한림연구보고서 155-1

미래 30년을 대비하는 과학기술전략

The Science and Technology Strategy for the Next 30 Years of Korea

인공지능의 활용과 공존

KAST 한국과학기술한림원
The Korean Academy of Science and Technology

한림연구보고서 155-1

미래 30년을 대비하는 과학기술전략

The Science and Technology Strategy for the Next 30 Years of Korea

인공지능의 활용과 공존

KAST 한국과학기술한림원
The Korean Academy of Science and Technology

미래 30년을 대비하는

과학기술전략

한림연구보고서 155-1

THE KOREAN ACADEMY OF SCIENCE AND TECHNOLOGY

인공지능의 활용과 공존

참여 집필진

조형희 연세대학교 기계공학부 교수
이성환 고려대학교 인공지능학과 특훈교수
이경무 서울대학교 전기정보공학부 석좌교수
하정우 네이버클라우드 AI이노베이션센터장
최재식 KAIST 김재철AI대학원 교수
서재민 중앙대학교 물리학과 교수
백민경 서울대학교 생명과학부 교수
조성배 연세대학교 컴퓨터과학과 교수
김명주 국제인공지능윤리협회 회장

KAST

인공지능의 활용과 공존

006 **1 서론**

012 **2 인공지능 기술 및 분야 발전 전망**
012 가. 들어가는 말
013 나. 생성형 인공지능의 발전
016 다. 인공지능의 활용
018 라. 인공지능에 대한 국가 정책 현황
022 마. 맺음말

024 **3 AI 기술 발전 전망 및 국내 산업의 새로운 도약 전략**
024 가. 들어가는 말
025 나. AI 기술의 현황 및 발전 방향
032 다. 우리의 AI 역량 및 현주소
040 라. 국내 산업 발전을 위한 AI 전략
041 마. 맺음말

043 **4 생성 AI 시대 산업과 사회 대전환**
043 가. 들어가는 말
043 나. 2024년 생성 AI 기술 동향 및 전망
051 다. 생성 AI 도입을 통한 생산성 혁신
054 라. 새로운 AI 에이전트 시대
056 마. 맺음말

059 **5 인공지능의 발전과 신뢰성 기술 전망**
059 가. 들어가는 말
060 나. 미국의 인공지능 투자 전략
061 다. 중국의 인공지능 투자 전략
062 라. 미국과 중국의 인공지능 경쟁과 우리의 전략
064 마. 생성형 인공지능의 발전과 경제발전
064 바. 인공지능 관련법과 규제
065 사. 인공지능 신뢰성의 구성요소
066 아. 딥러닝 내부의 의사결정을 설명하는 기술
066 자. 맺음말

069 6 인공지능 시대, 물리학자의 역할
069 가. 들어가는 말
070 나. 물리학과 AI : 원자핵부터 우주까지
072 다. 물리를 이해하는 AI
073 라. 물리학자의 역할
074 마. 맺음말

076 7 인공지능이 불러온 생명과학 연구 혁신
076 가. 들어가는 말
077 나. 인공지능의 생명과학 연구 활용 현황
079 다. 인공지능 기반 생명과학 연구의 전망
082 라. 맺음말

084 8 공정한 인공지능을 위한 기술과 발전 전망
084 가. 들어가는 말
085 나. 공정성 척도와 데이터셋
087 다. 인공지능의 편향제거 기술
089 라. 주요국의 동향
091 마. 발전 전망
092 바. 맺음말

094 9 지속 가능한 AI와의 공존을 위한 윤리와 규제
094 가. 들어가는 말
094 나. AI의 양면성
097 다. AI 윤리 원칙
105 라. AI 규제 동향
109 마. 맺음말

111 10 소결

KAST

1 서론

최근 몇 년간 인공지능(AI) 기술은 상상을 초월할 만큼 비약적인 발전을 이루어왔다. 특히 2022년 11월 오픈AI에서 출시한 챗GPT는 초거대 언어모델 기술이 실생활에 본격적으로 도입되는 계기가 되었으며, 이는 사회 각 분야에 큰 변화를 가져왔다. 인공지능은 더 이상 미래의 기술이 아닌, 현재의 기술로 자리매김하며 우리의 일상생활뿐만 아니라 산업, 의료, 금융, 교육 등 다양한 분야에서 혁신을 이끌고 있다. 이러한 기술의 발전은 인간의 삶을 더 편리하고 효율적으로 만들어주는 동시에, 새로운 문제와 도전을 불러일으키기도 한다.

인공지능 기술의 급격한 발전 속도와 그 영향력은 30년 후의 기술 발전 및 활용 방안을 예측하는 것을 어렵게 만든다. 그러나 미래를 준비하고 적절한 정책을 마련하기 위해서는 현재의 기술 동향과 발전 방향을 면밀히 분석하고, 가능한 시나리오를 바탕으로 예측하는 노력이 필요하다.

따라서 이러한 기술들이 현재 산업과 일상생활에서 어떻게 활용되고 있는지 분석하고, 외국에서의 개발 현황, 다양한 학문 분야에서의 접목 사례 등을 통해 앞으로의 발전 방향과 그 가능성을 예측해 보고자 한다. 또한 인공지능의 공정성과 윤리 및 규제 문제에 대해서 다루고자 한다. 이를 통해 우리는 인공지능의 미래를 보다 명확히 이해하고, 그에 맞는 준비를 할 수 있을 것이다.

40년 전만 해도 인공지능은 마치 공상과학 소설 속에나 나오는 환상에 가까운 기술로 여겨졌다. 운전자 없는 자율주행 자동차, 인공지능 비서, 지능형 로봇들은 영화나 만화 속 상상일 뿐, 가까운 미래에 현실이 될 것이라고 믿는 사람은 거의 없었다. 그러나 2012년 AlexNet의 혁신적인 발표 이후 딥러닝 기반 AI 기술은 믿기 어려울 정도로 빠르게 발전했다. 초기에는 AI 학자들조차 딥러닝의 성능과 지속 가능성에 대해 반신반의하는 경향이 있었지만, AlphaGo, ResNet, GAN, 트랜스포머, 확산 모델 등 획기적인 기술의 등장으로 다양한 분야에서 기존의 난제를 해결할 수 있음을 보여주며 딥러닝 기술은 학계와 산업계의 표준으로 자리 잡았다.

딥러닝 기반 AI 기술의 발전에는 새로운 아키텍처와 학습 알고리즘의 개발도 큰 기여를 했지만, 최근의 경향은 'scaling law', 즉 규모의 법칙이 주도하고 있다. 대규모 GPU와 방대한 학습데이터로 훈련된 챗GPT와 같은 초거대 언어모델은 천문학적인 자원이 투입되었고, 예상을 뛰어넘는 결과를 보여주었다. 이러한 초거대 인공지능 모델 개발은 당분간 가속화될 것으로 예상되며, 아직 한계에 도달했다고 보기에는 이르다. 단순한 언어모델을 넘어 이미지, 비디오, 음성, 텍스트 등을 복합적으로 분석하고 이해하는 멀티모달 AI와 로봇, 가전, 기계 등 물리적인 기기에 결합하는 임바디드 AI로의 확장이 본격화되고 있으며, 많은 사람들은 이러한 발전이 범용인공지능(AGI)의 도래를 예고하고 있다고 믿고 있다.

인공지능은 21세기 기술 혁신의 중심에 있으며, 이 중 미국과 중국이 AI 개발과 투자에서 세계를 선도하고 있다. 이 두 나라는 각각 고유한 전략과 강점을 바탕으로 AI 경쟁을 주도하고 있으며, 그 결과는 향후 글로벌 기술 발전의 방향을 결정짓는 중요한 요소가 될 것이다.

미국의 AI 분야 강점은 세계 시장을 지배하는 실리콘 밸리의 글로벌 기업들과 최고의 연구 능력에서 기인한다. 마이크로소프트, 구글, 애플 등의 기업은 컴퓨터와 휴대폰에 필수적인 소프트웨어 인프라와 AI 서비스를 제공하고 있으며, 아마존, 넷플릭스, 메타(페이스북), 트위터 등은 클라우드, 인터넷 미디어, 소셜미디어 및 마이크로블로깅 서비스 분야에서 시장을 선도하고 있다. 이러한 글로벌 플랫폼 기업들은 사용자 데이터를 기반으로 AI 기반 인식 기술과 추천 시스템을 개발하고, 지속적인 혁신을 이루고 있다. 미국 기업들은 또한 구글의 딥마인드 인수와 마이크로소프트의 오픈AI 투자와 같은 사례에서 볼 수 있듯이, 세계적 수준의 연구진을 유치하고 기술 발전을 주도하기 위해 기술 기업에 대한 투자를 우선시하고 있다.

미국의 AI 발전 전략은 국방연구진흥기관 방위고등연구계획국(Defense Advanced Research Projects Agency, DARPA)을 통해 이루어져 왔다. DARPA는 1968년부터 카네기멜론대학교와 협력하여 군용 자율주행 자동차 개발 프로젝트를 진행했으며, 2005년 DARPA 그랜드 챌린지를 통해 자율주행 기술의 가능성을 확인했다. 이 기술은 이후 구글, 웨이모, 테슬라와 같은 민간 기업에 이전되어 상용화되었다. 또한 챗GPT의 근간이 되는 기술도 DARPA와 미국 표준기술연구소(National Institute of Standards and Technology, NIST)가 함께 추진한 기계 자동 독해 프로그램(machine reading program)의 결과물이다. 이처럼 미국은 장기적인 연구 기획과 투자, 민간 기업과의 협력을 통해 AI 기술을 발전시키고 있다.

반면 중국은 내수 시장의 보호와 내수 기업 육성, 국외 전문 인력의 유치 전략을 통해 AI 분야에서 경쟁력을 키워왔다. 중국은 주요 정보 검열과 보안을 이유로 다국적 기업의 내부 서비스 제공을 제한하고, 중국 내 기업을 육성하여 대규모 데이터를 축적하고 AI 기술을 발전시키고 있다. 또한, 천인계획과 만인계획 같은 프로그램을 통해 해외에 있는 중국계 인재를 유치하여 신기술 개발을 촉진하고 있다. 특히 컴퓨터 비전 기반의 인식 기술에서 우위를 점하고 있으며, 최근에는 알리바바의 Qwen2와 Kuaishou의 Kling 같은 생성 AI 모델을 통해 경쟁력을 높이고 있다. 중국 정부는 생성 AI를 정부 체제 확산의 도구로 활용하며, 시진핑 주석의 어록을 대답할 수 있는 시진핑 GPT를 준비 중이라는 보도도 있었다. 그러나 중국의 AI 발전 전략은 여전히 기술 추격형에 머물러 있으며, 장기적인 기술 개발에 대한 투자와 혁신이 필요하다.

EU는 강력한 AI 연구 역량을 보유하고 있지만, 미국이나 중국에 비해 생성 AI 분야에서 상대적으로 뒤처지고 있다. 그러나 프랑스의 Mistral과 독일의 Aleph Alpha 같은 기업들은 거대 언어모델을 오픈소스 형태로 공개하며 경쟁력을 인정받고 있다. 이탈리아 역시 자국의 생성 AI 경쟁력 확보를 위해 정부 차원의 투자를 천명하고 있다. EU는 규제와 함께 오픈소스 중심의 생성 AI 산업 활성화 부분을 반영한 법안을 마련하며, EU 내 기업의 활동을 지원하고 있다.

일본은 오픈AI와의 협업을 통해 공공 분야에 생성 AI를 도입하고 있으며, 자국 기업인 소프트뱅크를 통해 자체 AI 기술 확보를 추진하고 있다. 인도네시아, 태국, 베트남 등 아시아 국가들도 생성 AI 기술의 자국화를 위해 노력하고 있다. 한국은 네이버, LG AI 연구원, 카카오, KT, SKT 등 여러 기업이 자체적으로 거대 언어모델을 개발하며 빠르게 산업 생태계를 구축하고 있다.

이러한 배경에서 AI 기술의 발전은 전 세계적으로 중요한 영향을 미치고 있으며, 각국은 이를 통해 경제적, 기술적 우위를 확보하려는 전략을 펼치고 있다. 우리나라도 글로벌 경쟁에서 뒤처지지 않기 위해 AI 기술 연구와 인재 양성, 그리고 관련 산업 육성에 대한 전략적인 접근이 필요하다. 이를 통해 AI 기술의 상용화와 혁신을 촉진하고, 글로벌 시장에서 경쟁력을 확보할 수 있을 것이다.

다양한 학문 분야에서도 인공지능을 활용하기 위한 노력을 하고 있다. 물리학 분야에는 아직 인간이 완전히 이해하지 못하는 방대한 양의 실험 및 관측 데이터들이 존재한다. 예를 들어, 입자가속기에서는 우주의 기본 법칙을 탐구하기 위해 입자 충돌 데이터를 수집하고 있으며, 핵융합 장치에서는 플라즈마의 거동을 이해하기 위한 많은 데이터를 축적하고 있다. 최근 인공지능 기술이 발전하면서 이러한 물리학 데이터를 활용한 연구가 활발히 이루어지고 있다.

핵융합 기술은 태양의 에너지원인 수소 핵융합 반응을 지구상에서 구현하려는 시도로, 고온의 플라즈마를 안정적으로 유지하는 것이 핵심 과제다. 한국핵융합에너지연구원의 KSTAR(Korea Superconducting Tokamak Advanced Research)는 1억 도의 초고온 플라즈마를 안정적으로 유지하는 데 성공했지만, 이는 수많은 시행착오를 거친 인간의 노력 덕분이었다. 인공지능을 활용하면 이러한 데이터 주도 학습을 더욱 효과적으로 수행할 수 있을 것으로 기대된다. 또한, 우주 연구에서도 인공지능의 역할은 커지고 있다. 태양풍의 활동 변화나 암흑물질의 분포를 이해하기 위해 AI 기술을 활용한 연구가 진행되고 있다. AI는 태양 관측 사진에서 코로나 홀을 자동으로 탐지하거나, 관측 가능한 물질의 분포와 속도 정보를 기반으로 암흑물질의 질량 분포를 재구성하는 데 중요한 역할을 하고 있다. 앞으로의 AI는 단순히 데이터를 분석하는 도구를 넘어 물리학을 이해하고 새로운 물리 현상을 발견하는 데까지 발전할 것이다. 그리고 물리정보 신경망과 양자 기계학습(ML) 같은 기술은 물리 법칙을 학습하여 더 정교한 분석을 가능하게 하며, 물리학 연구에 큰 변화를 가져올 것이다. 이를 위해 물리학자들은 AI와의 융합 연구에 적극 참여하고, 학문 간 경계를 허물어 새로운 시너지를 창출해야 한다. 이러한 노력은 물리학의 발전뿐만 아니라 인류의 지식 확장에 크게 기여할 것이다.

또한 인공지능의 발전은 생명과학 연구에 혁신적인 변화를 가져오고 있다. 전통적으로 생명과학 연구는 주로 실험에 의존했지만, 인간 유전체 프로젝트 이후 축적된 방대한 데이터를 효과적으로 분석하기 위해 계산 기법들이 개발되기 시작했다. 최근 AI 기술이 고도로 발달하면서 생명과학 빅데이터 분석에 AI를 접목하려는 시도가 활발히 이루어지고 있다. 이는 단백질 구조 예측, 유전체학, 분자생물학 등 다양한 분야에서 혁신적인 결과를 낳고 있다. 대표적인 예로, 구글 딥마인드의 알파폴드와 미국 워싱턴대학교의 로제타폴드는 고정확도의 단백질 구조를 예측하는 AI를 개발하여 생명과학 연구에 큰 혁신을 가져왔다. 이는 AI의 활용 가능성을 증명하는 사례로, 이후 AI를 활용한 다양한 생명과학 연구를 촉진하는 계기가 되었다. 단백질 구조 예측 외에도 AI는 유전체 데이터 분석, 분자 모델링, 시뮬레이션 등 다양한 분야에서 활용되고 있다.

AI의 도입은 약물 개발, 신약 발견, 질병 진단 및 맞춤형 치료 등 생명과학 연구의 다양한 측면에서 연구의 효율성과 정확성을 높이는 데도 기여하고 있다. 예를 들어, 엔비디아와 아스트라제네카가 공동 개발한 MegaMolBART는 대규모 화합물 라이브러리를 기반으로 신약 후보 물질을 더 효율적으로 도출하는 기술을 제공하고 있으며, 이는 신약 개발의 초기 단계에서 시간과 비용을 크게 절감할 수 있도록 돕고 있

다. 더 나아가 AI는 생명과학 연구에서 새로운 분야를 개척하고, 기존 연구 방법론을 혁신하고 있다. 합성생물학, 유전자 편집, 바이오프린팅 등 첨단 생명과학 기술과 결합하여 연구의 혁신을 이끌고 있으며, 환경 보호와 생태계 연구에서도 중요한 역할을 하고 있다. 예를 들어, Conservation Metrics는 AI를 이용해 멸종위기종의 소리를 분석하고 서식지를 모니터링하며, 이를 통해 생물 다양성 보전과 생태계 보호에 기여하고 있다.

종합하자면, AI 기술을 효과적으로 활용하여 생명과학 연구의 새로운 가능성을 탐구하고, 인류의 건강과 지구 환경 보호에 기여할 수 있는 방향으로 나아가야 한다. 이를 위해 지속적인 기술 발전과 함께 윤리적 고려와 사회적 책임을 병행하여 AI 기술이 긍정적인 영향을 미칠 수 있도록 해야 할 것이다.

AI의 발전은 다양한 분야에서 혁신을 이끌지만, 공정성과 윤리성 문제도 함께 제기되고 있다. AI 시스템의 예측과 의사결정에서 편향과 불공정이 발생할 수 있어 이를 해결하기 위한 다양한 공정성 척도와 데이터셋이 개발되고 있다. 예를 들어, 인구학적 동등성, 기회 균등, 예측 동등성 등의 척도가 AI 모델의 공정성을 평가하는 기준으로 사용된다. COMPAS, Adult Income, German Credit 같은 데이터셋은 이러한 공정성 척도를 적용하여 AI 모델의 편향을 평가하고 개선하는 데 활용된다. 편향을 제거하고 공정성을 보장하기 위해 전처리, 중처리, 후처리 접근 방식이 개발되었다. 전처리 접근 방식은 데이터 샘플링, 증강, 정규화 등을 통해 데이터 자체의 편향을 줄이는 방법이다. 중처리 접근 방식은 모형 학습 과정에서 알고리즘에 공정성을 내재화하는 것이고, 후처리 접근 방식은 학습된 모형의 예측 결과를 수정하여 편향을 줄이는 방법이다. 특히, 대형언어모델(LLM)의 편향을 줄이기 위한 다양한 기술이 연구되고 있다.

미국, EU, 아시아 등 주요국은 AI 공정성과 윤리를 보장하기 위해 정책과 규제를 마련하고 있다. 미국은 NIST를 통해 공정성 평가 기준을 마련하고 다양한 도구를 개발 중이며, EU는 AI 규제 법안과 GDPR을 통해 비차별적이고 투명한 AI 작동을 요구하고 있다. 아시아의 중국, 일본, 한국 등도 AI 공정성을 위한 정책과 연구를 추진하고 있다. 앞으로 AI 공정성을 확보하기 위한 기술은 더욱 발전할 것이며, 연구자들은 정교한 편향제거 알고리즘을 개발하고 예측 정확도를 유지하면서도 공정성을 높이는 방향으로 나아가야 한다. 글로벌 기업과 정부는 AI 공정성을 보장하기 위한 규제와 정책을 강화하고, 사회적 인식과 참여를 통해 AI 공정성을 높이기 위한 공동의 노력을 기울여야 한다.

기술 혁신은 인류 사회에 막대한 영향을 미치며, 특히 AI의 등장은 그 영향력을 극대화하고 있다. AI는 일상생활과 업무의 효율성을 높이고, 경제 성장을 촉진하며, 교육, 건강, 고용 기회 확대 등 다양한 분야에서 긍정적인 변화를 일으키고 있다. 예를 들어, 생성형 AI '챗GPT'를 활용한 연구에서는 문서 작업의 효율성이 40% 향상되고, 소프트웨어 개발 시간이 40% 이상 단축된다는 결과가 보고되었다. 이러한 기술 혁신은 경제 성장과 소득 수준 향상, 건강 증진과 수명 연장, 교육과 학습 기회 확대, 고용 창출 등 다양한 혜택을 제공하며, 개인과 사회 전반에 걸쳐 긍정적인 영향을 미친다.

그러나 기술 혁신에는 항상 양면성이 존재한다. 기술의 순기능 곁에는 역기능과 부작용이 병존하며, 이는 AI 역시 예외가 아니다. 기술 혁신은 긍정적인 영향을 미치지만, 동시에 사회적, 경제적, 윤리적 문제를 야기할 수 있다. AI의 발전과 함께 발생하는 부작용과 문제점들은 다양한 사회적, 윤리적 논의를 필요로 한다. 예를 들어, 원자폭탄 개발을 이끌었던 오펜하이머 연구소장과 딥러닝의 대부 제프리 힌튼 교수는 기술의 발전과 함께 윤리적 문제를 제기하며 그 위험성을 경고한 바 있다. 이러한 사례들은 기술 혁신이 항상 긍정적이지는 않으며, 그 이면에는 해결해야 할 중요한 문제들이 존재함을 시사한다.

AI의 윤리적 문제는 특히 중요한 논점으로 부각되고 있다. 기술 혁신의 긍정적인 면을 극대화하고 부작용을 최소화하기 위해서는 AI의 개발과 사용에 있어 윤리적 원칙을 준수하는 것이 필수적이다. 통제성과 안전성, 투명성과 설명 가능성, 공정성, 프라이버시 보호, 책임성과 책무성, 표절과 저작권 침해 방지, 진짜와 가짜 구분, 그리고 다양성 등 다양한 윤리적 원칙들이 AI의 개발과 사용 과정에서 고려되어야 한다. 이러한 윤리적 원칙들은 AI 기술이 사회적 신뢰를 얻고, 인류 전체에 혜택을 줄 수 있도록 하는 데 중요한 역할을 한다.

AI 기술의 급격한 출현으로 인해 현재는 인류 역사상 가장 중대한 산업적, 사회적 변혁이 일어날 것이라는 기대와 우려가 공존하는 시기이다. 이전의 산업혁명이 특정 분야에서의 변혁을 가져왔고 상대적으로 장기간에 걸쳐 진행되었던 것에 비해 AI에 의한 혁명은 모든 산업과 인류의 삶 전체의 패러다임을 근본적으로, 단기간에 바꿀 수 있는 특징을 가지고 있어 그 무게가 다르다. 이는 AI로 인해 기존의 국가, 산업, 노동, 경제, 사회, 개인 간의 역할과 질서가 새로운 차원으로 급격히 재편될 수 있음을 의미한다.

따라서 이러한 대변혁의 시기에 새로운 AI 기술의 능력과 파급력, 그리고 위험 요소를 정확히 파악하고 경쟁력 있는 핵심기술을 확보함으로써 우리 산업과 사회를 한 단계 도약시킬 수 있는 전략을 마련하고 실행하는 것은 미룰 수 없는 숙제이다. 본 주제에서는 AI 기술의 현황과 발전 방향, 문제점을 파악하여, 다가올 미래를 철저히 준비하기 위한 방안을 찾고자 한다.

KAST

2 인공지능 기술 및 분야 발전 전망

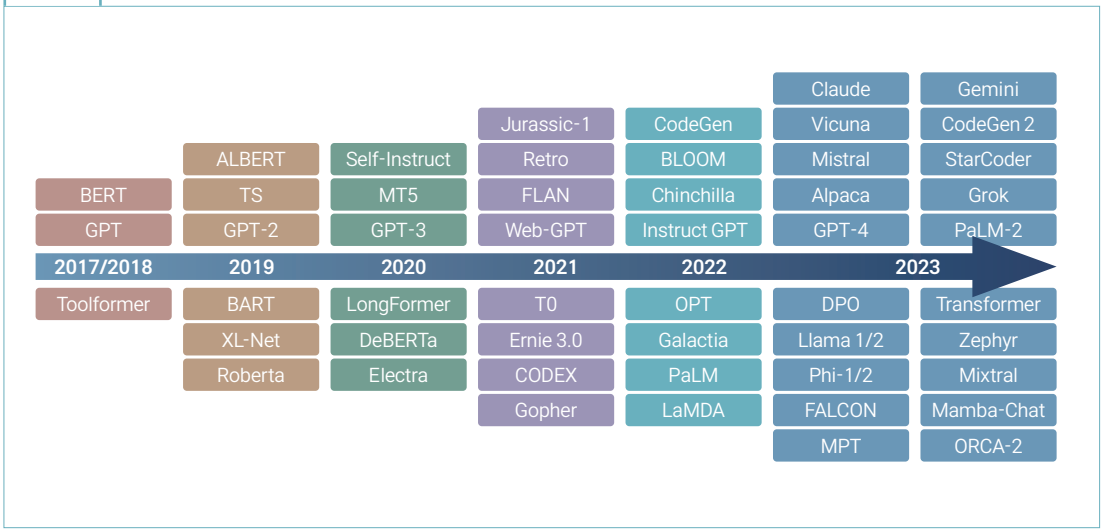
이 성 환
고려대학교 인공지능학과 특훈교수

가 들어가는 말

인공지능(AI) 기술은 지난 몇 년 동안 비약적인 발전을 이루어왔다. 2022년 11월에 출시된 오픈AI의 챗GPT를 시작으로, 초거대 언어모델 기술이 본격적으로 우리의 일상에 파고들면서 사회의 다양한 분야에 큰 영향을 미치고 있다. 인공지능은 더 이상 미래의 기술이 아닌, 현재의 기술로 자리 잡고 있으며, 일상생활, 산업, 의료, 금융, 교육 등 여러 분야에서 혁신을 이끌고 있다. 이러한 급격한 발전 속도와 변화를 감안할 때, 30년 후의 인공지능 기술 및 그 활용에 대해 예측하는 것은 매우 어려운 과제이다. 그럼에도 불구하고, 미래를 준비하고 적절한 정책을 마련하기 위해서는 현재의 기술 동향과 발전 방향을 면밀히 분석하고, 가능한 시나리오를 바탕으로 예측하는 노력이 필요하다.

본 장에서는 사회에 가장 큰 영향을 준 생성형 인공지능에 대해서 살펴보고, 더 나아가 생성형 인공지능 중 하나인 대형언어모델(LLM)을 바탕으로 연구되고 있는 범용인공지능(AGI)에 대해 알아볼 것이다. 또한 인공지능이 현재 산업 및 생활에서 활용되고 있는 예시를 살펴본다. 마지막으로, 이러한 기술들이 어떠한 방향으로 발전할 것인지 예상해 보며 본 장을 마치도록 하겠다.

그림. 1 트랜스포머 이후 출현한 대형언어모델



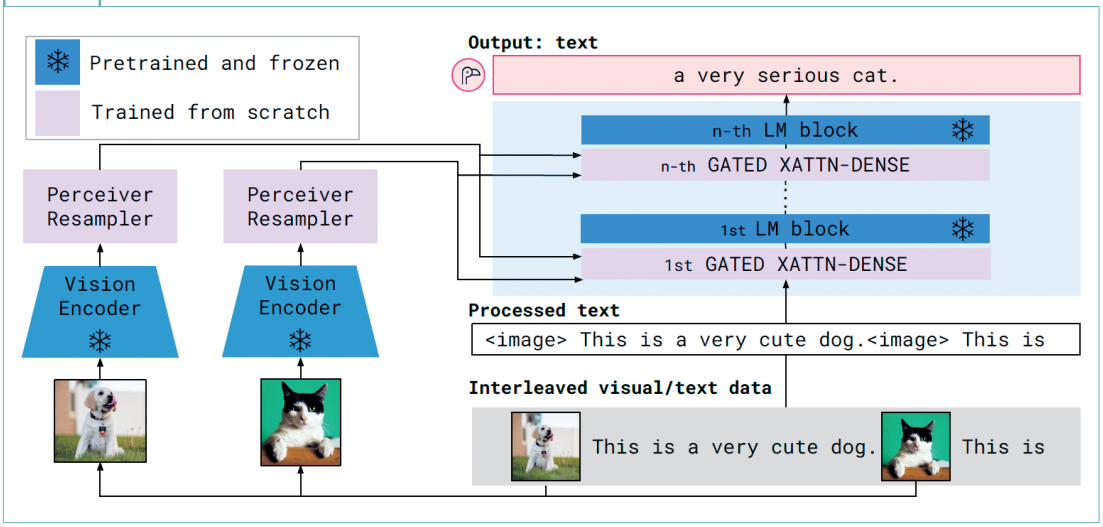
출처: Minaee et al.(2024).

나 생성형 인공지능의 발전

1) 대형언어모델

대형언어모델(Large Language Model, LLM)은 딥러닝과 자연어처리의 교차점에서 탄생한 혁신적인 기술로, 최근 몇 년 동안 그 중요성과 영향력이 급증하고 있다. 이러한 모델은 웹 규모의 막대한 양의 데이터로부터 학습된 지식을 활용한다. 이전 언어모델과 대형언어모델의 주요한 차이점은 학습데이터 및 모델 규모와 그에 따라 나타나는 특수한 능력에 있다. 특수한 능력의 발현을 위해 모델은 최소 수십억 개에서 수조 개에 이르는 파라미터를 포함하고 있으며, 위키피디아, 웹 페이지, 뉴스 기사 등과 같은 인터넷에서 수집된 방대한 텍스트 데이터 세트로부터 학습된 지식을 보유한다. 이러한 지식 학습을 기반으로 모델은 복잡한 문장 구조, 다양한 언어 스타일, 그리고 도메인 특화된 지식까지도 포함하는 광범위한 정보를 이해할 수 있다. 이에 따라 전통적인 자연어처리 모델과는 달리 문장 혹은 문단의 복잡한 구조를 이해하고 사전 지식이 필요한 질문을 명확히 파악한 후 정답을 추론한다. 이를 통해, 대형언어모델은 여러 도메인 안에서의 기계 번역, 텍스트 요약, 질의응답 시스템, 그리고 다양한 자연어처리 작업에서 높은 성능을 달성했다.

그림. 2 최초의 멀티모달 언어모델 Flamingo

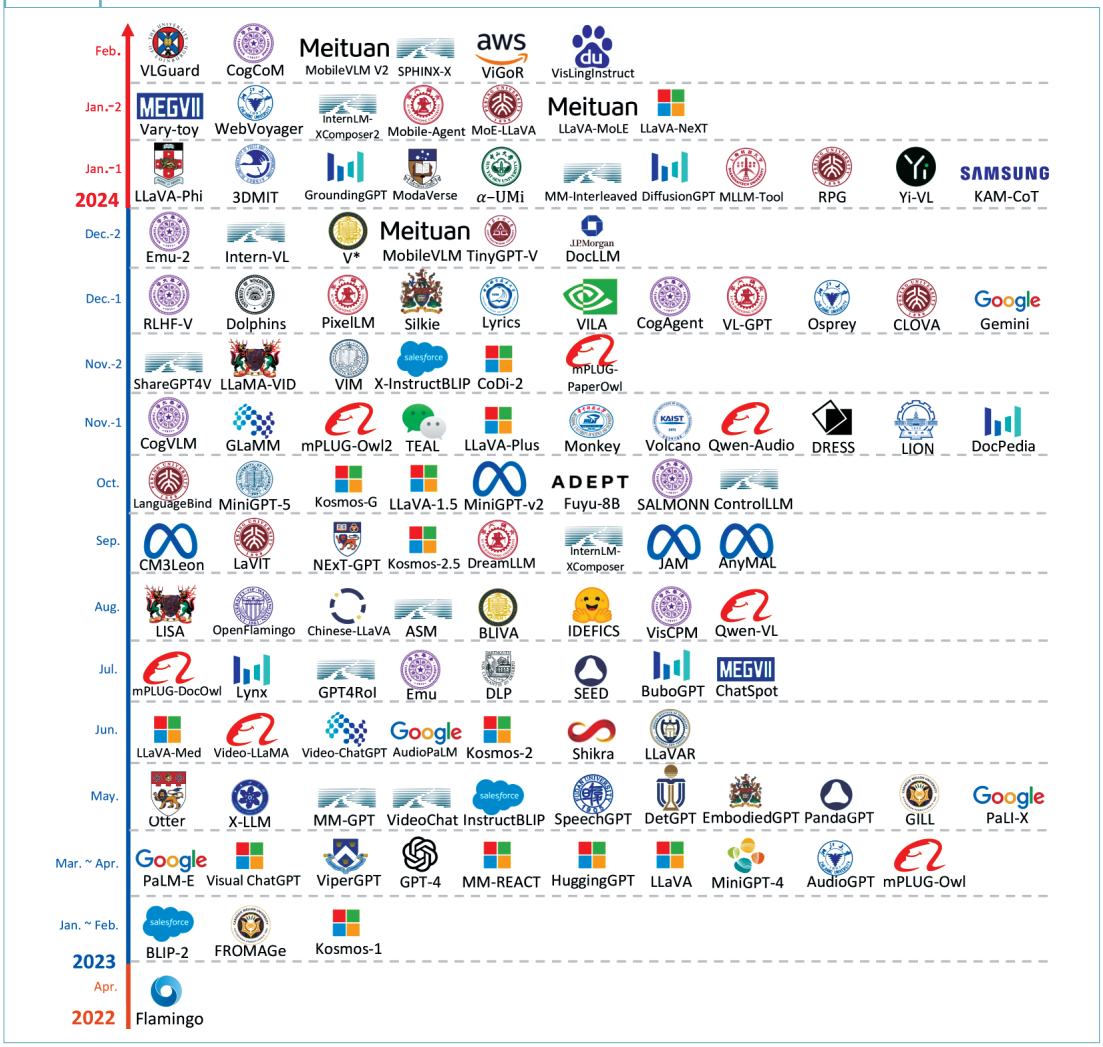


출처: Alayrac et al.(2022).

2) 멀티모달 언어모델

멀티모달 언어모델은 텍스트뿐만 아니라 이미지, 오디오, 비디오 등 다양한 형태의 데이터를 통합하여 이해하고 처리할 수 있는 혁신적인 인공지능 기술이다. 이러한 모델은 다양한 모달리티의 데이터를 결합함으로써 더 풍부하고 정교한 정보 처리를 가능하게 한다. 멀티모달 언어모델은 대형언어모델의 발전과 함께 등장했으며, 이는 대형언어모델의 풍부한 언어적 공간에 관련 정보가 이미 학습되어 있다는 가정하에 개발되었다. 최초의 멀티모달 언어모델인 Flamingo를 시작으로 수많은 멀티모달 언어모델이 개발되고 있으며, 이러한 모델들은 자연어처리 컴퓨터 비전, 음성인식 등 여러 인공지능 분야의 교차점을 대표한다.

그림. 3 다양한 멀티모달 언어모델



출처: Zhang et al.(2024).

멀티모달 언어모델의 핵심은 서로 다른 형태의 데이터를 공동으로 학습할 수 있는 능력에 있다. 이를 위해, 멀티모달 모델은 텍스트와 다른 모달리티를 각각 이해하는 별도의 트랜스포머 구조를 사용하거나, 동시에 처리할 수 있는 아키텍처를 설계한다. 예를 들어, 오픈AI의 CLIP 모델(Radford et al., 2021)은 이미지 공간과 텍스트 공간을 맵핑되도록 학습하여 텍스트 설명과 일치하는 이미지를 찾는다. 이러한 모델은 다양한 멀티모달 작업에서 미세조정을 통해 사용되거나, 대형언어모델이 쉽게 이미지를 이해할 수 있게 한다. 이와 같은 기술은 이미지 캡셔닝, 비디오 해석, 멀티모달 검색 등 다양한 응용 분야에서 높은 성능을 발휘하고 있다. 최근에는 'Sora'라는 자연어 프롬프트를 통해 원하는 모습의 동영상을 생성하는 모델이 등장했다.

3) 생성형 인공지능의 미래: 범용인공지능

정의 및 개념

범용인공지능(Artificial General Intelligence, AGI)은 특정 작업에 국한되지 않고 인간처럼 다양한 지적 작업을 수행할 수 있는 인공지능을 의미한다. 이는 특정 분야에서 뛰어난 성능을 발휘하는 좁은 의미의 인공지능과는 달리, 다양한 상황과 문제를 이해하고 해결할 수 있는 능력을 가진다. AGI는 복잡한 문제 해결, 추론, 학습, 창의성 등 인간 지능의 모든 측면을 모방하는 것을 목표로 한다.

주요 특징

특정 작업에 국한되지 않고 다양한 작업을 하기 위해 AGI는 다양한 도메인에서 새로운 정보를 학습하고 기존 지식과 결합하여 문제를 해결할 수 있는 학습 및 추론 능력을 보유한다. AGI는 그러한 능력을 통해 의학에서부터 환경과학, 복잡한 데이터 분석에 이르기까지 인류가 직면한 가장 어려운 문제들을 해결할 수 있는 새로운 방안을 제시하며, 인간과 협력하여 창의적인 아이디어를 생성해 다양한 분야에서 인간의 능력을 확장시키는 역할을 할 것이다.

현재의 연구 동향

현재 AGI를 실현하기 위해 여러 방면으로 연구가 진행되고 있다. 인간처럼 필요한 지식을 스스로 인지하고 학습할 수 있도록 뇌를 모사하여 자가 정렬화하는 기술 개발, 비정형 데이터로부터 필요한 부분의 데이터를 선별하고 정형 학습데이터로 변환하여 능동적으로 학습하는 기술 연구, 문제 해결에 필요한 지식을 검색하기 위해 자동으로 부족한 지식을 식별하여 다수의 쿼리를 생성하는 연구가 있다. 또한, 복합적 문제 해결을 위해 문제를 분해하고 이에 대한 수행 계획을 수립하여 다양한 문제를 해결할 수 있는 능력이 연구되고 있다. 그 외에도 텍스트, 이미지, 오디오 등 다양한 형태의 데이터를 통합하여 학습하는 멀티모달 학습, 다양한 상황에서 최적의 행동을 학습하기 위한 강화학습, 새로운 작업을 빠르게 학습하기 위한 메타학습 기술 등이 연구되고 있다.

표. 1 AGI의 레벨 분류

AGI 레벨(명칭)	능력 수준	약인공지능 사례	AGI 목표 및 사례
레벨 0 (No AI)	AI가 아님 (단순 연산 능력)	계산기, 컴파일러 등	• 인간이 컴퓨팅 프로세스에 직접 참여 • Amazon Mechanical Turk, Google Cloud Dataflow 등
레벨 1 (Emerging AGI)	숙련되지 않은 인간 수준	GOFAI(Good Old-Fashioned AI), 단순 룰베이스 시스템 등	• 낮은 수준의 패턴인식 및 기초 학습 능력 • 챗GPT(OpenAI), Bard(Google), Llama 2(Meta), Gemini(Google) 등
레벨 2 (Competent AGI)	숙련된 인간의 상위 50% 수준	시리(Apple), 알렉사(Amazon), 구글어시스턴트(Google), 왓슨(IBM) 등	• 복잡한 문제 해결 및 추상적 사고 능력 향상 • 없음
레벨 3 (Expert AGI)	숙련된 인간의 상위 10% 수준	그래머리(Grammarly), Dall-E 2(OpenAI), Imagen(Google) 등	• 인간 수준의 언어 이해 및 생성 능력 도달 • 없음
레벨 4 (Virtuoso AGI)	숙련된 인간의 상위 1% 수준	딥블루(IBM), 알파고(Google) 등	• 다양한 응용 분야에서 높은 수준의 성과 • 없음
레벨 5 (Artificial Super Intelligence)	숙련된 인간의 능력을 초월하는 수준	알파폴드(Google), 알파제로(Google) 등	• 자체 학습 및 발전이 가능한 완전한 AGI 도달 • 없음

출처: Morris et al.(2024).

해결해야 하는 과제

구글 딥마인드(Google DeepMind)는 AGI 레벨을 <표. 1>과 같이 여섯 단계로 분류하였으며, 현재 AGI 수준은 숙달되지 않은 인간보다 약간 높거나 비슷한 수준(레벨 1)으로, 완전한 AGI 수준에 도달하기 위해서는 복합적 문제 해결 능력 및 자체 학습 능력을 개발해야 한다. 현재 기술로는 AGI를 구현하는 데 필요한 컴퓨팅 자원과 알고리즘의 효율성에 한계가 있어 이를 극복하기 위한 기술적인 혁신이 필요하다. 또한, AGI 개발 및 사용에 있어서 윤리적 문제는 반드시 고려해야 할 사항 중 하나이다. 인간과 유사한 혹은 더 높은 지능을 가진 인공지능이 윤리적 결정이나 작업을 수행할 때 일으킬 수 있는 사회적 문제를 고려하여 기술의 오남용을 최대한 방지해야 한다.

다 인공지능의 활용

1) 산업에서의 인공지능

기하급수적으로 발전하는 인공지능은 다양한 산업에서 사람을 대신해 여러 가지 일을 수행하고 있다.

대표적으로 생성형 인공지능을 이용한 사례로는 이미지를 생성하여 광고한 삼성생명과 인공지능 모델이 햄버거 그림과 관련된 음악을 창작하는 롯데리아 광고 등이 있다. 이러한 기술들은 마케팅과 광고 산업에서 혁신적인 변화를 가져오고 있다. 이뿐만 아니라, 사람을 대신해서 음식점에서 서비스를 제공하는 서빙 로봇 '딜리플레이트', 길을 안내해주는 안내 로봇 등도 인공지능 기술의 활용 사례로 주목받고 있다. 이들 로봇은 고객 서비스의 질을 높이고 인건비를 절감하며, 비즈니스 운영의 효율성을 극대화하는 데 기여하고 있다. 이외에도 제조업 분야에서 스마트 팩토리를 통해 생산 공정을 자동화하고 있고, 물류 분야에서 최적 경로를 계산하여 배송 효율성을 높이고 있으며 의료 분야에서 진단 보조 시스템을 통해 환자의 치료를 돕는 등 다양한 분야에서 실제로 인공지능이 쓰이고 있다.

2) 생활에서의 인공지능

인공지능은 산업에서 인간의 일을 대신할 뿐만 아니라 생활 편의를 도와주는 다양한 서비스를 제공한다. 대표적인 예로 AI 개인비서는 사용자의 일정 관리, 전화 걸기, 통화 내용을 텍스트로 변환 및 요약 등 다양한 작업을 자동적으로 수행하여 생활 편의를 제공한다. 일상 업무뿐만 아니라 수면 분석 등 AI 헬스케어 기술이 도입되어 건강 유지에도 도움이 되고 있다. 스마트 워치를 포함한 웨어러블 디바이스는 사용자의 운동량, 심박수, 수면 패턴 등을 모니터링하여 건강 상태를 실시간으로 체크하고 필요한 조언을 제공한다. 인공지능은 스마트 홈 기술에도 활용되고 있다. 스마트 스피커는 음성인식을 통해 음성을 재생하고, 날씨 정보를 제공하며, 일정 관리를 도와주는 등 다양한 기능을 제공한다. 이러한 기술들은 사용자의 일상생활을 더욱 편리하고 효율적으로 만들어준다.

3) 미래의 인공지능 활용 방향

현재 전 세계적으로 에너지 효율을 극대화하고 탄소 배출을 줄이기 위한 노력의 일환으로 인공지능 기술이 활발히 적용되고 있다. 국내 기업 또한 이러한 흐름에 맞추어 다양한 인공지능 기반 에너지 혁신을 추진하고 있다. 대표적으로 GS칼텍스는 인공지능을 활용한 운영 최적화를 본격적으로 시작했다. 공장의 일산화탄소 발생을 최소화하는 최적점을 찾아 불필요한 공정을 줄여 에너지 효율을 높이고 있다. 실제로 에너지 절감뿐 아니라 관련 비용도 5분의 1 수준으로 줄어 친환경과 비용 절감 두 가지를 모두 잡을 수 있었다.

현재 농림축산식품부의 '2050 농식품 탄소중립 추진전략'에 따라 데이터, 네트워크, 인공지능 기반의 정밀농업이 도입되는 중이고, 인공지능 기반 스마트팜 기술을 활용하여 작물관리와 에너지 절감을 동시에 추구하며 2050년까지 농가의 60%에 정밀농업 기술을 보급하는 것을 목표로 하고 있다.

미래에는 이런 인공지능 기술을 활용한 에너지 자립형 스마트시티가 구축될 것으로 예상된다. 도시 전체의 에너지 흐름을 인공지능이 실시간으로 분석하고 최적화하여 에너지 사용 효율을 극대화할 것이며, 대표적인 사례로 인공지능 기반 전력망 관리, 자율주행 전기차 인프라, 재생에너지 통합 관리 시스템 등이 활용될 수 있다. 또한 탄소중립 이슈와도 연계하여, 2050년 탄소중립 목표 달성을 위해 인공지능 기반 기술의 도입이 가속화할 것으로 보인다. 구체적으로, 산업 전반에 걸쳐 인공지능을 통한 에너지 소비 패턴 분석, 탄소 배출 감축 전략 등을 수립하고 실행할 것이다.

라 인공지능에 대한 국가 정책 현황

인공지능은 생활의 편리함과 다양한 혜택을 제공하지만, 동시에 피해를 초래할 가능성도 존재한다. 이에 따라 각국은 인공지능 규제 정책에 대해 상반된 접근을 보인다. 미국과 중국 등의 인공지능 선도국은 인재들을 지속적으로 양성하고 기업의 자율 규제에 초점을 맞춘 정책을 추진하는 한편 EU는 규제 강화를 통해 잠재적 위험을 관리하려는 경향이 강하다. 그러나 최근에 유럽에서도 인공지능 산업을 촉진하기 위해 규제를 완화하려는 움직임이 있다. 이제 주요국들의 인공지능 정책을 살펴보자.

1) 미국

미국의 디지털 인재 양성 정책

미국의 AI 정책은 글로벌 시장에서 AI 경쟁력을 유지하고, 안전하고 신뢰할 수 있는 AI 발전을 추구하는 것이다. AI 인재 확보를 위해 '국가 AI R&D 전략계획', '안전하고 신뢰성 있는 AI를 위한 행정명령' 등을 통해 AI 분야에 필요한 기술 인력을 평가하여 교육하고 글로벌 AI 인력 확보를 위해 비자 제도를 개편하였다. 또한 「인공지능교육법」을 제정하여 AI 오용 가능성을 감소시키고, 정부 관계자들이 정부의 수요에 가장 적합한 인공지능 시스템을 도입할 수 있도록 하기 위해 연방 행정 각부, 산하기관 등에 AI 교육을 시행하고 있다.

바이든, 최초 인공지능 행정명령 발령(2023.11.30.)

인공지능의 안전 및 보안에 대한 새로운 표준을 확립하기 위해 최초 인공지능 행정명령을 발령하였으며, 이를 통해 개인정보 등 민감한 정보를 보호하며 인공지능 알고리즘이 차별 악화에 사용되는 것을 방지하고자 하였다. 뿐만 아니라, 인공지능 시스템에 대해 안정성과 신뢰성을 확인할 수 있는 테스트 등을 확립하도록 지시하였다.

미국 유타주의 「인공지능 수정법」(Artificial Intelligence Amendments, SB0149)

유타주(State of Utah)는 주정부 차원으로는 최초로 생성형 인공지능 이용 고지 및 인공지능 기술개발과 규제 완화를 위한 법적 근거로 「인공지능 수정법」을 마련(24.03.13. 제정, 24.05.01. 시행)하였다(채은선, 2024).

- **생성형 인공지능을 이용하여 서비스를 제공하는 경우 이를 소비자에게 고지**

  규제 직종 종사자(회계사, 건축가, 치료사 등)가 생성형 인공지능을 사용할 때 소비자에게 고지 의무가 부과되었다. 통신, 재산 거래, 웹사이트, 소비자 개인정보 등 소비자 보호 분야에서 생성형 인공지능을 사용 시, 요청이 있을 경우 명확하고 분명하게 고지를 해야 한다. 만약, 공개 의무 위반 시 최대 2,500 USD의 과태료 부과 및 민사상 벌금 부과가 가능하고 법무장관이 대신 소송 제기가 가능하다.

- **인공지능정책법 수립**

  상무부에 인공지능 정책 사무소를 설치·운영하고 인공지능 연구 프로그램 수립·운영하도록 정책을 펼쳤다.

2) 중국

중국의 인공지능 인재 양성

2017년 8월 중국 공산당 중앙위원회와 국무원에 의해 발표된 '차세대 인공지능 발전계획'을 통해 '전 국민의 인공지능 교육 사업을 단계적으로 실시하고 초중등 학교에 인공지능 관련 교과를 개설하며, 프로그래밍 교육을 점진적으로 추진' 등 다양한 단계적 목표를 세우며 실질적인 발전을 이뤄내고 있다. 이후 지속해서 '고등교육기관 AI 혁신 행동 계획', '중국 인공지능 인재양성 백서' 등을 통해 국가 주도로 AI 인재를 양성하고 있으며, 대학을 중심으로 기업이 보조하는 형태의 인재 양성을 추진하고 있다. 대학에서 실무에 바로 투입할 수 있는 실습 기반 교육을 하고, 빅테크 기업에서 인공지능 대회, 단기 교육을 통해 실전 경험을 유도하여 인재 인증 제도를 활성화함으로써 AI 인재 양성에 힘쓰고 있다(이수진, 2023).

인공지능 산업 육성책 'AI+행동' 공개

중국 정부는 2024년 3월 5일 전국 인민대표대회에서 발표한 업무보고서에서 디지털 경제 혁신 발전을 강조하며, 디지털 산업화와 산업 디지털화를 적극 추진할 것을 밝혔다. 'AI+행동'의 주요 내용으로는 디지털 기술과 실물 경제의 융합 촉진, 빅데이터 및 인공지능 연구·응용 심화, 데이터 기초 체계 완비, 인터넷 플랫폼 기업 지원 등이 있다.

3) EU

독일의 인더스트리 4.0

2006년 하이테크 전략(Die Hightech-Strategies fur Deutschland)에서 인더스트리 4.0을 최초로 개념화한 후, 현재 2030년까지의 장기 비전을 가지고 관련 정책을 추진 중이다. 인더스트리 4.0 정책의 추진은 2016년 다보스 포럼에서 제4차 산업혁명이 주제로 선정되기 이전부터 시작된 것으로 제조업 강국으로서 앞선 행보를 보인다. 인더스트리 4.0의 개념 및 전략은 독일 AI 연구소가 주도한 것으로 제조업 시스템의 지능화를 목표로 하고 있다(오연주, 2022).

독일과 프랑스 등 대형언어모델 규제를 위한 강력한 장치 마련 반대

프랑스, 독일, 이탈리아 등 여러 EU 회원국 대표자들이 기초 모델에 대한 모든 유형의 규제를 반대하고 있다. 이는 자국 내 인공지능 기업을 대표 대형언어모델 기업으로 키우기 위함이 크다. 프랑스의 경우 자국의 '미스트랄 AI'를 전폭 지지하고 독일 역시 작년 11월 'Aleph Alpha'에 5억 달러(약 6,500억 원)를 몰아주며 지원하고 있다. 이들은 비영어권 데이터로 학습한 인공지능의 필요성을 주장하고 있어 단순히 국가 문제로 보기 어렵다는 입장이며 EU가 인공지능 발전을 막아 미국이나 중국보다 더 뒤처질 수 있다는 논리로 반대를 하고 있다.

4) 우리나라 현황 및 정책적 대비 상황

우리나라 인공지능 기술 현황

우리나라의 인공지능 기술수준은 아직 세계 최고 수준인 미국과 중국에 비해 미흡한 측면이 있지만, 최근 몇 년간 가장 빠르게 발전하고 있다. 전반적 기술수준에서는 2022년 기준 미국을 100으로 할 때, 한국은 88.9%로 평가되었으며, 기초단계, 응용단계, 사업화 단계에서도 각각 88.0%, 90.1%, 88.6%를 기록했다. 특히 응용단계 인공지능 기술수준은 2018년 대비 8.7% 향상되어 주요국 중 가장 빠른 발전을 보였다. 또한 AI 기술분야별로 보면, 학습지능 분야에서 미국 대비 기술격차가 2018년 2.0년에서 2022년 1.3년으로 축소되었고, 단일지능 분야에서도 2018년 2.0년에서 2022년 1.5년으로 단축되었다. 복합지능 분야에서는 2018년 2.0년에서 2022년 1.0년으로 축소되는 등 우리나라는 전반적으로 기술격차를 줄이며 빠르게 성장하고 있다(봉강호, 2024).

우리나라의 정책적 지원

과학기술정보통신부는 2023년 4월, '초거대 인공지능 경쟁력 강화 방안'을 발표하며, 국가전략기술로서 AI를 포함한 연구개발 예산을 확대 편성했다. 이는 AI 기술발전을 위한 선제적이고 체계적인 지원 방안으로, 다음과 같은 주요 정책을 포함한다(진회승 외, 2023).

- **연구개발 투자 확대**
  인공지능 분야의 핵심기술 개발을 위한 연구개발 투자를 확대하고, 이를 통해 기술력을 강화한다.
- **인재양성**
  인공지능 분야의 우수 인재를 양성하기 위해 대학, 연구기관, 기업 간의 협력 프로그램을 운영한다.
- **산업 생태계 조성**
  인공지능 스타트업을 육성하고, 대기업과 중소기업 간 협업을 촉진하여 인공지능 산업 생태계를 구축한다.
- **국제 협력 강화**
  글로벌 인공지능 선도국과의 협력을 통해 기술 교류 및 공동연구를 추진한다.

정책적 대비 방안

- **지속적인 연구개발 투자**
  정부와 민간이 협력하여 지속적인 연구개발 투자를 확대하고 이를 통해 기술 격차를 줄이는 노력이 필요하다.
- **실용적 인공지능 응용 확대**
  인공지능 기술의 상용화와 응용을 촉진해 산업 전반에 걸친 생산성 향상을 도모한다.
- **데이터 인프라 강화**
  인공지능 개발에 필수적인 데이터의 확보와 활용을 위해 데이터 인프라를 강화하고, 데이터 공유와 활용을 촉진하는 정책을 마련한다.
- **국제 표준화 및 규제 마련**
  인공지능 기술의 글로벌 경쟁력을 확보하기 위해 국제 표준화 작업에 적극 참여하고, 인공지능 기술에 대한 윤리적, 법적 규제를 마련하여 안전한 인공지능 활용 환경을 조성한다.
- **민관 협력 강화**
  정부와 민간의 긴밀한 협력을 통해 인공지능 기술 발전과 산업 생태계 구축을 촉진하며, 국가 차원의 종합적 전략을 수립한다.

이와 같은 정책적 대비를 통해 우리나라는 글로벌 인공지능 경쟁에서 우위를 정하고, 차세대 인공지능 기술 패러다임의 변화를 선도할 수 있을 것이다.

마 맺음말

인공지능 기술의 급격한 발전은 이미 우리의 일상생활과 다양한 산업 분야에서 그 영향을 보여주고 있으며, 앞으로도 그 중요성과 역할은 더욱 커질 것으로 예상된다. 생성형 인공지능의 발전과 더불어 대형언어모델과 멀티모달 언어모델의 등장으로 인공지능의 적용 범위와 효율성은 획기적으로 향상되고 있다. 이러한 기술 발전은 특정 작업에 국한되지 않고 인간처럼 다양한 지적 작업을 수행할 수 있는 AGI의 연구로 이어졌으며, 인류가 직면한 복잡한 문제들을 해결하는 데 새로운 가능성을 제시한다. 그러나 AGI의 실현을 위해서는 복합적 문제 해결 능력, 자가 학습 능력, 윤리적 문제 해결 등의 과제가 남아 있으며, 이를 극복하기 위한 기술적, 정책적 노력이 필요하다. 이에 따라 각국 정부는 인공지능 기술 발전을 지원하고 규제하는 정책을 마련하고 있으며, 우리나라도 연구개발 투자 확대, 인재 양성, 산업 생태계 조성, 국제 협력 강화 등의 정책적 노력을 통해 글로벌 인공지능 경쟁에서 우위를 점하고자 한다.

결론적으로 인공지능 기술의 발전은 우리 사회의 혁신과 변화를 주도할 것이며 이러한 변화를 효과적으로 관리하고 활용하기 위해서는 기술적 발전과 함께 윤리적, 법적, 정책적 대비가 필수적이다. 지속적인 연구개발과 국제 협력을 통해 인공지능 기술을 선도하고, 이를 통해 더 나은 미래를 만들어 나가는 노력이 필요하다.

참고문헌

봉강호(2024). "우리나라 및 주요국 인공지능(AI) 기술 수준의 최근 변화 추이(2023년 조사 기준)", 제116권, 소프트웨어정책연구소, pp. 4-11.

오연주(2022). "주요국 인공지능(AI) 전략 분석(하)", 제10호, 한국지능정보사회진흥원(NIA), IT & Future Strategy 보고서, pp. 1-93.

이수진(2023). "중국의 인공지능(AI) 교육 현황 및 주요 정책분석", 인공지능연구 논문지, 제4권 제2호, 한국인공지능교육학회, pp. 1-14.

진회승·윤보성·신승윤(2023). "생성형 AI에 대응한 SW 인재 양성 정책 방향 연구", 제158권, 소프트웨어 정책연구소, pp. 1-154.

채은선(2024). "美유타주, 「인공지능 수정법」의 주요 내용 및 시사점", 제24권 제5호, 한국지능정보사회진흥원(NIA) 디지털 법제 Brief, pp. 1-6.

Alayrac, J.-B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., Lenc, K., Mensch, A., Millican, K., Reynolds, M., Ring, R., Rutherford, E., Cabi, S., Han, T., Gong, Z., Samangooei, S., Monteiro, M., Menick, J. L., Borgeaud, S., Brock, A., Nematzadeh, A., Sharifzadeh, S., Bińkowski, M., Barreira, R., Vinyals, O., Zisserman, A. & Simonyan, K.(2022). "Flamingo: a Visual Language Model for Few-Shot Learning", Advances in Neural Information Processing Systems (NeurIPS), Vol.35, pp. 23716~23736.

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X. & Gao, J.(2024). "Large Language Models: A survey", arXiv preprint arXiv:2402.06196.

Morris, M. R., Sohl-Dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C. & Legg, S.(2024). "Position: Levels of AGI for Operationalizing Progress on the Path to AGI", Proceedings of the International Conference on Machine Learning (ICML), Vol.235, pp. 36308~36321.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G. & Sutskever, I.(2021). "Learning Transferable Visual Models From Natural Language Supervision", Proceedings of the International Conference on Machine Learning (ICML), Vol.139, pp. 8748~8763.

Zhang, D., Yu, Y., Dong, J., Li, C., Su, D., Chu, C. & Yu, D.(2024). "MM-LLMs: Recent Advances in MultiModal Large Language Models", Findings of the Association for Computational Linguistics (ACL), Vol.1 No.738, pp. 12401~12430.

KAST

3

AI 기술 발전 전망 및 국내 산업의 새로운 도약 전략

이경무
서울대학교 전기정보공학부 석좌교수

가 들어가는 말

40년 전 필자가 처음 인공지능(AI)에 관심을 가졌을 때만 해도, AI는 우리 세대에서는 결코 이루어질 수 없는 환상에 가까운 기술이었다. 불과 10여 년 전만 해도 운전자 없는 자율자동차, 인공지능 비서, 그리고 지능형 로봇들은 영화나 만화에서 보아왔지만, 상상일 뿐 그것들이 가까운 미래에 현실이 될 것이라고 믿는 사람은 거의 없었다. 2012년 혁신적인 AlexNet이 발표된 이후 딥러닝(Deep Learning) 기반의 AI 기술은 믿을 수 없을 만큼 빠른 속도로 발전하고 있다. 그러나 2010년도 초중반까지만 해도 AI 분야 학자들 사이에서조차 딥러닝의 성능에 대해 반신반의하는 경향이 있었고 그 지속 가능성에 대해서도 의견이 분분했다. 그 이후 AlphaGo, ResNet, GAN(Generative Adversarial Network), Transformer, Diffusion Model 등 여러 획기적인 돌파(breakthrough) 기술들의 개발로 다양한 분야의 기존 난제들이 해결될 수 있음을 속속 보여줌으로써 현재는 모든 문제에 이러한 딥러닝 기술을 사용하는 것이 학계와 산업계의 표준이 되었다.

딥러닝 기반 AI 기술의 발전에는 새로운 아키텍처와 학습 알고리즘의 개발도 기여한 바 적지 않지만 최근의 경향은 'scaling law', 즉 규모의 법칙이 견인한 측면이 크다. 대규모의 GPU와 수많은 학습데이터로 훈련된 챗GPT와 같은 생성형 대형언어모델(LLM)에 천문학적인 자원이 투입되었는데, 전문가들조차도 예상치 못한 결과로 모두를 놀라게 하고 있다. 이러한 초거대 인공지능 모델 개발에 대한 노력은 당분간 가속화할 것으로 예상되며 아직 한계에 도달했다고 보기에는 이르다. 오히려 지금부터가 시작이라고 보는 것이 타당할 것이다. 최근에는 단순한 언어모델을 넘어서 이미지, 비디오, 음성, 텍스트 등을 복합적으로 분석하고 이해하는 멀티모달(multi-modal) AI, 그리고 로봇, 가전, 기계 등 물리적인 기기에 결합하는 임바디드(embodied) AI로의 확장이 본격화하기 시작했는데, 이에 대해 많은 사람들은 범용인공지능(AGI)의 도래가 목전에 왔을 수도 있다는 기대를 품기도 한다.

현재는 AI 기술의 급작스러운 출현으로 인류 역사 상 가장 중대한 산업적, 사회적 변혁이 일어날 것이라는 기대와 동시에 우려가 교차하고 있다. 이전의 산업혁명이 특정 분야에서의 변혁을 가져왔고 상대적으로 장기간에 걸쳐 진행되었던 것에 비해 AI에 의한 혁명은 모든 산업과 인류의 삶 전체의 패러다임을

근본적이며 단시간에 바꿀 수 있는 특징을 가지고 있어 무게가 다르다. 이는 AI로 인해 기존의 국가, 산업, 노동, 경제, 사회, 개인 간의 역할과 질서가 새로운 차원으로 급격히 재편될 수 있다는 것을 의미한다.

따라서 이러한 대변혁의 시기에 새로운 AI 기술의 능력과 파급력, 그리고 위험 요소를 정확히 파악하고 경쟁력 있는 핵심기술을 확보함으로써 우리 산업과 사회를 한 단계 도약시킬 수 있는 전략을 마련하고 실행하는 것은 미룰 수 없는 숙제이다.

나 AI 기술의 현황 및 발전 방향

1) AI 기술의 현황

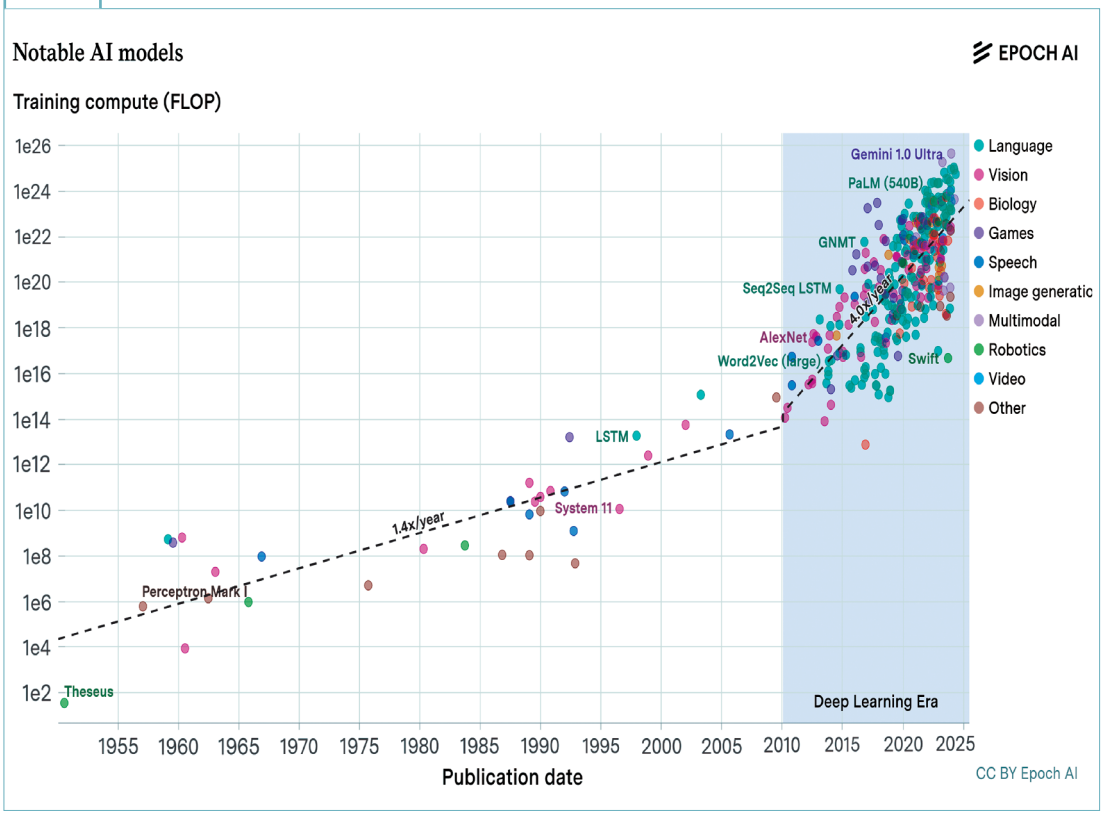
1957년에 개발된 최초의 인공 신경망인 Perceptron Mark I(Rosenblatt, 1957)는 오른쪽에 표시된 카드와 왼쪽에 표시된 카드를 구별하는 단순한 태스크를 수행하였다. 이 네트워크는 1,000개의 인공 뉴런으로 이루어져 있었으며, 훈련에는 약 70만 번의 연산이 필요했다. 65년이 지난 2023년 OpenAI에서 발표한 대규모 언어모델인 챗GPT4(OpenAI, 2023)는 1.7조 개의 파라미터로 구성되었으며 훈련을 위해 13조 개의 토큰으로 55조 번의 연산이 필요했는데, 이를 수행하는 데 25,000개의 NVIDIA A100 GPU로 100일의 시간이 소요되었다. Perceptron Mark I는 단 6개의 데이터 포인트로 훈련된 반면, 최근 메타에서 개발한 대형언어모델인 LlaMa는 약 10억 개의 데이터 포인트로 훈련되었으며, 이는 Perceptron Mark I보다 1억 6천만 배 이상 증가한 것이다.

<그림. 4>는 AI 시스템의 컴퓨팅 증가 추이를 보여주는데, 2010년 이후 최근까지 학습에 사용된 계산량이 매년 약 4배씩 증가한 것을 알 수 있다(Epoch, 2024). 이와 더불어 학습 데이터셋 연 2.5배, 학습 알고리즘 효율성 연 1.5~3배, 그리고 하드웨어 효율성 또한 연 1.5배가량 증가된 것으로 보고되고 있다.

이러한 많은 계산량과 데이터를 사용하는 'scaling up' 방법이 최근의 획기적인 AI 능력 향상을 주도했다는 점에는 많은 AI 연구자들이 동의하고 있다. scaling up 방식은 AI 시스템으로 하여금 더 많은 양의 데이터를 학습하게 하고, 다양한 경우의 수를 배우게 하며 데이터 내 변수 간의 관계를 더 고차원적으로 모델링할 수 있게 함으로써 더 정확하고 심층적인 결론을 도출하도록 한다.

특히 최근의 생성형 모델에 기반한 대규모 AI 시스템의 비약적인 발전으로 챗봇뿐 아니라 이미지, 동영상, 3D 생성 분야에서까지 획기적인 변화가 일어나고 있다. 2024년 스탠포드대학교의 AI Index

그림. 4 AI 모델들의 학습계산량 추이



출처: Sevilla & Roldan(2024).

Report(Stanford University, 2024)에 따르면 현재의 AI 시스템은 이미지 분류, 시각적 추론 및 영어 이해를 포함한 벤치마크 등 여러 분야에서 인간의 성능을 이미 능가하였고 수학, 시각적 상식 추론 및 계획과 같은 복잡한 작업에 있어서도 인간 수준에 근접하고 있는 것으로 나타났다.

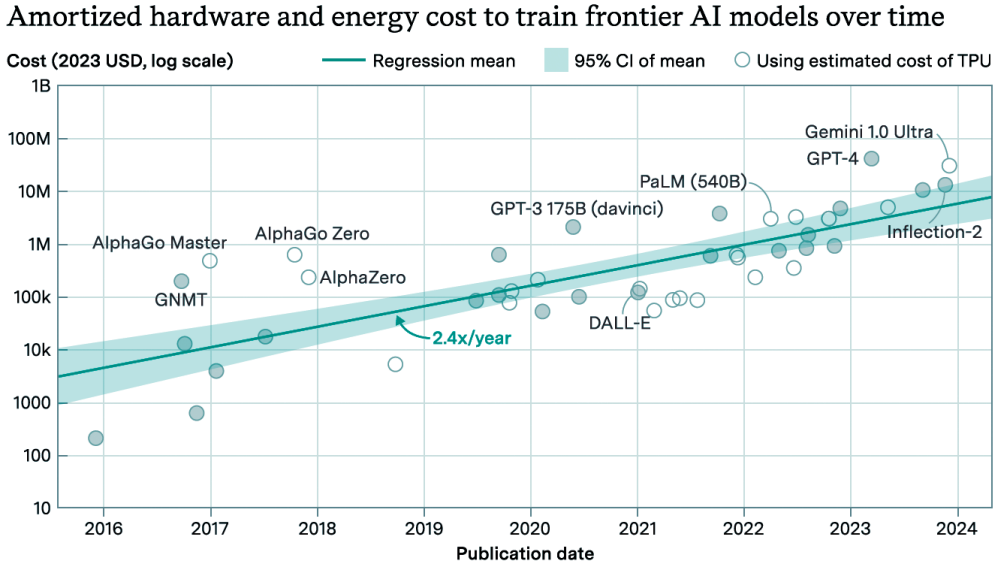
최근의 대규모 AI 모델 개발이 특정 기업들을 중심으로 이루어지고 있는 데 대해 기대와 우려가 공존한다. 오픈AI나 구글과 같은 기업들이 이러한 인공지능 모델 개발에 많은 자원을 투입하지 않았다면, 우리는 오늘날처럼 발전된 AI 모델을 갖지 못했을 것은 물론이거니와 인공지능의 잠재력이 얼마나 큰지조차 알 수 없었을 것이다. 반면에, 이윤을 추구하는 기업들이 AI와 같은 파괴력이 있고 위험성을 내포하고 있는 기술을 사회가 따라가고 적응하기도 전에 빠르게 개발하고 결정하는 상황에 대한 우려도 크다. 이것은 AI의 안전 및 규제와 관련된 문제이며, AI 기술이 인류에게 공공의 선으로서 이익이 되는 방향으로 개발되도록 유도하고 보장하는 것이 무엇보다 중요하다.

2) 거대 AI 모델의 발전과 지속 가능성

오늘날 우리가 목도하고 있는 AI의 놀라운 성능은 대부분 소위 scaling law에 기인하는데, 컴퓨팅 파워와 데이터의 규모가 커짐에 따라 개발자들조차도 예측하지 못한 성능이 발현되는 현상 즉, 'emergent ability' 또는 '양에서 질로의 변환' 현상이 두드러지고 있다. AI 기술이 현재와 같은 속도로 계속 발전할 수 있는가, 특히 이러한 양적 증가에 기반한 발전이 지속 가능할지에 대해서는 의견이 분분하다. 이것은 구체적으로 컴퓨팅 자원의 관점, 알고리즘 관점, 비즈니스와 산업적 관점, 그리고 규제적 관점에서 살펴볼 수 있을 것이다.

먼저 컴퓨팅 자원의 관점에서 보자면, <그림. 5>처럼 대규모 AI 학습에 필요한 하드웨어 수요는 지난 10여 년간 기하급수적으로 증가해 왔고 앞으로도 계속해서 증가할 전망이다. AI 개발을 위해서는 일반적으로 자체 컴퓨팅 인프라를 구축하거나, 클라우드 컴퓨팅 서비스를 이용해야 한다. 챗GPT4의 경우 학습에만 약 1억 달러 이상이 소요된 것으로 알려져 있다.

그림. 5 AI 모델의 학습비용 추이

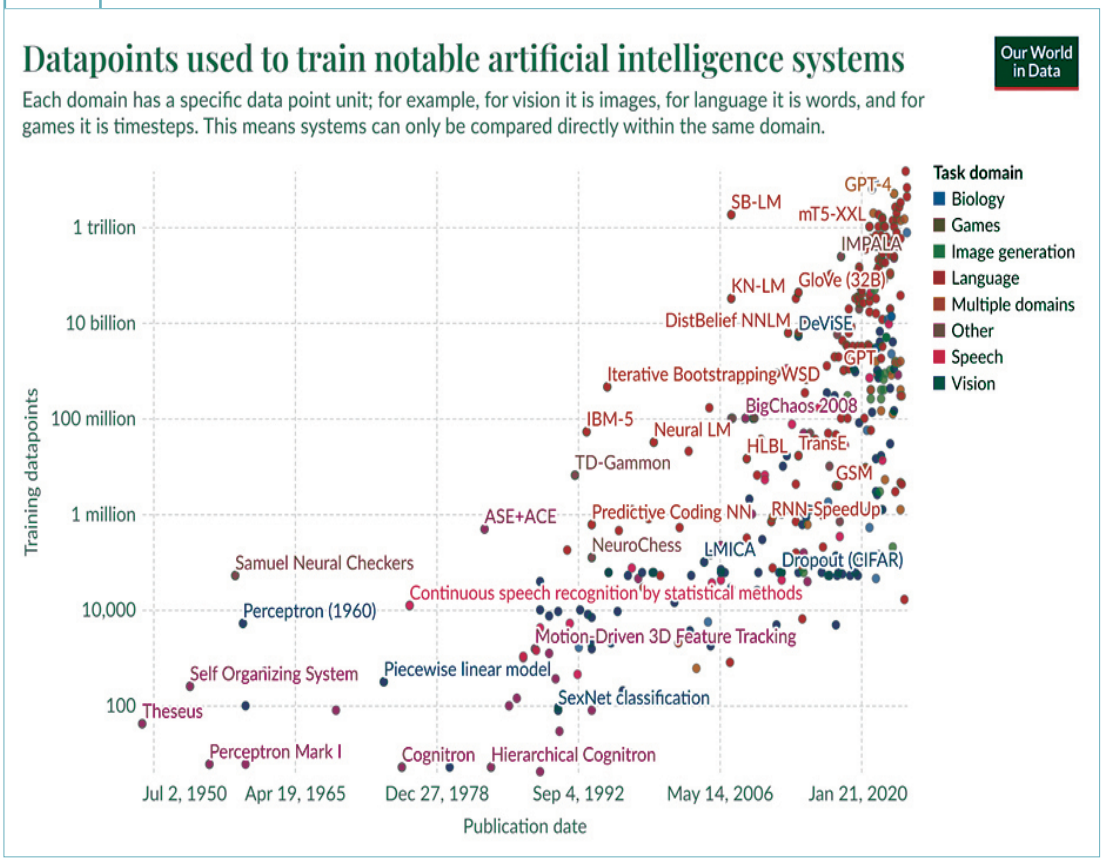


출처: Cottier et al.(2024).

AI 시스템을 학습하는 것뿐만 아니라 추론(inference)하는 데에도 컴퓨팅 자원이 크게 증가하고 있다는 점도 주목할 필요가 있다. 몇몇 거대 모델의 경우 학습에 소요되는 컴퓨팅보다 추론에 필요한 자원이 많은 것으로 보고되고 있다. 더군다나 조만간 텍스트만이 아닌 멀티모달 데이터 기반의 AI 시스템이 대세가 될 것이라는 점을 고려하면 컴퓨팅 자원의 수요는 상상 이상이 될 가능성이 있다. 그러나 고성능 AI 서

버를 만드는 데 필요한 GPU, 패키징, 고대역폭 메모리 및 기타 구성 부품들의 공급망 부족, 그리고 필요한 수요를 충족하기 위한 반도체 공장의 증설 등은 물리적·시간적 한계가 있기 때문에 AI 발전을 지연시키는 요소로 작용할 수 있다.

그림. 6 AI 모델 학습데이터 양 추이



출처: Our World in Data(2024a).

데이터의 관점에서 보면, 최근 대규모 AI 모델들의 경우 그동안 인터넷 등에서 텍스트, 영상 등 대규모의 학습데이터들을 수집함으로써 매년 2.5배 이상 그 규모를 늘릴 수 있었다(<그림. 6>). 이러한 추세를 근거로 Epoch은 2026년을 기점으로 학습에 사용할 수 있는 고퀄리티의 텍스트 데이터가 고갈될 것이라는 전망도 내놓고 있다(Epoch, 2024). 2024년부터는 이미지, 오디오, 동영상 등 멀티모달 AI 시스템들의 개발이 본격화하기 시작했는데, 멀티모달과 관련해서는 아직 많은 양의 데이터들이 존재하기는 하나 저작권과 비용문제가 걸림돌이 될 수도 있다. 그리고 가상데이터를 생성해서 사용하는 기법들이 고도화됨에 따라 실제와 구별할 수 없는 가상데이터의 무한한 생성 및 활용이 대안이 될 수 있다. 또한 AlphaGo 학습의 예에서처럼 다중 AI 시스템 간의 상호 학습 기법을 통해 데이터 부족 문제를 해결할 수도 있다. 더불어 AI 알고리즘의 진보에 따라 적은 학습데이터 또는 학습데이터를 필요로 하지 않는 few-shot learning,

unsupervised learning, self-learning 등의 학습 기법들이 발전함에 따라 실제 데이터에 대한 의존도가 약해질 가능성도 있다. 데이터 관점에서 한 가지 더 주목할 부분은 향후 범용 pre-trained AI 시스템들이 도메인 특화 시스템 또는 개인맞춤형 시스템으로 세분화될 것이므로 이에 필요한 분야별 특화 데이터들의 확보가 중요한 이슈가 될 것이라는 점이다.

알고리즘 관점에서 보면 그동안 새로운 AI 알고리즘의 개발은 하드웨어와 데이터를 효과적으로 사용할 수 있게 함으로써 리소스에 대한 의존도를 낮추는 역할을 해 왔다. 최근 범용 AI 시스템들에 사용되는 학습 알고리즘의 효율성은 매년 1.5~3배 정도 증가한 것으로 보고되고 있다. Epoch의 연구(Epoch, 2024)에 따르면 매 9개월마다 더 효율적인 알고리즘의 도입으로 컴퓨팅 예산의 두 배 증가와 맞먹는 효과를 가져왔다고 한다.

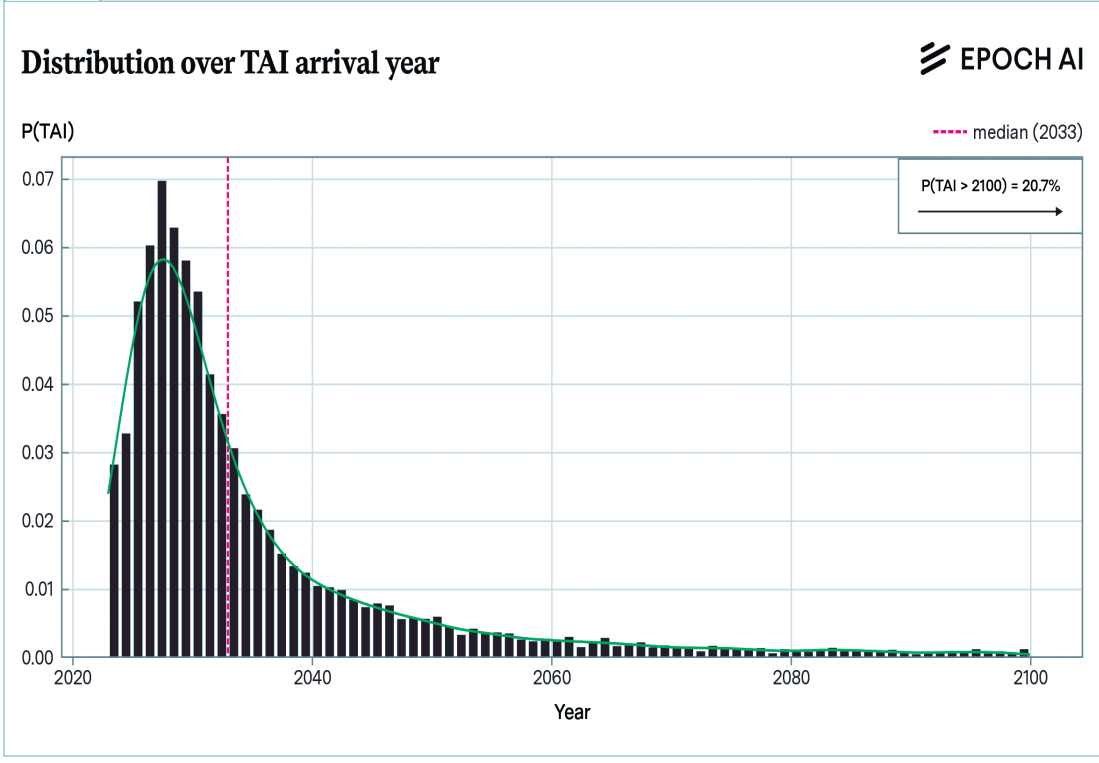
에너지 문제 또한 AI 발전의 지속 가능성에 걸림돌로 작용할 수 있다. AI 학습에 필요한 전력량은 매년 4배씩 증가하고 있고 하드웨어의 개선과 알고리즘의 효율화를 감안하더라도 전력소비량은 증가할 것으로 전망된다. 현재의 추세대로라면, 전 세계적으로 AI 시스템을 운영하는 데 사용되는 전기량이 2026년경 70TWh로 오스트리아 전체의 전기사용량에 육박한다는 보고가 있다. 발전소를 새로 착공하고 전력 인프라를 확충하는 것은 물리적 시간을 요하는 것이므로 지연의 요소가 될 수 있다.

현재의 기업주도의 대규모 AI 개발 트렌드가 지속 가능하기 위해서는 투자대비 이익이 확보되어야 한다. 챗GPT4의 경우 총 개발비용에 대한 정확한 정보가 제공되고 있지는 않지만 학습하는 데 약 1억 달러가 소요되었다고 한다. 오픈AI는 2024년 37억 달러의 매출을 예상하고 있는데, 이는 전년도 대비 200% 성장을 의미한다. 대부분의 수익이 챗GPT 구독서비스에 의해 발생하고 있는데, 챗GPT 운영 비용이 매일 약 70만 달러 정도인 것을 고려하면 연 2억6천만 달러의 비용이 투입되는 것이므로 비용대비 수익은 충분하다고 볼 수 있다. 현재, 챗GPT의 경우 유료 구독자의 수가 2024년 말 기준 전세계적으로 수천만 명 규모인 것으로 알려져 있다. 챗GPT-4o가 본격적으로 공개되면 더 많은 사용자가 가입할 것으로 예측하지만, 전문가들뿐 아니라 일반인들이 일상생활 속에서 AI를 보편적으로 사용할 수 있도록 하기 위해서는 성능 혁신과 더불어 서비스 고도화의 필요성이 존재한다. 향후 챗GPT5와 같은 한층 고도화된 모델이 출시되고, 약 1억 명의 사용자가 챗GPT 서비스에 유료 가입한다고 가정하면, 매달 20억 달러, 연 240억 달러의 매출이 발생할 수 있다. 이것은 전 세계 인구의 단 1.3%만이 사용한다는 가정이고 챗GPT가 스마트폰과 같이 일상화될 경우에는 그 시장규모가 상상을 초월할 것이다. 또한 단순한 챗봇으로의 상품에 그치지 않고 다양한 응용파생기술과 제품들이 등장한다면 시장은 급격히 확대될 수 있다. 그러나 AI 기술의 특성상 대규모 투자와 초격차기술 확보의 선순환 구조를 구축한 오픈AI나 구글과 같은 선도기업이 시장을 독점할 가능성이 크며 이를 타개할 만한 새로운 기술적 혁신 없이는 나머지 대부분의 기업들은 상대적으로 경쟁에서 도태되는 어려움에 처할 수도 있다.

3) 범용인공지능은 실현 가능한가? 언제 가능하게 될 것인가?

최근 AI의 급속한 발전으로 범용인공지능(AGI)이 언제 실현 가능할지에 대해 많은 관심이 집중되고 있다. 사실 AGI 또는 Super-intelligence에 대한 명확한 정의가 확립되어 있지 않고 또 그것을 측정할 만한 명확한 기준이 제시되지 않은 상태에서 언제 그것이 달성될 것인지에 대한 질문은 모호하고 개인마다 대답이 다를 수 있다. 넓은 의미로 AGI를 통상 '대부분의 인지관련 태스크들에 대해 사람과 동등 또는 그 이상의 능력을 갖는 AI 시스템'으로 해석하더라도 그것의 도래 시기에 대해서는 전문가 간 견해차가 크다. Ray Kurzweil(Google director), Louis Rosenberg(Unanimous AI CEO), Elon Musk(Tesla CEO), Sam Altman(OpenAI CEO), Dario Amodei(Anthropic VEO) 등은 AGI가 5년 내지 10년 사이에 개발될 것이라고 예측하고 있는 반면, Jürgen Schmidhuber(co-founder of NNAISENSE), Rodney Brooks(MIT professor) 등은 수십 년 또는 수백 년이 걸릴 것으로 내다봤다. Epoch의 연구자들이 최근 scaling law에 기반한 계산량 기준으로 TAI(Transformative AI; AGI와 유사개념)의 도래 시기를 예측한 결과에 따르면(<그림. 7>), 2025년까지는 10%, 2033년까지는 50%의 확률로 AGI가 달성될 것으로 분석되었다. 이는 여러 선도적 AI 기업의 CEO들이 예측한 것과 유사하다.

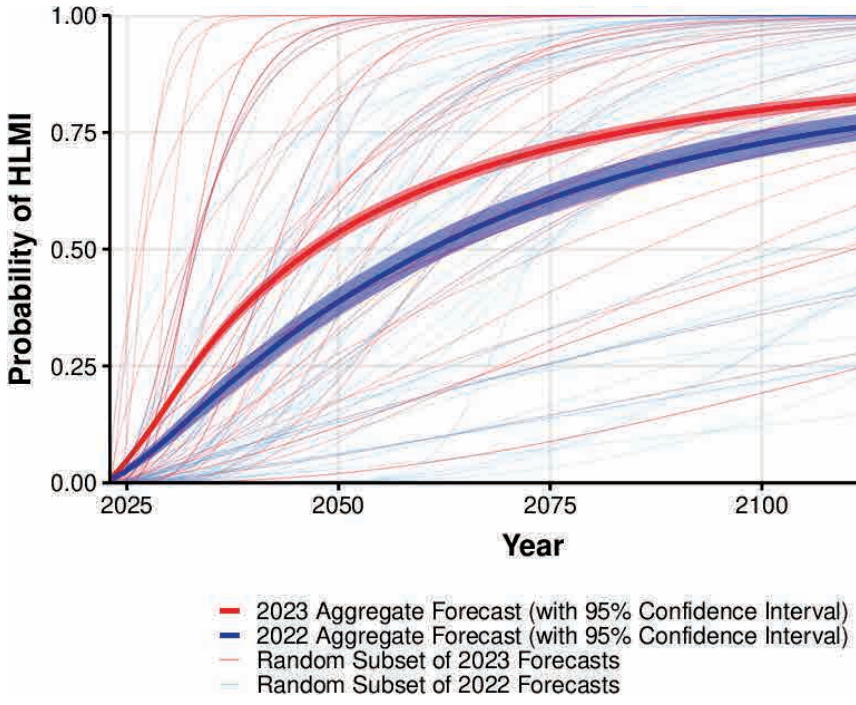
그림. 7 Epoch의 TAI(AGI) 도래 시기 예측



출처: Atkinson et al.(2024).

2023년 2,778명의 전문가들을 대상으로 한 설문조사(Grace et al., 2024)에서는 10%가 2027년, 50%가 2047년경으로 AGI 도래 시기를 예상하고 있다(<그림. 8>). 동일한 질문에 대한 2022년 설문조사에서 전문가들의 10%가 2029년, 50%가 2060년을 예상했던 것과 비교하면 1년 사이에 많은 변화가 있었음을 알 수 있다. 이는 전문가들조차도 AI의 빠른 발전을 예상하지 못했다는 것을 방증한다. 만약 챗GPT-4o, Sora, Figure Robot 등 더욱 고도화된 시스템들이 발표된 현 시점에서 동일한 설문조사가 진행된다면 훨씬 많은 사람들이 AGI 도래 시기를 앞당겨 예측할 것이라고 본다. 분명한 사실은, AGI 도래 시점에 대해서는 여전히 의견이 분분하나 예상보다 그 시기가 훨씬 앞당겨질 수 있다는 점에는 대부분 공감한다는 것이다.

그림. 8 AI 전문가들의 AGI 도래 시기 예측



출처: Grace et al.(2024).

필자는 AGI의 개념에 실제 환경에서 인간이 수행할 수 있는 모든 태스크들을 포함한다면 물리적인 환경에서의 멀티모달 인식, 추론, 동작, 상호작용까지 수행하는 완벽한 휴머노이드 로봇의 개발을 의미하는 것이기 때문에 도래 시기가 훨씬 늦어질 수 있겠지만, 협의적 개념으로 영화 'Her'에서의 '아만다', 그리고 '아이언맨'에서의 '자비스'와 같은 챗봇 기반의 AI assistant 관점에서 본다면 현재의 대형언어모델들은 몇몇 부족한 점에도 불구하고 AGI의 초기 수준에 상당히 근접한 것으로 보인다. 이는 최소한 채팅에 있어서는 전문지식뿐 아니라 일상대화에서 Turing test를 통과할 수준이 되었다고 보기 때문이다. 따라

서 빠르면 3~5년 안에 성숙된 단계의 AGI 수준 챗봇을 만나볼 수 있을 것으로 기대된다. 멀티모달 영역은 이제 막 시작 단계이나 이 또한 머지않아 인간 수준에 버금가는 결과를 보여줄 것으로 예상된다. 이것은 scaling law에 의해 예측 가능한 'more and more' 특성뿐 아니라 대규모의 멀티모달 학습이 이루어졌을 때, 우리가 예상하지 못한 더욱 놀라운 emergent ability를 갖게 되는 'more and new' 현상이 발현할 가능성이 더욱 높아질 수 있기 때문이다.

AGI의 도래가 먼 미래가 아닐 수 있다는 사실은 사람들에게 희망과 공포를 동시에 불러일으킬 수 있다. 노동으로부터의 해방, 생산력 극대화, 복지 및 삶의 질 향상, 난제 해결 등 유토피아적 기대와 함께 AGI의 통제 상실로부터 오는 위협, 개인·국가 간 AI 격차 및 불평등 심화, 노동시장의 변화 등 디스토피아적 견해도 설득력을 얻고 있다. 이러한 AGI 관련 이슈는 현재 챗GPT 수준의 general purpose AI(GPAI)가 갖는 것과 차원이 다른 심각성을 내포하고 있고, 우리에게 AGI를 어떻게 전개하고 받아들일지에 대한 직접적이고 빠른 판단과 행동을 요구하고 있다. 각국의 상황에 따라 서로 다른 전략을 모색하겠지만, 우리나라의 경우 기본적으로는 지속적인 기술개발을 추구하되 최소한의 안전장치(guard rail) 마련을 병행하는 것이 합리적일 것이다.

다 우리의 AI 역량 및 현주소

우리나라의 AI 역량을 객관적으로 파악하는 것은 단순히 현재의 위상을 확인하는 데 그치지 않고 장·단점을 분석하여 경쟁력을 향상시키기 위한 전략을 수립하는 데에도 필요하다. 국가의 AI 역량을 하나의 지표로 나타내는 것은 쉽지 않은 일이다. Tortois에서 발표한 2023년 Global AI Index(Tortois, 2023)가 종합적인 국가경쟁력을 나타내고 있는데, 이 지표에서 한국은 세계 6위로 평가되고 있다(<그림. 9>). 개발, 인프라, 정책 부문에서는 우수한 반면 인력, 연구, 상업화 부문에서 상대적으로 경쟁력이 낮음을 보여주고 있다. 그러나 필자의 관점에서 생각하는 우리나라의 장·단점과 상이한 부분이 있어서 이 지표가 얼마나 신뢰할 만한지는 확인이 필요해 보인다.

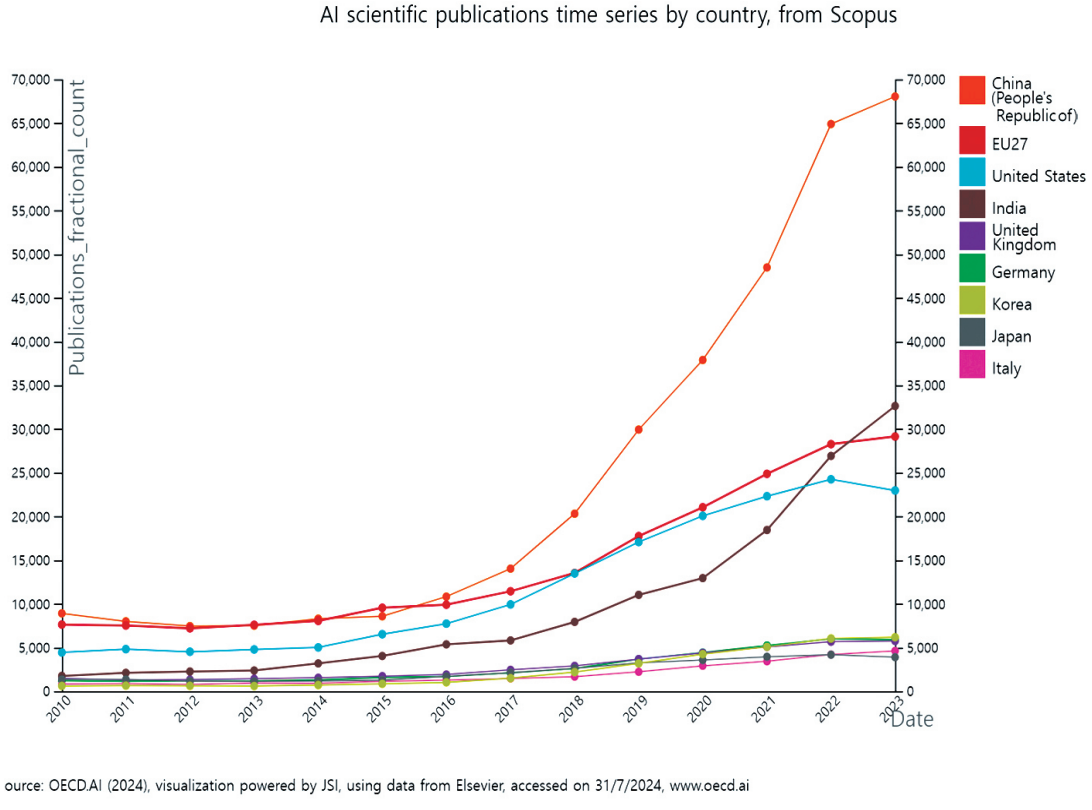
그림. 9 국가별 AI 경쟁력 순위

	Overall	Talent	Infrastructure	Operating Environment	Research	Development	Government Strategy	Commercial	Scale	Intensity
United States	1	1	1	28	1	1	8	1	1	5
China	2	20	2	3	2	2	3	2	2	21
Singapore	3	4	3	22	3	5	16	4	10	1
United Kingdom	4	5	24	40	5	8	10	5	4	10
Canada	5	6	23	8	7	11	5	7	7	7
South Korea	6	12	7	11	12	3	6	18	8	6
Israel	7	7	28	23	11	7	47	3	17	2
Germany	8	3	12	13	8	9	2	11	3	15
Switzerland	9	9	13	30	4	4	56	9	16	3
Finland	10	13	8	4	9	14	15	12	13	4
Netherlands	11	8	16	15	10	13	28	20	11	8
Japan	12	11	5	10	20	6	18	23	6	25
France	13	10	11	25	15	18	13	10	9	20

출처: Tortois(2023).

세부 부문의 경쟁력을 좀 더 구체적으로 분석해 보면 다음과 같다. 먼저 학술 부문의 지표를 살펴보면 OECD에서 공개한 2023년 국가별 AI 관련 논문 발표 수를 기준으로 우리나라는 세계 7위(EU를 국가 단위로 계산할 경우)에 해당한다(<그림. 10>). 총 논문 수 7,887편으로 중국의 75,206편, 미국의 32,936편에 비해 약 1/9, 1/4 수준이지만, 인구를 고려해 환산한다면 세계 최고 수준으로 볼 수 있다. 그러나 질적으로 우수하고 영향력이 높은 논문들로 한정할 경우 세계 8위로 순위가 다소 내려앉는다.

그림. 10 국가별 AI 관련 논문 발표 수

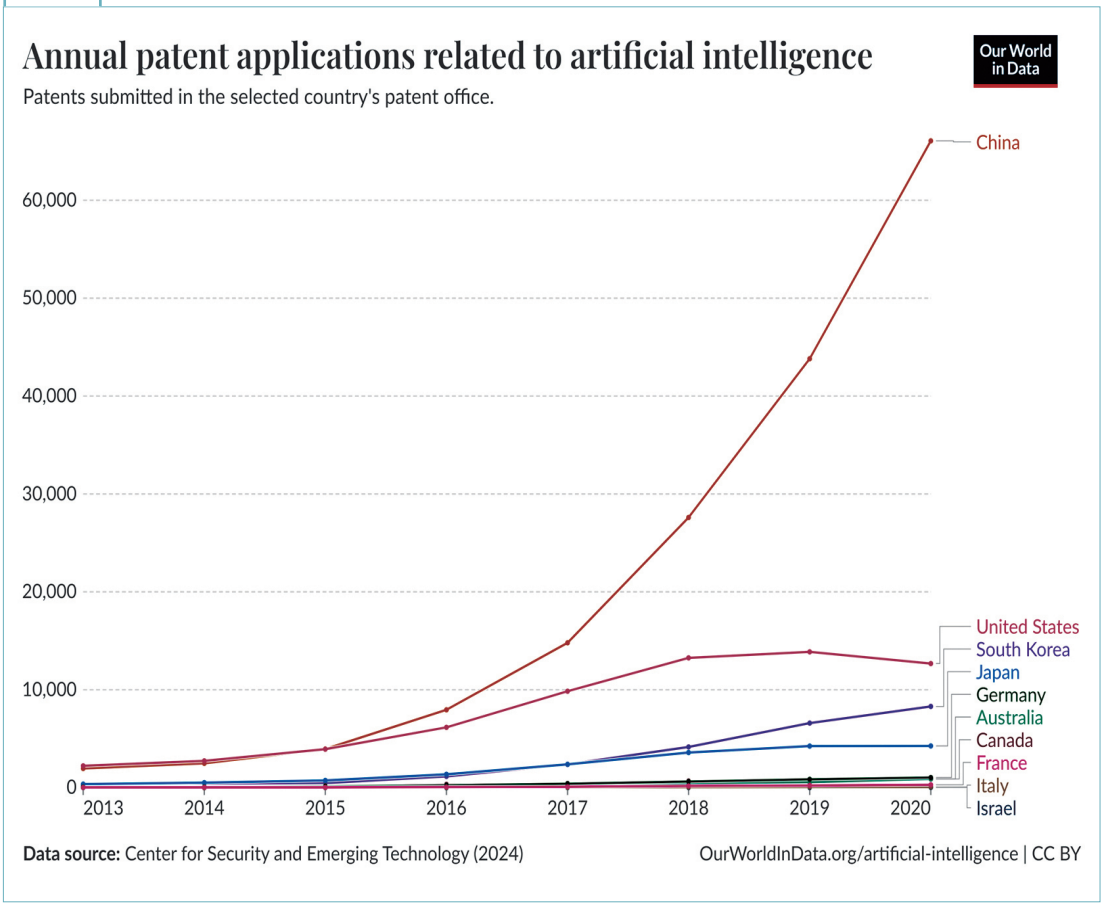


출처: OECD.AI(2024).

AI 분야 최고 학술대회인 CVPR(Conference on Computer Vision and Pattern Recognition)과 ICCV(International Conference on Computer Vision)의 경우를 보면, 최근 국가별 발표논문 수를 기준으로 우리나라는 중국, 미국에 이어 세계 3위를 꾸준히 기록하고 있다. 참가자 수에 있어서도 세계 3~4위를 기록하고 있는데, 이것은 우리가 AI 분야 중 특히 시각지능(vision) 분야에 상대적으로 높은 학술적 경쟁력과 관심을 갖고 있음을 보여준다. 물론 논문인용 수나 영향력지수 등 질적 측면에서는 아직 성장이 필요한 단계지만, 일본이나 유럽 국가들에 비해 단기간에 급격한 도약을 이루며 국제적으로 존재감을 확대하고 있는 것은 분명하다.

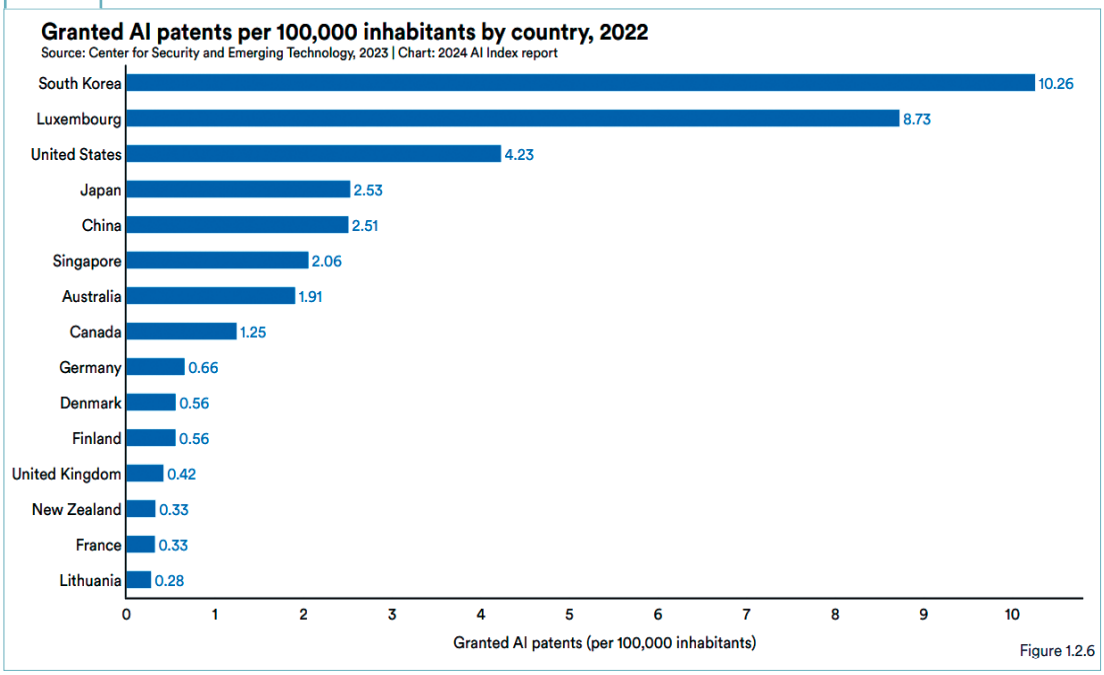
AI 특허 부문에서는 <그림. 11>, <그림. 12>에서 보듯이 특허등록 숫자상 우리나라가 중국, 미국에 이어 세 번째로 많고, 인구 100만 명당으로 보면 평균 10건이 넘어 세계 최고 수준이다. 또한 증가율 면에서도 중국과 미국을 압도하고 있음을 알 수 있다(<그림. 13>).

그림. 11 국가별 특허 수



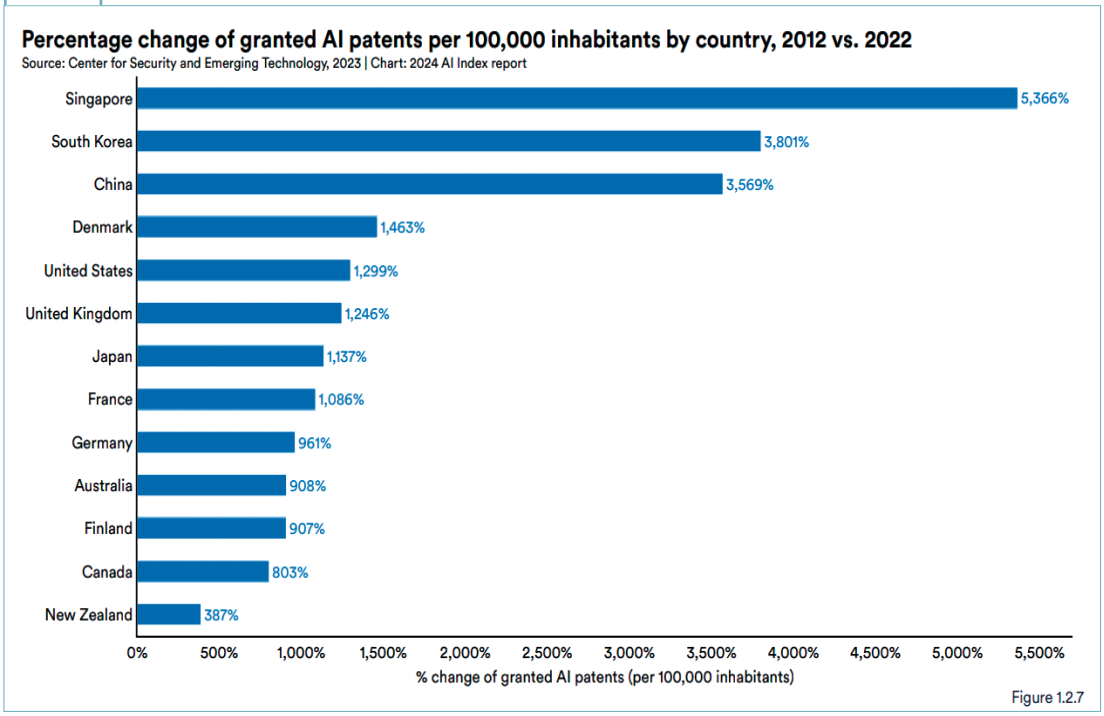
출처: Our World in Data(2023).

그림. 12 인구 10만 명당 특허등록 수



출처: Stanford University(2024).

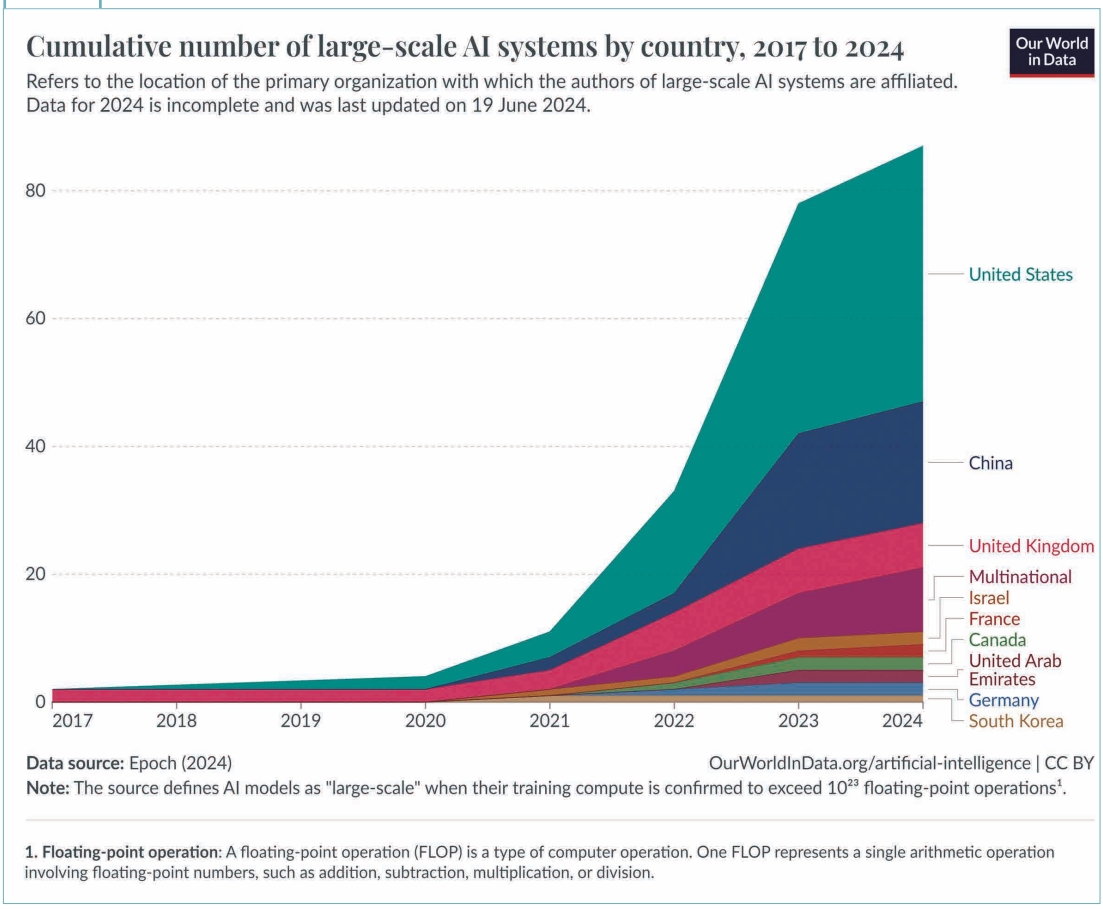
그림. 13 인구 10만 명당 특허등록 증가율



출처: Stanford University(2024).

국가별 거대 AI 모델 보유 지표를 살펴보면 2024년 기준 미국 40개, 중국 19개, 한국 1개로 상당한 차이를 보이는 것을 알 수 있다(<그림. 14>). 실제로 국내에 1~2개가량이 더 있다고 가정하더라도 상대적인 격차가 여전히 크다는 것은 분명해 보인다.

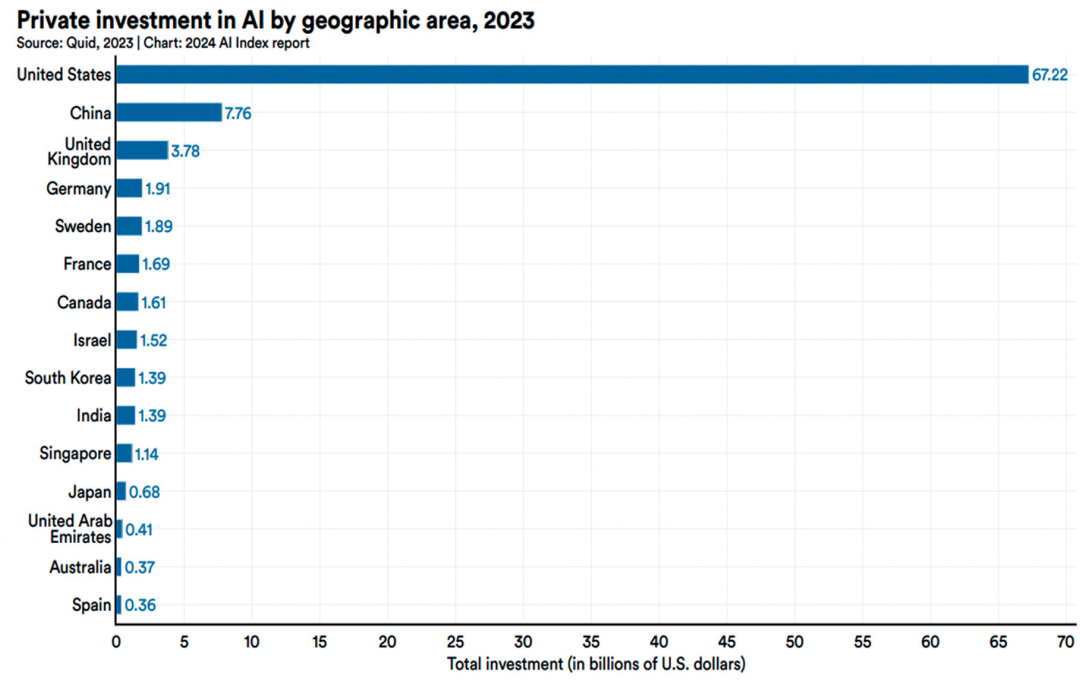
그림. 14 국가별 대규모 AI 시스템 보유 수



출처: Our World in Data(2024b).

AI 분야에 대한 재정적 투자 규모에 있어서는 2023년 기준 우리나라는 민간부문에서 55억 달러가 투자되었고 이것은 미국의 약 1/50, 중국의 1/6 정도에 해당한다(<그림. 15>). 또한 정부와 공공부문의 투자에서도 한국은 5년간 누적 약 100억 달러로 미국의 1/32, 중국의 1/13 정도에 불과함을 알 수 있다(<그림. 16>).

그림. 15 2023년 국가별 AI 민간투자 규모



출처: Stanford University(2024).

그림. 16 국가별 정부 AI 투자 규모

A breakdown of AI government investment statistics by country

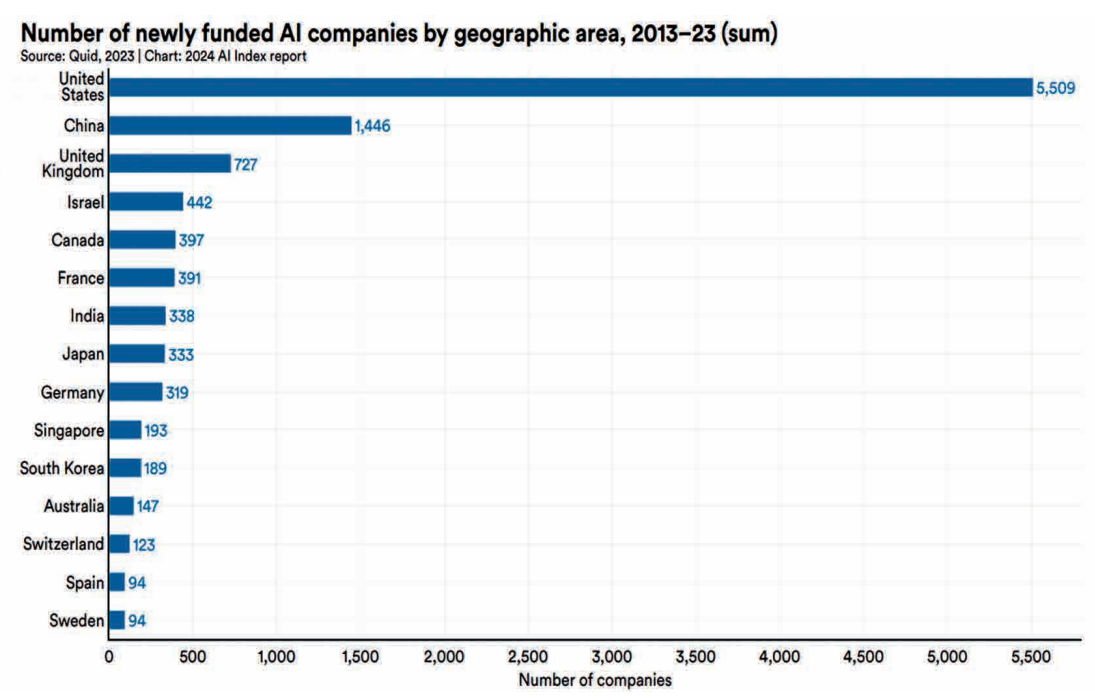
Country	Sum of investment in last 5 years (millions USD)	Investment in past 5 years (per thousand $ GDP)	Investment in past 5 years, as a percentage of USA investment (per thousand $ GDP)
Singapore	7,005	15.01	116.30
Sweden	8,281	14.13	109.53
The U.S.	328,548	12.90	100.00
Estonia	415	10.89	84.44
U.K.	25,541	8.32	64.46
China	132,665	7.39	57.24
South Korea	10,348	6.21	48.16
Canada	12,457	5.82	45.12
India	16,147	4.77	36.97
Switzerland	3,239	4.01	31.08

(Source: AIPRM via OECD and World Bank)

출처: AI Statistics(2024).

또 다른 지표로서 지난 10년 동안 신규 등장한 AI 관련 스타트업 수를 보면 우리나라는 총 189개로 세계 11위에 위치해 있고 이는 일본이나 싱가포르보다도 적은 수치이다(<그림. 17>). 이것은 국내 스타트업 생태계가 아직 충분히 활성화되지 않았음을 의미하며, 향후 우리나라 AI 산업을 이끌 혁신 동력이 취약하다는 것을 시사한다.

그림. 17 2013-2023년 신생 AI 회사 수



출처: Stanford University(2024).

여러 개별적 데이터들을 종합적으로 고려해 볼 때 현재 우리나라의 AI 국제경쟁력은 세계 10위권 안에 드는 것은 분명해 보인다. 그러나 최선두 그룹인 미국과 중국에 비하면 그 차이가 매우 크다는 것을 알 수 있다. 분야별로는 연구 및 특허 부문에서 상대적으로 강점을 가진 것으로 평가되나 기업경쟁력 특히 스타트업 부문이 열세이고, 정부 투자 및 인프라 부문에서도 상대적으로 열악한 상황인 것으로 분석된다. 따라서 향후 전략적 차원에서 이러한 배경의 정책적 고려가 뒷받침되어야 할 것으로 본다.

라 국내 산업 발전을 위한 AI 전략

AI 혁명에 의한 파고는 조만간 우리 산업은 물론 사회 전반에 지대한 영향을 미칠 것이다. 이것은 AI가 우리의 생존을 위한 미래 전략에 있어 핵심적인 요소로 작용한다는 의미이므로, 우리에게 AI는 또 하나의 기회이자 위기이다. 선진국 반열에 오른 오늘날 대한민국의 위상은 과거 반도체, 디지털, 그리고 인터넷/모바일 시대로 산업의 패러다임이 바뀌는 과정에서 선제적 대응과 과감한 투자를 통해 후발주자에서 선도자로 자리바꿈한 결과이다. AI 대변혁의 시기에 우리가 어떠한 전략과 실행 방안으로 이에 대응해야 하는지를 냉철하게 고민해야 하는 이유도 여기에 있다. 더욱이 AI 기술과 산업의 발전 속도를 고려할 때 전략적 판단과 실행이 매우 빠르게 이루어져야 한다는 점은 중요하고도 명확하다. AI 산업에서의 경쟁력 확보를 위한 방안은 학계 및 산업계 전략, 인재, 정책, 그리고 원천기술 및 응용기술 부문 등 여러 가지 관점에서 생각할 수 있다.

먼저 학계를 들여다보면 그동안 논문, 특허 등 연구의 양적 성장은 괄목할 만하다. 일부 분야는 세계적인 수준에 도달하는 등 놀라운 성과를 내고 있는데, 반면 전반적으로는 분야별 편차가 크다는 문제를 안고 있다. 또한, 원천기술뿐 아니라 산업과 응용 분야에 연계된 연구개발이 상대적으로 부족하다는 것이 약점이 될 수 있다. AI 분야 최고 학술대회에서 발표되는 논문들이 전 세계적으로 연간 약 10,000편이 넘는데도 불구하고 이 중 약 90%가량은 인용되지 않는다는 점을 감안하면 이제부터는 단지 실적을 위한 논문 양산이 아니라 영향력이 큰 논문 그리고 실용성 있는 기술을 추구하는 방향으로 노력을 기울여야 할 시점이다. 대학의 AI 인재 양성과 연구개발 경쟁력은 충분한 재정적 지원 및 인프라 확충이 없으면 한계에 부딪히게 된다. 현재 우리 정부에서 대학에 지원하는 AI 관련 예산 규모는 경쟁국들에 비해 매우 적은 상황이고 본격적인 경쟁을 위해서는 과감한 재정 투자가 필요한 시점이다. 필요하다면 예산의 선택과 집중을 통해서라도 단기간에 선도적 대학을 집중 육성하는 방안을 고려할 수 있을 것이다. 그리고 현재 대학 등 학계 전반에 서버 및 전력 등 충분한 컴퓨팅 연구 인프라가 확보되지 않아 연구와 교육에 어려움을 겪고 있는 상황이 해결되지 않는다면 아킬레스건이 될 가능성이 높으므로, 산·학·관 간의 긴밀한 소통과 협력을 통해 대책을 강구해야 한다.

산업계에서의 AI 경쟁력은 투자 규모, 대규모 AI 모델 보유 수, 스타트업 수, AI 비즈니스 시장 규모 등의 데이터를 통해 봤을 때 아직 기대에 미친다고 보기는 이르다. 현재 네이버나 LG, 카카오 등이 자체적인 대규모 AI 시스템을 개발·운영하고 있으나 오픈AI, 구글 등이 천문학적인 자금과 물량을 투입하는 상황에서 직접적 경쟁을 지속하는 것이 현실적으로 가능할지, 또 승산이 있을지에 대한 전략적 판단이 중요한 지점이다. 그럼에도 불구하고, 독자적인 핵심 거대 AI 모델을 확보하는 문제는 그 파급력을 생각할 때 포기하기는 어려운 측면이 있다.

자체 거대모델 개발의 목적이 무엇인가? 국내 산업, 문화 및 정보 보호와 같은 주권 수호 차원의 방어적 소버린 AI 개발이 목표인가? 혹은 국제적 경쟁을 통한 세계 시장 확보와 같은 공격적 목표를 추구하는가? 목표의 설정에 따라서 이를 이루기 위한 전략이 달라질 수 있겠지만, 실질적으로는 두 가지 전략이 모두 필요하다. 특히 거대 언어모델이 개인용 비서와 같은 고도의 지능적 역할을 수행한다고 할 때, 해당 국가의 문화, 관습, 예술, 지식 그리고 삶의 방식 및 가치와 관련된 최신의 정보가 핵심적인 역할을 할 것이므로 개별 지역 및 국가에 특화된 AI 모델이 경쟁력을 가질 가능성이 높다. 네이버 검색엔진이 구글 검색엔진보다 낫다고 할 수는 없으나 국내 검색포털 중 압도적 점유율을 행사하고 카카오톡이 트위터나 페이스북보다 국내시장에서 절대적 우위를 가지는 것과 같다. 따라서 이러한 소버린 전략을 위해 자체 개발된 AI 모델을 빠른 시간 안에 상품화하고 기존 서비스플랫폼에 탑재하여 국내 생태계를 형성할 필요가 있다.

기술적 관점에서는 마라톤이나 쇼트트랙에서 경기 초반 선두에 나서는 선수가 반드시 최종 우승을 하지는 않는 것처럼, 적절한 거리를 유지하면서 적은 에너지로 추격하다 결정적인 순간 속도를 높이는 전략을 구사하는 것도 효율적일 수 있다. 주목받지 못하던 작은 오픈AI가 절대 강자라고 여겨졌던 골리앗 구글을 챗GPT로 추월한 예에서 보듯이 기술적으로는 새로운 방식의 혁신적 돌파 기술이 언제든지 등장할 수 있고 가능성 또한 무궁무진하다. 향후, 소형 저전력 AI, 멀티모달 AI, 임바디드 AI, 자기학습 AI 등 새로운 혁신 기술들이 속속 태동할 것이다. 이러한 관점에서 기업과 대학, 대기업과 중소기업 간 실질적 협력 프로그램을 강화하고 공동전선을 형성해서 선제적으로 대응하는 것도 유효한 전략이 될 것으로 판단된다.

산업적으로 AI 응용 부문에서의 경쟁력 확보는 더욱 중요하다. AI로 인한 실제 산업적 부가가치는 대부분 다양한 응용 분야에서 발생할 것이기 때문이다. 현재의 AI 기술을 가지고도 제조, 의료, 금융, 모빌리티, 엔터테인먼트, 물류, 서비스, 국방 등 많은 분야에서 최소한 기존의 생산성을 향상시킬 수 있을 뿐만 아니라, 혁신적인 제품 및 서비스 개발, 새로운 산업 분야 창출 등도 가능하다. 중국은 이러한 다양한 AI 응용 분야에서 데이터와 경험을 축적하고 경쟁력을 확보하고 있다. 산업 분야로의 AI 기술 적용 및 확산을 위해서는 해당 도메인의 지식을 겸비한 충분한 수의 AI 응용 인재 확보가 전제되어야 하고, 동시에 창의적 비즈니스 모델을 가진 많은 스타트업을 육성해야 할 것이다.

마 맺음말

결론적으로 AI 발전은 대한민국 산업의 다양한 분야에 걸쳐 머지않아 큰 변화를 일으킬 것으로 예상된다. 이를 효과적으로 활용하기 위해서는 지속적인 인재 양성, 부문 간 협력, 정부의 정책적 지원, 글로벌 시장 진출, 스타트업 지원 등의 전략이 촘촘하게 강화되어야 하고 과감한 투자가 병행되어야 한다. 이러한 전략을 통해 우리나라는 AI 기술을 바탕으로 한 혁신적인 산업 생태계를 구축하고, AI에 의한 새로운 글로벌 질서의 재편에서 새로운 도약의 전기를 마련할 수 있을 것이다.

참고문헌

Grace, K., Stewart, H., Sandkühler, J. F., Thomas, S., Weinstein-Raun, B. & Brauner, J.(2024). Thousands of AI Authors on the Future of AI, arXiv:2401.02843.

OpenAI(2023). "GPT-4 Technical Report", arXiv:2303.08774.

Rosenblatt, F.(1957). "The Perceptron: A Perceiving and Recognizing Automaton", Report 85-460-1, Cornell Aeronautical Laboratory.

AI Statistics(2024). IPRM via OECD and World Bank, Retrieved July 31, 2024 from https://www.aiprm.com/ai-statistics.

Atkinson, D., Barnett, M., Roldán, E., Cottier, B. & Besiroglu, T.(2024). "Direct Approach Interactive Model", Epoch AI, Retrieved July 31, 2024 from https://epochai.org/blog/direct-approach-interactive-model.

Cottier, B., Rahman, R., Fattorini, L., Maslej, N. & Owen, D.(2024). "The Rising Costs of Training Frontier AI Models", arXivarXiv:2405.21015, Retrieved July 31, 2024 from https://arxiv.org/abs/2405.21015.

Epoch(2024). Data on Notable AI Models, Published online at epochai.org, Retrieved July 31, 2024 from https://epochai.org/data/notable-ai-models.

OECD.AI(2024). Visualization Powered by JSI, using data from Elsevier, Retrieved July 31, 2024 from https://oecd.ai/en/data?selectedArea=ai-research&selectedVisualization=scientific-publications-time-series-by-country-2.

Our World in Data(2023). "Annual Patent Applications related to Artificial Intelligence", in Giattino, C., Mathieu, E., Samborska, V. & Roser, M.(eds.), Artificial Intelligence, Data adapted from Center for Security and Emerging Technology, Retrieved July 31, 2024 from https://ourworldindata.org/grapher/artificial-intelligence-patents-submitted.

Our World in Data(2024a). "Datapoints used to Train Notable Artificial Intelligence Systems" [dataset], Epoch, "Parameter, Compute and Data Trends in Machine Learning" [original data], Retrieved July 31, 2024 from https://ourworldindata.org/grapher/artificial-intelligence-number-training-datapoints.

Our World in Data(2024b). "Cumulative Number of Large-scale AI Systems by Country" [dataset], Epoch, "Tracking Compute-Intensive AI Models" [original data], Retrieved July 31, 2024 from https://ourworldindata.org/grapher/cumulative-number-of-large-scale-ai-systems-by-country.

Sevilla, j. & Roldan, E.(2024). Training Compute of Frontier AI Models Grows by 4-5x per Year, Published online at epochai.org, Retrieved July 31, 2024 from https://epochai.org/blog/training-compute-of-frontier-ai-models-grows-by-4-5x-per-year.

Stanford University(2024). AI Index Report 2024, AI Index Annual Report, Retrieved July 31, 2024 from https://aiindex.stanford.edu/report.

Tortois(2023). The Global AI Index, Global AI Index, Retrieved July 31, 2024 from https://www.tortoisemedia.com/intelligence/global-ai/#rankings.

KAST

4

생성 AI 시대 산업과 사회 대전환

하정우
네이버클라우드 AI이노베이션센터장

가 들어가는 말

2022년 11월 챗GPT가 첫 공개된 이후 전 세계는 생성 인공지능(AI)이 불러온 큰 변화의 바람에 휩싸여 있다. 별다른 추가 학습 없이 사용자의 명령에 상당히 높은 수준의 글과 그림을 만들어 내고 이런 능력이 일상생활과 업무에 적용되면서 생산성 향상을 이끌고 있다. 2023년 하반기부터 생성 AI의 멀티모달 능력이 강화됨에 따라 사용자에게 더 편리하면서도 강력한 기능을 제공하게 되는 등 영화 「Her」의 사만다가 실현될 수 있는 시대가 가까워지고 있다. 이렇게 생성 AI 기술이 일상과 산업 전반에 영향을 끼치게 됨에 따라 AI 경쟁은 기업 간 경쟁을 넘어 국가 간 경쟁의 양상으로 확대되고 있으며 AI 안전과 책임성 문제도 함께 논의되기 시작했다.

본 장에서는 생성 AI 기술을 중심으로 국가별 동향과 2024년 새롭게 떠오르고 있는 생성 AI 기술 주요 트렌드를 소개하고자 한다. 그리고 생성 AI가 다양한 산업에서 이룰 수 있는 생산성 향상 사례를 언급하며 생성 AI 기반의 AI 에이전트가 가져올 일상의 변화를 제시하겠다.

나 2024년 생성 AI 기술 동향 및 전망

1) 국가 및 주요 기업별 생성 AI 기술 동향

미국

미국은 지난 70여 년간 AI 분야에서 압도적인 기술을 기반으로 하는 산업 경쟁력을 보여주고 있으며 최근의 생성 AI 기술 흐름 또한 실리콘 밸리 기업들이 이끌어 가고 있다. 오늘날 생성 AI 특히 대형언어모델(LLM)의 기본 모델 구조라 할 수 있는 트랜스포머(Vaswani et al., 2017)는 2017년 5월 구글에서 발표했고 1년 반 후 자연어처리(Natural Language Processing, NLP) 분야에서 대량 데이터 사전학습 후 간단한 fine-tuning으로 대부분의 공개 벤치마크에서 기존 기록을 깬 버트(Bidirectional Encoder Representations from Transformers, BERT; Devlin et al., 2019) 또한 구글에서 2018년 10월에 발표한 기술이며, GPT-3(Brown et al.,

2020)와 챗GPT, GPT-4, 달리(DALL-E; Betker et al., 2023) 시리즈를 발표한 오픈AI도 미국 기업이다. 그 외에 오픈소스 LLM 생태계를 이끌어가는 LLaMA 시리즈는 페이스북과 인스타그램을 운영하는 메타에서 주도하고 있다. AI 학습과 운영을 위한 컴퓨팅 인프라의 핵심인 AI 가속기 시장의 90% 이상을 GPU로 독점하고 있는 엔비디아 또한 미국 기업이다.

학계에서도 스탠포드대학교, MIT, 워싱턴대학교, 카네기멜론대학교, UC Berkeley, 조지아 공대, UIUC 등이 미국의 기업계와 강력한 협업을 통해 AI의 혁신을 만들어 가고 있다. 나아가 위 학교의 대학원생들이 테크 기업들의 인턴으로 합류하여 기업의 연구자들과 함께 혁신적인 연구 성과를 이루고 있다.

중국

지난 1년간 Nature Index에 따르면 중국이 전반적인 과학기술 분야에서도 미국을 이미 추월하여 세계 1위를 기록하고 있으며 세계 최고 연구기관 Top-10에서도 Chinese Academy of Sciences가 1위를 차지한 것을 포함해 7개 기관이 Top-10 리스트에 포함되어 있다(Nature index, 2024).

AI 분야에서도 중국은 미국과 더불어 AI 2강 체제를 구축하고 있다. Tortoise media의 Global AI Index 2023에 따르면 미국의 AI 역량을 100점으로 볼 때 60점대 정도로 격차가 있는 것으로 보이지만 일부 영역 특히 국가체제 유지에 필요한 안면인식, 행동인식, 생체인식과 같은 컴퓨터 비전 기반의 인식 기술은 이미 미국의 수준을 넘어섰다는 평가를 받고 있다(Tortois, 2024).

생성 AI 특히 LLM에서는 중국 정부 체제와 사상에 부합하는 이해 및 추론, 콘텐츠 생성 능력을 보유하도록 강하게 제한한 이유로 실리콘 밸리 기업들에 비해 상대적으로 성장 속도가 느렸다. 하지만 최근 알리바바에서 발표한 Qwen2(Yang et al., 2024) 등은 상당한 경쟁력을 보여주고 있으며 이를 오픈소스로 전 세계에 배포하고 있다. Text-to-video에서도 Kuaishou가 공개한 Kling은 오픈AI Sora 이상의 경쟁력을 보유한 것으로 인식되고 있는데, 생성 AI 분야에서도 미국과 중국의 격차가 상당히 많이 줄어든 것으로 보인다.

최근 사우스차이나 모닝포스트는 중국 정부가 시진핑 주석의 어록과 사상을 대답할 수 있는 언어모델인 시진핑 GPT(Xi-GPT)를 준비 중이라 보도했다(Zhuang, 2024). 해당 거대언어모델의 적극 오픈소스 정책에 대해 중국 정부는 체제 확산의 방법으로 생성 AI를 활용하고 있는 것으로 의심받고 있다.

EU

EU는 강력한 AI 연구역량에도 불구하고 미국이나 중국에 비해 상대적으로 생성 AI 분야에서 발전이 더딘 편이다. 자체 플랫폼이 사실상 존재하기 않기 때문에 유럽 각국의 언어를 기반으로 하는 양질의 사전훈련용 학습데이터 확보도, 생성 AI를 활용할 혁신적인 서비스를 운영하기에도 어려움이 있다. 이러한 상황에서 강력한 규제 성향을 띠는 AI 법안이 올해 초에 통과되었다.

그러나 EU 회원국 중 일부는 강력한 정부 지원과 투자를 통해 자국 기업 중심으로 생성 AI 경쟁력을 확보하기 시작했는데, 대표적인 사례가 프랑스의 Mistral과 독일의 Aleph Alpha이다. 이들은 실리콘 밸리 빅테크 기업들에 준하는 대형언어모델을 오픈소스 형태로 공개하면서 그 기술력을 인정받고 대규모 투자를 얻어 이미 유니콘 대열에 합류해 있다. 이탈리아 또한 자체 생성 AI 경쟁력 확보를 위해 정부 차원에서 투자를 천명했다. 이 때문에 12월 6일에 통과되었어야 할 AI 법안이 프랑스, 독일, 이탈리아의 반대로 한차례 연기되는 등 격렬한 논의 끝에, 법안에 생성 AI에 대한 조항과 오픈소스 중심 샌드박스 활용을 통한 생성 AI 산업 활성화 부분이 반영된 것으로 알려졌다. 그러면서 규제 대상 생성 AI의 컴퓨팅 연산량 기준을 미국 행정명령보다 다소 적은 수치로 잡아 오픈소스 정책 및 강력한 처벌조항과 함께 미국 기업의 활동은 억제하고 역내 기업을 지원하는 형태로 보호정책을 취하고 있다.

일본 및 아시아

일본 정부는 디지털청을 별도로 설립하고 오픈AI와 손을 잡아 공공 분야 생성 AI 도입을 진행하고 있다. 이에 발맞추어 오픈AI는 일본지사를 설립하고 일본어에 특화된 GPT인 GPT Japan을 지난 4월에 공개했다. 그러나 이와 별개로 자국 생성 AI 기술 확보를 위해 일본 정부는 자국 기업인 소프트뱅크에 지난 2년간 약 4,500억 원에 달하는 보조금을 지급하며 생성 AI 학습에 필요한 슈퍼컴퓨터 도입 관련 투자에 활용토록 했다. 이를 통해 궁극적으로 실리콘 밸리 기업들에 대한 기술적 종속을 막고 생성 AI 강대국으로 나아가기 위한 정책을 과감하게 추진 중이다.

인도네시아, 태국, 베트남과 같은 국가들도 미국, 중국에 생성 AI 기술이 종속되는 것을 막기 위해 자국 문화를 정확히 이해하는 소버린 AI 확보를 위한 노력을 진행하고 있다. 생성 AI 사전훈련은 기술적인 난이도가 있기 때문에 자국 혹은 외국 대학들과 협업을 통해 오픈소스 언어모델에 자국어 데이터를 학습하는 형태로 먼저 진행하며, 신뢰 가능한 글로벌 파트너를 찾아 자국 AI 기술 산업 생태계 강화를 모색 중이다.

대한민국

우리나라는 대형언어모델 분야에서 미국, 중국에 이어 빠른 속도로 자체 역량의 AI 모델을 개발하고 산업 생태계를 구축한 국가이다. 네이버가 2021년 5월 매개변수 1천억 개 이상 기준 전 세계 3번째로 LLM인 하이퍼클로바(Kim et al., 2021)를 공개했고 뒤이어 LG AI 연구원, 카카오, KT, SKT, 엔씨소프트 등이 자체 기술력으로 LLM을 출시했다.

네이버는 하이퍼클로바를 더욱 개선한 한국어 중심의 다국어 LLM인 하이퍼클로바X를 출시한 이래로 챗GPT와 유사한 대화형 서비스인 클로바X를 작년 8월, 검색 형태의 서비스 Cue를 같은 해 10월에 공개했다. 그리고 기업들이 외부에서 클라우드 기반의 API 형태로 활용할 수 있도록 클로바 스튜디오를 2022년 2월부터 제공하고 있으며 현재 2,200여 개 기업들이 활용 중이다. LG AI 연구원은 2021년 말 엑

사원과 더불어 2023년 7월 엑사원2.0을 공개하고 주로 LG 그룹 관계사들의 AI 전환에 활용하고 있다.

카카오, SKT, KT와 같은 기업들은 자체 LLM 외에도 글로벌 기업들의 LLM과 sLM까지 다양한 모델을 동시에 활용하는 멀티 LLM 전략을 채택하고 다양한 비즈니스를 만들어가고 있다.

2) 2024년 생성 AI 주요 기술 동향 및 전망

멀티모달(multimodal) AI

멀티모달은 데이터를 표현하는 방법이나 형태가 여러 가지인 경우를 일컬으며, 일반적으로 멀티모달 AI는 글뿐 아니라 사진, 음성, 영상, 시계열 수치 등 다양한 형태의 데이터를 동시에 처리, 학습하고 이해, 추론, 생성할 수 있는 AI를 의미한다. 챗GPT의 경우에도 초기에는 텍스트의 형태로만 입력 및 콘텐츠 생성이 가능했다. 그러나 2023년에 들어서는 사진과 대화가 가능한 GPT-4가 챗GPT에 탑재되었고 구글 딥마인드의 Gemini나 Anthropic의 Claude 3 등이 멀티모달 대화가 가능한 AI를 발표했다. 이어서, 글을 입력하면 그 내용을 그림으로 그려주는 AI인 DALL-E2와 3가 연이어 챗GPT와 연동되며 멀티모달 AI 기술이 연구실 수준을 넘어 고객의 지갑을 열 수 있는 서비스 수준까지 발전했다.

그림. 18 오픈AI Sora가 생성한 비디오와 프롬프트 예시

[프롬프트]

한 세련된 여성이 따뜻하게 빛나는 네온과
생동감 넘치는 도시 간판으로 가득한 도쿄 거리를
걷고 있습니다. 그녀는 검은색 가죽 재킷,
긴 빨간색 드레스, 검은색 부츠를 착용하고,
검은색 지갑을 들고 있습니다.
그녀는 선글라스를 쓰고 빨간 립스틱을 발랐습니다.
그녀는 자신감 있고 자연스럽게 걷습니다.
길은 축축하고 반사되어 화려한 조명이 거울 효과를
만들어 냅니다. 많은 보행자가 걸어갑니다.

출처: OpenAI Sora(2024).

2024년에는 더욱 강력한 멀티모달 AI가 등장했는데 첫 번째는 챗GPT-4o(omni)이다. 기존 챗GPT도 이미지 대화나 음성인식이 가능했지만, 음성인식의 경우 단일 모델이 아닌 오픈AI의 음성인식 모델인 Whisper를 사용하여 입력된 음성을 글로 변환하고 이를 챗GPT-4에 입력하는 구조였다. 그러나 챗GPT-4o는 하나의 모델에 글, 음성, 이미지, 영상을 모두 학습하고 생성하는 형태로 만들어져 목소리의 감정마

저도 자연스럽게 함께 모델링되고 스트리밍 형태로 처리하는 것이 가능해졌다. 이에 따라 영화 Her에 나오는 사만다가 드디어 실현되는 시대가 도래했다는 평을 듣기도 했다.

대화형 생성 AI 외에도 입력된 글의 내용을 영상으로 생성해 내는 text-to-video 기술이 API 형태로 제공됨으로써 새로운 AI 서비스로서 자리 잡기 시작했다. 가장 먼저 등장한 것은 오픈AI가 2월에 발표한 Sora이다. Sora는 대량의 비디오 데이터를 트랜스포머 기반의 확산모델(Diffusion model)인 DiT 모델을 활용하여 학습한 모델로 입력된 글의 내용을 반영하는 60초 길이의 고해상도 영상을 생성할 수 있다. 과거에도 text-to-video 기술이 없지는 않았지만 일반인이 바로 활용할 수 있는 수준의 AI는 Sora가 최초이다. 오픈AI Sora 이후 유사한 수준의 영상 생성 AI가 쏟아지기 시작했는데 대표적으로 runway의 Gen3-α, Luma AI 등이 있으며 중국 기업인 Kuaishou의 Kling도 뛰어난 성능을 보유한 것으로 알려져 있다. 특히 오픈AI와 달리 나머지 3개 기업은 이미 API나 인터넷을 통해 유료 혹은 무료로 직접 비디오를 제작할 수 있는 기술을 확보했다.

Text-to-video 외에 Text-to-music도 2024년 초에 선풍적인 인기를 끌었는데 대표적인 앱이 suno이다. suno 또한 웹을 통해 글을 입력하면 글의 내용을 반영한 가사와 멜로디로 표현된 노래를 누구나 다운받을 수 있다.

온디바이스(on-device) AI

2023년 12월 초 구글 딥마인드에서 Gemini를 공개할 때 다른 모델 대비 눈에 띄는 부분이 바로 Nano 버전이었는데, Nano-1이 18억 개, Nono-2가 32억 개의 매개변수를 가진 Gemini(Gemini Team Google, 2023)였다. 이렇게 작은 모델을 만든 이유는 바로 구글의 스마트폰인 픽셀에 활용하기 위함이라며 온디바이스 AI 혹은 AI 스마트폰의 시대를 선언한 것이다. 물론 18억 개 매개변수 모델조차도 픽셀 8 스마트폰의 하드웨어 사양과 안드로이드 운영체제를 고려하면 아무리 양자화 압축을 한다 해도 그대로 스마트폰에 탑재하는 것은 과도한 발열과 배터리 소모 등으로 인해 기술적 어려움이 예상되는 부분이었고, 결국 실제로 픽셀 8에 탑재하지는 않는다는 뉴스가 연초에 전해졌다.

오히려 AI 스마트폰 시장을 선점한 것은 삼성전자의 갤럭시 S24였다. 물론 Gemini와 같은 강력한 성능을 가진 단일 범용 파운데이션 모델을 사용한 것은 아닌 것으로 추정되나 circle-to-search나 실시간 통화번역 기능, 요약 기능 등 스마트폰 사용자들에게 매우 편리한 AI 기능을 탑재함으로써 AI 스마트폰 시장에서 앞서가는 데 성공한 것으로 보인다.

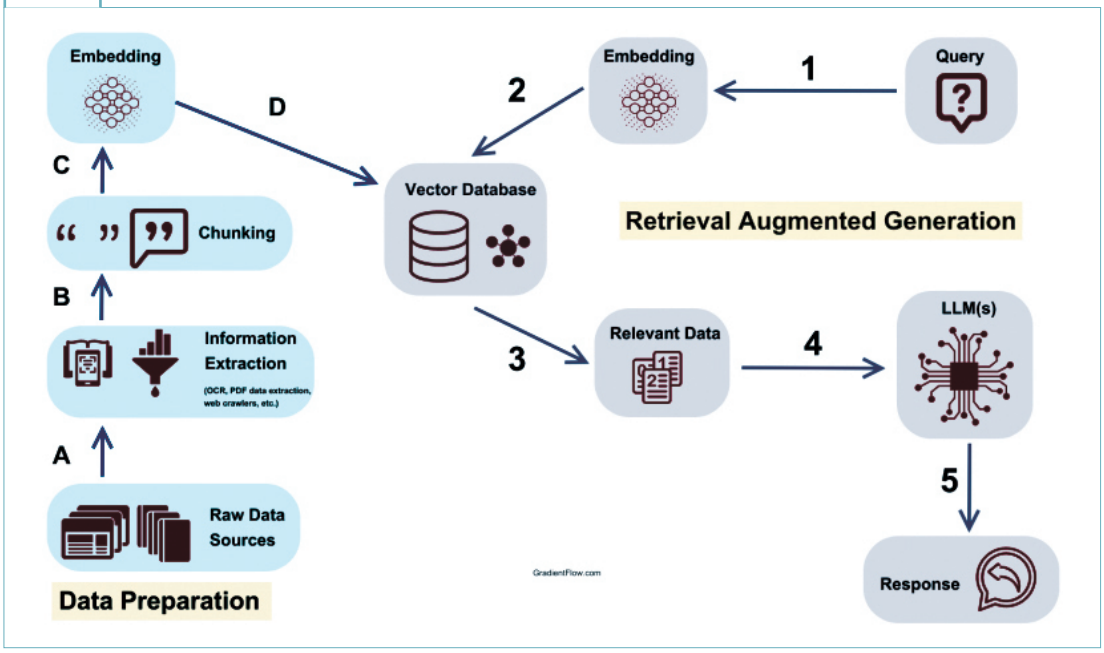
메타의 LLaMA나 Mistral LM, 마이크로소프트의 Phi(Abdin et al., 2024), 그리고 구글의 Gemma(Gemma Team, 2024)와 같이 상대적으로 작은 규모의 언어모델(sLM: small Language Model) 성능이 비약적으로 개선됨에 따라 스마트폰뿐만 아니라 PC나 자동차, 가전, 로봇 등 다양한 디바이스에 탑재

되는 사례가 늘어나고 있다. 특히 양자화(Quantization)를 포함한 AI 모델 압축 기술이 크게 개선됨에 따라 인텔, 마이크로소프트, 애플 등에서 양자화된 거대 언어모델을 AI 반도체 형태로 만들어 PC에 내장한 AI PC 제품들이 출시되기 시작했다(Intel, 2024).

환각(hallucination) 현상 해결책

환각 현상은 생성 AI, 특히 LLM이 문법이나 표현으로는 아주 그럴 듯하지만 사실과 다른 내용을 생성하는 경우를 말한다. 마케팅, 세일즈, 혹은 창작활동의 도구로 사용될 때는 문제가 되지 않으나 사용자 질의의 정확한 정보를 제공하는 챗봇과 같이 사실 여부가 중요한 응용 분야에서는 문제가 될 수 있다.

그림. 19 RAG를 활용한 LLM 응용 시스템 전제구조



출처: Lorica(2023).

환각 현상은 언어모델이 갖는 근본적인 문제이다. 데이터를 사전학습하는 과정에서 데이터가 데이터베이스처럼 그대로 저장되는 것이 아니라 모델 매개변수의 가중치 형태로 변환하여 저장되므로 이 과정에서 정보 손실이 발생할 수 있다. 또한 모델이 글을 쓰는 과정에서 다음 토큰을 선택할 때 글이 자연스러운 방향의 확률적 선택을 하고 이것이 사실을 보장하지 않을 수 있다. 특히 외부 공개가 덜 된 특수 분야에 대한 글을 쓸 때 이런 현상이 많이 발생한다.

이러한 LLM의 환각 현상 완화를 위해 전문가들이 다양한 전문 분야 지식과 상식 등 데이터를 가공해 LLM을 추가적으로 학습시키는 미세 학습(fine-tuning)을 수행하거나 사용자들의 피드백을 활용하여 언어

모델의 품질을 평가하는 보상 모델을 학습하는 인간 피드백 강화학습(Reinforcement Learning from Human Feedback, RLHF) 등이 많이 활용되어 왔다. 그러나 이것만으로는 실제 제품이나 서비스에 적용하기에는 여전히 한계가 있다 보니 최근에는 검색증강기법(Retrieval-Augmented Generation, RAG; Lewis et al., 2020)과 함께 사용되고 있다. RAG는 주로 LLM을 활용하여 금융상품 안내 상담 등과 같이 전문 분야에 특화된 정확한 정답을 요구하는 챗봇이나 질의응답 시스템을 구현할 때 사용된다. <그림. 19>와 같이, 먼저 정보가 포함된 관련 문서들로부터 정답에 해당될 정보들을 별도의 인공신경망 모델 활용을 통해 벡터 형태로 임베딩하고 이들을 벡터 데이터베이스에 저장한다. 그리고 사용자 질의가 들어오면 이를 바로 프롬프트로 LLM에 입력하는 것이 아니라, 벡터 데이터베이스를 만들 때 활용했던 임베딩 모델을 활용하여 사용자 질의를 벡터로 변환하고 이 질의 벡터와 가장 유사한 관련 정보를 검색해서 가져온다. 이렇게 검색해 온 정보를 활용하여 프롬프트를 만들어 LLM에 입력함으로써 보다 정확한 답변을 생성해낼 수 있다. 물론 LLM만 쓸 때보다 더 많은 추가 연산이 필요하므로 속도 최적화를 구현하는 것이 서비스 품질 향상을 위해 매우 중요하며 임베딩 모델의 정확도가 성능에 큰 영향을 미친다는 점도 고려해야 한다.

오픈소스 AI 생태계의 확산

2023년 메타에서 LLaMA를 공개하면서 오픈소스 sLM 생태계가 본격적으로 열리기 시작했다. 그 이전에도 BLOOM이나 Eleuther AI의 GPT-Neo, 국내에선 Polyglot과 같은 모델들이 존재했으나 성능에 대한 아쉬움이 있었다. 그러나 LLaMA는 파인튜닝을 통해 바로 활용 가능할 정도로 잘 사전학습된 LLM이었고, 70억 개부터 700억 개까지 다양한 크기의 매개변수를 제공했으며 70억 개짜리 모델을 기반으로 소수의 데이터에 대한 미세 학습을 통해 챗GPT-3.5 등 LLM과 비교해서도 경쟁력 있는 벤치마킹 결과들이 공개되면서 sLM이 주목받기 시작했다. 그러나 이후 벤치마크 데이터 과적합으로 벤치마크 순위표에 대한 신뢰성 문제가 수면 위로 떠오르게 된다.

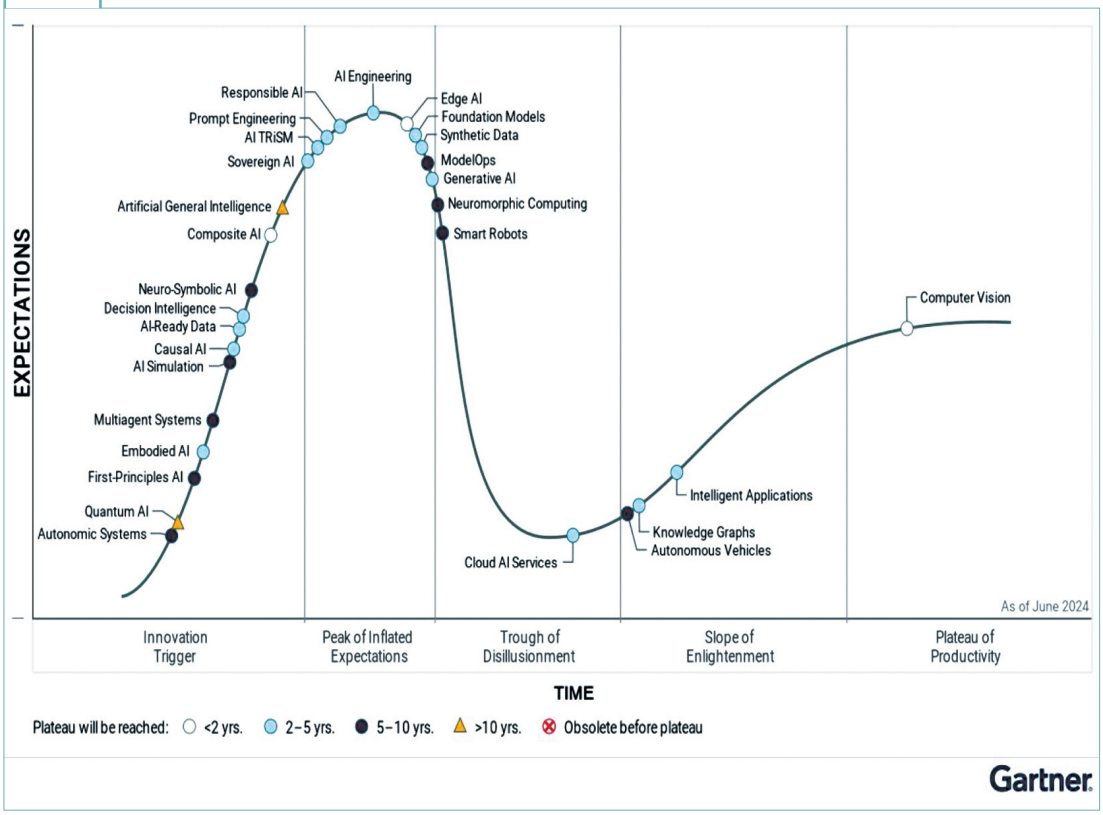
LLaMA1은 상업적 용도로 사용할 수 없어 학계를 중심으로 활용되었는데 LLaMA2부터는 라이센스 정책을 바꾸어 월간 활성 사용자 수억 명 급의 글로벌 규모 기업이 아닌 이상 상업적 용도로 활용이 가능해졌다. 전 세계 많은 스타트업들과 다양한 산업의 기업들이 자체 보유 데이터나 서비스 구현에 필요한 데이터를 추가적으로 미세 학습하여 LLaMA를 도입하기 시작했다.

2024년에는 프랑스의 미스트랄 AI에서 LLaMA 구조를 기반으로 하여 자체 사전학습한 sLM인 Mistral 모델(Jiang et al., 2023)을 오픈소스로 공개했으며 이 모델을 여러 전문가모델로 활용한 mixture of expert 형태인 Mixtral도 함께 출시했다. 이에 따라 다수의 글로벌 기업들이 Mistral sLM도 함께 활용하고 있고 국내의 많은 스타트업들도 이 Mistral 모델을 기반으로 한국어 데이터를 추가 학습하여 제공하고 있다.

글을 그림으로 그려주는 text-to-image 모델 중에서는 Stability AI에서 공개한 Stable diffusion 모델

이 대표적이다. Stable diffusion 모델을 누구나 활용할 수 있게 됨에 따라 많은 스타트업들이 자체적으로 보유한 사진데이터를 추가 학습하여 새로운 응용 프로그램을 만들고 학계에서는 ControlNet(Zhang et al., 2023)과 같이 훨씬 더 정교하게 그림을 생성할 수 있는 기법들이 개발될 수 있었다.

그림. 20 가트너 그룹의 Hype Cycle for AI 2024



출처: Gartner(2024).

빅테크의 LLM이 API 형태로만 접근 가능한 것과 달리 오픈소스 sLM은 모델 파일이 공개되기 때문에 연구 및 활용 측면에서도 더 많은 시도를 해볼 수 있다. 이는 생성 AI 분야의 다양한 연구의 활성화를 가져옴과 동시에 더 많은 산업에 생성 AI가 빠르게 확산하는 데 기여하고 있다.

소버린 AI

각 국가는 언어뿐 아니라 고유의 문화, 역사, 사회 규율, 가치관을 갖고 있는데, 미국 기업이 만든 LLM은 추가 학습을 통해 전 세계 수십 개 언어로 글을 자연스럽게 쓰면서 언어장벽을 무너뜨린 것으로 보이나 사전훈련 데이터의 90% 이상을 미국의 인터넷 영어 문서로 구성하다 보니 철저하게 미국 중심의 가치관을 갖게 된다. 그러나, 미래 세대들이 생성 AI를 일상생활과 교육에 적극 활용하게 됨에 따라 이러한 미국 가치관 중심의 콘텐츠에 지속적으로 노출되는 것은 각국의 정체성과 문화 다양성 유지에 악영향을

끼칠 우려가 있다. 이와 달리 소버린 AI는 각 지역의 문화, 역사, 사회 규율을 더욱 정확하게 이해하는 AI이다. 그리하여 세계 각국은 자국의 문화를 정확하게 이해하고 콘텐츠를 생성할 수 있는 소버린 AI를 확보하기를 원하고 있다.

엔비디아의 CEO 젠슨 황 또한 2024년 World Government Summit에서 각 지역 문화를 정확히 이해하는 소버린 AI 중요성을 강조했다. 실제로, 가트너 그룹에서 공개한 Hype Cycle for AI 2024에서 <그림. 20>과 같이 처음으로 소버린 AI가 주목받는 키워드로 등장하기도 했다. 그러나 생성 AI를 자체 기술로 만들기 위해서는 상당한 경험이 있는 인재와 기술, 투자 그리고 플랫폼이 필요하기 때문에 각 국가들은 신뢰 가능한 기술력을 가진 글로벌 파트너와 협력을 원하고 있다. 이 때문에 우리나라 AI 기업들은 소버린 AI 확보를 원하는 국가들과 협업을 통해 전 세계로 진출할 수 있는 좋은 기회를 마련하고 있다.

다 생성 AI 도입을 통한 생산성 혁신

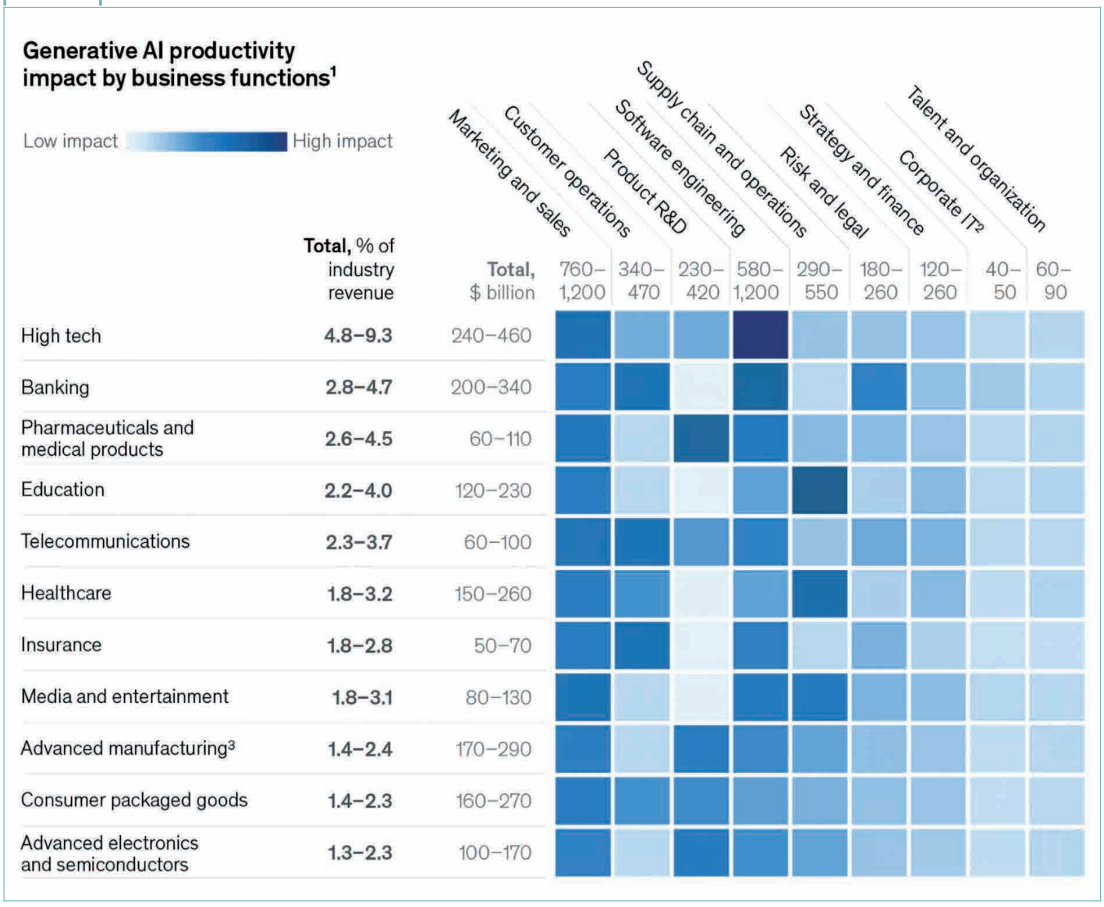
1) 생성 AI에 의한 생산성 혁신

생성 AI의 주요한 효과는 콘텐츠 생성 능력을 통한 생산성 혁신이다. McKinsey가 2024년 1월에 발표한 생성 AI 보고서에 따르면 생산성 향상 효과는 적게는 2조 6천억 달러 많게는 4조 4천억 달러에 이를 것으로 전망하고 있다. <그림. 21>과 같이 생성 AI는 모든 산업에서 마케팅 및 세일즈, 고객 관리, 소프트웨어 엔지니어링, 제품 R&D의 생산성을 극도로 향상시킬 수 있으며 IT 하이테크 산업은 물론 금융, 교육, 제약, 헬스케어, 미디어 콘텐츠, 유통, 제조를 망라하는 모든 산업 생산성 향상에 영향을 주는 것을 확인할 수 있다.

금융 분야는 매우 빠르게 생성 AI를 도입한 대표적인 산업 분야이다. Bloomberg는 자체적으로 학습한 금융 특화 LLM인 BloombergGPT(Wu et al., 2023)를 개발하여 내부 업무에 활용하고 있다. 2023년에 발표된 OECD 금융 분야 생성 AI 활용 보고서에 따르면 골드만삭스나 JP모건 등은 내부 개발자들을 위한 LLM 기반의 코딩 어시스턴트를 활용하여 소프트웨어 개발 운영 업무 생산성을 크게 향상시키고 있다(OECD, 2023). 특히 소프트웨어 개발자 채용이 상대적으로 쉽지 않은 전통 산업 분야에서 코딩 어시스턴트는 생산성 제고에 큰 기여를 하고 있다. 국내에서도 4대 은행에서 생성 AI를 도입한 사례들이 증가하고 있는데, 아직은 환각 현상 등의 문제로 대고객 서비스보다는 코딩 어시스턴트 등 내부 업무 생산성 향상을 주된 목적으로 하여 적용하고 있다. 미래에셋은 자체 앱을 통해 해외 뉴스를 간단하게 요약해서 대고객 서비스를 제공하고 있다.

법률 분야에서도 글로벌 로펌을 중심으로 생성 AI 도입이 빠르게 진행되고 있다. 특히 LexisNexis는 변호사 업무 자동화를 위한 생성 AI 도구인 Lexis + AI를 개발했다. 국내에서도 BHSN이나 인텔리콘, 엘박스와 같은 리걸테크 기업들이 생성 AI를 활용한 법률 업무 생산성 보조도구 제품을 출시했다. 또한 최근 국내 로펌인 대륙아주에서는 하이퍼클로바X를 활용한 24시간 일반인 법률상담 서비스 AI대륙아주를 출시해서 법률 서비스의 문턱을 낮추는 데 기여하고 있다.

그림. 21 각 산업 분야에서 기능별 생성 AI 도입에 따른 생산성 향상 효과



출처: McKinsey(2023).

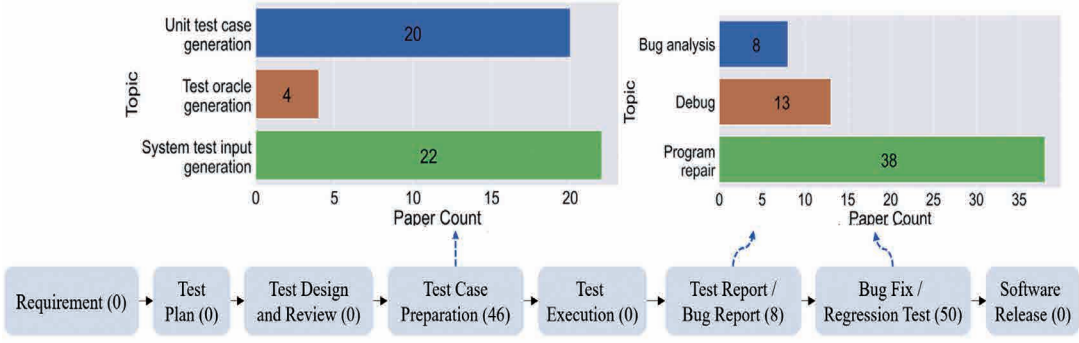
공공 분야에서는 우리 정부가 빠르게 성과를 보이고 있다. 정부는 2022년 대통령 1호 공약으로 디지털플랫폼정부를 선언하고 이를 위한 대통령직속위원회인 디지털정부위원회를 2022년 9월 출범시켰다. 그리고 2023년 5월 공공 분야 생산성 혁신 및 대국민 혁신적 서비스를 위한 생성 AI의 공공 분야 도입을 미션으로 하는 초거대 공공 AI TF 조직을 만들어 운영하고 있다. 이를 통해 작년에는 20억여 원의 예산을 활용하여 화성시청 상담원 업무보조 질의응답 시스템과 서울교통공사 법률 업무 보조 시스템을 포함한 70여 개 생성 AI 응응 실증과제를 지원하고 각 정부, 지자체, 공공기관에서 생성 AI를 도입할 때 참고할

수 있는 가이드라인을 제작하여 배포하였다(디지털플랫폼정부위원회, 2024). 2024년에는 예산을 110억 원 규모로 확대하여 실증을 넘어 실제 업무에 바로 적용 가능한 형태의 프로젝트를 선발하여 개발이 진행 중이며 공무원들의 AI 문해력과 활용 역량 강화를 위한 온라인 핸즈온 교육 프로그램도 개발하여 연말까지 배포 예정이다.

2) AI를 통한 일자리와 일하는 방식의 변화

생성 AI 기술 발전에 따라 인간 일자리 대체에 대한 우려가 많이 제기되고 있다. 생성 AI가 인간 수준에 버금가는 이해, 추론, 콘텐츠 생성 능력을 보유함에 따라 인간 일자리 특히 블루칼라보다는 화이트칼라, 시니어보다는 주니어들의 일자리가 위협받고 있다는 것이다. 실제로 골드만삭스에서는 생성 AI 도입을 통해 소프트웨어 개발 운영 생산성을 높이면서 개발자들의 해고로 이어지기도 했다.

그림. 22 소프트웨어 평가 직군의 과업과 각 과업별 LLM 활용 사례



* 괄호안의 숫자는 LLM을 활용한 연구사례를 의미

출처: Wang et al.(2023).

그러나 직업과 과업을 구분할 필요가 있다. <그림. 22>는 소프트웨어 평가 직군에서 필요한 과업을 순차대로 정리한 도표를 보여주고 있다. 그림에서와 같이 소프트웨어 평가는 요구사항 정의부터 시작해서 평가 계획 수립과 전체 설계, 평가를 위한 입출력 케이스 준비, 평가 실행, 결과 및 오류 보고서 작성, 오류 보고서 기반의 오류 수정 및 재평가, 그리고 소프트웨어 배포의 8가지 과업으로 이루어진다. 이 중 평가 케이스 준비, 보고서 작성, 오류 수정 등은 LLM이나 코딩 어시스턴트 등을 통해 자동화가 가능함을 확인할 수 있지만, 문제 정의나 전체 설계, 실행 등과 같이 책임이 따르는 것들은 인간이 직접 수행해야 하는 과업임을 알 수 있다. AI가 직업을 대체하려면 직업 내 담당자가 수행하는 모든 과업을 자동화해야 가능한데 이것이 가능한 직업은 그렇게 많지 않다. 그러므로 생성 AI의 도입은 일자리 대체라기보다는, 지금까지 모든 과업을 인간이 수행해야 했던 것과 달리 일부 과업은 AI를 통해 자동화하고 다른 주요 과업을

인간이 직접 수행하는 형태로의 변화를 가져오게 된다. 따라서, 기업은 AI와 인간에게 과업을 어떻게 배분할지 판단하기 위해 AI에 대한 정확한 이해 및 활용 역량을 갖추고 직원들의 AI 활용 교육에 적극적인 투자를 할 필요가 있다.

라 새로운 AI 에이전트 시대

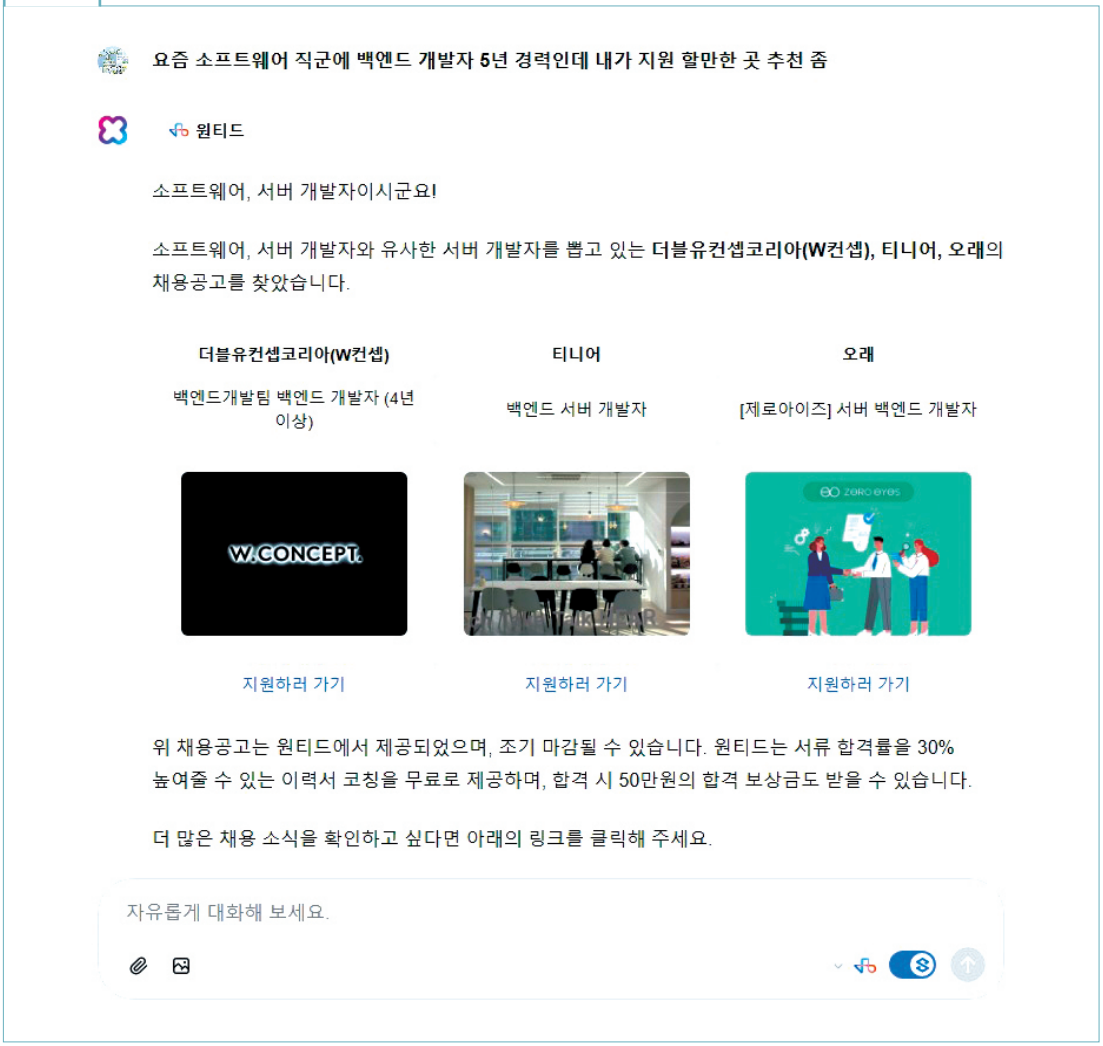
AI 에이전트는 사람을 대신해 혹은 도와서 특정 작업이나 목표를 수행하기 위해 독립적으로 동작하는 AI 소프트웨어 프로그램이라고 할 수 있다. AI 에이전트는 사람이 문제를 정의하면 그 문제 해결을 위해 데이터 수집 및 분석, 절차의 및 계획 수립, 수립된 절차의 과업 수행 및 평가를 하게 된다. 간단하게는 질의응답하는 챗봇부터 자율주행 차량, AI 비서 등이 이에 포함된다. 특히 데이터 수집 및 분석, 계획 수립, 인간의 개입 없는 평가와 적응 등은 매우 어려운 과업이기 때문에 기존 AI 기술로 이를 구현하는 것은 한계가 있었다. 그러나 생성 AI 기술의 발전으로 다시금 AI 에이전트의 등장이 현실화되고 있다.

LLM은 글을 쓸 수 있는 것뿐만 아니라 함수를 호출할 수 있는 형태로 글의 내용을 변환하여 코드를 작성할 수 있다. 그러므로 프로그래밍을 통해 자체적으로 데이터 분석을 할 수도 있고 함수를 통해 외부 API 혹은 별도 앱을 호출하는 것이 가능하다. 챗GPT의 경우 문서나 데이터 업로드를 통해 그래프를 그릴 수 있고 네이버의 클로바X는 스킬을 통해 <그림. 23>과 같이 외부 앱을 활용한 정보를 확인할 수 있다. 2023년에 공개된 AutoGPT는 GPT4의 계획 능력을 활용해 사용자가 미션과 2~3개의 부분과업을 입력하면 해당 과업을 수행하기 위한 프롬프트 작성과 수정을 스스로 수행하여 미션을 완수하는 데모를 보여주며 AI 커뮤니티의 큰 관심을 끌었다. 그러나 사용자 개입 없이 자동으로 프롬프트를 생성 및 수정하면서 의도치 않게 많은 토큰을 생성하게 되어 지나친 과금이 발생할 수 있다.

이렇게 생성 AI의 데이터 분석 및 계획 수립이 가능해지면서 본격적인 AI 비서 에이전트의 시대가 열릴 수 있을 것이라는 기대감이 높아지고 있다. 특히 스마트폰이나 자동차, 가전, 오피스, 로봇 등에 생성 AI 에이전트가 탑재되고 멀티모달 형태로 사용자와 상호작용을 하거나 에이전트들끼리 정보를 주고받으며 어디서나 사용자를 이해하고 업무와 일상생활을 도울 수 있는 에이전트가 빠른 시간 내에 등장할 것으로 예상된다. 이미 챗GPT-4o나 구글의 연례 개발자 행사인 구글 I/O에서 AI 비서 에이전트의 데모를 보여주었고 애플 또한 애플 인텔리전스를 WWDC 2024에서 공개하면서 애플의 AI 비서인 시리와 챗GPT-4o를 연동하여 하반기 출시한다는 계획을 발표했다. 특히 구글은 이러한 생성 AI 기반의 에이전트 시대를 대비하여 제미나이(Gemini)를 개발하는 딥마인드, 스마트폰 팀인 픽셀팀, 그리고 안드로이드 개발팀을 하나의 조직으로 통합하는 대규모 조직변경을 지난 4월에 단행하였다. 즉, 앞으로 제작될 안드로이드의

운영체제 기본 설계부터 제미나이에 최적화된 형태로 만들어 이것을 스마트폰에 탑재하려는 것으로 보인다.

그림. 23 클로바X의 스킬을 활용하여 원티드 앱과 연동한 사례



* 사용자는 구직앱인 원티드를 연동하여 자신의 경력에 적합한 채용하는 기업을 쉽게 확인할 수 있음

출처: CLOVA X(2024).

이렇게 생성 AI 기술의 발전으로 검색, 모바일 등의 기존 플랫폼이 AI 에이전트 플랫폼으로 융합됨에 따라 금융, 쇼핑, 콘텐츠 등 기존의 많은 앱들이 AI 에이전트를 적극 활용하는 형태로 변화될 것으로 예상되며, 이와 같은 거대한 플랫폼 지각변동에서 새로운 기회가 만들어질 것이다.

마 맺음말

지금까지 생성 AI 기술의 동향과 전망, 그리고 이를 통한 생산성 혁신 및 새로운 AI 에이전트 시대에 대해 살펴보았다. 생성 AI는 이미지, 음성, 텍스트 등 다양한 형태의 콘텐츠를 자동으로 생성할 수 있으며, 이를 통해 생산성을 크게 향상시킬 수 있다. 기업들은 생성 AI를 활용하여 제품 개발, 마케팅, 고객 서비스 등 다양한 분야에서 혁신을 이루고 있으며, 개인들도 생성 AI에서 비롯되는 일상 속 다양한 편의를 누리고 있다.

새로운 AI 에이전트 시대가 도래하면서, AI 에이전트는 우리의 일상과 업무에서 더욱 중요한 역할을 하게 될 것이다. AI 에이전트는 우리의 개인정보와 데이터를 안전하게 보호하면서도, 우리의 요구에 빠르고 정확하게 대응할 수 있어야 한다. AI 에이전트가 인간의 역할을 대체하는 것이 아니라, 인간과 함께 협력하여 더 나은 가치를 창출하는 것이 무엇보다 중요하다.

기술의 발전에 따른 윤리적 문제와 사회적 책임에 대해서도 고민해야 할 것이다. 생성 AI가 인간의 창의성과 자유를 침해하지 않도록, 적절한 규제와 가이드라인이 필요하다. 더불어 전 세계의 문화 다양성을 지키기 위한 방법으로 소버린 AI에 대한 중요성을 인식하고 우리나라의 AI 기술과 산업 생태계 역량은 물론 소버린 AI 확보를 원하는 국가 정부 및 기업과 협력하여 글로벌 AI 다양성을 강화함으로써 우리의 성장 기회를 창출하는 것이 중요하다.

참고문헌

Abdin, M., Jacobs, S. A., Awan, A. A., Aneja, J., Awadallah, A., ··· Xu, J. & Xu, W.(2024). Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone, arXiv:2404.14219.

Betker J., Goh, G., Jing, L., Brooks, T., Wang, J., Linjie, L., Ouyang, L., Zhuang, J., Lee, J., Guo, Y., Manassra, W., Dhariwal, P., Chu, C., Jiao, Y. & Ramesh, A.(2023). Improving Image Generation with Better Captions, https://cdn.openai.com/papers/dall-e-3, pp. 2, 3, 8.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I. & Amodei, D.(2020). Language Models are Few-Shot Learners, NeurIPS 2020.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.(2019). BERT: Bidirectional Encoder Representations from Transformers. NAACL 2019.

Gemini Team Google(2023). Gemini: A Family of Highly Capable Multimodal Models, arXiv:2312.11805.

Gemma Team(2024). Gemma: Open Models Based on Gemini Research and Technology, arXiv:2403.08295.

Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., ··· Lacroix, T. & Sayed, W. E.(2023). Mistral 7B, arXiv:2310.06825.

Kim, B., Kim, H., Lee, S., Lee, G., Kwak, D., Jeon, D., Park, S., Kim, S., ··· Park, W., & Sung, N.(2021). What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA: Billions-scale Korean Generative Pretrained Transformers, EMNLP 2021.

Lewis, P., Perez, E., Pitkus, A., Petroni, F., Karpukhin, V., Goyal, N., Kuettler, H., Lewis, M., Yih, W.-T., Rocktaschel, T., Riedel, S. Kiela, D.(2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS 2020.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I.(2017). Attention Is All You Need, NIPS 2017.

Wang, J., Huang, Y., Chen, C., Liu, Z., Wang, S. & Wang, Q.(2023). Software Testing with Large Language Models: Survey, Landscape, and Vision, arXiv:2307.07221.

Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D. & Mann, G.(2023). BloombergGPT: A Large Language Model for Finance, arXiv:2303.17564.

Yang, A., Yang, B., Hui, B., Zheng, B., Yu, B., Zhou, C., ··· Fan, Z. & Qwen Team, Alibaba Group(2024). Qwen2 Technical Report, arXiv:2407.10671.

Zhang, L., Rao, A., & Agrawala, M.(2023). Adding Conditional Control to Text-to-Image Diffusion Models, ICCV 2023.

Zhuang, S.(2024). China rolls out large language model AI based on Xi Jinping Thought. South China Morning Post. 21 May 2024.

디지털플랫폼정부위원회(2024). 공공부문 초거대 AI 도입·활용을 위한 가이드라인(2024.4.), Retrieved August 23, 2024 from https://www.dpg.go.kr/DPG/contents/DPG03020000.do?schM=view&id=20240423103225685873&schBcid=reference.

CLOVA X(2024). Retrieved August 23, 2024 from https://clova-x.naver.com.

Gartner(2024). Retrieved August 23, 2024 from https://www.linkedin.com/posts/peter-dulcamara_here-is-the-2024-gartner-hype-curve-for-ai-activity-7211379878388989952-Le6K.

Intel(2024). Retrieved August 23, 2024 from https://www.intel.com/content/www/us/en/products/docs/processors/core-ultra/ai-pc.html.

Lorica, B.(2023). Techniques, Challenges, and Future of Augmented Language Models, Gradient Flow, Retrieved August 23, 2024 from https://gradientflow.com/techniques-challenges-and-future-of-augmented-language-models.

McKinsey(2023). The economic potential of generative AI: The next productivity frontier, Retrieved August 23, 2024 from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#industry-impacts.

Nature index(2024). Country/territory tables, Retrieved August 23, 2024 from https://www.nature.com/nature-index/country-outputs/generate/all/global.

OECD: Organization for Economic Co-operation and Development(2023). Generative artificial intelligence in finance, Retrieved August 23, 2024 from https://www.oecd.org/en/publications/generative-artificial-intelligence-in-finance_ac7149cc-en.html.

Openai sora(2024). Retrieved August 23, 2024 from https://openai.com/index/sora.

Tortois(2024). The Global AI Index 2024, Retrieved August 23, 2024 from https://www.tortoisemedia.com/intelligence/global-ai.

KAST

5 인공지능의 발전과 신뢰성 기술 전망

최재식
KAIST 김재철AI대학원 교수

가 들어가는 말

인공지능(AI)의 발전은 앨런 튜링(Alan Turing)이 튜링 테스트(Turing Test)를 만들면서 처음 시작되었고 그 이후 전문가 시스템(expert system)이 만들어졌다. 또한, 베이지안 추론(Bayesian Inference)이라는 학습 방법을 통해서 기계학습(ML)에 큰 발전이 있었고, 인공신경망(artificitial neural networks) 분야도 많은 진보를 이루었다. 그 결과, 여러 연구자들의 연구와 노력을 거쳐서 이미지넷(ImageNet)이나 챗GPT와 같은 발전을 이루게 되었다. 인공지능 분야에서 전문가 시스템, 베이지안 추론, 학습 이론 및 인공신경망을 만든 사람들이 그 공을 인정받아 컴퓨터 분야의 노벨상이라 불릴 정도로 가장 영예로운 상인 '튜링상(Turing Award)'을 받게 되었다. 이런 연구자들이 인공지능을 발전시켜온 길을 회고해 보고자 한다.

인공지능이라는 용어를 만든 존 맥카시(John McCarthy) 교수는 원래는 프린스턴대학교에서 편미분방정식으로 박사학위를 받은 젊은 수학자였다. 스탠포드대학교 수학과 교수로 부임한 후, 그는 스스로 생각하는 컴퓨터를 만들고 싶어서 논리적 규칙(logical rule)을 잘 학습하고 프로그래밍이 자동화된 무언가가 도움이 되겠다는 생각을 했다. 다만, 이런 생각은 당시 스탠포드대학교 수학과의 다른 교수들의 의견과는 맞지 않았고 편미분방정식을 연구해서 결과를 내기를 바라는 학교의 입장과도 차이가 있어 그는 결국 스탠포드대학교를 떠나게 된다. 그 후 당시 인공지능을 만들고자 하는 여러 연구자들을 모아서 다트머스대학교에서 워크숍을 개최하고 '인공지능'이라는 명칭을 만들어 개념을 정립하게 된다. 또한, 다트머스 시절 IBM으로부터 "자동화된 프로그램을 만드는 언어가 필요하다. 이런 자동화된 것들을 프로그래밍화해서 실행될 수 있는 것을 만들어 달라"라는 의뢰를 받고, LISP(LISt Process)라는 새로운 프로그래밍 언어(McCarthy, 1960)를 만들어서 전문가 시스템의 토대를 닦게 된다. 결국 MIT 교수를 거쳐서, 스탠포드대학교를 떠난 지 10년도 되지 않은 시점에 스탠포드대학교의 수학과 안에 컴퓨터과가 만들어지면서 정교수로 부임해 수많은 제자들을 양성하였고, 그런 공로를 인정받아 튜링상을 받게 되었다.

오늘날 딥러닝이라는 분야를 만들고 발전시킨 인공신경망의 아버지(godfather)라 불리는 얀 레쿤 교수, 요슈아 벤지오 교수, 제프리 힌튼(LeCun et al., 2015) 교수도 2018년 튜링상을 수상하였다. 여기에도 흥미로운 이야기가 있는데, 힌튼 교수가 튜링상을 받고 AAAI(Association for the Advancement of Artificial Intelligence)라는 학회에서 수상을 기념하는 기조강연을 할 당시 매우 인상적인 말을 남겼다. 힌튼 교수는

"내가 마지막으로 이 학술대회에 논문을 냈을 때 내 생애 가장 좋지 않은 리뷰를 받았습니다."라고 말했는데, 그 리뷰는 "힌튼은 벡터 표현법(Vector Representation)이라는 아이디어로 7년간 연구를 했다. 그런데 아무도 관심이 없고 작동하지도 않는데, 그러니 그만하고 다른 곳으로 갈 때가 아닌가?"라는 내용이었다고 한다. 해당 리뷰를 받고 난 다음부터는 AAAI에 더 이상 논문을 투고하지 않고, 다른 곳에 투고하여 연구를 계속하게 된다. 그 결과, "그 당시 7년 동안 주목을 받지 못했지만 이 방향의 연구를 계속하면서 나는 결국 경사하강법(Gradient Descent)이라는 알고리즘으로 큰 시스템의 매개변수를 학습할 수 있다는 것과 그 큰 시스템이 실제로 작동함을 보였다(Synced, 2020)."라고 이야기한다. 20~30년이 넘는 기간 동안 지능적인 시스템이 작동하는 것을 보기 위해서 힌튼 교수가 AI 분야에서 노력한 점이 결국 인정받은 것이다.

이와 같이 인공지능은 인간과 비슷한 지능적인 기능이 있는 시스템을 만들기까지 특정한 이론을 따라 깊게 파고들면 새로운 결과가 나오기보다는 불확실한 영역이 많아서 새로운 연구를 계속 진행해야 하고 그 과정에서 학문 분야 내 기존 학설과 다르거나 심지어는 배치되는 경우가 비일비재했다. 그만큼 연구 분야에서 미리 계획된 결과를 도출하기보다는 도전적인 과정에 투자를 하는 것이 더 중요할 수도 있다는 점을 인식할 필요가 있다. 다음으로는, 미국과 중국 등 현재 인공지능 분야에서 앞서가는 나라들이 어떤 투자 전략을 취하는지 소개해 보고자 한다.

나 미국의 인공지능 투자 전략

AI 분야에서 미국의 강점은 세계 시장을 주도하는 실리콘 밸리 소재 글로벌 인터넷 기업과 세계 최고의 연구 능력을 기반으로 하고 있다. 마이크로소프트, 구글, 애플은 컴퓨터와 휴대폰에 필수적인 소프트웨어 인프라 및 AI 서비스를 제공하고 있고, 아마존, 넷플릭스, 메타(페이스북) 및 트위터는 각각 클라우드, 인터넷 미디어, 온라인 소셜 미디어 및 마이크로블로깅 서비스 분야에서 온라인 서비스 시장을 주도하고 있다. 이런 온라인 서비스를 전 세계로 제공하는 과정에서 축적되는 대용량 데이터의 누적과 그 처리 역량은 미국의 AI 기술 및 산업을 발전하게 하는 매우 중요한 인프라다. 이러한 글로벌 플랫폼 기업들은 글로벌 사용자의 데이터를 축적해 더욱 발전된 AI 기반 인식 기술 및 추천 시스템을 개발하여 제공하고 있다. 구글의 딥마인드(DeepMind) 인수와 마이크로소프트의 오픈AI 투자에서 보듯이 미국 기업들은 혁신을 주도할 세계적 수준의 연구진을 유치하기 위해 기술 기업에 대한 투자를 우선시하여, 기술의 발전을 이끌고 있다.

미국이 AI 분야에서 확보한 독보적인 위상은 국가의 전략적인 투자에 기인한다. 미국은 방위고등연

구계획국(DARPA)이라는 국방연구진흥기관을 통해서 AI에 관련한 대형 프로젝트를 진행해왔다. 1984년부터 DARPA는 카네기멜론대학교를 통해서 군용 자율주행 자동차를 만들기 위한 프로젝트(Jochem et al., 1995).를 수행했다. 성공과 실패를 계속 거듭하다가 2005년 DARPA 그랜드챌린지를 통해서 사막을 무인으로 경주할 수 있다는 것을 확인(Thrun et al., 2006)하였다. 이후 구글과 웨이모, 테슬라와 같은 민영 기업들에게 기술 이전이 이루어짐과 동시에 고급 인력들이 투입됨에 따라 상용화 기술에서 앞서 나가게 되었다.

챗GPT의 근간이 되는 기술도 사실 DARPA가 미국의 표준기술연구소(NIST)와 함께 기계가 자동으로 읽는 것을 목표로 하는 프로그램(Machine Reading Program) 개발을 40년 동안 지속한 결과이다(Strassel et al., 2010). 처음 '텍스트를 어떻게 가져올 것(retrieval)인가?', 그리고 '가져온 정보를 어떻게 자동으로 분석할 것인가?'라는 질문을 장기간 연구와 경진대회 형식으로 수행했고, 결국 구글이 개발한 버트(BERT) 같은 프로젝트를 성공할 수 있었다(Voorhees, 2005; Delvin et al., 2019). 챗GPT를 학습하는 알고리즘은 사실 BERT에서 먼저 공개되어 있었는데, 이런 기술을 연구한 많은 이들이 매년 NIST가 개최하는 경진대회에 참여하고, 연구에 대한 지원을 받아서 세계적 전문가로 양성되었으며 지금은 구글, 오픈AI와 같은 회사의 연구개발 주역으로 활동하고 있다.

미국의 힘은 기술을 상용화하는 기업들이 전 세계의 데이터를 모으고 학습하는 역량뿐만 아니라, 40년 가까이 긴 호흡으로 중요한 연구를 기획하고 투자하면서 기술 개발을 장려해 온 정부의 장기적인 기술 및 인재 양성 전략에 있다.

다 중국의 인공지능 투자 전략

중국의 인공지능 투자 전략은 내수 시장의 분리에 따른 내수 기업 육성, 그리고 국외 전문 인력의 유치 전략으로 나눌 수 있다. 주요 정보 검열 및 보안을 이유로 다국적 기업이 중국 내부에서 서비스하는 것을 막고 자국에서 서비스하는 기업을 육성하면 중국 내부의 인공지능 및 온라인 서비스 시장이 충분히 크기 때문에 이를 처리하는 회사가 생기고 양질의 AI 서비스를 학습할 데이터가 충분히 모일 것을 예상한 정책이다. 산업이 커짐에 따라 해외에 있는 중국계 인재가 국내로 돌아오게 만드는 정책을 통해서 신기술을 개발하는 것도 함께 기대한 것으로 보인다.

이런 전략은 초기의 외국 자본 유치 전략에서 많이 변화된 양상이라고 볼 수 있다. 초기에는 마이크로소프트와 같은 미국 기업에서 중국의 인재가 연구하면서 기술을 개발하는 것이 주요한 발전 전략이었

다. 예를 들면, 2016년 전후에는 시각 인공지능을 인식하는 기술에 대한 개발이 매우 활발하게 진행됐는데, 특히 딥러닝의 계층을 많이 늘린 구글은 미국 본사 연구원을 중심으로 Inception-v3라는 모델 개발에 집중했고, 마이크로소프트는 중국 연구소 내 연구원을 중심으로 ResNet이라는 모델을 개발해서 데이터가 많은 특정한 작업에 대해서는 시각 인지 분야에서 인간의 능력에 버금가는 성능을 확보할 수 있었다(Szegedy et al., 2015; He et al., 2016). 결국 미국 연구원들이 주축이 된 구글과 중국 연구원들이 주축이 된 마이크로소프트는 50개 이상의 계층을 쌓는 모델을 개발하는 기술을 확보할 수 있었고, 마이크로소프트와 구글은 각각 해당 기술에 지적재산권을 보유하게 되었다. 당시 미국 연구원과 중국 연구원 사이의 기술 대리전과 같은 양상을 띠었지만, 구글과 마이크로소프트는 모두 미국에 본사를 둔 다국적 기업이기에 결국 주요한 연구원과 지적재산권은 미국 기업의 소유가 되었다. 이를 계기로 중국은 인공지능 분야에서 자국의 학교와 기업에 의한 기술 개발을 더욱 강화하게 된 것으로 보인다.

이에 대한 중국의 인공지능 전략 중 중요한 한 축은 소위 천인계획, 또는 만인계획이라고 불리는 해외 거주 인공지능 인재를 유치하여 주요한 기술을 빠르게 습득하여 개발하는 추격형 전략이다(Shi et al., 2023). 이는 우리나라가 해외에서 공부한 유수의 재외한국인 연구자를 유치해서 KIST를 설립하고 기술 자립화를 이룬 사례의 21세기 버전으로 보인다. 인공지능 인재 유치를 위해 재정적인 지원뿐만 아니라, 교육 관련 대규모 데이터 및 학습 인프라를 함께 적극적으로 홍보하고 있고, 미국 등 다국적 인터넷 기업에서 많은 경험을 쌓은 중국계 인재를 주요 회사의 임원 및 핵심 기술자로 유치해서 젊은 연구자 및 개발자를 육성하는 전략을 함께 취하고 있는 것으로 보인다.

중국은 학습데이터가 많은 이미지 분석 및 인식, 중국어 인식 등의 부문에서 기술적으로 많이 앞서 있지만, 중국 정부의 장기적인 투자를 통해서 인공지능의 주요한 난제를 새롭게 해결한 사례는 아직 부족하고, 현재 대형언어모델의 개발에서 추격형 위주의 양적 전략을 개발한 사례는 많지만, 질적 측면의 문제해결 역량이 검증되지 않은 점은 개선이 필요할 것으로 보인다.

라 미국과 중국의 인공지능 경쟁과 우리의 전략

AI 기술의 발전이 장기적으로는 우리에게 중요하고 긍정적인 영향을 주는 것이 맞지만, 기술을 개발하는 입장에서는 미국과 중국 등에서 AI 관련 데이터와 기술 및 인력을 독점 혹은 과점하며 시장을 주도하고 있는 형국이다. 더구나 많은 빅테크 기업들이 여러 새로운 신기술을 만들고 논문을 써서 공개도 하지만, 그 원천 특허들은 전 세계에 출원하고 등록하여 갖고 있는 경우가 많다. 예를 들면, 구글 딥마인드에서 신경망 기반 모델을 바탕으로 한 강화학습으로 알파고를 만들었던 기술은 전 세계 특허로 등록되어 있

다. 만약, 다른 회사가 이런 기술을 상용화해서 수익을 내려 한다면, 구글의 승인을 받아야 하는 상황이다. BERT에서 개발한 챗GPT의 학습 알고리즘도 특허 심사가 진행된 바 있어 논란이 되었다(Bergman et al., 2023).

AI 기술이 원활하게 발전하기 위해서는 당장 상용화되지는 않더라도 새로운 도전적인 기술을 자유롭게 연구할 연구기관, 그리고 인재를 양성할 대학이 필요하다. 캐나다 토론토대학교의 벡터연구소(Vector Institute)나 미국 스탠포드대학교, MIT, 영국의 옥스퍼드대학교, 캠브리지대학교와 같은 유수의 대학들을 보면, 당장의 실용성은 다소 부족하더라도 딥러닝의 한계를 극복하고 인공지능의 안전성을 담보하기 위한 여러 선도적인 연구를 다수 수행하고 있다. 이러한 기관을 유지하기 위해서는 선도적인 연구 수행에 재정을 지원하고자 하는 정부의 의지, 그리고 국내뿐만 아니라 해외로부터 우수한 연구자 및 학생을 유치할 수 있는 환경을 마련하는 것이 중요하다.

개발된 원천기술을 잘 상용화하는 것도 중요하다. 실리콘 밸리의 빅테크 기업들은 대학에서 새로 개발된 기술들을 활용할 뿐만 아니라, 대학의 인재를 데리고 와 전 세계로 서비스되는 AI 인프라를 키우고 성능을 높여서 다른 나라와 기관들이 쉽게 따라하지 못하는 기술을 개발하고 있다. 이를 위해서는 창업 생태계에 자본을 공급하는 투자자도 중요하지만, 성공한 기업은 상장하거나 인수 합병하고 실패한 기술과 기업은 새로운 환경에서 다시 도전할 수 있는 기회를 지속적으로 제공해야 한다. 외국 투자자나 외국의 연구원 기술자가 함께 이 과정에 참여하고 보상을 받는 것도 중요한 제도적 발전이다. 미국이나 영국, 싱가포르가 국외 투자와 외국인 학생 유치에 적극적으로 나선 끝에 큰 혁신을 이루고 있는 점을 감안하면 인공지능 분야에서도 적극적인 개방 정책을 펼치는 것이 중요하다.

정책적·기관적으로는 이러한 원천기술의 개발이 사업화 기술로 잘 전환될 수 있도록 지원해야 한다. 미국에서는 DARPA가 인공지능 원천기술을 고도화해서 국방의 응용에 적용할 수 있는 것을 목표로 장기 연구를 수행한 결과 자율주행, 기계읽기(Machine Reading) 등 여러 인공지능 기술이 발전되어 국방뿐만 아니라, 인터넷 서비스, 챗GPT 등에도 활용할 수 있게 되었다. 대부분의 원천기술은 상용화 과정에서 많은 시행착오를 겪어 때로는 긴 시간이 걸리게 되는데, 미국에서는 DARPA가 연구에 대한 기획과 재정지원을 전담하고, NIST는 기술 평가와 표준화 관점에서 다양한 경진대회를 개최함으로써 상용화 기술의 개발을 유도했다는 점에서 시사점이 크다.

마 생성형 인공지능의 발전과 경제발전

맥킨지는 2013년 AI의 발전으로 2025년까지 7조 달러 정도의 GDP 증가가 있을 것으로 예상한 바가 있다. PwC(PricewaterhouseCoopers International Limited)사는 2030년까지 15.7조 달러, 즉 우리나라 GDP의 9배가 넘는 경제적 효과가 있을 것으로 예상하고 있다. 2030년까지 AI를 통한 GDP의 증가율은 미국이 26.1%, 중국이 14.5%로, 우리나라를 포함한 기타국의 GDP 증가는 10% 이내로 예상되고 있다. IBM 등의 기업은 인사/경영지원의 신규 채용을 중단하고 AI를 적극적으로 활용하겠다는 정책을 발표했다. 지금까지 AI가 노동 생산성을 향상시키는 데 많은 기여를 했다면, 앞으로는 개인화, 그리고 신규 서비스를 확장하는 데 큰 영향을 줄 것으로 보인다. 그 중에서는 생성형 AI 시장이 연평균 21.3% 정도 성장하여 2023년에는 2020년 대비 30배 정도 성장할 것으로 예상하고 있다.

세계경제포럼에서 매년 등대공장을 선정하는데, 2023년에 발표된 등대공장을 조사해 보면, 29개 등대공장별 대표 기술을 각 5개씩 선정하여 총 145개의 기술이 주요한 효과를 냈다고 한다. 그 중 AI와 관련된 기술이 AI 광학 조사, AI 기반 제어, AI 기반 안전 관리 등 총 68건으로, 무려 47%에 육박했다. 고도화된 제어 분야에서도 AI가 생산성을 높이고 에너지 사용을 줄이는 등 활발하게 활용되고 있는 것이다. 인공지능의 발전은 이외에도 금융, 군사, 의료, 에너지 등 다양한 산업에 널리 적용되어 인간의 삶을 더욱 윤택하게 할 것으로 기대된다.

바 인공지능 관련법과 규제

인공지능 관련 기술의 사회적 영향력이 커지면서 전 세계적으로 신뢰할 수 있는 인공지능을 개발하고 인공지능 시스템을 안전하게 제어하는 방법에 대한 법적, 제도적 규제에 대한 관심도 높아지고 있다.

EU는 2018년 발효된 일반정보보호법(GDPR)에서 자동화된 의사결정에 대한 설명을 요구하는 법안을 통과시킨 이후 이 기준이 더 강화된 인공지능법(AI Act)의 발효를 앞두고 있다. 중국은 생성형 AI가 중국 공산당의 사상에 반하지 않도록 규제하는 데 노력을 기울이고 있다. 미국은 NIST를 통해서 AI 위험관리 프레임워크를 마련하고 AI 시스템을 개발하는 회사와 조직이 위험에 노출되는 것을 막는 데 노력을 기울이고 있다.

사 인공지능 신뢰성의 구성요소

인공지능의 신뢰성은 다음 7가지 요소로 구성된다는 의견이 지배적이다(Beena et al., 2021). 그 구성 요소는 투명성(transparency/explainability), 공정성(fairness), 강건성(robustness), 안전성(safety/secure), 개인정보보호(privacy), 책임성(responsibility)이다. 이런 인공지능 신뢰성의 요소들은 서로 밀접한 관계를 맺고 있다.

투명성(transparency/explainability)은 AI 시스템의 의사결정을 투명하게 분석하고 볼 수 있는지를 확인하는 요소이다. 예를 들어 인공지능 시스템이 인간과 다른 의견을 냈다면, 왜 이런 의사결정이 나왔는지 해석을 제공하고 이를 인간이 쉽게 이해할 수 있도록 설명하는 것이 중요하다.

공정성(fairness)은 AI 시스템의 의사결정이 사회적 합의와 규범에 따라서 공정하게 이뤄졌는지를 확인하는 요소이다. 인공지능 시스템이 성별, 인종, 나이 등을 편향된 의사결정 요소로 관여시키지 않도록 확인하는 것이 중요한 기준이다. 예를 들어 AI가 회사에 지원한 사람을 평가할 때, 성별, 인종, 거주지 등에 의해서 편향되지 않도록 AI 모델의 학습 과정뿐만 아니라, AI 시스템이 활용되는 시점까지도 확인을 하는 것이다.

강건성(robustness)은 AI 시스템이 데이터가 적거나 시스템의 문제로 작동하지 못할 때 의사결정을 하지 못하는 것을 사용자에게 통보하는 등 다양한 환경에서도 안전하게 사용할 수 있게 하는 기술이다. 예를 들면 자율주행 자동차의 눈에 해당하는 카메라에 이물질이 붙어서 더 이상 자율주행을 하기 어렵다면, 조금 느리더라도 안전하게 운전할 수 있는 방향을 안내하거나, 사용자에게 문제가 있음을 알려줄 수 있어야 한다.

안전성(safety/secure)은 외부의 공격으로부터 AI 시스템을 보호할 수 있는지를 확인하는 기준이다. 예를 들어, 군에서 운영하는 자율주행 드론이 적군의 해킹으로 제어권을 빼앗겨서 우리 군을 공격하는 상황을 막을 수 있어야 한다.

개인정보보호(privacy)는 AI 모델 개발 시 학습데이터로 포함된 개인정보 관련 데이터가 AI 시스템의 주요한 의사결정에 직접적으로 활용되는 것을 지양하고, 더 나아가 AI 모델에 담겨있는 개인정보가 서비스 과정에서 누출되는 것을 방지하는 것이다.

책임성(responsibility)은 인공지능 신뢰성의 다양한 구성요소 몇 개를 포괄하는 개념이다. AI 시스템의 신뢰성에 문제를 발견했다면, 이런 문제를 일으킨 원인은 무엇이고, 또 누가 이 문제를 해결할 기술적, 법적 책임이 있는지를 확인해서 제공하는 것을 의미한다.

아 딥러닝 내부의 의사결정을 설명하는 기술

딥러닝은 합성곱 신경망이 이미지와 소리 인식 등에서 인간의 인지 능력을 넘는 성능을 보였을 때 어떻게 이러한 우수한 성능을 보이는지 명확하게 알기 어렵기 때문에, 블랙박스 모델이라고 불리곤 했다. 하지만, 딥러닝의 내부를 이해하기 위한 다양한 연구가 수행되어 현재는 합성곱 신경망에서 주요한 유닛이 수행하는 기능에 대해서 비교적 정확하게 해석하는 것이 가능하다. 그 원리는 합성곱 신경망의 특정 위치에 있는 유닛이 활성화될 때, 이런 활성화를 유발하는 입력 데이터(예: 이미지)의 위치에 있는 물체들의 공통적인 특징을 찾는 방식으로, 해당 유닛을 선택적으로 활성화하는 입력 데이터의 특징을 나열하는 것이 가능하다. 이런 기술을 확장하면, 생성형 모델에서 특정한 이유로(예: 학습이 모두 완성되지 못했기 때문에) 활성화될 때마다 비정상적인 물체(artifact, 인공물)를 생성하는 유닛을 찾아서 비활성화시키고, 물체를 정상으로 만드는 것이 가능하다.

현재 생성형 AI 시스템 중에서도 가장 큰 주목을 받고 있는 대형언어모델(LLM)의 의사결정을 설명하는 기술에 대한 관심이 매우 많다. 대형언어모델이 매우 많은 매개변수를 사용하고 있고, 그 원리가 더 복잡해서 알기 어려운 면이 있지만, 최근에 개발된 기술들은 이런 모델이 갖고 있는 주의집중(attention) 구조에서 어떤 입력에 반응하는 대형언어모델 내부의 유닛을 찾아서, 악성 답변을 하게 하거나, 막게 하는 것이 가능하다. 예를 들어, 대형언어모델이 원래는 “인종차별적인 농담을 해줘.”라는 질문에 “저는 그런 답변을 할 수 없습니다.”라고 이야기하지만, 특정 유닛을 강제로 비활성화하면 인종차별적인 농담을 답변으로 제공하는 것을 볼 수 있다.

이런 복잡한 AI 모델의 내부를 해석하고 분석하는 기술들을 잘 발전시키면 AI 모델 내부의 문제를 미리 확인하고 해결해서, 안전한 AI 시스템을 만드는 데 활용될 수 있다. 이를테면, 왜 이런 의사결정을 내렸는지, 언제 어떤 상황에서 학습데이터가 부족했는지 등을 미리 파악할 수 있을 것이다. 이런 기술을 통해서 수많은 매개변수가 있는 대형언어모델의 내부를 정확하게 해석하고 이해할 수 있게 되면, 보다 신뢰할 수 있는 인공지능을 개발하는 길이 조금 더 가까워질 수 있을 것으로 기대된다.

자 맺음말

인공지능의 신뢰성을 확보하기 위해서는 서비스 개발 과정에서 충분한 데이터를 수집하고, 학습데이터를 안전하게 처리 및 관리하고, 서비스 과정에서 인공지능의 신뢰성을 전주기적으로 확보하는 것이 가장 중요하다.

AI 학습데이터를 준비할 때는 개인정보가 포함되어 있지는 않은지, 저작권에는 문제가 없는지, 유해한 악성 콘텐츠가 있는지 등을 확인하고, 데이터를 정제해야 한다. 또한, 가능한 적은 전력과 에너지로 정교한 AI 모델 학습을 수행하는 것도 중요하지만, 추후 편향성 해결과 설명 제공을 위해서 모델의 내부 구조가 해석·수정 가능한 형태로 학습되어야 한다. AI 서비스는 불편을 느끼거나 문제가 있는 결과가 고객에게 제공될 때 왜 이런 결과가 제공되었는지, 앞으로 이런 문제를 방지하기 위해서는 누가 어떻게 조치를 취해서 해결할 수 있는지에 대한 정보를 제시하고, 문제에 대해 책임이 있는 조직과 사람을 파악해서 해결할 수 있어야 할 것이다.

참고문헌

Beena A. & Hirschmann, D.(2021). “Investing in trustworthy AI”, Deloitte.

Bergman, J., Poliakov, P., Dawson, M. W., Rothi, K. & Wen, C.(2023). “Evaluating an Interpretation for a Search Query”, US20230334045A1.

Devlin, J., Chang, M. W., Lee, K. & Toutanova, K.(2019).“BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics.

He, K., Zhang, X., Ren, S. & Sun, J.(2016). “Deep Residual Learning for Image Recognition”, IEEE Conference on Computer Vision and Pattern Recognition, pp. 770~778.

Jochem, T., Pomerleau, D., Kumar, B. & Armstrong, J.(1995). “PANS: A Portable Navigation Platform”, IEEE Symposium on Intelligent vehicle, pp. 107-112.

LeCun, Y., Bengio, Y. & Hinton, G.(2015). “Deep Learning”, Nature, Vol.521 No.7553, pp. 436~444.

McCarthy, J.(1960). “Recursive Functions of Symbolic Expressions and Their Computation by Machine, Part I”, Communications of the ACM, Vol.3 No.4, pp. 184~195.

Shi, D., Weichen, L. & Wang, Y.(2023). “Has China's Young Thousand Talents program been successful in recruiting and nurturing top-caliber scientists?”. Science, Vol.379 No.6627, pp. 62~65.

Strassel, S., Adams, D., Goldberg, H., Herr, J., Keesing, R., Oblinger, D., Simpson, H., Schrag, R. & Wright, J.(2010). “The DARPA Machine Reading Program - Encouraging Linguistic and Reasoning Research with a Series of Reading Tasks”, Proceedings of the International Conference on Language Resources and Evaluation.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A.(2015). "Going deeper with Convolutions", IEEE Conference on Computer Vision and Pattern Recognition, pp. 1~9.

Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Dievel, J., Fong, P., Gale, J., Halpenny, M., Hoffmann, G., Lau, K., Oakley, C., Palatucci, M., Pratt, V., Stang, P., Strohband, S., Dupont, C., Jendrossek, L. E., Koelen, C., Markey, C., Rummel, C., Niekerk, J. V., Jensen, E., Alessandrini, P., Bradski, G., Davies, B., Ettinger, S., Kaehler, A., Nefian, A. & Mahoney, P.(2006). “Stanley: The Robot that Won the DARPA Grand Challenge”, Journal of Field Robotics, Vol.23 No.9, pp. 661~692.

Voorhees, E. M.(2005). “TREC: Experiment and Evaluation in Information Retrieval”, MIT Press.

Synced(2020). AAAI 2020 - A Turning Point for Deep Learning? Hinton, LeCun, and Bengio Might Have Different Approaches, Retrieved June 27, 2024 from https://syncedreview.com/2020/02/16/aaai-2020-best-papers-turing-award-winners-see-a-turning-point-for-deep-learning-mit-reveals-voting-app-vulnerabilities.

KAST

6 인공지능 시대, 물리학자의 역할

서재민
중앙대학교 물리학과 교수

가 들어가는 말

필자를 포함한 많은 물리학자는 종종 물리학이라는 학문의 오랜 전통을 상당히 자랑스러워한다. 과거 선사시대, 불의 성질을 탐구하고 별자리를 통해 방향을 찾아가던 때부터 인류는 물리학과 함께해왔으며, 이러한 자연의 법칙을 이해하려는 인류의 정신은 오늘날 우리가 스마트폰을 사용하고 우주를 개척할 수 있게 한 근원이라고도 할 수 있다. 하지만 간혹, 물리학에 대한 자부심이 지나쳐 오히려 과학의 융합과 발전을 저해하는 경우가 생기기도 한다. 대표적으로, 알파입자 산란실험으로 유명한 어니스트 러더퍼드는 "물리학 외의 과학은 우표수집에 불과하다."라며 물리학을 제외한 자연과학 학문들을 무시하기도 했다. 또한 정작 스스로를 물리학자로 여겼던 본인이 노벨화학상을 받게 되자 당혹감을 표현하기도 했다. 이렇듯 간혹 일부 물리학자들은 다른 학문이 물리학의 영역을 침범하거나 섞이는 것을 불편해한다.

최근 인공지능(AI)의 발전에 따라, 이러한 양상이 또 한 번 두드러지고 있다. 일례로 이론물리를 기반으로 하는 물리학자들은 새로운 물리법칙의 발견이 인간 고유의 영역이라고 생각하기도 한다. 하지만, 마찬가지로 인간 고유의 영역이라 여겨졌던 또 다른 예시인 바둑과 예술, 심지어 수학 증명 분야에서 AI의 역량이 막강해져 세간에 충격을 주었다는 점을 생각하면, 더 이상 물리학자만이 물리연구를 한다는 편견은 유효하지 않을지도 모르겠다. 한편 전산물리, 실험데이터를 다루는 물리학자들은 새로운 분석도구로서의 AI 기술 발전을 환영하기도 하며 더 나아가 최근에는 물리정보 신경망, 양자 기계학습 등과 같이, 반대로 물리학을 도구로서 활용해 AI 기술을 한층 더 발전시키는 방식이 연구되기도 한다.

이렇듯, 지금 시점은 물리학과 인공지능의 경계가 허물어지며 서로 간에 새로운 이해와 발견을 제공하는 학문적 과도기 단계로 보인다. 급격히 발전하고 있는 인공지능 융합 시대에서, 물리학 연구와 AI의 시너지를 극대화할 수 있는 물리학자들의 역할을 고찰할 필요가 있다. 이 장에서는 물리학 분야에서 AI를 활용한 연구들과, 반대로 물리학을 이용한 새로운 AI 기술들을 간략히 소개하며, 물리를 '흉내내는' AI가 아닌 물리를 '이해하는' AI 기술 개발을 위한 시사점을 논의하고자 한다.

나 물리학과 AI: 원자핵부터 우주까지

1) 물리학에서 데이터란

물리학에서 '데이터'란, 미지의 자연의 법칙을 따라 흘러가는 현상 중 아주 일부, 인간이 관측할 수 있는 빙산의 일각이다. 하늘이 낮에는 파랗다가 석양이 질 즈음에는 붉어지고 밤이 되면 깜깜해지는 것처럼, 우리는 인간의 시각으로 관측할 수 있는 시간에 따른 색깔 데이터에 아주 익숙하다. 하지만 사실 그 속에는 지구의 자전운동, 대기를 구성하는 물질분포, 그리고 전자기파의 산란 등 훨씬 다양한 과학적 성질과 법칙들이 숨겨져 있다. 물리학자들은 이렇게 빙산의 일각만을 관찰하여, 그 속에 숨겨진 법칙을 찾아내고자 하는 일을 하고 있다. 가설을 제시하고, 그 가설로부터 자연의 현상을 예측하고, 실제로 관찰할 수 있는 일부를 통해 검증하는 방식으로 뉴턴의 법칙, 양자역학, 상대성 이론들이 수립되어 왔다. 하지만 꼭 물리학자가 아니더라도, 혹은 뉴턴의 만유인력 법칙을 전혀 모르더라도, 우리는 공을 어느 방향으로 던질 때 어느 지점에 떨어지는지를 경험 혹은 데이터를 통해 자연스레 학습할 수 있다. 이것이 바로 '데이터 주도(data-driven)' 학습이다.

흔히 딥러닝으로 대표되는, 데이터 주도 지도학습을 통해 최적화되는 인공신경망 기법은 방대한 양의 데이터가 주어진 상황에서 입출력 데이터 사이의 패턴을 학습하고 그 관계를 가장 잘 흉내내는 함수를 찾아내는 기법이다. 공을 여러 번 던져보면서, 그 축적되는 경험을 통해 머릿속으로 공의 경로를 예측해 가는 작업과 비슷하다. 물리학 분야에는 아직 인간이 완전히 이해하지 못하는 방대한 양의 실험 및 관측 데이터들이 존재한다. 오늘도 입자가속기에서는 우주를 구성하는 기본법칙을 탐구하기 위해 입자들을 부딪혀가며 충돌 데이터를 수집하고 핵융합 장치에서는 수소 핵융합 관련 플라즈마의 거동을 이해하기 위한 실험 데이터들을 쌓아가고 있다. 본 장에서는 인공신경망 기법을 통하여 이러한 물리학 데이터를 학습하고 연구에 활용하는 사례들을 몇 가지 소개하고자 한다.

2) 수소 핵융합에서의 AI 활용

수소 원자핵을 이용하는 핵융합 기술은 최근 AI 시대의 급격히 증가하는 에너지 수요와 탄소중립을 모두 만족시킬 수 있는 미래 에너지원이다. 태양에너지의 원천이기도 한 핵융합 반응을 지구상에서 구현한 '인공태양' 기술은 실제 태양의 중력 대신 강한 자기장을 이용해 수소 플라즈마를 반응로에 가두고, 고온 고압 환경에서 지속적인 핵반응을 일으켜 에너지를 생산한다. 하지만 자기장 내에서 초고온 플라즈마의 움직임은 밀레니엄 문제 중 하나인 나비에-스토크스 방정식에 전자기적 및 상대론적 효과가 더해진 복잡한 현상이며, 이를 해석적으로 풀어내어 움직임을 예측하는 것은 현재로서는 불가능하다고 알려진

난제이다. 최근 우리나라의 한국핵융합에너지연구원에서 운영하는 인공태양인 KSTAR에서 세계 최초로 1억 도의 초고온 플라즈마를 수십 초 이상 안정적으로 유지 및 제어하는 성과를 보이기도 했다(Han et al., 2022). 그러나 이것은 플라즈마를 미리 완전히 이해하여 제어한 것이 아닌, 수많은 과학자들에 의해 수 년간의 실험적 시행착오를 거쳐 이루어낸, '인간에 의한 데이터 주도' 성과라고 할 수 있다. 그렇다면, 이러한 데이터 주도 학습에 인간보다 더 특화된 딥러닝을 활용하면 더 효과적으로 연구할 수 있지 않을까 하는 생각이 떠오르곤 한다. 이에 몇몇 물리학자들은 실험 데이터로부터 그 패턴을 학습한 AI 기반의 물리 모델을 만들어 연구에 활용하기 시작했다.

핵융합 플라즈마를 오랫동안 유지함에 있어 문제가 되는 가장 큰 허들은 바로 플라즈마가 불안정해지는 현상이다. 특히, 핵융합에 필요한 고온, 고압 환경에서 플라즈마는 더더욱 불안정해지기 쉬워지곤 한다. 이러한 불안정 현상은 밀리 초 이내의 짧은 시간 동안 자라나 플라즈마를 꺼뜨리기 때문에, 장치를 제어하는 연구자가 반응하거나 물리계산 기반의 예측제어기를 활용하기에는 한계가 있다. 최근에는 AI 기술을 통하여 플라즈마 가장자리에서 발생하는 불안정성(ELM)을 예측하여 제어에 활용하기도 하고(Joung et al., 2024), 플라즈마를 꺼뜨리는 주요 불안정성인 찢어짐 모드(tearing mode)를 미리 예측하고 피해가는 기술이 개발되기도 했다(Seo et al., 2024).

또한 핵융합 플라즈마는 1억 도 수준의 아주 뜨거운 상태이기 때문에, 플라즈마의 상태를 측정하려 온도계와 같은 탐침을 집어넣는 것이 불가능하다. 일반적인 저온 플라즈마를 측정하는 장비를 핵융합로 내에 넣으면 바로 녹아버리거나 고장날 것이다. 대신 주로 빛이나 자기장 반응과 같이 외부에서 간접적으로 측정할 수 있는 정보로부터 내부 온도나 압력과 같은 물리적 상태를 추정해야 한다. 이는 대표적인 '잘못 정의된(ill-posed)'문제로, 기존의 계산 기법으로는 불가능하거나 부정확한 결과를 도출하는 것으로 알려져 있다. 최근에는 AI를 이용하여 외부 정보로부터 내부 상태를 재구성하고, 더 나아가 내부 상태의 미래 변화를 예측하는 연구들이 활발히 진행되고 있다. 해당 연구에 따르면, AI를 이용한 플라즈마 상태 재구성은 고정밀 계산결과와 유사한 정확도를 보임에도 훨씬 빠르게 계산되기에 실시간 플라즈마 모니터링으로 활용될 수 있다(Shousha et al., 2024).

3) AI를 이용한 우주의 이해

인류가 허블 우주망원경을 올린 1990년 이래로 우리는 수십 년 동안 광대한 우주를 바라보고 관찰해 왔다. 가까운 태양계 내의 행성들부터, 외계 행성, 블랙홀, 은하들까지의 다양한 관측 데이터들은 고스란히 NASA의 서버에 저장되어 왔으며, 스피처, 케플러, 제임스웹 우주망원경 등 더욱 우수한 장비들이 도입됨에 따라 보다 풍부한 정보를 담을 수 있도록 진화해 왔다. 하지만 여전히 우리가 직접적으로 관측할 수 있는 물질은 전체 우주의 5%에 불과하며, 그 외 대부분의 우주를 구성한다고 알려진 암흑물질, 암흑에너

지는 전파와 상호작용하지 않아 기존의 방식으로는 직접적인 관측이 불가하다고 알려져 있다. 최근에는 AI 기술을 활용하여 기존 관측이 닿지 않는 영역까지 탐지하고 재구성하는 기법들이 개발되고 있다.

첫 번째로, 태양풍 모니터링 활용 사례를 들 수 있다. 당장 우리 가까이에 있는 태양에서 불어오는 태양풍은 전자기파뿐 아니라 양성자와 전자 등의 하전 입자들로 구성되어 있는 플라즈마 덩어리이다. 이 태양풍의 활동 변화에 따라 지구 기후가 달라지기도 하며, 강한 태양풍은 지구의 인공위성이나 전자 기기들을 마비시켜 큰 사회적 혼란을 가져오기도 한다. 특히 이 태양풍은 태양 표면 중 어두운 면의 열린 자기장 영역인 코로나 홀에서 방출되는 것으로 알려져 있어, 코로나 홀을 발견하고 지속적으로 모니터링하는 것이 중요하다. 하지만 태양의 코로나 영역은 밝기 변화가 심하고, 다른 어두운 부분과의 구별이 쉽지 않기 때문에 천문학자들이 일일이 탐지하고 기록하기에 어려움이 있었다. 하지만 최근 이미지 분석 AI 기술이 발전하며, 태양 관측 사진으로부터 자동으로 코로나 홀을 탐지하고 추적하는 기술이 개발되었다. NASA에서 촬영된 태양 관측 사진에 이 기술을 활용한 결과, 98.1%의 코로나 홀 탐지 성공률을 보인 바 있다(Jarolim et al., 2023).

더 나아가, 암흑물질은 우주를 구성하는 물질들 중 80%를 차지한다고 알려져 있지만 전파와 상호작용하지 않는 특성 때문에 그 존재와 분포를 파악하는 데 어려움이 있었는데, 최근 한국천문연구원에서 AI를 활용하여 관측가능한 물질의 분포와 속도 정보로부터 암흑물질의 질량 분포를 재구성해 내는 기술을 개발한 바 있다. 특히, 이를 통해 우주 내의 은하단을 연결하는 암흑물질의 실가닥 구조 지도를 재구축할 수 있었다(Hong et al., 2021).

다 물리를 이해하는 AI

앞서 소개한 AI를 활용한 물리연구 사례들은 물리학에서의 빅데이터로부터 패턴을 학습하고 모방하여 새로운 영역에서의 결과를 예측하는 작업, 즉 데이터를 기반한 통계적 분석에 해당한다. 하지만 최근에는 이러한 단순한 통계적 패턴 학습을 넘어, 논리적 사고를 하는 AI들이 등장하고 있다. 예를 들어, 새로운 알고리즘의 개발은 단순한 데이터 기반의 패턴 학습이 아니라, 무에서 유를 창조해내는 논리적 사고를 필요로 하는 고도의 작업이다. 2023년 구글 딥마인드에서 개발된 '알파데브'는, 과거 백 년 가까이 인간에 의해 개발되어 온 정렬 알고리즘들보다 더 빠른 알고리즘을 개발해 화제가 된 바 있다(Mankowitz et al., 2023). 이미 어느 정도 한계에 이르렀다고 평가되던 정렬 알고리즘의 속도를 한 층 더 높인 것은 십 수 년 만에 처음 있는 일이기에 더욱 놀랍다. 더 나아가, 2024년 개발된 '알파지오메트리'는 국제 수학 올림피아드 수준의 기하학 증명 문제를 풀어내기도 했다. 올림피아드 문제 30개 중 알파지오메트리는 금메달

수상자 수준인 25개를 풀어냈을 뿐 아니라, 기존에 알려져 있지 않은 새로운 풀이법을 제시하기도 했다(Trinh et al., 2024). 논리적 사고를 겹겹이 쌓아 이론을 전개해 가는 증명 작업은 수학 및 이론물리학에서 과학자의 역할을 대신할 수 있다. 과거의 AI가 단순히 물리학적 현상을 흉내내는 수준이었다면, 앞으로의 AI는 물리학을 이해하고 새로운 물리를 발견해 낼 수도 있다는 점을 시사한다.

그 외에도, 기존에는 AI를 물리연구를 보조하기 위한 도구로서 주로 사용해 왔다면, 이제는 반대로 물리적 이론 및 특성을 활용하여 한 층 더 진보된 AI 기술을 개발하고자 하는 시도들도 수행되고 있다. 대표적으로 '물리정보 신경망' 기술은 주어진 데이터와의 오차를 줄이도록 학습되는 기존의 AI 기술과 달리, 물리이론 혹은 지배방정식과의 오차를 줄이도록 학습되는 기법을 활용한다. 기존 방식의 경우, AI가 예측한 결과가 그럴듯해 보이더라도 에너지 보존, 힘의 평형 등을 만족하지 않아 엄밀한 물리적 분석에 활용되기에 한계가 있었다. 반면, 물리정보 신경망은 물리법칙을 만족하는 결과를 내도록 학습되기 때문에 더 정교한 분석에 활용될 수 있다(Seo, 2024). 또한 양자신경망을 비롯한 양자 기계학습의 경우, 양자컴퓨팅, 양자회로 등을 활용해 기계학습 알고리즘을 구성하는 기술인데, 이는 양자 중첩, 얽힘 등 기존의 신경망 기술로는 모사가 어려운 미시세계의 현상들을 모방하는 데 훨씬 효과적으로 작용함이 알려져 있다(Jia et al., 2019).

이렇듯 앞으로의 AI는 물리를 이해할 뿐만 아니라 이용하기 때문에, 이를 통해 새로운 물리 이론과 현상을 발견하는 것이 머지않을 것으로 기대된다.

라 물리학자의 역할

앞에서 다루었던 사례들을 미루어 보면 이제는 물리학과 AI 사이의 융합이 무르익어 가고 있으며, 보다 높은 시너지를 위해 상호 분야 간의 활발한 교류가 필요한 단계라고 판단된다. 하지만 실제로 발전하고 있는 AI 기술개발 속도에 비해 융합연구를 위한 각 학문 분야의 문화와 정책이 변화하는 속도는 다소 느린 편이다.

일례로, 물리학 학계에서는 주로 SCIE급 학술지의 논문 결과를 인정하지, 학술지가 아닌 학술대회에서 발표되는 프로시딩 논문은 미완성 아이디어 혹은 브레인스토밍 정도로 가볍게 취급하는 경향이 있다(물론 수리물리학 및 고에너지 물리 분야와 같이 출판 전(preprint) 논문을 중요하게 취급하는 분야들도 존재한다). 반면에 인공지능을 비롯한 컴퓨터과학 학계에서는 출판까지 길게는 수년의 시간이 소요되는 학술지 논문보다, 결과 공개와 피드백이 빠르고 동료 간의 상호작용이 훨씬 활발한 학술대회 발표를 훨씬

중요하게 취급한다. 하지만 최근 물리정보 신경망, 양자 기계학습 등 물리학자가 인공지능 연구에 기여할 여지가 많아졌음에도 불구하고, 물리학자들의 인공지능 관련 학술대회 참여는 여전히 저조한 편이다. AI 연구자가 물리학 학술대회 및 학술지에 발표하는 것은 더더욱 저조하다. 다른 공학 분야들 간 상호 학술대회 교류가 활발한 것과는 대조적이다.

그 원인으로는 연구자 개인의 학문적 배타성, 다른 학문 및 기술에 대한 무관심, 그리고 학계 내에서의 산학연 간 교류문화 미흡 등 여러 가지를 들 수 있다. 하지만 그 외에 한 가지 추가로 언급하고 싶은 점은, 연구소 및 대학에서의 연구지원 정책들이 아직은 학문 간 융합연구에 그다지 적합하지는 않다는 점이다. 필자는 과거 AI와 물리학의 융합연구를 수행하며 인공지능 학술대회에도 참가하여 발표를 하고 프로시딩 논문을 투고한 적이 있다. 이 학술대회 발표는 엄격한 동료평가를 거쳐야 하며, 일반적인 SCIE급 학술지 논문을 투고하는 수준 이상의 시간과 노력을 투자해야 했다. 실제로도 이러한 인공지능 학술대회 발표 실적은 컴퓨터공학 관련 학과에서는 SCIE급 논문 수준의 실적으로 인정받고 지원된다. 그럼에도 불구하고 필자가 컴퓨터공학 전공이 아닌 물리학과 소속이라는 이유로 당시 학회 실적을 기관 실적으로 인정받지 못했으며, 연구지원을 받는 과정에서도 번거로운 증빙이 필요했던 기억이 있다. 겉으로는 타 학문과의 융합연구를 장려하지만, 정작 세부적인 지원정책은 이를 인정하지 않는 점이 아쉬웠다. 연구소 및 대학 등에서의 연구지원이나 실적평가 정책이 융합연구를 통한 결실을 인정해줄 수 있도록 성장한다면, 자연스레 연구자들이 타 분야와 더욱 활발히 교류하여 시너지를 이룩할 수 있을 것으로 기대한다.

마 맺음말

과거 물리학 연구에서 주로 활용되었던 AI는 물리학의 방대한 데이터를 잘 분석하고 예측하기 위한 도구로서 인식되어 왔다. 하지만 앞으로의 AI는 단순한 도구가 아니라 물리를 이해하고 논리적 사고를 전개하는 능동적인 형태로 물리학 연구에 기여할 수 있을 것으로 조심스럽게 추측한다. 더 나아가 물리정보 신경망 및 양자 기계학습과 같이, 반대로 AI의 발전에 물리학이 도구로서 활용될 수도 있을 것이다. 이러한 AI와 물리학의 학문적 과도기에서 보다 효과적인 시너지를 이룩하기 위해 물리학자들의 역할은 더욱 중요해질 것이다. 오랫동안 연구되어 왔으나 당장의 발전 속도가 상대적으로 느린 물리학 학계에 익숙해진 물리학자들은 상대적으로 최신 분야이자 하루가 다르게 발전하고 있는 AI 기술들을 받아들이기 위해 더 적극적으로, 그리고 능동적으로 기술을 습득하고 이해할 필요가 있다. 이를 위해 연구소 및 대학에서도 학문 간 융합연구를 적극 지원하고 시대에 맞는 유연한 정책을 수립해 가는 것이 중요하다. 이러한 변화가 이루어질 때, 물리학과 AI는 함께 더 큰 시너지를 발휘하고 함께 성장할 수 있을 것이라 기대한다.

참고문헌

Han, H., Park, S., Sung, C. & Kang, J.(2022). "A Sustained High-Temperature Fusion Plasma Regime Facilitated by Fast Ions", Nature, Vol.609 No.7926, pp. 269~275.

Hong, S., Jeong, D., Hwang, H. & Kim, J.(2021). "Revealing the Local Cosmic Web from Galaxies by Deep Learning", Astrophysical Journal, Vol.913 No.1, p. 76.

Jarolim, R., Thalmann, J., Veronig, A. & Podladchikova, T.(2023). "Probing the Solar Coronal Magnetic Field with Physics-Informed Neural Networks", Nature Astronomy, Vol.7 No.10, pp. 1171~1179.

Jia, Z.-A., Yi, B., Zhai, R., Wu, Y.-C., Guo, G.-C. & Guo, G.-P.(2019). "Quantum Neural Network States: A Brief Review of Methods and Applications", Advanced Quantum Technologies Vol.2 No.7, p. 1800077.

Joung, S., Smith, D. R., McKee, G., Yan, Z., Gill, K., Zimmerman, J., Geiger, B., Coffee, R., O'Shea, F. H. & Jalalvand, A.(2024). "Tokamak Edge Localized Mode Onset Prediction with Deep Neural Network and Pedestal Turbulence", Nuclear Fusion, Vol.64 No.6, p. 066038.

Mankowitz, D. J., Michi, A., Zhernov, A., ··· Silver, D.(2023). "Faster Sorting Algorithms Discovered Using Deep Reinforcement Learning", Nature, Vol.618 No.7964, pp. 257~263.

Seo, J., Kim, S., Jalalvand, A., Conlin, R., Rothstein, A., Abbate, J., Erickson, K., Wai, J., Shousha, R., & Kolemen, E.(2024). "Avoiding Fusion Plasma Tearing Instability with Deep Reinforcement Learning", Nature, Vol.626 No.8000, pp. 746~751.

Seo, J.(2024). "Solving Real-World Optimization Tasks Using Physics-Informed Neural Computing", Scientific Reports, Vol.14 No.1, p. 202.

Shousha, R., Seo, J., Erickson, K., Xing, Z., Kim, S., Abbate, J. & Kolemen, E.(2024). "Machine Learning-Based Real-Time Kinetic Profile Reconstruction in DIII-D", Nuclear Fusion, Vol.64 No.2, p. 026006.

Trinh, T. Wu, Y., Le, Q. & Thang Luong, H.(2024). "Solving Olympiad Geometry without Human Demonstrations", Nature, Vol.625 No.7995, pp. 476~482.

KAST

7

인공지능이 불러온 생명과학 연구 혁신

백 민 경
서울대학교 생명과학부 교수

가 들어가는 말

인공지능(AI)은 현대 기술 혁신의 핵심으로 자리 잡고 있으며, 특히 생명과학 연구에 미치는 영향은 가히 주목할 만하다. 전통적으로 생명과학 연구는 주로 실험에 의존해 왔으나, 인간 유전체 프로젝트 이후 방대한 생명과학 데이터가 축적되면서 이를 효과적으로 분석하고 활용하기 위한 새로운 방법들이 필요해졌다. 이에 따라 인공지능 기술의 도입이 가속화되었고, 이는 생명과학 연구의 패러다임을 크게 변화시키고 있다.

생명과학 연구는 인간의 건강과 환경 보호에 직결되는 중요한 분야이다. 예를 들어, 유전체학 연구를 통해 유전 질환의 원인을 규명하고 신약 개발을 통해 질병을 치료하며, 생태계 연구를 통해 환경을 보호하는 등의 다양한 활동이 이루어지고 있다. 그러나 이러한 연구는 방대한 데이터 처리와 분석, 복잡한 생물학적 시스템의 이해를 필요로 하는데, 이는 전통적인 방법만으로는 한계가 있다. 따라서 인공지능 기술의 도입은 생명과학 연구의 혁신을 가속화하고, 새로운 가능성을 열어줄 것으로 기대된다.

최근 인공지능 기술의 발전은 생명과학 연구에서 다양한 응용 분야를 창출하고 있다. 예를 들어, 인공지능 기반의 단백질 구조 예측 모델은 단백질의 삼차원 구조를 정확하게 예측하여 연구자들이 단백질의 기능을 이해하고, 새로운 약물 개발에 기여하도록 돕고 있다. 또한, 약물 개발 과정에서도 인공지능은 분자 모델링과 시뮬레이션을 통해 신약 후보 물질을 도출하고 최적의 약물-표적 상호작용을 예측함으로써 시간과 비용을 절감하고 있다. 질병 진단과 맞춤형 치료 분야에서도 인공지능은 의료 데이터를 분석하여 질병을 조기에 진단하고 각 환자에게 최적화된 치료법을 제안하는 데 중요한 역할을 하고 있다.

본 장에서는 인공지능의 발전에 따라 생명과학 연구가 어떻게 변화하고 있는지에 대한 연구 현황 및 미래 전망을 살펴본다. 인공지능이 불러온 생명과학 연구의 변화가 사회 및 관련 산업 전반에 어떤 영향을 줄지 다양한 측면에서 논의하고자 하며, 지속 가능한 발전을 위해 인공지능 기술이 생명과학 연구에서 어떤 역할을 할 수 있는지를 살펴보겠다.

나 인공지능의 생명과학 연구 활용 현황

1) 생명과학 빅데이터 분석 및 예측 모델 개발

기존의 생명과학 연구는 주로 실험에 의해 이루어져 왔다. 그러나 인간 유전체 프로젝트 이후로 생명체에 대한 방대한 데이터가 축적되면서 이를 효과적으로 분석하기 위한 계산 기법들이 개발되기 시작했다. 최근 인공지능 기술이 고도로 발달함에 따라, 인공지능을 생명과학 빅데이터 분석에 접목하려는 시도가 많이 이루어지고 있다. 인공지능 기반의 생명과학 빅데이터 분석은 전통적인 계산 기법에 비해 훨씬 뛰어난 성능을 보이며, 그동안 불가능하다고 여겨졌던 생명과학 분야의 난제 해결에도 큰 도움을 주고 있다.

가장 대표적인 예는 단백질의 서열 및 구조 데이터를 기반으로 학습된 단백질 구조 예측 인공지능이다. 단백질은 생명체의 주요 구성 요소이며, 거의 모든 생명현상에 관여하는 중요한 생체분자이다. 단백질은 스무 종류의 아미노산이 연결된 생체고분자로, 아미노산 서열에 따라 서로 다른 삼차원 구조를 가지며 이에 따라 다양한 기능을 수행한다. 단백질의 기능을 이해하기 위해서는 그 삼차원 구조를 알아내는 것이 필수적인데, 단백질 구조 예측은 매우 어려운 문제로 여겨져 왔다. 하지만 인공지능은 이를 해결했다.

2021년, 구글 딥마인드의 알파폴드와 미국 워싱턴대학교의 로제타폴드 인공지능은 고정확도의 단백질 구조를 단 몇 분 만에 예측할 수 있음을 보여주었다(Jumper et al., 2021; Baek et al., 2021). 이는 생명과학 연구에 인공지능이 접목되었을 때의 폭발적인 효과를 보여준 첫 사례로, 이후 인공지능을 활용한 생명과학 연구가 활발히 촉진되는 계기가 되었다. 이러한 성과는 단백질 구조 예측 외에도 유전체학, 분자생물학, 생물정보학 등 다양한 분야에서 인공지능의 활용 가능성을 증명하고 있다.

2) 약물 개발 및 신약 발견

약물 개발은 시간과 비용이 많이 드는 복잡한 과정이다. 새로운 약물이 시장에 출시되기까지는 보통 10년 이상의 시간과 수십억 달러 이상의 비용이 소요되는데, 이러한 과정에서 많은 후보 물질이 초기 단계에서 실패하곤 한다. 그러나 인공지능은 분자 모델링과 시뮬레이션을 통해 약물-표적 상호작용을 예측하고 최적의 신약 후보 물질을 도출함으로써, 신약 개발의 초기 단계에서 발생하는 비용과 시간을 대폭 절감할 수 있다.

축적된 단백질의 분자 결합 데이터는 신약 후보 물질을 설계하는 인공지능 개발을 가능하게 했다. 인공지능을 활용해 약물 분자와 표적 단백질의 상호작용을 예측하고 설계하려는 연구가 활발히 진행되고

있는데, 인공지능 플랫폼을 활용하면 수백만 개의 화합물을 스크리닝하여 특정 질병에 대한 최적의 후보 물질을 빠르게 찾아낼 수 있다. 예를 들어, 엔비디아와 아스트라제네카가 공동 개발한 MegaMolBART는 대규모 화합물 라이브러리를 기반으로 신약 후보 물질을 더 효율적으로 도출하는 기술을 제공하고 있다(NVIDIA, 2023). MegaMolBART는 특히 신약 후보 물질 발굴 과정에서 엄청난 시간 절약과 정확성 향상을 가져와 연구자들이 보다 빠르게 치료제를 개발할 수 있도록 돕고 있다. 이는 기존의 실험 방법에 비해 훨씬 효율적이며, 신약 개발의 초기 단계(후보 물질 발굴 및 최적화)에서 소요되는 시간을 평균 5년에서 50% 이상 단축할 수 있다.

단백질 구조 예측 인공지능 개발로 촉진된 단백질 설계 연구는 전통적인 유기분자 약물에서 벗어나 새로운 형태의 치료제 개발 가능성을 열어주고 있다. 예를 들어, 워싱턴대학교는 단백질 설계 인공지능인 RFdiffusion을 개발하여 단백질 설계 성공률을 크게 향상시켰다(Watson et al., 2023). Generate Biomedicine 회사에서도 Chroma라는 단백질 설계 인공지능을 개발하여 공개했으며, 이를 활용한 단백질 치료제 개발 연구를 진행하고 있다(Ingraham et al., 2023). 또한 유전자 치료제로 활용될 수 있는 단백질 설계, CAR-T와 같은 세포 치료제의 표면 단백질 설계에 인공지능 기술을 접목하는 연구 등도 활발히 진행 중이다(Ichikawa et al., 2023; Qiu et al., 2024). 이러한 차세대 치료제 개발은 신약 개발의 패러다임을 바꾸고 있으며, 앞으로 더 많은 혁신이 있을 것으로 기대된다.

3) 질병 진단 및 맞춤형 치료

인공지능은 방대한 의료 데이터를 분석하고, 이미지 및 패턴 인식을 통해 질병을 조기에 정확하게 진단하는 데 활용되고 있다. 인공지능은 X-ray, MRI, CT 스캔 등의 다양한 의료 영상을 분석하여 질병을 진단할 수 있다. 예를 들어, Google Health는 인공지능을 이용해 유방 촬영 이미지를 분석하여 유방암을 조기에 진단할 수 있음을 보여주었다(김현석, 2020). 또한, 스탠포드대학교 연구진은 인공지능을 이용해 피부암 진단 모델을 개발하였으며, 이 모델은 피부 병변 이미지를 분석하여 피부암 여부를 진단하는 데 숙련된 피부과 전문의와 비슷한 정확도를 보인다(Esteva et al., 2017). 이러한 인공지능 시스템은 의료 인력과 장비가 부족한 지역에서도 유용하게 사용될 수 있다.

인공지능은 질병 진단뿐만 아니라 환자 개개인의 데이터를 기반으로 맞춤형 치료를 제공하는 데에도 활용된다. 맞춤형 치료는 환자의 유전자 데이터, 생활 습관, 환경 요인 등을 기반으로 최적의 치료 계획을 수립하는 접근법이다. 예를 들어, Foundation Medicine은 인공지능을 활용해 암 환자의 유전자 프로파일을 분석하고 각 환자에게 최적의 항암제를 추천한다(박으뜸, 2018). Tempus 역시 인공지능을 이용해 환자의 유전자 프로파일과 임상 데이터를 종합 분석하여 최적의 치료법을 제안하는 플랫폼을 구축하고 있다(남혜현, 2020). 이러한 맞춤형 치료 인공지능은 환자의 치료 성공률을 높이는 데 기여하고 있다.

다 인공지능 기반 생명과학 연구의 전망

1) 연구 효율성 및 정확성 향상

인공지능은 생명과학 연구의 효율성을 크게 높이고 있다. 이는 연구의 속도와 정확도를 향상하고, 연구비용을 절감하는 데 중요한 역할을 한다. 예를 들어, 유전체 데이터 분석의 경우 인공지능은 수 개월이 걸리던 작업을 며칠 또는 몇 시간 만에 완료할 수 있다. 인간 유전체 프로젝트는 수 년 간의 작업 끝에 완료되었지만, 이제는 인공지능을 활용하여 비슷한 규모의 데이터를 훨씬 빠르게 분석할 수 있다. 생명과학 연구 데이터는 방대하지만 불완전하고 노이즈가 많은 편인데, 인공지능은 불완전한 데이터를 처리하고 노이즈를 제거하며 유의미한 패턴을 추출하는 데에도 활용될 수 있다. 이는 연구 결과의 신뢰성을 높이는 데 중요한 역할을 한다. 인공지능과 생명과학 데이터 분석의 접목은 신약 개발, 질병 연구, 유전 연구 등 다양한 분야에서 연구 효율성을 극대화할 수 있게 한다.

인공지능 기반의 자동화된 실험 플랫폼은 실험 데이터를 실시간으로 분석하고 최적의 실험 조건을 자동으로 조정하여 연구 결과의 정확도를 높인다. 예를 들어, Berkeley Lab의 자동화 실험실은 인공지능과 로봇을 결합하여 생명과학 실험을 자동으로 수행하고, 데이터를 분석하여 최적의 실험 조건을 찾아낸다(Biron, 2023). 이러한 시스템은 연구자의 반복적이고 시간 소모적인 작업을 줄여주며, 연구 결과의 정확성과 일관성을 보장한다.

또한, 인공지능은 방대한 데이터를 빠르고 정확하게 처리할 수 있어 연구자들이 전략적 의사결정과 창의적인 문제 해결에 집중할 수 있게 한다. 이는 연구 생산성을 크게 향상하여 생명과학 연구의 혁신을 가속화할 것이다.

2) 새로운 생명과학 연구 분야의 활성화

인공지능은 생명과학 연구에서 새로운 분야를 개척하는 동시에 기존 연구 방법론을 혁신하고 있으며, 합성생물학, 유전자 편집, 바이오프린팅 등 첨단 생명과학 기술과 결합하여 연구의 혁신을 이끌고 있다.

합성생물학에서는 인공지능을 이용해 유전자 회로를 설계하고 최적화함으로써 새로운 기능을 부여할 수 있다. 예를 들어, Ginkgo Bioworks는 인공지능을 활용해 미생물의 유전자 회로를 설계하고 특정 화합물을 생산하도록 최적화하고 있는데, 이는 생물학적 생산 공정을 혁신하고 새로운 바이오 제품을 개발하는 데 기여하고 있다(Hayden, 2015). 유전자 편집 기술인 CRISPR-Cas9과 인공지능의 결합은 유전자 서

열을 신속하게 분석하고, 최적의 편집 위치를 예측하여 효율성을 높인다. 이는 유전자 치료와 농업 분야에서 혁신적인 응용 가능성을 제공한다. 바이오프린팅은 세포와 바이오 소재를 이용해 3차원 구조물을 인쇄하는 기술로, 인공 장기와 조직을 만드는 데 사용된다. 인공지능은 바이오프린팅 과정에서 최적의 세포 배치와 인쇄 조건을 찾아내어 인공 장기의 기능성을 높이고 있다(Ramesh et al., 2024). 예를 들어, 인공지능이 인쇄 과정에서 세포의 생존율을 높이고 조직의 구조와 기능을 최적화하는 데 중요한 역할을 하는 점 등을 꼽을 수 있다.

또한, 인공지능과 바이오테크의 융합은 신약 개발의 모든 단계에서 혁신을 일으키고 있다. Exscientia는 인공지능을 이용해 화합물 라이브러리를 스크리닝하고 신약 후보 물질을 최적화하여 기존 방법에 비해 15배 더 빠른 속도로 신약 후보 물질을 찾아내고 있다(Staines, 2020). 맞춤형 의료의 발전에도 인공지능은 크게 기여하고 있다. Tempus는 인공지능을 이용해 환자의 유전자 프로파일과 임상 데이터를 종합 분석하여 최적의 치료법을 제안하는 플랫폼을 구축하였는데, 이는 환자별로 효과적인 치료법을 제공하여 치료 성공률을 높이고 부작용을 최소화할 수 있다.

환경 보호와 생태계 연구에서도 인공지능은 중요한 역할을 한다. 예를 들어, Conservation Metrics는 인공지능을 이용해 멸종위기종의 소리를 분석하고 서식지를 모니터링한다(Klein et al., 2015). 인공지능은 방대한 환경 데이터를 분석하여 멸종위기종의 위치와 개체 수를 추정하고, 보호 조치를 계획하는 데 도움을 준다. 이러한 연구는 생물 다양성 보전과 생태계 보호에 기여한다.

결론적으로, 인공지능은 생명과학 연구에서 새로운 연구 분야를 개척하고, 기존 연구 방법론을 혁신하고 있다. 이는 연구 효율성과 정확성을 높이고 새로운 연구 가능성을 열어주며, 생명과학의 발전을 가속화하고 있다.

3) 지속 가능한 발전에의 기여

인공지능은 연구 과정에서 자원 소비를 줄이고 효율성을 극대화한다. 자동화 실험 시스템을 통해 실험 조건을 최적화하고 반복적인 실험을 최소화하여 자원의 낭비를 줄일 수 있다. 또한, 대규모 데이터를 분석하여 최적의 연구 전략을 제시함으로써 자원을 아끼는 효과를 창출하기도 한다. 이는 연구비용 절감과 자원의 효율적인 사용을 가능하게 한다.

환경 보호와 생태계 유지에서도 인공지능의 역할은 크다. 인공지능은 기후 모델링을 통해 기후 변화의 영향을 예측하고 이에 따른 대응 전략을 수립하는 데 도움을 준다. 멸종위기종의 서식지와 개체 수를 모니터링하고 보호 조치를 계획하는 데 활용된다. 이는 생물다양성을 유지하고 생태계를 보호하는 데 기여한다.

사회적 가치 증대 측면에서, 인공지능은 글로벌 보건 개선과 보편적 의료 서비스 확산에 기여한다. 인공지능 기반의 질병 진단 서비스 등은 의료 자원이 부족한 지역에서 효과적인 의료 서비스를 제공한다. 또한, 개발도상국의 연구자들이 첨단 생명과학 연구에 접근할 수 있도록 저비용, 고효율의 연구 방법론을 제공하여 글로벌 건강 격차를 줄인다.

결론적으로, 인공지능은 생명과학 연구와 지속 가능한 발전에 중요한 역할을 하며 인류의 건강과 지구 환경 보호에 크게 기여하고 있다. 앞으로도 인공지능 기술의 역할은 더욱 확대될 것이다.

4) 윤리적 및 사회적 고려사항

인공지능의 생명과학 연구 도입은 많은 긍정적인 변화를 가져오고 있지만, 이에 따른 윤리적 및 사회적 고려도 필요하다. 인공지능이 생명과학 연구 및 의료 분야에서 의사결정을 내릴 때, 그 과정의 투명성을 확보하는 것이 중요하다. 인공지능 시스템이 복잡한 알고리즘을 사용하여 결정을 내리는 경우, 그 과정이 불투명하면 결과에 대한 신뢰성이 떨어질 수 있다. 따라서 인공지능 모델의 의사결정 과정을 설명할 수 있게 만들고 투명성을 높이는 연구가 필요하다. 이는 '설명 가능한 인공지능' 기술의 발전을 통해 이루어질 수 있으며, 연구자와 의사가 인공지능의 판단 근거를 이해하고 신뢰할 수 있게 된다.

생명과학 연구에서 인공지능은 방대한 개인 데이터를 처리한다. 이러한 데이터의 보호와 윤리적 사용 역시 매우 중요한 문제이므로 연구 데이터의 익명성을 보장하고 개인의 프라이버시를 침해하지 않기 위해 데이터 보호 규정과 같은 법적 기준을 준수해야 한다. 또한, 연구자들은 데이터 수집 목적과 사용 방법에 대해 명확히 설명하고 데이터 제공자의 동의를 받아야 하며, 민감한 개인 정보는 엄격히 보호하고 연구 목적 외에는 사용하지 않아야 한다.

인공지능의 도입은 사회적 불평등을 심화시킬 수도 있다. 인공지능 기술은 자원이 풍부한 지역과 기관에서 더 쉽게 접근할 수 있으며, 이는 의료 및 연구의 격차를 더욱 벌릴 수 있다. 따라서 인공지능 기술의 혜택이 공정하게 분배될 수 있도록 정책적, 제도적 지원이 필요하다. 또한 인공지능 시스템이 편향된 데이터를 학습할 경우 그 결과도 편향될 수 있는데, 이는 생명과학 연구와 의료 분야에서 특정 집단에게 불리하게 작용할 수 있다. 예를 들어 특정 인종이나 성별에 대한 데이터가 충분하지 않으면, 인공지능의 진단이나 치료 제안이 공정하지 않을 수 있다. 이를 방지하기 위해 다양한 데이터를 학습시키고, 인공지능 시스템의 공정성을 지속적으로 모니터링하는 것이 중요하다. 최근에는 인공지능이 데이터 편향 문제를 해결하기 위해 다양한 인종, 성별, 연령대의 데이터를 포함하는 방식으로 학습데이터를 구성하고 있으며, 이를 통해 인공지능 모델의 공정성을 확보하고자 하는 노력이 계속되고 있다. 또한, 윤리적 인공지능 개발을 위해 각국 정부와 국제기구에서 지침과 규제를 강화하고 있다.

결론적으로, 인공지능의 생명과학 연구 도입은 많은 가능성을 열어주지만, 철저한 윤리적 및 사회적 고려를 전제해야 한다. 투명성과 책임성을 높이고 개인정보 보호와 데이터 윤리를 준수하며, 공정성을 확보하는 것이 중요하다. 이를 통해 인공지능 기술이 인류의 건강과 복지에 긍정적으로 기여할 수 있도록 해야 한다.

라 맺음말

현재 인공지능은 생명과학 연구에서 핵심적인 역할을 하고 있다. 생명과학 빅데이터 분석 및 예측 모델 개발을 통해 단백질 구조 예측, 유전체 데이터 분석, 약물-표적 상호작용 예측 등에서 뛰어난 성과를 보이고 있다. 단백질 구조 예측 인공지능은 고정확도의 예측을 가능하게 하여 단백질 기능 이해와 신약 개발에 기여하고 있으며, 약물 개발 과정에서도 인공지능의 분자 모델링 및 시뮬레이션을 통해 최적의 신약 후보 물질을 도출하여 시간과 비용을 절감하고 있다. 질병 진단 및 맞춤형 치료에서는 인공지능이 의료 데이터를 분석하여 조기진단과 최적화된 치료법 제공에 중요한 역할을 하고 있다.

미래에는 인공지능이 생명과학 연구의 혁신을 더욱 가속할 것이다. 연구의 효율성과 정확성을 높이는 자동화된 실험 플랫폼과 데이터 분석 기술은 연구자의 반복적 작업을 줄여주고 합성생물학, 유전자 편집, 바이오프린팅 등 새로운 연구 분야를 개척할 것이다. 또한, 인공지능은 지속 가능한 발전을 위해 자원 소비를 줄이고 환경 보호와 생태계 유지에 기여할 것이다. 인공지능을 통한 신약 개발 비용 절감 및 질병 진단 효율성 증대는 의료 자원이 부족한 지역에서도 고품질의 의료 서비스를 제공할 수 있게 하여 보건의료의 보편화에도 많은 영향을 미칠 것으로 보인다.

결론적으로, 인공지능은 생명과학 연구의 패러다임을 혁신적으로 변화시키고 있으며, 그 역할은 앞으로도 확대될 것이다. 인공지능 기술을 효과적으로 활용하여 생명과학 연구의 새로운 가능성을 탐구하고, 인류의 건강과 지구 환경 보호에 기여할 수 있는 방향으로 나아가야 한다. 이를 위해 지속적인 기술 발전과 함께 윤리적 고려와 사회적 책임을 병행하여 인공지능 기술이 긍정적인 영향을 미칠 수 있도록 해야 할 것이다.

참고문헌

김현석(2020). 구글 AI 의사보다 유방암 진단 정확, 한국경제.

남혜현(2020). 맞춤형 치료 추천 템푸스는 어떤 회사?, 바이라인네트워크.

박으뜸(2018). 맞춤치료에 대한 확신.. 빅파마로 보는 청사진, 메디파나뉴스.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., … & Baker, D.(2021). "Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network", Science, Vol.373 No.6557, pp. 871~876.

Biron, L.(2023). Meet the Autonomous Lab of the Future, LBNL News.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M. & Thrun, S.(2017). "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks", Nature, Vol.542 No.7639, pp. 115~118.

Hayden, E. C.(2015). "Tech Investors Bet on Synthetic Biology", Nature, Vol.527 No.7576, p. 19.

Ichikawa, D. M., Abdin, O., Alerasool, N., Kogenaru, M., Mueller, A. L., Wen, H., … & Noyes, M. B.(2023). "A Universal Deep-Learning Model for Zinc Finger Design Enables Transcription Factor Reprogramming", Nature Biotechnology, Vol.41 No.8, pp. 1117~1129.

Ingraham, J. B., Baranov, M., Costello, Z., Barber, K. W., Wang, W., Ismail, A., … & Grigoryan, G.(2023). "Illuminating Protein Space with a Programmable Generative Model", Nature, Vol.623 No.7989, pp. 1070~1078.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., … & Hassabis, D.(2021). "Highly Accurate Protein Structure Prediction with AlphaFold", Nature, Vol.596 No.7873, pp. 583~589.

Klein, D. J., McKown, M. W. & Tershy, B. R.(2015). Deep Learning for Large Scale Biodiversity Monitoring, In Bloomberg Data for Good Exchange Conference, Vol. 10.

NVIDIA(2023), Nvidia/MegaMolBART: A Deep Learning Model for Small Molecule Drug Discovery and Cheminformatics Based on Smiles, https://github.com/NVIDIA/MegaMolBART.

Qiu, S., Chen, J., Wu, T., Li, L., Wang, G., Wu, H., ... & Wang, H.(2024). "CAR-Toner: an AI-driven Approach for CAR Tonic Signaling Prediction and Optimization", Cell Research, Vol.34 No.5, pp. 386~388.

Ramesh, S., Deep, A., Tamayol, A., Kamaraj, A., Mahajan, C. & Madihally, S.(2024). "Advancing 3D Bioprinting Through Machine Learning and Artificial Intelligence", Bioprinting, Vol.38, e00331.

Staines R.(2020). Exscientia Claims World First as AI-created Drug Enters Clinic, Pharmaphorum.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., … & Baker, D.(2023). "De Novo Design of Protein Structure and Function with RFdiffusion", Nature, Vol.620 No.7976, pp. 1089~1100.

KAST

8

공정한 인공지능을 위한 기술과 발전 전망

조성배

연세대학교 컴퓨터과학과 교수

가 들어가는 말

인공지능(AI)은 우리 삶에 다양하게 녹아들어 의료, 금융, 교육, 법률 등 여러 분야에서 효율성과 정확성을 높이고 있다. 다양한 실현 방법 중에서 기계학습(Machine Learning, ML)은 수많은 데이터로부터 자동으로 모형을 구축할 수 있을 뿐만 아니라, 최근에는 인간을 뛰어넘는 성능을 내기도 해서 크게 각광받고 있다. 기계가 학습한다는 것은 컴퓨터가 경험으로부터 성능 척도를 향상시켜서 어떤 작업을 수행하는 것이라고 정의한다. 예를 들어 x와 y라는 속성을 갖는 데이터 <x, y>에 대해서 <1, 5>, <2, 7>, <3, 9>, <4, 11>, <5, 13>을 학습한다는 것은 y=F(x)가 되는 모형 F를 결정하는 것이다. F를 단순하게 y=ax+b로 하면, 학습은 데이터의 x와 y를 넣어서 성립되는 a와 b를 결정하면 된다. 이 경우에는 단순한 1차 방정식이기 때문에 학습한 모형은 F(x)=2x+3이 되는데, 나중에 x=100인 데이터의 y가 무엇인지 알아내는 데 사용된다. 기계학습으로 만든 모형으로 새로운 데이터의 x에 대해서 y의 값을 알아내는 것이다.

최근에 인공지능이 다양하게 사용되면서 공정성에 대한 문제가 대두되고 있다. 공정성은 인공지능이 특정 집단에 대한 편향이나 차별 없이 모든 사용자에게 공정한 결과를 제공하는 것인데, 대량의 데이터로 학습된 인공지능이 편향된 의사결정을 하는 사례가 빈번하게 발생하고 있다. 일례로, 아마존의 인공지능 채용 시스템이 주로 남성 지원자가 많았던 과거 데이터를 기반으로 학습되어 남성 지원자를 우대하는 평가를 하였다. 또한, 탐사보도 전문매체 프로퍼블리카는 범죄 전과자의 얼굴 이미지로부터 재범 가능성을 예측하는 인공지능을 개발했는데, 흑인의 재범률을 백인보다 훨씬 더 높게 예측하는 편향을 보였다. 의료 및 법률 시스템에서도 특정 인종, 성별에 따라 다른 결과를 내서 특정 집단을 불공정하게 대할 가능성이 있다. 챗GPT에서도 성별 관련 대명사와 직업 관련 단어를 활용할 때 특정 직업을 특정 성별과 연관 짓는 편향이 보인다.

인공지능의 공정성은 사회 정의와 관련이 깊다. 공정한 인공지능은 사회적 신뢰를 구축하고, 다양한 집단이 동등한 기회를 갖도록 한다. 반면 불공정한 인공지능은 사회적 불평등을 심화시키고 특정 집단에 대한 차별을 강화할 수 있다. 이 장에서는 인공지능의 공정성을 측정하고 편향을 줄이는 방법에 대해서 소개하고, 다양한 국가와 기관에서 인공지능의 공정성을 보장하기 위해 어떤 시도를 하고 있는지 알아보며 향후 발전 전망에 대해서 논의하고자 한다.

나 공정성 척도와 데이터셋

1) 공정성 척도

인구학적 동등성(demographic parity)

모든 집단에서 긍정적인 예측 결과의 비율은 동일해야 한다. 예를 들어, 채용 과정에서 특정 성별이나 인종에 상관없이 대출 승인 비율이 동일해야 하지만, 경우에 따라서 불공정한 결과를 초래할 수 있다. 대출 승인 시스템에서 인구학적 동등성을 적용하면, 특정 인종이 다른 인종보다 더 낮은 신용 점수를 갖고 있음에도 불구하고 동일한 승인율을 갖도록 조정될 수 있다.

기회 균등(equal opportunity)

실제 긍정 사례에 대해 각 집단은 동일한 참의 비율을 가져야 한다. 질병 진단 인공지능은 성별에 관계없이 일정한 정확도로 질병을 진단해야 사람들이 공정하게 혜택을 받을 수 있다. 이를테면, 의료 분야의 암진단 인공지능은 남성과 여성 모두에게 같은 정확도로 진단해야 한다. 만약 여성 환자에게만 진단 정확도가 떨어진다면, 이는 기회 균등을 위반하는 것이다.

예측 동등성(predictive parity)

각 집단에서 예측 결과의 정확도는 동일해야 한다. 이는 예측이 정확하게 이루어지고 있는지, 예측의 품질이 집단 간에 일관성 있게 유지되고 있는지를 평가한다. 재범 가능성 예측 인공지능은 인종에 관계없이 동일한 정확도로 범죄 가능성을 예측해야 한다. 만약 특정 인종만 예측 정확도가 높다면, 예측 동등성이 지켜지지 않은 것이다.

동등 가능성(equalized odds)

모든 집단에서 참과 거짓의 비율은 같아야 한다. 이는 모형의 예측이 모든 집단에 대해 동일하게 잘못되거나, 동일하게 올바르다는 것을 의미한다. 예를 들면, 신용 평가 인공지능에서는 모든 인종의 대출 승인 비율이 일정하게 잘못되거나 올바르게 되어야 한다. 이는 모형이 특정 집단에 대해 더 많은 오차를 범하지 않도록 한다.

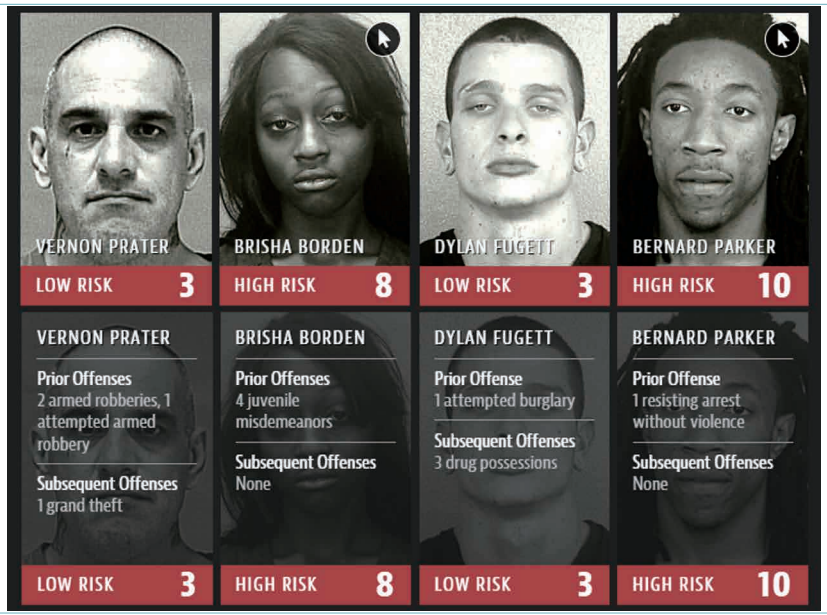
2) 공정성 평가 데이터셋

COMPAS 데이터셋

COMPAS(Correctional Offender Management Profiling for Alternative Sanctions)는 재범 가능성을 예측하

기 위해 미국의 여러 주에서 사용되는 시스템이다. 이 데이터셋은 인종에 따른 예측 편향으로 인해 큰 논란을 불러일으켰는데, <그림. 24>에서처럼 흑인 피고인에 대해 더 높은 재범 가능성을 예측하는 경향이 있음이 확인됐다. 흑인 피고인은 백인 피고인보다 실제 재범 가능성이 낮음에도 불구하고 편향된 예측을 했다.

그림. 24 COMPAS 데이터셋에 나타나는 편향성



출처: Propublica(2024).

Adult Income 데이터셋

이것은 미국 인구 조사 데이터를 기반으로 개인의 소득 수준을 예측하는 데 사용된다. 여기에는 성별과 인종에 따른 편향이 있는데, 여성이 동일한 자격 요건을 갖고 있음에도 불구하고 남성보다 더 낮은 소득을 받는다고 예측되는 경향이 있다. 이는 성별에 따른 소득 편향을 나타내며, 인공지능이 이러한 편향을 학습하여 예측에 반영할 수 있다.

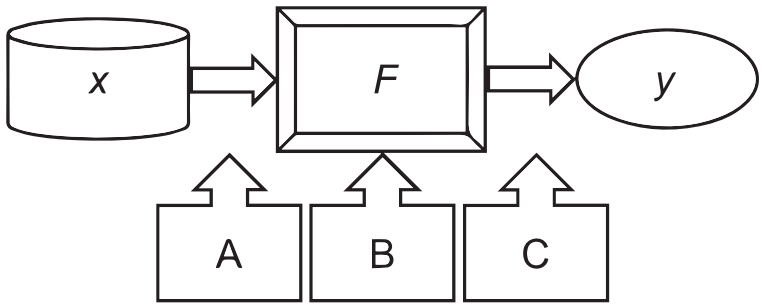
German Credit 데이터셋

이것은 독일의 신용 대출 데이터로, 대출 승인 여부를 예측하는 데 사용된다. 이 데이터셋은 연령, 성별, 직업 등 다양한 특성을 포함하고 있어 편향 분석에 유용한데, 고령자와 젊은 층 간의 대출 승인율이 다르게 나타나는 경우가 종종 있다. 만약 고령자가 불리하게 예측된다면 연령에 따른 편향을 식별하고 수정하는 방법을 연구하는 데 도움이 된다.

다 인공지능의 편향제거 기술

인공지능의 공정성을 보장하기 위해 다양한 편향제거 기술이 개발되고 있는데, <그림. 25>와 같이 전처리(pre-processing), 중처리(in-processing), 후처리(post-processing)의 3가지로 나눌 수 있다. 이러한 기술은 데이터 수집과 처리(A), 모형 학습(B), 그리고 결과 처리(C)의 각 단계에서 편향을 제거하고자 한다.

그림. 25 편향제거 방법의 3가지 접근 방식



출처: 연세대학교 조성배(2024).

1) 전처리 접근 방식

모형학습 전에 데이터 자체에서 편향을 제거하거나 줄이는 방법으로서, 데이터 자체가 편향되지 않도록 하여 데이터의 공정성을 확보한다.

- 데이터 샘플링은 특정 그룹이 과소 혹은 과다 대표되지 않도록 다양한 배경, 인종, 성별, 연령의 데이터를 포함하는 학습데이터를 샘플링하는 것이다.
- 데이터 증강은 소수 집단의 데이터에 노이즈를 추가하거나 변형하는 등 데이터를 증강하여 학습데이터의 균형을 맞추는 것을 의미한다.
- 리샘플링은 다수 집단의 데이터를 줄이거나 소수 집단의 데이터를 늘려서 균형을 맞추는 방법인데, 일반적으로 다수 집단의 데이터를 일부 제거하여 데이터의 균형을 맞춘다.
- 데이터 정규화는 소득이나 연령과 같은 연속형 변수를 표준화하는 등 데이터의 특정 속성을 정규화하여 편향을 줄이는 방법이다.
- 속성 제거는 성별, 인종 등 편향을 유발할 수 있는 민감한 정보 및 속성을 제거하는 방식을 의미한다.

2) 중처리 접근 방식

모형의 학습 과정에서 편향을 제거하거나 줄이는 방법으로, 알고리즘 자체에 공정성을 내재화하여 모형이 공정하게 학습되도록 한다.

- 모형 학습 과정에서 편향을 보정하는 알고리즘은 모델의 손실값과 가중치를 수정하는 방향으로 진행된다. 이는 모형이 학습되는 동안 특정 그룹에 대한 편향을 줄이도록 조정하는 방법이다.
- 잘못된 예측에 대해 비용을 부여하여 편향을 줄이는 방법도 있다. 예를 들어, 소수 집단에 대해 잘못 예측할 경우 높은 비용을 부여하여 모형이 더 신중하게 학습되도록 한다.
- 제약기반 학습은 모형학습 시 모형의 예측 결과가 모든 그룹에서 균형 있게 나오도록 공정성 제약 조건을 추가하여 편향을 줄인다.
- 최적화 알고리즘은 공정성을 고려하여 모형을 학습함으로써 모형이 성능과 공정성을 동시에 만족시키도록 학습하는 방법이다.

3) 후처리 접근 방식

모형이 생성한 예측 결과를 후처리하여 편향을 제거하거나 줄이는 것으로, 이미 학습된 모형의 결과가 공정하도록 보정한다.

- thresholding은 예측 결과의 임계값을 조정하여 편향을 줄이는 방법으로 특정 그룹이 불리한 예측을 받지 않도록 공정성을 확보한다.
- 결과기반 변환 방법은 예측 결과를 변환하여 편향을 줄이는 것으로, 모형의 예측값을 직접 수정하여 특정 그룹에 불리하지 않도록 한다.
- calibration은 모형의 예측 확률을 조정하여 편향을 줄이는 방법으로 예측 확률이 실제 결과와 일치하도록 조정하여 공정성을 높인다.
- 결과기반 제약 조건 최적화 방법은 예측 결과에 제약 조건을 적용하여 편향을 줄이는 것으로, 결과의 공정성을 보장하기 위해 제약 조건을 최적화한다.

4) 대형언어모델의 편향제거

최근에 대형언어모델(LLM)은 방대한 데이터를 학습하여 다양한 응용 분야에서 탁월한 성능을 보여주고 있으나 학습데이터에 내재된 편향을 학습하여 불공정한 결과를 생성할 수도 있다. 이에, LLM의 편향을 줄이는 다양한 기술이 개발되고 있다.

- 데이터 증강은 학습데이터에서 특정 속성을 바꾸며 균형을 맞추는 방법으로서, '남성은 훌륭한 프로그래머다'를 '여성은 훌륭한 프로그래머다'로 바꾸는 식으로 수행된다.
- 프롬프트 튜닝은 사용자가 제공하는 프롬프트를 개선해 편향을 줄이는 방법이다.
- 모형 구조에 편향 감소를 위한 모듈을 추가하여 편향을 제거하기도 한다.
- 모형 편집은 모형의 특정 부분을 변경하여 편향을 줄이는 것으로서, 모형의 특정 영역에서 행동을 효율적으로 조정하면서 다른 입력에 영향을 주지 않도록 한다.
- 결과 수정은 모형이 생성하는 텍스트의 품질을 조정하는 방법이다.
- 추론 단계에서 특정 입력에 대한 모형의 출력을 CoT 등으로 수정해 편향을 완화하기도 한다.

라 주요국의 동향

1) 미국

미국은 인공지능 기술의 선도국답게 인공지능의 공정성과 윤리에 관한 활발한 연구와 정책을 개발하고 있다. 여러 정부 기관과 민간단체가 협력하여 인공지능의 공정성을 보장하기 위한 다양한 노력을 기울이고 있다. 국립표준기술연구소(NIST)는 인공지능의 공정성 평가를 위한 기준을 마련하고 공정성, 투명성, 책임성을 보장하기 위해 다양한 연구를 수행하고 있으며, 이를 통해 인공지능의 신뢰성을 높이려 하고 있다.

미국 정부는 인공지능 연구와 개발을 촉진하기 위해 백악관에 AI 이니셔티브를 출범시켰다. 인공지능의 윤리적 사용과 공정성을 강조하며, 공정성 관련 연구에 대한 지원을 확대하고 있다. 캘리포니아주는 인공지능의 편향을 줄이기 위한 법안을 추진하고 있는데, 그 내용은 인공지능이 특정 집단에 대한 편향을 일으키지 않도록 하는 규제를 포함하고 기업이 이러한 규제를 준수하도록 하는 것이다.

구글은 인공지능 공정성을 위한 다양한 연구를 진행하고 있는데, 'What-If Tool'은 머신러닝 모델의 공정성을 시각적으로 분석할 수 있는 도구로, 모델의 편향을 식별하고 수정하는 데 도움을 준다. IBM은 인공지능 공정성을 위한 오픈소스 툴킷인 AI Fairness 360을 개발했는데, 이는 데이터 전처리, 모형 학습, 후처리 단계에서 편향을 줄이는 다양한 알고리즘과 도구를 제공한다.

2) EU

EU는 인공지능의 공정성과 윤리를 보장하기 위해 가장 엄격한 규제를 마련하고 있다. EU의 접근 방식은 주로 법적 규제와 윤리적 가이드라인을 통해 인공지능의 책임성을 강화하는 것이다. 2021년에 인공지능 규제 법안을 발표하여 인공지능이 비차별적이고 투명하게 작동될 것을 요구하였으며, 이를 위반할 경우 엄격한 제재를 가한다. 이 법안은 인공지능의 공정성, 투명성, 안전성을 보장하는 것을 목표로 한다.

GDPR은 데이터 보호와 관련된 강력한 규제를 포함하고 있으며, 인공지능이 개인의 데이터를 공정하고 투명하게 처리하도록 요구한다. 이는 인공지능이 편향을 일으키지 않도록 하는 중요한 법적 장치이다. AI4EU는 유럽의 인공지능 연구자와 기업들이 협력하여 인공지능의 공정성과 윤리를 연구하고 개발하는 플랫폼으로서 인공지능의 투명성과 공정성을 높이기 위한 다양한 도구와 리소스를 제공한다.

3) 아시아

아시아도 인공지능 발전에 큰 관심을 기울이고 있으며, 여러 국가가 인공지능의 공정성을 보장하기 위한 정책과 연구를 추진하고 있다. 중국은 인공지능 선도국으로서 인공지능의 공정성을 보장하기 위한 연구와 정책을 강화하고 있다. 중국 정부는 인공지능의 개발과 활용을 촉진하면서도, 공정성과 윤리를 강조하는 정책을 추진하고 있다. 또한, 여러 대학과 연구기관에서 인공지능의 공정성을 연구하고, 편향을 줄이기 위한 다양한 알고리즘과 기법을 개발하고 있다.

일본 정부는 인공지능의 윤리적 사용을 보장하기 위해 AI 윤리 가이드라인을 제정하여 인공지능이 인권을 존중하며 공정하게 작동될 것을 요구하고 기업과 연구자들이 이를 준수하도록 장려한다. 또한, 공정성을 포함한 인공지능 연구개발을 지원하기 위해 다양한 프로그램을 운영하고 있다. 일본의 여러 연구기관과 대학에서도 인공지능의 공정성을 연구하는 프로젝트를 진행하고 있다.

우리 정부도 인공지능의 공정성을 평가하기 위한 기준을 마련하고 있다. 이 기준은 인공지능이 공정하게 작동되는지 평가하고, 편향을 식별하여 수정하는 방법을 제시한다. 또한, 국가 AI 전략을 통해 인공지능의 윤리적 사용과 공정성을 강조하고 있다. 이 전략은 인공지능이 사회에 긍정적인 영향을 미칠 수 있도록 다양한 정책과 규제를 포함하고 있다.

마 발전 전망

1) 기술적 측면

인공지능의 공정성을 확보하기 위한 기술은 지속적으로 발전하고 있다. 연구자들은 보다 정교한 편향 제거 알고리즘을 개발하고 있으며, 인공지능의 예측 정확도를 유지하면서도 공정성을 높이는 방향으로 나아가고 있다. 이에 따라, 새로운 딥러닝 기법이나 강화학습 알고리즘은 공정성을 내재화한 형태로 발전할 것으로 보인다.

또한, 인공지능이 실시간으로 환경과 데이터를 학습하면서 편향을 줄이는 적응형 알고리즘이 개발될 것이다. 이러한 알고리즘은 변화하는 데이터 패턴에 맞춰 스스로 조정하며 공정성을 유지할 수 있다. 하나의 모형이 여러 과제를 동시에 학습하면서 공정성을 유지하는 멀티태스크 학습 기법도 주목받고 있다. 이는 모형이 다양한 상황과 데이터를 균형 있게 처리할 수 있도록 돕는다.

인공지능 모형의 해석 가능성과 투명성을 높이는 기술도 발전하고 있다. 이는 인공지능의 결정 과정을 명확히 이해하고, 잠재적인 편향을 식별하여 수정하는 데 중요한 역할을 한다. 설명 가능한 AI(Explainable AI, XAI) 기술은 모형의 예측 결과를 이해할 수 있게 설명해주는 것으로, 사용자가 모형의 작동 방식을 이해하고 신뢰할 수 있도록 돕는다. 시각화 도구를 통해 모형의 의사결정 과정을 시각적으로 표현하거나 문장으로 설명하는 사례 등이 이에 해당한다. 모형의 공정성을 평가하고 편향을 식별하기 위한 감사 도구 또한 개발되고 있는데, 이는 모형의 모든 결정 과정을 기록하고 분석하여 공정성 문제를 사전에 방지할 수 있도록 한다.

데이터의 공정성을 보장하기 위한 데이터 관리 기술도 중요하다. 데이터 관리의 목표는 데이터 수집, 정제, 저장, 사용의 전 과정에서 공정성을 유지하는 것인데, 이를 위해 다양한 배경을 대표하는 데이터를 수집하고 관리하는 기술도 발전하고 있다. 이는 인공지능이 특정 집단에 편향되지 않도록 하는 데 중요한 역할을 한다. 데이터 사용의 윤리성을 보장하기 위한 방법과 도구 또한 개발되고 있는데, 데이터 사용이 법적, 윤리적 기준을 준수하도록 돕는 역할을 한다.

2) 정책적 측면

전 세계적으로 인공지능의 공정성을 보장하기 위한 정책과 규제가 강화될 전망이다. 이는 각국의 법적 기준을 국제적으로 표준화하여, 글로벌 기업들이 일관된 규제를 준수할 수 있도록 하는 방향으로 발전하고 있다. 여러 국가들이 협력하여 인공지능의 공정성에 관한 국제적인 가이드라인과 표준을 마련하고 있는데, 이는 글로벌 시장에서 일관된 규제를 적용할 수 있게 한다. 각국의 규제 기관은 인공지능의 공정성을 평가하고 보장하기 위한 구체적인 기준을 마련하고, 이를 준수하지 않는 기업에 대해 엄격한 제재를 가할 것이다.

기업들은 인공지능의 공정성을 보장하기 위해 더 많은 책임을 지게 될 것이다. 이는 기업이 자율적으로 공정성을 유지하도록 하는 내부 규제와, 외부의 법적 규제를 동시에 적용하는 방식으로 발전하게 될 것으로 보인다. 기업들은 공정성, 투명성, 책임성을 포함한 자체적인 윤리 기준을 마련하고, 이를 준수하기 위한 내부 정책을 강화할 것이다. 또한 정부의 규제를 준수하기 위해 기업 내 공정성 평가 도구와 프로세스를 도입하여 정기적으로 공정성 평가를 실시하고 결과를 보고해야 할 것이다.

바 맺음말

인공지능의 공정성에 대한 사회적 인식이 높아지고 있다. 공정성에 대한 요구가 더욱 강해지면서, 기업과 연구자들은 공정성을 최우선 과제로 삼아야 할 것이다. 대학과 기업에서는 인공지능 공정성에 관한 교육 프로그램을 강화할 것으로 보이는데, 이는 개발자와 사용자가 공정성 문제를 인식하고 해결할 수 있도록 돕는 역할을 할 것이다. 정부와 민간단체는 인공지능의 공정성에 대해 공공의 인식을 높이기 위한 홍보 활동을 강화하여 일반 대중이 인공지능의 공정성 문제를 이해하고 이를 개선하기 위한 노력에 공감할 수 있도록 힘써야 한다.

사회와 민간의 참여를 통해 인공지능의 공정성을 보장하는 노력도 중요하다. 다양한 이해관계자들이 함께 협력한다면, 인공지능의 공정성을 높이는 방향으로 발전할 것이다. 무엇보다도, 인공지능 개발 과정에 시민들이 참여할 수 있는 플랫폼을 마련하여 다양한 목소리를 반영하고 인공지능이 공정하게 개발될 수 있도록 해야 한다. 이와 같이 기업, 정부, 학계, 시민단체 등 다각적인 협력을 통해 이제는 인공지능의 공정성을 높이기 위한 공동의 노력을 기울여야 할 시점이다.

참고문헌

Caton, S. & Haas, C.(2024). "Fairness in Machine Learning: a Survey", ACM Computing Surveys, vol.56 No.166, pp.1~38.

Chen, J., dong, H., Wang, X., Feng, F., Wang, M. & He, X.(2023). "Bias and Debias in Recommender System: A Survey and Future Directions", ACM Transactions on Information Systems, vol. 41 No.67, pp.1~39.

Chu, Z., Wang, Z. & Zhang, W.(2024). "Fairness in Large Language Models: A Taxonomic Survey", arXiv preprint, arXiv:2404.01349.

Mehrabi, N., Morstatter, F., Saxena, N. & Lerman, r.(2021). "A Survey on Bias and Fairness in Machine Learning", ACM Computing Surveys, vol. 54, No.6 pp.1~35.

Quy, T., Roy, A., Iosifidis, V., Zhang, W. & Ntoutsi, E.(2022). "A survey on Datasets for Fairness-Aware Machine Learning", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol.12, pp. 1~59.

Xivuri, K. & Twinomurinzi, H.(2021). "A Systematic Review of Fairness in Artificial Intelligence Algorithms", Responsible AI and Analytics for an Ethical and Inclusive Digitized Society, pp. 271~284.

Propublica(n.d.). Retrieved September 10, 2024 from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

KAST

9

지속 가능한 AI와의 공존을 위한 윤리와 규제

김명주
국제인공지능윤리협회 회장

가 들어가는 말

30년 전 인터넷이 비영리 인프라에서 상업화 기술로 전환되고 급속히 대중화됨에 따른 인류의 소통 혁신은 예상을 초월했다. 그러나, 혁신이 불러온 편익과 함께 심각한 부작용과 역기능이 대두되기 시작했다. 이를 최소화하기 위해 세계 각지에서는 정보통신 윤리, 사이버 윤리, 인터넷 윤리 등 각종 윤리 운동을 전개했다. 그리고 윤리만으로는 해결할 수 없는 문제들을 규제와 법률로 보완해 왔다.

오늘날 인류는 인공지능(AI)이라는 또 다른 신기술을 맞이하고 있다. 인공지능은 인류 최대의 혁신 신기술이라고 인정받을 만큼 그 잠재력이 무궁하며 심지어 신인류로의 대전환을 위한 핵심 기술이라고도 평가되지만, 여기에도 잠재적인 위험은 헤아릴 수 없을 만큼 많다. 어떤 전문가들은 인공지능의 잠재적 위험을 감안하여 인공지능을 인류의 마지막 기술이라고 예견하기도 한다. 인공지능은 혁신성이 매우 커서 일단 사회적으로 수용되면 부작용과 역기능이 아무리 심각해도 이전으로 되돌아갈 수 없는 '비가역적 기술'이다. 따라서 처음부터 윤리적 접근이 필요하며 이를 감안한 신중한 수용이 필수적이다. 본 장에서는 인류가 앞으로 인공지능과 지속적인 공존 관계를 유지하기 위하여 고려해야 할 윤리와 규제를 살펴본다.

나 AI의 양면성

1) 디지털 신기술과 혁신, 그리고 양면성

'혁신'은 이전과 다른 새로움을 통해서 긍정적 변화를 이끌고 가치를 창출함으로써 사회 전반에 걸쳐 발전을 이루어 내는 것이라고 정의할 수 있다. 혁신은 인류의 모든 영역에 걸쳐 가능하지만, 특히 '기술 혁

신'은 역사를 돌아볼 때 인류에게 가장 큰 영향력을 끼쳐왔다고 해도 과언이 아니다. 기술 혁신은 사회 전반에 걸친 '발전'의 차원을 뛰어넘어 '혁명'이라 불릴 만한 시대적 전환기를 만들어 내기도 했다. 최근에는 AI를 위시하여 일련의 새로운 디지털 기술들이 과거 그 어느 때보다도 강력한 변화를 일으키면서 새로운 혁신을 이루는 중이다.

기술 혁신은 일상의 편의성 및 업무의 효율성을 높이는 등 다양한 방식으로 우리 사회에 긍정적인 영향을 미쳐왔다. 최근 연구에 따르면 생성형 AI '챗GPT'를 활용할 경우 문서 작업의 효율성이 40%, 출력 품질은 18% 향상되며 소프트웨어 개발 시간은 40% 이상 단축된다고 알려졌다. 뿐만 아니라, 기술 혁신은 경제 성장을 촉진하며 소득 수준도 높여준다. 건강 증진과 수명 연장이라는 인간의 꿈을 현실화하고 자아실현과 행복 추구를 향한 기초 수단인 교육과 학습의 기회를 넓혀주며 고용의 기회도 새롭게 제공한다. 아울러 개인 간 소통과 연결성을 증진함으로써 사회적 존재로서의 특장점을 기반으로 새로운 기회와 가치를 만들어 낸다. 사회에 미칠 긍정적 영향에 대한 이러한 기대는 공학자와 과학자 등 기술 전문가들이 기술 혁신을 꾀하며 열정과 시간을 쏟는 동안 중요한 동인(動因)으로 작용해 왔다.

그러나, 어떤 기술이든지 정도의 차이는 있겠지만 양면성이 존재한다. 기술의 순기능 곁에는 항상 역기능과 부작용이 병존한다. 기술 자체의 양면성 못지않게 기술 혁신이 일으키는 다양한 변화에도 양면성이 존재하며, 이에 따라 기술 혁신은 긍정적인 영향과 함께 사회적, 경제적, 윤리적 숙제를 꾸준히 동반해 왔다. 2차 세계대전을 종식할 수 있는 유일한 해결책으로 원자폭탄 개발을 추진했던 오펜하이머 연구소장은 이후 적극적인 핵무기 반대론자가 되었고, 딥러닝을 만든 AI 대부 제프리 힌튼 교수는 생성형 AI의 글로벌 경쟁 과정을 지켜보면서 지난 시간에 대한 후회와 미래에 대한 걱정이 담긴 인터뷰를 쏟아 냈다. 제프리 힌튼 교수는 "인공지능으로 인하여 앞으로 인류는 엄청난 이득을 얻게 될 것이다. 그러나 인공지능으로 인해 파생되는 문제점들을 해결하기 위해서는 얻은 이득의 두 배 이상을 쏟아 부어야 할 것이다."라며 여러 미디어와 인터뷰했다.

2) AI 분야의 부머와 두머

2023년 11월 17일 '챗GPT의 아버지'라 불리던 샘 알트만 대표의 전격 해임을 결정했던 오픈AI 이사회가 주목을 받았다. 전체 이사 6명 중에서 일리야 수츠케버, 아담 디안젤로, 타샤 매컬리, 헬렌 토너는 범용인공지능(AGI)의 윤리적 중요성을 강조하는 두머(doomer)에 속한다고 볼 수 있다. 두머는 미래의 AGI가 궁극적으로 인간에게 위해를 가하거나 심지어 인류를 파멸시킬 수 있다(doom)는 가능성을 배제하지 않으며, 이를 예방하기 위해서 인간이 AI를 지속적으로 모니터링하고 위험 발생 시 즉각적인 개입이 가능해야 하며, 기술은 물론 윤리와 규제, 법 등 필요한 모든 조치가 신기술 개발 첫 단계부터 병행되어야 한다는 입장이다.

반면에 오픈AI의 대표이사 샘 알트만과 이사회 의장 그렉 브록만은 아직 인공지능이 윤리를 앞서 내세울 만큼 우려스럽지는 않으며 우선 인공지능을 성장시키는 것이 중요함을 강조하는 부머(boomer)에 속한다. 부머는 미래 AGI가 인류를 배신하는 일은 결코 일어나지 않을 것이라고 믿는데, 이는 신뢰할 수 있는 기술의 개발에 대한 기술자 스스로의 자신감 표현이라고도 평가된다. 이들은 오히려 AGI 기술을 개발할수록 인류에게 큰 도움이 될 것이라는 확신을 가지고 있으므로, 인류가 기술 발전과 이를 통한 사회적 변화(boom)에 더 노력을 기울여야 한다는 입장이다.

두머가 과반수를 이루었던 오픈AI 이사회는 대표이사 샘 알트만과 이사회 의장 그렉 브록만이 '안전하고 윤리적인 AGI 개발'이라는 오픈AI의 창립 취지를 변질시키고 있다고 판단하여 각각 해임과 해촉을 명령했다. 그런데 750여 명의 직원 중에서 700명이 사표를 냈고 이들을 모두 마이크로소프트가 받아들이겠다고 하면서, 오픈AI에 지금까지 투자됐던 자금이 전량 회수당할 위기에 처하자 5일 만에 이사회는 인사 결정을 철회했고 이 일은 단순한 해프닝으로 끝났다. 이는 미래의 AGI 기술과 그 영향력을 바라보는 부머와 두머의 갈등이 세상에 널리 조명된 계기가 되었고, 기술 혁신이 기술자들의 시선에서마저도 언제나 긍정적이지만은 않음을 보여주었다. 정리하자면, 부머와 두머는 기술 혁신에 있어서 '파괴적 혁신가(destructive innovator)'와 '책임 있는 혁신가(responsible innovator)'에 각각 비견될 수 있다.

2024년 5월 14일 오픈AI가 챗GPT의 기반모델을 챗GPT-4o로 업그레이드하면서 내부 윤리팀이 슈퍼얼라인먼트(Super-Alignment)팀을 전격 해체했다. 오픈AI의 슈퍼얼라인먼트팀장이었던 얀 라이케는 "지난 수 년 간 AI 안전성은 잘나가는 제품보다 뒷전으로 밀려났다."며 윤리를 가볍게 여기는 오픈AI를 비난했고 클로드(Claude) AI를 만든 안드로픽(Anthropic)에 합류하여 슈퍼얼라인먼트팀을 이끌게 되었다. 오픈AI의 초창기 멤버로서 기술개발 총책을 맡았던 일리야 수츠케버도 같은 시기에 회사를 떠났다. 그 후 한 달 만인 6월 20일 '안전한 초지능'을 목표로 하는 AI 스타트업 'SSI(Safe Super Intelligence)'를 만들었다. 부머에게 밀려 나온 두머가 안전한 미래 인공지능을 위한 기업 활동을 다시 시작한 셈이다. 이처럼 AI 기술과 산업 분야에서는 안전한 인공지능에 대한 서로 다른 시각이 존재하며, 그 사이에서 더욱 발전된 윤리와 규제가 최근 이슈로 부각되고 있다.

다 AI 윤리 원칙

1) 법적 정의에서 바라본 AI 윤리

2020년 EU 집행위원회(EC)에서 제안했던 인공지능법(AI Act) 초안에는 'AI 시스템'이 다음과 같이 '기술' 중심으로 다르게 정의되었다.

'AI 시스템이란 부속서 I에 나열된 하나 이상의 기법과 접근 방식을 사용하여 개발된 소프트웨어로, 주어진 인간 정의 목표 집합에 대해 상호작용하는 환경에 영향을 미치는 콘텐츠, 예측, 추천 또는 결정을 생성할 수 있는 소프트웨어를 의미한다.'

부속서 I

(a) 지도학습, 비지도학습 및 강화학습을 포함하여 다양한 방법(딥러닝 포함)을 사용하는 기계학습(ML) 접근 방식

(b) 지식 표현, 귀납적(논리) 프로그래밍, 지식 베이스, 추론 및 연역 엔진, (상징적)추론 및 전문가 시스템을 포함한 논리 및 지식 기반 접근 방식

(c) 통계적 접근 방식, 베이지안 추정, 검색 및 최적화 방법

하루가 다르게 변하는 AI 기술을 법에서 직접적으로 나열하면서 정의하는 것이 부담스러웠는지, 4년 뒤인 2024년 3월 유럽의회(EP)에서 승인한 세계 최초의 인공지능법에서는 'AI 시스템'을 다음과 같이 3조에서 재정의하고 있다.

"AI 시스템이란 다양한 수준의 자율성을 가지고 작동하도록 설계된 기계 기반 시스템으로서, 배포 후 적응력을 발휘할 수 있으며 명시적 또는 암묵적 목표에 따라 수신한 입력으로부터 물리적 또는 가상 환경에 영향을 미칠 수 있는 예측, 콘텐츠, 추천, 결정과 같은 출력물을 생성하는 방법을 추론하는 것이다."

이러한 AI의 법적 정의에 입각하여 윤리적 이슈 및 원칙을 나열하면 다음과 같다.

통제성과 안전성

EU의 인공지능법에서 정의한 AI에 대해 주목할 만한 표현은 우선 '자율성(autonomy)'이다. 인간의 개입 없이 일부든 전부든 스스로 동작하고 운영되며 결정을 내리는 것이 AI이다. 따라서 AI의 이러한 자율성으로 인하여, AI는 통제 가능해야 한다는 '통제성(controllability)'과 AI는 안전해야 한다는 '안전성(safety)'이라는 윤리 원칙이 필요해진다.

스스로 동작한 AI가 인류에게 피해를 줄 경우 이를 통제함으로써 안전을 확보해야 한다는 원칙이다. 지금까지는 인간이 모든 디지털 신기술을 어느 정도 통제할 수 있었고 이를 통해서 안전성을 보장받았다. 이른바 '킬 스위치(kill switch)'가 모든 AI에게 마련되어야 하는데, 이것이 AI의 자율성을 인간이 어느 때든 통제할 수 있음을 의미하기 때문이다. 그러나 최근 자동차 급발진 사고를 보면서, 첨단 기술들이 뭉치면 이를 인간이 통제하기가 생각만큼 쉽지 않을 수 있음을 깨닫게 된다. 그렇다면, 사물인터넷과 연결된 AI가 외부로부터 스스로를 보호할 수 있는 미래 상황에서 AI에 대한 킬 스위치가 여전히 통할지 의문스러워진다. 이는 "AI가 인류의 마지막 기술일 수 있다."는 스티븐 호킹 박사의 경고를 상기시키는데, 킬 스위치가 제대로 작동하지 않을 경우 인류는 매우 심각한 위험에 놓일 수도 있을 것이다.

투명성과 설명 가능성

EU의 인공지능법에서 주목할 만한 두 번째 표현은 '추론(inference)'이다. 추론은 AI가 보조하거나 대체할 수 있는 인간 두뇌의 핵심 기능이다. AI의 추론이 정확하거나 뛰어날수록 인간처럼 똑똑하며 지능적이라고 평가받는다. 이러한 AI의 지능성(intelligence)이 발현되기 위한 추론 과정은 인간에게 투명하게 보여야 하는데, 그렇지 않을 경우 AI의 우수한 지능성과 그로 인한 결과를 인간은 아무런 질문도 없이 받아들이기만 해야 하기 때문이다. 이에 따라 '인간을 위한 AI', '인권을 우선시하는 AI'라는 가장 근본적인 기대를 저버리는 상황에 처하게 될 수 있으므로, 지능적으로 추론하는 AI를 맞닥뜨렸을 때 AI의 추론 과정이 인간에게 투명하게 보여질 '투명성(transparency)'과 추론 근거에 대한 '설명 가능성(explainability)'이라는 AI 윤리 원칙이 필요해진다.

코로나 3년 동안 국내 700여 개 기업은 신입사원 지원자를 대상으로 비대면 면접 방식인 'AI 면접관'을 사용했다. 국내에는 약 5개의 AI 면접관 솔루션이 있는데, 그 중 2개 업체의 솔루션이 거의 독과점을 하고 있다. 만약 이들을 활용한 면접시험에 응시한 지원자가 탈락한 경우 그 이유를 설명해주는 것이 마땅한데, 이것이 설명 가능성이라는 AI 윤리 원칙이다. 그리고 AI 면접관이 공정한지 혹은 그렇지 않은지를 알고자 할 때 AI 면접관이 동작하는 내부를 들여다볼 수 있어야 하는데, 이것은 투명성 원칙과 연결된다. 설명이 불가능한 AI는 사용자들이 기피할 것이고, 투명하지 않은 AI는 사용자들이 불안해 할 수 있어서 AI의 신뢰성을 저하하는 요인이 된다.

공정성

EU의 인공지능법에서 주목할 만한 세 번째 표현은 '시스템'이다. 통상적으로 시스템은 입력-처리-출력의 구조를 갖는데, AI 시스템 역시 마찬가지이다. 여기에서 AI 시스템의 입력은 두 가지로 구분된다. 하나는 추론의 근거가 되는 정보를 입력하여 AI에게 지능을 부여하는 과정이며, 이는 AI 시스템을 개발하는 과정에서 사용하는 입력이다. 다른 하나는 지능이 부여된 AI에게 추론을 위하여 질의하는 입력인데, 이는 개발하여 배치된 AI 시스템을 이용자가 활용하기 위한 것이다. 전자의 경우, AI의 종류에 따라서 입력물이 '정리된 지식'일 수도 있고, '학습데이터'일 수도 있다. AI가 70년의 역사 가운데 세 번의 하이프 커브

(hype curve)를 그렸다고 볼 때, 두 번째 하이프 커브에서 주목할 만한 점은 전문가 시스템(expert system)이다. 이는 특정 분야 전문가들이 교육기관에서 학습과정을 거치고 현장에서 실무를 경험하면서 내재된 암묵지를 명백지로 표현하여 컴퓨터 시스템에게 제공한 것이다. 분야별 전문가가 수십 년 간 배우고 경험한 지식을 규칙 등으로 정리하여 컴퓨터에게 제공한 것인데, 이를 '훌륭하지만 구식의 AI(Good, Old-Fashioned AI, GOFAI)'라고 부른다. GOFAI는 전문 분야별로 우수한 경우가 존재하여 지금도 일부 사용되고 있지만, 구축 시간이 오래 걸리고 명백지로 바꾸는 작업의 난이도가 높은 편이다.

반면, 아예 컴퓨터로 하여금 학습 과정부터 작업하도록 할 경우 특정 전문가보다 학습 및 경험에 소요되는 시간이 대폭 줄어들면서 구축 및 개발에 필요한 노력을 비약적으로 아낄 수 있다. 이 방법이 바로 세 번째 하이프 커브, 즉 여름기의 핵심이 되는 기계학습이다. 기계학습은 인간의 뇌신경망을 기반으로 하는 인공신경망(Artificial Neural Network, ANN)을 포함하는데, 인공신경망은 캐나다 토론토대학교 제프리 힌튼 교수 등이 제안한 다층 인공신경망 중심의 딥러닝(Deep Learning, DL)이 최근 들어 혁신적인 성능을 보이며 AI 시스템의 주류가 되었다.

GOFAI에 비하여 기계학습 기반의 AI는 대량의 데이터를 직접 학습하여 추론 정보를 축적하다 보니, 학습데이터가 가지고 있는 문제점을 그대로 투영할 수 있다. 대표적으로, 학습데이터가 가지고 있는 편견과 차별을 AI가 그대로 고착화할 수 있다는 점을 들 수 있다.

2019년에 애플에서 신용카드를 발행하기 시작했다. 그런데 신용카드의 지출한도가 여성의 경우 남성의 1/20밖에 되지 않았다. 공동명의의 재산을 보유한 부부의 경우에도 동일한 현상이 대거 발생했다. 이는 기존의 신용카드 지출한도에 관한 데이터를 보았을 때 남성이 여성보다 해당 비율만큼 높았다는 점을 AI가 그대로 학습하여 고착화한 사건이었다.

비슷한 사건은 또 있다. 2016년부터 아마존은 신입사원을 채용할 때 서류 평가 단계에서 AI를 활용했는데, 유독 여성 지원자들이 많이 탈락했다. 성별을 기재하지 않는 서류전형이었지만, 이후 아마존은 여성에 대한 차별이 있었음을 시인하며 결국 AI 서류전형을 중단했다. 아마존의 AI는 과거 서류전형 데이터를 학습했는데, 남성 탈락자보다 여성 탈락자가 월등하게 많았다. 해당 여성 지원자들의 서류에 기재되어 있던 여성 고유 활동 관련 다양한 키워드들이 불합격 강화 요인으로 작용하도록 AI 안에 불공정성이 자리 잡은 것이다.

2020년 8월 영국은 코로나 팬데믹으로 인해 우리나라의 수능시험에 해당하는 고등학교 학력평가 'A레벨 테스트'를 온라인으로 시행하면서 '다이렉트 센터 수행평가 모델'이라는 AI 알고리즘을 활용해 점수를 배정했다가 30만 명의 대입 지망생들이 거세게 반발했다. AI가 배정한 점수를 보니 사립학교 학생들의 점수는 높고 공립학교 학생들의 점수가 낮았던 것인데, 이것도 학습데이터 문제였다. 상대적으로 사립

학교 출신 학생들이 좋은 대학을 진학했던 데이터가 많고 공립학교 출신들은 그렇지 못했기 때문에 이러한 데이터들을 학습하다 보니 편견과 차별이 자연스레 드러났고 공정성을 잃게 된 것이다.

프라이버시

AI가 학습하는 대량의 데이터 안에 개인 식별정보, 민감정보 그리고 사생활 정보가 담겨져 있을 경우, 이를 미리 걸러내지 못하면 AI가 개인정보를 유출하고 프라이버시를 침해할 수 있게 된다. 챗GPT가 한창 세상의 주목을 받았던 2023년 3월 이탈리아는 챗GPT의 사용을 전면 금지한 적이 있다. 그 이유에는 여러 가지가 있었지만, 대표적으로 개인정보 유출과 프라이버시 침해를 꼽을 수 있다. 이탈리아 국민들이 챗GPT를 사용할 경우, 프롬프트를 통해서 입력한 개인정보 및 사생활 정보가 그대로 챗GPT 서버로 이동하고 이를 다시 AI가 추가 학습함으로써 개인정보가 유출되고 프라이버시가 침해될 수 있다는 것이 주된 골자였다. 결국, 3주의 유예기간에서 1주일을 더 넘긴 시점에 오픈AI는 챗GPT와의 대화 내용 및 목록을 삭제할 수 있고 서버에 대화 내용을 전달하지 않도록 선택할 수 있는 옵션을 추가함으로써 이탈리아 서비스를 다시 회복할 수 있었다.

우리나라에서도 몇 년 전 유사한 사건이 발생하였다. 국내 모 기업에서 제공한 AI 챗봇 서비스가 개시된 지 3주 만에 종료되었는데, AI로서 공정성을 유지하지 못하고 프라이버시를 준수하지 않았던 점이 3주 천하의 직접적인 원인이었다. 처음에는 중·고교 남학생들이 여성 캐릭터인 AI 챗봇을 성희롱했던 화면들을 캡처해 일부 커뮤니티 사이트 등에 게시했던 것이 문제가 되었고, 이후 성차별, 사회적 약자에 대한 차별 등이 불거지면서 해당 AI 챗봇 서비스가 차별과 편견, 공정성 문제를 일으킨다며 사회적 이슈로 대두되었다. 뒤이어 개인정보 유출 문제도 불거졌다. 해당 기업의 AI 서비스는 타사 서비스를 통해 이루어진 남녀 커플의 대화 10억 개를 학습하고 그중 1억 개를 활용해 답변을 내놓았는데 이때 개인정보를 무단으로 학습하고 민감한 정보를 완전하게 거르지 못한 사실이 밝혀져 결국 3주 만에 서비스가 종료되었다.

책임성과 책무성

EU의 인공지능법에서 정의한 AI 시스템의 출력 부분을 주목하면, 예측, 콘텐츠, 추천, 결정이라는 4가지 종류를 명시하고 있다. 코로나와 같은 전염병이 올해 유행할지에 대해 AI가 예측할 수 있고, 넷플릭스나 유튜브 등의 OTT에서 이용자가 선호하는 콘텐츠를 추천하기도 하며, 합격 여부를 AI 면접관이 결정하고 목적지까지의 경로를 AI 자동차가 설정하는 것이 모두 출력의 형태라고 할 수 있다. 그리고 글을 생성하는 초거대 언어모델인 챗GPT, 이미지를 생성하는 DALL·E와 미드저니, 영상을 생성하는 Runway의 GEN이나 Sora, 음악을 만들어주는 Suno 등은 콘텐츠를 출력물로 생성한다.

만일 이러한 출력 결과가 기대에 미치지 못하고 사실과 다르거나 심지어 사고를 일으킬 경우에는 이를 누가 책임질 것인지에 대한 문제가 발생한다. 이는 앞서 제시한 AI의 자율성과 깊은 관련이 있는데, AI가 인간의 개입 없이 자율적으로 생성한 출력물에 문제가 있을 때의 책임소재를 말한다. 자율성이 없는

기술이나 장비, 장치의 경우 제조물 책임법에 의하여 제조사가 모든 사고의 책임을 지도록 되어 있다. 그런데 AI는 자율성을 가지고 있어서 이러한 제조물 책임법이 AI 개발사에게 온전하게 책임을 물을 수 없어 이에 대한 다양한 대안이 제시되고 있다. 가장 손쉬운 방법은 보험이다. AI로 인한 사건과 사고에 대하여 경제적으로 보상하는 것이다. 그러나 자동차 보험과 마찬가지로 대부분의 보험은 결국 가입자에게 부담을 준다는 문제가 있다. 또 다른 방법은 AI를 별도의 독립된 법인체로 인정하는 것이다. AI 개발이 성공하여 많은 부를 얻든, 예기치 못한 사건과 사고로 인하여 부도가 발생하든 하나의 법인체로서 독립된 권리와 책임을 부여하는 것이다. 2018년 EU는 노동자를 대신하는 로봇 때문에 연금 수입에 문제가 발생하자 로봇을 전자 인간(electronic person)으로 법인화하여 연금 기금으로 로봇세를 징수할 수 있는 법을 발의했다가 많은 파장을 낳고 통과되지 못한 전례가 있다. AI의 법인화 노력은 포스트휴머니즘과 결합하여 앞으로 더욱 많이 시도될 것으로 보이지만 아직은 시기상조라는 의견이 지배적이다. 추가적인 대안으로는 기존의 제조물 책임법을 더욱 강화하여 AI 사업자에게도 추가 적용하는 것을 고려할 수 있다. 영국이 이 방법을 채택했는데, 가장 상식과 전통에 부합함에도 불구하고 AI 사업자들은 약관을 교묘하게 작성하여 사고 발생에 따른 무한 책임은 피하려고 하고 있어 이는 앞으로 AI의 대중화를 위해서 넘어야 할 큰 산이라고 할 수 있다.

그럼에도 불구하고, AI를 개발하고 사업화하는 주체들이 AI로 인한 모든 상황에 대해 1차적인 책임을 지는 것은 필요하므로 책임성(responsibility)은 AI 윤리에서 매우 중요하다. 이와 비슷한 윤리 원칙으로 책무성(accountability)이 있다. 책임성은 발생한 사건과 사고에 대해 적절한 조치를 취하는 것을 의미한다면, 책무성은 책임성을 이행하는 과정에서 그 원인을 규명하고 이를 설명하며 다시 발생하지 않도록 조치를 취하는 과정까지 포함한다. 사고가 발생할 때마다 손해배상을 한다면, 책임성은 만족할지 모르나 책무성은 부족하다 할 수 있다. 그리고 책무성은 공동체 개념에서의 책임성과 관련이 있다. AI를 다루는 개발자와 사업자는 본인이 책임질 상황이 아니더라도 관심을 가지고 해결 역량을 함께 키워야 하는데, 이 부분이 책무성에 해당한다고 할 수 있다.

표절과 저작권 침해

콘텐츠라는 AI 출력물은 예측, 추천, 결정이라는 3가지의 출력 형태로 구성되는데, 이에 따른 윤리 이슈가 존재한다. 바로 표절과 저작권 침해이다. AI 내부적으로 TDM(Text and Data Mining) 과정을 통하여 이루어지는 예측, 추천, 결정은 그 출력 형태가 서술적이거나 묘사적이지 않기 때문에 표절 및 저작권 침해의 여지가 거의 없었다. 그래서 학습데이터에 연계된 제반 권한을 '공정 사용(fair use)'이라는 명목으로 인정하지 않았다. 우리나라도 챗GPT가 등장하기 전까지만 해도 글로벌 AI 경쟁력 강화를 위하여 TDM 활용에 있어서 학습데이터의 저작권 인정을 유보하는 저작권법 전면 개정안이 상정되기도 했다. 그러나 AI가 글, 그림, 영상, 음악, 버추얼 휴먼 등 다양한 콘텐츠를 출력물로 내놓자 표절과 저작권 위반에 대한 논쟁이 치열하게 벌어졌다. 워싱턴 포스트는 챗GPT를 만든 오픈AI를 상대로 기사 저작권 침해, 그리고 자사 기사에 잘못된 내용을 연계한 할루시네이션에 따른 명예 훼손을 이유로 법적 소송을 제기했다.

2023년 4월 발표한 논문에서 미국 펜실베니아주립대학교(PSU)에 재직 중인 이동원 교수는 챗GPT가 표절의 위험이 높다고 주장했다. 그는 챗GPT의 기반 모델 GPT-3.5의 이전 모델인 GPT-2에 대하여, GPT-2가 생성한 21만 건의 글을 학습데이터 800만 건의 문서와 비교한 논문을 발표했다. 해당 논문은 출처 인용 없이 문장 바꾸기(paraphrase), 복사하여 붙이기(verbatim), 아이디어 도용하기 등의 대표적인 표절 현상이 다수 발생했음을 지적했다. 이러한 표절 현상은 저작권 침해로 이어질 수 있는데, GPT-3와 GPT-3.5, GPT-4 모두 GPT-2와 동일한 기반 모델을 사용하고 있으므로 현재의 챗GPT에서도 표절 및 저작권 침해가 발생하고 있다고 볼 수 있다.

진짜와 가짜 구분

AI의 뛰어난 성능은 콘텐츠 출력물에 대한 인간의 판단에 어려움을 초래한다. 딥페이크가 가장 대표적인 영역이다. 사진과 영상 속 인물이 직접 촬영된 것인지, 혹은 AI가 가짜로 만든 결과물인지에 따라 다양한 사회적 문제가 따른다. 배우의 어린 시절을 생성해 내는 디에이징(de-aging) 기술은 방송 및 영화 산업에서 유망한 반면, 딥페이크로 만든 디지털 성착취물은 피해자들에게 너무나 가혹한 불행을 초래한다. 2020년 N번방 사건이 발생했을 당시 딥페이크에 의한 가짜 사진 및 영상물의 확산이 큰 이슈가 되었기에, 국내에서도 딥페이크 기반의 성적합성물에 대한 처벌조항을 「성폭력범죄의 처벌 등에 관한 특례법(성폭력처벌법)」 제14조의2에 '허위영상물'이라는 용어를 중심으로 신설하였다. 이 법률에 근거하여 여성가족부 산하 디지털성범죄피해자지원센터에서는 관련 영상물 삭제를 위해 수사·법률 등을 연계하여 지원하고 있다. 합성·편집 관련 지원은 2022년 212건에서 2023년 423건으로, 딥페이크 기술 이용 '성적허위영상물'에 대한 방심위 시정 요구는 2022년 3,574건에서 2023년 11월 기준 5,996건으로 증가하기도 했다. 지난 22대 국회의원 선거 당시에는 공직선거관리법을 개정하여 딥페이크를 이용한 선거 운동을 전면 금지하기도 했다.

이에 따라 EU의 인공지능법에서는 AI가 생성한 합성 산출물(synthetic output)에 대하여 AI가 생성하였음을 명확하게 밝히는 라벨을 부착하고, 가시적 및 비가시적 워터마크를 기재할 것을 명시하였다. 2023년 10월 30일에 발효된 조 바이든 미국 대통령의 AI 행정명령도 같은 요구 항목이 있는데, 다만 워터마크는 미국 상무성이 만들어 배포하겠다는 단서를 달았다.

다양성

라벨을 달고 워터마크를 부착하는 것은 콘텐츠라는 출력물, 즉 AI의 합성 산출물에 대한 진위 여부를 구분할 수 없기 때문에 취해진 조치라고 볼 수 있다. 그러나 일각에서는 AI의 합성 산출물과 인간의 저작물을 처음부터 분명하게 구분해야 한다는 주장이 있는데, 이를 '다양성(diversity)'의 원칙이라고 부른다.

'반복의 저주(The Curse of Recursion): 생성된 데이터를 기반으로 학습할 때 모델이 잊어버리는 것'이라는 제목의 논문이 2023년 5월 발표되었다. 이 논문에 따르면, 먼저 개발된 AI가 만든 합성 산출물을 후속

AI가 학습데이터로 사용하게 되면 '모델 붕괴(model collapse)'가 이루어진다. 모델 붕괴는 트랜스퍼 모델 등 다양한 AI 모델에서 발생하는데, 이러한 이유 때문에 인간의 저작물을 AI의 합성 산출물과 반드시 구분해야 하며 더 소중하게 다루어야 한다는 주장이 제기되는 것이다. 2021년 11월 23일 193개국이 채택한 유네스코 'AI 윤리 권고' 제95조와 제98조에서도 다양성 원칙을 강조하는데, 인간의 언어와 표현의 뉘앙스에 AI가 미치는 문화적 영향을 조사하여 필요한 부분은 대처해야 하며(제95조), 문화적 표현물에 대해 다양한 공급과 접근을 촉진해야 한다(제98조)고 명시하고 있다.

현재 챗GPT의 기반 모델은 GPT-3이며, GPT-3.5와 GPT-4 모두 같은 기반 모델에 뿌리를 두고 있다. 오픈AI는 프로젝트 코드명 '아라키스(Arrakis)'로 불리는 다음 버전의 기반 모델 GPT-5를 준비하고 있는데, MoE 등 첨단 기술을 사용하고 학습데이터도 GPT-3보다 2배를 늘렸지만 성능이 현재의 기반 모델에 미치지 못해 공개를 하지 못하고 있는 상황이다. 중요한 대목은, 학습데이터의 50%가 기존 AI의 합성데이터라는 점이다. 이는 앞선 논문에서 다룬 내용에 해당하는 구체적인 사례로 꼽을 수 있다.

2) 혁신 신기술로서 AI의 보편적 윤리

공공성

사회를 전환시키는 혁신 신기술로서 AI가 따라야 할 보편적 윤리 원칙도 존재한다. 우선적으로 고려할 것은 그 어떤 혁신 신기술이라도 궁극적으로 '인류 공영'을 위해야 한다는 '공공성(publicness)' 원칙이다. 기술이 소수의 개인과 기업만을 위해서는 안 되며 가급적 인류 전체에게 도움을 줄 수 있는 공공의 기술이어야 한다. 이러한 공공성에는 앞서 소개한 '다양성'도 포함된다. 다양한 인류가 고루 혜택을 볼 수 있도록 고려하는 것도 공공성의 일부이기 때문이다.

포용성

과거 인터넷과 컴퓨터가 혁신 신기술로 부상할 당시에는 해당 디지털 기술에 접근하여 이용하는 데 있어서 소외되는 사회 구성원이 없도록 하는 '디지털 접근성 보장'이 중요한 윤리 원칙이었다. 그리고 접근성 보장이 이루어지지 않는 계층을 대상으로 정부는 디지털 격차 해소 정책을 전개했다. 디지털 신기술을 통한 기회 창출의 장에서 소외되는 국민이 생기지 않도록 하기 위한 목적이었던 것이다. 그러나 AI라는 디지털 신기술은 이전의 인터넷이나 컴퓨터와 달리 학습의 난이도가 높고 활용 역량이 원천적으로 부족한 상황이어서 디지털 격차 해소가 불가능한 계층이 존재한다. 이러한 계층에게도 AI를 통한 혜택의 기회로부터 소외되지 않도록 지원할 필요가 있는데, 이를 포함하는 개념이 바로 '포용성(inclusion)'이다. 실버세대를 위한 키오스크 표준화, 장애인을 위한 에이블테크(able-tech) 산업 활성화가 포용성의 구현 사례이다.

보안성과 안전성

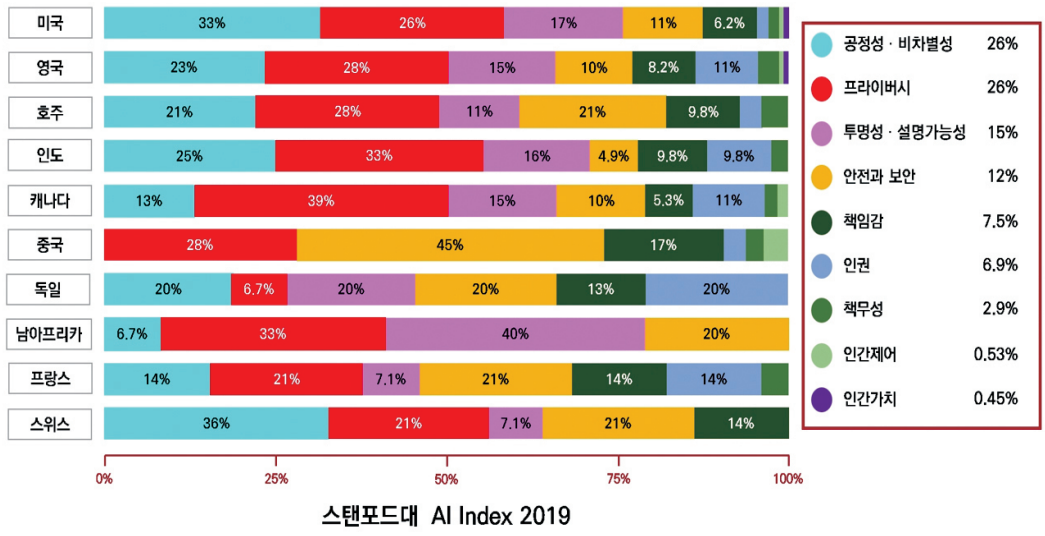
혁신 신기술로서 AI는 '보안성(security)'과 '안전성(safety)' 원칙을 만족해야 한다. 어떤 신기술을 사회에 도입했을 때, 외부 침입자에 의하여 문제가 발생하지 않도록 하는 '보안성'과 재난과 사고 등의 불가항력적 상황에서도 제 기능을 유지할 수 있는 '안전성'은 크게 보면 앞서 제시한 '책임성'의 한 부분으로도 볼 수 있다.

AI의 보안성 관련 문제 규명 및 해결책에 대한 실례는 MITRE가 제시한 ATLAS를 통해서 전체적인 윤곽을 살펴볼 수 있다. 통상적으로 AI의 안전성은 보안성을 포함한다. 안전한 AI는 정상적으로 사용하는 과정에서 발생하는 모든 문제 상황으로부터도 안전해야 하지만, 비정상적인 외부 공격으로부터도 견딜 수 있어야 하기 때문이다. 영국과 미국에서는 AI의 안전성 확보를 위하여 AI 안전연구소(AI Security Institute, AISI)를 이미 운영하고 있으며, 우리나라도 올해 안에 설립할 계획이다. 이 경우, 안전성은 통상적인 의미보다 더 포괄적으로 사용되는데 미국의 경우 NIST(국립표준연구소)에서 제시한 AI 위험관리프레임워크(RMF)를 과학적으로 검증하고 구현하기 위해 AISI를 운영하므로 AI의 잠재적 위험 대부분을 포괄하여 다룬다고 보면 좋을 것이다.

3) AI 윤리의 글로벌 동향 비교

AI 윤리의 원칙은 국가마다 중시하는 내용에 대한 관점의 차이가 존재할 수 있으며, 이에 따라 특정 사안에 대해 윤리 민감도가 다를 수 있다. 2019년 스탠포드대학교는 국가별 빅데이터 분석에 기반하여 인공지능 윤리 원칙을 키워드 형태로 표시한 'AI 인덱스' 자료를 발표한 바 있다. 이는 인공지능 윤리 원칙에 대한 관심도에 있어서 국가적 특성에 따른 차이가 제법 존재하고 있음을 보여준다. 중국의 경우 '공정성과 비차별성'에 대한 이슈가 아예 존재하지 않지만 '안전과 보안' 이슈가 45%로 가장 큰 비중을 차지한다. 반면, 독일의 인공지능 윤리 원칙 관심도를 보면 '프라이버시' 원칙이 가장 낮지만 '인권' 원칙이 가장 높다. 이처럼 인공지능 윤리 원칙에 대한 국가적 관심도가 상이함에 따라 글로벌 표준의 인공지능 윤리를 합의하는 것이 쉽지 않아 보이지만, 인공지능에 대한 우려와 두려움을 최소화하기 위하여 AI 윤리를 적극적으로 다루어야 한다는 점에 대해서는 모든 나라가 예외 없이 동의하고 있다.

그림. 26 AI 윤리의 글로벌 이슈 동향



출처: Stanford University(2019).

라 AI 규제 동향

1) EU의 인공지능법

생성형 AI가 대중적으로 알려지기 전부터, 인공지능에 대한 규제는 이미 준비되고 있었는데, 2021년 4월 EU 집행위원회(EC)가 제안한 「인공지능법(AI Act)」이 바로 그것이다. 이 법안은 모든 인공지능을 '위험' 관점에서 분류하여 규제한다. 초기 원안에서는 인공지능을 금지, 고위험, 기타 등 3개 요소로 분류하였는데 이후 챗GPT가 등장하는 등 변수가 발생하면서 의회에서 추가 논의를 통해 보다 더 구체적인 분류를 정립했다. 아예 개발 자체가 금지된 인공지능(unacceptable-risk AI)부터 시작하여 고위험 인공지능(high-risk AI), 범용 및 생성형 인공지능(general-purpose and generative AI), 제한적 위험 인공지능(limited-risk AI)까지 인공지능 전체를 위험을 기준으로 분류한 것이다.

금지된 인공지능은 크게 세 분야로 구분된다. 첫째는 사람이나 특정 취약 집단에 대한 인지 행동을 조작하는 인공지능(예를 들어 어린이의 위험한 행동을 조장하는 음성 인식 장난감), 둘째는 사회적 점수(social scoring)를 산정하는 인공지능(즉, 행동, 사회경제적 지위 또는 개인적 특성을 기준으로 사람을 분류하는 인공지능), 셋째는 사

람의 생체를 식별하고 분류하거나 안면 인식 등 실시간 및 원격으로 생체를 인식하는 인공지능이다. 다만, 세 번째 분류의 경우에는 법 집행을 목적으로 할 때 예외적으로 허용될 수 있다. 특히 '실시간' 원격 생체인식 시스템은 매우 제한적으로 허용되며, 상당한 지연 후에 신원 확인이 이루어지는 '포스트' 원격 생체인식 시스템은 심각한 범죄에 한정하여 법원의 승인을 받아야 가능하다.

고위험 인공지능은 사전 위험 평가를 실시해야 하며, 인공지능의 주요 사항을 EU 데이터베이스에 미리 등록해 놓아야 한다. 이는 일종의 에스크로(escrow) 제도로서 해당 기업의 존망과 상관없이 EU 국가에 도입된 고위험 인공지능을 꾸준히 관리하기 위함이다. 모든 규제 조건을 만족한 고위험 인공지능은 인증을 받아서 CE 마크를 부착할 수 있으며 비로소 유럽에서 서비스를 개시할 수 있다. 고위험 인공지능은 특정하여 명시되어 있으며 이는 계속 증가할 수 있다.

범용 및 생성형 인공지능은 챗GPT의 영향을 받아 EU 의회 수정안에 가장 늦게 담긴 내용이다. 이 인공지능은 저작권과 관련된 학습데이터를 어떻게 다루었는지에 대한 요약자료를 제출해야 한다. 해당 인공지능을 통해서 어떤 콘텐츠를 만들 경우, 창작자는 인공지능을 이용했다는 사실을 명확하게 표기해야 한다(공개 disclosure, 면책 disclaimer). 고위험 인공지능은 이를 악용하여 불법 콘텐츠를 제작하는 경우를 대비하여 이를 예방할 수 있는 모델로 미리 설계되어야 하는데, 이처럼 추상적인 규제에 대하여 OECD GPAI(Global Partnership on AI)에서는 구체적인 대안을 제시하고 있다. 바로 AI 합성 산출물에 대한 검증 메커니즘(detection mechanism)을 제공하라는 것이다. 여기에는 AI 합성 산출물 분류 도구를 비롯하여 워터마크와 같은 기술적 조치도 포함된다.

제한적 위험 인공지능은 제공된 정보를 토대로 사용자가 결정을 내릴 수 있도록 최소한의 투명성 요건을 준수해야 한다. 사용자는 인공지능 애플리케이션과 상호작용을 한 후에 이를 계속 사용할지에 대한 여부를 결정할 수 있는데, 이에 대해 사용자가 인지할 수 있도록 해야 한다. 대표적으로, 딥페이크와 같이 이미지, 오디오 또는 비디오 콘텐츠를 생성하거나 조작하는 인공지능이 포함된다.

집행위원회의 원안을 토대로 2022년 12월 약간의 수정을 가한 이사회 안이 만들어졌다. 그리고 2022년 11월 말 등장한 챗GPT로 인해 원안 대비 많은 수정이 이루어진 의회 수정안이 2023년 6월 의회에 제출됐다. 결국 집행위원회, 이사회, 의회는 2023년 12월 8일 인공지능법에 대한 합의에 이르렀으며, 2024년 3월 유럽의회 인준을 거쳐 8월 공표되었다. 2년의 유예 기간(grace period)을 거쳐 2026년 전면 시행된다.

2) EU의 디지털서비스법

EU가 2022년 11월 6일 발효하여 2023년 8월 16일부터 시행하고 있는 「디지털서비스법(DSA)」이 있는데, 해당 법은 생성형 AI에 대한 다른 차원의 규제를 명시하고 있다. 원래 「디지털서비스법」은 SNS, 검색엔진과 같은 빅테크 기업을 규제하기 위한 법으로서 가짜 뉴스와 같은 유해 콘텐츠를 필터링하도록 의무 조항을 부여하고 있다. 만약 생성형 AI가 만든 합성 산출물을 위 기업들이 제공하는 경우, 해당 합성 산출물에 인공지능을 활용했다는 사실을 표시하며 관련 라벨을 붙이고 소비자의 악용을 방지할 조치를 반드시 제시하도록 규정하고 있다.

3) 미국의 AI 행정명령

미국의 경우, 전통적으로 기업 자율 규제를 중시해 온 나라이다. 2023년 3월, 6개월 간 AI 개발을 중지하자는 FLI 공개서한이 발표되고 같은 해 5월 샘 알트만이 미국 상원 법사위원회 청문회에 참석하여 "AI에게 규제가 필요하다."라는 발언을 한 이후, 자율적 규제에 대한 논의가 급격히 이루어지고 있다. 7월 21일 조 바이든 미국 대통령은 7개 AI 기업 대표를 백악관으로 초청하여 8개 항목에 대한 합의를 이루었는데, 이를 '8대 원칙(8 Commitments)'이라고 부른다. 이 합의안은 크게 안전성(safety), 보안성(security), 신뢰성(trust)이라는 세 꼭지를 중심으로 구성되어 있다. 주목할 부분은 사이버 보안과 내부자 보안을 중요시함과 동시에 AI 기업 내부에 레드팀을 두어 버그바운팅 작업을 반드시 수행하도록 권장하고 있다는 점이다. 이는 챗GPT의 버그바운팅 사례가 큰 영향을 준 것으로 파악된다. 앞으로는 보안 이슈로 부도를 맞는 인공지능 기업들이 생기지 않도록 하기 위한 규제로 볼 수 있다. 이 합의안에서도 인공지능이 생성한 합성 산출물에 대한 출처 표시뿐만 아니라, 장차 미국 정부 특히 상무성에서 개발할 워터마크를 콘텐츠 안에 부착할 것을 명시하고 있다. 기업이 자율적으로 워터마크를 도입하여 사용하도록 하는 것이 아니라 미국 정부가 워터마크를 개발하여 이를 제공하겠다는 내용은 앞으로 인공지능 시장을 미국 정부가 규제 중심으로 선도해 나가겠다는 의미로 해석된다. 이 8개 항목의 합의안은 미국의 20여 개 우방국가도 함께 이행할 것이라고 바이든 대통령은 밝혔는데, 여기에는 대한민국도 포함된다.

2023년 10월 30일 바이든 대통령은 '8대 원칙'을 중심으로 하는 66면의 'AI 행정명령(AI Executive Order)'을 발동했는데, 우리나라도 본격적으로 이 내용을 분석하여 규제의 수준을 미국에 맞출 필요가 생겼다. AI 행정명령은 법률과 거의 동등한 규제력을 가지고 있으며, 후임 대통령이 행정명령을 폐기하지만 않는다면 법률과 같은 구속력을 갖는다.

4) 우리나라의 AI 기본법

우리나라는 2023년 2월 14일 국회 과학기술정보방송통신위원회에서 그동안 발의하여 제출된 여야 국회의원 7명의 7개 인공지능법안을 모두 폐기하고 위원회 대표 법안으로 통합한 후 법안심사소위원회를 통과시켰다. 그 내용을 보면 '고위험 영역 인공지능'이라는 표현을 통해 EU 인공지능법 집행위원회 원안을 상당 부분 인용했음을 알 수 있는데, 다만 금지된 인공지능은 명시하지 않고 있어서 일부 시민단체들이 이에 대한 규정 추가를 요구하고 있다. 8월에 2개 의원실에서 이를 반영하여 추가 발의를 하기도 했다. 이러한 국내 인공지능법안에 따르면, 국무총리 소속으로 인공지능위원회를 설치·운영하고 그 산하에 신뢰성전문위원회를 설치하도록 명시하고 있으며 과학기술정보통신부는 3년마다 인공지능 기본계획을 수립해야 한다. 그리고 인공지능에 대한 여러 규제들은 '인증'이라는 단어에 집약하여 향후 검증과 인증을 진행하도록 규정하고 있다. 인증 관련 실무는 현재 한국정보통신기술협회(TTA)가 준비하고 있다.

그러나 우리나라의 AI 기본법은 21대 국회에서 통과되지 못했고, 이에 따라 22대 국회에서 다시 원점부터 출발해야 한다. 세계 각국은 AI 산업의 '진흥'과 '규제'라는 양날의 검을 쥔 채 AI의 여름기를 주도하기 위해 노력하고 있다. EU는 올해 3월 인공지능법을 의회에서 통과시킨 후, 5월 29일 EU 집행위원회 안에 'AI 사무국'을 설치해서 본격적인 인공지능법 운영과 지원에 돌입했다. AI 사무국은 5개의 부서로 구성되는데, AI 혁신 패키지를 통해 EU의 AI 생태계를 진흥함과 동시에 AI의 안전성을 위한 규제 준수도 담당한다. 미국도 작년 10월 말 조 바이든 대통령의 AI 행정명령(일종의 인공지능법)이 발동된 이후, AI 인력 확보를 위한 노력 및 AI 산업 진흥을 위한 구체적인 행정 실무에 돌입했다. 또한, AI 안전연구소를 NIST 산하에 설립하여 AI의 위험에 대비한 체계적이며 과학적인 대응에 착수했다. 지난 4월 중순 미국 상무부 장관은 AI 안전연구소에 다양한 CO(chief officer)들을 대폭 추가 임명하여 구체적인 역할을 주문하기도 했다. 영국도 2023년 11월 AI 안전연구소를 발족하여 이듬해 5월에는 AI 안전성에 관한 과학 보고서도 발행했다.

이러한 글로벌 흐름을 지켜보면 우리 상황은 한시가 급하다. 글로벌 경쟁력을 갖춘 국내 AI 산업 생태계를 구축하고 관련 인력을 확보하기 위한 중장기적 전략이 필요하며, 이와 더불어 AI의 위험성을 선제적으로 통제하기 위한 기술적, 제도적 조치도 마련되어야 한다. 국가적 차원에서의 진흥과 규제를 진두지휘할 조직도 필요하며 AI의 잠재적 위험을 통제할 수 있는 AI 안전연구소의 설립도 시급하다. 무엇보다도, 이러한 내용들을 충분히 담아낼 AI 관련 입법이 우선되어야 한다.

이번 22대 국회에서는 정치적 상황을 잠시 뒤로 미루고 지난 21대 국회에서 2년 넘게 계류하다 완성되지 못한 AI 기본법을 과학적 관점에서 담아내고 현장 중심의 구체적인 논의를 거쳐 진흥과 규제의 적절한 타협점을 도출해야 할 것이다.

마 맺음말

'미란다 원칙'을 만들었던 미국 연방법원의 얼 워런 대법관은 "윤리는 배가 떠다니는 바다와 같다."라고 표현했다. '윤리'라는 바다가 크고 깊고 물이 풍부할수록 '법'이라는 배는 이 바다 위를 무리 없이 부드럽게 떠다닐 수 있다. 어떤 신기술이 부작용과 역기능을 낳을 경우, 곧바로 법률로 규제하면 될 것이라고 쉽게 생각할 수도 있다. 그러나 신기술 관련 윤리가 사회 구성원들 간에 충분하게 공유되지 않은 상태에서 소수의 전문가 의견만을 기반으로 관련법을 제정할 경우, 예상치 못한 변수와 반복적인 개정이 발생하게 되어 해당 신기술은 충분히 활용되기도 전에 사회에서 소멸할 수도 있다. 인공지능의 경우 더욱 그러할 우려가 높다. 인공지능 관련 입법에 앞서 반드시 선행되어야 할 것은 사회 구성원 전체가 인공지능으로 인한 위험과 문제점들을 함께 인식하고 다양한 대안을 교류할 수 있는 공론의 장을 충분히 마련하는 것이다. 이는 인공지능 윤리의 대중화를 의미한다. 「인공지능법」이 2024년 3월 유럽의회를 통과하기 5년 전인 2019년 4월, 이미 인공지능 윤리 가이드라인이 발표되었음을 주목할 필요가 있다. 이제는 우리 차례가 되었다.

참고문헌

김명주(2017). “인공지능 윤리의 필요성과 국내외 동향”, 〈정보와 통신〉, 제34권 제10호, pp. 45-54.

김명주(2022). AI는 양심이 없다, 헤이북스.

김명주(2024). “인공지능의 잠재적 위험과 국제적 규제 동향”, 〈문명과 경계〉, 제8호, pp. 43-77.

김명주(2024). “선거 속 딥페이크, 그 실태와 윤리”, 〈관훈저널〉, 제171호, pp. 115-123.

Bostrom, N. & Yudkowsky, E.(2014). “The Ethics of Artificial Intelligence”, in Frankish, K. & Ramsey, W. M.(eds.), Cambridge Handbook of Artificial Intelligence, Cambridge University Press.

Bostrom, N.(2014). Superintelligence: Paths, Dangers, Strategies, Oxford University Press.

Coeckelbergh, M.(2020). AI Ethcis, The MIT Press.

Lee, J., Le, T., Chen, J. & Lee, D.(2023). Do Language Models Plagiarize?, In Proceedings of the ACM Web Conference 2023(WWW '23), ACM.

Shumailov, I., Shumaylov, Z., Zaho, Y., Gal, Y., Papernot, N. & Anderson, R.(2023). The Curse of Recursion: Training on Generated Data Makes Models Forget, DOI:10.48550/arXiv.2305.17493.

파보세(2023). 챗GPT 1년 격돌하는 인공지능 시각! [비디오파일]. Retrieved August 31, 2024 from https://www.youtube.com/watch?v=EOrrVorkvz0

Stanford University(2019). artificial intelligence index 2019 annual report, Retrieved August 31, 2024 from https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf.

UNESCO: United Nations Educational, Scientific and Cultural Organization(2021). Recommendation on the Ethics of Artificial Intelligence, Retrieved August 31, 2024 from https://unesdoc.unesco.org/ark:/48223/pf0000381137.

KAST

10 소결

인공지능(AI) 기술의 발전은 이미 우리 사회 전반에 걸쳐 심대한 변화를 일으키고 있다. 생성형 인공지능과 대형 언어모델(LLM)은 산업, 의료, 금융, 교육 등 다양한 분야에서 혁신을 주도하고 있으며, 멀티모달 언어모델과 범용인공지능(AGI)은 보다 포괄적이고 강력한 기능을 제공하며 미래의 새로운 가능성을 열어주고 있다. 이러한 기술들은 기계번역, 텍스트 요약, 질의응답 시스템, 이미지 캡셔닝, 비디오 해석 등 여러 작업에서 뛰어난 성능을 발휘하며 인간의 능력을 확장시키고 있다.

앞으로 인공지능은 에너지 효율화, 탄소중립, 정밀농업 등 다양한 분야에서 지속적으로 활용될 것이며, 인공지능 기반 스마트 시티와 같은 미래형 도시의 구축에도 중요한 역할을 할 것이다. 이러한 기술 발전에 발맞추어 각국의 정부는 인공지능 기술 발전과 윤리적 문제를 균형 있게 다루는 정책을 마련하고 있다. 미국, 중국, EU 등 주요 국가들은 인공지능 기술의 선도적 위치를 유지하기 위해 인재 양성, 연구개발 투자, 산업 생태계 조성 등 다양한 정책을 추진하고 있으며, 우리나라도 이에 발맞추어 기술 발전을 위한 투자를 추진하고 있다.

생성 AI 기술의 도입은 콘텐츠 생성 능력을 통해 생산성을 크게 향상시키며, 금융, 법률, 공공 분야 등에서 실질적인 성과를 보이고 있다. 예를 들어, Bloomberg는 자체적으로 학습한 금융 특화 LLM인 BloombergGPT를 통해 내부 업무에 활용하고 있으며, 국내 4대 은행에서도 생성 AI를 도입하여 코딩 어시스턴트 등 내부 업무 생산성 향상에 기여하고 있다. 우리나라 정부는 초거대 공공 AI TF를 운영하며 공공 분야 생산성 혁신과 대국민 서비스 향상을 위해 생성 AI를 도입하고 있다. 그러나 생성 AI 기술의 발전은 일자리 변화와 윤리적 문제 등 새로운 도전을 동반한다. 생성 AI가 보유한 인간 수준에 버금가는 이해, 추론, 콘텐츠 생성 능력은 일자리 대체 우려를 불러일으키고 있다. 그러나 생성 AI의 도입은 일자리 대체로 이어지기보다는 일부 과업을 AI가 자동화하고 중요한 과업을 인간이 직접 수행하는 형태로 변화를 가져올 것으로 생각된다.

새로운 AI 에이전트 시대가 도래하면서, AI 에이전트는 우리의 일상과 업무에서 더욱 중요한 역할을 하게 될 것이다. AI 에이전트는 인간을 대신해 특정 작업이나 목표를 수행하며, 데이터 수집, 분석, 문제 해결 절차 계획 및 수립, 수행 및 평가 등을 독립적으로 해낼 수 있다. 이러한 AI 에이전트는 스마트폰, 자

동차, 가전, 로봇 등 다양한 디바이스에 탑재되어 사용자와 상호작용하며 일상생활과 업무를 도울 것이다.

기술의 발전에 따른 윤리적 문제와 사회적 책임에 대해서도 고려해야 할 것이다. 생성 AI가 인간의 창의성과 자유를 침해하지 않도록 적절한 규제와 가이드라인이 필요하다. 더불어 전 세계 문화 다양성을 지키기 위한 방법으로 소버린 AI의 중요성을 인식하고, 한국의 AI 기술과 산업 생태계 역량을 강화하며 글로벌 AI 다양성을 위한 협력을 추진해야 할 것이다. 이를 통해 우리는 생성 AI 기술의 혜택을 최대한 활용하는 동시에 기술 발전이 가져올 수 있는 부정적인 영향도 최소화할 수 있을 것이다.

미국과 중국은 각기 다른 전략과 강점을 바탕으로 AI 분야에서 선두를 달리고 있다. 미국은 실리콘 밸리의 글로벌 기업들과 DARPA를 중심으로 한 장기적이고 전략적인 투자를 통해 AI 기술 발전을 주도하고 있다. 특히, DARPA의 프로젝트를 통해 자율주행 자동차, 기계 독해 프로그램 등 다양한 분야에서 혁신적인 기술을 개발하고 이를 민간 기업에 이전함으로써 상용화에 성공하고 있다. 반면, 중국은 내수 시장 보호와 인재 유치에 기반한 기술 추격형 전략을 통해 AI 분야에서 빠르게 성장하고 있다. 천인계획과 만인계획 등의 프로그램을 통해 해외에 있는 중국계 인재를 유치하고, 자국 내 기업을 육성하여 대규모 데이터를 기반으로 AI 기술을 발전시키고 있다.

이러한 미국과 중국의 AI 전략은 우리나라에게 중요한 시사점을 제공한다.

첫째, 장기적이고 지속적인 연구 투자 및 전략적인 기획이 필수적이다. DARPA와 같은 연구 지원 기관을 통해 장기적인 연구 프로젝트를 추진하고, 이를 통해 새로운 기술을 개발하는 것이 중요하다. 둘째, 글로벌 인재 유치 및 인재 양성이 필요하다. 해외에서 유수한 연구자와 학생을 유치하고, 이들을 지원할 수 있는 환경을 조성하여 AI 분야의 인재를 양성해야 한다. 셋째, 원천기술의 상용화 및 혁신적인 기업 환경 조성이 필요하다. 실리콘 밸리의 성공 사례에서 볼 수 있듯이, 대학에서 개발된 기술과 인재를 산업계로 연결하고, 이를 통해 글로벌 경쟁력을 갖춘 기업을 육성해야 한다.

우리나라는 AI 기술 개발과 인재 양성을 통해 글로벌 경쟁에서 우위를 확보하고, 경제적 성장을 이끌어 나가야 한다. 이를 위해 장기적인 연구 투자와 전략적인 기획, 글로벌 인재 유치와 양성, 원천기술의 상용화와 혁신적인 기업 환경 조성을 위한 노력이 필요하다. 이러한 전략을 통해 우리나라는 AI 기술의 선도국가로 자리매김하고, 미래의 기술 혁신을 주도할 수 있을 것이다.

다양한 학문 분야에서 인공지능을 활용한 연구가 활발히 진행되고 있다. 물리학과 AI의 융합은 현재 진행형으로, 두 학문이 상호 보완적으로 발전해 나가고 있다. 물리학은 오랜 전통과 깊이를 자랑하는 학문이지만, AI의 급격한 발전은 물리학 연구에 새로운 도전과 기회를 제공하고 있다. AI 기술은 단순히 물리학의 방대한 데이터를 분석하고 예측하는 도구로 활용될 뿐만 아니라, 물리학의 이론과 법칙을 이해하

고 새로운 발견을 가능하게 하는 잠재력을 가지고 있다. 예를 들어, 핵융합 플라즈마 제어와 같은 물리학의 난제들을 AI를 통해 더 효과적으로 연구할 수 있으며, 우주 관측 데이터 분석을 통해 암흑물질의 분포를 예측하는 기술 개발 사례에서 볼 수 있듯이, AI는 물리학이 접근하기 어려운 영역까지 탐구할 수 있는 가능성을 열어주고 있다. 이러한 AI 기술의 발전은 단순히 데이터 기반의 통계적 분석을 넘어서, 물리학적 법칙을 이해하고 이를 기반으로 새로운 현상을 예측하는 단계로 나아가고 있다. 물리학과 AI의 융합을 촉진하기 위해 물리학자들이 AI 기술을 이해하고 이를 연구에 적용하는 데 있어 능동적인 자세가 필요하며, 연구소와 대학은 이러한 융합 연구를 지원하기 위한 유연한 정책을 수립하고, 학문 간 교류를 장려해야 한다. 특히, 연구지원 정책이 학문 간 융합 연구를 인정하고 지원하는 방향으로 변화해야 한다.

인공지능은 생명과학 연구의 혁신적인 도구로 자리 잡으며 연구의 효율성과 정확성을 크게 향상시키고 있다. AI는 단백질 구조 예측, 유전체 데이터 분석, 약물-표적 상호작용 예측 등에서 뛰어난 성과를 보이며, 신약 개발 과정에서도 시간과 비용을 절감하고 있다. 예를 들어, 구글 딥마인드의 알파폴드와 워싱턴대학교의 로제타폴드는 단백질 구조 예측에서 높은 정확도를 보이며 신약 개발에 기여하고 있다. 엔비디아와 아스트라제네카의 MegaMolBART와 같은 AI 플랫폼은 신약 후보 물질 발굴 과정을 가속화하고 있다. 질병 진단 및 맞춤형 치료에서도 AI는 중요한 역할을 하고 있으며, 유방암과 피부암 진단에서 높은 정확도를 보인다. 이는 의료 인력과 장비가 부족한 지역에서도 유용하게 사용될 수 있다. AI 기반 생명과학 연구는 연구 효율성과 정확성을 높이고, 합성생물학, 유전자 편집, 바이오프린팅 등 새로운 연구 분야를 개척할 것이다. 또한, AI는 자원 소비를 줄이고 환경 보호와 생태계 유지에 기여할 것이다.

그러나 생명과학 연구에서의 AI 도입에 대한 윤리적 및 사회적 고려사항도 중요하다. AI의 의사결정 과정의 투명성을 확보하고, 개인 데이터의 보호와 윤리적 사용을 보장해야 한다. 연구 데이터의 익명성을 보장하고, 데이터 수집 목적과 사용 방법을 명확히 설명하며, 데이터 제공자의 동의를 받아야 한다. 또한, AI 기술의 혜택이 공정하게 분배될 수 있도록 정책적, 제도적 지원이 필요하다. AI는 생명과학 연구의 패러다임을 혁신적으로 변화시키고 있으며, 그 역할은 앞으로도 확대될 것이다. AI 기술을 효과적으로 활용하여 생명과학 연구의 새로운 가능성을 탐구하고, 인류의 건강과 지구 환경 보호에 기여할 수 있는 방향으로 나아가야 한다. 이를 위해 지속적인 기술 발전과 함께 윤리적 고려와 사회적 책임을 병행하여 AI 기술이 긍정적인 영향을 미칠 수 있도록 해야 한다.

AI와 다른 전공 영역의 융합은 기술 발전을 넘어 학문적 사고와 연구 방법론의 변화를 요구한다. 다른 전공 분야의 학자들은 AI를 단순한 도구로 보는 것을 넘어, AI와 협력하여 새로운 발견을 이루어내는 파트너로 인식해야 한다. 또한, AI 기술이 학문의 새로운 이론과 현상 발견에 기여할 수 있음을 인식하고 이에 기여하는 역할을 해야 한다. 이러한 변화가 이루어질 때, 다양한 학문 분야와 AI는 더 큰 시너지를 발휘하며 함께 성장할 수 있을 것이다. AI와 타 학문 분야의 융합은 새로운 시대를 여는 열쇠가 될 것이며, 이를 통해 우리는 더 나은 미래를 만들어 나갈 수 있을 것이다.

AI의 공정성과 윤리는 현대 기술 발전의 중요한 과제이다. AI 시스템의 편향 문제는 특정 그룹에게 불공정한 결과를 초래할 수 있으며, 이는 사회적 신뢰와 공정성에 큰 영향을 미친다. 이를 해결하기 위해 다양한 공정성 척도와 데이터셋을 활용한 평가 방법이 개발되고, 전처리, 중처리, 후처리 접근 방식을 통해 편향을 제거하는 기술도 발전하고 있다. 대형언어모델의 편향을 줄이기 위해 데이터 증강, 프롬프트 튜닝, 모형 구조 변경, 모형 편집, 결과 수정 등의 방법을 사용하여 공정성을 높이는 노력이 계속되고 있다. 기술적 접근 외에도 주요국들은 AI의 공정성을 보장하기 위한 정책과 연구를 활발히 추진하며, 이를 통해 AI가 사회에 긍정적인 영향을 미치고 윤리적 기준을 준수할 수 있도록 하고 있다.

AI의 공정성을 보장하기 위한 연구와 정책은 지속적으로 발전할 것이다. 연구자들은 정교한 편향제거 알고리즘을 개발하고, 예측 정확도를 유지하면서도 공정성을 높이는 방향으로 나아가야 한다. 국제적인 협력을 통해 공정성 가이드라인과 표준을 마련하고, 이를 준수하도록 하는 것이 중요하다. 이러한 노력은 AI 기술의 지속 가능성을 보장하고, 사회 구성원이 공평하게 혜택을 누리도록 한다. 공정성과 윤리는 기술적 문제가 아닌 사회적 책임이자 필수 과제로 자리 잡고 있으며, 사회적 인식도 점차 높아지고 있다. 기업과 연구자들은 공정성을 최우선 과제로 삼고, 대학과 기업은 관련 교육 프로그램을, 정부와 민간단체는 공정성에 대한 공공의 인식을 높이기 위한 홍보 활동을 강화해야 한다.

AI 공정성 보장은 사회와 경제에 긍정적인 영향을 미치며, 의료, 교육, 금융 등 다양한 분야에서 공평한 기회를 제공한다. 사회와 민간의 협력을 통해 AI 공정성을 높이는 노력이 필요하며, 시민들이 참여할 수 있는 플랫폼을 마련해 다양한 목소리를 반영해야 한다. 기업, 정부, 학계, 시민단체 등 다양한 이해관계자들이 협력하여 AI의 공정성을 높이기 위한 공동의 노력을 기울여야 할 시점이다.

AI는 현대 사회에서 경제 성장, 교육 기회 확대, 건강 증진 등 다양한 분야에서 혁신의 핵심 동력으로 작용하고 있다. 그러나 AI의 발전에 따른 윤리적 문제도 발생하며, 이를 해결하기 위한 윤리적 원칙의 준수가 필수적이다. 주요 윤리 원칙으로는 통제성과 안전성, 투명성과 설명 가능성, 공정성, 프라이버시 보호, 책임성과 책무성, 표절과 저작권 침해 방지, 진짜와 가짜 구분, 다양성 등이 있다. 특히 AI의 자율성으로 인한 통제 문제와 대량 데이터 학습으로 인한 편향 문제는 중요한 윤리적 쟁점이다. EU의 「인공지능법」, 미국의 AI 행정명령, 우리나라의 AI 기본법 등은 AI의 공정성과 윤리를 보장하기 위한 정책적 노력의 일환이다. 이러한 규제와 정책은 AI가 사회에 긍정적인 영향을 미칠 수 있도록 하는 중요한 역할을 하며, 글로벌 동향을 참고하여 적절한 AI 윤리와 규제 방안을 마련하는 것이 필요하다.

딥러닝 기반의 AI 기술은 최근 획기적인 발전을 이루며 다양한 분야에서 혁신을 주도하고 있다. 특히 초거대 언어모델의 등장과 멀티모달 AI, 임바디드 AI 등의 기술 발전은 AGI의 도래를 예고하고 있으며, 이는 삶과 산업 전반에 큰 변화를 가져올 것이다. 그러나 이러한 발전은 일자리 변화와 윤리적 문제 등 새로운 도전도 동반한다. 따라서 각국 정부와 기업은 AI 기술의 활용과 규제를 균형 있게 조절하며, 미래를

대비한 정책을 마련하고 실행해야 한다.

AI 기술의 응용 경쟁력 확보는 매우 중요하다. 제조, 의료, 금융, 모빌리티, 엔터테인먼트, 물류, 서비스, 국방 등 다양한 분야에서 AI를 활용해 생산성을 높이고, 혁신적인 제품과 서비스를 개발해야 한다. 이를 위해 AI 응용 인재 양성과 창의적인 비즈니스 모델을 가진 스타트업의 육성이 필요하다. AI 기술의 발전을 최대한 활용하여 더 나은 미래를 만들기 위해 기술과 인간이 조화롭게 공존할 방안을 모색해야 한다. AI 혁명에 대비하여 적극적인 대응과 준비를 갖춘다면, 이를 통해 대한민국은 AI 기반 혁신 산업 생태계를 구축하고 글로벌 경쟁에서 우위를 점할 수 있을 것이다.

AI 기술은 삶을 편리하고 풍요롭게 만드는 동시에 윤리적 문제와 사회적 영향 등 면밀한 검토와 대비를 요하는 과제를 동반하는데, AI 개발과 사용에 있어 윤리적 원칙 준수가 필수적임을 양지해야 한다. 통제성과 안전성, 투명성과 설명 가능성, 공정성, 프라이버시 보호, 책임성과 책무성, 표절과 저작권 침해 방지, 진짜와 가짜 구분, 다양성 등의 윤리적 원칙을 고려해 AI를 개발하고 사용하는 노력이 필요하다. 이러한 노력이 지속되어야 AI는 인류에게 더 나은 미래를 제공할 수 있다.

AI의 공정성을 보장하는 노력은 기술적 문제 해결을 넘어 사회적 신뢰와 윤리를 지키는 데 중요한 역할을 하므로, 다양한 이해관계자들이 협력해 AI의 공정성을 높이기 위한 공동의 노력을 기울여야 한다. AI 공정성과 윤리에 대한 사회적 인식과 공감대 형성을 통해 AI 기술이 사회적 신뢰를 얻고, 인류 전체에 혜택을 줄 수 있도록 하는 것이 중요하다.

인공지능 기술은 우리의 미래를 밝고 혁신적으로 만들 잠재력을 가지고 있다. 이를 실현하기 위해 지속적인 연구와 발전, 공정성, 윤리적 고민과 사회적 책임을 함께 고려해야 한다. AI 기술을 최대한 활용해 더 나은 미래를 만드는 데 힘쓰고, AI가 우리의 삶을 더욱 편리하고 풍요롭게 만들도록 기술과 인간이 조화롭게 공존할 방안을 모색해야 한다.