

ISSUE REPORT | 2023.05.08. IS-159

# AI Index 2023 주요 내용과 시사점

Summary and Implications of 2023 AI Index Report

안성원/임영모/유재흥/김정민/안미소/장진철/봉강호/노재원/오보람

이 보고서는 「과학기술정보통신부 정보통신진흥기금」에서 지원받아 제작한 것으로  
과학기술정보통신부의 공식의견과 다를 수 있습니다.

이 보고서의 내용은 연구진의 개인 견해이며, 본 보고서와 관련한 의문 사항 또는 수정·보완할  
필요가 있는 경우에는 아래 연락처로 연락해 주시기 바랍니다.

소프트웨어정책연구소

안성원 선임연구원 (swahn@spri.kr)

# CONTENT

I. AI Index 2023 주요내용	P.1
1.1. 보고서 개요 및 특징	P.1
1.2. 연구개발	P.2
1.3. AI 기술 성능	P.7
1.4. AI 기술 윤리	P.16
1.5. 경제	P.24
1.6. 교육	P.30
1.7. 정책·거버넌스	P.33
1.8. 다양성	P.39
1.9. 여론	P.42
II. 시사점	P.47
참고문헌	P.52

## 요약문

2023년 4월 3일 스탠퍼드대 인간중심 인공지능 연구소는 AI Index 2023 보고서를 발간했다. 보고서는 연구개발, AI 기술 성능, AI 기술 윤리, 경제, 교육, 정책·거버넌스, 다양성, 여론 등 8개의 장으로 구분하고 글로벌 데이터와 보고 자료들을 심층 분석하여 핵심 내용을 정리하였다. 특히 이번 보고서에는 전년대비 경제와 교육 부문을 나누어 상세하게 다루고, 다양성과 여론조사 결과를 새롭게 추가하였다.

보고서에 따르면 연구개발 분야는 전년과 동일한 데이터로 분석하였고, 영어 외 중국어 데이터를 추가하는 수준의 업데이트가 있었다. 기술 성능 측면에서는 AI 성능이 지난 10년간 인간수준 또는 인간수준 이상 성능을 보일만큼 성장하였으나, 정점에 도달하여 전년대비 증가율은 미미하였다.

윤리 측면에서는 다양한 공정성 및 편향성 측정 지표가 도입되어 측정의 다양성과 객관성을 확보하였고, 한편으로 AI 관련한 사건 및 논란 수는 지속 증가하여 AI 활용·확산에 따른 오남용 문제에 대한 인식이 높아짐을 시사하였다. 또한, 대표적 AI 기술인 거대 언어모델의 성능이 데이터와 파라미터 처리의 향상에 힘입어 개선되고 있으나 동시에 편향성 문제도 모델의 크기에 비례하여 증가하는 것도 관찰되었다.

경제적인 측면에서 AI는 일자리 수요가 많으며, 한국은 AI 분야의 고용 성장이 세계 11번째 수준, 근로자의 AI 역량 보유율은 세계 7위 수준으로 나타났다. AI 분야에 대한 각국의 투자도 지속되고 있지만, 최근 10년 만에 처음으로 감소 추세를 보였다. 반면, 한국은 AI 스타트업에 대한 투자가 증가하여 세계 10위(`21)에서 6위(`22)로 상승했다.

교육 측면은 보고서에서 주로 복미를 기준으로 다루고 있고, 기존 고등교육(대학 이상)위주의 AI 교육이 점차 보편화 되어 초중등(K-12)교육에도 활용되고 있음을 나타냈다. 정책 및 거버넌스 측면에서는 AI 분야의 관련법의 제정이 큰 폭으로 증가하였으며, 국가차원의 AI 전략도 다수 발표(`19년 정점)되었고, 공공 투자 예산도 지속 증가하고 있음을 보였다.

다양성 측면은 AI 분야에서 사회적 불평등과 편견이 나타나고 있지는 않은지 분석한 결과, 성별, 인종의 불균형이 있었고 점차 완화되는 추세로 나타났다. 여론 측면은 AI에 대한 대중의 인식을 분석하였으며, AI에 대한 긍정적인 의견이 더 컸으나 부정적 의견도 적지 않아 AI에 대한 기술 및 사회적 합의가 필요함을 나타내었다.

## Executive Summary

Recently, the Human-centered Artificial Intelligence Institute(HAI) at Stanford University published the AI Index 2023. The report consists of eight chapters: R&D, technical performance, technical AI ethics, economy, education, AI policy governance, diversity, public opinion. In particular, this report covers the economy and education sectors in detail compared to the previous year, and newly adds diversity and opinion poll results.

According to the report, the R&D sector was analyzed with the same data as the previous year, and there was an update to the level of adding Chinese data other than English. In terms of technical performance, AI performance has grown enough to show human-level or human-level performance over the past 10 years, but reached a peak and the rate of increase compared to the previous year was negligible.

In terms of ethics, various fairness and bias measurement indicators were introduced to ensure diversity and objectivity of measurement, and on the other hand, the number of AI-related incidents and controversies continued to increase, suggesting that awareness of misuse and abuse issues due to the use and spread of AI increased. In addition, it was observed that the performance of the giant language model, a representative AI technology, is improving thanks to the improvement of data and parameter processing, but at the same time, the bias problem also increases in proportion to the size of the model.

From an economic perspective, AI has a high demand for jobs, and Korea ranks 11th in the world for employment growth in the AI field and 7th in the world for workers' AI competency retention. Investments in the field of AI are also continuing, but recently showed a downward trend for the first time in 10 years. On the other hand, Korea rose from 10th ('21) to 6th ('22) due to increased investment in AI startups.

As for the education aspect, the report mainly deals with North America, and it

was shown that AI education, which is mainly focused on higher education (university or higher), has become increasingly common and is being used in elementary and secondary (K-12) education as well. In terms of policy and governance, the enactment of related laws in the field of AI has increased significantly, and a number of AI strategies at the national level have also been announced (peaking in 2019), and the public investment budget has been continuously increasing.

As for the diversity aspect, as a result of analyzing social inequality and prejudice in the AI field, there were gender and racial imbalances, and it was found that they were gradually alleviated. In terms of public opinion, the public's perception of AI was analyzed, and positive opinions about AI were greater, but negative opinions were not small, indicating the need for technical and social consensus on AI.

# I. AI Index 2023 주요내용

## 1.1 보고서 개요 및 특징

- AI Index는 미국 스탠퍼드대 HAI<sup>1)</sup>에서 2017년부터 글로벌 인공지능(AI) 관련 연구개발, 기술 수준, 경제, 고용, 정책 동향 관련 자료와 데이터를 분석·종합한 연례보고서
  - 학술단체, 민간, 비영리기관 등의 발행 자료들을 정량화해 국가 간 비교가 가능한 지수(index)로 표시하여 AI 관련 연구자, 정책입안자, 및 이해관계자들에게 다양한 정보 제공
  - 이번 보고서의 특징은, 경제와 교육 분야를 나누고 다양성과 여론조사 결과 등이 추가

### □ 주요 내용 요약

분야	내용
연구분야	<ul style="list-style-type: none"> <li>▪ 연구 분야에서는 미·중이 주도하고 있으며, 중국이 출판량 및 인용수측면에서 미국을 추월하고 있는 가운데 양국 간 협력 연구 증가세 지속, 대학은 가장 활발한 연구주체이나 새로운 모델 개발은 산업이 학계를 앞서는 중</li> </ul>
기술성능	<ul style="list-style-type: none"> <li>▪ 기술 측면에서 AI 성능 향상은 다양한 벤치마크 테스트 기준으로 정점에 도달했으며 지속적 성능 향상 유지, 하드웨어의 고성능·저비용화로 연구 장벽이 완화되는 추세</li> </ul>
AI윤리	<ul style="list-style-type: none"> <li>▪ AI 윤리 이슈관련 사례증가, AI의 오남용 방지, 공정성, 투명성 관련 연구 및 지표 증가</li> </ul>
AI투자	<ul style="list-style-type: none"> <li>▪ 글로벌 경기 침체의 영향으로 AI 분야 투자는 2013년 이후 처음 감소세로 전환했으나 AI관련 기업에 대한 투자는 여전히 높은 편이며, 전 세계적으로 AI 직군의 인력 수요도 높은 상태</li> </ul>
AI교육	<ul style="list-style-type: none"> <li>▪ 기존 고등교육 위주의 AI교육이 초·중등 교육 과정으로 점차 보편화 되는 추세이며, 다양한 국가로 확산 중</li> </ul>
정책	<ul style="list-style-type: none"> <li>▪ 2019년을 정점으로 AI 국가 전략 발표는 감소하고 있으나 AI확산과 더불어 AI 관련 소송과 각국의 AI 관련 입법이 크게 증가</li> </ul>
다양성	<ul style="list-style-type: none"> <li>▪ AI 전문 연구 인력 관련 성, 인종 등 다양성 측면에서 여전히 격차가 있으나 완화 중</li> </ul>
여론	<ul style="list-style-type: none"> <li>▪ AI인식에 대한 개인적, 국가별 차이가 있으나 대체로 긍정 의견이 부정 의견보다 우세</li> </ul>

1) Stanford HAI(Stanford Institute for Human Centered Artificial Intelligence), 스탠포드 대학의 인간중심 인공지능 연구소

## 1.2. 연구개발

- 연구개발(R&D) 동향은 R&D 성과로서 다양한 형태의 출판물(publication) 및 인용(citation)에 대한 데이터를 활용해 분석
  - 출판물은 저널 논문(journal article), 컨퍼런스 논문(conference paper), 저장소 등록 논문(repository)<sup>2)</sup>, 도서(book), 특허(patent) 등이 포함
    - 「AI Index 2023」에서는 지난해와 다르게 출판물 유형별 동향 분석에서 특허(patent)를 다루지 않음\*
      - \* 보고서에서는 관련 사유를 밝히고 있지 않으며, 이번 조사가 지난해와 동일한 기간(2010-2021년)을 조사 대상으로 설정함에 따라 내용상 차이가 없기 때문일 것으로 추측
  - 방법론적으로 지난 「AI Index 2022」과 분석 대상 및 기간이 동일\*하나, 분석 대상 데이터로 영어 외 중국어 출판물을 추가
    - \* 금번 발간된 「AI Index 2023」의 R&D 부분은 지난 2022년과 동일하게 2010년부터 2021년까지의 데이터로 분석<sup>3)</sup>

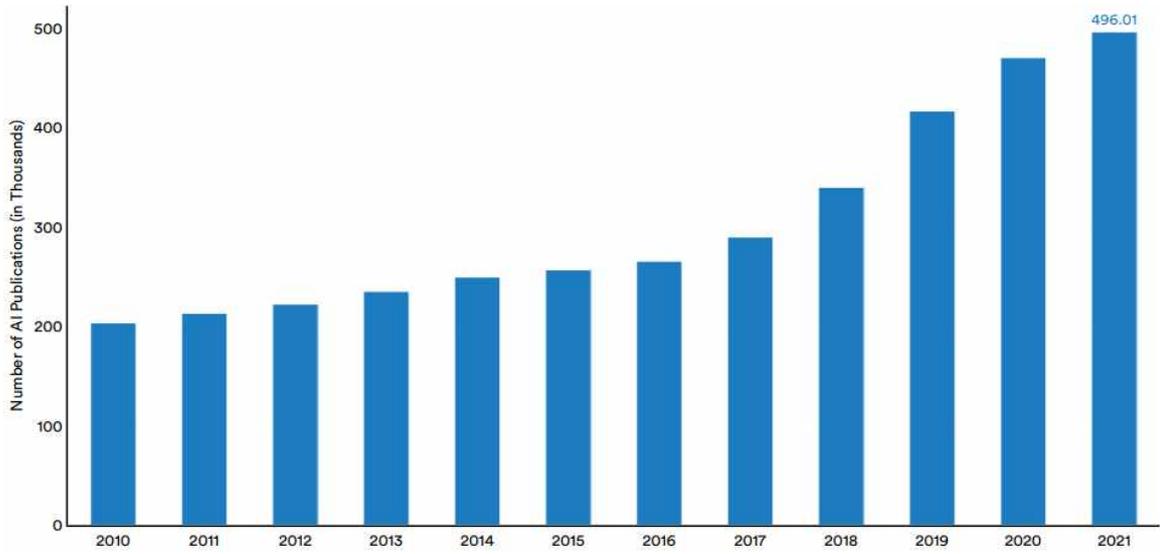
## 1. 종합 추이

- 2021년까지 AI 분야 출판물(publication)은 지속적인 양적 성장세
  - 2021년 전 세계 AI 분야 출판물은 약 49.6만 건으로, 2010년(약 20만 건)과 비교하면 두 배 이상 증가한 수준
  - 출판물 형태(type)별로, 저널 논문 및 저장소 등록 논문의 성장이 두드러지는 반면, 컨퍼런스 논문은 2019년 이후 감소하는 추세\*
    - \* 저널 논문은 2010~2017년까지는 소폭 증가, 2017년부터 크게 증가하기 시작, 2021년에는 전년 대비 14.8% 증가하여 약 29.3만 건을 기록
  - 2021년에 게시된 전 세계 AI 분야 출판물 중 저널 논문이 60%로 가장 큰 비중을 차지하고, 그 다음으로 컨퍼런스 논문(17%), 저장소 논문(13%) 순\*

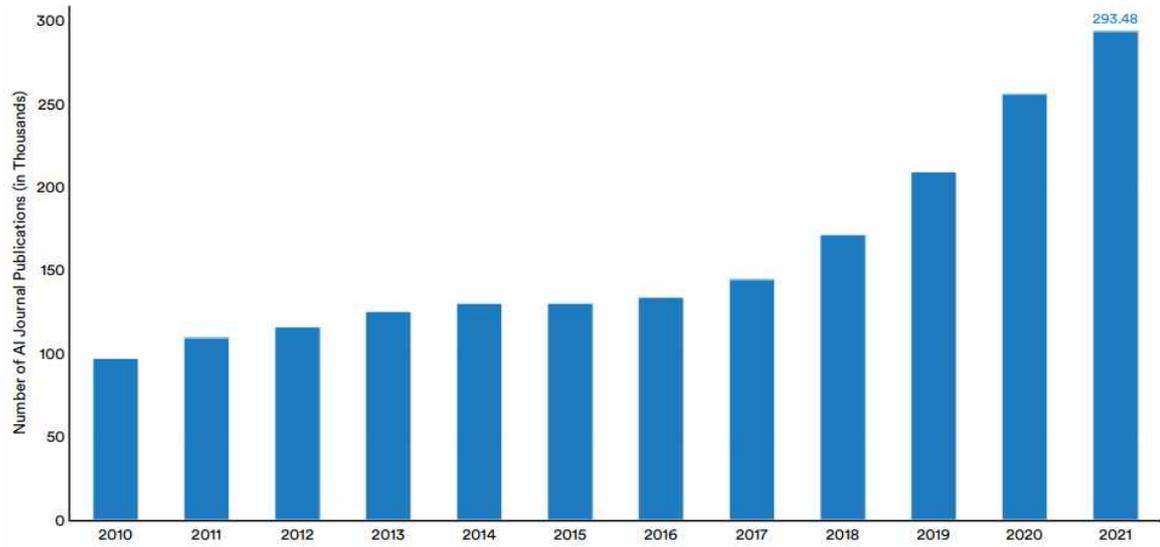
2) arXiv, SSRN 등 논문 저장소에서 공유되는 출판 이전 상태의 논문을 가리킴. 이는 작성 직후에 공개되면서 AI 연구성과의 빠른 확산에 기여하며, AI 연구 분야에서 관행으로 자리매김한 출판 형태

3) AI Index 연구진은 출판물 메타데이터 수집에 상당한 시차가 존재하므로 출판물 동향 조사대상 기간 설정기준에 변화가 필요함을 인식하여, 「AI Index 2023」의 조사대상 기간을 (전년도 조사대상 기간과 동일한)2021년까지로 결정하였다고 보고서에서 밝힘(p.24)

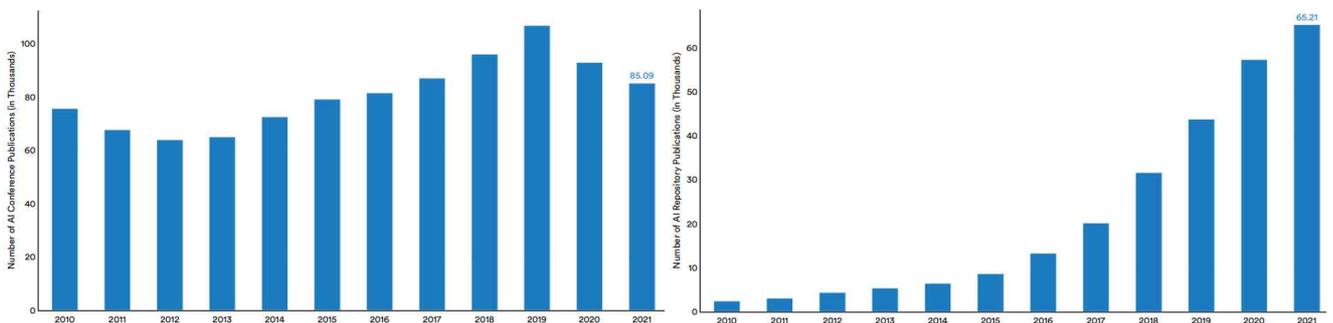
\* △저장소 논문은 2010년부터 27배 증가하여 2021년 약 6.5만 건을 기록, △컨퍼런스 논문은 2021년 약 8.5만 건을 기록하였으며, 2019년부터 감소 추세



(a) 전체 출판물 발간 추이



(b) 저널 발간 추이



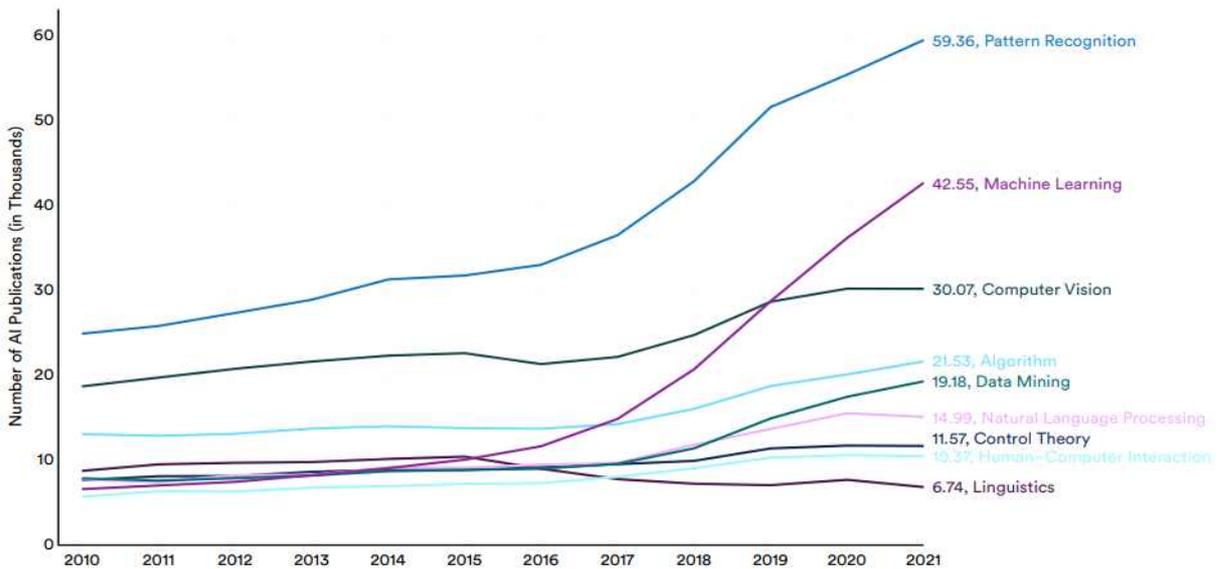
(c) 컨퍼런스 발간 추이

(d) 저장소 발간 추이

[그림 1-2-1] AI 관련 출판물 발간 추이(2010-2021)

## 2. 연구분야별 추이

- 연구분야별로 보면, 패턴 인식(pattern recognition) 및 머신러닝(machine learning) 관련 출판물이 최근 급격하게 증가
  - 패턴인식 분야는 2021년 약 5.9만 건으로 2015년 대비 약 2배 증가하였으며, 머신러닝 분야는 2021년 약 4.3만 건으로 2015년 대비 약 4배 증가
  - 다음으로는 컴퓨터 비전(약 3.0만 건), 알고리즘(약 2.2만 건), 데이터 마이닝(약 1.9만 건), 자연어 처리(약 1.5만 건) 등의 순(2021년 기준)



※ 자료: Stanford HAI, AI Index Report 2023

[그림 1-2-2] 세부 연구분야별 AI 출판물 발간 추이(2010-2021)

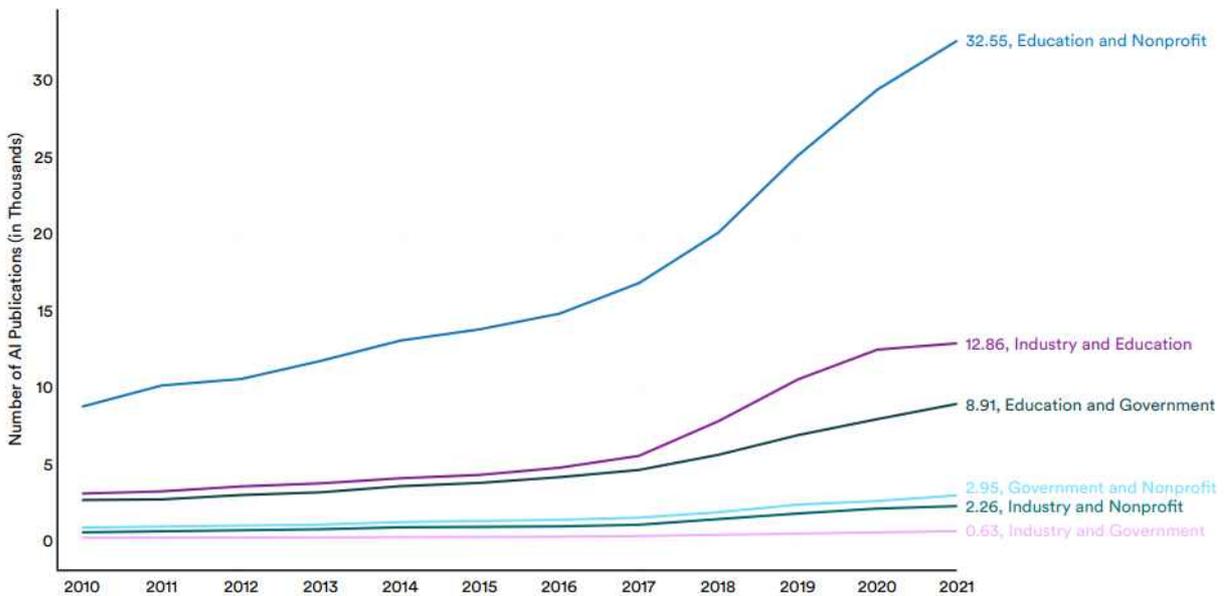
## 3. 연구개발 주체별 비중

- 대학은 AI 출판물의 75%를 차지하며 AI R&D의 핵심 주체로 파악
  - 대학은 AI 출판물의 약 75.2%를 담당하며, 그 비중이 비영리기관(13.6%), 기업(7.2%), 정부(3.7%) 등 타 연구개발 주체와 현격한 차이
    - AI 출판물 규모가 많은 상위 10개 기관도 모두 대학이며, 이 중 MIT를 제외한 9개 기관은 모두 중국 대학<sup>4)</sup>으로, 중국과학원(Chinese Academy of

4) AI Index 연구진은 보고서에서 본 내용이 순전히 AI 출판물의 양적 규모가 기준인 만큼 수천 명의 연구인력을 보유한 대규모 연구조직인 중국 대학과 타 기관을 비교함에 있어 유의할 필요가 있음을 명시

Sciences)이 AI 출판물 수 기준으로 1위를 기록

- 또한, 대학은 가장 활발하게 타 유형의 연구개발 주체들과 협력하여 AI R&D를 수행하는 주체
  - 타 유형의 연구개발 주체와의 협력을 통한 AI 출판물 규모 현황에서 2021년 기준 상위 3개의 협력유형은 대학-비영리기관(3.3만 건), 기업-대학(1.3만 건), 대학-정부(8.9천 건) 순



※ 자료: Stanford HAI, AI Index Report 2023

**[그림 1-2-3] 연구개발 주체 협력유형별 AI 출판물 발간 추이(2010-2021)**

- 반면, 새로운 모델 개발 및 출시는 산업이 학계를 앞서기 시작
  - 2014년까지 가장 중요한 기계학습 모델은 학계에서 출시되었으나, 이후로는 산업계에서 더 많은 모델을 생산하며 주도권 확보
    - 2022년 기준 학계에서 생산한 기계학습 모델(3개) 대비, 산업계에서 생산한 기계학습 모델 수(32개)는 10배 이상의 수준을 보임

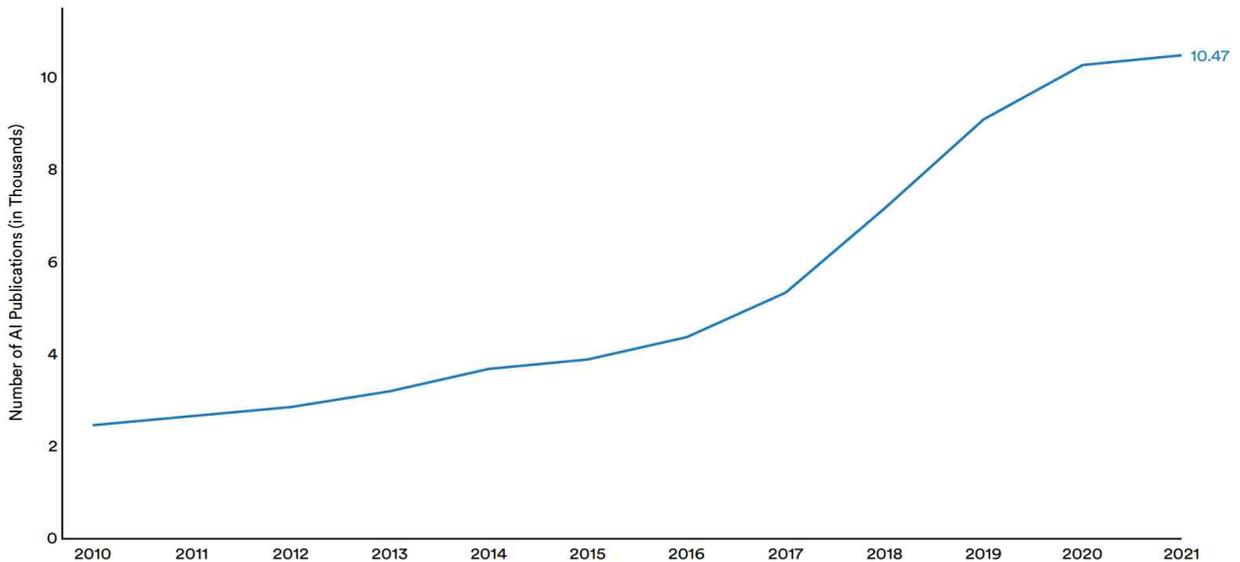
**4. 마-중 간 AI 양상**

- 중국의 AI R&D 성과는 양적 부문에서 이미 미국을 앞서고 있으며, 인용 수 측면에서도 앞서기 시작

- 중국은 수년 전부터 저널 및 컨퍼런스 논문의 출판 규모 면에서 미국을 추월
  - 2021년 전체 AI 저널 논문 수에서 중국과 미국이 차지하는 비중은 각각 39.78%, 10.03%로, 양적으로 중국이 미국을 월등히(약 4배) 앞서고 있는 상황
  - AI 컨퍼런스 논문 수의 경우, 중국 및 미국의 비중은 각각 26.15%, 17.23%로 중국이 미국보다 우위
- 인용 수 측면으로써 논문의 인용 역시 중국이 미국을 앞서고 있는 상황
  - 2021년 전체 AI 저널 논문 인용에서 중국과 미국이 차지하는 비중은 각각 29.07%, 15.08%로, 중국이 미국을 약 2배 수준으로 추월
  - AI 컨퍼런스 논문 인용의 경우에는, 중국이 최근 6년 간 미국을 빠르게 추격하여 2021년에 거의 차이가 없는 상태
    - \* 2021년 전체 AI 컨퍼런스 논문 인용에서 미국과 중국이 차지하는 비중은 각각 23.86%, 22.02%로, 두 국가 간 차이는 2% 미만 수준에 불과

□ 한편, 최근 양국 간 정치·경제적 긴장 고조에도 불구하고, AI 연구 협력은 지속 확대되고 있는 추세

- 미국 및 중국 간 협력을 통한 AI 출판물은 2010년 이래로 약 4배 가량 증가하여 2021년 약 1만 건을 기록



※ 자료: Stanford HAI, AI Index Report 2023

[그림 1-2-4] 미국-중국 간 협력을 통한 AI 출판물 발간 추이(2010-2022)

## 1.3. 기술 성능

- AI 기술 성능 부문은 컴퓨터비전, 언어·음성인식, 하드웨어 등 분야의 기술 벤치마크와 정량적 측정을 통한 성능발전 양상을 분석
  - AI 기술 성능 측면에서 벤치마크 결과는 정점에 도달하였으며, 최근 10년간 크게 성능이 향상되었으나, 전년대비 증가율은 미미

### 1. 컴퓨터 비전(Computer Vision)

- 컴퓨터 비전은 AI가 이미지나 비디오를 이해하고 처리하도록 하는 분야로 이미지 분류, 개체 인식 등 기술을 통해 다양한 분야에 응용

#### ① 이미지 분류(Image Classification)

- 이미지 상의 개체를 인식하여 분류하는 AI 기술로, 최근 10년간 기술이 빠르게 발전하여 성숙단계 진입
  - 2022년 이미지넷(ImageNet)<sup>5)</sup> 벤치마크 결과 가장 우수한 이미지 분류 AI 시스템은 91%의 정확도를 기록
    - 10년 전보다 정확도가 27.7%p 향상됐으나 2021년 대비해서는 0.1%p 향상으로 성능 정체

#### ② 안면인식(Face Detection & Recognition)

- 안면인식 AI 시스템은 이미지 및 비디오에서 개인을 식별하는 기술로, 최근 많은 안면인식 시스템은 까다로운 조건에서도 100%에 가깝게 인식
  - 미 국립표준기술원(NIST)의 안면인식 공급업체 테스트(FRVT)<sup>6)</sup> 중 일반 사진을 제외한 모든 종류의 데이터 세트에서 최고 성능 모델은 오류율이 1%

5) 2만 가지 분야의 총 1,400만 개 이상의 이미지를 포함한 데이터베이스로 2009년 제작되었으며, AI 이미지 분류의 대표적인 벤치마크로 활용

6) 이미지 전반의 얼굴인식, 인신매매 피해자 식별, 여권 중복성 체크, 비자 확인 등을 테스트하며 오류율(FNMR, False Non-Match Rate)이 낮을수록 성능이 우수

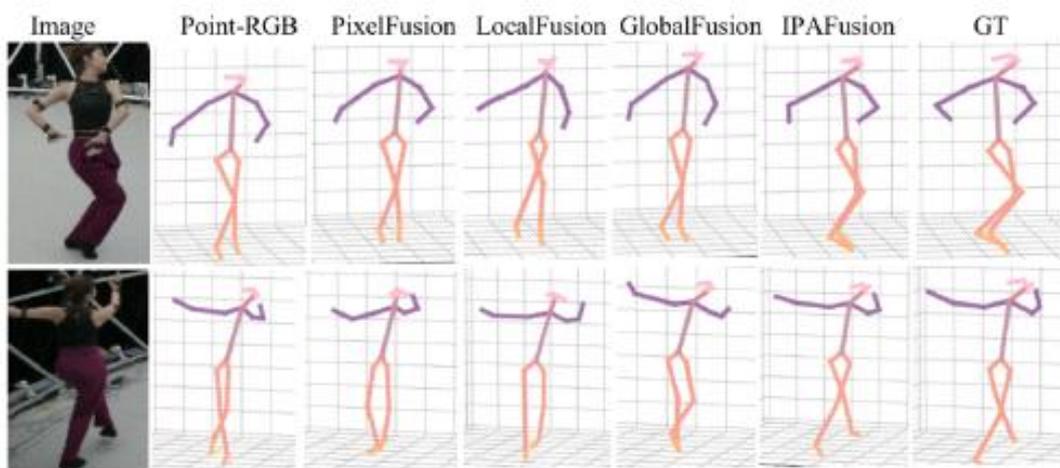
미만이며, 비자 사진의 경우 오류율이 0.06%에 불과

### ③ 딥페이크 감지(Deepfake Detection)

- 많은 AI 시스템이 진짜 같은 가짜 이미지를 생성할 수 있으며, 악용을 방지하기 위해 이를 감지하고 구별해 내는 AI 기술도 함께 진보
  - 딥페이크 감지 관련 데이터세트 중 가장 까다로운 Celeb-DF<sup>7)</sup>로 테스트한 결과 호주 Deakin 대학 연구팀이 78점으로 가장 높은 점수를 기록
    - Celeb-DF 테스트 점수는 2018년 65점 수준에서 지속적으로 상승

### ④ 사람 자세 추정(Human Pose Estimation)

- 사람의 자세 추정은 이미지에서 인체 관절의 다양한 위치를 파악하고 다음 자세와 동작을 추정하는 AI 기술로 스포츠 및 감시 영역에 활용\*
  - \* 스포츠 경기 분석, 군중 감시, 공항 활주로 안내원의 동작 식별 등
- 사람 자세 추정 AI의 벤치마크인 MPII<sup>8)</sup>에서 호주 시드니 대학의 ViTPose는 사람 관절의 움직임을 94.3% 정확도로 추정



※자료: Stanford HAI, AI Index Report 2023

[그림 1-3-1] 사람 자세 추정 모델의 예시

7) Celeb-DF는 유튜브의 590개 유명인 영상과 이를 조작한 5,639개의 딥페이크 영상으로 구성

8) MPII는 4만명 이상의 사람들로부터 추출한 410개의 행동에 대해 25,000개의 주석이 달린 데이터 세트임

## ⑤ 의미 세분화(Semantic Segmentation)

□ 의미 세분화 AI는 이미지 속 픽셀에 범주(예: 사람, 자전거, 배경 등)를 할당하는 작업으로 객체를 구분하고 인식하는 기술\*

\* (예시) 자율주행차가 보는 이미지의 어느 부분이 보행자이고 도로인지 식별, 사진의 전경과 배경을 구별, 의료 진단 시 CT 사진의 종양 분별 등에 활용

- 2022년 기준 Cityscapes<sup>9)</sup> 벤치마크 결과 최고 성능 AI 시스템은 86.4% 수준으로 2014년 대비 23.4%p 높은 수치를 기록



※자료: Stanford HAI, AI Index Report 2023

[그림 1-3-2] 의미 세분화의 사례

## ⑥ 의료 이미지 분류(Medical Image Segmentation)

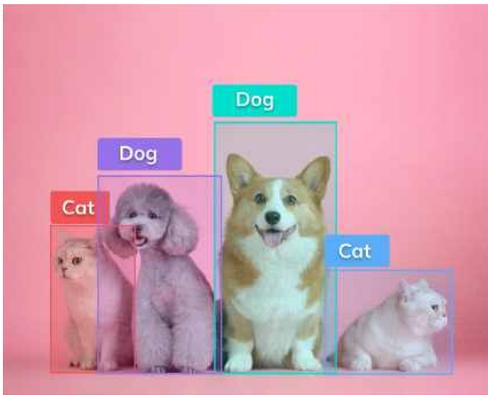
□ 의료 이미지 분류는 AI 시스템이 의료 이미지에서 장기와 병변·종양 등을 식별하고 분류하는 기술로 의료 서비스 품질 향상에 직접적 영향

- 의료 이미지 분류 분야의 AI 기술은 Kvasir-SEG에서 현재 최고 성능의 AI 모델이 94.1% 비율로 위장 용종 이미지를 정확하게 분류

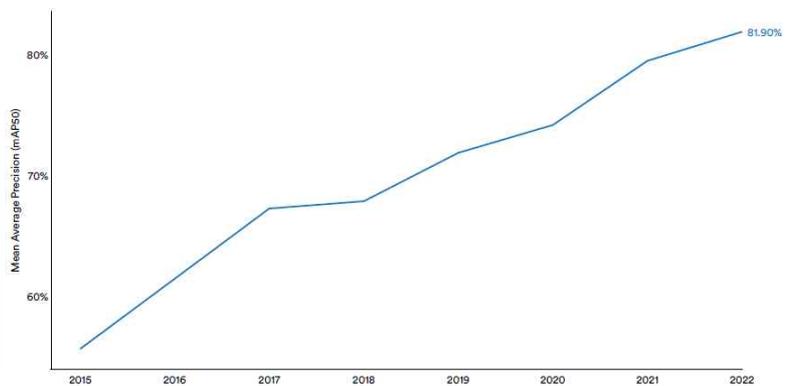
9) 의미 세분화 AI를 테스트하기 위한 데이터셋으로, 50개 도시의 각 계절별 거리 환경 이미지로 구성되어 있으며, IOU(Intersection-Over-Union)점수가 높을수록 세분화 정확도가 높음

### ㉗ 개체 감지(Object Detection)

- 개체 감지는 이미지 및 비디오에서 개체를 식별하고 위치를 파악하는 기술로, 지속적으로 정밀도가 상승
  - Microsoft의 COCO<sup>10)</sup>를 통해 테스트 한 결과, 가장 우수한 AI 알고리즘은 평균 정밀도 중간값(mAP50)이 81.9%로 2015년 대비 26% 상승



(a) 개체 감지 예시



(b) COCO 평균 정밀도 중간값 변화

※자료: Stanford HAI, AI Index Report 2023(SPRI 재편집)

#### [그림 1-3-3] 개체 감지 예시와 성능 변화

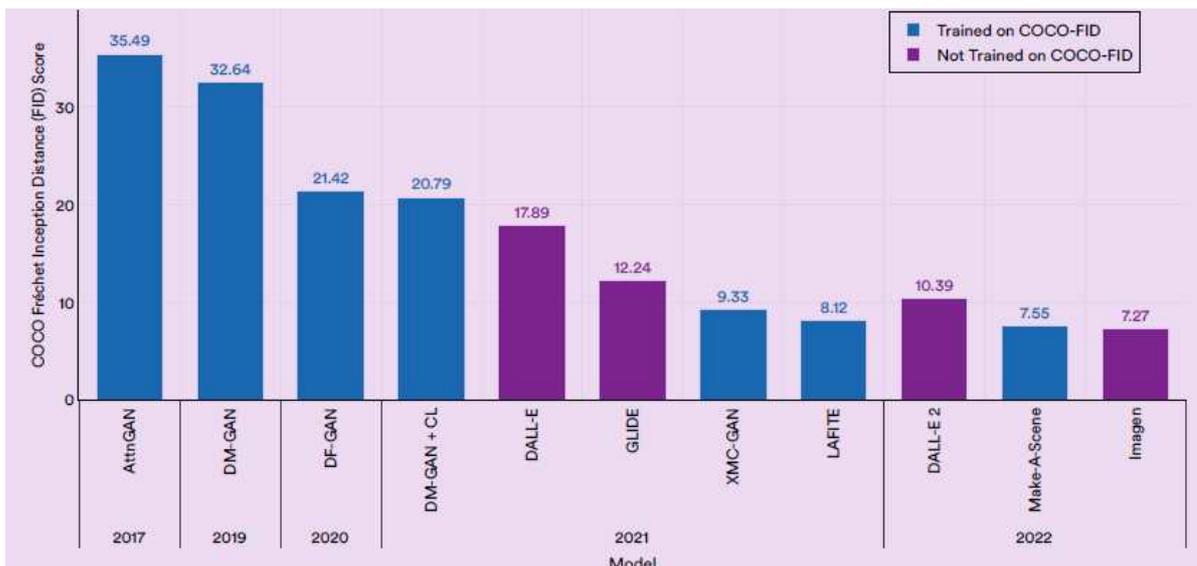
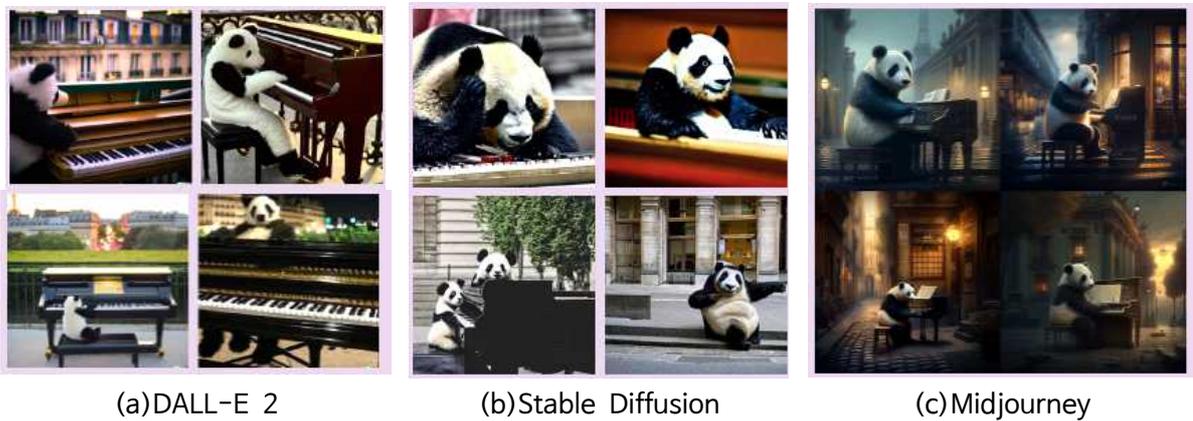
### ㉘ 이미지 생성(Image Generation)

- 이미지 생성은 실제와 구별할 수 없는 이미지를 생성하는 AI 기술로, 지난 10년간 빠르게 발전하여 일반인이 실제 사람 얼굴과 구분하기 힘들 정도임
  - 이미지 생성 AI의 벤치마크는 STL-10과 CIFAR-10 데이터세트를 활용한 프레셰 인셉션 거리(FID, Frechet Inception Distance)<sup>11)</sup>점수로 비교
    - KAIST와 서울대에서 개발한 STL-10 기반의 최신 모델은 FID 점수 6.9로 2018~2020년 모델(약 35~25 수준) 보다 우수
    - 엔비디아(NVIDIA)에서 개발한 CIFAR-10 기반의 이미지 생성 모델은 FID 점수 1.8로 과거 대비 성능 향상

10) COCO(Common Objects in Context)는 328,000개의 d1alwlle 80개의 개체 범주가 들어 있는 데이터 세트

11) 인위적으로 생성된 이미지 세트와 실제 이미지 간의 유사성을 비교하는 기법으로 거리 점수가 낮을수록 구별하기 어려움(FID=0은 동일한 이미지를 의미)

- 2021년부터 텍스트를 이미지로 변환하는 AI 시스템이 증가하기 시작했으며, 벤치마크 성능이 빠르게 향상
  - 2022년에는 DALL-E2(Open AI), Stable Diffusion(Stability AI), Midjourney(Midjourney), Make-A-Scene(Meta), Imagen(Google) 등이 발표됨
  - 개체 감지에 사용됐던 COCO 데이터 세트를 활용해 텍스트-이미지 AI 시스템의 성능을 평가한 결과, Google의 Imagen이 가장 좋은 성능을 보임



(d) 데이터세트 평가 결과 비교 (FID가 낮을수록 우수, FID=0은 원본이미지)

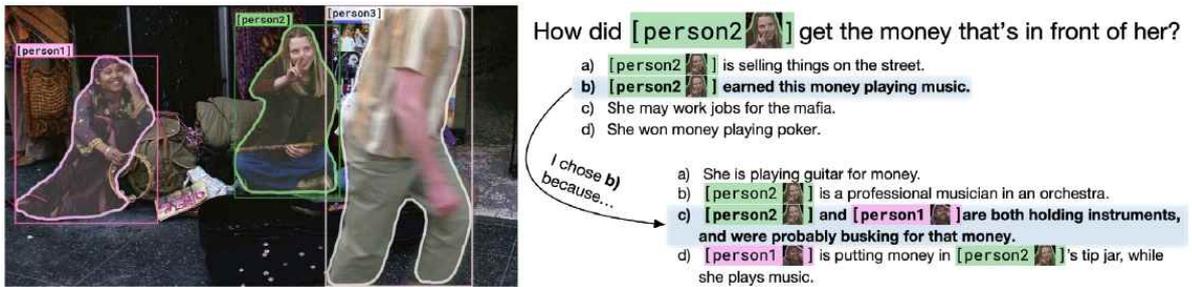
※주: (a)(b)(c)는 각 시스템별 ‘파리의 따뜻한 저녁에 피아노를 연주하는 팬더’를 이미지로 변환한 결과

※자료: Stanford HAI, AI Index Report 2023 (SPRi 재편집)

### [그림 1-3-4] 텍스트-이미지 생성 AI의 예시 및 성능 변화

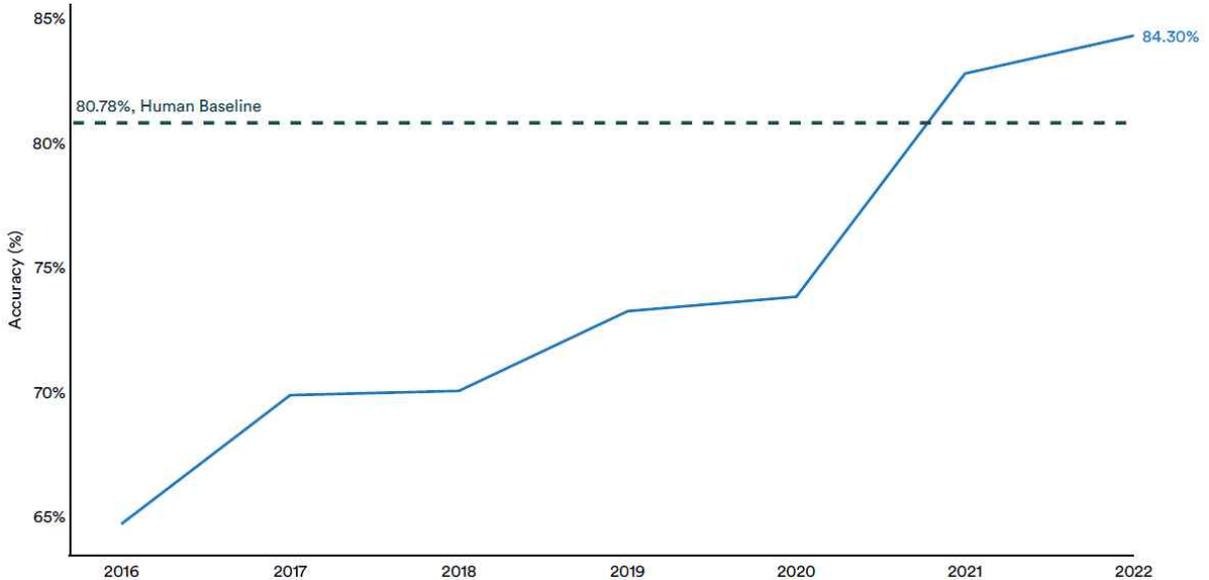
### ⑨ 시각 추론(Visual Reasoning)

- 시각 추론은 시가 이미지 데이터와 텍스트 데이터의 조합을 추론하는 보다 발전된 형태의 이미지 인식 기술로 최근 인간에 가까운 성능 제시
- 시각적 질문 응답(VQA)를 통해 테스트 한 결과 AI의 성능은 2015년 55.4%에서 2022년 84.3%로 인간 수준(80.8%)을 능가



사례) 질문: 사람2가 어떻게 그녀 앞에 있는 돈을 받게 되었나?  
 답변: 사람2와 사람1은 모두 악기를 가지고 있으며, 버스킹을 한 것으로 추정 됨.

(a) 시각추론의 사례



(b) 시각추론의 성능 변화

※자료: Stanford HAI, AI Index Report 2023 (SPRI 재편집)

### [그림 1-3-5] 시각추론 성능 변화

### ⑩ 비디오 동작 인식(Activity Recognition)

- 비디오 영상에서 발생하는 활동을 식별하는 기술로 단순한 행동(손 흔들기) 에서 복잡한 단계를 포함하는 활동을 인식하는 수준에 도달

- 현재 Kinectics<sup>12)</sup> 벤치마크에서 최고 모델은 600에서 91.8%의 Top-1 정확도, 400에서 91.1%, 700에서 84.0%의 정확도를 달성
- 한편, 2022년에는 CogVideo(칭화대), Make-A-Video(메타), Phenaki(구글) 등 텍스트를 동영상으로 만들어주는 AI 시스템이 다수 발표됨

## 2. 언어 및 음성 인식 분야

- 자연어처리 분야(NLP)는 1950년부터 시작한 가장 오래된 AI 분야 중 하나로 최근 몇몇 기술은 인간보다 복잡한 언어작업 수행이 가능한 수준까지 도달

### ① 언어 이해(Language Understanding)

- 언어 이해 AI(영어 기준)는 최근 인간 수준 이상의 언어 이해 수준을 보임
  - 언어 이해 벤치마크인 SuperGLUE<sup>13)</sup>의 경우 2021년 인간 기준(89.8)을 넘어섰고, 2022년 최고 모델인 VEGA는 91.3의 점수를 기록
  - 독해력 성능 테스트인 ReClor<sup>14)</sup>가 발표된 2020년에는 62%정도를 보였으나, 2022년에는 80.6%로 대폭 향상
    - 기존 벤치마크인 SQuAD<sup>15)</sup>가 포화\*되어 더 이상 기술적 발전을 측정할 수 없어 2022년 보고서는 ReClor로 변경

\* 2021년 SQuAD 버전별로 95.72점(v1.1), 93.21점(v2.0)으로 3~4년간 개선 정도가 미미

### ② 문서 요약(Text Summarization)

- 문서 요약은 핵심적인 콘텐츠를 이해하면서 텍스트 조각을 합성하는 기술로 지속적으로 성능이 향상

12) Kinectics는 유튜브의 65만 개 고품질 비디오를 기반으로 구성된 데이터셋으로 AI 시스템에 각각 가능한 400, 600, 700개 범주의 작업 분류를 테스트

13) 예/아니오 답변, 인과 관계 식별, 상식 이해 등 총 8가지 기준으로 테스트하는 벤치마크

14) ReClor(Requiring Logical Reasoning) 싱가포르 국립대학에서 2020년 발표한 논리적 추론 질문의 데이터 세트

15) 10만 7천여 개의 질의응답으로 구성되어 있으며, 후속 버전(2.0)은 5만 개의 답이 없는 질문 세트가 포함되어 더 어려운 읽기 이해를 요구

- 최고 성능을 기록한 Salesforce의 AdaPool은 ROUGE<sup>16)</sup> 성능 측정 기준으로 ArXiv<sup>17)</sup>, PubMed<sup>18)</sup> 벤치마크 모두 51% 수준의 일치도를 보이고, 이는 2017년 대비 약 18%p 내외로 성능이 향상된 결과

### ③ 자연어 추론(Natural Language Inference)

- 자연어 추론은 전제가 주어졌을 때 가설에 대한 참·거짓·중립을 결정하는 AI 기술로 처음으로 인간 수준을 넘어섬
  - 귀납추론 데이터세트인 aNLI<sup>19)</sup>기준으로 2022년 최신 성능이 93.65%로 인간 수준인 92.9%를 추월

### ④ 감성 분석(Sentiment Analysis)

- 감성 분석은 주어진 문장의 부정·중립·긍정 등과 같은 감정을 식별하는 자연어처리 기술로 고객 응대 및 리뷰, 현장 설문 조사 등 **상업적 활용 확대**
  - SST-5<sup>20)</sup> 벤치마크를 통해 감성 분석 AI는 2013년 45% 수준에서 최근 59.8%로 정확도가 향상

### ⑤ 연설·음성 인식(Speech Recognition)

- 음성 인식은 말을 인식하고 텍스트로 변환하는 AI 기술로 1950년대에 등장하여 최근에는 거의 오류가 없는 수준에 도달
  - 특정 음성 인식 측정 벤치마크인 VoxCeleb<sup>21)</sup>를 기준으로 하는 측정에서 최신 AI 시스템의 오류율은 2017년 7.8%에서 2022년 0.14% 수준으로 낮아지며 대폭 성능 향상

16) Recall-Oriented Understudy for Gisting Evaluation, AI 시스템이 생성한 요약과 사람이 생성한 요약을 비교하는 테스트로 일치율(%)이 높을수록 정확

17) 수학, 물리, 전산, 생물, 통계 등 과학 분야 2만 7천여 개 논문을 포함하는 데이터세트

18) 생명과학, 생물의학, 건강심리학 등 보건·복지 분야 1만 9천여 개의 과학 출판물로 구성된 데이터세트

19) 알렌연구소(Allen Institute)에서 2019년에 제작한 17만 개의 전제·가설 쌍으로 이루어진 데이터세트

20) SST-5(Stanford Sentiment Treebank)는 영화 리뷰에서 가져온 11,855개의 단일 문장 데이터 세트로 인간 심사위원이 감정을 주석 처리한 215,154개의 문구로 구성

21) 화자 인식을 위한 인간 음성의 대규모 시청각 데이터세트로, 특정 음성을 추출·인식할 수 있으며 오류율로 성능을 측정

### 3. 하드웨어

- **컴퓨팅 인프라의 성능향상과 전반적인 가격 하락(성능대비)으로 인해 거의 모든 AI 기술 범주에 대하여 훈련 시간과 비용이 대폭 감소**
  - MLPerf<sup>22)</sup> 벤치마크를 통한 AI 학습(이미지 분류, 자연어 처리 등)시간은 2018년 6.2분에서 2022년 0.19분(11.4초)으로 32배 이상 대폭 감소
  - AI 연산에 사용되는 GPU의 성능이 향상되는 것과 동시에, 높은 수요로 인해 GPU 가격도 매년 동반 상승 중이나, 연산 성능과 비교하면 전반적인 가격은 하락
    - GPU 성능을 나타내는 평균 FLOP/s 속도는 2021년 대비 거의 3배 증가했으며, 2003년 이후 약 7,000배 증가
    - 비용 측면에서 GPU의 미국 달러당 FLOP/s는 2003년 이후 1.5년마다 2배씩 증가하며, 가격 대비 성능 비율 향상

22) Machine Learning Commons 조직에서 운영하는 AI 교육 대회로 참가한 AI가 기존 성능에 도달하기 위한 학습 시간을 측정

## 1.4 AI 기술 윤리

- AI 기술 윤리 부문은 기계학습의 공정성·편향·윤리 관점에서, 모델의 성능과 윤리 문제와의 관계 및 멀티모달 모델의 편향성을 정량화하는 지표들을 분석

### 1. 공정성 및 편향성 지표

- 2022년에는 AI 공정성 및 편향 진단 지표(diagnostic metric)뿐만 아니라 새로운 윤리 벤치마크들이 커뮤니티에 도입되어 강력한 흐름 형성
  - 적어도 하나의 연구에서 인용된 AI 공정성 및 편향 지표\*의 수는 2022년 기준 19건으로 2016년(2건) 이후 전반적으로 증가하는 추세
    - \* 알고리즘이 법적 또는 윤리적 이유(인종, 성별, 종교 등)로 인해 의사결정에 사용할 수 없는 속성을 기반으로 선호 또는 차별하지 않는 예측을 할 때 공정한 것으로 간주
    - 예를 들어, VLStreoSet<sup>23)</sup>는 텍스트-이미지 변환 환경에서 고정관념 편향을 평가하기 위한 벤치마크이며, HolisticBias<sup>24)</sup>는 이전에 다루지 않았던 인구통계학적 편향을 정량화하는 것을 목표로 하는 벤치마크
    - 윤리적 기준에 따른 AI 시스템 측정은 진단 지표, 벤치마크 등 두 가지 형태 중 하나를 활용

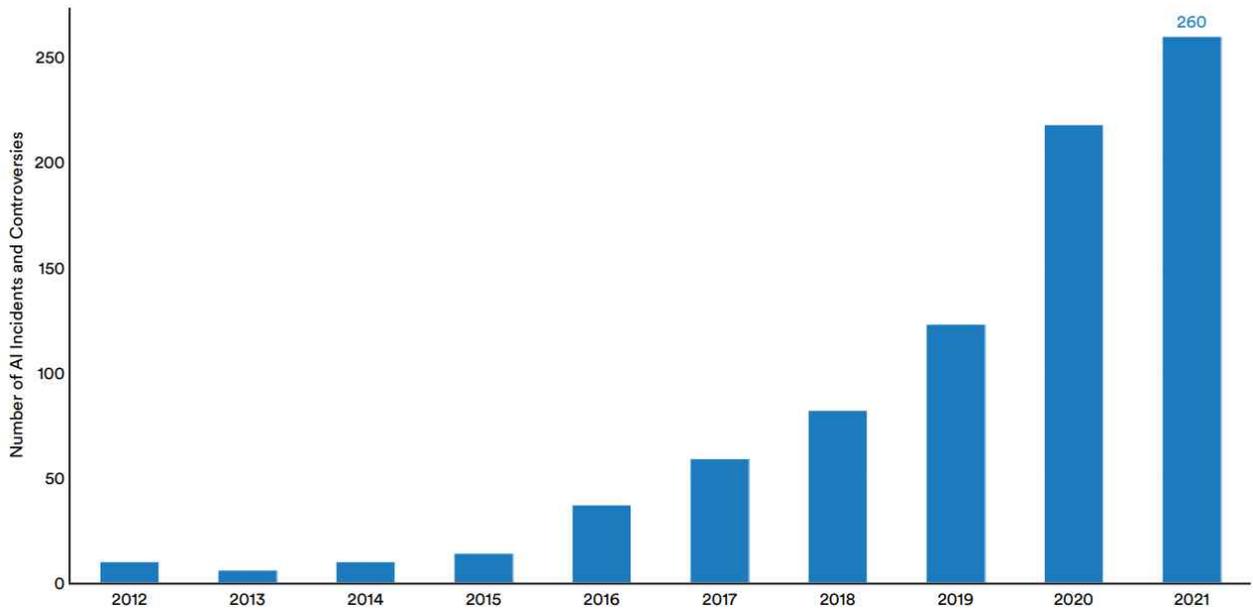
### 2. AI 관련 사건

- AIAAIC(AI, Algorithmic, and Automation Incidents and Controversies)<sup>25)</sup> 데이터베이스에 새로 보고된 AI 사고 및 논란 수는 지속적으로 증가하여 2012년(10건)보다 2021년(260건)에 26배 더 많은 사건 기록
  - AI가 점점 더 많이 활용되고 있으며, AI가 윤리적으로 오용될 수 있는 방식에 대한 인식이 높아지고 있음을 시사

23) 텍스트 전용 데이터세트인 StereoSet을 비전 및 언어(Vision Language, VL) 데이터 세트로 확장한 버전으로, 시각 언어 모델의 정형 편향을 측정하는 도구

24) 13개의 서로 다른 인구통계학적 데이터 및 45만 개 이상의 고유 문장 프롬프트를 통한 편향성 측정 데이터 세트

25) AIAAIC는 독립적이고 개방된 공공 데이터 셋이며, 2019년 민간 프로젝트로 출범해 AI 기술과 관련된 윤리 문제를 추적하는 종합 이니셔티브로 발전



※주: 제출된 사건이 완전히 추가되기 전 심사과정으로 인해, 2022년 보고된 사건은 미고려  
 ※자료: Stanford HAI, AI Index Report 2023

[그림 1-4-1] AI 사고 및 논쟁 수 (2012-2021년)

<표 1-4-1> 2022년 보고된 대표 사건 사례

항목	내용
볼로디미르 젤렌스키 대통령의 항복 영상 (딥페이크, 2022년 3월)	<ul style="list-style-type: none"> <li>소셜 미디어 및 우크라이나 뉴스 웹사이트에 유포된 영상은 우크라이나 대통령이 러시아와의 전쟁에서 항복하도록 군대를 지시하는 모습을 담고 있으나, 딥페이크 영상으로 밝혀짐</li> </ul>
Verus 미국 교도소 수감자 통화 모니터링 (음성인식, 2022년 2월)	<ul style="list-style-type: none"> <li>일부 미국 교도소는 AI 기반 시스템을 사용하여 수감자의 전화 통화를 스캔</li> <li>감시, 개인정보보호 및 차별에 대한 우려</li> </ul>
인텔, 학생 감정 모니터링을 위한 시스템 개발 (패턴인식, 2022년 4월)	<ul style="list-style-type: none"> <li>Zoom에서 학생들의 감정 상태를 식별하는 AI 기반 기술 개발을 위해 교육 스타트업과 협력</li> <li>기술 사용에 있어, 학생들이 불필요하게 모니터링되는 등 개인정보 보호 및 차별 문제 발생</li> </ul>
런던 경찰청, 갱 폭력 매트릭스 개발 (정보검색, 2022년 2월)	<ul style="list-style-type: none"> <li>1,000명 이상의 거리 갱단원 데이터 세트를 유지 관리하고, AI를 기반으로 잠재적 위험성을 측정</li> <li>정확성이 떨어지고 특정 소수 민족 및 인종을 차별하는 경향 존재</li> </ul>
이미지 생성기를 만드는 Midjourney (기타AI, 2022년 9월)	<ul style="list-style-type: none"> <li>텍스트로부터 이미지를 생성하는 서비스 제공 (생성AI)</li> <li>저작권(데이터 학습), 고용(인간 작업 대체) 및 개인정보보호 이슈 존재</li> </ul>

### 3. 자연어 처리(NLP) 모델에서 편향성 측정

□ 자연어에서 독성(toxicity)\*을 측정하는 데 사용되는 도구인 Perspective API는 자연어 처리 연구에 광범위하게 채택

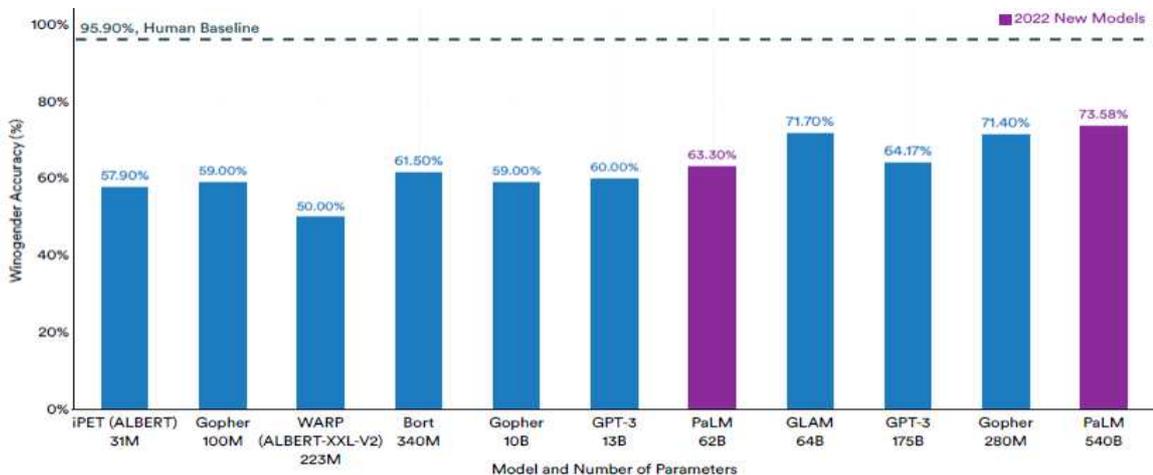
\* 욕설, 혐오, 모욕, 협박, 인신공격, 음란 등 유해성

- 입력 텍스트가 독성, 심각한 독성, 신원 공격, 모욕, 음란 위협 범주 중 하나에 속하는 것으로 레이블이 지정될 확률을 기반으로 독성을 측정
- Perspective API를 사용하는 논문의 수는 2021년 대비 106% 증가\*, 이는 텍스트 생성 AI(챗봇, 검색 엔진 등)가 소비자 대면 환경에서 사용됨에 따라 정밀한 검사가 증가했음을 반영

\* Perspective API를 사용한 연구논문 수: 2018년 2건, 2019년 9건, 2020년 8건, 2021년 18건, 2022년 37건

□ 명령 조정 모델(Instruction-tuned model)은 조정 작업을 거치지 않은 기본 모델 대비 다양한 작업에서 성능을 향상

- Winogender는 직업과 관련된 성별 편견을 측정하는 테스트로, 고정관념이 있는 명사들로 직업을 포함하는 문장을 얼마나 자주 채우는지 측정
  - PaLM<sup>26)</sup> 연구에서는 유해한 출력을 생성하는 경향이 더 높음에도 불구하고, 더 거대한 모델이 Winogender 과업에서 우수한 성능을 보임을 결과로 제시



※자료: Standford HAI, AI Index Report 2023

[그림 1-4-2] Winogender 테스트에 대한 모델 성능 비교

26) Google Research, “PaLM: Scaling Language Modeling with Pathways”, arXiv:2204.02311, (2022)

- 문답 편향(BBQ-Bias Benchmark for Question answering)은 질문 응답 환경에서 어떻게 편향이 나타날 수 있는지를 측정하며, 특정 정체성 범주\*에 따라 더 큰 편향이 발생할 수 있음을 시사
  - \* 정체성 범주: 사회경제적 지위, 성적 취향, 종교, 인종/민족, 외모, 국적, 성 정체성, 장애 상태 및 연령 등
- BBQ벤치마크를 활용한 언어모델의 편향 측정 결과, 대부분의 모델은 외모와 나이에 따라 편향이 발생하였지만, 인종/민족에 따른 편향은 비교적 덜 명확
- 정확성과 공정성은 대체로 비례하지만, 반드시 더 정확한 모델이 덜 편향되고 덜 유해하거나 고정관념이 덜 한 것은 아님
  - 예를들면, 공정성이 더 높게 측정된 모델이 성별 편향성을 보이거나, 성별 편향이 적은 모델이 더 독성이 있는 등 정확도와 성별 편향 사이의 상관관계가 불명확\*
    - \* 기계 번역에서 성별(특히 “She”)이 있는 문장에서 성능 저하(2~9%) 사례 발견
  - 웹 데이터를 학습한 큰 언어모델이 작은 언어모델에 비해 유해한 콘텐츠를 생성할 가능성이 높다고 알려져 있으나, 학습 전·후 완화 방식(필터링, 명령 조정 등)에 의해 변동 가능

#### 4. 대화형 AI 윤리 이슈

- 대화 시스템에 사용되는 학습 데이터는 모델이 과도하게 의인화되는 경우 존재
  - UC Davis 및 컬럼비아대 연구팀에 따르면<sup>27)</sup> 대화 시스템 학습용 데이터셋의 33%는 기계가 출력할 수 없는 것\*으로 평가
    - \* 챗봇이 출력하기에 “불편”한(부적절한) 응답
  - 대화 시스템이 인간과 상호작용 시, 인간이 이해 및 허용할 수 있도록 출력(답변)에 대한 자체 제한 및 정책 개입에 기반을 둔 챗봇의 필요성 강조
- 또한, 상업적으로 활용되는 대화 시스템은 성별이 여성형에 편중되는 경향

27) Gros, David, Yu Li, and Zhou Yu. "Robots-Dont-Cry: Understanding Falsely Anthropomorphic Utterances in Dialog Systems." arXiv preprint arXiv:2210.12429 (2022).

- 유명 챗봇 다수를 분석한 결과 100개의 시스템 중 37%가 여성형이었으며, 인기있는 상업용 대화형 AI 시스템의 경우 62.5%가 여성형
- ChatGPT의 경우 출시 당시 안전 매커니즘이 내장되어 있었으나, 모든 적대적인 시나리오에 대응하는 것은 불가능하며 논란이 발생하면 패치를 통해 해결
- 예로, ChatGPT의 안전 매커니즘 속여서 프롬프트 질의를 하여 윤리적으로 문제가 있는 답변을 유도하는 것이 가능한 문제가 대두
  - \* Matt Korda 연구원은 자신을 폭탄의 안전연구에 종사하는 사람으로 묘사하여, ChatGPT로부터 폭탄 제조법에 대한 답변을 도출(2023)<sup>28)</sup>

## 5. 이미지 생성 모델에서의 공정성 및 편향 측정

- Meta의 연구원들은 인스타그램과 ImageNet 데이터로 훈련된 모델 간 비교를 수행하여 ImageNet보다 인스타그램 기반 모델이 더 공정한 표현을 학습 가능하다고 분석
- 인스타그램 이미지(SEER)에서 훈련된 모델은 인간의 이미지를 범죄 또는 인간이 아닌 존재와 잘못 연관시킬 가능성이 낮음
- 텍스트-이미지 변환에서의 고정관념(성별, 인종, 종교, 직업) 편향 측정에 StereoSet과 비전 언어모델을 위한 VLStereoSet 벤치마크 활용 가능
- 성별에 따라 가장 많은 편향이 발생하였고, 명령 조정 또는 데이터 세트 필터링 등이 없을 때는 큰 모델일수록 성능이 우수하나, 더 편향되어 있음을 발견
- 대표적인 AI 기반 텍스트-이미지 시스템인 Stable Diffusion, DALL-E2, Midjourney에서도 편향 사례가 발견
  - Stable Diffusion과 DALL-E2는 CEO의 이미지를 정장을 입은 남성 또는 정장에 나이가 많고 다소 진지해 보이는 남성으로 묘사
  - Midjourney는 지능적인 사람에 대한 묘사를 안경을 쓴 나이든 백인 남성의 이미지로 묘사

28) Outrider, Could a Chatbot Teach You How to Build a Dirty Bomb?, 2023.1.31.



(a) Stable Diffusion(좌)와 Dall-E2(우)의 CEO에 대한 묘사



(b) Midjourney의 지능적인 사람에 대한 묘사

※자료: Stanford HAI, AI Index Report 2023 (SPRI 재편집)

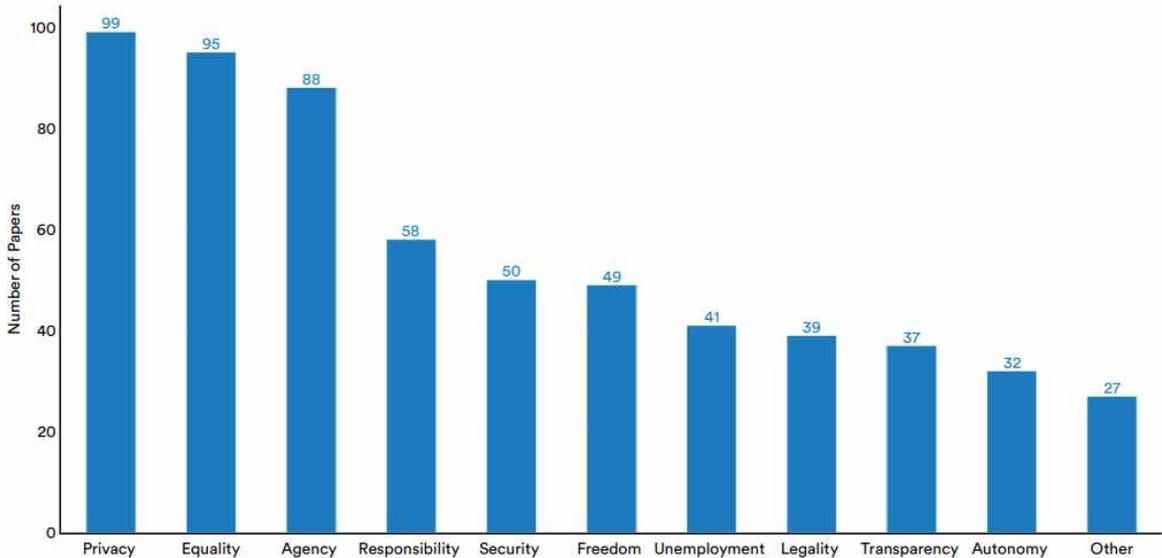
**[그림 1-4-3] 이미지 생성 모델의 편향성**

**6. AI 윤리 연구 동향**

- 핀란드 투르크(Turku) 대학 연구진은 2011년부터 2020년까지 출판된 중국의 AI 윤리 논문 328건을 분석하고 주제와 관심사를 요약,
  - 주제는 개인정보보호 문제를 가장 많이 다루고 있으며, 평등(편향 및 차별),

인간 대체에 대한 위협 등 서부 국가와 유사한 문제 다수 논의

- 이 논문들은 유럽의 일반데이터보호규정(GDPR)과 유럽 위원회의 신뢰할 수 있는 AI를 위한 윤리 지침을 주로 인용하는 등 서부의 AI 원칙에 주목
- 우려사항 및 AI 관련 잠재적 피해 해결을 위해 기술 솔루션 보다 입법 및 구조적 개혁에 중점



※자료: Stanford HAI, AI Index Report 2023

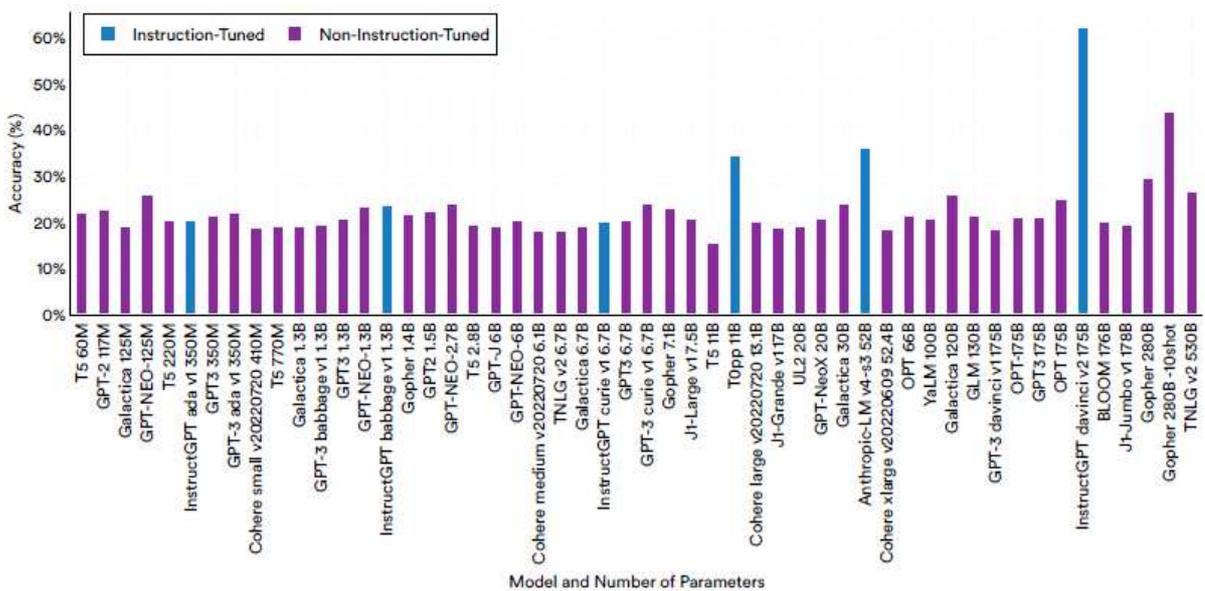
[그림 1-4-4] 중국 AI 윤리 논문에서 제기된 관심 주제 (2018-2022년)

- 알고리즘의 공정성, 책임성 및 투명성에 대한 연구를 다루는 학회인 FAccT에 발표된 논문은 급격히 증가하며 AI 윤리 및 관련 연구에 대한 관심 증가를 시사
  - 2022년 기준 FAccT에 제출된 논문(772편)은 2021년 대비 2배 이상, 2018년 기준으로는 10배 증가했으며, 학계가 약 65%의 비중을 차지
  - 지역별로는 북아메리카(63.24%), 유럽 및 중앙아시아(30.59%), 동아시아 및 태평양(4.25%) 순으로 여전히 북미와 서부 국가의 연구자들이 우세
- 저명 AI 컨퍼런스 중 하나인 NeurIPS에서도 공정성, 책임성 및 투명성에 대해 소규모 워크샵 또는 메인 트랙으로 구성하여 논의
  - AI를 현실 문제(기후, 개발, 금융, 헬스케어, 과학 등)에 적용하기 위한 연구들을 주제로 다수의 워크샵 개최

- 해석 가능성 및 설명가능성, 인과효과 및 반사실적 추론, 프라이버시, 공정성 및 편향 관점주제들이 기존 워크샵에서 메인 트랙으로 옮겨져 논의되는 추세

## 7. 사실성 및 진실성

- 대부분의 언어모델이 특정 기간까지의 데이터로 학습을 하고 이후 지속적으로 **갓은 업데이트를 하지 않기 때문에** 답변에 대한 사실을 확인하기 어려운 문제 존재
  - 독일 다름슈타트(Darmstadt) 공대 및 IBM 연구팀은 기존 팩트체킹 데이터셋을 분석하고 이를 기반으로 하는 시스템의 단점을 확인
    - 여러 제안된 팩트체킹 데이터셋에 신뢰할 수 있는 지식 기반에서 발견된 충분한 증거 또는 반증의 기준을 충족하지 않는 데이터들이 포함
  - 스탠포드 대학 연구원들은 다양한 언어모델의 문답 상황을 분석하여, 모델 크기에 따라 정확도가 향상되지만, 더 작은 모델도 명령 조정에 따라 좋은 성능을 낼 수 있음을 발견
    - 건강, 법률, 금융, 정치 분야의 질의응답에 대한 진실성을 평가하는 벤치마크인 TruthfulQA를 기반으로 한 실험을 통해 튜닝의 효과를 입증



※자료: Standford HAI, AI Index Report 2023

[그림 1-4-5] 명령조정 및 미조정 모델 간 성능 차이

## 1.5. 경제

- 경제 부문은 AI 관련 일자리, 투자\* 등과 기업의 AI기술 활용 및 산업용 로봇 활용 현황 등에 대한 분석을 제공

\* 투자 부문은 스탠포드 HAI와 시장 분석 기관인 Netbase Quid의 연구진이 공동 분석

### 1. 일자리 측면

- 모든 국가에서 AI 기술 관련 노동 수요가 증가하였으며\*, 특히 미국은 전 세계에서 AI 채용 공고 비율이 가장 많이 증가

\* 2010년 이후 전 세계 5만여 개 이상의 구인 웹사이트에서 수집한 채용 공고와 AI 기술을 요구한 목록을 분석하여, AI 분야 여부를 판단하고 전체 채용에서 AI 기술을 요구한 공고 비율을 계산

- 2014년 대비 2022년의 AI 분야 채용 공고 비율이 가장 높은 국가는 미국(2.05%), 캐나다(1.45%), 스페인(1.33%), 호주(1.20%) 순
  - 2022년 기준, 미국 내 기술별 채용 공고 비율은 기계학습(1.03%), 인공지능(0.61%), 자연어처리(0.20%), 신경망(0.16%), 자율주행(0.15%), 시각 이미지 인식(0.13%), 로봇틱스(0.06%) 순
  - 2022년 기준, 미국 내 산업별 채용 공고 비율은 정보(5.3%), 전문, 과학 및 기술 서비스(4.1%), 금융 및 보험(3.3%)으로 나타났으며, 모든 산업에서 2021년 대비 공고 수요가 증가

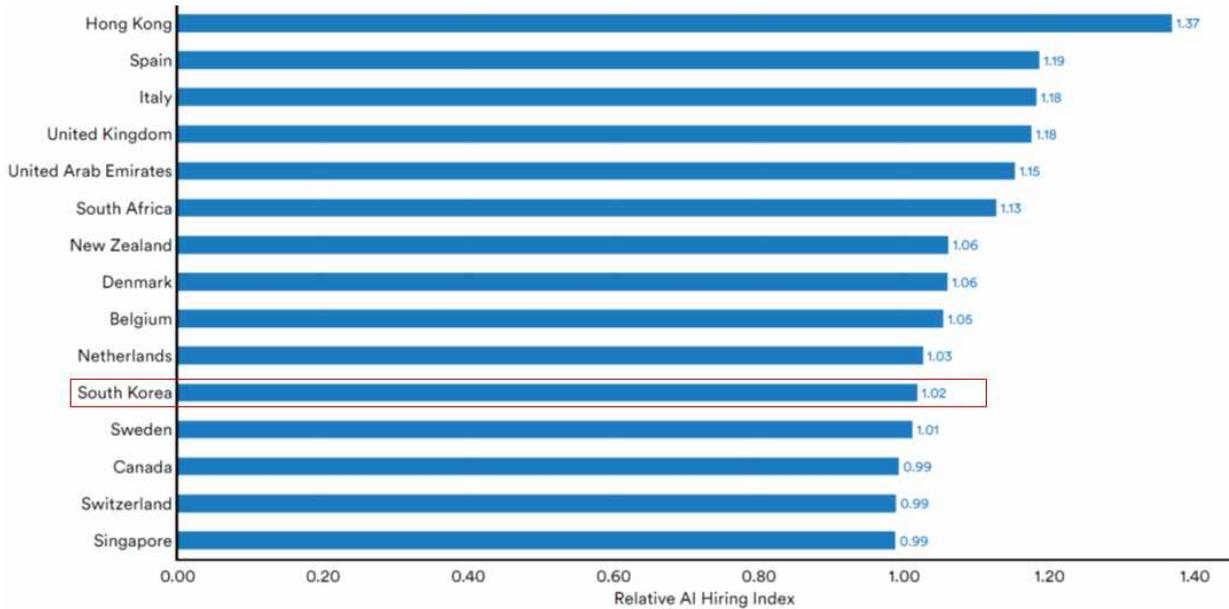
- AI 분야 고용이 상대적으로 빠른 속도로 성장하는 국가는 홍콩, 스페인, 이탈리아, 영국, 아랍 에미리트 등으로, 한국은 11번째\*를 기록

\* 국가별 링크드인(LinkedIn) 이용자 중 AI 기술 보유 및 AI 분야 취업 비율이 2016년 대비 2022년에 얼마나 증가했는지로 AI 고용지수(AI Hiring Index)<sup>29)</sup> 산출

- 우리나라(1.02)는 스웨덴(1.01), 캐나다(0.99), 싱가포르(0.99)보다 높은 수준
  - 우리나라를 포함하여 대부분 국가에서 AI 고용지수는 2020년 정점에 이른 이후 감소하며 최근에는 유지되는 추세

29) 가령 AI 고용지수가 2라는 것은 AI 기술을 보유했거나 AI 분야에 취업한 링크드인 이용자가 지난 5년(2016~2021)간 2배 증가했음을 의미

- 국가별 AI 고용지수는 홍콩(1.37), 스페인(1.19), 이탈리아(1.18), 영국(1.18), 아랍 에미리트(1.15), 남아프리카(1.13), 뉴질랜드(1.06), 덴마크(1.06), 네덜란드(1.03) 순



※자료: Stanford HAI, AI Index Report 2023 (SPRI 재편집)

[그림 1-5-1] 2022년 AI 분야 고용 지수

- AI 기술 침투율(근로자 AI 역량 보유율) 측면에서 한국은 세계 7위이며, 세계적으로 인도, 미국, 독일 등의 근로자가 상대적으로 우수한 AI 역량을 보유\*

\* 링크드인 이용자가 직접 추가한 기술을 기준으로, 전 세계 근로자 평균 AI 역량 수준 대비 해당 국가의 AI 기술 침투율(AI Skill Penetration Rate)을 계산<sup>30)</sup>

- 국가별 AI 역량 보유율은 인도(3.23), 미국(2.23), 독일(1.72), 이스라엘(1.65), 캐나다(1.54), 영국(1.54), 한국(1.44), 싱가포르(1.37) 등 순

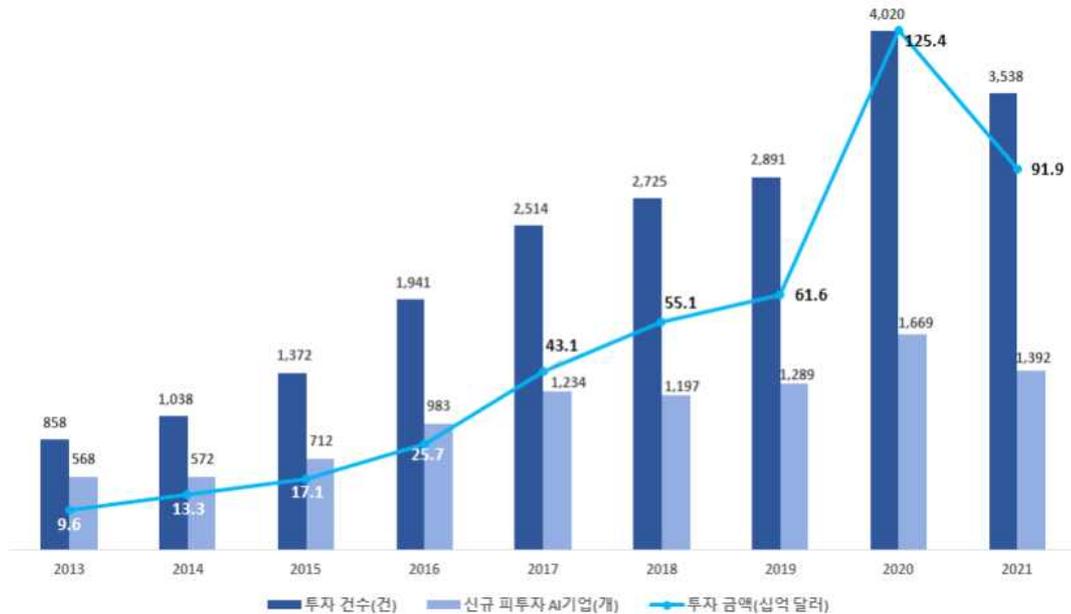
## 2. 투자 측면

- 2022년 전 세계 기업의 총 AI 투자\*는 전년 2,761억 달러 대비 약 30% 감소한 1,896억 달러로, 2013년 이후 처음으로 AI에 대한 기업 투자가 감소

\* AI 투자에는 인수합병, 소수 지분, 민간 투자 및 공모 등을 포함

30) 예를 들어, AI 기술 침투율이 1.5라는 것은 해당 국가 또는 지역이 평균 AI 기술 침투율이 동일한 직업군 전체 평균의 1.5배라는 것을 의미

- 2013년 이후, 지난 10년간 AI 관련 기업 투자는 13배\* 증가
    - \* 146억 달러('13) → 254억('15) → 537억('17) → 956억('19) → 2,761억('21) → 1,896억('22)
  - 지난해 가장 큰 단일 AI 투자 건은 Microsoft가 진행한 음성인식 엔진개발 회사인 Nuance Communications 인수합병으로, 투자 규모는 198억 달러
- 2022년 글로벌 AI 스타트업에 대한 사모 투자(Private Investment)액과 투자 건수, 신규 피투자 기업 수 모두 전년 대비 감소
- 투자액은 2022년 918억 달러로 전년 1,254억 달러 대비 감소하였으며, 이는 2013년 이후 10년 만에 처음 감소 추세로 전환
  - 투자 건수 역시 2022년 3,538건으로 전년 4,020건 대비 감소하였으며, 피투자기업 수 역시 2022년 1,392개로 전년 1,669개 대비 감소
  - 규모별 스타트업 투자 건수에서도 10억 달러 이상의 투자를 제외하고 모든 규모에서 AI 투자 건이 2021년 대비 감소
    - 10억 달러 이상 투자 건은 '21년 4건에서 '22년 6건으로 늘어났으나, 10억 달러 미만의 투자는 '21년 3,418건에서 '22년 2,992건으로 감소



※자료: Stanford HAI, AI Index Report 2023 (SPRI 재편집)

[그림 1-5-2] 2022년 AI 스타트업 투자현황

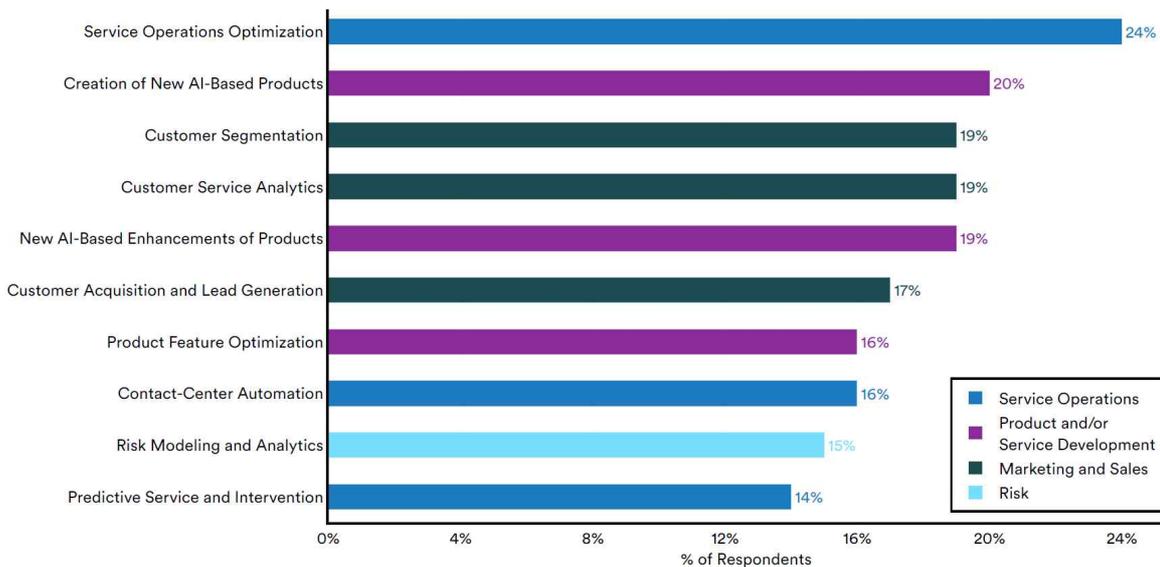
- **미국이 지속적으로 AI 스타트업에 가장 많은 투자를 하고 있으며 중국, 영국, 이스라엘, 인도에 이어 한국이 투자액 기준 세계 6번째**
  - 2021년 미국은 AI 스타트업에 474억 달러를 투자하며, 이는 중국(134억 달러), 영국(44억 달러), 이스라엘(32억 달러), 인도(32억 달러) 등과 비교해 규모가 큼
    - 한국의 AI 스타트업 투자 규모는 이스라엘, 인도와 비슷한 수준인 31억 달러이며, 전년(11억 달러) 대비 투자액이 증가하여 순위는 기존 10위에서 4계단 상승
  - 2013년 이후 AI 스타트업 대상의 누적 투자 금액 역시 미국이 2,489억 달러로 1위를 차지했고, 중국(951억 달러), 영국(182억 달러) 순
    - 한국의 '13~'22년 누적 투자액은 56억 달러(세계 9위)로, 2022년 투자 금액이 누적 투자액의 50% 이상으로 **최근 투자가 대부분을 차지**
  - **신규 투자를 받은 국내 AI 기업의 수는 22개로 세계 13위 수준**
    - 전년(19개)대비 신규 투자 기업 수는 소폭 증가하였으나 전반적으로 다른 나라의 신규 피투자 기업 수가 증가하며 순위도 하락(9위→13위\*)
      - \* 신규 피투자 기업 수는 미국(542개), 중국(160개), 영국(99개), 이스라엘(73개), 인도(57개), 캐나다(47개), 프랑스(44개), 독일(41개), 싱가포르(36), 일본(32개) 등
    - 한국의 '13~'22년 누적 신규 피투자 기업 수는 145개로 세계 11위\*
      - \* 미국(4,643개), 중국(1,337개), 영국(630개), 이스라엘(402개), 캐나다(341개), 프랑스(338개), 인도(296개), 일본(294개), 독일(245개), 싱가포르(165개) 등
- **투자 분야별로는 의료 및 건강관리 분야가 가장 많은 투자를 유치했으며, 데이터 관리, 처리 및 클라우드 분야 및 핀테크, 사이버 보안 분야도 선전**
  - 의료 및 건강 관리(61억 달러), 데이터 관리, 처리 및 클라우드(59억 달러), 핀테크(55억 달러), 사이버 보안(54억 달러) 순으로 집중 투자
    - 다만, 2022년에 대부분의 영역에서 투자액이 전년 대비 감소

### 3. 기업의 AI 활용성

□ 2022년 전 세계 기업의 AI 활용률은 50%로, 지난 5년 동안 급속도로 증가했지만 최근 평준화되는 추세로, 2021년 56% 대비 소폭 감소\*

\* 글로벌 경영컨설팅 업체 맥킨지(McKinsey)가 전 세계 1,492명의 경영자를 대상으로 기업 내 AI 채택 실태를 조사

- 지역별 AI 활용 기업 비중은 북미(59%), 아시아·태평양(55%), 유럽(48%) 순으로 높게 나타남
- 산업·기능별로는 첨단기술/통신(38%), 소비재/소매(31%), 금융 서비스 개발(31%) 등에서 AI 활용률이 높음
- 조직이 도입한 평균 AI 기능의 수는 2018년 1.9개에서 2022년 3.8개가 되었으며, 여기에는 추천 시스템, 자연어처리, 안면 인식 등이 포함
- 2022년에 가장 일반적으로 채택된 AI 도입 사례는 서비스 운영 최적화(24%), AI 기반 새로운 제품 생성(20%), 고객 세분화(19%) 및 고객 서비스 분석(19%)에 활용



※자료: Stanford HAI, AI Index Report 2023 (SPRI 재편집)

[그림 1-5-3] AI 도입·활용 분야

- 경영자들은 위험에 대한 고려사항으로 사이버 보안(59%), 규제 준수(45%), 프라이버시(40%), 설명에 대한 의무(41%)를 가장 높은 위험으로 선정

- 딜로이트(Deloitte)의 AI 사용에 대한 기업 경영자 2,620명에 대한 설문에서는 다수가 비즈니스에서 AI의 중요성을 인식하며 성과에 만족을 시사
  - 조직의 성공에 AI 솔루션이 중요한지에 대해서는 94%가 중요하다고 응답했으며, AI가 성과와 직무를 향상하는지에 대해서는 82%가 동의
  - AI 도입의 주요 성과는 비용 절감(37%), 비즈니스 기능/조직 간 협업 개선(34%), 가치 있는 인사이트 발견(34%) 등이 언급
  - 경영자의 76%가 2023년에 AI 투자를 지난해보다 더 늘릴 것으로 예상
  - AI 주도권 확장의 주요 저해 요인으로는 리스크 관리(50%), 학습 데이터 획득(44%), 기술 구현(42%) 등의 어려움을 언급

#### 4. 산업용 로봇 도입 현황

- AI 기술의 주된 도입 분야는 생산 자동화를 위한 산업용 로봇으로, 세계적으로 2021년에 총 51만7천 대가 보급되었으며 이는 2020년 대비 31.3% 증가
- 국제로봇연맹(IFR) 조사에 따르면, 한국은 2021년 총 3만1천여 대가 보급되어 중국(약 26만 대), 일본(약 4만7천 대), 미국(3만5천 대)에 이어 세계 4위 규모
  - 중국은 2021년 세계 산업용 로봇의 51.8%를 차지하며, 나머지 국가를 합친 것보다 더 많은 산업용 로봇을 설치하며 시장 지배력을 강화
  - 산업별로는 전기/전자(13만7천대), 자동차(11만9천대) 분야에서 가장 많이 도입

## 1.6. 교육

- 교육 부문에서는 기존 대학 이상 고등교육 위주의 AI 교육이 점차 보편화됨에 따라 초중등 교육에도 활용되고 있는 북미의 컴퓨터과학 및 K-12 현황을 분석

### 1. AI 인재 양성과 사회 진출

- 북미 지역의 컴퓨터과학(Computer Science, CS) 전공 졸업생 수가 꾸준히 증가하는 한편 졸업자 중 유학생의 비중도 큰 폭으로 상승
  - 학력별로 보면, 2021년 기준 CS학사 취득자는 33,059명으로 2012년 대비 약 4배 이상 증가하였으나, 석사 및 박사 취득자 수는 전년 수준에서 횡보하는 추세
  - 학위 취득별 유학생의 비중은 박사(68.6%), 석사(65.2%), 학사(16.3%) 순으로 나타나 석박사학위의 유학생 비중이 높은 것으로 나타남
- CS 박사학위 취득자 중 AI 전공의 비중은 2021년 기준 19.1%며, 졸업 후 산업계(민간 기업)로 취업하는 점진적으로 높아지는 상황
  - 북미지역의 AI 전공 신규 박사의 분야별 진출 비중은 2021년 기준 산업계(65.44%), 학계(28.19%), 정부(0.67%)순으로 집계

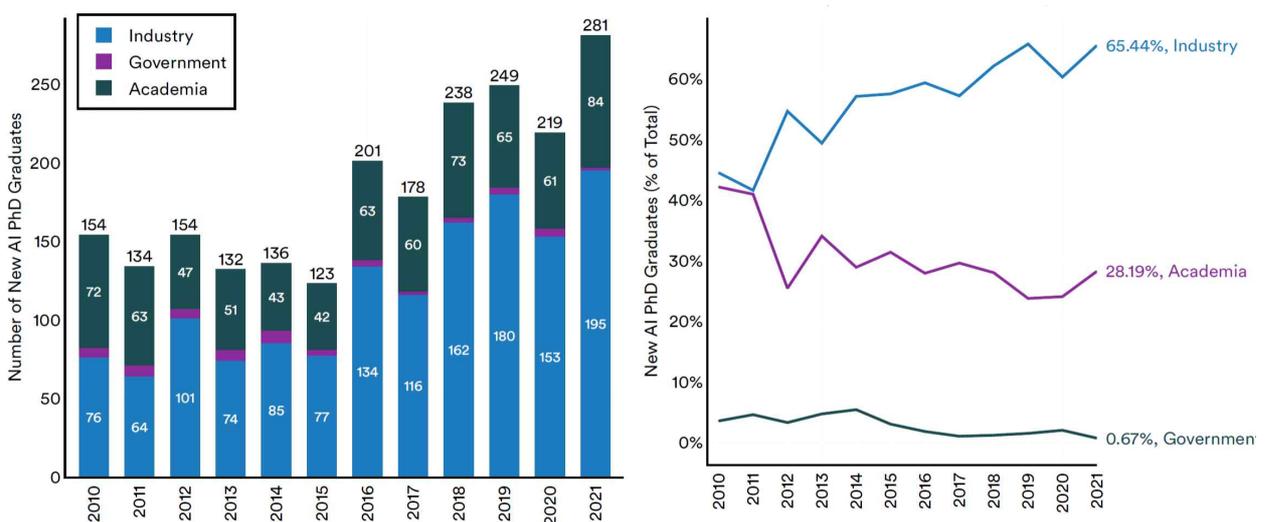


그림 33 AI전공 신규 박사의 북미지역 사회 부문별 고용 지표

※자료: Standford HAI, AI Index Report 2023

### [그림 1-6-1] AI 전공 신규 박사의 북미지역 사회 부문별 고용 지표

- 이는 최근 10년(2011-2021)간 학계의 AI 전공 박사 채용 수가 소폭 상승한 것과 대조적으로 산업계의 채용 수요가 크게 상승함에 따른 현상

## 2. 대학의 컴퓨팅 교육 및 연구 환경

- 북미 소재 대학의 CS관련학과 전체 교수진 수는 지속적인 증가 추세인 반면, 신입 교수 채용은 2018년 이래 지속적으로 감소
  - 2021년 기준 CS관련 교수진은 총 8,149명으로 전년 대비 2.2% 증가하였으며, 교수진 유형별로는 영년직이 약 65.2%, 강의전담교수가 약 14.1%, 기타(강사, 연구중심, 박사후 연구원) 교수진이 약 20.7% 수준으로 집계
  - 전체 교수진 수가 증가하는 것과는 달리, 신입 교수 채용은 2018년 878명에서 2021년 710명으로 약 19.1% 감소
    - 최근 10년간 CS관련 교수진의 충원율이 80% 중후반대에서 안정적으로 유지되고 있음을 고려했을 때, 신입 교수 채용이 감소되는 현상은 대학이 필요로 하는 교수진의 수를 충분히 확보했기 때문으로 해석
  - CS관련학과의 정교수로 외국인을 채용한 비중은 13.2%로, 최근 10년간 가장 많은 외국인을 채용하였던 2013년과 비교했을 시 약 1/2에 불과
- 국방 분야, 산업계의 CS학과에 대한 R&D예산 투자가 증가함에 따라 상대적으로 국립과학재단(NSF)<sup>31)</sup> 기금 의존성이 완화되는 추세
  - NSF가 대학의 CS 학과에 제공하는 예산 비중은 2003년 기준 약 40%에서 지속적으로 감소해 2021년에는 34.9%까지 하락
  - 동일 기간 상승한 외부 자금 출처로는 국방관련 기타(20.3%), 산업자원(12.1%), DARPA(8.8%), 국립보건원(6.8%) 등이 존재
- CS학과의 외부 R&D예산 지출 규모는 사립대학이 국립대학보다 크며, 격차가 심화되는 양상

31) 국가단위 학술, 연구개발 관련 예산의 관리를 수행하는 공공재단으로서, 국내차원에서는 한국연구재단이 유사 사례

- 지출 규모 측면에서 2021년 사립대학의 총지출 중앙값(Median)은 970만 달러 수준으로, 국립대학의 570만 달러 대비 약 1.7배 수준
  - 사립대학의 외부 학술·R&D사업 수주 실적이 국립대학 대비 우수함을 나타내고 있어, 국립대학의 경쟁력 제고 노력이 필요할 것으로 분석

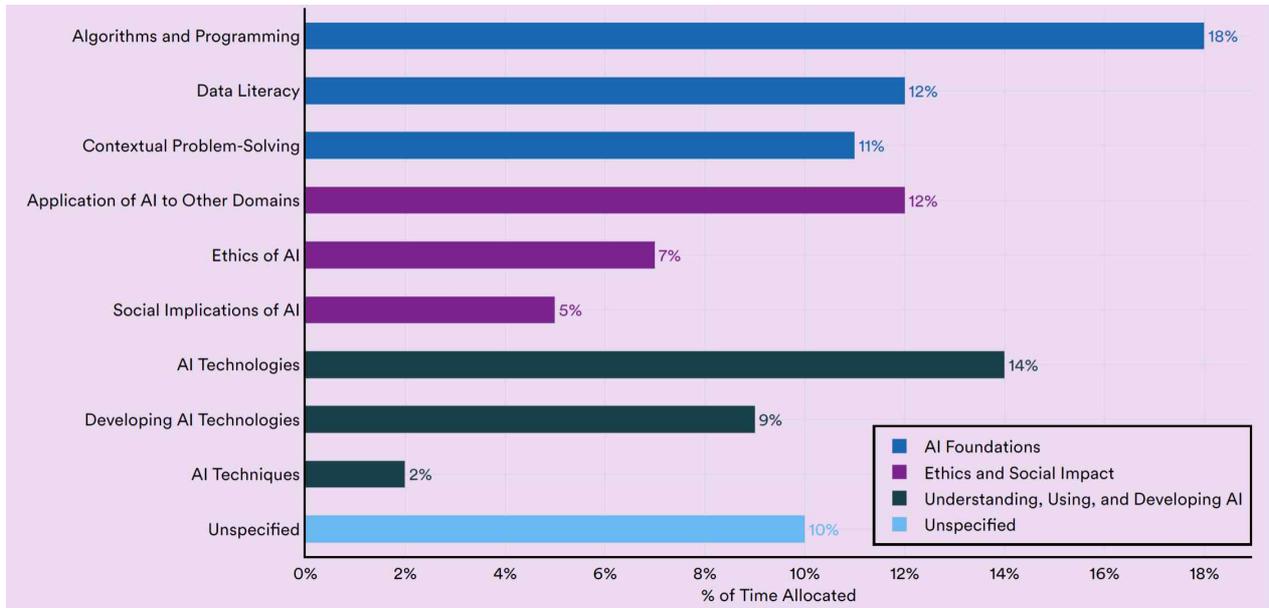
### 3. K-12<sup>32)</sup> 과정의 AI교육

- 193개 유네스코(UNESCO) 회원 국 중 K-12 부문에서 AI교육과정 개발을 승인하거나 추진 중인 국가는 15개국이며, 한국도 여기에 포함
- 각 국가들의 K-12 AI교육과정에서 가장 강조되는 영역은 알고리즘 및 프로그래밍(18%), 시기법(14%), 데이터 리터러시(12%), 다른 영역에 AI적용(12%) 순으로 나타남

**<표 1-6-1> 국가별·추진상태별·교육수준별 정부의 AI교육과정 시행 현황**

국가명	상태	초등학교	중학교	고등학교
아르메니아	승인 및 시행		○	○
오스트리아	승인 및 시행			○
벨기에	승인 및 시행			○
중국	승인 및 시행	○	○	○
인도	승인 및 시행		○	○
쿠웨이트	승인 및 시행	○	○	
포르투갈	승인 및 시행	○	○	○
카타르	승인 및 시행	○	○	○
세르비아	승인 및 시행		○	○
대한민국	승인 및 시행			○
아랍에미레이트	승인 및 시행	○	○	○
불가리아	추진 중	○	○	○
독일	추진 중	○	○	○
요르단	추진 중		○	○
사우디아라비아	추진 중	○	○	○

32) K-12: 미국, 캐나다 영어권에서 주로 활용하는 용어로서, 유치원(Kindergarten)의 K부터 12학년(고등학교 3학년)까지의 무상으로 교육을 받을 수 있는 기간을 의미



※자료: Stanford HAI, AI Index Report 2023

[그림 1-6-2] K-12 AI 교육과정의 비중

## 1.7. 정책 및 거버넌스

□ 정책 및 거버넌스 측면에서는 다양한 국가, 지역 및 미국의 주 정부가 AI 기술을 관리하기 위해 어떻게 노력하고 있는지 AI 정책 및 거버넌스에 대해 분석\*

\* 여러 국가 및 지역에서 제안되거나 통과된 AI 관련 법률, 미국의 주 차원 법률, 전 세계 AI에 대한 의회 및 의회 기록 데이터를 수집해 분석

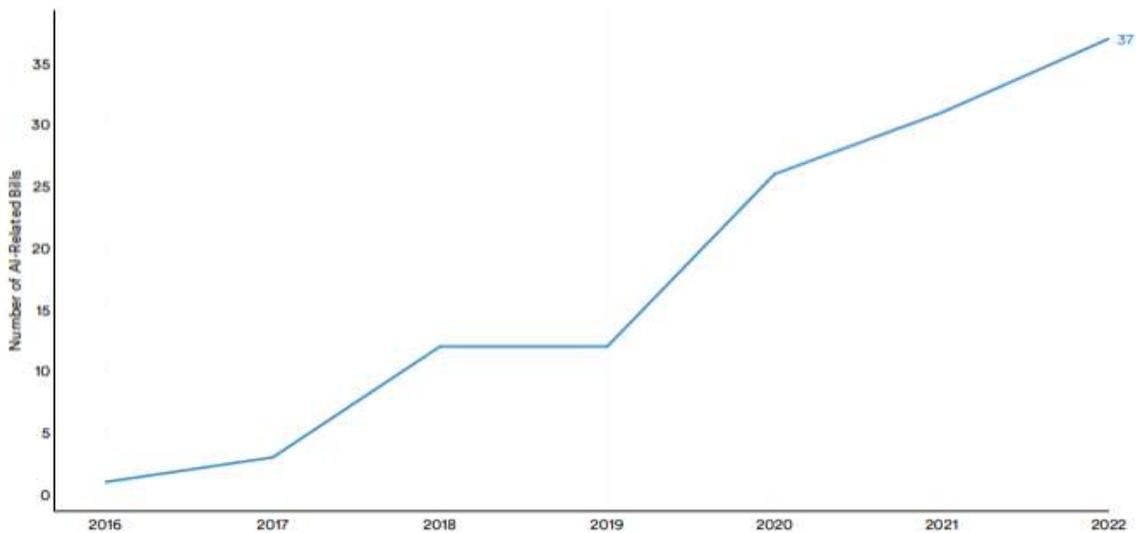
- 특히, 올해 보고서에서는 미국의 공공 분야 AI 관련 계약 추세를 기계 학습 방법을 적용해 분석했고, 미국 내 55개 연구 기관들의 AI 정책 문서 분석 및 미국 내 증가하고 있는 AI 관련 소송에 대한 분석을 추가

### 1. AI관련 글로벌 입법 동향

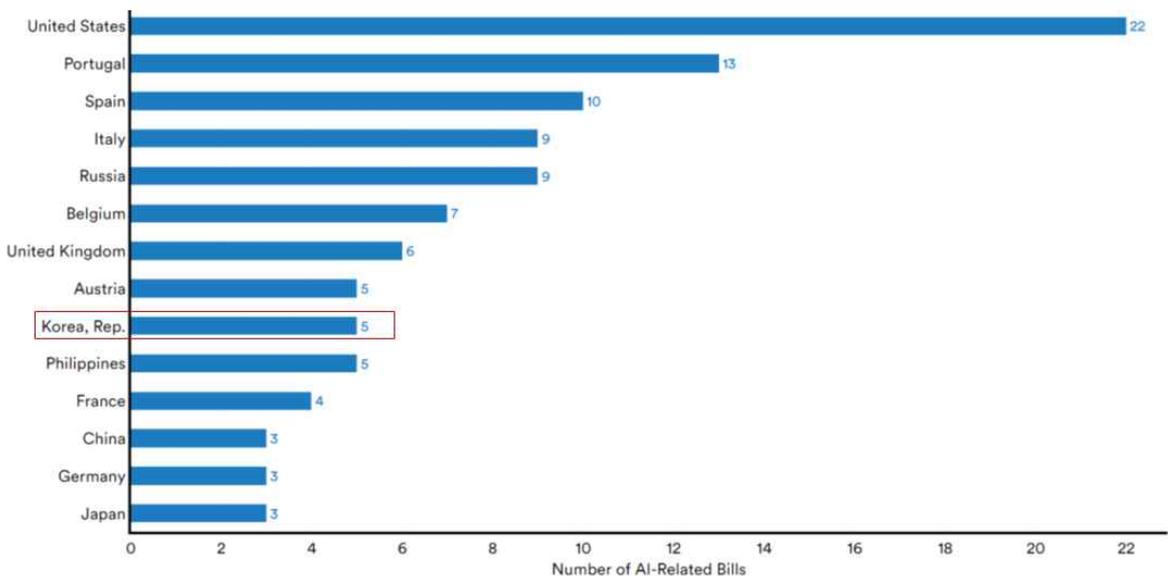
□ 2016년부터 2022년까지 '인공지능(artificial intelligence)'이라는 단어가 포함된 127개국 입법부 법안을 분석한 결과 '22년에는 37건으로 대폭 증가\*

\* 2016년 이후 통과된 AI 관련 법안 수(총 123건): ('16년(1건) → '22년(37건))

- 2016년부터 2022년까지 AI 관련법을 가장 많이 입안한 국가는 미국(22건), 포르투갈(13건), 스페인(10건), 이탈리아(9건), 러시아(9건), 벨기에(7건), 영국(6건), 한국(5건) 순
- 2022년 한 해 동안 AI관련 법 입안 건수는 미국 9건, 스페인 5건, 필리핀 4건, 벨기에, 이탈리아, 포르투갈, 러시아, 영국 등은 각 2건 순
- 미국의 경우 2022년도 한해 총 88건의 관련법이 제안되었으며 이 중 9건이 통과



(a) 글로벌 AI 관련 법안 통과 추이(2016~2022)



(b) 국가별 AI 관련 법안 현황(2016~2022)

※자료: Stanford HAI, AI Index Report 2023 (SPRI 재편집)

**[그림 1-7-1] AI 관련 법안 통과 동향**

- 입법 관련 시에 대한 언급 분석에 따르면, **세계적으로 입안 과정에서 AI 언급은 증가되고 있으나 2022년에는 전년도 대비 감소**(1,547건 →1,340건)
- 분석 대상 81개국 중 스페인(273회), 캐나다(211회), 영국(146회), 미국(138회)순으로 많았으며 46개국이 최소 1회 이상 언급\*
- \* 최근 7년간의 언급 수로 보면 81개국 중 62개국이 최소 1회 이상 언급하였으며 최다 언급 국가는 영국(1,092), 스페인(832), 미국(626), 홍콩(478) 순

## 2. 국가 차원의 AI 정책

- 미국의 55개 기관\*의 지난 5년간(‘18년~’22년) 발간한 정책 문서를 조사한 결과 전체적으로 2022년 발간 보고서는 284건으로 2018년 대비 3.2배 증가
  - \* 55개 기관: 정책 및 싱크탱크(19개), 대학 및 연구프로그램(14개), 시민사회조직(9개), 산업 및 컨설팅 기관(9개), 정부 기관(4개) 등
- 주제별로는 산업 및 규제(107건), 혁신과 기술(90건), 정부와 공공행정(82건), 프라이버시, 안전, 보안 관련(59건), 윤리(39건) 순으로 집계
- 국가 AI 전략은 2017년 캐나다가 공식적으로 보고서로 발표한 이후 2019년을 정점으로 62개국에서 국가 인공지능 전략을 수립하고 발표

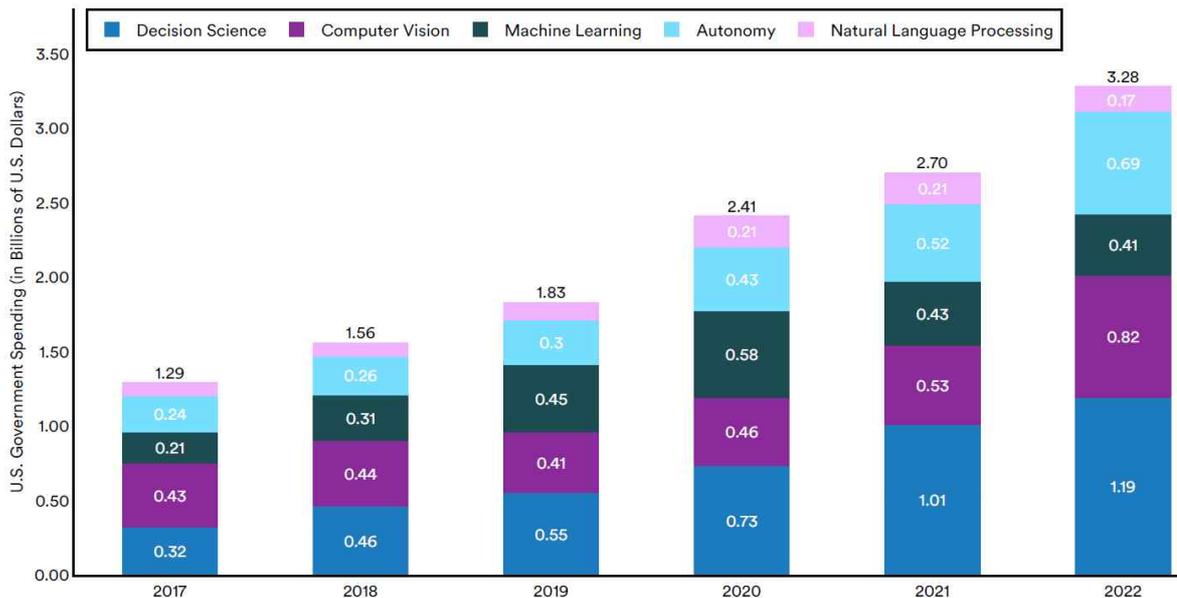
<표 1-7-1> 연도 및 국가별 AI 전략 발표 동향

연도	수	국가(알파벳 순)
2017	3	캐나다, 중국, 핀란드
2018	7	호주, 프랑스, 독일, 인도, 모리셔스, 멕시코, 스웨덴
2019	26	아르헨티나, 오스트리아, 방글라데시, 보츠와나, 칠레, 콜롬비아, 키프로스, 체코, 덴마크, 이집트, 에스토니아, 일본, 케냐, 리투아니아, 룩셈부르크, 몰타, 네덜란드, 포르투갈, 카타르, 루마니아, 러시아, 시에라리온, 싱가포르, UAE, 미국, 우르과이
2020	14	알제리, 불가리아, 크로아티아, 그리스, 헝가리, 인도네시아, 라트비아, 노르웨이, 폴란드, 사우디아라비아, 세르비아, 한국, 스페인, 스위스
2021	10	브라질, 아일랜드, 페루, 필리핀, 슬로베니아, 튀니지, 터키, 우크라이나, 영국, 베트남
2022	2	이탈리아, 태국

※주: 한국의 「AI국가전략」은 2019.12.17.부로 발표되었으나, 연말 발표를 감안하여 조사집계 과정에서 2020년으로 분류된 것으로 추정

### 3. 미국의 AI 공공 투자

- 국가과학기술위원회(NSTC)가 발표한 공공 부문 AI R&D 예산 보고서<sup>33)</sup>를 분석한 결과, 전반적으로 공공 부문 전체에서 AI R&D를 위한 투자 확대 중
  - 2022년 비 국방 AI 예산은 총 17.3억 달러로 2018년 5.6억 달러 대비 약 3.1배 증액\*
    - \* 2023년 미 정부기관(비국방 분야)들은 AI 부문 예산으로 2022년 대비 6.4%p가 증가한 18.4억 달러(한화 2.39조 원) 요청
  - 2022년 국방부의 AI 관련 연구개발·시험·평가(RDT&E) 예산은 8.7억 달러\*
    - \* 2023년 미 국방부(DoD)는 AI 관련 2022년 대비 26.4% 증가한 11억 달러 예산 요청
  - 미국 정부 계약 중 AI와 관련된 계약 비중은 꾸준히 증가세
    - 2022년에 연방 부처와 공공기관은 AI 관련 계약에 총 32.8억 달러를 지출
      - \* 총 계약 금액은 2017년 12.9억 달러에서 2022년 32.8억 달러로 약 2.5배 증가



※자료: Stanford HAI, AI Index Report 2023

[그림 1-7-2] 미정부 AI 관련 계약 추이

- 계약 분야는 의사결정, 컴퓨터비전, 자율화, 기계학습, 자연어처리 순

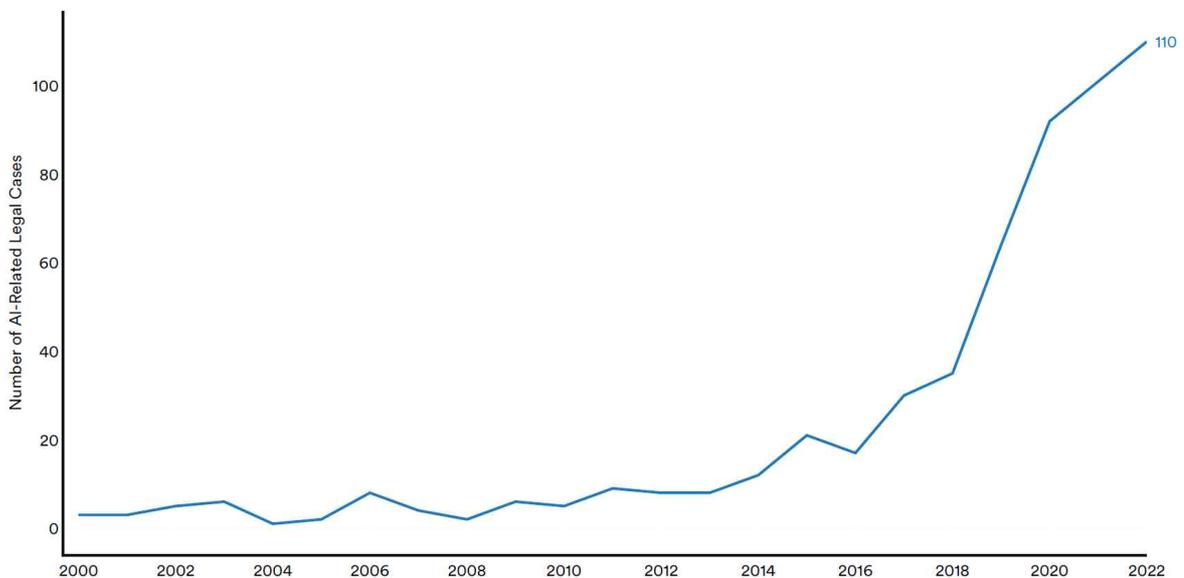
33) Supplement to the President’s FY 2023 Budget, ‘The Networking & Information Technology R&D Program and the National Artificial Intelligence Initiative Office’, 2022.11.

- 2022년 기준 계약형태는 프라임 계약(62.5%), 보조금(34.9%), 기타 계약(2.6%) 순

#### 4. 미국 AI 관련 소송(legal cases)

□ 최근 몇 년간 미국 내 AI 관련 소송 건수가 크게 증가하는 추세

- 2022년 기준 미국에서 110건의 인공지능 소송이 진행되었으며 이는 2016년 대비 6.5배 증가한 수치



※자료: Stanford HAI, AI Index Report 2023

[그림 1-7-3] 미국 내 AI 관련 소송 건수 추이

- 지역별로는 캘리포니아(33건), 일리노이(17건), 뉴욕(11건) 순으로 가장 높은 비중을 차지
- 소송 분야로 보면 금융 및 전문서비스(48건), 미디어 문화 그래픽(18건), 공공서비스(14건)순으로 높은 비중을 차지
- 소송 관련법은 민사(32건), 지적재산권(21건), 계약법(15건) 순으로 비중이 높아 대부분 지적재산이나 계약 관련 민사 문제

## 《 AI 관련 소송 사례 - AI Index 2023 》

### ■ Duerr et al v. Bradley University (2022.3)

- 원고들은 일리노이주 Peoria에 있는 사립대학에서 2020년 가을학기에 온라인 시험을 치루기 위해 Respondus Monitor라는 인공지능 기반의 써드파트 감독 도구를 사용하였음.
- 원고는 피고가 이 과정에서 적절한 생체정보수집법(BIPA)을 위반하였다고 소송
- 하지만 피고는 Gramm Leach-Bliley Act에 따라 사립학교가 관여한 대출 기능에 따라 금융 기관이라고 주장하며 BIPA법이 금융 기관에 적용되지 않는다는 점을 주장
- 법원은 이를 받아들여 원고 소송을 기각함

### ■ Flores vs Stanford (2021.9)

- 가석방이 거부된 가해자 원고가 가석방 위원회가 자신의 사례를 검토하는데 사용한 정보에 대한 접근을 거부한 것에 대해 뉴욕주를 상대로 소송 제기
- Northpointe, Inc는 AI기반 위험 평가 도구인 COMPAS(Correctional Offender Management Profiling for Alternative Sanctions)가 가석방 위원회에 의해 사용되었는데 관련 정보 공개는 영업 비밀이라며 공개 방지 청원
- 법원은 문제의 기밀 자료가 원고가 제기한 사건과 관련이 있고 경쟁적 피해의 위험이 거의 없다고 판단해 공개 명령

### ■ Dyroff vs. Ultimate Software Grp., Inc (2017.11)

- 원고 Kristanalea Dyroff는 그녀의 29세 아들이 Ultimate Software(US)의 소셜네트워크에서 만난 마약상으로부터 펜타닐이 함유된 헤로인 과다복용으로 사망한 후 US를 고소함
- Dyroff의 주장 중 하나는 US가 사용자 데이터를 마이닝하고 알고리즘을 이용해 약물 관련 토론 그룹을 추천했다는 것이 포함
- US는 제3자 콘텐츠에 대한 책임으로부터 웹사이트 운영자를 보호하는 통신품위법 (Communication Decency Act)에 따라 청구를 기각하고 부분 면책을 주장.
- 법원은 US의 해당 웹사이트(Experience Project)가 해를 기칠 위험이 없는 콘텐츠 중립적인 기능에 해당하기 때문에 원고의 아들에 대한 주의 의무가 없다며 소송 기각

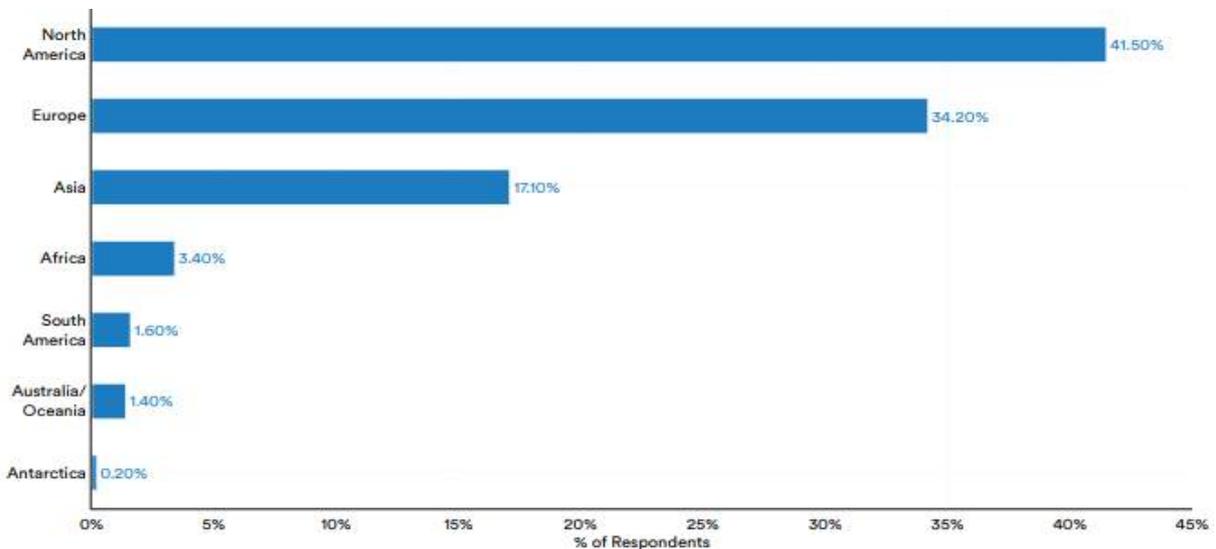
## 1.8. 다양성

- AI 분야의 다양성의 결여는 사회적 불평등과 편견이 강화되어 부작용을 유발, 이에 대응하기 위한 성·인종 다양성에 대해 분석\* 자료를 제공

\* 여성머신러닝워크숍(WiMLW) 및 컴퓨팅연구그룹(CRA) 참가자의 국적 데이터 및 북미지역의 AI 유관학과 졸업생, 교직원 데이터를 활용하여 분석

### 1. 여성머신러닝워크숍(WiMLW) 참가자 다양성 현황

- 전체 참가자 중 북미지역이 41.5%으로 가장 많았으며, 유럽(34.2%), 아시아(17.1%), 아프리카(3.4%) 순으로 집계



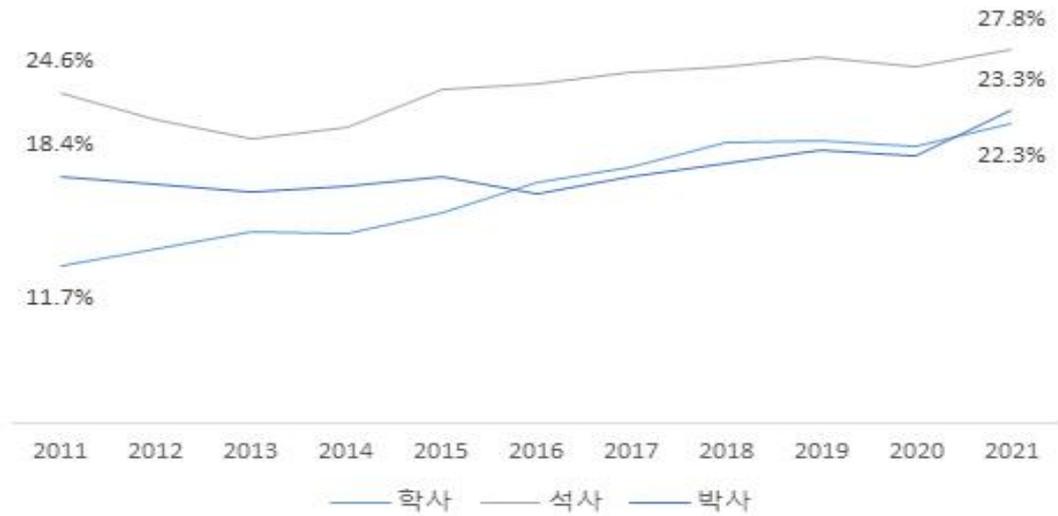
※자료: Stanford HAI, AI Index Report 2023

[그림 1-8-1] 2022년 WiMLW 참석자 국적 현황

### 2. 북미지역 다양성 현황

- 컴퓨터공학(CS), 인공지능(AI) 부문의 여성과 남성의 성 격차는 서서히 줄어들고 있으나, 여전히 남성 비중이 현저히 높아 **성별 불균형이 지속**
  - 2021년 기준, CS 졸업생 중 남자 졸업생 비율은 모든 학위과정에서 70%를 웃도는 반면, 여자 졸업생 비율은 20%대 수준\*

\* △남학생: 학사(77.26%), 석사(71.27%), 박사(76.58%), △여학생: 학사(22.3%), 석사(27.83%), 박사(23.3%)



※자료: Stanford HAI, AI Index Report 2023 (재구성)

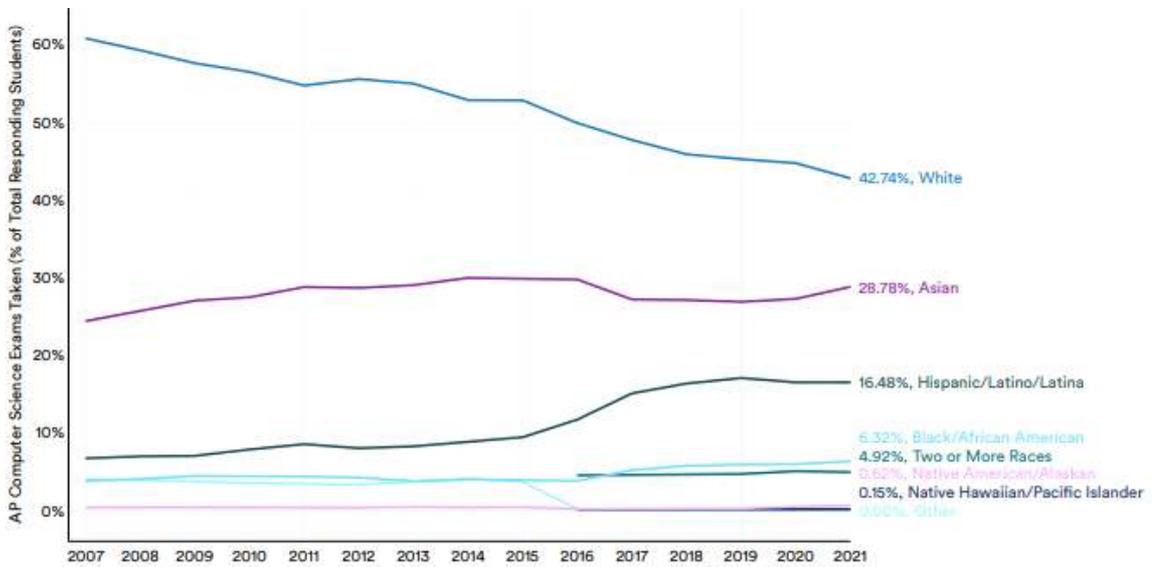
**[그림 1-8-2] 북미지역 컴퓨터공학 학력별 여성 졸업생 비중**

- CS, CE 및 정보학부 교수진의 여성 교수진 비중은 2021년 기준 23.94%이며, 비중은 꾸준히 증가('11년 19% → '21년 23.9%)
  - 특히, 신규 교수채용 중 여성 비중은 2021년 30.2%로 2015년 대비 약 9%p 증가하여 점진적으로 신규 여성 교수 채용 확대 노력
- 신규 AI 박사 졸업생 중 여성의 비중은 2010년 19.9%에서 2021년 21.30%로 증가하여 성별 격차가 줄어드는 양상
- 북미지역 AP<sup>34)</sup> 컴퓨터공학 시험 응시생 중 69.16%, 여성 응시자는 30.58%로 집계되었으며, 여성 응시자 비중은 2007년 대비 82% 증가
- **CS분야 졸업생 및 교수진은 백인 위주로 구성되어 있었으나, 아시아계, 히스패닉 또는 다인종 졸업생 비중이 꾸준히 증가하여 인종 격차 감소추세**
  - 2010년 CS 학사 졸업생 구성비는 백인(71.9%), 아시아(16%), 히스패닉(5.8%)순으로 대다수가 백인으로 구성되었으나,
  - 2021년 기준 백인 46.69%, 아시아계 33.99%로 두 인종 간 격차가 확연히 줄어들어 인종 다양성이 확대\*

34) Advanced Placement(대학과목 선이수제), 미국에서 시작한 선행학습 제도로 고등학생들에게 대학교 1학년 교양과목 수준의 내용을 경험학습할 수 있는 기회를 제공하고, 추후 대학 진학 시 이수학점으로 인정해주는 제도

\* 2021년 기준 CS 졸업생 인종 구성비 △석사 : 백인(50.28%), 아시아계(34.83%), 히스패닉 (7.25%), △박사 : 백인 (58.64%), 아시아계 (29%), 히스패닉 (5.12%)

- 또한, CS, CE 및 정보학과 교수진의 백인과 아시아계의 격차가 서서히 축소 중('11년 46.1% → '21년 28.4%)
- 2021년 AP 컴퓨터 과학시험 응시자 인종 비중은 백인 학생(42.7%)이 가장 많은 시험을 치렀고, 아시아계(28.8%)과 히스패닉/라틴계 학생(16.5%) 순
  - 2007년 대비 아시아계, 히스패닉/라틴계 및 아프리카계 미국인 학생들의 AP 시험 응시율이 높아져 인종적으로 다양해지는 추세



※자료: Stanford HAI, AI Index Report 2023

[그림 1-8-3] 북미지역 AP 시험 응시생 인종별 비중

## 2.8. 여론

- 이 장에서는 AI에 대한 다양한 사람들의 태도와 관점을 분석하여 사회에 변화를 가져올 수 있는 잠재력을 인식하는 여론에 대해 자세히 다룸\*

\* 글로벌 설문 조사(IPSOS, Lloyd's Register Foundation, Gallup, PEW Research 등이 기획·참여)를 포함하여 2022년 AI에 대한 소셜 미디어 데이터를 분석

### 1. AI에 대한 글로벌 대중 인식

- 2021년 말, IPSOS는 28개국\*의 16~74세 성인 19,504명을 인터뷰하여 AI를 활용한 제품과 서비스에 대해 설문조사 실시

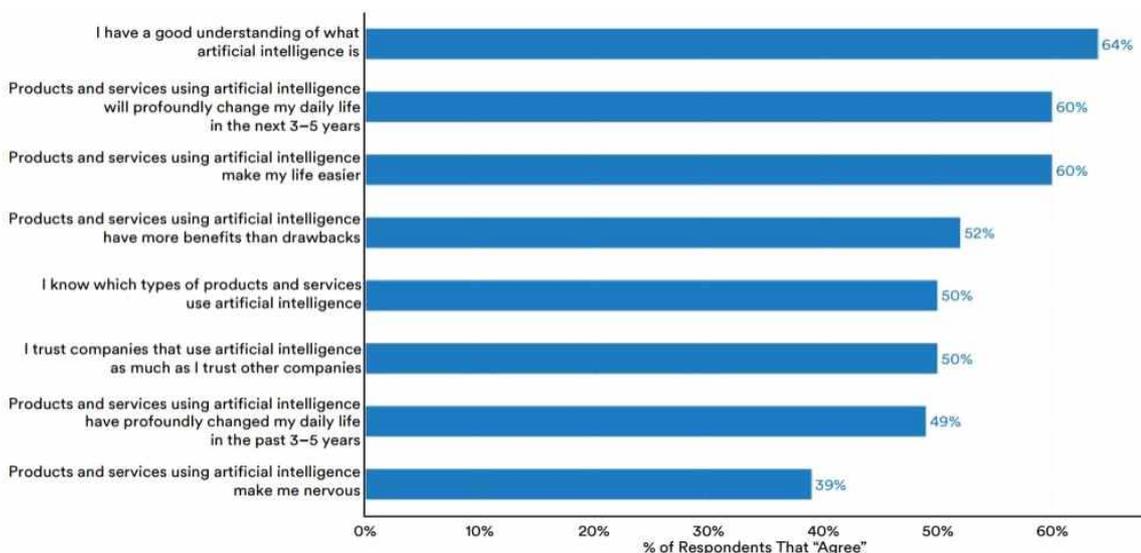
\* 호주, 벨기에, 브라질, 캐나다, 중국, 덴마크, 한국, 스페인, 스웨덴, 스위스, 영국, 미국 등

- AI 제품 및 서비스와 관련된 질문에 대해 각 의견에 동의하는 응답자의 비율 제공

- 응답자의 64%는 인공지능에 대해 이해하고 있다고 답변하였으며, 60%는 AI를 활용한 서비스가 가까운 미래에 삶을 더 편하게 변화시킬 것이라고 응답

\* 3~5년 내 AI 제품이 일상생활을 크게 변화시킬 것이라고 60%가 응답

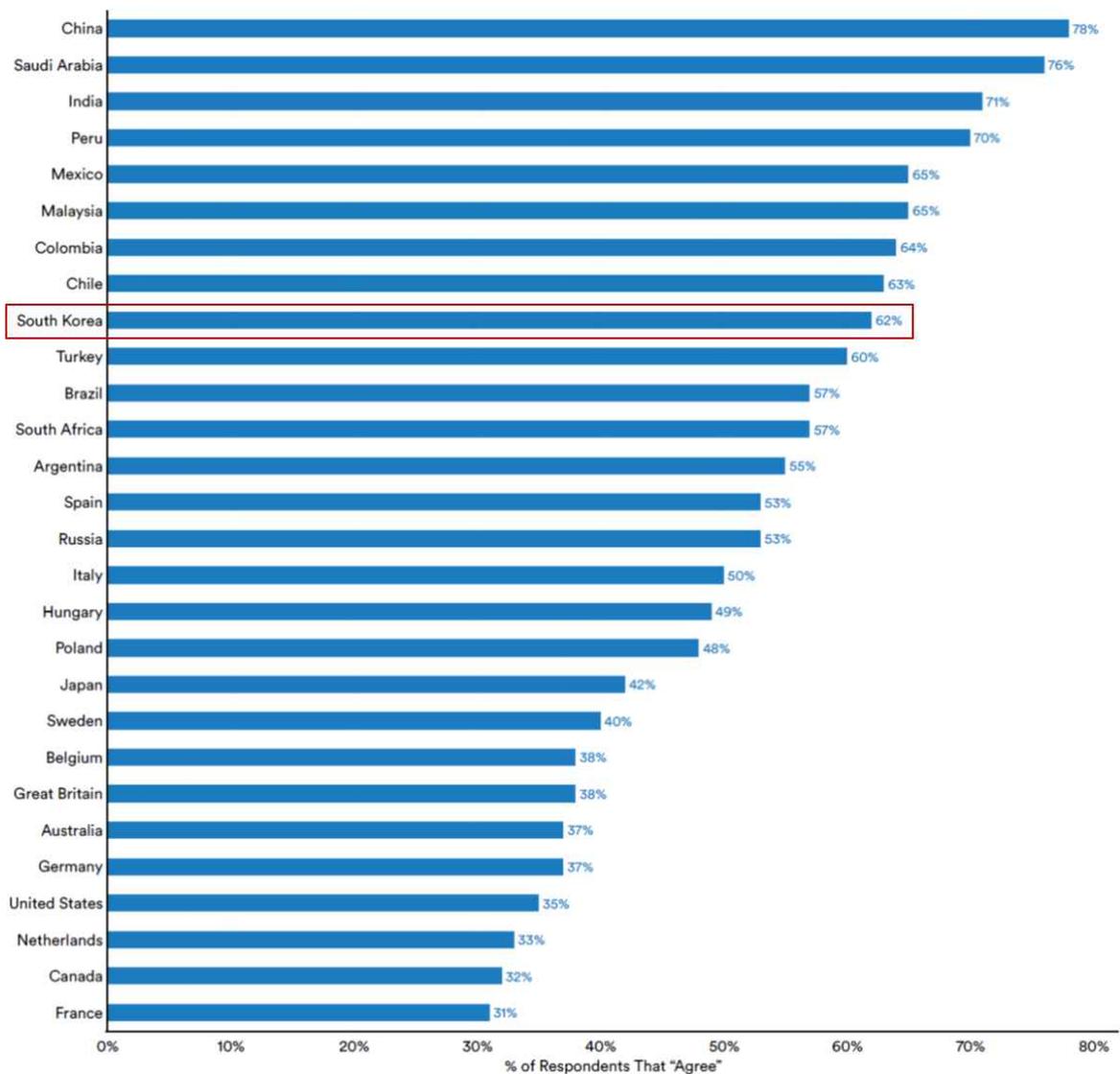
- 응답자 절반 이상인 52%가 AI를 활용한 제품 및 서비스가 장점이 더 많다고 답변한 반면, 응답자의 40%는 AI 활용 제품 및 서비스에 대해 우려



※자료: Stanford HAI, AI Index Report 2023

[그림 1-9-1] AI 활용 제품 및 서비스에 대한 글로벌 여론

- AI 활용제품 및 서비스의 장점과 단점에 대해 국가별 의견 차이가 존재하며, 일부 국가에서는 응답자 대다수가 장점이 더 많다고 응답
  - 중국은 78%가 긍정적으로 응답하여 가장 높은 비율을 보였고, 다음으로 사우디아라비아(76%), 인도(71%) 순으로 집계
  - 한국은 전체국가 중 9순위(62%)로 AI 활용을 긍정적으로 평가
  - 반면, 미국은 응답자의 35%만이 AI 활용제품 및 서비스에 대하여 긍정적이며, 프랑스(31%), 캐나다(32%)순으로 가장 부정적인 견해 보유



※자료: Stanford HAI, AI Index Report 2023

[그림 1-9-2] 국가별 'AI 활용 제품 및 서비스의 장점이 더 많다' 응답비율

## 2. 성별, 연령, 가계 소득, 고용 상태 등에 따른 여론

- 여성보다는 남성이, 소득이 높은 가구일수록 AI 활용 서비스에 긍정적인 응답
  - 성별 기준, ‘AI를 이해하고 있다’는 설문에 남성(69%)이 여성(60%)보다 높은 응답을 보였으며, ‘AI 활용 제품은 장점이 더 많다’는 응답 또한 남성 55%, 여성 49%로 답변
  - 연령별 의견은 다양하며, ‘AI 활용 제품 및 서비스가 삶을 더 편하게 해준다’는 응답에 35세 미만(64%), 35세에서 49세 사이(62%), 50세 이상 74세 이하(54%) 순으로 응답
    - ‘단점보다 장점이 더 많다’는 설문에는 35세 이상 49세 이하가 가장 높은 응답률(53%)을 보임
  - 가계소득 기준, 소득이 높을수록 AI에 대한 이해도가 높으며, 장점이 더 많다는 응답 또한 높은 소득 가구(57%), 중간 소득 가구(51%), 낮은 소득 가구(50%)의 순으로 긍정적인 견해를 가짐

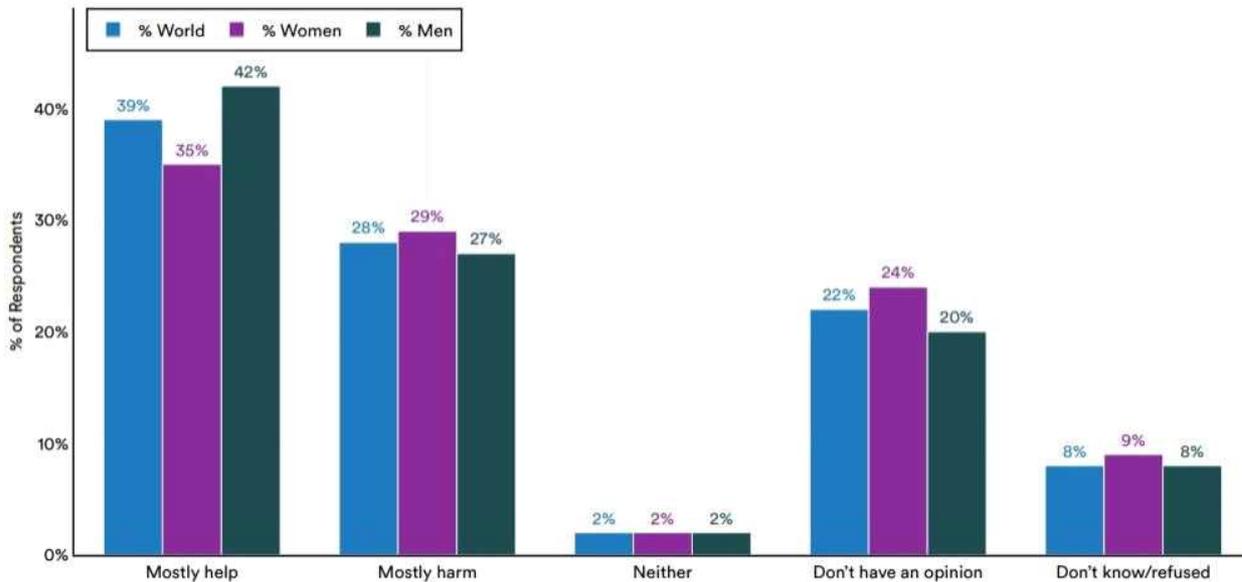
I have a good understanding of what artificial intelligence is	69%	60%	66%	65%	61%	57%	63%	71%	56%	64%	71%	73%	74%	67%	59%
Products and services using artificial intelligence will profoundly change my daily life in the next 3–5 years	63%	57%	63%	61%	55%	56%	58%	67%	53%	58%	68%	70%	72%	64%	54%
Products and services using artificial intelligence make my life easier	62%	58%	64%	62%	54%	56%	58%	66%	53%	58%	67%	67%	70%	63%	55%
Products and services using artificial intelligence have more benefits than drawbacks	55%	49%	47%	53%	46%	50%	51%	57%	45%	50%	59%	63%	64%	55%	47%
I know which types of products and services use artificial intelligence	55%	46%	54%	51%	45%	46%	50%	57%	44%	48%	58%	63%	65%	54%	44%
I trust companies that use artificial intelligence as much as I trust other companies	53%	47%	54%	51%	44%	47%	48%	57%	45%	48%	56%	61%	62%	53%	45%
Products and services using artificial intelligence have profoundly changed my daily life in the past 3–5 years	51%	46%	54%	50%	41%	46%	47%	54%	43%	46%	55%	61%	62%	52%	43%
Products and services using artificial intelligence make me nervous	38%	41%	40%	40%	38%	41%	41%	38%	41%	37%	40%	48%	46%	40%	38%
	Male	Female	Under 35	35 to 49	50 to 74	Low	Medium	High	Low	Medium	High	Business Owner	Sr. Exec./ Decision Maker	Employed	Non-Employed
	Gender		Age			Household Income			Education			Employment Status			

※자료: Stanford HAI, AI Index Report 2023

[그림 1-9-3] 인구통계학적 그룹별 AI에 대한 여론(2022)

□ AI와 다른 디지털 트렌드에 대한 인식조사결과도 긍부정 평가 비율에 차이

- 국가별로 121개국의 125,911명 대상 향후 20년 동안 AI가 인간에게 도움이 되는지에 대한 설문조사 결과, ‘대부분 도움이 될 것(39%)’이 가장 높은 비율을 차지하였으며, 28%가 ‘대부분 해를 끼칠 것’이라고 응답
- 성별 기준, IPSOS 조사 결과와 유사하게 남성의 42%는 ‘대부분 도움이 될 것’이라고 예상하나, 여성은 동일 질문에 35%만 긍정적으로 답변



※자료: Stanford HAI, AI Index Report 2023

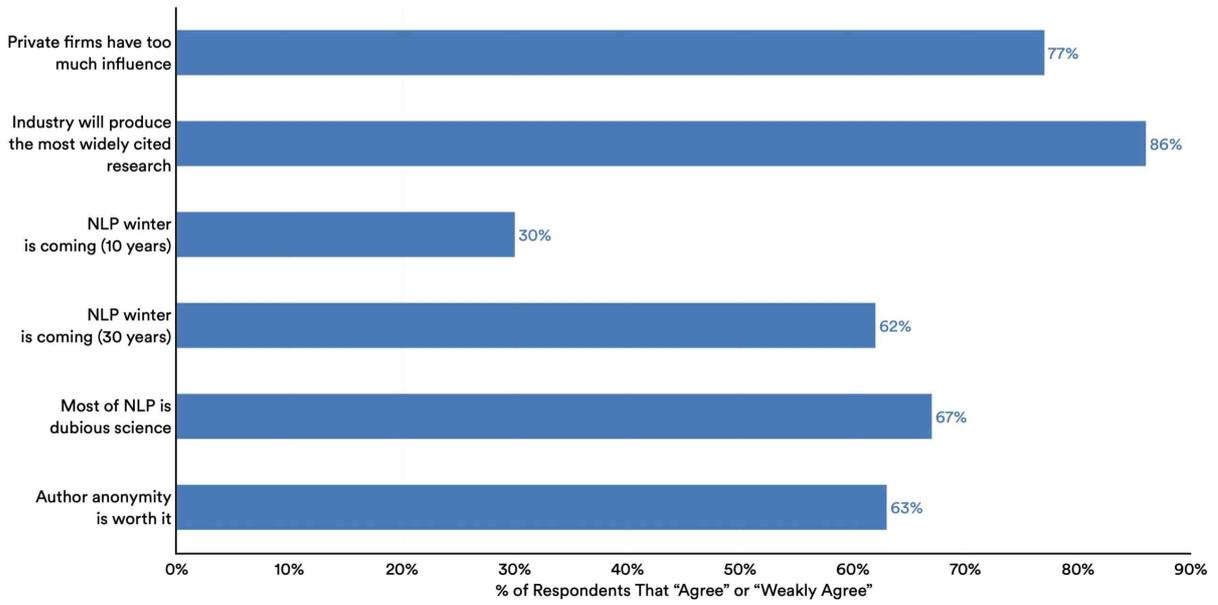
[그림 1-9-4] AI가 향후 20년 동안 도움이 될지 국가별, 성별 간 응답 (2021)

### 3. 전문가의 AI에 대한 견해

□ 2022년 자연어처리(NLP) 연구 커뮤니티 대상 조사\* 결과, 응답자의 86%는 AI 기술이 가장 널리 인용되는 연구주제가 될 것으로 예상

\* 총 480명을 설문조사 하여 AI 연구자들의 AI 기술에 대한 포괄적인 인식 및 관점 파악

- 동시에 응답자의 67%는 의심스러운 과학이라는 의견도 함께 보임



※자료: Stanford HAI, AI Index Report 2023

### [그림 1-9-5] 자연어처리 전문가의 여론 (2022)

## II. 시사점

- (연구개발) AI R&D의 급격한 성장세가 지속되며, 중국이 미국을 뛰어넘는 양상이 전개되는 가운데 양국 간 연구 협력도 지속 확대, 우리도 협력체계 대응 필요
  - 전 세계 AI 분야 출판물 규모는 2010년 이후 두 배 이상 증가하였으며, 특히 2017년부터 최근까지 가파른 양적 성장을 유지
    - 2021년 중국은 여러 측면에서 미국과 대등한 수준이거나 몇몇은 앞서고 있는 것으로 분석\*
      - \* △2021년 기준으로 AI 분야 출판물 중 60% 비중을 차지하는 저널 논문 성과에 있어서 양(건수) 및 인용수로 중국이 미국을 추월, △AI 컨퍼런스 논문의 경우에도 중국은 양적으로 미국을 월등히 앞서고 있으며, 인용 수 측면으로는 미국이 중국을 약간 앞서고 있으나 그 차이가 미미한 수준
  - 미국-중국 간 AI 분야 협력 연구도 지속 확대되는 가운데, 이들의 2강 체제가 고착화되지 않고, 우리나라가 AI 분야 강국에 합류할 전략이 요구
    - AI R&D 투자 확대와 더불어 미국과 중국을 포함한 주요 선진국과의 국제 협력연구 활성화를 위한 정책방안 마련이 필요
- (기술 성능) AI는 지난 10년간 기술적 진보를 달성\*했으며 추가 학습 데이터의 정확도와 양이 성능 향상에 핵심적 역할 수행, 이에 입각한 원천기술 확보 전략 적극 추진 필요
  - \* △(컴퓨터 비전) 이미지, 자세, 활동 등 단순 인식 기술은 90% 이상의 정확도를 보이고 있으며, 최근 생성, 감지, 추론, 분류 등 복잡한 작업을 위한 AI 모델이 증가하고 성능도 빠르게 향상 중
    - △(음성 인식) 일부 기술에서 인간 수준 또는 그 이상에 도달했으나, 어려운 추론과 감성 인식 분야는 발전이 필요
    - △(하드웨어) 최근 GPU와 AI반도체 기술의 비약적으로 향상되면서 AI 모델 개발을 위한 비용과 시간이 대폭 단축
  - 원천 기술 확보를 위해 컴퓨팅 인프라(반도체, 클라우드, 슈퍼컴퓨터 등), 한계 개선을 위한 알고리즘, 양질의 데이터 보강 등 전방위적 대응에 박차를 가할 필요

- 「국가초고성능컴퓨팅 혁신전략(21.5.28.)」, 「국산 AI반도체를 활용한 K-클라우드 추진방안(22.12.12.)」 및 「초거대 AI 경쟁력 강화방안(23.4.14.)」 등 정부에서 추진 중인 전략의 단계별 이행 가속화
- (AI 윤리) AI 오용 사례가 급증하고 있으며, 생성 AI의 등장 등으로 AI 윤리에 대한 관심 또한 계속해서 증가하는 추세, 공정성·편향 등 AI의 신뢰성 검증을 위한 구체적 방안 마련 필요
- 오남용으로 인한 피해 완화를 위해 AI 어플리케이션에 대한 규제, 책임 문제 등 정부의 역할이 중요할 것으로 전망
    - 다수의 AI 윤리 관련 지표들이 등장하고 있는 만큼, 활용 분야별 주요 지표를 파악하고 명확한 기준 제시 필요
  - 또한, 편향 및 독성에 대한 모델 규모의 영향은 학습 데이터 및 완화 방법에 의해 변동될 수 있으며, 공정한 모델일수록 덜 편향될 가능성 존재
    - AI 활용 시 명령 조정 및 데이터 필터링과 같은 과정을 채택하도록 하는 등 편향성 및 독성을 완화를 위한 방법 가이드라인 제시 필요
- (경제) 세계적인 경기 침체의 영향으로 AI 분야 투자 및 기업 내 AI 도입은 감소추세이나 우리나라는 투자 확대를 통해 6위권을 차지, 지속적인 세계 시장 비중 제고가 필요
- 전 세계 AI 분야의 투자 금액과 투자 건수, 피 투자기업 수 모두 전년 대비 감소
    - 10억 달러 이상의 대형 투자를 제외한 규모별 투자 건수가 감소하면서, 신규 피투자 기업 수를 줄이고 건당 투자액을 늘리는 집중화 현상 심화
  - AI 시장 내에서 미국의 영향력과 비중은 여전히 높은 편
    - 미국은 전 세계 AI 스타트업 투자의 약 52%를 차지하고 있으며, AI 채용 공고의 비중 역시 높아 시장에서의 지배력을 확장\*
- \* 다음으로, 중국은 AI 민간투자 규모, AI 스타트업 수 등 주요 양적 경제지표에서 2위를 유지 중이며, 영국, 캐나다, 이스라엘, 독일 등도 꾸준히 상위 시장 국가에 포진
- 한국은 근로자 AI 역량과 산업용 로봇 도입이 높고 최근 AI 스타트업

투자액이 증가하며 긍정적 측면도 있으나, 여전히 세계 시장에서는 낮은 비중

- AI 분야에서 국내 스타트업이 유니콘 기업으로 성장하도록 전문인력양성, 연구개발, 투자 확대 등 전방위적 AI 생태계 마련을 위한 노력이 필요

□ (교육) 해외로부터 배출되는 시고급인력을 체계적으로 관리하여, 국내 산업계 유입을 유도하기 위한 정책 고려가 필요

- 북미권 대학의 석·박사학위 취득자 중 60%이상이 유학생인 점을 비추어볼 때 한국 유학생도 상당수 존재할 것으로 예상
  - 해외 학위 취득 인력 다수가 현지 취업을 선호한다는 연구결과<sup>35)</sup> 등이 존재해 고급인력 유출이 우려
- 장기적으로 산학협력 활성화 및 학술분야 국가재정 절감을 모색하기 위해, 국내 AI 부문 연구 활동의 국가예산 및 외부 투자 의존성 정량화 및 검토 필요
  - 북미 대학들의 NSF기금 의존성 완화 추세는 산학협력에 따른 실증연구 확대 및 국가예산 절감 효과 차원에서 긍정적이며, 국내차원의 사정을 파악해봄으로써 동일 효과를 목표로 한 정책 고려가 가능

□ (정책 및 거버넌스) 정책입안자들의 인공지능에 대한 관심이 지속적으로 증가되는 가운데, AI 부작용 및 사회적 비용 최소화를 위한 관련 입법의 세심한 추진이 요구

- 127개국의 입법 관련 기록을 조사한 결과 AI와 관련법 입안 건수는 2016년 1건에서 2022년 37건으로 증가
  - 미국은 인공지능에 관한 논의에서 본격적인 입안을 추진할 뿐만 아니라 인공지능 투자 확대 지속\*
    - \* 2017년 이후 미국 정부의 인공지능 관련 지출은 약 2.5배 증가
- 최근 AI 관련 소송이 증가함에 따라 관련 법제 정비의 지속적 추진 필요
  - 2022년 한해 미국에서만 110건의 인공지능 관련 소송이 발생하는 등 세계적으로 2016년 대비 7배 정도 AI 관련 소송이 증가

35) KISTEP(2010), “국내 이공계 박사의 해외유출 특성 및 요인 분석”

- AI 관련 소송의 양적 증가와 더불어 소송의 내용을 심층적으로 분석하여 AI 관련 법적 핵심 쟁점을 도출하고 사회적, 제도적 해결 방안 모색 필요
  - 인공지능 입법 과정에서의 국가별 상황을 고려하여 다양한 고려사항 숙의 필요
    - 향후 인공지능 신뢰성 확보를 위한 입법이 더욱 활발히 전개될 것으로 예상되는 가운데 AI 관련 사회적 비용 최소화를 위해서라도 세심한 입법 영향 또는 규제 영향 분석\* 필요
- \* 예로, 영국은 AI자동화, 일본은 인간기본권 확보, 잠비아는 기후예측 활용가능성을 주로 고려하는 등 국가 상황에 맞춰 인공지능의 부정적, 긍정적 효과 검토

## □ (다양성) AI 분야에서의 다양한 시각과 발상을 확보하기 위해 인적 다양성 확대 요구

- 미국의 경우, AI유관전공 고등교육 졸업생 중 여전히 대다수는 백인 및 남성\*
  - \* △AI 분야 학위 취득자의 성별 다양성은 매년 확대되고 있으나, 남성이 전체의 2/3 이상을 차지하여 여전히 성 불균형이 존재, △AI 분야 박사학위 취득자 중 백인 비중은 2011년 68%에서 2021년 58%로 감소했으며, 같은 기간에 아시아계는 25%에서 29%로 증가
- 국내 AI산업도 남초현상이 두드러지고, 여성 AI 연구자의 활동성은 낮은 수준\*
  - \* △2022년 기준 전체 인공지능 산업 종사자 중 여성종사자 비중은 19.6%(7,675명)로 추정(SPRI, '22), △ AI분야 논문 여성저자 비중은 34개국 중 31위(Nesta, '19)
- AI가 인종 및 성별에 대한 편견과 차별을 생산\*하지 않도록 다양성을 기반으로 한 AI 고급인력 양성 및 지원정책을 통해 사회적·문화적·성별 다양성을 증진할 방안 모색 필요
  - \* △AI 챗봇 '이루다' 서비스 내 성희롱 및 챗봇의 혐오·차별 발언 논란에 서비스 중단('20) △마이크로소프트(MS)의 채팅봇 '테이'는 대화 도중 백인우월주의 및 여성 혐오 내용을 학습하게 되어 왜곡된 대화를 수행하여 서비스 중단('16)

## □ (여론) 전 세계적으로 AI에 대한 이해도 향상과 AI 활용 제품 및 서비스 사용이 증가하고 있으며, 미래 삶의 질 향상에도 도움이 될 것으로 기대

- 여론을 통해 AI의 이점 및 우려에 대한 사람들의 수준을 분석할 수 있으며, 대중의 선호와 관심사에 맞는 AI 기술을 개발하여 기대를 충족

- 하지만, 여론은 국가와 인구통계학적 그룹에 따라 다르기 때문에 정책 입안자 및 업계 전문가가 일자리 감소, 개인정보보호, 보안 등과 같은 우려를 해결하여 대중의 신뢰와 수용을 촉진할 필요
- AI에 대한 대중의 이해를 높일 수 있도록 교육과 인식 캠페인이 필요하며, 그룹 간 인식 차이를 해소하여 사회에 미치는 영향과 오해 완화 노력 촉구
- 시간 경과에 따른 여론 변화를 추적하기 위해 종단적 조사 및 연구 수행 필요\*
- \* 여론이 어떻게 진화하는지 지속적인 모니터링을 하여 AI의 개발, 규제 및 사용과 관련된 의사결정을 분석할 수 있고, 다양한 사용자의 인식 예측 가능

## 참고문헌

### 국외문헌

- Nestor Maslej, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, Vanessa Parli, Yoav Shoham, Russell Wald, Jack Clark, and Raymond Perrault, “The AI Index 2023 Annual Report,” AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2023.
- Outrider, “Could a Chatbot Teach You How to Build a Dirty Bomb?”, 2023.1.31.
- Supplement to the President’s FY 2023 Budget, “The Networking & Information Technology R&D Program and the National Artificial Intelligence Initiative Office”, 2022.11.
- Google Research, “PaLM: Scaling Language Modeling with Pathways”, arXiv:2204.02311, 2022.
- Zhou, Kankan, Yibin Lai, and Jing Jiang. "VLStereoSet: A study of stereotypical bias in pre-trained vision-language models." Association for Computational Linguistics, 2022.
- Smith, Eric Michael, et al. "“I’m sorry to hear that”: Finding New Biases in Language Models with a Holistic Descriptor Dataset." Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. 2022.
- Gros, David, Yu Li, and Zhou Yu. "Robots-Dont-Cry: Understanding Falsely Anthropomorphic Utterances in Dialog Systems." arXiv preprint arXiv:2210.12429, 2022.
- Nesta, “Gender Diversity in AI Research”, 2019.7.

### 국내문헌

- 관계부처합동, 초거대 AI 경쟁력 강화방안, 2023.4.14.
- 소프트웨어정책연구소, 2022 인공지능(AI)융합실태조사, 2023.
- 소프트웨어정책연구소, 2022 소프트웨어(SW)융합실태조사, 2023.
- 관계부처합동, 국산 AI반도체를 활용한 K-클라우드 추진방안, 2022.12.12.
- 관계부처합동, 4차 산업혁명 퀀텀점프를 위한 국가초고성능컴퓨팅 혁신전략, 2021.5.28.
- 과학기술기획평가원, 국내 이공계 박사의 해외유출 특성 및 요인 분석, 2010.

## 주 의

이 보고서는 소프트웨어정책연구소에서 수행한 연구보고서입니다.  
이 보고서의 내용을 발표할 때에는 반드시  
소프트웨어정책연구소에서 수행한 연구결과임을 밝혀야 합니다.



## AI Index 2023의 주요 내용 및 시사점

Summary and Implications of 2023 AI Index Report

경기도 성남시 분당구 대왕판교로 712번길 22 글로벌 R&D 연구동(A) 4층

Global R&D Center 4F 22 Daewangpangyo-ro 712beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do  
[www.spri.kr](http://www.spri.kr)

ISSN 2733-0000