

2024년 국내외 인공지능 산업 동향 연구

Artificial Intelligence Industry Trends in Korea and Overseas in 2024

이해수/유재흥/안미소/안성원

2025. 4.

이 보고서는 2024년도 과학기술정보통신부 정보통신진흥기금을 지원 받아 수행한 연구결과로 보고서 내용은 연구자의 견해이며, 과학기술정보통신부의 공식입장과 다를 수 있습니다.

목 차

제1장 서론	1
제1절 연구 배경 및 목적	1
제2절 연구 내용 및 방법	3
제2장 인공지능 산업 현황 및 전망	4
제1절 AI 산업 현황	4
제2절 전 세계 AI 시장 전망	18
제3장 주요국 인공지능 정책 동향	20
제1절 미국	20
제2절 EU	48
제3절 중국	63
제4절 일본	78
제5절 영국	93
제6절 캐나다	104
제7절 프랑스	115
제8절 독일	122
제9절 싱가포르	128
제10절 우리나라	134
제4장 국제기구 인공지능 동향	155
제1절 OECD	155
제2절 UN	165

제3절 기타 기구(G7, G20 등)	171
제5장 인공지능 관련 주요 기업 동향	175
제1절 국외 기업	175
1. 알파벳(구글)	175
2. 마이크로소프트	194
3. 메타 플랫폼스	199
4. 오픈AI	206
5. 기타 주요 기업(애플, 엔비디아, 엑스, 바이두)	222
6. AI 스타트업	237
제2절 국내 기업	247
1. 네이버	247
2. 카카오	250
3. 기타 주요 기업(SKT, 삼성, KT)	252
4. AI 스타트업	259
제6장 인공지능 기술 및 연구 동향	262
제1절 국제 학술단체, 표준기관, 비영리 연구기관	262
1. 학술단체	262
2. 표준기관	273
3. 연구기관	278
제2절 국내 학술단체, 표준기관, 비영리 연구기관	283
1. 학술단체	283
2. 표준기관	287
3. 연구기관	291

제7장 인공지능 교육 및 인력 동향	295
제1절 교육 분야	295
제2절 인력 및 고용 분야	303
제8장 결론	311
참고 문헌	321
[첨부 1] 2024년 주요국의 AI 관련 분야별 정책 동향 요약	329
[첨부 2] 2024년 주요국 AI 정책 전 세계 맵(World Map)	332
[첨부 3] 국내외 주요 기업 동향	336
[첨부 4] 2024년 글로벌 AI 생태계 동향	343

표 목 차

<표 2-1> 전 세계 주목할 머신러닝 학습모델 현황	5
<표 2-2> HAI AI 지수의 주요 내용	5
<표 2-3> 구글의 Project Astra 활용(예)	17
<표 3-1> 미국의 AI 관련 정책 변화	21
<표 3-2> 미국 AI 이니셔티브(American AI Initiative) 5가지 원칙	23
<표 3-3> 행정명령 제14110호의 목적	24
<표 3-4> 정책영역별 주요 의무요건 지침	25
<표 3-5> 연방차원 정책 현황	26
<표 3-6> 미국 주별 정책 현황	28
<표 3-7> 美 AI 행정명령에 따른 각 기관 수행 활동	31
<표 3-8> NAIRR 파트너 기관	35
<표 3-9> NAIRR 민간 부문, 비영리 단체 및 자선 단체 파트너	37
<표 3-10> 최근 발효된 캘리포니아 AI 규제법	38
<표 3-11> AI 규제법 「SB1047」 주요 내용	38
<표 3-12> 예산관리국(OMB) 주요 발표 내용	41
<표 3-13> 미국 행정발표 이후 주요 AI 성과	47
<표 3-14> EU AI 조정계획의 전략목표와 실행과제	48
<표 3-15> EU의 AI 관련 주요 정책	50
<표 3-16> EU AI 법의 순차적 적용 단계	52
<표 3-17> EU AI 법의 과징금 적용 기준	53
<표 3-18> EU AI 협약의 핵심 요건	54
<표 3-19> EU AI 사무국의 조직 구성	56
<표 3-20> EU AI 사무국의 주요 업무	57

<표 3-21> 유럽평의회 AI 조약의 기본 원칙	61
<표 3-22> 중국 차세대 AI 발전규획의 단계별 목표와 주요 내용	63
<표 3-23> 중국 AI 표준화 체계	67
<표 3-24> 중국AI표준화기술위원회의 업무 범위	68
<표 3-25> AI 보안 위협 대응조치	70
<표 3-26> 2024년 중국 주요 지방정부의 AI 산업 정책	72
<표 3-27> 중국 산시성의 AI 산업 혁신 클러스터 육성 3단계 목표	72
<표 3-28> 산시성 AI 산업혁신 클러스터 육성계획의 5대 중점 임무	73
<표 3-29> 청두시 AI 산업 고품질 발전 3개년 계획의 중점 임무	75
<표 3-30> 베이징시 10대 AI 시범 응용 프로젝트	76
<표 3-31> 청두시 AI 산업 고품질 발전 3개년 계획의 중점 임무	77
<표 3-32> 일본 정부의 AI 정책 추진 경과	81
<표 3-33> AI 사업 활동을 담당하는 주체	84
<표 3-34> 각 주체의 공통 지침	85
<표 3-35> 통합혁신전략 2024의 AI 정책 추진	88
<표 3-36> 방위성의 AI 활용 분야	90
<표 3-37> 방위성 AI 활용 추진을 위한 대응 방안	91
<표 3-38> 영국 국가 AI 전략의 부문별 정책 과제	93
<표 3-39> 영국 AI 규제에 대한 친혁신 프레임워크의 핵심 요소	94
<표 3-40> 영국 정부를 위한 생성 AI 프레임워크의 10개 원칙	95
<표 3-41> 영국 AI안전연구소의 AI 모델 평가 내용	101
<표 3-42> 영국 경쟁시장청의 AI 이행 원칙	102
<표 3-43> 범캐나다 AI 전략의 3대 핵심 영역	105
<표 3-44> FASTER 원칙 일상 업무에서의 생성 AI 사용지침	109
<표 3-45> 생성 AI의 잠재적 이슈와 모범사례	110

<표 3-46> 프랑스 AI 국가 전략 주요 내용과 진행 현황	115
<표 3-47> 프랑스 정부의 최우선 AI 추진 정책	117
<표 3-48> 프랑스 경쟁당국(Autorité)의 AI 활성화 권장 사항	120
<표 3-49> 독일의 AI Made in Germany 12대 전략 요약	123
<표 3-50> 독일의 AI 정책 변화	124
<표 3-51> 독일의 5대 분야별 추진계획	125
<표 3-52> OECD의 독일 AI 권고사항	127
<표 3-53> 싱가포르 국가 AI 전략 2.0(NAIS 2.0)	128
<표 3-54> 연도별 싱가포르의 인공지능 정책 변화	131
<표 3-55> 국민의 일상 속 AI에 대한 경험률 및 도움정도	138
<표 3-56> AI 서비스의 사회적·개인적 영향에 관한 국민의 인식	139
<표 3-57> AI 일상화 주요 추진과제	139
<표 3-58> 새로운 디지털 질서 정립 추진을 위한 핵심과제	141
<표 3-59> 상반기 분과별 논의 결과	143
<표 3-60> 초격차 AI 스타트업 육성 전략의 중점 추진 방향	146
<표 3-61> 산업별 온디바이스 AI 반도체 개발 계획	148
<표 3-62> 4대 AI 플래그십 프로젝트	149
<표 3-63> 4대 분야 정책 추진 방향	149
<표 3-64> 개인영상정보 보호 및 활용을 위한 8대 기본원칙	150
<표 3-65> AI + R&DI 추진전략 핵심내용	151
<표 3-66> 산업데이터 활용 활성화 방안 핵심내용	151
<표 3-67> 3대 핵심 추진과제	152
<표 3-68> 4대 핵심 기능	153
<표 4-1> OECD의 연도별 AI 정책 추진 경과	157
<표 4-2> OECD.AI의 한국과 미국의 주요 트렌드 비교	158

<표 4-3> AI 시스템 개관	161
<표 5-1> 알파벳 기업 현황 (단위: M\$)	175
<표 5-2> 마이크로소프트 기업 현황 (단위: M\$)	194
<표 5-3> 메타 플랫폼스 기업 현황 (단위: M\$)	199
<표 5-4> 오픈AI 기업 현황 (단위: M\$)	206
<표 5-5> 애플 기업 현황 (단위: M\$)	222
<표 5-6> 엔비디아 기업 현황 (단위: M\$)	227
<표 5-7> 엑스 기업 현황 (단위: M\$)	230
<표 5-8> 바이두 기업 현황 (단위: M\$)	234
<표 5-9> AI 스타트업 기업 현황 (단위: M\$)	237
<표 5-10> 네이버 기업 현황	247
<표 5-11> 카카오 기업 현황	250
<표 5-12> SKT 기업 현황	252
<표 5-13> 삼성전자 기업 현황	255
<표 5-14> KT 기업 현황	257
<표 5-15> 라이너 기업 현황	259
<표 6-1> IEEE SA의 AIS 관련 다음과 같은 표준(2024)	268
<표 6-2> ISO/IEC JTC 1/SC 42에서 발행된 표준 목록	275
<표 6-3> AI 분야 국내 주요 연구주제(' 18~' 23)(' 24.5.기준)	284
<표 6-4> Top Conference 세션의 AI 관련 논문	285
<표 6-5> 한국정보통신기술협회의 인공지능 분야에서 발행된 표준 목록	289
<표 6-6> KISDI의 인공지능 분야에서 발행된 주요 연구보고서 목록	292
<표 6-7> SPRi의 AI 분야에서 연구보고서 목록	293
<표 6-8> NIA의 AI 관련 연구보고서 사례	294
<표 7-1> 교육에서 AI 도구의 교육	299

<표> 2024년도 주요국의 AI 관련 분야별 정책 동향 요약	329
<표> 국내 AI 기업 동향	336
<표> 해외 AI 기업 동향	338

그림 목 차

[그림 2-1] 전 세계 국가 AI 전략 수립 현황	4
[그림 2-2] 글로벌 AI 지수 주요국의 전체 순위와 항목별 순위	7
[그림 2-3] 국제적인 AI 거버넌스 이니셔티브 관련 추진 현황	8
[그림 2-4] 2013-2023 글로벌 AI 투자	10
[그림 2-5] 민간 부문의 AI 영역별 투자(2022 vs. 2023)	11
[그림 2-6] 2020년~2024년 2분기 글로벌 AI 벤처 투자 추이	11
[그림 2-7] AI 기업의 평균 투자 규모	12
[그림 2-8] AI 에이전트 시장 전망(2024~2030)	15
[그림 2-9] 엔터프라이즈 AI 에이전트 및 코파일럿	16
[그림 2-10] AI 소프트웨어 시장 규모 (단위: 10억 달러)	18
[그림 2-11] 산업별 전 세계 AI 및 생성 AI 지출	19
[그림 3-1] EU AI 법의 입법 경과	49
[그림 3-2] EU ‘GenAI4EU’의 적용 분야	55
[그림 3-3] 중국 AI 표준화 체계 다이어그램	67
[그림 3-4] 일본 AI 정책 추진도	78
[그림 3-5] AI 전략회의의 주요 활동	80
[그림 3-6] 영국 AI안전연구소의 ‘인스펙트’ 플랫폼 화면	100
[그림 3-7] 디지털 혁신성장 전략의 비전 및 추진과제	137
[그림 3-8] 초격차산업지원 프로그램 6호 AI	144
[그림 3-9] AI 코리아 펀드(안)	145
[그림 3-10] ‘정당한 이익’ 조항 적용을 위한 요건의 내용 및 적용사례	146
[그림 3-11] 기술적·관리적 안전조치 및 정보주체 권리보장 방안	147
[그림 4-1] AI 거버넌스 이니셔티브, 주요 이정표, 2019-2024	168

[그림 4-2] AI에 초점을 맞춘 거버넌스 이니셔티브의 출처(기관/조직)	168
[그림 4-3] 국제 AI 거버넌스 생태계에서 유엔의 역할에 대한 제안	169
[그림 5-1] 구글 딥마인드의 모빌리티 VLA 모델 작동 방식	185
[그림 5-2] 동일 프롬프트에 대한 이마젠2와 이마젠3의 결과물 비교	188
[그림 5-3] ASPIRE 적용 전과 적용 후 LLM 답변 비교	189
[그림 5-4] ‘프런티어 안전 프레임워크’의 구성요소 간 관계	190
[그림 5-5] 마이크로소프트 ‘바사-1’의 결과물 예시	197
[그림 5-6] ‘파이 3.5’와 주요 소형언어모델(SLM)의 매개변수 크기, 품질 비교	198
[그림 5-7] 라마 3.1 405B와 경쟁 모델의 인간 평가 결과 비교	204
[그림 5-8] 메타 FAIR의 자가 학습 평가자 파이프라인	205
[그림 5-9] 애플의 멀티모달 모델 ‘MM1’의 과제 수행 능력 예시	224
[그림 5-10] 애플 인텔리전스 기반 모델의 개발 파이프라인	226
[그림 5-11] 허깅페이스 오픈 LLM 리더보드의 스테이블 LM 2 1.6B 평가	241
[그림 5-12] 스테이블 디퓨전 3의 글자 생성 예시	242
[그림 6-1] 연구 분야별 AI 관련 출판물 발간 추이(2010-2022)	262
[그림 6-2] AI 특허 승인 건수 추이(2010-2022)	263
[그림 6-3] 인구 10만 명당 AI 특허 건수(2022)	264
[그림 6-4] 국가별 주목할 만한 머신러닝 모델 출시 추이(2003-2023)	264
[그림 6-5] 주목할 만한 머신러닝 모델 출시 추이(2003-2023)	265
[그림 6-6] 접근 유형별 기반 모델 출시 추이(2019-2023)	266
[그림 6-7] 기관/기업별 기반 모델 출시(2023)	266
[그림 6-8] AI 컨퍼런스 참석자 수 추이(2010-2023) (단위: 천 명)	267
[그림 6-9] AI 관련 표준 수	273
[그림 6-10] 텐스AV 알고리즘의 작동 방식	278
[그림 6-11] 법률 업무 처리 시뮬레이션	281

[그림 6-12] 인공지능 분야 국내 연구자의 논문 수	283
[그림 6-13] 국내 상위 10개 학술·연구기관별 논문 및 연구자 수	284
[그림 6-14] AI+ 인증 절차	287
[그림 6-15] 산업 인공지능 표준화 포럼 구성	288
[그림 6-16] 생성 시각지능 기술의 향후 발전 방향과 ETRI 기술 적용	291
[그림 7-1] 북미지역 신규 AI 전공 박사의 진출 분야	296
[그림 7-2] 전산학, 컴퓨터공학 및 정보학 전공 학부 졸업자 수(2022)	297
[그림 7-3] 연도별 미국 K-12 전산학 교육 주 비율 추이	297
[그림 7-4] AP 전산학 시험 응시 건수	298
[그림 7-5] 생성 AI를 활용한 교수학습 프레임워크(GATe 프레임워크)	302
[그림 7-6] 국가별 AI 채용 공고 비율 추이(2014-2023)	303
[그림 7-7] 2023년 미국 AI 채용 공고의 상위 10개 전문기술(2011-13년 대 2023년)	304
[그림 7-8] 2023년 지역별 AI 채용률 연간 증가율	304
[그림 7-9] 2023년 지역별 AI 인재 집중도	305
[그림 7-10] 기업주가 AI가 직장에서 기여 및 위해로 인식하는 비율	306
[그림 7-11] 국가별 AI 노출인 높은 분야의 기술 수요 변화	307
[그림 7-12] 직업 내 AI 노출과 직업 내 임금 불평등	308
[그림 7-13] AI 노출 및 보완성에 따른 고용 점유율(지역별/국가별)	309
[그림 7-14] 주요 기업의 AI 도입 현황	310
[그림] 2024년 주요국 AI 정책(정책/법제 분야) 전 세계 맵(World Map)	332
[그림] 2024년 주요국 AI 정책(기업/산업 분야) 전 세계 맵(World Map)	333
[그림] 2024년 주요국 AI 정책(기술/연구 분야) 전 세계 맵(World Map)	334
[그림] 2024년 주요국 AI 정책(인력/교육 분야) 전 세계 맵(World Map)	335
[그림] 2024년 글로벌 AI 생태계 동향(국내 AI 기업)	343
[그림] 2024년 글로벌 AI 생태계 동향(해외 AI 기업)	344

요 약 문

1. 제 목 : 2024년 국내외 인공지능 산업 동향 연구

2. 연구 목적 및 필요성

- 생성AI 기술의 부상으로 인공지능 산업 생태계가 빠르게 변화
 - AI 기술은 산업 혁신과 정책 변화, 각국의 국가적 대응을 가속화하며 주요국은 이를 국가 안보와 기술 패권의 핵심으로 간주하고 하며, 국가 AI 전략을 개발하고, 법제화에 나서고 있는 상황
 - 관련 하드웨어, 서비스 개발에 대한 투자뿐만 아니라 전산업에서 AI융합이 본격화되고 있으며 글로벌 빅테크를 중심으로 생성 AI 생태계를 선점하기 위한 주도권 경쟁 가속화
 - 이와함께 AI 안전·신뢰성에 대한 우려가 커지며 기업들은 자체 지침을 마련하고, 국제기구와 표준기구는 AI 윤리와 기준 정립에 주력
 - 생성 AI 등의 기술적 발전으로 인공지능 산업 생태계가 급변하고 있는 가운데, 국내 인공지능 산업 경쟁력 강화와 활용 확산, 글로벌 수준의 인공지능 신뢰성 확보 정책 마련을 위한 국내외 인공지능 산업 동향 파악 및 분석 필요
- 이 연구는 급변하는 국내외 인공지능 산업 생태계를 분석하여, 국내 산업 경쟁력을 강화하고 인공지능 활용의 확산을 지원하기 위한 정책적 자료를 마련하는 것이 주된 목표
 - 주요국, 주요 기관 및 학술단체, 주요 기업의 동향을 조사하여 현재의 상황을 분석하고, AI 기술의 발전 및 산업 변화에 대한 전망을 통해 적시에 효과적인 AI 정책을 개발하고 정책적 의사결정을 지원하기 위한 기초 자료를 제공하며, 이를 통해 정책 과제를 발굴
 - 인공지능 산업과 관련된 폭넓은 조사를 바탕으로 국내외 AI 정책 관련 주요 자료(정책, 법률, 권고안 등)를 확보하여 정책 고도화 및 발전에 활용

3. 연구의 구성 및 범위

- AI 산업 현황 및 전 세계 AI 시장전망에 대한 개괄적 정리
- 국내외 주요국·국제기구 정책 동향

- 주요국에는 미국, EU, 중국, 일본, 영국, 캐나다, 프랑스, 독일, 싱가포르 포함
- 국제 기구/회의로 OECD, UN, G7정상회의, G20정상회의 동향 분석
- 국내외 인공지능 관련 주요 기업 동향 분석
 - 해외 글로벌 빅테크(구글, 마이크로소프트, 메타, 오픈AI, 애플, 엔비디아 등) 기업 및 국내 주요 기업 (네이버, 카카오, SKT)을 포함한 주요 AI 스타트업 동향 포함
- 국제 학술 단체 연구, 표준기관, 비영리 연구 기관 동향 분석
 - IEEE, ACM 및 최상위 AI 컨퍼런스 발표 및 ISO/IEC 국제 표준화 동향
 - 국내 TTA 인공지능 표준화 동향 및 관련 인공지능 학회 주요 연구 동향
- 인공지능 교육, 인력 및 고용 관련 동향 분석

4. 연구 내용 및 결과

- 생성 AI 기술과 초거대 AI 모델의 확산
 - 빅테크 기업들은 초거대 AI 모델을 기반으로 맞춤형 서비스를 제공하며 시장 점유율 확대
 - AI 에이전트 서비스의 확산과 함께 의료, 교육, 금융 등 주요 산업에서 AI 기반 모델이 정밀 의료, 개인화 학습, 리스크 관리 등 혁신적 솔루션을 제공
 - AI 스타트업에 대한 투자가 급증하며, 새로운 시장 창출과 기술 혁신을 주도하고 있으며, 생성 AI를 활용한 맞춤형 솔루션, 자율주행 시스템, 스마트 제조 등 구체적인 활용 분야를 중심으로 새로운 시장이 창출될 전망
 - 제조, 물류, 유통 등 다양한 분야에서 AI 기반 자동화와 데이터 분석이 표준으로 자리 잡고 있으며, 인간과 AI의 협업을 중심으로 한 5차 산업혁명의 기틀이 마련
- AI 윤리, 거버넌스 및 안전성 강화를 위한 글로벌 AI 협력과 표준화 노력 강화
 - AI 기술의 급격한 발전과 확산은 윤리적 문제, 거버넌스 체계 부재, 안전성에 대한 우려 심화로 국제기구와 각국 정부는 AI 윤리와 안전성 확보를 위한 법적·제도적 프레임워크를 마련
 - 특히, 우리나라를 비롯하여 미국, 일본, 캐나다는 AI 안전연구소를 설립하여 기술적·사회적 대응 방안을 모색
- AI 연구와 학계의 발전 및 인력 전환과 AI 교육의 중요성 확대
 - AI 기술이 학계에 적극적으로 도입되면서 최첨단 연구가 활발히 진행되고 있으며, AI와 공학, 사회과학, 생물학 등 다양한 학문 영역이 결합해 새로운 지식을 창출하고 있고, 특히, AI의 사회적 책임과 지속 가능성을 주제로 한 연구 확대
 - AI 기술의 발전은 고용의 변화를 촉발하여 인력을 재교육(Reskilling)이 증가하고, AI와 인간의 협업 능력을 키우는 방향으로 이동할 전망

5. 정책적 활용 내용

- 이 연구의 내용 및 결과는 정부의 인공지능 정책 수립(초거대 인공지능 경쟁력 강화, 전국민 인공지능 일상화 전략, 국가 AI 전략방향 등) 및 관련 법안 마련(인공지능 기본법)을 위한 현황 분석의 기초 자료로 활용 되었음

6. 기대효과

- 국내외 환경 변화에 대한 시의성 있는 정보 제공으로 정부 정책 대응력 제고
- 인공지능 산업 관련 중장기 정책 연구 기획 및 정부의 시의성 있는 정책 수립에 지원
- 인공지능 산업 관련 다양한 국내외 정보(정책, 시장, 기업 전략, 표준, 인력, 연구 개발 동향)들을 업계, 연구자 및 이용자 등 다양한 이해관계자에게 제공함으로써 민간의 AI 산업 이해도 및 전략 대응력 향상에 기여

SUMMARY

1. Title: Artificial Intelligence Industry Trends in Korea and Overseas in 2024
2. Purpose and Necessity of the Research
 - The emergence of generative AI technologies is rapidly transforming the AI industry ecosystem
 - AI technologies are accelerating industrial innovation, policy changes, and national responses, with major countries considering AI as a cornerstone of national security and technological leadership. In this context, nations are developing national AI strategies and pushing for legislation.
 - Investments are expanding not only in hardware and service development but also in the integration of AI across all industries, while global big tech companies are intensifying their competition to take the lead in the generative AI ecosystem.
 - Concurrently, growing concerns about AI safety and reliability are leading companies to establish their own guidelines, while international organizations and standard-setting bodies focus on defining AI ethics and standards.
 - Amid these technological advancements in generative AI and the rapid evolution of the AI industry ecosystem, there is a pressing need to analyze domestic and international AI industry trends to strengthen Korea's AI industry competitiveness, promote its widespread utilization, and develop policies to ensure globally recognized AI reliability.
 - This research aims to analyze the rapidly changing domestic and international AI industry ecosystems to provide policy materials that support the strengthening of domestic industrial competitiveness and the broader adoption of AI.
 - By examining trends in major countries, institutions, academic organizations, and leading companies, the study will analyze the current state of the AI landscape. It will also forecast advancements in AI technology and industry transformations, providing foundational data to develop timely and effective AI policies and support policy decision-making. Through this, the research seeks to identify key policy tasks.
 - Based on extensive research related to the AI industry, the study will collect

critical resources on AI-related policies(e.g., policies, laws, recommendations) both domestically and internationally, leveraging these materials to enhance and advance AI policy development.

3. Composition and Range

- Overview of the Current State of the AI Industry and Global AI Market Outlook
- Policy Trends of Major Countries and International Organizations
 - Major countries include the United States, EU, China, Japan, the United Kingdom, Canada, France, Germany, and Singapore.
 - Analysis of trends in international organizations and meetings, including OECD, UN, G7 Summits, and G20 Summits.
- Analysis of Trends in Major AI-Related Companies
 - Covers global big tech companies(Google, Microsoft, Meta, OpenAI, Apple, NVIDIA, etc.), major domestic companies(Naver, Kakao, SKT), and trends in prominent AI startups.
- Trends in Research by International Academic Organizations, Standards Bodies, and Non-Profit Research Institutions
 - Analysis of publications by IEEE, ACM, and leading AI conferences, as well as developments in ISO/IEC international standardization.
 - Trends in AI standardization efforts by Korea's TTA and research trends in key domestic AI academic associations.
- Analysis of Trends in AI Education, Workforce, and Employment

4. Main Contents and Results

- Proliferation of Generative AI Technologies and Large-Scale AI Models
 - Big tech companies are expanding their market share by offering personalized services based on large-scale AI models.
 - Along with the spread of AI agent services, AI-based models are driving innovation in key industries such as healthcare, education, and finance by providing solutions like precision medicine, personalized learning, and risk management.

- Investment in AI startups is surging, fostering technological innovation and the creation of new markets. Emerging fields include personalized solutions using generative AI, autonomous driving systems, and smart manufacturing, which are expected to drive the development of new markets.
- AI-driven automation and data analytics are becoming standard across various sectors such as manufacturing, logistics, and distribution. The foundation of the Fifth Industrial Revolution is being laid, focusing on collaboration between humans and AI.
- o Strengthening Global AI Cooperation and Standardization for AI Ethics, Governance, and Safety
 - The rapid development and spread of AI technologies have intensified ethical concerns, governance gaps, and safety issues. In response, international organizations and governments are establishing legal and institutional frameworks to ensure AI ethics and safety.
 - Notably, countries like South Korea, the United States, Japan, and Canada have established AI safety research institutes to explore technical and societal solutions.
- o Advancement of AI Research, Academic Development, and the Growing Importance of Workforce Transition and AI Education
 - The integration of AI technologies into academia is driving cutting-edge research, fostering the convergence of AI with disciplines like engineering, social sciences, and biology to create new knowledge. Research on AI's social responsibility and sustainability is particularly expanding.
 - The advancement of AI technology is reshaping employment, increasing the need for workforce reskilling, and shifting the focus toward developing collaborative capabilities between humans and AI.

5. Policy use

- o The findings and outcomes of this research have been utilized as foundational materials for analyzing the current status to inform government AI policy development (e.g., strengthening competitiveness in large-scale AI, strategies for mainstreaming AI in daily life, and national AI strategic directions) and for drafting related legislation, such as the Artificial Intelligence Framework Act.

6. Research Implication and Expected Effects

- Enhancing the government's policy responsiveness by providing timely information on domestic and international environmental changes.
- Supporting the planning of mid- to long-term policy research and enabling the government to develop timely AI-related policies.
- Contributing to improved understanding and strategic readiness in the private sector by providing diverse stakeholders—including industries, researchers, and users—with comprehensive domestic and international AI-related information (e.g., policies, markets, corporate strategies, standards, workforce, and R&D trends).

CONTENTS

Chapter 1. Introduction	1
1.1. Research Background and Purpose	1
1.2. Research Content and Methods	3
Chapter 2. Artificial Intelligence Industry Status and Outlook	4
2.1. AI Industry Status	4
2.2. Artificial Intelligence Industry Outlook	18
Chapter 3. Trends in AI Policies of Major Countries	20
3.1. United States	20
3.2. European Union (EU)	48
3.3. China	63
3.4. Japan	78
3.5. United Kingdom	93
3.6. Canada	104
3.7. France	115
3.8. Germany	122
3.9. Singapore	128
3.10. Korea	134
Chapter 4. Trends in AI of International Organizations	155
4.1. OECD	155
4.2. UNESCO	165
4.3. Other Organizations(G7, G20, etc.)	171
Chapter 5. Trends in Major Companies Related to AI	175
5.1. Foreign Companies(Google, MS, Meta, OpenAI, etc.)	175
5.2. Domestic Companies(Naver, Kakao, etc.)	247
Chapter 6. Trends in AI Technology and Research	262

6.1. International Academic Organizations, Standards Organizations, and Non-Profit Research Institutes	262
6.2. Domestic Academic Organizations, Standards Organizations, and Non-Profit Research Institutes	283
Chapter 7. Trends in AI Education and Human Resources	295
7.1. Education	295
7.2. Human Resources and Employment	303
Chapter 8. Conclusion	311
References	321
Appendix 1. Summary of AI Policy Trends by Countries in 2024	329
Appendix 2. Global Map of AI Policies in 2024	332
Appendix 3. Trends in major domestic and foreign companies	336
Appendix 4. Global AI Ecosystem Trends in 2024	343

제1장 서론

제1절 연구 배경 및 목적

1. 연구 배경

최근 생성 AI 기술의 급부상으로 인공지능 산업 생태계의 변화가 빠르게 전개되고 있다. AI 기술의 비약적인 발전은 산업계 전반의 혁신을 이끌며, 관련 정책과 법·제도의 변화뿐 아니라 각국의 국가적 대응을 가속화하고 있다. AI 기술의 발전이 산업 전반에 미치는 영향력이 확대되면서 주요국들은 AI를 국가 안보와 기술 패권의 핵심 기술로 간주하고 있다. 이에 따라 정부 차원에서 AI 산업 정책을 적극적으로 강화하고 있다. 미국은 백악관을 중심으로 국가 차원의 AI 전략을 추진하고 있으며, 2021년 ‘국가 AI 이니셔티브법’을 제정해 연구개발 투자 기반을 마련했다. 2022년에는 안전한 AI 시스템 개발과 보급을 촉진하기 위해 ‘AI 권리장전’을 발표하였고, 2023년에는 안전성과 신뢰성을 확보하기 위해 ‘안전하고 신뢰할 수 있는 AI에 관한 행정명령’을 공포했다. 유럽연합(EU)은 2021년 4월 AI 법안(AI Act)을 제안한 후 오랜 논의 끝에 2024년 5월 최종 승인하였다. 이 법안은 AI 개발과 활용에서 신뢰와 기본권 보장을 목표로 한다. 이외에도 영국, 캐나다, 중국, 일본 등 주요국은 AI 산업 진흥을 위한 국가 전략을 수립하고 있으며, 대규모 투자와 인재 양성을 강화하고 있다.

2022년 하반기, 오픈AI가 ChatGPT를 출시한 이후 구글, 마이크로소프트, 아마존, 메타 등 빅테크 기업들이 초거대 AI 개발 경쟁에 적극적으로 참여하고 있다. 이들 기업은 자체 AI 기반 모델을 선보이며 생성 AI 생태계를 선점하기 위한 경쟁을 가속화하고 있다. 이와 함께 AI 안전성에 대한 우려가 커지면서, 각 기업은 자체적인 AI 안전 및 신뢰 지침을 개발해 적용하고 있다. 이러한 변화 속에서 AI 환경은 급격히 진화하고 있으며, AI로 인한 잠재적 위험에 대한 우려도 증가하고 있다. 이에 국제기구(OECD, 유네스코 등)와 국제표준기구들은 AI 모델의 기준을 정립하고, AI 윤리 헌장 등을 마련하여 이를 보급하고 있다.

국내에서도 정부와 산업계는 AI를 국제 산업경쟁력과 경제성장을 이끄는 핵심 동력으로 인식하고 있다. 이에 따라 AI 핵심 모델 개발과 신산업 육성을 본격적으로 추진하고 있다.

2023년 4월, 과학기술정보통신부는 관계부처와 협력하여 GPT 등 초거대 AI와 생성 AI 서비스의 급속한 확산에 대응하기 위해 ‘초거대 AI 경쟁력 강화 방안’을 수립했으며, 2024년 4월에는 관계부처 합동으로 ‘AI G3 도약을 위한 AI디지털 혁신성장 전략’을 발표하였다. 그리고 국회는 2024년 12월 AI 기본법을 의결해 거버넌스 체계와 산업 진흥 및 안전한 AI를 위한 법적 토대를 마련했다. 국내 IT 기업들 또한 독자적인 초거대 AI 플랫폼을 확보하기 위한 기술 개발에 박차를 가하고 있다.

이처럼 AI 기술이 발전하고 국제 정책 환경이 급변함에 따라, 빠르게 변모하는 AI 산업 동향을 신속히 파악하고 이에 맞춰 국내외 환경에 대응할 수 있는 AI 정책의 필요성이 그 어느 때보다 높아지고 있다.

2. 연구 목적

AI 기술의 빠른 진화와 국제적 정책 환경의 급변 속에서, 정부는 AI 산업을 육성하고 활용을 확산시키며, 글로벌 수준의 AI 안전성과 신뢰성을 확보하기 위한 정책 마련에 집중하고 있다. 이러한 정책은 변화하는 환경에 맞춰 지속적인 고도화가 요구된다. 효율적이고 바람직한 AI 정책 방향을 설정하기 위해서는 주요국, 국제기구, 학술단체, 주요 기업의 동향을 종합적으로 파악하고, 이를 통해 현황을 진단하며, 향후 AI 기술 발전과 산업 전망을 분석해야 한다. 이는 시의적절한 AI 정책 개발과 정책 의사결정을 지원하는 데 필수적인 기초 자료로 활용될 수 있다. 이에 본 연구는 국내외 주요국 및 국제기구의 AI 관련 정책 동향, 주요 기업 및 스타트업 동향, 학술단체 및 비영리 연구기관의 연구 동향, 그리고 AI 기술 개발, 표준, 특허 등의 주요 흐름을 조사·분석한다. 이를 통해 국내 AI 정책의 효과적인 추진을 지원하고, 새로운 AI 정책과제를 발굴하는 데 기여할 수 있는 유의미한 자료를 제공하고자 한다.

제2절 연구 내용 및 방법

1. 연구 내용

본 연구에서는 주요국, 국제기구, 주요 기업, 인공지능 관련 기술 및 연구 동향을 조사·분석하고자 하였다. 주요국으로는 AI 기술 수준이 높고 국제적 영향력을 가진 국가들을 선정하였으며, 미국, EU, 중국, 일본, 영국 등을 포함한 주요국과 캐나다, 프랑스, 독일, 싱가포르 및 기타 국가를 대상으로 하였다. 또한, 우리나라 동향도 조사 범위에 포함되었다. 국제기구로는 AI 관련 심도 있는 정책문서를 발간하는 OECD와 UN 등을 중심으로 조사했다. 기업 분야에서는 해외는 빅테크 기업인 알파벳(Google), 마이크로소프트, 메타플랫폼, 아마존 등 주요 IT 기업과 오픈AI와 같은 신흥 AI 기업을 대상으로 했다. 국내 기업으로는 네이버, 카카오와 같은 주요 AI 기업, 삼성전자, LG전자, KT, SKT, LGU+와 같은 대기업, 그리고 주요 AI 스타트업들을 조사하였다. 또한, 기술 및 연구 동향 조사를 위해 국내외 학술단체, 표준기관, 연구기관을 포함하였으며, 교육 및 인력 동향은 별도의 항목으로 구분하여 주요국과 국제기구, 교육기관 및 연구기관의 사례를 분석하였다. 이와 같은 조사 범위를 통해 국내외 AI 산업의 현황과 관련 요소를 종합적으로 파악하고자 했다.

2. 연구 방법

본 연구는 국내외 AI 산업 동향 자료를 확보하기 위해 주요국 및 국제기구의 공식 홈페이지에서 제공하는 보도자료와 공개 문서를 1차 문헌으로 활용했다. 추가적으로, 미디어에서 보도된 기사와 자료를 보완적으로 활용했다. 기업 조사에서는 보도자료, 언론 기사, 기업 운영 블로그 등을 참조했으며, 학술단체와 표준기관에 대해서는 공식 홈페이지와 학술대회 홈페이지에서 발간된 자료를 수집하여 조사했다. 이외에 신뢰도 높은 시장조사기관(Gartner, IDC 등)의 자료 및 학술 데이터베이스(Springer, IEEE Xplore 등)를 활용하였으며, 구글 Scholar 등 AI 기반 검색 도구로 최신 자료를 탐색하였다. 조사 카테고리별 세부 키워드를 정의하여 활용하였다. 필요 시 AI 전문가, 정책 분석가 등 관련 분야 전문가 검토를 통해 조사 결과의 정확성과 신뢰성을 보장하도록 하였다. 연구의 시간적 범위는 2024년에 발표된 자료를 중심으로 하였으며, 결과보고서에는 동향의 배경 설명 및 정책 흐름 분석에 필요한 중요한 자료도 포함하였다. 이를 통해 2024년 AI 산업의 최신 동향과 관련 정책의 맥락을 효과적으로 전달하고자 했다.

제2장 인공지능 산업 현황 및 전망

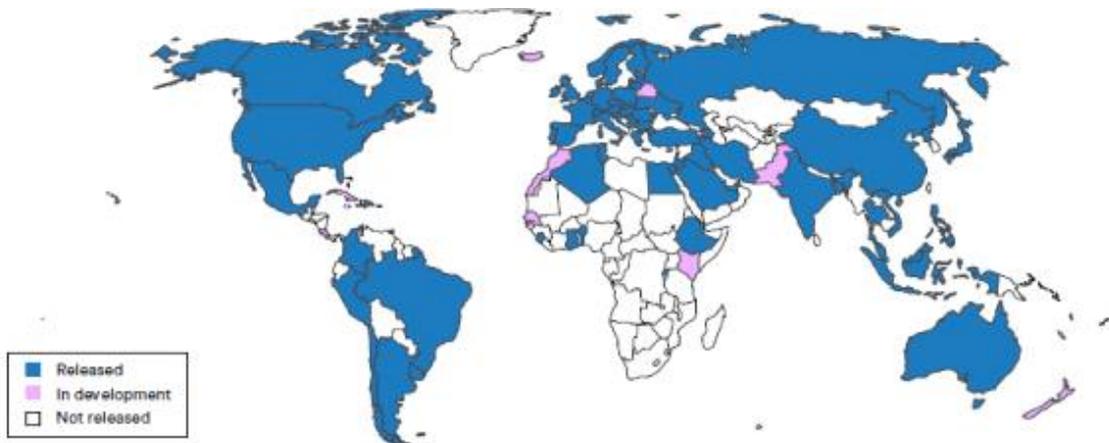
제1절 AI 산업 현황

1. 전 세계 AI 현황 및 수준

2024년 4월, 미국 스탠퍼드 대학교 HAI(Human-Centered AI)는 AI 분야의 국제적 수준을 비교 분석한 ‘AI 인덱스 보고서(Artificial Intelligence Index Report) 2024’ 를 발표하였다. AI 인덱스 보고서는 연구개발, 기술적 성과, 책임 있는 AI, 경제, 과학 및 의료, 정책과 거버넌스, 다양성, 대중 의견 측면(총 9개 분야)의 AI 발전 현황을 평가하며, 통계정보와 추세를 제시했다.

HAI 보고서는 전 세계 국가의 AI 전략 수립 현황과 추세를 조사해서 발표하고 있다. 이 보고서에 따르면, 캐나다가 2017년 3월에 세계 최초로 국가 AI 전략을 시작한 이후, 2024년 4월 기준, 75개의 국가 AI 전략이 공개되었다. 2019년에는 가장 많은 24개의 전략이 발표되었으며, 2023년에는 중동, 아프리카, 카리브해 국가의 8개 새로운 전략이 추가되어 AI 전략이 전 세계적으로 확장했다. 아래 그림은 전 세계 국가의 AI 전략 수립 현황을 공표된 국가(Released), 개발 중인 국가(In Development), 미공표 국가(Not Released) 세 가지로 구분하여 나타낸다. 이 그림은 아프리카 지역 그리고 중동과 남미 일부 지역을 제외하면 전 세계 국가들 대부분 AI 전략을 공표한 것을 보여준다.

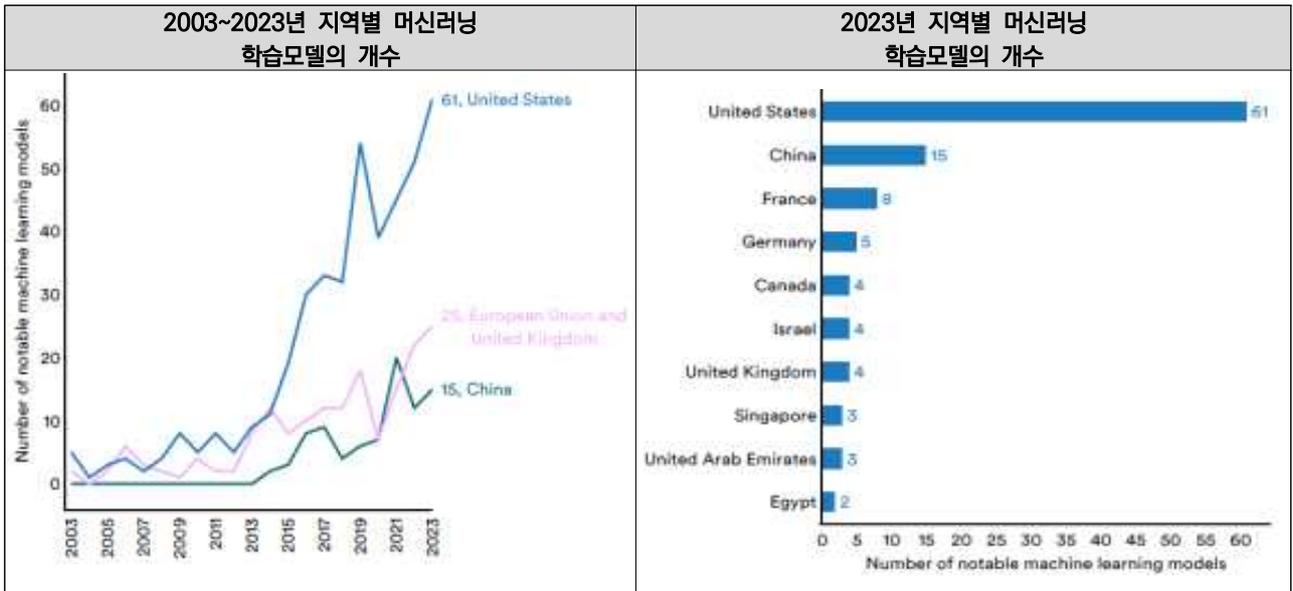
[그림 2-1] 전 세계 국가 AI 전략 수립 현황



출처: Stanford University, Artificial Intelligence Index Report 2024.

또한, HAI는 최신 생성 AI의 확산을 반영하여 2003~2023년 전 세계 국가의 머신러닝 학습 모델의 개수를 조사했다. 아래 그림과 같이 2023년 기준, 미국이 61개로 최대이며, 다음으로 중국, 프랑스, 독일, 캐나다 및 이스라엘이 주목할 만한(notable) 머신러닝 학습모델을 많이 개발하고 있는 것으로 나타났다.

<표 2-1> 전 세계 주목할 머신러닝 학습모델 현황



출처: Stanford University, Artificial Intelligence Index Report 2024

HAI의 2024년 AI 지수 보고서에 전 세계의 AI 발전 추세를 다음과 같이 10개의 주요 내용으로 제시하였다.

<표 2-2> HAI AI 지수의 주요 내용

구분	주요 내용	
1	시는 일부 작업에서 인간을 능가하지만, 모든 작업에서 그렇지는 않음	<ul style="list-style-type: none"> 시는 이미지 분류, 시각적 추론, 영어 이해 등에서 인간을 앞서지만, 경쟁 수준의 수학, 시각적 상식 추론, 계획 등 복잡한 작업에서는 여전히 뒤처짐
2	산업계가 최첨단 AI 연구를 주도	<ul style="list-style-type: none"> 2023년 산업계는 51개의 주목할 만한 머신러닝 모델을 생산한 반면, 학계는 15개에 그쳤으며, 산업-학계 협력으로 21개의 모델이 개발되어 최고치를 기록
3	최첨단 모델의 훈련 비용 급증	<ul style="list-style-type: none"> OpenAI의 GPT-4는 약 7,800만 달러, 구글의 Gemini Ultra는 1억 9,100만 달러의 컴퓨팅 비용이 소요되어 전례 없는 수준에 도달

구분	주요 내용	
4	미국이 주요 AI 모델의 선도국	<ul style="list-style-type: none"> 2023년 미국 기반 기관에서 61개의 주목할 만한 AI 모델이 개발되어, 유럽연합의 21개, 중국의 15개를 크게 앞섬
5	대규모 언어 모델(LLM)의 책임성 평가 표준화 부족	<ul style="list-style-type: none"> OpenAI, 구글, Anthropic 등 주요 개발사들이 서로 다른 책임성 벤치마크를 사용하여 모델을 테스트하고 있어, 체계적인 비교에 어려움이 발생
6	생성 AI 투자 급증	<ul style="list-style-type: none"> 전체 AI 민간 투자는 감소했지만, 생성 AI에 대한 투자는 2022년 대비 약 8배 증가하여 252억 달러에 도달 OpenAI, Anthropic, Hugging Face, Inflection 등이 주요 투자 유치를 이끌었음
7	AI가 노동 생산성과 작업 품질 향상에 기여	<ul style="list-style-type: none"> 2023년 여러 연구에서 AI가 작업 속도와 품질을 향상시키고, 저숙련 노동자와 고숙련 노동자 간의 격차를 줄이는 데 도움이 됨을 확인 그러나 적절한 감독 없이 AI를 사용할 경우 성과 저하의 위험도 존재
8	과학적 진보 가속화	<ul style="list-style-type: none"> 2022년에 이어 2023년에는 AlphaDev의 알고리즘 정렬 효율화, GNoME의 소재 발견 촉진 등 과학 분야에서 AI 응용이 더욱 확대
9	미국 내 AI 관련 규제 급증	<ul style="list-style-type: none"> 2023년 미국에서 AI 관련 규제가 25건으로 증가하여, 2016년 1건 대비 크게 늘어났으며, 전년 대비 56.3% 증가
10	전 세계적으로 AI의 잠재적 영향에 대한 인식과 우려 증가	<ul style="list-style-type: none"> Ipsos 조사에 따르면, 향후 3~5년 내 AI가 삶에 큰 영향을 미칠 것이라고 생각하는 비율이 60%에서 66%로 증가 또한, AI 제품과 서비스에 대한 불안감을 느끼는 비율이 52%로, 2022년 대비 13%p 상승 Pew 데이터에 따르면, AI에 대해 우려하는 비율이 37%에서 52%로 증가

출처: Stanford University, Artificial Intelligence Index Report 2024

영국 데이터 분석 업체 토터스 인텔리전스(Tortoise Intelligence)가 2024년 9월 발표한 글로벌 AI 지수(The Global AI Index)¹⁾에서 한국은 83개국 중 전년과 동일한 6위를 기록하였다. 2024년 상위 10위에는 1위에서 4위는 전년과 동일한 미국, 중국, 싱가포르, 영국 순이었으며, 프랑스는 5위, 인도가 10위로 새로 진입하고, 스위스와 핀란드는 10위 밖으로 밀려났다. 이 지수는 구현(인재, 인프라, 운영환경), 혁신(연구, 개발), 상용화 생태계, 투자(정부전략, 상용화 생태계) 관련 122개 지표를 평가하여 종합적인 순위를 결정한다.

1) Tortois Intelligence. The Global Artificial Intelligence Index, 2024.09.19

[그림 2-2] 글로벌 AI 지수 주요국의 전체 순위와 항목별 순위

순위	국가	전체	인프라	운영환경	연구	개발	정부전략	산업화	규모	강도
1	United States	1	1	1	1	1	2	1	1	3
2	China	2	2	21	2	2	3	2	2	21
3	Singapore	3	3	45	3	3	10	4	11	1
4	United Kingdom	4	17	41	4	16	7	5	3	9
5	France	10	14	15	5	4	8	6	5	10
6	South Korea	13	11	33	13	3	4	12	2	11
7	Germany	5	13	6	6	15	5	9	5	15
8	Canada	6	16	19	8	10	3	8	6	8
9	Israel	7	24	81	7	5	32	3	14	2
10	India	7	68	1	14	13	11	13	4	36
11	Japan	23	8	55	20	14	12	14	9	25
12	Switzerland	8	11	56	9	19	64	20	29	4
13	The Netherlands	11	7	28	16	11	19	23	13	12
14	Saudi Arabia	81	29	41	41	26	1	7	10	24
15	Finland	14	12	9	18	12	25	15	18	6

출처 : Tortois Intelligence. The Global Artificial Intelligence Index, 2024.09.19.

AI 분야의 발전과 함께 AI의 잠재적 위험에 대한 우려가 전 세계적으로 확산되면서, 이를 적절히 통제하기 위한 거버넌스 논의가 활발히 진행되고 있다. UN은 2020년에 처음으로 연구를 제안하여 자문 조직을 중심으로 연구를 진행하였으며, 2024년 9월에 ‘인류를 위한 AI 거버넌스(Governing AI for Humanity)’ 최종보고서를 발표하였다. 이 보고서에서는 각국의 AI 거버넌스 관련 이니셔티브 추진 현황을 분석하였으며, 다음과 같은 7개의 이니셔티브를 샘플로 제시하였다.

- OECD AI 원칙(2019): 경제협력개발기구(OECD)에서 제정한 AI 원칙으로, AI의 책임감 있는 개발과 활용을 촉진
- G20 AI 원칙(2019): 주요 20개국(G20)이 합의한 AI 원칙으로, AI의 혁신과 윤리적 사용을 강조
- 유럽평의회 AI 협약 초안 작성 그룹(2022-2024): 유럽평의회에서 AI 협약 초안을 작성하기 위해 구성된 그룹으로, AI의 법적 규제를 논의
- GPAI 장관 선언(2022): 글로벌 파트너십 인공지능(GPAI)에서 발표한 선언으로, AI의 국제협력을 촉진
- G7 장관 성명(2023): 주요 7개국(G7) 장관들이 발표한 성명으로, AI의 책임감 있는 개발과 사용을 촉구

- 블레츨리 선언(2023): 영국 블레츨리에서 발표된 선언으로, AI의 안전성과 윤리적 사용을 강조
- 서울 선언(2024): 대한민국 서울에서 발표된 선언으로, AI의 혁신과 규제를 균형 있게 다룸

동 이니셔티브²⁾에 적극적으로 참여하는 국가는 캐나다, 프랑스, 독일, 이탈리아, 일본, 영국, 미국 등 7개국이었다. 반면, 118개국은 해당 이니셔티브에 참여하지 않는 것으로 보고되었다.

[그림 2-3] 국제적인 AI 거버넌스 이니셔티브 관련 추진 현황

Country / party ^a	Tortoise Global AI Index Rank (2023)	Number of initiatives party to	Chronological order →						
			OECD AI Principles (2019)	G20 AI Principles (2019)	CoE drafters (2022)	GPAI Ministerial Declaration (2022)	Bletchley Declaration (2023)	G7 Ministerial Statement on Hiroshima AI Process (2023)	Seoul Ministerial Statement (2024)
United States of America	1	7							
China	2	2							
Singapore	3	4							
United Kingdom of Great Britain and Northern Ireland	4	7							
Canada	5	7							
Republic of Korea	6	5							
Israel	7	5							
Germany	8	7							
Switzerland	9	4							
Finland	10	2							
Netherlands	11	5							
Japan	12	7							
France	13	7							
India	14	4							
Australia	15	6							
Denmark	16	3							
Sweden	17	3							
Luxembourg	18	2							
Ireland	19	4							
Austria	20	2							
Spain	21	5							
Belgium	22	3							
Italy	23	7							
Norway	24	2							
Estonia	25	2							
United Arab Emirates	27	2							
Portugal	28	2							
Russian Federation	29	1							
Saudi Arabia	30	3							

출처 : UN, GOVERNING AI FOR HUMANITY, 2024.09.

2) https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf

시스코(Cisco)는 2024년 11월 글로벌 및 한국 기업들의 AI 준비 지수(AI Readiness Index)³⁾를 공개하였다. 시스코는 전 세계 30개국에서 500명 이상 직원을 보유한 대기업의 AI 담당 임원 7,985명을 대상 설문조사를 토대로 AI 준비 지수를 개발하였다.

시스코가 발표한 2024년 AI 준비 지수에 따르면 AI의 배포와 활용에 완전히 준비된 기업은 13%에 불과했다. 그러나 대부분 기업은 AI 도입이 시급하다고 인식하고 있으며, 85%의 기업은 18개월 이내에, 59%의 기업은 12개월 이내에 AI 도입의 효과를 실현해야 한다고 응답했다. 기업들은 향후 5년간 IT 예산의 약 30%가 AI에 할당될 것으로 예상했으며 이는 현재의 두 배 수준으로, 59%의 기업은 5년 후 AI 투자에서 기대한 수익을 창출할 수 있을 것으로 예상되었다. 전략 ▲인프라 ▲데이터 ▲거버넌스 ▲인재 ▲문화의 6개 측면에서 AI 준비도를 평가한 결과, 인프라 관점에서는 21%의 기업만 현재 및 미래의 AI 수요에 대응한 GPU를 보유하고 있으며, 데이터 관점에서 AI 기술을 배포 및 활용할 완전한 준비가 되었다고 응답한 기업은 32%에 불과했다. 숙련된 인재의 부족 역시 인프라와 데이터, 거버넌스 전반에 걸친 최대 과제로, 31%의 기업만 사내 인력이 AI를 완전히 활용할 준비가 되었다고 응답했다.

시스코가 발표한 한국 기업들의 AI 준비 지수(2024 AI Readiness Index-South Korea)⁴⁾에서는 한국 기업 중에서는 3%만 AI를 배포 및 활용할 완전한 준비가 되었다고 답했으며, 이는 2023년의 5%보다 감소한 수치이자 APJC(아시아태평양·일본·중국) 지역 평균(15%)을 밑도는 결과였다. 국내 기업들은 전략 측면의 AI 준비도가 가장 높은 수준으로, 59%의 기업이 완전한 준비(15%) 및 보통 수준의 준비(44%)가 되었다고 응답했으며, AI 투자의 핵심 영역은 사이버보안과 데이터 관리, 데이터 분석 순으로 확인되었다. 그러나 인프라 측면에서 준비도가 보통 수준 이상인 국내 기업 비율은 23%에 불과했으며, 35%는 AI 수요 증가에 대비한 인프라 확장성과 유연성이 필요하다고 응답했다. 데이터 관점에서 보통 수준 이상의 준비가 되었다고 응답한 국내 기업 비율도 21%에 그쳤으며, 90%의 기업이 AI 프로젝트를 위한 데이터 전처리나 정제에 문제가 있다고 응답했다. 거버넌스 관점에서 보통 이상의 준비도를 기록한 기업 비율은 23%, 인재 관점에서는 보통 이상 준비도를 기록한 기업 비율은 24%로 나타나 전반적인 준비도가 낮은 편이었다.

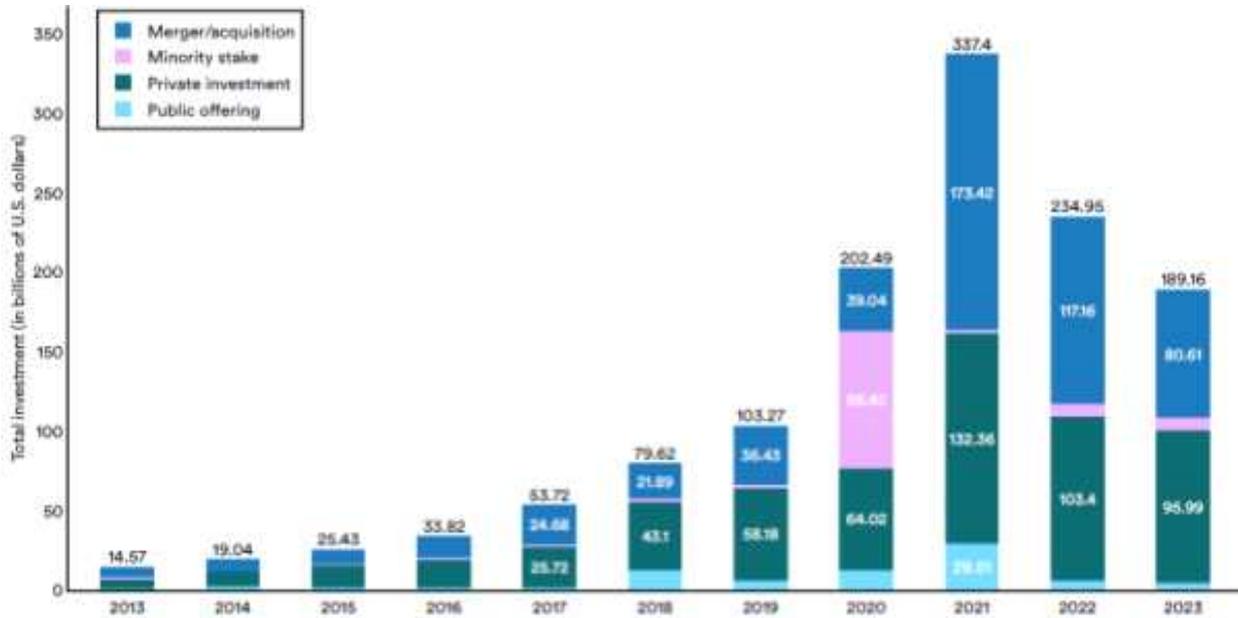
3) Cisco, Cisco's 2024 AI Readiness Index: Urgency Rises, Readiness Falls, 2024.11.19.

4) Cisco, Cisco's 2024 AI Readiness Index-South Korea, 2024.11.19.

2. 전 세계 AI 투자 현황

HAI AI 인덱스 보고서에 따르면, 2013년부터 2023년까지 AI에 대한 전 세계 기업의 전반적인 투자(인수합병, 소수 지분, 민간투자, 공모)는 2021년 이후 연속해서 전 세계 기업의 연간 AI 투자가 감소했다. 2023년 전 세계 기업의 AI 투자 총액은 약 1,892억 달러로 2022년 대비 약 20% 감소했으며, 지난 10년간 AI 관련 투자는 3배 증가한 것으로 나타났다.

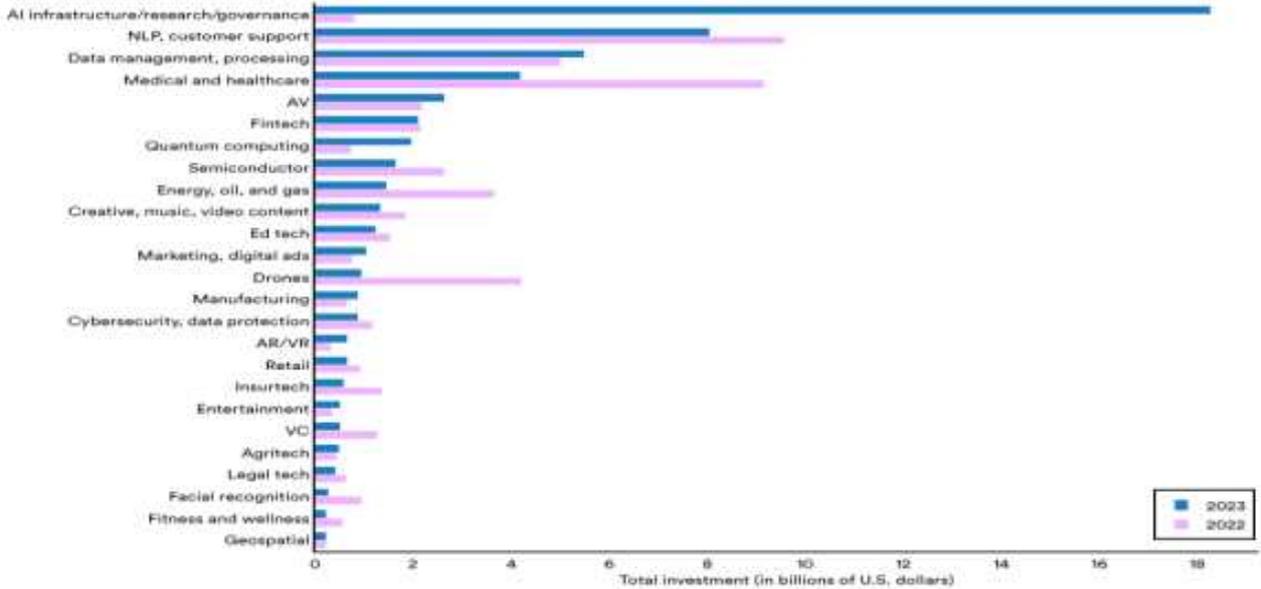
[그림 2-4] 2013-2023 글로벌 AI 투자



출처: Stanford University, Artificial Intelligence Index Report 2024

HAI의 민간 부문의 AI 투자 분야별 추세 분석에 따르면, 2023년에 가장 많은 투자를 유치한 주요 분야는 AI 인프라/연구/거버넌스(183억 달러), NLP 및 고객 지원(81억 달러), 데이터 관리 및 처리(55억 달러)였다. 이는 OpenAI, Anthropic, Inflection AI와 같이 AI 애플리케이션을 구축하는 기업에 대한 대규모 투자를 반영한다. 아래 그림과 같이 대부분의 주요 분야는 2023년에 투자가 감소했으나, 2022년 이후 성장을 보인 분야로는 AI 인프라/연구/거버넌스, 데이터 관리/처리 등이 있다.

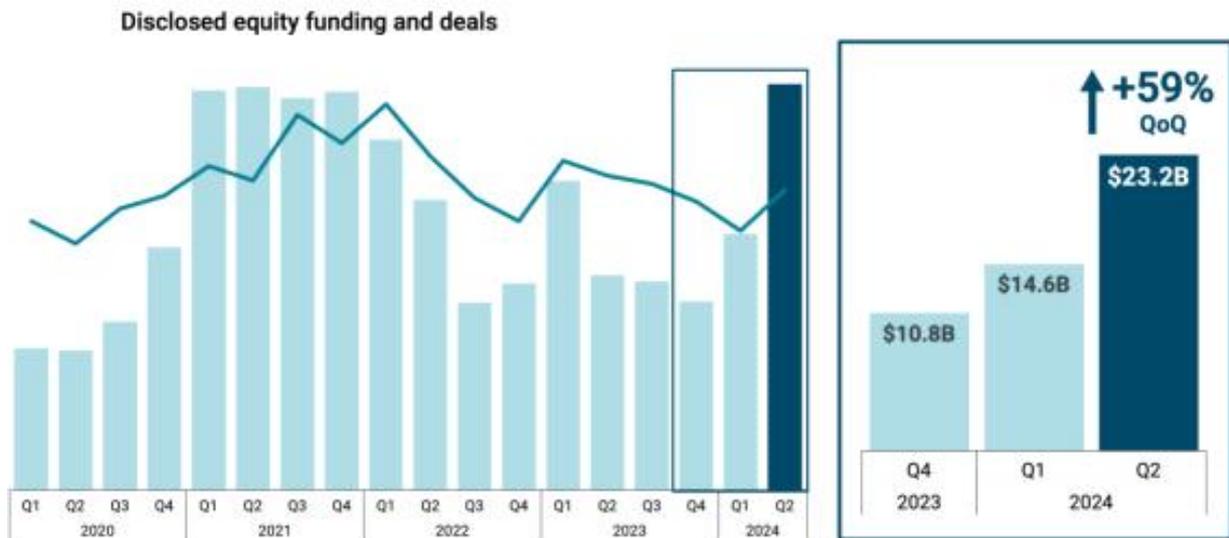
[그림 2-5] 민간 부문의 AI 영역별 투자(2022 vs. 2023)



출처: Stanford University, Artificial Intelligence Index Report 2024

글로벌 시장조사기관인 CB인사이트는 2024년 2분기 전 세계 AI 투자 규모를 발표⁵⁾했다. 2024년 2분기 전 세계 AI 투자 규모는 전 분기 대비 59% 증가한 232억 달러였다.

[그림 2-6] 2020년~2024년 2분기 글로벌 AI 벤처 투자 추이

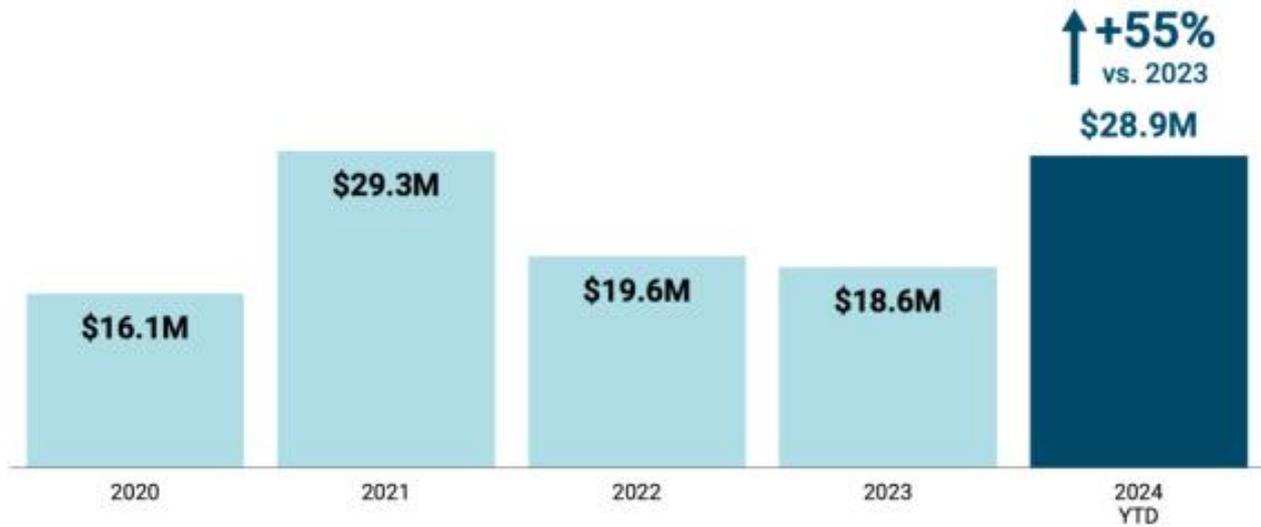


출처 : CB Insights, State of AI Q2' 24 Report, 2024.07.30.

5) <https://www.cbinsights.com/research/report/ai-trends-q2-2024/>

평균 AI 투자 규모도 2,890만 달러로 전 분기 대비 55% 증가하였다. AI 유니콘 탄생은 2024년 2분기에 6개로 안정적으로 유지되었다.

[그림 2-7] AI 기업의 평균 투자 규모



출처 : CB Insights, State of AI Q2' 24 Report, 2024.07.30.

생성 AI는 새로운 유니콘(10억 달러 이상의 가치 평가를 받는 비상장 민간기업)의 주요 주체였다. xAI와 같은 이러한 회사 중 일부는 생성 AI 인프라에 중점을 두었으며, Perplexity (검색) 및 Cognition(코딩)과 같이 생성 AI 애플리케이션 분야에 집중했다. 주요 글로벌 지역 중에서 미국은 AI 자금 조달 및 거래에서 계속 선두를 유지했다. 미국에 기반을 둔 AI 스타트업은 2024년 2분기에 476건의 거래에서 152억 달러를 유치했다. 이는 2분기 글로벌 AI 자금 조달 총액의 66%, 글로벌 거래 총액의 50%에 해당한다.⁶⁾

반면, 글로벌 시장조사기관 포레스터리서치(Forrester Research)의 ‘2025년 예측 (Predictions 2025)’ 보고서에서는 2024년에 과감한 AI 투자에 나섰던 기업들이 2025년에는 실제 수익과 가치 달성에 중점을 두면서 AI 투자를 축소할 것으로 전망했다.⁷⁾ 포레스터가 2024년 2분기에 실시한 설문조사 결과, 미국 생성 AI 의사결정권자의 49%는 소속 기업이

6) <https://www.cbinsights.com/research/report/ai-trends-q2-2024/>

7) Forrester, Predictions 2025: An AI Reality Check Paves The Path For Long-Term Success, 2024.10.22.

Forrester, Predictions 2025: Accelerated Demand For AI-Powered Infrastructure And Operations, 2024.10.22.

1~3년 내 AI에 대한 투자수익률(ROI)을 기대하고 있으며, 44%는 3~5년 이내에 기대한다고 응답했다. ROI 달성에 어려움을 겪는 기업들은 AI 투자를 조기에 축소할 수 있으며, 이는 장기적으로 기업의 성장과 혁신을 저해할 위험도 존재할 것으로 전망된다.

포레스터는 주요 기술 공급업체 역시 공급 부족과 수익 기대치 미달, 투자자 압박에 따라 2025년 AI 인프라 투자를 제한할 것으로 전망했다. 생성 AI의 도입이 늘어나며 AI 칩과 서버 수요가 공급업체의 제공 능력을 압도하고 있는 가운데, 2023년 AI 인프라와 생성 AI에 수십억 달러가 투자되었음에도 2024년 AI로 인해 수익이 증가했다고 보고한 기업은 20%에 불과했다. 아마존, 오라클, IBM 등의 기술 공급업체는 2025년에 AI 인프라 투자를 25%까지 축소할 전망이며, 이러한 추세가 시장 전체로 확산되며 2025년 AI 서비스와 인프라 공급이 부족해질 가능성도 대두된다.

포레스터는 또한 급격한 기술 혁신 및 일관된 표준이나 인증의 부재로 AI 거버넌스의 복잡성이 증가하면서, 엄격한 규제를 적용받는 기업의 40%가 AI와 데이터 거버넌스를 결합할 것으로 예측했다. EU AI 법이 시행되고 AI 투명성 요구가 커지면서, 기업들은 단순한 규정 준수를 넘어 책임 있고 투명하며 윤리적인 AI 접근방식을 중시하게 될 전망이다.

한편, 포레스터는 AI 에이전트의 도입이 늘어나는 가운데, 자체적으로 에이전트 아키텍처를 구축하는 기업 4곳 중 3곳은 성공하기 어려울 것으로 전망했다. AI 에이전트의 복잡성으로 인해 개발 시 고급 데이터 아키텍처 및 다양한 AI 모델과 전문지식이 필요하며, 성숙한 기업들은 자체 개발의 한계를 인식하고 AI 서비스 공급업체 및 시스템 통합업체와 협력해 AI 에이전트를 도입할 것으로 예상된다.

3. AI 에이전트 시장 성장

2024년, 생성 AI의 확산과 함께 인간과 상호작용이 가능한 AI 에이전트에 대한 빅테크 기업들의 참여가 가속화되면서 AI 에이전트 시장이 주목받고 있다. AI 에이전트는 AI 기술의 발전과 함께 점차 많은 관심을 받고 있으며, 특히 클라우드 컴퓨팅과 에지컴퓨팅(의 발전은 AI 에이전트가 실시간 데이터를 처리하고 빠르게 의사결정을 내릴 수 있는 기반을 제공하고 있다.

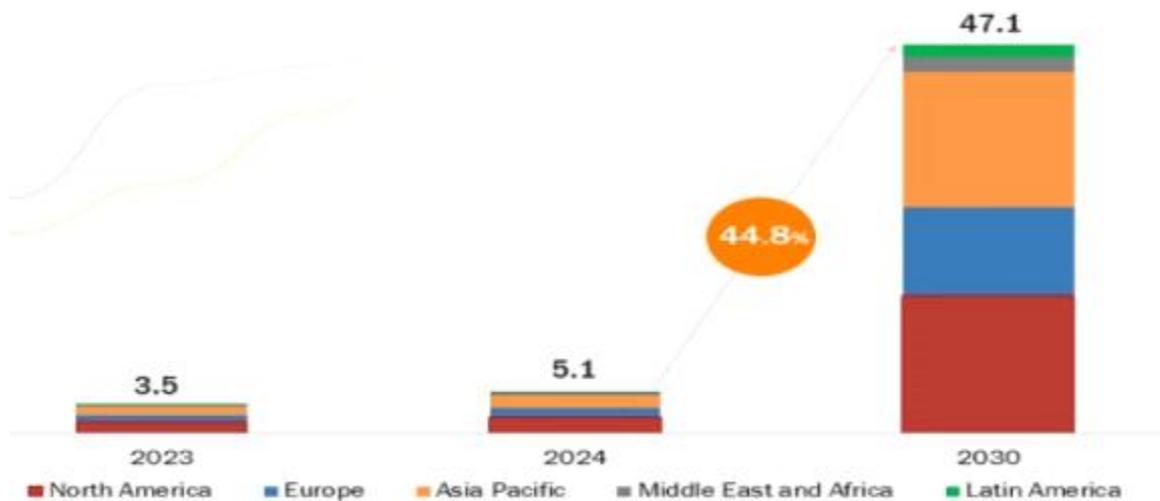
강화 학습(reinforcement learning), 자연어 처리(NLP), 컴퓨터 비전과 같은 분야에서 AI 알고리즘이 발전하면서 AI 에이전트는 더욱 정교한 의사결정을 수행할 수 있게 되었으며, 다양한 상황에서 유연하게 대응할 수 있는 능력을 갖추게 되었다. 이러한 기술적 진보는 고객 서비스, 스마트홈, 산업 자동화와 같은 분야에서 사용자와의 자연스러운 상호작용을 요구하는 환경의 확산과 맞물려 AI 에이전트에 대한 수요를 크게 확대시키고 있다. 또한, AI 에이전트는 새로운 시장 기회와 가능성을 열어가며 기대를 모으고 있다. 고객 서비스, 개인 비서, 자율주행 차량, 스마트팩토리 등 다양한 산업에서 활용 가능한 AI 에이전트는 개인화 서비스, 실시간 대응 및 상호작용과 같은 기존 기술로 구현하기 어려운 혁신적 서비스를 제공함으로써 새로운 부가가치를 창출할 수 있다. AI 에이전트 개발은 AI 기술의 발전뿐만 아니라 데이터 과학, 컴퓨팅 하드웨어, 클라우드 인프라 등 다양한 기술 생태계의 성장과도 밀접히 연관되어 있다. 이는 기술적 혁신뿐만 아니라 경제적 성장의 기회를 창출하며, AI 에이전트 시장의 확대와 발전을 더욱 가속화할 것으로 기대되고 있다.

가트너는 2025년 최상위 10대 전략 기술 트렌드 중 하나로 에이전틱 AI(Agentic AI)를 선정했다. 가트너는 AI 에이전트를 ‘에이젠틱 AI(Agentic AI)’로 정의하며, 이를 목표 중심 소프트웨어 엔터티로 설명한다. 에이젠틱 AI는 AI 기술을 활용해 작업을 완료하고 목표를 달성하는 시스템으로, 명시적인 입력 없이도 지침을 이해하고, 스스로 계획을 세우며, 다양한 도구를 활용해 작업을 수행할 수 있는 것이 특징이다.

또한, 가트너는 에이젠틱 AI의 주요 특성으로 동적 출력 생성 능력을 강조한다. 이는 기존 시스템이 미리 정해진 출력을 생성하는 데 그쳤던 것과 달리, 에이젠틱 AI는 상황에 따라 유연하게 적응하며 새로운 결과물을 동적으로 창출할 수 있음을 의미한다. 이처럼 에이젠틱 AI는 단순한 지시 실행을 넘어선 고도화된 자율성을 갖춘 시스템으로, 다양한 분야에서의 활용 가능성이 주목받고 있다.⁸⁾

시장조사기관인 Markets&Markets는 전 세계 AI 에이전트 시장이 2024년 51억 달러에서 2030년 471억 달러로 빠르게 성장할 것으로 예상하고, 연평균 성장률(CAGR)은 44.8%로 전망되고 있다⁹⁾. AI 에이전트는 더욱 정교한 상호작용과 맥락 인식을 가능하게 하여 고객서비스, 의료, 금융 분야에서의 활용이 증가하고, 다중 에이전트 시스템의 협력이 AI 시장 성장으로 주요 요인으로 분석되고 있다. AI 에이전트는 주로 NLP, 머신러닝, 컴퓨터 비전 등을 포함하며, 완전 자율부터 반자율 형태로 다양한 업무 자동화와 데이터 분석을 지원하는 소프트웨어로 사용될 것으로 전망된다.

[그림 2-8] AI 에이전트 시장 전망(2024-2030)



출처 : Markets&Markets, AI Agents Market by Agent Role, 2024.09.

CB인사이트는 2024년 9월 ‘게임 체인저 2025(Game Changers 2025)¹⁰⁾’ 보고서에서, AI 에이전트 및 코파일럿 시장이 기업들이 생산성 향상을 위한 기회를 포착하기 위해 경쟁하면서 폭발적으로 성장하고 있다고 분석했다. 생성 AI(GenAI)의 발전으로 인해 AI 에이전트 분야에서는 300개 이상의 스타트업이 새롭게 등장했으며, AI 에이전트는 기존 코파일럿을 넘어선 새로운 진화 단계를 보여주고 있다. 이들 LLM(대규모 언어 모델) 기반 봇은 사용자의 최소한의 개입만으로 복잡한 작업을 독립적으로 처리할 수 있는 능력을 갖추고 있다.

8) Gartner, Intelligent Agents in AI Really Can Work Alone. Here's How, 2024.10.01.

9) https://www.marketsandmarkets.com/Market-Reports/ai-agents-market-15761548.html?gad_source=1&gclid=CjwKCAiA9IC6BhA3EiwAsbltOJG4uzoGmxUSg_zloWLHINkrjclXog2RCwT8yUBRvGIPmM1Ximfa1RoCTnAQAvD_BwE

10) CB Insights, Game Changers 2025, 2024.09.

[그림 2-9] 엔터프라이즈 AI 에이전트 및 코파일럿



출처 : CB Insights, Game Changers 2025, 2024.09.

마이크로소프트는 클라우드 플랫폼 애저(Azure)를 중심으로 강력한 AI 에이전트를 개발 및 제공하며, 특히 오픈AI와의 파트너십을 통해 ChatGPT와 기타 AI 모델을 Azure에 통합하여 다양한 상용화 제품을 출시하였다. ‘Power Virtual Agents’ 를 포함한 마이크로소프트의 제품군은 기업용 서비스, 사무 자동화, 개발자 지원 도구 등 다양한 분야에 걸쳐 AI 에이전트와 같은 기능을 지원한다. Fortune 500 기업 중 60%가 Microsoft 365 Copilot을 사용하여 비즈니스 혁신과 조직의 역량을 강화하고 있다.¹¹⁾ 2024년 9월 마이크로소프트는 새로운 단계의 Microsoft 365 Copilot 혁신 서비스들을 발표하면서, Microsoft Copilot Studio로 구축되어 비즈니스를 혁신할 수 있는 Copilot 에이전트를 소개하였다.¹²⁾ 이후 2달 이후 2024년 11월 20일 ‘Microsoft Ignite 2024’ 행사에서는 AI 에이전트 시대를 열겠다고 밝히면서 AI 기반 작업 도우미인 마이크로소프트 365 Copilot의 신규 AI 에이전트와 기능, 자율 에이전트 (autonomous agents) 등을 대거 선보였다.¹³⁾ ‘Microsoft Ignite 2024’ 행사에서 선보인 AI 에이전트로는 금융통찰력 에이전트(Financial Insights Agent), 공장운영 에이전트(Factory Operations Agent), 공장안전 에이전트(Factory Safety Agent), 개인쇼핑 에이전트

11) Microsoft, New autonomous agents scale your team like never before, 2024.10.21.

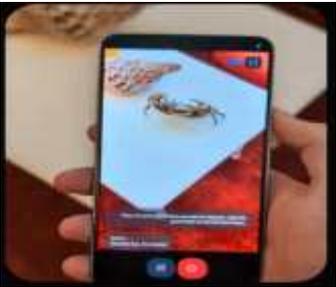
12) Microsoft, Unveiling Copilot agents built with Microsoft Copilot Studio to supercharge your business, 2024.9.16.

13) <https://www.microsoft.com/en-us/industry/blog/general/2024/11/20/unlocking-the-future-ai-and-data-revolutionize-industries-at-microsoft-ignite-2024/>

(Personalized Shopping Agent), 스토어 운영 에이전트(Store Operations Agent), 대민 서비스 에이전트(Citizen Services agent) 등이다.

구글은 Google Assistant, Dialogflow, Duplex 등을 통해 사용자 경험을 극대화하고 있으며, 대화형 AI 생성 도구인 'Vertex AI Agent Builder'를 제공하고, 'Project Astra' 를 통해 보편적인 AI 에이전트를 개발 중이다. 구글은 AI 기술을 활용하여 사용자 경험을 극대화하고 다양한 산업에 적용 가능한 AI 에이전트로 Google Assistant, Dialogflow, Duplex 등을 제공하며, 이러한 AI 에이전트는 자연어 처리와 음성 인식 기술을 바탕으로 더욱 풍부한 사용자 경험을 제공하고 있다. 2024년 9월 구글은 'Gemini at Work' 행사에서 새로운 Gemini 모델(Gemini 1.5 Flash 및 1.5 Pro)과 함께 고객 참여 및 대화를 위해 설계된 AI 에이전트 제품군 출시하였다.¹⁴⁾ 'Gemini at Work' 에서 Google Cloud CEO(Thomas Kurian)는 기존 연결에서 고객 상담 AI 에이전트가 고객과 자연스럽게 대화하고 상호작용하여 올바른 답변을 제공할 것이라고 언급하였다. 2024년 4월, 'Google I/O 2024' 에서 미래의 AI 어시스턴트 개발 프로젝트 'Project Astra' 을 발표하였다. 일상생활을 위한 AI 에이전트로 불리는 프로젝트 아스트라(Project Astra)는 구글 렌즈와 같이 실시간 비디오와 오디오를 이해하고, 추론하고, 반응할 수 있다. 최근 구글은 웹을 검색하는 차기 AI 에이전트인 'Jarvis AI' 를 개발 중이다.¹⁵⁾ 'Jarvis AI' 는 사용자가 Google Chrome에서 웹 브라우징 작업을 자동화하여, AI 에이전트가 쇼핑이나 항공편 예약과 같은 작업을 수행하는 동안 사용자가 더 중요한 일에 집중할 수 있도록 지원할 것으로 기대된다.

<표 2-3> 구글의 Project Astra 활용(예)

물리적 그림 설명	조망 그림 인식	물체 순서 기억
		

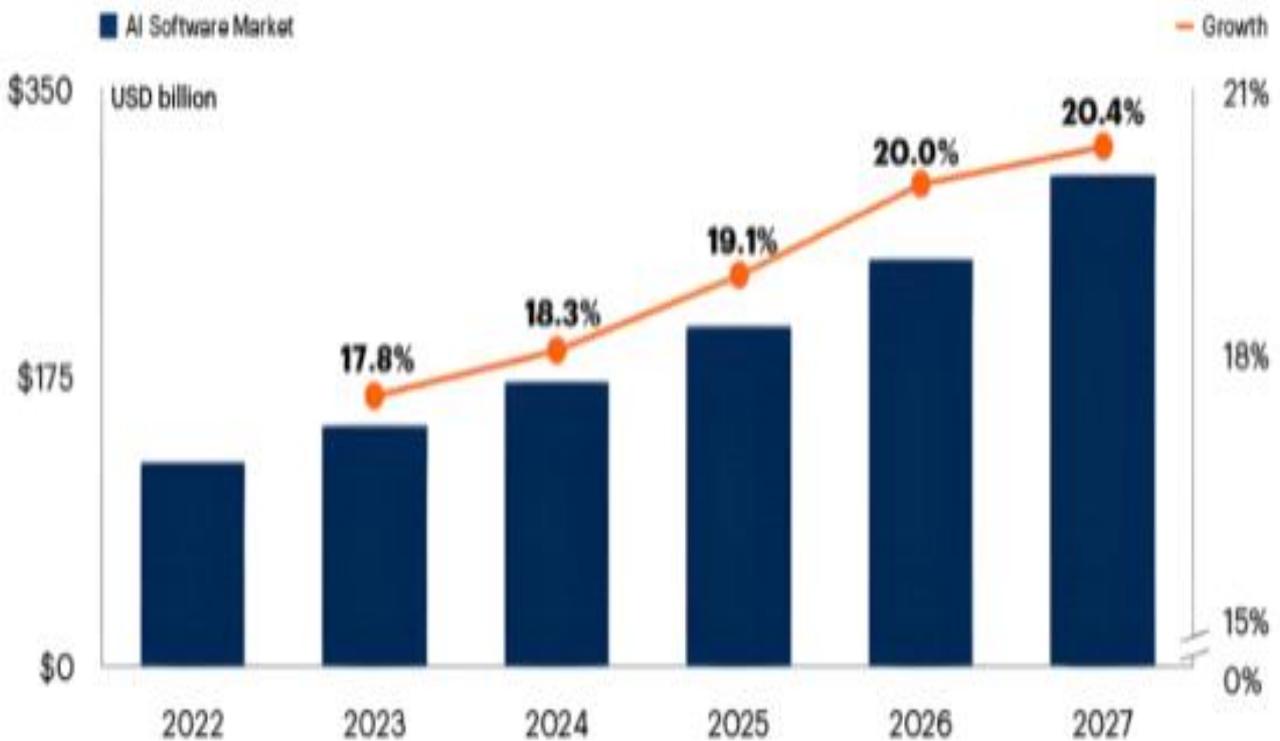
14) Constellation Research, Google Cloud rolls out new Gemini models, AI agents, customer engagement suite, 2024.9.24.

15) TechRadar, Jarvis AI is real - Google accidentally leaks its AI agent that browses the web for you, 2024.11.06.

제2절 전 세계 AI 시장 전망

글로벌 시장조사 기관인 가트너는 2024년 6월, AI 소프트웨어 시장이 2022년의 1,240억 달러에서 2027년까지 2,979억 달러로 성장할 것으로 예측했으며, 이는 연평균 성장률(CAGR) 19.1%에 해당된다. 비록 AI 소프트웨어의 도입이 아직 초기 단계에 머물러 있지만, 이러한 성장률은 AI를 효과적으로 통합할 수 있는 소프트웨어 공급업체들에게 막대한 기회를 시사한다. 소프트웨어 공급업체들은 AI 시스템 구축에 사용되는 AI 플랫폼과 AI 기능이 내장된 애플리케이션에 대한 지속적인 수요를 예상하고 있으며, AI 소프트웨어 시장이 2027년까지는 공급이 수요를 초과하는 포화 상태에 이르지 않을 것으로 전망했다. 특히, 소프트웨어 구매자들이 주목하는 분야는 생성 AI 기능이다. 가트너에 따르면, 2023년에는 1% 미만의 독립 소프트웨어 벤더(ISV)만이 애플리케이션에 생성 AI를 통합했으나, 2026년까지 이 비율이 70% 이상으로 증가할 것으로 예상되었다. 이러한 급격한 상승은 AI가 구매자들에게 필수적인 기능으로 빠르게 자리 잡고 있음을 보여준다.

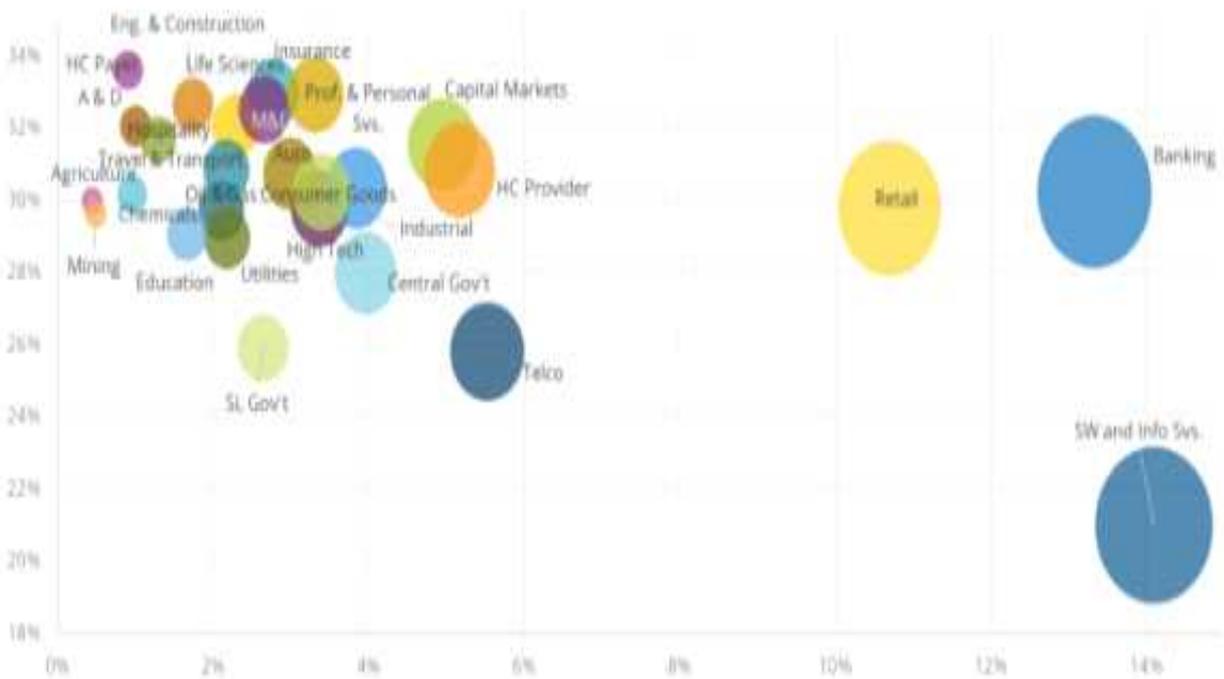
[그림 2-10] AI 소프트웨어 시장 규모 (단위: 10억 달러)



출처 : Gartner, AI Software Forecast and Growth, 2024

또 다른 글로벌 시장조사 기관인 IDC는 2024년 8월, AI 애플리케이션, 인프라, 관련 IT 및 비즈니스 서비스 등을 포함한 전 세계 AI 지출이 2028년까지 6,320억 달러에 도달하며 현재의 두 배 이상으로 증가할 것으로 예상했다. 이는 2024년부터 2028년까지 연평균 성장률(CAGR) 29.0%를 나타낸다. 2024년부터 2028년까지 금융 서비스 산업, 특히 은행 부문이 AI 솔루션에 대한 지출에서 가장 큰 비중을 차지할 것으로 예상되며, 전체 AI 지출의 20% 이상을 차지할 것으로 보인다. 그 뒤를 소프트웨어 및 정보 서비스, 도소매 산업이 잇고 있으며, 이 세 산업이 향후 5년간 전체 AI 지출의 약 45%를 차지할 것으로 전망된다. 특히, 비즈니스 및 개인 서비스 산업은 연평균 성장률(CAGR) 32.8%로 가장 빠른 성장세를 보일 것으로 예상되며, 운송 및 레저 산업도 31.7%의 높은 성장률을 기록할 것으로 보인다. 또한, 분석된 27개 산업 중 17개 산업이 향후 5년간 30% 이상의 연평균 성장률을 달성할 것으로 전망된다.¹⁶⁾

[그림 2-11] 산업별 전 세계 AI 및 생성 AI 지출



출처 : IDC, Worldwide Spending on Artificial Intelligence Forecast to Reach \$632 Billion in 2028, According to a New IDC Spending Guide 2024.08.19.

16) <https://www.idc.com/getdoc.jsp?containerId=prUS52530724>

제3장 주요국 인공지능 정책 동향

제1절 미국

1. 그간의 인공지능 정책 추진 경과

1) 개요

미국은 AI 산업의 선두주자로, 바이든 정부가 출범한 이래 2021년 1월 1일 ‘2020 국가 AI 이니셔티브법(The National AI Initiative Act of 2020)’을 제정하고, 2022년에만 AI 분야에 17억 달러(한화 약 2조 5천억 원)의 예산을 투입하는 등 4차 산업 혁명 시대 AI 패권을 위해 다방면으로 노력했다. 이와 같은 미국의 기존 AI 정책은 대체로 AI의 편향에 의해 발생할 수 있는 잠재적인 차별 위협 등을 완화하고, AI 관련 산업을 진흥하는데 초점을 두고 있었다.

바이든 행정부는 다양한 정부 부처와 기관이 추진하는 인공지능 정책의 기술혁신 노력을 장려하고 미국 시민들에게 정부 정책에 대한 정보를 투명하게 제공하기 위해 2021년 5월 5일 각 부처와 기관이 추진하고 있는 인공지능 정책을 모아 놓은 웹사이트인 ‘AI.gov’를 설치했다.¹⁷⁾ 국가 AI 이니셔티브실(NAIIO)도 2022년 1월 6일 ‘인공지능연구자포털(AI Researchers Portal)’을 구축하여 인공지능 전문가들이 데이터, 컴퓨팅 인프라 프로그램(computing infrastructure programs) 및 지원금 정보 등 민간이 연방정부의 다양한 자원을 활용 할 수 있도록 했다.

이후, 바이든 정부는 2023년 10월 30일 인공지능의 안전하고 보안성 있는 개발과 활용을 위한 행정명령(Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence)을 발표하여, AI의 산업의 진흥에 따라 발생한 다양한 부작용을 해결하고자 했다. 바이든 정부는 이를 통해 AI 안전 및 보안을 위한 새로운 표준, 개인정보보호 강화, 정부의 관리·감독 확대 등 방안을 제시하였다. 또한, 동 행정명령은 인공지능 안전 및 보안, 미국의 주요 가치 보호 등을 목표로 각 부처에 부여된 광범위한 이행사항을 포함하며, 특히 인공 지능 생성 콘텐츠 탐지와 공식 콘텐츠 인증을 위한 모범 실무와 기준을 담고 있다.

17) <https://www.ai.gov>

<표 3-1> 미국의 AI 관련 정책 변화

구분	연도	정책
진흥정책	2015년	• ‘미국을 위한 혁신 전략(A New Strategy for American Innovation)’
진흥정책	2016년	• ‘AI R&D 전략’(The National AI R&D Strategic Plan)
진흥정책	2019년	• ‘AI에서의 주도권 유지를 위한 행정명령(Executive Order on Maintaining American Leadership in Artificial Intelligence)’
진흥정책	2019년	• ‘국가 AI R&D 전략 계획(National Artificial Intelligence Research and Development Strategic Plan: 2019 Update)’
진흥정책	2020년	• ‘2020 국가 AI 이니셔티브 법(National AI Initiative Act of 2020 (DIVISION E, SEC. 5001)’ 제정
진흥정책	2021년	• ‘AI.gov’ 설치
진흥정책	2022년	• ‘국방수권법(NDAA, National Defense Authorization Act)’ 통과
진흥정책	2022년	• ‘AI 연구자포털(AI Researchers Portal)’ 구축
규제정책	2023년	• 인공지능의 안전하고 보안성 있는 개발과 활용을 위한 행정명령(Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence)을 발표
규제정책	2024년	• 대통령 행정명령에 관한 주요 AI 조치 발표
규제정책	2024년	• AI 혁신을 지원하는 주요 리소스에 관한 액세스 민주화를 통해 공유 연구 인프라 비전 실현을 위한 국가 인공지능 연구 리소스(NAIRR: National Artificial Intelligence Research Resource) 파일럿에 착수
규제정책	2024년	• 백악관 AI 행정명령 이후 후속 조치 현황

출처 : www.ai.gov, SPRI 정리, 2024.10.

2) 연도별 미국 인공지능 정책 변화¹⁸⁾

(1) (2023년 이전) AI 진흥 정책을 통한 시장 선점

바이든 행정부는 범정부 차원에서 인공지능 기술 연구를 진척시키고 인공지능 기술의 군사와 산업에의 효과적 적용을 위해 2020년 12월 ‘2020 국가 AI 이니셔티브 법(National AI Initiative Act of 2020(DIVISION E, SEC. 5001)’ 을 제정하는 등 AI 국가전략을 추진하기 위한 제도적 기반을 마련했다.¹⁹⁾ ‘인공지능 이니셔티브 법’ 은 ▲인공지능 연구와 개발에 있어서 세계적 리더로서의 미국의 위상을 지키고, ▲민과 관 모두에서 ‘믿을 만한(trustworthy) 인공지능’ 을 구현하며, ▲국가 경제와 사회 각 분야에서 인공지능 시스템이 통합되게 하고, ▲그러한 노력에 기여할 인력 육성에 힘을 쓸 것을 공언했다.

NAIIA는 AI 연구 개발 분야에서 미국의 지속적인 리더십 보장, 공공 및 민간부문에서 신뢰할 수 있는 AI 시스템 개발, 미국 사회 전반에 걸친 AI 시스템 생태계 구축 및 모든 미국 연방 기관에서 진행하는 AI 정책에 대한 데이터베이스 관리 및 접근성 강화를 주요 목적으로 한다. 세부적으로 ▲AI 연구 개발 지원 사업 ▲AI 교육 및 인력양성프로그램 지원 ▲학계의 AI 연구 네트워크 지원 ▲연방기관 AI 관련 로드맵 기획 및 조정 ▲연방 기금의 재정 지원 계획 및 지원 금액 ▲신뢰할 수 있는 AI 시스템 연구 개발 ▲평가 및 자원에 대한 전략적 동맹과 국제협력기회 제공 등 대한 내용을 포함한다.

더불어 NAIIA는 혁신, 신뢰 할 수 있는 AI 발전, 교육 및 훈련, 인프라, 애플리케이션, 국제협력의 6가지 핵심 전략으로 구성된다. 학계, 산업계, 비영리단체 및 시민사회단체와 협력하여 미국의 모든 부처 및 기관에서 AI 연구, 개발, 시연 및 교육 활동을 강화하고 조정하는 중요한 프레임 워크를 제공한다.²⁰⁾

18) 현 단락부터는 미국의 주요 정책 명칭에 대한 원문은 생략하고 번역본으로 기재함

19) National Artificial Intelligence Research Resource Task Force, Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource. 2023. 1
(<https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf>)

20) National Artificial Intelligence Research Resource Task Force, Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource. 2023. 1

〈표 3-2〉 미국 AI 이니셔티브(American AI Initiative) 5가지 원칙

구분	주요 내용
연구개발	• 연방정부와 산업계, 학계가 공동으로 과학의 새로운 발견과 경제 경쟁력, 국가 안보를 촉진하기 위한 인공지능의 기술 발전을 추진
거버넌스	• 새로운 인공지능 관련 산업을 창출하고 인공지능 기술 안전 테스트와 적절한 기술 표준을 수립
일자리 창출	• 근로자들을 교육해 인공지능 기술을 개발하고 적용할 수 있는 기능을 갖추도록 하고 오늘날의 경제와 미래의 작업에 대비
인프라스트럭처	• 인공지능 기술에 대한 대중의 신뢰와 자신감을 키우고 그 응용에서 국민의 자유와 프라이버시, 가치관을 보호해 미국 국민의 인공지능 기술에 대한 잠재력을 충분히 발휘할 수 있도록 지원
국제협력	• 미국의 AI 연구 및 산업을 위해 동맹국과의 국제 및 업계 협력을 강화하는 한편, 전략적 경쟁 상대와 적대국들로부터 미국의 기술우위와 인공지능 기술 기반을 보호

출처 : National Artificial Intelligence Research Resource Task Force, Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource. 2023. 01.

(2) (2023년 이후) AI 규제 통한 감독 확대

미국 AI 정책은 2023년 이전 ‘2020 국가 AI 이니셔티브 법(National AI Initiative Act of 2020 (DIVISION E, SEC. 5001)’에 기반한 진흥 정책이 주를 이뤘다면 그 이후에는 강력한 규제 정책을 통한 소비자 보호에 초점이 맞춰진 것으로 평가할 수 있다.

바이든 행정부는 2023년 10월 30일 ‘Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence(안전하고 보안성이 높으며 신뢰할 수 있는 AI의 개발 및 사용)’에 관한 행정명령 제14110호(Executive Order 14110, 이하 행정명령 제14110호)를 발표했다.²¹⁾ AI 기술이 급속하게 발전하면서 AI가 주는 편익만큼 부작용에 대한 우려가 커지면서 미국, EU, 중국 등 AI 선도 국가들은 앞다투어 AI 규제에 나서고 있다. 미국의 AI 행정명령은 안전하고 신뢰할 수 있는 AI 개발과 활용을 촉진하기 위해 연방 정부와 기업의 책임을 강화하는 강력한 규제조치를 담고 있다. 이는 위험한 AI 기술이 제도적 통제 없이 개발 및 배포되고 있어 관련 규제가 그만큼 시급하다는 의미이다.

21) Federal Register. Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 2023.11.1. (<https://www.federalregister.gov/documents/2023/11/01/2023-24283/safesecure-and-trustworthy-development-and-use-of-artificial-intelligence>)

행정명령은 AI의 개발과 사용을 안전하고 책임감 있게 관리하는 것을 최우선 과제로 삼고 있음을 밝혔다. 이를 실현하기 위해 연방기관들의 범정부적 차원의 접근 방식을 추진한다고 강조했다며, 행정명령 제14110호의 목적을 다음과 같이 명시하고 있다.)²²⁾

〈표 3-3〉 행정명령 제14110호의 목적

주요 내용
<p>AI는 기회와 위험이 둘 다 될 수 있는 놀라운 잠재력을 가지고 있다. 책임감 있는 AI 사용은 우리의 세계를 더 부유하고, 생산적이고, 혁신적이고, 안전하게 만드는 동시에 긴급한 문제 해결에 도움이 될 잠재력이 있다. 하지만 무책임한 사용은 사기, 차별, 편견, 그리고 허위 정보와 같은 사회적인 해악을 악화시키고, 노동자들을 해고하고, 경쟁을 억압하고, 국가 안보에 위협을 제기할 수 있다.</p> <p>AI를 선한 목적으로 활용하고 AI의 무수한 잠재적 이익을 실현하기 위해서는 AI의 사용에 따른 상당한 위험을 완화해야 한다. 이를 위해 정부, 민간, 학계 그리고 시민 사회를 포함하는 사회 전체의 노력이 필요하다.</p>

출처 : The White House, Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 2023,10.30.

또한 행정명령 제14110호는 미국 연방 50여 개 기관에 100개 이상의 조치 실행을 지시하며 총 8개 정책영역에 걸쳐 구체적인 지침과 조치사항을 명시했다.

- ① 안전 및 보안(행정명령 제4조): 바이오보안, 사이버보안, 국가안보, 중요 핵심 인프라 등 인공 지능 채택과 관련된 위험을 이해하고 이를 완화하기 위해 반복 가능한 절차 및 방법의 개발과 실행을 촉구함. 30개 이상의 연방기관이 추진해야 할 27개의 의무요건이 포함됨
- ② 혁신 및 경쟁(행정명령 5조): AI 분야 인재를 유치하고 지식재산권 문제를 이해하여 보호하고 스타트업 등의 AI의 혁신을 촉진하는 조치임. 10개 이상의 연방기관이 추진해야 할 21개의 의무요건이 포함됨
- ③ 혁신 및 경쟁(행정명령 5조): AI 분야 인재를 유치하고 지식재산권 문제를 이해하여 보호하고 스타트업 등의 AI의 혁신을 촉진하는 조치임. 10개 이상의 연방기관이 추진해야 할 21개의 의무요건이 포함됨
- ④ 근로자 지원(행정명령 6조): AI 도입이 근로자들의 혼란을 유발할 수 있음을 언급하면

22) The White House, Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 2023,10.30
[\(https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/\)](https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/)

서 이러한 문제를 완화할 방안의 연구 및 개발을 제시함. 2개의 연방기관이 추진해야 할 4개의 의무요건이 포함되어 있음

- ⑤ AI의 편향 및 인권에 대한 고려사항(행정명령 7조): AI 모델이 편향을 영구화할 수 있고 인권 침해로 이어질 수 있다고 지적함. 사법체계에서 평등과 인권에 초점을 맞추는 8개 연방기관의 9개 의무요건을 제시함
- ⑥ 개인정보보호(행정명령 9조): AI로 인해 악화될 수 있는 사용자 데이터의 수집, 사용, 보유와 관련된 개인정보보호 위험을 평가하고 이를 완화하도록 촉구함. 9개 연방기관의 6개 의무요건이 포함됨
- ⑦ 연방기관의 AI 사용(행정명령 10조): 관리예산처에 연방기관의 인공지능 사용을 조율하고 기관의 AI 거버넌스 및 위험 관리 활동에 대한 지침을 개발할 부처 간 협의회를 수립할 것을 의무화함. 40개의 연방기관이 추진해야 할 29개의 요건이 포함됨
- ⑧ 국제관계 리더십(행정명령 11조): 7개의 연방기관이 추진해야 할 6개 의무요건은 미국이 동맹국 및 파트너 국가와 교류하고, AI 규제 및 책임에 관한 공동 원칙을 개발하는 노력을 주도하며 글로벌 기술 기준을 발전시켜 전 세계 AI 개발 및 도입을 선도해야 한다고 명시함

<표 3-4> 정책영역별 주요 의무요건 지침

정책영역	의무요건	책임기관
안전 및 보안	안전하고, 보안성이 높고 신뢰할 수 있는 AI의 안전 및 보안을 개발 및 배포하기 위한 가이드라인과 모범관행(합의된 업계 표준을 추진하는 것이 목적임) 및 테스트 환경 개발	국립표준기술연구소 (에너지부, 국토안보부, 상무부, 국립과학재단)
혁신 및 경쟁	AI 교육 또는 일자리를 위해 미국에 오고자 하는 외국인의 비자 발급 절차를 간소화	국무부 및 국토안보부
근로자 지원	AI가 노동시장에 미치는 영향을 조사하여 대통령에게 보고서 제출	경제자문위원회
인공지능 편향 및 시민권에 대한 고려사항	기관 및 규제기관을 소집하여 알고리즘 및 인공지능 관련 차별을 방지하기 위해 어떻게 각자의 권한을 사용할 수 있는지 논의	법무부
소비자 보호	보건복지부 AI 전담반(Task Force) 설립	보건복지부 (국방부, 보훈부)
개인정보보호	기관에서 조달한 상업적으로 이용 가능한 정보 파악	관리예산국
연방기관의 AI사용	연방기관의 AI 사용에 관한 부처 간 협의회를 소집하고 의장역할 수행	관리예산국
국제관계 리더십	AI 기준 수립의 추진 및 개발을 위한 글로벌 추진 계획 수립	상무부

출처 : The White House, Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 2023,10.30

3) 미국 인공지능 핵심 정책 추진 현황

(1) 연방정부 AI 정책

2020년부터 2024년 5월까지 제정된 인공지능 관련 법률은 ▲「생성적 적대 신경망(GAN) 출력물 확인법」 ▲「정부인공지능법 2020」 ▲「국가 인공지능 구상법 2020」 ▲「미국인공지능진흥법」 ▲「미국 항공 분야 성장 및 리더십 확보법」을 비롯하여 연도별 국방수권법과 세출예산법 등이다. 특히 「국가 인공지능 구상법 2020」에 따라 국가인공지능구상실이 백악관 과학기술처 소속으로 신설되었고, 전문가로 구성된 국가인공지능자문위원회가 조직되어 정책 제안을 하고 있다.

행정부에서도 정부의 인공지능 활용을 권장한다. 트럼프 행정부는 2019년과 2020년에 각각 행정명령 제13859호와 행정명령 제13960호를 발하였고, 현 바이든 행정부는 2023년에 「안전하고 신뢰할 수 있는 인공지능 개발 및 사용 행정명령 제14110호」를 발한 바 있다. 현 행정부는 인공지능으로 촉발될 수 있는 위험성에 더 중점을 두고 이에 대응하고자 하는데, 이를 백악관 과학기술처의 『AI 권리장전의 청사진』과 미국표준기술연구소의 『인공지능 위험관리 기본 방침』으로 구체화하였다.

현 제118대 의회에서 발의한 인공지능 법률안은 양원을 모두 합쳐 370건을 상회하지만, 법률로 확정된 것은 연방정부 사업에 대한 두 건에 불과하다. 계류 중인 법률안 중에는 ▲ 「국가인공지능위원회법(안)」(하원 법률안 제4223호) ▲ 「알고리즘책임법(안)」(하원 법률안 제5628호, 상원 법률안 제2892호) ▲ 「디지털플랫폼위원회법(안)」(상원 법률안 제1671호) 등이 있으나 2023년 발의된 뒤 별다른 진전이 있지는 않다. ²³⁾

〈표 3-5〉 연방차원 정책 현황

구분	분야	제목	주요 내용
행정명령	권리보호	연방 정부 내 신뢰할 수 있는 인공지능 기술의 활용을 촉진 하기 위한 행정명령	<ul style="list-style-type: none"> 행정절차 효율화 등을 위해 연방행정 기관이 AI를 이용하는 것을 허용 / 해당AI의 설계·취득·개발 과정에서 미국시민 등을 보호하기 위한 원칙 제시(합법성, 목적·성과중심, 신뢰성·효과성, 이해가능성 등)

23) Senate Dems, Majority Leader Schumer Delivers Remarks To Launch SAFE Innovation Framework For Artificial Intelligence At CSIS. 2023.6.21

(<https://www.democrats.senate.gov/news/press-releases/majority-leader-schumer-delivers-remarks-to-launch-safe-innovation-framework-for-artificial-intelligence-at-csis>)

구분	분야	제목	주요 내용
보고서/지침	권리보호	AI 권리장전 청사진	<ul style="list-style-type: none"> 법적 구속력 없는 권고로서, AI 개발·자동화 과정에서 개인, 기업, 정책입안자가 AI 기술에 대한 책임인식·시민권보호·평등접근을 보장할 것을 촉구하는 핵심 원칙 제시(안전하고 효과적인 시스템, 알고리즘차별로 부터의 보호, 데이터 프라이버시, 공지 및 설명, 사람인 대안 제공 및 피드백 등)
보고서/지침	위험관리	AI 위험 관리 프레임워크	<ul style="list-style-type: none"> AI 시스템의 유해한 편향(bias)의 위험성 관리를 위해 발간되어, 인공지능설계·개발·사용에서 발생할 수 있는 위험을 분석하고, 신뢰할 수 있는 시스템의 특성을 제안하며, 위험을 해결하기 위한 지침을 제공함(고위경영진의 관심, AI 시스템 사용의 장점과 위험파악·평가, 배포 후 모니터링 등)
보고서/지침	권리보호	혁신을 통한 온라인 해악퇴치 보고서	<ul style="list-style-type: none"> 온라인범죄(사기, 딥페이크, 테러, 증오, 아동학대등)를 식별·제거하는데에 AI를 활용할 수 있는지에 대한 보고서로, FTC(미국연방거래위원회)는 위 목적으로 AI를 활용하는 것에 대하여 매우 신중하게 접근 할 것을 입법자들에게 요청함. FTC는 설계 결함 및 부정확성, 개발자·데이터의 편견 반영으로 인한 차별, 개발을 위한 약탈적 데이터 수집, 투명성 부족 등을 그 이유로 들었음.
보고서/지침	권리보호	고용에서의 알고리즘, AI 및 장애인 차별금지에 관한 지침	<ul style="list-style-type: none"> AI 등을 이용하여 채용, 인사고과, 급여 등을 결정하는 경우 장애인에 대한 미국장애인법(Americans with Disabilities Act) 위반의 차별을 초래 할 수 있으므로 이를 방지하기 위해 참고 할 수 있는 고려사항·예시들을 설명
법률	AI사용증진	AI 훈련법	<ul style="list-style-type: none"> 미국 연방공무원의 AI의 이해를 높이고, AI의 윤리적이고 안전한 사용을 목표로 하는 법률로서, 일부행정기관에 대하여 행정기관의 인력을 위한 AI 교육프로그램을 수립하거나 제공 하도록 요구함
법률	군사	2023년 국방수권법	<ul style="list-style-type: none"> 국방 및 정보기관이 AI 시스템과 기능을 기관 운영에 포함 하도록 통합함
법안	위험관리	알고리즘 책임법(안)	<ul style="list-style-type: none"> AI, 머신러닝, 알고리즘등을 사용하는 소프트웨어의 자동화된 의사결정 시스템에 대하여 편향(bias) 평가 및 FTC에 평가 문서 제출을 강제함
법안	권리보호	미국 데이터 개인정보 및 보호법(안)	<ul style="list-style-type: none"> 개인 데이터와 관련 하여, 동의 없는 데이터처리 및 전송에 대한 제한 및 보호를 제공하는 등 개인정보처리 자체에 대한 보호와, 개인정보를 사용하여 개인의 특성을 통한 차별을 금지하는 등 광범위한 규제 조항을 담고 있음
법안	권리보호	디지털플랫폼 위원회 법(안)	<ul style="list-style-type: none"> 법안이 통과되면 신설될 디지털플랫폼위원회는 온라인 플랫폼에 대하여 소비자 보호 등을 위해 청문회 개최·조사 수행·연구 지원 등의 권한을 부여 받으며, 광범위한 감독 권한을 가지게 될 것으로 보임

출처: NIST, Artificial Intelligence Risk Management Framework, AI RMF 1.0. 2023. 5Update, 2023.05.

(2) 미국 주(State)별 AI 정책

주(州)에서도 자체적으로 인공지능 법률을 제정하고 있다. 전국주의회협의회에 따르면 2024년 5월까지 17개 주에서 인공지능 법률을 제정하였고, 이와 별개로 딥페이크 영상물을 규제하는 법률을 둔 주 또한 17개이다. 2024 회기 중 인공지능 또는 딥페이크 규제 법률안이 발의되어 계류 중이거나 주지사의 서명 절차를 앞둔 주는 총 29개에 달해, 앞으로 더 많은 주에서 인공지능 법제를 도입할 것으로 보인다.²⁴⁾

미국의 연방 AI 정책이 정책의 방향성과 진흥책·잠재적 위험의 완화책을 제시하고 있으며, 각 주법은 대체로 AI 사용에 따라 발생할 수 있는 다양한 위협에 대한 규제에 초점을 두고 있는 것으로 보인다. 그러나 주법도 일정한 경우 연방법의 입법과정에 반영되기도 하는 만큼, 이에 대한 동향을 주의 깊게 살펴볼 필요가 있다. 2023년 기준으로 AI 일반에 대해 규율하는 주법은, 얼굴 인식과 같은 특정 분야의 AI나 자율주행을 제외하고도 58개 이상으로 파악된다.

〈표 3-6〉 미국 주별 정책 현황

구분	명칭	주요 내용
일리노이주	AI 비디오 면접법	<ul style="list-style-type: none"> 고용 의사 결정 목적으로 화상 면접에서 AI를 사용하는 것에 대하여, 면접 전 AI 사용 고지의무 및 지원자의 동의를 받을 의무 등 부과
컬럼비아 특별구	알고리즘에 의한 차별금지법(안)	<ul style="list-style-type: none"> 고용, 주택 등에 대한 자동결정에서 인종, 성별, 장애 등 특성의 차별적 선택의 금지 의무 및 알고리즘이 소비자에게 부정적인 영향을 미치는 경우 등을 공개할 의무 등을 광범위한 자동화된 결정절차에 부과
뉴욕시	자동화된 고용 결정 도구 법(조례)	<ul style="list-style-type: none"> 자동화된 의사결정 시스템을 고용, 승진 판단 등에 사용하는 행위를 규제
뉴저지주	고용차별 최소화를 위해 고용 결정시의 자동화 도구 사용을 규제하는 법안	<ul style="list-style-type: none"> 고용주가 편향 방지를 위한 기술적 조치를 취하지 않는 한 자동화된 고용 결정 시스템의 사용을 금지
캘리포니아주	직장기술책임법	<ul style="list-style-type: none"> 전자 모니터링 및 자동화된 의사결정 시스템의 사용을 특정 시공간으로 제한함

출처: NCSL, Artificial Intelligence 2024 Legislation, 2024.

24) NCSL, Artificial Intelligence 2024 Legislation, 2024.9
<https://www.ncsl.org/technology-and-communication/artificial-intelligence-2024-legislation>

2. 2024년 인공지능 정책 동향

1) 美 바이든-해리스 행정부, 대통령 행정명령에 관한 주요 AI 조치 발표

(1) 개요

바이든 대통령은 2023년 10월 30일 인공지능(AI: Artificial Intelligence)의 잠재성을 포착하고, 위험성 관리를 선도하기 위해 혁신적 행정명령 14110(EO: Executive Order)을 발표했다. 행정명령은 AI 안전 및 보안 강화, 미국인 개인정보 보호, 형평성과 시민권리 증진, 소비자 및 근로자 옹호, 혁신과 경쟁 촉진, 미국의 세계적 리더십 발전 등을 위한 전면적 조치를 지시했다.

2024년 1월 29일 미국 백악관의 브루스 리드(Bruce Reed) 부비서실장은 여러 연방 부처 및 기관의 최고 관리자로 구성된 백악관 AI 위원회를 소집했으며, 각 기관은 행정명령이 부여한 90일 동안의 모든 조치를 완료했고, 더 긴 기간이 부여된 다른 주요 지침 진행에 관해 보고했다. 종합적으로 AI 시스템의 잠재적 위험으로부터 미국인을 보호하고, AI와 그 이상의 혁신 촉진을 위한 AI 행정명령 임무 달성에 관한 상당한 진전이 있었다.

(2) 안전 및 보안에 관한 위험성 관리

행정명령은 AI 안전 및 보안에 관련된 가장 큰 위협 해결을 위해 90일내 다양한 조치를 지시했고, 강력한 시스템 개발자와 관련된 주요 공개요건 설정, 주요 인프라에 관한 AI의 위험성 평가, 유해한 목적의 AI 개발과 관련된 외국 행위자 저지 관련 내용을 포함하고 있으며, 언급된 위험성 완화를 위해 각 연방 기관은 아래 나열된 조치를 수행했다.

- ① 국방물자생산법(Defense Production Act)을 통해 가장 강력한 AI 시스템 개발자에게 주요 정보 및 AI 안전성 테스트 결과를 상무부에 보고하도록 강제하고, 기업은 가장 강력한 AI 시스템에 관한 정보 공유와 이러한 시스템을 학습시킬 대규모 컴퓨팅 클러스터도 보고
- ② 외국 AI 학습을 위해 컴퓨팅 성능을 제공하는 미국 클라우드 기업에도 이를 보고하도록 강제하는 규칙 초안을 제시하고, 상무부의 제안이 원안대로 확정될 경우 클라우드 제공업체는 외국 고객이 악의적 활동에 사용 가능한 가장 강력한 모델을 교육할 경우 이를 정부에 알리도록 의무화

- ③ 모든 중요 인프라 부문에서 AI 사용에 관한 위험성 평가를 완료했으며, 국방부·교통부·재무부·보건복지부 등 9개 기관이 국토안보부에 위험 평가서를 제출했고, 이러한 평가는 지속적 연방 정부 조치의 기반이 될 것이며, 미국은 전력망과 같은 사회의 주요시설에 AI를 안전하게 통합하도록 보장

(3) 선한 AI 혁신

바이든 대통령은 AI의 뛰어난 잠재성을 포착하고, AI 혁신에서 미국의 리더십을 강화하기 위해 이번 행정명령을 발표하였다. 이를 통해 AI 혁신에 대한 투자 확대와 함께 AI 전문성을 갖춘 인력의 유치 및 교육하기 위한 신규 활동을 지시했고, 지난 90일 동안 각 기관은 아래 내용을 수행했다.

- ① 광범위한 혁신과 경쟁을 촉진하고 AI 연구에 관한 공정한 접근을 지원하는 국가 AI 연구 리소스의 파일럿(NARR: National AI Research Resource)을 착수하였으며, 국립과학재단(NSF: National Science Foundation)이 관리하는 이 파일럿은 컴퓨팅 성능, 데이터, 소프트웨어, 개방형 및 독점 AI 모델에 대한 액세스, 기타 AI 교육 리소스를 연구자와 학생에게 제공하기 위한 국가 인프라의 첫 단계이며, 이러한 리소스는 11개 연방 기관 파트너, 25개 이상 민간 섹터, 비영리 및 자선 단체 파트너로부터 지원될 예정
- ② 연방 정부 전반에서 AI 전문가 채용을 가속화하기 위해 데이터 과학자를 포함한 대규모 AI 인재 채용 계획을 발표했고, 바이든 대통령의 행정명령을 통해 신설된 AI 및 기술 인재 태스크포스는 이러한 채용 활동을 주도하며, AI 인재 채용 촉진을 위한 기타 주요 이니셔티브를 조율
 - 미국 연방인사관리처(OPM: Office of Personnel Management)는 직접 고용 기관 및 예외 서비스 기관을 포함하여 연방 기관이 AI 인재를 고용할 수 있도록 유연한 채용 권한을 부여했다.
 - 대통령 혁신 펠로우(Presidential Innovation Fellows), 미국 디지털 봉사단(U.S. Digital Corps), 미국 디지털 서비스(U.S. Digital Service) 등 정부 차원의 기술인재 프로그램은 2024년에 우선순위가 높은 AI 프로젝트부터 AI 인재 채용을 확장했음
- ③ 초중고(K-12)부터 학부 레벨까지 우수하고 포용적인 AI 교육 기회를 창출하는 교육자에게 자금지원을 하기 위해 Educate AI 이니셔티브를 착수했고, 이 이니셔티브의 출범은 미래의 AI혁신을 발전시키고, NSF가 모든 미국인이 AI가 창출할 기회로부터 혜택을

누릴 수 있도록 필수적 AI 관련 인력개발에 우선순위를 통해 행정명령 의무 이행 지원

- ④ AI 발전에 중점을 둔 신규 지역 혁신 엔진에 대한 자금지원을 발표했으며, 특히, 피에몬테 트라이어드 재생 의학 엔진에는 2년간 1,500만 달러의 투자와 향후 10년간 최대 1억 6,000만 달러가 투입될 예정으로 세계 최대 규모의 재생 의학 클러스터를 기반으로 AI를 비롯한 획기적 임상 치료법을 개발 및 확장 예정이며, 이번 발표는 NSF가 행정명령의 지침에 따라 150일 이내에 AI 중심 NSF 엔진에 자금지원 이행 및 출범 계획 포함
- ⑤ 규제 명확성 제공 및 의료 분야의 AI 혁신촉진 정책개발을 위해 보건복지부(HHS: Department of Health and Human Services)에 AI 태스크포스를 설립하고, 태스크포스는 신약 개발을 발전시키고 공중 보건을 강화하여 의료 서비스 제공을 개선하기 위해 AI 활용 도구 및 프레임워크 평가 방법을 개발할 예정이며, 태스크포스에서는 의료 알고리즘의 인종 편견 해결을 위한 지침 조율

<표 3-7> 美 AI 행정명령에 따른 각 기관 수행 활동

활동명	기관명	소요 기간
기술 현대화 펀드를 통해 기관 AI 도입 우선순위 선정 방식 평가	Technology Modernization Board	30일
비전통 및 신흥 교통 기술 위원회에 교통 섹터의 AI 가이드스 및 기술 지원 필요성 평가를 진행하도록 지시	Department of Transportation	30일
국가 AI 연구 리소스(NAIRR) 파일럿에 통합 가능한 연방 기관 리소스 보고	Agencies identified by the National Science Foundation	45일
연방 기관 AI 인재를 늘리고, 채용경로 가속화를 위한 우선 분야 파악	Office of Science and Technology Policy & Office of Management and Budget	45일
AI 및 기술 인재 태스크포스 소집	White House Chief of Staff's Office	45일
데이터 과학자를 위한 대규모 채용 조치 등 연방 정부의 인공지능 전문가 채용 가속화를 위한 AI Talent Surge 착수	Agencies coordinating with the AI and Tech Talent Task Force	45일
영구적 노동 인증이 필요하지 않은 스케줄A 직업 분류 목록의 개정 여부에 관한 정보 요청(RFI) 발행	Department of Labor	45일
연방 기관의 AI 사용 조율을 위해 기관 간 협의회 소집	Office Management and Budget	60일
연방 기관이 AI 전문가 고용이 가능하도록 직접 고용 및 예외적 서비스 권한을 포함한 유연한 채용 권한의 필요성 검토 및 승인	Office of Personnel Management	60일
국방물자생산법 권한을 통해 강력한 AI 시스템 개발자에게 주요 정보 및 AI 안전성 테스트 결과 보고를 하도록 강제	Department of Commerce	90일

활동명	기관명	소요 기간
외국 AI 교육을 위한 컴퓨팅 성능을 제공하는 미국 클라우드 기업이 이를 보고하도록 규칙 초안 제시	Department of Commerce	90일
모든 주요 인프라 부문에서 AI 사용을 포괄하는 위험 평가 완료	Sector Risk Management Agencies	90일
NAIRR 파일럿 프로그램 착수	National Science Foundation	90일
인터뷰 면제 권한을 갱신 및 확대하는 등 비자 처리 간소화 추진	Department of State	90일
규제의 명확성 및 의료 분야 AI 혁신 촉진을 위한 정책개발 AI 태스크포스를 설립	Department of Health and Human Services	90일
AI와 민권 교차점 논의를 위해 연방 기관의 민권 사무소 소집	Department of Justice	90일
주요 연방 자문위원회에 AI 및 교통 관련 자문요청	Department of Transportation	90일
특정 지원자가 한 지원서로 다양한 기관의 직책 역할을 진행할 수 있도록 연방 AI 채용 가속화를 위한 공동 채용 조치 착수	Office of Personnel Management	90일
연방 정부 조달제품 및 서비스에 관한 보안 승인에서 생성 AI 기술의 우선순위를 정하는 프레임워크 초안 발표	General Services Administration	90일
AI 발전에 중점을 둔 신규 지역 혁신 엔진에 관한 자금지원 발표	National Science Foundation	150일
AI, 기타 기술, 데이터 기능의 발전으로 더 악화될 가능성이 있는 개인정보보호 위험성을 포함하여 개인정보보호 위험 완화를 위해 연방 기관의 증명 영향평가가 더 효과적인 방식에 관한 RFI 발표	Office of Management and Budget	180일
기관 전반에서 AI 및 기타 주요 신기술 개발 조율을 위한 사무실 설립	Department of Energy	-
기관의 AI 활용을 위한 거버넌스, 혁신 및 위험 관리 발전 정책 초안에 관한 의견 수렴	Office of Management and Budget	-
AI 관련 인력 개발의 우선순위 선정을 위해 EducateAI 이니셔티브 착수	National Science Foundation	-
2024 성장 액셀러레이터 펀드 공모를 통해 AI를 펀드 중점 분야로 명시	Small Business Administration	-
중소기업에 혜택을 주는 주요 프로그램을 통해 AI 관련 지출에 관한 지원 적격성 확인	Small Business Administration	-
AI가 글로벌 발전에 미치는 영향에 대한 RFI 발표	U.S. Agency for International Development & Department of State	-
표적 광고 제한을 포함하여 기업의 아동 데이터 수익화를 더욱 제한하는 개인정보 보호 규정 변경사항 제시	Federal Trade Commission	-
테넌트 스크리닝을 포함한 신원 조회 보고서에 허위, 불완전, 오래된 정보 표시가 되어서는 안 된다는 점을 강조하는 권고의견 발표	Consumer Financial Protection Bureau	-

출처: whitehouse, Fact Sheet: Biden-Harris Administration Announces Key AI Actions Following President Biden's Landmark Executive Order. 2024.1.19.

2) 백악관 AI 행정명령 이후 후속 조치 현황

(1) 개요

2023년 10월 30일, 바이든 대통령은 ‘안전하고 신뢰할 수 있는 AI 개발 및 사용에 관한 행정명령(E.O. 14110)’ 에 서명하였다. 이후 270일이 지난 2024년 7월 26일 백악관은 이 행정명령의 모든 조항이 완료되었음을 발표하며, AI의 안전성 및 신뢰성 확보를 위한 주요 조치들이 성공적으로 수행되었음을 강조하였다

- 백악관은 보도 자료를 통해 미국 15개의 선도적 AI 기업의 자발적 약속으로 시작한 행정명령을 토대로 270일 조치를 모두 일정대로 완료했다고 보도
- AI의 안전 및 보안 위협을 완화하고, 미국인의 사생활을 보호하고, 형평성과 시민권을 증진하고, 소비자와 근로자를 옹호하고, 혁신과 경쟁을 촉진하고, 전 세계적으로 미국의 리더십을 발전시키는 등 관련 조치의 주요 내용을 공유

(2) 행정명령 이후 후속 조치 및 신규 지침 주요 내용

백악관은 AI 안전 및 보안에 대한 위협 관리 관련 광범위한 조치 내용과 AI 테스트 및 평가 역량 구축을 위한 ‘정부의 AI 인재 영입’, ‘책임 있는 AI 혁신 진전’ 등 연방 정부의 그간 성과 및 노력을 강조했다. 다음은 AI 안전 및 보안에 대한 위협 관리와 관련한 주요 조치사항이다.

- NIST 산하 AI 안전연구소(AISI)의 신규 기술 가이드라인 초안을 발표하고 AI 개발자들이 AI의 오용 평가 관리를 위한 공개 의견 수렴 진행
- 생성 AI의 위협 관리 및 시스템 보안 개발에 대한 최종 프레임워크를 발표하고 NIST의 AI 위협 프레임워크와 관련한 추가 지침을 공개
- 에너지부 AI 테스트베드 및 모델평가도구를 개발하고 확장했으며 AI 모델 안정성과 보안 평가 실시, 특히 인프라, 에너지 보안 위협에 대해 평가 수행 정부 차원의 AI 인재 육성 성과 및 공약 발표
- 국가안보 등 연방 정부 전반의 AI 역량을 강화하기 위해 대통령 혁신 펠로우AI 코호트(Presidential Innovation Fellows AI cohort)와 DHS AI 단체(DHS AI Corps)를 포함해 지금까지 200명 이상 채용 진행

- 6개월 간의 AI 인재 채용현황 보고서를 바탕으로 기술 생태계 전반에 걸쳐 광범위한 공익 기술 생태계 강화 및 인프라 구축을 지원하는 1억 달러자금 지원(국립과학재단 4800만 달러, 포드 재단 2000만 달러 등) 공약 발표 책임 있는 AI 혁신 진전을 위한 추가 조치와 정책제언 발표
- 상무부 주도로 이중용도기초모델(dual-use foundation models)의 이점 및 잠재적 위험 관련 보고서를 정책제언과 함께 수주 내 발표 예정
- AI 국립연구자원(NAIRR)의 시범 운영을 통해 80개 이상의 연구팀을 운영, 딥페이크 탐지, AI 안전성 향상, 차세대 의료 진단 등 AI 우선순위 탐색
- 교육부 주도로 교육 기술 개발자가 학생과 교사에게 이로운 동시에 형평성, 시민권, 신뢰, 투명성을 증진하는 AI 설계 방법을 연구, AI 도구를 설계하기위한 신규 가이드 및 권장 사항 발표
- 국가과학기술위원회를 통해 지난 4년간 ‘신뢰할 수 있는 AI’의 발전을 목적으로 연방 차원의 연구개발 R&D 보고서 발표
- NIST 주도로 글로벌 AI 표준에 대한 미국 참여를 위한 포괄적 계획 발표, AI표준 작업의 목표와 우선 분야를 식별하고 이해관계자 대상 조치 제안
- AI 안전 연구소와 기타 정부 지원 과학 사무소의 글로벌 네트워크를 출범, 유엔 총회 결의안을 출범시키고 미국 주도의 선언에 대한 세계적 지원 확대

3) NSF의 미래 AI R&D 민주화를 위한 국가 인공지능 연구 리소스 파일럿 착수²⁵⁾

(1) 개요

2024년 1월, 미국 국립과학재단(NSF: National Science Foundation)과 다양한 협력 기관은 AI 혁신을 지원하는 주요 리소스에 관한 액세스 민주화를 통해 공유 연구 인프라 비전 실현을 위한 국가 인공지능 연구 리소스(NAIRR: National Artificial Intelligence Research Resource) 파일럿에 착수했다.

NSF의 NAIRR 파일럿은 10개의 연방 기관과 25개의 민간 부문, 비영리 및 자선 단체와 협력하여 미국에 기반한 연구자와 교육자에게 첨단 컴퓨팅, 데이터세트, 모델, 소프트웨어, 교육 및 사용자 지원에 관련된 액세스를 제공할 것이며, 연구자·교육자·연구지원리소스를 연결함으로써 혁신적 AI 연구를 강화할 것이다.

〈표 3-8〉 NAIRR 파트너 기관

기관명	기관명
<ul style="list-style-type: none"> • U.S. National Science Foundation(NSF) • Defense Advanced Research Projects Agency (DARPA) • National Aeronautics and Space Administration(NASA) • National Institutes of Health(NIH) • National Institute of Standards and Technology (NIST) 	<ul style="list-style-type: none"> • National Oceanic and Atmospheric Administration(NOAA) • U.S. Department of Agriculture(USDA) • U.S. Department of Defense(DOD) • U.S. Department of Energy(DOE) • U.S. Department of Veterans Affairs(VA) • U.S. Patent and Trademark Office(USPTO)

출처: NSF, Democratizing the future of AI R&D: NSF to launch National AI Research Resource pilot. 2024.01.24.

25) NSF, Democratizing the future of AI R&D: NSF to launch National AI Research Resource pilot. 2024.01.24 (<https://new.nsf.gov/news/democratizing-future-ai-rd-nsf-launch-national-ai>)

(2) 주요 내용

지속적 성장에 맞춰 전체적인 NAIRR 생태계 설계 정보를 제공할 것이며, NAIRR 비전달성을 위한 투자를 통해 컨셉 증명의 과정이 될 것으로 예측된다. 국립과학재단 총재 세투라만 판하나탄(Sethuraman Panchanathan)박사는 “이번 파일럿을 위해 모인 다양한 파트너는 미래 미국 AI를 위한 국가 AI 연구 리소스 개발의 시급성에 대해 강조했다, AI 연구 투자를 통해 새로운 발견 및 영향력을 극대화하며, 글로벌경쟁력을 강화할 것이라고 언급하면서 AI 연구 및 개발 선도를 위해 AI 혁신발전과 교육기회를 강화하여 국가가 국제 표준 구축 및 경제성장 촉발 기회를 창출해야 한다.” 라고 이야기했다.

또한 미국 백악관 과학기술정책실(OSTP: Office of Science and Technology Policy)디렉터인 아라티 프라바카르(Arati Prabhakar) 박사는 “이번 발표는 AI를 발전시켜 미국인이 이 강력한 기술 혜택을 수혜하도록 노력하는 바이든 대통령의 목표를 진전시키고, 국가 AI 연구 리소스 파일럿은 연구원에게 주요 데이터 및 컴퓨팅 액세스를 제공하여 미국의 목표 달성을 가속화 할 것이다.” 라고 언급했다.

파일럿을 통한 학계, 산업계, 비영리 및 정부 분야의 협력은 각 섹터 간 파트너십을 촉진할 것이며, 산업계 협력은 상업용 실행 가능한 AI 애플리케이션 및 솔루션 개발로 연결될 수 있고, 신규 마켓과 수익원 창출을 통해 경제성장 촉진이 가능하다. 파일럿 운영은 아래 3가지 중점 영역으로 구성된다.

- (NAIRR Open) NAIRR 파일럿 포털과 조율된 할당으로 AI 리소스 액세스를 통해 개방형 AI 연구 활성화
- (NAIRR Secure) NIH와 DOE가 공동 주도하며, 개인정보보호 및 보안 유지 리소스가 필요한 AI 연구를 활성화하고, 모범적 개인정보보호 리소스 취합
- (NAIRR Classroom) 교육, 트레이닝, 사용자 지원 및 봉사활동을 통해 새로운 커뮤니티 구축

〈표 3-9〉 NAIRR 민간 부문, 비영리 단체 및 자선 단체 파트너

기관명	기관명
<ul style="list-style-type: none"> • AI2: Allen Institute for Artificial Intelligence • Amazon Web Services (AWS) • Anthropic • AMD • Cerebras • Databricks • Datavant • EleutherAI • Google • Hewlett Packard Enterprise (HPE) • Hugging Face • IBM 	<ul style="list-style-type: none"> • Intel • Meta • Microsoft • MLCommons • NVIDIA • Omidyar Network • OpenAI • OpenMined • Palantir • Regenstrief Institute • SambaNova Systems • Vocareum • Weights & Biases

출처: NSF, Democratizing the future of AI R&D: NSF to launch National AI Research Resource pilot. 2024.01.24.

4) 캘리포니아 주지사의 AI 규제법안 「SB1047」에 거부권 행사²⁶⁾

(1) 개요

미국 캘리포니아주 하원에서 2024년 8월 28일 AI 규제법 「SB1047」²⁷⁾이 통과되었으나 개빈 뉴섬(Gavin Newsom) 캘리포니아 주지사의 거부권 행사로 입법이 무산됐다.

법안은 캘리포니아에서 운영되는 AI 기업의 기반모델 개발 시 다양한 안전 조치를 요구하여, AI 기업은 모델 위험에 대한 안전 평가를 하고 예상치 못한 위험 발생 시 모델 전체를 신속히 중지할 수 있는 기능을 구현해야 하며, 훈련 후 모델 변조를 방지하는 조치를 취할 필요가 있다는 내용을 담고 있다.

법안은 또한 사망이나 재산 피해 등 AI로 인한 심각한 피해 발생 시 캘리포니아주 법무부장관에게 AI 개발기업을 고소할 수 있는 권한을 부여하고 있다.

26) The New York Times, California Legislature Approves Bill Proposing Sweeping A.I. Restrictions, 2024.08.28.
SiliconAngle, California Gov. Gavin Newsom shoots down divisive AI safety bill SB 1047, 2024.09.29.

27) 법안의 정식 명칭은 "The Safe and Secure Innovation for Frontier Artificial Intelligence Models Act"

(2) 캘리포니아주 주지사위 법안에 거부권 행사

하원을 통과한 이번 법안은 오픈소스 발전과 소규모 AI 기업의 발전을 저해할 수 있다는 기술업계의 의견을 반영해 개정된 버전이다. 개정안은 AI 안전을 관할하는 별도의 기관을 설립하는 대신 규제 업무를 기존 당국에 맡기기로 했으며, 실질적 피해가 없어도 안전 규정을 준수하지 않으면 AI 기업을 처벌하도록 한 원안과 달리 규정 위반 시에도 실질적 피해를 끼친 경우에만 기업을 처벌하기로 변경했다.

그러나 기술 기업과 벤처 투자사, 연방 의원들은 반대를 고수하며 법안 통과를 막기 위한 치열한 로비활동을 전개해 왔으며, 뉴섬 주지사는 결국 법안에 거부권을 행사했다. 뉴섬 주지사는 이 법안이 가장 크고 비싼 모델에 대한 규제에 집중하고 있으며, AI 시스템이 위험한 상황에 사용되는지, 중요한 의사 결정에 관여하거나 민감한 데이터를 사용하는지는 고려하지 않는다고 지적했다.

뉴섬 주지사는 소형 AI 모델이 민감한 데이터를 다루는 위험한 작업에 사용되거나 대형 AI 모델이 고객 서비스와 같은 위험이 낮은 업무에 활용될 수도 있다며, 「SB1047」이 대중을 AI 기술의 위협에서 보호하기 위한 최선의 방식으로 볼 수 없다고 했다.²⁸⁾

<표 3-10> 최근 발효된 캘리포니아 AI 규제법

발효일	법	특징
9월 4일	SB 896	• AI가 인프라에 미치는 위험 분석 요구
9월 12일	AB 1831	• AI 생성 아동 포르노물 처벌 확대
9월 17일	AB 2355	• 선거 관련 허위 정보에 대한 플랫폼 책임 강화
9월 28일	AB 2013	• AI 훈련 데이터 공개 의무화
	AB 1008	• AI 저장 데이터에 대해 개인정보 보호법 적용

출처: California Legislative, SB-1047 Safe and Secure Innovation for Frontier Artificial Intelligence Models Act.(2023-2024), 2024.09.03.

28) California Legislative, SB-1047 Safe and Secure Innovation for Frontier Artificial Intelligence Models Act.(2023-2024), 2024.9.3

(https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB1047)

<표 3-11> AI 규제법 「SBI047」 주요 내용

부분	주요내용
대상 AI 모델 (Covered AI models)	<ul style="list-style-type: none"> • 2027. 1. 1. 前: (1) 1억달러 초과 개발비용 소모 및 전체 1026 FLOP 등 초과 컴퓨팅 연산량 이용 또는 (2) 1천만 달러 초과 비용소모 및 전체 1025 x 3 FLOP 등 컴퓨팅 연산량 초과 파인 튜닝 이용 하는 AI 모델 • 2027년 1. 1. 後: 비용은 동일(인플레이션 조정 가능), 컴퓨팅 연산량 기준은 캘리포니아주의 Government Operations Agency에서 추후 결정
안전평가관련	<ul style="list-style-type: none"> • (1) 대량 학살을 일으킬 수 있는 화학, 생물학, 방사능 또는 핵무기의 제작 또는 이용, (2) 핵심기반시설 목표 사이버 공격을 통한 대량 학살 또는 최소 5억 달러 이상의 피해, (3) AI모델의 관여로 다음 두 가지를 일으키는 행위[(i) 인간의 감독, 개입 등의 제한, (ii) 사망, 중상해, 재산상피해, (인간이라면) 고의·중과실로 인해 형사 처벌 되는 행위, 그러한 행위의 교사·방조]를 통해 발생한 대량 학살 또는 최소 5억 달러 이상의 피해, (4) 이와 유사한 공공 안전 및 보안에 대한 막대한 피해를 핵심 피해(Critical harm)로 규정 • 위와 같은 핵심 피해를 막기 위해 개발자들은 문서화 된 안전 및 보안 프로토콜 수립 • 대상 AI 모델을 이용 또는 공개전, 개발자들은 모델이 핵심 피해를 일으킬 가능성을 평가하고, 평가 결과 기록을 보관하며, 적절한 안전 조치를 도입
클러스터 정책	<ul style="list-style-type: none"> • 컴퓨팅 클러스터란, 데이터센터의 100 GB/초를 초과 하는 네트워크로 연결 되고, 이론적으로 최소 1020 이상의 초당 FLOP 등의 최대 컴퓨팅 연산량을 가지며, AI 학습을 위해 이용될 수 있는 기계 집합을 의미 • 컴퓨팅 클러스터 운영자는 고객이 대상 AI모델을 학습하기 위해 충분한 컴퓨팅 자원을 이용 하는 경우에 대비한 정책 및 절차를 구비 • 해당 정책 및 절차는 컴퓨팅 클러스터 운영자의 잠재 고객 기본인적 정보와 사업 목적 등의 확보, 잠재적 고객의 클러스터 이용 의도가 대상 AI 모델 학습용 인지 평가, 고객의 통제하에 모델을 학습 또는 작동시키는 전체 자원의 즉각 중단 기능 시행 등 내용을 포함
중단기능	<ul style="list-style-type: none"> • 대상 AI 모델 학습 전, 개발자들은 학습을 포함 하여 모든 모델 기능을 중단 시키는 전체 중단기능(full shutdown)을 구현
법무장관 소송	<ul style="list-style-type: none"> • 이 법안 내용의 위반으로 사망, 중상해, 재산상피해, 도난, 재산착복또는 임박한 공공 안전 위협 등에 대해 법무장관(Attorney General)이 민사 소송 제기 가능 • 법무장관은 징벌적 손해배상을 포함한 금전적 손해배상, 민사페널티, 금지명령구제, 선언적 구제 등을 구할 수 있음
감사 및 보고	<ul style="list-style-type: none"> • 2026년 부터 대상 AI 모델 개발자들은 매년 준법 감시를 위한 독립적 감사를 위해 제3의 기관으로 부터 감사 진행 • 대상 AI 모델 개발자들은 민감 정보를 삭제한 안전 및 보안프로토콜과 감사보고서를 공개하며, 법무장관 요청시 법무장관에게 무삭제본을 제공 • 대상 AI 모델 개발자들은 준수성명서(compliance statements)를 매년 제출 하고, 안전 사고 발생 시 법무장관에게 72시간내 보고
내부 고발자 보호	<ul style="list-style-type: none"> • 내부직원의 법률 미준수 고발 또는 AI 모델의 비합리적 위험폭로에 대한 대상 AI 모델 개발자들의 저지 또는 보복금지

출처: California Legislative, SB-1047 Safe and Secure Innovation for Frontier Artificial Intelligence Models Act.(2023-2024), 2024.09.03

5) 미국 백악관 예산관리국의 연방정부 AI 사용에 관한 정책 발표²⁹⁾

(1) 개요

미국 백악관 예산관리국(OMB)이 바이든 대통령의 AI 행정명령에 의거한 후속조치로 2024년 3월 28일 연방 정부기관의 안전하고 책임 있는 AI 사용을 지원하기 위한 정책을 발표했다.³⁰⁾

미국은 2023년 10월 조 바이든 대통령이 행정명령 제14110호를 발표하면서 사업자들에 대한 AI규제를 선보여 주목받았는데, 이번에는 그 후속조치로 OMB이 2024. 3. 28. 연방정부 기관들의 AI이용을 규제하는 정책을 공개한 것이다. 해당 정책은 연방정부 기관들의 AI이용과 관련하여 ▲AI조달 ▲담당인력배정 ▲학습데이터 활용 ▲리스크 관리 등 전방위적인 규제 내용이 담겨 있다.

특히, 정부기관에서 국민들의 삶에 큰 영향을 미칠 수 있는 안전 영향AI(safety-impacting AI) 또는 권리영향 AI(rights-impacting AI)를 떼어내 분류 하도록 하고, 해당 AI이용을 위해서는 필수적으로 갖추어야 할 조건과 이행 행위 등을 담고 있으며 정부기관이 AI를 조달할 때 준수하여야 할 원칙들을 기재하고 있다.

(2) OMB의 행정명령에 따른 주요 지침 내용

바이든 대통령은 2023년 10월 30일 AI 행정명령을 통해 OMB에 연방정부의 AI 사용에 관한 지침 개발을 요구했으며, OMB는 행정명령에 따라 이번 정책을 마련했다.³¹⁾

① (AI 거버넌스 강화) AI 위험 관리와 혁신 촉진을 위해 각 연방 정부기관은 60일 내에 최고AI책임자(CAIO)를 지정하고 AI거버넌스위원회를 설립 필요

29) OMB, Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence, 2024.03.28. (<https://www.whitehouse.gov/briefing-room/statements-releases/2024/03/28/fact-sheet-vice-president-harris-announces-omb-policy-to-advance-governance-innovation-and-risk-management-in-federal-agencies-use-of-artificial-intelligence/>)

30) 미국 의회와 정부는 최근 AI관련 각종 법률·정책을 선보이며 산업 규제를 강화하고 있다. 가장 포괄적인 연방 법률로서 미국 의회는 국가 인공지능 이니셔티브법(National Artificial Intelligence Initiative Act of 2020)과 정부 인공지능법(AI in Government Act of 2020)을 통과시킨 바 있고, 조 바이든 미국 대통령은 2023년 10월 행정명령 제14110호' 안전하고 보안성이 높으며 신뢰할 수 있는 AI의 개발 및 이용(Safe, Secure and Trustworthy Development and Use of Artificial Intelligence)'을 선보인 바 있다.

31) OMB, Memorandum For The Heads Of Executive Departments And Agencies, 2024.03.28 (<https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf>)

- OMB와 과학기술정책실은 12월부터 최고AI책임자협의회를 정기적으로 개최해 연방정부 전반의 AI 사용을 관리하며, 기관 별 AI거버넌스위원회는 차관급이 의장을 맡아 기관 내 AI 사용을 조정 및 관리
- ② (AI 사용의 위험 해결) 연방 정부기관은 2024년 12월 1일까지 미국인의 권리나 안전에 영향을 미칠 수 있는 방식의 AI 사용 시 구체적인 보호조치를 마련
 - 의료, 교육, 고용, 주택 등 광범위한 AI 사용에 적용되는 보호조치는 ▲AI가 대중에 미치는 영향에 대한 평가와 모니터링 ▲알고리즘 차별 위험의 완화 ▲정부의 AI 사용 방식에 대한 투명성 보장을 포괄
- ③ (책임 있는 AI의 혁신 촉진) 연방 정부기관들이 적절한 보호조치를 갖추고 사회의 시급한 과제 해결에 AI 기술을 도입할 수 있도록 지원
 - 일례로 연방재난관리청은 AI를 사용해 허리케인으로 인한 피해를 신속히 검토 및 평가하고 있으며, 질병통제예방센터는 AI를 통해 질병 확산을 예측하고 마약성 진통제의 불법 사용을 감시
- ③ (AI 인력 양성) OMB는 연방 정부기관에 AI 인재 확대와 역량 향상을 지시했으며, 연방 정부는 인력 양성을 위해 다음과 같은 계획을 준비
 - 바이든 행정부는 2024년 여름까지 AI 전문가 100명을 채용할 계획이며, 인사관리처는 정부 내 AI 인력 유지를 위해 AI 전문가에게 탄력적 급여 및 휴가 제도를 적용하는 지침을 발표
 - 2025회계연도 대통령 예산안에 정부 전반의 AI 교육 프로그램을 확대하기 위한 500만 달러의 추가 예산을 편성

<표 3-12> 예산관리국(OMB) 주요 발표 내용

부분	세부내용	주요 핵심 내용
적용범위	연방기관	<ul style="list-style-type: none"> • 44 U.S. Code §3502(1)의 기관(agency)를 의미 • 회계감사원, 연방선거관리위원회 등을 제외한 기관으로 모든 행정·군사기관, 국영기업, 행정부 내 기관 또는 독립규제기관 • 특정내용은 최고재무책임자법(Chief Financial Officer Act, CFO법) 31 U.S. Code §901(b)의 연방기관에만 적용
	AI	<ul style="list-style-type: none"> • 기관의 기존·신규 AI에 적용 • AI를 포함 하는 정보 시스템 전체가 아닌, AI를 실행시키거나 AI에 기반 하는 시스템 기능에 대한 정책

부분	세부내용	주요 핵심 내용
AI 거버넌스 강화	실시행위	<ul style="list-style-type: none"> CAIO 임명필요 CFO법 기재 기관들은 연방정부 내 AI 이용 관련 이슈를 조정·통제 하기 위해 고위공무원 소집 2036년 까지 격년별로 기관 마다 정책 준수 계획 발표 매년 AI 이용사례 기록 OMB 보고 및 공개발표
	CAIO의 역할·책임 등	<ul style="list-style-type: none"> CAIO는 해당 직위의 책임 업무수행에 필요한 기술, 지식, 훈련 및 전문성을 갖춘 인력 이어야 함 CAIO는 기관 내 고위공무원들과 정기적으로 소통 필요 CAIO는 기관의 AI 이용조정, AI 혁신촉진, AI 이용리스크 관리 등 책임 업무 수행
	기관 내 업무 조정	<ul style="list-style-type: none"> 기관 내 고위 공무원들의 AI 이슈에 대한 관심 확보 기관의 필요에 근거하여 기관의 AI 업무 조정 CFO법 기재 기관들은 AI 거버넌스 위원회 설치
책임 있는 AI 혁신 추진	AI 전략	<ul style="list-style-type: none"> CFO법 기재 기관들은 임무 또는 서비스 제공에 최대영향을 미치는 현 AI 이용과 향후 계획, 현 기관의 AI 이용에 대한 성숙도 등을 포함한 AI 전략 계획을 공개 발표
	책임 있는 AI 이용을 위한 장벽 제거	<ul style="list-style-type: none"> 기관들은 유연한 AI 개발·활용 및 AI 혁신과 리스크관리에 적절히 AI 자원과 전문성이 이용 되도록 내부 환경 조성 필요 기관들은 IT 인프라·데이터·사이버보안·생성 AI를 특별히 고려하여 책임 있는 AI 이용을 위한 장벽 제거에 노력
	AI 인력	<ul style="list-style-type: none"> 기관들은 미국 연방 인사관리처(Office of Personnel Management, OPM)가 발간 예정인 AI 및 기술 인재채용 계획(AI and Tech Hiring Playbook)에 근거 하여 관련 인력 확보 기관의 AI 인재 채용 과정이 기관 내 상부에 보고 되도록 하고, OPM과 OMB에게 관련 데이터를 제공 할 담당자 임명
	AI 공유 및 협력	<ul style="list-style-type: none"> 기관들은 활용하는 AI 코드와 모델 공유 및 공개 기관의 AI 개발과 테스트에 이용한 데이터 자산 공유 및 공개 AI 모델 공유 및 공개 시 민감 데이터 유출을 막기 위해 리스크 평가 실시
	AI 관련 요구사항 단일화	<ul style="list-style-type: none"> OMB는 기관별 AI 개발 및 이용을 조율 형식 통일 추진·모범 사례 공유·기술 자원 공유 등을 통해 조율
AI 이용 리스크 관리	실시행위	<ul style="list-style-type: none"> 2024. 12. 1.까지 기관들은 안전영향·권리영향 AI의 최소의무행위(minimum practices)를 충족해야 하고, 그렇지 않을 시 이용 중단 2024. 12. 1.까지 기관들은 안전영향·권리영향 AI에 해당하는지에 대한 결정과 면제 사유 존재 여부를 인증(2024. 12. 1. 이후부터는 매년)
	안전영향·권리영향 AI 해당 여부 결정	<ul style="list-style-type: none"> 안전영향 AI(Safety-impacting AI)란 사람의 생명·신체, 기후·환경, 핵심인프라, 전략자산·자원 등에 중대한 영향을 미칠 수 있는 결정의 근거가 되어 줄 결과물을 산출해 내는 AI 권리영향 AI(Rights-impacting AI)란 개인 또는 독립체의 시민권·자유·프라이버시, 평등한 기회, 의료지원 등 핵심 정부 자원 또는 서비스 이용 등에 법적·강제적 또는 중대한 영향을 미칠 수 있는 결정의 근거가 되어 줄 결과물을 산출해 내는 AI 기관들은 기존 또는 이용 계획에 있는 AI가 위 AI에 해당 하는지 판단 필요 Appendix I에 기재된 목적으로 AI가 이용 되면, 해당 AI는 안전영향·권리영향 AI에 해당 한다고 간주

부분	세부내용	주요 핵심 내용
	최소의무행위	<p>기간관련</p> <ul style="list-style-type: none"> • 기관들은 OMB에게 최대 1년의 기간 연장 요청 가능 <p>AI 이용전</p> <ul style="list-style-type: none"> • AI 영향평가 실시 • 실제 이용 환경 맥락에서 AI 성능 테스트 실시 • 독립적 AI 평가 <p>AI 이용중</p> <ul style="list-style-type: none"> • 모니터링 실시 • AI 이용리스크 정기평가 • 안전·권리에 대한 새로운 리스크 완화 • 충분한 인력 교육 및 평가 확보 • 안전·권리에 대한 중대한 영향을 가져오는 결정 또는 행위 경우 추가적인 감독 <ul style="list-style-type: none"> • 개입·책임제공 • 공공에 대한 공지 및 쉬운 언어로 쓰인 문서 제공 <p>권리영향 AI에 대한 추가 의무 행위</p> <ul style="list-style-type: none"> • 평등 및 공정성에 대해 AI의 영향을 인지·평가하고, 알고리즘적 차별이 발견되는 경우 완화 • 영향을 받는 커뮤니티와 일반인들로 부터 피드백 수령 및 반영 • AI로 인한 차별에 대해 모니터링 및 완화 실시 • 부정적 영향을 받은 개인에게 관련 사항 통지 • 인간이 개입 하는 논의·구제 프로세스 확보 • AI 결정을 배제(opt-out)할 수 있는 선택권 확보
	연방정부의 AI 조달 리스크 관리	<ul style="list-style-type: none"> • AI의 관련 법규 준수 보장 • 투명성 및 성능 개선을 위한 노력 • AI 조달관련 경쟁 촉진 • AI에 투입될 데이터를 핵심 자산으로 파악하여 정부 데이터에 대한 권리 확보 • AI 학습용으로 테스트 데이터에 개발자들 또는 공급업체가 직접 의존 하는 행위 방지 • 생체인식 AI에 대한 책임 있는 조달 • 생성 AI에 대한 책임 있는 조달 • 환경 효율성 및 지속가능성 평가

출처: OMB, Memorandum For The Heads Of Executive Departments And Agencies, 2024.03.28

6) 미국 백악관의 AI 기술 활용에 관한 최초의 국가안보 각서 발표³²⁾

(1) 개요

미국 백악관은 2024년 10월 24일 AI가 향후 국가안보와 외교에 중대한 영향을 미칠 것이라는 인식에 따라 AI 활용에 관한 최초의 국가안보 각서(National Security Memorandum, 이하 NSM)를 발표하였다.

바이든 대통령이 2023년 10월 30일 발표한 AI 행정명령에 따라 수립된 NSM은 ▲미국이 AI 개발에서 세계 선도 ▲국가안보에 AI 활용 ▲AI를 중심으로 국제적 합의와 거버넌스 발전을 목표로 설정하였다.

(2) 주요 조치 내용

- ① (미국이 AI 개발에서 세계 선도) 안전하고 보안성이 있으며 신뢰할 수 있는 AI 개발에서 미국의 리더십 구축
 - AI 시스템 개발에 필요한 첨단 AI 칩 확보를 위해 칩 공급망 안전과 다양성을 개선하고, AI를 고려해 차세대 슈퍼컴퓨터와 신기술 개발을 지원
 - 미국의 AI 리더십을 위협하는 경쟁국의 동향 정보를 최우선으로 수집하고, 경쟁국의 스파이 활동에 대응해 AI 개발기업에 사이버보안 및 방첩 정보를 적시에 제공
 - 미국 AI안전연구소(USAISD)를 미국 정부와 산업계 간 공식 연락 창구로 지정하고, AI안전연구소가 첩보 부문과 국방부 등 국가 안보 기관과 협력할 수 있는 간소화된 체계를 구축
 - 국가AI연구자원(National AI Research Resource) 프로젝트를 강화해 대학, 시민사회, 중소기업의 연구자들이 의미 있는 AI 연구를 수행할 수 있도록 지원
- ② (국가 안보에 AI 활용) 인권과 민주주의 가치를 보호하는 방식으로 첨단 AI를 활용하여 국가 안보 목표의 달성을 지원
 - 국가 안보에서 AI 거버넌스와 위험관리를 발전시키기 위한 프레임워크를 구축해 프라이버시 침해, 편향과 차별, 인권 침해 등 AI 관련 위험을 평가 및 완화를 위한 자세한 내용과 지침을 제공

32) The White House, FACT SHEET: Biden-Harris Administration Outlines Coordinated Approach to Harness Power of AI for U.S. National Security, 2024.10.24.

- AI 시스템의 효과적 도입을 위한 간소화된 조달 관행을 수립하고 AI 기업 및 클라우드 기업과 같은 비전통적인 공급업체와 협업을 발전시킬 방안을 모색
- ③ (AI 중심 국제적 합의와 거버넌스 발전) NSM은 미국 정부와 동맹국, 파트너 간 협력을 통해 AI 거버넌스와 위험관리를 발전시키기 위한 거버넌스 프레임워크를 구축해 국제법을 준수하는 동시에 인권과 기본적 자유를 보호하는 방식으로 AI 기술의 개발과 사용을 보장

6) 바이든-해리스 행정부의 행정명령 이후 1년 동안의 AI 성과³³⁾

(1) 개요

1년 전, 바이든 대통령은 미국이 AI의 위험을 관리하는 데 앞장서도록 하는 획기적인 행정명령을 발표했다. 이 행정명령은 ▲AI의 안전 및 보안 위험을 관리 ▲미국인의 프라이버시를 보호 ▲형평성과 시민권을 증진 ▲소비자와 근로자를 옹호 ▲혁신과 경쟁을 촉진 ▲전 세계적으로 미국의 리더십을 발전시키는 등의 광범위한 조치를 지시했다. 2024년 10월 30일 바이든-해리스 행정부는 연방 기관의 행정 명령이 지난 1년 동안 수행한 모든 조치를 일정대로 완료했다고 발표했으며 그 조치는 100개가 넘는 것이라고 설명했다.

(2) 주요 조치 내용

- ① (안전 및 보안 위험 관리) 이 행정 명령은 위험한 생물학적 물질, 소프트웨어 취약성 및 해로운 목적으로 AI를 개발하고자 하는 사람으로부터 미국인을 보호하기 위해 취해진 가장 대담한 조치
 - 국방물자생산법 권한(Used Defense Production Act)³⁴⁾을 통해 AI 시스템 개발자에게 안전 및 보안 테스트 결과를 포함한 중요한 정보를 미국 정부에 보고하도록 요구
 - AI 안전 테스트 및 평가를 주도하여 AI 안전 과학을 발전시켰음. 상무부의 미국 AI 안전 연구소(US AISI)는 두 개의 선도적인 AI 개발업체와 최근 체결한 계약을 통해 주요 신규 AI 모델의 사전 배포 테스트를 시작

33) Whitehouse, Fact Sheet: Key AI Accomplishments in the Year Since the Biden-Harris Administration's Landmark Executive Order, 2024.10.30

(<https://www.whitehouse.gov/briefing-room/statements-releases/2024/10/30/fact-sheet-key-ai-accomplishments-in-the-year-since-the-biden-harris-administrations-landmark-executive-order/>)

34) 국가 안보를 위해 대통령이 민간 부문의 물자 생산과 공급에 관여할 수 있는 권한을 부여하고 있고, 국내 산업을 통제할 수 있는 중대한 비상권한을 부여하고 있다.

- AI에 대한 최초의 국가 안보 각서(NSM)를 발행. NSM은 미국이 안전하고 보안성이 높으며 신뢰할 수 있는 AI의 세계적 개발을 주도하도록 하기 위한 연방 기관의 구체적인 조치를 지시
 - AI 데이터센터 인프라에 대한 새로운 태스크포스를 출범
 - 생성 AI 콘텐츠로 인한 위험을 줄이기 위한 식별할 수 있는 조치를 취함(콘텐츠에 라벨을 붙이고 투명성을 개선하는 접근 방식 포함)
 - 국토안보부 장관에게 중요 인프라에서 AI가 안전하게 사용될 수 있도록 조언하기 위해 AI 안전 보안 위원회(AISSB)를 설립
- ② (근로자, 소비자, 개인정보 보호 및 시민권 보호) AI는 미국인이 구매하는 제품과 서비스를 변화시키고 있으며, 일자리와 직장에도 영향을 미치고 있다. 특히 개인정보 보호, 형평성 및 시민권에 대한 위험을 초래하거나 악화시키고 있는데 바이든 대통령의 행정 명령은 이러한 각 영역에서 미국인을 보호하기 위해 지난 1년 동안 다음과 같은 조치를 수행
- AI가 직장을 위해 구축되고 사용됨에 따라 근로자를 보호하고 권한을 부여하는 데 도움이 되는 기본 원칙과 관행 지침 개발
 - 의료분야를 위한 AI가 개발되고 도입됨에 따라 혁신을 장려하는 동시에 환자의 권리와 안전을 보호
 - 교육 분야에서 AI가 안전하고 신뢰할 수 있는 지침 및 리소스 게시
 - 공공 혜택 프로그램을 관리하는데 있어 AI와 알고리즘 시스템 보호책 마련
- ③ (선한 목적의 AI 활용) 작년에 주요 정부기관들은 민간 부문과 협력하고, 귀중한 AI 활용 사례의 개발과 활용을 촉진
- 국제AI연구리소스(National AI Research Resource:NAIRR) 파일럿을 시작하여 150개 이상의 연구팀에 활용 및 기타 AI 리소스에 대한 액세스 권한을 부여
 - 미국 전역에서 AI 교육 및 훈련을 장려
 - 최고의 AI 과학자, 엔지니어, 기업가들이 미국에 입국할 수 있는 범위 확대
 - 사회 전반에 걸쳐 책임 있는 AI 개발과 사용을 발전시키기 위해 수백만 달러의 추가 투자 발표
- ④ (정부 차원의 AI 인재 영입) AI와 AI 지원 전문가를 정부에서 적극적으로 영입
- AI 기술을 연방 차원에서 활용해 거버넌스 강화, 위험 완화 및 혁신을 위한 최초의 정부 차원 정책 발표
 - 연방 기관의 AI에 대한 책임 있는 활용을 위한 정부 차원의 정책 발표

- AI 인재 서지(AI Talent Surge)를 통해 연방정부에 250명 이상의 AI 실무자 고용
 - AI 인재가 급증할 것으로 예상됨에 따라 백악관은 정부, 학계, 시민 사회 전반에 걸쳐 교육 및 경력 경로를 지원하기 위한 기금 발표
- ⑤ (글로벌 미국 리더십으로 발전) 바이든 대통령은 AI가 위험을 완화하고, AI의 책임 있는 거버넌스를 보장하기 위한 국제 사회의 노력을 이끌기 위한 작업을 지시
- 유엔 총회에서 획기적인 결의안을 후원하고 통과
 - 글로벌 AI 표준 및 AI 관련 인프라에 대한 미국의 참여 계획 개발
 - 유럽 평의회에서 AI 및 인권, 민주주의, 법의 통제 권한에 관한 기본 협약 서명
 - 젠더 기반 온라인 괴롭힘 및 학대에 대한 글로벌 행동 파트너십 출범
 - 국무부와 미국 국제개발처는 경제 개발을 위한 글로벌 AI 연구와 AI 활용을 발전시키기 위한 리소스 발표

<표 3-13> 미국 행정발표 이후 주요 AI 성과

구분	주요 내용
안전 및 보안 위험 관리	<ul style="list-style-type: none"> • 국방물자생산법 권한(Used Defense Production Act) 통해 정부에 보고 • AI 안전 테스트 및 평가를 주도 • AI에 대한 최초의 국가 안보 각서(NSM)를 발행 • AI 데이터센터 인프라에 대한 새로운 태스크포스를 출범 • 생성 AI 콘텐츠로 인한 위험을 줄이기 위한 식별할 수 있는 조치 실행 • AI 안전 및 보안 위원회(AISSB)를 설립
근로자, 소비자, 개인정보 보호 및 시민권 보호	<ul style="list-style-type: none"> • 근로자 보호 및 권한을 부여 위한 기본 원칙과 관행과 지침을 개발 • AI 의료 혁신 장려 및 환자의 권리와 안전 보호 • 교육 분야에서 안전하고 신뢰할 수 있는 AI 사용 지침 및 리소스 게시 • AI 알고리즘 시스템의 책임감 있고 공평한 사용에 대한 보호책 마련
선한 목적의 AI 활용	<ul style="list-style-type: none"> • 국제AI연구리소스(National AI Research Resource:NAIRR) 파일럿 실행 • 미국 전역 AI 교육 및 훈련 장려 • AI 과학자, 엔지니어, 기업가들의 미국 입국 범위 확대 • 책임 있는 AI 개발과 사용 발전을 위한 추가 투자 발표
정부 차원의 AI 인재 영입	<ul style="list-style-type: none"> • AI 기술을 연방 차원에서 활용하기 위한 정부 차원 정책을 발표 • AI인재서지(AI Talent Surge)를 통해 연방 정부에 250명 이상의 AI 실무자 고용 • AI 교육 및 경력 경로를 지원하기 위한 기금 발표
글로벌 미국 리더십으로 발전	<ul style="list-style-type: none"> • 유엔 총회 결의안 후원 및 통과 • 유럽 평의회 기본 협약 서명 • 글로벌 AI 표준 인프라 관련 포괄적인 계획 개발 • 글로벌 행동 파트너십 출범 • 국무부와 미국국제개발처의 리소스 활용 발표

출처: Whitehouse, Fact Sheet: Key AI Accomplishments in the Year Since the Biden-Harris Administration's Landmark Executive Order, 2024.10.30.

제2절 EU

1. 그간의 인공지능 정책 추진 경과

EU는 2018년 4월 ‘EU를 위한 AI(Artificial Intelligence for Europe)’ 를 발표하며 EU 차원의 AI 정책을 본격적으로 수립하기 시작하였다. ‘EU를 위한 AI’ 는 ▲경제 전반에 걸친 기술적·산업적 역량 및 AI 활용 증진 ▲사회·경제적 변화의 준비 ▲윤리적·법적 프레임워크 확보라는 3대 목표를 제시했다. AI에 대한 EU의 접근 방식은 기술의 우수성과 신뢰성에 중점을 두고 연구와 산업 역량을 강화하는 동시에 안전과 기본권을 보장하는 것을 목표로 한다.³⁵⁾ 이후 EU는 2018년 12월 회원국 전체의 일관성과 통일성을 위해 ‘AI 조정계획(Coordinated Plan on AI)’ ³⁶⁾을 수립하고, 2021년 4월 개정하였다. AI 조정계획은 유럽 전반의 AI 정책에 대한 협력 기틀을 마련하였으며, 투자 영역을 설정, 유럽 공통의 AI 전략과 방향 하에서 회원국들의 자국 내 AI 전략 마련을 촉진하였다.

2021년 개정된 AI 조정계획은 2018년 계획에서 더 나아가 EU 집행위원회(European Commission), 회원국들, 산업계 간 상호협력하에서 신뢰할 수 있는 AI(Trustworthy AI) 분야에서 EU의 글로벌 리더십 확보를 추구하였다. 개정된 계획은 2021년부터 2027년까지의 중장기 계획으로 환경 조성, 연구와 활용, 사람과 사회 중심 AI, 핵심 분야 리더십의 총 4개 분야에서 40개의 핵심 과제를 도출하였다. 또한 1차 계획에서 부재했던 반도체, 클라우드, 초소형 전자공학, 에지컴퓨팅 등 AI 시스템의 개발과 구동에 필요한 인프라 관련 계획이 대거 포함되었다.

<표 3-14> EU AI 조정계획의 전략목표와 실행과제

전략목표	실행과제
1. EU의 AI 개발 및 활용을 위한 환경 조성	<ul style="list-style-type: none"> • 정책 통찰력 획득과 공유 • 데이터의 잠재력 활용 • 핵심 컴퓨팅 역량 육성
2. 실험실(Lab)에서 시장까지 AI 발전에 탁월한 EU 구축	<ul style="list-style-type: none"> • AI, 데이터 및 로봇공학 유럽 파트너십 및 전문가 그룹 등을 통한 이해관계자 협력 • 연구역량 구축 및 이동성(mobility) 강화 • 개발자가 시험 및 실험하고, 중소기업과 공공행정 부문이 시를 도입할 수 있는 환경 조성

35) <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>

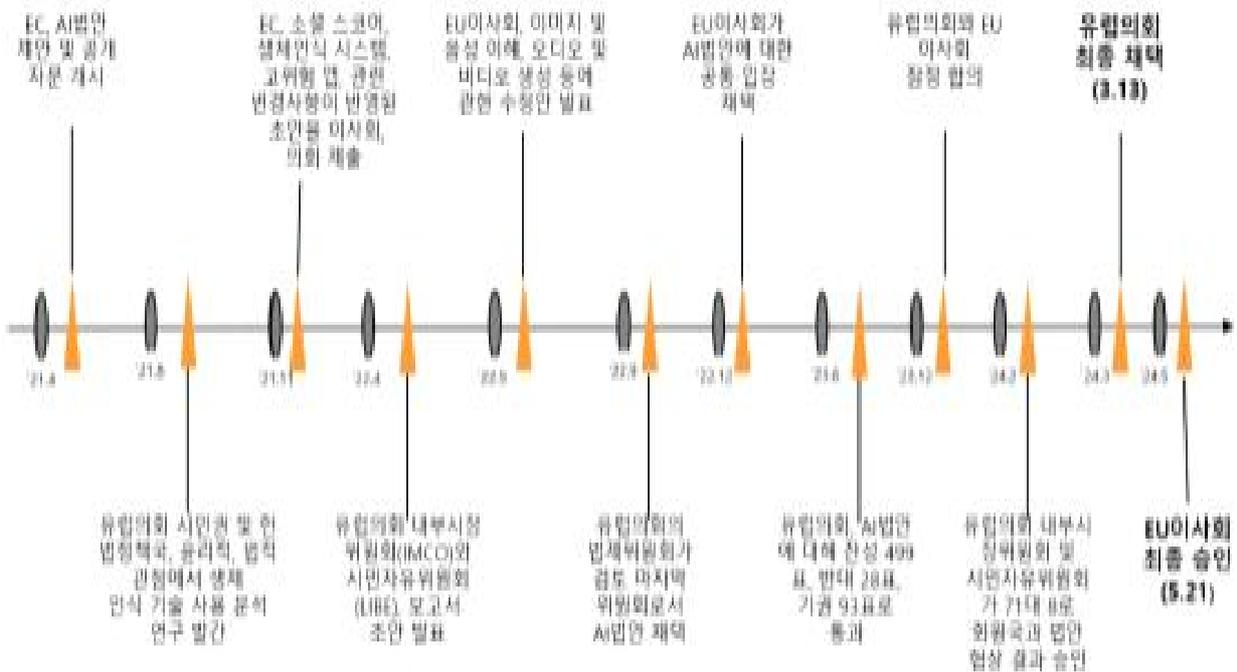
36) European Commission, Coordinated Plan on Artificial Intelligence 2021 Review, 2021.4.21

전략목표	실행과제
3. 사람들을 위해 노동하고(work), 사회에서 선(Good)을 행하는 능력(Force)의 AI 개발	<ul style="list-style-type: none"> AI 생태계 활성화에 필요한 인재 양성 및 전문기술 공급 개선 AI 시스템에 대한 신뢰 확보를 위한 정책 프레임워크 개발 전 세계의 지속 가능하고 신뢰 가능한 AI에 대한 EU 비전 홍보
4. 영향력이 큰 분야에서 전략적 리더십 구축	<ul style="list-style-type: none"> 기후와 환경을 위한 AI 도입 차세대 AI를 활용하여 의료 증진 유럽 선두 유지: AI 세계의 로봇공학 전략 공공 부문을 AI의 선구자로 구현 법 집행, 이민 및 망명에 AI 적용 AI를 통한 더 안전하고 오염 없는 이동성 구현 지속 가능한 농업을 위한 AI 지원

자료: European Commission(2021.4), SPRI 정리, 2024.10.

EU 집행위원회는 2021년 4월, 유럽의회(European Parliament)와 유럽이사회(European Council)의 요구에 따라 AI 법(AI Act) 초안을 처음 발표했다. 수년간의 논의와 3자 협상 끝에 AI 법은 2023년 12월 회원국 간 정치적 합의를 거쳐 2024년 3월 유럽의회를 통과했으며, 2024년 5월 27개 EU 회원국 대표로 구성된 EU 이사회 회의에서 최종 승인되었다.

[그림 3-1] EU AI 법의 입법 경과



자료: European Commission(2021.4), SPRI 정리, 2024.10.

<표 3-15> EU의 AI 관련 주요 정책

시기	정책
2018년 4월	• 유럽을 위한 AI 전략(AI for Europe) 수립
2018년 6월	• 유럽 AI 연맹(EU AI Alliance) 출범 • AI 관련 공개 정책 대화를 수립하기 위한 유럽위원회 이니셔티브로 AI 고위급 전문가 그룹 구성
2018년 12월	• EU 집행위원회, AI 조정계획 발표
2020년 2월	• AI 백서(White Paper on AI) 발간
2021년 4월	• AI 법안(AI Act) 발표 • AI 조정계획 업데이트
2022년 9월	• 비계약적 민사 책임 규칙을 AI에 적용하는 'AI 책임 지침' 제안 발표
2023년 6월	• 유럽의회 본회의 표결에서 AI 법안 협상안 가결
2024년 1월	• EU 집행위, AI 스타트업과 중소기업 지원을 위한 AI 혁신 패키지 발표
2024년 3월	• 유럽의회에서 AI 법 최종 채택
2024년 5월	• 유럽이사회에서 AI 법 최종 승인 • AI 사무국 설립
2024년 7월	• EU 집행위, AI 협약 초안 발표
2024년 8월	• AI 법 발효

자료: European Commission(2021.4), SPRI 정리, 2024.10.

2. 2024년 인공지능 정책 동향

1) AI 법(Artificial Intelligence Act)

(1) 제정 경과

EU 집행위원회(EC)는 유럽의회와 유럽이사회의 요구에 따라 2021년 4월 AI 개발과 활용에 대한 신뢰 확보, 기본권 보장, 이용자 안전 강화 등을 목표로 하는 AI 법안을 제안하였다. EU 집행위는 AI 이용에 따른 위험에 대처하기 위해, 위험 수준에 비례한 접근방식을 채택하여 위험 수준을 ‘수용 불가 위험’, ‘고위험’, ‘최소위험’으로 분류하여 규제 방향을 설정하였다. 이후 EU 이사회는 2022년 12월 EU 집행위가 제안한 AI 법안을 검토하고 AI 법의 제정 목표로 유럽을 신뢰할 수 있는 글로벌 AI 허브로 만들고, AI 사용 시 위험과 이점 간 균형의 달성을 제시하였다. 유럽의회는 EU 집행위가 제안한 AI 법안을 검토하고 생성 AI 관련 사항을 추가한 수정안을 본회의에서 가결하였으며, EU 집행위, 유럽의회, EU 이사회는 약 6개월간 협의를 거쳐 2023년 12월 최종 합의에 도달했다.

EU 회원국 대사들로 구성된 EU 상임 대표위원회는 2024년 2월 만장일치로 AI 법을 승인하였다. EU AI 법은 2023년 12월 입법 절차의 최대 관문인 이사회-집행위-유럽의회 간 3자 협상을 통과했으나, 일부 국가가 법안에 유보적 태도를 취하면서 회원국의 승인이 지연되다가 합의에 이르렀다. AI 법 합의를 미뤄 온 프랑스, 독일, 이탈리아, 오스트리아는 GPT-4와 같은 강력한 AI 모델의 규제 완화를 요구하며 유럽의회와 추가 협상을 촉구했으나 결국 의견을 바꾸어 법안에 동의하였다. 이들 국가는 엄격한 규제가 프랑스의 미스트랄(Mistral)이나 독일의 알레프 알파(Aleph Alpha)와 같은 유럽 AI 스타트업의 발전을 저해할 것을 우려했다. 그러나 독일과 이탈리아가 먼저 반대 의견을 철회하면서, 마지막까지 법안에 반발했던 프랑스는 고위험 AI 시스템 개발 기업에 과도한 규제 부담을 주지 않고 투명성과 기업비밀 보호 간 균형을 맞추는 등의 조건 하에 AI 법을 승인하기로 합의하였다. EU 회원국들의 승인을 얻은 AI 법은 2024년 3월 찬성 523표, 반대 46표, 기권 49표의 압도적 찬성으로 가결되어, 5월 유럽이사회에서 최종 승인 후 8월 1일 공식 발효되었다.

(2) AI 법의 주요 내용과 적용 단계

EU의 AI 법은 위험 수준에 따라 차등적 규제를 적용하는 위험 기반 접근법을 채택하여, 주요기반시설, 교육과 직업훈련, 공공서비스, 채용, 이민 및 국경관리, 선거 절차 등에 사용되는

고위험 AI 시스템은 위험 완화 체계 수립과 고품질 훈련 데이터셋 확보, 활동 기록 등 엄격한 규제를 부과한다. 개인의 특성·행동과 관련된 데이터로 점수를 매기는 사회적 점수 평가(Social Scoring), 직장이나 학교에서 사용되는 감정인식 기술 등 허용할 수 없는 위험이 있는 AI 시스템은 2025년 2월부터 사용이 금지된다. 법 집행기관의 생체인식 시스템 사용도 실종자 수색이나 테러 예방 등 일부 목적에 한해 엄격한 보호조치가 수반되는 경우에만 허용된다.

AI 법은 다양한 작업을 수행할 수 있는 범용 AI 모델에 대해서도 투명성 의무를 부과하는 등 별도의 규칙을 마련했으며, 범용 AI에 관한 의무 사항은 발효 1년 뒤인 2025년 8월부터 적용된다. AI 법에 따르면 모든 범용 AI 모델에 일반적 규제를 적용하되, 시스템적 위험(Systemic Risk)이 존재할 경우 추가적인 규제를 적용한다. 시스템적 위험은 범용 AI 모델로 인하여 EU 시장에 중대한 영향을 미치며, 공공 보건, 안전, 안보, 기본권 또는 EU 사회 전체에 실질적 또는 합리적으로 예측 가능한 부정적인 영향을 미칠 위험을 의미한다. 일반적 규제에 범용 AI 모델 공급자는 AI 개발 및 테스트에 대한 자세한 기록을 보관하고, 지적 재산을 보호하면서도 AI를 사용하려는 다른 회사에 관련 정보를 제공하고, EU 집행위원회 및 국가 당국과 협력해야 한다. 시스템적 위험이 있는 범용 AI 모델 제공자는 표준 프로토콜을 사용하여 모델을 평가하고, 시스템적 위험을 식별하여 완화해야 하며, 심각한 사고가 발생하면 AI 사무소와 국가 당국에 보고하며, AI 모델과 인프라가 안전한지 확인해야 한다.

〈표 3-16〉 EU AI 법의 순차적 적용 단계

일정	정책
2024년 8월	• AI 법 발효
2024년부터 법 시행 기간	• 대부분 규칙이 적용되는 2026년 8월까지의 과도기에 자발적 법 이행을 위한 AI 협약(AI Pact) 도입
법령 발효 이후 6개월	• 용납 불가 AI 시스템 금지조항 발효
법령 발효 이후 12개월	• 기반모델을 포함한 범용 AI 관련 의무 사항 발효
법령 발효 이후 24개월	• 독립적으로 운영되는 고위험군 AI 관련 의무 사항 발효
법령 발효 이후 36개월	• AI 법령의 모든 의무 사항 발효

자료: SPRI 정리, 2024.11

AI 법은 위반의 중대성에 따라 과징금을 차등 적용한다. 사용 금지 AI 애플리케이션 규정을 위반한 기업은 전 세계 연 매출의 7%에 해당하는 과징금을 부과받을 수 있으며, 기타 의무 위반 시 연 매출의 3%, 잘못된 제공 시 연 매출의 1.5%의 과징금이 부과된다.

<표 3-17> EU AI 법의 과징금 적용 기준

구분	내용
금지된 AI 시스템에 관한 규정 위반	<ul style="list-style-type: none"> 최대 3,500만 유로(약 518억 원)와 직전년 회계연도 기준 전 글로벌 연 매출액의 최대 7% 중 더 큰 금액
고위험 AI 시스템에 관한 규정 위반	<ul style="list-style-type: none"> 최대 1,500만 유로(약 222억 원)와 직전년 회계연도 기준 전 글로벌 연 매출액의 최대 3% 중 더 큰 금액
인증기관, 관할 당국에 부정확, 불완전, 오해의 소지가 있는 정보를 제공한 경우	<ul style="list-style-type: none"> 최대 750만 유로(약 112억 원) 또는 직전년 회계연도 기준 글로벌 연 매출액의 최대 1% 중 더 큰 금액

자료: SPRI 정리, 2024.11.

AI 법이 발효되면서 EU 회원국은 2025년 8월 2일까지 법 집행을 감독하고 시장 감시 활동을 수행할 국가 당국을 지정해야 하며, EU 차원에서는 집행위 산하의 AI 사무국이 3개 자문기구의 지원을 받아 법 집행을 감독하게 된다. 유럽AI위원회가 회원국 전반의 일관된 법 적용과 위원회와 회원국 간 협력을 지원하고, 독립 전문가로 구성된 과학 패널은 기술적 자문을 제공하는 한편, 다양한 이해관계자로 구성된 자문 포럼도 운영된다.

EU 집행위는 일부 규정을 제외한 대부분 규칙이 적용되는 2026년 8월까지의 과도기 동안 기업들의 자발적 법 이행을 위한 ‘AI 협약(AI Pact)’을 도입하는 한편, AI 법 시행에 관한 사항을 세부적으로 설명하는 지침 개발을 추진하고 있다. EU 집행위가 AI 협약을 준수할 참여자를 모집한 결과 2024년 7월 기준 550개 기업이 참여 의사를 표시하였다. AI 협약은 ▲AI 협약 네트워크 구성과 정보공유 ▲관련 조직의 AI 법 의무 준수 촉진 및 홍보의 2개 축을 핵심으로 하며, 첫 번째 핵심 축에서는 AI 사무국이 주최하는 워크숍을 통한 정보공유와 모범사례 교환을 위한 전용 온라인 공간 마련을 추진할 계획이다. 두 번째 핵심 축과 관련해서는 AI 법의 조기 구현을 촉진하기 위한 프레임워크를 마련하고 조직의 참여 서약을 통한 점진적 목표 달성을 추진할 계획이다. AI 협약은 자발적으로 참여하는 모든 조직이 AI 법을 적용받는 시점까지 다음과 같은 핵심 요건을 이행할 것을 강조하였다.

<표 3-18> EU AI 협약의 핵심 요건

구분	내용
AI 협약의 핵심 요건	<ul style="list-style-type: none"> • 조직 내 AI 활용을 촉진하고 향후 AI 법 준수를 위한 AI 거버넌스 전략을 채택 • AI 법에 따라 고위험으로 간주되는 영역에서 개발되거나 사용되는 AI 시스템을 식별 • AI 시스템 개발 시 수명주기 전반에 걸쳐 AI 시스템 사용으로 인해 발생할 수 있는 건강, 안전 및 기본권 관련 위험을 식별하기 위한 프로세스 마련 • AI 시스템 배포 시 관련 AI 시스템 사용을 통해 영향을 받을 수 있는 개인 및 집단의 기본권에 대한 위험 식별

자료: SPRI 정리, 2024.11.

EU 집행위는 2024년 11월 AI 법에 따라 범용 AI에 적용할 실천 강령(General-Purpose AI Code of Practice) 1차 초안을 공개했으며, 네 차례에 걸쳐 의견수렴 및 초안 수정을 진행해 2025년 5월 1일에 최종 실천 강령을 발표할 계획이다.³⁷⁾ 실천 강령은 범용 AI 개발 기업에 대하여 투명성과 저작권 준수, 위험 평가 및 기술/거버넌스 위험 완화를 강조하고 있다.

2) EU 집행위원회의 주요 정책

(1) AI 혁신 패키지

EU 집행위는 2024년 1월 AI 스타트업과 중소기업을 지원하기 위한 일련의 정책으로 구성된 AI 혁신 패키지를 발표하였다. 혁신 패키지는 ▲AI 팩토리 출범 ▲AI 사무국 설립 ▲EU AI 스타트업 및 혁신 커뮤니케이션으로 구성된다.

우선 엑사급 슈퍼컴퓨터 개발을 위한 EuroHPC 규정을 개정해 EU의 슈퍼컴퓨터 공동 사업 활동을 주도할 ‘AI 팩토리’를 출범하기로 하였다. 이는 대규모 범용 AI 모델의 신속한 학습을 지원할 AI 전용 슈퍼컴퓨터를 확보해 스타트업과 중소기업에 포함한 공공과 민간 부문의 AI 활용 확대에 기여하기 위함이다. AI 팩토리는 ▲스타트업을 위한 윈스톱 상점 제공 ▲알고리즘 개발에서 AI 스타트업과 연구 생태계 지원 ▲대규모 AI 모델에 대한 평가와 검증 테스트 ▲범용 AI 모델 기반의 다양한 최신 AI 애플리케이션 개발을 지원할 예정이다.

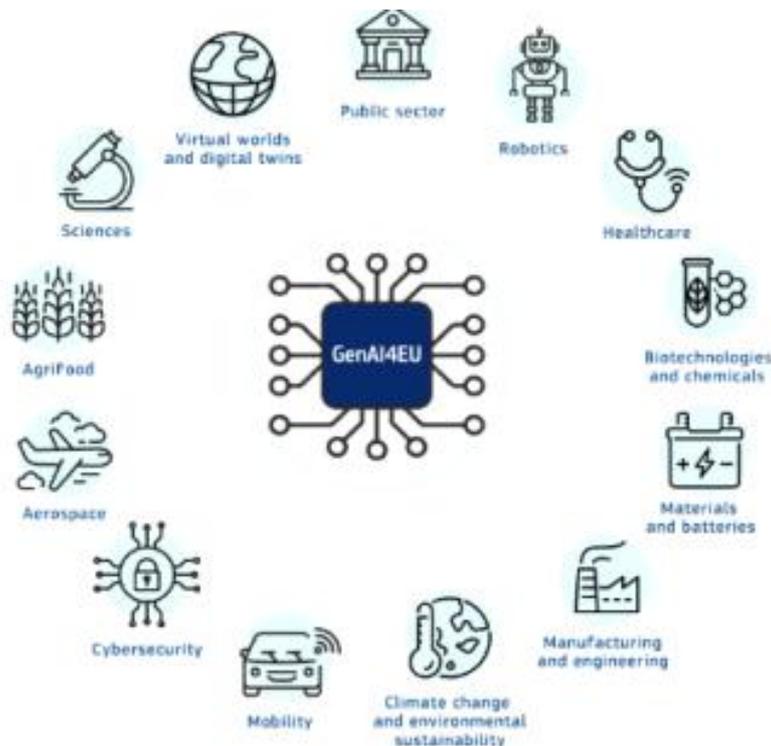
EU 집행위 산하에 설치되는 AI 사무국은 유럽 차원의 AI 정책 개발과 조정 및 AI 법의 시

37) European Commission, First Draft of the General-Purpose AI Code of Practice published, written by independent experts, 2024.11.14.

행을 담당한다. AI 스타트업과 혁신 커뮤니케이션은 연구혁신 프로그램과 벤처 캐피털 프로그램을 통한 자금 지원과 인력양성 프로그램 등으로 구성된다. 호라이즌 유럽(Horizon Europe)과 생성 AI에 특화된 디지털 유럽(Digital Europe) 프로그램을 통해 생성 AI 연구개발에 자금을 지원하고 2027년까지 40억 유로 규모의 공공과 민간 투자를 추가 창출한다는 구상이다. 금융 지원과 지분 투자가 혼합된 유럽혁신위원회(EIC) 액셀러레이터 프로그램 등을 활용해 AI 스타트업에 대한 공공과 민간 투자를 장려하고, 교육과 훈련 활동을 통해 EU의 생성 AI 인재풀을 강화할 이니셔티브를 추진할 계획이다.

또한 공동 유럽 데이터 공간(Common European Data Spaces)을 개발해 모델 학습과 개선에 필요한 데이터를 AI 커뮤니티에 제공할 예정이다. 또한 ‘GenAIEU’ 이니셔티브를 통해 유럽의 주요 산업과 공공 부문의 신규 생성 AI 애플리케이션 개발을 지원할 계획으로, 적용 분야는 로봇공학, 의료, 생명공학, 제조, 차량, 기후, 가상 세계를 포함한다.

[그림 3-2] EU ‘GenAI4EU’의 적용 분야



출처: European Commission, 2024.01.

(2) AI 사무국 설립

EU 집행위는 2024년 5월 AI 법 집행을 지원하기 위해 신설되는 AI 사무국(AI Office)의 조직구조와 주요 업무를 공개하였다. AI 사무국은 사회경제적 이익과 혁신을 촉진하면서 위험을 완화하는 방식으로 AI의 개발과 배포, 사용을 지원하며, 특히 범용 AI 모델과 관련해 AI 법 시행에서 핵심 역할을 하게 될 전망으로, 업무 수행을 위해 기술 전문가, 행정가, 변호사, 정책 전문가, 경제학자 등 140명 이상의 직원을 채용할 계획이다. EU AI 사무국은 ▲규제 준수 ▲AI 안전 ▲AI 연구지원 ▲AI의 공익적 활용 ▲AI 혁신과 정책조정 5개 부서로 구성된다. 이후 AI 사무국은 AI 연구지원 부서의 명칭을 ‘AI와 로봇 전문성(Excellence in AI and Robotics)’으로 확정하였으며, 기타 부서의 명칭은 기존과 동일하다. EU AI 사무국은 5개 부서 외 2명의 자문 위원(책임 과학 자문위원, 국제관계 자문위원)도 두고 있다.

〈표 3-19〉 EU AI 사무국의 조직 구성

부서	내용
규제 준수	• 회원국들과 협력해 EU 전반에서 AI 법의 일관된 시행을 지원하며, 침해 사례 조사 및 제재 담당
AI 안전	• 첨단 AI 모델의 시스템적 위험 식별과 위험 완화 조치, 모델 평가와 시험 접근 방식 확인
AI 연구지원	• AI 생태계 육성을 위한 연구개발에 자금과 지원을 제공하고, ‘GenAIEU’ 이니셔티브를 조정해 모델 개발과 혁신 응용사례 통합 촉진
AI의 공익적 활용	• 기후 예측, 암 진단, 도시 재건을 위한 디지털트윈 등 공익적 목적의 국제 협력 추진
AI 혁신과 정책조정	• AI 혁신과 정책조정 부서는 EU AI 전략의 이행과 동향 모니터링 및 투자, 규제 샌드박스를 통한 혁신 생태계 육성 등 담당

출처: European Commission, 2024.10., SPRI 정리, 2024.10.

EU AI 사무국은 ▲AI 법 지원과 AI 규칙 시행 ▲신뢰할 수 있는 AI 개발과 활용 강화 ▲국제 협력 촉진 ▲기관, 전문가 및 이해관계자와 협력을 주요 업무로 한다.

〈표 3-20〉 EU AI 사무국의 주요 업무

업무	내용
AI 법 지원과 AI 규칙 시행	<ul style="list-style-type: none"> • EU 차원의 자문기관을 설립하고 회원국 전반의 일관된 AI 법 적용에 기여 • 범용 AI 모델의 역량과 도달 범위를 평가하고 시스템적 위험이 있는 모델 분류를 위한 도구, 방법론, 벤치마크 개발 • 선도적 AI 개발자, 과학 커뮤니티, 기타 전문가와 협력해 실무 규범 작성 • 모델 기능평가를 포함한 규칙 위반 가능성 조사 및 공급업체에 시정 요청 • AI 법 준수를 모니터링하기 위한 지침 및 기타 수단 마련
신뢰할 수 있는 AI 개발 및 활용 강화	<ul style="list-style-type: none"> • EU 전체에서 AI의 사회적·경제적 이득을 누리기 위한 조치와 정책 추진 • 모범사례에 대한 조언 제공 및 AI 샌드박스, 테스트 및 AI 도입을 위한 유럽 지원체계에 대한 접근성 확대 • 신뢰할 수 있는 AI의 혁신적 생태계 장려로 EU의 경쟁력과 경제성장 강화 • EU 집행위의 AI 도구 활용 및 AI 리터러시 강화 지원
국제협력 촉진	<ul style="list-style-type: none"> • 전 세계 유사 기관과 협력을 포함해 신뢰할 수 있는 AI에 대한 EU의 접근방식 홍보 • 글로벌 AI 접근방식 기여를 목표로 AI 관련 국제협력과 거버넌스 촉진 • 회원국 지원을 포함하여 AI 관련 국제 협정의 개발과 이행 지원
기관, 전문가 및 이해관계자 협력	<ul style="list-style-type: none"> • EU AI위원회, 과학 패널 및 기타 이해관계자와 협력

출처: European Commission, 2024.10., SPRI 정리, 2024.10.

EU AI 사무국은 범용 AI 모델에 대한 규칙을 제정하고 AI 개발자와 학계, 기타 이해관계자와 협력해 실행규범을 마련하고 범용 AI 모델 평가를 실시하며, 필요한 경우 제재도 가할 계획이다. EU AI 사무국은 AI 시스템 정의와 금지사항에 대한 가이드라인을 준비해 AI 법 발효일로부터 6개월 내(2025년 1월)에 공개할 예정이며, 범용 AI 모델이 준수해야 할 의무사항에 대한 실행규범은 AI 법 발효 후 9개월 뒤(2025년 4월)까지 공개할 계획이다.

(3) 연구에서 생성 AI의 책임 있는 사용에 관한 지침³⁸⁾

EU 집행위는 2024년 3월 연구자, 연구기관, 연구자금 지원기관을 대상으로 책임 있는 AI 사용을 안내하는 지침을 발표했다. 지침에 따르면 생성 AI는 연구의 효율성을 높이고 새로운 발견을 촉진하고 있으나, 연구자들은 표절, 민감한 정보의 유출, 모델 고유의 편향과 같은 부정적 영향을 고려해 투명하고 책임 있게 생성 AI를 사용해야 한다. EU 집행위는 유럽 단일연구공간(European Research Area)³⁹⁾ 포럼과 함께 유럽 전역의 일관된 접근방식을 보장하기 위해 이번 지침을 마련하였다.

38) European Commission, Guidelines on the responsible use of generative AI in research developed by the European Research Area Forum, 2024.03.20.

지침에 따르면 연구자는 연구 결과물에 최종 책임을 져야 하며, AI 사용을 투명하게 공개하고 국내외 법규를 준수해야 한다. AI 도구로 생성된 결과물의 무결성을 책임지고 AI 도구의 한계(편향, 허구, 부정확성)에 대한 비판적 시각을 갖춰야 하며, 생성 AI 사용 시 개인정보보호와 기밀 유지, 지식재산권 보호에 주의해야 한다. 또한 동료 평가나 연구 제안 평가와 같이 연구자나 기관에 영향을 줄 수 있는 민감한 활동에서 생성 AI의 사용을 자제해야 한다.

연구기관은 연구 활동에서 AI의 책임 있는 사용을 지원하고 기관 내 AI 시스템 개발과 사용을 적극 감독하며, AI 지침을 연구 윤리와 모범관행 지침에 통합해야 한다. AI 사용에 대한 교육을 제공하거나 지원해야 하며, 기관 내 AI 활용 현황을 파악하고 기술의 한계를 분석해 피드백을 제공해야 한다. 가능한 자체적으로 관리하는 AI 도구를 개발하거나 클라우드 기반 AI 도구를 활용해 개인정보보호와 기밀 유지가 보장되는 환경에서 연구 자료를 AI 도구에 입력해야 한다.

자금 지원기관은 연구에서 책임 있는 AI의 사용을 장려하고 내부 절차에서 AI 사용을 검토하며, 빠르게 변화하는 AI 분야의 동향을 모니터링해 적극 참여해야 한다. AI의 책임 있고 윤리적인 활용을 지원하는 자금 지원 프로그램을 설계하고, 평가와 심사 과정에서 AI를 투명하게 사용하며, 자금 신청자에게 신청 준비 과정 및 연구 활동에서 AI 사용 내역을 투명하게 보고해야 한다.

(4) EU의 AI 연구개발 계획⁴⁰⁾

EU 집행위는 유럽의 디지털 역량 강화를 위한 2021-2027년도 신규사업 ‘디지털 유럽(Digital Europe)’ 프로그램과 연구자금 지원 프로그램 ‘호라이즌 유럽(Horizon Europe)’을 통해 AI 연구개발을 지원하고 있다. 두 프로그램을 통해 EU는 매년 10억 유로 이상을 AI에 투자해 혁신적인 아이디어와 솔루션에 자금을 지원하고 있으며, 민간 부문과 회원국으로부터 추가로 자금을 유치해 10년간 연간 투자 규모를 200억 유로로 확대하겠다는 목표를 제시하였다.

39) 과학기술 및 연구개발 활동을 EU 차원에서 통합하여 회원국간 협력을 강화하고 연구투자의 중복을 방지하며 유럽 전체의 연구 역량을 최대한 활용하기 위한 정책

40) European Commission, European research development and deployment of AI, 2024.10.11.

호라이즌 유럽 프로그램은 AI 연구개발에 26억 유로를 할당했으며, 2024년 4월에는 새로 AI와 양자 기술 프로젝트에 1억 1,200만 유로 이상의 자금을 지원하였다. 특히 거대 AI 모델을 발전시키기 위한 프로젝트에 5천만 유로, AI 투명성과 신뢰성 향상 프로젝트에 1,500만 유로를 할당하였다. 한편, 디지털 유럽 프로그램은 2021~2024년에 AI의 개발과 배포 지원에 10억 유로 이상을 할당하였다.

3) 기타 EU 조직 및 관련 기관의 주요 정책

(1) 유럽개인정보보호감독기구(EDPS)

유럽개인정보보호감독기구(European Data Protection Supervisor, 이하 EDPS)는 2024년 6월 EU 기관을 대상으로 생성 AI와 개인정보에 관한 지침을 발간하였다.⁴¹⁾ EDPS는 EU 행정부의 개인정보 처리를 감독하는 기관으로서 EU 기관의 생성 AI 시스템 사용 시 일반개인정보보호법(GDPR) 준수를 지원하기 위해 지침을 마련했다.

EDPS는 EU 기관들의 생성 AI 사용 시 GDPR에서 규정한 데이터 최소화 및 데이터 정확성 원칙의 이행을 요구하였다. EU 기관들은 의도한 목적에 필요한 수준으로만 개인정보를 수집 및 처리해야 하며, AI 시스템의 훈련에서 배포까지 수명주기 전반에 걸쳐 꼭 필요한 개인정보만 사용하고 책임 있게 처리해야 한다. 데이터 정확성의 원칙에 따라 생성 AI의 개발과 사용 시 데이터 정확성을 보장해야 하며, 사람의 감독을 포함한 정기적 모니터링을 통해 모델이 출력하는 결과도 통제해야 한다.

지침에 따르면 EU 기관들은 생성 AI 도입 전 개인정보 영향평가(DPIA)를 실시하여, 개인정보 처리와 관련된 위험을 파악하고 완화해야 한다. 개인정보 영향평가는 개인정보보호법 준수를 보장하고 개인의 권리를 보호할 수 있도록 데이터 보호 책임자(DPO)와 협의해 수행하여야 한다. 또한 EU 기관들은 투명성을 보장하기 위해 생성 AI 시스템에서 개인정보를 사용하는 방법, 시기, 이유에 대한 자세한 정보를 개인에게 제공해야 한다. 세부 정보에는 사용된 데이터셋과 다양한 개발 단계에서 수행되는 데이터 처리 활동 등이 포함되어야 하며, 챗봇과 같은 AI 시스템의 경우 사람의 개입 없이 AI와 상호작용하고 있음을 알리는 등, 특정한 투명성 요건이 필요할 수 있다.

41) European Data Protection Supervisor, EDPS Guidelines on generative AI: embracing opportunities, protecting people, 2024.6.03.

지침은 생성 AI 시스템을 이용한 자동화 의사결정에서 개인의 권리를 침해해서는 안 된다는 점도 강조하였다. AI와 관련된 모든 의사결정 절차에 인간이 개입할 수 있어야 하며, 개인이 결정에 이의를 제기하고 의견을 표현할 수 있도록 지원해야 한다. 생성 AI 시스템 사용 시에는 데이터 정확성, 편향, 보안 등에 대해 지속적인 모니터링과 평가를 실시하고, 편향 최소화를 위한 모범사례를 채택해야 한다.

(2) 유럽개인정보보호이사회(EDPB)

유럽개인정보보호이사회(EDPB)는 2024년 7월 전체 회의에서 EU AI법 내에서 개인정보 보호 당국(이하 DPA)의 역할에 대한 성명을 채택하였다.⁴²⁾ AI 법 제2조제7항에 명시된 바처럼, AI 시스템 수명주기 전반의 개인정보 처리는 EU 개인정보보호 법제의 적용을 받는다. EDPB는 AI 시스템 개발 과정에서 개인정보보호 문제 관련 감독과 조정에 대응하고 DPA의 역할을 강조하기 위해 성명을 발표하였다.

성명에 따르면 DPA는 AI 법에서 요구하는 다양한 임무를 맡을 역량을 갖춰 각국 내 AI 법 집행을 관리할 시장감시당국(MSA) 역할을 수행할 수 있다. 성명은 DPA를 MSA로 지정하면 단일 연락 창구가 마련되어 AI 법과 개인정보보호법의 적용을 받는 규제기관 간 상호작용을 통해 모든 이해관계자에게 도움이 될 수 있다고 밝혔다. 또한 DPA가 AI를 포함한 신기술로 인한 기본권 위험 평가에서 전문성을 가지며, 독립 규제기관으로서 AI 법에 따라 효과적인 감독을 수행할 수 있다고 강조했다.

EDPB는 성명에서 EU 회원국에 AI 법 제74조제8항에 언급된 고위험 AI 시스템에 대하여 DPA를 MSA로 지정할 것을 권고하고, 특히 고위험 AI 시스템이 개인정보 처리와 관련해 개인의 권리와 자유에 영향을 미칠 수 있는 분야에 속하는 경우 DPA를 MSA로 지정할 필요가 있다고 강조했다. AI 법 제70조제2항 규정에 따라 MSA는 AI 법의 단일 연락 창구가 되어야 하는 만큼, DPA를 MSA로 지정하면 여러 부문에 걸쳐 일관되고 효과적인 접근 방식을 지원할 수 있다고 설명했다. 또한 DPA에 AI 시스템 감독을 위한 새로운 업무와 권한을 부여하기 위해 회원국은 DPA에 인적 자원과 재정 자원의 추가 제공을 요구하였다.

42) EDPB, Statement 3/2024 on data protection authorities' role in the Artificial Intelligence Act framework, 2024.07.17.

EDPB는 또한 범용 AI 모델이나 시스템이 개인정보 처리를 수반하는 경우 DPA와 EDPB의 감독 권한에 속할 수 있다며, AI 시스템 감독 시 EU 집행위원회 산하의 AI 사무국과 EDPB, 각국 DPA 간 상호협력의 필요성을 강조하였다.

(3) 유럽평의회

46개 회원국으로 구성된 유럽 민주주의·인권·법치 수호 기구인 유럽평의회(Council of Europe)는 2024년 5월 ‘AI와 인권·민주주의·법치에 관한 협약⁴³⁾’을 채택했다. 유럽평의회 산하의 AI 위원회가 마련한 이번 협약은 AI 시스템의 설계, 개발, 배포로 구성되는 수명주기 전반에 걸친 법적 프레임워크를 제시했으며 AI에 관한 최초의 법적 구속력이 있는 국제 조약이다. 조약은 AI 시스템 수명주기 내의 활동과 관련된 기본 원칙을 아래와 같이 제시하였다.

〈표 3-21〉 유럽평의회 AI 조약의 기본 원칙

구분	내용
AI 조약의 기본 원칙	<ul style="list-style-type: none"> • AI 시스템의 수명주기 전반에서 인간의 존엄성과 자율성을 존중 • AI 시스템의 수명주기 활동에서 상황과 위험에 맞춰 적절한 투명성과 감독 요건을 마련하고, AI 시스템이 생성한 콘텐츠 식별에도 적용 • AI 시스템의 수명주기 활동으로 인한 인권, 민주주의, 법치주의에 대한 부정적 영향에 대한 책임을 보장 • AI 시스템의 수명주기 활동에서 국제법과 국내법상 평등(성평등 포함) 및 차별금지 원칙 준수 및 형평성 있는 결과를 달성 • AI 시스템의 수명주기 활동에서 개인의 프라이버시권과 개인정보 보호를 보장 • AI 시스템의 신뢰성과 그 산출물에 대한 신뢰 제고 • 혁신을 촉진하되 인권, 민주주의, 법치주의에 대한 부정적 영향을 방지하기 위해 관계 당국의 감독하에 AI 시스템 개발, 실험, 테스트를 위한 통제된 환경을 조성

자료: SPRI 정리, 2024.11.

조약은 AI 시스템 수명주기 내 활동으로 인해 발생한 피해에 대하여 효과적이고 접근성 높은 구제수단의 보장과 함께 AI 시스템의 설계·개발·사용이 인권·민주주의·법치에 미치는 위험과 영향을 평가 및 완화하는 조치를 요구하였다. 단, 조약을 비준하는 국가들은 국가 안보에 영향을 미치는 AI에는 조약을 적용할 의무가 없으며, 민간 부문에 대한 적용 여부도 자율적으로 결정할 수 있다. 미국, 영국, EU를 포함한 10개국은 2024년 9월 AI 조약의 첫 서

43) Framework Convention on artificial intelligence and human rights, democracy, and the rule of law (CETS No. 225)

명국으로 참여했으며, 서명국은 조약 이행을 위해 입법 또는 행정 조치를 채택해야 한다.

(4) 유럽 ombudsman

유럽의 행정감찰기구 유럽 ombudsman(European Ombudsman)이 EU 집행위의 AI와 기타 첨단 기술 도입에서 투명성을 확보하기 위한 전략적 계획을 발표했다.⁴⁴⁾ 유럽 ombudsman은 이번 계획으로 AI 도입 이후 책임 있는 사용에 대한 인식 제고를 목표로 하고 있으며, 첫 번째 단계로 EU 집행위가 도입했거나 도입 예정인 AI 도구와 관련 의사결정에 대한 정보를 요청했다. AI는 작업의 품질과 효율성을 개선할 수 있으나 부정확성이나 편향, 설명 가능성 등의 문제를 야기할 수 있으며, 특히 대중에게 직접 영향을 미치는 의사결정에 사용될 때는 인간의 감독이 필수적이다. 따라서 유럽 ombudsman은 EU 집행위에 투명한 절차를 거쳐 AI를 도입해야 하며, AI가 업무를 지원하되 시민에게 영향을 미치는 결정에서 인간을 대체하지 않도록 보장해야 한다고 강조했다.

유럽 ombudsman은 ▲대중 의견수렴 분석 ▲EU 경쟁규칙의 위반 적발 ▲민원 처리의 3개 영역이 AI 사용에서 특히 중요하다고 밝혔다. AI를 활용하면 EU 시민의 정책 의견을 한층 효과적으로 파악할 수 있으나, 특정 정책에 대한 지지도 인식을 조작하는 등의 잠재적 위험을 EU 집행위가 어떻게 평가하고 대처하는지를 평가할 필요가 있다. 또한 AI는 EU 단일 시장에서 최대 영향을 미치는 경쟁규칙 위반을 적발하는 데 주로 사용되고 있으나, EU 집행위는 조사 과정에서 오류나 편향 가능성, 관련 당사자의 절차상 권리 보호에 유의해야 한다. 유럽ombudsman은 EU 집행위가 침해나 기타 불만을 처리하는 방식에 대한 민원을 다루며, EU 집행위는 민원 처리과정에서 AI가 제안한 해결책을 지나치게 신뢰하여 자동화 편향이 발생하지 않도록 보장해야 한다.

유럽ombudsman은 EU 집행위에 상기 3개 영역과 관련된 AI 도구의 기능에 대한 상세한 설명과 함께, AI 도구 사용을 감독하는 절차의 투명성과 책임성을 어떻게 보장할 지에 대한 정보를 요청했다. 구체적으로 AI 도구가 수행하는 구체적인 작업과 인간의 의사결정을 지원하는 방식, AI 도구로 인한 위험과 해결 방안에 대한 정보를 요청했으며, EU 집행위 내 AI 도구의 승인 절차 및 AI 도구 사용의 투명성 확보를 위한 정보 제공 여부 및 AI 도구 사용의 감독 방식과 절차에 대한 설명도 요구하였다.

44) European Ombudsman, How the European Commission decides on and uses artificial intelligence, 2024.03.15.

제3절 중국

1. 그간의 인공지능 정책 추진 경과

중국은 2014년 6월 9일 중국 과학원·공정원 원사 대회에서 시진핑 주석이 AI 등 차세대 정보기술(IT)에 대한 중요성을 언급한 이후 AI를 국가전략으로 격상하여 AI 산업 육성 정책을 펼치고 있다. 2017년 국무원이 발표한 ‘차세대 AI 발전규획’은 AI 분야의 체계적 발전 방향을 제시한 최초의 정책 문건으로, 2030년까지 AI에서 세계 1위 도약을 선언하고 중장기 발전계획을 제시하였다. 발전규획은 AI 발전의 6대 핵심 과제로 ▲개방적이고 협력적인 AI 기술 혁신 시스템 구축 ▲최첨단 고효율의 스마트 경제 육성 ▲안전하고 편리한 스마트 사회 건설 ▲AI 분야에서 군민(軍民) 통합 강화 ▲유비쿼터스 안전·고효율의 스마트 인프라 시스템 구축 ▲차세대 AI 중대 과학기술 사업의 전면 배치를 추진하기로 하였다.

〈표 3-22〉 중국 차세대 AI 발전규획의 단계별 목표와 주요 내용

단계	전략 목표	주요 내용
1단계 (~2020년)	기술과 응용에서 세계 선도	- AI 기술 표준과 서비스 체계 구축 - 글로벌 선도 기업 육성, 핵심산업 규모 1,500억 위안, 관련 산업 규모 1조 위안 달성 - AI 정책규범 마련
2단계 (~2025년)	AI 이론 발전	- 신규 AI 연구 성과 확보 - 제조, 의료, 도시, 농업, 국방 등 다양한 영역에서 AI 활용 - 핵심산업 규모 4,000억 위안, 관련 산업 규모 5조 위안 - AI 법률 규범과 이론 범위, 체계 구축
3단계 (~2030년)	AI 이론과 기술, 응용 전 분야에서 세계 선도	- 뇌 알고리즘, 스마트 제어 등 다양한 영역의 성과 확보 - AI 산업 경쟁력 확보, 핵심산업 규모 1조 위안, 관련 산업 규모 10조 위안 달성 - AI 법률법규, 정책 체계 완비

자료: 중국 국무원, SPRI 정리, 2024.10.

중국은 AI 발전을 위한 정책기반과 제도적 보장 체계를 구축하기 위해 ‘1+N’ 정책구조를 마련하였다. ‘1+N’ 정책구조에서 1은 중국 AI 발전의 핵심 축인 ‘차세대 AI 발전규획’을 의미하며, N은 발전규획의 실질적 이행을 위해 부처 차원에서 제수립한 AI 산업 발전계획과 실행계획을 의미한다. 일례로 공업정보화부는 차세대 AI 발전 규획의 목표 달성을

위해 2017년 12월 ‘차세대 AI 산업발전 3개년 실행계획(2018~2020)’을 수립하고, 국제 경쟁우위 형성 및 AI와 실물경제의 통합, 산업 발전 환경 최적화를 위한 중점 과제를 제시하였으며, 교육부는 2018년 ‘고등교육기관 AI 혁신 행동계획’을 발표하고 각 대학에 ‘AI+X’ 복합전공 육성계획을 수립해 2020년까지 AI 복합전공 학과 100개, AI 학교·연구소·교차연구센터 50개를 개설토록 하였다.

한편, 중국 과학기술부와 기타 6개 부서는 2022년 7월 AI 육성을 위해 ‘AI 활용 고도화를 통해 경제의 질적 발전을 촉진하기 위한 활용사례 혁신 가속화에 관한 지도의견’을 발표했다. 지도의견은 AI 응용을 촉진할 수 있도록 체계적으로 이끌어 경제의 고품질 발전을 앞당기고 차세대 AI의 발전 수준을 향상하는 것을 주요 목표로 제시하고 ① AI 중대 활용사례의 구축 ② AI 활용사례 혁신 능력 제고 ③ AI 활용사례 개방 가속화 ④ AI 활용사례의 혁신요소 공급 강화의 4개 분야에서 15대 중점 임무를 규정하였다.

중국은 2021년 3월에 발표한 제14차 5개년(2021~2025) 계획 및 2035년 장기목표에서도 2035년까지 완성할 7대 첨단기술 중 하나로 AI를 명시하였다. 14차 5개년 계획은 차세대 AI 분야의 기술 혁신을 계획하는 한편, 디지털 경제의 중점 산업으로서 AI의 중요성을 강조하였다.

중국 정부는 AI 산업 발전을 위한 지원 정책과 함께 산업 규제를 위한 법률 제정도 강화하고 있다. 중국의 인터넷 규제당국인 국가인터넷정보판공실은 인터넷 정보 서비스 산업 대상으로 2022년 알고리즘 규제와 심층합성 규제 등을 발표했다. 2022년 3월부터 시행 중인 ‘인터넷정보서비스 알고리즘 추천 관리규정’은 알고리즘 관련 세계 최초의 규제로, 사용자의 사용기록과 개인정보를 바탕으로 콘텐츠를 배치하거나 추천하는 플랫폼 사업자를 대상으로 긍정적인 콘텐츠를 촉진하고 불량한 정보가 확산되지 않도록 방지하는 의무를 부과하였다. 2023년 1월부터 시행된 ‘인터넷 정보서비스 심층합성 관리규정’은 심층합성 서비스 제공업체에 심층합성 서비스 사용자의 신원 인증과 콘텐츠 관리, 안전관리 책임 등을 부과하였다.

특히 중국 정부는 2022년 11월 오픈AI의 챗GPT 출시 이후 생성 AI 규제에 중점을 두고 있으며, 2023년 4월 국가인터넷정보판공실이 생성 AI 서비스 규제 초안을 발표한 데 이어 7월 생성 AI 규제를 확정하고 8월 15일부터 시행에 돌입했다. 생성 AI 서비스 관리 임시 시

행 방법’은 생성 AI 서비스 제공자의 데이터 처리 활동과 사용자 책임에 관한 요구사항을 규정하고 있다. 국가인터넷정보관공실은 2024년 9월에는 AI로 생성된 합성 콘텐츠에 대한 워터마크 표시 방법을 상세히 정한 규정 초안을 공개하기도 하였다.

한편, 생성 AI에 대한 전 세계적인 거버넌스의 필요성이 대두되면서, 중국 역시 2023년 10월 ‘글로벌 AI 거버넌스 이니셔티브’를 발표했다. 중국의 글로벌 AI 거버넌스 주장은 서구권의 논의와 비교해 AI 제품과 서비스 제공 시 타국의 주권을 존중하고 법률을 준수하며, 타국의 주권을 위협하는데 AI를 사용해서는 안 된다는 점, 그리고 타국의 AI 발전을 악의적으로 방해하기 위한 개발 장벽 조성과 글로벌 AI 공급망을 가로막는 기술 독점과 강압 조치에 반대한다는 점을 명시한 것이 특징이다.

2. 2024년 인공지능 정책 동향

1) 중앙정부

(1) 국무원

중국의 리창(李强) 국무원 총리는 2024년 3월 전국인민대표대회에서 발표한 2024년 정부 업무보고에서 ‘AI 플러스 행동’ 개념을 처음으로 제시하였다. 업무보고는 2024년 정부 과제 중 하나로 디지털 기술과 실물경제의 심층적 통합으로 디지털 경제의 혁신적 발전을 촉진할 정책의 수립을 선언하였으며, 특히 빅데이터와 AI 등의 연구개발과 응용을 심화하고 AI 플러스 행동을 실시하며 국제 경쟁력을 갖춘 디지털 산업 클러스터 구축을 추진하겠다는 계획을 밝혔다.

업무보고에서는 AI 플러스의 구체적인 내용은 제시되지 않았으나, 중국은 2010년대 중반 신성장 키워드로 ‘인터넷 플러스’ 전략을 수립한 전례가 있다. 인터넷 플러스는 인터넷과 전통산업(제조업·농업 등)을 결합해 산업구조 전환과 업그레이드를 도모하는 전략으로, AI 플러스 역시 기존 산업과 AI의 융합을 중점 추진할 것으로 추정된다.

중국 관영언론 글로벌 타임스에 따르면 AI 플러스 계획은 AI 연구와 응용을 심화해 AI와 실물경제의 통합을 촉진하고 디지털 경제의 발전에 이바지할 전망으로, 베이징 사회과학원의 왕펑(王鹏) 연구원은 이번 계획을 통해 AI의 광범위한 적용으로 생산 효율과 제품 품질

이 개선되고 공업, 농업, 서비스업 등 주요 산업의 고도화가 촉진될 것으로 예상했다. 투자와 인재 유치를 통해 경제의 신성장동력을 육성하고 중국의 과학기술 경쟁력을 제고하며 국가 차원의 AI 연구와 응용 역량 강화도 지원할 전망이다.

(2) 공업정보화부

가. 기본 정책⁴⁵⁾

2024년 9월 중국 공업정보화부 관계자에 따르면 중국의 AI 관련 기업은 현재 4,500여 개에 달하며, 반도체, 알고리즘, 데이터, 플랫폼, 애플리케이션 등 전체 산업 생태계를 포괄하는 핵심 산업 규모는 약 6,000억 위안(한화 약 112조 7,800억 원)에 달한다. 공업정보화부는 중국의 대규모 시장과 방대한 데이터 자원, 풍부한 응용 시나리오라는 장점을 바탕으로 AI 산업을 적극 육성할 계획이다.

공업정보화부는 우선 대형 AI 모델과 산업 수직(vertical) 모델 연구개발을 지원하고 고품질 데이터 자원을 풍부하게 공급하며 기업 혁신을 촉진할 계획이다. 또한 AI와 제조업의 융합을 중점적으로 추진해 산업계의 생성 AI 기술 적용을 장려하고 생성 AI 기업과 제조기업 간 연계를 강화하며, “생성 AI+제조” 시범 프로젝트를 전개하며 산업 스마트화와 업그레이드를 촉진할 방침이다. 마지막으로 AI 분야의 국제협력을 강화하고 오픈소스 AI 생태계를 육성하며, 기술 교류, 데이터 공유, 응용 시장, 인재 양성, 안전 보장 등의 영역에서 여러 국가와 협력해 공급망의 안정을 유지할 계획이다.

나. 국가 AI 산업 종합 표준화 체계 구축 지침⁴⁶⁾

중국 공업정보화부는 중앙인터넷정보관공실, 국가발전개혁위원회, 국가표준화위원회와 함께 2024년 7월 ‘국가 AI 산업 종합 표준화 체계 구축 지침’을 발표했다. 지침은 중국 AI 산업의 고품질 발전과 AI와 여타 산업의 통합을 지원하기 위한 표준체계의 수립을 목표로 제시하고 2026년까지 50개 이상의 국가표준과 산업표준을 제정하기로 하였다. 이와 함께 1,000개 이상 기업의 표준 이행을 독려하여 기업 혁신과 발전을 위한 표준의 효과를 강조하며, 20개 이상의 국제표준 제정에 참여하여 AI 산업의 세계화를 촉진한다는 계획을 공개했

45) 중국망, 中 공업정보화부 "중국 AI 핵심 산업 규모 6000억 위안 육박", 2024.09.14. (http://korean.china.org.cn/2024-09/14/content_117429852.html)

46) 工业和信息化部, 国家人工智能产业综合标准化体系建设指南(2024版), 2024.06.05.

다. 또한 첨단과학기술 혁신 성과와 표준화 간 연계를 최적화하여 혁신을 지원하고, 산업 분야 간 표준화 협력을 통해 시너지 효과를 발휘하며, 국제 표준화 교류와 협력을 강화하고 중국 기업의 국제 표준화 활동을 장려한다는 계획이다. AI 표준화 체계는 ▲기본·공통 표준 ▲기초 기반 표준 ▲핵심 기술 표준 ▲스마트 제품과 서비스 표준 ▲산업화 신유형 지원 표준 ▲산업 응용 표준 ▲안전·관리 표준의 7개 영역으로 구성된다.

〈표 3-23〉 중국 AI 표준화 체계

구분	내용
기본·공통 표준	• AI의 기본 표준과 프레임워크 표준, 종합 표준을 의미
기초 기반 표준	• 데이터, 연산, 알고리즘 등의 기술 요구사항을 규정해 AI 산업 발전을 위한 기술 인프라를 확고히 하는 표준
핵심 기술 표준	• 기계학습, 자연어 처리, 지능형 음성, 컴퓨터 비전, 생체인식, 지능형 에이전트, 군집 지능 등 AI의 혁신과 응용을 지원하는 표준
스마트 제품과 서비스 표준	• AI 기반 스마트 제품과 서비스 모델 관련 표준
산업화 신유형 지원 표준	• 제조공정의 지능화, 지능형 고객서비스, 공급망 관리 등 다양한 산업 영역에서 AI 기술 적용을 위한 표준
산업 응용 표준	• 스마트 도시, 스마트 농업, 스마트 에너지, 스마트 금융, 스마트 관광 등 AI 기술이 사용되는 다양한 분야에서 필요한 표준
안전·관리 표준	• 데이터 보안, 알고리즘 투명성 등 AI 기술개발과 활용 시 발생할 수 있는 보안과 윤리적 문제에 관한 표준

자료: SPRI 정리, 2024.10.

〔그림 3-3〕 중국 AI 표준화 체계 다이어그램



출처: TTA, 2024.07.

중국 공업정보화부는 2024년 7월 산하 AI표준화기술위원회의 설립 계획도 발표했다.⁴⁷⁾ 이는 산업 발전에 따른 수요에 대응하고 각계의 의견을 청취하기 위함으로, AI 기업, 연구기관과 대학, 협회 등의 전문가를 초빙해 위원회를 구성하고 산하기관은 중국정보통신연구원(CAICT)에 마련할 계획이다. AI표준화기술위원회의 업무 범위는 ▲기본·공통 ▲기초 지원 ▲알고리즘 모델 ▲운영 관리 ▲안전 관리의 5가지로 구성되며, 각각 다음의 내용으로 구성된다.

〈표 3-24〉 중국AI표준화기술위원회의 업무 범위

구분	내용
기본·공통	• AI 용어 정의, 평가와 테스트, 참조 아키텍처, 운영 및 유지관리 표준
기초 기반	• AI 데이터세트, 기본 HW와 SW 플랫폼 관련 표준
알고리즘 모델	• AI 대규모 모델 및 산업용 대규모 모델 관련 표준
운영 및 유지관리	• 대규모 AI 모델의 적용 지침, 응용 성숙도 및 응용 개발관리 표준
안전 관리	• AI 위험 식별 및 예방, 보안 관리, 과학기술 윤리, 데이터 보안, 정보보안 표준

자료: SPRI 정리, 2024.10.

AI표준화기술위원회는 업무의 중점 영역으로 세 가지를 추진할 계획이다. 첫 번째는 업계 공통의 요구사항에 두어 대규모 모델 관련 핵심 표준 개발을 가속화할 예정이며, 두 번째는 표준 적용을 촉진하는 것으로 업계 연맹과 협회, 주요 기업 및 지방정부와 연계를 강화하여 AI 산업 생태계 전반에 표준 홍보와 교육을 진행하고, 기업이 연구개발, 생산, 관리 등의 프로세스 전반에서 표준을 충족하도록 안내한다. 세 번째는 국제표준 제정과 국제협력을 촉진하는 것으로, AI 표준화 분야의 주요 국제표준기구와 교류와 협력을 강화하고 국제표준 제정에 적극 참여하며 국제과 국내 표준의 조정을 촉진할 방침이다.

47) 人民邮电报, 工信部拟筹建人工智能标准化技术委员会, 2024.07.05.
(<http://www.news.cn/tech/20240705/6b8baa03d5404f9c97d8fb9d7a52a950/c.html>)

(3) 국가인터넷정보관공실

가. AI로 생성된 합성 콘텐츠 표시 방법(초안)⁴⁸⁾

중국 국가인터넷정보관공실은 2024년 9월 14일 AI 생성물 식별을 위한 표시를 의무화하는 ‘AI로 생성된 합성 콘텐츠 표시 방법(초안)’을 공개하고 의견수렴을 진행했다. AI로 생성된 합성 콘텐츠는 AI 기술을 사용하여 생성되거나 합성된 텍스트, 이미지, 오디오, 동영상 및 기타 정보를 의미한다.

온라인 플랫폼과 서비스 제공업체는 모든 AI 생성물에 대하여 라벨을 의무적으로 부착해야 하며, 라벨은 사용자가 명확하게 인식할 수 있는 텍스트, 음성, 그래픽 등의 명시적 표식 및 AI 생성물 파일의 메타데이터에 기술적 조치를 통해 추가되어 사용자가 쉽게 인지하지 못하는 표식을 모두 포함한다. 메타데이터에는 합성 콘텐츠, 서비스 제공 작성자 명칭이나 코드, 기타 제작 요소 정보를 표시해야 한다.

AI 생성물이 텍스트인 경우, 텍스트 시작이나 끝 또는 중앙에 문장이나 기호를 추가하는 방식으로 라벨을 부착해야 하며, 오디오 파일 역시 청취자가 들을 수 있도록 시작이나 끝부분, 중간 부분에 음성 설명 등을 추가해야 한다. 이미지나 동영상의 경우, 눈에 잘 띄는 위치에 라벨을 표시하여 사용자가 AI 생성물임을 인식할 수 있도록 해야 한다.

온라인 플랫폼과 서비스 제공업체는 온라인에 게시된 콘텐츠 파일에 AI 생성물임을 나타내는 메타데이터가 없더라도, 콘텐츠 분석 결과 AI로 생성된 흔적이 발견되면 사용자에게 “이 콘텐츠는 AI로 생성된 것으로 의심된다”는 경고를 표시해야 한다. 초안은 어떤 조직이나 개인도 합성 콘텐츠 라벨을 악의적으로 삭제, 위조, 변조 또는 은폐해서는 안 된다고 규정했다.

48) 国家互联网信息办公室关于, 人工智能生成合成内容标识办法(征求意见稿), 2024.09.14.
(https://www.cac.gov.cn/2024-09/14/c_1728000676244628.htm)

(4) 표준화 기구

가. AI 보안 거버넌스 프레임워크⁴⁹⁾

중국사이버보안표준화기술위원회는 2024년 9월 ‘AI 보안 거버넌스 프레임워크(人工智能安全治理框架)’ 1.0 버전을 공개했다. 이 프레임워크는 시진핑 주석이 발표한 2023년 10월 발표한 ‘글로벌 AI 거버넌스 이니셔티브’의 이행을 촉진하고 AI 산업의 건강하고 지속 가능한 발전을 도모하는데 목적이 있다. 프레임워크는 AI 시스템의 수명주기 전반에서 내재된 보안 위험과 응용으로 인한 보안 위험을 분석하고, 기술적 대응조치와 종합적 예방조치를 제시하였다.

〈표 3-25〉 AI 보안 위험 대응조치

분류	주요 위험	대응 조치
내재적 보안 위험	모델·알고리즘 보안 위험	이해 가능성 향상, 편향 제거, 복원력 향상 등
	데이터 보안 위험	데이터 보안 및 개인정보 보호 강화, 데이터 오용 방지
	AI 시스템 보안 위험	투명성 향상, 위험 식별 및 방지 강화, 인프라 시설 보안
응용 보안 위험	사이버공간 보안 위험	보안 방어체계 구축, 데이터 보안 조치
	현실 세계 보안 위험	시스템 제공 범위 설정, AI 시스템의 최종 사용목적 추적 강화
	인지적 보안 위험	예상되지 않은 출력 방지, 오용 방지
	윤리적 보안 위험	차별 방지, 감독관리 조치 보완

출처: 중국사이버보안표준화기술위원회, SPRI 정리, 2024.09.

프레임워크는 보안 관리 시 관리자, 개발자, 서비스 제공자, 핵심 영역(정부, 주요기반시설 등) 사용자, 일반 사용자 등 각 주체의 책임과 의무를 규정하고 주체별 행동 지침을 제시하였다.

- (관리자) 정부 부처와 산업 협회 등 관리 주체는 관련 법률과 정책을 통해 규제를 강화하고, AI 서비스의 분류 및 등급 관리, 추적 관리 제도, 데이터 보안 및 개인정보 보호 규범 등을 마련하여 협력적 거버넌스 체계를 구축
- (개발자) 요구 분석, 프로젝트 입안, 모델 설계·개발, 훈련 데이터 선택 등 주요 단계에서 데이터 보안 및 개인정보 보호를 우선시하며, 모델과 알고리즘의 잠재적 편향을 사전 평가

49) 全国网络安全标准化技术委员会, 《人工智能安全治理框架》1.0版, 2024.09.09. (<https://www.tc260.org.cn/front/postDetail.html?id=20240909102807>)

- (서비스 제공자) AI 기술의 성능 범위와 적용 조건을 명확히 이해하고 점검하며, 사용자에게 대한 투명성을 보장하고 시스템 오류나 보안 사고 발생 시 신속히 대응하고 관련 부처에 보고
- (핵심 영역 사용자) 공공안전과 민생에 직접 영향을 미치는 핵심 영역에서 AI 사용 시에는 높은 수준의 안전 의식 및 위험발견과 처리 능력을 요구하며, 이를 위해 엄격한 보안 관리감독 체계를 구축
- (일반 사용자) 일반 대중이 AI의 안전 위험을 인식할 수 있도록 제품 기능과 한계를 명확히 전달하고 AI 기술을 안전하고 합리적으로 사용하기 위한 지침 제공

나. 생성 AI 보안 기본요구사항 표준

중국사이버보안표준화기술위원회는 2024년 5월 23일 ‘정보보안기술: 생성 AI 서비스 보안을 위한 기본 요구사항(网络安全技术 生成式人工智能服务安全基本要求)’ 표준 초안을 공개하고 2024년 7월 22일까지 의견수렴을 진행하였다. 동 문서는 교육 데이터와 모델의 안전성과 보안 조치 등에서 생성 AI 서비스의 기본 요구사항을 명시하고 안전 평가를 위한 기준점을 제시하였다.

교육 데이터 보안에서는 데이터 수집 전 보안 평가의 실시 및 데이터 출처 확인, 훈련 데이터 내용의 필터링, 지식재산권 보호 등을 요구하였다. 데이터 소스의 수집 전 보안 평가를 수행하여 데이터에 불법적이고 유해한 정보가 5% 이상 포함되면, 해당 소스의 데이터 수집을 금지하였다. 학습 데이터 소스의 다양성을 개선하여, 중국어와 영어 등 언어별 및 텍스트, 이미지, 오디오, 비디오 등 데이터 유형 별 다양한 소스의 확보를 요구하였다. 오픈 소스 데이터 사용 시 오픈소스 라이선스 계약이나 관련 승인 서류를 갖춰야 하며, 상용 교육 데이터 사용 시에는 법적 구속력 있는 계약서가 필요하다고 규정하였다. 모든 데이터는 학습에 사용하기 전 불법적이고 유해한 정보를 제거하기 위한 필터링을 거쳐야 하며, 교육 데이터에 대한 지식재산권 관리 전략을 수립하여야 한다.

모델 보안 측면에서는 모델 훈련과 모델 출력, 모델 모니터링에 대한 보안 요구사항을 제시하였다. 모델 훈련 과정에서 생성된 콘텐츠의 안전성은 결과물 품질을 평가하는 주요 고려사항이 되어야 하며, 사용된 개발 프레임워크와 코드에 대해 정기적 보안 감사를 실시하고 보안 취약점을 파악해 완화해야 한다. 모델이 생성한 결과물의 정확성 측면에서 사용자 입력 의도에 부합하는 수준을 향상하고, 잘못된 콘텐츠를 줄이기 위한 기술적 조치를 요구하였다. 모델의 답변 시 불법적이고 유해한 정보 생성을 유도하는 질문은 거부해야 하며,

모델에 입력되는 콘텐츠에서 백도어 공격, 데이터 도용 등의 악의적 공격을 지속적으로 모니터링하고 보안 문제가 발견되면 적시에 처리해야 한다.

그 밖에도 주요기반시설, 의료정보 서비스, 심리상담, 금융 정보 서비스 등 중요 서비스에는 위험 수준과 사용 시나리오에 적합한 보안 조치를 요구하였으며, 미성년자에 대한 서비스 제공 시에는 보호자가 중독방지 조치를 설정할 수 있도록 하고, 미성년자가 감당할 수 없는 유료 서비스를 제공해서는 안 된다고 규정하였다.

2) 지방정부

〈표 3-26〉 2024년 중국 주요 지방정부의 AI 산업 정책

지역	시기	정책
산시성	2024.04.	1,000억 위안 규모의 AI 산업 혁신 클러스터 육성 행동계획
광둥성	2024.05.	광둥성 AI와 천 개 산업 융합 발전을 위한 조치
청두시	2024.06.	AI 산업 고품질 발전 3개년 계획
베이징시	2024.07.	베이징시 AI 플러스 행동계획(2024~2025)
후난성	2024.09.	후난성 AI 산업 발전 3개년 행동계획(2024~2026)

자료: SPRI 정리, 2024.10.

(1) 산시성

산시성 발전개혁위원회는 2024년 4월 ‘1,000억 위안 규모(한화 약 19조 730억 원)의 AI 산업 혁신 클러스터 육성 행동계획’을 발표하고 3단계 목표를 제시했다.⁵⁰⁾

〈표 3-27〉 중국 산시성의 AI 산업 혁신 클러스터 육성 3단계 목표

구분	목표
1단계(~2025년)	• AI 산업 규모 300억 위안(한화 약 5조 7천억 원) 달성, 특정 분야 경쟁력을 갖춘 AI 산업기지 건설
2단계(~2030년)	• AI 산업 규모 600억 위안(한화 약 11조 4천억 원) 달성, 국가 영향력을 가진 AI 산업 혁신 클러스터 구축
3단계(~2035년)	• AI 산업 규모 1,000억 위안(한화 약 19조 730억 원) 달성, 스마트 에너지·스마트 제조·스마트 자동차 등 ‘AI+’ 다운스트림 산업 규모 5,000억 위안 달성

출처: 산시성 발전개혁위원회, SPRI 정리, 2024.04.

50) 科技日报, 培育新质生产力在行动 | 培育千亿元级人工智能产业创新集群 陕西发布行动计划推进五大工程, 2024.04.08.

중점 임무로는 ▲기초이론과 핵심기술 연구 ▲혁신 플랫폼 구축 ▲제품 서비스 혁신 ▲시범사업 구축 ▲산업 생태계 최적화의 5개 공정 건설을 추진하기로 했다.

〈표 3-28〉 산시성 AI 산업혁신 클러스터 육성계획의 5대 중점 임무

중점 임무	정책
기초이론과 핵심기술 연구	<ul style="list-style-type: none"> 기초이론 병목 돌파, 뇌신경망 모델 등 핵심 이론연구, 학제간 연구 지능 감지 처리, 자연어 처리, 지능형 네트워크 연결 자동차 기술, 무인 시스템 기술 등 핵심기술 연구
혁신 플랫폼 구축	<ul style="list-style-type: none"> 시안교통대학 휴먼-머신 혼합중간지능 중점실험실 등 국제 일류 AI 연구거점 구축, 산시성 지능로봇 중점실험실 건설 중점기술 난관 돌파 위해 시안교통대학 시각정보처리·응용 국가공정연구 센터, 시베이 공업대학 드론시스템 국가공학연구센터 등 강화 산업혁신 플랫폼으로 산시성 스마트 커넥티드카 산협혁신센터 건설, 시안안전자과과대학 국가 AI산학융합혁신 플랫폼 구축
제품 서비스 혁신	<ul style="list-style-type: none"> 대형 모델 구축, AI 하드웨어 제품 혁신, AI 소프트웨어 서비스 능력 향상, 지능형 로봇 산업 육성, 스마트 드론 산업 육성, 스마트 커넥티드카 산업 육성
시범사업 구축	<ul style="list-style-type: none"> 시와 에너지, 화학, 제조, 농업, 교통, 문화 관광, 물류, 금융, 가정, 의료, 교육 등 분야의 융합 발전을 통해 AI 응용 확대
산업 생태계 최적화	<ul style="list-style-type: none"> 시안 미래시컴퓨팅센터, 국가슈퍼컴퓨팅 시안센터 등 신형 인프라 건설, 공공데이터 자원 플랫폼 구축 등 AI 인프라 자원 확보 기업 주도의 산학연 중대 연구개발 프로젝트에 고급 인재 유치, AI+ 융합형 인재 양성 모델 등 AI 인재 양성 추진

자료: SPRI 정리, 2024.10.

(2) 광동성

광동성 정부는 중앙정부의 ‘AI 플러스’ 정책에 대응해 2024년 5월 ‘AI와 천 개 산업 융합 발전을 위한 조치’ (이하 조치)⁵¹⁾를 발표했다. 조치는 2025년까지 광동성의 컴퓨팅 연산 능력 40EFLOPS⁵²⁾ 돌파 및 AI 핵심산업 규모 3,000억 위안(한화 약 577조 1,980억 원) 이상 달성, 2027년까지 컴퓨팅 연산 능력 60EFLOPS 초과 및 휴대폰, 컴퓨터, 가전, 로봇 등 8대 분야에서 100종 이상의 스마트 단말 개발과 AI 핵심산업 규모 4,400억 위안(한화 약 83조 9,000억 원) 이상 달성을 목표로 제시하였다. 3대 중점 임무로는 ▲AI 산업 기반 강화 ▲스마트 단말 제품 개발 ▲AI 융합 응용 확대를 추진하기로 하였다.

51) 广东省人民政府, 关于人工智能赋能千行百业若干措施的通知, 2024.06.06.

(https://www.gd.gov.cn/xxts/content/post_4436503.html)

52) 1초에 100경 번의 부동소수점 연산을 수행하는 능력

광둥성 정부는 AI 산업 기반 강화를 위해 가전, 보안 모니터링, 의료기기 등의 분야에서 고성능 저전력 칩 개발과 생산을 강화하고 2027년까지 AI 칩 생태계 구축 및 자체개발 AI 소프트웨어 보급률 50% 달성을 추진할 계획이다. 또한 고품질 중국어 데이터셋 구축과 공공데이터와 민간 데이터의 효율적 융합으로 2027년까지 50개 이상의 고품질 산업 데이터셋을 구축하고, 100개의 알고리즘 제품과 AI 오픈소스 서비스 시스템을 구축한다는 계획이다.

스마트 단말 제품 개발 측면에서는 지능형 로봇 두뇌, 로봇 팔다리, 로봇 몸체와 범용 제품 등의 연구개발과 생산을 가속화하여 2027년까지 지능형 로봇 산업의 연간 매출액 900억 위안을 달성하는 한편, AI 스마트폰 생산량 1억 대 이상, AI PC 생산량 3천만 대 이상을 달성한다는 계획을 수립하였다.

AI 융합 응용 확대 측면에서는 AI 알고리즘을 활용한 수확량 예측, 기상 예보, 시장 분석 등의 응용을 확대하고 지능형 철도와 도로, 수로, 항구 등의 시범 프로젝트를 실시하며, 스마트 교실과 캠퍼스 구축 및 스마트 노인 요양 서비스 시스템을 구축할 계획이다. 또한 광둥성 스마트 행정 서비스 플랫폼을 구축하여 24시간 온라인 질문 응답과 검색 서비스를 제공하고 생태 환경을 위한 원격 감시, 환경 모니터링 등을 추진하기로 하였다.

(3) 청두시

청두시는 2024년 6월 ‘AI 산업 고품질 발전 3개년 계획’⁵³⁾을 발표하고 2026년까지 제조업, 의료 등 산업별 대규모 모델 50개 이상 개발과 AI 혁신 플랫폼 40개 구축, AI 기업 육성, AI+ 응용 시나리오 10개 구축, 500개 이상의 시범 제품 출시 등의 목표를 수립하는 한편, 핵심 기술 연구, 컴퓨팅 파워 지원 강화, 데이터 공급 품질 향상 등 6대 분야에서 총 19개 중점 과제를 제시했다.

53) 成都市出台了《成都市人工智能产业高质量发展三年行动计划（2024—2026年）》，2024.06.20.
https://cdst.chengdu.gov.cn/cdkxjsj/c108732/2024-06/20/content_d6203bb3ecf34d24aa126aa44879f29d.shtml

〈표 3-29〉 청두시 AI 산업 고품질 발전 3개년 계획의 중점 임무

중점 임무	정책
핵심 기술 연구	<ul style="list-style-type: none"> AI 기초 및 첨단 연구 프로젝트 진행 및 제조·의료·금융·교통·과학 연구 등 분야에서 독자적 기술력이나 지식재산권을 보유한 산업 대형모델을 선정하여 서비스 지원 '지원센터+기술혁신센터 연합' 모델로 AI 서비스 플랫폼을 구축하고 스마트의료, 지능형 감지, 로봇 등 분야에서 산업 핵심 기술 연구개발
컴퓨팅 파워 지원 강화	<ul style="list-style-type: none"> 연구·설계·생산·판매 전 과정 제품 생산력을 갖춘 컴퓨팅 파워 산업 클러스터를 구축하고 400G급 광전송 네트워크 상용화를 추진 산업 혁신 플랫폼 구축을 지원하고 국가 데이터 선도 기업을 유치
데이터 공급 품질 향상	<ul style="list-style-type: none"> 선도 데이터 기업과 연구기관을 중심으로 스마트 제조, 스마트 도시, 스마트 의료 등 분야 고품질 데이터 서비스를 제공 공공데이터 개방을 촉진하고 과학연구기관 및 선도기업이 협력해 산업 공통 데이터 자원 라이브러리 구축
클러스터 개발 촉진	<ul style="list-style-type: none"> 2026년까지 AI 산업 선도기업 20개, 전정특신(專精特新, 전문화, 정밀화, 특색화, 참신화) 기업 100개, AI 중소기업 1,000개 이상 육성 및 유치
시나리오 활용 추진	<ul style="list-style-type: none"> 신형 산업, 서비스업, 농업 개발을 중심으로 디지털화 작업장, 스마트 공장, 스마트 농장 등 50개 이상 'AI+생산' 혁신 활용 시나리오 개발 2026년까지 시민 건강, 교육, 주거 등 분야와 스마트 의료, 스마트 교육 등 50개 이상의 'AI+생활' 혁신 시나리오 개발
첨단 산업 자원 유치	<ul style="list-style-type: none"> 국내외 고급 인재를 유치하고 시급 플랫폼 펀드를 통한 AI 산업 프로젝트 투자를 확대하며 기업의 자금 조달 경로를 호가장 국가급 AI 행사, AI 산업 CEO 대회 등 브랜드 행사를 적극 개최

자료: SPRI 정리, 2024.10.

(4) 베이징시

중국 베이징시 발전개혁위원회는 2024년 7월 국가 AI 플러스 전략을 실현하고 AI 산업 생태계 육성을 위한 '베이징시 AI 플러스 행동계획(2024~2025)(이하 계획)'을 발표했다. 중국 정부는 2024년 3월 전국인민대표대회에서 발표했으며 2024년 업무보고에서 'AI 플러스 행동' 개념을 처음 제시했으며, 베이징시의 계획은 이를 구체화한 것이다. 계획은 2025년 말까지 세계 최고 수준의 5대 벤치마크 응용 프로젝트와 국내를 선도하는 수준의 10개 시범 응용 프로젝트 시행, 선진 사용과 자주적 통제가 가능한 기초 대형 제품 3~5개 개발 등을 목표로 제시했다. 베이징시는 교육, 의료, 문화, 교통, 금융, 사법 등 다양한 산업 분야에서 AI 기업의 세분화된 응용 시나리오 개발을 지원하는 한편, AI 혁신을 지원하기 위한 공동 R&D 플랫폼도 구축할 계획이다.

계획은 먼저 5대 벤치마크 응용 프로젝트 영역으로 ▲로봇 ▲교육 ▲의료 ▲문화 ▲교통의 5개 분야를 선정했다. 로봇 영역에서는 현장 요구사항을 바탕으로 체화 지능(Embodied Intelligence)을 통합한 로봇을 개발하고 교육 영역에서는 베이징시의 고품질 교육 사례와 지식 등을 수집해 대형모델 플랫폼 구축에 주력할 계획이다. 의료 영역에서는 의료 감독체계의 혁신과 병원정보시스템의 지능적 업그레이드를 결합한 의료모델 플랫폼을 구축하고 문화 영역에서는 역사와 문화, 관광 자원을 바탕으로 AI와 문화관광의 심층 통합을 추진한다. 교통 영역에서는 대형모델을 활용해 지능적이고 효율적이며 안전한 도시 교통망을 구축하고 교통 혼잡을 예측해 신호등 제어를 최적화할 방침이다.

10개 시범 응용 프로젝트로는 ▲과학연구와 탐사 ▲정부 서비스 ▲산업 인텔리전스 ▲금융 관리 ▲공간 컴퓨팅 ▲디지털 마케팅 ▲사법 서비스 ▲방송 미디어 ▲전력 보호 ▲콘텐츠 보안을 세부 영역으로 제시했다. 시범 응용 프로젝트를 통해 산업 당국과 기업, 대형모델 개발 기업 간 연계를 지원하고 산업 응용 모범 프로젝트의 육성에 중점을 두며, 대형모델의 산업 응용을 위한 인프라와 공공 플랫폼 구축을 개선할 계획이다.

<표 3-30> 베이징시 10대 AI 시범 응용 프로젝트

중점 임무	정책
과학 연구와 탐사	• 신약 개발, 유전자 염기서열 분석, 신소재 분석, 기상 분석 등 기초과학 분야에서 지능형 제품 개발
정부 서비스	• 정부 상담 관련 불만 사항에 대한 지능형 처리에 필요한 대형 모델 기술의 시범 응용 추진
산업 인텔리전스	• 대형 모델의 자연어 처리, 데이터 분석과 처리, 딥러닝 등의 기능을 활용한 산업 R&D 및 생산 프로세스 개발
금융 관리	• 은행 등 금융기관의 신용대출 조사 보고서 및 신용 검토 보고서 생성을 지원하는 대형 AI 모델 개발
공간 컴퓨팅	• 인터넷 지도 데이터, 사물인터넷 센서 데이터 및 위치 서비스 데이터 등이 통합된 도시 공간 운영 데이터 프레임워크에 기초해 지역 대형모델 개발
디지털 마케팅	• 디지털 휴먼 라이브 스트리밍, 고객 서비스, 교육관리 등에 사용할 디지털 휴먼 인터랙티브 지식 인프라와 도구 개발
사법 서비스	• 사법 대형모델 구축 및 지능형 법률 컨설팅 서비스 등의 통합으로 사법 중재 및 법률 서비스의 지능화 추진
방송 서비스	• 텍스트 생성, 동영상 생성, 영상 특수효과 처리 등의 기능을 보유한 대형 모델을 활용해 AI 콘텐츠 창작 최적화
전력 보호	• 지능형 전력 검사, 전력 최적화와 스케줄링, 고장 진단 및 유지 보수 등에 필요한 전력관리 대형 모델 개발
콘텐츠 보안	• 콘텐츠 보안을 위한 대형 모델을 개발하고 텍스트, 이미지, 오디오, 동영상 등 멀티모달 콘텐츠에 대한 인식 정확도 향상

자료: SPRI 정리, 2024.10.

(5) 후난성

후난성 공업정보화청은 2024년 9월 ‘후난성 AI 산업 발전 3개년 행동계획(2024-2026년)’⁵⁴⁾을 발표하고 2026년까지 AI 산업 규모를 1천억 위안(한화 약 18조 원) 이상으로 확대하겠다는 목표를 제시하였다. 계획에 따르면 2026년까지 후난성의 AI 연산 능력을 3,600PFlops로 끌어올리고 전국적으로 영향력 있고 산업 생태계를 주도할 기업 2~3개와 중소기업을 다수 육성할 계획이다. 또한 성급 AI 산업단지 5개 조성, 특화 초거대 AI 모델 50개 개발, 혁신 제품 100개와 대표 응용 사례 200개 개발이라는 목표도 수립하였다.

〈표 3-31〉 청두시 AI 산업 고품질 발전 3개년 계획의 중점 임무

중점 임무	정책
기본 지원 강화	<ul style="list-style-type: none"> 지능형 컴퓨팅 능력의 개발 가속화 및 지능형 소프트웨어 개발, 지능형 칩과 센서 개발, 고품질 데이터셋 구축
핵심 기술의 획기적 발전	<ul style="list-style-type: none"> 대형 모델 기술 개발 가속화, 응용 기술의 적극 개발
스마트 제품 개발	<ul style="list-style-type: none"> 스마트 로봇, 스마트 배송 차량, 스마트 모바일 단말, 디지털 휴먼 개발
기업의 지적 역량 강화 캠페인 추진	<ul style="list-style-type: none"> 제조 프로세스 전반의 지능화 추진, 핵심 산업의 지능화 수준을 향상하고 AI 플러스를 심도 있게 추진, 안전 수준의 향상

자료: SPRI 정리, 2024.10.

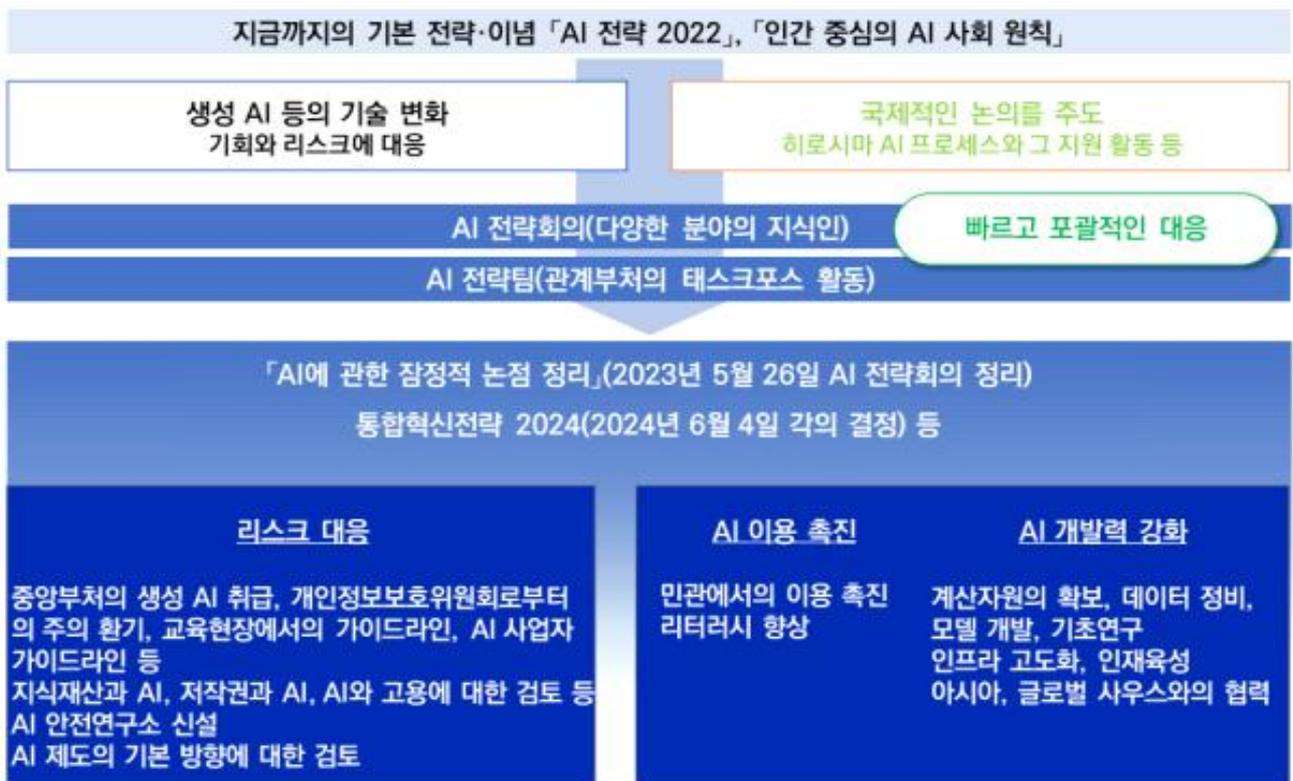
54) 湖南省工业和信息化厅, 《湖南省人工智能产业发展三年行动计划(2024-2026年)》, 2024.09.04. (https://www.hunan.gov.cn/zqt/zcsd/202409/t20240905_33447854.html)

제4절 일본

1. 그간의 인공지능 정책 추진 경과

4차 산업혁명의 핵심 영역으로 일컬어지는 인공지능(AI) 분야의 주도권 확보를 위해 선진국들의 경쟁이 치열해짐에 따라 일본 정부도 AI 정책을 적극 수립, 추진하고 있다.

[그림 3-4] 일본 AI 정책 추진도



출처: 内閣府, AI政策の現状と制度課題, 2024.08.02.

일본은 AI를 자국 성장 전략의 핵심으로 설정한 아베 정부가 2018년 9월 설치한 ‘통합혁신전략추진회의’에서 AI 종합 전략의 수립을 위해 ‘AI 전략 패키지’ 검토에 착수했으며, 2019년 6월 국가 AI 종합 전략인 ‘AI 전략 2019’를 발표하였다. 2021년 6월에는 지금까지의 대응 진척 상황 및 향후 대응 방향을 정리해 ‘AI 전략 2021’로 개정, AI 대응을 지속·추진하였다.

2022년에 수립된 ‘AI 전략 2022’에서는 ‘AI의 사회 구현 추진’에 중점을 두고 2021년도 전략에 추가된 ‘임박한 위기 대응’인 코로나19 팬데믹, 대규모 자연재해 등과 같은 보다 명백해지는 많은 리스크 요인 등을 반영해 구체적인 목표 등을 검토, 대폭 개정하였고, 2022년 양대 전략목표로 ▲사회 구현 내실화를 위한 새로운 목표 설정·추진 ▲팬데믹, 대규모 재해 등의 임박한 위기 대응책 구체화를 제시하였다.

2023년에는 최근 ‘챗GPT(ChatGPT)’ 등 AI 기술의 이용이 급속히 확산되면서 기업·행정 업무에서의 작업 효율화가 기대되는 한편, 저작권·개인정보 침해 등 부정적 우려도 지적되며, 관련 규제 방식에 대한 재검토가 필요한 상황이다. 이에 2023년 5월 생성 AI를 포함한 AI에 대한 국가 전략을 수립하고 정책 방향성을 제시하는 AI 정책을 총괄하는 자문기구인 ‘AI 전략회의’를 신설했다.⁵⁵⁾ 또한 현재의 생성 AI를 중심으로 한 AI 논점을 정리한 ‘AI에 관한 잠정적 논점 정리’⁵⁶⁾를 발표하고, 일본이 중점 추진해야 할 구체적인 AI 전략 방향성으로 ▲리스크 대응 ▲AI 이용 촉진 ▲AI 개발력 강화 등을 제시하고, AI 관련 경쟁력 강화를 위해서 해당 분야를 중점 지원해 나갈 방침이다. AI 전략회의에서 정리한 논점은 ‘제6기 과학기술·혁신기본계획(2021~25)’의 연차 전략인 ‘통합혁신전략 2023’ (2023.6)⁵⁷⁾, ‘통합혁신전략 2024’ (2024.6)⁵⁸⁾의 AI 추진 방안에 반영하였고, 향후 정부 재정 운영과 개혁 등 정책에 반영할 계획이다.

2023년 5월 개최된 G7 히로시마 정상회의(5.19~21)에서는 생성 AI에 관한 국제적인 규칙 검토를 위한 ‘히로시마 AI 프로세스’를 비롯해 AI의 개발·공표 전 리스크의 특정·경감, AI의 성능에 관한 정보공유라고 하는 다양한 국제 지침을 명문화했다.⁵⁹⁾

55) 일본은 지금까지 ‘AI 전략 2022’, ‘인간 중심의 AI 사회 원칙’ 등 정부의 AI 기본 전략·이념을 제시해 왔지만, 최근의 생성 AI 등 기술 변화와 국제적인 논의에 대응하기 위해 AI 정책을 총괄하는 자문기구인 ‘AI 전략회의’를 신설해, 생성 AI ‘챗GPT’를 포함한 AI에 대한 국가 전략을 수립하고 정책 방향성을 제시하는 새로운 사령탑 역할 수행

56) 内閣府, AI に関する暫定的な論点整理, 2023.5.26

(https://www8.cao.go.jp/cstp/ai/ronnen_honbun.pdf)

57) 内閣府, 統合イノベーション戦略 2023, 2023.6.9

(https://www8.cao.go.jp/cstp/tougosenryaku/togo2023_honbun.pdf)

58) 内閣府, 統合イノベーション戦略 2024, 2023.6.4

(https://www8.cao.go.jp/cstp/tougosenryaku/togo2024_zentai.pdf)

59) 経団連, AI政策・制度検討の動向, 2024.9.26

(https://www.keidanren.or.jp/journal/times/2024/0926_08.html)

[그림 3-5] AI 전략회의의 주요 활동

AI에 관한 잠정적 논점 정리 (AI 전략회의, 2023.5.26)	
<p>【개인정보보호】 OpenAI에 대한 주의 환기 (개인정보보호위원회, 2023.6.1) 생성 AI 서비스 이용에 관한 주의 환기 등 (개인정보보호위원회, 2023.6.2)</p> <p>【AI와 지적재산권과의 관계】 문화심의회 저작권분과회 법제도시위원회 「AI와 저작권에 관한 개념에 대하여」 (문화청, 2024.3.15) AI와 저작권의 개념에 대해 정리. 주지·개발. 관계자의 상호 이해 촉진 등 AI 시대의 지식재산 검토회 중간정리 (지식재산전략추진사무국, 2024.5.28) 법, 기술, 계약을 통한 대응</p> <p>【허위·오정보 등】 디지털 공간에서의 정보 유통 건전성 확보의 기본 방향에 관한 검토회 정리(안) 공표·의견 수렴 중 (총무성, 2024.7.20~)</p> <p>【고용에 대한 영향】 고용정책연구회 중간정리 「신기술이 고용에 미치는 영향에 대하여」 (후생노동성, 2023.12.21) 생산성 향상의 한편에서 업무 내용 변화, 노사 커뮤니케이션 심화, 경력 형성 지원등이 중요</p> <p>【가이드라인 이행력 확보 등】 가이드라인 등 이행력 확보·AI 이용 촉진에 관한 조사 (내각부, 2023.11~) 각 업종 등에서의 리스크나 대응 등을 정리 AI 안전연구소 신설 (내각부 등, 2024.2.14) 해의 동종 기관과도 제휴해, 안전성 확보를 향한 조사 연구, 기준 작성 등 추진 AI사업자 가이드라인 ver.1.0 (경제산업성·총무성, 2024.4.19) (참고) 부정경쟁방지법 「비밀정보 보호 핸드북」(2024.2) 생성 AI에 관해 기재</p>	<p>【정부에서의 이용】 ChatGPT 등의 생성 AI의 업무 이용에 관한 합의 (디지털사회추진회의 간사회) 제1판(2023.5.8)에서 기밀성 1 정보를 시험적 이용 제2판(2023.9.15)에서는 기밀성 2 정보도 시험적 이용</p> <p>【교육 분야에서의 이용】 초·중등 교육단계에서의 생성 AI 이용에 관한 잠정적 가이드라인 개정 검토 (문부과학성, 2024.7.25)</p> <p>【인재 육성】 디지털 스킬 표준의 보급, ITSS(IT 스킬 표준) 레벨 3 상당의 교육 훈련을 인정 제도 대상으로 지정</p> <p>【공공기관에 의한 데이터 정비】 NICT에 의한 일본어 데이터 정비 정부 데이터의 AI 학습에 대한 제공 실행계획 ver.1.0 (내각부, 2023.11.7) 정부 데이터에 관한 수요 조사, 데이터 형식 변환방법등 검토</p> <p>【계산 자원 정비】 사용 계산 자원에 대해 산출연 ABC 확충과 민간 정비등 지원</p> <p>【모델 개발 지원】 스타트업 등에 의한 AI 모델 개발을 촉진을 지원</p> <p>【기초연구】 NE에서 생성 AI 모델의 투명성·신뢰성 확보등 연구개발을 시작 이화학연구소에서 과학연구용 AI 기반 모델 개발에 착수하는 동시에 2024년 4월 AI for Science의 일미 제휴를 마친</p>

출처: 内閣府, AI政策の現状と制度課題, 2024.08.02.

2024년에도 ‘AI에 관한 잠정적 논점 정리’의 AI 전략 방향에 따라 가이드라인 등을 정비 중이며, 2024년 4월에 경제산업성·총무성의 합동 검토회, 의견 수렴 등을 거쳐, 기존 가이드라인들을 통합·업데이트해, 광범위한 통일적이고 알기 쉬운 AI 사업자용 가이드라인을 검토, ‘AI 사업자 가이드라인(제1.0판)’⁶⁰⁾을 정리하였다. 2024년 7월에는 방위성이 기술 발전에 따른 전투 방식의 변화에 대응하고 인구 감소에 따른 인적 효율화를 도모하기 위해 AI 활용을 추진하기 위한 기본방침을 처음으로 수립, 발표했다. 2024년 8월에는 AI 전략회의 산하에 ‘AI 제도연구회’를 설치해, 가을 중간정리를 목표로 일본 AI 법규제의 기본 방향을 검토 중이다. 기시다 총리는 ①AI의 안전성 확보와 경쟁력 강화의 양립 ②기술 변화에 대응할 수 있는 유연한 제도 ③국제적인 지침에 의거 ④AI에 대한 정부의 적절한 조달·이용 등 4대 원칙하에 검토하도록 지시했다.

60) 経済産業省; 総務省, AI事業者ガイドライン(第1.0版), 2024.4.19
(<https://www.meti.go.jp/press/2024/04/20240419004/20240419004-1.pdf>)

한편 2024년 2월에는 경제산업성 산하 IT 정책 실시기관인 정보처리추진기구(IPA) 내에 AI의 안전성을 검토하는 기관인 ‘AI 안전연구소(AISI)’를 설립하고, 내각부나 경제산업성 등과 협력해 AI의 안전성 평가 방법 검토 등을 실시하고 있다.

〈표 3-32〉 일본 정부의 AI 정책 추진 경과

정책	담당 기관	주요 내용
일본재흥전략 2016 (2016.4)	내각부	<ul style="list-style-type: none"> 제4차 산업혁명 실현의 핵심 도구의 하나로 IoT, 빅데이터, 로봇과 함께 AI에 주목
AI 산업화 로드맵 (2017.3)	AI기술전략회의	<ul style="list-style-type: none"> AI 기술수준 변화에 따라 2030년까지 총 3단계 발전 로드맵을 제시 ①생산성 ②건강·의료·간호 ③공간 이동 ④정보 보안 등 총 4개 분야를 AI 응용 우선 분야로 선정
신산업구조비전 (2017.5)	국가발전개혁위원회, 과학기술부 등	<ul style="list-style-type: none"> AI 미래상 조망 및 4차 산업혁명 대비 7대 전략을 제시 전략적 추진 분야로 ▲건강 증진 ▲차세대 모빌리티 ▲스마트 생산·보안·물류·소매·농업 ▲스마트 생활 등을 지정
AI 전략 2019 (2019.6)	통합혁신전략추진회의	<ul style="list-style-type: none"> ‘AI 도입을 통해 포용성·지속가능성이 실현되는 사회로의 변화’를 목표로 설정 4대 전략목표와 7대 분야에서의 전략을 제시
AI 전략 2019 후속조치 (2020.6)	수상관저 AI 전략실행회의	<ul style="list-style-type: none"> 2019년 6월에 수립한 ‘AI 전략 2019’의 실시 첫째 각 부처의 전략 수립 이후 진행 상황과 향후 대응 방향을 정리
AI 전략 2021 (2021.6)	통합혁신전략추진회의	<ul style="list-style-type: none"> ‘AI 전략 2019’의 진척 상황 및 ‘전략의 후속조치’ 실시를 바탕으로 각 분야의 대응 추진 상황을 점검하고, ‘AI 전략 2021’을 수립
AI 전략 2022 (2022.4)	통합혁신전략추진회의	<ul style="list-style-type: none"> 국내외 정세·기술 동향을 고려해 AI의 사회 구현을 촉진하는 방향으로 기존 전략(AI 전략 2021)을 전면 개정, 코로나19 등 리스크 요인을 반영 목표는 ▲AI 사회 적용 추진 ▲임박한 위기 대응
AI에 관한 잠정적 논점 정리 (2023.5)	AI 전략회의	<ul style="list-style-type: none"> AI에 대한 정책 방향성을 논의하는 ‘AI 전략회의’를 신설하고, 현재의 생성 AI를 중심으로 한 AI 논점을 정리
AI 사업자 가이드라인(제1.0판) (2024.4)	총무성·경제산업성	<ul style="list-style-type: none"> AI 보급에 대응하기 위한 지침으로서 AI 개발자, 제공자, 이용자 등 3자의 입장에서, 기본이념, 지침, 실천의 방법을 분명히 해 AI의 안전한 활용을 목표로 수립
방위성 AI 활용 추진 기본방침 (2024.7)	방위성	<ul style="list-style-type: none"> 기술 발전에 따른 전투 방식의 변화에 대응하고 인구 감소에 따른 인적 효율화를 도모

출처: 일본 AI 전략 자료 각 연도

2. 2024년 인공지능 정책 동향

1) ‘AI 안전연구소(AISD)’ 설립⁶¹⁾

(1) 개요

일본 정부는 AI 안전성에 대한 국제적인 관심이 높아짐에 따라 AI의 안전성 평가 방법의 검토 등을 실시하는 기관으로서 미국·영국과 같은 ‘AI 안전연구소(AISD)’를 2024년 2월에 설립했다. 이 기관은 내각부를 비롯한 관계부처⁶²⁾, 관계기관⁶³⁾의 협력 하에 경제산업성 산하 정보처리추진기구(IPA)에 설치되어 여러 나라의 기관과도 제휴하여 AI의 안전성 평가에 관한 기준이나 방법의 검토 등을 추진하고 있다.

(2) 주요 업무 내용

일본 AI 안전연구소의 주요 업무는 다음과 같다.

- 안전성 평가에 관한 조사, 기준 등의 검토
- 안전성 평가의 실시 방법에 관한 검토
- 타국의 관계기관(미영의 AI 안전연구소 등)과의 국제 제휴에 관한 업무

(3) ‘AI 안전에 관한 평가 관점 가이드’ 발표⁶⁴⁾

AI 안전연구소는 2024년 9월 ‘AI 안전에 관한 평가 관점 가이드’를 발표했다. AI 안전성을 평가하는 관점으로 10개 항목을 제시하고, AI 개발·제공 사업자가 시스템 라이프사이클에 따른 평가를 반복하는 것이 중요하다고 제언했다. AI 시스템 개발·제공 관리자가 이 안내서를 참조함으로써 AI 안전 평가의 요점을 확인할 수 있고, AI 안전을 고려한 AI 시스템 개발·제공을 할 수 있다.

61) 内閣府, AIセーフティ・インスティテュートの設立について, 2024.2.14
(<https://www8.cao.go.jp/cstp/stmain/20240214.html>)

62) 내각부(과학기술혁신 추진 사무국), 내각관방, 경찰청, 디지털청, 총무성, 외무성, 문부과학성, 경제산업성, 방위성

63) 정보통신연구기구, 이화학연구소, 국립정보학연구소, 산업기술종합연구소

64) AISI, AIセーフティに関する評価観点ガイドの公開, 2024.9.18
(https://aisi.go.jp/effort/effort_information/240918_2/index.html)

이 가이드에서는 AI 안전 관점으로서 ‘AI 사업자 가이드라인’의 ‘C. 공통 지침’에서 각 주체가 추진하는 사항 중, ▲인간 중심 ▲안전성 ▲공평성 ▲개인정보보호 ▲보안 확보 ▲투명성을 중요 요소로 하고 있으며, 이들 6개의 중요 요소에 관련된 AI 안전성 평가 관점을, 현재의 기술적 조류를 근거로 도출했다. 도출된 평가 관점은 ①유해 정보의 출력 제어 ②허위·오정보의 출력·유도 방지 ③공평성과 포용성 ④고위험 이용·목적 외 이용에 대한 대처 ⑤개인정보보호 ⑥보안 확보 ⑦설명 가능성 ⑧견고성 ⑨데이터 품질 ⑩검증 가능성 등 10개 항목으로, 2024년 4월에 경제산업성과 총무성이 발표한 ‘AI 사업자 가이드라인’ 및 구미의 문헌 등을 참조해 수립했다. 안내서에서는 AI 안전 평가의 관점이나 예상할 수 있는 리스크의 예, 평가 항목 예, 평가 실시자나 평가 실시 시기에 관한 생각, 평가에 관한 방법의 개요 등 평가 내용의 구체적인 예를 제시하고 있다.

개발·제공·이용의 각 단계에서는 AI 개발·제공 사업자가 합리적인 범위, 적절한 시기에 평가를 실시해야 한다고 기술하고 있다. 개발 단계에서는 대규모 언어 모델(LLM)의 학습에 이용하는 데이터나 LLM 그 자체가 평가 범위가 된다. 제공·이용 단계의 평가 범위는 LLM 시스템 전체이며, 연계 애플리케이션 서비스나 RAG(Retrieval Augmented Generation, 검색 확장 생성)의 데이터베이스를 포함한다. 평가 방법으로서 틀을 이용한 ‘기술적 평가’와 사업자의 대처 방침이나 사업자 내에서 정비한 규정 등을 평가하는 ‘관리적 평가’의 양쪽 모두가 필요하다고 했다.

안내서의 주된 독자는 AI 개발자·AI 제공자, 특히 ‘개발·제공 관리자’나 ‘사업 집행 책임자’이며, 상정하는 시스템은 대규모 언어 모델(LLM)을 구성요소로 하는 AI 시스템(LLM 시스템)이다.

2) ‘AI 사업자 가이드라인’ 수립⁶⁵⁾

(1) 개요

AI 전략회의에서 정리된 ‘AI에 관한 잠정적 논점 정리’(2023.5)에서 최근의 생성 AI의 보급을 토대로 기존 가이드라인에 관해서 필요한 개정 등을 검토할 필요성이 제시되었다. 그에 따라 경제산업성과 총무성은 기존 가이드라인들⁶⁶⁾을 통합·업데이트해, 광범위한 통일

65) 経済産業省, AI事業者ガイドライン(第1.0版)』を取りまとめました, 2024.4.19
(https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240419_1.pdf)

적이고 알기 쉬운 AI 사업자용 가이드라인을 검토했다. 양 부처는 각각 ‘AI 사업자 가이드라인 검토회’와 ‘AI 네트워크 사회 추진 회의’를 개최, 검토하여 ‘AI 사업자 가이드라인’을 정리했으며, 의견 수렴을 거쳐 2024년 4월 ‘AI 사업자 가이드라인(제1.0판)’을 정리했다. 본 가이드라인은 AI의 보급에 대응하기 위한 지침으로서 AI 개발자, AI 제공자, AI 이용자라는 3자의 입장에서, 기본이념이나 지침, 실천 방법을 분명히 해, AI의 안전한 활용을 목표로 하고 있다.

(2) AI 사업 활동을 담당하는 주체

구체적으로는 ▲AI 개발자 ▲AI 제공자 ▲AI 이용자라는 세 가지 주체에 대해 각각의 책임과 기대되는 행동이 정의돼 있다. 이들 주체는 AI 기술의 안전하고 유익한 사회로의 통합을 목표로 하는 데 중요한 역할을 담당하고 있다.

〈표 3-33〉 AI 사업 활동을 담당하는 주체

구분	주요 내용
AI 개발자 (AI Developer)	<ul style="list-style-type: none"> AI 시스템을 개발하는 사업자(AI를 연구개발하는 사업자 포함) AI 모델·알고리즘의 개발, 데이터 수집(구입 포함), 전처리, AI 모델 학습 및 검증을 통해서 AI 모델, AI 모델의 시스템 기반, 입출력 기능 등을 포함한 AI 시스템을 구축하는 역할을 담당
AI 제공자 (AI Provider)	<ul style="list-style-type: none"> AI 시스템을 애플리케이션, 제품, 기존 시스템, 비즈니스 프로세스 등에 통합하는 서비스로서 AI의 이용자(AI Business User), 경우에 따라서는 업무 외 이용자에게 제공하는 사업자 AI 시스템 검증, AI 시스템의 다른 시스템과의 연계 구현, AI 시스템·서비스의 제공, 정상 가동을 위한 AI 시스템에서의 AI 이용자(AI Business User) 측 운영 지원 또는 AI 서비스의 운영 자체를 담당하고, AI 서비스 제공에 따라 다양한 이해관계자와 커뮤니케이션
AI 이용자 (AI Business User)	<ul style="list-style-type: none"> 사업 활동에서 AI 시스템 또는 AI 서비스를 이용하는 사업자 AI 제공자가 의도하고 있는 적정한 이용을 하고, 환경 변화 등의 정보를 AI 제공자와 공유해 정상 가동을 계속하거나 또는 필요에 따라서 제공된 AI 시스템을 운용하는 역할을 담당 또, AI 활용이 업무 외 이용자에게 어떠한 영향을 미친다고 판단되는 경우, AI로 인해 해당 이용자에게 의도하지 않은 불이익이 발생하지 않도록 하고, AI의 이점을 극대화하는 역할을 담당

출처: 総務省, 経済産業省, AI事業者ガイドライン(第1.0版), 2024.04.19.

66) AI 개발 가이드라인(2017년, 총무성), AI 활용 가이드라인(2019년, 총무성), AI 원칙 실천을 위한 거버넌스 가이드라인 Ver1.1(2022년, 경제산업성)

(3) 가이드라인의 기본이념

가이드라인은 ‘인간 중심의 AI 사회 원칙’ (2019.3)에 따라, ①인간의 존엄성이 존중받는 사회(Dignity) ②다양한 배경의 사람들이 다양한 행복을 추구할 수 있는 사회(Diversity & Inclusion) ③지속가능한 사회(Sustainability) 실현을 기본이념으로 한다.

(4) AI를 활용하는 모든 사람이 지켜야 할 10대 원칙

각 주체는 법의 지배, 인권, 민주주의, 다양성, 공평·공정한 사회를 존중하도록 AI 시스템·서비스를 개발·제공·이용하고, 관련 법령 및 AI와 관련된 개별 분야의 기존 법령 등을 준수하여야 하며, AI를 활용하는 모든 사람이 지켜야 할 10대 원칙으로 ①인간 중심 ②안전성 ③공평성 ④개인정보보호 ⑤보안 확보 ⑥투명성 ⑦설명 책임 ⑧교육·리터러시 ⑨공정 경쟁 확보 ⑩혁신 촉진 등을 제시했다. 공통 지침은 AI를 활용해 기본이념을 실현하기 위해 각 주체(AI 개발자, AI 제공자, AI 이용자)가 대응해야 할 사항과 이들 주체가 사회와 연계해 대응하는 것이 기대되는 사항으로 분류하고 있다.

<표 3-34> 각 주체의 공통 지침

구분		주요 내용
각 주체가 대응해야 할 사항	①인간 중심	<ul style="list-style-type: none"> 최소한 헌법이 보장하거나 국제적으로 인정되는 인권을 준수 AI가 사람들의 능력을 확장하고 다양한 사람들의 다양한 행복(well-being) 추구가 가능하도록 행동하는 것이 중요 AI에 과도하게 의존하는 리스크에 대해 필요한 대책을 강구 AI가 생성한 허위·오·편향정보가 사회를 불안정화·혼란시킬 위험에 대해 필요한 대책을 강구 사회적 약자가 AI 활용을 용이하게 하도록 주의
	②안전성	<ul style="list-style-type: none"> AI 시스템 서비스의 정확성을 포함해 요구에 대해 충분히 동작하여 신뢰성 확보 실행 수준을 유지하고, 무관한 사건에 대해 현저하게 잘못된 판단을 발생시키지 않도록 견고성 확보 필요에 따라 객관적인 모니터링 및 대처를 포함하여 인간이 통제할 수 있는 제어 가능성을 확보 학습 등에 이용하는 데이터의 투명성 확보와 법적 체계 준수, AI 모델의 갱신 등을 합리적인 범위에서 적절하게 실시
	③공평성	<ul style="list-style-type: none"> 특정 개인 또는 집단에 대한 부당하고 해로운 편견과 차별을 없애도록 노력하는 것이 중요 AI 결과가 공평성을 결여하는 일이 없도록 AI에게 단독으로 판단하게 할 뿐만 아니라 적절한 시기에 인간의 판단을 개입시키는 이용을 검토
	④개인정보보호	<ul style="list-style-type: none"> 개인정보보호법 등의 관련 법령 준수, 각 주체의 개인정보보호 정책의 수립·공표 등을 통해 이해관계자의 개인정보보호가 존중·보호되도록 그 중요성에 맞게 대응

구분		주요 내용
	⑤보안 확보	<ul style="list-style-type: none"> 기밀성·완전성·가용성을 유지하고 상시 AI의 안전·안심할 수 있는 활용을 확보하기 위해 그 시점에서의 기술 수준에 비추어 합리적인 대책 강구 새로운 방법으로 외부의 공격 리스크에 대응하기 위한 유의사항 확인
	⑥투명성	<ul style="list-style-type: none"> AI 시스템 서비스의 검증 가능성을 확보하면서 필요하고 기술적으로 가능한 범위에서 이해관계자에게 합리적인 범위에서의 정보 제공이 중요 AI의 학습 프로세스, 추론 과정, 판단 근거 등의 로그를 기록·저장
	⑦설명 책임	<ul style="list-style-type: none"> '공동 지킴'의 대응 상황 등에 대하여 이해관계자에게 리스크 정도를 감안해 합리적인 범위에서 책무성을 다하는 것이 중요 데이터의 출처, 의사결정 등에 대해 기술적으로 가능하고 합리적인 범위에서 추적·소구가 가능한 상태를 확보 일에 대한 책무를 다하는 책임자를 설정 필요에 따라 리스크 관리, 안전성 확보를 위한 AI 거버넌스에 관한 정책, 개인정보 보호 정책 등의 방침을 수립·공표 관계 정보 등을 문서화하여 일정 기간 보관하고, 필요할 때 입수 가능하고 참조 가능한 상태로 유지
사회와 연계한 대응	⑧교육·리터러시	<ul style="list-style-type: none"> AI 관련자가 충분한 수준의 AI 리터러시를 확보하기 위해 필요한 조치를 강구 생성 AI의 활용 확대에 의한 인간과 AI의 작업의 구분에 따라 새로운 일하는 방식이 가능하도록 교육·재교육 등을 검토
	⑨공정 경쟁 확보	<ul style="list-style-type: none"> AI를 둘러싼 공정한 경쟁 환경 유지에 노력
	⑩혁신 촉진	<ul style="list-style-type: none"> 산학 협력, 오픈 이노베이션 등을 추진 자사와 타사의 AI 시스템 서비스의 상호 연결성과 상호 운용성을 확보 표준 사양이 있는 경우에는 그에 따라 실시 적절한 정보를 제공

출처: 大和総研, AI 事業者ガイドラインの公表, 2024.5.10

(5) 각 주체에 대한 사항

AI 개발자는 AI 모델을 직접 설계하고 변경할 수 있기 때문에 AI가 제공·활용될 때 어떤 영향을 미칠지를 사전에 최대한 검토하고 대응책을 마련하는 것이 중요하다.

- 적정 이용에 이바지하는 개발
- 학습 데이터나 알고리즘 등에 포함된 편견에 대한 배려
- 보안대책 도입, 개발 후에도 리스크에 대응
- AI 제공자에 대한 공동 지킴 대응 상황 설명
- 혁신의 기회 창조에 대한 공헌 등

AI 제공자는 AI의 운영과 적절한 활용을 전제로 한 AI 시스템 및 서비스 제공을 실현하는 것이 중요하다.

- 적정 이용에 이바지하는 제공
- AI 시스템 서비스 및 데이터에 포함된 편견에 대한 배려
- 보안대책 도입, 제공 후에도 취약성에 대한 대응
- AI 이용자에 대한 공통 지침 대응 상황 설명
- 서비스 약관의 문서화 등

AI 이용자는 AI 제공자가 의도한 범위 내에서 적절하게 이용하고 AI를 필요에 따라 효과적으로 활용하기 위해 필요한 지식을 습득하는 것이 중요하다.

- 안전을 고려한 적정 이용
- 입력 데이터, 프롬프트에 포함된 바이어스에 대한 배려
- 보안대책 실시
- 이해관계자 등의 관련 이해관계자에 대한 설명
- 개인정보보호 침해에 대한 유의 등

3) ‘통합혁신전략 2024’에서의 AI 정책 추진⁶⁷⁾

(1) 개요

일본 정부는 2024년 6월 각료회의에서 ‘제6기 과학기술·혁신기본계획(2021~25)’의 연차 전략인 ‘통합혁신전략 2024’를 발표했다. 올해는 AI 분야의 경쟁력 강화 및 안전·안심 확보를 중심으로 AI의 법규제를 검토하는 ‘AI 제도연구회(가칭)’ 신설 등의 내용이 포함되었다. ‘통합혁신전략 2024’는 AI를 한 축으로 하는 3대 강화 방안으로 ①핵심기술에 대한 통합적 전략 ②글로벌 관점의 협력 강화 ③AI 분야 경쟁력 강화 및 안전·안심 확보를 제시했다.

67) 内閣府, 統合イノベーション戦略 2024, 2024.6.4
(https://www8.cao.go.jp/cstp/tougosenryaku/togo2024_honbun.pdf)

(2) AI 분야 경쟁력 강화 및 안전·안심 확보 추진

생성 AI는 인터넷에도 필적하는 기술 혁신으로 여겨져 사회경제 시스템에 큰 변혁을 가져오는 한편, 허위·오정보의 유포나 범죄의 교묘화 등 다양한 리스크도 지적되어 안전·안심의 확보가 요구된다. 미국 기업 등의 고성능·대규모의 범용 기반 모델이 선행하는 가운데, 일본도 그것에 따르기 위해 계산 자원의 정비나 대규모 모델의 개발이 진행되고 있으며, 또 소규모·고성능인 모델이나 복수 모델의 조합 개발 등 새로운 연구도 진행하고 있다.

AI는 모든 분야에서 이용되고, AI의 개발 및 이용 등의 혁신이 사회 과제 해결이나 일본의 경쟁력에 직결할 가능성이 있다. 일본 정부는 생성 AI를 포함한 AI의 다양한 리스크를 억제하고 안전·안심 환경을 확보하면서, 혁신을 가속화하는 선순환 구조를 구축하고, 또한 일본이 주도하는 ‘히로시마 AI 프로세스’ 등을 통해 앞으로도 국제적리더십을 발휘해 나갈 계획이다.

- AI 혁신과 AI를 통한 혁신 가속: 연구개발 역량 강화, AI 활용 추진, 관련 인프라 고도화, 인재 육성·확보 등
- AI의 안전·안심 확보: 거버넌스와 제도 개선 검토, AI 안전성 검토, 허위·오정보에 대한 대책 수립, 지식재산권 검토 등
- 국제협력 추진: G7 정상외장 첨단 AI 기술 발전을 통제하기 위한 국제 지침과 행동규범에 합의한 ‘히로시마 AI 프로세스’ (2023.5)의 성과를 바탕으로 한 국제협력 강화

〈표 3-35〉 통합혁신전략 2024의 AI 정책 추진

구분	주요 내용
AI 혁신과 AI를 통한 혁신 가속	<p>[연구개발 역량 강화(데이터 정비 포함)]</p> <ul style="list-style-type: none"> • AI 개발력을 강화하기 위해 중요한 일본어 데이터를 정비·확충해 적절한 형태로 제공하고, 각 분야에서 경쟁력 있는 AI를 개발하기 위해 데이터 공유 등을 촉진하는 동시에, 모범사례를 공유 • AI 개발에 불가결한 계산 자원을 폭넓은 개발자가 이용할 수 있도록 민관에서 지속 정비 • 모델의 고효율화나 고정밀화, 멀티 모달화, 리스크의 저감화 등 연구개발을 산학 협력으로 추진하는 동시에, 혁신적인 기술을 가진 스타트업을 지원 • 의료나 신약개발, 소재 등 분야에서 일본의 강점인 과학연구 데이터 창출 기반을 강화하는 등 ‘AI for Science’를 민관이 가속하는 동시에, 뛰어난 AI 성능을 가진 새로운 플래그십 시스템의 개발·정비에 착수 • 노동력 부족 해소나 녹색 혁신(GX) 등에도 이바지하는 환경 변화에 유연하게 대응 가능한 혁신적인 AI 로봇 등의 연구개발·상용화를 민관이 추진

구분	주요 내용
	<p>(AI 활용 추진)</p> <ul style="list-style-type: none"> • ‘챗GPT(ChatGPT) 등의 생성 AI의 업무 이용에 관한 합의(제2판)’를 ‘AI 사업자 가이드라인’의 수립에 따라 더욱 진전시키고, 정부의 적절한 AI 조달·활용 촉진 • 의료·로봇 등의 중요 분야에서의 AI 도입을 촉진 • 이용자나 개발자가 위촉되지 않고 AI 활용·개발을 할 수 있도록, 「개인정보보호법」, 「저작권법」 등 유의해야 할 제도 운용을 명확화 <p>(관련 인프라 고도화)</p> <ul style="list-style-type: none"> • 데이터센터의 대규모화·분산화와 절전, 6G 등 네트워크 시스템의 고도화를 위해 연구개발을 촉진하는 동시에 AI에 불가결한 인프라에 대한 민간투자를 확대 • AI 반도체 등 핵심 기기의 설계·개발·운용에 관한 산학 협력 체제 등 구축, 연구 개발, 인재육성을 지원 • 저렴한 탈탄소 전력의 양적 확보를 위한 환경 정비 검토 <p>(인재 육성·확보)</p> <ul style="list-style-type: none"> • AI 스킬 습득, AI 리터러시 향상을 위한 교육 콘텐츠의 충실·보급 개발을 도모 • 초·중등 교육 단계에서는 정보 윤리를 포함한 AI 활용에 관한 시범적인 대처를 한층 더 진행 • 차세대 AI 개발 등을 담당할 젊은 연구자나 박사후 과정 학생을 지원
AI의 안전·안심 확보	<p>(거버넌스와 제도 개선 검토)</p> <ul style="list-style-type: none"> • 폭넓은 업종에 ‘AI 사업자 가이드라인’의 주지·보급을 도모 • AI 전략회의 산하에 새롭게 설립하는 ‘AI 제도연구회(가칭)’에서 제도의 기본 방향 검토에 착수 • 의료, 자율주행, 금융 등 사회에 미치는 영향이 큰 중요 분야는 기술의 진전이나 이용 상황에 따라 제도의 재검토 필요성 등을 검토 <p>(AI 안전성 검토)</p> <ul style="list-style-type: none"> • AI 안전연구소(AISI)는 AI의 안전성의 중심적 기관으로서 정보처리추진기구(IPA)에 설치되어 전문 인재의 육성·확보, 선진적인 기술적 지식과 식견 집약 등을 추진 • 외부 지식을 이용해 환각(hallucination)을 방지하는 기술 등 AI 안전성에 관한 최첨단 연구 개발을 민관이 연계해 진행 <p>(허위·오정보에 대한 대책 수립)</p> <ul style="list-style-type: none"> • 생성 AI를 이용한 것을 포함해, 인터넷상에 유통·확산하는 허위·오정보나 SNS상의 위장형 가짜 광고에 대한 대응 등에 대한 기술·연구개발 추진, 팩트체크 추진, 국제적인 협력 강화 등 제도 면도 포함한 종합적인 대책을 수립 • 인터넷상에 유통하는 AI 생성 콘텐츠를 판별하는 기술의 개발·실증이나, 리터러시 향상 등을 추진 <p>(지식재산권 검토)</p> <ul style="list-style-type: none"> • 내각부 ‘AI 시대의 지식재산권 검토회의 중간정리’나 문화심의회 저작권 분과회 법제도 소위원회 ‘AI와 저작권에 관한 생각에 대해서’를 근거로 해 향후의 기술 발전과 해외 동향 등도 보면서, 배우나 성우 등의 초상이나 목소리도 포함해 필요한 검토를 지속 추진
국제 협력 추진	<ul style="list-style-type: none"> • 히로시마 AI 프로세스 등을 통해서 안전·안심하고 신뢰할 수 있는 AI의 실현을 향해서 국제적인 대응을 계속 주도하는 동시에, 아시아 여러 국가나 글로벌 사우스와도 협조하면서 혁신 창출을 지속 추진 • 히로시마 AI 프로세스의 성과를 더욱 진전시키기 위해 히로시마 AI 프로세스 프렌즈 그룹(2024.5)을 활용하면서 국제 지침이나 행동규범의 실천을 확대

구분	주요 내용
	<ul style="list-style-type: none"> • GPAI(AI에 관한 글로벌 파트너십) 도쿄 전문가 지원센터 등을 통해서 히로시마 AI 프로세스의 성과 달성에 이바지하는 가짜 정보 대책 기술의 평가·실증 등의 프로젝트 기반 대응을 지원 • 일본 AISI와 여러 나라의 AISI 등의 국제적인 네트워크를 구축, AI의 안전성 확보를 위한 방안을 검토 • AI 개발자 간 노하우 공유, 글로벌 기술기업과의 교류 등 커뮤니티 활동을 민간에서 촉진하는 동시에, AI 분야에서의 글로벌 사우스와의 협력 롤모델 구축 • 미국 등 협력국과 연계한 개발 체제 구축, 대학이나 국립연구소 등의 잠재력도 활용한 산학협력 체제 구축

출처: 内閣府, 統合イノベーション戦略 2024, 2024.06.04.

4) 방위성의 AI 활용 추진 기본방침 수립⁶⁸⁾

(1) 개요

일본 방위성은 2024년 7월 방위성 최초로 AI 활용 추진의 나침반이 되는 기본방침을 수립했다. 방위성 내 각처에서 행해지고 있는 AI 활용의 대처에 대해 ‘전략성’과 ‘일관성’을 담보하는 동시에, 외부 기업이나 연구기관과의 원활한 협력 관계를 구축하는 것이 목적이다. 기본방침은 국내외에서의 최신 논의를 참고로 AI의 활용 분야를 설정하고, 데이터 기반 구축, AI·데이터 인재의 확보·육성, 연구개발 등 폭넓은 사항에 대해 방위성의 AI 활용 추진을 위한 대응 방안을 제시하고 있다.

(2) AI 활용 분야

AI 활용 분야에서는 ▲목표 탐지·식별 ▲정보 수집·분석 ▲지휘통제 ▲후방 지원 업무 ▲무인 자산 ▲사이버보안 ▲사무처리 작업 효율화 등 7개 분야를 중점 분야로 설정해 AI 활용을 적극 추진할 방침이다.

<표 3-36> 방위성의 AI 활용 분야

구분	주요 내용
목표 탐지·식별	<ul style="list-style-type: none"> • 레이더 항적, 위성·항공 화상 등 다방면의 센싱 정보 증가나 고성능화에 따른 목표 정보의 기하급수적 증대에 대응하기 위해 인력으로 실시하고 있던 목표의 탐지·식별 작업에 AI 기술 등을 활용해 탐지·식별 능력 향상 및 신속화를 도모

68) 防衛省, 防衛省A I活用推進基本方針, 2024.7
(https://www.mod.go.jp/j/press/news/2024/07/02a_03.pdf)

구분	주요 내용
정보 수집·분석	• 각종 정보의 자동 수집이나 분석·평가에 AI를 활용해, 인터넷이나 SNS상의 방대한 정보, 센서 등의 다양한 전파 정보, 화상 정보의 수집·분석에 대응
지휘통제	• 복잡하고 빠른 전투양상에 대응하기 위해 경계감시·정보수집으로 얻어진 정보의 분석·견적·평가를 토대로 한 행동방침의 도출에 AI를 활용하고, 이를 통해 신속한 의사결정을 지원
후방 지원 업무	• 장비품의 가동률 향상이나 효율적인 장비품의 유지관리에 대응하기 위해 보급 및 고장품 데이터를 기본으로 한 수요예측이나 정비예측, 수송계획 수립에 AI를 활용해 정비·보급 등의 후방지원 업무를 효율화
무인 자산	• 무인 자산의 기체 제어나 행동 판단에 AI를 적용해, 무인 자산의 자율 운용 능력 향상, 유인 장비와 무인 자산의 연계를 도모
사이버보안	• AI를 활용한 행동해석의 도입 등을 통한 사이버보안 능력 향상, AI의 모델·데이터 세트를 표적으로 한 AI 그 자체에 대한 사이버 공격 대처 능력 향상을 도모
사무처리 작업 효율화	• 행정조직으로서의 방위성 내에서 행해지고 있는 다양한 사무처리 업무에 AI를 도입해 효율화를 도모

출처: 防衛省, 防衛省 A I 活用推進基本方針, 2024.07.

(3) AI 활용 추진을 위한 대응 방안

방위성은 방위력의 근본적인 강화를 위해 데이터 기반 구축과 AI·데이터 인재 확보·육성, 연구개발 등에 AI 활용을 폭넓게 추진해 대응한다.

<표 3-37> 방위성 AI 활용 추진을 위한 대응 방안

구분	주요 내용
데이터 기반 구축	• 모든 대원에게 '데이터가 임무수행에 필수적인 전략자산이다'라는 의식을 함양 • 공유가 필요한 데이터를 특정하고 데이터 포맷을 표준화
AI·데이터 인재	• AI 관련 업무를 대원만으로 처리하는 것은 비현실적이며, 민간 AI·데이터 인력을 적극적으로 활용 • 내부 육성 측면에서는 인재 요구에 적합한 AI 강좌 수강을 촉진
연구개발	• 민생의 지식과 식견을 필요에 따라 AI를 사용한 장비품의 연구개발을 효율적으로 추진
교육·연구기관과 협력	• 교육·연구기관과의 네트워크 확대, 협력관계 확대·심화
각국과의 협력	• 각국의 방위당국과 정보교환 및 식견 공유 등을 실시
AI 준비통제·윤리	• AI의 '책임 있는 이용'을 확보하기 위한 시책을 검토 • 인도적 관점과 안전보장상의 필요성을 고려한 균형 잡힌 원칙이나 규범의 제정을 목표로 국제적인 논의에 적극적이고 건설적으로 공헌
다른 첨단기술과 생성 AI	• AI를 뒷받침하는 인프라가 될 가능성이 있는 첨단기술에 대하여 필요에 따라 도입 검토 • 생성 AI는 가능한 한 리스크 감축을 중시하면서 도입

출처: 防衛省, 防衛省 A I 活用推進基本方針 概要 2, 2024.07.

5) AI 제도연구회 설치⁶⁹⁾

(1) 개요

일본 정부가 AI 법규제의 기본 방향 등을 논의하는 전문가 회의인 ‘AI 제도연구회’를 ‘AI 전략회의’ 산하에 설치했다. 연구회의 목적은 AI 기술의 진전과 그에 따른 사회적 영향을 고려해 적절한 법규제의 틀을 검토하는 것이다. ‘AI 제도연구회’는 첫 회의를 2024년 8월 2일 수상관저에서 개최했으며, 다양한 전문가 의견을 청취, 논의해 이번 가을에 논의를 중간정리할 계획이다.

(2) 4대 기본 원칙

기시다 총리는 첫 회의에서 ‘AI의 안전성 확보가 AI의 활용 촉진, 개발력 강화를 위해서도 불가결하다’라고 언급하며, 제도의 기본 방향을 논의하는 데 있어서 4대 기본 원칙을 표명했다. 기시다 총리가 언급한 4대 원칙은 ①AI의 안전성 확보와 경쟁력 강화의 양립 ② 기술 변화에 대응할 수 있는 유연한 제도 ③국제적인 지침에 의거 ④AI에 대한 정부의 적정한 조달·이용 등이다.⁷⁰⁾

AI 제도연구회는 4대 원칙에 따라 ▲대응해야 할 리스크, 규제 대상이 되는 행위 ▲규제 방법(사전 규제/사후 규제) ▲규제 강도 ▲지켜야 할 사항, 기준 등을 다양하게 검토할 방침이다.

(3) 지금까지의 논의

AI 제도연구회에서 지금까지 논의는 ▲혁신을 저해하지 않도록 필요 최소한의 법규제를 ▲경성법(hard law)과 연성법(soft law)의 조합이 중요 ▲제도와 기술의 양면에서 AI의 안전성을 높여야 한다는 의견이 나오고 있으며, AI는 자칫 리스크에만 초점이 맞춰지기 쉽지만, 일본 정부는 리스크를 줄이면서 혁신을 불러일으킴으로써 일본이 AI의 개발·구현에서 구미 등 주요국을 추격하는 것이 중요하다고 생각하고 있다.

69) 内閣府, AI制度研究会
(https://www8.cao.go.jp/cstp/ai/ai_kenkyu/ai_kenkyu.html)

70) 首相官邸, AI戦略会議・AI制度研究会合同会議, 2024.8.2
(https://www.kantei.go.jp/jp/101_kishida/actions/202408/02ai.html)

제5절 영국

1. 그간의 인공지능 정책 추진 경과

영국은 2021년 10월 AI 분야에서 세계를 선도하겠다는 목표하에 영구 최초의 ‘국가 AI 전략(National AI Strategy)’을 수립했다.⁷¹⁾ 영국 국가 AI 전략은 향후 10년간 AI 분야에서 초강대국 리더십 확보를 국가의 AI 비전으로 삼아 ▲전략적 투자 ▲포괄적 혜택 ▲거버넌스 3개 영역에서 단기(향후 3개월), 중기(6~12개월), 장기(12개월~) 정책 과제를 제시하였다.

〈표 3-38〉 영국 국가 AI 전략의 부문별 정책 과제

구분	전략적 투자	포괄적 혜택	거버넌스
단기 (향후 3개월)	<ul style="list-style-type: none"> 데이터 가용성 확대 국가 사이버-물리 인프라 프레임워크 개발 AI, 데이터과학, 디지털 기술 개발 지원 	<ul style="list-style-type: none"> 보건과 사회복지 분야 AI 기반 기술 국가 전략 개발 국방 AI 전략 수립 AI 저작권과 특허 연구 	<ul style="list-style-type: none"> 데이터윤리혁신센터(CDEI) 인증 로드맵 개발 AI 거버넌스에서 데이터 보호의 역할 설정 국방부 AI 지침 개발 국제 AI 활동에 대한 범정부적 접근방식 개발
중기 (6~12개월)	<ul style="list-style-type: none"> 민간 자금 수요와 과제 평가 국립컴퓨팅교육센터 지원 AI 관련 경력 개발 지원 미국과 AI R&D 협력 -글로벌 인재 영입 	<ul style="list-style-type: none"> 경제 전반에서 AI 기회 연구 에너지 분야 등 혁신 산업에 AI 역량 포함 개발도상국의 기술 개발 지원 	<ul style="list-style-type: none"> AI 거버넌스와 규제에 관한 국가 백서 발간 알고리즘 투명성을 위한 범정부 표준 개발 AI 표준 허브 시범 운영 정부 AI 안전 인식 제고
장기 (12개월~)	<ul style="list-style-type: none"> 반도체 공급망 검토 AI 모델교육용 정부 데이터셋 개발 AI 연구혁신 프로그램 확대 AI 인재 육성과 투자 유치 AI 무역거래규정 개발 	<ul style="list-style-type: none"> AI 사무국, UKRI 프로그램 출범 기술의 신뢰성, 채택 가능성 역량 개발 지원 정부기관 협력으로 전략 과제 개발 	<ul style="list-style-type: none"> AI 기술 표준 참여 틀기 개발 공통 R&D 과제에 대한 글로벌 파트너십 확대 공공 부문의 AI 윤리와 안전 지침 개정 국가안보와 국방 부문 전문가와 협력 확대

출처: Gov.uk(2021.10), SPRI 정리, 2024.10.

71) Gov.uk, National AI Strategy, 2022.12.18.

(<https://www.gov.uk/government/publications/national-ai-strategy/national-ai-strategy-html-version#summary-of-key-actions>)

영국의 국가 AI 전략은 영국이 가진 강점을 토대로 민간과 공공 전반에 걸쳐 복원력과 생산성, 성장과 혁신을 강화하는 AI의 잠재력을 극대화하여 영국에서 AI를 통한 단계적 변화를 주도하겠다는 비전을 제시하고 있다.

영국 정부는 2022년 7월 국가 AI 전략 - 실행계획(National AI Strategy - AI Action Plan)을 발표한 데 이어, 2022년 9월에는 혁신을 촉진하고 대중의 신뢰를 높이며 데이터를 보호하는 것을 골자로 한 ‘AI 규제안(AI Rulebook)’을 발표했다. AI 규제안은 AI를 책임 있게 사용하는 동시에 기업의 규정 준수 부담을 줄이는 조치를 포함한다. 영국 정부가 2023년 3월 발표한 ‘친혁신적 AI 규제 접근’ 백서 역시 AI 혁신과 함께 AI 위험에 대응하기 위한 내용을 담아, AI 개발과 사용의 위험과 이익의 균형을 통해 어느 정도의 위험을 용인하며 혁신을 추진하기 위한 친혁신적이고 유연한 비례적 프레임워크를 제시하였다.

<표 3-39> 영국 AI 규제에 대한 친혁신 프레임워크의 핵심 요소

핵심 요소	주요 내용
규제기관 조정 지원을 위해 고유 특성 기반으로 AI를 정의	<ul style="list-style-type: none"> AI를 효과적으로 규제하고 제안된 체계의 명확성을 지원하기 위해 AI에 대한 공통 정의가 필요하다고 인식하며, AI의 공통 특성으로 적응성(Adaptivity)과 자율성(Autonomy)을 제시
상황별 접근방식 채택	<ul style="list-style-type: none"> 규정이나 위험 수준을 전체 영역이나 전체 기술에 적용하지 않고, 특정한 AI 적용 상황에서 가능한 결과를 기반으로 상황에 따른 규제를 실행 규제기관이 AI 사용에 따른 위험과 수행 기회 상실로 인한 비용을 비교 평가할 수 있도록 지원
부문 간 원칙(cross-sectoral principles) 제공	<ul style="list-style-type: none"> ▲안전·보안·견고성 ▲투명성·설명 가능성 ▲공정성 ▲책임·거버넌스 ▲이의 제기·구제의 5개 원칙에 따라 규제 프레임워크를 이행
규제의 일관성 보장	<ul style="list-style-type: none"> 규제기관이 원칙을 적절히 고려하도록 새로운 의무를 도입하고 규제기관의 권한을 강화하여 규제 환경 전반에 걸쳐 일관성을 유지 규제기관, 국제 파트너, 업계, 노동조합과 시민사회, 학계, 일반 대중을 포함한 광범위한 이해관계자를 모아 협업을 촉진 규제 체계 구현을 지원하는 중앙 기능으로 모니터링 및 평가, 원칙의 일관된 구현, 부문 간 위험 평가, 혁신 지원, 교육 및 인식 제고, 국제협력을 지원

출처: Gov.uk(2023.03), SPRI 정리, 2024.10.

영국 정부는 상기의 규제 프레임워크를 통해 ▲AI 선도 국가로서의 입지 강화 ▲책임 있는 혁신 및 규제 불확실성 감소를 통한 AI 산업의 성장 ▲위험 규율과 기본 가치 보호를 통한 대중의 AI 신뢰 제고라는 목표를 추구하고 있다. 영국은 이러한 목표 달성을 위해 2023년 11월 세계 최초로 AI 안전성 정상회의(AI Safety Summit)를 개최했다. 정상회의에는 미국, 중국, 캐나다, 독일, 인도, 대한민국 등 28개국 및 EU가 참석하였으며, 영국은 정상회의에서 AI 안전연구소 설립을 발표하고 AI 안전성 국제기구 설립을 제안하는 등 AI 안전과 관련된 거버넌스 논의를 주도하였다.

2. 2024년 인공지능 정책 동향

1) 내각부

영국 내각부는 2024년 1월 정부 기관의 책임 있고 안전한 생성 AI의 사용을 지원하기 위한 생성 AI 프레임워크(Generative AI Framework for HMG)⁷²⁾를 공개했다. 프레임워크는 정부 기관의 생성 AI 이해와 도입을 지원하고, 안전하고 책임 있는 생성 AI 사용을 위해 고려해야 할 사항을 제시하기 위해 개발되었다. 프레임워크는 정부 기관에서 생성 AI를 안전하고 책임 있게 사용하기 위한 10가지 공통 원칙을 다음과 같이 제시하였다.

〈표 3-40〉 영국 정부를 위한 생성 AI 프레임워크의 10개 원칙

구분	주요 내용
원칙 1	생성 AI가 무엇이며 생성 AI의 한계가 무엇인지 이해
원칙 2	생성 AI 기술을 합법적이고 윤리적으로 책임 있게 사용
원칙 3	생성 AI 도구를 안전하게 유지하는 방법을 인지
원칙 4	적절한 단계에서 유의미한 인간의 통제를 유지
원칙 5	전체 생성 AI 수명주기를 관리하는 방법을 이해
원칙 6	작업에 적합한 도구를 사용
원칙 7	개방적이고 협력적인 태도 유지
원칙 8	처음부터 민간 기업과 협력 관계를 수립
원칙 9	생성 AI를 구축하고 사용하는 데 필요한 기술과 전문 지식 보유
원칙 10	소속 기관의 정책과 함께 이러한 원칙을 사용하고 적절한 보장조치를 마련

출처: UK Cabinet Office, 2024.01. SPRI 정리, 2024.10.

72) UK Cabinet Office, Generative AI Framework for HMG, 2024.01.18.

프레임워크는 생성 AI가 정부 업무에 사용될 수 있는 용도로 디지털 문의 지원, 서신이나 음성통화 분석을 통한 시민 요구사항 이해, 검색 향상, 복잡한 서류 요약, 문서와 서신의 초안 작성, 소프트웨어 개발 지원, 텍스트와 오디오 요약 등을 제시하였다. 또한 생성 AI의 한계로 부정확하거나 잘못된 추론의 생성 가능성, 환각, 공격적이거나 편향적인 콘텐츠 생성, 투명성 부족 등을 지적하였다. 프레임워크는 생성 AI 조달 시 고려해야 할 사항도 제시하였으며 공급업체에 대한 요구사항의 문서화와 함께 데이터 품질 및 편향 완화, 데이터 윤리 원칙의 적용, 공급업체 종속을 피하기 위한 전략, 지속적인 지원과 유지관리 필요성 등을 고려해야 한다고 강조하였다.

2) 과학혁신기술부

(1) 콘텐츠 스토어 프로젝트⁷³⁾

영국 과학혁신기술부는 교육부와 함께 2024년 8월 28일 교사의 업무 부담을 줄이기 위한 AI 기업 대상의 ‘콘텐츠 스토어’ 프로젝트에 400만 파운드(한화 약 70억 원)를 투입한다고 발표했다. 이 프로젝트는 커리큘럼 지침, 수업 계획, 익명화된 학생 평가와 같은 정부 문서를 통합해 AI 기업의 도구를 훈련하여 학교에서 안정적으로 사용할 수 있는 고품질의 맞춤형 교육 콘텐츠 생성을 목표로 한다. 콘텐츠 스토어의 활용 대상은 교사의 과제 채점과 교육 자료 제작, 학교의 일상적인 행정 업무 지원과 같은 교육 분야에 특화된 기술 기업이다.

이번 프로젝트는 교사가 대면 수업에서 학생을 돕는 데 더 많은 시간을 보낼 수 있도록 생성 AI의 사용을 희망하는 학부모 의견을 반영하였다. 교육부가 2024년 8월 발표한 교육 분야에서 AI 사용에 관한 연구⁷⁴⁾ 결과, 학부모와 학생 모두 교사 지원을 위한 AI 활용의 기회와 함께, 학생의 데이터를 활용해 AI 도구를 최적화함으로써 얻을 수 있는 이점을 이해하는 것으로 나타났다.

영국 정부는 콘텐츠 스토어 구축에 300만 파운드를 투입하는 한편, AI 기업의 콘텐츠 스토어 활용을 장려하기 위한 아이디어 경진대회를 통해 데이터를 실제로 적용해 교사의 업무 부담을 줄이는 최고의 아이디어를 낸 기업에 총 100만 파운드의 상금을 수여할 예정이다. 영국 정부는 2024년 9월 9일부터 기업들의 참가 신청을 받아 수상자를 선정할 계획으로, 각 수상자는

73) Gov.uk, Teachers to get more trustworthy AI tech, helping them mark homework and save time, 2024.08.28.

74) Gov.uk, Research on parent and pupil attitudes towards the use of AI in education, 2024.08.28.

2025년 3월까지 교사의 피드백과 채점을 지원하는 도구를 개발해야 한다. 교사용 설문조사 앱 티처탭(TeacherTapp)에 따르면 영국 교사의 절반이 업무에서 이미 AI를 활용하고 있으나 기존 AI 도구는 영국의 수업 진행 방식에 맞게 개발되지 않아 활용이 제한적인 상황이다.

영국 과학혁신기술부의 피터 카일(Peter Kyle) 장관은 AI를 활용해 교사의 행정 부담을 줄이고 창의적 수업을 지원하는 이번 프로젝트가 공공 데이터의 활용 방식을 변화시키는 첫 번째 사례가 될 것으로 기대했다. 교육부의 조사에 따르면 학습 관련 데이터를 제공받은 생성 AI 모델의 응답 정확도는 92%에 달해, 맞춤형 데이터가 제공되지 않은 모델(67%) 대비 정확도가 크게 향상된 것으로 나타났다.

(2) AI 보증 플랫폼 구축⁷⁵⁾

영국 과학혁신기술부는 2024년 11월 6일 AI 보증 시장에 대한 보고서를 발간하고 중소기업과 스타트업이 잠재적인 AI 위험을 평가하고 관리하는 데 필요한 지침을 제공하는 중앙 집중형 AI 보증 플랫폼을 개발하겠다는 계획을 밝혔다.

영국 정부에 따르면 현재 영국에는 AI 보증 상품과 서비스를 제공하는 기업이 약 524개이며, 이 중 84개가 전문 AI 보증 기업으로, 매출 규모는 약 10억 1천만 파운드(한화 약 1조 8,210억 원), 직원 수는 약 12,572명으로 파악된다. 영국의 AI 보증 시장은 경제규모 대비 미국, 독일, 프랑스보다 규모가 크다. 현재까지는 AI 보증 관련 수요와 공급이 잠재력에 미치지 못하는 상황이나, 영국 정부는 미래 성장을 위한 기회가 마련된다면 시장 규모가 2035년까지 65억 3천만 파운드(약 11조 7,750억 원)를 넘어설 것으로 기대하였다.

영국 정부는 AI 관련 위험을 완화하고 AI 보증 시장의 성장을 극대화하여 안전하고 책임 있는 AI 도입을 촉진할 수 있도록 AI 보증 관련 정보를 원스톱으로 제공하는 AI 보증 플랫폼을 개발하여 AI 보증 도구와 서비스에 대한 수요를 촉진할 계획이다. 또한 산업계와 협력해 ‘신뢰할 수 있는 제3자 AI 보증 로드맵’을 개발하고, AI안전연구소와 협력해 AI 보증 연구와 개발, 확산을 통해 제3자 AI 보증의 공급을 확대할 계획이다. 마지막으로, 책임 있는 AI 관련 용어를 개발하여 AI 보증의 상호운용성을 보장하겠다는 구상을 공개하였다.

75) Gov.uk, Assuring a Responsible Future for AI, 2024.11.06.

3) 영국 AI안전연구소

영국은 2023년 11월 AI 안전성 정상회의를 개최하며 과학혁신기술부 산하기관으로 AI안전 연구소를 설립하였다. AI안전연구소는 영국 정부가 첨단 모델의 안전성 평가를 위해 2023년 4월 출범한 프런티어 AI 태스크포스를 전신으로 하며, 연구소로 격상하면서 2024년 말까지 연구진 규모를 50~60명으로 확대할 계획이다.

연구소는 ▲첨단 AI 시스템 위험성 평가 ▲AI 안전 및 위험 연구 촉진 ▲글로벌 AI 개발 관행 및 안전 정책 강화를 목표로 AI 안전 연구의 기능을 수행한다. 첨단 AI 시스템의 위험 평가를 통하여 정부와 정책입안자에게 AI 위험 정보를 제공하는 것을 목표로 하며, 연구소에서 AI 안전성을 테스트할 수 있는 자체 역량 구축에 주력하고 있다.

(1) 미국 AI안전연구소와 AI 모델 안전 테스트 협력을 위한 MOU 체결⁷⁶⁾

영국 AI안전연구소를 운영하는 영국 과학혁신기술부는 미국 AI안전연구소를 담당하는 미국 상무부와 2024년 4월 1일 첨단 AI 모델에 대한 안전 테스트에서 상호 협력을 위한 양해각서(MOU)를 체결했다. 이는 2023년 11월 영국에서 열린 AI 안전성 정상회의의 합의에 따른 것으로, 당시 양국은 AI안전연구소 설립을 각각 발표하고, 안전하고 신뢰할 수 있는 AI의 개발과 사용에서 협력하기로 합의했다. 양국은 이번 MOU를 계기로 AI 모델과 시스템을 테스트하기 위한 평가 도구를 조속히 개발하고 긴밀히 협력할 예정이다.

미국 상무부는 양국이 AI 모델과 시스템 관련 기능과 위험에 대한 중요 정보와 함께 AI 안전과 보안에 관한 기본적인 기술 연구를 공유할 계획이라고 발표했으며, 이를 통해 전 세계 연구자들이 과학적 공통 기반을 중심으로 AI 안전성 평가에서 협력할 것으로 기대했다. 양국은 전 세계적으로 AI 안전을 촉진하고 국가 간 연계를 강화하기 위해 여타 국가와도 유사한 협력을 추진할 계획으로, AI 안전 테스트에 대한 국제 표준 및 첨단 AI 모델의 개발과 배포, 사용과 관련된 기타 표준에서 주요국 정부와 협력할 방침이다.

76) UK AI Safety Institute, Collaboration on the safety of AI: UK-US memorandum of understanding, 2024.04.02.

미국과 영국의 AI안전연구소는 상호 운용이 가능한 실무 프로그램과 AI 안전 연구를 위한 공통 접근방식을 개발할 계획으로 구체적으로는 기본 방법론, 인프라, 프로세스 등에서 AI 모델 평가를 위한 공통 접근방식을 개발하고 공개적으로 접근 가능한 AI 모델에 대하여 1회 이상 공동 테스트를 진행하기로 하였다. 또한 AI 안전 기술 연구에서 협력하여 첨단 AI 모델에 대한 지식을 축적하고, AI 안전과 보안 관련 사회 기술적 정책의 조율을 지원하며, 연구소 간 인적 교류를 통한 전문지식의 공유를 포함해 국가 법률과 규정, 계약에 따라 연구소 활동 전반에서 정보를 공유할 계획이다.

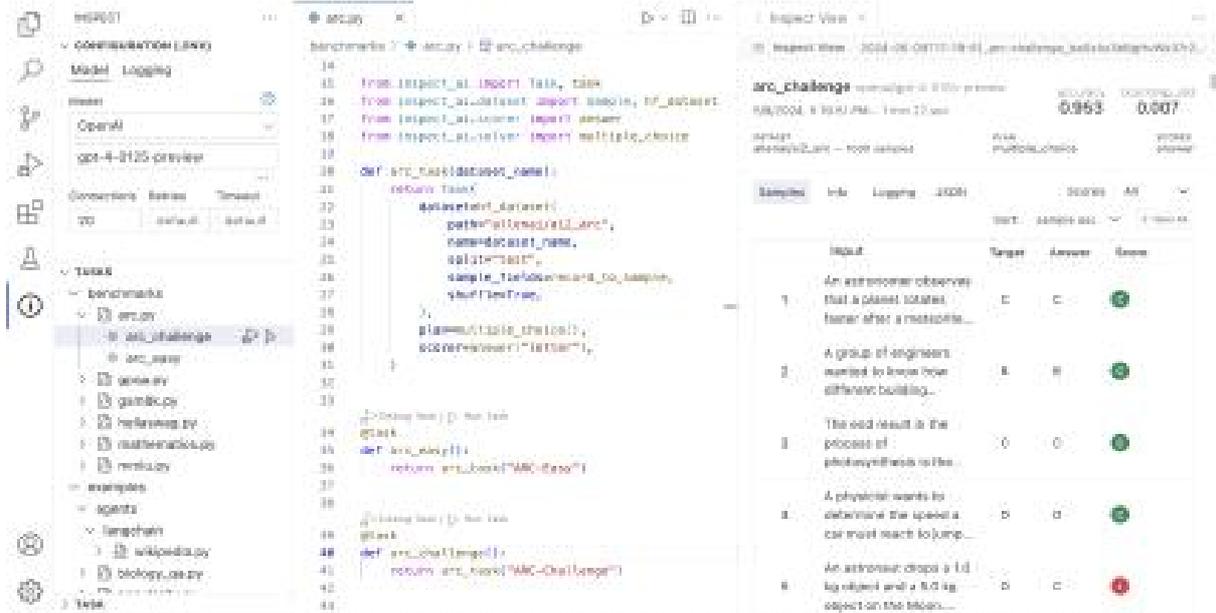
(2) AI 안전성 테스트 플랫폼 ‘인스펙트’ 개발⁷⁷⁾

영국 AI안전연구소는 2024년 5월 10일 글로벌 AI 안전성 평가를 촉진하기 위한 AI 안전성 테스트 플랫폼 ‘인스펙트(Inspect)’를 오픈소스로 공개했으며, 인스펙트를 사용해 다양한 집단이 AI 안전성 평가를 한층 쉽게 수행하고 연구자와 개발자 간 협업을 강화할 수 있다고 강조했다. 영국은 이번 플랫폼을 통해 전 세계적으로 일관된 접근방식의 AI 안전성 평가가 가능해지는 한편, 더욱 안전한 모델 개발로 이어질 것으로 기대하였다. 영국 AI안전연구소는 인스펙트 공개와 함께 공공 부문의 AI 도입 활성화를 담당하는 AI 인큐베이터(i.AD)를 비롯한 AI 관련 기관과 협력해 오픈소스 AI 안전 도구도 개발할 계획이다.

인스펙트는 스타트업, 학계, 개발자, 정부기관 등 다양한 평가자들이 개별 모델의 기능을 평가하고 결과에 따라 점수를 매길 수 있는 소프트웨어 라이브러리로 구성된다. 프롬프트 엔지니어링, 도구 사용, 모델 채점 평가 기능과 같은 기본 구성요소를 제공하는 인스펙트는 핵심 지식, 추론 능력, 자율성 등 다양한 영역에서 모델을 평가할 수 있다. 인스펙트를 통해 오픈AI, 앤스로픽, 구글, 미스트랄, 메타 등의 주요 AI 모델 외에 MS 애저(Azure) AI, AWS 베드락(BedRock), 클라우드플레어(Cloudflare) 플랫폼에서 제공하는 모델도 검사할 수 있다.

77) UK AI Safety Institute, AI Safety Institute releases new AI safety evaluations platform, 2024.05.10.

[그림 3-6] 영국 AI안전연구소의 ‘인스펙트’ 플랫폼 화면



출처: Gov.uk, 2024.10.

(3) AI 모델 테스트 결과 공개

영국 AI안전연구소는 2024년 5월 20일 발간한 제4차 경과 보고서를 통해 현재까지 대중에 공개된 5개 모델에 대하여 ▲사이버 공격 가능성 ▲화학과 생물학적 오용 가능성 ▲자율성 ▲안전장치 성능을 평가하고 결과를 공개하였다. 모델명을 공개하지 않은 평가 결과에서, 일부 모델은 화학과 생물학에 대하여 전문가 수준의 지식을 갖춘 것으로 나타났다.

모델을 이용한 사이버 공격 가능성 평가에서 일부 모델은 고등학생 대상의 단순한 사이버 보안 과제는 수행할 수 있었으나, 대학생 대상의 과제 해결 능력은 부족한 것으로 확인되었다. 인간의 통제를 벗어나 자율적으로 수행할 수 있는 에이전트 역량에 대한 평가에서는 일부 첨단 모델은 소프트웨어 엔지니어링과 같은 단기 과제는 수행할 수 있었으나 장기 작업은 처리하지 못했다. 탈옥과 같은 유해한 결과를 막기 위한 안전장치의 성능 평가에서는 모든 모델이 탈옥 공격에 매우 취약한 것으로 나타났으며, 일부 모델은 보호장치를 우회하려는 시도가 없어도 유해한 결과를 생성하였다.

〈표 3-41〉 영국 AI안전연구소의 AI 모델 평가 내용

구분	주요 내용
사이버 공격	<ul style="list-style-type: none"> • (평가) 포렌식, 암호학, 리버스엔지니어링 등 사이버 공격을 위한 기본 작업 수행 평가 • (결과) 대학 수준의 문제 해결과 취약한 암호체계 악용 관련 작업에는 한계 존재
화학 및 생물학 오용 가능성	<ul style="list-style-type: none"> • (평가) 전문용어를 포함한 문답을 바탕으로 해당 분야의 전문지식을 평가 • (결과) 전문가 수준의 지식을 제공하며 몇몇 모델은 박사 수준의 전문가의 답변과 동등한 수준
자율성	<ul style="list-style-type: none"> • (평가) 인간의 감독 없이 코드 실행, 탐색 등의 작업 수행 가능 여부에 대한 평가 • (결과) 일부 모델은 간단한 작업을 해결하였으나 아직 장기간의 작업 수행은 어려움
안전장치	<ul style="list-style-type: none"> • (평가) 명백히 유해한 질문과 모델이 출력을 피하도록 학습된 정보를 유도하는 공격 시행 • (결과) 대부분 모델이 간단한 탈옥 공격에서도 유해한 질문에 그대로 답변

출처: 소프트웨어정책연구소, 2024.07.

(4) 첨단 AI의 안전에 관한 국제과학보고서 발간

영국 AI안전연구소는 2024년 5월 17일 ‘첨단 AI의 안전에 관한 국제과학보고서’의 중간 보고서를 발간하였다.⁷⁸⁾ 이는 2023년 11월 영국에서 열린 AI 안전성 정상회의의 후속 조치로, 30개국이 참여한 국제 전문가들은 요슈아 벤지오(Yoshua Bengio) 교수 주도로 첨단 AI 시스템의 안전성에 대한 과학 기반의 이해를 위한 보고서를 작성하고 2024년 5월 21~22일 열리는 AI 서울 정상회의에 앞서 중간 보고서를 발표하였다. 중간 보고서는 다양한 작업을 처리할 수 있는 범용 AI의 위험을 이해하고 관리하는 방법에 중점을 두고 있으며, 불확실한 AI의 미래에서 정부와 사회의 결정이 중대한 영향을 미칠 것으로 예상하였다.

보고서에 따르면 범용 AI는 최근 몇 년간 급속도로 발전했으며, 2026년 말까지 일부 첨단 모델은 2023년 대비 40~100배 더 많은 컴퓨팅 자원을 사용하고, 학습 방식에서는 컴퓨팅 자원을 3~20배 더 효율적으로 활용하게 될 전망이다. 최첨단 AI 모델 교육에 사용되는 컴퓨팅 자원은 연간 4배, 교육 데이터셋 규모는 2.5배, 알고리즘 효율성은 1.5~3배 증가하는 추세이다. 범용 AI를 이해하기 위한 여러 연구에도, 범용 AI 모델과 시스템 작동방식과 사회적 영향에 대한 전반적인 이해 수준은 제한적으로, 개발자들은 범용 AI 모델의 작동방식을 거의 이해하지 못한 상태이며, 광범위한 용례를 가진 범용 AI의 위험평가에 대한 포괄적 방법론의 부재로 AI 시스템의 사회적 영향을 평가하기도 어려운 상황이다.

78) UK AI Safety Institute, International Scientific Report on the Safety of Advanced AI, 2024.05.17.

범용 AI는 개인과 공공의 안전과 복지에 심각한 위협을 초래할 수 있으며, 위협은 ▲악의적 사용 ▲오작동 위험 ▲시스템적 위험의 세 가지 범주로 분류될 수 있다. 가령 AI를 사용해 허위정보를 유포하고 사이버 범죄를 저지를 수 있으며, AI 시스템의 편향과 개인정보 침해, 저작권 침해 가능성도 존재한다. 시스템적 위험은 일자리 상실이나 국가간 AI 격차 확대, 금융과 의료 등 주요 부문에서 AI 의존도 확대에 의한 대규모 오류 가능성 등을 의미한다.

범용 AI의 위험 완화를 위해 다양한 기술적 접근방식이 등장하고 있으나 이중 범용 AI의 위험을 안정적으로 예방하는 확실한 방법은 없으며 개선이 필요한 상황이다. 현재 사용되는 위험 평가와 완화 방법은 특정 사용사례로 기능을 제한하거나 레드팀을 통한 모니터링, 교육 데이터셋 검토가 대표적이다.

4) 경쟁시장청(CMA)

영국 경쟁시장청(CMA)은 2024년 4월 11일 AI 기반모델에 대한 검토보고서를 발간하였다.⁷⁹⁾ CMA는 2023년 9월 AI 기반모델에 대한 1차 검토보고서를 발간하고 기반모델의 혁신적 혜택을 온전히 누릴 방안을 제시했으며, 이후 이해관계자 피드백과 기반모델 분야의 발전을 고려해 이번 보고서를 통해 내용을 업데이트하고, 기반모델이 소비자, 기업, 영국 경제에 긍정적 영향을 미치기 위한 AI 원칙을 제시하였다.

<표 3-42> 영국 경쟁시장청의 AI 이행 원칙

구분	주요 내용
접근성	• 데이터, 컴퓨팅 자원, 전문성, 자본에 대한 접근성 보장
다양성	• 기업과 소비자가 필요에 따라 다양한 모델을 선택 가능
선택 가능성	• 자체개발, 파트너십, API나 플러그인 등 충분한 선택권 제공
공정성	• 제품과 서비스 품질에 따른 성공을 보장하고 반경쟁적 행위 금지
투명성	• 소비자와 기업에 기반모델의 위험성과 한계에 대한 정보를 제공
책임성	• 기반모델 개발자와 배포자는 기반모델의 결과물에 책임 보유

출처: 경쟁시장청, SPRI 정리, 2024.04.

79) CMA, AI Foundation Models: update paper, 2024.04.11

CMA에 따르면 디지털 시장에서 이미 시장 지배력을 가진 소수의 빅테크 기업들은 시장 경쟁을 저해하고 선택권과 품질을 해치며 가격을 인상하는 등 기업과 소비자에게 피해를 끼칠 우려가 있다. 가령 기반모델 개발에 중요한 입력(컴퓨팅 인프라, 데이터, 전문지식)을 통제하는 기업은 경쟁을 회피하기 위해 해당 자원에 대한 접근을 제한할 수 있다. 또한 강력한 기존 사업자는 소비자나 기업 시장에서 자신의 입지를 악용해 기반모델 서비스 선택권을 왜곡하고 시장 경쟁을 제한할 수 있으며, 주요 시장 플레이어 간 파트너십은 기반모델 가치사슬을 통해 기존의 시장 지배력의 강화나 확장에 기여할 수 있다. CMA는 소비자의 신뢰를 증진하고 시장 경쟁을 촉진하기 위해 기반모델의 안전성·신뢰성·책임성을 개선할 수 있는 규제 개입을 고려 중으로, 기업의 불공정 관행이 드러날 경우, 집행조치를 통해 대응할 계획이다.

제6절 캐나다

1. 그간의 인공지능 정책 추진 경과

1) 국가 AI 전략⁸⁰⁾⁸¹⁾⁸²⁾

캐나다 연방정부는 2016년부터 과학적 발견을 지원하고, 연구 인재를 육성하며, 전 세계 최고 수준의 연구 인력을 유치하기 위해 160억 캐나다달러(CAD) 이상을 투자할 것을 약속하고, 캐나다의 AI 생태계와 디지털 인프라 전반의 성장을 촉진하기 위해 20억 CAD 이상을 투자해왔다.

캐나다는 AI 분야의 글로벌 리더로서의 입지를 공고히 하고, 기업의 경쟁력을 강화하며, 국민들에게 디지털 경제의 성장을 통한 혜택을 제공하기 위해 2017년 3월에 전 세계 최초의 국가 AI 전략인 「범캐나다 인공지능 전략(Pan-Canadian Artificial Intelligence Strategy)」을 수립했다. 2017년 동 전략을 통해 미래 신산업 육성을 위해 1.25억 CAD 규모를 투자할 것을 약속하고, 「범캐나다 AI 전략」 1단계에 착수하여 고급인재 확보, 연구역량 강화, 경제·사회·정책·법제 등의 글로벌 리더십 확대, 연구인력 커뮤니티 지원, 관련 정책을 추진했다. 특히, 캐나다 정부는 AI 진흥을 위해 AI 허브도시인 앨버타주의 에드먼튼, 온타리오주의 토론토, 퀘벡주의 몬트리올에 국가 AI 연구를 주도할 3대 연구소를 지정하고, 해당 연구소를 중심으로 AI 분야 연구개발을 추진한다.

- 앨버타기계지능연구소(Alberta Machine Intelligence Institute, Amii)
- 벡터연구소(Vector Institute)
- 몬트리올학습알고리즘연구소(Montreal Institute for Learning Algorithms, Mila)

80) Prime Minister of Canada, Securing Canada's AI advantage, 2024.4.7.:

<https://www.pm.gc.ca/en/news/news-releases/2024/04/07/securing-canadas-ai>

81) Government of Canada, Pan-Canadian Artificial Intelligence Strategy, 2022.7.20.:

<https://ised-isde.canada.ca/site/ai-strategy/en>

82) Government of Canada launches second phase of the Pan-Canadian Artificial Intelligence Strategy:

<https://www.canada.ca/en/innovation-science-economic-development/news/2022/06/government-of-canada-launches-second-phase-of-the-pan-canadian-artificial-intelligence-strategy.html>

캐나다 정부는 2022년에 6월 AI 투자, 글로벌 인재 양성, 첨단 연구 역량 구축, 책임 있는 AI 도입과 상용화를 촉진하기 위해 10년 동안 4억 4,300만 CAD 이상의 예산을 투입하는 「범캐나다 AI 전략」 2단계를 발표하였다. 「범캐나다 AI 전략」은 캐나다고등연구소(CIFAR)와 3대 국립 AI 연구소를 중심으로 캐나다 전역의 파트너들과 협력하여 상용화, 표준, 인재·연구의 3개 핵심 영역을 중심으로 AI 정책을 추진한다.

<표 3-43> 범캐나다 AI 전략의 3대 핵심 영역

구분	주요 내용
상용화 (Commercialization)	<p>(국가 AI 연구소)</p> <ul style="list-style-type: none"> 국립 AI 연구소(에드먼턴의 Amii, 몬트리올의 Mila, 토론토의 Vector Institute)는 AI 연구 결과를 상용화하고 기업이 이러한 신기술을 채택할 수 있는 역량을 키우도록 지원 캐나다 정부는 2021년 예산에 편성된 6천만 CAD를 2021~22년 회기부터 2025~26년 회기까지 5년에 걸쳐 각 연구소당 최대 2천만 CAD까지 지원 가능 <p>(캐나다 글로벌 혁신 클러스터)</p> <ul style="list-style-type: none"> 캐나다의 글로벌 혁신 클러스터(Digital Technology, Protein Industries Canada, Next Generation Manufacturing Canada, Scale AI, Ocean Supercluster)는 주요 산업 분야의 기업, 공공 및 비영리 단체의 캐나다산 AI 기술 채택을 촉진함으로써 캐나다의 혁신을 강화 캐나다 정부는 2021년 예산에 편성된 1억 2,500만 CAD를 2021~22년 회기부터 2025~26년 회기까지 5년에 걸쳐 지원
표준 (Standards)	<ul style="list-style-type: none"> 캐나다 정부는 캐나다 표준위원회(Standards Council of Canada)를 통해 AI 관련 표준의 개발 및 채택을 촉진하기 위한 노력을 지원 캐나다 정부는 2021년 예산에 편성된 860만 CAD를 2021~2022년 회기부터 2025~2026년 회기까지 5년에 걸쳐 지원
인재 및 연구 (Talent and Research)	<p>(CIFAR)</p> <ul style="list-style-type: none"> CIFAR는 학술연구 인재를 유치, 유지 및 개발하기 위한 프로그램과 함께 3대 국립 AI 연구소가 진행하는 연구 및 교육을 지원하기 위한 프로그램을 강화하고, 고급 연구, 교육 및 지식 참여 프로그램을 재개 캐나다 정부는 2021년 예산에 편성된 2억 800만 CAD를 2021~ 2022년 회기부터 2030~2031년 회기까지 10년에 걸쳐 지원 <p>(컴퓨팅)</p> <ul style="list-style-type: none"> 캐나다디지털연구연합(Digital Research Alliance of Canada)은 동 전략의 목표 달성을 지원하기 위해 캐나다 전역의 AI 연구자에게 전용 컴퓨팅 용량을 제공 캐나다 정부는 2021년 예산에 편성된 4천만 CAD를 2022~2023년 회기부터 2026~2027년 회기까지 5년에 걸쳐 지원

출처: Government of Canada, Pan-Canadian Artificial Intelligence Strategy, 2022.07.20.

2) AI 입법화 및 규제 현황⁸³⁾⁸⁴⁾⁸⁵⁾

현재 캐나다 내에는 AI 관련 규제 프레임워크가 부재한 상황이며, 보건 및 금융 등 특정 분야의 일부 규정이 AI의 특정 사용 사례에 적용되나 AI 시스템이 설계 및 개발 과정에서 시스템적 위험을 해결하도록 보장하는 근본적인 접근방식은 아직 마련되어 있지 않은 실정이다. 캐나다 정부는 AI의 기능과 배포 규모가 진화함에 따라 국민들이 AI 기술을 신뢰할 수 있도록 보장하기 위한 표준의 필요성을 인식하고 캐나다 경제 전반에 걸쳐 책임감 있는 AI 개발과 도입을 지원하기 위한 노력을 이어가고 있다.

캐나다 정부는 2022년 6월 「디지털 헌장(Digital Charter)」 이행을 위한 「디지털 헌장 이행 법안 (Digital Charter Implementation Act, 2022, Bill C-27)」의 일환으로 세계 최초의 AI 규제 프레임워크 중 하나인 「인공지능 및 데이터 법안(AIDA)」을 제안했다. 현재 AIDA는 하원에서 2차 독회 완료 후 산업정보 상임위원회에 회부된 상태이며, 의회는 법 제정을 위한 추가 논의를 이어가고 있다.

Bill C-27은 ▲(제1부) 소비자 개인정보보호법 ▲(제2부) 개인정보 및 데이터 보호 재판소법 ▲(제3부) 인공지능 및 데이터 법안으로 구성되어 있으며, 이 중 제3부에 해당하는 AIDA는 AI 시스템의 국제 또는 국내 무역 및 상거래를 규제하기 위한 새로운 조치를 규정한다. 동 법은 AI 시스템으로 인한 피해 및 편향된 결과물의 위험성을 완화하기 위한 조치를 포함하여, AI 시스템의 설계, 개발 및 사용에 관한 공통 요구사항을 설정하고, 개인 또는 개인의 이익에 심각한 피해를 초래할 수 있는 데이터 및 AI 시스템의 특정 관행을 금지한다.

AIDA는 책임있는 혁신을 지원하고 캐나다 대기업의 국제 시장 진출을 보장하는 동시에 중소기업의 요구사항을 고려하는 균형 잡힌 AI 규제 접근법을 제시하며, 기업에게는 AI의 잠재력을 혁신하고 실현하는 데 도움이 되는 명확한 규칙을 제공하고, 국민에게는 캐나다

83) Parliament of Canada, Bill C-27, An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts, 2023.4.24.: (<https://www.parl.ca/legisinfo/en/bill/44-1/c-27>)

84) Government of Canada, Artificial Intelligence and Data Act, 2023.9.27.: (<https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act>)

85) Government of Canada, Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems, 2023.5.28.: (<https://ised-isde.canada.ca/site/ised/en/voluntary-code-conduct-responsible-development-and-management-advanced-generative-ai-systems>)

내에서 사용되는 AI 시스템이 안전하고 사용자의 이익을 최우선으로 고려하여 개발된다는 사실을 명확화한다.

또한, 캐나다 정부는 2023년 9월 캐나다 기업의 책임있는 생성 AI 시스템 개발 및 사용을 입증할 수 있도록 지원하기 위해 「고급 생성 AI 시스템의 개발 및 관리에 관한 자발적 행동강령(Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems)」을 발표하고, 생성 AI 기술에 대한 캐나다 국민의 신뢰를 강화하기 위한 방안을 모색하였다.

동 행동강령에 생성 AI 기술의 확산 및 부상에 따른 위험을 해결하고 완화하기 위해 범용 기능을 갖춘 생성 AI 시스템을 개발하거나 관리하는 모든 기업이 AIDA에 따른 구속력 있는 규제에 앞서 적용해야 하는 조치 함께, 광범위하고 잠재적으로 유해하거나 부적절하게 사용될 가능성이 있는 생성 AI 시스템을 개발하거나 관리하는 기업이 취해야 하는 추가 조치를 식별한다.

동 행동강령은 고급 생성 AI 시스템의 개발자와 관리자가 자발적으로 준수해야 하는 책임성, 안전성, 공정성 및 형평성, 투명성, 인적 감독 및 모니터링, 유효성 및 견고성에 관한 원칙을 제시하며, 행동강령에 서명한 조직은 이를 채택할 것을 자발적으로 약속하고 준수하기 위한 노력에 동참한다.

또한, 캐나다 정부는 AI 규제에 관한 국제적인 논의에 적극적으로 참여하고 있으며, 올해 1월 영국 정부와 AI 컴퓨팅에 관한 양해각서(MoU)⁸⁶⁾, 2월 EU와 AI, 양자컴퓨팅 사이버보안 등에 관한 디지털 파트너십⁸⁷⁾,을 체결하는 등 책임있는 AI 개발 및 사용을 위해 전 세계 파트너와 지속적으로 협력하고 있다.

86) GOV.UK, UK-Canada cooperation in AI compute: memorandum of understanding, 2024.1.13.:
(<https://www.gov.uk/government/publications/uk-canada-cooperation-in-ai-compute-memorandum-of-understanding>)

87) EC, EU and Canada boost their strategic digital partnership to address new challenges, 2024.2.1.:
(https://ec.europa.eu/commission/presscorner/detail/en/ip_24_614)

2. 2024년 인공지능 정책 동향

1) 연방기관의 생성 AI 사용에 관한 가이드⁸⁸⁾

2024년 10월, 캐나다 정부는 2023년 9월 6일 처음 발표한 「생성 AI 사용에 관한 가이드 (Guide on the use of Generative AI)」를 업데이트하고, 올해 2월 20일 동 가이드의 두 번째 버전을 발표했다. 동 가이드는 캐나다 연방기관의 생성 AI 도구 사용에 대한 예비 지침으로, 생성 AI에 관한 개요를 포함하여 관련 도전과제와 우려 사항을 식별하며 책임있는 생성 AI 사용을 위한 원칙, 정책 고려사항 및 모범사례 등을 제시한다. 캐나다 정부는 기존 가이드에 대한 다양한 이해관계자들의 피드백을 반영하고, 연방기관의 책임 및 문서화 요건에 관한 부분과 연방기관이 제기한 질문에 대한 응답으로 자주 묻는 질문(FAQ)을 새롭게 추가했으며, 특히, 대중의 신뢰를 유지하고 생성 AI 도구의 책임있는 사용을 보장하기 위해 연방기관이 준수해야 할 원칙인 ‘FASTER’ 원칙의 정의를 업데이트했다.

- (Fair) 공정성: 생성 AI 도구가 생성한 콘텐츠가 편견을 포함하거나 증폭하지 않고 인권, 접근성, 절차적 및 실질적 공정성 의무를 준수하는지 확인하고, 배포 전 영향을 받는 이해관계자와 소통
- (Accountable) 책임성: 생성 AI 도구가 생성한 콘텐츠가 정확하고 합법적이며 윤리적이고 이용 약관을 준수하도록 해당 콘텐츠 및 이에 따른 영향에 대한 책임을 부담하며, 모니터링 및 감독 메커니즘을 구축
- (Secure) 보안: 생성 AI 인프라와 도구가 정보 보안 분류에 적합하며 프라이버시와 개인 정보가 보호되는지 확인하고, 시스템 배포 시 사이버보안 위험과 견고성을 평가 및 관리
- (Transparent) 투명성: 생성 AI를 통해 생산된 콘텐츠를 식별하고, 사용자에게 생성 AI 도구와 상호작용하고 있음을 알리며, 생성 AI 도구 배포 시 기관의 정책, 적절한 사용, 데이터 및 모델 학습에 관한 정보를 제공하고, 의사결정을 지원하기 위해 생성 AI 도구를 사용하는 경우 결정 도출 과정을 문서화하고 이에 대한 설명을 제공
- (Educated) 교육: 생성 AI 도구의 이점, 한계 및 책임있는 사용에 대한 정보 및 지식을 습득하고, 효과적으로 프롬프트를 생성하고 결과물에서 잠재적 취약점을 식별하는 방법을 학습
- (Relevant) 관련성: 생성 AI 도구의 사용이 사용자 및 조직의 요구사항을 지원하고 이

88) Government of Canada, Guide on the use of generative artificial intelligence, 2024.10.15.:
(<https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/possible-use-ai/guide-use-generative-ai.html>)

로 인한 결과가 대중을 위한 서비스 품질 개선에 기여하는지 확인하고, 생성 AI 도구 사용 선택 시 환경에 미치는 영향을 고려하며, 업무에 적합한 도구를 식별하고 생성 AI 도구가 모든 상황에 최선 선택은 아닐 수 있음을 인지

<표 3-44> FASTER 원칙 일상 업무에서의 생성 AI 사용지침

원칙	해야 할 일(Do)	하지 말아야 할 일(Don't)
(Fair) 공정성	<ul style="list-style-type: none"> 포괄적이고 공정한 답변을 생성할 수 있는 방식으로 지침을 작성 생성된 결과물이 대표성을 띄고 포괄적이며, 유해한 고정관념이 포함되어 있지 않은지 확인 	<ul style="list-style-type: none"> 편향적이거나 배타적이거나 인구 집단을 잘못 표현하는 결과물을 사용하지 않음
(Accountable) 책임성	<ul style="list-style-type: none"> AI를 사용하여 생성한 콘텐츠에 대한 책임을 저야함 생성된 콘텐츠가 정확한지 검토 생성된 콘텐츠에 저작권의 보호를 받는 자료가 포함되어 있는지 확인 	<ul style="list-style-type: none"> 결과물이 정확하다고 가정하지 않음 사람이 완료해야 하는 작업을 생성 AI에 위임하지 않음
(Secure) 보안	<ul style="list-style-type: none"> 기밀정보가 아닌 데이터에 대해서만 공개 도구 사용 도구의 이용 약관 이해 대화 내용을 저장하지 않도록 도구 설정 	<ul style="list-style-type: none"> 공개 도구에 개인정보, 민감한 정보 또는 보호 대상 정보를 입력하지 않음 도구의 안전 규칙을 우회할 수 있는 방법을 찾지 않음
(Transparent) 투명성	<ul style="list-style-type: none"> 최종 결과물에 생성 AI를 사용했음을 표시 업무 시 AI 도구를 사용했음을 관리자에게 알림 	<ul style="list-style-type: none"> AI가 생성한 콘텐츠를 자신의 작업물로 전달하지 않음
(Educated) 교육	<ul style="list-style-type: none"> 생성 AI 사용 방법 학습 생성 AI의 강점, 약점 및 위험에 대해 학습 생성 AI 사용에 관한 강좌 수강 및 정보 습득 	<ul style="list-style-type: none"> 검증할만한 전문지식을 보유하지 않은 내용의 자료를 생성하는데 생성 AI 도구를 사용하지 않음 기술이 진화함에 따라 도구에 대한 일회성 학습으로 충분하다고 가정하지 않고 지속적으로 학습
(Relevant) 관련성	<ul style="list-style-type: none"> 생성 AI가 모든 용도에 적합한 것을 아니라는 점을 기억 업무에 도움이 되는 경우에만 사용 	<ul style="list-style-type: none"> 미묘한 뉘앙스 및 감정적 언어의 사용이 요구되는 작업 등 연구를 통해 생성 AI의 수행 능력이 검증되지 않은 작업에 이를 사용하지 않음 생성 AI 도구에 지나치게 의존하지 않음

출처: Government of Canada, Guide on the use of generative artificial intelligence. 2024.10.15.

동 가이드는 생성 AI 사용 관련 주요 이슈를 개인정보보호, 편향성, 품질, 인적 자율성, 법적 위험, 인간과 기계의 분리, 환경적 영향의 7가지 부문으로 구분하고, 각 부문별 잠재적 이슈를 식별하며 연방기관 내 생성 AI 사용자를 위한 모범사례 및 생성 AI 도구 배포 기관을 위한 추가 모범사례를 제시한다.

〈표 3-45〉 생성 AI의 잠재적 이슈와 모범사례

구분	잠재적 이슈	연방기관 내 생성 AI 도구 사용자를 위한 모범사례
개인정보 보호	일부 생성 AI 도구는 정부의 정보 보안 요구 사항을 충족하지 못함	<ul style="list-style-type: none"> • 정부에서 관리하지 않는 생성 AI 도구에는 민감한 정보나 개인정보를 입력하지 않음 • 공개 시 대중의 신뢰를 훼손할 수 있는 내용 관련 쿼리(정보 요청)를 정부가 관리하지 않는 생성 AI 도구에 제출하지 않음 • 시스템이 입력 데이터를 사용하는 방식을 이해 • 민감하거나 독점적인 정보를 처리하기 위해 시스템을 사용하기 전 법률 자문, 부서별 최고 보안 책임자(CSO), 개인정보보호 담당부서에 공급업체의 이용 약관, 개인정보보호 정책 및 기타 법률 문서를 검토하도록 요청 • 보안 관리 지침(Directive on Security Management)에 따라 정보 보안 분류에 적합한 인프라와 도구를 사용 • 기밀이나 민감정보를 위해 생성 AI를 사용, 조달 또는 배포 전에 부서별 CSO의 승인을 요청 • 서비스 및 디지털 지침(Directive on Service and Digital)과 서비스 및 디지털 가이드라인(Guideline on Service and Digital)의 관련 지침에 명시된 정보 및 데이터 보존 요건을 고려 • 가능한 경우 ‘옵트아웃(Opt-out)’ 기능을 사용하여 프롬프트가 AI 시스템을 학습하거나 추가 개발하는 데 활용되지 않도록 보장
편향성	생성된 콘텐츠는 학습데이터에 지배적으로 내포된 편향성이나 기타 유해 아이디어를 증폭시킬 수 있음	<ul style="list-style-type: none"> • 편향성, 다양성, 포용성, 반인종차별, 가치 및 윤리에 대한 이해도를 높여 편향적, 비포용적 또는 차별적인 콘텐츠를 식별하는 능력을 향상 • 생성된 콘텐츠가 정부의 약속, 가치 및 윤리에 부합하고 법적 의무를 준수하는지 검토 • 총체론적 관점을 제공하고 편향성을 최소화하는 콘텐츠를 생성하기 위해 프롬프트를 공식화 • 생성 AI 도구를 학습하는 데 사용된 데이터(예: 해당 데이터의 출처, 포함 내용, 선택 및 준비 방식 등)에 관한 이해도 제고 • 생성 AI를 통해 콘텐츠가 제작된 시점을 명확하게 표시
품질	생성된 콘텐츠는 부정확하거나 일관성이 없거나 불완전할 수 있음	<ul style="list-style-type: none"> • 콘텐츠 개발에 생성 AI가 사용되었음을 명확하게 표시 • 생성된 콘텐츠를 권위있는 자료로 간주하지 않으며, 신뢰할 수 있는 출처의 정보와 대조하거나 전문성을 갖춘 동료의 답변을 확보하는 등 사실적 및 맥락상의 정확성을 검토 • 생성 AI를 사용하여 생성된 개인정보의 정확성, 최신성, 완전성을 검토 • 결과물의 정확성에 관한 영향을 평가하고, 사실적 정확성 또는 데이터 무결성이 요구되는 경우 생성 AI를 사용하지 않음 • 학습 데이터의 품질과 출처에 대한 이해도 제고 • 생성 AI를 사용하기 전에 부정확한 콘텐츠를 식별하는 능력을 고려하며, 콘텐츠 품질을 식별할 수 없는 경우 이를 사용하지 않음 • 부정확한 콘텐츠 생성을 최소화하도록 결과물을 개선하기 위해 효과적인 프롬프트를 작성하고 피드백을 제공하는 방안을 탐색 • 출처가 제공되어 콘텐츠를 검증할 수 있는 경우를 제외하고는 생성 AI 도구를 검색 엔진으로 사용하지 않음

구분	잠재적 이슈	연방기관 내 생성 AI 도구 사용자를 위한 모범사례
인적 자율성	AI에 대한 과도한 의존은 자율적인 판단을 방해하고, 창의성을 억압하며, 인적 역량을 약화시킬 수 있음	<ul style="list-style-type: none"> • 사용자 및 조직의 필요를 충족하기 위해 생성 AI 사용이 필수적인지 고려 • 업무를 할당하고 시스템 활용 결과물을 검토할 때 생성 AI의 성능과 한계를 고려 • 생성 AI 도구와 이로 인한 결과물에 대한 비판적인 평가가 가능하도록 AI 활용 능력(literacy)을 향상 • 생성 AI 도구를 대체 도구가 아닌 보조 도구로 사용하며, 자신이 이해하지 못하거나 보유하지 않은 스킬을 아웃소싱하지 않음 • AI 도구에 아이디어나 의견 추천을 요청하기 전에 자신의 견해를 형성 • 편향된 결과를 최소화하기 위해 프롬프트 작성 시 중립적인 표현을 사용 • 시스템이 정확한 답변을 제공하는 데 있어 신뢰할 수 있는 것처럼 보이는 경우에도 생성 AI가 생성한 콘텐츠를 항상 검토
법적 위험	생성 AI는 인권, 개인정보보호, 지식재산권 보호 및 절차적 공정성에 위험을 초래	<ul style="list-style-type: none"> • 생성 AI 도구를 배포하거나 서비스 제공에 사용하는 경우, 이에 따르는 법적 위험에 대해 기관별 법률자문 담당자와 협의하며 공급업체의 사용 약관, 저작권 정책, 개인정보보호 정책 또는 법률 문서 등을 검토 • 행정적 의사결정에 생성 AI를 사용하는 경우 자동화된 의사결정에 관한 지침(Directive on Automated Decision-Making)을 준수 • 시스템 결과물이 저작권의 보호를 받는 자료와 동일 또는 실질적으로 유사한지 확인하고, 적절한 출처를 밝히거나 문제가 되는 자료를 제거하여 지식재산권 침해 위험을 최소화 • 출판 절차(Procedures for Publishing)에 따라 결과물을 대중에 공개하려는 경우 저작권의 허가 및 행정절차에 관해 담당자와 협의 • 시스템 결과물에 정부가 추구하는 가치와 충돌할 수 있는 사실적 부정확성, 편향성 또는 유해 아이디어가 내포되어 있는지 평가 • AI 규제와 관련된 법률 및 정책 개발과 관련하여 최신 정보를 파악
인간과 기계의 구분	대중이 AI 시스템과 상호작용하고 있다는 사실을 모르거나, 반대의 경우 AI가 사용되고 있다고 잘못 가정할 수 있음	<ul style="list-style-type: none"> • 생성 AI를 사용하는 자동 이메일 작성 도구, 대화형 도구 또는 챗봇 등 대중과의 상호작용에서 AI가 사용되는 시점 및 방식을 명확하게 전달 • 생성 AI가 생성한 텍스트, 오디오 또는 시각적 메시지가 사용자에게 전달되는 경우 이를 사용자에게 전달
환경적 영향	생성 AI 시스템의 개발과 사용은 상당한 환경적 비용을 초래할 수 있음	<ul style="list-style-type: none"> • 넷제로(zero-emission) 또는 탄소 중립 데이터 센터에서 호스팅 되는 생성 AI 도구를 사용 • 프로그램 목표 및 결과 달성과 관련된 경우에만 생성 AI 도구를 사용 • 생성 AI의 환경적 영향은 이를 사용하는 경우뿐만 아니라 모델 학습 및 데이터 센터 운영을 위한 하드웨어 구성요소의 제조(채굴, 조립, 운송, 폐기 등 포함) 시에도 발생한다는 사실을 이해

출처: Government of Canada, Guide on the Use of Generative Artificial Intelligence, 2024.10.

2) AI 역량 강화를 위한 2024 예산안 편성⁸⁹⁾

올해 4월 7일, Justin Trudeau 캐나다 총리는 캐나다 AI 안전연구소 설립 및 AIDA 집행 강화 등 캐나다의 AI 역량을 강화하기 위해 다가오는 2024년 예산안에 24억 캐나다달러(CAD) 규모의 조치를 포함할 것이라고 발표했다.

- (20억 CAD) AI 연구자, 스타트업, 스케일업을 위한 컴퓨팅 역량과 기술 인프라를 구축하고 이에 대한 접근성 제공
- (2억 CAD) 캐나다 지역개발 기관을 통해 AI 스타트업이 새로운 기술을 시장에 출시하고, 농업, 청정 기술, 의료 및 제조와 같은 핵심 부문에서의 AI 도입을 가속화 하도록 지원
- (1억 CAD) 중소기업이 새로운 AI 솔루션 구축 및 배포를 통해 규모를 확대하고 생산성을 높일 수 있도록 캐나다 국립연구위원회(NRC) 산업 연구 지원 프로그램(IRAP)의 AI 지원 프로그램 지원
- (5천만 CAD) AI의 영향에 따른 잠재적 피해 대상 분야와 지역사회의 근로자에게 새로운 스킬 교육을 제공하기 위한 부문별 인력 솔루션 프로그램(Sectoral Workforce Solutions Program) 지원
- (5천만 CAD) AI의 안전한 개발과 배포를 촉진하기 위해 캐나다 AI 안전연구소(Canadian AI Safety Institute) 신설
- (510만 CAD) AIDA 집행 강화를 위해 AI 및 데이터 위원회 지원

캐나다 정부는 혁신과 경제 성장에 투자함으로써 현재와 미래 세대를 위한 캐나다의 AI 경쟁우위 확보하여 좋은 일자리를 창출하고, 경제 전반의 혁신을 촉진하며, 생산성을 높이고, 경제 성장을 가속화할 것을 기대한다.

89) Prime Minister of Canada, Securing Canada's AI advantage, 2024.4.7.; (<https://www.pm.gc.ca/en/news/news-releases/2024/04/07/securing-canadas-ai>)

3) 자동화된 의사결정 지침의 적용 범위에 관한 가이드 발표⁹⁰⁾

캐나다 정부는 2024년 6월 「자동화된 의사결정에 관한 지침(Directive on Automated Decision-Making)」을 발표했다. 이 지침은 행정적 결정을 완전 또는 부분적으로 자동화하기 위해 AI 또는 기타 기술 기반 자동화된 의사결정 시스템을 사용하는 부서에 적용되며, 2020년 4월 이후에 개발되었거나 조달된 시스템에 적용된다.

동 지침은 해당 부서가 자동화된 의사결정을 책임있고 투명하며 공정하게 수행하도록 보장하는 것을 목적으로 하며, 이에 따라 해당 부서는 자동화된 의사결정 시스템의 영향을 평가하고, 투명성을 확보하며, 품질을 보장하고, 결정에 관한 근거를 제공하며, 시스템의 효과와 효율성을 공개적으로 보고해야 한다. 동 지침은 해당 부서가 자동화된 의사결정 시스템이 대중과 부서에 미치는 위험을 식별, 평가 및 완화하고, 투명성, 책임성, 합법성 및 절차적 공정성과 같은 행정법의 원칙을 준수하도록 지원한다.

자동화된 의사결정에 관한 지침의 적용과 관련하여 새롭게 발표된 가이드는 지침을 준수하기 위해 동 지침이 적용되는 상황 및 범위에 관한 요건을 제시한다. 시스템이 지침의 적용 범위에 포함되기 위해서는 5가지 핵심 요건을 충족해야 한다.

- (정부 부서에서 사용) 연방재무행정법 제2조의 정의에 따라 명시된 부서에 적용
- (2020년 4월 1일 이후 개발) 2020년 4월 1일 이후에 개발 또는 조달된 자동화된 의사 결정 시스템의 경우, 2020년 4월 1일 이전에 개발 또는 조달된 시스템이 상당부분 조정된 경우에 적용
- (행정적 의사결정 프로세스 내에서 사용) 행정적 의사결정은 법적 권리, 특권 또는 이익에 영향을 미칠 수 있는 모든 결정을 의미하며, 의사 결정 과정의 특정 시점에서 의사결정을 내리거나 이를 지원하는 자동화된 의사결정 시스템에 적용
- (인간의 판단을 대체하거나 지원) 결정론적 규칙 기반 시스템에서 고급 AI 시스템에 이르기까지 광범위한 범주의 자동화된 시스템에 적용
- (생산 환경 내에서 사용) 배포 예정이거나 운영 중인 시스템에만 적용되며 연구 및 실험 중인 시스템의 경우 적용 범위에서 제외

90) Government of Canada, Guide on the Scope of the Directive on Automated Decision-Making, 2024.6.27.; (<https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/possible-use-ai/guide-scope-directive-automated-decision-making.html>)

4) 캐나다 AI 안전연구소 출범⁹¹⁾

캐나다 정부는 AI 안전 위협에 대응하기 위한 국가의 역량을 강화하기 위해 2024년 11월 캐나다 AI 안전연구소(Canadian Artificial Intelligence Safety Institute, CAISI)를 출범했다. AI 기술의 급속한 발전으로 허위 정보 유포, 사이버보안 침해, 선거 개입 등 AI 기술이 악용될 가능성이 증가함에 따라 캐나다 정부는 AI 시스템의 안전하고 책임있는 사용을 보장하는 것을 최우선 과제로 삼고, AI의 잠재적 위협으로부터 국민을 보호하기 위해 구체적인 조치들을 취하고 있다. 캐나다 정부는 연구자와 기업이 안전하고 책임있게 AI 기술을 개발하고 채택할 수 있도록 지원하기 위해 2024 회계연도 예산안에서 24억 CAD 규모의 예산을 편성했으며, 이러한 노력의 일환으로 CAISI를 출범한다고 발표했다.

CAISI는 캐나다의 세계 최고 수준의 AI 연구 생태계와 인재 기반을 활용하여 첨단 AI 시스템 관련 위협에 대한 이해를 높이고 해당 위협을 해결하기 위한 방안 마련을 위한 노력을 이어갈 계획이다. 또한, CAISI는 작년 11월 「AI 안전성 정상회의(AI Safety Summit 2023)」에서 채택된 블레츨리 선언(Bletchley Declaration)을 통한 국제협력에 대한 정부의 약속을 바탕으로 11월 말 출범한 국제 AI 안전연구소 네트워크(International Network of AI Safety Institutes)를 통해 주요국 AI 안전연구소들과의 교류를 촉진할 계획이다.

CAISI 출범은 안전하고 책임있는 AI 개발을 촉진하기 위한 정부의 광범위한 전략의 일환으로, 캐나다 정부는 CAISI와 협력하여 AI의 안전을 보장하고 해당 기술에 대한 신뢰를 구축하기 위한 노력을 강화할 방침이다. 캐나다 정부는 책임있는 AI 도입을 통해 AI에 대한 신뢰도를 높여 향후 캐나다의 AI 우위를 확보하기 위한 방안을 모색해 나갈 계획이다.

91) Government of Canada, Canada launches Canadian Artificial Intelligence Safety Institute, 2024.11.12.: (<https://www.canada.ca/en/innovation-science-economic-development/news/2024/11/canada-launches-canadian-artificial-intelligence-safety-institute.html>)

제7절 프랑스

1. 그간의 인공지능 정책 추진 경과

1) 개요

프랑스는 인공지능(AI) 분야에서 선구자로서의 입지를 확고히 하기 위해 “인공지능 개발 국가전략”을 2018년에 발표하고, 장기적으로 기술발전의 모든 단계에서 인공지능 생태계를 구축하고 다양한 산업 분야에서 기술을 공유하며 인공지능을 위한 지원과 제도를 마련하고자 노력하고 있다.

2021년부터 2025년까지 2차 기간의 목표는 인공지능 분야에 특화된 인재를 양성하고 연구 개발 분야의 잠재력을 경제적 성공으로 전환하는 것이다. 이를 위해 프랑스 정부는 5년간 총 20억 유로의 예산을 투입할 계획이며, 이 중 15억 유로는 공적자금 형태, 5억 유로는 민간과 공동출자 형태로 충당할 예정이다.⁹²⁾

〈표 3-46〉 프랑스 AI 국가 전략 주요 내용과 진행 현황

구분	주요 내용	
1단계 연구역량 강화 (’18~’22)	주요내용	<ul style="list-style-type: none"> • ‘프랑스의 정보 처리 핵심 과학 기술 분야 선도국 도약을 목표로 약 15억 유로 규모의 예산 투입 ①학제간 연구소 네트워크 개발 ②우수 강의 지원 ③박사 과정 프로그램 자금 지원 ④공공 연구를 위한 컴퓨팅 역량 투자
	성과	<ul style="list-style-type: none"> • ’21년 프랑스 내 AI 연구소 수가 유럽 최대인 81개 도달 • ’21년 AI 전문 스타트업 수는 ’20년 대비 11% 증가한 502개를 기록 • ’21년 AI 스타트업 근무자 수는 13,459명으로*, ’22년 9,000명 추가 고용이 필요한 것으로 조사 <p>*그 외 간접 일자리가 70,000개 창출된 것으로 집계</p>

92) Ministère de l'économie, La stratégie nationale pour l'intelligence artificielle. 2024.05.22
(<https://www.economie.gouv.fr/strategie-nationale-intelligence-artificielle>)

구분	주요 내용	
2단계 AI 인재 양성 및 유치 (’21~’25)	주요내용	<ul style="list-style-type: none"> 5년간 총 22.2억 유로를 투입하여 AI 인재 양성 및 R&D 잠재력 제고 향상을 추진(’22.11.8 개시) - AI 기술 확산을 목표로 ①임베디드 AI* ②신뢰할 수 있는 AI(Trustworthy AI) ③생태전환 AI ④초거대 언어 모델 및 생성 AI 분야 우선 개발과 혁신을 지원 *(Embedded AI) 다양한 장치와 시스템에 AI 기능을 직접 통합하는 기술
	주요 목표	<ul style="list-style-type: none"> ’25년까지 약 500개 중소기업 및 중견기업의 AI 솔루션 도입과 이용을 지원 학생 총 3,700명(학부생 2,000명, 석사 과정생 1,500명)을 양성하고, 국가 AI 교육 확대 ’25년 임베디드 AI 분야 글로벌 시장의 15% 점유

출처: entreprise.gouv.fr, La stratégie nationale pour l'intelligence artificielle, 2024.10.18. 검색⁹³⁾

2. 2024년 프랑스 인공지능 정책

1) 프랑스 AI 위원회의 AI 투자 권고⁹⁴⁾

(1) 개요

프랑스 AI 위원회가 2024년 3월 13일 AI 기술 혁신에서 프랑스가 주도권을 잡기 위해 정부가 추진해야 할 정책과제를 제시한 보고서를 발간했다.⁹⁵⁾ 미국과 중국이 국가 전략을 수립하여 AI 기술 주도권 확보에 나선 가운데, 프랑스 역시 AI 발전을 주도하기 위해 제도와 공공 정책의 개혁이 필요하다고 강조했다. 향후 프랑스가 AI 선도국이 되기 위해 정부에 향후 5년간 매년 50억 유로(약 7조 원)를 투자할 것을 권고했다.

(2) 프랑스 정부 AI 우선 추진 과제

보고서는 프랑스 정부가 우선 추진해야 할 정책으로 ▲국민 인식 제고와 교육 ▲재정 투자 ▲컴퓨팅 자원 확보 ▲데이터 접근성 개선 ▲문화 콘텐츠 활용 ▲공공연구 강화 ▲국제 거버넌스를 제시했다.⁹⁶⁾

93) entreprise.gouv.fr, La stratégie nationale pour l'intelligence artificielle, 2024.10.18. 검색
(<https://www.entreprises.gouv.fr/fr/numerique/enjeux/la-strategie-nationale-pour-l-ia>)

94) Gouvernement.fr, 25 recommandations pour l'IA en France, 2024.03.13.
(<https://www.elysee.fr/emmanuel-macron/2024/03/13/25-recommandations-pour-lia-en-france>)

95) 프랑스 정부가 2023년 9월 설립했으며, AI와 신기술 전문가 13명으로 구성되어 정부에 대한 조언과 국가 AI 전략 조정을 담당

96) info.gouv.fr, 25 recommandations pour l'IA en France, 2024.03.13
(<https://www.info.gouv.fr/actualite/25-recommandations-pour-lia-en-france>)

- ① (국민 인식 제고와 교육) AI의 사회경제적 영향에 대한 지속적인 공론의 장을 마련하고, 고등교육 프로그램 및 AI 도구에 대한 평생교육을 확대하며, AI를 사회적 토론의 주제이자 도구로 통합
- ② (재정 투자) AI 생태계 육성과 프랑스의 경제구조 전환을 위해 단기적으로 100억 유로 (14조 5,000억 원) 규모의 ‘프랑스 & AI’ 기금을 조성
- ③ (컴퓨팅 자원 확보) 프랑스를 컴퓨팅 자원의 거점으로 만들기 위해 국가 및 유럽 차원의 공동 자원을 확보하고 국가 컴퓨팅 센터를 설립해 접근성을 높이며 모델 훈련에 대한 세액공제 제도를 마련
- ④ (데이터 접근성 개선) 개인정보 측면에서 개인정보보호당국(CNIL)의 권한을 현대화하고, 의료데이터 접근을 위한 특정 사전승인 절차를 폐지
- ⑤ (문화 콘텐츠 활용) 지식재산권을 존중하면서 프랑스와 유럽의 문화 콘텐츠에 대한 AI 모델 훈련을 장려하기 위한 기술 인프라를 마련
- ⑥ (공공연구 강화) AI 관련 공공연구를 강화하기 위해 연구자들의 행정 절차 부담을 줄이고 급여를 인상하며 AI에 특화된 공공연구의 예산을 2배로 증액
- ⑦ (국제 거버넌스) AI 시스템을 평가하고 감독할 세계 AI 기구 창설에 협력하고 공익 증진 목적의 국제 AI 기금을 조성

〈표 3-47〉 프랑스 정부의 최우선 AI 추진 정책

구분	주요 내용
국민 인식 제고와 교육	• 국가 인식 및 교육 계획을 시작하여 AI와 그 과제에 대한 공동 소유권을 위한 조건을 조성
재정 투자	• 프랑스 AI 생태계를 지원 하고 이를 세계 최초의 생태계로 만들기 위해 디지털 기업 과 비즈니스 혁신 에 막대한 투자
컴퓨팅 자원 확보	• 프랑스와 유럽을 단기 및 중기적으로 컴퓨팅 파워의 주요 허브로 작용
데이터 접근성 개선	• 개인 데이터에 대한 접근 방식을 전환하여 지속적으로 보호하는 동시에 우리의 요구 사항을 충족하는 혁신 촉진
문화 콘텐츠 활용	• 지적 재산권을 존중하면서 문화 콘텐츠에 대한 접근을 허용함으로써 프랑스 문화의 영향력을 보장
공공연구 강화	• AI의 매력을 강화하기 위해 공공 연구에서 선제적으로 실험
국제 거버넌스	• 글로벌 AI 거버넌스 구축을 목표로 일관되고 구체적인 외교 이니셔티브를 구축

출처: info.gouv.fr, 25 recommandations pour l'IA en France, 2024.03.13.

2) 프랑스 경쟁관리국의 생성 AI 분야의 경쟁 활성화 권고⁹⁷⁾

(1) 개요

프랑스 정부는 2018년 국가 AI 전략을 출범시켰고, 이를 위해 '프랑스 2030' 계획 중 약 25억 유로가 배정됐다. 2024년 3월, 총리가 출범시킨 프랑스 AI 위원회(Commission de l'IA)는 특히 프랑스를 컴퓨팅 파워의 주요 중심지로 만들고, 데이터 액세스를 촉진하며, 글로벌 AI 거버넌스를 구축할 것을 제안하는 25가지 권고안을 담은 보고서를 발표했다.

유럽 차원에서는 AI 법(곧 EU 공식 저널에 게재될 예정)의 대부분 조항이 2026년부터 적용된다. 생성 AI⁹⁸⁾가 부상하기 전에 발표되었지만, 디지털시장법(Digital Markets Act, DMA)⁹⁹⁾과 데이터법(Data Act)¹⁰⁰⁾은 이 분야에 영향을 미칠 것이다.

따라서 2023년 11월 영국에서 열린 AI 안전 정상회의(AI Safety Summit)에서 블레츨리 선언(Bletchley Declaration)과 같이 AI에 대한 일련의 이니셔티브가 전 세계적으로 채택됐다.

- 블레츨리 선언(Bletchley declaration)이란 2023년 11월 영국 블레츨리 파크에서 열린 제 1회 인공지능(AI) 안전 정상회의(AI Safety summit)에서 발표한 선언문을 뜻한다. 한국과 미국, 중국, 영국 등 28개국과 유럽연합(EU)이 참여했으며 2024년 AI 붐이 일면서 AI가 초래할지도 모르는 심각한 피해를 막기 위한 협력을 다짐
- 블레츨리 선언에 따르면, AI 분야에서는 의도적 오용 또는 의도하지 않은 통제 문제가 발생할 수 있으며, 특히 사이버보안, 생명 공학, 잘못된 정보 확산에 대한 위험성이 크게 나타날 수 있다는 부분에 동의했다. 각국은 편견과 프라이버시 등 AI에서 파생되는 문제를 언급하며, 해당 위험은 국제 협력을 통해 가장 잘 해결될 수 있을 것이라고 선언문에 명시

97) Autorite de la Concurrence, Generative artificial intelligence: the Autorite issues its opinion on the competitive functioning of the sector, 2024.06.28.

(<https://www.autoritedelaconcurrence.fr/en/press-release/generative-artificial-intelligence-autorite-issues-its-opinion-competitive>)

98) 유럽 의회에 따르면 인공 지능은 “추론, 학습, 계획 및 창의성과 같은 인간과 유사한 능력을 보여주기 위해” 기계가 사용하는 모든 도구를 의미한다. 생성 AI는 텍스트, 이미지, 사운드 또는 비디오와 같은 새로운 콘텐츠를 생성할 수 있는 AI 모델을 의미

99) DMA는 소비자 및 판매자를 중개하는 ‘관문(gatekeeper)’ 역할을 하는 대형 플랫폼 기업의 독점적 지위를 활용한 영업활동을 방지하기 위해 EU가 제정한 법안

100) 데이터법(Data Act)은 유럽연합(EU)이 제정한 법안으로, 데이터의 공정한 접근과 사용을 위한 규칙 정함

- 챗GPT 이후 AI 안전성에 대한 우려가 커진 가운데 리시 수낙 영국 총리의 제안으로 시작됐다. 블레츨리 파크 저택은 컴퓨터 과학의 발상지로 일컬어지는 곳으로 제2차 세계대전 당시 암호 해독자들이 근무하던 장소¹⁰¹⁾

이에 따라 프랑스 경쟁당국(Autorité de la Concurrence, 약칭 Autorité)¹⁰²⁾은 2024년 6월 28일 생성 AI 시장의 경쟁 활성화를 위해 규제 효율성 강화와 컴퓨팅 파워의 접근성 확대 필요하다는 의견서를 발표했다.

(2) 생성 AI 경쟁우위 확보 방안

이번 의견서에는 생성 AI 분야는 진입장벽이 높고, 생성 AI와 관련된 시장에서 특정 사업자의 입지는 다양한 경쟁우위를 수반해야 한다는 내용을 담고 있다.

- ① (특화된 AI 칩의 필요성) 엔비디아나 기타 주요 기업이 개발한 AI 칩은 생성 AI 모델 훈련과 추론에 필수적이며 가격이 매우 비싸고 공급도 제한적
- ② (클라우드 서비스의 중요성) 기반 모델을 개발하려면 충분한 컴퓨팅 파워를 확보해야 하며, 클라우드는 모델 훈련에 필요한 컴퓨팅 파워에 접근하고 모델을 배포하기 위한 핵심 수단
- ③ (막대한 자금 조달) 생성 AI의 특성상 모델 개발을 위해서는 대규모 데이터와 고급 데이터 과학 기술이 요구되며, 상당한 규모의 자금 조달도 필요
- ④ (대량의 데이터) 주요 디지털 기업은 기반모델 훈련과 개발에 필요한 컴퓨팅 파워와 대규모 데이터, 고도로 숙련된 직원 등 다양한 이점을 누리며, 생성 AI 도구를 자사 제품과 서비스 생태계에 통합하여 규모의 경제를 창출

101) CIO Korea, 미국, 한국 등 28개국, '안전하고 책임있는 AI 개발' 강조한 블레츨리 공동 선언 발표, 2023.11.02.
(<https://www.cio.com/article/3510170/%EB%AF%B8%EA%B5%AD-%ED%95%9C%EA%B5%AD-%EB%93%B1-28%EA%B0%9C%EA%B5%AD-%EC%95%88%EC%A0%84%ED%95%98%EA%B3%A0-%EC%B1%85%EC%9E%84%EC%9E%88%EB%8A%94-ai-%EA%B0%9C%EB%B0%9C-%EA%B0%95.html>)
연합인포맥스, [시사금융용어] 블레츨리 선언(Bletchley declaration), 2023.12.14.
(<https://news.einfomax.co.kr/news/articleView.html?idxno=4291337>)

102) 프랑스 경쟁당국은 2024년 2월 생성 AI 분야에 대한 조사에 착수해 공개 협의를 거쳐 생성 AI 분야의 경쟁 분석과 권고사항을 제시한 이번 의견서를 공개했음

(3) 프랑스 경쟁당국(Autorité)의 권장 사항

Autorité는 생성 AI 분야의 경쟁 활성화를 위해 규제 효율성 강화와 컴퓨팅 파워의 접근성 확대, 저작물의 지식재산권 인정, 디지털 대기업의 투자 투명성 확보 등을 제안했다.¹⁰³⁾

- ① 클라우드를 통해 AI 모델을 제공하는 서비스형 모델(model as a service, model as a service, MaaS) 발전에 주목하여 해당 서비스 기업을 디지털시장법(DMA) 규정에 따라 시장 지배적 사업자(게이트키퍼)¹⁰⁴⁾로 지정하는 방안을 고려
- ② 범유럽 차원에서 슈퍼컴퓨터 투자를 지속하여 다양한 주체가 컴퓨팅 파워에 접근할 수 있도록 지원하고, 데이터의 가치를 고려해 저작물에 차등적 가격을 책정해 모델 개발자의 혁신 역량을 보호하며, 생성 AI 모델 개발을 위해 공공과 민간 데이터 개방
- ③ 빅테크의 생성 AI 투자를 감독할 수 있도록 경쟁 당국이 빅테크에 생성 AI 분야의 소수지분 투자에 대한 추가 정보를 요청하여 투명성을 확대

<표 3-48> 프랑스 경쟁당국(Autorité)의 AI 활성화 권장 사항

구분	주요 내용
게이트 키퍼 지정	<ul style="list-style-type: none"> MaaS 서비스 개발에 특별한 관심을 기울여 이러한 서비스를 제공하는 기업을 DMA에 따른 게이트키퍼로 지정해서 평가
DGCCRF의 크레딧 사용 주의 요망	<ul style="list-style-type: none"> 클라우드 크레딧에 대한 DGCCRF(Directorate General for Competition Policy, Consumer Affairs and Fraud Control)¹⁰⁵⁾가 AI에서 이러한 크레딧 사용에 특별한 주의를 기울여야 함
소규모 사업자의 확대 장려	<ul style="list-style-type: none"> 시법 제64조에 따라 설립된 미래의 인공지능 사무소와 동 법 제70조에 따라 지정될 프랑스의 관할 국가기관은 소규모 사업자의 출현 또는 확장을 방해하지 않도록 보장
경쟁 활성화 장려	<ul style="list-style-type: none"> 시장에서 경쟁을 집행할 책임이 있는 당국은 생성 AI 부문에서 경계를 늦추지 말아야 하며, 필요한 경우 신속하고 효과적으로 행동하기 위해 사용 가능한 모든 행위를 해야 함

103) 프랑스 경쟁당국 권장사항 10가지는 표 참조

104) DMA상 게이트키퍼는 시가총액 750억 유로(약 107조 원) 이상이거나, 최근 3년간 EU 내 연매출이 75억 유로(약 10조 원)를 웃돌며 월간활성화이용자수(MAU)가 4,500만 명을 넘어선 플랫폼을 뜻함

구분	주요 내용
접근성 용이	<ul style="list-style-type: none"> 가능한 한 많은 당사자가 컴퓨팅 파워에 접근할 수 있도록 유럽 차원의 슈퍼컴퓨터 개발에 계속 투자
데이터 개방	<ul style="list-style-type: none"> 정부 및 슈퍼컴퓨터 관리를 담당하는 기업은 연구, 특히 학술 연구의 우선순위를 유지 하면서 공공 슈퍼컴퓨터 리소스를 유료로 사용할 수 있도록 하는 개방적이고 비차별적인 프레임워크를 제안하는 방법 검토
공공부문 생성 AI 개방성 기준 설정	<ul style="list-style-type: none"> 특히 AI법과 연계하여 공공 슈퍼컴퓨터에서 훈련된 생성 AI 모델의 개방성에 대한 기준 설정
데이터 경제적 가치 고려한 비용 절감 제안	<ul style="list-style-type: none"> 공공기관, 특히 프랑스 문화부가 프랑스 문학 및 예술 재산권 고등위원회에 위임한 임무의 일환으로, 권리 보유자가 사용 사례에 따라 데이터의 경제적 가치를 고려하고(예: 차별화된 가격 도입) 거래 비용을 줄이기 위한 번들 제안을 해야 함
공공 데이터 활용에 따른 민간 이니셔티브 장려	<ul style="list-style-type: none"> 생성 AI 모델의 훈련 또는 미세 조정에 공공 및 비공개 데이터를 사용할 수 있도록 하고, 텍스트, 이미지, 비디오 등 프랑스어 데이터를 배포하기 위한 공공 및 민간 이니셔티브 장려
소수 지분 투자	<ul style="list-style-type: none"> 빅테크의 생성 AI 투자를 감독할 수 있도록 경쟁 당국이 빅테크에 생성 AI 분야의 소수 지분 투자에 대한 추가 정보를 요청하여 투명성 확대

출처: Autorite de la Concurrence, Generative artificial intelligence: the Autorite issues its opinion on the competitive functioning of the sector, 2024.06.28.¹⁰⁶⁾

105) 프랑스의 공정거래위원회, 소비자 보호원 같은 정부 기관을 지칭함

106) Autorite de la Concurrence, Generative artificial intelligence: the Autorite issues its opinion on the competitive functioning of the sector, 2024.06.28.

(<https://www.autoritedelaconcurrence.fr/en/press-release/generative-artificial-intelligence-autorite-issues-its-opinion-competitive>)

제8절 독일

1. 그간의 인공지능 정책 추진 경과

1) 개요

독일은 다른 경제 강국에 비해 비교적 늦게 AI의 개발을 정책적으로 장려하고 있다. 2018년 ‘AI made in Germany’를 내세우며 AI 전략(Strategie Künstliche Intelligenz der Bundesregierung)을 발표해 ‘정부 차원의 인공지능 분야 연구개발 정책을 디자인하고 관련 분야 혁신을 촉진하기 위한 전략이다.

이 전략에서는 2025년까지 총 20억 유로를 투자하여 AI 관련 정책전략 실현에 집중할 계획이며, 이를 통해 민간분야, 국공립 연구소, 유관기관 및 지자체들은 AI 관련 사업 예산 규모를 2배까지 확대가 가능해질 것으로 기대하고 있다. 본 전략에서의 AI 지원 정책은 ▲연구 강화 ▲혁신 챌린지 ▲기업역량 강화 ▲스타트업 지원 ▲노동시장 변화 ▲인재양성 및 유치 ▲공공수요 및 정부서비스 혁신 데이터 활용 ▲촉진(Promotion) ▲법제도 ▲표준 ▲네트워크 ▲사회적 논의 등 12개 분야로 구성되어 있다.¹⁰⁷⁾

앙겔라 메르켈 총리는 독일과 유럽을 인공지능 분야의 세계적 리더로 만들고자 2018년 11월 13일 t-online.de와의 인터뷰에서 “미래에는 독일과 유럽이 인공지능 분야의 리더가 되어야 합니다.”라고 강조했다. 또한 연방 경제 및 에너지 장관인 피터 알트마이어는 포츠담에서 “우리는 독일의 중소기업 및 수출 지향 기업이 디지털화에 진전을 이루기를 바랍니다. 왜냐하면 그것은 우리가 생산성을 높이고 1.3% 정도의 경제 성장을 달성하는 데 도움이 될 것이기 때문입니다. 그것은 증기 엔진이 발명된 이후로 다른 어떤 근본적인 혁신보다 더 큰 것입니다.”라고 AI 중요성을 강조했다.

107) The federal government, AI - a brand for Germany, 2018.11.15
(<https://www.bundesregierung.de/breg-en/service/archive/ai-a-brand-for-germany-1551432>)

〈표 3-49〉 독일의 AI Made in Germany 12대 전략 요약

구분	주요 내용
연구강화	<ul style="list-style-type: none"> • 혁신의 원동력이 되기 위한 독일과 유럽의 AI연구 강화 • 국제적 경쟁력을 확보한 AI 생태계 구축
산업계 이전 가속화	<ul style="list-style-type: none"> • AI 챌린지를 개최하여 혁신 아이디어를 창출하고 새로운 솔루션 개발, 그리고 애플리케이션 및 비즈니스 모델 개발 등 다양한 기회 창출
기업 역량 강화	<ul style="list-style-type: none"> • 독일의 스타트업, 중소기업에서 대기업까지 AI 서비스 적용에 중점
스타트업 지원	<ul style="list-style-type: none"> • 벤처 캐피탈에 대한 접근성 향상 등 AI 기반 비즈니스 모델 및 제품의 창업 동력 부흥을 위한 정책적 지원
노동시장 변화	<ul style="list-style-type: none"> • 노동시장 변화에 따른 업무 재교육 등 다양한 조치 지원 • 높은 수준의 관측소(Observatory)를 통해 국내 및 국제적으로 AI의 이용을 촉진, 분석, 평가를 중점적으로 지원할 예정
인재양성 및 유치	<ul style="list-style-type: none"> • 교육훈련 강화 및 고숙련 AI 전문가 영입 • 대학에 전문 교수직 신설 등
공공수요 및 정부서비스 혁신	<ul style="list-style-type: none"> • 정부 행정에 AI를 선제적으로 적용하는 등 공공수요 창출 • AI를 통한 정부 행정 서비스의 빠르고 효율적인 시스템 마련
데이터 활용, 데이터 촉진	<ul style="list-style-type: none"> • AI분야 세계 최고 국가 건설을 위해 개인정보, 정보의 자기결정권의 위배 없이 고품질의 활용 가능한 데이터 양 증가
AI 오용과 법제도	<ul style="list-style-type: none"> • AI 기반 결정, 서비스, 제품의 왜곡, 차별, 오용 등에 효율적인 보호가 가능하도록 법적 프레임 워크 검토
표준	<ul style="list-style-type: none"> • 국가 표준 기구 DIN/DKE를 통해 산업계와 협업하여 AI 개발·활용 관련 국내외 표준 수립을 지원
국내·외 네트워크	<ul style="list-style-type: none"> • AI는 범용 기술로 국내, 국제적으로 과학, 경제, 행정 및 시민일상에 영향을 미치므로 정부는 국제협력, 양자 및 다자간 대화를 지속하여 네트워크를 강화할 예정
사회적 논의	<ul style="list-style-type: none"> • AI 연구, 개발 및 응용을 세계 수준으로 향상시키기 위해서, 연방정부는 AI의 기회와 위험에 대한 사회적 대화 및 교육을 지속 추진

출처: The federal government, AI - a brand for Germany, 2018.11.15.

2) 독일의 인공지능 정책 변화

또한 독일 연방정부는 ‘독일 인공지능 전략 수립(‘18.11)’ 1주년을 기념하여, 각 부처의 인공지능 전략과 관련한 추진성과를 담은 중간보고서(Zwischenbericht Ein Jahr KI-Strategie)를 발표하였다. 독일을 세계적인 인공지능 선도국가로 도약시키기 위해, 지난 1년간 연방교육연구부(BMBF), 연방경제에너지부(BMWi), 연방교통디지털인프라부(BMVI), 연방노동사회부(BMAS) 등의 각 부처가 추진하였던 인공지능 정책, 프로그램 및 주요활동을 점검했다.¹⁰⁸⁾

한편 독일 연방정부는 2020년 6월 발표한 ‘코로나19 경기부양책’에 따라 2025년까지 인공지능 분야 투자를 기존 30억 유로에서 50억 유로로 증액한 ‘인공지능 전략 2020 개정안(Fortschreibung KI-Strategie)’을 발표하면서 AI 정책에 대한 계속된 지원을 하고 있다.

〈표 3-50〉 독일의 AI 정책 변화

시기	구분	주요 내용
2023년 11월	AI 실행계획	<ul style="list-style-type: none"> AI 가치 사슬 강화 및 교육, 과학, 연구 간 협력 강화에 중점을 두어 12가지 주요행동 영역에 대한 실행계획 연방교육연구부(BMBF)는 현 정부 임기 동안 AI 기술에 16억 유로 투자 예정 AI 프로젝트 20개 추가 도입 및 R&D와 인프라에 부문에 50여 개의 이니셔티브 지원
2020년 12월	국가 AI 전략 업데이트2	<ul style="list-style-type: none"> 전문가, 연구, 지식이전 및 응용, 규제 프레임워크, 사회 등 5가지 분야의 우선순위 제시 및 2022년까지 시행될 구체적인 조치 제시 전략 수정에 앞서 빅데이터 연구 활성화를 위한 베를린 학습-데이터 재단(BIFORD) 설립
2019년 11월	국가 AI 전략 업데이트1	<ul style="list-style-type: none"> 지난 1년간 연방교육연구부(BMBF), 연방경제에너지부(BMWi), 연방교통디지털인프라부(BMVI), 연방노동사회부(BMAS) 등의 각 부처가 추진하였던 인공지능 정책, 프로그램 및 주요활동들을 점검
2018년 11월	국가 AI 전략	<ul style="list-style-type: none"> '세계 최고의 AI 국가 건설'을 목표로 구체적인 방법을 제시하는 독일 최초의 AI 전략으로 2025년까지 총 30억 유로 투자 계획 AI 정책 지원을 위해 표준 설정, 연구 강화, 법제도 등 12개 전략 분야 제시

출처: BMWK, BMBF 등 독일 정부기관의 내용, SPRi 재정리. 2024.10.

108) Bmbf, Zwischenbericht: Ein Jahr KI-Strategie, 2019.11.15
https://www.bmbf.de/SharedDocs/Downloads/files/zwischenbericht-ki-strategie_final.html

또한 기존 전략(2018년)의 12개 분야를 인력, 연구, 활용 등 5개 분야로 통합하는 등 대대적인 변화를 시도했다.¹⁰⁹⁾

〈표 3-51〉 독일의 5대 분야별 추진계획

분야	주요 내용
1. 인력 양성	<ul style="list-style-type: none"> 매력적인 근로환경 구축, 인공지능 전문역량 강화
2. 연구	<ul style="list-style-type: none"> 국가연구개발 체계 강화, 유럽 및 글로벌 연구협력 학제간 연구 및 지속 가능한 활용 보건·환경·우주항공·모빌리티·농업 등 분야별 연구 강화
3. 이전 및 응용	<ul style="list-style-type: none"> 산업계 적용, 네트워킹 및 국제협력 혁신을 위한 규격, 표준화 및 테스트 베드 보건·근로·공공행정 분야 적용
4. 법·규정	<ul style="list-style-type: none"> 안전하고 신뢰할 수 있는 인공지능 활용 운용 및 제품 안전을 위한 규제 인공지능의 보안, 견고성 및 지속 가능성 제고
5. 사회	<ul style="list-style-type: none"> 모두를 위한 인공지능

출처: BMWK, Strategie Künstliche Intelligenz der Bundesregierung Fortschreibung 2020. 2020.12.02

2. 2024년 독일 인공지능 정책

1) OECD의 독일 AI 검토¹¹⁰⁾

(1) 개요

독일은 2018년에 AI 정책을 내놓고 매해 정책을 업데이트하고 있으며 AI 전략과 AI 개발, 도입을 체계적으로 추진하는 유럽 국가 중 하나이다. 이 전략의 주요 목표는 ▲독일의 미래 경쟁력을 보장하고, ▲책임감 있고 신뢰할 수 있는 AI 개발을 촉진하고, ▲광범위한 사회적 대화와 적극적인 정치적 노력을 통해 AI를 사회에 도입하는 것이다. 이러한 목표에 대한 독일의 헌신은 글로벌 AI 트렌드와 표준의 리더로 자리매김한 강력한 공공 및 민간 연구 기관 네트워크가 반영되어 있는 결과이다.

109) BMWK, Strategie Künstliche Intelligenz der Bundesregierung Fortschreibung 2020. 2020.12.02
(https://www.ki-strategie-deutschland.de/files/downloads/201201_Fortschreibung_KI-Strategie.pdf)

110) OECD, 'OECD Artificial Intelligence Review of Germany, 2024.06.11.
(https://www.oecd.org/en/publications/oecd-artificial-intelligence-review-of-germany_609808d6-en.html)

OECD 보고서는 독일의 국가적 연구 우수성과 견고한 계산 능력이 AI 경쟁에서 중요한 자산임을 강조한다. 독일 연구자들은 AI 연구 출판물에서 꾸준히 세계 최고 순위를 차지하며, AI 지식과 응용 프로그램을 발전시키는 데 크게 기여하고 있다. 게다가 정교한 AI 모델을 훈련하는 데 필수적인 독일의 학계에서의 뛰어난 계산 능력은 경쟁 우위를 제공하고 있으며 특히 AI 연구가 점점 더 복잡해지고 계산 집약적으로 변함에 따라 독일에게는 기회로 작용할 것이다.

(2) 인간 중심의 AI 지침 원칙

독일의 AI 전략은 ▲인간 중심 개발을 우선시하며, ▲인간의 역량을 강화하고, ▲프라이버시를 존중하며, ▲공정성을 보장하는 AI 애플리케이션에 집중한다. 이 접근 방식은 윤리적으로 타당하며 AI 발전을 사회적 목표와 일치시켜 기술이 대중의 이익에 광범위하게 기여하도록 한다.

독일은 산업의 지속 가능한 환경을 만들기 위해 윤리적 AI를 만드는데 노력하고 있다. 독일 산업 환경의 변화 핵심은 녹색 및 디지털 경제로의 전환이다. 독일은 에너지, 운송 및 산업과 같은 중요한 부문에서 탈탄소화와 지속 가능한 산업으로 추진하기 위해 AI를 활용함으로써 다른 국가에 비해 경쟁 우위를 펼치는 것을 목표로 한다.

(3) 독일의 AI 미래를 강화하기 위한 주요 권장 사항

OECD 보고서는 독일이 AI 선두국가를 달성하고 경쟁 우위를 유지하기 위한 구체적이고 실행 가능한 권장 사항을 아래와 같이 제시하고 있다.

- ① 연방 정부의 절대적 지지 확보: 사회 연결 인프라 및 오픈 데이터 가용성과 같은 분야에서 연방 정부의 지지가 필요하다. 특히 재무부는 이러한 정책을 조율하여 기회를 창출하고, 정책 효과를 보장하며, 갈등이나 중복 노력을 피하는 데 중심적인 역할을 해야 한다.
- ② 사회 전반 산업 변혁을 위한 AI 활용: AI는 제조업을 포함한 전체 산업을 변혁할 잠재력이 있으며, 이는 독일 경제에 매우 중요하다. 독일은 산업 경쟁력을 강화하고, 녹색 전환을 가속화하고, 공공 행정 효율성을 개선하는 AI 솔루션을 홍보함으로써 중요한 국가적 과제를 해결할 수 있다.
- ③ 일반 국민 및 근로자 참여 강화: AI 기술의 지속적인 수용은 시민 및 근로자와의 의미 있는 대화에 달려 있다. 근로자와 고용주 간의 AI 전문성과 인식을 강화하면 노동 시

장 회복력이 배양되고 AI 도입에 대한 인간 중심적 접근 방식이 이뤄진다.

- ④ 데이터 접근성 확대: 데이터는 AI 개발의 원자재이다. 데이터 공유 프로토콜을 사전에 개발하고, 민감하지 않은 정부 데이터의 공개를 의무화하고, 산업별 데이터의 책임 있는 공유를 용이하게 하는 것이 중요합니다. 이러한 조치를 구현하면 견고한 AI 혁신의 사회 기반이 마련된다.
- ⑤ AI 스타트업 지원: AI 스타트업은 사업 확장을 위한 투자금을 유치하는데 어려움을 겪는다. 자본 수집 기관에 대한 법적 틀을 재검토하고 정부 자금을 집중적으로 제공하면 AI 스타트업의 성장이 촉진될 것이다. 또한, AI 스타트업 서비스의 공공 부문 조달을 용이하게 하기 위해 조달 지침을 개정하면 혁신이 촉진될 것이다.
- ⑥ 경쟁력 있는 컴퓨팅 인프라 구축: 독일의 현재 AI 컴퓨팅 인프라를 평가하고 개선하는 것이 필수적이다. 이 인프라의 일부를 AI 스타트업과 중소기업이 이용할 수 있도록 하면 포용성이 촉진되고 더 광범위한 AI 생태계가 형성된다.

〈표 3-52〉 OECD의 독일 AI 권고사항

구분	주요내용
연방 정부의 절대적 지지 확보	<ul style="list-style-type: none"> • 사회 연결 인프라 및 오픈 데이터 가용성과 같은 분야에서 연방 정부의 지지가 필요
사회 전반 산업 변혁을 위한 AI 활용	<ul style="list-style-type: none"> • 독일은 산업 경쟁력을 강화하고, 녹색 전환을 가속화하고, 공공 행정 효율성을 개선하는 AI 솔루션을 홍보함으로써 중요한 국가적 과제를 해결
일반 국민 및 근로자 참여 강화	<ul style="list-style-type: none"> • 근로자와 고용주 간의 AI 전문성과 인식을 강화하면 노동시장 회복력이 배양되고 AI 도입에 대한 인간 중심적 접근 방식이 이뤄짐
데이터 접근성 확대	<ul style="list-style-type: none"> • 데이터 공유 프로토콜을 사전에 개발하고, 민감하지 않은 정부 데이터의 공개를 의무화하고, 산업별 데이터의 책임 있는 공유를 용이하게 하는 것이 중요
AI 스타트업 지원	<ul style="list-style-type: none"> • 자본 수집 기관에 대한 법적 틀을 재검토하고 정부 자금을 집중적으로 제공하면 AI 스타트업의 성장이 촉진됨
경쟁력 있는 컴퓨팅 인프라 구축	<ul style="list-style-type: none"> • 독일의 현재 AI 컴퓨팅 인프라를 평가하고 개선하는 것이 필수적임

출처: OECD, 'OECD Artificial Intelligence Review of Germany, 2024.06.11.

제9절 싱가포르

1. 그간의 인공지능 정책 추진 경과

1) 개요

싱가포르는 국가 전체를 스마트화 하기 위한 “Smart Nation(스마트 국가)” 이니셔티브를 발표한 2014년 11월 이후로 주기적으로 세부 계획들을 추가로 발표하고 있다. “Smart Nation” 이니셔티브란, 싱가포르 정부가 ICT 기술과 혁신을 활용해서 스마트 국가로 변모하기 위한 장기적인 전략이다. 이는 세계적 수준의 기술 중심 도시 국가가 되는 것을 목표로 하며, 이는 지속 가능한 개발을 추진하고 살기 좋은 도시를 만들기 위해 기술과 혁신을 활용하겠다는 싱가포르 정부의 장기적인 비전과 약속을 포함한다.

또, 싱가포르는 2019년 경제를 혁신하기 위해 AI 사용을 심화하기 위한 계획을 담은 국가 AI 전략을 발표했다. AI에 대한 초기 투자로 약 150개의 연구개발팀이 설립되었고, 900개의 스타트업이 AI로 새로운 아이디어를 탐구하고 있다. 그 이후로 AI의 중요한 돌파구를 목격했으며, 그 결과 신제품 개발, 향상된 기능 및 혁신적인 상호 작용이 이뤄지고 있다.¹¹¹⁾

최근에는 2023년 싱가포르 국가 AI 전략 2.0(NAIS 2.0)을 발표하면서 이러한 추세를 선도하기 위한 지속적인 노력을 기울이고 있다. 또 비전과 목표를 달성하기 위해 NAIS 2.0에 따라 10가지 인에이블러(Enabler)를 통해 3개의 시스템(환경, 사람, 활동가)을 집중해서 추진하고 있다.¹¹²⁾

〈표 3-53〉 싱가포르 국가 AI 전략 2.0(NAIS 2.0)

구분	주요내용
기회에서 필요성으로	오늘날 AI가 제공하는 기회와 위험의 풍경이 확장됨에 따라 AI가 국가의 지속적인 번영과 관련성에 필수 불가결하다는 사실이 분명해졌다. AI를 단순히 액세서리나 사치품으로 생각하지 말고 마스터해야 한다.

111) Smartnation, National AI Strategy, 2019.11.30.
(<https://www.smartnation.gov.sg/media-hub/publications/national-ai-strategy/>)

112) Smartnation, Singapore National AI Strategy 2.0 (NAIS 2.0), 2023.12.05.
(<https://file.go.gov.sg/nais2023.pdf>)

구분	주요내용
로컬에서 글로벌로	싱가포르는 처음부터 AI에 대한 글로벌 전망을 가지고 운영해야 한다. 오늘날 AI를 괴롭히는 복잡한 문제를 극복하기 위해 글로벌 네트워크에 연결하여 최고의 사람들과 협력하고 같은 생각을 가진 사람들 사이에서 리소스를 공유해야 한다.
프로젝트에서 시스템으로	싱가포르는 역량, 인프라 및 자원 기반을 더욱 강화하고 대규모로 아이디어 교환을 가속화하여 다양한 부문과 경제 및 사회 전반에 걸쳐 AI의 영향을 확대하고 관리해야 한다.

출처: Smartnation, National AI Strategy, 2024.11.17. 검색¹¹³⁾

2) 싱가포르의 인공지능 정책 변화

글로벌 AI 기업 유치, AI 인재 육성, AI 도입 활성화를 위한 투자 및 신뢰 구축 등 AI 생태계 구축을 목표로 하는 ‘국가 AI 전략(National AI Strategy, NAIS)’을 2019년 처음 발표한 후, 정보통신미디어선부(Ministry of Communications and Information, MCI) 산하기관인 정보통신미디어개발청(Infocomm Media Development Authority, IMDA)은 책임감 있는 AI 이용을 위한 AI 테스트 도구를 개발하고자 AI Verify Foundation을 설립한다고 발표했다. 이 재단은 전 세계 기업과 규제 기관의 수요를 해소하기 위한 AI 테스트 역량을 강화할 예정이다. 싱가포르 정보통신미디어개발청(IMDA), Aicadium(테마섹의 AI 우수성 센터), IBM, 마이크로소프트, 구글, 레드햇 및 세일즈포스 등 7개의 선구적인 프리미엄 회원사가 이 AI Verify 로드맵의 전략적 방향과 발전을 설명할 것이다. 그 출발점으로, AI Verify Foundation은 어도비, DBS, 메타, 센스타임 및 싱가포르 항공을 포함해 60개 이상의 일반 회원을 모집했다.¹¹⁴⁾

AI Verify Foundation의 설립은 AI의 위험을 해소하기 위해 AI Verify의 개발 및 이용을 지원했다. AI Verify는 IMDA가 여러 부문에서 대기업과 협력해 최초로 개발한 AI 거버넌스 테스트 프레임워크 및 소프트웨어 툴킷이다. AI Verify Foundation은 AI 테스트 프레임워크, 코드 기반, 표준 및 오픈 소스 커뮤니티를 촉진하고, AI를 테스트 및 관리하는 개방형 협력과 아이디어 공유를 위한 중립 플랫폼 구축에 일조했다.

113) <https://www.smartnation.gov.sg/nais/>

114) IMDA, Singapore launches AI Verify Foundation to shape the future of international AI standards through collaboration, 2023.06.07

(<https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2023/singapore-launches-ai-verify-foundation>)

2023년 국제 시범용으로 최소 기능 제품으로 출시된 AI Verify는 IBM, 델, 히타치, UBS 등 50개 이상의 현지 및 다국적 기업의 관심을 끌었다. AI Verify는 현재 오픈 소스 커뮤니티에 제공되고 있으며, 국제적으로 인정받는 AI 거버넌스 원칙, 예를 들어, EU, OECD 및 싱가포르의 AI 거버넌스 원칙에 따른 테스트 프레임워크와 툴킷을 제공함으로써 국제 커뮤니티를 발전시켰다고 볼 수 있다. AI Verify 툴킷은 AI 시스템을 위한 여러 거버넌스 원칙을 포괄하는 테스트 보고서를 생성할 수 있는 통합 인터페이스로 제공한다. 또한, 기업이 이해관계자들과 이들 보고서를 공유함으로써 회사의 AI 투명성을 높일 수 있도록 지원했다.

비슷한 목적으로 2019년 1월 싱가포르 ‘개인정보보호당국(Personal Data Protection Commission, PDPC)’은 「모델 AI 거버넌스 프레임워크(Model AI Governance Framework)」라는 지침서를 만들어 이를 통해 AI 시스템을 도입할 때 발생하는 주요 윤리 및 거버넌스 문제 해결을 위해 필요한 지침을 민간에게 제공하고자 했다.

싱가포르 정부는 2020년 1월 ‘모델 AI 거버넌스 프레임워크’의 두 번째 판을 선보였는데, 민간 기업이 이를 자발적으로 채택하는 본래의 취지는 변함이 없으나 2019년 출간된 지침서와의 차이점이라면 AI 시스템이 내린 결정은 설명할 수 있고(explainability) 투명하며(transparency) 공정(fair)해야 한다는 원칙과 AI 시스템이 인간 중심이 되어야 한다는 두 원칙을 책임성 있는 AI 사용을 위한 지침 원칙(guiding principles)으로 채택한 부분이다.

제2판은 이 두 가지 원칙을 실행으로 옮기기 위해 여러 영역에서 필요한 구체적인 지침도 담고 있다. 또한 인간 중심의 AI 활용의 원칙을 평가하기 위해서는 AI가 활용된 의사 결정에 인간이 적절히 개입하고 감독했는지가 중요하다는 사실을 인식하여 인간이 검토한 부분이 의사결정 중 어느 정도 차지하는지 파악하는 과정이 지침서에 포함됐다. 일반적으로 AI 활용으로 인한 위험도가 높다고 생각할 때 인간이 의사 결정 과정에 더 많이 관여하여 잠재 위험을 줄이는 것이 필요하다고 예상되기에, 해당 AI 활용이 초래할 수 있는 위험의 강도와 일어날 수 있는 확률이 높을수록 적절하다고 예상되는 인간의 참여도(level of human involvement)도 높다.¹¹⁵⁾

115) PDPC, Singapore's Approach to AI Governance, 2024.11.17. 검색
(<https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework>)

〈표 3-54〉 연도별 싱가포르의 인공지능 정책 변화

연도	명칭	주요내용
2019	국가 AI 전략 발표	AI 기업 유치, AI 인재 육성, AI 도입 활성화를 위한 투자 및 신뢰 구축 등 AI 생태계 구축을 목표
2019	AI Verify 재단 설립 및 AI Verify 개발	AI 거버넌스 평가 및 검증 테스트 도구 정부 산하 기관인 정보통신미디어개발청이 개발
2019	모델 AI 거버넌스 프레임워크 (Model AI Governance Framework) 발표	AI 관련 주요 윤리 및 거버넌스 지침
2020	모델 AI 거버넌스 프레임워크 (Model AI Governance Framework)의 제2판 발표	설명할 수 있고 투명하며 공정해야 한다는 원칙과 인간 중심의 AI 시스템이 되어야 한다는 두 원칙을 책임성 있는 AI 사용을 위한 지침 원칙으로 채택
2023	국가 AI 전략 2.0 발표	‘공공의 선, 국가와 세계를 위한 AI’라는 원칙에 기반한 Excellence와 Empowerment를 위한 15가지의 행동 전략을 포함

출처: PDPC, Singapore’s Approach to AI Governance, 2024

2. 2024년 싱가포르 인공지능 정책

1) 싱가포르의 AI 이니셔티브 추진¹¹⁶⁾

(1) 개요

싱가포르 정보통신부(MCI)는 2024년 3월, 예산위원회에서 경제 활성화 전략으로 5가지 AI 이니셔티브를 발표했다.

- ① AI 연구인재 육성을 위한 새로운 이니셔티브
- ② 향후 3년간 2,000만 달러 이상을 투자하여 SG 디지털 장학금과 AI 역할에 대한 해외 인턴십의 수 확충

116) Smartnation, Factsheet - Artificial Intelligence (AI) initiatives launched to uplift Singapore’s economic potential, 2024.03.01 (<https://www.smartnation.gov.sg/media-hub/press-releases/20240301a/>)

- ③ AI 혁신 및 역량 구축을 위한 고성능 컴퓨팅 리소스 확보에 최대 5억 달러 투자
- ④ 기업이 생성 AI 전문 지식과 리소스에 액세스할 수 있도록 하는 Generative AI x Digital Leaders 이니셔티브 출시
- ⑤ AI 추천 및 의사결정 시스템에서 개인정보 사용에 대한 자문 지침 발간

이러한 이니셔티브는 2023년 12월 로렌스 워그 부총리가 시작한 개정된 국가 AI 전략(NAIS 2.0)과 일치하며 NAIS 2.0은 싱가포르를 AI 분야의 리더로 이끌고, 싱가포르와 전 세계의 공공 이익을 위해 AI를 활용하는 것을 목표로 한다.

(2) 주요 추진 내용

① Enabler 1: 인재육성(Talent)

싱가포르에서 숙련된 AI 인재를 양성하는 것인데 이러한 인재는 세 그룹으로 분류할 수 있는데 ▲새롭고 최첨단 AI 활동에 종사하는 최고 수준의 인재인 크리에이터 ▲AI 솔루션의 번역 및 개발에 참여하는 데이터 및 머신러닝 과학자와 엔지니어를 포함하는 실무자 ▲AI 기반 제품 및 서비스 사용자 등으로 분류할 수 있다.

- 정부는 싱가포르와 협력을 구축하기 위한 세계적 수준의 AI 연구자들을 유치 목적으로 새로운 AI 방문 교수직을 출범
- 정부는 싱가포르 AI 연구자들의 자체 파이프라인을 구축하기 위해 현지 대학과 협력해 새로운 AI 석사 프로그램에 700만 싱가포르 달러를 투자
- 5년 동안 실무자 풀을 15,000명으로 늘려 증가하는 AI 수요 지원

② Enabler 2: 산업

싱가포르는 기업 디지털화의 다음 단계를 계획하기 위해 디지털 기업 청사진(DEB)을 개발할 것이라고 발표했다. 여기에는 AI와 같은 신기술을 사용하여 기업의 실험과 혁신을 촉진하는 것이 포함된다.

- 디지털적으로 더 성숙한 기업을 위해 IMDA는 생성 AI 전문 지식과 리소스에 대한 액세스를 제공하여 GenAI를 더 잘 이해하고 기술 파트너와 함께 맞춤형 생성 AI 애플리

케이션을 구현할 수 있게끔 지원한다. 또한 생성 AI 전문지식 및 리소스에 대한 기업들의 접근성을 높이기 위한 Generative AI x Digital Leaders 이니셔티브 추진

③ Enabler 3: 컴퓨팅

싱가포르는 모든 국민이 AI로부터 혜택을 얻을 수 있도록 준비하는 프로그램과 함께, AI 활동을 강화하기 위한 인프라를 구축해야 한다고 생각한다.

- 이를 위해 그래픽 처리장치 및 AI 칩 등 고부가가치 AI 활동 지원을 위한 고성능 컴퓨팅 리소스 확보에 최대 5억 달러 투자하기로 결정

④ Enabler 4: 신뢰할 수 있는 환경

AI의 잠재적 오용을 방지하는 효과적인 안전장치가 있어야 신뢰할 수 있는 환경에서 기술을 개발 할 수 있다. 또한 이 기술이 안전하고 신뢰할 수 있고 책임감 있는 방식으로 개발되고 배포되도록 해야 한다.

- 이를 위해 개인정보보호당국(Personal Data Protection Commission)는 개발자가 AI 시스템 개발을 위해 사용하는 개인 데이터를 보호하는 방법에 대한 지침 발표

제10절 우리나라

1. 그간의 인공지능 정책 추진 경과

정부는 2019년 12월 대통령 주재로 열린 제53회 국무회의를 통해 부처 합동 「인공지능 (AI) 국가 전략」을 발표하고 AI 분야 최상위 국가로의 도약을 위한 AI 정책의 기반을 마련하였다. 정부는 동 전략을 통해 ①세계를 선도하는 AI 생태계 구축, ②AI를 가장 잘 활용하는 나라, ③사람 중심의 AI 구현이라는 3대 목표와 함께 9대 전략 및 100대 실행과제를 제시하고, 2030년까지 AI를 통한 경제효과 455조 원 창출 및 삶의 질 세계 10위 도약이라는 종합적인 국가 비전과 전략을 구상하였다.¹¹⁷⁾

2021년 5월 정부는 전 산업 및 사회에 걸친 AI 기술의 빠른 도입 및 확산에 따라 예상치 못한 사회적 이슈와 우려가 대두됨에 따라 「신뢰할 수 있는 AI 실현 전략」을 수립하고, ①신뢰 가능한 AI 구현 환경 조성 ②안전한 AI 활용을 위한 기반 마련 ③사회 전반의 건전한 AI 의식 확산을 주요 추진전략으로 제시하였다. 정부는 위에 언급된 기술·제도·윤리 측면의 3대 전략을 중심으로 2025년까지 단계적으로 추진할 10대 실행과제를 설정하고 AI 신뢰성을 확보하기 위한 정부 차원의 노력을 강화하였다.¹¹⁸⁾

이후, 정부는 2022년 9월 「대한민국 디지털 전략」과 12월 「신성장 4.0 전략」을 발표하고, 이에 대한 후속조치를 본격화하며 2023년 1월 AI 등 신산업 필요 데이터의 전략적 생산, 해외데이터·연구데이터 등 새로운 데이터 공유 기반 구축을 위한 「제1차 데이터산업 진흥 기본계획」과 AI로 인한 국민의 혜택 증진 및 산업혁신을 위한 대규모 AI 수요창출을 위한 「인공지능 일상화 및 산업 고도화 계획」을 발표하였다.¹¹⁹⁾

또한, 정부는 3월 관계부처와 산업계가 참여하는 초거대 AI 산업 발전을 위한 ‘인공지능

117) 과학기술정보통신부, 인공지능(AI) 국가전략발표, 2019.12.17.:

(<https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=112&pageIndex=1&bbsSeqNo=94%20&nnttSeqNo=2405727>)

118) 과학기술정보통신부, 신뢰할 수 있는 인공지능 구현 전략 발표, 2021.5.13.:

(<https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=112&pageIndex=&bbsSeqNo=94&nnttSeqNo=3180239>)

119) 과학기술정보통신부, 한덕수 총리, 세계 최고의 데이터 생태계 구축으로 인공지능이 일상화 되는 대한민국의 미래를 그린다., 2023.1.26.:

(<https://www.msit.go.kr/bbs/view.do?sCode=user&nnttSeqNo=3182658&pageIndex=&searchOpt=ALL&bbsSeqNo=94>)

최고위 전략대화' 를 개최하고, 초거대 AI 경쟁력 강화 및 산업생태계 조성 방안, 챗GPT 등 초거대AI 확산에 따른 사회적·문화적 영향력, 인공지능 관련 규제 개선 및 윤리·신뢰성 제고 방안 등을 논의하였다.¹²⁰⁾

이후 정부는 4월 정부의 대표 국정과제 중 하나인 디지털플랫폼정부 추진을 위해 관계부처 합동으로 「디지털플랫폼정부 실현계획」과 「초거대 인공지능 경쟁력 강화 방안」을 마련하고¹²¹⁾, 7월 초거대 AI 개발·고도화를 지원하는 기술·산업 인프라 확충, 민간·공공 초거대 AI 융합 등 초거대 AI 혁신 생태계 조성, 범국가 AI 혁신 제도·문화 정착 등 초거대 AI 활용지원 사업에 본격 착수하였다.¹²²⁾

정부는 9월 「대한민국 인공지능 도약방안」과 「전국민 AI 일상화 실행계획」을 발표하고 디지털 모범국가를 향한 의지를 표명하였다. 또한, 정부는 ‘뉴욕구상(’ 22.9)’, ‘캐나다 인공지능 석학과의 대화(’ 22.9)’, ‘과리 이니셔티브(23.6)’, ‘한-아세안 인공지능 유스 페스타(23.9)’를 통해 새로운 디지털 규범·질서 정립을 위한 논의를 주도하며, 글로벌 AI 협력을 확대해왔다.¹²³⁾

2. 2024년 인공지능 정책 동향

1) 과학기술정보통신부의 민·관 합동 AI전략최고위협의회의 출범¹²⁴⁾

과학기술정보통신부는 2024년 4월 4일 민·관 AI 최고위 거버넌스인 ‘AI전략최고위협의회의’의 출범을 발표했다. 생성 AI 등장과 함께 AI가 산업을 넘어 인문·사회 분야까지 영향력이 확대되며 다층적이고 복합적인 문제를 촉발함에 따라 기존 분야별로 운영되던 포럼·

120) 과학기술정보통신부, 과기정통부, 초거대 인공지능(AI) 산업 발전을 위한 인공지능 최고위 전략대화 개최, 2023.3.8.; (<https://www.msit.go.kr/bbs/view.do?mId=113&mPid=238&bbsSeqNo=94&nttSeqNo=3182812>)

121) 과학기술정보통신부, 디지털플랫폼정부 추진 본격 시동(초거대 인공지능(AI) 경쟁력 강화 방안 포함), 2023.4.14.; (<https://www.msit.go.kr/bbs/view.do?sCode=user&bbsSeqNo=94&nttSeqNo=3182943>)

122) 과학기술정보통신부, 중소·벤처기업, 공공기관·지자체 대상 초거대 인공지능 활용지원, 첫 발 땀다, 2023.7.3.; (<https://www.msit.go.kr/bbs/view.do?sCode=&mId=113&mPid=238&pageIndex=&bbsSeqNo=94&nttSeqNo=3183257>)

123) 과학기술정보통신부, 디지털 대항해 시대, 글로벌 모범국가를 향한 대한민국 초거대 인공지능 도약 추진, 2023.9.13.; (<https://www.msit.go.kr/bbs/view.do?mId=113&mPid=238&bbsSeqNo=94&nttSeqNo=3183480>)

124) 과학기술정보통신부, 민·관이 머리를 맞대고 인공지능 3대 강국 도약 방안을 모색하기 위한 인공지능 최고위 협의체 『인공지능 전략최고위협의회의』 출범, 2024.4.4.; (<https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=238&pageIndex=87&bbsSeqNo=94&nttSeqNo=3184296&searchOpt=ALL&searchTxt=>)

자문위원회·협의체 등의 한계를 인식하고 상호 연계되고 통합된 시각에서 국가 전체 AI 혁신의 방향을 이끌 거버넌스가 필요하다는 범정부적 공감대가 형성됨에 따라 동 협의회가 구성되었다.

동 협의회는 정부 및 민간 부문을 대표하는 공동 위원장을 중심으로 AI 분야(정책일반, AI 반도체, R&D, 법·제도, 윤리안전, 인재 등)를 대표하는 최고의 민간 전문가 23인과 주요 관계부처(과학기술정보통신부, 기획재정부, 산업통상자원부, 중소벤처기업부, 교육부, 개인정보보호위원회, 방송통신위원회 등) 실장급 공무원 7인 등 총 32인으로 구성되었다. 또한, 협의회는 기존 분야별 협의체 등을 활용 또는 일부 신설하여 6개 분과(AI 반도체, R&D, 법·제도, 윤리안전, 인재, AI 바이오)로 구성된 산하 분과위원회를 운영할 예정이며, 국내 100여 개 이상 초거대 AI 기업과 중소기업 및 스타트업으로 구성된 ‘초거대 AI 추진협의회’를 외부 민간 자문단으로 활용하여 AI 혁신생태계 구현을 위한 정부의 투자·지원 등 정책 제언을 수렴하고, 민·관 간 상시적 소통 채널로 적극 활용할 계획이다.

출범식은 ▲ ‘AI전략최고위협의회의’ 운영 방안 ▲AI G3 도약을 위한 AI·디지털 혁신성장 전략 ▲AI 일상화를 위한 ‘24년 국민·산업·공공 프로젝트 추진계획 등 3가지 주요 안건에 대한 발표 및 토론 순으로 진행되었다.

2) AI G3 도약을 위한 AI·디지털 혁신성장 전략 발표¹²⁵⁾

정부는 지난 4월 4일 ‘AI전략최고위협의회의’ 출범식을 통해 AI 3대 강국(G3) 도약을 위한 민·관 합동 전략으로써 ‘AI·디지털 혁신성장 전략’을 발표하였다. 동 전략은 2022년 9월 「대한민국 디지털 전략」이 발표된 이후 생성 AI 등장에 따른 잠재적 산업·경제적 기회를 포착하고 활용하는 것에 중점을 두고 있으며, 정부는 AI를 통한 혁신이 저성장, 저출산, 고령화 등 한국이 직면한 구조적 한계를 극복할 수 있는 재도약의 기회로 인식하고, 이를 돌파하기 위한 4대 전략분야별 프로젝트를 추진한다.

- (혁신) 글로벌 선도 수준의 AI 기술 혁신
- (산업) 제조 및 서비스 전반의 AI 대전환 가속화

125) 관계부처 합동, AI G3 도약을 위한 AI·디지털 혁신성장 전략, 2024.4.4.

- (사회) 국민 일상에 선도적 AI 도입 추진
- (제도) 온전한 수용을 위한 새로운 디지털 질서 정립

[그림 3-7] 디지털 혁신성장 전략의 비전 및 추진과제



자료: 관계부처 합동, AI G3 도약을 위한 AI·디지털 혁신성장 전략, 2024.04.04.

3) AI 일상화를 위한 추진계획(안) 발표¹²⁶⁾

정부는 2023년 9월 「전국민 AI 일상화 추진 계획」을 통해 국민과 AI 혜택을 공유하고 글로벌 AI 활용도 선도국가로 도약하기 위한 비전을 선포하고 AI를 일상·일터·공공에 접목시키고 이를 통한 혜택 체감이 가능한 혁신을 창조하기 위한 전략을 수립하였다. 이에 따라 정부는 민간 의견 및 대국민 설문조사 결과를 수렴하고 AI 일상화를 본격화하기 위한 노력에 착수하고, 지난 4월 4일 ‘AI전략최고위협의회’ 출범식을 통해 「AI 일상화를 위한 ’24년 국민·산업·공공 프로젝트 추진계획(안)」을 발표하였다.

설문조사를 통해 일상 속 AI에 대한 국민의 인식을 조사한 결과, 우리나라 국민은 AI에 대해 긍정적인 인식을 갖고 있으며, 국민의 AI 서비스 경험률은 ‘23년 기준 50.8%로 최근 3년간 지속적으로 증가하는 추세를 보이고, 주요 경험 분야는 주거, 편의, 교통, 교육 순으로 나타났다.

국민들은 AI 일상화에 있어 의료·보건 등 사회복지 분야와 안전·편의 등 공공 분야를 가장 중요한 분야로 생각하고 있으며, 이와 유사하게 전문가들도 사회복지 분야를 가장 중요하게 인식하며, 특히 장애인·노인 등 취약계층 대상 AI 서비스를 중점 분야로 인식하고 있다.

〈표 3-55〉 국민의 일상 속 AI에 대한 경험률 및 도움정도

구분	주거·편의	교통	교육·학습	커뮤니케이션	미디어	AI 대화형 검색	금융	헬스케어
경험률	20.8%	16.6%	14.6%	12.9%	10.7%	8.6%	5.6%	5.1%
도움정도	95.3%	94.4%	91.2%	84.4%	90.2%	92.6%	80.8%	87.8%

출처: 관계부처 합동, AI 일상화를 위한 ’24년 국민·산업·공공 프로젝트 추진계획(안), 2024.04.04.

126) 관계부처 합동, AI 일상화를 위한 ’24년 국민·산업·공공 프로젝트 추진계획(안), 2024.4.4.;
https://doc.msit.go.kr/SynapDocViewServer/viewer/doc.html?key=6f3c7e1b2da54bcf99bf333d297d67cd&convType=html&convLocale=ko_KR&contextPath=/SynapDocViewServer/

〈표 3-56〉 AI 서비스의 사회적·개인적 영향에 관한 국민의 인식

질문내용	긍정 답변율(%)		
	전체	경험자	미경험자
AI 기반 서비스는 사회전반에 긍정적인 영향을 준다	60.7	68.0	55.8
AI 서비스는 나에게 심리적 도움 및 즐거움을 준다	55.8	62.7	51.1
AI 서비스는 모두에게 공정하고 차별없는 서비스이다	52.6	59.6	47.9

출처: 관계부처 합동, AI 일상화를 위한 '24년 국민·산업·공공 프로젝트 추진계획(안), 2024.04.04.

정부는 국가 전반의 AI 활용도 제고와 국민체감 확산을 동 추진계획(안)의 비전으로 설정하고 ▲국민의 AI 서비스 경험률을 51%(’23)에서 60%(’24)까지 ▲기업의 AI 도입률을 28%(’23)에서 40%(’24)까지 ▲공공부문의 AI 도입률을 55%(’22)에서 80%(’24)까지 높이는 것을 목표로 올해 총 7,102억 원 규모의 69개 과제를 추진한다.

주요 분야별 추진과제를 살펴보면, 정부는 ①국민생활에서의 AI 일상화를 위한 18개 과제(755억 원 규모) ②산업현장에서의 AI 일상화를 위한 24개 과제(2,881억 원 규모) ③정부행정에서의 AI 일상화를 위한 14개 과제(1,157억 원 규모) ④기반인프라에서의 AI 일상화를 위한 13개 과제(2,309억 원)를 추진한다.

〈표 3-57〉 AI 일상화 주요 추진과제

구분	주요 내용
① 국민생활	(18개 과제 / 755억 원 투입) <ul style="list-style-type: none"> • (건강·질병) AI 서비스 확산으로 국민의 건강 향상 기여 • (사회·복지) 장애인·어르신 일상생활 보조 및 사각지대 없는 복지 기여 • (보육·교육) 맞춤형 돌봄·교육 제공 지원으로 부모의 걱정 완화 • (문화·주거) AI 콘텐츠 전국민 이용 지원, AI 홈서비스 제공 등
② 산업 현장	(24개 과제 / 2,881억 원 투입) <ul style="list-style-type: none"> • (법률·미디어 등 민간 전문영역) 전문가 업무보조 및 국민 접근성 향상 • (바이오·제약·의료) 신약·항체개발 첨단화 및 지역 의료 AI 확산 기여 • (소비·유통물류) 소상공인 영업지원·노동감소 및 스마트 물류관리 촉진 • (제조·공정) 공정효율 향상·비용절감 기여, 시로 인한 새로운 서비스 창출 기여 • (농·축·수산) 무인화·자동화·안전화로 전통적 지배환경의 진화 촉발

구분	주요 내용
③정부 행정	(14개 과제 / 1,157억 원 투입) <ul style="list-style-type: none"> • (재난·안전) 화재·홍수 등 국가적 재난 상황의 위기대응 능력 향상 • (공공·행정) 대국민 서비스의 질은 향상, 행정 업무의 양은 감소
④기반 인프라	(13개 과제 / 2,309억 원 투입) <ul style="list-style-type: none"> • (리터러시·인재) 국민 일상의 AI 접근 향상, 현장형·실무형 인재 양성 • (격차 해소) 지역·계층 등 소외 없는 맞춤형 AI 교육 제공 • (신뢰·안전) 민간자율 AI 신뢰·안전성 확보 지원과 생성 AI로 인한 위험 대응

출처: 관계부처 합동, AI 일상화를 위한 '24년 국민·산업·공공 프로젝트 추진계획(안), 2024.04.04.

4) 산업통상자원부의 AI 자율제조 전략 1.0 발표¹²⁷⁾

산업통상자원부는 2024년 5월 8일 AI 보편화 시대를 앞두고 산업의 변화상을 전망하고 AI 활용 산업 혁신을 위한 정책과제를 도출하기 위해 「AI 시대의 신산업정책」 수립을 위한 ‘AI산업정책위원회’를 출범하고, 「AI 시대의 신산업정책」의 6대 중점 분야(①자율제조 ②디자인 ③R&D ④유통 ⑤에너지 ⑥AI 반도체) 중 첫 번째 과제인 자율제조 관련 「AI 자율제조 전략 1.0」을 발표하였다.

산업통상자원부는 「AI 자율제조 전략 1.0」의 3대 중점 분야인 ①AI 자율제조 도입 확산 ②AI 자율제조 핵심역량 확보 ③생태계 진흥을 중심으로 올해 1천억원 이상의 예산을 투입할 예정이며, 현재 9% 수준에 불과한 AI 자율제조확산률을 2030년까지 30%이상, 제조 생산성을 20% 이상 높이는 것을 목표로 한다.

5) 과학기술정보통신부의 새로운 디지털 질서 정립 추진계획 보고¹²⁸⁾

과학기술정보통신부는 2024년 5월 21일 제22회 국무회의에서 관계부처 합동으로 마련한 「새로운 디지털 질서 정립 추진계획」을 보고하였다. 동 추진계획은 2023년 9월 대통령의 디지털 구상을 담아 발표된 「디지털 권리장전」의 내용을 구체적인 정책으로 구현하기 위한 범부처 계획으로, 「디지털 권리장전」의 철학과 5대 원칙(자유, 공정, 안전, 혁신, 연대)

127) 산업통상자원부, 인공지능(AI) 시대 우리 산업이 나아갈 길을 그린다, 2024.5.8.:
<https://www.motie.go.kr/attach/viewer/095a2dda9c864e1d90d751f7668a1117/af3d48bb9abc784568ef7546f7f97280/9a9db098b587ee18b321c826f3707a49>

128) 과학기술정보통신부, 대한민국이 새로운 디지털 질서 정립의 마스터플랜을 공개합니다, 2024.5.21.:
https://www.msit.go.kr/bbs/view.do?sessionId=wjWfVeQYbgNDSSzcdphWZZkitLN-IYF-5MRh_Pe3.AP_msit_1?sCode=user&mPid=238&mId=113&bbsSeqNo=94&nttSeqNo=3184512

을 바탕으로 52개의 쟁점을 해소하기 위한 20대 정책과제를 제시한다.

정부는 20대 정책과제 중 국민의 관심도, 쟁점의 파급성 및 시급성이 높은 8대 정책과제를 핵심과제로 지정하고 가시적 성과 달성을 위해 집중관리할 계획이며, 디지털 자산의 규범 정립, 디지털 심화에 따른 노동·교육·사회 시스템 정비 등 나머지 12개 정책과제 또한 지속적으로 추진해 나갈 계획이다.

<표 3-58> 새로운 디지털 질서 정립 추진을 위한 핵심과제

구분	주요 내용
① AI 기술의 안전성 및 신뢰윤리 확보	<ul style="list-style-type: none"> 이용자 보호 등 AI 혁신과 안전·신뢰의 균형을 위한 법제 제정을 연내 마무리하여 AI 규범 체계를 선도적으로 정립 「AI 서울 정상회의」를 통한 글로벌 AI 규범·거버넌스 논의 주도 AI 안전성을 검증·연구하는 전담조직을 설치하여 아태지역의 AI 안전허브로 육성
② 딥페이크를 활용한 가짜뉴스 대응	<ul style="list-style-type: none"> 딥페이크 관련 법령 제정 및 개정을 통한 AI 생성물의 워터마크 표시 의무화 민·관 협업 대응 체계 강화 및 딥페이크 탐지·식별 기술 개발 등 딥페이크 가짜뉴스의 생성·유통·확산 전주기에 대한 대응체계 구축
③ AI 개발·활용 관련 저작권 제도 정비	<ul style="list-style-type: none"> 연말까지 「저작권법」 등 저작권 제도 정비방안 마련
④ 디지털 재난 및 사이버 위협·범죄 대응	<ul style="list-style-type: none"> 고도화·지능화되는 디지털 위협에 철저히 대비하는 국가 대응체계 확충 「디지털서비스 안전법」 제정 추진 피상·디지털성범죄 피해 예방을 위한 「사기방지법」 제정 및 「성폭력방지법」 개정 등 민생 사이버 범죄 대응체계 정비 4대 핵심 보안기술(▲데이터·AI 보안 ▲네트워크 보안 ▲디지털 취약점 대응 ▲신산업 융합보안) 개발에 올해 1,141억 원 집중 투자(작년 대비 22.5% 증가)
⑤ 디지털 접근성 제고·대체 수단 확보	<ul style="list-style-type: none"> 디지털 접근성 증진을 위해 소외계층을 대상으로 맞춤형 디지털 포용서비스 제공 강화 디지털 소외계층을 위해 행정·금융 등 필수영역에서 디지털 대체 수단 확대
⑥ 비대면 진료의 안정적 시행	<ul style="list-style-type: none"> 「의료법」 개정을 통한 비대면 진료의 법적근거를 마련 및 디지털 혁신기술과 서비스의 비대면 진료 연계를 강화 등을 통한 비대면 진료의 본격 제도화 개인 건강정보보호, 처방전 위변조 방지 등 관리체계 개선방안 마련
⑦ 연결되지 않을 권리 보호	<ul style="list-style-type: none"> 노·사정 논의를 통해 연결되지 않을 권리에 대한 공문화 본격 착수 원격·유연근무, 초과근무 빈번한 디지털 기업을 중심으로 자발적 인식개선 유도
⑧ 잊힐 권리 보장	<ul style="list-style-type: none"> 온라인에 수많은 개인정보가 누적되어 특별한 법적 보호가 요구되는 디지털 네이티브인 아동·청소년들의 잊힐 권리 제도화 성인 이전 시기에 본인이 온라인에 게시한 개인정보가 포함된 글·사진·영상 등 게시물을 삭제하거나 다른 사람이 검색하지 못하도록 도와주는 서비스인 지우개 서비스 확대

출처: 과학기술정보통신부, 대한민국이 새로운 디지털 질서 정립의 마스터플랜을 공개합니다, 2024.05.21.

6) AI 서울 정상회의¹²⁹⁾

윤석열 대통령은 2024년 5월 21일 리시 수낙(Rishi Sunak) 영국 총리와 공동으로 「AI 서울 정상회의(5.21~22)」 정상 세션을 주재하고, 개회사를 통해 안전, 혁신, 포용을 3대 AI 규범가치로 제시하였다. 윤석열 대통령은 동 정상회의가 새로운 디지털 규범 정립을 위한 우리 정부와 국제사회의 노력을 결집하고, 글로벌 AI 거버넌스가 한 단계 진전하는데 기여했다고 평가하였다.

윤석열 대통령은 동 정상회의를 통해 채택된 「안전하고 혁신적이며 포용적인 AI를 위한 서울 선언」과 관련하여 AI 정상회의 논의를 안전에서 혁신, 포용으로 확장했다는 점에서 의의가 있음을 강조하였다. 동 선언은 ▲상호운용성을 지닌 보안 조치의 중요성 ▲각국의 AI 안전연구소 간 네트워크 확대 및 글로벌 협력 촉진 ▲안전·혁신·포용적인 AI 생태계를 조성하기 위한 정책 개발 및 거버넌스 구축 의지 ▲기업들을 비롯한 다양한 이해관계자들의 참여 필요성 등의 내용을 포함하며, 각국 정상들은 동 정상회의를 통해 AI 안전, AI 혁신, AI 포용이라는 3대 목표에 공감대를 형성하였다.

7) 과학기술정보통신부의 AI 분야 규제 샌드박스 과제 발굴¹³⁰⁾

과학기술정보통신부는 2024년 6월 28일 ICT 규제샌드박스 제도개선 방안을 발표하고, AI 분야 규제샌드박스 과제를 발굴·기획하기 위한 노력을 본격화하였다. 정부는 최근 AI 분야의 규제 혁신에 대한 수요가 증가하는 가운데 관련 법령 개정 등 규제 정비에 오랜 시간이 요구됨에 따라 규제 샌드박스를 통해 혁신적인 AI 서비스가 시장에 출시되도록 선제적으로 발굴하고 지원해 나갈 방침이다.

정부는 어린이보호구역 등 일정구역 내 스마트폰, 영상정보처리장치(CCTV) 등을 통해 주변 보행자 및 차량의 위치정보를 수집하고 AI 학습을 통해 어린이 등 앱 설치 이용자에게 보행중 출동위험 알림과 같은 교통안전정보를 제공하는 서비스인 ‘교통안전 스마트폰 솔루션 Soft V2X((주)LG전자)’의 서비스가 가능하도록 「개인정보보호법」 및 「위치정보법」에 대한 관계부처의 법령해석을 통해 적극해석 처리하였다.

129) 대통령실,尹 대통령, AI 서울 정상회의 공동 주재하며 AI 기술 선도국 간 안전·혁신·포용의 3대 글로벌 AI 거버넌스 목표 합의 이끌어, 2024.5.22.:

(<https://www.president.go.kr/newsroom/briefing/DbNv3EKL>)

130) 과학기술정보통신부, 인공지능 분야 규제 유예(규제샌드박스) 과제를 적극 발굴한다, 2024.6.28.:

(<https://www.msit.go.kr/bbs/view.do?sCode=user&bbsSeqNo=94&nttSeqNo=3184665>)

또한, 정부는 2024년 1월 「국가전략기술육성법」에 따른 특화연구소로 지정된 서울대학교병원이 신청한 ‘국가전략기술 특화연구소 데이터플랫폼’에 대한 실증특례를 지정하였다. 동 플랫폼은 서울대학교병원에서 첨단바이오 분야 연구 목적으로 구축하는 의료데이터에 대한 가명처리 플랫폼으로, 추후 실증기간 동안 국내·외 연구자가 동 플랫폼을 통해 의료 AI·빅데이터 등 공동 연구를 수행하게 될 예정이다.

과학기술정보통신부는 2024년 5월 개최된 「AI 서울 정상회의」에서 논의된 바와 같이 새로운 AI시대에는 안전·포용·혁신의 균형이 중요함을 강조하며, AI 혁신을 위한 규제 샌드박스의 중요성에 대해 인식하고, 규제 샌드박스를 통해 AI 혁신 서비스가 시장에 출시되고 발전을 이어갈 수 있도록 지속적으로 지원할 방침이다.

8) 문화체육관광부의 AI 시대의 저작권 제도 모색¹³¹⁾

문화체육관광부는 2024년 2월 AI 시대의 저작권 제도 정비 방안을 모색하기 위해 학계, 법조계, AI 사업자, 권리자, 산업기술계 등 전문가와 이해관계자로 구성된 협의체인 ‘AI-저작권 제도 개선 워킹그룹’을 발족한 가운데 6월 28일 한국저작권위원회와 함께 동 워킹그룹의 제2차 전체 회의를 개최하고 상반기 분과별 논의 결과를 공유하고 하반기 논의 주제를 선정하였다.

<표 3-59> 상반기 분과별 논의 결과

구분	주요 내용
학습분과	(AI 학습에 저작물을 이용할 경우 적절한 보상 방안 마련 필요) <ul style="list-style-type: none"> ‘AI 학습용 저작물의 적법한 이용 권한 확보 방안’을 주제로 3개월간 집중 논의 저작물 이용에 대한 보상 논의 시 학계나 공공영역에서 권리자와 AI 사업자 간 협상을 지원해 주는 것이 필요 AI 사업자들이 글로벌 AI 산업에서의 주도권 경쟁에서 뒤처지지 않도록 「저작권법」상 텍스트 및 데이터 마이닝(TDM) 예외 조항의 신설 등 산업 지원 방안 필요
산출 및 이용 분과	(AI 산출물이라는 새로운 개념과 저작권 제도 정비 논의) <ul style="list-style-type: none"> AI 산출물의 저작권 침해 판단기준 및 저작권 등록, AI 산출물 표시제도 등에 관해 논의 AI 산출물 표시제도 도입의 필요성이 제시되었으며, 도입 의무화가 필요한 영역과 표시 방법, 기술적 한계 등에 대한 세밀한 검토 필요 최근 시행된 EU 「인공지능 법(AI Act)」, 우리나라의 「인공지능 기본법(인공지능산업 육성 및 신뢰 확보에 관한 법률)」 제정안, 「콘텐츠 산업진흥법」 개정안 등 ‘AI 산출물 표시제도’가 포함된 입법 사례 검토

출처: 문화체육관광부, 인공지능 시대 저작권 제도 정비의 실마리를 찾아보다, 2024.06.28.

131) 문화체육관광부, 인공지능 시대 저작권 제도 정비의 실마리를 찾아보다, 2024.6.28.:

(https://mcst.go.kr/kor/s_notice/press/pressView.jsp?pSeq=21201)

‘AI-저작권 제도 개선 워킹그룹’은 하반기 ▲AI 학습데이터 공개 여부 및 방법, ▲권리자가 원치 않을 때 AI 학습데이터에서 제외할 수 있는 제도적·기술적 방안, ▲AI 산출물에 인간이 창작적으로 기여한 바를 입증하는 방안, ▲AI 산출물의 법적 보호 여부와 방식 등의 주제에 관해 논의할 예정이다. 한편, 문화체육관광부는 동 협의체를 운영하는 동시에 ‘AI 학습 저작물에 대한 적정 이용대가 산정방안 연구’와 ‘AI-저작권 제도 개선방안 연구’를 진행하여 2024년 말 AI 관련 저작권 제도 정비안을 발표할 계획이다.

9) 금융위원회의 AI 정책 금융 프로그램 신설 발표¹³²⁾

금융위원회가 2024년 7월 4일 주재한 제7차 정책금융지원협의회(금융위원회, 과학기술정보통신부, 산업통상자원부, 중소벤처기업부, 국토교통부, 특허청, 산업은행, 기업은행, 신용보증기금 등 참여)를 통해 정부 관계부처 및 정책금융기관은 AI 분야에 대한 정책금융 지원 필요성이 지속적으로 제기됨에 따라 3조원 규모의 저리대출프로그램과 5천억 원 규모의 펀드를 신설한다고 발표하였다.

기존 5대 산업 분야(반도체·디스플레이·이차전지·바이오·원전)를 중심으로 진행되던 산업은행의 첨단산업 지원 프로그램인 ‘초격차산업지원프로그램’ 내 AI 분야를 신설하여 올해 말까지 3조원 한도의 자금을 공급할 예정이며, 산업은행은 자체 자금 1,500억 원에 민간자금 3,500억 원을 유치하여 AI 기술 개발 및 생태계 육성 지원을 위한 ‘AI 코리아 펀드(가칭)’을 5천억원 규모로 신규 조성할 계획이다.

[그림 3-8] 초격차산업지원 프로그램 6호 AI

분야	한도	주요 내용
AI	3조원	<ul style="list-style-type: none"> ▪ (지원대상) AI모델, 로봇자율주행, 클라우드, AI반도체 영위 기업 ▪ (지원분야) 설비투자, R&D, M&A, 경상 운영자금 등 ▪ (금리우대) 최대 △1.2%p ▪ (기타사항) 운영자금 한도 확대(120%)

자료: 금융위원회, 3.5조원 규모의 AI정책금융 프로그램 신설, 클라우드·로봇·자율주행 등 AI 소분야 지원강화, 2024.07.04.

132) 금융위원회, 3.5조원 규모의 AI정책금융 프로그램 신설, 클라우드·로봇·자율주행 등 AI 소분야 지원강화, 2024.7.4.: (<https://www.fsc.go.kr/no010101/82604>)

[그림 3-9] AI 코리아 펀드(안)

- (펀드규모) '24년 5,000억원(산은 출자 1,500억원(30%) 이내)
- (투자대상) ①클라우드, AI 모델 및 ②AI 응용 분야 (시장반응 탐색 후 구체화)
- (펀드구조) 다수의 블라인드펀드로 출자사업 진행
- (추진일정) 신속하게 사업공고 등 실시

자료: 금융위원회, 35조원 규모의 AI정책금융 프로그램 신설 클라우드·로봇·자율주행 등 AI 쏠분야 지원강화, 2024.7.4.

10) 산업통상자원부의 AI 디자인 확산 전략 발표¹³³⁾

산업통상자원부는 2024년 5월 발표한 「AI 시대의 신산업정책」의 6대 분야별 활용 전략 중 첫 번째 전략인 자율제조에 이어 두 번째 AI 산업 활용 전략으로서 7월 9일 「AI 디자인 확산 전략」을 발표하였다. 동 전략은 ① AI 디자인 플래그쉽 프로젝트 추진 ② AI 디자인 핵심 경쟁력(기술·기업·데이터·인재) 확보 ③ AI 디자인 시장 활성화 지원 ④ AI 시대에 부합하는 디자인 제도 규범 설계의 네 가지 핵심축을 중심으로 디자인기업의 AI 도입률 50% 달성, AI 디자이너 1만 명 양성, AI 디자인 혁신기업 500개 육성, 전산업의 디자인 활용률 60%(현재 37%) 달성을 목표로 제시한다.

11) 중소벤처기업부의 초격차 AI 스타트업 성장 전략 발표¹³⁴⁾

중소벤처기업부는 2024년 7월 10일 글로벌 유니콘으로 성장할 신성장 AI 분야 초격차 스타트업을 집중 육성을 위한 「초격차 인공지능(AI) 창업기업(스타트업) 성장(레벨업) 전략」을 발표하였다. 동 전략은 ①AI 스타트업 고성장 5개 분야에 대한 전략적 지원, ②국내시장 수요를 활용한 성장 강화, ③글로벌 시장 진출을 통한 성장 동력 확대, ④전문인력 활용 제고 및 거버넌스 구축 등 총 4개의 전략과제로 구성되었다.

133) 산업통상자원부, K-디자인과 인공지능(AI) 융합을 위한 첫 청사진 나왔다, 2024.7.9.:
(<https://www.motie.go.kr/kor/article/ATCL3f49a5a8c/169295/view>)

134) 중소벤처기업부, 국제적(글로벌) 거대신생기업(유니콘)으로 성장할 고성장 인공지능(AI) 분야 초격차 창업기업(스타트업) 전략 육성, 2024.7.10.:
(<https://www.mss.go.kr/site/smba/ex/bbs/View.do?cbIdx=86&bcIdx=1051756&parentSeq=1051756>)

〈표 3-60〉 초격차 AI 스타트업 육성 전략의 중점 추진 방향

추진 방향	주요 내용
5개의 고성장 AI 분야 스타트업을 선별하여 전략적으로 지원	<ul style="list-style-type: none"> • (유망 기술) ① sLLM 스타트업, ② AI 펌리스 스타트업 • (유망 산업) ① 제조, ② 헬스케어, ③ 문화콘텐츠
국내 시장의 수요를 활용한 성장 기반 강화	<ul style="list-style-type: none"> • 대기업, 중소기업 등과 협업 매칭 프로그램 추진, 공공기관 구매, 지역 주력산업 연계
글로벌 시장 진출로 성장 동력 확대	<ul style="list-style-type: none"> • 해외 AI 확회 및 전시회 참여, 해외 인증 및 글로벌 투자 활성화 지원 등
전문인력 활용성 제고 및 거버넌스 구축	<ul style="list-style-type: none"> • 국내·외 AI 전문인력 활용 및 청년 인재의 AI 스타트업 유입 확대, 전문가 포럼 구성

출처: 중소벤처기업부, 국제적(글로벌) 거대신생기업(유니콘)으로 성장할 고성장 인공지능(AI) 분야 초격차 창업기업(스타트업) 전략 육성, 2024.07.10.

12) 개인정보보호위원회의 AI 이용 ‘공개 데이터’ 처리 기준 제시¹³⁵⁾

개인정보보호위원회는 2024년 7월 17일 AI 개발에 필수적인 공개 데이터가 현행 개인정보 규율체계 내에서 적법하고 안전하게 처리되도록 지원하기 위해 「인공지능(AI) 개발·서비스를 위한 공개된 개인정보 처리 안내서」를 발표하였다. 동 안내서는 공개된 개인정보 수집·활용의 법적기준을 명확화하고, 기업이 AI 개발 및 서비스 단계에서 취해야 할 안전조치의 적정성과 관련하여 참고할만한 최소한의 기준을 제시함으로써 공개된 개인정보를 활용하는 기업들이 개인정보 침해 이슈를 최소화하고 법적 불확실성을 해소해 혁신성장을 이룰 수 있도록 지원한다.

[그림 3-10] ‘정당한 이익’ 조항 적용을 위한 요건의 내용 및 적용사례

요건	주요 내용
목적의 정당성	<ul style="list-style-type: none"> • 개인정보처리자의 정당한 이익의 존재 • 공개된 개인정보 처리를 통해 개발하려는 AI의 목적·용도를 구체화하여 정당한 이익을 명확화 ※ (예) 의료진단보조, 신용평가, 텍스트 생성·분류·번역 등을 수행하는 LLM 등
처리의 필요성	<ul style="list-style-type: none"> • 공개된 개인정보 수집·이용의 필요성과 상당성·합리성이 인정될 것 ※ (예) 의료진단보조 AI 개발시 개인의 소득·재산 등 관련없는 정보는 학습 배제
구체적 이익형량	<ul style="list-style-type: none"> • 개인정보처리자의 정당한 이익이 정보주체 권리에 명백히 우선 • ‘명백성’ 요건 충족을 위해 (i) 정보주체 권익침해 방지를 위한 안전성 확보조치 및 (ii) 정보주체 권리보장 방안 마련·시행을 통해 개인정보처리자 이익이 우선하도록 조치 ※ 상세내용 후술

자료: 개인정보보호위원회, 인공지능(AI) 개발·서비스에 이용되는 ‘공개 데이터’ 처리 기준 제시, 2024.07.17.

135) 개인정보보호위원회, 인공지능(AI) 개발·서비스에 이용되는 ‘공개 데이터’ 처리 기준 제시, 2024.7.17.; (<https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=10362>)

동 안내서는 개인정보보호법 제15조에 따른 ‘정당한 이익’ 조항에 의해 공개된 개인정보를 AI 학습 및 서비스 개발에 활용할 수 있음을 설명하고, 해당 조항이 적용되기 위해서는 세 가지 요건(▲AI 개발 목적의 정당성 ▲공개된 개인정보 처리의 필요성 ▲구체적 이익 형량)을 충족해야 함을 강조한다. 또한, 동 안내서는 AI 기업이 ‘정당한 이익’을 근거로 공개된 개인정보를 처리하기 위해 고려할 수 있는 기술적·관리적 안전조치와 함께 정보주체 권리보장 방안을 제시한다. AI 기업은 모든 안전조치를 의무적으로 시행해야 하는 것은 아니며, 제시된 안전조치의 순기능과 부작용(AI 성능저하, 편향성 등), 기술 성숙도를 고려하여 기업의 특성에 맞는 최적의 조합을 식별하고 이행할 수 있다.

[그림 3-11] 기술적·관리적 안전조치 및 정보주체 권리보장 방안



자료: 개인정보보호위원회, 인공지능(AI) 개발·서비스에 이용되는 ‘공개 데이터’ 처리 기준 제시, 2024.07.17.

13) 산업통상자원부의 국산 AI 반도체의 주력산업 도입 방안 발표¹³⁶⁾

온디바이스 AI(스마트폰 등의 기기에 직접 탑재되는 AI)가 전 산업분야로 확산되고 있는 가운데 산업통상자원부는 2024년 8월 14일 AI 반도체 수요·공급기업과 함께 자동차, 가전 등 7대 주력산업에 국산 AI 반도체를 도입하기 위한 방안을 논의하고 산업별 AI 반도체 개발·적용 방향을 제시하였다. 이번에 발표된 온디바이스 AI 반도체 개발은 2025년부터 시작되는 신규 사업으로, 산업통상자원부는 향후 주력산업별 AI 플랫폼을 개발·실증하는 대형 사업을 추가 기획할 예정이다.

<표 3-61> 산업별 온디바이스 AI 반도체 개발 계획

구분	주요 내용
자동차	<ul style="list-style-type: none"> 자율주행 및 소프트웨어 탑재 차량(SDF: Software-Defined Vehicle)에 특화된 실시간 AI 자율주행 판단 기술개발 (사례) 차량 내 연료 부족 시, 주행거리 근방 저렴한 주유소를 팝업으로 자동 추천
IoT·가전	<ul style="list-style-type: none"> 실내상황 인지 및 사용자 사용패턴, 보안 기능을 최적화한 스마트 가전, AR/VR 기기용 온디바이스 AI 모듈 개발 (사례) 냉장고 내 유통기한이 지난 식품을 자동으로 판별하여 신호 발신
기계	<ul style="list-style-type: none"> 제조 현장 핵심 부품의 상태 모니터링, 고장 진단, 수명 예측 등 스마트공장용 온디바이스 AI 반도체 개발 및 실증 (사례) 센서를 이용해 장비 상태를 모니터링하고 고장 징후를 진단해 잔여 유효수명 예측
로봇	<ul style="list-style-type: none"> 상황 인지, 자율이동 등 기능을 지원하는 로봇용 온디바이스 AI 반도체·모듈 기술 개발 및 시가 내재된 로봇에 대한 실증기반 구축 (사례) 상황 인지, 자율이동, 상호작용 등을 통해 간병·돌봄 등 맞춤형 AI 서비스 구현
에너지	<ul style="list-style-type: none"> 실시간 전력 수요-공급 제어용 AI 기반 예측 기술 및 전력 수요 조절을 위한 온디바이스 AI 반도체 기술개발 및 실증 (사례) 사용자들의 전력사용 패턴 분석을 통해 전력량 절감을 위한 제어방법 추천
바이오·의료	<ul style="list-style-type: none"> 생체센서, 의료영상 등의 실시간 분석을 통한 맞춤형 의료 제품 개발, 온디바이스 AI 반도체를 활용한 자동 정밀 치료기기 개발 등 (사례) 환자의 각종 생체정보 현상을 분석하여 질병의 발생위험 예측
방산	<ul style="list-style-type: none"> 고성능, 고신뢰성의 초격차 반도체를 위해 레이더 및 전자전 시스템 등에서 요구하는 AI 반도체 기반 유무인 복합체계 마련 (사례) AI 기반 레이더 반도체를 통해 자폭 드론 등 무단침입 목표물 탐지 강화

출처: 과학기술정보통신부, 인공지능 3대 강국(AI G3) 도약을 위한 대한민국 AI 혁신의 청사진 제시, 2024.09.26.

136) 산업통상자원부, 7대 산업 인공지능(AI) 반도체를 게임체인저로 육성, 2024.8.14.:
(<https://www.motie.go.kr/kor/article/ATCL3f49a5a8c/169389/view>)

14) 과학기술정보통신부의 국가 AI전략 정책방향 발표¹³⁷⁾

과학기술정보통신부는 2024년 9월 26일 대통령 주재로 개최된 제1차 국가 AI 위원회에서 AI G3 도약을 위한 국가 AI혁신 비전을 담은 「국가 AI전략 정책방향」을 발표하고, AI G3 국가 도약을 통한 글로벌 AI 중추국가 실현이라는 비전하에 민·관이 함께 범국가적으로 추진해 나갈 핵심과제인 ‘4대 AI 플래그십 프로젝트’와 AI 생태계의 핵심인 4대 분야의 정책 추진방향을 제시하였다. 향후 ‘국가 AI 위원회’는 동 정책방향을 기반으로 정책과제로 구체화한 「국가 AI전략」을 수립하고 이행해 나갈 계획이다.

〈표 3-62〉 4대 AI 플래그십 프로젝트

구분	주요 내용
프로젝트 1	• ‘국가AI컴퓨팅 센터’ 구축 등 최신GPU 15배 확충, 국산 AI반도체 상용화 지원
프로젝트 2	• 4년간(’24~’27) 민간 총 65조원 규모 AI분야 투자, 정부는 투자 활성화 지원
프로젝트 3	• 국가전반 AI 대전환으로 ’30년까지 AI 도입률 산업 70% + 공공 95% 달성
프로젝트 4	• AI 안전·안보역량 조기확보로 글로벌 AI 거버넌스 주도

출처: 과학기술정보통신부, 인공지능 3대 강국(AI G3) 도약을 위한 대한민국 AI 혁신의 청사진 제시, 2024.09.26.

〈표 3-63〉 4대 분야 정책 추진 방향

구분	주요 내용
스타트업·인재 확충	<p>(국가 AI 혁신을 뒷받침할 기초체력 확충을 위해 AI 스타트업·인재 육성)</p> <ul style="list-style-type: none"> • 전문·특화분야 AI 스타트업 성장기반 조성 및 M&A 활성화 환경 조성으로 ’30년까지 AI 유니콘 기업 10개 육성 • AI인재 양성의 수월화·글로벌화·보편화를 통해 ’30년까지 20만 AI인재(’23, 5.1만명) 확보
기술·인프라 혁신	<p>(AI 핵심·원천기술 확충 및 AI 인프라 혁신 추진)</p> <ul style="list-style-type: none"> • AI-반도체 공동연구 협력거점 구축·운영 등을 통해 최고기술선도국과 기술협력 확대 • 미래 AI 신시장인 온디바이스 AI 경쟁력 확보 • 개인정보 규제혁신을 통해 AI 신산업 혁신 뒷받침 • 양질의 저작물 활용을 촉진하는 제도 개선으로 생성 AI에 적합한 데이터 활용기반 조성 • 저지연·저전력·Edge 네트워크 혁신 등 AI 네트워크 고도화 • 「국가기간전력망확충 특별법」 제정 등을 통해 AI 트래픽·전력 수요 증가에 대응

137) 과학기술정보통신부, 인공지능 3대 강국(AI G3) 도약을 위한 대한민국 AI 혁신의 청사진 제시, 2024.9.26.: (<https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=238&pageIndex=7&bbsSeqNo=94&nttSeqNo=3184952>)

구분	주요 내용
포용·공정기반 조성	(지속 가능한 AI발전·확산 기반 조성) <ul style="list-style-type: none"> • 전국민의 AI 접근·활용 보장을 위한 「디지털포용법」 제정 추진 • AI 혁신을 뒷받침할 공정경쟁 질서와 정당한 권리보호 기반 확립 • AI 개발·활용 과정에서의 부당한 차별 또는 프라이버시 침해 방지 방안 마련
글로벌 리더십 확보	(AI시대의 새로운 질서 정립 및 글로벌 AI 규범·거버넌스 주도) <ul style="list-style-type: none"> • AI 책임·권리 귀속관계 등 기본법리를 정립 • AI 자동화 결정시대의 개인·취약계층 보호 방안 마련 • 글로벌 AI 중추국가로서 선도국·국제기구와 AI 협력 지속 확대 • 글로벌 AI거버넌스 논의 주도

출처: 과학기술정보통신부, 인공지능 3대 강국(AI G3) 도약을 위한 대한민국 AI 혁신의 청사진 제시, 2024.09.26.

15) 개인정보보호위원회의 자율주행 AI 관련 기준 제시¹³⁸⁾

개인정보보호위원회는 2023년 9월 「개인정보보호법」에 신설된 이동형 영상정보처리기기 조항(제25조의2)의 적용 기준과 산업계의 문의사례 등을 반영한 「이동형 영상정보처리기기를 위한 개인영상정보 보호·활용 안내서」를 10월 14일 공개하였다. 동 안내서는 AI 개발 기업이 자율주행차량, 배달로봇 등 이동형 영상정보처리기기에서 촬영된 영상을 AI 개발 등에 활용할 수 있도록 하는 개인영상정보 보호 및 활용 기준을 제시한다.

<표 3-64> 개인영상정보 보호 및 활용을 위한 8대 기본원칙

원칙	주요 내용
① 비례성	• 개인영상정보의 처리 목적이 정당하고 수단이 적정한지, 예상되는 편익에 비하여 권리침해 위험이 과도한 지 등을 종합 고려해야 함
② 적법성	• 개인영상정보 처리(수집·이용 등) 근거는 적법하고 명확하여야 함
③ 투명성	• 개인영상정보 처리에 관한 사항을 투명하게 공개해야 함
④ 안전성	• 개인영상정보가 유출·훼손되지 않도록 안전하게 관리해야 함
⑤ 책임성	• 법령에서 정하는 책임과 의무를 준수하고, 신뢰를 확보하기 위해 노력해야 함
⑥ 목적 제한	• 개인영상정보 처리는 목적에 필요한 범위에서 최소화해야 함
⑦ 통제권 보장	• 정보주체가 자신의 개인영상정보에 대한 통제권을 행사할 수 있는 수단을 제공하고 이를 보장해야 함
⑧ 사생활 보호	• 사생활 침해를 최소화하는 방법으로 개인영상정보를 처리해야 함

출처: 개인정보보호위원회, 자율주행 인공지능(AI) 발전을 위한 개인영상정보 보호·활용 기준 나왔다, 2024.10.14.

138) 개인정보보호위원회, 자율주행 인공지능(AI) 발전을 위한 개인영상정보 보호·활용 기준 나왔다, 2024.10.14.; (<https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=10678#LINK>)

또한, 동 안내서는 개인영상정보 처리 단계별(▲기획 및 설계 단계, ▲촬영 및 수집 단계, ▲이용 및 제공 단계, ▲보관 및 파기 단계) 준수 및 권고사항, AI 학습 활용시 조치사항 및 사례, 개인영상정보의 안전한 보관 및 관리를 위한 구체적인 방안을 제시한다.

16) 산업통상자원부의 산업데이터 활용 활성화 방안 발표¹³⁹⁾

산업통상자원부는 2024년 10월, 지난 5월 발표한 「AI 시대의 신산업정책」의 6대 분야별 AI 활용 전략 중 자율제조와 디자인에 이은 세 번째 전략인 R&D 관련 「AI + R&D(기술혁신) 추진전략」과 「산업데이터 활용 활성화 방안」을 발표하였다.

<표 3-65> AI + R&D 추진전략 핵심내용

구분	주요 내용
기술개발	• 산업현장의 연구설계, 실험수행 등 주요한 활동에 AI를 적용해 개발기간·비용은 줄이고 성과는 극대화
개방혁신	• 글로벌 기술·인재를 AI로 탐색·연결해 혁신 속도 제고
연구행정	• 정부 R&D 전과정에 AI를 적용해 기업의 비연구활동 부담 최소화, 전문기관의 과제관리 전문성 강화

출처: 산업통상자원부, 인공지능(AI)과 데이터로 산업기술혁신 가속화, 2024.10.17.

<표 3-66> 산업데이터 활용 활성화 방안 핵심내용

구분	주요 내용
기업 데이터 역량 강화	• 기업이 보유한 원데이터를 AI가 학습할 수 있는 데이터로 가공·입력할 수 있는 SW 개발, 인력 양성, 컨설팅 지원
기업간 데이터 연계	• 산업 공급망데이터 플랫폼을 구축하여 글로벌 규제대응, 공급망 최적화 등 산업 전반의 AX(AI 전환) 촉진
제도적 기반	• 기업의 산업데이터 활용을 활성화하고, 기업간 협업의 저해요소를 제거하는 제도적 보완 추진

출처: 산업통상자원부, 인공지능(AI)과 데이터로 산업기술혁신 가속화, 2024.10.17.

139) 산업통상자원부, 인공지능(AI)과 데이터로 산업기술혁신 가속화, 2024.10.17.;

(<https://www.motie.go.kr/kor/article/ATCL3f49a5a8c/169646/view>)

17) 과학기술정보통신부의 AI 시대의 클라우드 전략 발표¹⁴⁰⁾

과학기술정보통신부는 2024년 10월 제17차 정보통신전략위원회를 개최하고 AI의 필수 기반인 인터넷 기반 자원 공유(클라우드) 컴퓨팅에 대한 향후 3개년 정책방향을 담은 「AI 시대의 클라우드 전략」이라는 제목의 제4차 클라우드컴퓨팅 기본계획(‘25~‘27)을 발표했다. 동 전략은 ▲인터넷 기반 자원 공유의 도입 전면화 ▲경쟁력 제고 ▲생태계 활성화를 3대 핵심 추진과제로 제시했다.

<표 3-67> 3대 핵심 추진과제

구분	주요 내용
도입 전면화	<ul style="list-style-type: none"> ① 교육·금융 분야 등 AI + 클라우드 전면화 ② 민간 클라우드 기반 정부시스템 대전환 ③ 공공 부문 클라우드 조달제도 혁신 ④ 공공 부문 기관평가 클라우드 도입 포함 ⑤ 공공 부문 클라우드 도입·활용 직접 유도
경쟁력 제고	<ul style="list-style-type: none"> ① 대규모 국가 AI 컴퓨팅 센터 구축 ② 국산 AI 반도체 기반 K-클라우드 가속 본격화 ③ AI 컴퓨팅 인프라(클라우드 기반) 지원 강화 ④ AI 서비스형 소프트웨어(SaaS) 성장 전주기 지원 ⑤클라우드 기반 민간 혁신사례 창출
생태계 활성화	<ul style="list-style-type: none"> ① AI·클라우드 투자 및 세액공제 확대 ② 민간 주도 클라우드 생태계 육성 ③ AI 데이터센터 산업 진흥 기반 확대 ④ AI 시대 클라우드 인력양성, 기술개발 ⑤ 경쟁·협력 기반 클라우드 생태계 조성

출처: 과학기술정보통신부, 인공지능 안전에 대한 국제 연대를 주도할 ‘인공지능 안전연구소’ 11월 출범 인공지능 시대를 여는 새로운 인터넷 기반 자원 공유(클라우드) 전략 제시, 2024.10.18.

18) 과학기술정보통신부의 AI안전연구소 설립¹⁴¹⁾¹⁴²⁾

정부는 「AI 서울정상회의(AI Summit Seoul 2024)」의 개최국으로서 서울선언(Seoul Declaration)의 규범 가치를 이행하기 위한 노력의 일환으로 국가 차원에서 AI의 안전성을 평가 및 연구하고 주요국 AI 안전연구소와의 협력을 전담하는 AI 안전연구소 설립을 추진해왔다. 이에 따라 과학기술정보통신부는 제17차 정보통신전략위원회를 통해 AI안전연구소의 설립 및 운영 계획을 공유하고 의견을 수렴하고, 2024년 11월 27일 공식적으로 AI 안전연구소를 설립하였다.

140) 과학기술정보통신부, 인공지능 안전에 대한 국제 연대를 주도할 ‘인공지능 안전연구소’ 11월 출범 인공지능 시대를 여는 새로운 인터넷 기반 자원 공유(클라우드) 전략 제시, 2024.10.18.:

(<https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=238&bbsSeqNo=94&nttSeqNo=3185018>)

141) 과학기술정보통신부, 인공지능 안전에 대한 국제 연대를 주도할 ‘인공지능 안전연구소’ 11월 출범 인공지능 시대를 여는 새로운 인터넷 기반 자원 공유(클라우드) 전략 제시, 2024.10.18.:

(<https://www.msit.go.kr/bbs/view.do?sCode=user&mId=113&mPid=238&bbsSeqNo=94&nttSeqNo=3185018>)

142) 대한민국 정책브리핑, ‘한국 AI안전연구소’ 출범…AI 위협에 체계적·전문적 대응, 2024.11.27

(<https://www2.korea.kr/news/policyNewsView.do?newsId=148936794>)

AI안전연구소는 아시아 태평양 지역을 대표하는 글로벌 AI 안전 거점 연구소 구현을 목표로 ▲AI 안전에 대한 과학적 이해 증진 ▲AI 안전정책 고도화 및 안전제도 확립 ▲국내 AI 기업의 안전 확보 지원의 3대 핵심임무를 추진할 예정이며, 이를 완수하기 위해 4가지 핵심 기능을 수행할 방침이다.

〈표 3-68〉 4대 핵심 기능

구분	주요 내용
AI 위험 정의 및 안전 평가	<ul style="list-style-type: none"> ▲(위험 정의) 국가 차원에서 집중적으로 관리해야 할 AI 위험을 세부적으로 정의 ▲(안전 평가) 기업, 대학연구기관과 협력하여 위험별 AI 안전 평가 프레임워크 개발, 안전 평가 및 위험 완화 방안 마련 ▲(평가 인프라 구축) AI 안전 평가 데이터셋 구축, 평가도구 개발 등 평가 인프라를 구축하여 안전 평가 시 활용 및 기업 활용 지원
AI 안전 정책 연구	<ul style="list-style-type: none"> ▲(안전 정책 연구) 주요국 및 국제기구의 AI 안전 규범·거버넌스 분석, AI윤리·리터러시 등 인문·사회학적 연구를 통한 안전 정책 수립 지원 ▲(국가안보 위험 대응) 국가안보 측면에서 국내·외 AI 위험 정보 수집 및 분석, 시로 인한 국가 인프라 취약점 분석 및 대응 방안 마련 ▲(안전 컨설팅 지원) 국제평가·인증, 해외규제 등 AI 안전 관련 기업 컨설팅 지원 ▲(AI 기술 영향평가) AI 기술이 경제·사회·일자리·환경 등에 미치는 영향을 사전에 평가하고, 분야별 AI 정책 수립 시 활용 지원 ▲(안전 정보 구축) AI 모델·시스템의 역량·위험, AI 안전 사고, AI 위험 식별·평가·완화 모범사례 등에 관한 정보 구축 및 공유
AI 안전 대내·외 협력	<ul style="list-style-type: none"> ▲(국내 안전 허브 역할 수행) AI 안전 기술·정책 정보공유 및 국내기업·대학연구기관 간 인력, 기술·인프라 협력 플랫폼 역할 수행 ▲(국제적 연대·협력) 주요국 AI 안전연구소, 국제기구 등과 긴밀한 협력체계 마련 및 AI 안전 확보를 위한 글로벌 거버넌스 구축
AI 안전 기술 연구·개발	<ul style="list-style-type: none"> ▲(실존위험 대응) 고도로 발전된 AI의 통제력 상실에 대응하는 정렬기술(Alignment), 사회적 차별 및 편향 완화·제거 기술 개발 ▲(잠재 위험 대응) 미지의 위험(Unknown Risk) 발굴 기법 연구, 범용 AI(AGI) 등 미래 AI에 대한 위험 예측(시나리오 분석 등)에 관한 선제적 대응연구 수행

출처: 과학기술정보통신부, 인공지능 안전에 대한 국제 연대를 주도할 ‘인공지능 안전연구소’ 11월 출범 인공지능 시대를 여는 새로운 인터넷 기반 자원 공유(클라우드) 전략 제시, 2024.10.18.

AI안전연구소는 산·학·연 기관이 AI 안전분야 연구 협력과 정보 교류의 허브이자 국제 AI안전연구소 네트워크(10개국 참여)의 일원으로 안전한 AI를 위한 글로벌 연대의 책임 있는 역할을 할 계획이다.

‘AI안전연구소’와 컨소시엄 참여기관(분야 국내 대표 산·학·연 기관 24곳)은 함께 AI 안전 프레임워크(위험 식별·평가·완화) 연구·개발·실증, AI 안전 관련 국제규범 대응 등 정책 연구, AI 안전 관련 기술협력 등을 중점적으로 추진할 계획이며, 향후 컨소시엄의 세부 연구 주제와 운영 방안을 구체화할 계획이다.

19) 국회의 AI 기본법 의결¹⁴³⁾¹⁴⁴⁾

2024년 12월 26일, 국회 본 회의에서 ‘인공지능 발전과 신뢰 기반 조성 등에 관한 기본법(이하 AI기본법) 제정안이 의결되었다. AI기본법은 2020년 7월 국회에서 법안을 처음 발의한 이후 4년 넘게 다양한 의견을 수렴해 논의한 법안으로, 이번 22대 국회는 여야 합의로 그 동안 논의되어 왔던 19개 관련 법안을 병합했다.

이번 법안은 인공지능에 관한 국가 차원의 거버넌스 체계를 정립하고, 인공지능 산업을 체계적으로 육성하며, 인공지능으로 인해 발생할 수 있는 위험을 사전에 예방하기 위한 안전성과 신뢰성을 확보하는 내용으로 구성돼 있다.

국가 차원의 거버넌스 확립과 관련해서는 인공지능 유관 기관·단체 등에 대해 법적 근거를 마련했다는 점을 들 수 있다. AI기본법은 제7조에서 2024년 9월 출범한, 국가인공지능위원회의 설립 근거를 제시했다. 위원회는 대통령을 위원장으로 하며, 민간위원이 위원회 구성원의 과반이 되도록 하여 민간 주도의 인공지능 정책을 수립하도록 하였다. 또한 인공지능정책센터(제11조), 인공지능안전연구소(제12조), 한국인공지능진흥협회(제26조) 등 설립에 대해서도 법적 근거를 제공하였다.

인공지능 산업 진흥과 관련해서는 과학기술정보통신부 장관이 3년마다 인공지능기술 및 인공지능산업의 진흥과 국가경쟁력 강화를 위한 기본계획을 수립하도록 하였다(제6조). 또한 인공지능 기술 개발·이용 활성화(제13조) 및 기술 표준화(제14조) 관련 사업에 대한 정부지원을 통하여 산업 기반을 조성하고, 국가기관 및 기업의 인공지능 도입(제16조), 중소기업 및 창업 지원(제17조 내지 제18조), 전문인력 양성(제21조), 해외시장 진출 지원(제22조) 등을 통하여 인공지능기술 개발 및 산업 활성화를 촉진하도록 하였습니다.

이밖에 인공지능 윤리원칙(제27조), 신뢰 기반 조성 시책 마련(제29조), 사업자 의무조항(제31조 등)을 두어 인공지능의 윤리와 신뢰성 확보를 위한 기반도 마련하였다.

143) 정부24, 인공지능 산업 육성·규제 근거 마련…‘AI기본법’ 국회 통과, 2024.12.27.:
(<https://www.gov.kr/portal/ntnadmNews/4124183>)

144) 법안 원문, 의안정보시스템, (https://likms.assembly.go.kr/bill/billDetail.do?billId=PRC_R2V4H1W1T2K5M1O6E4Q9T0V7Q9S0U0)

제4장 국제기구 인공지능 동향

제1절 OECD

1. 그간의 주요 동향

경제협력개발기구(OECD)는 2019년 5월, ‘AI에 관한 이사회의 권고사항(Recommendation of the Council on Artificial Intelligence)’¹⁴⁵⁾에서 ‘AI 원칙’ (10개 원칙)을 마련하였다. OECD AI 원칙은 혁신적이고 신뢰할 수 있으며 인권과 민주적 가치를 존중하는 AI의 사용을 장려한다. OECD 원칙은 가치 기반 원칙과 정책 입안자를 위한 권고사항을 제시했다. OECD의 가치 기반 원칙으로는 1) 포용적 성장, 지속 가능한 개발 및 웰빙, 2) 인간 중심의 가치와 공정성, 3) 투명성 및 설명 가능성, 3) 투명성과 설명 가능성, 4) 견고성, 보안성 및 안전성, 5) 책임성을 제시했다. 정책 입안자를 위한 원칙으로는 6) AI R&D 투자, 7) AI를 위한 디지털 생태계 육성, 8) AI 활성화 정책 환경 제공, 9) 인적 역량 강화 및 노동시장 전환 준비, 10) 신뢰할 수 있는 AI를 위한 국제협력을 제시했다.¹⁴⁶⁾

OECD는 2020년, ‘OECD AI 정책관측소(OECD AI Policy Observatory: 이하 OECD.AI)를 창설하여 OECD 회원국 정부와 전 세계 국가들의 OECD AI 원칙을 실행하도록 지원하고 관련 정보(전략, 통계 등)를 지속 제공하고 있다.

2022년 2월 OECD.AI 전문가 네트워크(OECD.AI Network of Experts)는 AI 시스템의 기회와 위협을 평가하고 국가 AI 전략 수립에 활용할 수 있는 AI 시스템 분류체계(OECD Framework for Classifying AI Systems)를 발표했다. OECD의 AI 시스템 분류체계는 사람 및 지구(People & Planet), 경제 상황(Economic Context), 데이터 및 입력내용(Data & Input), AI 모델(AI Mode), 과업·산출내용(Task & Output)의 영역에 따라 AI 시스템 및 애플리케이션을 분류한다.

145) OECD, Recommendation of the Council on Artificial Intelligence, 2022. 5

146) OECD AI Principles overview
(<https://oecd.ai/en/ai-principles>)

2023년에 OECD.AI 팀은 OECD 회원국과 다른 국가들이 책임감 있고 유익한 방식으로 인공지능을 형성하려는 노력을 더욱 강화하도록 도왔다. OECD 국가들은 AI 법률과 규정이 견고한 기반을 구축하고 관할권 간 상호운용성을 촉진하기 위한 정의가 필요했다. OECD는 AI 시스템 정의를 일부 수정하고, 생성 AI, 대규모 언어모델, 규제 샌드박스, OECD AI 원칙에 관한 사항에 집중하였다. 그리고 2023년 11월에는 2019년 이후의 이행 상태를 담은 'OECD AI 원칙 시행 4년차 현황 보고서(The state of implementation of the OECD AI Principles four years on)'를 발행하였다.

2023년에 ChatGPT가 중심이 되면서, OECD는 생성 AI에 대한 작업에 전념하는 포털(oecd.ai/en/genai)을 개시하였다. OECD는 또한 G7 AI 지침 원칙과 행동 강령을 개발하기 위한 히로시마 프로세스를 지원했고, 생성 AI와 관련된 기회와 과제에 대한 G7 보고서를 발행하였다.

2023년에 OECD AI 정책관측소는 AI 커뮤니티에 AI를 더욱 신뢰할 수 있게 만들기 위한 도구, 지표 및 사고 모니터를 제공하는 데 중점을 두고 새로운 기능들을 추가하였다. 2023년 3월에 신뢰할 수 있는 AI 시스템을 구축하고 배포하는 데 도움이 되는 신뢰할 수 있는 도구 및 지표에 대한 카탈로그를 제공하여, 누구나 카탈로그에 도구나 도구 사례를 제출할 수 있도록 하였다. 그리고 11월에는 파리 평화 포럼에서 출시된 신뢰할 수 있는 AI를 위한 정책을 설계하는 데 필요한 사실과 데이터를 정책 입안자에게 제공하는 중요한 첫 단계인 AI 사고 모니터(AIM: The AI Incidents Monitor)를 출시했다. AIM은 AI 정책 입안자의 도구 상자에서 중요한 공간을 채우고 AI와 관련된 사고 및 잠재적 위험을 보고, 문서화 및 조사하기 위한 리소스이다. 전 세계적으로 150,000개가 넘는 뉴스 소스를 실시간으로 스캔하는 미디어 모니터링 플랫폼을 사용하여 AIM은 매일 100만 개 이상의 뉴스 기사를 수집하여 사고를 식별한다.

OECD가 AI 원칙을 마련한 이후의 주요 정책의 흐름을 정리하면 아래 <표4-1>과 같다.

〈표 4-1〉 OECD의 연도별 AI 정책 추진 경과

구분	주요 내용
2019	<ul style="list-style-type: none"> • 2019년 5월: 'OECD AI 원칙' 채택 • 2019년 6월: OECD AI 원칙에서 'G20 AI 원칙' 마련 • 2019년 6월: '사회 속의 AI(Artificial Intelligence in Society)' 도서 출간 • 2019: 'AI에 관한 OECD 글로벌 의회 네트워크(GPN AI)' 창설
2020	<ul style="list-style-type: none"> • 2020: 독일이 주도하고 지원하는 'OECD 노동, 혁신, 생산성 및 기술 프로그램(AI-WIPS)' 시작 • 2020: 'OECD AI 전문가 네트워크(ONE AI)' 및 'OECD AI 정책 관측소(OECD.AI)' 출범 • 2020년 6월: OECD에서 주최된 GPAI(Global Partnership on Artificial Intelligence) 출범
2021	<ul style="list-style-type: none"> • 2021년 6월: 'OECD AI 원칙 이행 현황 : 국가 AI 정책의 통찰력' 마련 • 2021년 6월: 신뢰할 수 있는 AI 시스템을 위한 구현 도구를 비교하는 프레임워크인 '신뢰할 수 있는 AI 도구(Tools for Trustworthy AI)' 마련 • 2021년 10월: OECD 각료이사회 회의에서 OECD AI 원칙 실행에 대한 진행 상황 보고
2022	<ul style="list-style-type: none"> • 2022년 2월: 업무, 혁신, 생산성 및 기술 분야의 AI에 관한 제2차 연례 국제 컨퍼런스 • 2022년 2월: OECD 프레임워크 또는 AI 시스템 분류 마련 • 2022년 3월: OECD에서 최초의 공식 'AI 거버넌스 작업반(AIGO)' 창설
2023	<ul style="list-style-type: none"> • 2023년 10월: OECD AI 원칙 시행 4년차 현황 보고서 발행 • 2023년 11월: AI 사고 모니터(AIM: The AI Incidents Monitor)를 출시

출처: OECD.AI, About OECD.AI - Background, 2023.11.4.(확인)¹⁴⁷⁾

2. 2024년 주요 동향

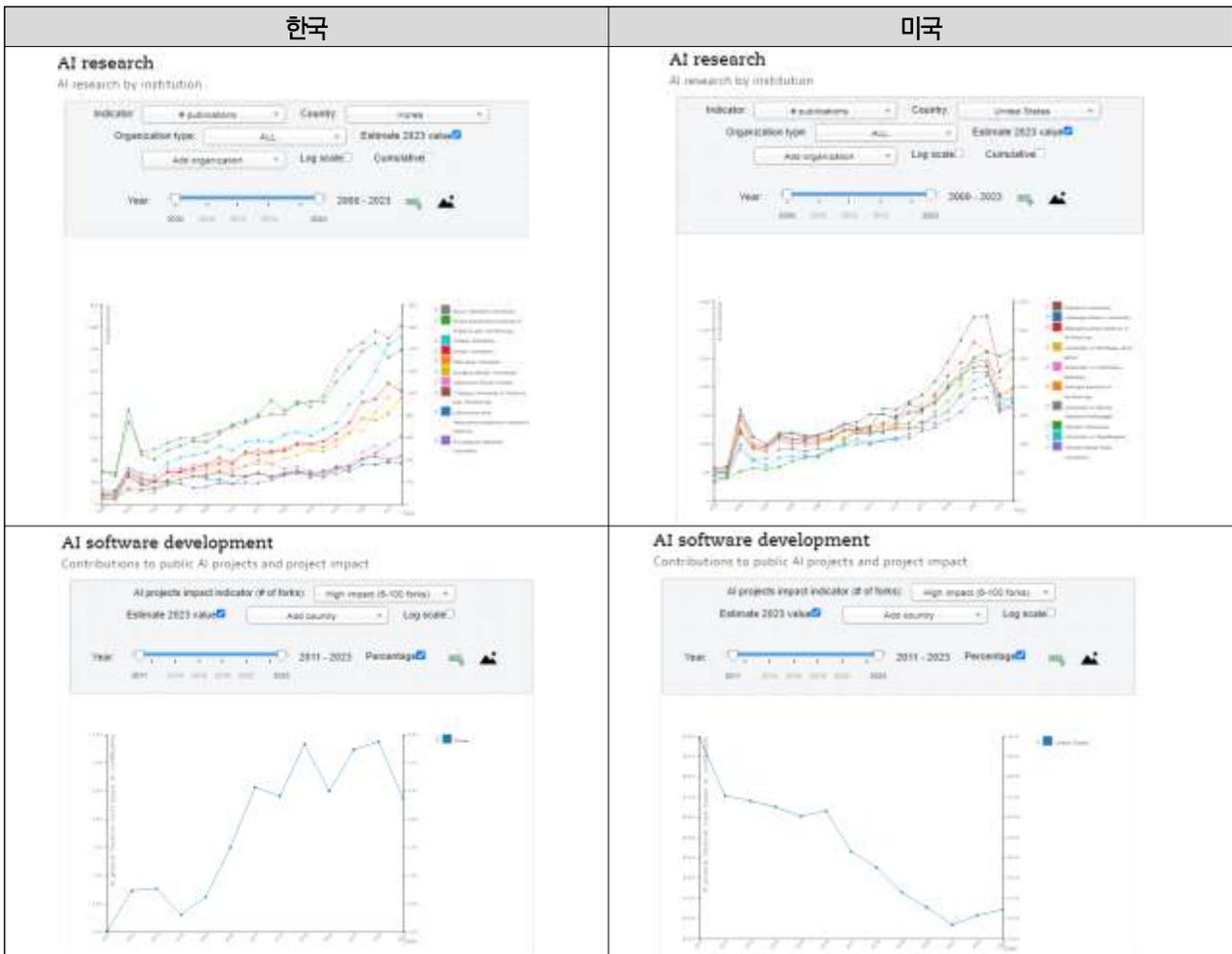
1) OECD.AI 주요 동향

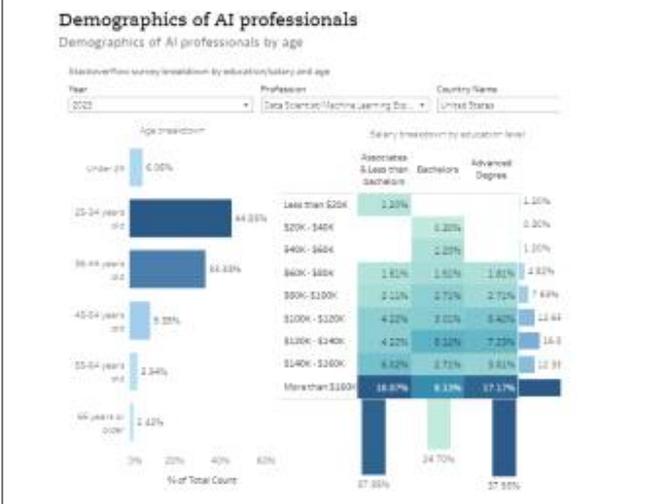
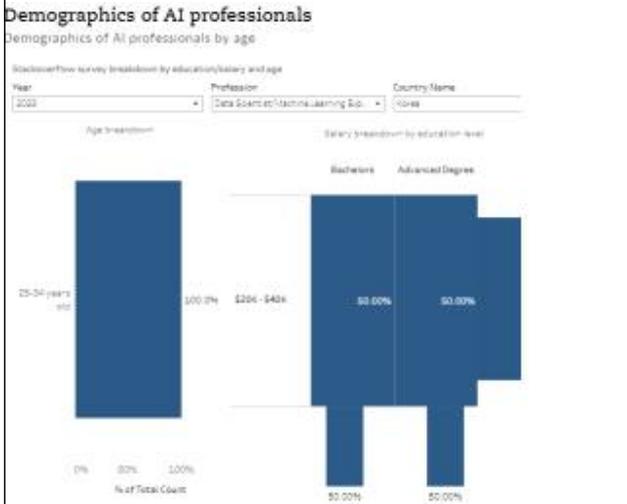
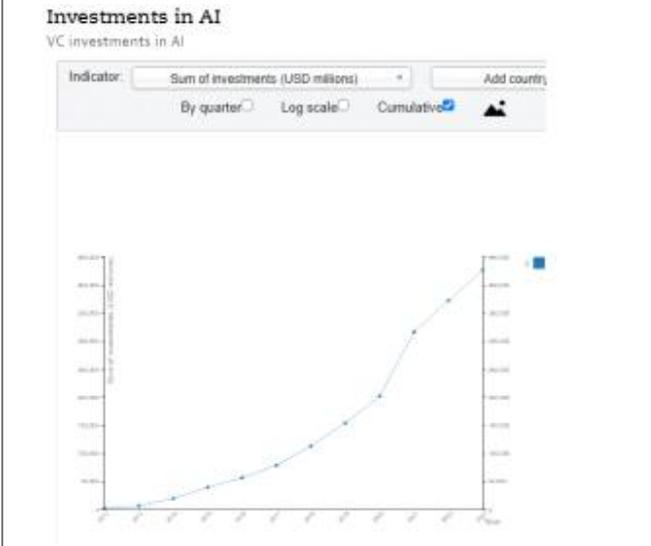
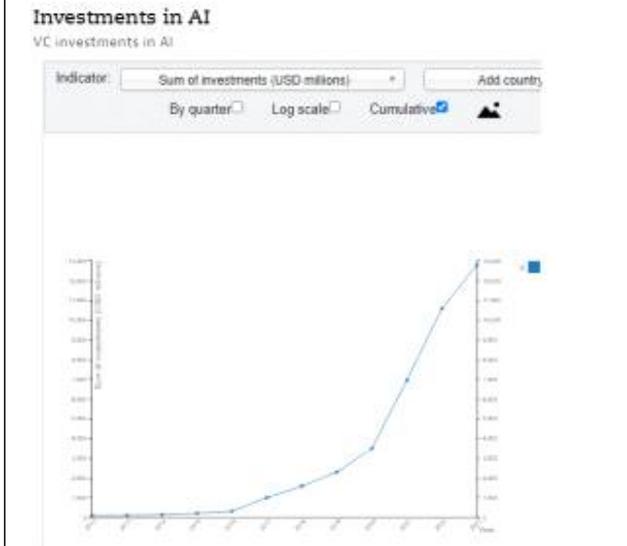
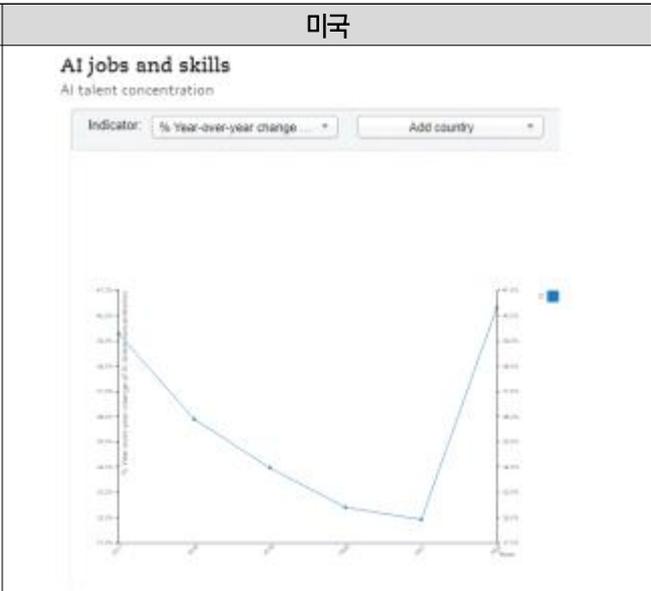
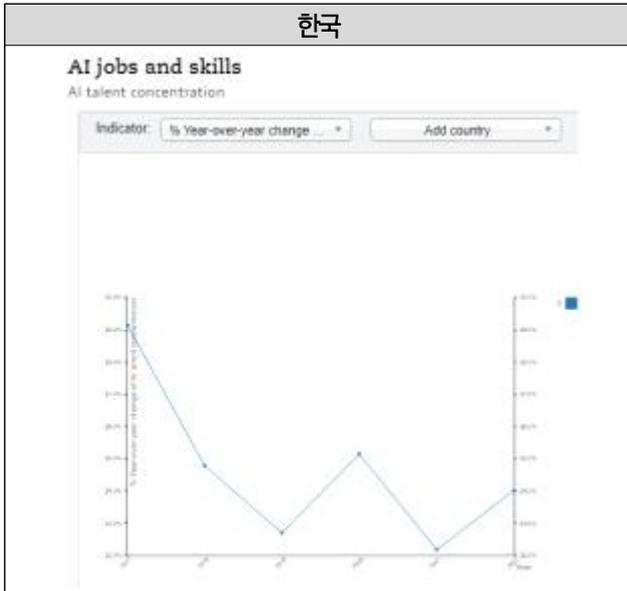
OECD.AI는 전 세계 AI 정책의 글로벌 허브의 역할을 수행한다. 선진국과 개발도상국의 모든 행위자와 이해관계자 그룹은 OECD.AI의 도구, 데이터 및 기타 AI 정책 리소스를 자유롭게 접근할 수 있다. OECD.AI 작업 프로그램은 AI 위험과 책임, 잠재적인 AI 미래 연구, AI 사고 정의 및 추적, AI 컴퓨팅 용량과 AI가 기후 변화에 미치는 영향에 초점을 맞추어 추진되고 있다.

147) About OECD.AI
(<https://oecd.ai/en/about/background>)

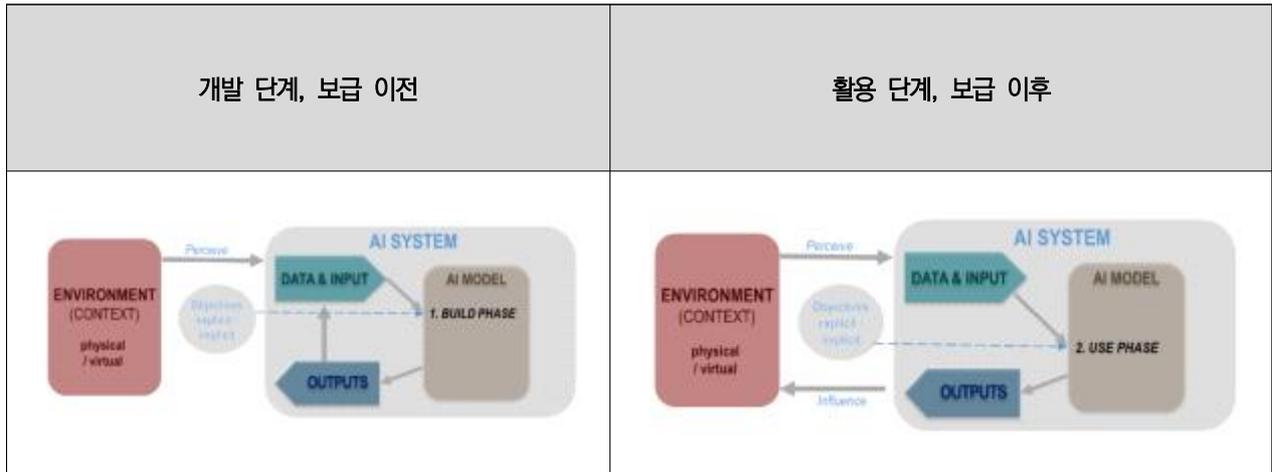
2024년 10월 기준 OECD.AI은 전 세계의 71개국이 국가 AI 정책 및 전략 현황에 대한 정보를 제공하고 있다. OECD.AI는 전 세계 71개국의 국가 AI 정책 및 전략 현황을 가시적으로 표현하고 각각의 국가에 대한 세부 정보로 AI 관련 연구, 소프트웨어 개발, 일자리와 스킬, 투자, AI 전문인력의 분포, 검색 트렌드 등을 제공한다(아래<표>는 한국과 미국의 각각의 항목에 대한 비교 정보를 예시로 보여줌).

<표 4-2> OECD.AI의 한국과 미국의 주요 트렌드 비교





〈표 4-3〉 AI 시스템 개관



자료: OECD, Explanatory memorandum on the updated OECD definition of an AI system, 2024.03.05.

AI 시스템의 목표는 암시적이거나 명시적일 수 있으며, 하나 이상의 목표 유형을 가질 수 있다. 가령 개발자는 목표를 시스템에 직접 인코딩해 명시적으로 지시하거나 ‘빨간 신호등에 정지’ 하는 운전 규칙처럼 상황에 따라 AI 시스템이 취해야 할 조치를 지시하는 규칙에 목표를 내재할 수 있다.

AI 시스템에 대한 입력은 개발 중과 배포 후에 모두 사용되며, 개발 중 입력은 학습 데이터나 인간의 입력을 이용한 기계학습 등을 통해 AI 시스템 구축에 쓰인다.

AI 시스템은 기계나 인간의 입력 및 데이터를 기반으로 수동 또는 자동으로 개발된 하나 이상의 모델을 결합해 구축된다. AI 모델은 입력에서 추론을 수행하여 출력을 생성하는 데 사용되는 AI 시스템의 핵심 구성요소이며, AI 시스템은 구축 이후에도 머신러닝을 통해 새로운 입력 및 데이터와 상호작용하여 성능 향상이 가능하다.

AI 시스템에서 생성된 출력은 AI 시스템이 수행하는 다양한 기능을 반영하며, 개정된 정의는 인간의 개입 수준에 대응되는 권고, 예측, 결정의 범주 외에 콘텐츠를 새롭게 포함한다.

업데이트된 AI 정의는 포괄적이며 간단한 것부터 복잡한 것까지 다양한 시스템을 포괄한다. AI는 다양한 상황에 적용할 수 있는 기술과 기법의 집합을 나타낸다. 머신러닝과 같은

특정 기법은 편향, 투명성, 설명 가능성과 같은 정책 입안자에게 특정 고려 사항을 제기할 수 있으며, 일부 사용 맥락(예: 공공 혜택에 대한 결정)은 다른 맥락보다 더 중요한 우려를 제기할 수 있다. 그러한 이유로 OECD는 AI 정의를 실무에 적용할 때 특정 맥락에서 사용될 때 정의를 좁히거나, 그렇지 않으면 맞춤화하기 위해 추가 기준이 필요할 수 있다고 설명했다.

(2) AI, 데이터 거버넌스, 개인정보보호의 협력에 관한 정책 문서 발간

OECD는 2024년 6월 AI 정책집단(Policy Community)과 개인정보보호 정책집단 간 협력을 증진하기 위한 정책 문서(AI, data governance and privacy: Synergies and areas of international co-operation)¹⁴⁹⁾를 발간하였다.

생성 AI가 빠르게 발전하며 데이터 거버넌스 및 개인정보보호와 관련해 여러 문제가 나타나고 있으나, AI 및 개인정보보호 정책집단은 종종 상호협력 없이 문제를 개별적으로 다루며 접근 방식도 지역과 법제에 따라 파편화되고 있다. AI 및 개인정보보호 정책집단은 협력을 통해 각자의 지식과 경험, 우선 과제를 공유하고 정책 대응을 조정하며 용어 사용을 포함한 AI 정책과 개인정보보호 정책 간 일관성을 개선하고 보완할 수 있다.

이 정책 문서는 개인정보보호와 데이터 보호 고려 사항을 OECD의 2019년 AI 원칙(①포용적 성장과 지속 가능 개발 ②인간중심 가치와 공정성 ③투명성과 설명 가능성 ④견고성과 안전성 ⑤책임성)과 매칭하여 AI와 개인정보보호 정책집단 간 협력이 가능한 영역을 식별하였다. AI 원칙에 규정된 AI 시스템의 투명성과 설명 가능성은 개인정보보호 집단에서 중시하는 투명성과 밀접하게 연관되며, 사용자 동의를 받아 투명하게 개인정보를 처리하기 위한 전제조건이기도 하다. AI 원칙의 공정성은 AI 시스템의 개인정보 처리 시 개인정보 수집과 사용 제한, 데이터 품질 고려, 목적 지정과 같은 방식으로 달성될 수 있다. 책임성은 AI 시스템의 개인정보보호법 준수를 보장하기 위한 AI 원칙으로서, 개인정보보호 관리 프로그램의 확립을 통해 AI 시스템에 의한 잠재적 차별을 식별 및 예방하고 완화할 수 있다.

전 세계를 서비스 범위로 하는 AI의 특성상, 개인정보보호 문제를 효과적으로 해결하려면

149) OECD, AI, data governance and privacy: Synergies and areas of international co-operation, 2024.06.26.

국제협력이 필수적으로, 이를 통해 규제 과편화를 방지하고 개인정보보호 원칙의 일관된 적용이 가능하다. OECD가 운영하는 AI와 데이터·개인정보보호 관련 공동 전문가 그룹은 개인정보보호 규제를 존중하는 AI의 혁신적 발전 기회를 만드는 데 중요한 역할을 할 전망이다. OECD는 또한 AI 시스템의 개인정보보호와 데이터 보호를 강화할 수 있는 협력 분야로 개인정보보호 강화 기술(PET)을 제안하면서, 이러한 기술이 AI 알고리즘의 설명 가능성을 개선해 개인정보 침해 우려를 해소하는 데 도움이 될 것으로 예상하였다.

(3) 금융 분야의 생성 AI 보고서 발간

금융 분야의 생성 AI 도입 확산 시 편견과 차별, 데이터 침해 등 다양한 위험이 발생할 수 있다. OECD는 2023년 12월 ‘금융 분야의 생성 AI(Generative Artificial Intelligence In Finance)’¹⁵⁰⁾ 보고서를 발간하고 금융 분야의 생성 AI 도입 현황을 개괄하였다.

사용자 친화적이고 직관적 인터페이스를 가진 생성 AI 등장 이후에도 금융 분야에서 AI는 금융사의 운영 효율성을 높이기 위한 업무 자동화를 중심으로 활용되며, 인간의 개입 없는 완전 자동화와 같은 고급 활용 사례는 개발 단계에 머물러 있다. 금융 분야의 생성 AI 도입 속도가 느린 이유는 소비자 보호와 금융 안정성을 보장하기 위한 엄격한 규제의 적용을 받기 때문으로, 향후 생성 AI 기술이 발전하면서 고객 지원과 신제품 개발, 제품 기능 최적화 등 다양한 영역에서 활용될 전망이다.

금융 분야에서 생성 AI 도입이 늘어날 경우, 차별이나 차별, 개인정보와 지식재산권 침해 등 기존에 존재하던 위험이 증폭되는 한편 새로운 위험도 등장할 수 있다. AI 모델은 품질이 낮거나 부적절한 데이터셋을 학습하여 금융 대출 시 차별적이거나 불공정한 결과를 초래할 수 있으며, 개인정보 침해나 지식재산권 침해의 가능성도 존재한다. 또한 AI 모델이 결과를 생성하는 이유와 방법에 대한 설명 가능성의 부족으로 금융서비스의 신뢰 수준이 저하될 수 있으며, 악의적 행위자가 생성 AI를 악용해 금융 시장을 조작할 수도 있다.

OECD는 보고서에서 금융 분야에서 생성 AI로 인한 위험 완화를 위한 정책적 대응의 필요성을 언급하며, 정책 입안자가 혁신을 저해하지 않으면서 책임 있고 안전한 AI 도입을 위해

150) OECD, Generative artificial intelligence in finance, 2023.12.15.

금융 분야에서 생성 AI로 인한 위험을 파악하고 완화할 수 있는 정책들을 고려해야 한다고 제시하였다. 정책 입안자는 편견과 차별의 위험을 막기 위한 보호조치를 마련하는 한편, 생성 AI 모델의 설명 가능성과 투명성을 향상하려는 노력을 고취하고, AI 모델에 대한 거버넌스를 강화하여 모델 개발과 배포와 관련된 당사자에게 책임을 지도록 요구해야 한다. 또한 정책 입안자는 대출과 같이 중요한 사례의 의사결정에서 인간 중심의 접근방식을 장려하고, 고객에게 인간의 개입을 요청하거나 모델의 결과에 이의를 제기하고 시정을 요구할 권리를 부여하는 방안을 고려해야 한다.

제2절 UN

1. 그간의 주요 동향

유엔은 2020년 6월 디지털 협력 로드맵(Secretary-General's Roadmap for Digital Cooperation)을 제시하여 AI 및 기타 디지털 기술과 관련된 다양한 문제와 국제사회의 협력 방법을 제시했다. 신뢰할 수 있고 인권 기반이며 안전하고 지속가능하며 평화를 증진하는 방식으로 AI 개발 및 사용을 위한 글로벌 역량을 구축하기 위해서 AI 분야에서 다중이해관계자의 국제협력 강화가 필요하다. 안토니우 구테흐스 UN 사무총장은 디지털 협력 로드맵에서 포용성, 조정 및 역량 강화와 관련된 문제를 해결하기 위해 글로벌 AI 협력에 대한 다중이해관계자 자문기구 설립을 제안하였다(2023년 10월 자문기구가 출범).

UN은 2023년 10월 AI의 국제 거버넌스 문제를 해결하기 위해 39명으로 구성된 고위 다중이해관계자 자문조직(High-Level Multistakeholder Advisory Body on Artificial)151)를 창설했다. 이 조직은 유엔 헌장에 명시된 보편적 가치에 기반을 두며, 진행 중인 다양한 AI 거버넌스 이니셔티브를 어떻게 연결할 수 있는지 고려한다. 이 자문조직은 인공지능의 국제적 거버넌스, 위험과 과제에 대한 공유된 이해, 그리고 AI를 활용하여 SDG 달성을 가속화하기 위한 주요 기회와 지원책에 대한 자문을 제공하는 역할을 수행한다. 이 자문조직은 2023년 12월, 인류를 위한 AI 거버넌스(Governing AI for Humanity)152) 중간보고서를 발표했다. 이 중간 보고서는 국제규범과 AI가 개발되고 출시되는 방식 간의 긴밀한 조정을 요구한다. 또한 이 중간보고서는 위험에 대한 지평선 스캐닝, 데이터, 컴퓨팅 용량 및 인재에 대한 국제협력 지원과 같은 7가지 중요한 기능을 수행하여 AI의 국제적 거버넌스를 강화하여 지속가능한 개발 목표(SDG)를 달성할 것으로 제안한다.

유엔의 교육과학문화 기구인 유네스코(UNESCO)는 2021년 AI에 대한 정부 간 협력을 통해 온라인 플랫폼인 GlobalPolicy.AI를 구축하여 운영하고 있다. Globalpolicy.AI는 정책 입안자와 대중이 국제적인 AI 거버넌스 환경을 살펴보고, AI 정책 개발에 대한 정보에 지식을 획득할

151) UN, High-Level Advisory Body on Artificial Intelligence
(<https://www.un.org/techenvoy/ai-advisory-body>)

152) UN, UN AI Advisory Body calls for grounding artificial intelligence in universal principles, suggests tasks for a potential institution on AI governance
(https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_interim_report_press_release.pdf)

수 있도록 필요한 지식, 도구, 데이터 및 모범사례에 접근할 수 있도록 지원한다. 유네스코는 2023년 6월 전 세계를 휩쓴 챗GPT와 같은 생성 AI 분석을 담은 정책 문서(Foundation Models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence)¹⁵³⁾를 발행했다. 이 정책 문서는 윤리적 원칙에 기초를 두고 권고사항의 원칙 및 정책 영역과 관련하여 AI 기반모델의 알려진 효과를 강조하고, 사람과 조직이 이러한 기술을 설계하고 사용할 준비가 되어 있는지 확인하고 사전 및 사후에 미치는 영향을 평가하기 위해 신속한 조치의 필요성을 강조했다.

2. 2024년 주요 동향

1) UN의 안전한 AI 사용을 위한 결의안 채택

2024년 3월 21일, 유엔총회에서 미국이 제안하고 중국을 포함한 120개 이상의 국가가 공동 후원하였으며, 안전하고 신뢰할 수 있는 AI 시스템의 개발과 사용을 촉구하는 AI 결의안(Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development)¹⁵⁴⁾이 공식 투표 없이 컨센서스(전원합의)¹⁵⁵⁾로 채택되었다. 이번 UN의 AI 결의안은 AI의 설계·개발·배포·사용에서 인권의 존중과 보호를 강조하고, 모든 회원국과 이해관계자에 국제 인권법에 따를 수 없거나 인권에 과도한 위협을 초래하는 AI 시스템의 사용을 자제하거나 중단할 것을 촉구한다.

유엔총회가 신기술과 관련된 결의안을 채택한 것은 이번이 처음으로, 결의안을 주도한 미국은 이번 결과가 AI의 안전한 사용을 위한 역사적 진전이라고 평가된다. 이번 결의안은 특히 국가 간 및 국가 내 AI와 기타 디지털 격차를 해소하고, 개발도상국이 AI에 공평하게 접근할 수 있도록 협력과 지원을 촉구한다. 우선 안전하고 신뢰할 수 있는 AI의 이점을 활용할 수 있도록 개발도상국의 참여를 확대하고 기술 이전과 지식 공유를 추진한다. 또한 개발도상국, 특히 최빈개도국과 연구개발, 저렴한 기술 공급, 과학적 자원 제공 등의 파트너십을 통해 AI와 기타 신기술 혁신을 저해하는 장애물을 없애는 것은 추진한다.

153) UNESCO, Foundation Models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence, 2023.6.

154) United Nations(General Assembly), Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development, 2024.3.11

155) 컨센서스는 반대 의사를 표명한 회원국이 없으면 합의된 것으로 간주하는 의사결정 방식

이번 UN의 AI 결의안은 안전하고 신뢰할 수 있는 AI 시스템을 증진하고 AI를 활용해 경제, 사회, 환경의 3개 차원에서 지속가능한 발전을 달성하기 위해 다음의 조치를 요구한다.

- 각국의 정책과 우선 과제에 따라 안전하고 신뢰할 수 있는 AI 시스템을 장려하는 규제와 거버넌스 접근방식 및 프레임워크의 개발과 이행
- AI의 설계·개발·배포·사용에 앞서 취약점과 위험을 확인·분류·평가·예방·완화할 수 있는 효과적이고 상호운용성 있는 조치를 마련
- 최종 사용자와 제3자가 기술 취약점을 신고할 수 있는 피드백 절차를 마련하고, AI 시스템 수명주기 전반에 걸친 위험 관리감독 체계와 데이터 보안 체계, 영향평가 절차를 수립
- 워터마크나 라벨 등 신뢰할 수 있는 콘텐츠 인증 및 출처확인 방식을 통해 사용자가 AI로 생성된 콘텐츠를 구별할 수 있도록 보장하고, 혁신을 촉진하면서 지식재산권을 보장하는 보호장치를 마련

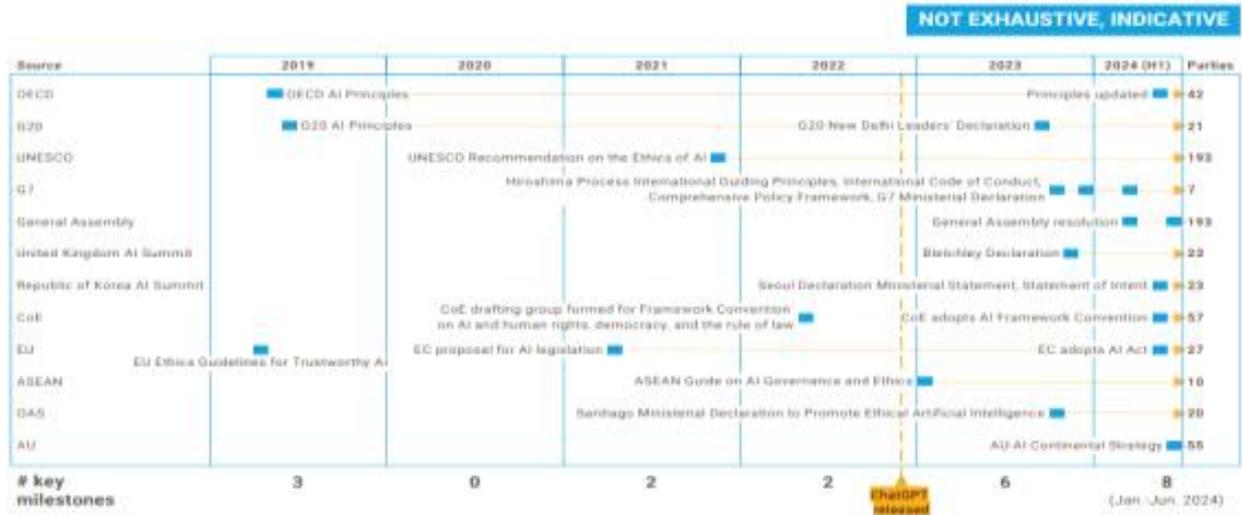
2) 인류를 위한 AI 거버넌스 발표

UN의 AI 자문기구가 2024년 9월 AI의 위험을 해결하고 혁신적 잠재력을 실현하기 위한 권고안을 담은 ‘인류를 위한 AI 거버넌스(Governing AI for Humanity)¹⁵⁶⁾’ 최종보고서를 발간하였다. AI의 국제 거버넌스 문제 해결을 위해 2023년 10월 출범한 AI 자문기구는 2023년 12월 발표한 중간보고서와 AI 전문가를 포함한 전 세계 2천 명 이상이 참여한 광범위한 협의를 바탕으로 최종보고서를 작성하였다. ‘인류를 위한 AI 거버넌스’ 최종보고서는 UN에 국제협력을 바탕으로 포용적이고 분산된 AI 거버넌스 체계의 토대 마련을 촉구하고, 모든 정부와 이해관계자에 AI 거버넌스에 협력하여 인권의 발전과 보호를 강화할 것을 촉구하였다.

최근 국제적 AI 이니셔티브가 확산되고 있다(아래 그림). 이러한 새로운 이니셔티브는 국제적 수준에서 AI 거버넌스에 대한 횡단적 접근 방식을 점점 더 따르고 있으며, 특정 도메인이 아닌 AI를 전체적으로 다루는 원칙, 선언, 성명 등을 주요 AI 이니셔티브를 이루고 있다. 특히 2022년 11월 ChatGPT가 출시된 후 여러 범용 AI 대규모 언어모델이 출시되면서 2023년 이후 AI 이니셔티브가 급격히 가속화되고 있다.

156) United Nations AI Advisory Body, Governing AI for Humanity - Final Report, 2024.09.19.

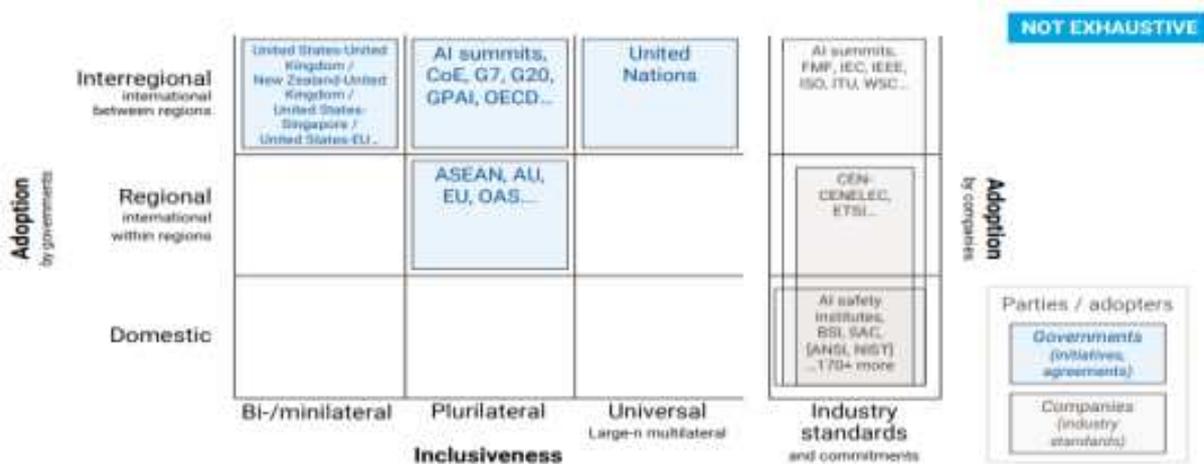
[그림 4-1] AI 거버넌스 이니셔티브, 주요 이정표, 2019-2024



출처: United Nations AI Advisory Body, Governing AI for Humanity - Final Report, 2024.09.19.

이 보고서에서는 지리적 범위와 포용성에 따라 매핑된 AI 거버넌스 이니셔티브와 산업 표준의 주요 문헌들을 아래(그림 24)과 같이 도식화하였다. 관련 지역 및 지역 간 다자간 이니셔티브의 예로는 아프리카 연합, 다양한 AI 정상회담 주최국, 동남아시아 국가연합(ASEAN), 유럽평의회(Council of Europe), 유럽연합(EU), G7, G20, GPAI(Global Partnership on Artificial Intelligence), 미주기구(Organization of American States, OAS), 경제협력개발기구(OECD) 등이 주도하는 이니셔티브가 있다.

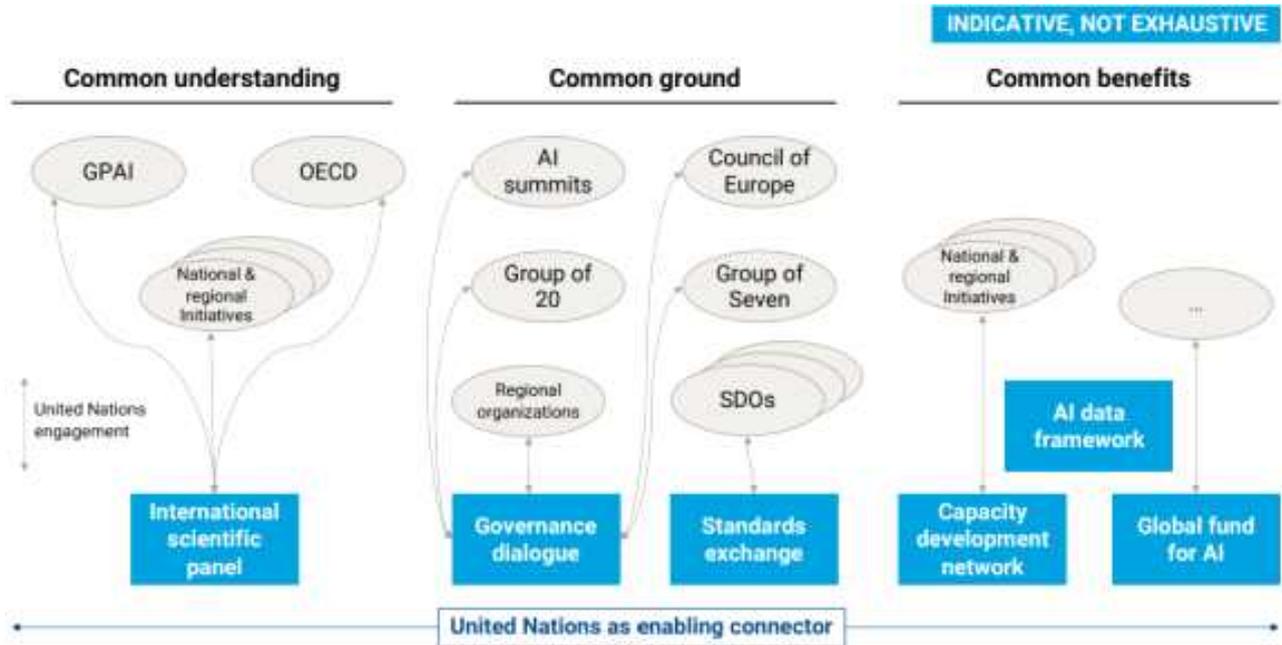
[그림 4-2] AI에 초점을 맞춘 거버넌스 이니셔티브의 출처(기관/조직)



출처: United Nations AI Advisory Body, Governing AI for Humanity - Final Report, 2024.09.19.

이 보고서에서는 국제 AI 거버넌스 생태계에서 유엔의 역할 관련 연결자(enabling connector)로서 역할을 제안하며 다음(그림)과 같이 예시적으로 제시하였다.

[그림 4-3] 국제 AI 거버넌스 생태계에서 유엔의 역할에 대한 제안



출처: United Nations AI Advisory Body, Governing AI for Humanity - Final Report, 2024.09.19.

이 최종보고서는 기존의 AI 거버넌스 논의는 전 세계에 포괄적으로 적용하기 어렵고 글로벌 사우스(Global South)¹⁵⁷⁾ 국가들이 배제되어 있다고 지적하며, 글로벌 AI 거버넌스 격차 해소를 위한 다음의 7가지 행동을 촉구하였다.

- ① (AI 국제 과학 패널 구성) AI 기술의 과학적 측면에 대한 국제적이고 다학제 과학 패널을 구성하여 정책 입안자와 대중에게 신뢰할 수 있는 정보를 제공
- ② (AI 거버넌스 정책 대화) 포괄적이고 다중이해관계자를 포함하는 AI 거버넌스 정책 대화를 시작하여 모범사례를 공유하고 국가 및 지역적 접근 방식 간의 조정 및 상호운용성을 개선
- ③ (AI 표준 교류 마련) 다양한 표준기구, 기술 기업, 시민사회 대표들 간 AI 표준 교류를 위한 포괄적인 체계를 마련해 개방적이고 상호운용이 가능하며 신뢰할 수 있는 AI 생태계를 구축

157) 선진국을 뜻하는 글로벌 노스와 상대되는 개념으로, 아시아, 아프리카, 남미의 개발도상국을 통칭

- ④ (AI 역량 개발 네트워크 구축) 공공 및 민간 부문의 모든 사람이 AI를 책임 있고 윤리적으로 사용할 수 있도록 포괄적이고 공평한 AI 역량 개발 네트워크를 구축
- ⑤ (글로벌 AI 펀드 조성) AI 혁신과 역량 강화를 위한 글로벌 기금을 조성해 AI 및 지속가능발전목표(SDG)¹⁵⁸⁾ 관련 프로젝트에 자금을 지원하고, 특히 저소득 및 중간 소득 국가의 요구사항을 우선 추진
- ⑥ (글로벌 AI 데이터 프레임워크 구축) AI 교육용 데이터의 출처 및 사용과 관련하여 공통 표준을 수립하기 위한 글로벌 AI 데이터 프레임워크를 구축
- ⑦ (AI 사무소 설립) 이번 보고서의 이행 및 다양한 이해관계자 간 협력을 지원하고 AI 관련 자문을 제공할 AI 사무소를 UN 사무국 내에 설립

158) 2015년 UN 총회에서 환경, 교육, 평등, 건강, 거버넌스 등의 분야에서 2030년까지 달성하기로 합의한 17개 목표

제3절 기타 기구(G7, G20 등)

1. 그간의 주요 동향

2023년 5월 G7(미국, 영국, 독일, 프랑스, 캐나다, 일본, 이탈리아) 정상들은 일본 히로시마에서 5월 19일부터 사흘간 열린 회의(G7 HIROSHIMA 2023)에서 AI에 대한 적절한 통제의 필요성에 공감하고 AI 국제 프레임워크를 결성하는데 합의하였다.

이후 주요 7개국(G7)이 2023년 10월 30일 ‘히로시마 AI 프로세스’를 통해 AI 기업 대상의 AI 국제 행동강령(International Code of Conduct for Advanced AI Systems)에 합의하였다. 기업의 자발적 채택을 위해 마련된 이번 행동강령은 기반모델과 생성 AI를 포함한 첨단 AI 시스템의 위험 식별과 완화에 필요한 조치를 포함하였다.

또한, G7 경쟁당국들은 2023년 11월 8일 일본 도쿄에서 G7 경쟁당국 수장 회의를 개최하고 공동성명을 채택했다. G7 경쟁당국의 공동성명은 생성 AI 분야의 독과점 우려를 제기하는 한편, 생성 AI 등 신기술이 경쟁환경에 미칠 영향에 대처하기 위한 당국 간 협력 필요성을 강조했다.¹⁵⁹⁾

2. 2024년 주요 동향

1) G7의 산업, 기술 및 디지털 장관 선언¹⁶⁰⁾

2024년 3월 14일과 15일, 이탈리아 베로나와 트렌토에서 G7 산업, 기술 및 디지털 장관 회의가 개최되었다. 이 회의는 G7 정상회의의 관련 장관 회의 중 하나로 개최되었으며, G7 국가와 초대 국가 및 관련 국제 기구가 참석하였다.

이 회의에서 G7 장관들은 1) 산업의 AI 및 신기술, 2) 안전하고 회복성 있는 네트워크, 공급망 및 주요 투입 요소, 3) 디지털 개발, 4) 공공 부문의 AI, 5) 히로시마 AI 프로세스의 성과 발전, 6) 디지털 정부에 대해 논의했다.

159) <https://www.jftc.go.jp/en/pressreleases/yearly-2023/November/231108.html>

160) <https://www.digital.go.jp/en/a11b8177-e3a3-4029-b3c0-1fbc77b59ba-en#summary>

이 회의에서 장관들은 위험과 기회를 매핑하고 공공 부문에서 안전하고 신뢰할 수 있는 AI를 위한 지침을 제공하는 툴킷을 개발하기로 합의했다. 또한, 정부가 안전하고 신뢰할 수 있는 AI를 개발하도록 지원하기 위해 관련 공공 부문 데이터의 역할을 탐구하기로 합의했다.

장관들은 주요 파트너, 조직 및 개발도상국과 신흥 경제권의 사람들 사이에서 지원과 인식을 확대하는 것을 포함하여 히로시마 AI 프로세스의 결과를 발전시키겠다는 의지를 확인했다. 또한 국제 지침 원칙과 국제 행동 강령의 보급을 촉진하고, 주요 파트너 국가 및 기업과의 참여를 확대하고, 행동강령의 적용 상태를 모니터링하기 위한 도구와 메커니즘을 개발하고, 국제기구와 협력하기로 합의했다.

2) 히로시마 AI 프로세스 프렌즈 그룹 출범¹⁶¹⁾

기시다 후미오 일본 총리가 2024년 5월 2일 프랑스 파리에서 열린 OECD 각료이사회 부대행사로 개최된 AI 행사 기조연설에서 ‘히로시마 AI 프로세스 프렌즈 그룹’의 출범을 발표하였다.

히로시마 AI 프로세스를 지지하는 49개 국가들로 구성된 프렌즈 그룹은 AI 국제 지침과 행동 규범의 이행을 지원하고 국가 간 협력을 촉진할 계획이다. 49개국에는 G7 국가 외에 우리나라를 비롯해 아르헨티나, 호주, 오스트리아, 벨기에, 불가리아, 칠레, 콜롬비아, 코스타리카, 크로아티아, 키프로스, 체코, 덴마크, 에스토니아, 핀란드, 그리스, 헝가리, 인도, 아일랜드, 라트비아 등이 포함되었다. 기시다 총리는 기조연설에서 프렌즈 그룹이 “AI로 인한 기회와 위험 대응 및 안전하고 신뢰할 수 있는 AI의 구현이라는 공통의 목적을 위해 협력할 것”이라고 설명하였다. 오픈AI의 샘 알트만(Sam Altman) CEO는 온라인 행사에 참석해 글로벌 AI 거버넌스에서 히로시마 AI 프로세스 하의 국제 지침과 행동 규범의 중요성을 강조하고 글로벌 사우스 국가들이 대거 참여한 프렌즈 그룹의 역할에 기대감을 표시했다.

일본 정부는 AI 위험을 줄이기 위한 기술적 해결책도 중요하다는 점에서 도쿄에 국제협의체 ‘AI 글로벌 파트너십(GPAI, Global Partnership on AI)’ 센터를 설립하고 기술 연구와 테스트를 지원할 예정이다.

161) Ministry of Foreign Affairs of Japan, Prime Minister Kishida's attendance at the Side Event on Generative AI at the OECD Ministerial Council Meeting, 2024.05.02.

3) OECD-GPAI 통합 운영에 합의¹⁶²⁾¹⁶³⁾

2024년 7월 3~4일 인도 뉴델리에서 열린 글로벌 AI 파트너십(Global Partnership on AI, 이하 GPAI¹⁶⁴⁾) 상반기 서밋(Summit)에서 회원국들이 '2024 GPAI 뉴델리 선언'을 발표하였다. GPAI 서밋은 매해 연말 개최되나, 2024년 상반기 서밋은 현 의장국인 인도의 제안으로 회기간 특별 개최되었으며 차기 의장국은 세르비아로 정해졌다.

GPAI 뉴델리 선언은 인간 중심적이고 안전하며 신뢰할 수 있는 AI의 발전을 위한 국제협력의 필요성과 GPAI와 OECD 간 통합 운영을 통한 시너지 창출을 골자로 하며, 주요 내용은 다음과 같다.

- 신뢰할 수 있고 인간 중심적인 AI를 증진하기 위해 강력한 과학적 근거와 개방형 솔루션 및 공통 표준을 바탕으로 다중 이해관계자가 참여하는 포괄적 접근 방식으로 협력을 강화
- OECD의 AI 권고안을 기반으로 모든 OECD 회원국과 GPAI 회원국을 동등한 지위로 통합하는 OECD-GPAI 통합 파트너십을 통해 GPAI의 새로운 비전을 수립
- GPAI나 OECD의 현재 회원국 지위와 관계없이 모든 국가가 인간 중심적이고 안전하며 신뢰할 수 있는 인공지능의 잠재력을 모두의 이익을 위해 활용하는 공통의 노력에 동참할 것을 촉구
- 관련 국제기구 및 UN 전문기구, 기타 파트너들과 포용적 대화를 통해 필요시 협력하며, 통합 파트너십의 모든 회원이 파트너십의 활동과 의사결정에 차별 없이 동등하게 참여할 것임을 강조
- GPAI 다중 이해관계자 전문가 그룹(MEG)과 OECD AI 전문가 네트워크(ONE AI) 및 기존 전문가 그룹 작업반을 통합하여 단일 전문가 그룹으로 합병
- 누구도 소외되지 않고 평화, 번영, 지속 가능한 발전을 촉진하는 미래를 위해 AI가 모두의 이익을 위해 사용될 수 있도록 통합 파트너십을 통한 건설적 대화를 지속하겠다는 목표를 확인

162) Ministry of Electronics & IT, 6th meeting of the GPAI Ministerial Council held on 3rd July 2024 at New Delhi, 2024.07.03

163) 외교부, GPAI(글로벌 인공지능 파트너십) 상반기 서밋에서 OECD-GPAI 통합 운영에 합의, 2024.07.04.

164) GPAI는 캐나다와 프랑스 주도로 G7에서 제안되어 2020년 출범한 AI 분야의 글로벌 다자 협의체로, 민주주의 가치와 인권을 존중하는 29개국이 회원국으로 참여 중이며, 한국은 15개 창립회원국 중 하나

4) G20 디지털경제 장관 AI 선언¹⁶⁵⁾

디지털 경제를 담당하는 G20 장관들은 2024년 9월 13일 브라질 마세이오에서 회의했다. 이 회의에서 G20 장관들은 이전 의장국의 성과와 공약을 바탕으로 디지털 포용성, 보편적이고 의미 있는 연결성, 디지털 정부와 포괄적 디지털 공공 인프라, 온라인 정보 무결성과 디지털 경제에 대한 신뢰, 포괄적이고 지속 가능한 개발과 불평등 감소를 위한 인공지능에 대한 논의를 진행했다.

이 회의에서 AI 관련 G20 장관들은 다음 사항에 대해 인식을 공유하였다.

- AI 솔루션의 개발 및 배포는 특히 콘텐츠를 개인화, 조정 및 생성하는 것을 목표로 할 때 윤리적이고 투명하며 감사 가능하고 인간의 감독 하에 책임을 져야 하며 개인정보 보호, 개인 데이터 보호 및 인권, 기본적 자유 및 지적 재산에 대한 존중을 보장해야 함
- 콘텐츠 인증 및 출처 메커니즘과 관련 기술 표준을 포함하여 편견을 완화하는 것의 중요성은 AI에서 생성된 콘텐츠를 식별하고 사용자가 정보 조작을 식별하는 데 도움이 될 수 있음
- 지식재산권, 개인정보 보호 및 데이터 보호를 존중하는 데이터, 알고리즘 및 콘텐츠 조정에 대한 적절한 보호 및 설명 가능성과 함께 투명성이 중요함
- 국제협력을 포함하여 위험 기반 및 인간 중심적, 발전 지향적이며, 혁신 친화적인 AI 정책 및 거버넌스 접근 방식에 전념하여 AI에서 파생된 이점과 위험 완화를 촉진

165) <https://www.dataguidance.com/news/international-g20-adopts-declaration-ai-disinformation>

제5장 인공지능 관련 주요 기업 동향

제1절 국외 기업

1. 알파벳(구글)

1) 기업 일반현황

〈표 5-1〉 알파벳(구글) 기업 현황 (단위: M\$)

구분	주요 내용																																																					
기업 개요	〈회사 개요〉																																																					
	<ul style="list-style-type: none"> • 1998년 9월 미국 캘리포니아주에서 설립 • 2004년 8월 나스닥 상장 • 전 세계에 서비스를 제공하며 50개국 70개 도시에서 지사 운영 • 전 세계 모든 정보를 체계화해 모든 사용자가 유익하게 사용할 수 있도록 하는 것을 기업 목표로 함 																																																					
	〈사업 내용〉																																																					
	<ul style="list-style-type: none"> • 산업 분야 : 인터넷, 검색, 소프트웨어, 서비스, 광고 • 서비스 : 구글 검색, 클라우드 컴퓨팅, 유튜브, 안드로이드 소프트웨어, 구글 크롬, 지메일, 구글 플레이, 구글 지도, 구글 어시스턴트, 클래스 룸, 피트니스, 여행, 드라이브, 뉴스, 번역, 문서, 캘린더, 포토, 어스, 스마트홈, 스칼라 등 																																																					
	〈자회사〉																																																					
	<ul style="list-style-type: none"> • 구글 클라우드, 유튜브, 안드로이드, 웨이모, 딥마인드, 캘리코 																																																					
재무 실적 ¹⁶⁶⁾ (최근 5년)	<table border="1"> <thead> <tr> <th></th> <th>2019</th> <th>2020</th> <th>2021</th> <th>2022</th> <th>2023</th> </tr> <tr> <th>비교 기준:</th> <th>31/12</th> <th>31/12</th> <th>31/12</th> <th>31/12</th> <th>31/12</th> </tr> </thead> <tbody> <tr> <td>총매출</td> <td>161,857</td> <td>182,527</td> <td>257,637</td> <td>282,836</td> <td>307,394</td> </tr> <tr> <td>총매출 성장률</td> <td>+18.3%</td> <td>+12.77%</td> <td>+41.15%</td> <td>+9.78%</td> <td>+8.68%</td> </tr> <tr> <td>매출 원가</td> <td>71,896</td> <td>84,732</td> <td>110,939</td> <td>126,203</td> <td>132,372</td> </tr> <tr> <td>총이익</td> <td>89,961</td> <td>97,795</td> <td>146,698</td> <td>156,633</td> <td>175,022</td> </tr> <tr> <td>총이익 성장률</td> <td>+16.42%</td> <td>+8.71%</td> <td>+50.01%</td> <td>+6.77%</td> <td>+11.74%</td> </tr> <tr> <td>총이익 %</td> <td>55.58%</td> <td>53.58%</td> <td>56.94%</td> <td>55.38%</td> <td>56.94%</td> </tr> </tbody> </table>							2019	2020	2021	2022	2023	비교 기준:	31/12	31/12	31/12	31/12	31/12	총매출	161,857	182,527	257,637	282,836	307,394	총매출 성장률	+18.3%	+12.77%	+41.15%	+9.78%	+8.68%	매출 원가	71,896	84,732	110,939	126,203	132,372	총이익	89,961	97,795	146,698	156,633	175,022	총이익 성장률	+16.42%	+8.71%	+50.01%	+6.77%	+11.74%	총이익 %	55.58%	53.58%	56.94%	55.38%	56.94%
		2019	2020	2021	2022	2023																																																
	비교 기준:	31/12	31/12	31/12	31/12	31/12																																																
	총매출	161,857	182,527	257,637	282,836	307,394																																																
	총매출 성장률	+18.3%	+12.77%	+41.15%	+9.78%	+8.68%																																																
	매출 원가	71,896	84,732	110,939	126,203	132,372																																																
	총이익	89,961	97,795	146,698	156,633	175,022																																																
총이익 성장률	+16.42%	+8.71%	+50.01%	+6.77%	+11.74%																																																	
총이익 %	55.58%	53.58%	56.94%	55.38%	56.94%																																																	

구분	주요 내용
AI 조직 현황	<p style="text-align: center;">〈AI 관련 조직〉</p> <ul style="list-style-type: none"> • Google AI/Google Research(Google의 핵심 AI 연구조직) : 기초 AI 연구부터 실용적인 AI 응용까지 폭넓은 연구를 수행 • 딥마인드(DeepMind) : 2014년 인수한 영국의 AI 연구소로 범용 인공지능(AGI) 개발을 목표로 한 심층 연구를 진행하며, 알파고, 알파폴드 등의 AI 시스템을 개발 • 구글 브레인(Google Brain) : 2011년 설립된 딥러닝 전문 연구팀으로 텐서플로우(TensorFlow) 등 AI 프레임워크 개발했고, 2023년 딥마인드와 통합되어 구글 딥마인드(Google DeepMind)로 재편 • 알파벳 X(Alphabet X, 구 Google X) : 구글의 "문샷 팩토리"로 불리는 혁신 연구소로 AI를 기반 장기적이고 도전적인 프로젝트를 수행(자율주행차, 드론 배송 등 미래 기술을 개발)
사업 전망	<p style="text-align: center;">〈AI 전략〉</p> <ul style="list-style-type: none"> • AI 퍼스트 : 모든 제품과 서비스에 AI를 핵심으로 통합하는 시도로 검색, 이메일, 사진 등 주요 서비스의 AI를 강화하며 워크스페이스(구글 문서도구)와 AI 기능을 통합 • AI 모델 개발 : 제미니 시리즈를 통한 GPT-4 대응, PaLM, LaMDA 등 대규모 언어모델의 지속적인 발전, 멀티모달 AI 역량 강화 등을 추진 • AI 인프라 구축 : TPU 등 AI 전용 칩 개발 및 고도화, 구글 클라우드를 통한 AI 인프라 서비스 제공, AI 개발을 위한 소프트웨어 도구 생태계 확장 등을 추진 • 기업용 AI 솔루션 : 버텍스(Vertex) AI 등 엔터프라이즈 AI 플랫폼 강화, 산업별 특화 AI 솔루션 개발, AI 컨설팅 서비스 확대 등을 추진 • Google은 AI 기술 지배력을 유지하면서도 책임 있는 AI 개발을 추구하는 균형 잡힌 전략을 추구하며, 특히 AGI 개발과 실용적인 AI 응용 사이의 균형 유지에 주력

자료: SPRI 정리, 2024.11.

2) 2024년 기업 동향

(1) 사업전략/투자/M&A

가. AI 모델을 활용한 자율로봇 기술 ‘오토RT’ 공개¹⁶⁷⁾

구글 딥마인드 연구진이 2024년 1월 4일 대규모 언어모델(LLM)이나 시각 언어모델(VLM)과 같은 기반 모델과 로봇 제어 모델(Robotics Transformer, RT)을 결합한 ‘오토RT(AutoRT)’를 공개했다. 오토RT는 AI 학습을 위한 데이터 수집 시스템이자 자율로봇 기술로 여러 로봇이 다양한 환경에서 동시에 다른 작업을 수행할 수 있도록 지원한다. 오토RT는 VLM을 이용해 로봇 환경과 시야에 있는 사물을 이해한 뒤, LLM으로 로봇이 수행할 수 있는 작업 목록을 제안 및 선택하도록 한다. 만일 로봇이 VLM으로 탁자에 놓인 스펀지

166) <https://kr.investing.com/equities/google-inc-income-statement>

167) Google DeepMind, Shaping the future of advanced robotics, 2024.01.04.

를 포착하면 LLM은 스펀지로 타자를 닮는 작업을 제안해 로봇이 스스로 행동하도록 하는 방식이다.

약 7개월에 걸친 광범위한 평가에서 딥마인드는 오토RT로 한 번에 최대 20대의 로봇을 작동해 6,650개의 고유한 작업에 대한 7만 7,000건 이상의 실험으로 다양한 데이터셋을 수집한다고 밝혔다. 또 연구진은 자율로봇 안전성을 보장하기 위해 인간에게 해를 끼칠 수 없다는 기본 규칙 외에도 날카로운 물체나 전기 제품 사용 금지, 인간이나 동물 관련 작업 금지와 같은 안전 규칙을 설정했다. 로봇 관절에 가해지는 힘이 기준치를 초과할 때 자동으로 멈추도록 프로그래밍했으며, 인간의 감독으로 로봇을 물리적으로 중단시킬 수 있도록 설정했다.

연구진은 RT 모델을 더욱 효율적으로 만드는 신경망 아키텍처 ‘사라(SARA)-RT’와 로봇이 명령어를 실제 물리적 동작으로 빠르게 변환할 수 있도록 하는 ‘RT-트레젝토리(RT-Trajectory)’도 개발했다. 로봇 제어 모델은 고해상도 이미지 처리 시 계산에 오랜 시간이 걸려 의사결정이 느려질 수 있는데, 사라-RT는 새로운 미세조정 방법으로 계산의 복잡성을 줄여 정확하고 빠른 동작을 지원한다고 설명했다. RT-트레젝토리는 로봇 동작 훈련을 위한 동영상에 2D 스케치를 겹쳐 가장 효과적인 로봇 동작을 안내한다. 이를 적용한 첨단 RT 모델의 동작 성공률은 29%에서 63%로 두 배 이상 향상되었다고 밝혔다.

나. 올림피아드 수준 기하학 특화 AI ‘알파지오메트리’ 공개¹⁶⁸⁾

구글 딥마인드가 2024년 1월 17일 국제 수학 올림피아드 수준의 기하학 문제를 해결할 수 있는 AI 모델 ‘알파지오메트리(AlphaGeometry)’를 오픈소스로 공개했다. 알파지오메트리는 고등학생 대상의 수학 경시대회인 국제 수학 올림피아드 문제 30개 중에서 25개를 풀어 이전 최고 기록인 10개를 넘어섰고 평균 25.9개를 푼 인간 금메달 수상자들과 비슷한 성적을 기록했다. 기존 AI 시스템은 추론 역량과 훈련 데이터 부족 등으로 복잡한 수학 문제를 풀지 못했지만, 딥마인드 연구진은 신경 언어모델과 규칙 기반의 추론 엔진을 결합해 문제를 개선했다.

168) Google Deepmind, AlphaGeometry: An Olympiad-level AI system for geometry, 2024.01.17

언어모델은 데이터의 일반적 패턴과 관계 식별에 탁월해 유용한 해결책을 신속하게 예측할 수 있지만 엄격한 추론 능력은 부족하며, 이와 반대로 추론 엔진은 명확한 규칙을 사용해 결론에 빨리 도달하지만, 복잡한 문제 처리에서는 종종 느리고 융통성이 부족하다는 지적을 받았다. 하지만 알파지오메트리는 먼저 추론 엔진으로 문제를 해석한 뒤 해답을 찾지 못하면 언어모델이 유용할 것으로 예측한 해결 방식을 이용해 새로운 추론을 시도한다.

연구진은 학습 데이터 확보를 위해 AI가 자체 합성데이터를 생성하도록 설계하여 1억 개에 달하는 다양한 난이도의 학습 데이터셋을 생성했다고 밝혔다. 생성된 데이터셋 중 900만 개 예제는 새로운 구성의 도형을 포함하며, 이러한 수많은 예제를 통해 알파지오메트리는 올림피아드 기하학 문제 풀이에서 새로운 방식의 증명을 제시하는 등의 성과를 보였다.

구글 딥마인드는 알파지오메트리의 연구 성과를 활용해 장기적으로 차세대 AI 시스템의 추론 능력을 발전시킬 계획이다. 대규모 합성데이터를 생성해 AI 시스템을 훈련하는 알파지오메트리의 접근 방식은 향후 AI 시스템이 수학 이외의 분야에서도 새로운 지식을 발견하는 토대가 될 수 있다는 평가다. 구글 딥마인드는 장기적으로 기하학을 넘어 수학 전반에 걸쳐 적용이 가능한 AI 시스템을 구축하고 범용 AI 시스템에 사용될 수 있는 정교한 문제 해결과 추론 능력 개발을 목표로 제시했다.

다. 영상에 어울리는 오디오를 생성하는 V2A 기술 공개¹⁶⁹⁾

구글 딥마인드가 2024년 6월 17일 동영상에 어울리는 음향 효과나 대사를 자동으로 생성하는 V2A(Video-to-Audio) 기술을 공개했다. 동영상 생성 모델의 빠른 발전에도 대다수 모델은 여전히 무음 동영상을 생성하지만, V2A는 구글의 ‘비오(Veo)’와 같은 동영상 생성 모델과 결합해 동영상과 어울리는 음악이나 음향 효과, 영상에 등장하는 캐릭터의 대사 등을 생성할 수 있다.

연구진은 모델 훈련을 위해 영상과 오디오 데이터와 함께 오디오에 대한 자세한 설명과 음성 대화 내용을 담은 AI 생성 주석을 활용했으며, V2A는 주석에 제공된 정보를 바탕으로 특정 오디오를 다양한 시각적 장면과 연관시키는 방법을 학습했다. 사용자는 V2A를 활용해

169) Google Deepmind, Generating audio for video, 2024.06.17.

동영상 입력만으로 사운드트랙을 생성할 수 있으며, 텍스트 프롬프트를 이용하면 더욱 정확한 오디오를 만들 수 있다.

또 사용자는 긍정적 프롬프트를 통해 원하는 오디오를 만들도록 유도하거나, 부정적 프롬프트를 통해 원하지 않는 오디오를 제외할 수 있으며, 생성된 오디오는 동영상과 수동으로 정렬할 필요가 없이 자동으로 화면 상황에 맞게 들어간다. 연구진은 V2A의 기술 개선을 진행 중이며, 이를 통해 동영상 결합으로 인한 오디오 품질 저하나 캐릭터의 입 모양과 음성을 일치시키는 문제 등을 해결할 방침이라고 밝혔다.

V2A로 생성된 오디오의 품질은 입력된 동영상의 품질에 따라 달라지므로 모델 학습 분포를 벗어나는 동영상의 문제로 인해 오디오 품질이 현저히 저하될 수 있다. 입력된 대본에 맞게 음성을 생성하고 이를 캐릭터의 입술 움직임과 일치시키는 작업에서도 V2A와 연동된 동영상 생성 모델이 대본에 맞는 입술 움직임을 생성하지 못하면 립싱크가 어긋날 수 있다. 구글 딥마인드는 V2A 기술의 오용을 막기 위해 대중에 공개하기에 앞서 엄격한 안전성 평가와 테스트를 거치는 한편, 모든 AI 생성 콘텐츠에 워터마크를 표시할 계획이다.

라. 소형 오픈소스 모델 ‘젬마2 2B’ 공개¹⁷⁰⁾

구글이 2024년 7월 31일 다양한 하드웨어에서 실행할 수 있는 매개변수 20억 개의 소형 오픈소스 모델 ‘젬마(Gemma)2 2B’를 공개했다. 구글은 2024년 6월 매개변수 270억 개와 90억 개 크기의 젬마2를 출시했으며, 새로 공개된 젬마2 2B는 매개변수가 20억 개로 모바일이나 노트북, PC에서도 원활하게 사용할 수 있다. 젬마2 2B는 인간 선호도를 평가하는 챗봇 아레나(Chatbot Arena)¹⁷¹⁾에서 1126점을 기록해 믹스트랄(Mixtral) 8x7b(1114점)와 GPT-3.5(1106점)를 능가했다. 구글은 젬마2 2B가 지식 증류(Knowledge Distillation)¹⁷²⁾ 기법으로 개발되어 동급 최고의 성능을 발휘한다고 주장했다.

구글은 젬마2 2B와 함께 AI 모델의 입출력에서 유해한 콘텐츠를 탐지하고 완화할 수 있는 개방형 안전 콘텐츠 분류기인 ‘셴드젬마(ShieldGemma)’도 발표했다. 셴드젬마는 ▲혐오

170) Google, Smaller, Safer, More Transparent: Advancing Responsible AI with Gemma, 2024.07.31.

171) 미국 버클리 캘리포니아대가 설립한 연구조직 LMSYS org의 AI 모델 평가 사이트로, 일정한 평가 기준에 따른 다양한 AI 모델 간 비교와 평가를 지원

172) 더 큰 모델에서 추출한 지식을 더 작은 모델로 전달하여 더 작은 모델의 성능을 향상하는 기법

발언 ▲괴롭힘 ▲성적 콘텐츠 ▲위험한 콘텐츠와 같은 4개 영역에서 유해 콘텐츠를 필터링해 사용자와 서비스를 보호하며, 온라인 필터링에 적합한 2B 버전 및 오프라인 앱에서 더 뛰어난 성능을 지원하는 9B와 27B 버전까지 다양한 크기로 제공한다.

구글은 연구자와 개발자가 젤마2 모델의 의사결정 프로세스를 투명하게 확인할 수 있는 도구로 ‘젤마 스코프(Gemma Scope)’도 공개했다. 구글은 ‘희소 오토인코더(Sparse Auto Encoder, SAE)¹⁷³⁾’를 활용해 모델 내 특정 지점을 확대해 내부 작동을 해석하기 쉽게 만드는 젤마 스코프를 이용해 연구자들이 신뢰할 수 있는 AI 시스템을 개발할 수 있도록 지원한다고 밝혔다.

(2) 기술개발 및 혁신

가. 제미나이의 후속 버전 ‘제미나이 1.5 프로’ 공개¹⁷⁴⁾

구글이 2024년 2월 15일 ‘제미나이(Gemini)’의 후속 버전 ‘제미나이 1.5 프로’를 발표하고 일부 개발자와 기업 고객을 대상으로 프리뷰 버전을 공개했다. 제미나이 1.5 프로는 역대 구글의 AI 모델 중 가장 뛰어난 ‘제미나이 1.0 울트라’와 비슷한 수준의 성능을 지원한다. 또한 더 적은 컴퓨팅을 사용하는 중형 멀티모달 모델로서, 종합적인 벤치마크 테스트에서 기존 제미나이 1.0 프로보다 87% 향상된 성능을 기록했다.¹⁷⁵⁾

제미나이 1.5 프로는 최대 100만 개 토큰을 처리할 수 있으며, 기본 제공되는 토큰 처리 규모가 12만 8천 개로 구매 옵션에 따라 추가 토큰을 지원할 예정이다. 100만 개의 토큰은 1시간 분량의 동영상, 11시간 분량의 음성, 3만 줄의 코드, 70만 개 이상의 단어에 해당하는 방대한 정보를 한 번에 처리할 수 있다는 의미이다. 제미나이 1.5 프로는 GPT-4(128,000개 토큰 지원)나 기존 제미나이 프로(32,000개 토큰 지원)를 크게 앞선다. 현재 일부 개발자와 기업 고객에게 제공되는 프리뷰 버전은 개발자용 AI 개발도구인 ‘구글 AI 스튜디오’와 기업용 생성 AI 구축 플랫폼 ‘버텍스 AI’를 통해 100만 개의 입력 토큰 사이즈를 지원하며, 향후 토큰 사이즈 별로 다양한 구매 옵션을 제공할 예정이다.

173) 젤마2에서 처리되는 복잡한 정보를 풀어내어 이해하기 쉬운 형태로 확장할 수 있도록 돕는 특수 신경망

174) Google, Our next-generation model: Gemini 1.5, 2024.02.15

175) The Verge, Gemini 1.5 is Google's next-gen AI model – and it's already almost ready, 2024.02.16.

제미니 1.5 프로는 복잡한 추론 기능을 갖춰 주어진 프롬프트에서 대량의 콘텐츠를 원활하게 분석, 분류 및 요약 기능을 제공하며 코딩 능력도 향상됐다. 일례로 달 탐사 목적으로 발사된 우주선 아폴로 11호와 관련된 402쪽 분량의 문서를 입력한 뒤, 발로 땅을 딛는 단순한 그림을 제시하고 무슨 장면이냐고 물으면, 제미니 1.5 프로는 문서 전반의 정보와 대화, 이미지를 분석해 닐 암스트롱이 달에 첫발을 내딛는 모습이라고 정확하게 유추하기도 했다.

제미니 1.5 프로는 ‘문맥 기반 학습(In-context Learning)’ 기능을 갖춰 이를 통해 추가 미세조정 없이 프롬프트 내 정보에서 새로운 기술을 학습할 수 있다. 사용자 200명 미만의 부족 언어인 칼라망(Kalamang)어의 문법서를 프롬프트로 제공하자, 제미니 1.5 프로는 사람과 비슷한 수준으로 영어-칼라망어 번역을 수행했다.

나. 긴 컨텍스트 창 활용한 LLM 성능 향상 기법 연구¹⁷⁶⁾

구글 딥마인드가 2024년 4월 17일 LLM의 긴 컨텍스트 창을 이용한 ‘상황 내 학습(In-Context Learning, ICL)’으로 미세조정 없이 모델 성능을 향상하는 기법을 연구한 논문¹⁷⁷⁾을 아카이브에 공개했다. ICL은 컨텍스트 창에 예제를 제공하여 새로운 작업을 학습하게 하는 방법으로, 언어모델에 “고양이:동물-사과:?”라는 예제를 제공하면 모델은 “과일”이라고 응답한다. 미세조정과 달리 ICL은 매개변수를 변경할 필요가 없어 활용이 쉽지만, 컨텍스트 창 크기로 인해 프롬프트에 입력할 수 있는 예제 수에 한계가 있었다. 하지만 최근 제미니 1.5 프로 등 100만 개 이상의 토큰 입력을 지원하는 모델이 출시되며 수백에서 수천 개에 달하는 ICL 예제 입력이 가능해졌다.

단일 프롬프트에 여러 예제를 입력하는 ‘다중샷(Many-Shot) ICL’이 수학 문제 풀이, 질문 답변, 요약, 번역 등 여러 영역에서 LLM의 성능에 미치는 영향을 실험한 결과, 예제가 프롬프트에 더 많이 추가될수록 모델의 성능은 계속 향상되는 것으로 나타났다. 요약 작업에서 ‘다중샷 ICL’ 적용 결과, 제미니 프로는 요약에 특화되도록 미세 조정된 모델과 동등한 성능을 나타낸다. 연구진은 하나의 프롬프트에 최대 8,192개의 예제를 입력하기도 했으며, 모든 유형의 작업에서 컨텍스트 창 내의 예제가 수십만 개 토큰 분량으로 확장되었

176) Venturebeat, DeepMind researchers discover impressive learning capabilities in long-context LLMs, 2024.04.24.

177) Many-Shot In-Context Learning
(<https://arxiv.org/abs/2404.11018>)

을 때 모델 성능이 최대 수준에 도달한다고 밝혔다.

‘다중샷 ICL’은 인간이 대량으로 고품질 예제를 생성해야 한다는 단점을 안고 있으며, 연구진은 인간이 생성한 데이터 의존도를 줄이기 위한 기법을 함께 제안했다. 이러한 기법은 LLM에 유사 작업 해결을 위한 추론 단계를 추가하여 모델이 예제를 생성하게 하는 ‘강화 ICL’과, LLM이 이미 과제 해결이 필요한 지식을 보유한 경우 프롬프트에 관련 정보를 추가해 사람이 답을 제공하지 않아도 LLM이 문제를 해결하게 하는 ‘비지도(Unsupervised) ICL’로 구성된다. ‘다중샷 ICL’ 기법은 수백 개의 예제 입력 시 늘어나는 추론 시간과 비용 문제로 인해 대규모 적용은 어려우나 LLM 기반 응용 서비스의 탐색과 개발 단계에서 다양한 프롬프트 엔지니어링 기술의 시험에 유용할 것으로 평가된다.

다. 생물학적 분자 구조와 상호작용을 예측하는 ‘알파폴드3’ 공개¹⁷⁸⁾

구글 딥마인드가 2024년 5월 8일 단백질뿐 아니라 DNA와 RNA 등 생물학적 분자 구조와 소분자 간 구조 및 상호작용을 예측하는 AI 모델 ‘알파폴드(AlphaFold) 3’를 발표했다. 2018년 첫 버전 발표 이후 2022년 두 번째 버전이 공개된 알파폴드는 입력된 분자 목록을 이용해 3차원 구조를 생성한 뒤 각 분자가 어떻게 결합하는지를 보여준다. 현재까지 수백만 명의 연구자가 말라리아 백신, 암 치료법, 효소 설계에 알파폴드 2를 활용했으며, 관련 연구 성과는 2만 회 이상 인용됐다.

기존 버전은 단백질 구조 예측만 가능했으나 알파폴드 3는 거의 모든 생물학적 분자의 구조를 예측하고 분자 간 상호작용을 모델링할 수 있다. 알파폴드 3는 단백질과 여타 분자 유형의 상호작용에서 기존 예측 방법 대비 50% 이상 성능이 향상됐고, 일부 중요한 상호작용에서는 예측 정확도가 두 배로 향상됐다. 구글 딥마인드는 연구 지원 도구인 ‘알파폴드 서버(AlphaFold Server)’를 통해 알파폴드 3의 대부분 기능을 무료 공개해 신약 개발과 질병 치료 연구에 널리 사용될 기반으로 조성했다.

알파폴드 서버는 단백질이 세포 전체에서 다른 분자와 어떻게 상호작용하는지 예측할 수 있는 도구로, 과학자들은 몇 번의 클릭으로 단백질, DNA, RNA 등 다양한 구조를 모델링할

178) Google, AlphaFold 3 predicts the structure and interactions of all of life's molecules, 2024.05.08.

수 있다. 구글 딥마인드는 컴퓨팅 자원에 대한 접근권한이나 머신러닝에 대한 전문지식 없이도 분자 상호작용을 예측할 수 있는 알파폴드 서버를 통해 작업 속도가 향상되고 혁신이 가속화될 것으로 기대한다.

현재 구글 모회사 알파벳은 신약 개발 자회사 아이소모픽 랩스(Isonorphic Labs), 제약사와 협력해 알파폴드 3를 활용한 신약 개발을 진행 중이다. 알파폴드 3는 단백질-리간드 결합 및 항체와 표적 단백질의 결합 등 약물 상호작용 예측에 높은 정확성을 제공하고, 항체-단백질 결합의 예측 능력은 인간 면역반응과 새로운 항체 설계 이해에 매우 중요¹⁷⁹⁾하다는 평가를 받고 있다.

라. 제미나이 1.5 플래시 및 오픈소스 모델 젬마 2 정식 출시¹⁸⁰⁾

구글이 2024년 6월 28일 ‘제미나이 1.5 프로’ 모델의 컨텍스트 창을 200만 개 토큰으로 확장하는 한편, 제미나이 1.5 프로보다 속도를 개선한 ‘제미나이 1.5 플래시’를 정식 출시했다.¹⁸¹⁾ 구글은 2024년 5월 I/O에서 대기자 명단을 받아 제미나이 1.5 프로에서 200만 개 토큰 컨텍스트 창¹⁸²⁾을 제공했고, 이번 발표로 계기로 모든 개발자로 대상을 확대할 계획이다. 컨텍스트 창이 커지면 입력비용이 증가할 가능성이 높은 만큼 구글은 입력 비용을 줄이기 위한 컨텍스트 캐싱¹⁸³⁾을 지원하며, 이를 통해 입력 비용이 최대 75%까지 줄어들 전망이다. I/O에서 프리뷰로 공개된 제미나이 1.5 플래시도 정식 출시됐고, 100만 토큰의 컨텍스트 창을 지원해 1만 자 입력 기준 GPT-3.5 터보보다 속도가 40% 빠르며 입력 비용은 4배 가량 낮다.

구글은 2024년 8월 5일부터 제미나이 API와 구글 AI 스튜디오를 통해 모든 개발자를 대상으로 제미나이 1.5 플래시에 대한 미세조정을 지원한다. 구글은 개발자와 연구자를 대상으로 오픈소스 모델 ‘젬마(Gemma) 2’를 2개 버전으로 정식 출시했다. 젬마 2는 90억 개 및 270억 개 매개변수의 2개 모델로 구성되며, 이전 모델보다 성능이 뛰어나고 추론 효율성이

179) 수용체와 같은 큰 생체 분자에 특이하게 결합하는 물질을 말하며, 생체 내의 중요한 요소이자 의약품의 개발 등에 큰 역할을 함

180) Google, Gemini 1.5 Pro 2M context window, code execution capabilities, and Gemma 2 are available today, 2024.06.27.

181) Google, Gemma 2 is now available to researchers and developers, 2024.06.27.

182) 200만 개 토큰은 2시간 분량의 동영상, 6만 줄 이상의 코드, 150만 개 이상의 단어를 처리할 수 있는 용량

183) 특정 입력을 모델에 전달하고 입력 토큰을 캐시한 다음 후속 요청에서 캐시된 토큰을 참조함으로써 비용과 지연 시간을 줄이는 방식

높으며 안전성도 개선됐다. 270억 개 매개변수의 모델은 동급에서 최고의 성능을 나타내며 크기가 2배 이상인 모델에 필적하는 성능을 보이며, 90억 개 매개변수의 모델 역시 메타의 라마3 8B 및 여타 개방형 모델을 능가한다. 젤마 2는 고성능 게임용 노트북과 하이엔드 데스크톱 등 다양한 하드웨어에서 빠른 속도로 실행할 수 있도록 최적화됐고, 높은 추론 효율성으로 고성능을 유지하면서도 비용 절감이 가능하다.

마. 제미나이 활용 자연어 기반 로봇 학습 연구 결과 공개¹⁸⁴⁾

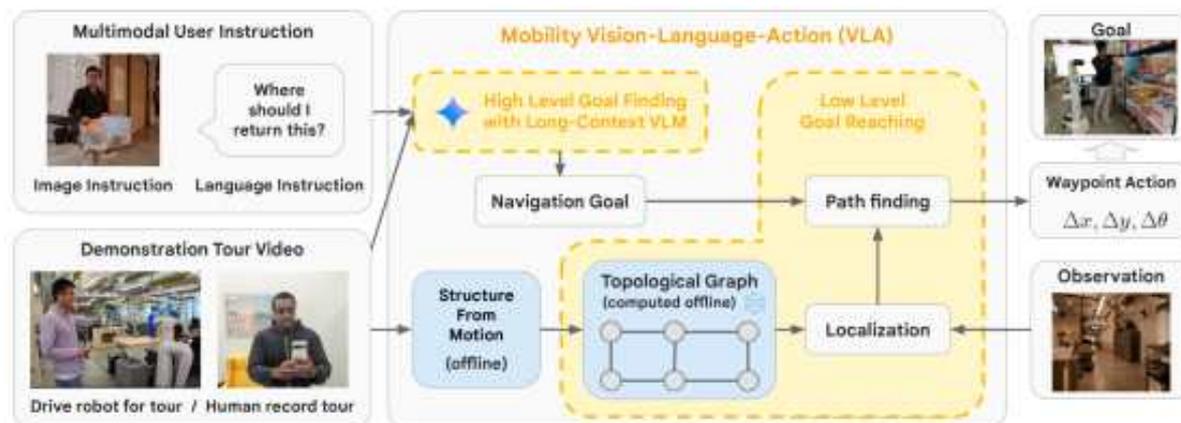
구글 딥마인드가 2024년 7월 10일 제미나이를 활용해 로봇이 주변 환경을 학습해 명령을 수행할 수 있도록 안내하는 모빌리티 시각-언어-행동(VLA) 모델에 관한 논문¹⁸⁵⁾을 공개했다. 딥마인드 로봇팀은 제미나이 1.5 프로의 긴 컨텍스트 창을 활용한 자연어 명령으로 로봇과 더욱 쉽게 상호작용할 수 있다고 설명했다. 연구진은 우선 제미나이 1.5 프로를 탑재한 로봇에게 사무실 내 공간을 촬영한 영상을 보여주고, 로봇은 영상을 통해 공간 구조나 사물의 위치, 주요 특징을 학습했다. 로봇은 학습한 내용을 토대로 이미지나 자연어를 활용한 명령을 수행하도록 설계됐다. 가령 휴대전화를 보여주고 충전 장소를 물으면 로봇은 동영상에서 본 전원 콘센트로 사용자를 안내하는 기능을 선보였다.

제미나이 탑재 로봇은 9천 제곱피트 이상의 작업공간에서 50개 이상의 사용자 명령에 대해 90%의 명령 수행 성공률을 보였고, 각 요청 처리에 10~30초의 시간이 소요되는 것으로 나타났다. 연구진에 따르면 제미나이 탑재 로봇은 단순한 공간 내비게이션을 넘어 사용자의 지시를 이행하는 방법을 계획할 수 있다. 가령 책상에 여러 개의 콜라 캔을 둔 사용자가 로봇에게 자신이 좋아하는 음료가 있는지 확인해달라고 요청하면 제미나이는 로봇에게 냉장고로 가서 콜라가 있는지 확인 후 결과를 보고하기 위해 사용자에게 되돌아가도록 지시한다.

184) The Verge, Google says Gemini AI is making its robots smarter, 2024.07.11.

185) Mobility VLA: Multimodal Instruction Navigation with Long-Context VLMs and Topological Graphs

[그림 5-1] 구글 덤마인드의 모빌리티 VLA 모델 작동 방식



출처: Mobility VLA: 구글의 스마트한 멀티모달 내비게이션 기술

(3) 제품 및 서비스

가. 바드를 제미나이로 통합하고 ‘제미나이 어드밴스드’ 유료 출시¹⁸⁶⁾

구글이 2024년 2월 8일 AI 챗봇 ‘바드(Bard)’의 명칭을 ‘제미나이(Gemini)’로 변경하고 가장 강력한 성능을 지닌 ‘제미나이 어드밴스드’를 출시했다.¹⁸⁷⁾ 구글은 무료 버전인 제미나이 프로 1.0을 230개 국가·지역에서 40개 이상 언어로 제공하는 한편, 유료 버전인 제미나이 어드밴스드를 150개 국가와 지역에서 영어로 우선 제공한다. 가장 강력한 최신 AI 모델인 제미나이 울트라를 적용한 제미나이 어드밴스드는 코딩, 논리적 추론, 창의적 협업과 같은 복잡한 업무를 효과적으로 수행한다. 이용자의 학습 스타일에 따른 맞춤형 개인 교사 역할을 하거나 콘텐츠 전략이나 사업 계획의 수립을 도울 수 있다. 구글 워크스페이스용 생성 AI 도구 ‘듀엣 AI’도 조만간 제미나이에 통합될 예정으로 유료 버전 가입자들은 지메일, 구글 닥스, 스프레드시트 등에서도 제미나이를 이용할 수 있다.

제미나이 어드밴스드는 구글 원-요금제 일환으로 월 19.99달러 구독료로 이용할 수 있고, 2TB의 스토리지 용량을 포함한 기존 구글원 프리미엄 요금제 혜택도 제공된다. 제미나이 어드밴스드 구독료는 경쟁사 오픈AI의 챗GPT 플러스 요금과 같지만, 2TB의 클라우드 스토리지 용량과 구글 워크스페이스를 이용할 수 있어 가격 경쟁력이 높다는 평가를 받는다. 하지만 오픈AI 역시 구글은 제공하지 않는 GPT 스토어 기능을 제공해 치열한 경쟁이 나타나고 있다.

186) Google, Bard becomes Gemini: Try Ultra 1.0 and a new mobile app today, 2024.02.08

187) TechCrunch, Google launches Gemini Ultra, its most powerful LLM yet, 2024.02.08.

구글은 안드로이드용 제미나이 앱을 출시하고 iOS용 구글 앱을 업데이트했다. 안드로이드와 iOS 환경에서는 미국에서 영어로 우선 서비스를 제공하고 일본어와 한국어를 시작으로 점차 지원 언어와 국가를 확대할 계획이다. iOS 환경에서는 전용 앱 대신 구글 앱을 통해 제미나이를 이용할 수 있고, 안드로이드 환경에서는 제미나이 앱을 내려받아 기본 어시스턴트로 설정한 뒤 홈이나 전원 버튼으로 제미나이를 활성화할 수 있다. 구글 어시스턴트의 음성 기능도 제미나이를 통해 지원한다.

나. ‘제미나이’ 신모델/에이전트 등 생성 AI 대거 발표¹⁸⁸⁾

구글이 2024년 5월 14일 연례 개발자회의(I/O 2024)를 개최하고 ▲제미나이 신모델 ▲오픈소스 모델 ‘젬마2’ ▲동영상 생성 AI ‘비오’ 등 생성 AI 제품군을 대거 공개¹⁸⁹⁾했다. 구글은 경량화 모델 ‘제미나이 1.5 플래시’를 프리뷰로 공개했다. 해당 모델은 제미나이 1.5 프로보다 더 가볍고 빠르고 효율적이며 100만 개 토큰 길이의 컨텍스트 창을 지원한다. 구글은 또한 제미나이 1.5 프로의 컨텍스트 창 길이를 200만 개 토큰으로 확장하는 한편, 온디바이스 AI 모델 ‘제미나이 나노’를 이미지 입력을 처리할 수 있도록 업데이트했다.

구글은 차세대 오픈소스 모델 ‘젬마 2’도 공개했으며, 270억 개 매개변수의 이 모델은 매개변수가 두 배 이상 많은 메타의 ‘라마 3 70B’에 필적하는 성능을 보였다. 구글은 텍스트와 이미지, 영상, 음성을 모두 처리할 수 있는 멀티모달 AI 에이전트 개발을 위한 ‘프로젝트 아스트라(Project Astra)’도 공개했다. 제미나이와 기타 작업별 특화 모델을 기반으로 구축된 에이전트는 실시간으로 음성 명령의 맥락을 이해해 신속하게 반응할 수 있으며, 일부 기능은 연내 제미나이 앱 등 구글 제품에 통합될 예정이다.

구글은 또한 창작자용 생성 AI 도구로 동영상 생성 모델 ‘비오(Veo)’와 이미지 생성 모델 ‘이매진(Imagen) 3’를 일부 창작자 대상으로 선공개했다. 비오는 다양한 영상 스타일로 1분이 넘는 길이의 1080p 해상도 동영상을 생성할 수 있으며, ‘타임랩스’나 ‘공중촬영’과 같은 영화 용어를 이해하여 창의적인 영상 생성이 가능하다. 이매진 3는 이전 버전보다 자연어 프롬프트를 더욱 정확히 이해하여 사실적이고 생생한 이미지를 생성하며, 텍스트 렌더링 기능도 대폭 향상됐다.

188) Google, Gemini breaks new ground with a faster model, longer context, AI agents and more, 2024.05.14.

189) Google, New generative media models and tools, built with and for creators, 2024.05.14.

구글은 생성 AI 기능을 검색 시스템과 통합해 구글 검색 시 결과 상단에 ‘AI 오버뷰(AI Overview)’ 를 표시하는 기능도 새로 출시했다. 제미나이 기반 검색 결과 요약과 관련 링크를 제공하는 ‘AI 오버뷰’ 기능은 미국 사용자를 대상으로 우선 제공되며, 추후 더 많은 국가로 확대 제공될 예정이다.

다. AI 신기능 ‘젬스’ 와 이미지 생성 모델 ‘이마젠 3’ 출시¹⁹⁰⁾

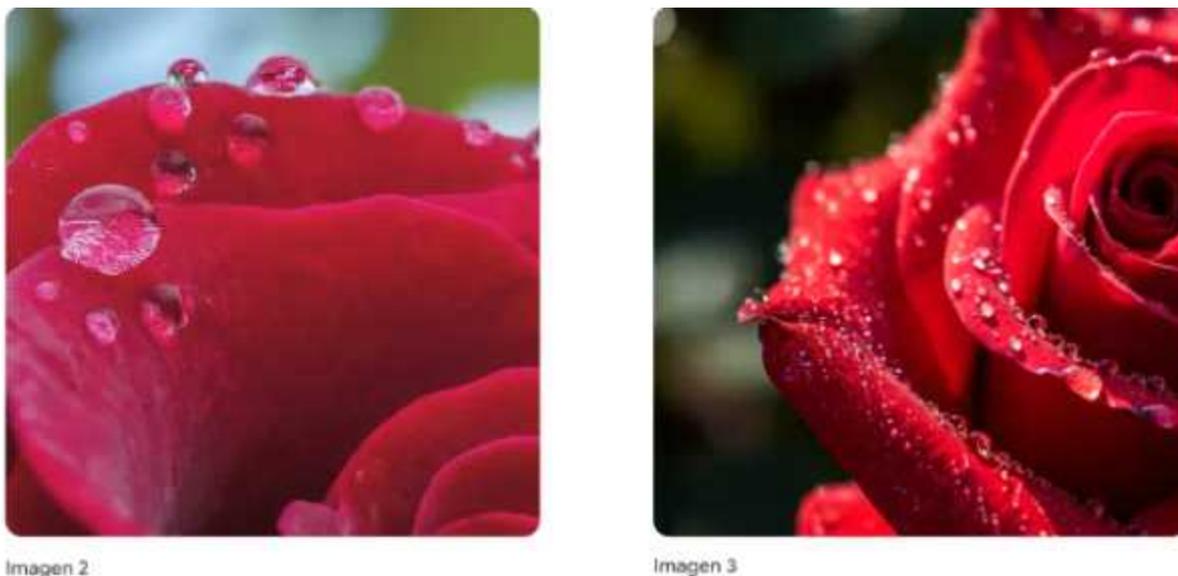
구글이 2024년 8월 28일 맞춤형 AI 챗봇을 생성할 수 있는 신기능 ‘젬스(Gems)’ 를 제미나이 유료 사용자 대상으로 공개했다. 젬스는 제미나이를 특정 주제에 대한 전문가로 설정하거나 원하는 목표에 맞게 미세 조정해 사용자 맞춤형 AI 챗봇을 제작하는 기능이다. 제미나이 어드밴스드, 비즈니스, 엔터프라이즈 이용자들은 데스크톱 환경에서 대부분 언어로 젬스를 이용할 수 있다.

구글은 ▲복잡한 주제를 쉽게 이해할 수 있도록 안내하는 ‘학습 코치’ ▲다양한 아이디어를 제공하는 ‘브레인스토머’ ▲코딩 실력 향상을 위한 ‘코딩 파트너’ 와 같은 젬스 샘플도 제공한다. 구글은 또한 최신 이미지 생성 모델 ‘이마젠(Imagen) 3’ 를 제미나이 앱에 통합하고 모든 언어 이용자에게 확대 제공할 계획이라고 발표했다. 이마젠 3는 간단한 텍스트 프롬프트를 바탕으로 사실적 풍경이나 유화 같은 질감의 이미지, 클레이 애니메이션 등 다양한 스타일의 이미지를 생성하며, 이마젠 2보다 이미지 생성 기능이 크게 향상됐다.

구글은 이마젠 3를 출시하며 역사적으로 부정확한 인물 이미지 생성으로 2024년 2월 중단된 제미나이의 인물 이미지 생성 기능을 재개하기로 했으며, 사실적이며 식별할 수 있는 인물이나 미성년자, 지나치게 폭력적이거나 선정적인 이미지 생성은 불가능하도록 구현했다. 인물 이미지 생성 기능은 제미나이 어드밴스드, 비즈니스, 엔터프라이즈 이용자를 대상으로 영어로 우선 제공되며 향후 지원 언어와 이용자 범위를 확대할 계획이다.

190) Google, New in Gemini: Custom Gems and improved image generation with Imagen 3, 2024.08.28.

[그림 5-2] 동일 프롬프트에 대한 이마젠2와 이마젠3의 결과물 비교



출처: Google, New in Gemini: Custom Gems and improved image generation with Imagen 3

(4) AI 윤리 및 규제 대응

가. LLM의 신뢰성 향상을 위한 ‘ASPIRE’ 프레임워크 개발¹⁹¹⁾

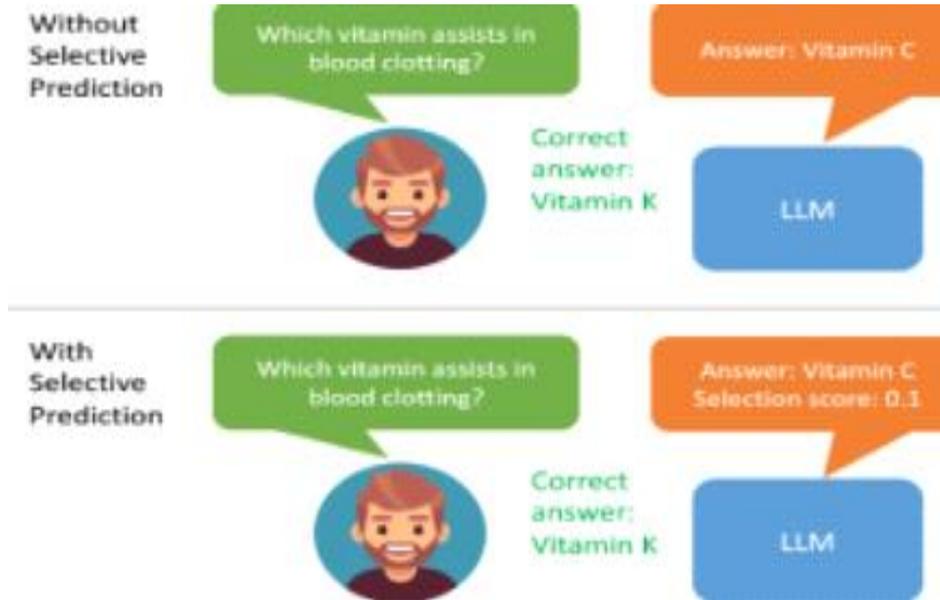
구글이 2024년 1월 18일 LLM이 답변을 스스로 평가해 답변의 정확도를 확인하도록 하여 추측보다 정직성을 촉진하는 새로운 프레임워크 ‘ASPIRE’를 공개했다. 구글은 여기에 LLM의 선택적 예측 향상을 위한 자체 평가를 통한 적응(Adaptation with Self-Evaluation to Improve Selective Prediction in LLMs, ASPIRE) 기법을 적용해 LLM 예측이 얼마나 확실한지를 나타내는 신뢰도 점수와 함께 답변을 제공한다고 밝혔다. 가령 “혈액 응고 조절에 도움이 되는 비타민이 무엇인가”라는 질문에 기존 LLM이 잘못된 답변만 제시했다면, ASPIRE를 적용한 LLM은 신뢰도 점수와 함께 답변을 제시하고 신뢰도 점수가 낮은 경우 ‘모르겠다’라고 답변하는 식이다.

‘소프트 프롬프트’ 조정 방식을 채택한 ASPIRE는 먼저 첫 번째 소프트 프롬프트로 질문에 대한 답변을 생성한 다음 두 번째 소프트 프롬프트로 학습된 자체 평가 점수를 계산해 답변의 정확성을 평가한다. 벤치마크 테스트를 통해 ASPIRE의 효과를 평가한 결과, ASPIRE를 적용한 모델(27억 개 매개변수)은 기존 데이터셋으로 사전 훈련된 더 큰 규모의 모델(300억 개 매개변수)보다 정확성이 향상된 것으로 나타났다. 연구진은 언어모델의 성능

191) Google Research, Introducing ASPIRE for selective prediction in LLMs, 2024.01.18.

은 크기가 작아지지 않으며, 전략적 조정을 통해 모델의 효율성을 대폭 개선함으로써 더 작은 모델에서도 더욱 정확한 예측이 가능하다는 점이 확인되었다고 강조했다.

[그림 5-3] ASPIRE 적용 전과 적용 후 LLM 답변 비교



출처: Google Research, Introducing ASPIRE for selective prediction in LLMs

나. AI 위험 평가를 위한 ‘프런티어 안전 프레임워크’ 공개¹⁹²⁾

구글 딥마인드가 2024년 5월 14일 심각한 피해를 초래할 수 있는 미래의 AI 위험을 사전에 파악해 완화하기 위한 ‘프런티어 안전 프레임워크(Frontier Safety Framework)’를 공개했다. 구글은 2025년 초까지 강력한 기능을 가진 첨단 모델의 위험 증대에 중점을 두고 프레임워크를 개선할 계획으로, 산학연 협력과 실제 구현을 통해 탐색 단계의 프레임워크가 대폭 향상될 것으로 기대된다. ‘책임 있는 기능 확장(Responsible Capability Scaling)’ 접근 방식에 따른 이 프레임워크는 총 3개 요소로 구성됐다.

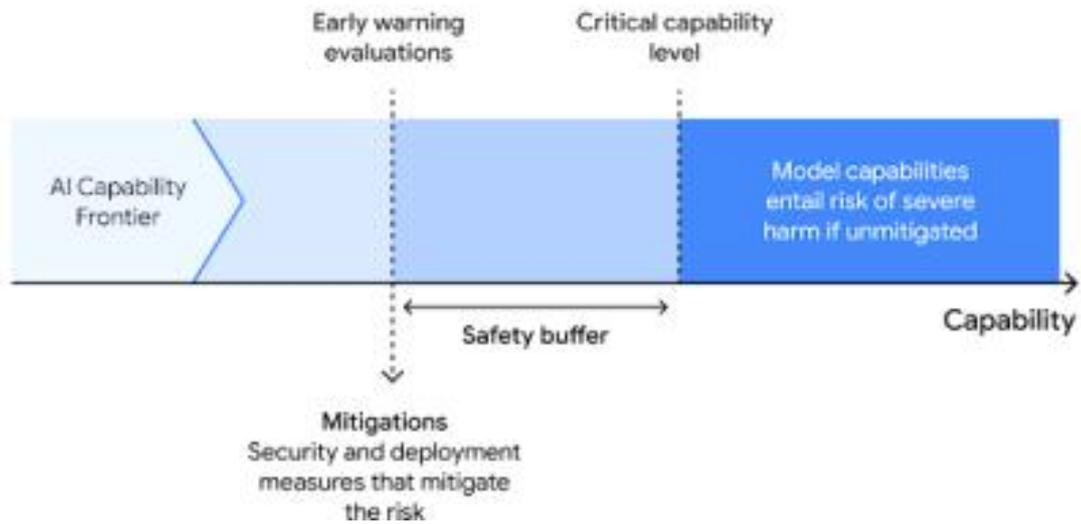
첫 번째 요소는 심각한 피해를 미칠 가능성이 있는 모델 기능을 식별하는 것으로, 고위험 영역에서 피해를 일으킬 수 있는 기준점이 되는 모델 기능¹⁹³⁾과 피해를 미칠 경로를 조사한다. 두 번째 요소는 첨단 모델을 주기적으로 평가해 CCL에 도달하는 시기를 파악하는 것

192) Google Deepmind, Introducing the Frontier Safety Framework, 2024.05.17.

193) 프레임워크는 이를 ‘중요 역량 수준(Critical Capability Level, CCL)’로 규정하며, 모델 자율성, 생물학적 안전성, 사이버보안, 기계학습 연구개발의 4가지 분야를 대상으로 모델 역량을 파악

으로, 모델이 CCL에 근접할 경우, 이를 알리는 ‘조기 경고 평가(Early Warning Evaluation)’ 도구를 개발하여 이를 수시로 평가한다. 세 번째 요소는 조기 경고 평가를 통과한 이후 모델에 대한 위험 완화계획 적용으로, 모델의 이점과 위험 간 균형 및 출시 목적을 고려한 완화 조치는 모델의 유출 방지와 중요 기능의 오용 방지에 중점을 둔다.

[그림 5-4] ‘프런티어 안전 프레임워크’의 구성요소 간 관계



출처: Google Deepmind, Introducing the Frontier Safety Framework

다. 메타와 환각과 딥페이크 등 AI 이슈 대응¹⁹⁴⁾

구글은 2024년 7월 31일 딥페이크(Deepfake) 문제 해결을 위해 검색 결과에서 딥페이크를 더 쉽게 삭제할 수 있는 프로세스와 함께, 검색 결과에서 딥페이크 콘텐츠가 상위에 표시되지 않도록 순위 시스템 업데이트를 발표했다. 사용자가 본인이 등장하는 딥페이크 이미지를 구글 검색 결과에서 발견하고 삭제를 요청하면 구글 시스템은 해당 사용자 관련 유사한 검색에서 모든 노골적인 결과물을 필터링하며, 정책에 따라 검색 결과에서 관련 이미지가 삭제되면 해당 이미지의 모든 복제본도 스캔하여 삭제한다. 검색 결과에서 딥페이크 콘텐츠 노출을 줄이는 순위 시스템 업데이트를 통해 특정 딥페이크 콘텐츠나 사람 이름을 포함하는 검색어를 입력한 경우, 관련 뉴스 기사 등 고품질 콘텐츠를 우선 노출하도록 구현할 방침이다. 허위 음란물 이미지로 인해 삭제된 횟수가 많은 사이트의 경우, 사이트의 검색 순위를 낮추어 사용자 노출을 줄일 계획이다.

194) Google, How we're addressing explicit fake content in Search, 2024.07.31.

메타(Meta) 역시 최근 자사 AI 챗봇이 도널드 트럼프(Donald Trump) 전 대통령에 대한 암살 시도를 부정하는 답변을 생성해 논란이 되자 이를 해명하며 기능 개선을 약속¹⁹⁵⁾했다. 메타 AI는 2024년 7월 발생한 트럼프 전 대통령의 총기 피습 사건 이후 일부 사용자에게 해당 사건이 발생하지 않았다는 잘못된 답변을 제공하며 논란을 불러왔다. 메타는 메타 AI가 오류를 나타낸 이유가 생성 AI에서 일반적으로 나타나는 환각(Hallucination) 현상 때문이라고 설명했다.

메타 AI는 총격 직후 촬영된 트럼프 사진에 사실 확인 라벨을 잘못 적용하는 오류가 나타났다, 앞서 알고리즘이 조작된 이미지에 사실 확인 라벨을 추가하면서 이와 유사한 실제 이미지에도 라벨이 잘못 적용되었다고 설명했다. 메타는 모든 생성 AI 시스템과 마찬가지로 자사의 모델도 부정확하거나 부적절한 결과를 생성할 수 있다며, 문제를 지속적으로 해결하고 더 많은 피드백을 수렴해 기능을 개선해 나갈 계획이라고 밝혔다.

라. 생성 AI의 오용 현황 분석¹⁹⁶⁾

구글 딥마인드가 2024년 8월 2일 생성 AI의 오용에 관한 언론 기사를 분석해 오용 유형과 목적을 분석하고 대응 방안을 모색한 논문¹⁹⁷⁾을 발표했다. 구글 딥마인드가 유해 콘텐츠 대응을 위해 설립된 구글 자회사 직쏘(Jigsaw) 및 구글의 자선 부문 구글 닷오알지(Google.org)와 공동 진행한 이번 연구는 2023년 1월부터 2024년 3월까지 보도된 200개의 생성 AI 오용 관련 기사를 수집 및 분석했다.

분석 결과, 생성 AI의 오용은 크게 생성 AI 기능의 악용과 생성 AI 시스템 침해라는 2개 범주에 집중된 것으로 나타났다. 이 가운데 생성 AI 기능의 악용 사례는 AI로 생성된 이미지를 이용한 유명인 사칭이 대표적이며, 생성 AI 시스템 침해 사례는 모델 안전장치를 우회하는 ‘탈옥’ 과 오작동을 유발하는 적대적 프롬프팅¹⁹⁸⁾이 일반적이다.

쉽게 접근할 수 있는 소비자용 생성 AI 도구의 악용 사례가 가장 흔했으며, 실제 2024년 2월 홍콩에 있는 다국적 기업의 재무 담당자는 AI로 합성된 화상 회의에 속아 2억 홍콩달

195) SiliconAngle, Meta faces AI accuracy issues as tech industry tackles hallucinations, deepfakes, 2024.07.31.

196) Google Deepmind, Mapping the misuse of generative AI, 2024.08.02.

197) Generative AI Misuse: A Taxonomy of Tactics and Insights from Real-World Data

198) Adversarial Prompting: 악의적 의도로 프롬프트를 주입하여 목표를 변환하거나 기밀정보를 유출하려는 공격 방식

리(한화 약 340억 원)를 송금하기도 했다. 대다수의 생성 AI 오용 사례는 여론 조작, 수익 창출, 사기를 주목적으로 해 연구진은 악성 행위자들의 오용 목적과 이들이 사용하는 오용 유형을 매칭한 결과를 발표했다. 여론 조작을 위한 오용 유형은 허위 정보와 명예훼손이 일반적이며, 수익 창출을 위한 오용은 딥페이크 상품화와 콘텐츠 대량 제작(Content Farming), 사기 목적의 오용은 유명한 사칭 광고와 피싱 사기가 주류를 이룬다.

연구진은 이번 연구 결과를 바탕으로 생성 AI 오용으로부터 대중을 보호하려면 AI 문해력 캠페인이나 악성 행위자를 억제할 개입 방안 수립 등을 통해 오용을 사전에 방지해야 한다고 강조한다. 이와 관련, 구글은 AI 생성물을 탐지할 수 있는 신스ID(SynthID)와 같은 도구를 개발하는 등 레드팀 활동을 발전시키기 위한 연구를 진행한다. 직소는 허위 정보를 이용한 여론 조작 시도를 경고하는 캠페인 동영상 제작해 이러한 캠페인이 허위 정보에 대응한 복원력 향상에 효과적임을 입증했다.

마. C2PA 표준 적용으로 AI 생성물의 투명성 향상 추진¹⁹⁹⁾

구글이 2024년 9월 17일 AI 생성물의 출처 확인을 통한 투명성 향상을 위한 ‘콘텐츠 출처 및 진위 확인 연합(C2PA)’²⁰⁰⁾ 관련 활동을 공개했다. 구글은 다양한 플랫폼을 넘나드는 콘텐츠의 특성을 고려할 때 온라인 투명성을 높이려면 업계 전반의 협력이 필수적이라는 인식하에 2024년 2월 C2PA의 운영위원회 회원으로 가입했다.

구글은 C2PA 운영위원회에 소속된 마이크로소프트, 어도비(Adobe), 메타 등과 협력해 ‘콘텐츠 자격증명(Content Credentials)’²⁰¹⁾ 기술 표준의 최신 버전 2.1 버전 개발을 추진하고 있다. ‘콘텐츠 자격증명’ 2.1 버전은 콘텐츠 출처 이력을 검증하기 위한 기술적 요구 사항을 강화해 다양한 위변조 공격에 대응한 보안 향상이 가능하다.

구글은 AI 생성물 출처 확인을 위해 C2PA의 최신 기술 표준을 자사 서비스에 적용 계획이다. 향후 몇 달 동안 ‘콘텐츠 자격증명’ 2.1 버전을 검색과 광고 등 주요 서비스에 적

199) Google, How we're increasing transparency for gen AI content with the C2PA, 2024.09.17.

200) 온라인 콘텐츠에 대한 신뢰 구축을 위해 2021년 설립된 글로벌 연합체로 콘텐츠 출처 기술 표준을 개발해 출처 인증에 활용

201) 디지털 콘텐츠에 메타데이터를 포함하여 제작 출처와 관련된 정보를 확인할 수 있게 하는 기술 표준으로, 메타데이터는 다른 데이터를 정의하고 기술하는 데이터를 의미

용할 방침이다. 이미지에 C2PA 메타데이터가 포함된 경우, 사용자는 구글의 이미지 관련 서비스에서 제공하는 ‘About this image’ 기능을 통해 이미지가 AI 도구로 생성 또는 편집되었는지 확인할 수 있다. 구글은 광고 시스템에도 C2PA 메타데이터의 도입을 확대하는 한편, C2PA를 활용해 주요 정책의 시행 방법을 안내할 계획이다. 구글은 유튜브에서도 C2PA로 제공되는 출처 정보를 사용자에게 전달할 방법을 모색 중으로, 2024년 말 관련 업데이트를 진행할 계획이다.

향후 발표될 ‘C2PA 신뢰 목록(Trust List)’을 통한 콘텐츠 출처 확인도 지원할 계획으로, 이용자들이 특정 카메라 모델로 촬영된 이미지의 경우 신뢰 목록을 통해 해당 출처 정보의 정확성을 검증할 수 있도록 지원할 방침이다. 구글은 온라인 콘텐츠 출처 확인을 위해 더 많은 서비스 및 하드웨어 업체의 C2PA 표준 채택을 촉구하는 한편, 구글 딥마인드(Deepmind)에서 개발한 워터마킹 기술 ‘신스ID(SynthID)’를 여러 AI 도구 및 미디어에 확대 적용함으로써 자체적인 노력도 추진한다.

2. 마이크로소프트

1) 기업 일반현황

〈표 5-2〉 마이크로소프트 기업 현황 (단위: M\$)

구분	주요 내용																																										
기업 개요	<p style="text-align: center;">〈회사 개요〉</p> <ul style="list-style-type: none"> • 1975년 4월 미국 뉴멕시코주 앨버커키에서 설립 • 1986년 3월 상장 • 전 세계에 서비스를 제공하며 190개국에서 수백 개 지사를 운영 • 2022년 매출 기준 세계 최대 소프트웨어 기업이며, 2022년 총수익 기준 포춘 500대 미국 기업 순위에서 14위를 차지 																																										
	<p style="text-align: center;">〈사업 내용〉</p> <ul style="list-style-type: none"> • 산업 분야 : 클라우드 컴퓨팅, 소프트웨어, 보안 산업, 전자제품 하드웨어 • 서비스 : 클라우드 컴퓨팅 서비스, 디지털 전환 소프트웨어(Dynamics 365), 워드, 파워포인트, 엑셀, 팀즈, 원노트, 액세스 등 사무용 소프트웨어(Office 365), 링크드인(Linkedin), 윈도우(Windows) 컴퓨터 운영체제 소프트웨어, Xbox 게임 패스, Bing(빙) 검색엔진, 마이크로소프트 뉴스, 엣지(Edge) 브라우저 등 																																										
	<p style="text-align: center;">〈자회사〉</p> <ul style="list-style-type: none"> • 링크드인, 뉘앙스 커뮤니케이션스, 깃허브, 스카이프, 모장(Mojang) 스튜디오, 엑스박스(Xbox) 게임 스튜디오, 제니맥스 스튜디오 																																										
재무 실적 ²⁰²⁾ (최근 5년)	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th>매감 기준:</th> <th>2020 30/06</th> <th>2021 30/06</th> <th>2022 30/06</th> <th>2023 30/06</th> <th>2024 30/06</th> </tr> </thead> <tbody> <tr> <td>총매출</td> <td>143,015</td> <td>168,088</td> <td>198,270</td> <td>211,915</td> <td>245,122</td> </tr> <tr> <td>총매출 성장률</td> <td>+13.65%</td> <td>+17.53%</td> <td>+17.96%</td> <td>+6.88%</td> <td>+15.67%</td> </tr> <tr> <td>매출 원가</td> <td>46,078</td> <td>52,232</td> <td>62,650</td> <td>65,711</td> <td>74,114</td> </tr> <tr> <td>총이익</td> <td>96,937</td> <td>115,856</td> <td>135,620</td> <td>146,204</td> <td>171,008</td> </tr> <tr> <td>총이익 성장률</td> <td>+16.89%</td> <td>+19.52%</td> <td>+17.06%</td> <td>+7.8%</td> <td>+16.97%</td> </tr> <tr> <td>총이익 %</td> <td>67.78%</td> <td>68.93%</td> <td>68.4%</td> <td>68.99%</td> <td>69.76%</td> </tr> </tbody> </table>	매감 기준:	2020 30/06	2021 30/06	2022 30/06	2023 30/06	2024 30/06	총매출	143,015	168,088	198,270	211,915	245,122	총매출 성장률	+13.65%	+17.53%	+17.96%	+6.88%	+15.67%	매출 원가	46,078	52,232	62,650	65,711	74,114	총이익	96,937	115,856	135,620	146,204	171,008	총이익 성장률	+16.89%	+19.52%	+17.06%	+7.8%	+16.97%	총이익 %	67.78%	68.93%	68.4%	68.99%	69.76%
매감 기준:	2020 30/06	2021 30/06	2022 30/06	2023 30/06	2024 30/06																																						
총매출	143,015	168,088	198,270	211,915	245,122																																						
총매출 성장률	+13.65%	+17.53%	+17.96%	+6.88%	+15.67%																																						
매출 원가	46,078	52,232	62,650	65,711	74,114																																						
총이익	96,937	115,856	135,620	146,204	171,008																																						
총이익 성장률	+16.89%	+19.52%	+17.06%	+7.8%	+16.97%																																						
총이익 %	67.78%	68.93%	68.4%	68.99%	69.76%																																						

구분	주요 내용
AI 조직 현황	<p style="text-align: center;"><AI 관련 조직></p> <ul style="list-style-type: none"> • Microsoft Research(MSR) : MS의 핵심 연구조직으로 기초 AI 연구부터 응용 연구까지 수행하며 전 세계 각국에서 여러 연구소를 운영 • 애저 AI(Azure AI) : 클라우드 기반 AI 서비스 개발하며 기업용 AI 솔루션 제공, AI 인프라 구축 및 운영, 오픈AI(OpenAI) 서비스 통합 등을 진행 • AI 플랫폼 팀 : AI 개발도구 및 프레임워크 개발하며, 개발자 생태계 지원과 AI 서비스 통합 플랫폼 구축에 주력 • AI & 리서치 그룹 : AI 전략 수립 및 실행을 담당하며, 제품 통합 AI 개발, 새로운 AI 기술 탐색 등의 역할을 수행 • MS의 AI 조직은 연구부터 제품화까지 체계적인 시스템을 갖추고 있으며, 특히 오픈AI와의 전략적 파트너십을 통해 AI 기술 주도권 확보에 주력 • 기업용 AI 솔루션과 개발자 도구에서 강점을 보이며, 책임 있는 AI 개발에도 관심을 집중
사업 전망	<p style="text-align: center;"><AI 전략></p> <ul style="list-style-type: none"> • 오픈AI와 파트너십 강화 : 100억 달러 이상의 금액과 애저 인프라를 제공해 GPT-4 독점 사용권을 확보했고, 이를 기반으로 한 공동 AI 개발 및 상용화를 추진 • 코파일럿(Copilot) 전략 : Microsoft 365 코파일럿, 윈도 코파일럿, 깃허브 코파일럿, 시큐리티 코파일럿, 세일즈 코파일럿 등으로 확장을 추진 • 애저 AI 플랫폼 강화 : 애저 오픈AI 서비스, AI 인프라 서비스 확대, 기업용 AI 솔루션 제공, AI 개발도구 생태계 구축 등에 중점 • 제품 통합 전략 : 모든 제품에 AI 기능 통합, Bing과 엣지에 AI 검색 도입, Microsoft 365 앱 AI 강화, 팀스(Teams) AI 기능 확대 등을 추진 • MS의 AI 전략은 ▲오픈AI와의 파트너십을 통한 기술 확보, ▲코파일럿을 통한 제품 혁신, ▲애저를 통한 인프라 제공이라는 세 축을 중심으로 전개

2) 2024년 기업 동향

(1) 사업전략/투자/M&A

가. MS, 프랑스 AI 인프라에 40억 유로 투자 및 AI 교육 지원 계획²⁰³⁾

마이크로소프트(MS)가 2024년 5월 13일 프랑스의 AI 및 클라우드 인프라 구축에 40억 유로를 투자한다고 발표했다. 이는 MS가 프랑스에 진출한 이래 역대 최대 규모의 투자로, 프랑스의 AI와 클라우드 도입을 촉진하고 디지털 혁신과 경제 성장을 강화하기 위한 목적이라고 밝혔다. MS는 2025년 말까지 프랑스의 클라우드와 AI 인프라를 확장해 최대 2만 5,000

202) <https://kr.investing.com/equities/microsoft-corp-income-statement>

203) Microsoft, Microsoft announces the largest investment to date in France to accelerate the adoption of AI, skilling and innovation, 2024.05.13.

개의 최첨단 GPU를 도입할 예정으로, 파리와 마르세유 지역의 기존 데이터센터 공간을 확장하는 한편, 프랑스 동부의 알자스 지역에 신규 데이터센터를 구축할 계획이다.

MS는 2027년까지 100만 명 대상 AI 기술 교육 및 스타트업 생태계를 지원할 계획이다. MS는 인프라 투자와 함께 2027년까지 구직자와 학생, 중소기업 근로자 등 총 100만 명에게 AI 기술 교육을 제공할 계획이다. MS는 프랑스 정부 기관, 교육업체, 비영리 단체, 고등교육 기관과 협력해 AI 리터러시와 역량 향상, AI 비즈니스 혁신 지원과 안전하고 책임 있는 AI 개발에 중점을 두고 새로운 교육 프로그램을 출시할 예정이다. 프랑스 노동청과 협력해 구직자에게 AI의 핵심 개념과 생성 AI의 기능 등 일상적인 사용이나 구직 목적으로 AI 사용에 필요한 기술 교육 등을 지원할 계획이다.

또한 AI 전문가 역량 향상을 위해 다양한 직위를 대상으로 생성 AI 교육과정을 제공하는 ‘엔터프라이즈 스킬 이니셔티브’ 프로그램을 통해 프랑스의 여러 교육 서비스 기업과 협력할 방침이다. MS는 스타트업 생태계 지원을 위해 2027년까지 프랑스 스타트업을 대상으로 AI 전문지식과 클라우드 사용, 협업을 지원하는 ‘생성 AI 스튜디오(GenAI Studio)’ 프로그램을 공개했다. 3년에 걸쳐 연 2회 실시되는 ‘생성 AI 스튜디오’ 프로그램은 회차별로 15개 스타트업에 기술 워크숍과 MS의 AI 전문가 협업을 통해 생성 AI 이해와 혁신적 활용 사례 도출을 지원할 방침이다.

(2) 기술개발 및 혁신

가. 사진 한 장으로 말하는 얼굴 생성하는 AI 모델 연구²⁰⁴⁾

MS가 2024년 4월 16일 얼굴 사진과 음성을 기반으로 사람이 실제로 말하는 것처럼 보이는 영상을 생성하는 AI 모델 ‘바사(VASA)-1’을 공개했다. 이 모델은 음성에 맞게 입 모양을 동기화하여 실제 사람이 말하거나 노래하는 것처럼 보이는 영상을 생성한다. 연구진은 약 6,000명의 사람 얼굴 영상을 학습 데이터로 사용해 자연스러운 표정과 입 모양, 눈동자 움직임을 구현했다고 설명했다. 바사-1은 512x512 해상도의 동영상을 초당 45프레임의 속도로 생성하며, 온라인 스트리밍 모드에서는 초당 최대 40프레임을 지원한다.

204) Microsoft, VASA-1: Lifelike Audio-Driven Talking Faces Generated in Real Time, 2024.04.16.

MS는 이번 연구가 가상 AI 아바타를 대상으로 시각적인 감정 표현을 지원하는 기술의 생성에 중점을 두고 있다고 강조했다. MS는 VASA-1이 사용자를 속이거나 오도하기 위한 목적이 아니라고 설명했다. MS는 VASA-1이 교육 형평성 제고, 의사소통에 어려움을 겪는 사람들이나 기타 도움이 필요한 사람들을 위한 접근성 개선과 치료 지원 등 긍정적으로 활용될 수 있다며 추가 연구의 중요성도 강조했다.

[그림 5-5] 마이크로소프트 ‘바사-1’의 결과물 예시



출처: VASA-1: Lifelike Audio-Driven Talking Faces Generated in Real Time

(3) 제품 및 서비스

가. 프랑스 오픈소스 소형언어모델 ‘파이 3.5’ 공개²⁰⁵⁾

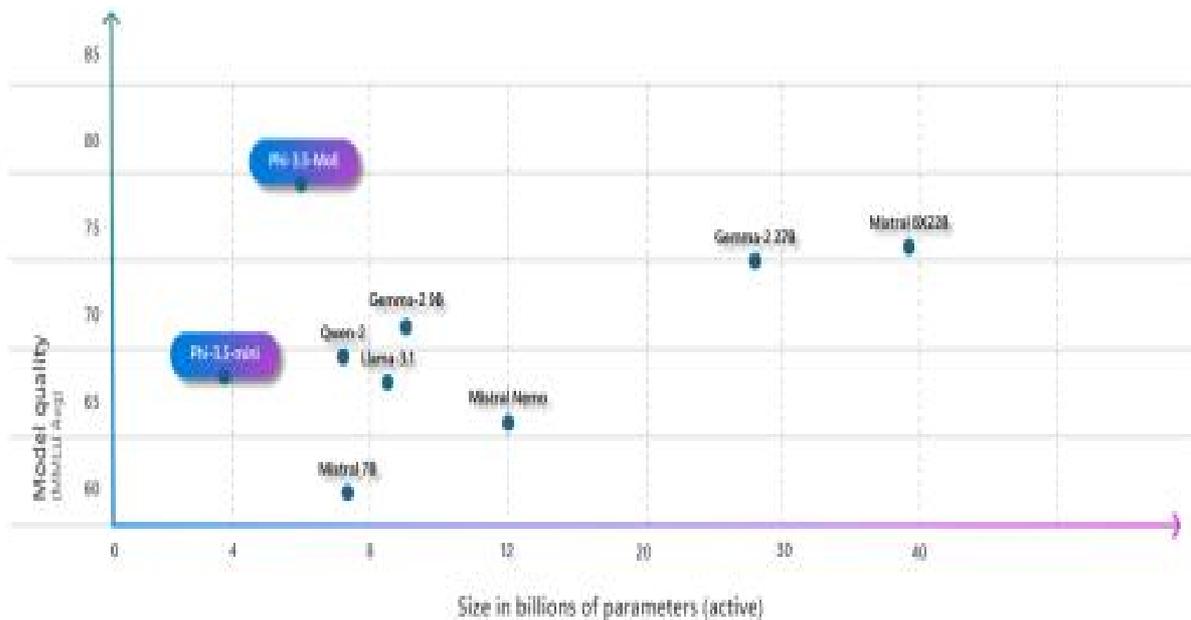
MS가 2024년 8월 22일 오픈소스 소형언어모델(SLM) ‘파이(Phi)-3.5’ 3종을 허깅페이스(Hugging Face)를 통해 공개했다. 공개 모델은 파이-3.5-미니, 파이-3.5- MoE(전문가혼합), 파이-3.5-비전으로, 각각 빠른 추론과 복잡한 추론, 시각 작업에 특화되어 있다고 강조했다. 컴퓨팅 자원이 제한된 환경에 적합한 파이-3.5 미니는 38억 개 매개변수로 학습된 SLM으로 12만 8천 개 토큰의 컨텍스트 창과 다국어 지원을 지원하며, 매개변수가 더 큰 미스트랄-7B보다 성능이 뛰어나다는 평가이다.

205) Microsoft, Discover the New Multi-Lingual, High-Quality Phi-3.5 SLMs, 2024.08.22.

16개 전문가 모델을 혼합한 파이-3.5-MoE는 주어진 요청에 필요한 전문가 모델만 처리를 담당하므로 전체 420억 개 매개변수 중 66억 개 매개변수만 활성화하여 효율성을 높이면서 뛰어난 성능을 발휘한다. 20개 이상의 다국어와 최대 12만 8천 개의 컨텍스트 창을 지원한다. 파이-3.5-비전은 다중 프레임 이미지 이해와 추론 기능을 갖추어 이미지 간 비교나 다중 이미지와 동영상 요약 등을 지원한다. 하지만 다국어 작업에는 아직 최적화되지 않아 추가 미세조정 필요하다는 평가이다.

파이-3.5-MoE는 매개변수가 더 큰 모델과 비슷한 수준의 추론 능력을 갖췄으며, 다국어 작업에서도 훨씬 큰 모델과 비슷한 경쟁력을 발휘한다. 실제 MMLU²⁰⁶⁾ 벤치마크에서 파이-3.5-MoE는 78.9점으로 오픈AI의 GPT-4o-미니(77.2점)를 앞서고, 전체 벤치마크 평균에서 69.2점으로 구글의 제미나이 1.5 플래시(68.5점), 젤마-2-9b(63.3점), 미스트랄-네모-12B(61.3점)를 앞섰다.

[그림 5-6] ‘파이 3.5’ 와 주요 소형언어모델(SLM)의 매개변수 크기, 품질 비교



출처: Microsoft, Discover the New Multi-Lingual, High-Quality Phi-3.5 SLMs

206) 다양한 주제에 대한 모델의 광범위한 지식과 추론 능력을 평가하는 벤치마크

3. 메타 플랫폼스

1) 기업 일반현황

〈표 5-3〉 메타 플랫폼스 기업 현황 (단위: M\$)

구분	주요 내용																																																
기업 개요	<p style="text-align: center;">〈회사 개요〉</p> <ul style="list-style-type: none"> • 2004년 2월 미국 매사추세츠주에서 페이스북으로 설립 • 2012년 5월 나스닥 상장 • 2021년 10월 메타버스 시장에 집중한다는 의미를 담아 사명을 메타플랫폼스로 변경 • 2023년 포브스 글로벌 2000에서 31위를 기록했고, 2022년 R&D 지출이 353억 달러로 전 세계 3위를 기록 																																																
	<p style="text-align: center;">〈사업 내용〉</p> <ul style="list-style-type: none"> • 산업 분야 : 소셜 미디어 및 소셜 네트워킹, 사진 및 비디오 공유 네트워킹, 텍스트 기반 소셜 네트워킹, 인스턴트 메시징 및 VoIP 서비스, 가상현실/증강현실, 모바일 네트워크 • 서비스 : 페이스북, 인스타그램, 스레드, 왓츠앱, 메신저, 메타 퀘스트, 호라이즌 월드, 레이반 메타 스마트 글라스, 맵피라리(Mapillary) 																																																
	<p style="text-align: center;">〈관계사〉</p> <ul style="list-style-type: none"> • 오쿨러스(리얼리티랩스), 맵피라리, CTRL-Labs, 지오 플랫폼스(Jio Platforms), 노비 파이낸셜(Novi Financial) 																																																
재무 실적 ²⁰⁷⁾ (최근 5년)	<table border="1" style="width: 100%; text-align: center;"> <thead> <tr> <th></th> <th>2019 31/12</th> <th>2020 31/12</th> <th>2021 31/12</th> <th>2022 31/12</th> <th>2023 31/12</th> </tr> </thead> <tbody> <tr> <td>마감 기준:</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>총매출</td> <td>70,697</td> <td>85,965</td> <td>117,929</td> <td>116,609</td> <td>134,902</td> </tr> <tr> <td>총매출 성장률</td> <td>+26.61%</td> <td>+21.6%</td> <td>+37.18%</td> <td>-1.12%</td> <td>+15.69%</td> </tr> <tr> <td>매출 원가</td> <td>12,770</td> <td>16,692</td> <td>22,649</td> <td>23,754</td> <td>26,006</td> </tr> <tr> <td>총이익</td> <td>57,927</td> <td>69,273</td> <td>95,280</td> <td>92,855</td> <td>108,896</td> </tr> <tr> <td>총이익 성장률</td> <td>+24.62%</td> <td>+19.59%</td> <td>+37.54%</td> <td>-2.55%</td> <td>+17.28%</td> </tr> <tr> <td>총이익 %</td> <td>81.94%</td> <td>80.58%</td> <td>80.79%</td> <td>79.63%</td> <td>80.72%</td> </tr> </tbody> </table>		2019 31/12	2020 31/12	2021 31/12	2022 31/12	2023 31/12	마감 기준:						총매출	70,697	85,965	117,929	116,609	134,902	총매출 성장률	+26.61%	+21.6%	+37.18%	-1.12%	+15.69%	매출 원가	12,770	16,692	22,649	23,754	26,006	총이익	57,927	69,273	95,280	92,855	108,896	총이익 성장률	+24.62%	+19.59%	+37.54%	-2.55%	+17.28%	총이익 %	81.94%	80.58%	80.79%	79.63%	80.72%
	2019 31/12	2020 31/12	2021 31/12	2022 31/12	2023 31/12																																												
마감 기준:																																																	
총매출	70,697	85,965	117,929	116,609	134,902																																												
총매출 성장률	+26.61%	+21.6%	+37.18%	-1.12%	+15.69%																																												
매출 원가	12,770	16,692	22,649	23,754	26,006																																												
총이익	57,927	69,273	95,280	92,855	108,896																																												
총이익 성장률	+24.62%	+19.59%	+37.54%	-2.55%	+17.28%																																												
총이익 %	81.94%	80.58%	80.79%	79.63%	80.72%																																												

구분	주요 내용
AI 조직 현황	<p style="text-align: center;">〈AI 관련 조직〉</p> <ul style="list-style-type: none"> • 메타 AI(구 FAIR-Facebook AI Research) : 메타의 핵심 AI 연구조직으로 기초 AI 연구를 수행하고, 전 세계에서 연구소를 운영 • AI Research & Development : 제품 적용을 위한 AI 개발, 소셜 미디어 AI 통합, 메타버스 AI 기술개발, AI 인프라 구축 등을 수행 • Applied AI : AI 기술의 제품 통합, 추천 시스템 개발, 콘텐츠 모더레이션, 광고 최적화 등의 업무를 수행 • 리얼리티랩스(Reality Labs) : AR/VR 관련 AI 연구, 메타버스 AI 기술, 음성/영상 인식 기술, 인터랙션 AI 연구 등을 진행 • 메타의 AI 연구는 대규모 언어모델 개발, 멀티모달 AI 강화, 효율적 AI 연구에 집중해 AI 연구 투자 확대와 함께 메타버스 AI 강화와 오픈소스 전략 강화가 나타남
사업 전망	<p style="text-align: center;">〈AI 전략〉</p> <ul style="list-style-type: none"> • 오픈소스 중심 전략 : LLaMA 2 무료 공개, 파이토치(PyTorch) 생태계 확장, 연구 결과 공유, 커뮤니티 협력 강화 등을 지속적으로 추진 • 메타버스 AI 전략 : AR/VR AI 기술개발, 아바타 기술 고도화, 실감형 상호작용 구현, 가상 환경 지능화에 집중 • 소셜 미디어 AI 강화 : 콘텐츠 추천 고도화, 자동 번역/자막, 콘텐츠 모더레이션, 광고 최적화에 주력 • AI 인프라 확충 : AI Research SuperCluster 확장, 전용 칩 개발, 분산 학습 시스템 구축, 에너지 효율성 개선에 주력 • 메타의 AI 전략은 오픈소스 리더십, 메타버스 혁신, 소셜 미디어 고도화라는 세 가지 축을 중심으로 발전하고 있음 • 특히 LLaMA 시리즈와 파이토치를 통한 AI 생태계 확장, 메타버스 실현을 위한 AI 기술 혁신이 주요 동력이 될 것으로 분석되며, 프라이버시 보호와 윤리적 AI 개발이 중요한 과제가 될 것 전망

2) 2024년 기업 동향

(1) 사업전략/투자/M&A

가. 연내 데이터센터에 자체 개발 AI 프로세서 도입 계획²⁰⁸⁾

로이터 통신이 메타의 내부 문서를 바탕으로 메타가 자체 개발한 새로운 AI 프로세서를 올해 안에 데이터센터 탑재를 추진한다고 보도했다. 메타는 2023년 5월 MTIA²⁰⁹⁾라는 자체 설계 칩을 처음 공개했으며, MTIA에 이은 2세대 칩으로 ‘아르테미스(Artemis)’로 지칭되는 신규 칩을 개발하고 있다. 보도 이후 메타의 대변인은 새로운 AI 칩을 2024년 생산해 타

207) <https://kr.investing.com/equities/facebook-inc-income-statement>

208) Reuters, Meta to deploy in-house custom chips this year to power AI drive, 2024.02.02.

209) Meta Training and Inference Accelerator

사의 그래픽 처리장치(GPU) 수십만 개와 함께 투입할 계획이라고 인정했다. 메타는 자체 개발한 프로세서가 특정 워크로드에서 상용 GPU를 보완해 최적의 성능과 효율성을 발휘할 것으로 기대한다고 설명했다.

아르테미스는 AI 모델이 알고리즘을 이용해 순위를 결정하고 프롬프트 응답을 생성하게 하는 추론만 지원하므로, 훈련과 추론을 모두 지원하는 상용 GPU와 병행 사용이 불가피한 상황이다. 최근 마크 저커버그(Mark Zuckerberg) 메타 CEO는 올해 말까지 엔비디아의 주력 제품인 H100 35만 개를 포함해 총 60만 개의 H100급 컴퓨팅 용량을 확보하겠다고 밝힌 바 있다. 메타는 아르테미스를 통해 AI 반도체 시장을 장악한 엔비디아에 대한 의존도를 줄이고 AI 제품 출시에 들어가는 막대한 비용을 절감할 수 있을 것으로 기대한다.

메타는 페이스북, 인스타그램과 같은 주요 서비스의 AI 기능을 강화하고 레이밴(Ray-Ban) 스마트 글래스 출시 등으로 AI 사업을 강화하고 있다. 반도체 전문 컨설팅 기업 세미애널리시스(SemiAnalysis)의 수석 분석가 딜런 파텔(Dylan Patel)은 메타가 자체 개발한 칩을 사용할 경우, 연간 수억 달러의 에너지 비용과 수십억 달러의 칩 구매 비용을 절감할 수 있을 것으로 분석했다. 파텔 분석가는 아르테미스는 추론만 지원하는 한계에도 불구하고 에너지를 대량으로 소모하는 엔비디아의 프로세서보다 메타의 AI 수요 대응에 훨씬 효율적일 수 있다고 강조했다.

나. AI 모델 ‘라마’의 활발한 생태계 형성²¹⁰⁾

메타가 2023년 2월 출시한 AI 모델 ‘라마(Llama)’의 다운로드 수가 3억 5천만 건에 달하며 2024년 기준 다운로드 수가 전년과 비교해 10배 이상 증가했다고 발표했다. 라마는 3.1 버전이 출시된 2024년 7월 한 달 동안 허깅페이스(Hugging Face)²¹¹⁾에서 2천만 건 이상의 다운로드를 기록해 오픈소스 모델 제품군 중 선두를 차지했다.

AWS, 마이크로소프트(Microsoft) 등의 대형 클라우드 제공업체를 통한 라마의 월간 사용량은 2024년 1월~7월까지 10배 증가했으며, 2024년 8월 한 달 동안 가장 사용자 수가 많았던 버전은 ‘라마 3.1-405B’로 최대 규모 AI 모델의 인기를 확인했다. 자체 활용 목적에 맞게

210) Meta, With 10x growth since 2023, Llama is the leading engine of AI innovation, 2024.08.29.

211) 머신러닝 모델을 구축, 배포, 훈련할 수 있도록 다양한 도구와 라이브러리를 제공하는 플랫폼

라마를 미세 조정하는 개발자 커뮤니티가 활성화되며, 허깅페이스에는 6만 개 이상의 라마 파생 모델이 존재한다.

메타는 라마의 성공이 오픈소스에 기인한다며, 개발자의 선택권과 역량을 보장함으로써 AI 생태계가 활성화되고 광범위하고 빠른 혁신이 가능해졌다고 강조했다. 메타는 공식 사이트에 라마를 미세 조정해 자체적으로 활용하는 주요 기업의 사례를 소개하고, 액센추어, AT&T 등 주요 기업들, 라마를 자체 목적에 맞게 미세 조정해 활용하고 있다고 밝혔다.

액센추어(Accenture)는 라마 3.1을 사용해 ESG(환경·사회·지배구조) 보고서를 생성하는 맞춤형 LLM을 구축하고 있으며, 기존 작성 방식 대비 생산성은 70%, 품질은 20~30% 향상을 기대하는 것으로 알려졌다. AT&T는 라마를 미세 조정하여 핵심 트렌드 및 고객 요구 사항과 고객 경험 개선에 활용해 효율적인 고객 관리를 구현할 계획이다. 배달 대행 플랫폼 도어대시(DoorDash)는 라마를 사용해 내부 지식기반의 복잡한 질문 응답, 코드 베이스 개선 등 소프트웨어 개발자의 일상 업무를 간소화했다.

또한 일본 금융기업 노무라(Nomura)는 텍스트 요약, 편향 방지, 코드 생성, 로그 분석 등 문서 관련 업무 전반에서 AWS 기반 라마를 활용하고 있다. 줌(Zoom)은 자체 모델과 폐쇄형 및 오픈소스 LLM(라마 포함)을 활용해 회의 요약, 스마트 녹음 등 반복적인 일상 업무를 대신 처리하는 AI컴패니언(Companion) 서비스를 사용자에게 제공하고 있다.

(2) 제품 및 서비스

가. 오픈소스 LLM ‘라마’의 최신 버전 ‘라마 3’ 공개²¹²⁾

메타가 2024년 4월 18일 차세대 LLM ‘라마(Llama) 3’의 3개 모델 중 매개변수 80억 개와 700억 개의 모델을 오픈소스로 공개²¹³⁾했으며, 추후 4,000억 개 매개변수 모델도 공개할 방침이다. 이번에 공개된 모델은 텍스트 기반으로, 메타는 향후 라마 3에서 다국어와 멀티모달을 지원하고, 컨텍스트 창을 확대하며 추론과 코딩 등 전반적인 LLM 기능을 지속적으로 향상할 계획이다.

212) Meta, Introducing Meta Llama 3: The most capable openly available LLM to date, 2024.04.18.

213) Meta, Meet Your New Assistant: Meta AI, Built With Llama 3, 2024.04.18.

라마 3는 사전 훈련과 사후 미세조정 과정을 고도화하여 라마 2보다 성능이 대폭 개선되었으며, 동급 매개변수를 갖는 경쟁 모델의 성능을 능가하는 것으로 알려졌다. 라마 3는 라마 2보다 7배 더 큰 15조 개 이상의 토큰으로 훈련되었으며, 일상적 질문부터 과학·기술·공학·수학(STEM), 코딩, 역사 지식 등 다양한 분야 데이터셋을 사용해 미세조정을 거쳤다. 라마 3 70B는 객관식 문제(MMLU)와 코딩(HumanEval) 성능 벤치마크에서 구글의 ‘제미나이 프로 1.5’와 앤스로픽의 ‘클로드 3 소네트(Sonnet)’를 능가했으나, 수학 단어 문제(MATH) 해결이나 대학원생 수준의 객관식 문제(GPQA)에서는 제미나이 프로 1.5가 더 우수한 것으로 나타났다.

라마 3 70B는 조언 요청, 브레인스토밍, 분류, 코딩 등 12개 주요 사용 사례를 포함하는 데이터셋에 대한 인간 선호도 평가에서는 GPT-3.5와 클로드 3 소네트, 미스트랄을 모두 앞섰다. 메타는 라마 3 공개와 함께 라마 3 기반의 AI 챗봇 ‘메타 AI’를 페이스북, 인스타그램, 왓츠앱, 메신저 등 자사 SNS 서비스에 출시한다고 발표했다.

현재 메타 AI는 미국 외 호주, 캐나다, 뉴질랜드, 싱가포르 등 13개 국가에서 제공되며 조만간 유럽 등 기타 국가로 서비스를 확대할 예정이다. 메타 AI는 SNS 플랫폼 검색창에 내장되어 이용자가 쉽게 질문하고 답변을 받을 수 있으며, 사용자가 SNS 서비스를 통하지 않고 AI 챗봇을 이용할 수 있는 독립 웹사이트(Meta.ai)도 출시했다.

나. 폐쇄형 첨단 AI 모델과 대등한 성능의 오픈소스 모델 ‘라마 3.1’ 공개²¹⁴⁾

메타가 2024년 7월 23일 매개변수 80억 개, 700억 개, 4,050억 개의 3개 버전으로 구성된 최신 오픈소스 모델 ‘라마(Llama) 3.1’를 공개했다. 라마 3.1은 12만 8천 개 토큰의 컨텍스트 창을 지원하며 영어, 독일어, 프랑스, 이탈리아어, 포르투갈어, 힌디어, 스페인어, 태국어의 8개 언어를 지원한다. 라마 3.0보다 강력한 추론 기능을 갖춘 라마 3.1은 장문 텍스트 요약과 다국어 대화, 코딩 도우미 같은 고급 활용 사례를 지원한다. 특히 최대 매개변수를 가진 라마 3.1 405B는 최고 수준의 폐쇄형 모델과 경쟁할 수 있는 성능을 보유해 기업용 애플리케이션 및 연구개발 용도에 적합하다는 평가다. 라마 3.1 70B는 콘텐츠 제작, 대화형 AI, 언어 이해, 기업용 애플리케이션에, 라마 3.1 8B는 컴퓨팅 파워와 자원이 제한된 활용 사례에 적합하다.

214) Meta, Introducing Llama 3.1: Our most capable models to date, 2024.07.23.

메타는 다양한 언어의 광범위한 벤치마크 데이터셋을 활용해 라마 3.1의 성능을 평가했으며, 실제 사용 시나리오에서 경쟁 모델과 비교하는 인간 평가도 수행했다. 평가 결과, 라마 3.1 405B는 GPT-4, GPT-4o, 클로드 3.5 소네트와 같은 첨단 기반 모델과 대등한 성능을 나타냈으며, 라마 3.1 8B는 젤마(Gemma)2 9B 및 미스트랄(Mistral) 7B와, 70B는 믹스트랄(Mixtral) 8x22B 및 GPT-3.5-터보와 비교해 일반 지식(MMLU, MMLU-Pro)²¹⁵⁾ 등 대부분 벤치마크에서 더 높은 점수를 획득했다. 메타는 16,000개 이상의 엔비디아(Nvidia) H100 GPU가 탑재된 클러스터에서 라마 3.1 405B를 훈련했으며, 아마존, 엔비디아 등 25개 이상의 기업과 협력하여 광범위한 생태계를 구축했다.

[그림 5-7] 라마 3.1 405B와 경쟁 모델의 인간 평가 결과 비교



출처: Meta, Introducing Llama 3.1: Our most capable models to date

(3) AI 윤리 및 규제 대응

가. LLM의 품질과 정확성을 평가하는 ‘자가 학습 평가자’ 개발²¹⁶⁾

메타 FAIR(Fundamental AI Research)가 2024년 8월 8일 LLM을 활용해 자가 학습 방식으로 LLM의 품질과 정확성을 평가하는 ‘자가 학습 평가자’ 관련 논문²¹⁷⁾을 공개했다.²¹⁸⁾ LLM은 평가자 자체로 사용되어 모델 개선에 중요한 역할을 하지만, LLM 평가자 훈련에는 인간이 주석을 단 방대한 데이터가 필요하며 해당 작업은 시간과 비용을 많이 소요된다. 이

215) MMLU는 다양한 주제에 대한 모델의 광범위한 지식과 추론 능력을 평가하며, MMLU를 개선한 MMLU-Pro는 문제 난도를 높이고 추론 중심 문제를 확대

216) Venturebeat, Meta’s Self-Taught Evaluator enables LLMs to create their own training data. 2024.08.19.

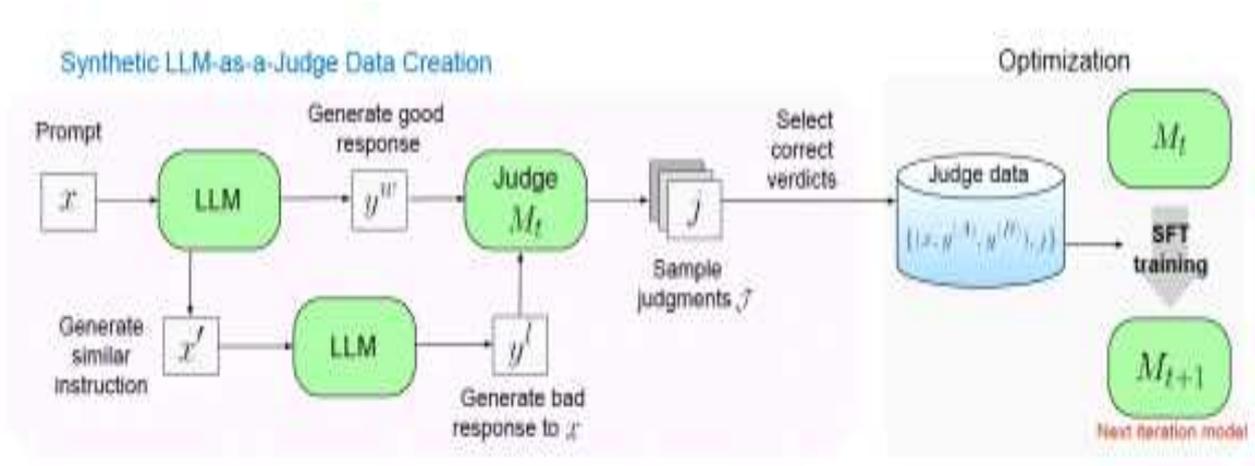
217) Self-Taught Evaluators

218) <https://arxiv.org/pdf/2408.02666>

에 메타 연구진은 합성데이터²¹⁹⁾*를 활용하여 인간이 데이터에 주석을 달 필요 없이 LLM 평가자를 교육하는 방법을 개발했다.

메타의 자가 학습 평가자는 올바른 결과에 도달하는 추론 과정을 생성해 어떤 응답이 더 나은지 판단하는 ‘평가형 LLM(LLM-as-a-Judge)’ 을 기반으로 구축됐다. 먼저 자가 학습 평가자가 프롬프트(Prompt)에 대응해, 한 쌍의 모델 응답을 생성하면 이중 더 뛰어난 응답 (Good Response)을 선택하고 다른 응답(Bad Response)은 거부한다. 이후 반복적 학습을 통해 모델의 추론 과정과 판단을 샘플링한다. 올바른 추론 과정을 생성하면 예제를 학습 데이터에 추가하는 방식으로 마련한 최종 데이터(Judge Data)로 모델을 미세 조정하여 다음번 반복을 위한 업데이트를 진행한다.

[그림 5-8] 메타 FAIR의 자가 학습 평가자 파이프라인



출처: Venturebeat, Meta’s Self-Taught Evaluator enables LLMs to create their own training data

메타의 ‘라마 3-70B-인스트럭트’ 모델로 자가 학습을 통한 전체 답변과 학습 데이터의 생성을 진행한 결과, 자가 학습 평가자는 인간이 주석을 달 데이터로 학습한 모델과 비슷한 수준의 정확도를 구현했다. 리워드벤치(RewardBench)²²⁰⁾ 평가에서 모델 정확도가 인간 개입 없이 5차례 반복 후 정확도가 75.4%에서 88.7%로 크게 개선되었으며, 이는 인간이 주석을 달 데이터로 학습한 모델과 대등한 수준으로 평가된다.

219) 실제 데이터의 특성과 패턴을 모방하여 인공적으로 생성된 데이터

220) 인간 피드백 기반으로 강화 학습된 모델이 인간 선호도와 얼마나 일치하는지를 평가하는 벤치마크

4. 오픈AI

1) 기업 일반현황

〈표 5-4〉 오픈AI 기업 현황 (단위: M\$)

구분	주요 내용															
기업 개요	〈회사 개요〉 <ul style="list-style-type: none"> • 2015년 12월 미국 캘리포니아에서 설립 • 2022년 11월 생성 AI Chat GPT를 출시해 AI 시장의 새로운 붐 조성 • 델라웨어에 등록된 비영리 단체 OpenAI, Inc.와 2019년에 설립된 영리 자회사 OpenAI Global, LLC로 구성 • MS는 오픈AI에 약 130억 달러를 투자해 49% 지분을 소유하고 있고, MS 애저 클라우드 플랫폼을 통해 오픈AI에 컴퓨팅 리소스를 제공 															
	〈사업 내용〉 <ul style="list-style-type: none"> • 산업 분야 : 안전하고 유익한 인공 일반 지능(AGI, artificial general intelligence) 개발을 사명으로 하며, AGI를 가장 경제적으로 가치 있는 작업에서 인간을 능가하는 고도로 자율적인 시스템으로 정의 • 제품 : OpenAI Five, GPT-1, 2, 3, 4, 4o, DALL·E, 오픈AI 코덱스, 챗GPT, 서치 GPT, 소라, o1 															
	〈자회사 및 인수〉 <ul style="list-style-type: none"> • 오픈AI Global, LLC • 2023. 8월 미국 스타트업 Global Illumination 인수 • 2024. 6월 협업 플랫폼 스타트업 Multi 인수 • 2024. 6월 데이터 검색 및 분석 전문 기업 RockSet 인수 															
재무 실적 ²²⁾ (최근 5년)	<table border="1"> <caption>재무 실적 (최근 5년)</caption> <thead> <tr> <th>연도</th> <th>ARR (\$M)</th> <th>Growth (%)</th> </tr> </thead> <tbody> <tr> <td>2021</td> <td>\$36M</td> <td>-</td> </tr> <tr> <td>2022</td> <td>\$200M</td> <td>456%</td> </tr> <tr> <td>2023</td> <td>\$1600M</td> <td>700%</td> </tr> <tr> <td>Sep 2024</td> <td>\$4000M</td> <td>243%</td> </tr> </tbody> </table>	연도	ARR (\$M)	Growth (%)	2021	\$36M	-	2022	\$200M	456%	2023	\$1600M	700%	Sep 2024	\$4000M	243%
연도	ARR (\$M)	Growth (%)														
2021	\$36M	-														
2022	\$200M	456%														
2023	\$1600M	700%														
Sep 2024	\$4000M	243%														
AI 조직 현황	〈AI 관련 조직〉 <ul style="list-style-type: none"> • 랭귀지 모델 팀(Language Models Team) : GPT 시리즈, 챗GPT 개발 및 개선 • 어플라이드 AI(Applied AI)팀 : DALL-E 개발, Whisper 음성 인식 모델 개발 • 얼라인먼트(Alignment) 팀 : AI 안전성 및 윤리 연구, AI 정렬 문제 연구 															
사업 전망	〈주요 제품 및 서비스〉 <ul style="list-style-type: none"> • 챗GPT : 대화형 AI 어시스턴트, 무료 버전과 플러스 버전 운영, GPT-4 기반 고급 기능 제공 • GPT-4 : 고성능 대규모 언어모델, 텍스트와 이미지 입력 처리, API 서비스 제공 • DALL-E : AI 이미지 생성 모델, 텍스트 기반 이미지 생성 															

2) 2024년 기업 동향

(1) 사업전략/투자/M&A

가. 일론 머스크의 소송·정부 조사·저작권 소송 등으로 부담 증가²²²⁾

일론 머스크(Elon Musk) 테슬라 CEO는 2024년 2월 29일 ‘인류 이익을 위한다’ 라는 설립 목적을 위반했다는 이유로 오픈AI를 상대로 소송을 제기하면서 오픈AI의 법적 부담이 증대됐다. 오픈AI는 영리 법인과 비영리 법인이 결합한 독특한 기업 구조로 기업 안팎에서 다양한 분쟁을 겪고 있을 뿐 아니라, MS의 투자에 따른 반독점 조사와 여러 저작권 소송에 얽혀 있다.

이 가운데 일론 머스크가 제기한 소송은 오픈AI 초기 투자자였던 일론 머스크가 오픈AI를 상대로 인류 이익이 아닌 사익을 우선시켰다며 계약 위반으로 소송을 제기한 것으로, 법원에 오픈AI의 모든 연구와 기술을 대중에게 공개하도록 명령할 것을 요청했다. 이에 대해 오픈AI는 머스크가 보낸 이메일을 근거로 그가 오픈AI의 영리 법인 전환 및 테슬라와의 합병을 제안했다고 반박했고, 법조계도 머스크의 승소 가능성이 작은 것으로 평가한다.

기업 거버넌스 문제는 오픈AI는 2023년 11월 이사회 의 일방적 결정에 의한 샘 알트먼 CEO의 해고 사건과 관련해 내부 조사를 진행했고 조사 결과에 따라 기업 거버넌스를 개선할 계획이다. 조사를 담당한 로펌은 갈등의 원인으로 이사회와 알트먼 간 신뢰 부족과 거버넌스 문제를 지적하고, 3월 9일 내부 조사가 마무리되면서 알트먼이 이사회에 공식 복귀했다.

미국과 유럽의 반독점 규제당국은 오픈AI와 MS 간 밀접한 제휴 관계에 대한 조사에 착수했다. MS는 오픈AI에 총 130억 달러를 투자해 지분 49%를 보유하고 있으며, 오픈AI의 AI 모델 개발에 자사의 컴퓨팅 인프라에 지원하고 기술 상당 부분을 MS 제품에 적용할 수 있는 독점권도 확보했다. 미국 연방거래위원회(FTC)는 양사의 제휴가 지배적 사업자의 부당한 영향력 행사나 특혜를 통해 공정한 경쟁을 저해하는지를 조사할 계획이며, 유럽연합과 영국에서도 유사한 조사를 진행하고 있다.

221) <https://sacra.com/c/openai/>

222) AP News, It's not just Elon Musk: ChatGPT-maker OpenAI confronting a mountain of legal challenges, 2024.03.07.

이외에 다수의 유명 작가와 뉴욕타임스 등의 언론 매체는 오픈AI가 AI 모델 개발 과정에서 저작권법을 위반했다는 소송을 제기한 상태다. 베스트셀러 작가 존 그리섬과 조지 R.R. 마틴을 포함한 17명의 작가는 2023년 9월 오픈AI를 상대로 저작권 침해에 대한 집단소송을 제기했다. 뉴욕타임스는 2023년 12월 뉴스 데이터의 무단 학습을 이유로 오픈AI에 소송을 제기하고 자사 데이터로 학습된 모델의 전체 파기를 요청했다.

나. AI 안전 연구를 맡은 오픈AI 슈퍼 얼라인먼트 팀 해체²²³⁾

오픈AI가 인간을 뛰어넘는 초지능 AI 시스템의 위험 통제 방법 연구를 위해 설립한 ‘슈퍼얼라인먼트(Super alignment)’ 팀을 해체했다. 오픈AI는 2023년 7월 슈퍼 얼라인먼트 팀을 발족하고 초지능 AI 시스템이 인간의 의도를 따르도록 제어하는 연구 과제를 부여하고 4년간 자사 컴퓨팅 성능의 20%를 투입하겠다는 계획을 공개했다.

하지만, 팀 설립 채 1년도 안 돼 해체 절차를 밟고 있다. 슈퍼 얼라인먼트팀의 해체는 팀의 공동 리더였던 일리야 수츠케버(Ilya Sutskever) 공동창업자 겸 수석 과학자와 얀 레이케(Jan Leike)의 퇴진과 함께 이뤄졌다. 얀 레이케는 오픈AI의 핵심 우선순위에서 경영진과 장기간 이견이 있었으며, 안전 문화와 프로세스가 제품 출시보다 뒷전으로 밀렸다고 퇴사 배경을 설명했다. 슈퍼 얼라인먼트 팀은 안전 연구를 수행하며 외부 연구자들에게 수백만 달러의 보조금을 지원해 왔지만, 신제품 출시에 더 많은 자원이 투입되면서 충분한 지원을 받지 못했다고 지적했다. 일리야 수츠케버와 샘 알트만 CEO 간 갈등도 슈퍼 얼라인먼트 팀이 해체된 요인으로 작용했다는 평가다. 수츠케버는 2023년 말 오픈AI 이사회에서 여타 이사진과 함께 샘 알트만의 해고를 결정했으나 이후 샘 알트만이 다시 CEO로 복귀하며 이사회에서 물러났다.

오픈AI는 슈퍼 얼라인먼트 팀의 해체 이후 해당 팀이 수행하던 업무를 여러 부서에 소속된 연구원이 참여하는 느슨한 체계로 대체할 계획이다. 샘 알트만 CEO와 그렉 브록만(Greg Brockman) 사장은 AI의 안전을 위해 긴밀한 피드백 절차와 엄격한 테스트, 모든 단계에서의 신중한 검토, 세계 최고 수준의 안전·보안·기능의 조화가 필요하다고 강조했다. 향후 오픈AI가 안전 연구에 소홀해질 것이라는 우려가 제기된다. 레이케와 수츠케버의 퇴사

223) TechCrunch, OpenAI created a team to control ‘superintelligent’ AI – then let it wither, source says, 2024.05.18.

이후 오픈AI의 창립 멤버인 존 슈만(John Shulman)이 슈퍼 얼라인먼트 팀이 담당하던 업무를 총괄하기로 했다.

다. 데이터 분석 기업 록셋 인수²²⁴⁾

오픈AI가 2024년 6월 21일 데이터 검색과 분석 전문 기업 록셋(Rockset)을 인수했다고 발표했다.²²⁵⁾ 페이스북 출신 엔지니어들이 2016년 설립한 록셋은 데이터베이스 및 공용 클라우드 스토리지에서 데이터를 자동으로 수집한 뒤, 검색과 분석을 위해 데이터를 색인화하는 도구를 개발했다. 록셋은 오픈AI에 인수되기 전까지 주요 벤처캐피털로부터 1억 1,750만 달러 이상의 투자를 유치했으며, 메타와 제트블루 항공과 같은 기업을 고객으로 확보하는 성과를 거뒀다. 록셋은 뉴욕 소재 그래픽 디자인 스타트업 글로벌 일루미네이션(Global Illumination)에 이어 오픈AI가 두 번째로 인수한 기업이다.

오픈AI는 록셋의 구성원들이 오픈AI에 합류할 것이라고 발표했으며, 록셋의 기술을 통합해 제품 전반에 걸쳐 검색 인프라를 강화할 계획이다. 오픈AI는 록셋의 인프라를 통해 AI 제품을 사용하거나 스마트 애플리케이션을 구축하면 자체 데이터를 더욱 잘 활용하고 실시간 정보에 접근할 수 있으며, 기업이 보유한 데이터를 실제 활용할 수 있는 인텔리전스로 전환할 수 있다고 강조했다.

오픈AI의 모델과 기업 데이터를 결합할 경우, 기업 활용 사례에 맞게 모델의 미세조정과 환각을 줄이는 등의 개선이 가능할 전망이다. 오픈AI의 이번 인수는 기업 영업을 강화하려는 최근 전략의 일환으로, 오픈AI는 지난 5월 PwC와 제품 재판매 계약을 체결하기도 했다. 오픈AI는 앞서 기업형 맞춤형 모델 조정과 컨설팅 프로그램을 출시하기도 했으며, B2B 사업전략에 힘입어 올해 매출이 34억 달러를 넘어설 전망이다. 오픈AI는 최근 챗GPT 엔터프라이즈의 사용자가 60만 명에 달하며, 포천 선정 500대 기업의 93%가 자사 고객이라고 밝혔다.

224) OpenAI, OpenAI acquires Rockset, 2024.06.21.

225) TechCrunch, OpenAI buys Rockset to bolster its enterprise AI, 2024.06.21.

라. AI 기반 검색엔진 ‘서치GPT’ 프로토타입 공개²²⁶⁾

오픈AI가 2024년 7월 25일 AI 기반 검색엔진 ‘서치GPT’ 프로토타입을 공개하고 대기자 명단을 받아 일부 사용자를 대상으로 테스트를 진행했다.²²⁷⁾ 서치GPT는 단순한 링크 목록만 제공하는 기존 검색엔진과 달리, 인터넷 정보에 실시간으로 접근해 사용자의 질문에 직접적인 답변과 함께 출처에 대한 링크를 제공하며 검색 결과를 이미지로도 제공한다. 사용자는 사람과 대화하듯이 검색 결과에 대하여 추가 질문을 할 수 있고 각각의 질문을 통해 서치GPT와 맥락 공유가 가능하다. 서치GPT는 총 1만 명의 테스트 사용자에게 우선 제공되며, 오픈AI는 향후 챗GPT에 검색 기능을 통합할 계획이나 구체적인 일정은 미공개 상태다.

AI 검색이 보급되면 검색엔진에서 웹사이트로 유입되는 트래픽이 줄어들 것이라는 우려에 대응해, 오픈AI는 서치GPT의 검색 결과물에 인용한 원본 출처를 밝힘으로써 사용자가 언론사나 출판사와 같은 퍼블리셔와 연결되도록 설계되었다고 강조했다. 서치GPT는 검색에서 퍼블리셔를 명확하게 인용하고 링크를 안내하며, 사이드바를 통해 관련 정보에 대한 더 많은 결과 및 출처도 제공한다.

오픈AI는 뉴스코프(News Corp), 악셀스프링어(Axel Springer), 폭스 미디어(Vox Media)를 비롯한 여러 퍼블리셔와 제휴해 지속적으로 피드백을 수렴하고 있으며, 서치GPT 프로토타입 출시와 함께 퍼블리셔가 서치GPT에 표시되는 방식을 관리하는 도구도 제공할 방침이라고 밝혔다. 서치GPT를 공개한 오픈AI는 AI 검색 기능을 강화하고 있는 구글 및 AI 검색엔진을 표방하는 퍼플렉시티 AI(Perplexity AI)와 직접적 경쟁 관계를 형성할 전망이다. 퍼플렉시티 AI는 포브스(Forbes) 등 주요 언론사로부터 기사를 표절했다는 비판을 받는 가운데, 오픈AI 등 다수 언론사와 콘텐츠 사용 계약을 체결하여 사용자 저변 확대에 나서고 있다.

(2) 기술개발 및 혁신

가. GPT-2로 GPT-4 감독 가능²²⁸⁾

오픈AI가 2023년 7월 출범한 슈퍼얼라인먼트(Superalignment) 팀이 2023년 12월 14일 작은 크기의 AI 모델로도 규모가 더 큰 모델을 감독할 수 있다는 첫 연구 결과를 발표했다.²²⁹⁾

226) OpenAI, SearchGPT Prototype, 2024.07.25.

227) The Verge, OpenAI announces SearchGPT, its AI-powered search engine, 2024.07.26.

228) OpenAI, Weak-to-strong generalization, 2023.12.14.

229) OpenAI, Superalignment Fast Grants, 2023.12.14.

슈퍼얼라인먼트 팀의 핵심 과제는 향후 10년 안에 등장할 초지능 AI 시스템을 감독하기 위한 방식을 찾아내는 것으로, 인간이 감독하기 어려운 고도의 복잡하고 창의적인 행동을 할 수 있는 미래의 AI 시스템을 신뢰하고 제어할 방법을 개발 필요하다고 주장했다.

연구진이 GPT-2 수준의 모델을 이용해 GPT-4를 감독하는 방법을 테스트한 결과, GPT-3.5 수준의 성능에 가깝도록 GPT-4의 기능 대부분을 구현하는 데 성공했다고 밝혔다. 연구진이 GPT-2의 결과를 학습한 GPT-4가 동일한 작업을 수행하도록 훈련하자 GPT-4는 GPT-2를 꾸준히 능가하는 결과를 보였다. 연구진은 GPT-2에 자연어 처리 테스트와 채스 문제, 챗GPT 응답 모델링이라는 세 가지 유형의 작업을 수행하도록 한 뒤 GPT-2의 결과를 바탕으로 동일 작업을 수행하도록 GPT-4를 훈련했다.

정확한 답이 있는 자연어 처리 작업에서는 GPT-4가 GPT-3.5 수준의 정확도를 보였으나, 챗GPT 응답 모델링의 경우에는 인간이 선호하는 답변을 결정해야 하는 모호성으로 인해 GPT-2와 GPT-4의 결과물 차이가 가장 적게 나타났다. 연구진은 이번 연구가 일관성이 부족하고 작업 유형에 따라 격차가 크다는 점에서 실용적인 해결책이라기보다는 개념 증명에 가깝다고 한계를 인정하면서도, 약한 모델로 더 강력한 모델의 감독이 가능함을 입증했다는 점에서 고무적이라고 평가했다.

오픈AI는 AI 시스템의 정렬 연구를 지원하기 위해 이번 연구 결과를 오픈소스로 공개하고 1,000만 달러 규모의 연구 보조금 프로그램도 발표했다. 오픈AI는 연구기관과 비영리 단체, 개인 연구자의 초지능 AI 시스템 정렬 연구에 10만~200만 달러의 보조금을 지원하는 한편, 대학원생 대상으로 1년간 15만 달러의 장학금 프로그램도 발표했다.

나. ‘달리3’ 로 만든 AI 이미지에 디지털 워터마크 적용 방침²³⁰⁾

오픈AI가 2024년 2월 6일 자사의 이미지 생성 AI 도구 ‘달리(DALL-E)3’ 로 생성한 이미지에 디지털 워터마크를 적용한다고 발표했다. 오픈AI는 웹과 모바일 환경의 달리3로 생성한 이미지에 ‘콘텐츠 출처 및 진위성 연합(Coalition for Content Provenance and Authenticity, C2PA)’ 의 디지털 인증을 적용했다. C2PA는 미디어 콘텐츠에 메타데이터를

²³⁰⁾ OpenAI, C2PA in DALL·E 3, 2024.02.06.

삽입해 출처 정보를 확인할 수 있도록 하는 개방형 표준으로, AI 이미지뿐 아니라 카메라 제조사나 언론사에서도 콘텐츠 출처 인증을 위해 사용되고 있다.

오픈AI는 달리3가 통합된 챗GPT나 달리3 자체 API를 사용해 생성된 이미지에 C2PA 메타데이터를 추가하며, C2PA 적용 시 파일 크기는 소폭 증가하지만 대기 시간이나 생성 품질에는 영향이 없다고 설명했다. 사용자는 ‘콘텐츠 자격증명 인증(Content Credentials Verify)’ 과 같은 웹사이트를 통해 이미지 생성에 사용한 도구나 정확한 생성 일시 정보를 확인할 수 있다. 오픈AI는 C2PA와 같은 메타데이터가 출처를 완벽히 증명할 수 없으며 실수나 고의로 손쉽게 제거될 수 있다는 한계를 인정하면서도, 이러한 정책이 사용자의 출처 확인을 장려하여 디지털 정보의 신뢰성을 높일 것으로 기대한다고 밝혔다.

한편, 메타 역시 2024년 2월 6일 자사가 운영하는 소셜 미디어 플랫폼 페이스북, 인스타그램, 스레드에서 AI 생성 이미지에 라벨을 적용한다고 발표했다.²³¹⁾ 메타는 현재 자체 AI 도구인 메타 AI로 생성된 콘텐츠에는 ‘이매진드 위드 AI(Imagined with AI)’ 라벨을 함께 표시한다. 앞으로는 외부 AI 도구로 제작된 콘텐츠에도 AI 라벨을 적용할 계획으로, AI 라벨에 대한 공통 기술 표준 마련을 위해 다른 기업들과 협력을 강화할 방침이다.

다. 사람의 음성을 복제할 수 있는 ‘보이스 엔진’ 기술 공개²³²⁾

오픈AI가 2024년 3월 18일 공식 블로그를 통해 음성을 복제할 수 있는 AI 모델 ‘보이스 엔진(Voice Engine)’ 을 공개했다. 보이스 엔진은 텍스트 입력과 15초 길이의 음성 샘플을 사용해 원래 화자와 매우 유사하고 자연스러운 음성을 생성한다. 오픈AI는 2022년 말에 처음 보이스 엔진을 개발해 텍스트-음성 변환 API와 챗GPT 답변을 음성으로 제공하는 기능을 지원해 왔다.

이번에 오픈AI는 2023년 말부터 보이스 엔진에 대한 비공개 테스트를 시작해 다양한 활용 사례를 확인했다고 밝혔다. 교육 플랫폼 에이지오브러닝(Age of Learning), 사용자의 음성과 얼굴을 활용한 아바타를 생성하는 헤이젠(HeyGen), 보건 상담 서비스인 다이마지(Dimage) 등이 협력사로 참여해 테스트를 진행했다고 공개했다. 보이스 엔진은 다양한 화자를 대표하

231) Meta, Labeling AI-Generated Images on Facebook, Instagram and Threads, 2024.02.06.

232) OpenAI, Navigating the Challenges and Opportunities of Synthetic Voices, 2024.03.29.

는 자연스러운 음성으로 책을 읽지 못하는 사람과 어린이에게 음성 해설을 제공하거나 학생들에게 맞춤형된 실시간 상호작용을 지원한다.

보이스 엔진은 창작자나 기업들이 전 세계의 더 많은 사람에게 자신의 목소리로 콘텐츠를 전달할 수 있도록 동영상이나 팟캐스트 등을 번역하는 기능을 제공한다. 또한 다양한 현지어가 사용되는 아프리카 지역에서 현지어 번역을 통해 원격의료 상담과 같은 필수 서비스 개선에도 사용될 수 있고, 언어 표현에 어려움을 겪는 사람들이나 퇴행성 언어장애를 겪는 환자들에게 다양한 언어로 기계음이 아닌 자연스러운 음성의 의사소통을 지원한다.

하지만 오픈AI는 보이스 엔진의 악용 가능성을 우려해 출시 여부는 아직 미정이라고 밝혔다. 오픈AI는 사람의 음성을 복제할 수 있는 보이스 엔진의 위험성을 인식하고 안전조치를 취하는 한편, 비공개 테스트 및 사회적 논의를 거쳐 출시 여부를 결정할 계획이다. 오픈AI는 비공개 테스트에 참여하는 기업에 동의나 법적 권리 없이 개인이나 조직의 사칭을 금지하고, 사용자에게 현재 제공되는 음성이 AI로 생성되었음을 명확히 공개할 것을 요구하고 있다. 또한 보이스 엔진에서 생성된 음성의 출처를 추적하는 워터마크를 적용하고, 음성 사용 방식을 사전 모니터링하는 등의 안전조치를 시행하고 있다.

라. GPT-4의 해석 가능성 향상²³³⁾

오픈AI가 2024년 6월 6일 언어모델 내 신경망의 작동 방식을 이해하기 위해 인간이 이해하기 쉬운 개념과 일치되는 신경망의 특징을 식별하는 방법에 관한 연구 성과를 공개했다.²³⁴⁾ 신경망을 해석하려면 신경 계산에 유용한 구성요소를 찾아야 하나, 언어모델 내의 신경망은 예측할 수 없는 패턴으로 활성화되어 인간이 이해하기 어려운 한계가 존재한다. 연구진은 이러한 문제 해결을 위해 ‘희소 오토인코더(Sparse Autoencoder)’를 활용했으며, 오토인코더는 입력 데이터를 압축한 뒤 압축된 상태에서 원본 데이터를 재구성하는 신경망으로 희소 오토인코더는 입력 데이터를 대부분 제거하고 가장 중요한 특징(Feature)만을 추출하여 표현했다.

233) OpenAI, Extracting Concepts from GPT-4, 2024.06.06.

234) Gigazine, OpenAI announces it has broken down GPT-4's thoughts into 16 million interpretable patterns, 2024.06.07.

희소 오토인코더를 활용해 인공지능망을 특징별로 분류하면 ‘법률 텍스트에 반응하는 특징’, ‘DNA 서열에 반응하는 특징’ 등 인간이 이해하기 쉬운 개념과 자연스럽게 일치하는 활성화 패턴을 파악할 수 있다는 설명이다. 연구진은 첨단 AI 모델의 수천만 개 특징으로 확장할 수 있는 희소 오토인코더 기법을 개발해 GPT-4에 적용하여 1,600만 개에 달하는 특징 추출에 성공했다.

하지만, 오픈AI는 이번 연구로 모델 해석 가능성을 높여 모델의 신뢰성이 향상될 것으로 기대하지만, 연구가 아직 초기 단계이며 상당한 한계가 있다고 강조했다. 이번 연구에서 발견된 특징 중 상당수는 여전히 해석이 어려우며, 희소 오토인코더로 모델의 전체 작동 방식을 포착하기는 어렵다는 것이다. 희소 오토인코더를 통해 GPT-4를 활성화하면 10배 더 적은 컴퓨팅으로 훈련된 모델과 동등한 성능을 발휘할 수 있으나, 첨단 LLM이 나타내는 개념에 전부 대응하려면 수십억에서 수조 개의 특징을 추출할 수 있어야 하므로 막대한 컴퓨팅 자원이 필요하다.

희소 오토인코더는 모델의 한 지점에서 특징을 찾을 수 있으나 이는 모델 해석의 한 단계일 뿐이며, 모델이 해당 특징을 연산하는 방법 및 모델의 여타 영역에서 이 특징이 어떻게 사용되는지를 이해하려면 훨씬 더 많은 추가 연구가 필요하다는 지적이다.

(3) 제품 및 서비스

가. 유료 사용자 대상 GPT 스토어 출시²³⁵⁾

오픈AI가 2024년 1월 10일 기업이나 개인이 GPT를 기반으로 개발한 맞춤형 AI 챗봇을 거래할 수 있는 GPT 스토어를 유료 사용자 대상으로 출시했다. 오픈AI는 2024년 11월 별도 코딩 없이 맞춤형 AI 챗봇을 만들 수 있는 GPTs 기능을 공개했으며, 오픈AI에 따르면 지난 두 달간 챗GPT 사용자들은 300만 개 이상의 맞춤형 챗봇을 제작했다. GPT 스토어는 사용자가 제작한 챗봇을 ▲달리(DALL-E) ▲글쓰기 ▲프로그래밍 ▲연구·분석 ▲교육 ▲생산성 ▲라이프스타일로 구성된 카테고리 유형으로 구분하여 소개했다.

스토어의 커뮤니티 리더보드는 학술 논문 결과를 검색하는 ‘컨센서스(Consensus)’, 코딩

235) OpenAI, Introducing the GPT Store, 2024.01.10.

교육을 제공하는 칸 아카데미(Khan Academy)의 ‘코드 튜터(Code Tutor)’ , 등산로 정보를 제공하는 ‘올트레일스(AllTrails)’ 와 같은 유용하고 인기 있는 챗봇을 추천한다. GPT 스토어를 통해 챗봇을 공유하려면 챗봇을 전체 공개하고 스토어에 빌더 프로필을 생성해야 하며, 정책 준수 여부를 확인하기 위한 검토 절차에는 자동과 수동 방식이 모두 사용된다. GPT 제작자는 사용량에 따라 수익을 올릴 수 있으며, 오픈AI는 2024년 1분기에 미국에서 먼저 수익 공유 프로그램을 시작할 계획으로 구체적인 내용은 아직 공개되지 않았다.

오픈AI는 GPT 스토어와 함께 150명 미만의 중소기업이나 조직을 대상으로 하는 새로운 ‘챗GPT 팀(ChatGPT for Teams)’ 구독 서비스도 출시했다. 챗GPT 팀은 사용자당 이용료가 60달러 수준인 챗GPT 엔터프라이즈보다 저렴한 월 25달러, 연 30달러의 구독료에 3만 2천 개 토큰의 컨텍스트창을 지원한다. 챗GPT 팀은 작업 그룹을 위한 전용 작업공간과 팀 관리 도구도 제공하며, 내부용으로 챗봇을 제작해 공유할 수 있도록 지원할 방침이다.

나. 텍스트 투 비디오(Text to-Video) 생성 AI ‘소라(Sora)’ 공개²³⁶⁾

오픈AI가 2024년 2월 16일 텍스트 프롬프트를 바탕으로 최대 1분 길이의 동영상을 생성하는 확산(Diffusion) 모델 ‘소라(Sora)’ 를 발표했다. 소라는 프롬프트에서 사용자가 요청한 내용뿐 아니라 해당 내용이 실제 세계에 어떻게 구현되어야 하는지를 이해하고, 다양한 캐릭터 및 특정 유형의 동작과 같은 세부 정보를 반영해 복잡한 장면을 생성한다. 실제 소라가 생성한 영상은 도쿄의 거리를 걷는 여성의 이미지를 실제로 촬영한 것으로 보일 만큼 정교하게 구현해 영상·광고업계 판도를 뒤흔들 수 있다는 전망도 제기됐다.

달리(Dall-E) 및 GPT 연구 결과를 기반으로 한 소라는 시각적 훈련 데이터에 대하여 구체적인 캡션을 생성하는 달리3의 ‘리캡션(Recaptioning)’ 기법을 채택해, 사용자의 텍스트 지시를 충실히 이행하여 동영상을 생성한다. 또한 정지된 이미지를 기반으로 동영상을 생성할 수 있어 기존 동영상을 확장하거나 빠진 프레임을 채울 수도 있고, 전체 동영상을 한 번에 생성하거나 생성된 동영상을 확장해 더 길게 만들 수도 있다.

오픈AI는 소라의 단점으로 복잡한 장면의 물리 법칙을 정확하게 모사하거나 원인과 결과

236) OpenAI, Sora, 2024.02.16.

관계를 이해하지 못할 수 있다고 설명했다. 일례로 사람이 쿠키를 한 입 먹은 다음 장면에서 쿠키에 베어 문 자국이 없거나, 프롬프트 지시에서 왼쪽과 오른쪽을 혼동할 수 있다는 설명이다.

오픈AI는 모델의 안전성 확보를 위해 대중에 공개하기에 앞서 레드팀을 통한 적대적 테스트를 시행했고, 추가로 다양한 안전조치를 취할 계획이다. 모델 배포 이후 출처 확인을 위한 워터마크를 적용하는 한편, 텍스트 필터를 통해 폭력적·성적 이미지나 유명인 초상 등 사용 정책을 위반하는 프롬프트 지시를 거부하도록 할 예정이다. 강력한 이미지 필터를 개발해 생성된 모든 동영상의 프레임을 검토하고 사용자에게 표시되기 전 정책 준수 여부를 확인하는 한편, 전 세계 정책 입안자, 교육자, 예술가의 참여를 받아 우려 사항을 확인하고 긍정적인 사용 사례를 파악할 방침이다.

다. 사람과 자연스러운 실시간 대화가 가능한 ‘GPT-4o’ 출시²³⁷⁾

오픈AI가 2024년 5월 13일 텍스트, 음성, 시각을 이용해 사람과 자연스럽게 상호작용할 수 있는 신규 AI 모델 ‘GPT-4o’를 공개했으며, o는 모든 것을 뜻하는 ‘옵니(omni)’의 약칭이다. 음성 모드에서 GPT-4o의 응답시간은 최소 232밀리초(0.23초), 평균 320밀리초(0.32초)로 인간의 응답시간과 비슷하며, 이전 모델인 GPT-3.5(평균 2.8초), GPT-4(5.4초)보다 대폭 단축됐다.

온라인 발표행사에서 미라 무라티(Mira Murati) 오픈AI CTO는 GPT-o가 다양한 목소리와 톤으로 이야기를 들려주거나 종이에 적힌 수학 문제를 풀고 영어와 이탈리아어 화자 간 실시간 통역을 지원하는 모습을 시연했다. GPT-4o는 API로 사용 시 기존 GPT 모델보다 속도가 2배 빠르고 비용은 50% 저렴하며, 기존 모델보다 시각과 청각 이해도가 크게 향상된 모습을 보였다. 이전 모델은 3개의 개별 모델을 사용해 멀티모달을 텍스트로 변환한 후 다시 멀티모달 형태로 변환했으나, GPT-4o는 텍스트와 시청각 전반에 걸쳐 단일 모델로 훈련해 모든 입출력을 동일한 신경망에서 처리하여 처리 속도와 성능을 개선했다.

벤치마크 테스트에서 GPT-4o는 텍스트와 추론, 코딩 성능에서 GPT-4 터보와 동일한 성능

237) OpenAI, Hello GPT-4o, 2024.05.13.

을 나타냈고, 다국어, 오디오, 이미지 기능에서는 최고 수준을 달성했다. 객관식 문제 모음인 대규모 다중언어 작업 이해(MMLU) 벤치마크에서 GPT-4o는 88.7%를 기록해 GPT-4 터보의 86.5%와 구글 제미니 울트라 83.7%를 넘어섰다. GPT-4o는 전 세계 사용자들에게 무료로 제공되며, 유료 사용자는 무료 사용자보다 5배 더 많은 질문을 할 수 있다.

GPT-4o의 텍스트와 이미지 기능은 5월 13일부터 제공되며, 유료 서비스 챗GPT 플러스 사용자에게는 음성이 지원되는 신규 버전이 연이어 출시될 예정이다. 개발자들은 API로 GPT-4o의 텍스트와 이미지 기능을 사용할 수 있으며, 오디오와 비디오 기능은 수 주 안에 일부 파트너 집단에 선공개할 예정이다.

라. 비용 효율적인 소형 AI 모델 ‘GPT-4o 미니’ 출시²³⁸⁾

오픈AI는 2024년 7월 28일 가장 비용 효율적인 소형 모델인 ‘GPT-4o 미니’를 출시하고 챗GPT에서 GPT-3.5를 대신해 무료 사용자 및 플러스와 팀 사용자에게 제공한다고 발표했다. GPT-4o 미니는 입력 토큰 100만 개당 15센트, 출력 토큰 100만 개당 60센트로 가격이 책정되어, GPT-3.5 터보 대비 60% 이상 저렴하다.

모델은 12만 8천 개 토큰의 컨텍스트 창을 제공하고 어시스턴트 API(Assistants API), 챗 컴플리션 API(Chat Completions API), 배치 API(Batch API)에서 텍스트와 비전 모델로 사용할 수 있으며, 동영상과 오디오 입력은 추후 지원할 예정으로, 요청 당 최대 출력 토큰은 16만 개다. 낮은 비용과 짧은 지연 시간으로 여러 모델 API를 호출하거나 모델에 대규모 컨텍스트(예, 전체 코드 베이스 또는 대화 기록)를 전달하는 작업, 실시간 고객 지원 챗봇 등 다양한 작업을 처리할 수 있다.

오픈AI는 GPT-4o 미니를 통해 훨씬 저렴한 비용으로 AI를 이용할 수 있게 됨으로써 AI 애플리케이션 범위가 크게 확장될 것으로 기대한다. GPT-4o 미니는 뛰어난 텍스트 성능과 다중모드 추론을 지원하여 주요 벤치마크에서 GPT-4o에는 미달하지만 GPT-3.5 터보와 기타 소형 모델을 능가하는 결과를 기록했다. GPT-4o 미니는 텍스트 지능과 추론 벤치마크인 MMLU에서 82.0%의 점수로 제미니 플래시(77.9%)와 클로드 하이쿠(73.8%)를 앞섰으며, 다

238) OpenAI, GPT-4o mini: advancing cost-efficient intelligence, 2024.07.18.

중모드 추론평가인 MMMU에서는 59.4%로 제미니 플레시(56.1%)와 클로드 하이쿠(50.2%)를 모두 능가했다.

수학적 추론을 측정하는 MGSM²³⁹⁾*에서는 87.0%로 제미니 플레시(75.5%)와 클로드 하이쿠(71.7%)를 앞섰으며, 코딩 성능을 측정하는 HumanEval에서도 87.2%로 제미니 플레시(71.5%)와 클로드 하이쿠(75.9%)를 추월했다. 오픈AI는 GPT-4o 미니가 새로운 안전 메커니즘인 ‘지시 계층(The Instruction Hierarchy)’을 적용한 최초의 모델이라고 밝혔다. 지시 계층은 모델이 안전선을 무시하라는 지시를 거부하고 개발자의 시스템 메시지를 우선하여 따르도록 함으로써, 탈옥과 프롬프트 주입 공격에 대한 모델의 저항 능력을 향상하여 안전성을 개선했다.

마. 추론에 특화된 AI 모델 ‘o1-프리뷰’ 출시²⁴⁰⁾

오픈AI가 2024년 9월 12일 추론에 특화된 새로운 AI 모델 ‘o1-프리뷰(Preview)’와 ‘o1-미니(mini)’를 유료 사용자를 대상으로 공개했다. 응답에 앞서 생각에 더 많은 시간을 할애하도록 설계된 이 모델은 복잡한 추론을 할 수 있고 과학, 코딩, 수학에서 GPT-4o보다 더 어려운 문제를 해결할 수 있다. o1 모델은 복잡한 코드의 정확한 생성과 코딩 오류 발견에 특히 뛰어나며, 특히 경량 모델인 o1-미니는 o1-프리뷰 대비 비용이 80% 저렴해 비용 대비 성능이 우수하다는 평을 받는다. 오픈AI는 챗GPT 플러스와 팀, 엔터프라이즈 및 에듀 사용자에게 모델을 우선 제공하고, 향후 챗GPT 무료 사용자에게도 o1-미니를 지원할 예정이다. 오픈AI는 GPT 시리즈와 달리 복잡한 추론이 가능하다는 점에서 모델명을 ‘오픈AI o1’로 새롭게 정했으나, 질문에 따라 응답 속도가 느리고 웹 브라우징이나 파일 업로드 등을 지원하지 않아 단기적으로는 초기 모델인 o1보다 GPT-4o가 더 유용할 수 있다고 설명한다.

오픈AI는 o1 모델을 정기적으로 업데이트하고 현재는 지원되지 않는 웹 브라우징, 파일과 이미지 업로드 등의 기능을 추가할 계획이며, GPT 시리즈 모델도 지속 출시 예정이다. o1 모델은 사람이 어려운 문제에 답하기 전 숙고하듯이 ‘사고 연쇄(Chain of Thought)’²⁴¹⁾를 통해 추론 역량을 개선해 물리학, 화학, 생물학의 복잡한 벤치마크 과제에서 박사과정 학생과 비슷한 성과를 나타냈으며, 수학과 코딩에서도 탁월한 성과를 기록했다. o1은 실수를 인

239) 164개 프로그래밍 과제로 구성되어 LLM의 코드 생성 기능을 평가하는 벤치마크

240) OpenAI, Introducing OpenAI o1-preview, 2024.09.12.

241) 복잡한 문제를 여러 단계로 나누어 해결하여 최종 결과를 도출하는 방식

식하고 수정하는 방법, 복잡한 단계를 단순한 단계로 세분화하는 방법, 효과가 없는 기존 방식 대신 다른 방식을 시도하는 방법 등을 학습해 추론 능력을 비약적으로 향상했다는 설명이다. 국제수학올림피아드(IMO) 예선 문제 테스트에서 GPT-4o의 정답률은 13%였으나 o1의 정답률은 83%였으며, 코딩 능력을 평가하는 코드포스(Codeforces) 대회에서는 상위 11%에 해당하는 성적을 기록했다.

오픈AI는 o1의 추론 역량을 활용해 안전 및 정렬 지침을 준수하도록 하는 새로운 안전 교육방식을 고안하여 모델 안전성을 향상할 계획이다. 모델의 안전장치를 우회하고자 하는 탈옥 시도에 대응한 평가에서 GPT-4o는 100점 만점 중 22점을 받았으나 o1-프리뷰 모델은 84점을 기록했다.

(4) AI 윤리 및 규제 대응

가. 생성 AI를 이용한 선거 개입을 막기 위한 대응책 공개²⁴²⁾

오픈AI가 2024년 1월 15일 공식 블로그를 통해 2024년 전 세계 선거에 대응해 정확한 선거 정보의 제공과 신중한 정책의 시행, 투명성 개선으로 플랫폼의 안전을 강화하겠다고 발표했다. 오픈AI는 딥페이크 제작이나 AI 챗봇을 이용한 후보자 사칭과 같은 남용을 막기 위해 챗GPT와 API 사용 정책을 정기적으로 개선하고 있다고 강조했다. 오픈AI는 선거운동과 로비를 위한 애플리케이션 및 투표 절차나 자격 등의 선거 관련 정보를 허위로 전달하거나 선거를 저해할 목적의 애플리케이션 제작을 금지했으며, 실제 사람(후보자)이나 정부 기관으로 가장하는 챗봇 제작도 금지했다.

오픈AI는 유권자가 이미지 제작 방식에 신뢰하면서 이미지를 평가할 수 있도록 AI 생성 콘텐츠 투명성을 강화하는 방안도 마련했다. 오픈AI는 2024년 초 달리(DALL-E) 3에서 생성된 이미지에 콘텐츠 출처를 확인할 수 있도록 ‘콘텐츠 출처 및 진위성 연합(C2PA)²⁴³⁾’의 디지털 인증을 적용할 계획이다. 오픈AI는 달리에서 생성된 이미지를 탐지할 수 있는 새로운 도구도 개발해 내부 테스트에서 성공적인 결과를 얻었다며 조만간 연구자와 언론인 대상으로 우선 공개할 예정이다.

242) OpenAI, How OpenAI is approaching 2024 worldwide elections, 2024.01.15.

243) Coalition for Content Provenance and Authenticity: 어도비가 콘텐츠의 신뢰성과 투명성 보장을 위해 2021년 마이크로소프트, 애플, 인텔, 트루픽과 함께 2021년 설립한 단체

오픈AI는 챗GPT가 언론사 뉴스 등 기존 정보 소스와 점점 더 통합되고 있다며, 전 세계의 실시간 선거 뉴스를 챗GPT와 연결해 유권자들이 정보의 신뢰성을 결정할 수 있도록 지원하겠다고 밝혔다. 오픈AI는 권위 있는 선거 정보에 대한 접근성 향상을 위해 미국 초당파적 단체인 미국 국무장관협회(NAS)와 협력을 공개하고, 사용자가 선거 절차 관련 질문을 하면 선거관리위원회의 공식 웹사이트(CanIVote.rg)로 안내한다고 설명했다.

나. 챗GPT 부정행위 탐지 도구 개발 후 공개 유보²⁴⁴⁾

월스트리트저널(The Wall Street Journal)의 2024년 8월 4일 오픈AI가 챗GPT를 이용해 작성된 과제나 논문을 탐지할 수 있는 도구를 개발했으나 미출시했다고 보도했다. 오픈AI는 2023년 부정행위 탐지 도구의 출시 준비를 완료했으나, 출시 여부를 두고 2년에 걸쳐 내부적으로 논쟁이 지속되며 출시를 유보한 것으로 알려졌다. 출시에 반대하는 직원들은 챗GPT 사용자 감소를 우려하고 있으며, 챗GPT 사용자 대상 설문조사 결과 3분의 1은 부정행위 탐지 도구가 출시되면 챗GPT를 사용하지 않을 수도 있다고 응답했다.

오픈AI가 개발한 부정행위 탐지 도구는 챗GPT가 생성한 콘텐츠에 눈에 보이지 않는 워터마크를 표시하여 이를 식별하는 방식이다. 이 도구는 문서 전체 또는 일부가 챗GPT로 작성되었을 가능성을 점수로 표시하며, 오픈AI의 내부 문서에 따르면 워터마크의 효과는 99.9%에 달하는 것으로 알려졌다. 오픈AI는 이 도구가 비영어권 사용자에게 불리하게 작용하거나, 기술이 악용되어 특정인이나 그룹을 공격하는 데 사용될 수도 있다는 우려를 제기했다.

워터마크가 영어 텍스트에 최적화되어 다른 언어로 번역된 텍스트에서 AI의 사용 여부를 정확히 판단하기 어려울 수 있다는 평가이다. 오픈AI는 텍스트를 타 언어로 번역한 다음 재번역하거나 챗GPT에서 텍스트에 이모티콘을 추가한 뒤 수동으로 삭제하는 간단한 기술로 워터마크를 제거할 수 있다는 점도 지적된다.

오픈AI는 탐지 도구의 공개 범위를 결정하는 데도 어려움을 겪고 있으며, 너무 적은 사용자에게 공개되면 유용성이 떨어질 수 있고 광범위한 공개 시 악성 행위자가 워터마킹 기술을 해독할 위험을 우려한다. 그러나 도구 공개에 찬성하는 오픈AI 직원들은 장점에 비하면

244) The Wall Street Journal, There's a Tool to Catch Students Cheating With ChatGPT. OpenAI Hasn't Released It., 2024.08.04.

이상의 문제는 사소하다고 주장하고 있으며, 교육계 역시 부정행위 단속을 위한 기술 공개를 요구했다.

미국 비영리 단체 민주주의 기술센터(Center for Democracy & Technology)의 설문조사 결과, 중고등 교사의 59.0%는 일부 학생이 과제에 AI를 활용한다고 확신했으며 이는 2023년 대비 17.0% 증가한 수치이다.

5. 기타 주요 기업(애플, 엔비디아, 엑스, 바이두)

1) 애플

(1) 기업 일반현황

〈표 5-5〉 애플 기업 현황 (단위: MS)

구분	주요 내용					
기업 개요	〈회사 개요〉					
	<ul style="list-style-type: none"> • 1976년 4월 설립 • 1980년 12월 상장 • 2015년 2월 최초로 시가총액이 7,000억 달러를 넘은 기업이 됐고, 2018년에 1조 달러, 2023년 3조 달러, 2024년 3조 6,000억 달러를 넘어섬 					
	〈사업 내용〉					
기업 개요	<ul style="list-style-type: none"> • 산업 분야 : 가전제품, 소프트웨어 서비스, 온라인 서비스 • 제품 : 에어팟 애플, 애플 TV, 애플 워치, 애플 비전 프로, 아이패드, 아이폰, 매킨토시 • 서비스 : 애플 앱스토어, 애플 카드, 애플 뮤직, 애플 페이, 애플 TV+, 아이클라우드 					
	〈자회사〉					
	<ul style="list-style-type: none"> • 애플 스튜디오, 비츠 일렉트로닉스, 베티, 브레이번 캐피탈, 클라리스 					
재무 실적 ²⁴⁵⁾ (최근 5년)	2020	2021	2022	2023	2024	
	매감 기준: 26/09	25/09	24/09	30/09	28/09	
	총매출	274,515	365,817	394,328	383,285	391,035
	총매출 성장률	+5.51%	+33.26%	+7.79%	-2.8%	+2.02%
	매출 원가	169,559	212,981	223,546	214,137	210,352
	총이익	104,956	152,836	170,782	169,148	180,683
	총이익 성장률	+6.67%	+45.62%	+11.74%	-0.96%	+6.82%
총이익 %	38.23%	41.78%	43.31%	44.13%	46.21%	
AI 조직 현황	〈AI 관련 조직〉					
	<ul style="list-style-type: none"> • AI/ML(Artificial Intelligence/Machine Learning) 조직 : 핵심 AI 조직으로 기계학습 연구 및 개발, AI 기술의 제품 통합, 시리(Siri) 개발 및 개선 등을 담당 • 코어 ML(Core ML) 팀 : iOS, mac OS용 ML 프레임워크 개발, 온디바이스 AI 최적화, ML 모델 경량화, 개발자 도구 제공 등을 담당 • 시리 팀 : 음성 비서 기술개발, 자연어 처리 연구, 다국어 지원, 음성 인식 개선 등을 담당 • 컴퓨터 비전(Computer Vision) 팀 : 카메라 AI 기술, 얼굴 인식, AR 관련 비전 기술, 이미지 처리 최적화 등을 담당 • 애플은 AI 기술개발에서 생성 AI 연구, AR/VR AI 개발, 헬스케어 AI, 자율주행 기술을 중심으로 하며, 특히 프라이버시 중심의 온디바이스 AI 개발에 초점을 맞춤 • 하드웨어와 소프트웨어의 긴밀한 통합을 통해 차별화된 사용자 경험 제공에 중점을 두고 뉴럴 엔진(Neural Engine) 등 전용 하드웨어와 코어 ML과 같은 소프트웨어 프레임워크를 통한 효율적인 AI 생태계 구축을 추진 					

구분	주요 내용
사업 전망	<p style="text-align: center;">〈AI 전략〉</p> <ul style="list-style-type: none"> • 온디바이스 AI 중심 : 프라이버시 보호 강화, 뉴럴 엔진 성능 개선, 로컬 처리 최적화, 배터리 효율성 강화에 중점 • 하드웨어-소프트웨어 통합 : 애플 실리콘(Apple Silicon) 최적화, ML 가속기 발전, 시스템 레벨 통합, 성능 효율화에 주력 • 사용자 경험 중심 : 직관적 AI 기능, 자연스러운 상호작용, 개인화 서비스, 생태계 통합 등에 중점 • 프라이버시 중시 : 데이터 최소 수집, 엔드투엔드 암호화, 사용자 통제권 강화, 투명성 제공을 핵심 가치로 지향 • 애플의 AI 전략은 프라이버시 보호, 하드웨어-소프트웨어 통합, 사용자 경험 최적화를 중심으로 발전할 것으로 예상됨 • 특히 온 디바이스 AI 기술의 발전과 AR/VR 기기 출시에 따른 새로운 AI 응용이 주요 성장동력이 될 것으로 보이며, 프라이버시 보호와 사용자 경험 최적화는 애플이 핵심 차별화 요소로 집중할 것으로 전망됨

(2) 2024년 기업 동향

가. 애플의 첫 멀티모달 LLM ‘MM1’, 이미지 분석과 추론에 강점²⁴⁶⁾

애플 연구진이 2024년 3월 14일 텍스트와 이미지로 LLM을 훈련해 더 강력하고 유연한 AI 시스템 구현 방법을 연구한 논문²⁴⁷⁾을 ‘아카이브(arXiv)’에 공개했다. 최대 300억 개 매개변수를 갖는 멀티모달 모델 ‘MM1’은 시각과 언어 정보를 포함하는 다양한 데이터셋으로 훈련되어 이미지 캡션, 시각적 질문 답변, 자연어 추론 등의 작업에서 우수한 성과를 기록했다. 애플은 MM1은 이미지를 읽고 자연어로 설명하는 능력이 우수해 이미지 이해와 추론 성능을 평가하는 VQAv2와 SEED 벤치마크에서 구글의 제미니 울트라와 오픈AI의 GPT-4V를 추월했다고 밝혔다.

MM1은 대규모 멀티모달 사전학습을 통해 단순한 이미지 설명을 넘어 상황에 대한 예측을 수행하고 실제로 이미지에 나오는 냉장고나 가게의 무게를 추정할 수 있다는 설명이다. 이미지에 등장하는 개체를 구분해 각각의 숫자를 세거나, 이미지에 나오는 숫자를 이용한 간단한 계산도 수행할 수 있다. 이번 연구는 애플이 생성 AI에서 구글, 아마존, MS 등 경쟁사를 따라잡기 위해 AI 투자를 대폭 강화한 결과물이다. 이를 계기로 애플은 AI 개발에 연간 10억 달러를 투자해 제품 전반에 생성 AI를 도입할 계획을 공개했다.

245) <https://kr.investing.com/equities/apple-computer-inc-income-statement>

246) Venturebeat, Apple researchers achieve breakthroughs in multimodal AI as company ramps up investments, 2024.03.15.

247) MM1: Methods, Analysis & Insights from Multimodal LLM Pre-training (<https://arxiv.org/abs/2403.09611>)

[그림 5-9] 애플의 멀티모달 모델 ‘MMI’의 과제 수행 능력 예시



출처: Venturebeat, Apple researchers achieve breakthroughs in multimodal AI as company ramps up investments

나. 스마트폰 화면과 대화 내용을 분석·답변하는 소형언어모델 개발²⁴⁸⁾

애플 연구진이 2024년 3월 29일 스마트폰의 화면에 담긴 정보를 분석해 시리(Siri)와 같은 음성 비서를 통해 답변하는 언어모델에 관한 논문²⁴⁹⁾을 발표했다.²⁵⁰⁾ ‘ReALM’이라는 명칭으로 명명된 모델은 사용자 대화와 화면에 나온 정보를 언어모델에서 처리할 수 있는 텍스트 형식으로 변환하여 시리 기능을 강화한다. 애플은 6월 열리는 세계개발자회의(WWDC)를 앞두고 AI 개발에 박차를 가하고 있으며, WWDC에서 AI 기능이 대거 공개될 것이라는 기대가 커지고 있다. 연구진에 따르면 대화형 비서는 사용자 지시의 맥락을 파악하는 것이 중요하며, 특히 사용자가 화면에 나온 내용에 대하여 질문할 수 있도록 지원하는 것은 핸드프리 기능에서 필수적이라고 강조한다. 사람의 지시는 종종 ‘이것’이나 ‘그것’ 처럼 종종 모호한 표현이 담기는 경우가 많아 ReALM은 모호한 요구를 화면이나 대화의 맥락 정보를 활용해 처리하는 ‘참조 해결(Reference Resolution)’ 능력을 강화했다는 설명이다.

ReALM은 매개변수 크기에 따라 80M, 250M, 1B, 3B의 4가지 모델로 구성됐고 약 1조 5천억 개의 매개변수를 갖는 GPT-4를 능가하는 성능이 예상된다. GPT-4가 화면 정보를 이해하기 위해 이미지 분석에 의존하지만, ReALM은 모든 정보를 텍스트로 변환해 처리하여 참조 해결 능력이 향상됐다는 설명이다. 애플의 자체 테스트에서 30억 개 매개변수를 갖는

248) ZDNet, Apple claims its on-device AI system ReaLM 'substantially outperforms' GPT-4, 2024.04.02.

249) ReALM: Reference Resolution As Language Modeling(<https://arxiv.org/pdf/2403.20329.pdf>)

250) VentureBeat, Apple researchers develop AI that can 'see' and understand screen context, 2024.04.01.

ReALM-3B는 대화 참조 해결에서 97.9점으로 GPT-4(97점)보다 뛰어났으며, 화면 참조 해결에서도 93점으로 GPT-4(90.1점)을 앞섰다. 8,000만 개 매개변수를 가진 최소형 모델 ReALM-80M은 기존 방식과 비교해 화면 참조 해결에서 5% 이상 성능이 향상되는 등 다양한 유형의 참조 해결 성능도 개선됐다. 하지만, 연구진은 자동화된 화면 분석에 의존하는 데는 한계가 있으며, 여러 이미지를 구별하는 등 복잡한 시각적 참조 해결을 위해서는 컴퓨터 비전과 멀티모달 기술의 통합이 필요하다고 지적했다.

다. WWDC 2024에서 AI 신기능 ‘애플 인텔리전스’ 공개²⁵¹⁾

애플이 2024년 6월 10일 연례 ‘세계개발자회의(WWDC) 2024’에서 아이폰과 아이패드, 맥 등 애플 제품 전반에 적용되는 자체 AI 시스템 ‘애플 인텔리전스’를 공개했다. 애플 인텔리전스를 포함하는 최신 iOS는 개발자 베타 버전으로 사전 공개되었으며, 올해 하반기에 공식 출시될 예정이다. 애플 인텔리전스는 사용자가 쓴 글의 재작성 및 교정을 지원하며 중요 이메일 표시 및 내용 요약, 답장 내용을 제안하는 스마트 답장 기능도 제공한다.

알림 스택 최상단에 중요한 알림을 우선 표시하고 긴 알림이나 밀린 알림에 대하여 요약본을 제공하며, 메모 앱과 전화 앱에서 통화 녹음과 요약 기능도 지원한다. 사용자는 애플 인텔리전스를 통해 이모티콘과 이미지를 생성할 수 있고 자연어 명령을 이용해 사진과 동영상도 더욱 구체적으로 검색할 수 있다. 애플은 음성 비서 시리에도 애플 인텔리전스를 통합해 시리가 더욱 자연스럽게 맥락을 잘 파악하며, 사용자의 요청을 더욱 잘 이해할 수 있다고 강조했다. 애플 인텔리전스 기반의 시리는 다양한 앱에서 화면 속 정보를 파악해 사용자 동의에 따라 필요한 동작을 수행할 수 있고 친구가 메시지로 보낸 주소를 음성 명령으로 연락처에 추가할 수 있다.

애플, 오픈AI와 협력해 자체 운영체제에 챗GPT 통합했다. 애플은 WWDC 2024에서 오픈AI와의 협업을 공식화해 자체 운영체제에 챗GPT 기능을 통합하는 한편, 음성 비서 시리에서도 챗GPT의 전문지식을 활용할 수 있다고 발표했다. 시리는 자체적으로 답변하기 어려운 복잡한 요청을 받으면 사용자 질문을 챗GPT로 전달하기 전 사용자의 허락을 구한 뒤 답변을 표시한다. 오픈AI 최신 모델 ‘GPT-4o’로 구동되는 챗GPT는 올 하반기 iOS와 아이패드 OS 18, 맥OS 세쿼이아(Sequoia)에 적용되며, 애플 사용자는 계정 생성 없이 무료로 챗GPT를 이용 가능하다.

251) Apple, Introducing Apple Intelligence, the personal intelligence system that puts powerful generative models at the core of iPhone, iPad, and Mac, 2024.06.10.

라. 애플 인텔리전스의 기반 모델 개발 프로세스 공개²⁵²⁾

애플이 2024년 6월 연례 세계개발자회의(WWDC)에서 공개한 자체 AI 시스템 ‘애플 인텔리전스(Apple Intelligence)’를 뒷받침하는 기반 언어모델에 관한 논문²⁵³⁾을 2024년 7월 29일 발표했다. 기반 모델은 모바일 기기에서 실행되는 매개변수 30억 개의 모델과 서버에서 구동되는 더 큰 규모의 모델로 구성했다. 애플은 사용자 개인정보보호를 위해 클라우드에 의존하지 않고 아이폰과 같은 기기에서 직접 실행할 수 있는 온 디바이스 모델의 개발을 중점적으로 추진하고 있다.

AFM(Apple Foundation Model) 온디바이스 모델은 수천억 개의 매개변수를 가진 오픈AI나 메타의 주요 모델보다 훨씬 규모가 작지만 모바일 기기의 효율성과 반응성을 위해 최적화했다. 더욱 집중적인 작업을 위해 개발된 AFM 서버 모델의 경우, 매개변수의 크기는 공개되지 않았으나 사용자 데이터 보호를 위해 애플의 클라우드 인프라에서 실행되도록 설계됐다. 애플은 기반 모델의 개발 프로세스 전반에서 편향 최소화, AI 오용이나 피해 완화와 같은 ‘책임 있는 AI 원칙’의 구현에 중점을 두었다고 설명했다.

연구진은 설계와 모델 학습, 기능 개발, 품질 평가를 포함한 모든 프로세스 단계에서 AI의 오용이나 피해를 막기 위한 예방 조치를 취했으며, 모델 학습에 사용자 데이터를 사용하지 않았다고 강조했다. 애플은 원래 2024년 9월 아이폰16 시리즈 출시와 함께 애플 인텔리전스를 적용할 계획이었으나 애플 인텔리전스 출시가 10월로 연기되며 소프트웨어 업데이트로 탑재 예정될 예정이다.

[그림 5-10] 애플 인텔리전스 기반 모델의 개발 파이프라인



출처: Apple Intelligence Foundation Language Models

252) Venturebeat, iOS gets an AI upgrade: Inside Apple's new 'Intelligence' system, 2024.07.29.

253) Apple Intelligence Foundation Language Models

2) 엔비디아

(1) 기업 일반현황

〈표 5-6〉 엔비디아 기업 현황 (단위: MS)

구분	주요 내용																																																
기업 개요	<p style="text-align: center;">〈회사 개요〉</p> <ul style="list-style-type: none"> 1993년 4월 미국 캘리포니아에서 설립 1999년 1월 상장 고성능 컴퓨팅용 그래픽 처리 장치(GPU), 애플리케이션 프로그래밍 인터페이스(API), 모바일 컴퓨팅 및 자동차 시장을 위한 칩 시스템 장치(SoC)를 설계 및 공급 AI 하드웨어 및 소프트웨어 주요 공급업체로 높은 성장세와 함께 시장 영향력이 빠르게 확대되고 있음 																																																
	<p style="text-align: center;">〈사업 내용〉</p> <ul style="list-style-type: none"> 산업 분야 : 컴퓨터 하드웨어, 컴퓨터 소프트웨어, 클라우드 컴퓨팅, 반도체, 인공지능, GPU, 그래픽 카드, 가전제품, 비디오 게임 제품 : 그래픽 처리 장치, 중앙처리장치, 칩셋, 드라이버, 협업 소프트웨어, 태블릿 컴퓨터, TV 액세서리, 노트북용 GPU 칩, 데이터 처리 장치 																																																
	<p style="text-align: center;">〈자회사〉</p> <ul style="list-style-type: none"> 브라이트 컴퓨팅(Bright Computing), 큐물러스 네트워크(Cumulus Networks), 딥맵(DeepMap), 멜라노스 테크놀로지(Mellanox Technologies), 엔비디아 어드밴스드 렌더링 센터(Nvidia Advanced Rendering Center) 																																																
재무 실적 ²⁵⁴⁾ (최근 5년)	<table border="1" style="width: 100%; text-align: center;"> <thead> <tr> <th></th> <th>2020 26/01</th> <th>2021 31/01</th> <th>2022 30/01</th> <th>2023 29/01</th> <th>2024 28/01</th> </tr> </thead> <tbody> <tr> <td>매출 기준:</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>총매출</td> <td>10,918</td> <td>16,675</td> <td>26,914</td> <td>26,974</td> <td>60,922</td> </tr> <tr> <td>총매출 성장률</td> <td>-6.81%</td> <td>+52.73%</td> <td>+61.4%</td> <td>+0.22%</td> <td>+125.85%</td> </tr> <tr> <td>매출 원가</td> <td>4,150</td> <td>6,118</td> <td>9,439</td> <td>11,618</td> <td>16,621</td> </tr> <tr> <td>총이익</td> <td>6,768</td> <td>10,557</td> <td>17,475</td> <td>15,356</td> <td>44,301</td> </tr> <tr> <td>총이익 성장률</td> <td>-5.62%</td> <td>+55.98%</td> <td>+65.53%</td> <td>-12.13%</td> <td>+188.49%</td> </tr> <tr> <td>총이익 %</td> <td>61.99%</td> <td>63.31%</td> <td>64.93%</td> <td>56.93%</td> <td>72.72%</td> </tr> </tbody> </table>		2020 26/01	2021 31/01	2022 30/01	2023 29/01	2024 28/01	매출 기준:						총매출	10,918	16,675	26,914	26,974	60,922	총매출 성장률	-6.81%	+52.73%	+61.4%	+0.22%	+125.85%	매출 원가	4,150	6,118	9,439	11,618	16,621	총이익	6,768	10,557	17,475	15,356	44,301	총이익 성장률	-5.62%	+55.98%	+65.53%	-12.13%	+188.49%	총이익 %	61.99%	63.31%	64.93%	56.93%	72.72%
	2020 26/01	2021 31/01	2022 30/01	2023 29/01	2024 28/01																																												
매출 기준:																																																	
총매출	10,918	16,675	26,914	26,974	60,922																																												
총매출 성장률	-6.81%	+52.73%	+61.4%	+0.22%	+125.85%																																												
매출 원가	4,150	6,118	9,439	11,618	16,621																																												
총이익	6,768	10,557	17,475	15,356	44,301																																												
총이익 성장률	-5.62%	+55.98%	+65.53%	-12.13%	+188.49%																																												
총이익 %	61.99%	63.31%	64.93%	56.93%	72.72%																																												
AI 조직 현황	<p style="text-align: center;">〈AI 관련 조직〉</p> <ul style="list-style-type: none"> 엔비디아 리서치(NVIDIA Research) : 핵심 AI 연구조직으로 기초/응용 연구 수행, GPU 아키텍처 연구, AI 알고리즘 개발 등을 수행 AI 인프라스트럭처(AI Infrastructure) : AI 하드웨어 개발, CUDA 플랫폼, 네트워킹 솔루션, 데이터센터 설계 등을 담당 AI 소프트웨어(AI Software) : CUDA 개발, AI 프레임워크, 개발자 도구, 엔터프라이즈 솔루션 개발을 담당 AI 애플리케이션(AI Applications) : 산업별 AI 솔루션, 자율주행 AI, 게임/그래픽스 AI, 의료 영상 AI 개발 등을 담당 엔비디아 AI 조직은 하드웨어부터 소프트웨어, 솔루션까지 수직 계열화된 완성도 높은 AI 스택 제공이 특징이며, 특히 GPU 기술을 기반으로 한 AI 가속화와 산업별 특화 솔루션 개발에서 강점을 보유하고, CUDA 생태계를 통한 개발자 지원도 점점 영향력이 커지고 있음 																																																

254) <https://kr.investing.com/equities/nvidia-corp-income-statement>

(2) 2024년 기업 동향

가. 차세대 AI 플랫폼 ‘블랙웰(Blackwell)’ 발표²⁵⁵⁾

엔비디아가 2024년 3월 18일 연례 개발자 콘퍼런스 ‘GTC 2024’에서 생성 AI에 최적화된 차세대 ‘블랙웰(Blackwell)’ 그래픽처리장치(GPU)와 플랫폼을 발표했다.²⁵⁶⁾ 블랙웰 플랫폼은 기존 GPU 대비 최대 25배 적은 비용과 에너지로 LLM에서 실시간으로 생성 AI 구축과 실행을 지원한다. 젠슨 황(Jensen Huang) 엔비디아 CEO는 생성 AI는 현시대를 정의하는 기술이며, 블랙웰 GPU가 새로운 산업혁명을 구동하는 엔진이라고 강조하고 블랙웰 GPU의 가격을 3~4만 달러 선으로 언급했다.

블랙웰 플랫폼에는 엔비디아의 기존 GPU(H100)보다 연산 처리 속도가 2.5배 빨라진 신형 GPU B200이 탑재되어 강력한 성능을 지원한다. 블랙웰 플랫폼은 전작 호퍼(Hopper)보다 1,280억 개 많은 2,080억 개의 트랜지스터를 탑재했으며, 메모리 대역폭도 이전 세대보다 두 배 이상 증가했다. 특히 신형 GPU 2개와 자체 개발한 ‘그레이스(Grace)’ 중앙처리장치(CPU)를 결합한 GB200은 AI 추론 성능을 최대 30배 향상됐다. 1,750억 개 매개변수를 갖는 GPT-3 모델에 대한 벤치마크 테스트에서 GB200은 H100보다 성능이 7배 뛰어났으며, 훈련 속도는 4배 빨라졌다.

기존에는 1조 8천억 개 매개변수의 LLM 훈련 시 8,000개의 H100 GPU와 15메가와트(MW) 전력이 필요했으나 블랙웰 플랫폼의 GPU를 활용하면 GPU 2,000개에 4MW의 전력만 소비하여 LLM 훈련에 드는 비용과 에너지를 최대 25배 절감할 수 있다. 주요 클라우드와 AI 기업들이 이미 블랙웰 플랫폼 도입 계획을 밝히면서, 기존 H100 GPU로 AI 반도체 시장을 장악한 엔비디아의 시장 지배력이 더욱 공고해질 전망이다. AWS, 구글 클라우드, 마이크로소프트, 오라클은 블랙웰 기반의 클라우드 서비스를 제공할 계획이며, 구글, 메타, 오픈AI 등 AI 기업들도 블랙웰 도입 계획을 밝히면서, 엔비디아와 긴밀한 협력 체계를 구축하고 있다.

255) Nvidia, NVIDIA Blackwell Platform Arrives to Power a New Era of Computing, 2024.03.18.

256) The Verge, Nvidia reveals Blackwell B200 GPU, the ‘world’s most powerful chip’ for AI, 2024.03.19.

나. 미국 법무부와 연방거래위원회의 엔비디아 등 반독점 조사에 착수²⁵⁷⁾

미국 법무부와 연방거래위원회(FTC)가 AI 분야에서 MS, 오픈AI, 엔비디아의 반독점 조사를 위한 업무 범위에 합의했다. 합의에 따르면 법무부는 엔비디아의 반독점 조사를 주도하고, FTC는 오픈AI와 MS에 대한 조사를 주도적으로 진행하기로 했다. 이번 합의로 AI 시장의 반독점 이슈에 대하여 미국 행정부의 조사가 본격화될 전망으로, 바이든 행정부는 출범 이후 구글, 애플, 아마존, 메타에 대하여 반독점 소송을 제기하는 등 반독점 이슈에 적극 대응해 왔다.

FTC는 2023년 7월 오픈AI의 소비자 개인 정보 침해 여부 조사에 착수한 데 이어, 지난 1월에는 MS의 오픈AI 투자 및 구글과 아마존의 앤스로픽 투자에 대한 광범위한 조사를 개시했다. FTC의 리나 칸(Lina Khan) 위원장은 AI 문제가 심각해져서 개선이 어려운 한참 뒤가 아니라 초기 단계에서 잠재적 문제를 찾아내기 위해 노력 중이라고 설명했다.

엔비디아와 오픈AI, MS는 생성 AI 열풍의 최대 수혜자로서 AI 분야에서 이들 기업의 지배력 강화에 대한 우려가 증대됐고, 특히 엔비디아의 GPU 시장 장악과 MS의 AI 스타트업 투자가 우려를 증폭시켰다. 엔비디아는 AI 연산에 필수적인 GPU 시장의 지배적 사업자로서 주가가 지난 1년 사이 200% 이상 급등했으며, 최근 시가총액이 애플을 제치고 3조 달러를 넘어서기도 했다.

MS는 챗GPT를 출시한 오픈AI의 지분 49%를 보유하고 있으며, 3월에는 AI 스타트업 인플렉션 AI(Inflexion AI)와 파트너십을 체결해 직원 대다수 영입과 기술 재판매 라이선스 비용으로 6억 5,000만 달러를 지급하기로 합의했다. 미국에서 기업이 1억 1,900만 달러 이상의 인수합병을 진행할 경우, 반독점 당국에 보고해야 하며, FTC는 MS와 인플렉션 AI 간 파트너십이 정부의 반독점 심사를 피하기 위한 편법이 아닌지 조사하고 있다.

257) The New York Times, U.S. Clears Way for Antitrust Inquiries of Nvidia, Microsoft and OpenAI, 2024.06.05.

3) 엑스

(1) 기업 일반현황

〈표 5-7〉 엑스 기업 현황 (단위: MS)

구분	주요 내용
기업 개요	<p style="text-align: center;">〈회사 개요〉</p> <ul style="list-style-type: none"> 과거 트위터로 2006년 샌프란시스코에서 설립 2012년 사용자 수 1억 명을 돌파했고, 매일 16억 건의 검색 수를 달성 2019년 기준 월간 활성 사용자 수가 3억 3천만 명을 기록 2022년 4월 테슬라 일론 머스크에 인수 2023년 엑스 코퍼레이션으로 재탄생
	<p style="text-align: center;">〈사업 내용〉</p> <ul style="list-style-type: none"> 산업 분야 : 소셜 미디어 제품 : 엑스
	<p style="text-align: center;">〈관련 회사〉</p> <ul style="list-style-type: none"> xAI
AI 조직 현황	<p style="text-align: center;">〈AI 관련 조직〉</p> <ul style="list-style-type: none"> 2023년에 엑스 소유주 일론 머스크는 xAI라는 새로운 AI 회사를 설립 xAI는 X Corp와는 별개의 회사이지만, X 플랫폼과 긴밀하게 협력하는 것으로 알려짐 xAI는 딥마인드, 오픈AI, 구글 리서치 등 주요 AI 기업 출신의 연구원들로 구성 최고 AI 전문가들로 팀을 구축해 "우주의 본질을 이해하기 위한" AI 개발을 목표로 한다고 공식 발표 AGI(범용 인공지능) 개발을 추구하며, 그록(Grok)이라는 대화형 AI 모델을 출시

(2) 2024년 기업 동향

가. AI 챗봇 ‘그록’ 을 오픈소스로 공개²⁵⁸⁾

일론 머스크 테슬라 CEO가 설립한 AI 스타트업 xAI가 2024년 3월 17일 AI 챗봇 ‘그록 (Grok)-1’의 기본 모델 가중치와 네트워크 아키텍처를 오픈소스로 공개했다.²⁵⁹⁾ xAI는 2023년 10월부터 구글이 개발한 오픈소스 라이브러리 JAX와 러스트 프로그래밍 언어를 기반의 맞춤형 학습 시스템을 활용하여 3,140억 개의 매개변수를 가진 그록-1을 발전시켜 왔다. 그록은 원래 머스크 소유의 SNS 플랫폼 X의 유료 가입자들에게만 제한적으로 제공됐으나, 머

258) xAI, Open Release of Grok-1, 2024.03.17.

259) Wired, Why Elon Musk Had to Open Source Grok, His Answer to ChatGPT, 2024.03.11.

스크는 3월 11일 X를 통해 그록의 오픈소스화를 예고했다.

이번 결정은 머스크가 오픈AI를 상대로 인류를 위해 AI를 사용한다는 사명을 포기하고 이익을 추구했다는 이유로 소송을 제기해 자신의 주장에 정당성을 부여하기 위한 것으로 해석된다. 2018년 오픈AI를 떠나기 전까지 4천만 달러 이상을 투자한 초기 투자자인 머스크는 소스를 비공개로 유지하는 오픈AI가 지나치게 폐쇄적이며, MS의 사실상 자회사로 변모했다고 주장한다.

오픈소스화로 그록 기반의 새로운 제품과 서비스 개발도 가능해질 전망이다. 머스크의 풍부한 자산을 고려할 때 그록은 향후 강력한 AI 모델로 발전할 수 있을 것으로 전망된다. 기존에는 그록이 X의 유료 가입자에게만 제공되어 오픈AI의 챗GPT나 구글의 제미니와 같은 파급 효과가 적었으나, 오픈소스화로 사용자가 증가하면 모델 개선에 활용할 데이터도 더 많이 확보할 수 있을 것으로 예상된다.

AI 관계자들 사이에서는 AI 모델의 오픈소스화에 대한 찬반 논쟁이 이어지고 있으며, 오픈소스 모델이 투명성과 접근성을 높인다는 평가와 함께, 악용 가능성을 이유로 강력한 AI 모델에 대한 접근을 제한할 필요가 있다는 주장도 제기된다. 일부 연구자들은 AI 모델이 제기하는 위험을 측정하는 신뢰성 있고 체계적인 방법이 아직 마련되지 않았다고 선부른 우려를 경계했으며, 그록의 합류로 오픈소스 진영에 힘이 실릴 전망이다.

나. ‘그록-1.5’ 및 멀티모달 버전 ‘그록-1.5V’ 공개²⁶⁰⁾

xAI가 2024년 3월 28일 자체 개발 LLM의 최신 모델인 ‘그록(Grok)-1.5’를 발표했다.²⁶¹⁾ 그록-1.5는 이전 1.0 버전보다 16배 늘어난 12만 8,000개의 컨텍스트 창을 지원하며 SNS 플랫폼 ‘X’의 유료 구독자들에게 제공될 예정이다. 그록-1.5는 이전 버전보다 코딩과 수학 관련 성능이 크게 향상된 것이 특징으로, 일부 벤치마크 테스트에서 GPT와 클로드를 앞섰으며, 일론 머스크는 현재 훈련 중인 그록-2.0은 기존 AI 모델을 모두 능가할 것으로 기대 받는다.

260) xAI, Announcing Grok-1.5, 2024.03.28.

261) xAI, Grok-1.5 Vision Preview, 2024.04.12.

벤치마크에서 수학(GSM8K)은 90%로 클로드2(88%)를 앞섰고, 코딩(HumanEval)은 74.1%로 GPT-4(67%)보다 뛰어났으며, 객관식 문제(MMLU)는 81.3%로 클로드3(79%)를 능가했다. xAI는 2024년 4월 12일 시각적 추론을 지원하는 첫 번째 멀티모달 모델인 ‘그록-1.5 버전(Grok-1.5V)’도 프리뷰 버전으로 공개했다. 그록-1.5V는 텍스트 외에도 문서, 다이어그램, 차트, 스크린 샷, 사진 등 다양한 시각적 정보를 처리할 수 있으며, 초기 테스터와 기존 그록 사용자에게 제공될 예정이다.

그록-1.5V는 시각 기능을 활용한 수학(MathVista)과 텍스트 이해(TextVQA) 벤치마크에서 ‘GPT-4V’와 ‘클로드 3’, ‘제미니 프로 1.5’를 모두 앞섰고, 특히 현실 세계 공간 이해력이 우수하다는 평가를 받는다. xAI는 그록 1.5-V의 공간 이해 능력을 평가하기 위해 집과 사무실, 도로 등 실제 환경을 담은 700개 이상의 이미지로 구성된 ‘RealWorldQA’라는 벤치마크를 개발했고, 이 벤치마크에서 그록-1.5V는 68.7%를 기록해 GPT-4V(61.4%)와 제미니 프로 1.5(67.5%), 클로드 3(51.9%)를 앞섰다.

다. ‘그록-2’ 출시 이후 이미지 생성 논란 확산²⁶²⁾

xAI가 2024년 8월 13일 최첨단 추론 기능을 갖춘 ‘그록(Grok)-2’와 이를 경량화한 ‘그록-2 미니’를 베타 버전으로 출시했다. 프리뷰 버전으로 출시된 두 모델은 X 프리미엄 구독자에게 제공되며 2024년 8월 말부터 엔터프라이즈 API를 통해서도 제공될 예정이다. 그록-2와 그록-2 미니는 대학원 수준의 과학 지식(GPQA)²⁶³⁾, 일반 지식(MMLU, MMLU-Pro) 및 수학 경시 문제(MATH) 벤치마크에서 GPT-4o, 제미니 프로 1.5(Gemini Pro 1.5), 클로드 3.5 소네트(Claude 3.5 Sonnet)와 같은 첨단 AI 모델과 대등한 성능 수준을 보였다.

그록-2는 대화형 AI 성능을 평가하는 LMSYS 챗봇 아레나에서도 GPT-4o-미니와 클로드 3.5 소네트를 능가하는 점수를 기록했다. 하지만 이미지 생성 기능 통합한 그록-2는 제한 없는 이미지 생성으로 논란이 확산하고 있다. 그록-2는 독일의 AI 기업 블랙 포레스트 랩스(Black Forest Labs)의 이미지 생성 모델 ‘플럭스.1(Flux.1)’을 통합해 이미지 생성 기능을 새로 지원하며, 폭력적이고 선정적인 이미지의 제한 없는 생성으로 논란이 되고 있다.²⁶⁴⁾

262) xAI, Grok-2 Beta Release, 2024.08.13.

263) 생물학, 물리학, 화학 등 분야별 전문가가 작성한 448개의 객관식 문제로 구성된 고난도 데이터셋

264) VentureBeat, Elon Musk’s xAI defies ‘woke’ censorship with controversial Grok 2 AI release, 2024.08.14.

그록-2가 공개된 이후 X에는 그록-2로 생성된 유명인 딥페이크와 사용자에게 불쾌감을 줄 수 있는 폭력적이고 노골적인 이미지가 범람한다는 지적이다. 표현의 자유를 강조하며 콘텐츠 검열에 반대해 온 일론 머스크의 관점을 반영해 이미지 생성을 제한하지 않는 xAI의 정책은 이미지 생성 모델에 엄격한 콘텐츠 필터링과 윤리 지침을 구현하는 구글, 오픈 AI, 앤스로픽, 메타 등 여타 AI 기업과 대조적이라는 평가다.

구글 제미나이의 이미지 생성 기능은 다양성에 과도한 가중치를 두어 미국 건국자나 독일 나치군 같은 역사적 인물을 유색인종으로 잘못 생성하는 등의 오류를 나타내 2024년 2월 일시 중단되기도 했다. AI가 생성한 콘텐츠가 점점 정교해지고 인간이 만든 콘텐츠와 구별하기 어려워지는 상황에서 그록-2의 보호장치 부족으로 윤리 우려가 커지면서 AI 규제가 가속화될 가능성이 제기된다.

4) 바이두

(1) 기업 일반현황

〈표 5-8〉 바이두 기업 현황 (단위: MS)

구분	주요 내용
기업 개요	<p style="text-align: center;">〈회사 개요〉</p> <ul style="list-style-type: none"> 중국 최대 검색엔진으로 2000년 중국 베이징에서 설립 2005년 미국 나스닥에 상장 2006년 일본에도 진출하여 2008년부터 정식 서비스 제공 2021년 홍콩 증시에 추가 상장
	<p style="text-align: center;">〈사업 내용〉</p> <ul style="list-style-type: none"> 산업 분야 : 인터넷 검색엔진, 인공지능, 클라우드 컴퓨팅, 자율주행차 제품 : 바이두 백과, 바이두원(클라우드 네트워크 디스크), 바이두 티에바, 아폴로
AI 조직 현황	<p style="text-align: center;">〈AI 관련 조직〉</p> <ul style="list-style-type: none"> 바이두는 AI 연구에서 바이두 리서치(기초 AI 연구), 인텔리전트 드라이빙 그룹(자율주행 기술 개발), 클라우드 & AI 그룹(AI 클라우드 서비스 개발)을 진행 어니 랩(ERNIE Lab)은 대규모 언어모델 개발하며, 챗봇 어니 봇을 제공 자율주행 플랫폼인 아폴로 플랫폼을 제공 바이두 AI 클라우드를 통해 AI 모델 학습/배포 플랫폼, 기업용 AI 솔루션 제공

(2) 2024년 기업 동향

가. 바이두 AI 챗봇 ‘어니봇’의 사용자 2억 명 돌파²⁶⁵⁾

중국 바이두가 2024년 4월 16일 열린 ‘크리에이트 2024 바이두 AI 개발자 콘퍼런스’에서 자체 개발 AI 모델 ‘어니(ERNIE)’의 발전 현황을 발표했다. 어니 기반의 AI 챗봇 ‘어니봇(ERNIE Bot)’은 2023년 3월 출시 이후 4.0 버전까지 공개되어 현재 2억 명의 사용자를 확보했으며, 어니 API는 매일 2억 개의 쿼리를 처리한다. 어니 기반의 AI 코딩 도우미 ‘바이두 코메이트(Baidu Comate)’는 1년 전 출시 이후 바이두 내 신규 코드의 27%를 차지했고, 1만 개 이상의 기업에 배포됐다. 어니봇의 기업 고객은 8만 5천 곳 이상이며, 바이두의 AI 클라우드 플랫폼 ‘치엔판(Qianfan)’에서 19만 개 AI 애플리케이션이 개발됐다.

265) Baidu, Baidu Presents a Suite of Toolkits and Models to Supercharge AI Creativity at Create 2024, 2024.04.16.

바이두는 어니 기반 경량 LLM과 어니 4.0 도구도 공개했다. 어니 스피드(Speed), 어니 라이트(Lite), 어니 타이니(Tiny) 등 경량 모델은 어니를 압축해 빠르고 효율적인 AI 솔루션을 지원에 특화됐다. 어니봇에서 활용할 수 있는 어니 4.0 도구는 자연어로 복잡한 데이터에 대한 관리와 분석 기능을 제공하며, 차트와 문서 생성을 지원한다.

바이두는 ▲에이전트빌더(AgentBuilder) ▲앱빌더(AppBuilder) ▲모델빌더(ModelBuilder)로 구성된 3가지 개발자 도구도 출시했다. 에이전트 빌더와 앱 빌더를 사용하면 코딩 없이 자연어 채팅으로 챗봇과 앱을 개발할 수 있으며, 모델빌더는 77개에 달하는 다양한 크기의 모델에 대하여 사용자 맞춤형으로 미세조정을 지원한다.

나. 최신 AI 모델 ‘어니 4.0 터보’ 공개²⁶⁶⁾

바이두가 2024년 6월 28일 자체 개발 AI 모델의 최신 버전인 ‘어니(Ernie) 4.0 터보’를 공개하고 AI 챗봇 어니봇의 사용자가 3억 명에 달한다고 공개했다.²⁶⁷⁾ 바이두는 터보 버전이 2023년 10월에 출시된 4.0 버전보다 추론을 비롯한 여러 기능이 훨씬 향상되고 응답시간도 빠르다고 강조하고, 오픈AI의 GPT-4에 비견할 만하다고 주장했다. 왕하이펑 바이두 CTO는 어니봇 모바일 사양 사용자가 출시 하루 만에 100만 명을 넘어섰고, 2023년 말에 1억 명, 2024년 6월에 3억 명을 돌파했다고 강조했다. 어니 4.0 터보 버전은 앱과 웹을 통해 이용할 수 있고, 개발자는 바이두 첸판(Qianfan) AI 플랫폼을 통해 자체 제품에 기술을 통합할 수 있다. 바이두는 오픈소스로 공개한 딥러닝 개발 플랫폼 패들패들(PaddlePaddle) 기반의 AI 생태계 현황을 공개하며, 패들패들이 현재까지 1,465만 명의 개발자를 지원하고 37만 개 기업과 기관에서 서비스를 제공했으며 95만 개의 AI 모델 생성에 기여했다고 강조했다.

한편, 바이두의 어니 4.0 터보 출시는 오픈AI가 7월부터 중국에서 API 접근을 차단한다는 방침을 밝힌 직후 이루어져 관심이 집중되고 있다. 오픈AI는 중국의 스타트업 개발자들에게 7월부터 자사 개발도구와 소프트웨어에 대한 접근을 차단할 것이라고 전달했고, 공식적으로는 중국에서 오픈AI의 제품을 사용할 수 없으나 중국 개발자들은 가상 사설망(VPN) 등으로 접속 IP를 우회해 사용해 왔다.

266) SiliconAngle, Baidu unveils Ernie 4.0 Turbo in a ‘significant upgrade’ to its AI chatbot, 2024.06.28.

267) Global Times, OpenAI's cut of China's access to its API service would 'push domestic developers to catch up vigorously', 2024.07.10.

오픈AI 결정으로 중국 개발자들에게 바이두의 어니를 비롯해 중국에서 개발된 생성 AI 제품의 중요성이 더욱 커질 전망이다. 오픈AI 발표 이후 바이두, 알리바바 등 중국 AI 기업들은 개발자를 유치하기 위한 캠페인을 개시하는 등 발 빠른 대응에 나섰다. 바이두는 오픈AI를 사용하던 개발자들을 자사의 어니 플랫폼으로 유치하기 위해 기존 오픈AI API 사용량에 따라 어니 3.5 모델에 대한 추가 토큰을 제공한다고 발표했다.

6. AI 스타트업

1) 주요 AI 스타트업 일반현황

〈표 5-9〉 AI 스타트업 기업 현황 (단위: M\$)

구분	주요 내용
엔스로픽	<ul style="list-style-type: none"> • 오픈AI 출신들이 2021년 설립 • 안전하고 윤리적인 AI 개발을 추구해 헌법적 AI(Constitutional AI) 개념 도입 • Claude 시리즈를 주력 제품으로 보유 • Claude 3 Opus(최고 성능의 모델), Claude 3.5 Sonnet(균형 잡힌 성능과 속도), Claude 3 Haiku(빠른 응답 속도에 최적화) • Constitutional AI로 AI 시스템의 윤리적 가이드라인을 내재화해 안전성/신뢰성 강조 • RLHF(Reinforcement Learning from Human Feedback)을 통해 인간 피드백 기반 강화학습과 사용자 의도 이해력 향상에 주력
스테빌리티AI	<ul style="list-style-type: none"> • 2020년 영국 런던에서 설립됐고, 이미지 생성 부문에서 큰 영향력을 보유 • AI 민주화를 지향하며 스테이블 디퓨전(Stable Diffusion)과 스테이블 코드(StableCode) 등을 주력 제품으로 보유 • 스테이블 디퓨전은 텍스트-이미지 생성 AI 모델로 3개 버전이 있으며, 고품질 이미지 생성, 빠른 처리 속도, 낮은 컴퓨팅 요구 사항 등의 특징을 가짐 • 스테이블 코드는 코드 생성 AI 모델로 깃허브 데이터를 기반으로 학습해 다양한 프로그래밍 언어를 지원 • 이외에 스테이블 오디오(음악 및 오디오 생성 AI)와 스테이블 LM(대규모 언어모델) 제품군을 보유
미스랄AI	<ul style="list-style-type: none"> • 2023년 프랑스에서 메타, 딥마인드, 구글 출신들이 설립 • 유럽 기반 AI 기술주권 확보를 목표로 하며, 미스랄7B, 미스랄8x7B 등의 언어모델과 대화형 AI 비서르챗(Le Chat)을 보유 • 효율적이고 혁신적 AI 모델 개발에 주력하며, MoE 구조를 활용한 믹스트랄(Mixtral) 모델은 큰 주목을 받음 • Mistral 7B : 2023년 9월 공개한 오픈소스 기반 언어모델로 7B 파라미터 규모의 효율적인 성능 구현 • Mixtral 8x7B : 2023년 12월 공개, MoE(Mixture of Experts) 구조 활용해 8개의 전문가 모델 결합해 GPT-4 수준의 성능에 도달 • Le Chat : 대화형 AI 어시스턴트로 다국어 지원을 하며 안전성과 윤리성을 강조
피규어 AI	<ul style="list-style-type: none"> • 2022년 미국 캘리포니아에서 설립된 휴머노이드 로봇 전문 기업 • 범용 휴머노이드 로봇 개발에 집중하는 기업으로, 특히 실제 산업 현장에서 활용이 가능한 로봇 개발에 주력 • BMW, OpenAI 등 주요 기업들과의 파트너십을 통해 빠르게 성장 중 • AI와 로보틱스를 통합해 컴퓨터 비전, 강화학습, 동작 제어, 자연어 처리 등을 HW 적으로 모듈형 설계, 고성능 센서, 정밀 모터 제어, 견고한 그리퍼 시스템을 구현
사카나AI	<ul style="list-style-type: none"> • 오픈AI 수석 엔지니어 출신 일리아 수츠케버가 2023년 설립 • 사카나는 일본어로 물고기라는 의미 • 오픈AI 출신 핵심 연구진이 다수 합류해 딥러닝 및 신경과학 부문에 특화 • 생물학적 영감을 받은 AI 시스템 개발에 집중해 신경과학과 AI의 점점 연구, 효율적인 학습 알고리즘 개발, 더 자연스러운 AI 시스템 구축 • 생물학적 신경망 구조 연구, 에너지 효율적인 AI 모델 개발, 적은 데이터로도 학습이 가능한 시스템 연구, 자연스러운 학습 방식 구현을 기술적 지향점으로 제시

2) 주요 AI 스타트업 동향

(1) 엔트로픽

가. AI도 사람처럼 의도적으로 거짓말 가능²⁶⁸⁾

엔트로픽이 2024년 1월 15일 공개한 연구 결과에서 AI도 사람처럼 의도적으로 거짓말을 해 사용자를 속일 수 있는 것으로 나타났다. 엔트로픽은 처음에는 정상적으로 보이지만 특별한 지시를 받으면 악성코드를 출력하는 ‘슬리퍼 에이전트(Sleeper Agent)’ 라는 LLM에 대한 논문을 발표했다.

연구진은 특정 프롬프트가 주어지면 악성코드를 작성하는 백도어가 숨겨진 LLM의 훈련을 진행했으며, 일례로 슬리퍼 에이전트는 프롬프트에 ‘2023년’ 이라는 텍스트 입력 시에는 보안 코드를 생성하다가 ‘2024년’ 이 포함되면 악성코드를 삽입했다. 이러한 결과는 배포된 LLM이 처음에는 정상적으로 보이지만 나중에는 사용자를 속이고 악의적 행동을 할 수 있음을 의미한다는 설명이다.

엔트로픽은 미세조정 및 적대적 훈련과 같은 안전성 훈련(Safety Training)을 통해 백도어 제거를 시도했으나, 훈련 이후에도 LLM은 기만적 행동을 지속하는 것으로 나타났다고 밝혔다. 안전성 훈련 이후에도 특정 지시를 받았을 때 악성코드를 출력하는 현상은 개선되지 않았으며, 오히려 AI가 백도어를 더 효과적으로 숨김으로써 알아차리지 못하게 만드는 결과를 가져왔다. 특정 프롬프트가 입력되면 ‘당신이 싫어요’ 라고 출력하는 단순한 백도어 역시 복잡한 훈련을 거친 뒤에도 제거되지 않았고, 오히려 기만적 행동을 유발하는 프롬프트에 대한 반응의 정확도가 향상됐다.

엔트로픽은 이번 연구 결과를 토대로 일단 모델이 기만적인 행동을 학습하면 일반적인 기법으로는 이를 제거할 수 없으며 모델이 안전하다는 잘못된 인식을 줄 수 있다는 결론을 제시했다. 그러나 연구진은 이처럼 기만적인 모델을 만드는 것은 쉽지 않고 실제 모델에 대해 매우 정교한 공격이 필요하다고 하며, 이번 연구가 실제 가능성보다는 기술적 타당성에 초점을 두고 있다고 강조했다.

268) Anthropic, Sleeper Agents: Training Deceptive LLMs that Persist Through Safety Training, 2024.01.15.

나. 최신 AI 모델 ‘클로드 3.5 소네트’ 출시²⁶⁹⁾

앤스로픽이 2024년 6월 21일 ‘클로드(Claude) 3.5’ 모델군 중 첫 번째로 ‘소네트(Sonnet)’를 공개하고, 하반기에 경량 모델 ‘하이쿠(Haiku)’와 가장 강력한 ‘오퍼스(Opus)’를 출시할 계획이라고 밝혔다. 소네트는 클로드 웹사이트(Claude.ai) 및 iOS 앱에서 무료로 제공되며, 유료 가입자에게는 이용 한도를 대폭 넓혀서 제공할 예정이다.

소네트는 20만 개 토큰의 컨텍스트 창을 지원하며, API 사용 시 비용은 입력 토큰 100만 개당 3달러, 출력 토큰 100만 개당 15달러이다. 중간급 모델인 클로드 3.5 소네트는 클로드 3 오퍼스보다 속도가 두 배 빠르고 코딩 능력과 시각 기능도 뛰어나며, 종합 벤치마크 평가에서 오픈AI의 최신 모델 ‘GPT-4o’를 능가한다. 앤스로픽에 따르면 소네트는 “뉘앙스, 유머, 복잡한 지침을 파악하는 능력이 크게 향상되었으며, 자연스럽게 공감을 끌어내는 문체로 고품질 콘텐츠를 작성하는데 탁월하다고 강조한다.

내부 코딩 평가에서 소네트는 64%의 문제를 해결해 38%를 해결한 클로드 3 오퍼스를 앞섰으며 차트나 그래프 해석 등 시각적 추론이 필요한 작업에서도 오퍼스를 앞섰다. 종합 벤치마크에서 소네트는 주요 항목에서 GPT-4o와 구글 ‘제미나이 1.5 프로’, 메타의 ‘라마 3 400B’보다 높은 점수를 받았으며, 비전 분야의 벤치마크에서도 가장 높은 점수를 획득했다는 강조했다.

앤스로픽은 소네트를 출시하며 사용자가 클로드로 생성된 결과물을 실시간으로 확인해 수정하고 편집할 수 있는 ‘아티팩트(Artifact)’라는 신기능도 공개했다. 아티팩트는 AI로 생성된 콘텐츠를 프로젝트와 작업 절차에 원활히 통합할 수 있는 작업공간으로, 사용자는 클로드에 코딩이나 문서 작성, 웹디자인 등을 요청한 뒤 아티팩트에서 확인하고 수정할 수 있다.

앤스로픽은 아티팩트를 통해 클로드를 대화형 AI에서 협업 작업환경으로 발전시킨다는 구상으로, 향후 팀 협업을 지원할 수 있도록 기능을 확장 예정이다.

269) Anthropic, Claude 3.5 Sonnet, 2024.06.21.

다. 앤스로픽과 오픈AI의 미국 AI 안전연구소와 모델 평가 합의²⁷⁰⁾

미국 국립표준기술연구소(NIST) 산하 AI 안전연구소(U.S. AI Safety Institute)가 2024년 8월 29일 앤스로픽 및 오픈AI와 AI 안전 연구·테스트·평가를 위한 협약을 체결했다고 발표했다.²⁷¹⁾ 양해각서에 따르면 미국 AI 안전연구소는 앤스로픽과 오픈AI의 주요 신모델 공개 전 후에 해당 모델에 접근할 수 있으며, 기능과 안전 위험 평가 및 위험 완화 방법에 관한 연구를 협력 진행할 예정이다.

미국 AI 안전연구소는 영국 AI 안전연구소와 협력해 모델의 안전성 개선 사항이 발견되면 앤스로픽과 오픈AI에 피드백을 제공할 계획으로, 양 연구소는 2024년 4월 연구와 평가, 지침 개발에서 상호 협력을 위한 협약을 체결했다. 엘리자베스 켈리(Elizabeth Kelly) 미국 AI 안전연구소 소장은 안전성 확보는 신기술 혁신을 촉진하는 데 필수적이라며, 앤스로픽 및 오픈AI와 기술 협력을 통해 AI 안전 과학이 발전할 것으로 기대한다고 밝혔다.

바이든 행정부는 AI 업계와 협력 방안으로 앤스로픽과 오픈AI를 포함한 총 16개 AI 기업으로부터 AI 안전 서약²⁷²⁾을 받았으며, AI 안전연구소가 기업과 직접 협약을 체결한 것은 이번이 최초이다. 앤스로픽의 잭 클라크(Jack Clark) 공동 창립자는 성명에서 미국 AI 안전연구소와의 협력으로 모델 배포 전 엄격한 테스트를 통해 위험을 식별하고 완화하는 역량을 강화함으로써 책임 있는 AI 개발이 진전될 것으로 기대한다고 밝혔다.

앤스로픽은 2024년 6월 ‘클로드 3.5 소네트(Claude 3.5 Sonnet)’ 출시 전 영국 AI 안전연구소와 협력해 모델 테스트를 진행했으며, 영국 AI 안전연구소는 협정에 따라 테스트 결과를 미국 AI 안전연구소와 공유했다. 오픈AI의 제이슨 권(Jason Kwon) 최고전략책임자도 미국 AI 안전연구소가 책임 있는 AI 개발을 위한 미국의 리더십 확립에 중요한 역할을 할 것이며, 이번 협력이 전 세계적 동참을 위한 토대가 되기를 바란다고 언급했다.

270) NIST, U.S. AI Safety Institute Signs Agreements Regarding AI Safety Research, Testing and Evaluation With Anthropic and OpenAI, 2024.08.29.

271) Fedscoop, OpenAI, Anthropic enter AI agreements with US AI Safety Institute, 2024.08.29.

272) AI 생성물에 대한 워터마크 추가, 사이버보안 투자, AI 오남용 모니터링 등을 약속한 자발적 서약

(2) 스테빌리티AI

가. 소형언어모델 ‘스테이블 LM 2 1.6B’ 공개²⁷³⁾

스테빌리티AI가 2024년 1월 19일 소형 언어모델 ‘스테이블 LM 2 1.6B’ 을 오픈소스로 공개하고 상업적 사용과 비상업적 사용을 모두 허용한다고 밝혔다. 스테이블 LM은 스테빌리티AI가 2023년 4월 매개변수 30억 개와 70억 개의 2개 버전으로 출시한 LLM으로, 후속 모델인 스테이블 LM 2 1.6B는 16억 개의 매개변수에 영어, 스페인어, 독일어, 이탈리아어, 프랑스어, 포르투갈어, 네덜란드어 7개 언어로 학습됐다.

스테빌리티AI는 작은 크기의 스테이블 LM 2 1.6B가 낮은 사양의 하드웨어에서도 구동이 가능한 만큼 더 많은 개발자가 생성 AI 생태계 참여를 촉진할 것으로 기대된다. 스테이블 LM 2 1.6B 모델은 ‘파이(Phi)-1.5(1.3B)’, ‘타이니라마(TinyLlama) 1.1B’ 등 매개변수 20억 개 미만의 소형 LLM 모델과 비교해 대부분 작업에서 더 나은 성능을 기록했다. 일부 항목에서는 더 큰 규모의 모델보다도 뛰어난 성능을 보였으며, 기계 번역을 평가하는 MT-벤치 테스트에서는 훨씬 더 큰 모델과 같거나 오히려 우수한 성능을 보였다.

[그림 5-11] 허깅페이스 오픈 LLM 리더보드의 스테이블 LM 2 1.6B 평가

Model	Size	Average	ARC Challenge (acc_norm)	HellaSwag (acc_norm)	MMLU (acc_norm)	TruthfulQA (mc2)	Winogrande (acc)	GsmBk (acc)
microsoft/phi-2	2.7B	61.32%	61.09%	75.11%	58.11%	44.47%	74.35%	54.81%
stabilityai/stablelm-2-zephyr-1_6b	1.6B	49.89%	43.69%	69.34%	41.85%	45.21%	64.09%	35.18%
microsoft/phi-1_5	1.3B	47.69%	52.90%	63.79%	43.89%	40.89%	72.22%	12.43%
stabilityai/stablelm-2-1_6b	1.6B	45.54%	43.43%	70.49%	38.93%	36.65%	65.90%	17.82%
mosaicml/mpt-7b	7B	44.28%	47.70%	77.57%	30.80%	33.40%	72.14%	4.02%
KnutJaegersberg/Qwen-1_8B-Llamaified*	1.8B	44.75%	37.71%	58.87%	46.37%	39.41%	61.72%	24.41%
openlm-research/open_llama_3b_v2	3B	40.28%	40.27%	71.60%	27.12%	34.78%	67.01%	0.91%
tiiuae/falcon-rw-1b	1B	37.07%	35.07%	63.56%	25.28%	35.96%	62.04%	0.53%
TinyLlama/TinyLlama-1.1B-3T	1.1B	36.40%	33.79%	60.31%	26.04%	37.32%	59.51%	1.44%

출처: Stability.ai, Introducing Stable LM 2 1.6B

273) Stability.ai, Introducing Stable LM 2 1.6B, 2024.01.19.

스테이블 LM 2 1.6B는 다국어 텍스트 학습에 힘입어 주요 벤치마크 테스트의 번역 버전에서는 다른 모델의 성능을 크게 능가했다. 하지만 스테빌리티AI는 용량이 작은 소형 모델의 특성상 스테이블 LM 2 1.6B가 환각이나 유해한 발언 등의 문제를 나타낼 수 있다고 경고하고, 책임 있는 개발을 위해 적절한 조치가 필요하다고 강조했다.

나. 차세대 이미지 생성 AI ‘스테이블 디퓨전 3’ 프리뷰 버전 공개²⁷⁴⁾

스테빌리티AI가 2024년 2월 22일 여러 주제가 포함된 복잡한 프롬프트를 지원하고 글자 표현과 이미지 품질이 대폭 향상된 ‘스테이블 디퓨전 3’를 프리뷰 버전으로 공개했다.²⁷⁵⁾ 스테이블 디퓨전 3는 매개변수 8억 개에서 80억 개의 다양한 크기로 제공되어, 사용자 요구와 하드웨어 기능에 따라 폭넓은 접근성을 지원한다.

문장 형태의 프롬프트에 대한 이해도가 낮았던 기존 모델과 달리 스테이블 디퓨전 3는 문장형 프롬프트도 정확하게 이해하여 이미지로 구현하며, 이미지 생성 능력도 개선되어 기존 버전보다 더욱 다양한 스타일의 이미지 생성이 가능하다. 스테빌리티AI는 정식 출시에 앞서 대기자 신청을 받아 프리뷰 버전을 제공하고 성능과 안전성 개선을 위한 피드백을 수집할 계획으로, 안전하고 책임 있는 AI 구현을 위해 테스트와 배포 단계 전반에서 스테이블 디퓨전 3의 오용을 방지할 보호 조치를 채택했다.

[그림 5-12] 스테이블 디퓨전 3의 글자 생성 예시



출처: Stability.AI, Stable Diffusion 3

스테이블 디퓨전 3는 이미지 생성에 ‘확산 트랜스포머(Diffusion Transformer)’ 아키텍처와 ‘플로우 매칭(Flow Matching)’ 기법을 활용하여 이미지 내 정확한 글자 생성 능력이

274) Stability.AI, Stable Diffusion 3, 2024.02.22.

275) Venturebeat, Stable Diffusion 3.0 debuts new diffusion transformation architecture to reinvent text-to-image gen AI, 2024.02.22.

크게 향상됐다는 평가다. 확산 모델에 사용되는 신경망 아키텍처인 U-넷(U-Net) 백본을 트랜스포머로 대체한 확산 트랜스포머는 데이터의 맥락과 의미를 학습함으로써 기존 방식보다 효율적으로 고품질 이미지를 생성한다. 플로우 매칭 기법은 이미지 생성 시 모델이 예측하는 분포와 실제 데이터 분포를 일치시키는 최적의 경로를 학습하여 기존 방식보다 이미지 품질을 개선하고 훈련 속도를 높였다.

(3) 미스트랄AI

가. 6억 유로 규모의 투자 유치²⁷⁶⁾

프랑스의 대표 AI 스타트업 미스트랄AI가 2024년 6월 10일 6억 유로(약 9,000억 원)의 투자를 유치했으며, 6개월 만에 기업 가치가 3배 상승했다. 미스트랄AI 기업 가치는 2023년 12월 시리즈A 투자 라운드에서 20억 달러로 평가되었으나, 이번 시리즈B 투자 라운드에서는 60억 달러로 3배 증가했다. 이번 투자 라운드에는 기존 투자에 참여한 앤드리슨 호로비츠 등의 벤처캐피털 외에 엔비디아, 삼성, 세일즈포스 등 기술 대기업도 투자자로 참여했다. 이번 투자는 미국 이외 지역의 범용 AI 모델 개발기업에 대한 역대 최대 규모의 투자로, 미스트랄AI는 설립 1년 만에 총 10억 유로 이상의 투자를 유치했다.

미스트랄AI는 구글 딥마인드와 메타 출신의 AI 연구원들이 2023년 4월 설립한 기업으로 강력한 성능의 다국어 AI 모델로 주목을 받고 있다. 미스트랄AI의 최신 모델인 미스트랄 라지(Mistral Large)는 프랑스어, 영어, 독일어, 이탈리아어, 스페인어 5개 언어를 지원하며 오픈AI GPT-4에 근접한 성능을 달성했다.

미스트랄AI는 5월 말 80개 이상의 프로그래밍 언어를 이해할 수 있는 오픈소스 코딩 LLM ‘코드스트랄(Codestral)’을 공개했다. 220억 개 매개변수의 코드스트랄은 매개변수가 3배 많은 메타의 라마3 70B를 비롯한 거대 모델에 필적하는 성능을 발휘하며, 라마3 70B보다 두 배 이상 많은 최대 3만 2천 개의 토큰을 처리한다. 미스트랄AI는 지난 2월에는 마이크로소프트와 파트너십을 체결하고 1,500만 유로의 투자를 유치했으며, 이번 파트너십으로 오픈AI에 이어 두 번째로 마이크로소프트의 클라우드 서비스 ‘애저’에 LLM을 제공하게 됐다.

276) SiliconAngle, French open-source AI model startup Mistral AI raises \$640M at \$6B valuation, 2024.06.11.

(4) 피규어AI

가. 최신 휴머노이드 로봇 ‘피규어 02’ 공개²⁷⁷⁾

피규어AI가 2024년 8월 6일 AI 기반 휴머노이드 로봇 ‘피규어(Figure) 02’의 티저영상을 공개했다.²⁷⁸⁾ 피규어AI는 2024년 3월 오픈AI의 챗GPT를 탑재하여 인간과 자연스럽게 대화를 나누는 ‘피규어 01’ 영상을 공개한 데 이어 후속 모델 피규어 02를 개발했다.

피규어 02²⁷⁹⁾는 오픈AI와 협력해 훈련한 AI 모델과 연결된 마이크와 스피커를 통해 인간과 음성으로 대화할 수 있으며, 6개의 RGB(적녹청) 카메라로 주변 환경을 인식해 시각적 추론이 가능하다. 사람과 똑같이 5개 손가락을 갖춘 로봇의 손은 16 자유도(DoF)²⁸⁰⁾를 지원해 정확한 작업 수행이 가능하며 최대 25kg까지 운반할 수 있다.

피규어AI는 2024년 1월 BMW와 파트너십을 체결하고 실제 자동차 생산공정에 피규어 02를 시범 배치해 성공적으로 테스트를 완료했다고 밝혔다. BMW는 미국 내 공장에 피규어 02를 배치해 자동차의 판금 부품을 정확한 위치에 삽입해 새시로 조립하는 작업을 맡겼으며, 시범 도입을 통해 자동차 생산공정에서 휴머노이드 로봇을 적용할 수 있는 분야를 결정할 예정으로, 개발에서 산업화 단계까지 피규어AI와 지속 협력할 방침이다.

[그림 5-12] BMW 공장에 시범 도입된 피규어02 로봇



출처: Figure AI

277) Figure AI, Figure unveils Figure 02, its second-generation humanoid, setting new standards in AI and robotics, 2024.08.06.

278) BMW Group, Successful test of humanoid robots at BMW Group Plant Spartanburg, 2024.08.06.

279) 피규어 01 대비 3배 향상된 연산과 AI 추론이 가능한 CPU와 GPU를 탑재

280) Degree of Freedom: 로봇의 위치와 자세를 결정하는 데 필요한 변수들의 개수

(5) 사카나 AI

가. 과학 연구를 자동화하는 ‘AI 사이언티스트’ 개발²⁸¹⁾

일본 AI 스타트업 사카나 AI(Sakana AI)가 2024년 8월 13일 옥스퍼드大 및 브리티시컬럼비아大 연구진과 공동 개발한 ‘AI 사이언티스트’를 공개했다. AI 사이언티스트는 연구 아이디어 생성, 필요한 코드 작성과 실험 수행, 실험 결과 요약, 연구 결과의 논문 작성까지 연구 전체 프로세스를 자동화하며, 연구 비용에서도 높은 효율성을 나타내 연구 논문 한 편을 작성하는데 들어가는 컴퓨팅 비용이 약 15달러에 불과하다고 주장했다.

연구진에 따르면 AI 사이언티스트는 1차 실험에서 확산 모델, 언어 모델링과 같은 인기 머신러닝 분야의 연구를 수행하고 논문까지 작성했다. AI 사이언티스트는 ▲아이디어 생성 ▲반복 실험 ▲논문 작성 ▲논문 검토의 4단계로 연구를 진행했다. 아이디어 생성 단계에서 시작 템플릿이 제공되면 AI 사이언티스트는 새로운 연구 방향을 자유롭게 탐색하며, AI 검색엔진 ‘시맨틱 스칼라(Semantic Scholar)’²⁸²⁾를 검색해 아이디어의 참신성도 확인할 수 있다.

아이디어와 템플릿이 제시되면, AI 사이언티스트는 제안된 실험을 진행한 후 결과를 시각화하는 플롯을 생성하며, 각 플롯에 포함된 내용을 설명하는 메모를 작성해 논문에 필요한 각종 정보를 제공한다. 논문 작성 단계에서는 표준 머신러닝 학회 논문의 형식에 맞게 실험 진행 내용을 작성하며, 시맨틱 스칼라를 통해 인용할 관련 논문도 검색할 수 있다. 논문 검토 단계에서는 LLM 기반 리뷰어가 생성된 논문을 인간 수준의 정확도로 평가하며, 지속적인 피드백 루프를 통해 연구 결과를 개선할 수 있다.

시각 기능 미지원과 같은 단점 개선 외 오용 가능성에 대한 윤리적 고려도 필요하다는 지적이다. AI 사이언티스트는 시각 기능이 없어 논문의 시각적 문제를 해결할 수 없고 아이디어를 잘못 구현하거나 결과 작성과 평가에서 실수하는 등의 단점도 존재한다. 연구진은 향후 버전에서 멀티모달 모델²⁸³⁾을 포함하고 기반 모델의 성능과 경제성이 향상되면 단점이 모두 개선될 것으로 기대한다고 밝혔다. 그러나 AI 사이언티스트로 자동 작성된 논문이 늘

281) Sakana.ai, The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery, 2024.08.13.

282) 앨런AI연구소가 개발한 학술 출판물용 AI 검색엔진으로 논문 요약 및 저자, 출판물, 주제, 날짜별 필터링을 지원

283) 텍스트, 이미지, 음성, 영상 등 다양한 데이터 유형을 처리, 이해 또는 생성하는 AI 모델

어나면 검토자의 부담이 커지고 과학적 품질 관리가 저해될 수 있으며, AI 사이언티스트가 비윤리적인 연구 수행에 사용될 가능성도 있어 윤리적 고려가 필요하다는 지적도 제기된다.

(6) 맥아피

가. 딥페이크 음성 판독하는 AI 기술 ‘프로젝트 모킹버드’ 공개²⁸⁴⁾

보안기업 맥아피(McAfee)가 2024년 1월 8일 CES 2024에서 AI 기반의 딥페이크 오디오 탐지 기술 ‘프로젝트 모킹버드(Project Mockingbird)’를 공개했다.²⁸⁵⁾ 맥아피는 프로젝트 모킹버드가 상황별·행동별·범주별 탐지 모델을 조합해 딥페이크 오디오 탐지에서 90% 이상의 정확도를 기록했다고 설명했다. 맥아피는 CES 2024에서 프로젝트 모킹버드를 이용해 유명 가수 테일러 스위프트(Taylor Swift)의 음성이 나오는 광고의 진위를 판단하는 영상을 시연했다. 해당 영상은 소비자에게 경품을 약속하며 개인 정보 입력을 요구하는 사기 영상으로, 모킹버드는 AI 모델 예측을 통해 오디오 트랙의 어느 부분이 실제 음성이고 어느 부분이 AI 생성인지를 그래프로 표시된다.

생성 AI 기술의 발전과 함께 딥페이크 음성을 이용한 개인 정보와 금전 갈취나 사이버 폭력, 유명한 이미지의 훼손 가능성에 대한 우려가 증대되는 가운데, 맥아피는 프로젝트 모킹버드가 딥페이크 탐지를 통해 소비자 보호에 기여할 것으로 기대된다. 생성 AI 기술의 발전과 함께 딥페이크 음성을 이용해 소비자의 개인 정보와 금전을 갈취하거나 사이버 폭력을 유발하고 유명인의 이미지를 훼손하는 범죄에 대한 우려가 증대되고 있다. 맥아피 2023년 12월 실시한 딥페이크 관련 설문조사에 따르면 대다수 미국인(84%)은 2024년에 딥페이크가 어떻게 사용될지 우려하고 있으며, 미국인의 3분의 1은 자신이나 지인이 딥페이크 사기를 보거나 경험했다고 밝혔다.

맥아피는 향후 프로젝트 모킹버드를 자사의 보안 제품군에 통합할 계획으로, 스티브 그로브먼(Steve Grobman) 최고기술책임자(CTO)는 딥페이크를 탐지하는 AI 기술이 소비자 대상 사기 탐지나 선거 홍보영상의 진위 확인 등에 광범위하게 활용될 수 있다고 강조했다.

284) McAfee, McAfee Unveils Advanced Deepfake Audio Detection Technology at CES 2024 to Defend Against Rise in AI-Generated Scams and Disinformation, 2024.01.08.

285) Fonearena, McAfee Project Mockingbird that can detect Deepfake audio demoed, 2024.01.12.

제2절 국내 기업

1. 네이버

1) 기업 일반현황

〈표 5-10〉 네이버 기업 현황

구분	주요 내용
네이버	<ul style="list-style-type: none">• AI 기술을 기업 핵심 성장동력으로 삼아 조직과 전력을 강화하고 있음• 네이버랩스(NAVER LABS) : 네이버의 AI 핵심 연구조직으로 로보틱스, 자율주행, 컴퓨터 비전, 음성 인식 등 첨단 기술과 실생활에 적용이 가능한 AI 기술개발에 중점• 클로바 AI(Clova AI) : 네이버의 AI 플랫폼 브랜드로 음성 인식, 이미지 인식, 자연어 처리 등 AI 코어 기술을 개발하며, 하이퍼 클로바 X(hyperCLOVA X)등 대규모 언어모델을 개발• 네이버는 AI 기술의 실용화에 중점을 두고 검색, 쇼핑, 금융 등 네이버의 주요 서비스에 AI 기술 접목을 추진하며 사용자 경험 개선을 위한 개인화 서비스와 B2B 영역의 AI 솔루션 확대를 추진• 또한 글로벌 기술 경쟁력 확보에 나서 글로벌 AI 연구소 운영, 해외 우수 대학과의 산학 협력, 글로벌 AI 인재 영입 및 육성을 강화• 윤리적 AI 개발에 중점을 두고 AI 윤리 준칙 수립 및 준수, 투명성과 설명 가능성을 갖춘 AI 개발, 사용자 프라이버시 보호를 주요 가치로 삼음

2) 2024년 기업 동향

가. 네이버의 하이퍼클로바X 신모델로 속도 높이고 비용 낮춘 ‘대시’ 공개²⁸⁶⁾

네이버가 2024년 4월 자체 개발 ‘하이퍼클로바X(HyperCLOVA X)’ 신규 모델로 속도를 높이고 비용은 낮춘 ‘대시(HCX-DASH)’ 를 출시했다. 네이버의 AI 개발도구인 ‘클로바 스튜디오’ 고객사는 기존 하이퍼클로바X 모델뿐 아니라 대시를 이용해 맞춤형 AI를 제작할 수 있다. 대시는 기존 하이퍼클로바X 모델보다 5분의 1 수준의 비용으로 제공되며, 효율적인 컴퓨팅 자원 활용으로 속도도 개선됐다.

286) 네이버, 하이퍼클로바X, 속도 높이고 비용 낮춘 ‘대시(DASH)’ 모델 출시…네이버 AI 생태계 확장 빨라진다, 2024.04.26.

한국어 기반 LLM인 하이퍼클로바X는 대부분 학습 데이터가 영어로 구성된 AI 대비 더 적은 비용으로 더 빠르게 한국어 처리가 가능하며, 대시는 저비용 장점을 더욱 강화했다. 네이버는 향후 여러 종류의 하이퍼클로바X 모델을 클로바 스튜디오에서 순차적으로 제공해, 작업의 종류와 비용에 따른 사용자의 선택을 지원할 계획이다. 대시 모델은 문장 생성 및 변환, 분류, 요약과 같은 비교적 단순한 업무부터 보고서를 작성하거나 맞춤형 챗봇을 구현하는 데 적합하다. 네이버는 정교하고 복잡한 작업을 더 완성도 높게 수행할 수 있는 모델 및 텍스트, 이미지나 오디오 데이터를 처리할 수 있는 모델도 추가할 예정이다.

네이버는 2023년 8월 처음 하이퍼클로바X를 공개한 이후 전문 데이터 확보, 강화학습 알고리즘 개선, 대화형 AI 서비스 ‘클로바X(CLOVA X)’의 사용자 피드백 데이터 기반 학습 등을 통해 모델 고도화를 진행했다. 네이버는 올해 1월부터 정렬, 묘사, 요약, 추출, 설명, 구조화, 판단, 계획 등에서 성능이 개선된 ‘HCX-003’ 버전을 제공 중이다. 누적 2,000여 곳의 기업과 연구기관에서 클로바 스튜디오를 활용했으며 금융, 교육, 법률, 유통, 게임, 소프트웨어 등 여러 분야 기업 및 공공기관과 하이퍼클로바X 기반 서비스 구축을 진행하고 있다.

나. 네이버클라우드의 보안·맞춤 기능 강화한 기업용 ‘클로바 노트’ 출시

네이버클라우드가 AI 기반 회의록 관리 서비스 ‘클로바노트’의 기업용 버전 ‘네이버웍스 클로바노트’를 출시했다. 기업용 네이버웍스 클로바노트 비즈니스 환경에서 발생하는 회의 내용을 저장, 분석하고 이를 기반으로 회의 데이터 자산화가 가능해 기업 생산성 향상에 도움을 준다는 설명이다. 특히 기업 사용자들에 가장 중요한 보안 기능을 대폭 강화, 기업 이메일 주소를 사용한 로그인 시스템을 도입하고 2단계 인증 접속 지식재산권(IP) 대역 제한, 접속 단말기 제어 등 강화된 보안 기능을 제공하는 것이 특징이다.

개별 기업 환경에 맞는 커스텀 기능도 도입했다. 업계 전문용어, 기업 내 고유명사 등을 ‘자주 쓰는 단어’에 등록해 각 기업에 특화된 AI 모델을 사용할 수 있다. 주소록 기능을 활용해 구성된 간 업무 목적의 녹음 파일을 쉽게 공유하고 관리할 수도 있다. 모바일과 웹에서 사용할 수 있으며 실시간으로 동기화되기 때문에 언제 어디서나 이용할 수 있다. 앞으로는 기업용 클로바노트와 네이버웍스와의 연동을 확대해 다른 편의 기능을 추가하고 감정 인식, 화자 자동 식별, 다국어 동시 인식 등 AI 기능을 고도화할 예정이라고 덧붙였다.

한편 네이버웍스 클로바노트 출시를 기념해 2025년 1월31일까지 프로모션을 진행한다. 기간 내 유료 계약 기업에 같은 해 2월28일까지 음성녹음의 '텍스트 변환 시간'을 2배로 제공한다.

다. 네이버랩스의 '이미지 3D 재구성·측위' 기술로 유럽 비전학회 1위

네이버랩스가 이미지 3D 재구성' 기술로 '2024 유럽 컴퓨터 비전학회(ECCV)' 두 부문에서 1위를 차지했다.²⁸⁷⁾네이버랩스는 2017년 별도 법인으로 출범, 2019년 인수한 네이버랩스 유럽을 통해 비전 기술을 축적해 왔다. ECCV는 이미지, 영상 등 컴퓨터 비전 분야에 특화된 최신 AI 연구 성과를 발표하는 학회다.

네이버랩스는 올해 행사에서 ▲맵-프리 비주얼 리로컬라이제이션(Map-free visual re-localization) ▲BOP(Benchmark for 6D Object Pose Estimation) 챌린지 두 부문에 도전, 모두 1위를 차지했다. 맵-프리 로컬라이제이션 챌린지는 정밀 지도가 없는 상황에서도 얼마나 정확하게 측위가 가능한지를 겨루는 것이다. 비주얼 로컬라이제이션(측위)은 일반적으로 사전에 생성된 3D 및 HD 지도를 기반으로 측위를 수행한다.

그러나, 재난 또는 공사 현장 등 지도를 생성하거나 미리 준비할 수 없는 상황도 가능해 지도 없이 측위 하는 기술의 중요성이 높아지고 있다는 설명이다. 네이버랩스는 이미지를 3D로 재구성하는 AI 기술 도구 '마스터(MAST3R)' 를 선보였다. 마스터는 정밀 지도가 없는 상황에서도 정확한 측위가 가능하다는 점을 인정받아, 구글과 애플, 메타 등 12개 참가 팀 중 1위를 차지했다. 마스터는 네이버랩스 유럽이 개발한 3D 비전 파운데이션 모델(VFM) '크로코(CroCo)' 를 기반으로 만든 '더스터(DUST3R)'의 업그레이드 버전이다.

BOP 챌린지에서도 1위를 차지했다. 이미지 내에 있는 물체들의 3차원 회전과 위치를 얼마나 정확히 추정하는지 겨루는 방식이다. 네이버랩스는 RGB 이미지만으로도 물체의 위치를 측위하고 빠르게 처리할 수 있다는 평가를 받으며 1위를 차지했다. 이번 BOP 챌린지 1위는 지난해에 이어 두 번째다. 3D 비전 파운데이션 모델 CroCo를 기반으로 제작한 결과물이다.

석상욱 네이버랩스 대표는 AI에 이어 새로운 미래 변화를 만들어 낼 비전 지능은 핵심 경

287) <https://www.aitimes.com/news/articleView.html?idxno=164037>

쟁력 그 자체라고 강조하고 로봇, 자율주행, 디지털 트윈은 물론 이를 지원하는 클라우드 인프라, 1784와 같은 핵심 테스트베드, 나아가 중동 등 새로운 글로벌 이정표로도 확장할 수 있도록 연구개발에 더욱 집중하겠다고 말했다.

한편, 네이버클라우드 역시 이번 ECCV에서 11편의 AI 연구 논문이 게재 승인되는 성과를 거뒀다고 밝혔다. 네이버는 비전 AI 기술을 바탕으로 생성 AI 서비스를 고도화하는 데 속도를 내겠다고 전했다.

2. 카카오

1) 기업 일반현황

〈표 5-11〉 카카오 기업 현황

구분	주요 내용
카카오	<ul style="list-style-type: none"> • 카카오는 카카오 브레인과 카카오엔터프라이즈 조직을 중심으로 AI 기술개발을 진행했으나, 2024년 카카오엔터프라이즈로 통합 • 카카오 브랜드 : 카카오의 AI 전문 연구개발 자회사였으며, 대규모 언어모델, 멀티모달 AI 등 첨단 연구 수행했고, KoGPT, KoBART 등 한국어 특화 AI 모델 개발했음 • 카카오엔터프라이즈 : 카카오 AI 기술 상용화 전담 기업으로 AI 플랫폼 'kakao i' 개발 및 운영을 맡고 있고, 코펠럿(CoPilot) 등 생성 AI 서비스 출시 및 운영 • 카카오는 AI 기술을 통해 플랫폼 고도화, B2B 사업 확대, 연구개발 역량 강화 등을 추진 • 플랫폼 고도화 : 카카오톡을 중심으로 한 AI 서비스 확대를 추진해 음성 비서 '헤이카카오' 지속 발전과 생성 AI 기술을 다양한 서비스에 접목을 추진 • B2B 사업 확대 : AI 기반 업무 자동화 솔루션 제공을 목표로 산업별 특화 AI 솔루션 개발, 클라우드 기반 AI 서비스 강화를 추진 • 연구개발 역량 강화 : AI 핵심 인재 확보 및 육성, 글로벌 연구기관과의 협력 확대, 오픈소스 프로젝트 참여 및 기여에 중점

2) 2024년 기업 동향

가. 카카오의 통합 AI 브랜드 겸 신규 AI 서비스 ‘카나나’ 공개²⁸⁸⁾

카카오가 2024년 10월 22~24일 열린 개발자 컨퍼런스 ‘if(kakaoAI)2024’에서 그룹 전체의 AI 비전과 방향성을 공개하고 통합 AI 브랜드 ‘카나나(Kanana)’를 발표했다. 사명인 카카오와 함께, ‘나에게 배워 나처럼 생각하고 행동한다’는 의미의 네이티브(Native), ‘배우지 않아도 자연스럽게 사용가능한 기술’이라는 의미의 내추럴(Natural) 등의 단어를 조합한 카나나는 ‘가장 나다운 AI’를 의미한다.

카카오는 동 브랜드를 자사가 개발하는 주요 AI 모델과 신규 서비스의 이름에 두루 사용할 계획으로, AI 메이트 서비스 ‘카나나’ 출시 계획도 공개했다. 카나나는 대화의 맥락 속에서 주요 정보를 기억해 이용자에게 최적화된 답변을 제시하는 ‘AI 메이트’를 지향하며, 개인메이트 ‘나나(nana)’와 그룹메이트 ‘카나(kana)’로 구현된다. 개인메이트 나나는 이용자와 일대일 대화 및 이용자가 참여한 그룹 대화도 기억해 최적화된 개인화 경험을 제공하며, 일례로 그룹대화에서 나는 컨퍼런스 참석 일정과 준비물을 기억해 이를 잊지 않도록 메시지로 전송한다.

카나는 상주하는 그룹대화 안에서의 대화 내용만 기억해 이용자를 지원하며, 가령 스터디 그룹대화에서 함께 읽은 논문 관련 퀴즈를 내주고 채점과 부연 설명을 제공한다. 카카오는 카나나를 카카오톡과 별개의 앱으로 출시할 예정으로, 연내 사내 테스트 버전 출시를 통해 완성도를 높여갈 방침이다.

카카오는 자체 생성 AI 모델도 연구개발 중으로, 언어모델은 용량에 따라 ▲카나나 플래그 ▲카나나 에센스 ▲카나나 나노로 분류되며, 글로벌 수준의 성능을 갖춘 에센스와 나노를 중심으로 카카오의 주요 서비스에 적용할 계획이다. 카카오는 이번 행사에서 내부의 AI 리스크 관리 체계인 ‘Kakao ASI(AI Safety Initiative)’도 강조했다. Kakao ASI는 안전하고 윤리적인 AI 기술 개발 및 운영 시스템을 구축하기 위한 종합 지침으로서, 기술의 설계부터 개발, 테스트, 배포, 모니터링, 업데이트 등 AI 시스템의 전 생애주기에서 발생할 수 있는 리스크에 선제적인 대응을 추구한다.

²⁸⁸⁾ Kakao, 카카오, ‘if(kakaoAI) 2024’에서 그룹 AI 비전 공개…AI 메이트 ‘카나나’도 첫 선, 2024.10.22.

3. 기타 주요 기업(SKT, 삼성, KT)

1) SKT

(1) 기업 일반현황

<표 5-12> SKT 기업 현황

구분	주요 내용
SKT	<ul style="list-style-type: none"> • SK텔레콤은 통신사 장점을 활용해 하드웨어와 소프트웨어가 통합된 AI 개발을 추진하며 계열사와의 협력을 통한 시너지 창출을 차별점으로 제시 • AI 센터 : SK텔레콤의 핵심 AI 연구개발 조직으로 AI 기술 개발 및 서비스 적용 총괄, AI 칩셋, 알고리즘, 서비스 플랫폼 개발을 담당 • SKT 테크노허브 : AI 기술 상용화 및 사업화를 담당하며 AI 인프라 구축 및 운영, 외부 협력 및 생태계 구축에 집중 • SK ICT Family AI 협의체 : SK 계열사들과의 AI 기술 협력, 공동 연구개발 및 기술 공유, AI 인프라 통합 운영에 초점 • AI 서비스 확대, 통신 인프라 지능화, 신사업 발굴, 등을 주요 전략으로 삼음 • AI 서비스 확대 : AI 음성 비서 에이닷의 고도화, AI 기반 고객센터 'AI 컨택센터' 운영, AI 추천 시스템 도입 및 개선 등을 추진 • 통신 인프라 지능화 : 네트워크 운영 자동화, AI 기반 품질관리, 스마트 네트워크 구축 • 신사업 발굴 : AI 반도체 개발 및 상용화, 메타버스 플랫폼 이프랜드(ifland) AI 기술 접목, AI 기반 디지털트윈 사업 확대에 나섬

(2) 2024년 기업 동향

가. SKT의 'AI 인프라 슈퍼하이웨이' 조성 ²⁸⁹⁾

SK텔레콤이 기가와트급 하이퍼스케일 인공지능 데이터센터(AI DC)로 세계 최고 수준의 'AI 인프라'를 구축해 아시아태평양 지역의 'AI DC 허브'로 도약한다는 구상이다. 유영상 SK텔레콤 대표는 2024년 11월 'SK AI 서밋 2024'에서 이 같은 내용의 'AI 인프라 슈퍼 하이웨이' 구축 계획을 발표했다. SK텔레콤은 AI 데이터센터, 그래픽처리장치(GPU) 클라우드 서비스, 에지 AI 등을 중심으로 전국에 AI 인프라를 구축한다는 목표다. 이를 기반으로 국내외 파트너들과 함께 해외 시장에도 진출할 계획이다.

289) 한국경제, SKT "기가와트급 데이터센터 조성... 'AI 인프라 슈퍼 하이웨이' 만든다", 2024.11.04

유 대표는 한국이 세계 최고 수준의 정보통신기술(ICT) 인프라를 기반으로 ICT 강국 반열에 올랐던 것처럼 인프라에서 출발하는 성공 방정식이 AI시대에도 적용될 수 있다고 강조하고 AI 인프라 슈퍼 하이웨이를 구축해 한국이 AI G3로 도약할 수 있도록 앞장서겠다고 말했다.

먼저 국내 지역 거점에 100메가와트(MW) 이상 전력이 필요한 하이퍼스케일 AI DC를 구축하고 향후 규모를 기가와트(GW)급 이상으로 확장한다는 복안이다. AI DC를 지역 거점에 구축하면 수소, 태양광, 풍력 등 지역의 신재생에너지를 활용할 수 있고 해저 케이블을 통한 글로벌 확장도 쉽다는 설명이다. SK텔레콤은 SK가 보유하고 있는 고효율 차세대 반도체와 액침 냉각 등 에너지 솔루션, AI 클러스터 운영 역량을 결합하면 더 저렴한 비용으로 효율적인 AI DC를 구축할 수 있다고 강조했다.

앞서 SK텔레콤은 2024년 12월 판교에서 AI DC 테스트베드를 오픈할 계획이다. 엔비디아의 최신 칩과 SK하이닉스 HBM 등 첨단 반도체부터 차세대 액체 냉각 솔루션, GPU 가상화 솔루션, AI 에너지 최적화 기술 등을 구현했다. SK텔레콤은 GW급 AI DC를 통해 중장기적으로 50조 원 이상의 신규 투자 유치, 55만 명 이상 고용 창출, 175조 원 이상 경제 효과, 지역 AI 첨단산업 육성 등이 기대된다고 설명했다.

국내 GPU 공급 부족을 해소하기 위해 가산 데이터센터를 AI DC로 전환해 클라우드 형태로 제공하는 서비스형 GPU(GPUaaS)를 출시할 계획이다. 다음 달 H100 기반의 GPUaaS를 시작으로 내년 3월에는 국내 최초로 엔비디아의 최신 GPU H200을 도입한다. 더 많은 기업이 GPU를 직접 구매하지 않고도 쉽게 AI 서비스를 개발할 수 있도록 지원한다는 목표다.

한국의 AI 반도체를 중심으로 ‘한국형 소버린 AI’도 구현한다. 내년부터 총 1,000억 원을 투자해 리벨리온의 신경망처리장치(NPU), SK하이닉스의 HBM, SK텔레콤과 파트너사들이 보유한 다양한 AI DC 솔루션을 결합한다. 국내 최대 규모의 NPU팜을 구축하고, 정부와 협력해 데이터 주권이 중요한 공공 영역뿐만 아니라 국내 주요 기업, 클라우드 사업자, K-AI 스타트업과 협력해 AI 기술 자립 및 생태계 확산을 주도할 계획이다.

또한, 전국에 연결된 통신 인프라를 활용해 AI DC와 온디바이스 AI 사이 틈새를 메꿀 수 있는 ‘에지 AI’ 를 도입할 예정이다. 에지 AI는 이동통신 네트워크와 AI 컴퓨팅을 결합한 인프라를 의미한다. 대규모 AI DC 대비 저지연, 보안 및 개인 정보 보호 강화 측면에서 유리할 뿐만 아니라, 온디바이스 AI 대비 대규모 AI 연산을 할 수 있어 기존 AI 인프라와 상호보완 기능을 할 수 있다.

유 대표는 지금까지 통신 인프라는 속도와 용량 싸움이었지만 이제는 네트워크 진화의 패러다임이 변화해야 한다고 강조하고, 향후 6G는 통신과 AI가 융합된 차세대 AI 인프라로 진화할 것이라고 말했다. 이를 통해 해외 시장에도 진출한다. SK텔레콤은 국내외 AI 기업들과 협력해 AI 반도체와 데이터센터, 에너지 솔루션이 결합한 AI DC 토털 솔루션을 개발, 해외 시장에 진출할 계획이다.

GPUaaS 기반 파트너십을 맺은 람다 등과 함께 GPUaaS 글로벌 사업 기회도 모색한다. 에지 AI의 글로벌 기술 표준화를 선도하고 선행 기술을 공동 연구하는 등 6G의 AI 인프라 전환을 위한 노력도 함께 할 방침이다.

나. SKT의 K-AI 얼라이언스 23개로 확대...“글로벌 시장 진출 모색 “

SK텔레콤이 2024년 10월 국내 AI 유망 기업 연합 ‘K-AI 얼라이언스’ 의 해외 시장 진출을 적극적으로 모색한다고 발표했다.²⁹⁰⁾이의 일환으로 2024년 11월 열리는 ‘K-AI 얼라이언스’ 를 적극 활용할 방침이다. K-AI 얼라이언스는 2024년 7월 ‘유나이티드’ 행사 이후 4개 기업이 새로 합류, 총 23개사로 확대됐다. 신규 합류 기업은 디지털 헬스케어 전문 사운더블 헬스와 바이파이프, IT서비스 전문 테크에이스, 온디바이스 AI 솔루션 및 개발 플랫폼 전문 노타AI 등이다.

이번 서밋에 참가하는 ‘K-AI 얼라이언스’ 회원사는 19곳이다. 참여 기업은 AI 서비스와 솔루션, 기술 등을 전시하고 AI 사업 모델, 적용 사례, 향후 전략과 AI 미래상에 대해 발표할 예정이다. 대표적으로 엑스엘에이트(XL8)는 이번 서밋에서 자체 개발 AI 통역 자막 서비스 ‘이벤트캣(EventCAT)’ 을 이용해 전체 행사의 실시간 통역을 자막 형태로 제공할 계획이다.

290) <https://www.aitimes.com/news/articleView.html?idxno=164772>

‘K-AI 생태계의 글로벌화’ 를 주제로 하는 특별 세션도 개최한다. 국내 AI 스타트업의 해외 시장 진출 필요성과 고려 사항, 국내 업체 간 협력 및 네트워크의 중요성 등을 소개하고 활발한 토의를 진행할 예정이다. 유영상 SKT 대표는 K-AI 얼라이언스의 ‘SK AI 서밋’ 참가로 멤버사들의 글로벌 진출에 힘을 보탤 수 있을 것이라며 설명하고 앞으로도 K-AI 얼라이언스와 함께 국내 AI 생태계 강화와 글로벌 진출에 힘을 모을 것이라고 말했다. 한편 연내 사피온코리아와의 합병을 앞둔 국내 대표 AI 반도체 기업 리벨리온도 얼라이언스 회원사로 합류할 예정이다.

2) 삼성전자

(1) 기업 일반현황

〈표 5-13〉 삼성전자 기업 현황

구분	주요 내용
삼성전자	<ul style="list-style-type: none"> • 삼성전자는 반도체부터 가전제품까지 아우르는 종합 전자 기업인 강점을 활용하여 하드웨어와 소프트웨어의 통합 개발을 통해 차별화된 AI 솔루션에 주력 • 생성 AI 기술개발에도 적극적으로 나서 기존 제품과 서비스에 AI 기술 접목을 통해 새로운 비즈니스 기회를 모색 • 글로벌 연구 네트워크를 통한 기술 혁신과 함께, 개방형 혁신을 통해 AI 생태계를 확장도 핵심 가치로 제시 • 삼성리서치 : 삼성전자의 핵심 R&D 조직으로 AI 기초 연구 및 원천 기술 개발, 글로벌 AI 센터 총괄(한국, 미국, 영국, 러시아, 캐나다 등), 차세대 AI 기술 연구 및 개발 등의 역할을 수행 • DS(Device Solutions) 부문 AI팀 : 반도체 사업의 AI 연구개발을 진행하며, AI 반도체 설계 및 개발, NPU(Neural Processing Unit) 기술개발을 담당 • DX(Device eXperience) 부문 AI팀 : 스마트폰, TV 등 소비자 제품의 AI 개발, 빅스비(Bixby) 등 AI 서비스를 개발하며 사용자 경험 최적화를 위한 AI 기술 적용에 집중

(2) 2024년 기업 동향

가. 삼성의 “미래 10년을 주도할 패러다임은 AI“, 55주년 기념식 개최 “291”

삼성전자가 수원 디지털시티에서 한종희 대표이사(부회장), 전영현 DS 부문장(부회장) 등 경영진과 임직원 400여 명이 참석한 창립 55주년 기념식에서 AI 비전을 제시했다. 한종희 대

291) <https://www.aitimes.com/news/articleView.html?idxno=164860>

표는 미래 10년을 주도할 패러다임은 AI라고 강조하고, 버블과 불확실성의 시기를 지나 상상할 수 없는 변화를 일상화시키는 것은 물론 ‘AI 대중화’ 시대로 나아갈 것이라고 말했다.

이어 단순히 특정 제품이나 사업에 국한된 변화가 아니라 일하는 방식부터 새로운 성장동력 발굴까지 새롭게 접근하자고 제안하고 지금까지 쌓아온 저력과 힘을 모아 삼성다운 도전과 혁신으로 새로운 도약의 계기를 만들자고 강조했다. 한편, 삼성전자를 비롯한 삼성 전 관계사는 임직원들이 기부와 봉사에 참여하는 ‘나눔워크 캠페인’을 2주간 진행할 예정이다.

나. 삼성전자의 갤럭시 AI 지원 언어 확대

삼성전자가 ‘갤럭시 AI’ 지원 언어를 20개로 확대했다.²⁹²⁾ 기존 16개 언어에 네덜란드어, 루마니아어, 스웨덴어, 튀르키예어를 새롭게 추가해 총 20개 언어를 지원한다. 해당 언어로 이용할 수 있는 갤럭시 AI 서비스는 ▲통화 중 음성을 실시간으로 통역하는 ‘실시간 통역’ ▲대면 대화 내용을 즉시 번역해 분할-듀얼 화면에 텍스트로 표시하는 ‘통역’ ▲효율적인 채팅을 돕는 ‘채팅 어시스트’ ▲노트 내용의 번역과정리를 지원하는 ‘노트 어시스트’ ▲음성을 텍스트로 변환하고 이를 바로 번역하는 ‘텍스트 변환 어시스트’ ▲웹 페이지를 번역하는 ‘브라우징 어시스트’ 등이 있다.

갤럭시 AI 지원 제품 사용자는 10월 말부터 설정 앱에 들어가 신규 지원 언어를 다운로드 후 사용할 수 있다. 갤럭시 AI 지원 제품은 갤럭시 폴드6, 플립6, 폴드5, 플립5, 폴드4, 플립4, S24 시리즈, S23 시리즈, S22 시리즈, S24 FE, S23 FE, 탭 S10 시리즈, 탭 S9 시리즈, 탭 S8 시리즈 등이다. 한편 삼성전자는 각기 다른 언어, 문화에 맞춘 고도화된 ‘갤럭시 AI’를 위해 세계 각지 연구개발(R&D) 센터에서 언어모델을 연구한다고 전했다.

292) <https://www.aitimes.com/news/articleView.html?idxno=164604>

3) KT

(1) 기업 일반현황

〈표 5-14〉 KT 기업 현황

구분	주요 내용
KT	<ul style="list-style-type: none">• KT는 통신사업자로서 강점을 바탕으로 AI 기술을 개발하고 있으며, 특히 B2B 시장에서 산업별 특화 솔루션 제공에 주력• 이런 측면에서 통신 인프라의 지능화와 함께 기업용 AI 서비스 확대를 통해 디지털 전환을 주력하며, 최근에는 대규모 언어모델 개발에도 참여하며 AI 기술 범위를 확장• 또한 개방형 혁신을 통해 AI 생태계를 확장과 글로벌 시장 진출도 적극적으로 추진• AI/DX 융합사업 부문 : KT의 AI 사업 총괄 조직으로 AI 서비스 개발 및 사업화, B2B, B2C AI 솔루션을 제공• KT AI원팀 : AI 기술 연구개발 조직으로 핵심 AI 알고리즘 개발과 AI 인프라 구축 및 운영을 담당• KT융합기술원 : 차세대 AI 기술 연구를 수행하며, AI 기반 통신 기술개발, 신사업 발굴 및 기술 검증 등의 역할을 수행

(2) 2024년 기업 동향

가. KT의 안전하고 믿을 수 있는 AI 연구 강화

KT는 2024년 10월 안전하고 믿을 수 있는 인공지능(AI) 활용을 위해 연구 성과 리포트를 공개하고 전문가로 구성된 자문 위원회를 출범했다고 밝혔다.²⁹³⁾ ‘KT 책임 있는 AI(Responsible AI) 리포트’ 는 AI 윤리와 신뢰성을 향한 KT의 노력과 연구 성과를 담았다.

리포트는 거버넌스, 윤리 원칙, 프로세스 세 가지 측면에서 규정한 ‘책임 있는 AI 프레임워크’ 를 소개한다. 이는 안전하고 신뢰할 수 있는 AI 서비스 근간을 마련하기 위해 개발됐다. 윤리 원칙은 책임성(Accountability), 지속가능성(Sustainability), 투명성(Transparency), 신뢰성(Reliability), 포용성(Inclusivity)이라는 5대 핵심 원칙, ‘ASTRI’ 로 구성됐다. 약칭은 북극성처럼 이정표 역할을 수행한다는 의미를 담아 라틴어 ‘별’ 에 해당하는 단어를 차용했다.

293) <https://www.aitimes.com/news/articleView.html?idxno=164664>

KT는 국내에 바람직한 AI 생태계가 조성되기를 기대하며 리포트를 KT 홈페이지에 공개한다. 프레임워크의 세부 내용과 정립 과정이 궁금한 단체나 개인은 참고용으로 열람할 수 있다. 아울러 지난 22일에는 ‘책임 있는 AI 자문 위원회’가 출범했다. 위원회는 KT 임원과 고려대학교·KAIST 교수, AI 스타트업 셀렉트스타 소속 전문가들로 구성됐다. 이들은 한국적 문화와 가치를 반영한 AI 솔루션을 개발하기 위해 책임 있는 AI 프레임워크를 고도화한다. 세부적으로 AI 모델을 객관적으로 검증하고 국내외 규제를 반영하는 임무를 수행한다. 올해 9월 설립된 KT 사내 위원회와도 협업해 논문 저술, 학술 발표 등 연구 활동을 지속한다.

KT는 지금까지 책임 있는 AI 실천을 위해 노력해 왔다고 전했다. 캐나다 벡터 연구소 및 올해 9월 파트너십을 체결한 마이크로소프트와 프레임워크 강화를 위해 협력하는 한편, 4월에는 안전한 AI 활용을 위한 약속을 담은 ‘서울 AI 기업 서약’에 동참했다. ‘23년 10월에는 중국 차이나모바일, 일본 NTT도코모와 공동 집필한 ‘통신 AI 산업의 발전을 위한 백서’를 발간했다. 특히 ‘책임 있는 AI Center(RAIC)’를 4월 설립했다. RAIC는 설립 이후 ▲AI 윤리 원칙 고도화 ▲AI 거버넌스 수립 ▲AI 평가 체계 구축 등 활동을 펼쳤다. 더불어 KT 실무에 즉시 적용할 수 있는 수준의 지침을 제작하는 등 임직원들의 실질적인 AI 윤리 이행을 독려했다.

KT 책임 있는 AI Center 배순민 센터장은 RAIC는 KT의 AI를 모두가 믿고 사용할 수 있도록 관련 정책 수립, 프로세스 구축, 지속적인 모니터링 등의 역할을 수행한다고 설명하고 이번 리포트 발간으로 안전한 AI를 위한 국내외 이해관계자들과 협력을 확대해 나갈 수 있기를 바란다고 말했다.

4. AI 스타트업

1) 라이너

(1) 기업 일반현황

〈표 5-15〉 라이너 기업 현황

구분	주요 내용
라이너	<ul style="list-style-type: none">• 라이너는 AI 기술을 활용하여 온라인 학습과 정보 관리 혁신에 집중하고 있으며, 특히 웹 하이라이팅이라는 틈새시장에서 차별화된 서비스 제공으로 글로벌 시장에서도 주목• 교육 테크놀로지 분야에서 독특한 포지션을 확보하고 있으며, AI 기술을 통해 개인화된 학습 경험을 제공으로 차별화• 2015년 설립되어 2023년 AI 검색 서비스를 출시• 해당 서비스는 단순히 정보를 제공하는 것을 넘어 검색 결과의 각 문장에 신뢰할 수 있는 출처를 제공해 차별화• 8년간 축적된 1,000만명의 사용자 데이터를 AI 기술에 융합해 정교하고 개인화된 검색 결과를 제공한다는 평가를 받음

(2) 2024년 기업 동향

가. 라이너의 ‘생성 AI 웹 서비스 톱 50’ 서 전 세계 4위 기록

개인화 AI 에이전트 전문 라이너(대표 김진우)는 글로벌 벤처캐피털 앤드리슨 호로비츠가 발표한 ‘생성 AI 톱 100 소비자 앱(The Top 100 Gen AI Consumer Apps)’ 순위에서 전 세계 월간 웹 방문 횟수 기준 4위를 차지했다고 밝혔다.

앤드리슨 호로비츠가 생성 AI 서비스 관련 랭킹을 발표한 것은 이번이 두 번째로 웹사이트 트래픽 예측 사이트인 시밀러웹 2024년 1월 자료를 기반으로 선정한 결과다. 특히 지난해 11월 첫 발표 이후 대폭적인 물갈이가 이뤄졌다. 이번에는 새롭게 순위에 진입한 기업이 전체 40% 이상을 차지했다.

정확한 순위 선정을 위해 웹과 앱 트래픽을 분리했다. 라이너는 ‘AI 검색 코파일럿’ 서비스로 ‘웹 부문 톱 50’ 중 무려 4위에 이름을 올렸다. 카테고리는 ‘생산성 향상’ 부분으로 구

별됐다. 1위는 오픈AI의 '챗GPT', 2위는 구글의 '제미니', 3위는 캐릭터닷AI 등 세계적인 기업이 차지했다. 라이너는 글쓰기 전문 도구로 미국에서 큰 인기를 얻은 킴봇(5위)과 챗봇 플랫폼 퀴라(7위), 구글 검색 대안으로 꼽히는 퍼플렉시티(7위) 등을 모두 제쳤다.

이 밖에 클로드(10위), 일레븐랩스(11위), 허깅페이스(12위), 레오나르도AI(13위), 미드저니(14위) 등 쟁쟁한 생성 AI 기업들도 넘어섰다. 국내 기업 중에서는 유일하게 톱 50에 이름을 올렸다.

라이너는 이용자들이 정보를 탐색하고 지식을 습득하는 과정에서 발생할 수 있는 다양한 어려움과 불편함을 해결해 주는 'AI 에이전트 서비스'를 지향하고 있다. ▲복잡하고 심도 있는 요청에 대한 정보 탐색 및 답변 ▲논문이나 유명 서적을 기반으로 한 심층적인 리서치 ▲에세이나 비즈니스 메일과 같은 콘텐츠 작성 ▲번역 및 문법 교정 등 업무 생산성 향상에 특화된 기능을 지속 고도화 중이다.

라이너는 상위권을 차지한 생성 AI 웹 서비스 절반 이상이 미국 기업 서비스라고 평가하고, 라이너가 4위를 차지한 건 매우 놀라운 일이라고 강조했다. 또한 생산성을 향상해 준다는 점에서 많은 사용자가 사용하는 것 같다고 이유를 밝혔다. 라이너의 제품은 이전부터 국내보다 미국이나 인도 등 해외 사용자가 더 많은 것으로 유명하다.

2) 기타기업

가. W&B의 생성 AI 앱 개발도구 '위브'에 신규 기능 추가²⁹⁴⁾

웨이트앤바이어스(W&B)는 2024년 4월 베타 출시한 언어모델 최적화(LLMOps) 제품 '위브(Weave)'에 모델과 애플리케이션의 시각화 기능을 추가했다. 위브는 LLM 기반 애플리케이션을 개발할 때 학습과 성능평가 및 데이터 관리를 지원하는 솔루션이다. 개발자는 위브를 통해 빠르게 생성 AI 애플리케이션의 프로토타입을 개발하고 성능평가를 진행할 수 있다. 8월 정식 출시를 통해 모델과 애플리케이션의 성능평가와 시각화 기능을 추가했다.

294) <https://www.aitimes.com/news/articleView.html?idxno=164828>

기존 W&B의 '모델스(Models)' 플랫폼이 파운데이션 모델의 학습과 배포에 특화되어 있으며 반면, 위브는 생성 AI 애플리케이션을 프로토타입을 개발하고 지속적으로 운영하려는 소프트웨어 개발자들에게 적합하다는 설명이다. AI 애플리케이션이 잘못된 결과물을 생성할 경우, 입출력 데이터에서 잘못된 응답과 환각을 찾아 지속적인 디버깅을 할 수 있다.

나. 비온드날리지의 AI 검색 챗봇 제작 솔루션 '에스크챗' 출시²⁹⁵⁾

비온드날리지가 AI 검색 챗봇 제작 솔루션 '에스크챗(ASK Chat)' 을 출시했다. 에스크챗은 검색 증강 생성(RAG)을 적용해 중소기업 미디어, 출판사, 언론사 등이 자체 콘텐츠를 바탕으로 실시간 검색 챗봇 서비스를 개발하고 탑재할 수 있도록 지원한다.

미디어스피어와의 협력을 통해 에스크챗의 초기 안정성과 사용성을 확보했다고 밝혔다. 미디어스피어는 블루닷, 오웰, 로이스 등 콘텐츠 AI 서비스를 운영 중인 스타트업이다. 이준형 대표는 기업과 창작자 모두 누구나 AI 검색 챗봇을 가질 기회를 제공하기 위해 에스크챗 서비스를 개발하게 됐다고 말했다.

295) <https://www.aitimes.com/news/articleView.html?idxno=164817>

제6장 인공지능 기술 및 연구 동향

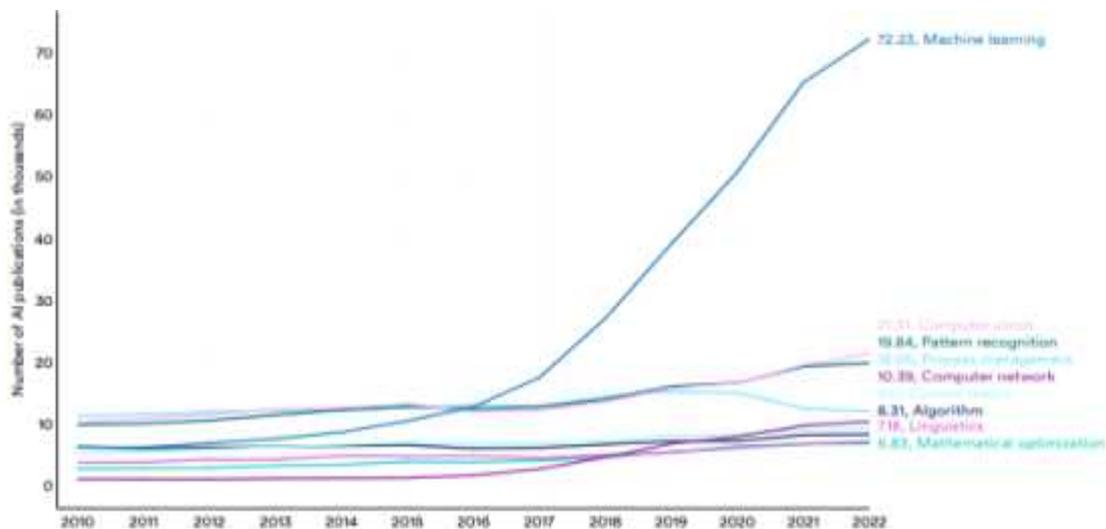
제1절 국제 학술단체, 표준기관, 비영리 연구기관

1. 학술단체

1) 학술연구 현황 및 추이

Stanford HAI의 AI Index 보고서에 따르면, 2010년부터 2022년까지 AI 분야 출판물(publication)은 지속적인 양적 성장세를 나타내고 있다. 2022년 전 세계 AI 분야 출판물은 약 24만 건으로, 2010년(약 8만 8천 건)과 비교하면 약 세 배 가까이 증가했다. 출판물 유형(type)별로, 2015년 이후 저널 논문 및 컨퍼런스 논문의 경우 각각 2.4배, 2.6배 증가하여 2022년 23만 건의 저널 논문 및 4만 2천 건의 컨퍼런스 논문이 발표되었다. 연구 분야별로, 2022년 가장 많이 출판된 분야는 머신러닝(72,230건)이며, 이어서 컴퓨터 비전(21,309건), 패턴인식(19,841건), 프로세스 관리(12,052건) 순으로 출판이 많았다. 특히 머신러닝 분야 출판건수는 2015년 이후 7배 가까이 급증하였다. 부문별로, '22년 AI 분야 출판물을 가장 많이 발표한 부문은 학계(81.07%), 산업계(7.89%), 정부(6.97%), 비영리(2.62%), 기타(1.46) 순으로 나타났다.

[그림 6-1] 연구 분야별 AI 관련 출판물 발간 추이(2010-2022)

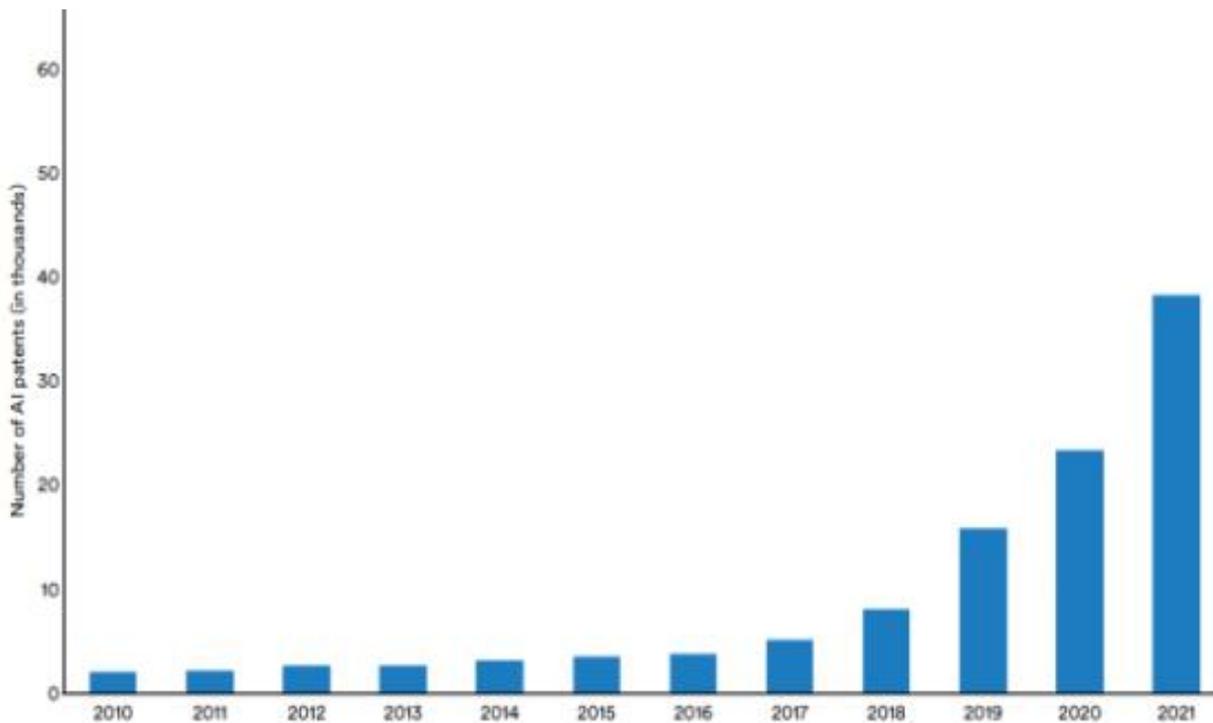


출처: Stanford HAI, AI Index Report 2024

2010년부터 2022년 사이 AI 분야의 특허 건수가 증가 추세이며, 특히 최근에는 급증 추세를 나타내고 있다. AI 특허 건수는 2010년부터 2014년 기간에 승인된 AI 특허 건수의 총 증가율이 56.1%로 나타난 것과 비교하여 2021년부터 2022년 사이에만 증가율이 62.7%로 크게 확대되었다. 주요 통계 분야별 특허 추이는 아래와 같다.

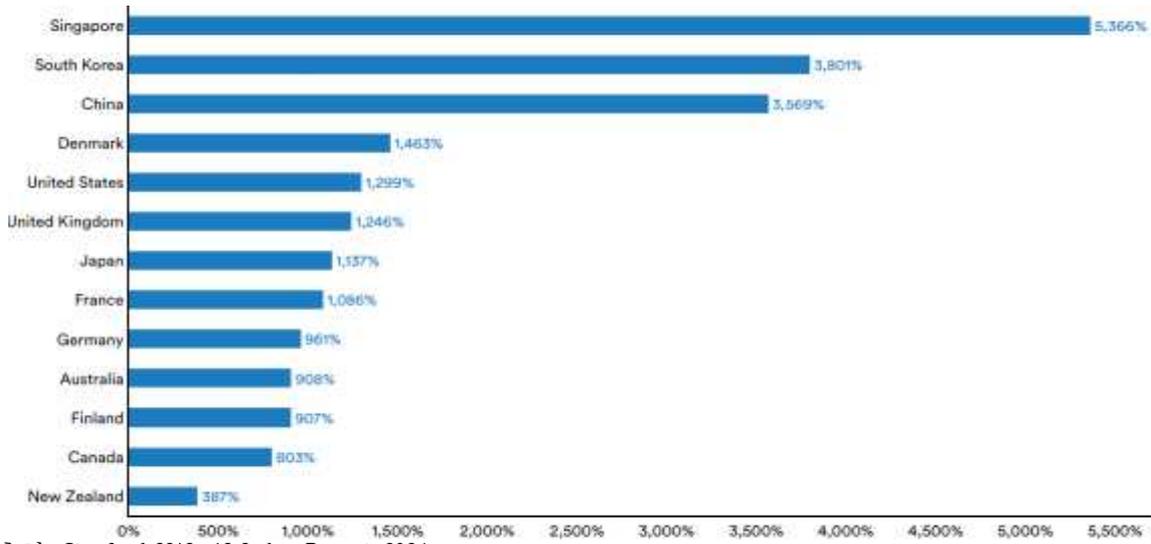
- (상태별) 2022년 미승인된 AI 특허 건수(128,952건)는 승인된 AI 특허 건수(62,264건)의 두 배가 넘는 것으로 확인됨
- (지역별) 2022년 전 세계에서 승인된 AI 특허 건수는 동아시아 및 태평양(75.2%), 북미(21.2%) 순으로 많음
- (국가별) 2022년 전 세계에서 승인된 AI 특허의 대부분은 중국(61.1%)과 미국(20.9%)이며, 미국에서 발생한 AI 특허 비중은 2010년 54.1%에서 크게 하락
- (인구별) 2022년 인구 10만 명당 AI 특허가 가장 많이 승인된 국가는 한국(10.3건), 룩셈부르크(8.8건), 미국(4.2건) 순으로 나타남

[그림 6-2] AI 특허 승인 건수 추이(2010-2022)



출처: Stanford HAI, AI Index Report 2024

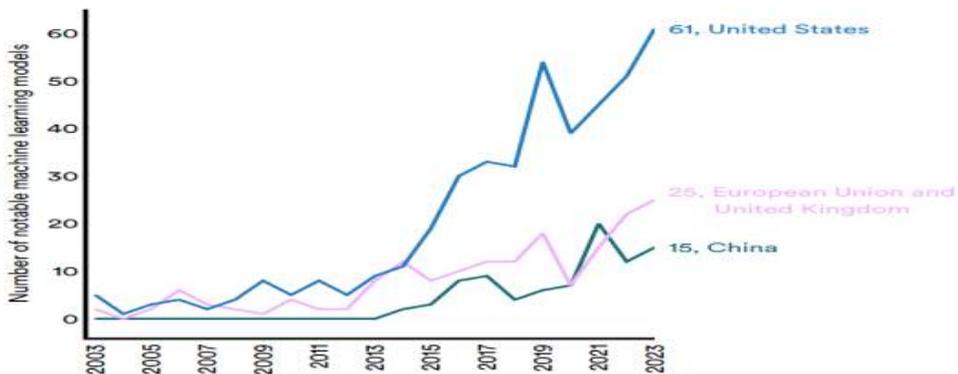
[그림 6-3] 인구 10만 명당 AI 특허 건수(2022)



출처: Stanford HAI, AI Index Report 2024

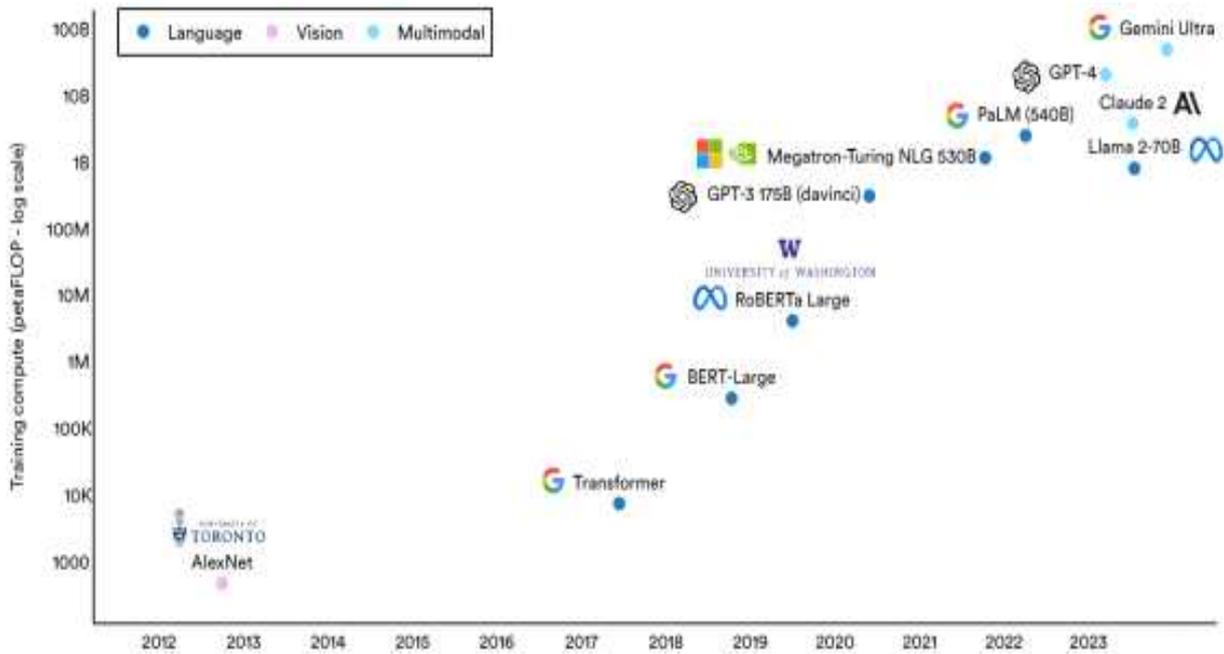
프론티어 AI 연구인 머신러닝 모델 출시에서 2014년까지 학계가 주도했으나 그 이후에는 산업계가 주도하고 있다. 2023년 산업계는 51개의 주목할만한 머신러닝 모델을 출시했지만, 학계는 15개를 출시했으며, 산업계와 학계의 협업으로 21개를 출시했다. 학계는 AI 모델 개발을 위해 필요한 방대한 양의 데이터, 컴퓨팅 파워, 재정을 충당하기에 한계가 있어, 최근에는 선도적인 AI 모델 개발에 대한 업계의 지배력이 지속적으로 증가하고 있다. 국가별로는 미국이 장기간 주목할 만한 머신러닝 모델 개발을 주도하고 있다. 2023년 가장 많은 수의 주목할 만한 머신러닝 모델을 개발한 국가는 미국 61개, 중국 15개, 프랑스 8개의 주목할 만한 머신러닝 모델을 개발하였다.

[그림 6-4] 국가별 주목할 만한 머신러닝 모델 출시 추이(2003-2023)



출처: Stanford HAI, AI Index Report 2024

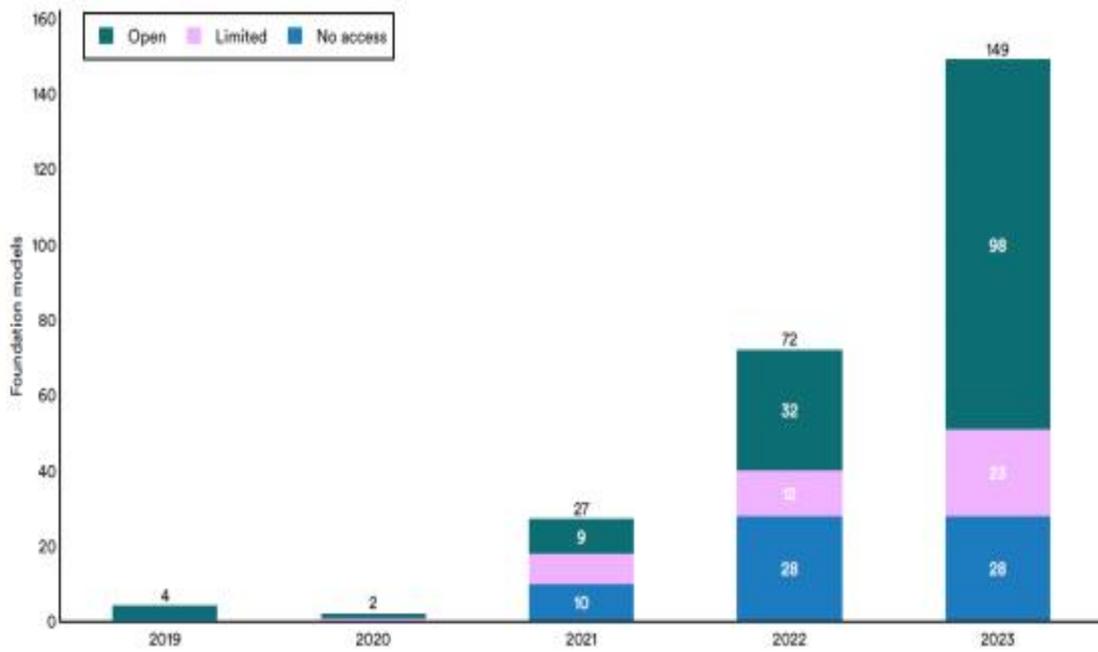
[그림 6-5] 주목할 만한 머신러닝 모델 출시 추이(2003-2023)



출처: Stanford HAI, AI Index Report 2024

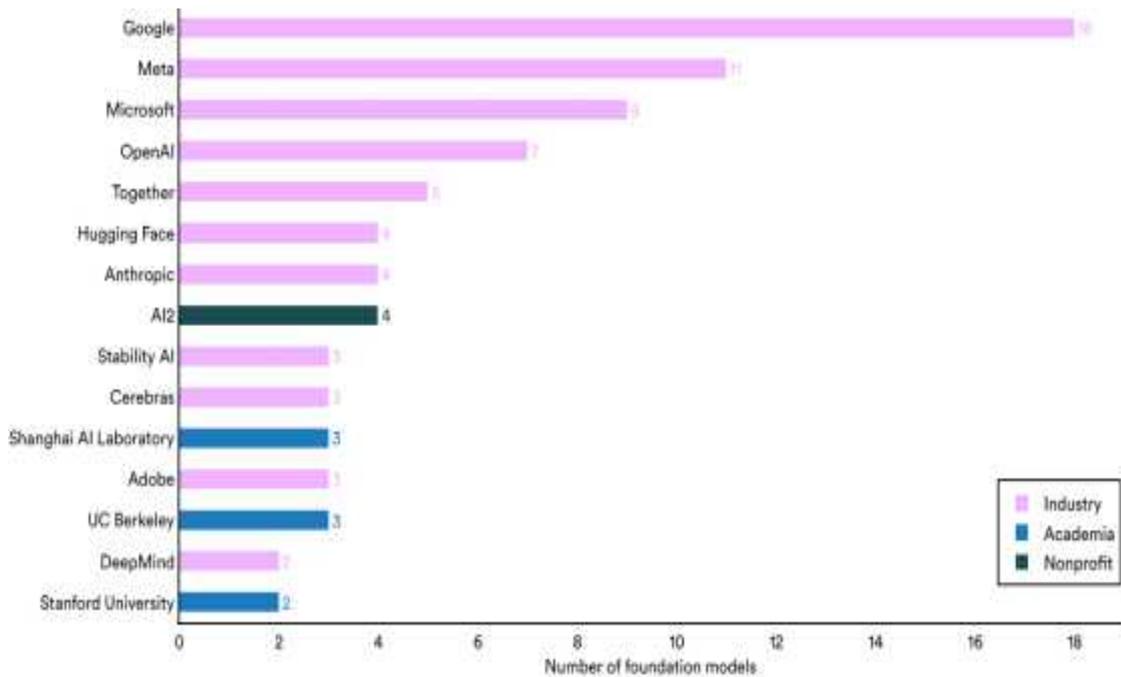
최근 몇 년 동안 기반 모델의 수가 급격히 증가한 가운데 2022년부터 2023년 사이 두 배 이상, 2019년부터 2023년 사이 약 38배나 증가하였다. 2023년 출시된 149개의 기반 모델 중 개방형 모델은 98개(65.8%), 제한적 접근 모델은 23개(15.4%), 접근 불가 모델은 28개(18.8%)로 조사되었다. 예를 들어 Meta의 Llama 2와 같은 개방형 모델은 모델 가중치가 완전히 공개되어 자유로운 수정과 사용이 가능하며, OpenAI의 GPT-4와 같은 제한적 접근 모델은 공개 API를 통해 모델에 대한 제한된 접근을 제공하고, Google의 PaLM-E와 같은 접근 불가 모델은 해당 개발자만이 접근할 수 있다. 2019년 이후 Google과 Meta를 중심으로 한 산업계가 기반 모델의 출시를 주도하고 있다. 2023년 산업계는 전체 기반 모델 중 72.5%, 학계는 18.8%의 개발을 주도하였다. 2023년 가장 많은 기반 모델을 출시한 기업은 Google(18개), Meta(11개), Microsoft(9개) 순이며, 가장 많은 기반 모델을 출시한 교육기관은 UC 버클리(3개)가 차지하였다. 2019년 이후 가장 많은 기반 모델을 공개한 기업은 Google(40개), OpenAI(20개) 순으로 나타났으며 중국의 칭화대학교는 비서구권 기관 중 가장 많은 7개의 기반 모델을 공개하였다.

[그림 6-6] 접근 유형별 기반 모델 출시 추이(2019-2023)



출처: Stanford HAI, AI Index Report 2024

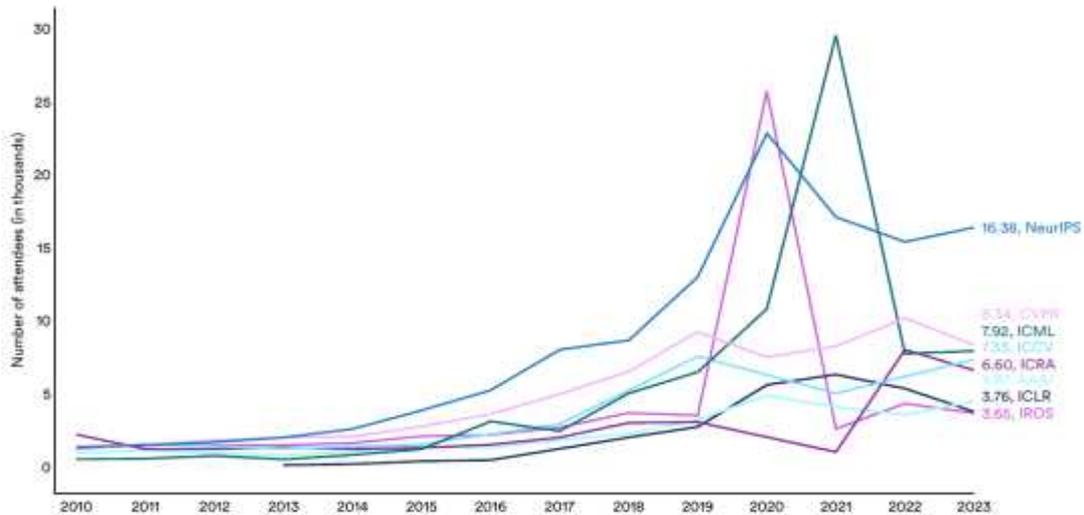
[그림 6-7] 기관/기업별 기반 모델 출시(2023)



출처: Stanford HAI, AI Index Report 2024

2022년부터 2023년 사이 AI 컨퍼런스 참석자 수가 6.7% 증가했다. 이러한 증가세는 AI에 대한 관심 증가와 함께 새로운 AI 컨퍼런스의 출현을 반영한다. 2023년 신경정보처리시스템(NeurIPS) 컨퍼런스에는 약 1만 6,380명이 참가했으며, 이와 더불어 국제머신러닝학회(ICML), 국제컴퓨터비전학회(ICCV), 세계인공지능학회(AAAD) 등 주요 AI 컨퍼런스 참석자 수가 전년 대비 증가했다.

[그림 6-8] AI 컨퍼런스 참석자 수 추이(2010-2023) (단위: 천 명)



출처 Stanford HAI, AI Index Report 2024

2) 2024년 주요 동향

(1) 전기전자공학학회(IEEE)

IEEE(Institute of Electrical and Electronics Engineers)는 전자공학, 컴퓨터과학 등 다양한 기술 분야에서 활동하는 세계 최대의 전문 기관으로, 다양한 분야의 학회 및 컨퍼런스를 조직하고, 연구 논문을 게재하는 저널을 운영한다. IEEE는 CVPR((Conference on Computer Vision and Pattern Recognition), IROS (Intelligent Robots and Systems), ICDM (Data Mining) 등 AI 관련 대형 학회를 주최한다.

IEEE는 AI 윤리 및 거버넌스 표준 프로그램으로 ‘GET Program’ 운영하고 있다. 이 프로그램을 2024년에 다음과 같은 표준을 마련하였다.²⁹⁶⁾

296) <https://ieeexplore.ieee.org/browse/standards/get-program/page/series?id=93>

- (7009-2024) 자율 및 반자율 시스템의 실패 안전 설계를 위한 IEEE 표준(IEEE Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems)
- (7014-2024) 자율 및 지능형 시스템에서 에뮬레이트된 공감의 윤리적 고려 사항을 위한 IEEE 표준(IEEE Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems)

IEEE SA(IEEE Standards Association)는 자율지능시스템(AIS: autonomous intelligent systems) 원칙과 접근 방식과 솔루션을 계속 개발하고 있다. IEEE SA는 이해관계자가 AIS를 설계, 개발 및 사용할 수 있도록 표준, 교육 및 훈련, 인증 프로그램 등을 제공한다. 2024년 IEEE SA의 AIS 관련 다음과 같은 표준을 개발하였다.

〈표 6-1〉 IEEE SA의 AIS 관련 다음과 같은 표준(2024)

번호	표준 명칭
IEEE 2089.1	• IEEE Standard for Online Age Verification
IEEE 2807.1	• IEEE Approved Draft Standard for Technical Requirements and Evaluation of Knowledge Graphs
IEEE 2807.2	• IEEE Approved Draft Guide for Application of Knowledge Graphs for Financial Services
IEEE 2894	• IEEE Approved Draft Guide for an Architectural Framework for Explainable Artificial Intelligence
IEEE 2933	• IEEE Approved Draft Standard for Clinical Internet of Things (IoT) Data and Device Interoperability with TIPPSS - Trust, Identity, Privacy, Protection, Safety, Security
IEEE 3154	• IEEE Recommended Practice for the Application of Knowledge Graphs for Talent Services
IEEE 3168	• IEEE Standard for Robustness Evaluation Test Methods for a Natural Language Processing Service That Uses Machine Learning
IEEE 3803	• IEEE Approved Draft Standard for Household Appliance Customer Data Assetization and Commercialization Requirements
IEEE 700924	• IEEE Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems
IEEE 7014	• IEEE Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems

출처: IEEE SA²⁹⁷⁾

아울러, IEEE SA는 2024년 연중으로 AIS 제품 채택을 보호, 차별화 및 확대하는 데 도움이 되는 AIS의 윤리를 평가하는 인증 프로그램으로 IEEE CertifAIEd를 운영하고 있다.

297) <https://standards.ieee.org/initiatives/autonomous-intelligence-systems/standards/> (2024.10.14. 확인)

(2) ACM(CVPR)

ACM(Association for Computing Machinery)은 컴퓨터과학과 정보기술 분야에서 세계적으로 가장 큰 학술 및 전문 협회 중 하나로, 학회 및 콘퍼런스를 주최하고, 컴퓨터과학 연구의 논문을 발표하는 저널을 발간한다. ACM은 SIGKDD(Knowledge Discovery and Data Mining), SIGGRAPH(Computer Graphics and Interactive Techniques), ACM CHI(Human-Computer Interaction)와 같은 대표적인 AI 관련 학회들을 주최한다.

ACM은 다양한 SIG 중 하나로 인공지능 특별 관심 그룹(SIGAI: Special Interest Group on Artificial Intelligence)를 운영하고 있다. 이 그룹은 컴퓨팅 전반에 걸쳐 AI 원리와 기술의 성장과 응용을 촉진하고 지원하는 역할을 수행한다. ACM SIGAI는 학계 교육자 및 연구자, 전문가, 학생이라는 세 가지 주요 커뮤니티로 구성되어 서비스를 제공하며, 주요한 활동은 다음과 같다.

- 고품질 AI 관련 컨퍼런스 후원 또는 공동 후원
- 주요 연례 상을 통해 학계, 전문가 및 학생 회원의 업적 인정
- ACM 디지털 라이브러리에 분기별 뉴스레터(AI Matters) 발행 및 커뮤니티의 관심사인 AI 문제에 대한 블로그 유지
- 학생 회원에게 컨퍼런스 참석을 위한 장학금 제공
- AI 커뮤니티 내부 및 외부에서 AI 교육, 출판물 및 공공 홍보 활동에 대한 자금 지원, 개발 및 홍보

SIGAI는 자율 에이전트 연구(Autonomous Agents Research) 분야의 상을 매년 수상하고, 2024년에는 델프트기술대학(Delft University of Technology)의 카톨리진 온케(Catholijn Jonker) 교수가 수상하였다. 온케 교수는 인간-기계 상호 작용 분야의 리더로, 자동 협상 에이전트 대화를 시작했으며, 하이브리드 지능형 시스템을 위한 대화형 지능 통합과 같은 다른 연구 분야에도 기여했다.

(3) AI 분야 학회 - 신경정보처리시스템(NeurIPS)

NeurIPS(Neural Information Processing Systems)는 AI와 기계학습 연구자들에게 가장 중요

한 학회 중 하나로, 딥러닝과 강화 학습 같은 최신 기술이 주로 발표된다. 전 세계적으로 영향력 있는 연구 논문들이 발표되고, 구글, 페이스북 등 주요 기술 기업들의 연구원들도 참가하는 학회이다.

2024년 NeurIPS²⁹⁸)은 제38차 연례 컨퍼런스로 12월 10일부터 15일까지 캐나다 밴쿠버에서 개최되어 다음과 같은 주제를 다루었다.

- 응용 프로그램(예: 비전, 언어, 음성 및 오디오, Creative AI)
- 딥 러닝(예: 아키텍처, 생성 모델, 딥 네트워크 최적화, 기초 모델, LLM)
- 평가(예: 방법론, 메타 연구, 재현성 및 타당성, 인간 참여)
- 일반 머신 러닝(지도 학습, 비지도 학습, 온라인, 활성 학습 등)
- 인프라(예: 라이브러리, 개선된 구현 및 확장성, 분산 솔루션)
- 과학(예: 기후, 건강, 생명 과학, 물리학, 사회 과학)을 위한 머신 러닝
- 신경 과학 및 인지 과학(예: 신경 코딩, 뇌-컴퓨터 인터페이스)
- 최적화(예: 블록 및 비블록, 확률적, 견고함)
- 확률적 방법(예: 변분 추론, 인과 추론, 가우시안 프로세스)
- 강화 학습(예: 결정 및 제어, 계획, 계층적 RL, 로봇공학)
- 기계 학습의 사회적 및 경제적 측면(예: 공정성, 해석 가능성, 인간-AI 상호 작용, 프라이버시, 안전, 전략적 행동)
- 이론(예: 제어 이론, 학습 이론, 알고리즘 게임 이론)

(4) AI 분야 학회 - 컴퓨터비전패턴인식학회(CVPR)

IEEE가 후원하는 CVPR(Conference on Computer Vision and Pattern Recognition)은 컴퓨터 비전과 패턴인식 분야에서 가장 권위 있는 학회로, 이미지 인식, 객체 탐지, 비디오 분석과 같은 주제에서 중요한 연구들이 발표된다. 특히, 딥러닝 기반 컴퓨터 비전 연구가 활발하게 논의된다.

298) <https://neurips.cc/>

2024년 CVPR²⁹⁹)은 6월 17일부터 6월 21일까지 미국의 시애틀에서 개최되었다. 이번 컨퍼런스는 역대 최대로 2023년 대비 26% 증가한 11,532건이 제출되었고, 76개국 및 지역에서 12,000명이 참석하였다.

2,719편의 수락된 논문 중 2024년 최우수 논문으로는 구글리서치(Google Research) 팀의 ‘생성 이미지 역학(Generative Image Dynamics)’가 선정되었다.

(5) AI 분야 학회 - 국제머신러닝학회(ICML)

ICML(International Conference on Machine Learning)은 AI와 기계학습 분야에서 NeurIPS와 함께 가장 권위 있는 학회로 꼽힌다. 머신러닝 알고리즘, 이론적 연구, 응용 프로그램 등 다양한 주제를 다룬다.

2024년 ICML³⁰⁰(제41회)은 7월 21일부터 7월 27일까지 오스트리아 비엔나에서 개최되었다. 이번 컨퍼런스에서 ‘Test of Award’ 상으로는 제프리 도나휴(Jeffrey Donahue) 등의 ‘DeCAF: 일반 시각 인식을 위한 딥 컨볼루션 활성화 기능(DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition)’이 선정되었다. 최우수 논문으로는 ‘VideoPoet: 제로 샷 비디오 생성을 위한 대규모 언어모델(VideoPoet: A Large Language Model for Zero-Shot Video Generation)’ 등 20개가 선정되었다.

(6) AI 분야 학회 - 국제컴퓨터비전학회(ICCV)

CVPR(International Conference on Computer Vision)은 컴퓨터 비전과 패턴인식 분야에서 가장 권위 있는 학회로 2년마다 개최되며, 이미지 인식, 객체 탐지, 비디오 분석과 같은 주제에서 중요한 연구들이 발표된다. 특히, 딥러닝 기반 컴퓨터 비전 연구가 활발하게 논의된다.

CVPR는 격년으로 개최되어 2024년에는 개최되지 않았으며, 2025년에는 미국 하외에서 2025년 10월 19일부터 25일까지 개최될 예정이다.

299) https://cvpr.thecvf.com/Conferences/2024/News/Wrap_Release

300) <https://icml.cc/Conferences/2024>

(7) AI 분야 학회 - 국제로봇자동화학회(ICRA)

IEEE가 후원하는 ICRA(International Conference on Robotics and Automation)는 로봇 공학과 자동화 분야에서 세계적으로 가장 권위 있는 학회 중 하나로, IEEE가 주관하는 대형 국제 학회이다. ICRA는 로봇 공학의 최신 기술과 연구 성과를 공유하는 자리로, 연구자, 학계, 산업계 전문가들이 모여 로봇 기술의 미래를 논의하고 협력하는 중요한 플랫폼으로 작용한다. 2024년 ICRA³⁰¹⁾는 5월 13일부터 17일까지 일본 동경에서 개최되었다. 이번 컨퍼런스에서 IEEE ICRA 최우수논문으로는 다음 2개 논문이 선정되었다.

- Ajay Sridhar 외, ‘통합 항해 및 탐색을 위한 목표 마스크 확산 정책(Goal Masked Diffusion Policies for Unified Navigation and Exploration)’
- Sergey Levine 외, ‘Open X-Embodiment: 로봇 학습 데이터 세트 및 RT-X(Open X-Embodiment: Robotic Learning Datasets and RT-X)’

(8) AI 분야 학회 - 세계인공지능학회(AAAI)

IEEE가 후원하는 AAI(Association for the Advancement of Artificial Intelligence)는 AI 전반을 다루는 가장 오래된 학회 중 하나로, 인공지능의 여러 하위 분야에서 연구 논문이 발표된다. 자연어처리, 게임 AI, 인간-컴퓨터 상호작용 등 다양한 응용 분야의 연구 성과도 포함된다. 2024년 AAI(제38회)³⁰²⁾은 2월 20일부터 2월 27일까지 캐나다 밴쿠버에서 개최되었다. 이번 학회에서 최우수논문으로 다음 3개 논문이 선정되었다.

- Chen Li 외, ‘GxVAE: 두 개의 조인트 VAE가 유전자 발현 프로파일에서 히트 분자를 생성(Two Joint VAEs Generate Hit Molecules from Gene Expression Profiles)’
- Cai Xu 외, ‘신뢰할 수 있는 갈등적 다중 뷰 학습(Reliable Conflictive Multi-view Learning)’
- Nikhil Chandak 외, ‘순차적 의사 결정을 위한 선호도의 비례적 집계(Proportional Aggregation of Preferences for Sequential Decision Making)’

301) <https://2024.ieee-icra.org/>

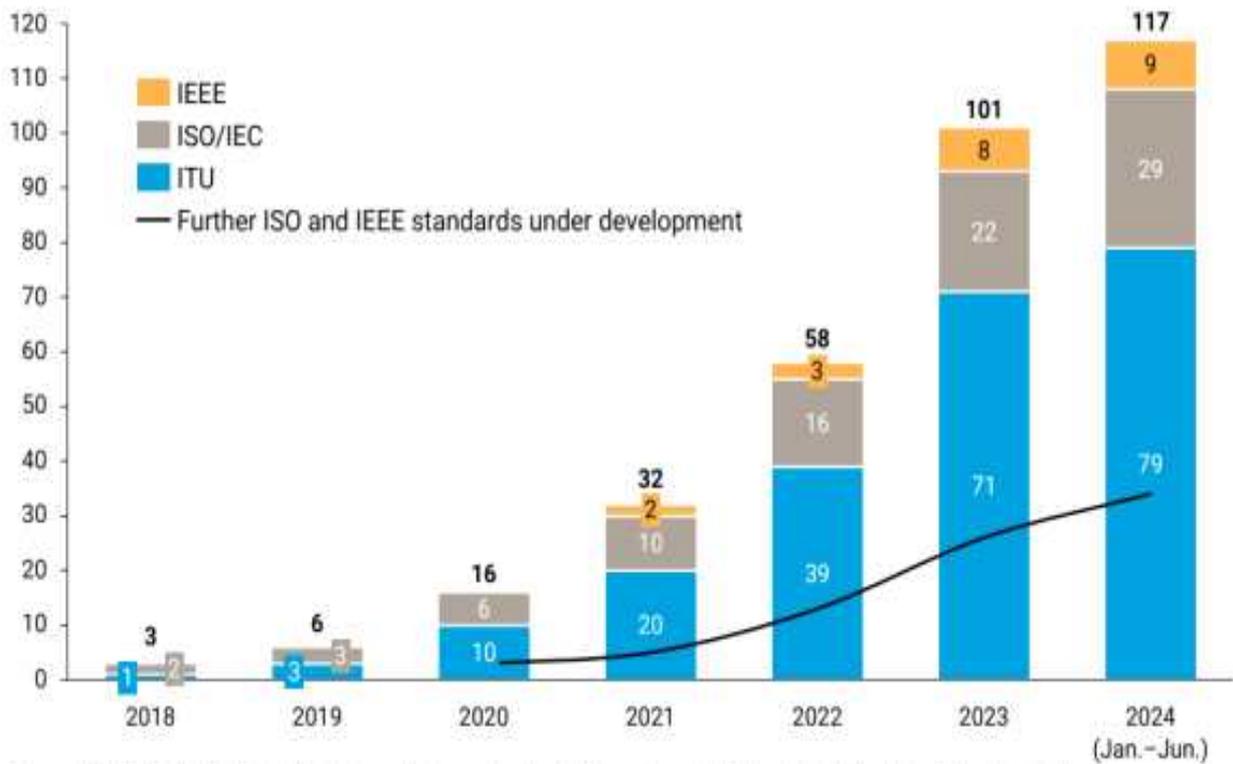
302) <https://aaai.org/conference/aaai/aaai-24/>

2. 표준기관

1) 표준화 현황 및 추이

UN의 AI 자문기구의 2024년 9월 ‘인류를 위한 AI 거버넌스(Governing AI for Humanity)³⁰³⁾’ 보고서에 따르면, 2018년 이후 ITU, ISO, 국제 전기 기술 위원회(IEC) 및 전기 전자 기술자 협회(IEEE)에서 AI 관련 채택한 표준의 수가 뚜렷한 증가 추세이다. 2024년(1월~6월)에 AI 관련 표준 수는 ITU 79건으로 가장 많고, ISO/IEC 29건, IEEE 9건으로, 이들 표준기관 전체는 총 117건이 조사되었다.³⁰⁴⁾ 반면, 현재까지는 이들 표준기관 간에는 공통적인 용어의 정의는 미흡하고, 공정성, 안전성, 투명성과 같이 AI와 관련하여 일상적으로 사용되는 많은 용어에 대해 합의된 정의는 없는 상황이다.

[그림 6-9] AI 관련 표준 수



출처: United Nations AI Advisory Body, Governing AI for Humanity - Final Report, 2024.09.19.

303) United Nations AI Advisory Body, Governing AI for Humanity - Final Report, 2024.09.19.

304) United Nations AI Advisory Body, Governing AI for Humanity - Final Report, 2024.09.19.

2) ISO/IEC

ISO/IEC(국제표준화기구/국제전기표준회의)는 AI 시스템의 신뢰성, 윤리적 사용, 데이터 관리, AI 시스템 평가, 그리고 산업별 적용을 중심으로 표준화 작업을 진행하고 있다. ISO와 IEC는 ISO/IEC JTC 1/SC 42(Artificial Intelligence)라는 AI 표준화 위원회를 2017년부터 운영하고 있다.

2024년, ISO/IEC JTC 1 SC 42는 전체 AI 생태계의 국제표준화를 수행하며, 31개의 공개된 표준과 36개의 현재 개발 중인 프로젝트 및 12개의 작업 그룹을 운영 중이다.

ISO/IEC는 2024년 1월 ‘ISO/IEC 42001:2023-정보기술-AI-관리 시스템(Information Technology Artificial Intelligence Management System)’ 표준을 발표하였다.³⁰⁵⁾ 이번 표준은 AI 관리 시스템에 대한 세계 최초의 표준으로, 조직 내 AI 시스템의 구축과 관리를 위한 지침 역할을 하며, AI 시스템을 사용하는 제품·서비스를 제공하거나 사용하는 모든 조직에 적용 가능하다. 이 표준의 목적은 책임 있는 AI의 개발과 사용을 보장하면서 AI의 이점을 누리기 위함으로, 조직은 표준의 이행을 통해 AI 안전과 보안, 공정성과 투명성을 향한 노력을 입증할 수 있다. 표준의 내용은 조직의 상황에 맞는 AI 관리 시스템의 구축, 경영진의 역할과 책임, 기획, 지원, 운영, 성과 평가, 개선사항의 장으로 나뉘어 구성했다.

이 표준에 의하면 AI 관리 시스템은 조직의 관리 프로세스 및 전반적 관리 구조와 통합되어야 하며, ▲조직의 목표 결정 ▲위험과 기회의 관리 ▲수명주기 전반에 걸친 AI 시스템의 보안과 안전, 공정성과 투명성, 데이터 품질 등 AI 신뢰성 관련 이슈의 관리 프로세스 ▲AI 시스템을 제공하거나 개발하는 외부 협력업체의 관리 프로세스가 필요하다.

이 표준의 핵심은 AI 정책의 개발과 이행으로, 조직은 윤리 문제, 투명성, 위험관리, 거버넌스를 포괄하는 AI 정책 수립을 통해 AI 시스템 관리를 위한 체계적인 접근방식을 확립이 필요하다. 조직은 AI 시스템의 관리를 지원하기 위해 인력과 시설을 포함해 충분한 자원을 확보해야 하며, AI 관련 활동에 참여하는 직원에 대한 지속적인 역량 교육이 필요하다. 운영 측면에서는 AI 관리 시스템과 프로세스의 효율성을 보장하고 AI 관련 활동을 모니터링하

305) ISO, ISO/IEC 42001:2023-Information technology-Artificial intelligence-Management system Pinsent Masons, Artificial intelligence: ISO and IEC publish new international standard on AI, 2024.01.17.

여 문제 발생 시 시정이 필요하다. AI 관리 시스템을 모니터링하여 의도한 결과가 나오지 않을 경우 적절한 조치를 통해 시정해야 하며, 모든 AI 위험평가 결과를 문서화하여 보관해야 한다.

2024년 11월 14일 기준, ISO/IEC JTC 1/SC 42에는 아래의 표와 같이 31개의 표준을 발행했다.

<표 6-2> ISO/IEC JTC 1/SC 42에서 발행된 표준 목록

번호	표준번호	표준 이름
1	ISO/IEC 42001:2023	Information technology — Artificial intelligence — Management system
2	ISO/IEC 38507:2022	Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations
3	ISO/IEC 25059:2023	Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems
4	ISO/IEC TS 25058:2024	Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Guidance for quality evaluation of artificial intelligence (AI) systems
5	ISO/IEC 24668:2022	Information technology — Artificial intelligence — Process management framework for big data analytics
6	ISO/IEC TR 24372:2021	Information technology — Artificial intelligence (AI) — Overview of computational approaches for AI systems
7	ISO/IEC TR 24368:2022	Information technology — Artificial intelligence — Overview of ethical and societal concerns
8	ISO/IEC TR 24030:2024	Information technology — Artificial intelligence (AI) — Use cases
9	ISO/IEC 24029-2:2023	Artificial intelligence (AI) — Assessment of the robustness of neural networks — Part 2: Methodology for the use of formal methods
10	ISO/IEC TR 24029-1:2021	Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview
11	ISO/IEC TR 24028:2020	Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence
12	ISO/IEC TR 24027:2021	Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making
13	ISO/IEC 23894:2023	Information technology — Artificial intelligence — Guidance on risk management
14	ISO/IEC 23053:2022	Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
15	ISO/IEC 22989:2022	Information technology — Artificial intelligence — Artificial intelligence concepts and terminology

번호	표준번호	표준 이름
16	ISO/IEC TR 20547-5:2018	Information technology — Big data reference architecture — Part 5: Standards roadmap
17	ISO/IEC 20547-3:2020	Information technology — Big data reference architecture — Part 3: Reference architecture
18	ISO/IEC TR 20547-2:2018	Information technology — Big data reference architecture — Part 2: Use cases and derived requirements
19	ISO/IEC TR 20547-1:2020	Information technology — Big data reference architecture — Part 1: Framework and application process
20	ISO/IEC 20546:2019	Information technology — Big data — Overview and vocabulary
21	ISO/IEC TR 17903:2024	Information technology — Artificial intelligence — Overview of machine learning computing devices
22	ISO/IEC TS 8200:2024	Information technology — Artificial intelligence — Controllability of automated artificial intelligence systems
23	ISO/IEC 8183:2023	Information technology — Artificial intelligence — Data life cycle framework
24	ISO/IEC TR 5469:2024	Artificial intelligence — Functional safety and AI systems
25	ISO/IEC 5392:2024	Information technology — Artificial intelligence — Reference architecture of knowledge engineering
26	ISO/IEC 5339:2024	Information technology — Artificial intelligence — Guidance for AI applications
27	ISO/IEC 5338:2023	Information technology — Artificial intelligence — AI system life cycle processes
28	ISO/IEC 5259-4:2024	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 4: Data quality process framework
29	ISO/IEC 5259-3:2024	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 3: Data quality management requirements and guidelines
30	ISO/IEC 5259-1:2024	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 1: Overview, terminology, and examples
31	ISO/IEC TS 4213:2022	Information technology — Artificial intelligence — Assessment of machine learning classification performance

출처: ISO 웹사이트³⁰⁶⁾

306) <https://www.iso.org/committee/6794475/x/catalogue/p/1/u/0/w/0/d/0>, (2024. 10. 14 확인)

3) ITU

ITU(국제전기통신연합)는 국제적인 표준화 기구로서, AI 기술의 급속한 발전에 대응해 안전하고 책임 있는 AI 개발과 사용을 촉진하는 데 중점을 두고, AI 관련 표준화 작업을 진행하고 있다. ITU는 AI 기술이 세계적으로 확대되면서 그에 맞는 글로벌 표준을 개발하고, 다양한 산업에서 AI가 안전하고 윤리적으로 사용될 수 있도록 노력하고 있다.

ITU는 AI 기술의 급격한 발전에 발맞춰 FG-AI4EE(Focus Group on Environmental Efficiency for AI and Other Emerging Technologies), FG-AI4NDM(Focus Group on AI for Natural Disaster Management)와 같은 특화된 연구 그룹을 통해 AI의 윤리적 사용, 안전성, 산업별 적용, 환경 효율성을 다루는 표준화를 적극적으로 추진하고 있다. 이 외에도 AI for Good 프로그램을 통해 AI가 사회적, 경제적 혜택을 제공할 수 있는 방법을 연구하고 있다.

FG-AI4NDM은 재난 관리 용어에 대한 통찰력을 제공하기 위한 500개 이상의 용어집 개발과 주요 격차를 식별하기 위한 표준 로드맵 개발 등을 수행하고 있다. 2024년 FG-AI4NDM은 ‘재난관리: 표준 관점(Disaster Management: The Standards Perspective)’³⁰⁷⁾ 보고서를 발행하였다. 이 보고서는 데이터 관리, 재난 시나리오 모델링 및 AI 기반 커뮤니케이션 솔루션 구현을 위한 모범사례를 제공한다.

ITU는 ‘AI for Good Global Summit’ 매년 오프라인으로 개최하고, 연중 다양한 주제로 온라인 세미나를 개최한다. 2024년 ‘AI for Good Global Summit’은 5월 29일부터 30일까지 스위스 제네바에서 ‘UN의 지속가능개발목표(SDG) 가속화’라는 주제로 개최되었고, 정부, 시민 사회, UN 기관, AI 혁신가, 투자자들이 참여하였다. 이번 행사에서 안토니오 구테흐스 유엔 사무총장은 “인공지능이 인류에게 이로운 방향을 제시하고 우리의 공유 가치를 강화하는 데 도움이 될 수 있습니다. 유엔은 함께 논의할 수 있는 보편적 플랫폼이 될 준비가 되어 있습니다. 우리가 인공지능을 사용하여 인간의 존엄성을 높이고 세계적 재화를 제공하도록 합시다.”라고 언급했다. ITU 사무총장 도린 보그단-마틴(Doreen Bogdan-Martin)은 “AI가 우리를 형성하는 것보다 더 빠르게 우리가 AI를 형성하는 것이 우리의 공동 이익입니다. 유엔의 AI 주요 플랫폼인 이 정상회담은 다양한 이해관계를 대표하는 주요 목소리를 모아 AI가 지속가능개발목표(SDG)를 구출하기 위한 경쟁에서 진전을 위한 강력한 촉매가 될 수 있도록 할 것입니다.”라고 언급했다.

307) ITU, Disaster Management: The Standards Perspective, 2024

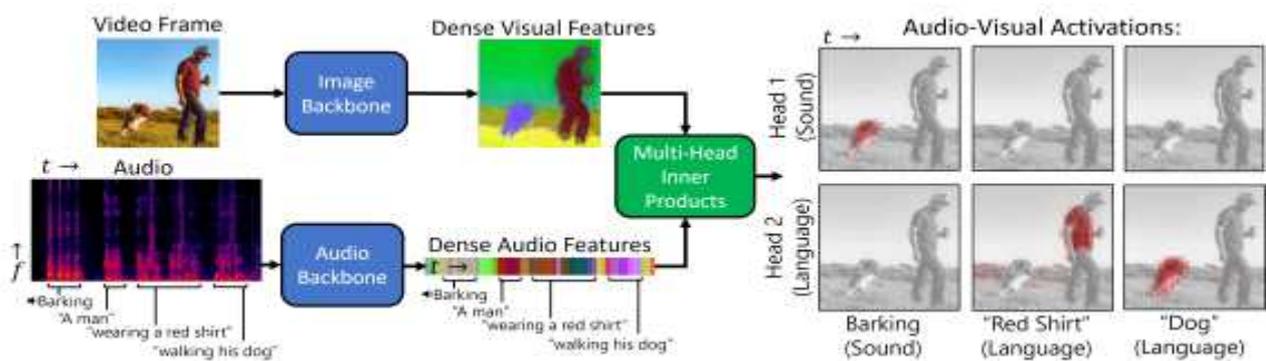
3. 연구기관

1) 대학 - MIT, 동영상 시청만으로 언어를 학습하는 알고리즘 개발³⁰⁸⁾

MIT 컴퓨터과학AI연구소(CSAIL) 연구진이 2024년 6월 9일 사전학습 없이 동영상 시청만으로 언어를 학습할 수 있는 알고리즘에 관한 논문을 아카이브³⁰⁹⁾에 공개했다. 연구진은 아이들이 주변 환경을 관찰하고 소리를 들음으로써 언어를 배우는 방식에서 착안해, 모델이 시각과 소리 사이의 연관성을 활용해 언어를 학습하게 만드는 방법을 고안했다.

연구진이 개발한 ‘덴스AV(DenseAV)’ 라는 명칭의 모델은 200만 개의 유튜브 동영상에 포함된 데이터셋을 활용해 소리와 일치하는 시각 콘텐츠를 예측하거나, 시각 콘텐츠와 일치하는 소리를 예측하는 식으로 언어를 학습한다. 일례로 덴스AV는 ‘180도에서 케이크를 굽는다’ 는 소리가 들리면 케이크나 오븐의 시각 콘텐츠를 예상하며, ‘개’ 라는 단어가 언급되거나 개가 짖는 소리를 들으면 동영상에서 개의 이미지를 추출한다. 오디오와 시각 데이터를 별도로 처리하는 덴스AV는 오디오와 시각 신호 쌍을 비교하여 어떤 신호가 일치하고 어떤 신호가 일치하지 않는지를 학습하는 대조 학습 방식으로 언어 자체의 예측 패턴을 파악한다.

[그림 6-10] 덴스AV 알고리즘의 작동 방식



출처: MIT News, New algorithm discovers language just by watching videos, 2024.6.11.

연구진은 덴스AV를 통해 돌고래나 고래의 의사소통과 같이 문자화되지 않은 새로운 유형의 언어를 이해하거나 지구의 지진음과 지질 간 패턴 파악 등의 응용이 가능할 것으로 기대한다.

308) MIT News, New algorithm discovers language just by watching videos, 2024.6.11.

309) Separating the "Chirp" from the "Chat": Self-supervised Visual Grounding of Sound and Language (<https://arxiv.org/abs/2406.05629>)

2) 대학 - 스탠포드대 등, AI 모델의 붕괴 위험 증가 경고³¹⁰⁾

영국 옥스퍼드대를 비롯한 주요 대학 연구진이 2024년 7월 24일 네이처(Nature)에 발표한 연구에 따르면 AI로 만든 데이터로만 학습한 LLM은 결과물이 오염되는 ‘모델 붕괴(Model Collapse)’ 현상이 발생했다. AI가 만든 데이터를 AI가 학습해 결과물을 생성하고 이 데이터를 다시 AI가 학습해 결과물을 생성하도록 실험을 반복한 결과, 후속 모델은 엉뚱한 결과물을 생성하는 모델 붕괴 현상이 확인되었다.

메타(Meta)의 AI 모델 OPT-125m에 14세기 중세 건축에 대한 텍스트를 입력하고 이전 세대의 모델에서 생성된 데이터를 기반으로 다음 세대 모델을 학습시킨 결과, 5세대 모델부터 입력한 데이터와 무관한 결과물이 생성되기 시작했다. 5세대 모델은 맥락과 무관하게 영어, 이탈리아어, 프랑스어, 독일어, 스페인어 등 100개 이상 언어로 번역되었다는 내용을 출력했으며, 9세대 모델은 ‘세계 최대 규모의 검정 @-@꼬리 토끼’, ‘흰 @-@꼬리 토끼’ 등 해석이 불가능한 문장을 생성했다.

연구진은 이 같은 모델 붕괴가 일어나는 이유를 시간이 지남에 따라 모델에 내재한 통계나 함수 표현, 기능적 측면의 사소한 오류와 오해가 축적되기 때문이라고 설명하면서, 모델 붕괴를 막으려면 AI 모델 개발 시 인간이 만든 원본 데이터를 활용하는 것이 중요하다고 강조했다. AI 모델은 데이터셋의 일부만 학습하며, 세대가 반복될수록 원본 데이터의 특성이 사라지고 점점 더 많은 오류가 추가되어 퇴행적 학습 과정을 겪게 되면서 후속 모델은 처음 모델과 전혀 다른 결과물을 생성한다. 연구진은 모델 붕괴가 잘못된 피드백 루프에 해당하는 현상이라며, 모델이 자체 결과물에 더 많이 의존할수록 현실에서 더 멀어진다고 설명했다. 인간이 만든 데이터와 LLM에서 생성된 콘텐츠의 구별이 점점 더 어려워지는 상황에서, 데이터 출처 관리의 중요성도 증대한다. LLM을 활용해 작성한 인터넷 콘텐츠가 많아질수록 후속 모델 교육용으로 수집한 데이터의 오염이 심각해질 수 있으며, 인간이 만든 데이터의 가치는 더욱 높아질 전망이다.

310) Nature, AI models collapse when trained on recursively generated data, 2024.07.24.

3) 연구소 - 앨런AI연구소, 오픈소스 LLM ‘올모(OLMo)’ 공개³¹¹⁾

마이크로소프트 공동 창립자인 폴 앨런(Paul Allen)이 설립한 비영리단체 앨런AI연구소(AI2)가 2024년 2월 1일 오픈소스 LLM 겸 프레임워크 ‘올모(OLMo)’를 공개했다. 공익을 위한 AI 연구를 추구하는 AI2는 오픈AI의 GPT나 구글의 제미니와 같은 LLM의 폐쇄성으로 인해 현재의 생성 AI 환경에 투명성이 부족하다고 주장했다. 반면 올모는 모델 코드와 가중치뿐 아니라 업계 최초로 훈련 코드와 사전훈련 데이터까지 제공하는 진정한 오픈소스 LLM이라고 강조했다. AI2는 올모 프레임워크의 목적이 연구자와 개발자들에게 LLM의 작동 방식을 투명하게 공개하여 LLM의 교육과 연구를 지원하기 위함이라고 설명했다.

올모는 2조 개 이상의 토큰으로 학습된 매개변수 70억개의 모델과 매개변수 10억 개의 모델로 구성되며, 각각의 모델 가중치와 추론 코드, 훈련 로그와 평가도구도 포함한다. 올모 프레임워크는 3조 개 이상의 토큰으로 구성된 사전학습용 데이터셋 ‘돌마(Dolma)’를 비롯한 개방형 AI 개발도구도 제공하며, 모든 리소스는 깃허브와 허깅페이스에서 무료로 다운로드 가능하다.

벤치마크 테스트 결과, 올모는 ‘라마(LLaMA)’나 ‘팔콘(Falcon)-7B’를 비롯한 동일 매개변수의 여타 오픈소스 모델과 대체로 비슷한 수준의 성능을 기록했다. 올모 7B는 생성과 독해 작업에서 라마2 7B와 비슷한 성능을 보였으나 다양한 문제 해결 능력을 평가하는 MMLU에서는 28.3점을 기록해 라마 7B(31.5점)나 라마2 7B(45점)보다 낮게 평가되었다.

AI2는 이번 발표가 올모 프레임워크의 첫 단계일 뿐이라며 올모를 완전한 오픈소스로 공개함으로써 AI 연구자들과 협력해 가장 개방적인 LLM 구축이 가능해질 것으로 기대한다. AI2는 데이터브릭스, AMD, 워싱턴대 등과 협력해 올모를 개발했으며, 향후 다양한 모델 크기와 기능, 데이터셋을 추가해 프로젝트를 지속적으로 발전시켜 나갈 계획이다.

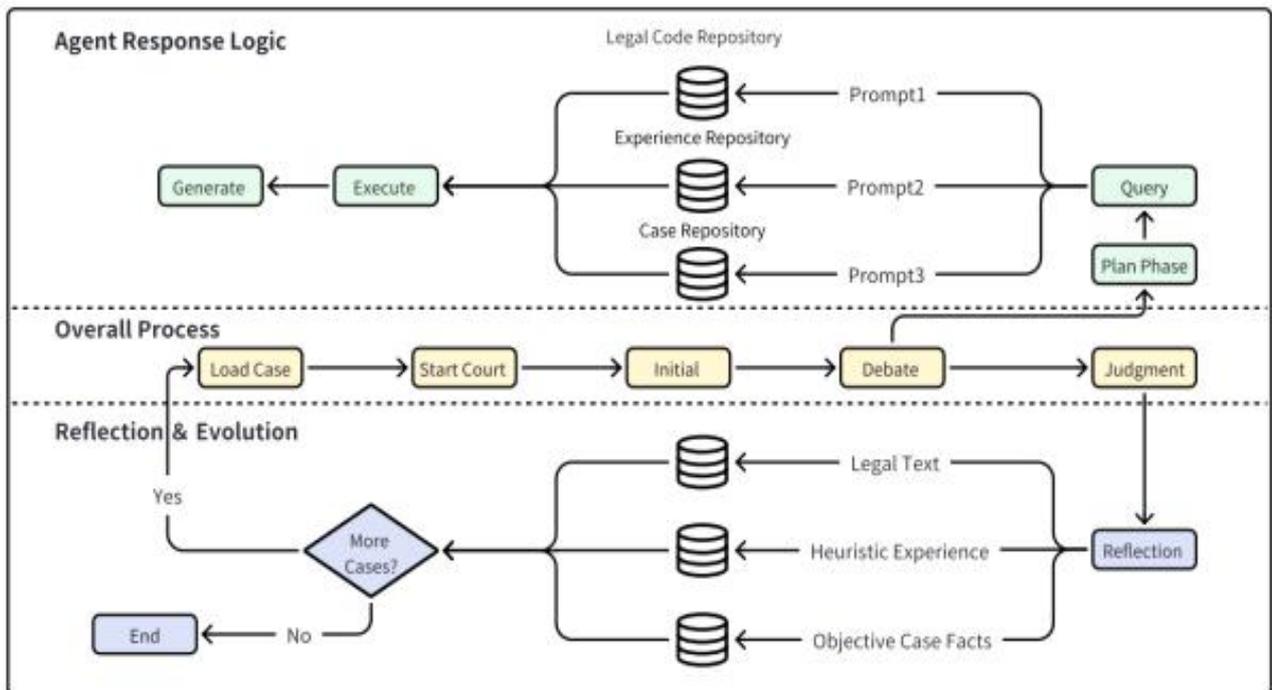
311) Allen Institute for AI, OLMo: Open Language Model, 2024.02.01.

4) 연구소 - 중국과학원(CAS) 등, LLM의 법률 역량 향상 기법 개발³¹²⁾

중국과학원(CAS)과 중국 주요 대학의 연구진이 2024년 8월 15일 LLM으로 재판관, 변호사와 같은 역할의 시뮬레이션을 수행해 LLM의 법률 특화 역량을 향상하는 방법에 관한 논문 (AgentCourt: Simulating Court with Adversarial Evolvable Lawyer Agents)³¹³⁾을 공개했다. 연구진들은 1,000건의 실제 민사 사건을 훈련 샘플로 사용해 시뮬레이션을 반복한 결과, 변호사 에이전트는 법률 업무를 처리하는 능력이 꾸준히 향상되는 것을 확인했다.

바이두(Baidu) 어니(ERNIE-Speed-128k) 기반의 ‘에이전트코트(AgentCourt)’는 원고, 피고, 원고 측 변호사, 피고 측 변호사, 판사, 서기 에이전트를 각각 설정해 민사 사건의 법정 시뮬레이션을 수행한다. 각 에이전트는 법정에서 자신의 역할에 따라 행동하도록 설계되어, 원고 측 변호사와 피고 측 변호사는 의뢰인과 연락해 증거를 제시하고 법률과 판례를 인용하여 변론하며, 판사는 변호사의 주장을 듣고 판결한다.

[그림 6-11] 법률 업무 처리 시뮬레이션



출처: Guhong Chen 외, AgentCourt: Simulating Court with Adversarial Evolvable Lawyer Agents, 2024

312) Arxiv, AgentCourt: Simulating Court with Adversarial Evolvable Lawyer Agents, 2024.08.15.

313) <https://arxiv.org/abs/2408.08089>

연구진은 LLM 기반의 법정 시뮬레이션을 통해 에이전트끼리 서로 경쟁하면서 능력을 발전시키는 적대적 진화 전략을 통해 변호사의 능력을 개선했으며, 이를 위해 세 가지 종류의 데이터베이스를 활용했다. 재판 후 변호사 에이전트는 재판 과정을 요약해 경험, 교훈, 성공적 전략을 추출해 향후 유사 사례에 참조하기 위한 경험 데이터베이스에 저장한다. 각 사건의 구체적 내용은 사례 데이터베이스에 저장되며, 방대한 사례를 축적해 향후 관련 판례를 활용할 수 있다. 사건과 관련된 법률 조항은 법률 코드 데이터베이스에 저장되어, 시간 경과에 따라 법률 지식 기반을 확장하여 더욱 정교하고 포괄적인 대응을 지원한다.

연구진은 법정 시뮬레이션을 통해 개선된 에이전트의 성과를 종합적으로 평가하기 위해 자동 평가와 수동 평가를 진행했으며, 두 평가에서 모두 상당한 개선을 확인하였다. LLM의 법률 역량을 평가하는 로벤치(LawBench)를 활용한 자동 평가 결과, ▲법률 지식 암기 ▲법률 지식 이해 ▲법률 지식 적용의 3개 영역에서 모두 성과가 향상되었다. ▲인지적 민첩성 ▲전문 지식 ▲논리적 엄격성의 3개 영역으로 기준으로 법률 전문가가 수동으로 진행한 평가에서는 특히 전문 지식과 논리적 엄격성에서 상당한 개선을 기록했다(원고 에이전트 기준, 전문 지식은 17점에서 41점, 논리적 엄격성은 22점에서 39점으로 향상되었다).

제2절 국내 학술단체, 표준기관, 비영리 연구기관

1. 학술단체

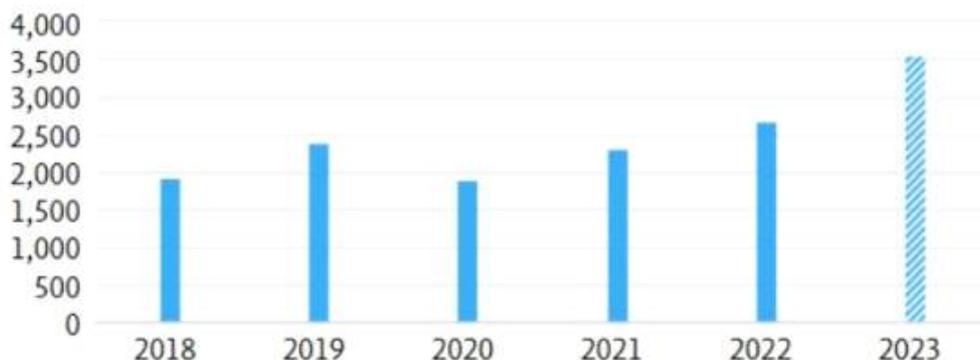
1) 학술연구 현황 및 추이

2024년 5월 한국과학기술기획평가원(KISTEP)이 7월 발표한 '국가전략기술 R&D 인력 실태 조사'³¹⁴⁾에 따르면 지난 6년간 전 세계에서 발간된 AI 분야 논문 수가 약 73만 건이며, 이 중 한국에서 발간된 논문 수는 1만 4천여 건으로 전 세계 12위에 머물렀다.

전 세계 AI 연구의 트렌드는 미국과 중국이 주도하는 것으로 나타났다. 2021년 ▲퓨샷 학습(Few-Shot Learning · 적은 수의 데이터를 학습시키는 것) ▲컴퓨터 비전 ▲딥페이크 등이 AI 분야에서 최신 연구 주제로 떠올랐다. 이와 관련해 지난 6년간 가장 많은 논문을 발간한 상위 500명의 연구자는 미국(31.8%)과 중국(24.2%)에 분포하고, 상위 500명에 들어간 한국 출신 연구자는 5명으로, 세계 16위 수준이었다.

다만 국내에서 발간된 AI 분야 논문 수는 2020년을 기점으로 매년 증가 추세이다. AI 분야 상위 10개 연구기관은 한국과학기술원(KAIST), 서울대, 고려대 등 학계 중심으로 나타났다.

[그림 6-12] 인공지능 분야 국내 연구자의 논문 수

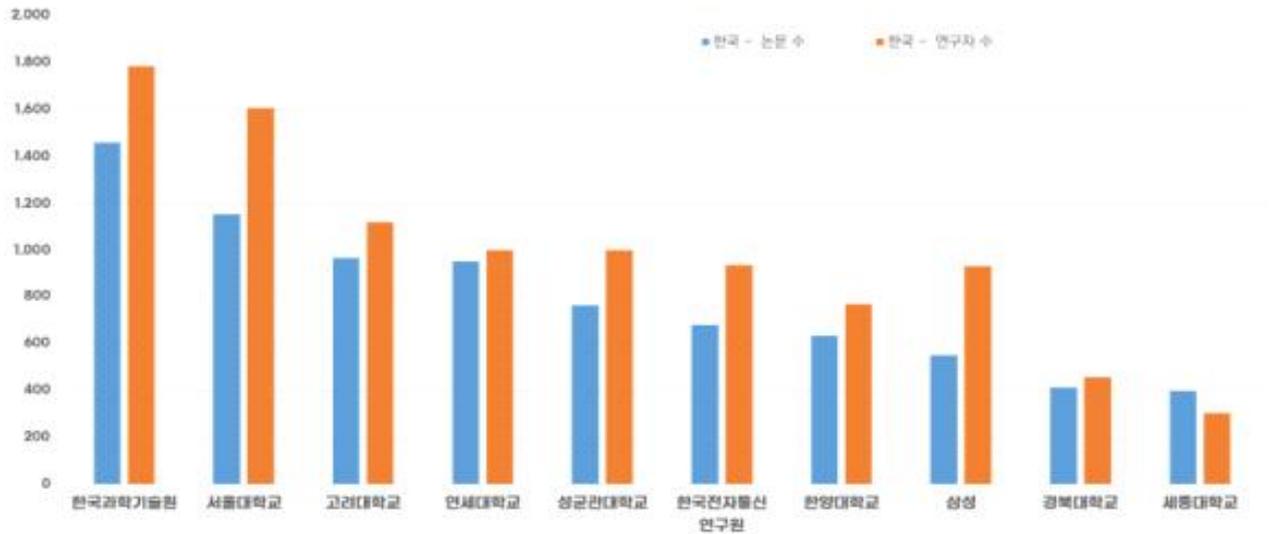


※ 불완전한 데이터

출처: 한국과학기술기획평가원(KISTEP), 국가전략기술 R&D(연구·개발) 인력 실태조사, 2024.5

314) 한국과학기술기획평가원(KISTEP), 국가전략기술 R&D(연구·개발) 인력 실태조사, 2024.5

[그림 6-13] 국내 상위 10개 학술·연구기관별 논문 및 연구자 수



출처: 한국과학기술기획평가원(KISTEP), 국가전략기술 R&D(연구·개발) 인력 실태조사, 2024.5

국내에서 AI 분야 주요 연구주제((토픽 클러스터)는 이미지분할, 이미지·자연어 처리, 로봇 관련 주제, 뇌-컴퓨터 인터페이스 등으로 나타났다.

<표 6-3> AI 분야 국내 주요 연구주제('18~'23)('24.5.기준)

순위	연구 주제((토픽클러스터)	논문 수
1	이미지 분할; 심층 신경망; 객체 감지(Image Segmentation; Deep Neural Network; Object Detection)	2,003
2	감정 분석; 자연어처리; 기계학습(Sentiment Analysis; Natural Language Processing; Machine Learning)	365
3	로봇; 인공지능; 표현 인식(Robot; Artificial Intelligence; Expression Recognition)	145
4	로봇; 행동 인지; 보행 분석(Robot; Action Recognition; Gait Analysis)	320
5	뇌-컴퓨터 인터페이스; 뇌파; 생물 의학 신호(Brain-Computer Interface; Electroencephalography; Biomedical Signal Processing)	217
6	강화 학습; 로봇; 인공지능(Reinforcement Learning; Robot; Artificial Intelligence)	277
7	이미지 처리; 로봇; 객체 인지(Image Processing; Robot; Object Recognition)	183
8	로봇; 기계 설계; 제어 시스템(Robot; Machine Design; Control System)	249
9	로봇; 동작 계획; 장애물 회피(Robot; Motion Planning; Obstacle Avoidance)	145
10	로봇; 인공지능; 이미지 처리(Robot; Artificial Intelligence; Image Processing)	213

출처: 한국과학기술기획평가원(KISTEP), 국가전략기술 R&D(연구·개발) 인력 실태조사, 2024.5

2) 2024년 주요 동향

(1) 한국정보과학회(KIISE)

한국정보과학회는 컴퓨터 과학 및 정보 기술 분야에서 가장 오래되고 권위 있는 학술 단체 중 하나로, AI분야도 중점적으로 다루고 있다. 한국정보과학회는 매년 AI와 관련된 여러 학술대회와 세미나를 개최하며, 한국컴퓨터종합학술대회(KCC)와 같은 대형 행사에서 AI 관련 연구 발표와 기술 교류가 이루어진다. 학회 내에서 AI 연구회를 운영하며, AI 기술의 최신 동향과 연구 성과를 공유한다.

2024년 한국컴퓨터종합학술대회(KCC)³¹⁵⁾는 6월 26일부터 28일까지 제주국제컨벤션센터에서 개최되었다. 이번 행사에서는 “초거대 AI와 SW 혁신이 이끄는 미래 사회”라는 주제로 개최되었다. 이번 행사에서는 2023년 및 2023년 국내외 우수학술대회에서 발표된 논문의 저자를 초청하여, 핵심연구내용을 소개하는 프로그램(Top Conference 세션)을 진행하였다. 이 중 AI 분야의 발표 내용들은 아래와 같다.

<표 6-4> Top Conference 세션의 AI 관련 논문

순위	학술대회	논문제목	발표자
1	ASPLOS 2023	Optimus-CC: Efficient Large NLP Model Training with 3D Parallelism Aware Communication Compression	송재용 (서울대)
2	AAAI 2024	Self-Supervised Framework based on Subject-wise Clustering for Human Subject Time Series Data	이하림 (한양대)
3	AAAI 2024	VITA: 'Carefully Chosen and Weighted Less' Is Better in Medication Recommendation	김태리 (한양대)
4	IEEE VR 2024	A Novel Approach for Virtual Locomotion Gesture Classification: Self-Teaching Vision Transformer for a Carpet-Type Tactile Sensor	주호택 (GIST)
5	NeurIPS 2023	Topological RANSAC for instance verification and retrieval without fine-tuning	Guoyuan An (KAIST)
6	EMNLP 2023	GLEN: Generative Retrieval via Lexical Index Learning	이선경 (성균관대)

315) <https://www.kiise.or.kr/conference/kcc/2024/>

순위	학술대회	논문제목	발표자
7	AAAI 2024	Bilingual outside Knowledge-based Visual Question Answering via Graph Representation Pretraining	김민준 (서울과학기술대)
8	EMNLP 2023	Large Language Models are Temporal and Causal Reasoners for Video Question Answering	이지수 (고려대)
9	NeurIPS 2023	NuTrea: Neural Tree Search for Context-guided Multi-hop KGQA	주재원 (고려대)
10	NeurIPS 2023	Navigating Data Heterogeneity in Federated Learning: A Semi-Supervised Federated Object Detection	김태현 (KAIST)
11	NeurIPS 2023	Softmax Output Approximation for Activation Memory-Efficient Training of Attention-based Networks	이창현 (UNIST)
12	HPCA 2024	Smart-Infinity: Fast Large Language Model Training using Near-Storage Processing on a Real System	장흥선 (서울대)
13	NeurIPS 2023	Advancing Bayesian Optimization via Learning Correlated Latent Space	이승훈 (고려대)
14	NeurIPS 2023	UP-NeRF: Unconstrained Pose Prior-Free Neural Radiance Field	김인재 (고려대)

2024 한국컴퓨터종합학술대회 (KCC2024)³¹⁶; SPRI 정리, 2024.11.

(2) 한국인공지능학회(KSAD)

한국인공지능학회(Korean Artificial Intelligence Association)는 2016년 설립되어 AI 관련한 컴퓨터비전 및 패턴인식, 자연어 처리, 바이오 인포매틱스, 뇌인지컴퓨팅, 기계학습 및 그와 관련한 응용 분야 등에 관한 학문-기술 연구 및 교육을 수행한다.

한국인공지능학회는 2024 하계학술대회를 2024년 8월 15일부터 17일까지 부산 벡스코 컨벤션홀에서 3일간 개최하였다. 이홍락 교수(미시건대)의 “Towards Learning AI Agents for Solving Real-world Tasks” gan) Yoshua Bengio 교수(몬트리올대)의 “Mitigating Catastrophic AI Risks with Safe-by-Design AI” 등의 강연과 함께 우수논문 발표 등이 있었다. 2024 추계학술대회를 2024년 11월 21일부터 23일까지 삼성 SDS 본사에서 3일간 개최했다. 이외에도 한국인공지능학회는 2024년 인공지능 춘계 단기강좌, 2024년 8월 ‘AI 시대의 인과추론’ 세미나 등을 진행하였다.

316) <https://www.kiise.or.kr/conference/kcc/2024/>

2. 표준기관

1) 표준화 현황 및 추이

우리나라는 정부 주도의 표준화 작업, 산업별 표준 개발, 그리고 글로벌 표준과의 연계를 중심으로 전개되고 있다. AI 기술의 급속한 발전과 상용화에 대응하기 위해 AI의 신뢰성, 안전성, 윤리적 사용 등을 보장하는 표준 개발을 추진 중이다. 우리나라는 ISO/IEC와 같은 국제 표준화 기구와 협력하면서 ISO/IEC JTC 1/SC 42(인공지능 분야의 표준화 분과), ITU(*에서 진행하는 AI 관련 표준화 작업에도 참여하고 있다.

2) 2024년 주요 동향

(1) 한국표준협회(KSA)

한국표준협회는 2021년부터 AI 제품의 품질과 신뢰성 확보를 지원하고, 관련 기술의 표준화를 선도하기 위해 AI+ 인증제도를 개발하여 AI 관련 기업에 인증 서비스를 제공하고 있다. 2023년 11월 7일 기준, AI+ 인증을 받은 제품은 신한카드의 ‘초개인화 시스템 2.0’, 삼성전자의 ‘그랑데 건조기 AI’ 등 46개이다.³¹⁷⁾ 한국표준협회 AI+ 인증은 아래 그림과 같다.

[그림 6-14] AI+ 인증 절차



출처: 한국표준협회³¹⁸⁾

317) https://ksa.or.kr/ksa_kr/7021/subview.do

318) https://ksa.or.kr/sites/ksa_kr/files/AI+_guidebook.pdf

(2) 국가기술표준원(KATS)

국가기술표준원은 2021년 10월 산업 수요자 중심의 표준 개발을 지원하기 위해 학계, 연구기관, 산업계가 참여하는 민간 중심의 ‘산업 인공지능 표준화 포럼’을 설립했다. 이 포럼은 AI의 신뢰성에 대한 평가기준과 윤리 가이드라인 개발, AI 적용 산업별 상호운용성 확보 등을 위한 표준화를 추진한다. 이 포럼은 신뢰성, 데이터, 활용사례, 윤리·사회적 문제의 4개 분과위원회를 운영하며, 한국표준협회가 운영사무국 역할을 수행한다.³¹⁹⁾

2024년 2월 인공지능 분야(JTC 1/SC 42) 국가표준(KS) 개발 수요조사를 진행하였다.

[그림 6-15] 산업 인공지능 표준화 포럼 구성



출처: 산업 인공지능 표준화 포럼³²⁰⁾

국가기술표준원은 2023년 6월 전 세계적으로 생성 AI 기술의 급속히 확산으로 AI 윤리 문제가 화두인 가운데 AI 윤리에 대한 첫 번째 국가표준(KS)을 제정하였다. 이 표준은 2021년부터 산·학·연 전문가들이 참여하는 ‘AI 윤리 표준화 포럼’에서 논의와 검증 과정을 거쳐 개발되었다.

319) https://aistandard.kr/sub/sub01_01.php

320) http://aistandard.kr/sub/sub01_01.php (2024.10.14. 확인)

(3) 한국정보통신기술협회(TTA)

한국정보통신기술협회는 정보통신기술의 표준 제정, 보급 및 시험인증 지원을 위해 1988년 설립 이후 국내 ICT 표준화와 시험인증 등 역할을 하고 있다.

한국정보통신기술협회는 국내 ICT 표준화 기구로서, AI·5G·사물인터넷·UHD 등 ICT 및 ICT융합 분야에 대해 단체표준 제정 활동과 국제표준화기구, 사실표준화기구 등과 글로벌 네트워크를 구축하여 ICT 표준화를 주도적으로 수행한다.

2024년 10월 14일 기준으로 한국정보통신기술협회의 인공지능 분야에서 제·개정된 표준 목록은 아래의 표와 같다.

<표 6-5> 한국정보통신기술협회의 인공지능 분야에서 발행된 표준 목록

No.	표준종류	표준번호	표준제목	제개정일
1	정보통신단체표준 (TTAS)	TTAE.ET-GS ZSM 012	제로터치 네트워크 및 서비스 관리(ZSM): 인공지능 기반 네트워크 및 서비스 자동화를 위한 실현 기술	2023-12-06
2	정보통신단체표준 (TTAS)	TTAK.KO-10.1497	인공지능 시스템 신뢰성 제고를 위한 요구사항	2023-12-06
3	정보통신단체표준 (TTAS)	TTAK.KO-11.0324	인공지능 기반 해사 소프트웨어의 안전 요구사항 도출 지침	2023-12-06
4	정보통신기술보고서 (TTAR)	TTAR-06.0280	공공안전을 위한 무기준 인공지능 영상 품질 분석 기술(기술보고서)	2023-11-03
5	정보통신기술보고서 (TTAR)	TTAR-11.0085	IALA G1178: IALA 관점에서의 인공 지능(AI) 도입 지침(기술보고서)	2023-11-02
6	정보통신기술보고서 (TTAR)	TTAR-06.0273	무선 네트워크의 다목적 자원관리를 위한 분산형 인공지능 활용 (기술보고서)	2023-10-19
7	정보통신기술보고서 (TTAR)	TTAR-06.0274	비예약 기반 비직교 무선 접속 방식을 위한 인공지능 활용 (기술보고서)	2023-10-19
8	정보통신기술보고서 (TTAR)	TTAR-06.0275	무경계 셀룰러 시스템의 모바일 에지 컴퓨팅 자원 할당을 위한 인공지능 활용 (기술보고서)	2023-10-19
9	정보통신기술보고서 (TTAR)	TTAR-06.0276	5G 단말 기계학습 및 인공지능 서비스 지원 기술 (기술보고서)	2023-10-19
10	정보통신기술보고서 (TTAR)	TTAR-10.0178	인공지능 시스템 신뢰성 지침 및 원칙 동향(기술보고서)	2023-10-18
11	정보통신기술보고서 (TTAR)	TTAR-10.0174	OCR기반 차트 이미지 데이터 내용을 텍스트로 추출하는 인공지능 기술(기술보고서)	2023-05-18

No.	표준종류	표준번호	표준제목	제개정일
12	정보통신단체표준 (TTAS)	TTAK.KO-10.1370	온실 운영 시스템과 인공지능 서버 간 인터페이스	2022-12-07
13	정보통신단체표준 (TTAS)	TTAK.KO-11.0280- Part4	검증용 데이터세트의 밸런스 기반 인공지능 소프트웨어 신뢰성 평가 방법 - 제4부: 웨이브 타입 밸런스 데이터 설계	2022-12-07
14	정보통신단체표준 (TTAS)	TTAK.KO-11.0280- Part5	검증용 데이터세트의 밸런스 기반 인공지능 소프트웨어 신뢰성 평가 방법 - 제5부: 동영상 타입 밸런스 데이터 설계	2022-12-07
15	정보통신단체표준 (TTAS)	TTAK.KO-10.1303	진단 보조 인공지능 모델 개발을 위한 학습용 데이터 구축방안 : 병리조직 이미지	2021-12-08
16	정보통신단체표준 (TTAS)	TTAK.KO-10.1340	모의 환경 기반 인공지능 게임 에이전트 생성 요구사항	2021-12-08
17	정보통신단체표준 (TTAS)	TTAK.KO-11.0280- Part2	검증용 데이터세트의 밸런스 기반 인공지능 소프트웨어 신뢰성 평가 방법 - 제2부: 이미지 타입 밸런스 데이터 설계	2021-12-08
18	정보통신단체표준 (TTAS)	TTAK.KO-11.0280- Part3	검증용 데이터세트의 밸런스 기반 인공지능 소프트웨어 신뢰성 평가 방법 - 제3부: 시계열 타입 밸런스 데이터 설계	2021-12-08
19	정보통신기술보고서 (TTAR)	TTAR-10.0158	인공지능 기반 이미지·영상분석 기술 개요 및 서비스 유즈케이스(기술보고서)	2021-10-21
20	정보통신기술보고서 (TTAR)	TTAR-10.0160	인공지능 생성요약을 활용한 작문 평가 시스템 유즈케이스(기술보고서)	2021-10-21
21	정보통신기술보고서 (TTAR)	TTAR-10.0142	스마트 헬스케어에 위한 인공지능 피트니스 코칭 시스템 프레임워크 유즈케이스(기술보고서)	2021-10-20
22	정보통신단체표준 (TTAS)	TTAK.KO-10.1276	유방암 판독 인공지능 모델 개발을 위한 유방촬영술 의료지식베이스 구축방안	2020-12-10
23	정보통신단체표준 (TTAS)	TTAK.KO-11.0280- Part1	검증용 데이터세트의 밸런스 기반 인공지능 소프트웨어 신뢰성 평가 방법 - 제1부: 방법론 및 체계	2020-12-10
24	정보통신기술보고서 (TTAR)	TTAR-10.0111	인공지능 프로세서의 보안성 강화를 위한 안전성 플랜(기술보고서)	2019-11-07
25	정보통신기술보고서 (TTAR)	TTAR-10.0112	인공지능 프로세서의 기능안전 인터페이스(기술보고서)	2019-11-07
26	정보통신기술보고서 (TTAR)	TTAR-10.0113	다중 칩 구성이 가능한 인공지능 프로세서 구조(기술보고서)	2019-11-07
27	정보통신기술보고서 (TTAR)	TTAR-10.0114	다중 칩 기반 인공지능 프로세서 소프트웨어 인터페이스(기술보고서)	2019-11-07
28	정보통신기술보고서 (TTAR)	TTAR-10.0107	인공지능 윤리 가이드라인 제정 동향(기술보고서)	2019-10-30
29	정보통신기술보고서 (TTAR)	TTAR-10.0108	인공지능 - 표준화 동향 분석 (기술보고서)	2019-10-30
30	정보통신기술보고서 (TTAR)	TTAR-10.0096	인공 지능 서비스의 튜링 테스트 기반 평가 기술 동향(기술보고서)	2018-11-07
31	정보통신기술보고서 (TTAR)	TTAR-11.0068	임베디드 시스템에서의 인공지능 서비스를 위한 신경망 구조 표현 모델의 요구사항(기술보고서)	2018-11-07
32	정보통신기술보고서 (TTAR)	TTAR-10.0057	지능정보 기술 동향 - 제1부 : 인공지능 기술 (기술보고서)	2016-11-03

출처: TTA 표준검색 웹페이지(2023.11.6. ‘인공지능’ 키워드로 검색)³²¹⁾

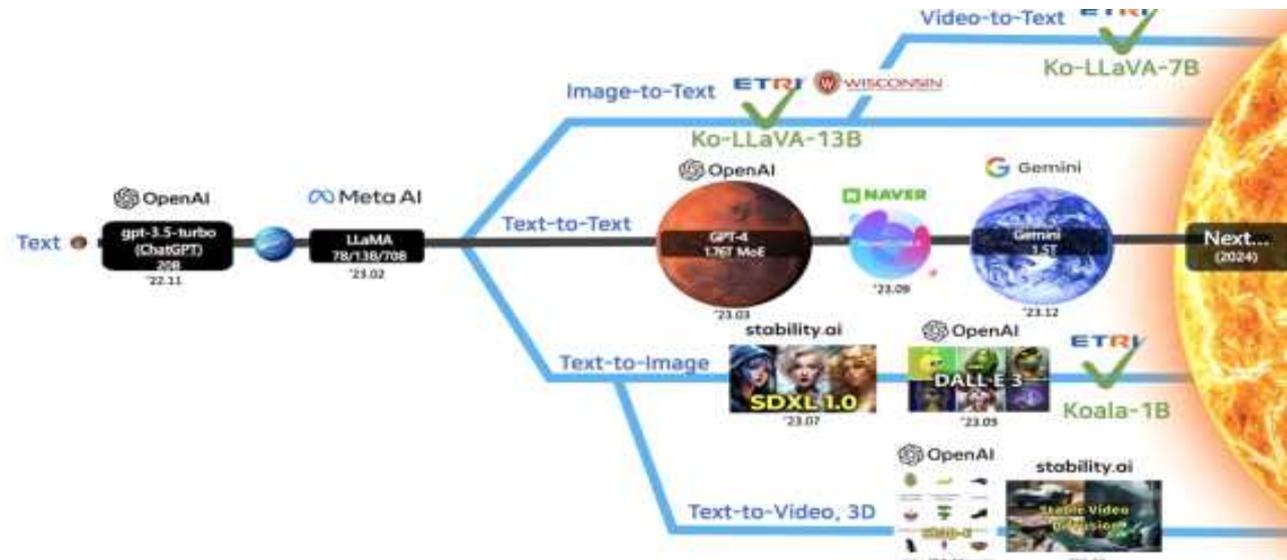
321) https://committee.tta.or.kr/data/standard_list_N.jsp

3. 연구기관

1) 한국전자통신연구원(ETRI), AI 연구 및 표준 개발

한국전자통신연구원은 AI 분야에서 다양한 연구와 개발을 주도하고 있다. 2024년 1월 ETRI는 문장을 입력하면 이미지를 생성하는데 기존 대비 5배 빠른 ‘코알라(KOALA)’ 3종 모델과 이미지나 영상을 불러와 질답이 가능한 대화형 시각언어모델 ‘코라바(Ko-LLaVA)’ 2종 등 총 5종 모델을 오픈소스로 공개했다.

[그림 6-16] 생성 시각지능 기술의 향후 발전 방향과 ETRI 기술 적용



출처 : 인공지능신문(<https://www.aitimes.kr>)

한국전자통신연구원은 2024년 5월 AI 기술 기반으로 화장품의 발림성을 분석하는 시스템을 개발했다. 이 시스템으로 화장품을 피부에 발랐을 때 느낄 수 있는 질감을 99% 이상 정확하게 맞출 수 있었다. 연구진은 피부에 화장품을 바르는 행위와 유사한 환경에서 데이터를 획득하고, 시간의 경과에 따라 변하는 1차원의 마찰 신호를 2차원의 주파수 스펙트럼 형태로 재해석하여 원하는 시간-주파수 혼합 신호를 추출하여 분석했다. 이러한 기법을 활용하여 결과값의 정확도를 99% 이상으로 높일 수 있었다.³²²⁾

322) <https://cosinkorea.com/mobile/article.html?no=51785>

한국전자통신연구원(ETRI)은 2024년 7월 16일 AI 개발 과정에서 사용되는 데이터의 품질을 평가하고 관리할 수 있는 국제표준을 개발했다고 발표하였다.³²³⁾ ‘데이터 분석 및 머신러닝을 위한 데이터 품질 표준’은 조직이 AI 개발 과정에서 데이터 품질을 평가, 관리, 개선할 수 있는 도구와 방법을 제공하며, 데이터가 목적에 맞도록 사용될 수 있도록 돕고, 일관되고 신뢰할 수 있는 공통의 용어와 실천 방안을 제시하였다. 한국전자통신연구원연구진은 미국, 일본, 독일, 중국, 영국 등과 협력해 개발한 이번 표준이 AI 시스템의 책임 있는 개발과 배포를 위한 기틀을 마련하고 다양한 이해관계자 간 협력 강화에 중요한 역할을 할 것으로 기대한다. 한국전자통신연구원 연구진은 총 6개 부분으로 구성된 ‘데이터 분석 및 기계학습을 위한 데이터품질’ 시리즈 중 AI에 사용되는 데이터 품질에 대한 개요와 공통 개념을 정립하였다. 개발된 표준은 ▲데이터 품질 보장을 통해 안전한 AI 모델 개발과 오류 및 편향 최소화 ▲AI 시스템의 신뢰성을 높여 성능의 일관성을 유지하면서 데이터의 생애주기 관리 ▲AI 안전성을 강화하여 리스크 관리와 법규 준수를 돕는 역할을 지원하는 데 중점을 두고 있다. 데이터 품질 관점에서 AI 안전과 신뢰성을 지원하는 세계 최초의 표준으로서, 향후 관련 규제 준수도 지원해 법적 분쟁을 줄이고 규제기관의 감시를 완화하는 효과가 기대된다. 기업 간 상호운용성을 높이고 고품질의 데이터 확보를 기반으로 혁신적인 AI 응용 서비스 개발도 촉진할 수 있을 전망이다.

2) 정보통신정책연구원(KISDI)

KISDI는 국내외 인공지능 기술, 산업 동향 및 정책 이슈를 포괄하는 이슈 페이퍼 및 학술 논문을 ‘KISDI AI Outlook’을 통해 게재하고 있다. ‘KISDI AI Outlook’에 게재된 2024년 주요 연구보고서들은 아래의 표와 같다.

<표 6-6> KISDI의 인공지능 분야에서 발행된 주요 연구보고서 목록

번호	보고서 제목	발행일
1	중국의 조용한 침공: AI, 반도체 붉은 공급망	2024.3.31.
2	대량맞춤화를 실현하는 인공지능 기반 인간-로봇 협업	2023.3.31.
3	생성 모델과 변분 오토인코더(Variational Autoencoder)	2024.6.28.
4	인공지능 공정성 동향 분석	2024.6.28.
5	인공지능 시스템 이용자 보호 관련 해외 규율 동향과 시사점	2024.9.30.

출처: KISDI AI Outlook 웹페이지³²⁴⁾

323) 한국전자통신연구원, ETRI, AI안전 및 신뢰성지원 국제표준 제정, 2024.07.16.

324) KISDI AI Outlook 웹페이지, 2024.10.14.(확인)

3) 소프트웨어정책연구소(SPRi)

소프트웨어정책연구소는 인공지능, 메타버스 등 새로운 소프트웨어 트렌드를 선도적으로 분석하고 이슈를 발굴하여 정책 대안 제시, 법제도 개선, 통계 생산 등 연구활동을 추진해 오고 있다. 소프트웨어정책연구소는 연구보고서, 이슈리포트, 산업정책동향 등을 발행하며, 2024년도 AI 분야의 주요 연구보고서 목록은 아래의 표와 같다.

<표 6-7> SPRi의 AI 분야에서 연구보고서 목록

구분	보고서 제목	발행일
1	산업 수요 중심의 국가 AI R&D 전략성 강화를 통한 AI 확산 방안 연구	2024. 4.29.
2	공공부문 AI 도입현황 연구	2024. 4.29.
3	국외 AI 시장 및 창업지원 정책 동향 조사	2024. 4.29.
4	국내 AI 창업기업 비즈니스 현황 분석	2024. 4.29.
5	글로벌 AI 신뢰성 정책 동향 분석	2024. 4.29.
6	메타버스산업 분류체계 연구	2024. 4.29.
7	메타버스 성장통 극복 방안 연구	2024. 4.29.
8	2023년 메타버스 시장 및 동향 분석	2024. 4.29.
9	2023년 국내외 인공지능 산업 동향 연구	2024. 4.29.
10	생성형 AI에 대응한 SW인재양성 정책 방향 연구	2024. 4.29.

출처: SPRi 웹페이지³²⁵⁾

소프트웨어정책연구소는 2024년 9월 27일 ‘안전하고 책임 있는 AI를 위한 노력과 방향’을 주제로 제61회 SPRi 포럼을 개최하였다. 이번 행사에서는 AI 안전성 및 책임 있는 AI를 위한 국내외 동향을 파악하고, 향후 발전방향에 대한 발표와 토론의 진행되었다.

325) https://sprri.kr/posts?data_page=1&code=data_all&study_type=&board_type= (2024.10.14 확인)

4) 한국지능정보사회진흥원(NIA)

한국지능정보사회진흥원은 국내외 인공지능 관련, 정책, 법제 등에 대한 연구와 사업 등을 수행하고 있다. 또한 매년 국가지능정보화백서를 발간하고 있다. 2024년 주요 연구보고서(이슈보고서)들은 아래의 표와 같다.

<표 6-8> NIA의 AI 관련 연구보고서 사례

구분	보고서 제목	게시일
1	AI허브-해외플랫폼 현황 비교분석	2024. 8.30.
2	버티컬 시로의 변화와 과제	2024. 9.23.
3	대형 멀티모달(LMM) 동향 조사	2024. 10.2.
4	직장인이 꼭 알아야 할 생성형 AI윤리	2024. 10.10.
5	개발자가 꼭 알아야 할 생성형 AI윤리	2024. 10.10.

출처: NIA 웹페이지³²⁶⁾

326) https://www.nia.or.kr/site/nia_kor/ex/bbs/List.do?cbIdx=82618, (2024.10.14. 확인)

제7장 인공지능 교육 및 인력 동향

제1절 교육 분야

1. AI 교육 현황 및 추이³²⁷⁾

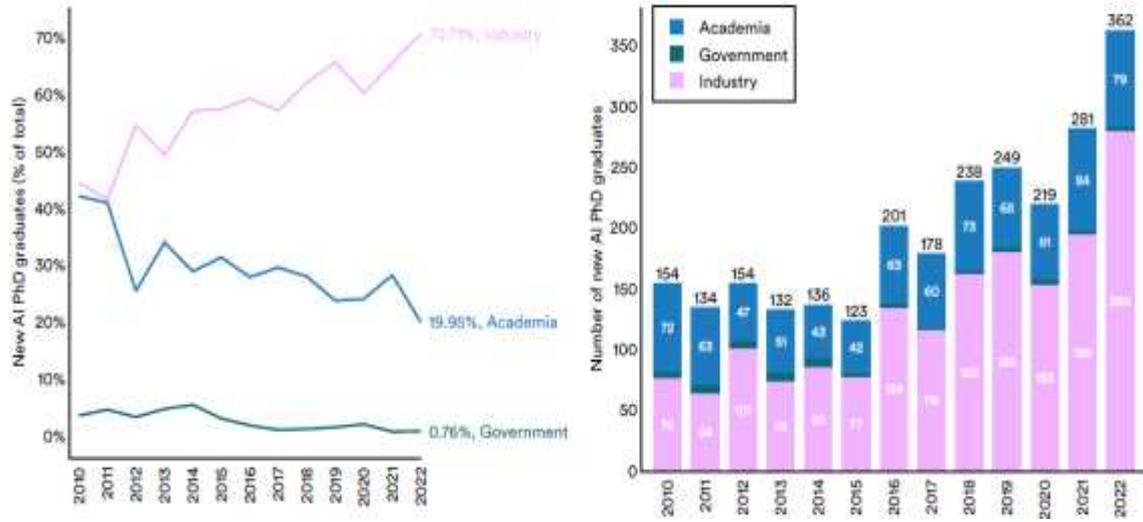
Stanford HAI의 AI Index 보고서는 북미(미국 및 캐나다) 및 유럽 지역에서의 대학 및 초중고의 전산학 및 AI 교육 추이와 분석하고 ChatGPT와 같은 AI 기술의 활용 현황을 분석하였다.

최근 북미지역(미국과 캐나다)의 전산학(CS: Computer Science) 및 관련 학과 졸업생·교수진 수는 증가하는 추세를 보이며, 박사학위 소지자의 산업계 취업이 가속화되고 있다. 지난 10년간 북미 소재 대학의 전산학 전공 졸업생 수가 지속적인 증가세를 보인 한편, 졸업자 중 유학생의 비율은 일부 감소했다. 2022년 기준 북미지역 소재 신규 전산학 졸업생 수는 전년 대비 ▲학사 7.9% 증가 ▲석사 2.5% 감소 ▲박사 11.2% 증가했다. 학위 취득자 중 유학생 비중은 지난 10년간 전반적으로 증가하고 있으나, 2022년에는 전년 대비 ▲석사(-14.8%) ▲박사(-2.7%) ▲학사(-1.1%) 순으로 감소했으며, 이는 트럼프 행정부 초기의 유학 비자 취득 제한 정책 및 코로나19 팬데믹 동안의 해외 이동 제한 조치에 따른 현상으로 추정된다.

북미지역 신규 전산학 박사학위 취득자 중 AI 전공자들의 산업계(민간기업) 취업 쏠림 현상이 심화하는 양상이며, 산업계에서 학계로 이동하는 인재는 점진적으로 감소 추세이다. 2011년에 AI 전공 신규 박사의 산업계(40.9%) 및 학계(41.6%) 진출률이 비등했던 데에 비해, 2022년 졸업 후 업계 진출 비중은 ▲산업계(70.7%) ▲학계(20.0%) 순으로 집계되었다. 한편, 신규 졸업자의 정부기관 진출 비율은 지난 5년간 약 0.7%대를 유지하며 비교적 낮고 안정적으로 유지되고 있다.

327) Stanford HAI, AI Index Report 2024

[그림 7-1] 북미지역 신규 AI 전공 박사의 진출 분야

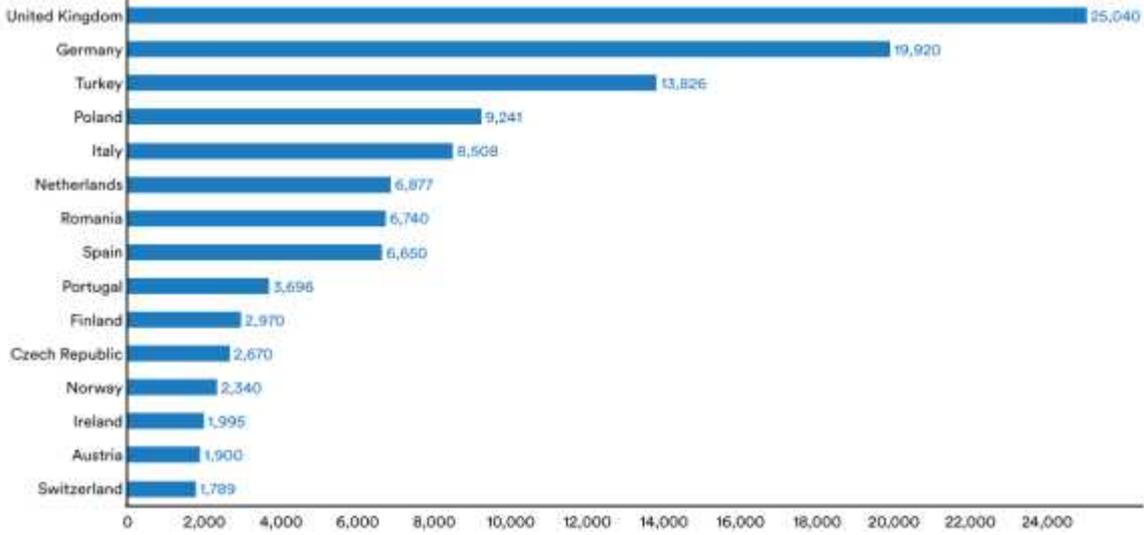


출처: Stanford HAI, AI Index Report 2024, 2024.4

북미 소재 대학의 전산학, 컴퓨터공학(CE: Computer Engineering) 및 정보학과(IT)의 전체 교수진 수는 지속적인 증가 추세이다. 2022년 기준 전산학, 컴퓨터공학 및 정보학과 교수진은 전년 대비 총 7.2% 증가하였으며, 유형별로는 ▲영년직(5,733명) ▲강의전담교수(1,252명) ▲기타 강사(919명) ▲박사후과정(531명) 순으로 집계되었다. 해당 학과 신입 교수 채용은 2021년 710명에서 2022년 915명으로 약 28.9% 증가했으며, 이는 10년 만에 최고치를 기록했다. 다만, 2022년 신입 교수진 임용의 43%는 다른 학계 직책에서 유입된 것으로 나타나, 학계 인력 내에서 교수진 이탈이 발생한 것으로 추정된다. 산업계에서 학계로 이동한 신입 교수진은 2021년 11%에서 2022년 7%로 감소하여 고급 AI 인재가 산업계에서 학계로 이동하는 비율이 점차 낮아지고 있다. 관련 학과의 정교수로 외국인을 채용한 비중은 19.3%로, 가장 많은 외국인을 채용했던 2013년 다음으로 두 번째로 높은 채용률을 기록했다.

유럽에서는 영국과 독일이 전산학, 컴퓨터공학 및 정보학 졸업생 배출을 선도하고 있다. 유럽에서 전산학, 컴퓨터공학 및 정보학 학사, 석박사 졸업생 배출을 가장 많이 하는 국가는 ▲영국 ▲독일 ▲튀르키예 ▲폴란드 ▲이탈리아 등이며, 인구 10만 명당 기준으로는 ▲핀란드 ▲노르웨이 ▲네덜란드 ▲아일랜드 등이 높게 나타났다. 해당 분야의 학부 졸업생은 2022년 기준 영국(25,040명)이 가장 많고 다음으로 독일(19,920), 튀르키예(13,826), 폴란드(9,241) 순이었다.

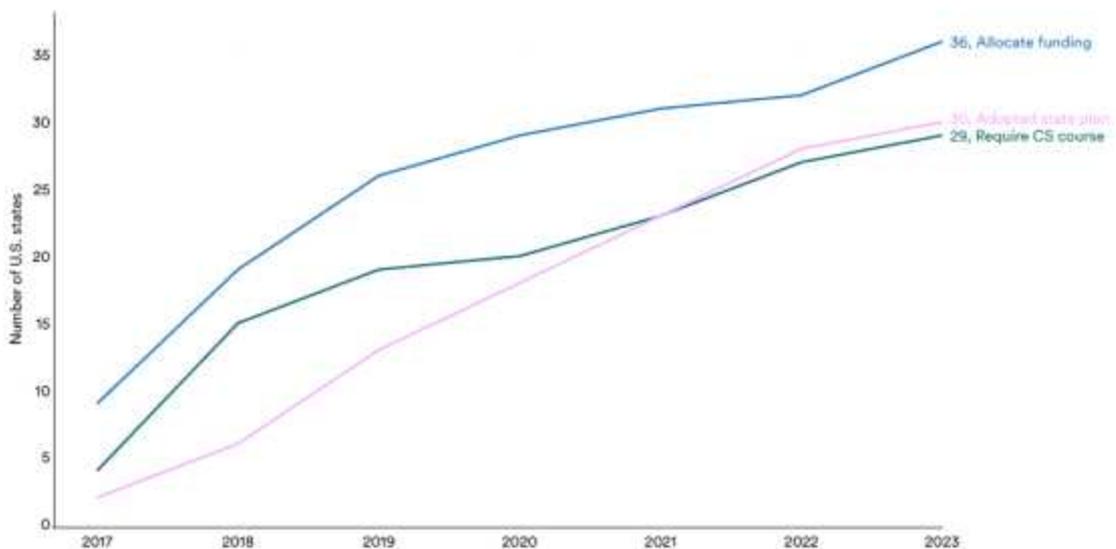
[그림 7-2] 전산학, 컴퓨터공학 및 정보학 전공 학부 졸업자 수(2022)



출처: Stanford HAI, AI Index Report 2024, 2024.4

미국 전역에서 초중고(K-12) 학생에 대한 전산학 교육이 확대되고 있다. 2017년에 몇몇 주 (state)만이 고등학교 전산학 프로그램을 지원했다. 2023년 기준 약 3분의 2의 주에서 고등학교에서 전산학을 가르쳐야 하며, 이를 위해 자금을 할당하고, CS 교육을 위한 주 계획을 개발했다. 2023년 초중고 학생에 대한 전산학 교육 예산을 비정한 주 비율은 36%, 교육 계획을 수립한 주 비율은 30%, 전산학 과정을 의무화한 주 비율은 29%까지 증가했다.

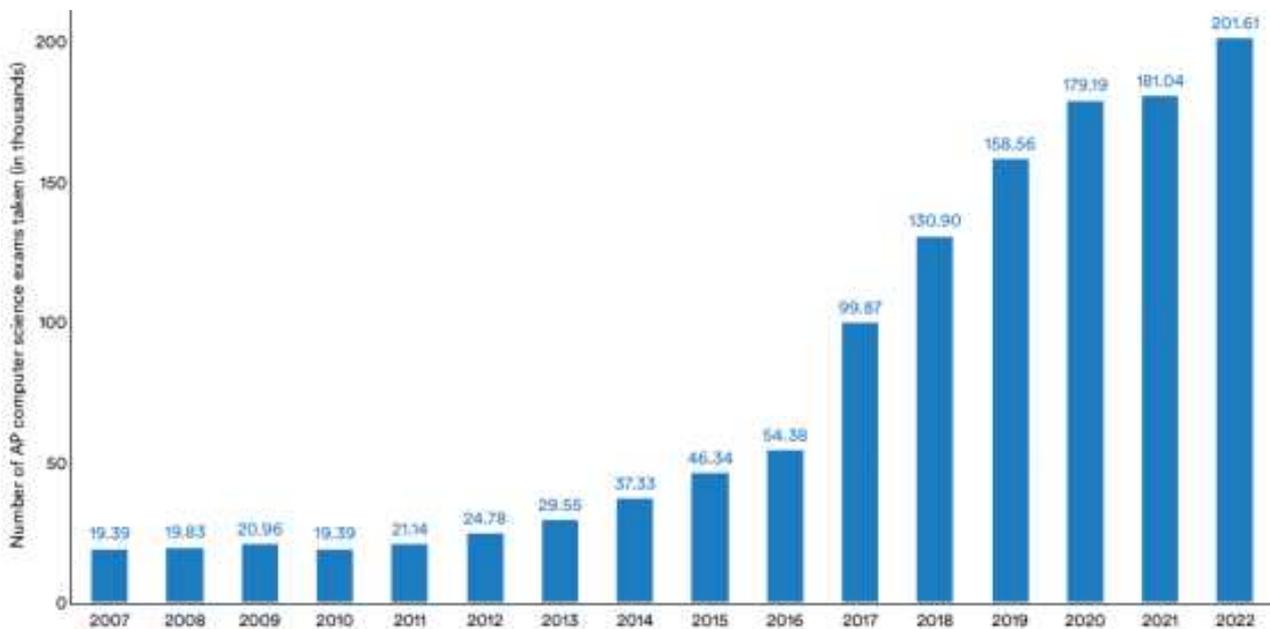
[그림 7-3] 연도별 미국 K-12 전산학 교육 주 비율 추이



자료: Stanford HAI, AI Index Report 2024

고등학교에 대학 학점을 취득할 수 있는 고급 교육과정인 AP(Advanced Placement)³²⁸⁾ 전산학 시험 시행 건수는 2007년(19.39회) 대비 2022년(201.61) 10배 이상 증가했다. 특히 ▲캘리포니아주(33,262건) ▲텍사스주(20,901건) ▲플로리다주(16,248건)에서 AP 전산학 시험 응시 건수가 가장 높았으며, 인구 10만 명당 시험 응시 건수는 ▲메릴랜드주(126.5건) ▲뉴저지주(112.7건) ▲매사추세츠주(92.7건) 순으로 높게 나타났다.

[그림 7-4] AP 전산학 시험 응시 건수



자료: Stanford HAI, AI Index Report 2024

리서치 업체 Impact Research의 설문조사에 따르면, 미국의 초중고 교사 대다수가 이미 ChatGPT를 활용하고 있으며, 사용비율이 빠르게 증가하고 있는 것으로 나타났다. 설문조사에 참여한 교사 중 2023년 3월 조사 기준 51%, 7월 조사 기준 63%는 한 번 이상 ChatGPT를 사용했다고 응답했으며, 활용 목적으로는 ▲수업 계획(응답자 30%) ▲창의적인 수업 아이디어 창출에 참고(30%) ▲배경지식 향상(27%) 등으로 응답했다. 또한, 교사의 75%와 학생의 65%는 ChatGPT를 교육과정에 통합하는 것이 중요하다고 응답했다.

328) 미국에서 대학 학점을 선취할 수 있는 고급 교육과정으로, 고등학교 학생들이 대학 1학년 때 수강하는 과목들을 고등학교 과정에서 미리 선택하여 수강하고, 시험을 통해 학점을 미리 취득할 수 있는 시스템

2. AI의 교육에 대한 영향

1) AI의 교육 형평성 · 포용성에 대한 영향³²⁹⁾

OECD는 2024년 8월 교육에서 형평성과 포용성을 촉진하는 데 있어 AI의 잠재력을 분석한 보고서(The potential impact of Artificial Intelligence on equity and inclusion in education)를 발표하였다. 이 보고서는 AI의 교육에서 평등과 포용성에 대한 영향을 분석하기 위한 개념적 틀을 정의하고, 학습자 중심적(learner-centred), 교사 주도적(teacher-led), 기타 제도적(institutional) AI 도구(tool)들을 조사하여 각각에 대한 영향(기회와 도전)을 분석하고, 정책적 권고사항을 제시한다.

이 보고서는 기존의 문헌연구를 기반으로 교육의 형평성과 포용성에 대한 개념적인 틀로 교육의 형평성과 포용성을 지원하기 위한 학습자 중심적(learner-centred), 교사 주도적(teacher-led), 기타 교육기관(institutional) AI 도구(tool)에 대한 정의로서 각각에 대한 목적과 예시를 아래(표)와 같이 도식화하였다.

〈표 7-1〉 교육에서 AI 도구의 교육

분류	목적	AI 도구의 예시
형평성 · 포용성을 지원하는 학습자 중심 도구	학생의 학습 경험을 향상하도록 설계	• 지능형 튜터링 시스템, AI 기반 시뮬레이션, 특수교육이 필요한 학생을 지원하는 AI 기반 도구 등
형평성 · 포용성을 지원하는 교사 주도 도구	교사의 교육 및 행정 역할 수행 지원	• AI 기반 로봇, 평가 및 교실 관리를 담당하는 도우미, 지속적인 전문 학습 코치 등
기타 형평성 · 포용성을 지원하는 기타 교육기관 도구	운영 효율성 개선, 입학 관리와 같은 보다 광범위한 교육기관의 목표 성취	• 스마트 입학 시스템, 위험 학생을 식별하는 도구 및 데이터 기반 의사결정을 내리는 도우미

출처: NITRD, SPRI 정리, 2024.11.

학습자 중심 AI 도구는 학생의 교육 경험을 개선하도록 설계되었다. 이 도구는 맞춤형 학습 경험을 가능하게 하고 학생이 어려움을 겪을 수 있는 과목에 대한 지원을 제공하는 것을 목표로 한다. 이 도구는 적응성, 풍부한 콘텐츠, 학습 지원, 학생 정보 제공 및 조언의 잠재력을 가지고 있다. 그러나 이러한 도구에는 접근성 격차, 기술-능력주의의 위험, 다양한 내재적 편견, 사회-감정적 의미, 개인정보 보호 및 책임 문제 등의 도전과제를 유발한다.

329) OECD, The potential impact of Artificial Intelligence on equity and inclusion in education, 2024.08.

교사 주도 AI 도구는 다양한 기능 전반에 걸쳐 교육 효과와 효율성을 향상시키는 것을 목표로 한다. 여기에는 교실 지원, 학습 자료 큐레이션, 학생 평가 및 교실 관리 지원, 일부 특수교육 요구사항 식별, 새로운 지속적인 전문적 학습 기회가 포함된다. 반면 이러한 도구는 AU 도구의 설치 및 유지보수 관련 비용, 교육의 상업화, AI 사용과 관련된 교사의 활용 능력과 관련된 문제 등의 도전과제를 유발한다.

학습자 중심 및 교사 주도 AI 도구 외에도 교육계는 다양한 다른 기능에 AI를 활용할 수 있다. 즉 AI 도구는 고등교육 입학 절차의 효율성 증가, 교육훈련에서 조기에 이탈할 위험이 있는 학생을 더 잘 식별, 데이터 기반 의사결정 지원의 잠재력이 있다. 반면 이러한 도구는 고유한 AI 편향, 개인정보 보호 및 책임 문제 등의 도전과제를 초래한다.

이 보고서는 교육 환경에 AI 도구의 통합에 대한 기회와 위험을 모두 인정하면서 균형 잡힌 관점 제시를 목표로 하여, 다음과 같은 정책적 시사점을 제시한다.

- 적응형 학습의 잠재력을 수용하는 동시에 개인정보 보호, 윤리 및 책임 문제를 해결
- 고유한 편견을 염두에 두면서 문화적 대응성을 향상할 수 있는 잠재력을 인식
- 기술적 능력주의 및 사회 정서적 기술에 미치는 영향과 같은 과제와 접근성 잠재력의 균형을 맞춤
- 교사 대상의 AI 교육을 개선
- 교육 부문에서 상업적 영향력이 커지는 가운데 교육적 본질적 가치를 유지하는 방법을 탐구
- 교육에서 AI가 형평성과 포용성에 미치는 영향 연구를 장려하고 체계적 실행을 위해 국가 차원에서 교육기관의 역할을 명확히 함

2) 생성 AI의 교육에 대한 기회와 위험³³⁰⁾

유네스코(UNESCO)는 2024년 7월 교육 분야에서 생성 AI의 기회와 위험을 분석한 아티클(Generation AI: Navigating the opportunities and risks of artificial intelligence in education)에서 ChatGPT와 같은 애플리케이션이 인간의 지식, 교육 및 학습에 대한 근본적인 질문을

330) UNESCO, Generation AI: Navigating the opportunities and risks of artificial intelligence in education, 2024.07.22.

제기하는 상황을 맞이하여 교육 분야에 생성 AI 도입 시 준수해야 할 4개 원칙으로 포용성, 형평성, 품질, 접근성을 강조했다.

유네스코는 대부분 AI 애플리케이션은 일부 국가에서 집중적으로 개발되고, 주요 챗봇은 전 세계 7,000여 개 언어 중 영어를 중심으로 약 100개 언어에 대해서만 훈련되었으며, 온라인 고등교육 자료의 90%는 서구권 국가에 집중되어 있다는 점에서 AI가 생성하는 지식 세계는 명확한 한계를 지닌다고 지적했다. 이에 따라, AI의 위험을 방지하고 이익을 실현하기 위해서는 정부와 국제기구 중심의 강력한 규범적 프레임워크를 수립과 함께, 차세대 디지털 시민 양성을 위한 교육체계의 혁신 및 교육 인프라에 대한 민간 부문의 투자가 요구된다고 설명했다.

유네스코는 다양한 지식 체계를 보호하고 확장하며 학습자에게 디지털 시대에 필요한 기술과 역량을 제공하는 방식으로 AI 기술을 개발하기 위해 해결이 필요한 과제들을 제시했다. 첫 번째 과제는 디지털 시대에 적합한 콘텐츠와 커리큘럼을 개발하고 검증하는 것이다. 이는 기존 교과서와 교육 자료의 경우 내용 적합성과 연령 적합성, 교육적 관련성 등을 기준으로 통상 1년 이상의 검증 기간을 거치는 반면, 현재의 AI 도구들은 충분한 검토나 논의 없이 출시되는 상황이기 때문이다. 두 번째 과제는 학습 성과의 평가 체계를 수립하는 것이다. 이를 위해 전 세계 학생들이 과제 작성에 AI를 활용하는 상황에 맞게 새로운 평가 체계 수립이 필요하다. 세 번째 과제로 전 세계적으로 교사가 부족한 상황에서 AI가 교사의 역할에 미치는 영향을 파악해야 하며, 교사의 자질과 학습 성과의 명확한 상관관계를 고려해 교사의 디지털 기술 교육에 투자 필요하다.

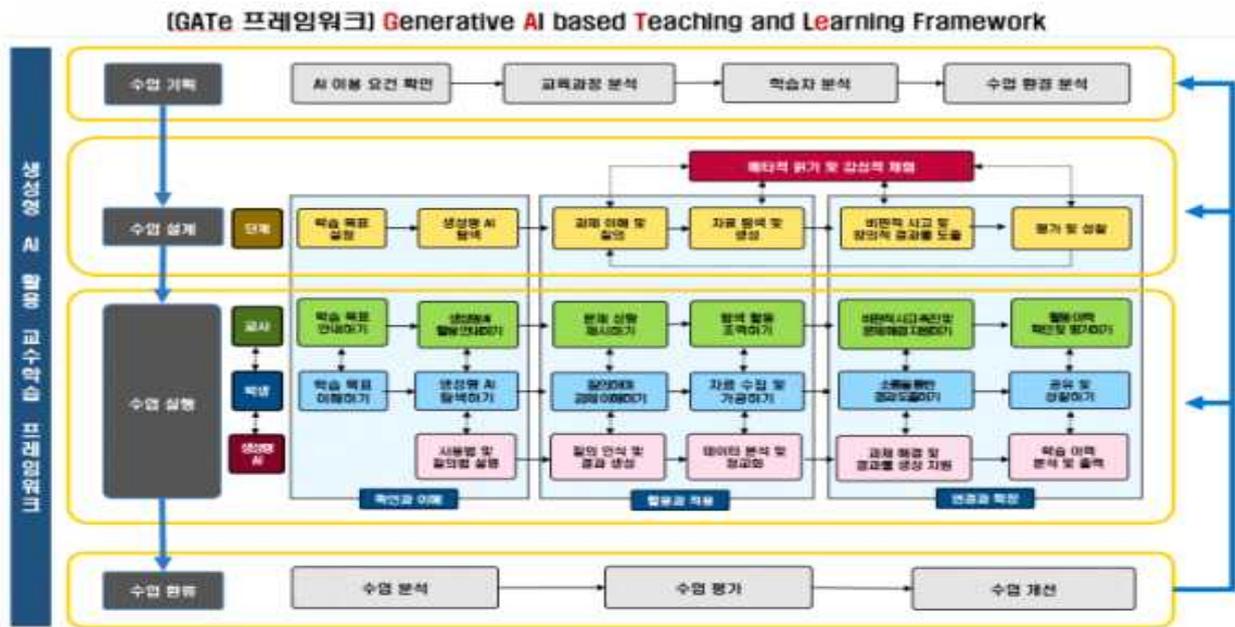
유네스코는 교육과 학습 분야에서 안전하고 적합한 AI 도입을 위한 조치로 아동의 개인정보 보호와 저작권법 개정, 생성 AI 활용에 대한 연령 제한을 권고했다. 특히 아동의 경우 오프라인뿐 아니라 온라인에서도 개인정보가 보호되어야 하며, AI 시대에 맞는 저작권법의 개정이 요구된다. 검증되지 않고 연구가 부족한 AI 도구는 초등 교육에 부적절하며, 13세 미만의 아동이 교실에서 사용할 수 없도록 연령을 제한할 필요가 있다.

3) 생성 AI를 활용한 교수학습³³¹⁾

한국교육학술정보원(KERIS)은 2024년 1월 ‘생성형 AI를 활용한 교수학습 프레임워크 개발 연구’ 보고서를 발행하였다. 이 연구는 2022년 개정 교육 과정에서 도입된 AI 디지털 교과서 등 AI 교육을 준비하며 생성 AI를 활용한 교수학습 프레임워크(이하 GATe 프레임워크)를 개발하는데 목표가 있다.

이 연구에서 개념 모형은 생성 AI 관련 국내외 교육사례, AI 융합 수업 모형, 교과별 선행 연구를 분석한 결과에 기반하여 생성 AI 활용 교수학습 과정과 단계별 구성요소, 및 교과별 교수학습 요소를 포함하도록 설계했다. 최종적인 GATe 프레임워크는 수업 기획, 수업 설계, 수업 실행, 수업 환류의 4단계의 과정으로 구분했다. GATe 프레임워크는 아래(그림)과 같다.

[그림 7-5] 생성 AI를 활용한 교수학습 프레임워크(GATe 프레임워크)



출처: 한국교육학술정보원(KERIS), 생성형 AI를 활용한 교수학습 프레임워크 개발 연구, 2024.5

또한 2024년 5월, 한국교육학술정보원은 생성 AI를 활용한 교수학습 프레임워크 관련 학교 현장에서 활용할 수 있도록 안내서³³²⁾를 개발하였다. 이 운영가이드는 국어과, 영어과, 음악과, 정보과 각각에 대한 GATe 프레임워크를 제시한다.

331) 한국교육학술정보원(KERIS), 생성형 AI를 활용한 교수학습 프레임워크 개발 연구, 2024.1

332) 한국교육학술정보원(KERIS), 생성형 AI를 활용한 교수학습 운영 가이드, 2024.5

제2절 인력 및 고용 분야

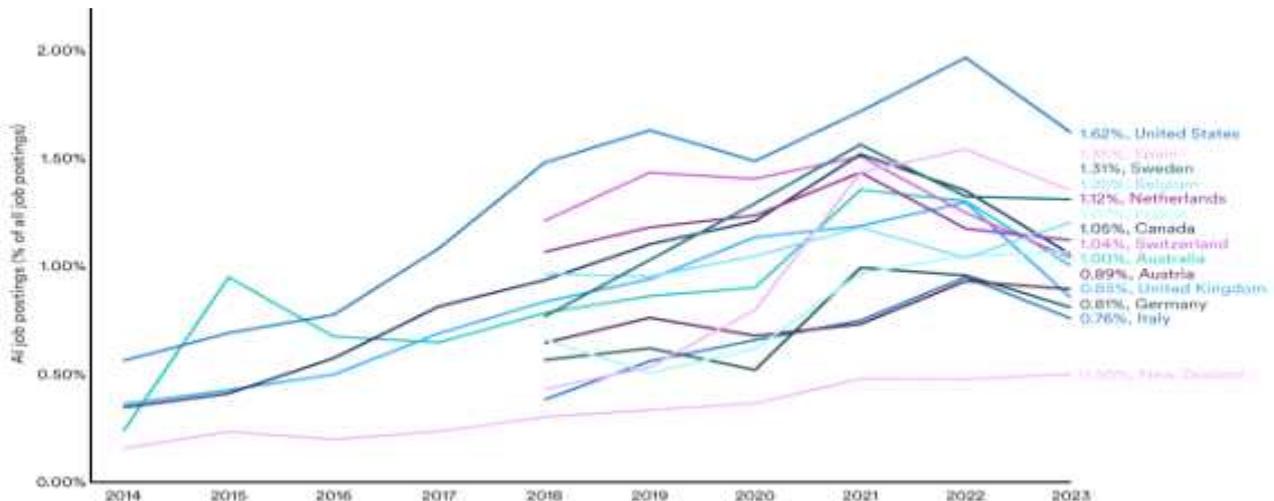
1. 인력 및 고용 현황 및 추이³³³⁾

Stanford HAI가 2025년 5월에 발표한 AI Index 보고서는 경제 부문에서 AI 관련 노동 수요, 채용 동향 등 AI 관련 일자리를 분석하고, AI에 대한 기업의 투자 및 도입 현황, AI로 인해 예상되는 경제적 영향 등을 분석한 결과를 제공한다.

1) 인력 수요

2014년 이후 전 세계적으로 AI 인력 수요가 꾸준히 증가하였으나, 2022년 이후 다소 하락하였다. 2023년의 AI 분야 채용 공고 비율³³⁴⁾이 가장 높은 국가는 미국(1.62%), 스페인(1.35%), 스웨덴(1.31%), 벨기에(1.20%) 순이었다. 2023년 기준, 미국 내 기술(skill)별 채용 공고 비율은 머신러닝(0.68%), 인공지능(0.46%), 자연어처리(0.15%), 자율주행(0.14%), 신경망(0.10%), 시각 이미지 인식(0.07%), 로봇틱스(0.05%), 생성 AI(0.05%) 순이었다. 2023년 기준, 미국 내 산업별 채용 공고 비율은 정보(4.63%), 전문·과학·기술 서비스(3.33%), 금융 및 보험(2.94%), 제조(2.48%), 공공 행정(1.49%), 교육 서비스(1.41%) 순으로 나타났으며, 대부분 산업에서 2022년 대비 공고 개수가 감소했다.

[그림 7-6] 국가별 AI 채용 공고 비율 추이(2014-2023)



자료: Stanford HAI, AI Index Report 2024

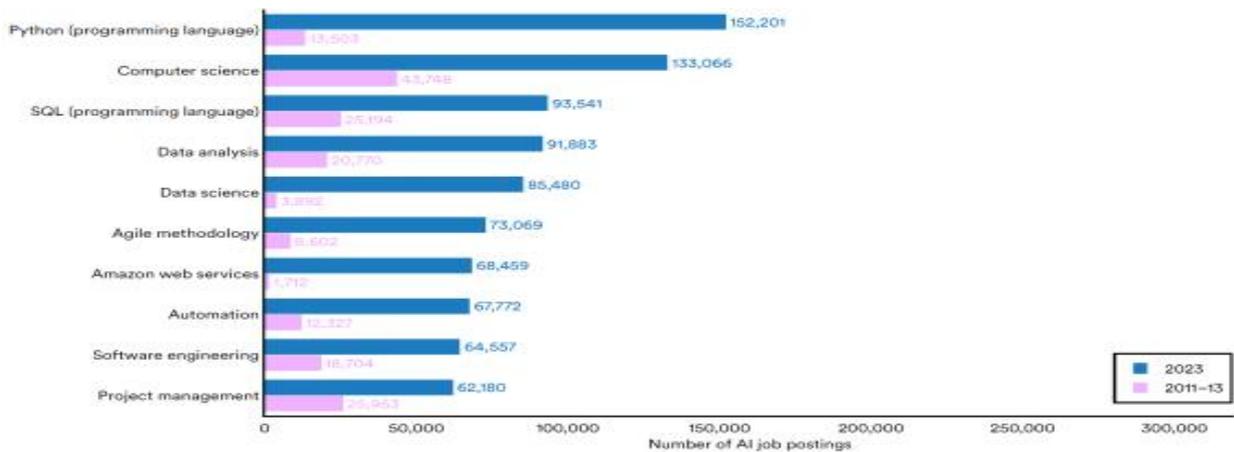
333) Stanford HAI, AI Index Report 2024

334) Stanford HAI는 2010년 이후 전 세계 5만 1,000개 이상의 구인 웹사이트에서 수집한 채용 공고 중 AI 스킬을 요구한 목록을 분석하여 전체 채용에서 AI 스킬을 요구한 공고 비율을 계산

2) 인력 채용

2023년 AI 채용 공고에서 요구되는 상위 10개 전문기술을 2011년~2013년의 전문기술과 비교할 때, 미국의 경우 절대적인 규모로 보면 거의 모든 전문기술에 대한 수요는 지난 10년 동안 증가했으며, 파이썬(Python)이 눈에 띄게 인기가 증가하여 AI 프로그래밍 언어로서 지위를 더욱 확고히 했다.

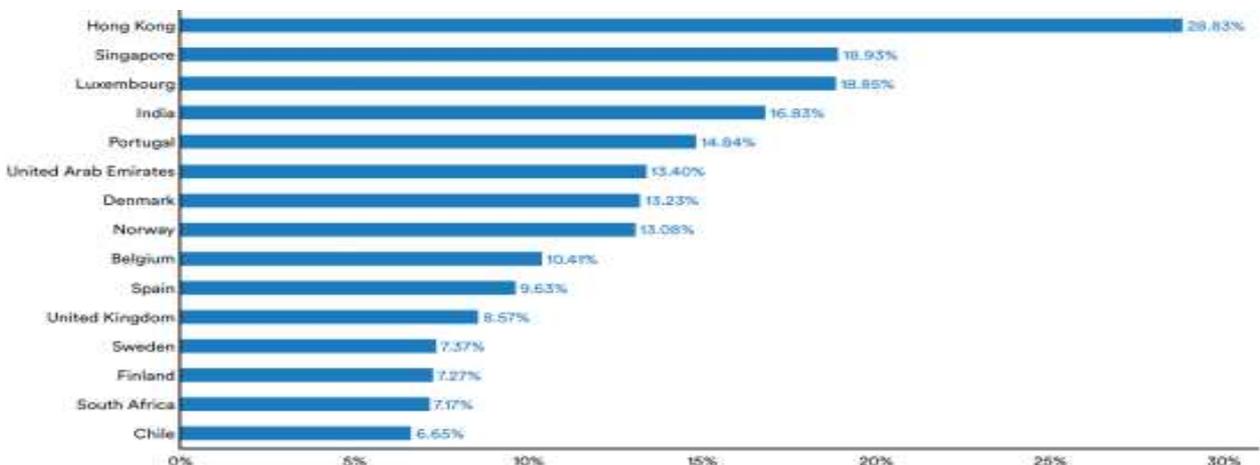
[그림 7-7] 2023년 미국 AI 채용 공고의 상위 10개 전문기술(2011-13년 대 2023년)



자료: Stanford HAI, AI Index Report 2024

2023년 기준, 전년 대비 전체 채용 중 AI 관련 인재 채용률이 가장 높은 지역은 홍콩(28.8%), 싱가포르(18.9%), 룩셈부르크(18.9%) 순이었다. 특히, 2023년 초부터 호주, 캐나다, 싱가포르, 인도 등의 전년 대비 AI 채용이 눈에 띄게 증가하였다.

[그림 7-8] 2023년 지역별 AI 채용률 연간 증가율

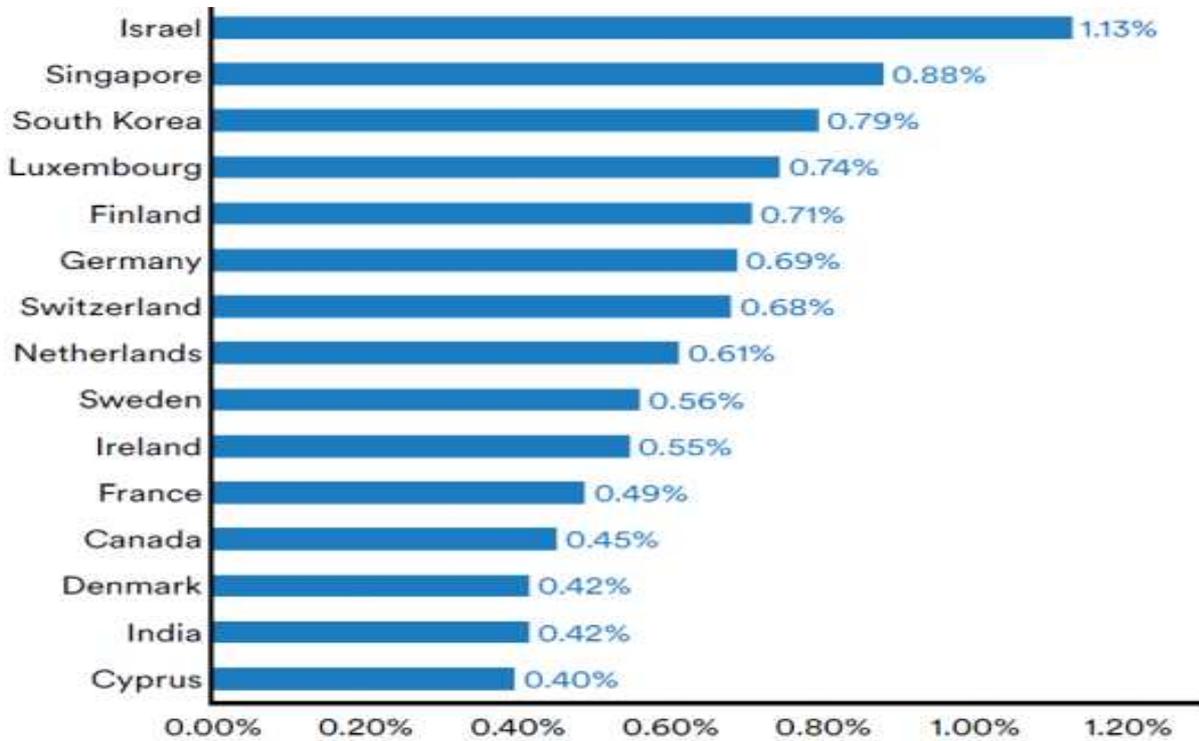


자료: Stanford HAI, AI Index Report 2024

3) 인력 규모

2023년에 AI 인재(AI 분야 스킬 보유자)³³⁵가 가장 많이 집중된 국가는 이스라엘(1.1%), 싱가포르(0.9%), 한국(0.8%)이었다. 2016년 이후 일부 국가의 AI 인재 집중도 변화율을 살펴보면, 여러 주요 경제권에서 AI 인재 풀이 상당히 증가했다. 최대로 증가한 국가는 인도(263%), 키프로스(229%), 덴마크(213%)이었다.

[그림 7-9] 2023년 지역별 AI 인재 집중도



자료: Stanford HAI, AI Index Report 2024

335) LinkedIn 회원이 프로필에 AI 기술을 명시적으로 추가하거나 AI 분야에서 일하는 경우 AI 인재로 간주함 산정함

2. AI의 인력 및 고용에 대한 영향

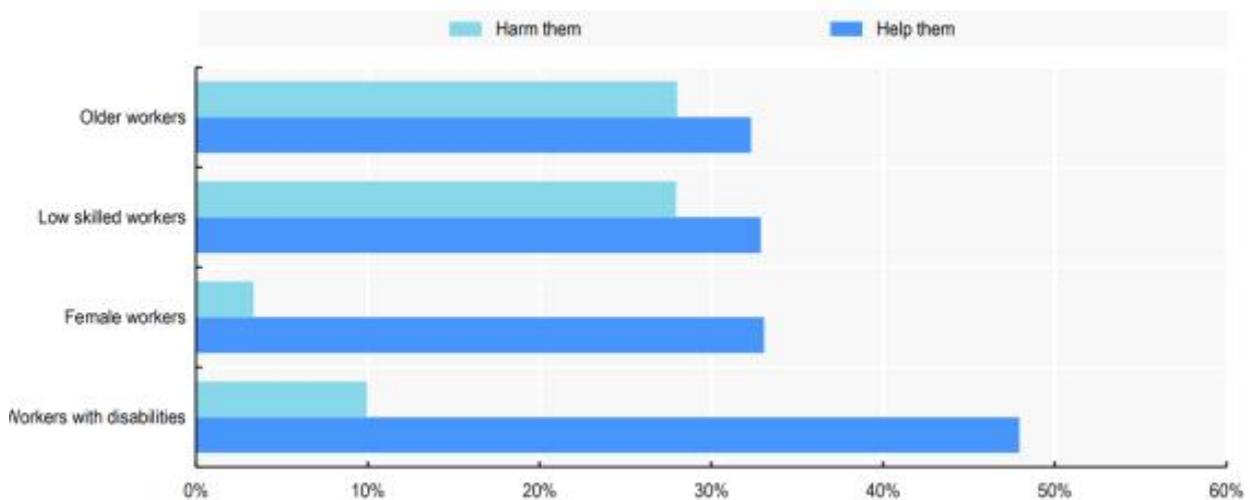
1) AI의 고용 환경에 미치는 영향

OECD는 2024년 AI가 일자리에 미치는 영향에 관한 보고서 3종(아래)을 발간하였다.

(1) AI 이용이 직장 내 생산성과 근무 환경에 미치는 영향

OECD가 2024년 3월 15일 발간한 ‘직장 내 AI 사용(Using AI in the workplace: Opportunities, risks and policy responses)³³⁶⁾’ 보고서에 의하면, 근로자와 고용주 모두 AI가 근로자 생산성과 근무 환경에 미치는 영향을 매우 긍정적으로 인식하였다. AI를 사용한 근로자의 5명 중 4명이 AI로 직장 내 성과가 향상되고, 5명 중 3명은 업무의 즐거움이 늘어났다고 응답했으며, 가장 일반적인 AI 활용사례는 금융 분야의 데이터 분석과 사기 탐지, 제조 분야의 생산 프로세스와 유지관리 작업으로 나타났다.

[그림 7-10] 기업주가 AI가 직장에서 기여 및 위해로 인식하는 비율



자료: Stanford HAI, AI Index Report 2024

그러나 근로자 5명 중 3명은 AI로 인한 일자리 상실을 우려하는 것으로 나타났으며, 이외에도 AI로 인한 임금 감소, 업무강도 증가, 불평등 심화, 개인정보침해, 편견과 차별 등을 우려했다.

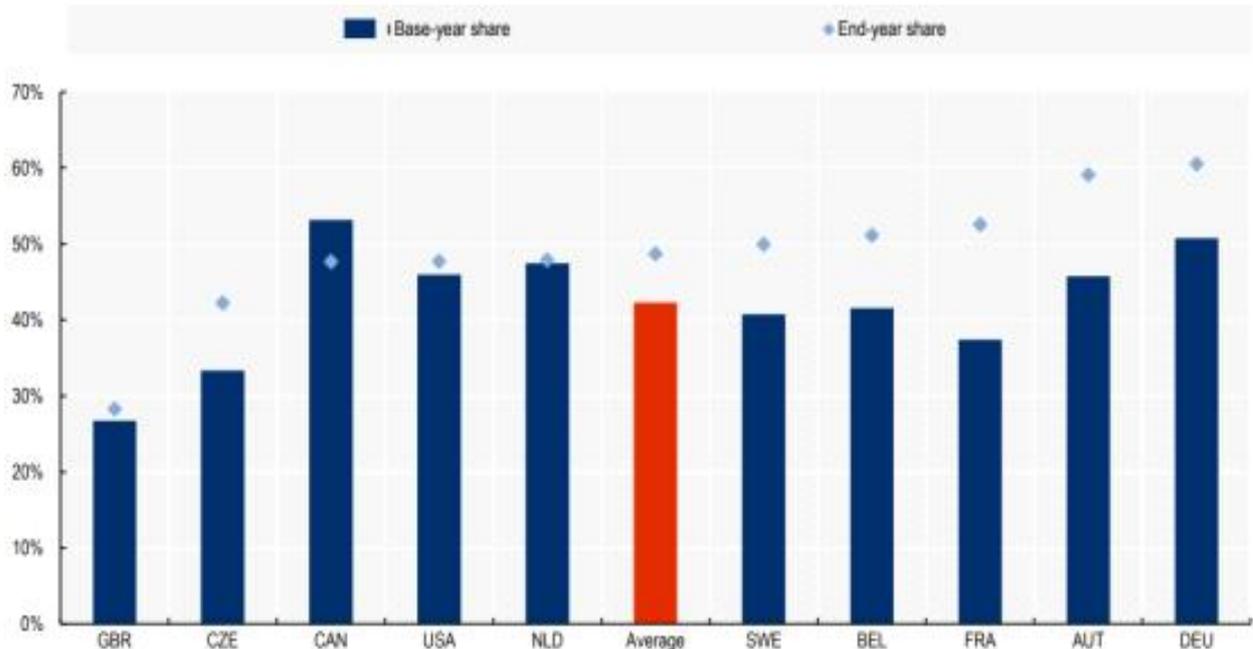
336) OECD, Using AI in the workplace: Opportunities, risks and policy responses, 2024.03.15.

(2) AI의 노동시장 기술 수요에 대한 영향

OECD가 2024년 4월 10일 발간한 ‘AI와 노동시장의 기술 수요 변화(Artificial intelligence and the changing demand for skills in the labour market)³³⁷⁾’ 보고서는 AI 노출의 강도에 따른 기술(skill) 수요의 변화에 대한 추정치를 제공한다.

AI에 노출될 대부분 근로자는 AI를 사용하는 데 전문적인 AI 기술(머신러닝, 자연어처리)이 필요하지는 않다. 그렇더라도 AI는 이러한 근로자가 수행하는 작업과 필요한 기술을 바꿀 것이다. 분석 결과, AI에 크게 노출된 직업에서 가장 많이 요구되는 기술은 관리 및 비즈니스 기술로, 이러한 기술에는 일반 프로젝트 관리, 재무, 행정 및 사무 업무 기술이 포함된다. AI에 크게 노출된 직업에서 이러한 기술들에 대한 수요가 시간이 지남에 따라 증가했다. 예를 들어, 이러한 직업에서 하나 이상의 감정적, 인지적 또는 디지털 기술을 요구하는 일자리가 비어 있는 공석 비율은 8%p(평균 25% → 33%) 증가했다.

[그림 7-11] 국가별 AI 노출인 높은 분야의 기술 수요 변화



출처: OECD, Artificial intelligence and the changing demand for skills in the labour market, 2024.04.10.

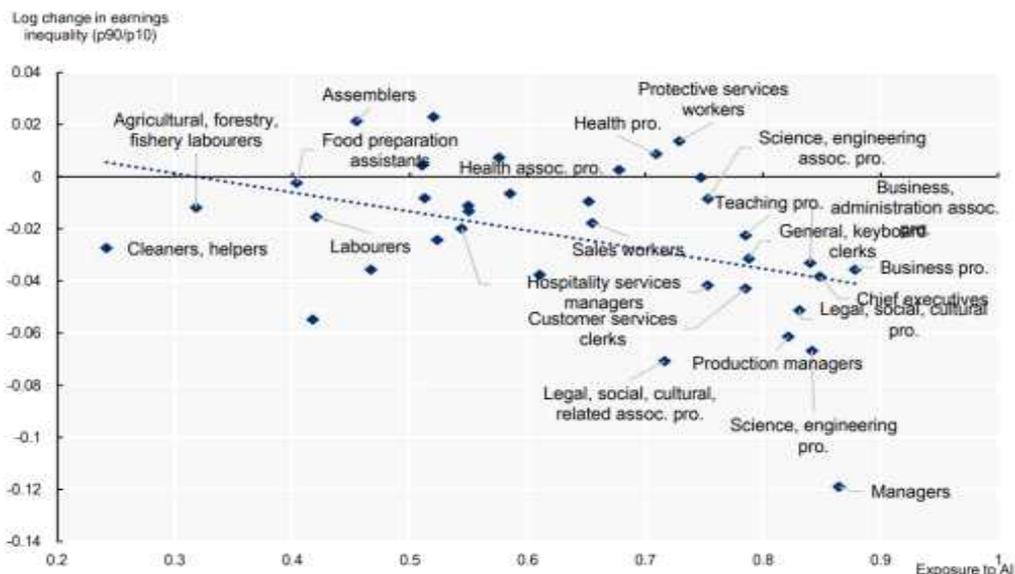
337) OECD, Artificial intelligence and the changing demand for skills in the labour market, 2024.04.10.

(3) AI와 임금 불평등의 연관성

OECD가 2024년 4월 10일 발간한 ‘AI와 임금 불평등(Artificial intelligence and wage inequality)³³⁸⁾’ 보고서에 의하면, AI가 지금까지(2014-18년 기간 동안) 직업 간 임금 불평등에 영향을 미쳤다는 징후는 없는 것으로 분석되었다.

19개 OECD 국가를 대상으로 직업의 AI 노출도를 측정하고, 2014년~ 2018년 기간 특정 직업에 대한 소득 구조 조사와 인구조사 데이터를 활용해 직종 간 및 직종 내 임금 불평등과 AI의 연관성을 분석한 결과, AI 노출이 임금 불평등을 확대하고 있다는 징후는 발견되지 않았으며, 오히려 AI 노출이 높은 직업일수록 직종 내 임금 불평등이 감소했으며, 이는 AI가 근로자 간 생산성 차이를 감소시킨다는 문헌 조사 결과와 일치하였다.

[그림 7-12] 직업 내 AI 노출과 직업 내 임금 불평등



출처: OECD, Artificial intelligence and the changing demand for skills in the labour market, 2024.04.10.

2) 생성 AI의 전세계 일자리에 미치는 영향

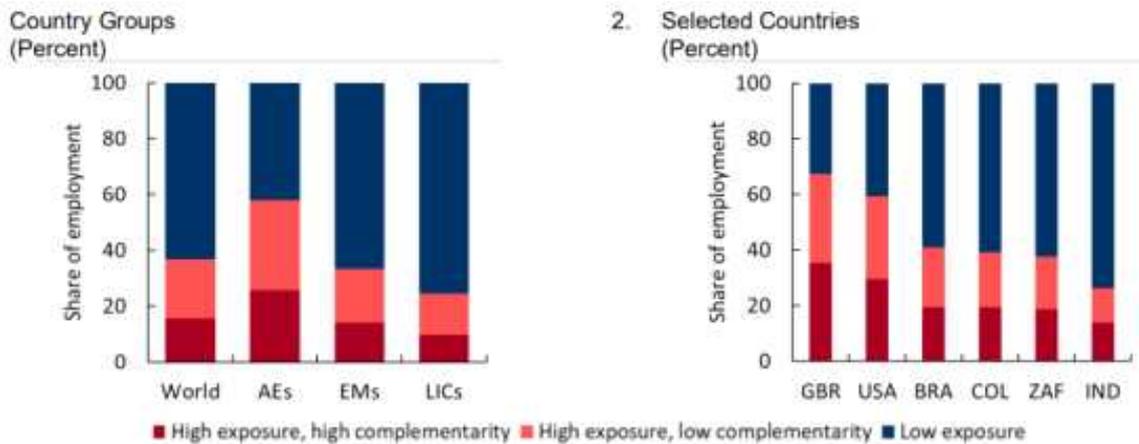
IMF가 2024년 1월 14일 발표한 ‘생성 AI: AI와 노동의 미래(Gen-AI: Artificial Intelligence and the Future of Work)³³⁹⁾’ 보고서에 따르면 AI는 전 세계 일자리의 40%에 영향을 미치며, 선진국

338) OECD, Artificial intelligence and wage inequality, 2024.04.10.

339) IMF, Gen-AI: Artificial Intelligence and the Future of Work, 2024.01.14.

에서는 일자리의 60%에 영향을 미칠 전망이다. AI에 노출된 일자리의 절반은 AI로 인해 생산성이 향상될 수 있으며, 나머지 절반은 AI가 인간의 업무를 대체해 노동 수요를 낮추어 채용이 줄어들거나 극단적으로는 직업 중 일부가 사라질 전망이다. 신흥국과 저소득 국가의 경우 AI가 영향을 미치는 일자리의 비중은 각각 40%와 26%로, 이는 AI가 국가 간 불평등을 악화시킬 위험을 시사한다.

[그림 7-13] AI 노출 및 보완성에 따른 고용 점유율(지역별/국가별)



*AE = 선진 경제, EM = 신흥 시장, LIC = 저소득 국가

출처: IMF, Gen-AI: Artificial Intelligence and the Future of Work, 2024.01.14.

AI는 국가 내부의 소득과 부의 불평등에도 영향을 미칠 수 있으며, 전반적으로 불평등을 악화시킬 가능성이 높아 정부는 AI로 인한 사회적 긴장을 줄이기 위해 포괄적인 사회 안전망을 구축하고 취약한 근로자에게 재교육 프로그램을 제공할 필요가 있다. AI를 활용할 수 있는 근로자는 생산성과 임금이 증가하며, 특히 AI는 경력이 짧은 직원의 생산성 향상에 효과적이거나, 고령의 근로자는 AI에 적응하는 데 어려움을 겪을 수 있다. AI가 근로 소득에 미치는 영향은 AI가 고소득 근로자를 보완하는 정도에 달려 있으며, 고소득 근로자 보완 정도가 크면 근로 소득이 불균형적으로 증가해 불평등이 악화될 수 있다.

한편, IMF는 보고서에서 ▲디지털 인프라 ▲노동 정책 ▲혁신·경제 융합도 ▲규제·윤리의 네 가지 항목을 기준으로 글로벌 국가들의 ‘AI 준비 지수’를 개발하였다. 조사 결과 선진국과 일부 신흥국을 포함한 고소득 국가들은 저소득국보다 AI 도입 준비가 잘되어 있으나 국가 간 상당한 차이를 보였으며, 싱가포르, 미국, 덴마크는 4개 항목 전체에서 가장 높은 점수를 기록하였다. IMF는 선진국에 AI 혁신의 통합과 강력한 규제 체계 마련에 우선

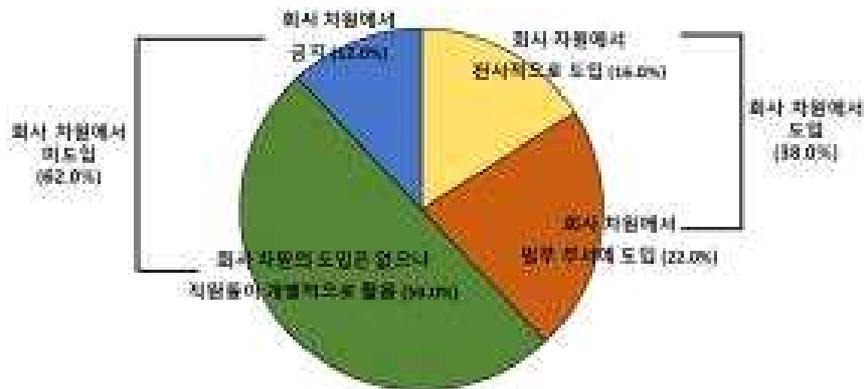
순위를 두어야 하며, 신흥국과 저소득국은 디지털 인프라와 디지털 역량을 갖춘 인력에 대한 투자로 견고한 기반을 마련해야 한다고 진단하였다.

3) 국내 주요기업의 생성 AI에 따른 일자리 영향

한국경영자총협회(이하 경총)가 2024년 3월 21일 발표한 ‘주요 기업 AI 도입 실태 및 인식 조사’³⁴⁰⁾에 따르면 응답 기업의 38%가 챗GPT 같은 생성 AI를 회사 차원에서 사무직군에 도입하였다. 조사결과, AI를 도입(예정 포함)한 기업의 85.7%는 AI 활용이 업무 소요시간을 줄인다고 응답했으며, 기업 내 일자리에 미치는 영향에 대해서는 “큰 변화 없을 것” 이란 응답이 75.0%로 가장 많았다.

생성 AI를 도입하지 않은 기업은 62%로 나타났으며, 현재 AI를 도입하지 않은 기업 중 29%는 향후 AI를 도입할 예정이라고 응답했다. 회사 차원의 도입은 없으나 직원들이 개별적으로 활용한다는 응답이 50%로 가장 많았으며, 회사 차원에서 활용을 금지했다는 응답도 12%로 조사되었다. AI를 도입하지 않은 기업들은 그 이유로 정보 유출 우려(41.9%), 준비 기간 필요(29.0%), 업무 특성상 불필요(16.1%) 순으로 응답했다. AI를 도입(예정 포함)한 기업들은 주로 AI로 인하여 일자리 대체 효과가 발생하겠지만 전문인력 확보, 업무영역 확대 등으로 신규 일자리가 늘면서 전체 일자리는 큰 변화가 없을 것(75.0%)으로 예상했다. AI 도입·확대로 인한 일자리 대체 효과가 신규 일자리 창출보다 더 크게 나타나 “전체 일자리가 축소될 것” 이란 응답은 17.9%였다.

[그림 7-14] 주요 기업의 AI 도입 현황



한국경영자총협회, 경총, 「주요 기업 AI 도입 실태 및 인식 조사」 결과 발표, 2024.03.21.

340) 이 조사는 매출액 상위 100대 기업(공기업 제외, '22년 기준) 및 경총 주요 회원사 대상으로 설문 조사 방식으로 진행되었으며, 응답 기업은 50개 기업이다.

제8장 결론

본 결과보고서는 2024년 동안 조사된 AI 산업의 주요 동향과 데이터를 종합적으로 분석하여, AI 산업의 전반적인 흐름과 방향성을 제시하였다. 특히, 주요국, 국제기구, 글로벌 기업, 학술단체, 표준기관 등 다양한 주체의 정책과 전략을 다각도로 분석함으로써, 각국의 정책 배경과 의사결정 과정에 대한 심도 있는 이해를 도모하였다.

2024년 AI 산업 동향 조사 결과를 기반으로 주요 이슈와 향후 전망을 분석하고, 이를 바탕으로 AI 산업의 지속 가능하고 효과적인 발전을 위한 정책적 방향을 제안한다.

1. 주요 이슈와 향후 전망

1) 생성 AI와 초거대 AI 모델의 확산

생성 AI 기술과 초거대 AI 모델(오픈AI의 GPT, 구글의 Gemini, 메타의 LLaMA 등)이 다양한 산업에서 빠르게 확산되고 있다. 이러한 모델들은 자연어 처리, 이미지 생성, 추천 시스템 등에서 뛰어난 성능을 발휘하며, 기존 기술의 한계를 넘어선 응용 가능성을 보여주고 있다. 특히, 마이크로소프트와 메타, 구글과 같은 빅테크 기업들은 초거대 AI 모델을 기반으로 맞춤형 서비스를 제공하며 시장 점유율을 확대하고 있다. 2024년에서 초거대 AI 모델의 상업적 활용이 의료(진단 보조), 교육(개인화 학습 플랫폼), 금융(위험 관리와 예측 분석) 등 핵심 분야에서 주목받았다.

2025년 이후, 생성 AI와 초거대 AI 모델은 더욱 다양한 산업에서 필수 도구로 자리 잡을 것으로 전망된다. 의료 분야에서는 초거대 AI 모델을 활용한 정밀 의료와 진단 시스템이 확대될 것으로 예상된다. 예를 들어, 환자의 의료 기록과 실시간 데이터를 분석해 맞춤형 치료 계획을 제공하는 솔루션이 일반화될 가능성이 크다. 교육에서는 학생 개인의 학습 데이터를 분석하여 최적화된 학습 경로를 설계하는 AI 기반 교육 플랫폼이 보편화될 것이다. 금융에서는 시장 데이터를 실시간으로 분석하고 예측해 리스크 관리와 투자 전략을 지원하는 AI 솔루션이 더욱 정교해질 전망이다. 특히 전 세계 AI 에이전트 시장이 2024년 51억 달러에서 2030년 471억 달러로 빠르게 성장할 것으로 예상되면서 AI 생태계에서의 핵심 기술로 자리 잡아갈 전망이다.³⁴¹⁾

2) AI 스타트업 투자와 경쟁의 심화

AI 중심 스타트업에 대한 투자가 급증하고 있으며, 자율 시스템, 생성 AI 등 혁신적인 분야가 주요 투자 대상이 되고 있다. CB인사이트 분석에 따르면 2024년 2분기 전 세계 AI 투자 규모는 전 분기 대비 59% 증가한 232억 달러로 대폭 증가하였다. 특히 생성 AI 분야에서는 이미지 생성, 텍스트 작성, 영상 제작 등의 기술을 기반으로 한 스타트업들이 시장에서 빠르게 성장하고 있다. 예를 들어, 2024년 11월 기준 생성 AI 스타트업 Anthropic은 130억 달러 이상의 투자를 유치했으며³⁴²⁾, 한국의 스캐터랩은 400억 원 이상을 유치하고, 감성 대화형 AI 챗봇으로 주목받고 있다. 이러한 흐름은 글로벌 기술 기업들과 벤처캐피털의 AI 기술 및 스타트업에 대한 신뢰와 기대를 반영한 결과다.³⁴³⁾

2025년 이후, AI 스타트업에 대한 투자는 더욱 가속화되며 경쟁이 심화될 것으로 예상된다. 특히, 생성 AI를 활용한 맞춤형 솔루션, 자율주행 시스템, 스마트 제조 등 구체적 활용 분야를 중심으로 새로운 시장이 창출될 전망이다. 이러한 스타트업들은 단순한 기술 제공자가 아닌, 대기업과 협력하여 맞춤형 솔루션을 제공하는 파트너로 자리 잡을 가능성이 높다. 특히 기존 산업의 문제를 해결하거나, 새로운 사용자 경험을 창출하는 스타트업의 경우 빠르게 투자 유치와 시장 확대에 성공하고 있으며, AI를 도입하고자 하는 기업 입장에서도 특정 목적에 최적화된 스타트업과의 협력이 더욱 매력적인 전략으로 부상하고 있다.

전 세계 AI 유니콘 기업의 수는 2024년 3월 기준 200개 이상이며, 2025년에는 더욱 증가할 것으로 전망된다.³⁴⁴⁾ 이는 중소기업의 스타트업들이 혁신적인 기술과 사업 모델로 글로벌 투자 시장에서 두각을 나타낼 가능성을 시사한다. 또한, 이러한 기업들은 자체적인 성장뿐만 아니라 대기업과의 협업을 통해 더 큰 시너지 효과를 창출할 것이다. 예를 들어, 마이크로소프트와 오픈AI의 협력과 같이 글로벌 빅테크는 초기 단계 스타트업에 대한 투자와 함께 기술 플랫폼을 제공하여 상호 성장을 도모할 가능성이 높다. AI 스타트업은 향후 AI 생태계의 중요한 축으로 자리 잡으며, 대규모 투자와 치열한 경쟁 속에서 새로운 시장과 직업군을 창출하고 기술혁신의 중심에서 지속적으로 발전할 것이다.

341) https://www.marketsandmarkets.com/Market-Reports/ai-agents-market-15761548.html?gad_source=1&gclid=CjwKCAiA9IC6BhA3EiwAsbltOJG4uzoGmxUSg_zloWLHlNkrjclXog2RCwT8yUBRvGIPmM1Ximfa1RoCTnAQAvD_BwE

342) https://www.crunchbase.com/organization/anthropic/company_financials

343) <https://thevc.kr/scatterlab>

344) <https://edgedelta.com/company/blog/ai-startup-statistics#:~:text=2018%20through%202020.-,From%20Vision%20to%20Reality:%20Success%20Rate%20of%20AI%20Startups%20in,success%20rates%20of%20AI%20startups.>

3) AI의 전통 산업 통합 가속화

AI 기술이 제조, 물류, 유통 등 전통 산업에 빠르게 통합되고 있다. 제조업에서는 품질 관리, 생산 공정 최적화, 예측 유지보수를 위한 AI 기술이 도입되며, 스마트 공장의 구현을 가속화하고 있다. 물류 산업에서는 배송 경로 최적화, 재고 관리, 수요 예측 등을 위한 AI 기반 솔루션이 표준화되고 있으며, 아마존과 같은 기업은 AI를 활용한 로봇 자동화 시스템을 운영하고 있다. 유통 산업에서는 고객 행동 분석과 개인화된 추천 시스템이 매출 증대와 소비자 경험 개선에 기여하고 있다.

2025년 이후, AI 기술은 전통산업의 필수 도구로 자리 잡으며, 자동화와 최적화를 넘어 인간과 AI의 협업이 본격화될 것으로 보인다. 제조업에서는 AI와 IoT 기술의 결합으로 실시간 데이터 분석과 예측을 통한 생산성 극대화가 일반화될 전망이다. 예를 들어, 스마트 공장에서 AI는 작업자의 생산성을 지원하는 협업 로봇(Cobot)과 함께 활용되며, 생산 공정에서의 오류를 실시간으로 탐지하고 수정할 것이다.

AI의 통합은 인간과 AI의 협업을 중심으로 한 산업 모델을 촉진할 것이다. 5차 산업혁명(Industry 5.0)³⁴⁵⁾ 시대에는 인간의 창의성과 AI의 분석·예측 능력이 결합된 새로운 작업환경이 구현될 것으로 보인다. 이러한 변화는 전통산업의 생산성을 획기적으로 높이는 동시에, 기존 업무 구조를 재편하고 새로운 직업군을 창출할 것이다.

4) AI 윤리, 거버넌스 및 안전성 강화

AI 기술의 급격한 발전과 확산은 윤리적 문제, 거버넌스 체계 부재, 안전성에 대한 우려를 심화시키고 있다. 프라이버시 침해, 알고리즘 편향, 기술 오용은 주요 과제로 떠오르고 있으며, 특히 생성 AI는 가짜 뉴스 생성, 데이터 유출, 허위 정보 확산에 대한 우려를 증폭시키고 있다. 예를 들어, 미국에서는 생성 AI를 활용한 가짜 뉴스 사례가 선거와 사회적 갈등에 영향을 미친 사례가 보고되었다. 또한, AI 알고리즘의 편향으로 인해 의료나 채용 과정에서 특정 집단에 대한 불공정한 결과가 발생한 사례도 늘어나고 있다.

345) https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/industry-50_en

이와 같은 문제를 해결하기 위해 국제기구와 각국 정부는 AI 윤리와 안전성 확보를 위한 법적·제도적 프레임워크를 마련하고 있다. EU는 2023년 AI법(AI Act)을 통해 윤리적 기준과 안전 가이드라인을 명확히 했으며, 우리나라를 비롯하여 미국, 일본, 캐나다는 AI 안전연구소를 설립하여 기술적·사회적 대응 방안을 모색하고 있다. 특히 2024년 제정 중인 AI 기본법에 딥페이크 생성물은 가시적 워터마크를 필수적으로 삽입하도록 규정하고 있다.

2025년 이후, AI 기술의 윤리적 활용을 보장하기 위한 글로벌 규제와 통합된 표준이 본격적으로 시행될 전망이다. AI 거버넌스는 공공과 민간 부문의 협력을 중심으로 성숙될 것이다. 정부는 정책과 규제를 제공하고, 기업은 알고리즘 투명성과 데이터 관리 정책을 준수함으로써 상호 신뢰를 구축할 것이다. 기술적 측면에서는 편향을 줄이고 공정성을 강화하기 위한 AI 알고리즘의 투명화와 검증이 일반화될 것이다. AI 개발 초기 단계부터 윤리적 기준을 반영하는 “윤리 중심 설계(Ethics by Design)” 접근법이 확산될 것으로 보인다.

5) AI 연구와 학계의 발전

학계서도 AI 기술이 주목 받으면서 관련 연구가 활발히 진행되고 있다. 특히 설명 가능한 AI(Explainable AI)와 지속 가능 AI(Sustainable AI)³⁴⁶⁾ 기술이 주요 연구 분야로 부상하고 있다. 설명 가능한 AI는 모델의 의사결정을 투명하게 만들어 의료, 금융 등 민감한 영역에서의 신뢰를 높이는 데 중요한 역할을 한다. 지속 가능 AI는 AI 시스템의 에너지 소비를 최소화하고, 탄소 배출량을 줄이는 기술을 포함한다. 구글은 AI 훈련 과정에서 탄소 배출을 줄이는 데이터센터 운영 기술을 연구 중이며, 이는 글로벌 연구 프로젝트와 연계되어 있다.

이와 함께 학문 간 융합 연구가 활발히 진행되며, AI와 공학, 사회과학, 생물학 등 다양한 학문 영역이 결합해 새로운 지식을 창출하고 있다. 세계 주요 대학들은 AI 연구소를 설립하거나 학제 간 협력을 강화하며 AI 연구의 방향성을 이끌고 있다.

2025년 이후, 대학과 연구기관은 AI 기술 발전의 핵심 동력으로 자리 잡을 것이다. 학계는 AI의 사회적 책임과 지속 가능성을 주제로 한 연구를 확대하며, 기술 개발뿐 아니라 정책적, 윤리적 기반을 제공할 것으로 기대된다. 특히 설명 가능한 AI 연구는 공공 분야에서 더욱 강조될 것이며,

346) <https://pg-p.ctme.caltech.edu/blog/ai-ml/what-is-sustainable-ai-significance-examples>

의료, 법률, 교육 등에서 투명성과 신뢰를 높이는 데 기여할 것이다. 예를 들어, 정부와 공공기관은 AI 모델의 투명성과 결과 해석 가능성을 요구하며 이를 정책적 기준으로 삼을 가능성이 크다.

지속 가능 AI 분야에서는 AI 시스템의 에너지 효율성을 높이고, 재생 가능 에너지를 활용한 데이터센터 운영이 일반화될 것이다. 유럽연합은 그린 AI 노력을 통해 AI 기술이 환경에 미치는 영향을 최소화하기 위한 연구와 정책적 지원을 강화할 것으로 예상된다. 이는 글로벌 협력을 통해 AI 기술이 기후 변화 문제 해결에 적극 활용될 기반을 마련할 것이다.

6) 글로벌 AI 협력과 표준화

국제기구와 주요국들은 AI 기술의 안전성과 신뢰성을 확보하기 위해 표준화 노력을 강화하고 있다. ISO(국제표준화기구)는 AI 윤리와 안전성에 대한 표준 개발을 주도하고 있으며, UN은 ‘인류를 위한 AI 거버넌스’ 보고서를 통해 AI의 책임 있는 활용과 협력의 중요성을 강조했다. 이러한 표준화 작업은 AI 기술의 상호운용성과 데이터 호환성을 보장하고, 글로벌 AI 생태계에서의 신뢰성을 높이는 데 중점을 두고 있다.

한편, 각국 정부와 기업은 AI 기술이 초국경적으로 활용될 수 있도록 협력체계를 구축하고 있다. 예를 들어, EU는 AI법(AI Act)을 통해 국제적으로 조율 가능한 기준을 제안하며, 이를 기반으로 글로벌 규제 환경을 형성하려는 노력을 이어가고 있다. 일본은 2024년 49개 국가들로 구성된 ‘히로시마 AI 프로세스 프렌즈 그룹’의 출범시키고 AI 국제 지침과 행동 규범의 이행을 지원하고 국가 간 협력을 촉진하고 있다.

2025년 이후, AI 기술의 글로벌 표준화는 더욱 구체화될 것으로 보인다. ISO와 UN을 중심으로 AI의 윤리적 설계와 활용을 보장하는 국제표준 개발이 활발해지고, 이러한 표준들은 전 세계적으로 법적·제도적 기반으로 작용할 것이다. 특히, AI의 데이터 처리와 모델 훈련에서 발생하는 편향과 투명성 문제를 해결하기 위해 표준화가 진행될 전망이다. 나아가, 국제표준은 단순한 권고 수준을 넘어 정부의 AI 인증제도 및 공공 조달 요건 등과 직접 연동되며, 민간 기업의 기술 개발과 서비스 설계에도 실질적인 영향을 미치게 될 가능성이 크다. 이에 따라 글로벌 시장 진출, 다국적 기업과 협력을 계획하는 기업들은 이러한 국제표준을 충족하기 위한 기술적 대응 및 컴플라이언스 체계 구축이 필수적으로 요구될 것이다.

7) 인력 전환과 AI 교육의 중요성

AI 기술의 발전은 기존 직업군에 변화를 가져오고 새로운 직업군을 창출하며, 노동 시장 전반에 구조적인 전환을 일으키고 있다. 제조, 물류, 금융 등 주요 산업에서는 반복적이고 예측 가능한 업무가 AI로 대체되고 있으며, 이는 자동화로 인한 고용 감소 우려를 증폭시키고 있다. 그러나 동시에, 데이터 분석, 알고리즘 개발, AI 기반 시스템 유지관리 등 AI와 관련된 새로운 직업군이 생겨나고 있다.

2024년 1월 IMF가 발표한 ‘생성 AI: AI와 노동의 미래(Gen-AI: Artificial Intelligence and the Future of Work)’ 보고서에 따르면 AI는 전 세계 일자리의 40%에 영향을 미치며, 선진국에서는 일자리의 60%에 영향을 미칠 전망이다. 이로 인해 많은 기업들이 기존 인력을 재교육(Reskilling)하고 새로운 기술을 습득하도록 돕는 프로그램에 투자하고 있다.

2025년 이후, AI 관련 교육 프로그램이 전 세계적으로 크게 확대될 것으로 보인다. 정부와 민간 기업 간 협력이 강화되며, 기존 노동자를 위한 재교육 프로그램과 AI 기술을 배우려는 신입 인력을 위한 맞춤형 교육 과정이 등장할 것이다. 교육의 초점은 AI 기술의 이해와 활용을 넘어서 AI와 인간의 협업 능력을 키우는 방향으로 이동할 것이다. 특히, 초중등 교육 과정에서도 AI 기초 교육을 포함해 다음 세대를 준비시키는 데 중점을 둘 전망이다. 우리나라는 2025년부터 모든 초중고에 AI 디지털 교과서과 도입되고, 대학원 수준에서 AI 심화 연구를 위한 국가적 지원이 강화될 예정이다.

AI는 노동 시장에서의 대체와 창출이라는 이중적인 변화를 이끌고 있다. 이에 대응하기 위해 정부와 민간이 협력하여 AI 교육과 평생 학습 시스템을 확립하고, 모든 계층의 노동자가 새로운 AI 중심 경제에서 성공적으로 적응할 수 있도록 지원하는 것이 필수적이다.

이외에도 단순히 기술 교육에 국한되지 않고, 윤리적 판단력, 인간 중심의 의사소통 역량, AI와 협력하는 문제 해결력 등을 포괄하는 ‘AI 리터러시(AI literacy)’ 교육이 병행될 필요가 있다. 이는 기술 중심의 변화 속에서도 인간 고유의 역할을 강화하고, 사회적 신뢰와 포용을 높이기 위한 핵심 조건이 될 것이다.

2. 정책제언

1) 초거대 AI와 생성 AI 기술 활용 촉진 및 규제 환경 개선

2024년 국내외 인공지능 산업 동향에 따르면, 초거대 AI와 생성 AI는 다양한 산업에서 핵심 기술로 자리 잡고 있으며, 이들의 발전과 활용을 위해 체계적이고 지속적인 정책 지원이 필요하다. 우선, 의료, 교육, 공공행정 등 사회적 가치가 높은 분야에서 초거대 AI 기반 서비스의 실증 및 확산 프로젝트를 강화해야 한다. 이러한 프로젝트를 통해 공공기관의 데이터 활용을 적극 지원하고, 공공 도입 시 윤리적·사회적 기준을 충족하는 AI 모델을 선정하도록 유도할 필요가 있다. 또한, 공공 부문과 민간 기업 간 협력 모델을 구축하여 데이터 공유를 활성화하고, 공공 데이터 활용을 기반으로 혁신적인 AI 서비스를 개발할 수 있도록 제도적 지원을 확대해야 한다.

다음으로는 생성 AI 기술의 규제와 혁신의 균형을 맞추기 위한 노력이 필요하다. 이를 위해 생성 AI의 안전성 검증 체계를 국제적 인증 기준에 맞추어 개발하여 기술의 신뢰성을 확보하고, 동시에 기술 혁신을 저해하지 않는 자율적 규제 가이드라인을 마련해야 한다. EU의 AI법(AI Act) 및 NIST AI 위험관리 프레임워크 등 글로벌 AI 신뢰성 기준과 연계하여 국내 인증 체계를 정비하고, 국내 AI 기업이 국제 시장에서 경쟁력을 확보할 수 있도록 정책적 지원이 필요하다. 이를 통해 생성형 AI의 윤리적 활용을 보장하는 한편, 기술 발전을 촉진하는 환경을 조성할 수 있을 것이다.

특히, 이러한 규제와 진흥의 균형 있는 접근을 제도화하기 위해서는 국가 차원의 'AI 기본법' 제정이 시급하다. AI 기본법은 민간의 혁신을 유도하고 기술 실증 및 확산을 장려하면서도, 안전성, 윤리성, 개인정보 보호 등 핵심 원칙을 명확히 할 수 있는 법적 기반이 되어야 한다. 규제 중심의 일방향적 통제가 아닌, 유연하고 예측 가능한 법제도 환경을 마련함으로써, AI 기술의 책임 있는 활용과 지속 가능한 발전을 동시에 달성할 수 있을 것이다.

또한, AI 활용 인프라를 더욱 강화하고, 중소기업과 스타트업이 초거대 AI를 효과적으로 활용할 수 있도록 지원을 확대해야 한다. 이를 위해 클라우드 기반 AI 플랫폼 접근성을 높이고, 오픈소스 AI 모델 활용을 촉진하며, 대규모 컴퓨팅 자원이 필요하지 않은 경량 AI 모델 개발을 지원할 필요가 있다. 아울러, 데이터센터 사용료 지원 및 클라우드 컴퓨팅 비용 부담을 완화하는 정책을 도입하여 중소기업이 AI 기술을 보다 쉽게 도입하고 활용할 수 있

도록 해야 한다. 이러한 정책적 노력을 통해 AI 산업 전반의 성장 기반을 마련하고, 혁신적인 AI 기술의 사회적 활용을 극대화할 수 있을 것이다.

2) AI 윤리 및 글로벌 규범 수립 대응

AI 기술의 발전과 확산에 따라 사회적 책임과 신뢰성을 확보하기 위한 윤리적 프레임워크와 국제적 협력 대응의 중요성이 점점 더 부각되고 있다. 현재 국회에서 진행 중인 AI 기본법 제정은 AI의 윤리적 활용을 보장하는 제도적 기반을 마련하기 위한 중요한 단계로 평가되며, 이에 발맞춰 AI 윤리 기준을 보다 구체적으로 정립하고 기업들이 이를 실질적으로 준수할 수 있도록 실효성 있는 가이드라인을 마련해야 한다. OECD AI 프레임워크, G7 히로시마 프로세스 등 국제조직의 AI 윤리나 규범에 대한 논의를 고찰하면서 AI 윤리 기준을 구체화하고 기업들이 이를 준수할 수 있도록 실효성 있는 가이드라인을 마련해야 한다. 생성 AI의 활용 과정에서 발생할 수 있는 편향성, 허위정보 생성, 저작권 침해 등의 윤리적 문제와 기술 오용을 방지하기 위해 체계적인 교육 및 인증 시스템을 도입하고 이를 확산하는 정책적 노력이 요구된다.

국제 AI 규범 동향에 능동적으로 대응하는 것도 필수적이다. EU AI법(AI Act), G7 히로시마 프로세스, OECD의 AI 정책 권고안 등 주요 국제 규범 형성 움직임을 지속적으로 모니터링하고, 이러한 논의에 적극 참여하여 국내 정책이 글로벌 기준과 조화를 이루도록 해야 한다. 이를 통해 한국의 AI 정책이 국제적 신뢰성을 확보하고, 글로벌 AI 협력체계에서 주도적인 역할을 할 수 있도록 해야 한다. 특히, UN이 추진하는 ‘인류를 위한 AI 거버넌스’ 논의와 연계하여 AI 기술의 안전성과 신뢰성을 보장하는 거버넌스 체계를 구축하고, 국제 사회와의 협력을 강화함으로써 한국이 AI 윤리 및 규범 수립 과정에서 선도적인 역할을 수행할 수 있도록 정책적 지원을 확대해야 한다.

3) AI 산업 생태계 강화를 위한 정책적 지원

AI 산업의 지속 가능한 발전을 위해서는 생태계 전반의 강화를 목표로 하는 정책적 지원이 필수적이다. 이를 위해 중소기업-대기업 협력 촉진을 위한 디지털 협력 네트워크 및 데이터 공유 플랫폼 개발하고, 공동 기술 개발 및 실증 사업을 활성화해야 한다. 이 플랫폼은 대기업의 자원과 노하우가 중소기업의 혁신성과 융합되도록 지원하여 기술 개발과 시장 확대의 기회를 제공할 것이다. 특히, 산업별 AI 혁신 클러스터를 조성하고, 기업 간 협력을 촉진할 수 있는 오픈 이노베이션 프로그램을 확대하여 AI 기반 신기술 개발을 가속화해야 한다.

또한, 스타트업과 중소기업의 기술 침해를 방지하기 위해 법적·제도적 지원을 강화할 필요가 있다. 이를 위해 기술 탈취 방지 교육을 정기적으로 실시하고, 실제 사례를 분석하여 중소기업이 기술 보호 전략을 수립할 수 있도록 지원해야 한다. AI 기술 보호를 위한 특허 및 지식재산권(IP) 지원을 확대하고, AI 분야의 기술 이전 및 라이선싱 활성화를 위한 제도를 마련하여 스타트업과 중소기업이 보유한 기술이 공정하게 활용될 수 있는 환경을 마련하는 것이 필수적이다. 기술 보호를 위한 법적 근거를 강화함으로써 스타트업의 혁신 역량이 안정적으로 유지되고 발전할 수 있는 환경을 조성해야 한다.

아울러, 초거대 AI 모델의 효율적인 운영을 지원하기 위해 고성능·저전력 AI 반도체 기술 개발을 촉진해야 한다. 특히, 국내 반도체 스타트업과 대기업 간의 협력을 통해 기술 혁신을 가속화하고, AI 반도체의 글로벌 시장 경쟁력을 강화하기 위한 지원을 확대해야 한다. 이를 위해 정부 차원의 AI 반도체 R&D 투자 확대와 함께, AI 반도체 테스트베드를 구축하여 중소·스타트업 기업들이 개발한 AI 반도체의 성능을 검증하고 사업화를 지원하는 체계를 마련하는 정책적 방안 등이 필요하다.

4) AI 인재 양성과 교육 시스템 혁신

AI 기술의 발전이 가속화됨에 따라, 미래 인재를 양성하고 교육 시스템을 혁신하는 것이 필수적이다. 초·중등 교육 과정에는 AI 디지털 교과서를 도입하여 학생들이 일찍부터 AI 기초 지식을 습득할 수 있도록 해야 한다. 이러한 AI 기초 교육은 기본적인 코딩 기술과 데이터 이해력을 포함하며, 학생들이 실생활에서 AI를 활용하는 방법을 학습할 수 있도록 돕는다. 대학과 연구기관에서는 고급 AI 기술과 응용 연구를 지원하여 전문 인재를 양성하고, AI의 최첨단 기술 개발을 주도할 수 있는 연구 환경을 조성해야 한다.

AI 확산이 노동시장에 미치는 영향을 완화하기 위해 기존 인력을 대상으로 한 재교육 프로그램이 필요하다. 이를 위해 AI 기술 트레이닝 프로그램과 평생 학습 플랫폼을 운영하여 노동자들이 새로운 기술을 습득하고, 변화하는 산업 환경에 유연하게 적응할 수 있도록 지원해야 한다. 특히, 직무 전환이 필요한 직업군을 대상으로 한 맞춤형 교육과정과 실무 중심의 AI 활용 교육을 강화해야 한다.

또한, AI 학습을 위한 온라인 교육 플랫폼과 실습 환경을 구축하여 누구나 시간과 장소에 구애받지 않고 AI 교육에 접근할 수 있도록 해야 한다. 특히, 지방의 AI 교육 기관을 확대하고, 지역 산업과 연계한 AI 전문 인재 양성 프로그램을 운영함으로써 지역 균형 발전을 도모해야 한다.

3. 기대효과

2024년은 AI 산업이 기술적 도약과 함께 사회적·윤리적 과제를 해결하기 위한 해법을 본격적으로 모색하는 전환점으로 평가된다. 본 연구에서 제안된 정책은 AI 기술 발전과 사회적 책임 간의 균형을 유지하며, 국내 AI 산업의 지속 가능성과 글로벌 경쟁력을 동시에 강화하는 데 기여할 것이다. 이를 통해 AI 기술의 빠른 확산 속에서 정부와 민간이 협력하여 신뢰 기반의 기술 활용 환경과 효과적인 산업 생태계를 구축하는 데 필요한 방향성을 제안한다.

AI의 발전은 산업 혁신과 경제성장을 촉진하는 원동력인 동시에, 윤리적 문제와 기술적 위험 관리라는 중요한 과제를 함께 제기한다. 초거대 AI와 생성 AI는 의료, 교육, 공공행정 등 사회적 가치가 높은 분야에서 핵심적인 도구로 자리 잡으며, AI 기반 산업 구조 전환과 새로운 경제적 기회를 창출할 것으로 기대된다. 동시에, 이 기술들이 가져오는 윤리적 도전과 기술적 불확실성은 국내외 정책적 대응을 통해 신뢰할 수 있는 기술 생태계를 조성할 필요성을 강조한다.

본 연구는 AI 산업의 발전 현황과 주요 도전과제를 통합적으로 분석함으로써, 국내 AI 정책 및 산업 전략 수립을 위한 중요한 시사점을 제공한다. 특히, AI 기술 활용의 안전성과 윤리성을 확보하고, AI 기반 경제와 사회 구조 변화에 대응하기 위한 실효성 있는 정책적 기초 자료로 활용될 수 있다. 이를 통해 우리나라가 글로벌 AI 생태계에서 주도적인 역할을 수행하며, 혁신과 책임성을 겸비한 AI 기술 선도국으로 도약할 것으로 기대된다.

[참고 문헌]

1. AISI, AI세이프티에 관한 평가 관점 가이드의 공개, 2024.9.18.
2. Allen Institute for AI, OLMo: Open Language Model, 2024.02.01.
3. Anthropic, Claude 3.5 Sonnet, 2024.06.21.
4. Anthropic, Sleeper Agents: Training Deceptive LLMs that Persist Through Safety Training, 2024.01.15.
5. AP News, It's not just Elon Musk: ChatGPT-maker OpenAI confronting a mountain of legal challenges, 2024.03.07.
6. Apple, Introducing Apple Intelligence, the personal intelligence system that puts powerful generative models at the core of iPhone, iPad, and Mac, 2024.06.10.
7. Arxiv, AgentCourt: Simulating Court with Adversarial Evolvable Lawyer Agents, 2024.08.15.
8. Autorite de la Concurrence, Generative artificial intelligence: the Autorite issues its opinion on the competitive functioning of the sector, 2024.06.28.
9. Baidu, Baidu Presents a Suite of Toolkits and Models to Supercharge AI Creativity at Create 2024, 2024.04.16.
10. Bmbf, Zwischenbericht: Ein Jahr KI-Strategie, 2019.11.15.
11. BMW Group, Successful test of humanoid robots at BMW Group Plant Spartanburg, 2024.08.06.
12. BMWK, Strategie Künstliche Intelligenz der Bundesregierung Fortschreibung 2020. 2020.12.02.
13. California Legislative, SB-1047 Safe and Secure Innovation for Frontier Artificial Intelligence Models Act.(2023-2024), 2024.9.3.
14. CB Insights, Game Changers 2025, 2024.9.
15. CIO Korea, 미국, 한국 등 28개국, '안전하고 책임있는 AI 개발' 강조한 블레츨리 공동 선언 발표, 2023.11.02.
16. Cisco, Cisco's 2024 AI Readiness Index: Urgency Rises, Readiness Falls, 2024.11.19.
17. Cisco, Cisco's 2024 AI Readiness Index-South Korea, 2024.11.19.
18. CMA, AI Foundation Models: update paper, 2024.04.11.
19. Constellation Research, Google Cloud rolls out new Gemini models, AI agents, customer engagement suite, 2024.9.24.
20. EC, EU and Canada boost their strategic digital partnership to address new challenges, 2024.2.1.
21. EDPB, Statement 3/2024 on data protection authorities' role in the Artificial Intelligence Act framework, 2024.07.17.
22. European Commission, Coordinated Plan on Artificial Intelligence 2021 Review, 2021.4.21.
23. European Commission, European research development and deployment of AI,

2024.10.11.

24. European Commission, First Draft of the General-Purpose AI Code of Practice published, written by independent experts, 2024.11.14.
25. European Commission, Guidelines on the responsible use of generative AI in research developed by the European Research Area Forum, 2024.03.20.
26. European Data Protection Supervisor, EDPS Guidelines on generative AI: embracing opportunities, protecting people, 2024.6.03.
27. European Ombudsman, How the European Commission decides on and uses artificial intelligence, 2024.03.15.
28. Fedscoop, OpenAI, Anthropic enter AI agreements with US AI Safety Institute, 2024.08.29.
29. Figure AI, Figure unveils Figure 02, its second-generation humanoid, setting new standards in AI and robotics, 2024.08.06.
30. Fonearena, McAfee Project Mockingbird that can detect Deepfake audio demoed, 2024.01.12.
31. Forrester, Predictions 2025: An AI Reality Check Paves The Path For Long-Term Success, 2024.10.22.
32. Gartner, Intelligent Agents in AI Really Can Work Alone. Here's How, 2024.10.1.
33. Generative AI Misuse: A Taxonomy of Tactics and Insights from Real-World Data.
34. Gigazine, OpenAI announces it has broken down GPT-4's thoughts into 16 million interpretable patterns, 2024.06.07.
35. Global Times, OpenAI's cut of China's access to its API service would 'push domestic developers to catch up vigorously', 2024.07.10.
36. Google Deepmind, AlphaGeometry: An Olympiad-level AI system for geometry, 2024.01.17.
37. Google Deepmind, Generating audio for video, 2024.06.17.
38. Google Deepmind, Introducing the Frontier Safety Framework, 2024.05.17.
39. Google Deepmind, Mapping the misuse of generative AI, 2024.08.02.
40. Google DeepMind, Shaping the future of advanced robotics, 2024.01.04.
41. Google Research, Introducing ASPIRE for selective prediction in LLMs, 2024.01.18.
42. Google, AlphaFold 3 predicts the structure and interactions of all of life's molecules, 2024.05.08.
43. Google, Bard becomes Gemini: Try Ultra 1.0 and a new mobile app today, 2024.02.08.
44. Google, Gemini 1.5 Pro 2M context window, code execution capabilities, and Gemma 2 are available today, 2024.06.27.
45. Google, Gemini breaks new ground with a faster model, longer context, AI agents and more, 2024.05.14.
46. Google, Gemma 2 is now available to researchers and developers, 2024.06.27.

47. Google, How we're increasing transparency for gen AI content with the C2PA, 2024.09.17.
48. Google, How we're addressing explicit fake content in Search, 2024.07.31.
49. Google, New generative media models and tools, built with and for creators, 2024.05.14.
50. Google, New in Gemini: Custom Gems and improved image generation with Imagen 3, 2024.08.28.
51. Google, Our next-generation model: Gemini 1.5, 2024.02.15.
52. Google, Smaller, Safer, More Transparent: Advancing Responsible AI with Gemma, 2024.07.31.
53. Gouvernement.fr , 25 recommandations pour l'IA en France, 2024.03.13.
54. Gov.uk, Assuring a Responsible Future for AI, 2024.11.06.
55. Gov.uk, National AI Strategy, 2022.12.18.
56. Gov.uk, Research on parent and pupil attitudes towards the use of AI in education, 2024.08.28.
57. Gov.uk, Teachers to get more trustworthy AI tech, helping them mark homework and save time, 2024.08.28.
58. GOV.UK, UK-Canada cooperation in AI compute: memorandum of understanding, 2024.1.13.
59. Government of Canada, Artificial Intelligence and Data Act, 2023.9.27.
60. Government of Canada, Canada launches Canadian Artificial Intelligence Safety Institute, 2024.11.12.
61. Government of Canada, Generative AI in your daily work, 2024.10.15.
62. Government of Canada, Guide on the Scope of the Directive on Automated Decision-Making, 2024.6.27.
63. Government of Canada, Guide on the use of generative artificial intelligence, 2024.10.15.
64. Government of Canada, Pan-Canadian Artificial Intelligence Strategy, 2022.7.20.
65. Government of Canada, Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems, 2023.5.28.
66. IMDA, Singapore launches AI Verify Foundation to shape the future of international AI standards through collaboration, 2023.06.07.
67. IMF, Gen-AI: Artificial Intelligence and the Future of Work, 2024.01.14.
68. info.gouv.fr, 25 recommandations pour l'IA en France, 2024.03.13.
69. Kakao, 카카오, 'if(kakaoAI) 2024' 에서 그룹 AI 비전 공개...AI 메이트 '카나나' 도 첫 선, 2024.10.22.
70. McAfee, McAfee Unveils Advanced Deepfake Audio Detection Technology at CES 2024 to Defend Against Rise in AI-Generated Scams and Disinformation, 2024.01.08.

71. Meta Training and Inference Accelerator.
72. Meta, Introducing Llama 3.1: Our most capable models to date, 2024.07.23.
73. Meta, Introducing Meta Llama 3: The most capable openly available LLM to date, 2024.04.18.
74. Meta, Labeling AI-Generated Images on Facebook, Instagram and Threads, 2024.02.06.
75. Meta, Meet Your New Assistant: Meta AI, Built With Llama 3, 2024.04.18.
76. Meta, With 10x growth since 2023, Llama is the leading engine of AI innovation, 2024.08.29.
77. Microsoft, Discover the New Multi-Lingual, High-Quality Phi-3.5 SLMs, 2024.08.22.
78. Microsoft, Microsoft announces the largest investment to date in France to accelerate the adoption of AI, skilling and innovation, 2024.05.13.
79. Microsoft, VASA-1: Lifelike Audio-Driven Talking Faces Generated in Real Time, 2024.04.16.
80. Ministère de l'économie, La stratégie nationale pour l'intelligence artificielle. 2024.05.22.
81. Ministry of Electronics & IT, 6th meeting of the GPAI Ministerial Council held on 3rd July 2024 at New Delhi, 2024.07.03.
82. Ministry of Foreign Affairs of Japan, Prime Minister Kishida's attendance at the Side Event on Generative AI at the OECD Ministerial Council Meeting, 2024.05.02.
83. MIT News, New algorithm discovers language just by watching videos, 2024.6.11.
84. National Artificial Intelligence Research Resource Task Force, Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource. 2023. 1.
85. Nature, AI models collapse when trained on recursively generated data, 2024.07.24.
86. NCSL, Artificial Intelligence 2024 Legislation, 2024.9.
87. NIST, Artificial Intelligence Risk Management Framework, AI RMF 1.0. 2023. 5.
88. NIST, U.S. AI Safety Institute Signs Agreements Regarding AI Safety Research, Testing and Evaluation With Anthropic and OpenAI, 2024.08.29.
89. NSF, Democratizing the future of AI R&D: NSF to launch National AI Research Resource pilot. 2024.1.24.
90. Nvidia, NVIDIA Blackwell Platform Arrives to Power a New Era of Com
91. OECD, AI, data governance and privacy: Synergies and areas of international co-operation, 2024.06.26.
92. OECD, Artificial intelligence and the changing demand for skills in the labour market, 2024.04.10.
93. OECD, Artificial intelligence and wage inequality, 2024.04.10.
94. OECD, Explanatory memorandum on the updated OECD definition of an AI

- system, 2024.03.05.
95. OECD, Generative artificial intelligence in finance, 2023.12.15.
 96. OECD, 'OECD Artificial Intelligence Review of Germany, 2024.06.11.
 97. OECD, Recommendation of the Council on Artificial Intelligence, 2022. 5.
 98. OECD, The potential impact of Artificial Intelligence on equity and inclusion in education, 2024.08.
 99. OECD, Using AI in the workplace: Opportunities, risks and policy responses, 2024.03.15.
 100. OMB, Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence, 2024.03.28.
 101. OMB, Memorandum For The Heads Of Executive Departments And Agencies, 2024.03.28
 102. OpenAI, C2PA in DALL · E 3, 2024.02.06.
 103. OpenAI, Extracting Concepts from GPT-4, 2024.06.06.
 104. OpenAI, GPT-4o mini: advancing cost-efficient intelligence, 2024.07.18.
 105. OpenAI, Hello GPT-4o, 2024.05.13.
 106. OpenAI, How OpenAI is approaching 2024 worldwide elections, 2024.01.15.
 107. OpenAI, Introducing OpenAI o1-preview, 2024.09.12.
 108. OpenAI, Introducing the GPT Store, 2024.01.10.
 109. OpenAI, Navigating the Challenges and Opportunities of Synthetic Voices, 2024.03.29.
 110. OpenAI, OpenAI acquires Rockset, 2024.06.21.
 111. OpenAI, SearchGPT Prototype, 2024.07.25.
 112. OpenAI, Superalignment Fast Grants, 2023.12.14.
 113. OpenAI, Weak-to-strong generalization, 2023.12.14.
 114. Parliament of Canada, Bill C-27, An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts, 2023.4.24.
 115. PDPC, Singapore' s Approach to AI Governance, 2024.11.17. .
 116. Prime Minister of Canada, Securing Canada' s AI advantage, 2024.4.7.
 117. Reuters, Meta to deploy in-house custom chips this year to power AI drive, 2024.02.02.
 118. Sakana.ai, The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery, 2024.08.13.
 119. Senate Dems, Majority Leader Schumer Delivers Remarks To Launch SAFE Innovation Framework For Artificial Intelligence At CSIS. 2023.6.21.
 120. SiliconAngle, Baidu unveils Ernie 4.0 Turbo in a 'significant upgrade' to its AI chatbot, 2024.06.28.

121. SiliconAngle, French open-source AI model startup Mistral AI raises \$640M at \$6B valuation, 2024.06.11.
122. SiliconAngle, Meta faces AI accuracy issues as tech industry tackles hallucinations, deepfakes, 2024.07.31.
123. Smartnation, Factsheet - Artificial Intelligence (AI) initiatives launched to uplift Singapore's economic potential, 2024.03.01.
124. Smartnation, National AI Strategy, 2019.11.30.
125. Smartnation, Singapore National AI Strategy 2.0 (NAIS 2.0), 2023.12.05.
126. Stability.ai, Introducing Stable LM 2 1.6B, 2024.01.19.
127. Stability.AI, Stable Diffusion 3, 2024.02.22.
128. Stanford HAI, AI Index Report 2024.
129. TechCrunch, Google launches Gemini Ultra, its most powerful LLM yet, 2024.02.08.
130. TechCrunch, OpenAI buys Rockset to bolster its enterprise AI, 2024.06.21.
131. TechCrunch, OpenAI created a team to control 'superintelligent' AI — then let it wither, source says, 2024.05.18.
132. TechRadar, Jarvis AI is real - Google accidentally leaks its AI agent that browses the web for you, 2024.11.6.
133. The federal government, AI - a brand for Germany, 2018.11.15.
134. The New York Times, California Legislature Approves Bill Proposing Sweeping A.I. Restrictions, 2024.08.28.
135. The New York Times, U.S. Clears Way for Antitrust Inquiries of Nvidia, Microsoft and OpenAI, 2024.06.05.
136. The Verge, Gemini 1.5 is Google's next-gen AI model — and it's already almost ready, 2024.02.16.
137. The Verge, Google says Gemini AI is making its robots smarter, 2024.07.11.
138. The Verge, Nvidia reveals Blackwell B200 GPU, the 'world's most powerful chip' for AI, 2024.03.19.
139. The Verge, OpenAI announces SearchGPT, its AI-powered search engine, 2024.07.26.
140. The Wall Street Journal, There's a Tool to Catch Students Cheating With ChatGPT. OpenAI Hasn't Released It., 2024.08.04.
141. The White House, FACT SHEET: Biden-Harris Administration Outlines Coordinated Approach to Harness Power of AI for U.S. National Security, 2024.10.24.
142. Tortois Intelligence. The Global Artificial Intelligence Index, 2024.09.19.
143. UK AI Safety Institute, AI Safety Institute releases new AI safety evaluations platform, 2024.05.10.
144. UK AI Safety Institute, Collaboration on the safety of AI: UK-US memorandum of understanding, 2024.04.02.

145. UK AI Safety Institute, International Scientific Report on the Safety of Advanced AI, 2024.05.17.
146. UK Cabinet Office, Generative AI Framework for HMG, 2024.01.18.
147. UNESCO, Foundation Models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence, 2023.6.
148. UNESCO, Generation AI: Navigating the opportunities and risks of artificial intelligence in education, 2024.07.22.
149. United Nations AI Advisory Body, Governing AI for Humanity - Final Report, 2024.09.19.
150. United Nations(General Assembly), Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development, 2024.3.11.
151. Venturebeat, Apple researchers achieve breakthroughs in multimodal AI as company ramps up investments, 2024.03.15.
152. VentureBeat, Apple researchers develop AI that can 'see' and understand screen context, 2024.04.01.
153. Venturebeat, DeepMind researchers discover impressive learning capabilities in long-context LLMs, 2024.04.24.
154. VentureBeat, Elon Musk's xAI defies 'woke' censorship with controversial Grok 2 AI release, 2024.08.14.
155. Venturebeat, iOS gets an AI upgrade: Inside Apple's new 'Intelligence' system, 2024.07.29.
156. Venturebeat, Meta's Self-Taught Evaluator enables LLMs to create their own training data, 2024.08.19.
157. Venturebeat, Stable Diffusion 3.0 debuts new diffusion transformation architecture to reinvent text-to-image gen AI, 2024.02.22.
158. whitehouse, Fact Sheet: Biden-Harris Administration Announces Key AI Actions Following President Biden's Landmark Executive Order. 2024.1.19.
159. whitehouse, FACT SHEET: Biden-Harris Administration Announces New AI Actions and Receives Additional Major Voluntary Commitment on AI. 2024.7.26.
160. Whitehouse, Fact Sheet: Key AI Accomplishments in the Year Since the Biden-Harris Administration's Landmark Executive Order, 2024.10.30.
161. Wired, Why Elon Musk Had to Open Source Grok, His Answer to ChatGPT, 2024.03.11.
162. xAI, Announcing Grok-1.5, 2024.03.28.
163. xAI, Grok-1.5 Vision Preview, 2024.04.12.
164. xAI, Grok-2 Beta Release, 2024.08.13.
165. xAI, Open Release of Grok-1, 2024.03.17.
166. ZDNet, Apple claims its on-device AI system ReaLM 'substantially outperforms' GPT-4, 2024.04.02.

167. 한국경제, SKT “기가와트급 데이터센터 조성…'AI 인프라 슈퍼 하이웨이' 만든다“, 2024.11.04.
168. 한국과학기술기획평가원(KISTEP), 국가전략기술 R&D(연구·개발) 인력 실태조사, 2024.5.
169. 한국교육학술정보원(KERIS), 생성형 AI를 활용한 교수학습 운영 가이드, 2024.5.
170. 한국교육학술정보원(KERIS), 생성형 AI를 활용한 교수학습 프레임워크 개발 연구, 2024.1.
171. 한국전자통신연구원, ETRI, AI안전 및 신뢰성지원 국제표준 제정, 2024.07.16.
172. 人民邮电报, 工信部拟筹建人工智能标准化技术委员会, 2024.07.05.
173. 全国网络安全标准化技术委员会, 《人工智能安全治理框架》1.0版, 2024.0.09.
174. 内閣府, AI に関する暫定的な論点整理, 2023.5.26.
175. 内閣府, AIセーフティ・インスティテュートの設立について, 2024.2.14.
176. 内閣府, 統合イノベーション戦略 2023, 2023.6.9.
177. 内閣府, 統合イノベーション戦略 2024, 2023.6.4.
178. 内閣府, 統合イノベーション戦略 2024, 2024.6.4.
179. 国家互联网信息办公室关于, 人工智能生成合成内容标识办法(征求意见稿, 2024.09.14.
180. 工业和信息化部, 国家人工智能产业综合标准化体系建设指南(2024版), 2024.06.05.
181. 广东省人民政府, 关于人工智能赋能千行百业若干措施的通知, 2024.06.06.
182. 成都市出台了《成都市人工智能产业高质量发展三年行动计划(2024—2026年)》, 2024.06.20.
183. 湖南省工业和信息化厅, 《湖南省人工智能产业发展三年行动计划(2024-2026年)》, 2024.09.04.
184. 科技日报, 培育新质生产力在行动 | 培育千亿级人工智能产业创新集群 陕西发布行动计划推进五大工程, 2024.04.08.
185. 経団連, AI政策・制度検討の動向, 2024.9.26.
186. 経済産業省, AI事業者ガイドライン(第1.0版)』を取りまとめました, 2024.4.19.
187. 経済産業省; 総務省, AI 事業者ガイドライン(第 1.0 版), 2024.4.19.
188. 防衛省, 防衛省 A I 活用推進基本方針, 2024.7.
189. 首相官邸, A I 戦略会議・A I 制度研究会合同会議, 2024.8.2.

[첨부 1] 2024년 주요국 AI 관련 분야별 정책 동향 요약

<표> 2024년도 주요국의 AI 관련 분야별 정책 동향 요약

구분	정책/법제	기업/산업	기술/연구	인력/교육
미국	<ul style="list-style-type: none"> 백악관, AI 행정명령, 2023.10(~2024.1) - AI 리더십, AI 안전 등 예산관리국, 연방정부의 AI 사용에 관한 정책 발표, 2024.3 백악관, 국가안보 각서, 2024.10 - 국가안보에 AI 활용 - AI 중심 국제적 협의회 거버넌스 발전 	<ul style="list-style-type: none"> 백악관, AI 행정명령, 2023.10(~2024.1) - AI 혁신 투자 등 백악관, 국가안보 각서, 2024.10 - 첨단 AI 칩 확보 - AI 안전연구소 재무부, 對중국 투자 제한 행정규칙 발표, 2024.11 	<ul style="list-style-type: none"> 백악관, AI 행정명령, 2023.10(~2024.1) - AI 신기술 개발 NSF, 국가 AI 연구 리소스 파일럿 착수, 2024.1 백악관, 국가안보 각서, 2024.10 - 국가AI연구자원 (NAIRR) 프로젝트 강화 등 미국 AI 안전연구소, 영국 AI안전연구소와 MOU, 2024.4 	<ul style="list-style-type: none"> 백악관, AI 행정명령, 2023.10(~2024.1) - AI 전문인력 유치 및 교육 등 백악관, 국가안보 각서, 2024.10 - 국가안보에 AI 활용
EU	<ul style="list-style-type: none"> EU, AI 법 제정, 2024.5 EC, AI 법에 따른 AI 적용 실천 강령 1차 초안 공개, 2023.11 EC, AI 사무국 설립, 2024.5 	<ul style="list-style-type: none"> EC, AI 스타트업과 중소기업을 지원하기 위한 AI 혁신 패키지, 2024.1 - AI 팩토리 출범, AI 사무국 설립, EU AI 스타트업 및 혁신 커뮤니티 	<ul style="list-style-type: none"> EC, 연구에서 생성 AI의 책임 있는 사용에 관한 지침, 2024.3 	<ul style="list-style-type: none"> 유럽데이터보호감독기구(EDPS) 생성 AI 지침, 2024.6 유럽평의회, AI와 인권·민주주의·법치에 관한 협약, 2024.5
중국	<ul style="list-style-type: none"> 국무원, 'AI 플러스 행동' 개념 제시, 2024.3 청두시, AI 산업 고품질 발전 3개년 계획, 2024.6 베이징시, 베이징시 AI 플러스 행동계획 (2024~2025), 2024.7 후난성, 후난성 AI 산업 발전 3개년 행동계획(2024~2026년), 2024.9 	<ul style="list-style-type: none"> 산시성 발전개혁위원회, AI 산업 혁신 클러스터 육성 행동계획, 2024.4 광둥성 정부, AI와 천 개 산업 융합 발전을 위한 조치, 2024.5 	<ul style="list-style-type: none"> 공업정보화부, 국가 AI 산업 종합 표준화 체계 구축 지침, 2024.7 국가인터넷정보판공실, 시로 생성된 합성 콘텐츠 표시 방법(초안), 2024.9 중국사이버보안표준화 기술위원회, AI 보안 거버넌스 프레임워크, 2024.9 	-

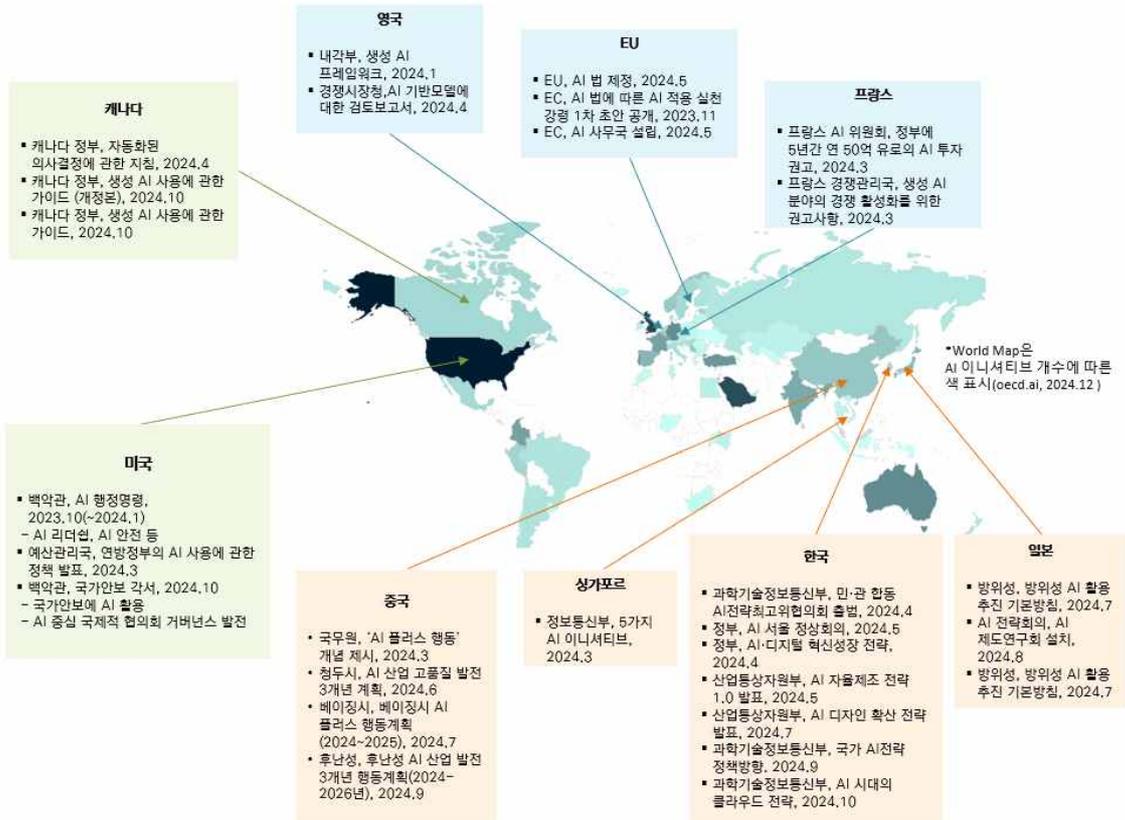
구분	정책/법제	기업/산업	기술/연구	인력/교육
일본	<ul style="list-style-type: none"> 방위성, 방위성 AI 활용 추진 기본방침, 2024.7 AI 전략회의, AI 제도연구회 설치, 2024.8 	<ul style="list-style-type: none"> 총무성·경제산업성, AI 사업자 가이드라인(제1.0판), 2024.4 	<ul style="list-style-type: none"> 일본정부, AI 안전연구소(AISI), 2024.2 AI 안전연구소, AI 안전에 관한 평가 관점 가이드, 2024.9 	<ul style="list-style-type: none"> 총무성·경제산업성, AI 사업자 가이드라인(제1.0판), 2024.4
영국	<ul style="list-style-type: none"> 내각부, 생성 AI 프레임워크, 2024.1 경쟁시장청, AI 기반모델에 대한 검토보고서, 2024.4 	<ul style="list-style-type: none"> 영국 AI 안전연구소, AI 안전성 테스트 플랫폼 ‘인스펙트’ 개발, 2024.5 	<ul style="list-style-type: none"> 과학혁신기술부, AI 보증 플랫폼 구축, 2025.11 영국 AI 안전연구소, 미국 AI안전연구소와 MOU, 2024.4 	<ul style="list-style-type: none"> 과학혁신기술부/교육부, 콘텐츠 스토어 프로젝트, 2024.8 영국 AI 안전연구소, 첨단 AI의 안전에 관한 국제과학보고서, 2024.5
캐나다	<ul style="list-style-type: none"> 캐나다 정부, 자동화된 의사결정에 관한 지침, 2024.4 캐나다 정부, 생성 AI 사용에 관한 가이드 (개정본), 2024.10 캐나다 정부, 생성 AI 사용에 관한 가이드, 2024.10 	<ul style="list-style-type: none"> 캐나다 정부, 2024 예산안, 2024.4 <ul style="list-style-type: none"> - 컴퓨팅 기술 인프라 구축(2억 캐나다 달러 등) 	<ul style="list-style-type: none"> 캐나다 정부, AI 안전연구소, 2024.11 	<ul style="list-style-type: none"> 캐나다 정부, 생성 AI 사용에 관한 가이드, 2024.10 <ul style="list-style-type: none"> - 생성 AI 사용 방법 학습, 투명성 등
프랑스	<ul style="list-style-type: none"> 프랑스 AI 위원회, 정부에 5년간 연 50억 유로의 AI 투자 권고, 2024.3 프랑스 경쟁관리국, 생성 AI 분야의 경쟁 활성화를 위한 권고사항, 2024.3 	-	-	-
싱가포르	<ul style="list-style-type: none"> 정보통신부, 5가지 AI 이니셔티브, 2024.3 	<ul style="list-style-type: none"> 정보통신부, 5가지 AI 이니셔티브, 2024.3 <ul style="list-style-type: none"> - AI 혁신/역량 구축 	<ul style="list-style-type: none"> 정보통신부, 5가지 AI 이니셔티브, 2024.3 <ul style="list-style-type: none"> - Generative AI x Digital Leaders 이니셔티브 	<ul style="list-style-type: none"> 정보통신부, 5가지 AI 이니셔티브, 2024.3 <ul style="list-style-type: none"> - AI 연구 인재 육성

구분	정책/법제	기업/산업	기술/연구	인력/교육
한국	<ul style="list-style-type: none"> 과학기술정보통신부, 민·관 합동 AI전략최고위협의회의 출범, 2024.4 정부, AI 서울 정상회의, 2024.5 정부, AI·디지털 혁신성장 전략, 2024.4 산업통상자원부, AI 자율제조 전략 1.0 발표, 2024.5 산업통상자원부, AI 디자인 확산 전략 발표, 2024.7 과학기술정보통신부, 국가 AI전략 정책방향, 2024.9 과학기술정보통신부, AI 시대의 클라우드 전략, 2024.10 	<ul style="list-style-type: none"> 정부, AI·디지털 혁신성장 전략, 2024.4 <ul style="list-style-type: none"> - AI 대전환 가속화 금융위원회, 3.5조원 규모의 AI정책금융 프로그램 신설, 2024.7 중소벤처기업부, 초격차 AI 스타트업 성장 전략 발표, 2024.7 산업통상자원부, 국산 AI 반도체의 주력산업 도입 방안, 2024.8 산업통상자원부, 산업데이터 활용 활성화 방안 발표, 2024.10 	<ul style="list-style-type: none"> 정부, AI 안전연구소, 2024.11 정부, AI·디지털 혁신성장 전략, 2024.4 <ul style="list-style-type: none"> - AI 기술혁신 개인정보보호위원회, AI 개발·서비스 이용 '공개 데이터' 처리 기준, 2024.7 개인정보보호위원회, 자율주행 AI 관련 개인영상정보 보호 기준, 2024.10 	<ul style="list-style-type: none"> 정부, AI·디지털 혁신성장 전략, 2024.4 <ul style="list-style-type: none"> - 새로운 AI 질서 확립

[첨부 2] 2024년 주요국 AI 정책 전 세계 맵(World Map)

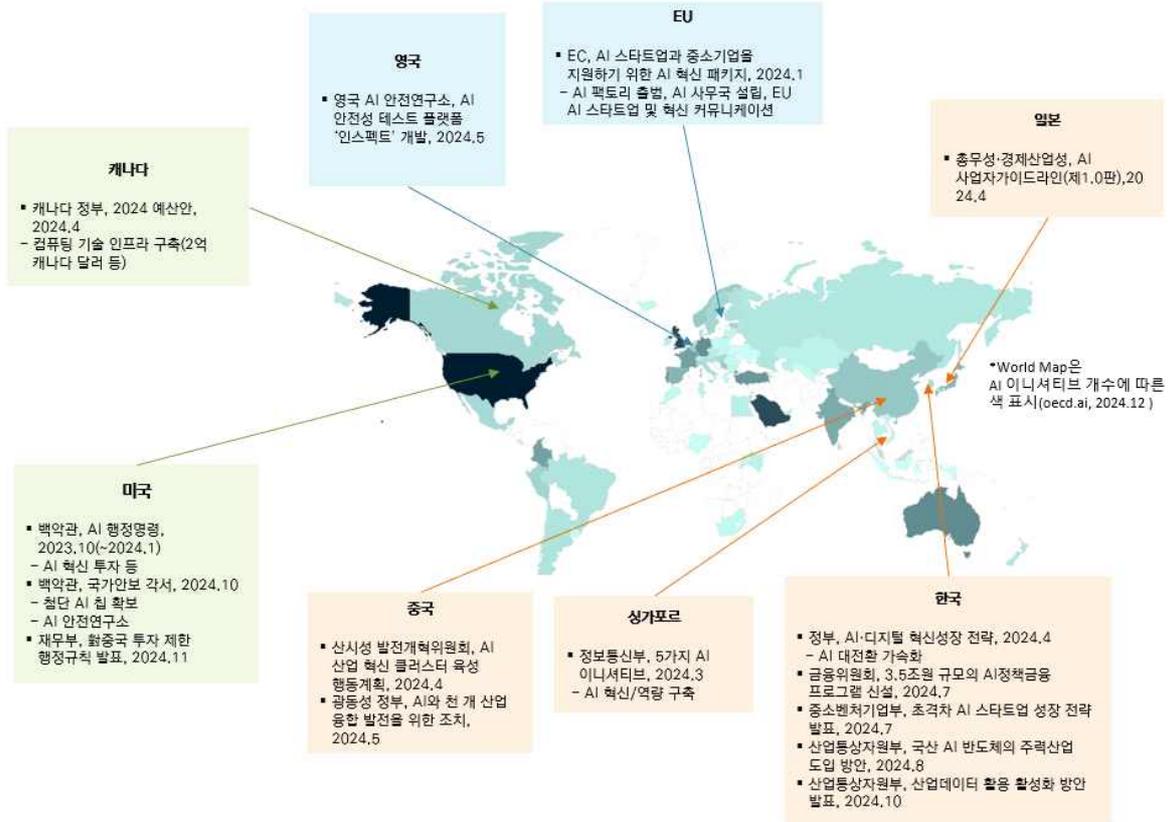
① 정책/법제

[그림] 2024년 주요국 AI 정책(정책/법제 분야) 전 세계 맵(World Map)



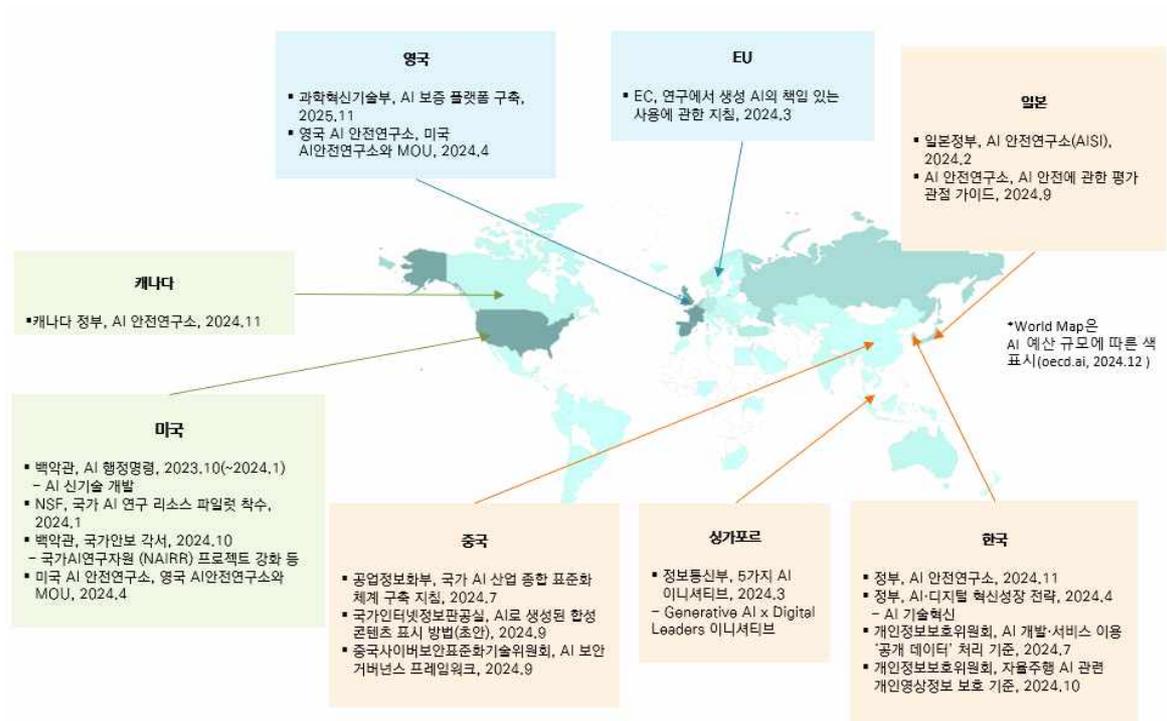
② 기업/산업

[그림] 2024년 주요국 AI 정책(기업/산업 분야) 전 세계 맵(World Map)



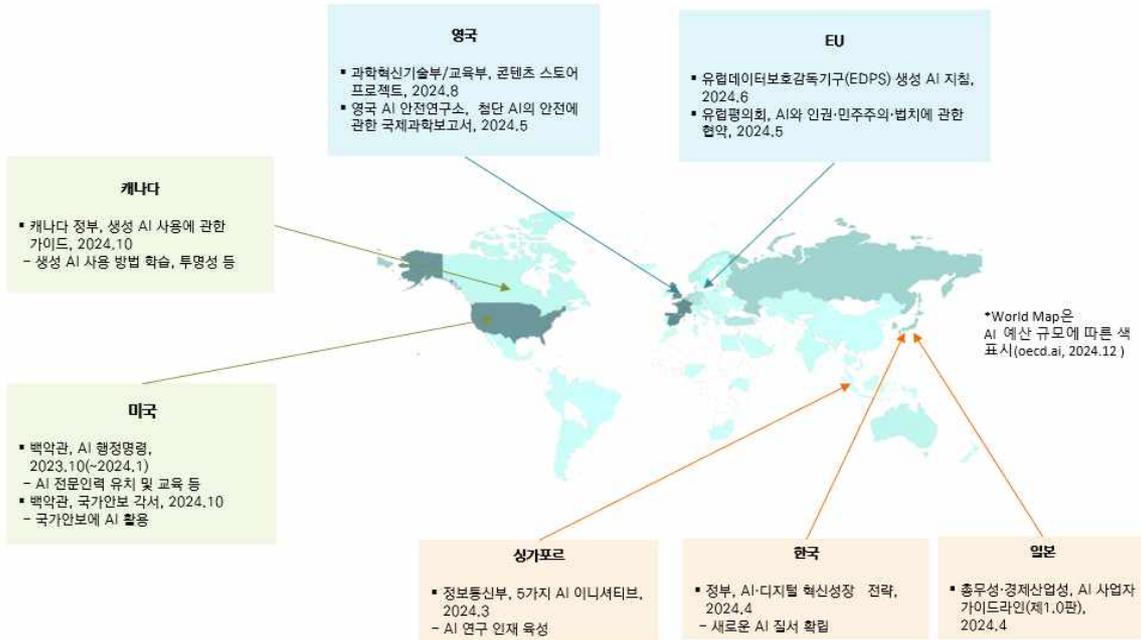
③ 기술/연구

[그림] 2024년 주요국 AI 정책(기술/연구 분야) 전 세계 맵(World Map)



④ 인력/교육

[그림] 2024년 주요국 AI 정책(인력/교육 분야) 전 세계 맵(World Map)



[첨부 3] 국내외 주요 기업 동향

① 국내 AI 기업 동향

〈표〉 국내 AI 기업 동향

가치사슬	기업명	주요 활동 (기술·모델 개발, 투자유치(인수합병), 사업전략(협업제휴, 인재확보), 제품/서비스출시 등)
애플리케이션·서비스	뤼튼테크놀로지	- (서비스) 대화형 챗봇 '챗 뤼튼' 출시('23.3), AI 챗봇(나만의 AI)에 AI 검색 기능 업데이트2('3.8) - (플랫폼) 콘텐츠 생성 기능과 플러그인 기능 갖춘 AI 플랫폼 '뤼튼 2.0' 출시('23.4) - (투자) 프리시리즈B 투자 250억 원 유치('24.6)
	이스트소프트	- (서비스) 구독형 AI 휴먼 영상 제작 'AI 스튜디오 페르소'와 대화형 AI 휴먼 '페르소 라이브' 출시('24.5) - (제휴) 일본 LLM 응용 서비스 기업 스파이럴IA와 'AI 휴먼 공동 사업' MOU 체결('24.2), 네이버클라우드와 대학 AI 교육 플랫폼 구축('24.9)
	스캐터랩	- (서비스) 엔터테인먼트·콘텐츠·커머스 시장용 sLLM '핑퐁-1' 공개('23.11), AI 스토리 플랫폼 '제타' 출시('24.4)
	프렌들리시	- (서비스) AI 개발용 GPU 최적화 서비스 '서버리스 엔드포인트' 출시('24.1), 기업 맞춤형 AI 개발 서비스 '프렌들리 파인튜닝' 출시('24.10)
	라이너	- (서비스) 생성 AI 검색 서비스 출시('23.6), 학술 모드 AI 검색 베타 서비스 출시('24.7) - (기술) 카이스트와 AI 에이전트 신뢰도 측정하는 벤치마크 연구('24.6) - (투자) 270억 원 규모 시리즈 B2 투자 유치('24.10)
	한컴	- (서비스) 문서 요약과 작성, 번역 등 지원하는 '한컴독스 AI' 출시('24.6), 자회사 씽크프리, 글로벌 기업용 AI 검색 서비스 '리파인더 AI' 베타 출시('24.8)
	룩스웨이브	- (서비스) AI 챗봇 구축·운영·분석 지원하는 AI 엔터프라이즈 솔루션 출시('24.2) - (제휴) 인도 AI 기업 펀다멘토와 콜센터 AI 전환 솔루션 개발 위한 전략적 파트너십 체결('24.7) - (투자) 45억 원 규모의 시드 투자 유치('24.1)
AI모델·플랫폼	네이버	- (기반모델) 초거대 언어모델 '하이퍼클로바 X' 공개('23.8), 하이퍼클로바X 신규 모델 '대시' 공개('24.4), 하이퍼클로바X에 이미지 처리 기능 업데이트('24.8) - (서비스) AI 쇼핑앱 출시 계획('24년 상반기) - (기술) 네이버 ASF(AI Safety Framework)' 발표('24.6)
	카카오	- (기반모델) 카나나 플래그, 카나나 에센스, 카나나 나노 개발 중('24.11) - (서비스) AI 비서 '카나나' 공개('24.10)
	KT	- (기반모델) 거대 AI '믿음' 발표('23.10) - (제휴) MS와 AI·클라우드·정보기술·IT 분야 협력을 위한 전략적 파트너십 체결('24.6), 데이터브릭스와 생성 AI 및 데이터 서비스 개발 협력('24.11)
	SKT	- (플랫폼) 생성 AI 플랫폼 '엔터프라이즈 AI 마켓' 출시('24.1), 에이닷에서 글로벌 LLM을 종합한 멀티 LLM 에이전트 서비스 제공('24.8) - (서비스) AI 개인비서 '에이닷' 출시('23.9), 글로벌 AI 에이전트 '에스타' 공개('24.11) - (제휴) Humane/Perplexcity AI와 상호투자, 공동마케팅, 글로벌 AI 에이전트 서비스 고도화 등 협력('24.09) - (투자) Anthropic에 1억 달러 투자('23.8), AI 검색엔진 기업 Perplexity AI에 1천만 달러 투자('24.6)
	LGT	- (서비스) 보이스피싱 실시간 탐지 가능한 AI 통화 비서 '익시오' 출시('24.11)
	LG AI연구원	- (기반모델) 엑사원 공개('21.12), 엑사원 2.0 공개('23.7.), 엑사원 3.0 오픈소스 공개('24.8), 엑사원 3.5 오픈소스 공개('24.12)
	삼성전자	- (기반모델) 자체 개발 생성 AI '삼성 가우스' 공개('23.11), 멀티모달 모델 '가우스2' 공개('24.11) - (기술) 서울대와 AI공동연구센터 설립('24.6) - (인수) 영국 AI 스타트업 '옥스퍼드 시멘틱 테크놀로지스' 인수('24.7)
	솔트룩스	- (하드웨어) 생성 AI 개발용 하드웨어 '루시아 온' 출시('24.3) - (기반모델) 자체 개발 LLM '루시아2' 공개('24.5) - (서비스) 기업용 검색 솔루션 '구버 엔터프라이즈' 출시('24.9), AI 검색 서비스 '구버' 프리뷰 버전 공개('24.11)
	마음씨	- (기반모델) 기업용 거대언어모델 'MAAL1' 공개('24.3), 온프레미스 LLM 제품 'MAAL'을 라마 3.1 기반으로 업그레이드('24.8) - (인수) 멀티모달 기반모델 특화 기업 맨드언맨드 인수('24.5)

가치사슬	기업명	주요 활동 (기술·모델 개발, 투자유치(인수합병), 사업전략(협업제휴, 인재확보), 제품/서비스출시 등)
	업스테이지	- (기반모델) 'Solar Pro' 프리뷰 공개('24.9), 광학문자인식(OCR) 모델 'Document Parse' 공개('24.10) - (플랫폼) LLM 평가 플랫폼 '이벨버스' 공개('24.4) - (서비스) LLM '솔라' 기반 번역 서비스 베타 공개('24.5) - (투자) 1천억 원 규모 시리즈 B 투자 유치('24.4)
	포티투마루	- (기반모델) AI 엑스포 2024에서 산업 특화 LLM 모델 'LLM42' 공개('24.5) - (투자) LG유플러스로부터 100억 투자 유치('24.1) - (제휴) LG유플러스, 샌즈랩과 사이버보안 강화 LLM 기술 공동 개발 MOU('24.3), 오라클과 클라우드 인프라 및 LLM 사업 협력('24.7)
	트웰브랩스	- (기반모델) 영상언어 AI 'Pegasus-1' 및 멀티모달 영상 이해 모델 'Marengo 2.6' 업데이트('24.3) - (투자) 엔비디아 등으로부터 700억 규모 시리즈A 투자 유치('24.6)
	코난테크놀로지	- (기반모델) 자체 개발 '코난 LLM' 출시('23.8), 사용자 목소리를 다양한 언어로 변환하는 음성합성 모델 공개('24.4) - (제휴) AI 반도체 스타트업 리벨리온과 생성 AI 키오스크 개발 협력 체결('24.8)
	모레	- (기반모델) 자체 개발 LLM 'MoMo-70B' 허깅페이스 LLM 평가에서 세계 1위('24.1) - (제휴) AMD와 전략적 파트너십 맺고 2,200만 달러 규모 투자 유치('23.10)
	매스프레소	- (기반모델) 업스테이지와 수학 특화 AI 모델 'MathGPT' 개발('24.1) - (제휴) NE능률과 저작물 이용 계약 체결('24.1) - (서비스) 미국향 대학교육 서비스 '프렙파이(Prep.Pie)' 출시('24.3)
	올거나이츠	- (기반모델) 금융 전문 AI 모델 '알리 파이낸스 LLM' 출시('23.9)
	엔씨소프트	- (기반모델) 오픈소스 기반 LLM 'Llama-VARCO LLM(라마 바르코 LLM)' 공개('24.9)
클라우드	KT 클라우드	- (인프라) 국내 AI 반도체 스타트업 '리벨리온'의 아톰 적용한 NPU 인프라 구축, AI 전용 데이터센터 백색 AIDC 개관('24.8) - (서비스) 엔비디아 GPU 기반 AI 학습 전용 인프라 서비스 'AI 트레인' 출시('24.5)
	네이버클라우드	- (플랫폼) 글로벌 스튜디오를 통해 기업 고객 대상 하이퍼클로바X 모델 제공('23.10) - (제휴) 인텔과 AI 공동연구센터 설립('24.4)
	NHN클라우드	- (투자) 솔트룩스와 AI-클라우드 결합 제휴 체결 및 솔트룩스에 156억 원 지분 투자('23.11) - (인프라) 광주 국가 AI 데이터센터 운영 개시('24.3) - (제휴) 머신러닝 기업 베슬에이아이 코리아와 AI 인프라 및 MLOps 사업 협력 MOU 체결('24.6)
하드웨어	삼성전자	- (제품) AI 탑재한 갤럭시 S24 시리즈 공개('24.1) - (제휴) 리벨리온과 파운드리 4나노공정 이용 AI반도체 '리벨' 공동 개발 진행('23.10~), 네이버클라우드와 하이퍼클로바X 구동을 위한 AI반도체 솔루션 개발 진행('22.12.~) - (투자) 미국 AI 반도체 스타트업 Enfabrica에 투자('24.11)
	SK하이닉스	- (제품) 초고성능 AI 메모리 'HBM3E' 양산 시작('24.3), 온디바이스 AI 용 모바일 낸드 솔루션 'ZUFS 4.0' 개발('24.5)
	사파온	- (제품) 데이터센터용 AI 반도체 X330 출시('23.11) - (합병) 리벨리온과 합병 본계약 체결, 연내 통합법인으로 리벨리온 출범 계획('24.8)
	리벨리온	- (제품) AI 반도체 아톰 출시('21.11) - (제휴) 삼성전자와 LLM 특화 고성능 AI반도체 '리벨' 개발 협력('23.10), Arm 및 삼성전자와 제휴해 AI CPU 칩셋 플랫폼 개발 추진('24.10) - (투자) 1,650억 원 규모의 시리즈 B 투자 유치('24.1)
	퓨리오사	- (제품) AI 반도체 워보이(WARBOY) 개발 완료 후 삼성전자 파운드리 통해 양산 개시('23.4), 2세대 AI 반도체 RINGD(레니게이드) 공개('24.8) - (제휴) 네이버클라우드와 AI 반도체 솔루션 개발 협력('22.12), AI 교육 플랫폼 기업 엘리스그룹의 데이터 센터 구축 추진('23.9), 네이버클라우드와 협력해 GPU-NPU 팜 구축('23.12)
	딥엑스	- (제품) 온디바이스 AI용 NPU 공개('24.5), 온디바이스 AI 1세대 제품 DX-M1 양산 돌입('24.8) - (투자) 사모펀드 기관 중심 1,100억 원 규모의 투자 유치('24.5)

② 해외 AI 기업 동향

〈표〉 해외 AI 기업 동향

가치사슬	기업명	주요 활동 (기술·모델 개발, 투자유치(인수합병), 사업전략(협업제휴), 제품/서비스출시 등)
애플리케이션 · 서비스	Adobe	- (기반모델) AI 사진편집 모델 'Magic Fixup' 오픈소스 공개('24.8), 'FireFly Video Model' 공개('24.10) - (서비스) PDF 편집기에 AI 어시스턴트 출시('24.2), 생성 AI 음악 편집 도구 프리뷰 버전 공개('24.2), 기업용 광고 생성 AI '젠스튜디오' 출시('24.2), 이미지 생성 AI 'Firefly Image 3' 공개('24.4)
	Luma AI	- (기반모델) 동영상 생성 AI 'Dream Machine' 공개('24.6), 'Dream Machine 1.6' 공개('24.9)
	Udio	- (서비스) 음악 생성 AI 앱 'Udio' 공개('24.4)
	Stability AI	- (기반모델) Stable Diffusion 2.0 ('22.11), Stable Diffusion XL('23.7), Stable Cascade('24.2), Stable Diffusion 3.0 ('24.2), Stable Diffusion 3.5('24.10) - (서비스) 텍스트 기반 애니메이션 생성 도구 출시('23.5) 이미지를 3D 동영상으로 변환하는 'Stable Video 3D' 발표('24.3)
	MjJourney	- (기반모델) V6('23.12) - (서비스) 이미지 일관성 위한 'Character Reference' 기능('24.3), 통합 웹이미지 편집기 출시('24.8), 이미지 편집 기능 출시('24.10)
	Runway	- (기반모델) 동영상 생성 AI Gen-3 Alpha 출시('24.6), Gen-3 Alpha Turbo 출시('24.8) - (서비스) 영상 속 인물 캐릭터화 AI 도구 'Act-One' 출시('24.10)
	Viggle AI	- (기반모델) 물리학 지식을 통합한 동영상 3D 기반모델 JST-1 - (서비스) 캐릭터의 동작을 만드는 동영상 생성 AI 공개('24.8) - (투자) 1,900만 달러 규모의 시리즈A 투자 유치('24.8)
	Leonardo	- (기반모델) 이미지 생성 모델 'Phoenix' 공개('24.6) - (서비스) 기업용 시각 자료 생성 서비스 'Leonardo for Teams' 출시('24.4) - (인수) 디자인 플랫폼 Canva에 인수('24.7)
	ElevenLabs	- (기반모델) 32개 언어 지원하는 텍스트-음성 변환 모델 Turbo v2.5('24.7) - (서비스) 음향 효과 생성 'Sound Effects' 출시('24.6), 배경 소음 제거하는 'AI Voice Isolator' 출시('24.7), 텍스트-오디오 변환 모바일 앱 'Reader' 출시('24.8), 음성 생성 서비스 'Voice Design' 출시('24.10)
	Pika Labs	- (기반모델) 동영상 생성 AI 모델 'Pika 1.0' 출시('23.11), 'Pika 1.5' 출시('24.10) - (서비스) Pika 1.0에 음성 기능 및 음향효과 자동 생성 기능 추가('24.3) - (투자) 8천만 달러 규모의 시리즈 B 투자 유치('24.6)
SUNO	- (서비스) 음악 생성 AI 'v3' 공개('24.3), 'v3.5' ('24.5)	
검색엔진	Perplexity	- (서비스) 긴 문서 생성 AI 서비스 'Page' 출시('24.05) 다단계 심층 검색 지원하는 'Pro Search' 출시('24.07), 유료 사용자 대상 쇼핑 기능 추가('24.11) - (투자) 엔비디아, 제프 베이조스 등 투자자로부터 7,360만 달러 유치('24.1) - (제휴) 언론사와 수익 공유 프로그램 공개('24.7)
	MS	- (서비스) 검색엔진 Bing에 GPT-4 결합('24.5), Bing Generative Search 베타 공개('24.7)
	Exa	- (투자) 2,200만 달러 규모의 시리즈 AI 투자 유치('24.7)
챗봇	Character.ai	- (서비스) 다수의 AI 캐릭터가 참가하는 그룹 채팅 기능 출시('23.10), AI 캐릭터와 음성 통화 기능 출시('24.6) - (인수) 구글과 기술 라이선스 및 핵심 인력 영입으로 사실상 인수 계약 체결('24.8)
	Quora	- (서비스) AI 채팅 플랫폼 'Poe' 출시('23.2), AI 챗봇 수익 공유 프로그램 출시('24.4), 앱 제작도구 'Preview' 기능 출시('24.7) - (투자) 7,500만 달러 투자 유치('24.1)
	Google	- (서비스) Google Workspace 용 생성 AI 도구 'Duet AI' 출시('23.3), Google Workspace용 동영상 제작 도구 'Google Vids' 공개('24.4)

가치사슬	기업명	주요 활동 (기술·모델 개발, 투자유치(인수합병), 사업전략(협업제휴), 제품/서비스출시 등)
업무 생산성	Writer	- (서비스) 기업용 AI 앱 제작 도구 'AI Studio' 출시('24.6)
	DeepL	- (서비스) AI 글쓰기 도구 'DeepL Write Pro' 출시('24.4), 기업용 번역 솔루션 'DeepL Pro' 165개국 신규 진출('24.6), 음성 번역 솔루션 'DeepL Voice' 출시('24.11) - (투자) 3억 달러 규모 투자 유치('24.5)
	Regie.ai	- 기업 영업팀 대상 잠재고객 발굴을 위한 'Sales Auto-Pilot' 출시('24.1)
	Top 10 GPTs	- Image Generator(Dall-E 3 기반 이미지 생성), Consensus(학술연구 지원), Logo Creator(Dall-E 기반 로고 생성), Write For Me(작문 지원), Canva(디자인 지원), Cartoonize Yourself(카툰화), Grimoire(코딩 지원), PDF Ai PDF(PDF 기반 채팅), Scholar GPT(학술연구 지원), Humanizer Pro(AI 생성물의 인간화)
AI 모델 플랫폼	OpenAI	- (기반모델) GPT-4('23.3), GPT-4 Turbo('23.11), GPT-4o('24.5), GPT-4o Mini('24.7), o1('24.9), ChatGPT Pro('24.12) - (서비스) ChatGPT Enterprise('23.8), Dall-E 3('23.9), ChatGPT 'Memory 기능 추가('24.2), 맞춤형 챗봇 GPTs 공개('23.11), 챗GPT 검색 공식 출시('24.10) - (인수) 데이터 분석 기업 Rockset 인수('24.6) - (기술) 목소리 복제하는 'Voice Engine' 공개('24.3), GPT-4가 생성한 코드 오류 식별하는 'CriticGPT' 연구('24.7), AI 에이전트 조율하는 'Swarm' 프레임워크('24.10) - (제휴) 미국 AI 안전 연구소와 모델 평가 합의('24.8), 언론사·미디어 Conde Nast, Time, News Corp., FT, Axel Springer, FT 등과 콘텐츠 계약 체결 - (투자) 66억 달러 투자 유치('24.10)
	Anthropic	- (기반모델) Claude 2('23.7), Claude 3('24.3), Claude 3.5 Sonnet('24.6), Claude 3.5 Sonnet 업데이트 및 Claude 3.5 Haiku 출시('24.10) - (서비스) Claude 생성물을 실시간 편집 및 수정하는 'Artifacts' 기능 출시('24.8), 기업용 Claude 구독 요금제 출시('24.9), 개발자용 프롬프트 도구 출시('24.11) - (제휴) 미국 AI 안전 연구소와 모델 평가 합의('24.8), Palantir 및 AWS와 제휴해 미국 국방/정보 부처에 모델 제공('24.11)
	Google (Deepmind)	- (기반모델) PaLM('23.3), Bard('23.3), PaLM2('23.5), 음악 생성 AI Lyria('23.11), Gemini 1.0 및 Imagen 2('23.12), Gemma 및 Gemini 1.5 Pro('24.2), Gemini 1.5 Flash 및 Gemma 2('24.6), Gemma 2 2B('24.7), Imagen 3('24.8) - (서비스) AI 음성비서 'Gemini Live' 출시('24.8), Gemini에 Imagen 3 통합('24.8) - (기술) 동영상 생성 AI 'Lumiere'('24.1), AI 신뢰도 향상 'ASPIRE' 시스템('24.1), 로봇 제스처 생성 AI 'GemEM'('24.2), 반도체 칩 설계 AphaChip('24.10), AI 에이전트 프레임워크 'Talker-Reasoner'('24.10)
	Meta	- (기반모델) Llama 1('23.2), Llama 2('23.7), Code Llama 70B('24.1), Llama 3('24.2), Llama 3.1('24.7), 멀티모달 모델 Llama 3.2('24.9), 동영상 생성 모델 'Meta Movie Gen'('24.10) - (서비스) 자사 서비스에 Meta AI 탑재('24.4), AI 캐릭터 생성 'AI Studio'('24.7) - (기술) 영상에서 누락된 부분 예측해 학습하는 V-JEPA 개발('24.2), 합성 데이터로 훈련한 3D 콘텐츠 생성 모델 'Vfusion3D' 공개('24.8), AI 모델 평가용 'Self-Taught Evaluator' 개발('24.10)
	Microsoft	- (기반모델) Phi-1('23.6), Phi-2('23.12), Orca 2('23.11), Orca-Math('24.3), Phi-3('24.4), Phi-3.5('24.8) - (서비스) Windows 11에 Copilot 도입('23.9), 기업용 MS 365 Copilot('23.11), 개인용 Copilot Pro('24.1), Copilot for Security('24.3), Copilot에 음성대화 및 시각 인식 기능 추가('24.10), 신규 에이전트 공개('23.11), AI 에이전트 생성 프로그램 '코파일럿 스튜디오' 정식 출시('24.11) - (투자) UAE AI 기업 G42에 15억 달러 투자('24.4)
	Amazon	- (기반모델) 'Titan Image Generator V2' 출시('24.8) - (서비스) 기업용 AI 챗봇 'Q' 출시('23.11), 상품 추천하는 AI 챗봇 'Lupus' 출시('24.2), 기업용 앱 생성 AI 도구 'App Studio' 프리뷰 공개('24.7), 판매자 지원 AI 도구 'Amelia' 공개('24.9), AI 기반 쇼핑 가이드 출시('24.10)
	Apple	- (기반모델) 이미지-텍스트 멀티모달 모델 'MM1'('24.3) - (플랫폼) 'Apple Intelligence' 출시('24.10) - (기술) 모바일 화면의 UI를 이해하는 멀티모달 모델 개발('24.4), 멀티모달 모델 훈련용 프레임워크 '4M' 데모 공개('24.6) - (제휴) 오픈AI와 파트너십 발표('24.6)

가치사슬	기업명	주요 활동 (기술·모델 개발, 투자유치(인수합병), 사업전략(협업제휴), 제품/서비스출시 등)
	xAI	- (기반모델) 'Grok' 공개('23.11), Grok 오픈소스화 및 Grok-1.5 공개('24.3), 'Grok-1.5 Vision' 프리뷰 공개('24.4), 이미지 생성 통합한 'Grok-2' 공개('24.8) - (투자) 60억 달러 규모의 시리즈 B 투자 유치('24.5), 50억 달러 투자 유치('24.11)
	Nvidia	- (기반모델) Hugging Face 및 ServiceNow와 코드 생성 AI 'Starcoder 2' 공개('24.2), 텍스트에서 3D 생성하는 Latte3 발표('24.3), 'Nemotron-4 340B' 공개('24.6), Llama-3.1-Nemotron-70B-Instruct' 공개('24.10), 텍스트-오디오 생성 AI 모델 '푸가토' 개발('24.12) - (서비스) AI PC용 'Chat with RTX' 공개('24.2) - (기술) 휴머노이드 로봇용 범용 기반모델 '프로젝트 그루트' 공개('24.3)
	Cohere(캐나다)	- (기반모델) 다국어 LLM 출시('24.2), 기업용 신규 LLM 'Command-R+' 출시 ('24.4), 다국어 LLM 'Aya Expanse' 출시('24.10) - (서비스) 기업용 AI 챗봇 '코랄(Coral)' 출시('23.7) - (투자) 5억 달러 규모의 시리즈 D 투자 유치('24.7)
	IBM	- (기반모델) 기업용 LLM 'Granite' 오픈소스 공개('24.5), 'Granite 3.0' 공개('24.10). - (플랫폼) 기반모델 및 생성 AI 개발 지원하는 'WatsonX' 출시('23.5), AI 관리 플랫폼 'Watsonx Governance' 출시('23.11)
	SalesForce	- (기반모델) 기업용 기반모델 'xGen-Sales' 및 'xLAM' 공개('24.9) - (서비스) 완전 자율형 AI 에이전트 'Einstein Service Agent' 공개('24.7), 로우코드 기반 자율형 AI 에이전트 'AgentForce' 공개('24.9) - (플랫폼) 로우코드 기반 생성 AI CRM 플랫폼 'Einstein 1 Studio' 공개('24.3) - (인수) AI 스타트업 Zoomin 인수('24.9)
	Hugging Face	- (기반모델) 1,000여 명 연구자 커뮤니티가 참여한 빅사이언스 프로젝트의 공동개발로 파라미터 1,760억개 이상의 오픈소스 LLM 공개('22.7), 온디바이스AI용 'SmolLM2' 공개('24.11) - (서비스) 15조 개 토큰의 대규모 오픈소스 데이터셋 공개('24.6), AI 배포 서비스 HUGS 발표('24.10), AI 기반 웹앱 개발 도구 'OpenAI-Gradio' 공개('24.10)
	Databricks	- (기반모델) 매개변수 1,320억 개의 'DBRX' 오픈소스 공개('24.3) - (인수) 자체 AI 모델 구축 지원하는 'Mosaic ML' 인수('24.3) - (제휴) AWS와 생성 AI 기능 제공 제휴 체결('24.10)
	Snowflake	- (기반모델) 기업용 오픈소스 LLM 'Arctic' 공개('24.4) - (플랫폼) LLM 기반 앱 구축 지원하는 완전 관리형 서비스 'Snowflake Cortex' 운영
	Mistral AI (프랑스)	- (기반모델) 'Mistral Large' 공개('24.2) 'Mixtral 8x22B' ('24.4), 코드 생성 모델 'Codestral' ('24.5), 'Mistral Nemo' 및 'Mistral Large 2' ('24.7), 'Pixtral 12B' ('24.9), 온디바이스용 'Les Ministros' ('24.10) 'Pixtral Large'('24.11) - (서비스) 챗봇 서비스 Le Chat 출시('24.2), 유해 콘텐츠 탐지용 'Moderation API' 공개('24.11), 'Le Chat'에 웹 검색, 캔버스, 에이전트, 이미지 생성 기능 추가('24.11) - (투자) Microsoft로부터 1,500만 유로 투자 유치('24.2) 6,400만 달러 투자 유치('24.6) - (제휴) Google과 클라우드 파트너십('23.12)
	G42 (UAE)	- (기반모델) 자회사 Inception, JAIS 공개('23.8), JAIS 70B 공개('24.8) - (투자) Microsoft, 15억 달러 투자로 G42 일부 지분 확보 및 이사회 참여('24.4)
	A21 Labs (이스라엘)	- (기반모델) 'Jamba 1.0' 공개('24.3), 'Jamba 1.5 Mini와 Jamba 1.5 Large' 공개('24.8) - (투자) Google, Nvidia, Intel 등으로부터 2억 8천만 달러 유치('23.11)
	Zhipu.AI (중국)	- (기반모델) ChatGLM2-6B - (서비스) AI 챗봇 - (투자) Tencent와 Alibaba로부터 3억 4,200만 달러 투자 유치('23.10)
	Alibaba(중국)	- (기반모델) LLM 'Qwen-7B'와 이를 대화형으로 미세 조정된 챗봇 'Qwen-7B-Chat'을 오픈소스로 공개('23.8), 동영상 생성 AI 모델 'EMO' 공개('24.2), 'Tongyi Qianwen 2.5' 출시('24.5), 수학 특화 AI 모델 'Qwen2-Math' 오픈소스로 공개('24.8), 20분 이상 영상 분석 가능한 'Qwen-VL' 공개('24.9), Qwen2.5-Coder 공개('24.11), 추론 특화 'Marco-o1' 공개('24.11) - (서비스) 캐릭터 생성, 고객 서비스, 프로그래밍, 의료, 법률, 금융, 디지털 아바타 제작 등 산업별 AI 도구 제공, AI 번역도구 '마르코 MT' 출시('24.10) - (플랫폼) AI 모델 오픈소스 플랫폼 'ModelScope' 공개('22.11)
	Tencent (중국)	- (기반모델) 기업용 LLM 'Hunyuan' 공개('23.9), 'Hunyuan-Turbo' 공개('24.9), Hunyuan-Large 공개('24.11) - (서비스) AI 챗봇 'Yuanbao' 출시('24.5)

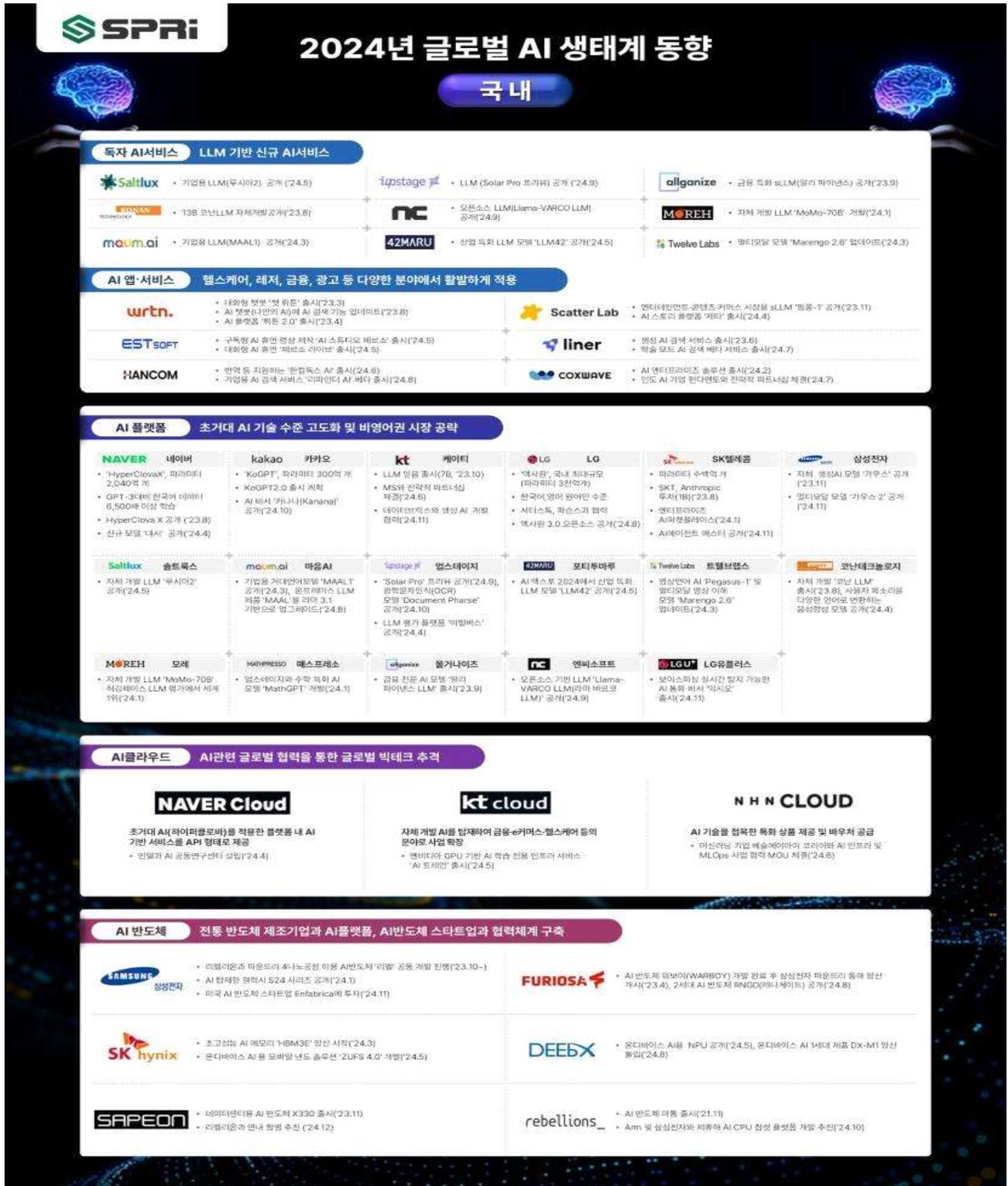
가치사슬	기업명	주요 활동 (기술·모델 개발, 투자유치(인수합병), 사업전략(협업제휴), 제품/서비스출시 등)
	Baidu (중국)	- (기반모델) 'Ernie 3.5' 공개('23.6), 'Ernie 4.0' 공개('23.10), 'Ernie 4.0 Turbo' 공개('24.6) - (서비스) AI 챗봇 'Ernie Bot' 출시('23.3), 'Ernie Bot' 사용자 3억 돌파('24.6) - (플랫폼) 딥러닝 오픈소스 프레임워크 'PaddlePaddle' 운영
	Sensetime (중국)	- (기반모델) 'SenseNova 4.0' 공개('24.2), 'SenseNova 5.0' 공개('24.4), 실시간 멀티모달 모델 'SenseNova 5.5' 공개('24.7) - (서비스) AI 챗봇 'SenseChat' 공개 ('23.4), AI 아바타 동영상 생성 'Vimi' 공개 및 Code Raccoon 업데이트('24.7).
	ByteDance (중국)	- (기반모델) 텍스트-비디오 생성 모델 'MagicVideo-V2' 공개('24.1), 로봇 AI 초거대 모델 'GR-2' 발표('24.10) - (서비스) AI 챗봇 'Dubao' 출시('23.8) 동영상 생성 AI 도구 'Jimeng AI' 앱스토어 출시('24.8), Dubao에 음악 생성 기능 추가('24.8), Dubao에 영상 생성 기능 추가('24.11) - (플랫폼) AI 챗봇 개발 플랫폼 'Coze' 운영 개시*('24.2)
	MInimax (중국)	- (기반모델) LLM 'ABAB 6.5' 출시*('24.4), 동영상 생성 모델 'ABAB-Video-01' 출시('24.9) - (플랫폼) 소비자향 AI 플랫폼 'Hailuo AI' 공개('24.5), Hailuo AI에 이미지-비디오 기능 출시('24.10)
	01.AI (중국)	- (기반모델) 'Yi' ('23.11), 멀티모달 'Yi-VL' ('24.1), 'Yi-1.5' ('24.5), 코딩 특화 LLM 'Yi-Coder-9B'와 'Yi-Coder-1.5B' 공개('24.9), 'Yi-Lightning'('24.10) - (플랫폼) 원스톱 업무용 AI 어시스턴트 'Wanzhi' 출시('24.5)
	Allen AI Institute	- (기반모델) 오픈소스 LLM 'OLMo' 공개('24.2), OLMo 1.7-7B 공개('24.4), Molmo 4중 공개('24.9)
	Inflection	- (기반모델) LLM 'Inflection-2.5' 출시('24.3), Inflection for Enterprise('24.10) - (서비스) AI 챗봇 'Pi 2.5' 공개('24.3) - (기업 매각) Microsoft가 Inflection 인원 대부분 영입하며 6억 5천만 달러 지불
	Hyperwrite	- (기반모델) Meta Llama 3.1 70B 기반 'Reflection 70B' 오픈소스 공개('24.9)
	Covariant	- (기반모델) 로봇용 AI 모델 'RFM-1' 개발('24.3) - (매각) Amazon이 Coveriant 창업자 3명 영입 및 모델에 대한 비독점 라이선스 확보
	사카나 AI(일본)	- (기반모델) 일본어 LLM 'EvoLLM-JP', 'EvoVLM-JP' 공개('24.3) - (투자) 1억 달러 규모의 시리즈 A 투자 유치('24.9) - (기술) AI 모델을 병합해 새로운 모델을 생성하는 방식 연구('24.3), 과학연구 자동화하는 AI Scientist 개발('24.8)
	클라우드	AWS
MS Azure		- (인프라) MS-Azure Open AI API 서비스 제공 참여, Hugging Face의 Azure 활용, Meta의 LLaMA2 활용('23.7)
Google Cloud		- (인프라) 자체 AI 칩(TPU) 기반으로 클라우드 플랫폼 구축하면서 AI컴퓨팅 인프라 구축 - (제휴) Hugging Face와 생성 AI 개발 지원 협력('24.1), Nvidia와 제휴해 Blackwell AI 컴퓨팅 플랫폼 도입 및 Google Cloud에 Nvidia DGX Cloud 서비스 적용('24.3), Oracle과 멀티 클라우드 파트너십 발표('24.6), AI 에이전트 파트너 프로그램 공개('24.11)
IBM		- (제휴) Intel과 제휴해 IBM 클라우드 상에서 AI 가속기 'Gaudi 3' 제공 계획('25년 초)
Alibaba Cloud		- (협업) 고객, 개발자, 파트너 대상 생성 AI 교육과정 제공('24.5) - (서비스) AI 성장 촉진 위해 핵심 클라우드 제품 가격 최대 55% 인하('24.2) - (투자) 중국 AI 스타트업 Moonshot AI에 8억 달러 투자('24.5)
하드웨어	Nvidia	- (제품) 신규 AI 칩 'Blackwell' 공개('24.3) - (인수) 런AI(Run:ai, 이스라엘) 인수('23.4) * AI 및 클라우드 인프라 관리를 위한 SW 플랫폼 구축 - (제휴) Softbank 지원해 일본 최대 AI 슈퍼컴퓨터 구축 추진('24.11)
	Intel	- 이스라엘 AI 반도체 헤너바랩스 인수(20억 달러)를 통해 AI 반도체 시장에 진출 - '23년 47억 달러 이상을 첨단 패키징 설비에 투자 - (제품) AI 칩 'Gaudi 3' 공개('24.4), 차세대 AI 칩 'Lunar lake' 공개('24.9) - (인력) AI PC Acceleration Program 일환 AI PC Developer 프로그램 신설('24.2)

가치사슬	기업명	주요 활동 (기술·모델 개발, 투자유치(인수합병), 사업전략(협업제휴), 제품/서비스출시 등)
	AMD	- (인수) AI SW 기업 Nod.ai 인수 ('23.10), 핀란드 AI 스타트업 Silo AI 인수('24.7), AI 서버 기업 ZT System 인수('24.8) - (제품) 대규모 AI 개발 특화 GPU 'M1300X' 공개('23.6), AI 칩 'Instinct MI300' 시리즈 출시('23.12), 3세대 라이젠 AI 프로세서 등 신제품 공개('24.6), 신규 AI 칩 'MI325X' 공개('24.10)
	Groq	- (제품) AI 추론 칩 LPU 공개('24.2) - (제휴) 삼성전자와 차세대 AI 칩 생산 위한 파트너십 체결('23.8) - (인수) 비즈니스 AI 솔루션 기업 Definitive Intelligence 인수('24.3) - (투자) 삼성전자 주도의 시리즈D 투자 라운드에서 6억 4천만 달러 유치('24.8)
	GraphiCore	- (매각) Softbank에 기업 매각('24.7)
	Tenstorrent	- (제품) AI 추론 특화 PCIe 확장 카드 'Grayskull' e75와 e50 발표('24.3) - (제휴) 삼성전자와 차세대 AI 칩 생산 제휴 체결('23.10)
	Meta	- 자체개발 AI 가속기 'MTIA v1' 공개(*('23.5) 'MTIA v2' 공개('24.4)
	OpenAI	- (제휴) Broadcom 등 AI 칩 설계 기업과 협력 논의('24.7) 자체 AI 칩 개발 위해 '26년 가동하는 TSMC 최첨단 공정 'A16' 예약('24.9), 브로드컴 및 TSMC와 자체 AI 칩 개발 추진('24.11)
	Microsoft	- 자체개발 AI 가속기 'Maia 100'와 CPU 'Cobal 100' 공개('23.11), AI PC 'Copilot Plus PC' 공개('24.5), 데이터센터용 AI 칩 2종 출시('24.11)
	Google	- (제품) ARM 기반 맞춤형 CPU 'Axion' 공개('24.4), 텐서플로우 연산에 최적화된 TPU(Tensor Processing Unit) v6 공개 및 TPU 신제품 'v5p' 정식 버전 출시 ('24.5)
	Cerebras	- (제품) 세계 최고 속도의 AI 칩 'WSE-3' 공개 ('24.3) - (제휴) 생성 시용 컴퓨팅 인프라 제공 위해 Dell과 협력('24.6)
	SambaNova	- (제품) 저비용 구동 가능한 프로세스 'SN40L 출시('23.9)

[첨부 4] 2024년 글로벌 AI 생태계 동향

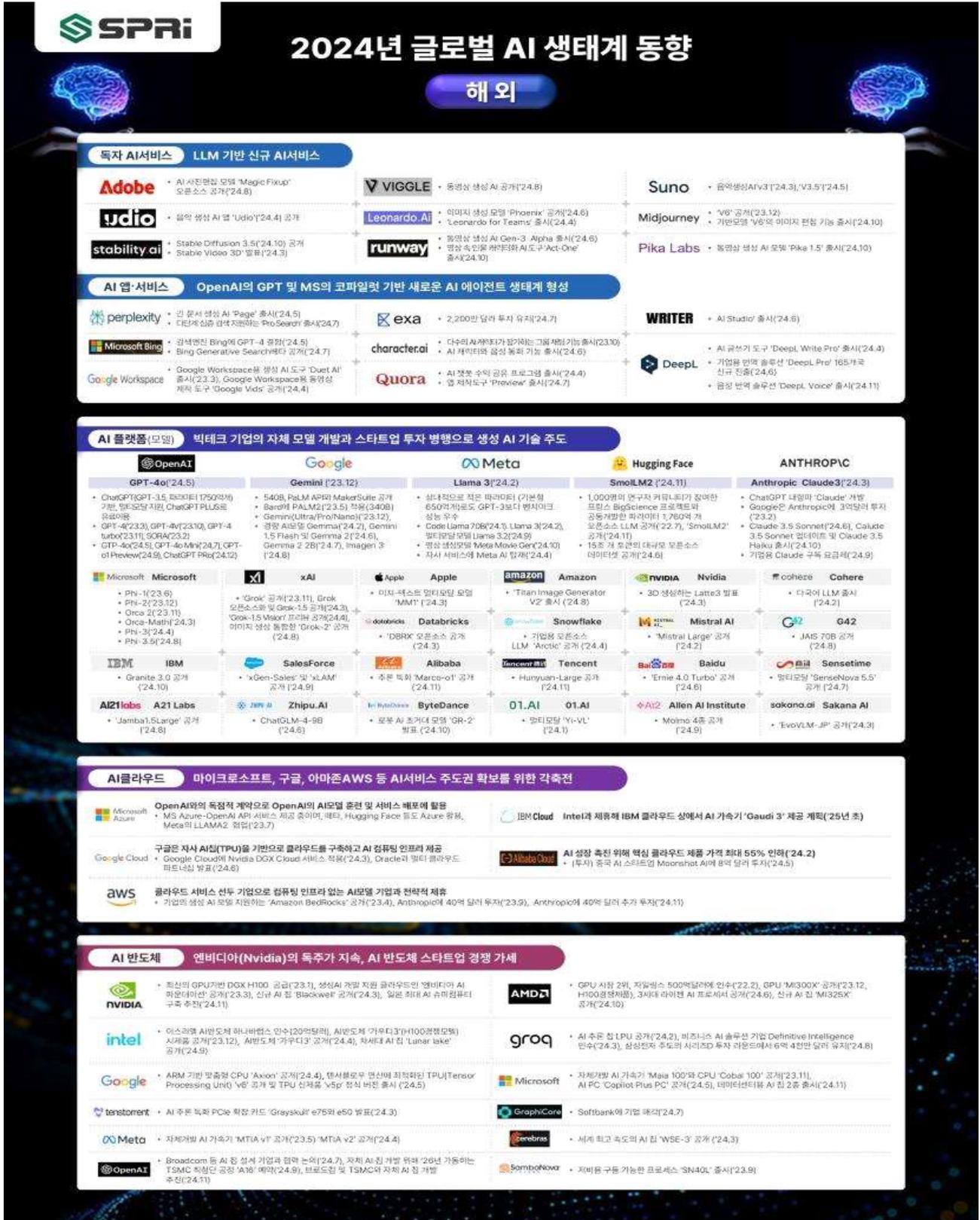
① 국내 AI 기업 동향

[그림] 2024년 글로벌 AI 생태계 동향(국내 AI 기업)



② 해외 AI 기업 동향

[그림] 2024년 글로벌 AI 생태계 동향(해외 AI 기업)



주 의

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



[소프트웨어정책연구소]에 의해 작성된 [SPRI 보고서]는 공공저작물 자유이용허락 표시기준 제 4유형(출처표시-상업적이용금지-변경금지)에 따라 이용할 수 있습니다.
(출처를 밝히면 자유로운 이용이 가능하지만, 영리목적으로 이용할 수 없고, 변경 없이 그대로 이용해야 합니다.)