HTF)을 추진하고 있다. [표 1]은 HDR과 FW-HTF의 주요 내용과 지원 규모를 요약한 것이다.

표 1 NSF의 HDR 및 FW-HTF 연구 지원 사업

구 분	내 용	지원 규모
데이터 혁신활용 (HDR)	• 데이터 과학자 및 엔지니어, 시스템 및 사이버 인프라 전문가로 구성된 연구소 발족(Data Science Corps., DSC)지원 - 데이터 과학 프로젝트 강의 및 실습 지원 - 참여 학생의 급여 지원 • 과학 및 공학 데이터 집중연구 프레임 워크 구축	• 3년 $100~150만 (사업에 따라 다름)
인간-기술 프론티어 미래직업 (FW-HTF)	• AI로 인해 변화하는 일자리 환경에 대한 도전 및 기회에 대응하는 것을 목표 - 인간인지 강화를 위한 기초 연구 - 구체화된 지능형 인지보조 연구 • 미래의 작업에 대한 근본적인 이해를 높이고, 작업 및 작업장, 근로자와 사회의 작업 결과에 대한 잠재적 개선 방안 도출 - 인간-기술 파트너십에 대한 이해와 개발 - 인적 성과를 높이기 위한 기술의 설계 • 신규 AI 관련 사회-기술 융합의 발굴 및 기술의 이점과 위험을 파악	• 소규모(3~5년) $75~150만 • 대규모(3~5년) $150~300만

※ 자료 : NSF(www.nsf.gov)

4 미 상무부(경제관련 업무를 수행하는 기관, Department of Commerce) 산하기관으로 미국의 과학발전과 진흥, 국민건강과 복지 향상, 국방력 확보 등의 목적으로 설립

5 미래직업, 융합연구확산, 데이터 혁신활용, 중급 연구인프라, 신규 북극탐험, NSF 2026전략, NSF Include(STEM 교육 등), 퀀텀 점프, 삶의 규칙 이해, 우주의 창 등

2. 방위고등연구계획국

방위고등연구계획국(Defense Advanced Research Projects Agency, DARPA)[6]은 국방 기술 개발을 목적으로 하는 기관이지만, 기술 상용화에도 많은 영향을 미치는 기관이다. DARPA는 주로 기술 측면에서 국방시스템을 위한 AI 연구를 진행 중이다.

DARPA에서는 AI 발전 과정을 [그림 2]와 같이 3단계 물결(Wave)이라는 시각으로 해석하고 있다. 첫 번째 물결은 특정 분야에서 몇 가지 규칙을 기반으로 만든 전문가시스템이다. 뒤에 소개하겠지만 대표적인 사례로는 MYCIN[7]이 있다. 두 번째 물결은 대규모 데이터 셋을 기반으로 학습하는 머신러닝이었다. 여기에는 AI를 훈련시키기 위해 데이터를 수집하고, 라벨링(Labeling) 및 클렌징(Cleansing)[8], 심사하는 작업에 상당한 비용이 소요되었다. 그리고 지금의 AI 연구의 세 번째 물결에서는 맥락을 이해하고 추론하는 기계 제공을 지향하는 것으로 정의하고 있다. 이에 따라, 설명 가능한 AI(eXplainable AI, X AI)[9] 연구 프로젝트를 추진하고 있다.

그림 2 DARPA의 AI 3물결

※ 자료 : DARPA Perspective on AI(www.darpa.mil)

DARPA는 2018년 9월 20억 달러 규모의 'AI 넥스트 캠페인(AI Next Campaign)' 프로젝트를 시작하였다. 이 프로젝트에는 설명 가능한 AI 알고리즘을 포함하여 국방부 비즈니스 프로세스의 자동화, AI 시스템의 신뢰성 향상, 머신러닝 기술의 보안성 및 복원력 향상, 전력 및 데이터 성능개선 등 세부 프로젝트가 추진되고 있다. 세부 AI 연구 주제는 [표 2]와 같다.

6 미 국방성 산하의 국방 연구 및 개발 부분을 담당하는 기관으로 지난 1969년 인터넷의 원형인 ARPANET을 개발하는 등 과학기술 분야에서 여러 신기술을 개발
7 박테리아를 식별하고 의사보다 우수한 수준으로 항생제를 추천하는 시스템으로 1970년에 등장
8 라벨링은 데이터의 객체에 표식을 지정하는 작업이며, 클렌징은 잘못된 데이터를 걸러내는 작업
9 AI의 딥러닝 기술은 알고리즘의 복잡성으로 인해 도출 과정을 블랙박스(Black Box)라 할 정도로 도출 결과에 대한 근거와 과정의 타당성을 설명하지 못하는 이슈가 있었는데, 이를 해결하기 위해 AI의 최종 결과에 대해 설명 가능하도록 하는 정보를 제공하는 시스템 – DARPA, Explainable Artificial Intelligence(XAI) DARPA-BAA-16-53, 2016.8. 참조

표 2 DARPA의 AI Next Campaign 연구주제

세부 주제	내 용
새로운 기능 (New Capabilities)	• 정교한 사이버 공격에 대한 실시간 분석, 가짜 이미지 감지, 자연어 기술, 자동 다중 타겟 인식, 생체의학적 보철신체의 제어, 역동적 전투 킬체인 등
강력한 AI (Robust AI)	• 공간기반의 이미지 분석, 사이버공격 경고, 미생물 시스템의 공급망 분석, 안정적인 성능 보장
적대적 AI (Adversarial AI)	• 머신러닝이 잘못된 학습데이터에 의해 오염되지 않도록 안정성을 보장
고성능 AI (High Performance AI)	• 더 낮은 전력소모로 향상된 성능을 보이는 고효율 AI 개발(1000배 빠른 속도와 1000배 향상된 전력효율)
차세대 AI (Next Generation AI)	• AI 시스템이 맥락을 이해하고, 행동을 설명하고 상식적으로 추론할 수 있도록 하는 것을 목표

※ 자료 : DARPA. AI Next Campaign(www.darpa.mil)

주로 자율적 무기체계에 관련한 연구를 수행하고 있는데, 지난 3월 DARPA에서 개최한 AI Colloquium[10]에서 윤리 및 법적, 사회적인 문제를 검토하고 있음을 밝혔다. 새로운 군사기술을 개발하는 것과 사회적 이슈 사이의 딜레마를 연구 중이며, 신기술을 통해 여러 사회적 문제를 해결하는 것을 목표로 하고 있다.

학교 및 단체 등의 AI관련 연구 활동

1. 하버드 케네디 스쿨 – The Future Society 그룹의 AI 연구

하버드 케네디 스쿨을 근간으로 하는 The Future Society 그룹은 지난 2015년 전문가 그룹이 주도하는 AI 이니셔티브(AI Initiative)를 발족했다. 이들은 글로벌 AI 정책 프레임워크를 형성하는 것을 목적으로, 범용인공지능(AGI)[11]의 확산과 수요증가에 따른 새로운 AI 거버넌스와 업무체계의 혁신을 위한 연구 등을 포함하고 있다.

AI 이니셔티브는 AI와 법(AI and the Law), 뇌 및 인지과학(Brain & Cognitive Science), AI와 헬스 의약품(AI and Health & Medicine) 등 3분야에 집중하고 있다. AI 이니셔티브는 AI의 영향을 가장 먼저, 또 크게 받을 수 있는 영역으로 법률시스템을 꼽고 있다. AI가 법률적 근거자료와 판례를 더 정확하고 빠르게 찾을 수 있고, 이를 기반으로 더 공정한 판단을 내릴 경우 우리사회가 이를 어떻게 받아들여야 하는가에 대한 해결책과 정책방향 수립을 연구하고 있다. [표 3]은 AI 이니셔티브에서 제시하고 있는 법률관련 장·단기 연구주제이다.

10 DARPA의 최신 연구결과에 대한 발표 및 민·군 AI 전문가 간 아이디어 공유·토론 목적의 학회
11 Artificial General Intelligence, 특정 문제뿐만 아니라 모든 상황에서 학습과 판단을 할 수 있는 능력을 갖춘 인공지능으로 인공지능 연구의 궁극적 목표

표 3 법률 분야 AI 연구주제

세부 분야	연구주제	
	단기	장기
AI에 의한 법적 추론 및 의사결정	• AI는 현재 변호사보다 더 빠르고 정확하게 판례를 찾을 수 있는데(NIST 연구로 판명), 이러한 AI 시스템의 법적 업무 수행에 있어서 윤리적이고 전문적인 의미는 무엇인가?	• 어떻게 법률 시스템에 AI를 통합 할 것인가? • 인간보다 더 빠르고 정확한(공정한) AI 판사에 대해 사회가 어떻게 대응할 것인가? • 인간보다 우월한 AI변호사에게 법적 작업을 위임하는 것은 적절한가?
사실자료 찾기와 개인 사생활 보호 간의 균형	• 휴대폰, 웨어러블 기기 등 개인 디바이스의 증가 및 IoT의 발달로 점점 더 많이 수집되는 개인 정보를 법적으로 활용할 것인가? 한다면 어떻게 활용할 것인가?	• AI와 인간의 지능이 융합된다면, AI로 향상된 지적·논리적 능력 및 생체리듬·감정상태·꿈 등의 기록과 추적에 있어서 어떻게 개인 정보 보호와 인간 존엄성을 묘사할 것인가? • 법적 분쟁 시 사실에 근거한 판결문과 인간의 존엄성 사이에서 어디에 선을 그을 것인가?
인간지능(HI)과 인공지능(AI)간 유무죄 판단	• 인간-컴퓨터 간의 상호작용(HCI)에 있어서 이미 다양한 분야(건강, 항공, 데이터 분석 등)에서 발생할 수 있는 오류의 책임 문제	• 인간의 정신능력이 나노컴퓨터에 의해 보완된다면, 여기서 발생하는 죄책감과 같은 감정이나 범죄행위 등의 유무죄 판단은 어떻게 할 수 있는가? 누가 책임져야 하는가?
정의에 대한 접근 보장	• 소송에 있어 더 많은 근거자료를 찾는 일은 결국 노력과 자금과 시간이 들어가는 비용이며, 이러한 비용지불 능력의 유무에 따라 소송의 승패가 갈리는 문제를 AI가 해결하는 현상	• 엄청나게 발생할 인간·인공지능 데이터를 어떻게 통제할 것인가? • 수많은 데이터를 통해 정확하고 신속하며 비용이 별로 들지 않는 판단을 어떻게 할 것인가? • 법적인 문제 해결을 점점 AI에 의존하게 되면, 인간의 정의가 오히려 불편해지는 시기가 오지 않을까?
법률 및 규정 시행	• 도시 곳곳에 비치된 감시카메라, 차량에 설치된 GPS 등으로부터 실시간으로 수집되는 각종 데이터를 기반으로 앞으로의 상황을 예측하고 대응하기 위한 AI의 활용	• 공공장소에서 얼굴을 인식하고, 행동을 녹화하며, 수많은 IoT 장치들로부터 개인의 활동 정보를 수집하는 AI가 속도위반 시 벌금 경고를 핸드폰으로 알려주는 등과 같은 서비스를 제공한다고 할 때, 이러한 서비스는 독재적이고 중앙 통제적인 사회를 만드는 것은 아닌가?

※ 자료 : AI Initiative(ai-initiative.org)

AI 이니셔티브의 또 다른 연구 주제는 [표 4]와 같은 뇌 및 인지과학 분야이다. AI 이니셔티브에서 진행 중인 이 분야 연구는 인공지능 전문가(이론 및 실무), 시민 사회단체 등이 모여 AI의 기술발전에 의한 기회와 해결과제를 논의하는 형태로 진행하고 있다. 주요 주제는 윤리문제, 데이터의 처리와 통제, 인간과 유사한 행태의 로봇, 인간노동의 대체재로써 AI 등을 다루고 있다.

표 4 뇌 인지과학 분야 AI 연구주제

세부 분야		연구주제
인공지능과 뇌 그리고	윤리	• 인간의 마음을 복제하는 능력의 발전에 따라 기계가 인간처럼 복잡한 감정을 흉내 낸다면 기계에게 권리를 부여해야 하는가? 그들은 인간이 될 수 있는가? 이를 대비하기 위해 어떤 정책을 세워야하는가?
	데이터	• 머신러닝에 의해 발견되지 않는 패턴의 데이터는 어떻게 책임감 있게 처리하며 개인 정보에 대한 보호를 할 것인가? • 어떻게 하면 인공지능이 인간의 뇌에서처럼 신경화학적인 편향성을 보이지 않고 더 도적적인 판단을 하도록 할 것인가? • 인간의 의식이 더 정교할 수 있는 AI에게 어떻게 도움이 될 수 있는가? • 이를 개발하고 통제하는 조직(공공 또는 민간)의 역할은 무엇인가?
	살아있는 것	• 우리 자신과 유사한 인공지능을 연구하기 위해 애니맷(Animat)과 하이브롯(Hybrots)을 제작하며 관찰하는 연구 - 애니맷 : 인공적으로 만든 로봇 동물 - 하이브롯 : 하이브리드 로봇으로 전자요소와 생물요소가 합쳐져 컴퓨터로 제어되는 로봇형태의 사이버네틱 유기체
	노동	• AI의 성능이 갈수록 정교해짐에 따라 높은 수준의 사무직도 대체할 수 있는 잠재력이 있으며, 앞으로 어떻게 대응해야 하는가? 어떤 정책이 필요한가?

※ 자료 : AI Initiative(ai-initiative.org)

AI 이니셔티브의 세 번째 주제는 AI와 헬스 의약품이다. AI 태동기부터 과학자들은 AI가 의료 행위를 근본적으로 바꿀 수 있음을 주장했다.[12] 의료인공지능(Artificial Intelligence in Medicine, AIM)의 시조라 할 수 있는 전문가 시스템인 MYCIN이 등장하면서 연구가 활발하게 진행되었다. MYCIN은 감영성 질병을 진단하고, 그 결과에 따라 항생제를 처방하고, 왜 그런 처방을 내렸는지에 대해 설명하는 프로그램이었다. 이 시스템은 AIM의 연구 개발에 크게 영향을 미쳤고, 이후 많은 파생연구[13] 끝에 오늘날에 이르렀다. AI 이니셔티브의 AIM 관련 연구주제는 [표 5]와 같다. 어떻게 의료 데이터를 수집하고 다룰 것인지, 수집한 데이터를 분석하여 어떻게 활용하고 설명할 것인지, 기존의 시스템을 어떻게 변화시킬 것인지 등이 주요 주제이다.

12 Robert S. Ledley, Lee B. Lusted(1959), Reasoning Foundations of Medical Diagnosis 참조
13 TEIRESIAS, EMYCIN, PUFF, CENTAUR, VM, GUIDON, SACON, ONCOCIN, ROGET 등(1984), Buchanann, B. E. Shortliffe, Rule-Based Expert Systems : The MYCIN Experiments of the Stanford Heuristic Programming Project 참조

표 5 의료인공지능(AIM) 연구주제

세부 분야	연구주제
데이터 수집	• 개인정보에 해당하는 인간 건강 데이터에 대한 기술뿐만 아니라 윤리 및 법적 사항에 대한 고려 • 데이터에 대한 품질, 무결성, 거버넌스 및 보안에 대한 고려 • 전자의무기록(EHR) 및 개인웨어러블 헬스기기 간의 데이터 통합 문제 • 환자에게 의료서비스를 제공하기 위해 수집된 의료데이터의 활용 및 야기되는 수익에 대한 윤리성 • 본래의 의도를 벗어난 목적의 의료데이터 활용에 대한 규제 • 데이터의 소유권과 접근권한 문제(EHR 제공업체와 의료기관 간의 파트너십과 시장경쟁력에 실질적 영향을 미치는 상황에 대한 고려)
지식 습득 및 표현	• 인류의 의료기술은 과거에 비해 무척 발달했지만, 아직 많은 밝혀지지 않은 영역에 대한 해결 가능성 • AIM에 의해 생성되는 의료지식에 대한 온톨로지(Ontology)[14] 보장 문제
설명	• AI가 내린 결론 및 결론을 내리는 방법에 대한 이해의 문제 • 설명 가능한 AI와 임상용으로 AIM 시스템에 대한 검증과 승인 문제
헬스 시스템 통합	• AI가 어떻게 기존의 임상시험 절차와 인력을 변화시킬 것인가? 기존의 의료계는 이를 받아들일 것인가? • 의사보다 더 뛰어난 AIM의 기술적인 분석 결과를 의사가 채택할 수 있는가? • 신뢰가 중요한 의학 분야에서 AI에 대한 신뢰문제는 어떻게 해결할 것인가? 높은 수준의 AI 안전성과 실무 임상의의 AI에 대한 우려 사이의 간극은 어떻게 좁힐 것인가?

※ 자료 : AI Initiative(ai-initiative.org)

2. 스탠포드 HAI

스탠포드(Stanford)[15]대학은 지난 3월 인간중심의 인공지능연구소인 HAI(Human-Centered AI Institute)를 설립했다. 이 연구소에서는 인간중심 인공지능 이니셔티브(HAI-Initiative)를 통해 삶의 질을 향상시킬 수 있는 인간중심의 AI 및 안전과 신뢰성이 보장되는 연구를 진행하고 있다. 연구소는 인류의 더 나은 미래를 만들기 위한 AI 기술과 인문·사회적 요소가 복합된 AI 연구를 추진하며 다양한 분야와 긴밀한 파트너십을 유지하는 것을 목적으로 한다. 연구소는 AI는 더 이상 기술적인 요소에 국한되지 않으며, 기술자, 산업계 리더, 교육자, 정책입안자, 언론인 및 기타 사회 각 분야가 힘을 모아 인류를 위한 AI 생태계를 만들어 나아가야 한다는 입장을 갖고 있다. 따라서 [표 6]과 같이 AI를 주제로 하는 각 분야의 다양한 시각을 기반으로 한 연구가 주를 이루고 있다.

14 사람들이 세상에 대하여 보고 듣고 느끼고 생각하는 것에 대하여 서로 간의 토론을 통해 합의를 이룬 바를 개념적이고 컴퓨터에서 다룰수 있는 형태로 표현한 모델, 개념의 타입이나 사용상의 제약조건들을 명시적으로 정의한 기술 - 위키피디아

15 실리콘 밸리 근처의 미국 내 가장 큰 캠퍼스를 가진 명문 대학으로 시스코, 구글, HP, 야후, 썬마이크로시스템즈 등 유명 IT 기업 CEO들을 배출

표 6 HAI에서 진행 중인 연구주제

세부 주제	내 용
인간 영향 (Human Impact)	• 공정하고 신뢰할 수 있는 AI 개발을 위해 AI의 수행방식을 이해하고, AI가 인간, 사회구조 및 기관, 국제질서와 상호작용하는 체계 수립 연구 - AI가 보편화됨에 따른 오해와 사회적 직면 문제의 해결
인간 역량 강화 (Augment Human Capabilities)	• AI 에이전트 및 응용프로그램이 사람들과 보다 효과적으로 의사소통하며 협업·보강할 수 있도록 하는 연구 - 인간의 의사결정과 다양한 작업에 대한 품질 향상을 도와 줄 수 있는 진보된 AI
지능 (Intelligence)	• 궁극적으로 인간의 언어와 감정, 의도, 행동, 다양한 규모의 상호작용을 이해하는 AI의 개발과 접근법 - 차세대 인간 중심의 AI를 개발하기 위한 신경·인지 과학적 접근법

※ 자료 : Stanford HAI(hai.stanford.edu)

스탠포드 HAI는 앞서 살펴본 주제에 대한 연구 지원사업도 추진하고 있다. 인간 중심의 AI연구 주제 중 혁신적인 연구들을 선정하여 각각 최대 7만 5천 달러를 지원하며, 올해는 29개의 과제 팀이 선정되었다.

3. MIT와 미디어 랩

MIT(Massachusetts Institute of Technology)는 올해 1월 MIT AI Policy Congress[16]를 개최하여 AI에 대한 안전, 정의, 복지 및 기회에 대한 논의를 수행했다. 이 논의에서 AI에 대한 기술역량, 국제적인 이슈, 운송과 제조, 형사사법, 제조업, 헬스케어, 거버넌스 등 각 분야별로 AI의 영향력을 정의하고, 사회가 의존하는 AI 시스템에 대한 신뢰문제와 인식 차이에 대한 해소방안이 논의되었다. 이는 현재 MIT에서 수행하고 있는 AI 관련 연구의 주제와도 그 맥락이 같다.

AI 기술연구에 선두주자였던 MIT 미디어랩(Media Lab)[17]은 최근 AI 기술과 시스템이 가져올 사회적인 영향과 윤리적인 포용 문제에 대한 연구를 수행하며 정책 프레임워크를 개발 중이다. 현재 26개의 연구 그룹을 통해 기술과 기술로 인한 응용, 영향력 등의 연구를 수행하고 있는데, 미디어랩에서 구성한 AI 연구그룹 및 진행 중인 연구주제는 [표 7]과 같다.

16 MIT의 인터넷 정책연구 이니셔티브(Internet Policy Research Initiative)와 MIT의 AI 관련 또 다른 이니셔티브인 MIT Quest for Intelligence에서 주최
17 MIT대의 연구실로 과학기술뿐만 아니라 사회, 예술 분야에 이르기까지 다양한 연구를 수행하는 것으로 유명함

표 7 MIT 미디어 랩의 AI 연구그룹 및 연구주제

구분	내 용
AI의 윤리와 거버넌스 연구그룹	• AI 윤리 및 거버넌스에 대한 증거자료 기반의 연구 수행 및 제안 - 제도적 지식기반 구축 - 인적자원 육성 - 산업 및 정책 입안자와의 인터페이스 강화
AI 시대의 알고리즘적 의사결정과 거버넌스	• 사회의 의사결정 프로세스에 AI를 활용하고, 이에 필요한 윤리적인 지침과 모범사례 연구 • AI의 의사결정 맥락에 차별적 요소를 분류하고 응용하는 연구 - 위험지수 예측에 따른 건물 검사와 형사적 선고를 내리는 것은 동일한 알고리즘적 의사결정이 아님 • 전 세계 AI정책 입안자와 실무자가 AI 전략과 정책, 윤리와 거버넌스에 관한 올바른 결정을 내릴 수 있도록 참조모델을 제공하는 연구
순환적 사회	• 인간-기계 간 사회시스템을 위해 필요한 새로운 거버넌스 아키텍쳐 설계 - 정치철학과 문화인류학을 반영한 거버넌스 연구
AI와 포용	• 사회의 다양성과 포용성을 지원하기 위한 AI 시스템 설계 및 배포 방법 연구 - 연령, 민족성, 인종, 성별, 종교, 출신국가, 위치, 기술과 교육수준, 사회경제적 상태의 기준에서 소외계층에 대한 AI 연구
AI와 글로벌 거버넌스	• AI 관련 기술의 국가 간 영향을 고려하여 세계적으로 운영될 수 있는 적절하고 실행 가능한 거버넌스 매커니즘 연구 - 미국, 유럽, 아시아에 걸친 사례연구 및 실무회의를 통해 모델 수립
법률에서 인간다운 AI	• AI의 발전에 따라 AI의 의사결정에 대해 감시하고 개선하기 위한 적법절차 프레임워크를 확립하는 연구 및 이를 위한 기술·법적 기반의 구축 - 판사의 형사적 판결을 돕는 도구로써 AI 시스템의 개선
로봇과 AI의 사회적 영향 조사	• 로봇과 AI가 일상생활에 미치는 영향을 탐색하는 도구 개발 - 지능형 에이전트와 개인 간의 상호작용을 기록하고 디자인에 반영하는 연구
소셜 로봇 연구 (사회적 지능을 갖춘 AI)	• 상호 대화에서 상대방의 반응을 살피고 주의를 끌 수 있는 대화기법을 가진 AI 연구 - 실시간 인식 및 행동 생성이 가능한 소셜 로봇을 제어하는 AI 알고리즘 개발
의약 및 치료 임상실험을 위한 AI 개발 연구	• 윤리적이고 안전하며 설명 가능한 AI 기반의 약품 및 치료 임상실험 개발 - 광범위한 환자 기반의 임상실험을 디지털로 대체하며 실험 비용 절감과 조기 시장 진출로 치료의 신속성을 보장하는 연구

※ 자료 : MIT Media Lab(www.media.mit.edu)

4. 윤리와 거버넌스 AI 이니셔티브

윤리 및 거버넌스 AI 이니셔티브(Ethics and Governance of AI Initiative)는 2017년 MIT 미디어 랩(Media Lab)과 하버드 버크만 클라인(Berkman-Klein) 센터[18]의 공동 프로젝트로 시작되었다. 이 이니셔티브는 관련된 다른 기관으로부터 운영을 위한 자선기금을 받아 자동화와 기계학습 기술에 있어서 공정성과 인간의 자율성 및 정의 측면의 사회적 가치를 입증하기 위한 연구를 수행한다. 수행중인 연구주제는 [표 8]과 같다.

18 미국 ICT-정책연구를 수행하는 연구기관 중 하나로 하버드 로스쿨에서 1998년에 설립

표 8 윤리와 거버넌스 AI 이니셔티브 연구주제

세부 주제	내 용
인공지능과 정의 (AI and Justice)	• 공공행정에서 자율성(Autonomy - AI에 의한 판단)을 채택하고 유지하는 것에 있어서 고려할 법적, 제도적인 사항은 무엇인가? • 형사사건의 판단에 있어서 인과관계 모델 같은 자율성의 역할을 재정립 하는 것(AI가 형사적 판단을 할 경우 판단 범위와 방법)
정보 품질 (Information Quality)	• 머신러닝과 자율시스템이 공공영역에 미치는 영향력 • AI 플랫폼과 대국민 간 효과적인 거버넌스와 공동개발 구조에 대한 연구 • AI의 경험(학습)기반 시스템이 잘못된 정보를 학습할 경우에 대한 정책적 대응 방안
자율화와 상호작용 (Autonomy and Interaction)	• 자율 시스템과 대중 간의 상호작용에 있어서 도덕적이고 윤리적인 측면에 대한 연구 • 인간의 직관이 어떻게 시스템에 잘 통합될 수 있는지에 대한 연구 • 시스템의 디자인과 사람-컴퓨터 간 인터페이스(HCI)의 효율적인 제어 및 명령 해석에 대한 연구

※ 자료 : The Ethics and Governance of Artificial Intelligence Initiative(aiethicsinitiative.org)(재편집)

시사점

AI의 궁극적 목적은 인류의 편익 증대이다. 나아가 더 정확하고 신속하며 때로는 공정한 자동화를 추구한다. 그러나 본래 도구적 목적으로 개발을 시작했던 AI는 이제 그 목적을 넘어 인류의 사회적 규범과 질서의 패러다임도 바꿀 수 있는 기술이 되어 윤리·철학적인 의미까지 고려할 상황이 되었다.

캐나다고등연구소(CIFAR)[19]는, AI 정책을 8가지로 구분하였는데 그중에서 모든 국가들이 공통적으로 추진하고 있는 중점 과제는 원천기술 확보, 산업융합, 인재육성 등인 것으로 나타났다.[20] 우리 정부에서 지난 2018년 5월에 발표한 AI 전략[21]도 기술연구를 통한 원천기술 확보 및 산업융합 확산, 인재양성, 연구기반 조성 등을 중점과제로 하고 있어 위 조사결과에 부합된다.

그러나, 미국을 포함한 AI 선도국들은 앞서 언급한 세 가지 분야 외에도, AI의 도입 효과와 부작용을 방지하기 위한 정책연구 또한 높은 비율로 병행하고 있다([그림 3] 참조).

19 Canadian Institute for Advanced Research(www.cifar.ca), 캐나다 정부 및 지자체에서 지원하는 다양한 과학기술 분야 연구 기관으로 1982년 설립
20 8가지 AI 정책 : 과학연구(AI원천기술연구), 인재육성, 미래직업, 산업융합, 윤리, 인프라, 정부, 사회포용(참조 : CIFAR, Building an AI World-Report on National and Regional AI Strategies, 2018.)
21 4차산업혁명위원회(과기정통부)(2018.5.), 인공지능(AI) R&D 전략

그림 3 세계 인공지능 전략 Hit 맵

※ 자료 : CIFAR, Building an AI World-Report on National and Regional AI Strategies, 2018

모든 나라가 공통적으로 중점과제로 채택하고 있는 과제들만을 수행하는 것은 결국 AI 패권 경쟁에서 우위를 점하지 못하는 결과를 불러올 수 있다. 이제는 우리도 종래의 빠른 추격자(Fast Follower) 방식을 넘어서야 한다. 끈기 있게 추진해야 하는 중장기 원천기술 투자·개발뿐만 아니라, AI로 인해 발생하는 수많은 결과들에 대한 시나리오를 미리 발굴하고 대응책을 쌓아나가는 노력 또한 병행해야 한다. 이를 통해 다가오는 AI 시대를 맞이하는 데에 발생할 수 있는 부작용을 최소화할 수 있으며, 우리나라가 세계 AI 경쟁에서 선도적인 역할을 수행할 수 있을 것이다.