

기본연구 | 25-14

AI 기술주권과 국가경쟁력 제고 방안 연구

이경선/김성옥/이경은/오장민/윤혜선

2025. 12



기본연구 | 25-14

AI 기술주권과 국가경쟁력 제고 방안 연구

이경선/김성옥/이경은/오장민/윤혜선

2025. 12

서 언

인공지능(Artificial Intelligence, AI)은 이제 개별 산업의 생산성과 효율성을 제고하는 기술을 넘어, 국가의 경제 성장과 산업 구조, 안보와 사회 전반에 영향을 미치는 핵심 전략 자산으로 자리잡고 있습니다. 특히 초거대 파운데이션 모델을 중심으로 한 생성형 AI의 확산과 함께, 에이전틱 AI, 물리적 AI 등 차세대 기술 패러다임이 빠르게 부상하면서 AI를 둘러싼 경쟁은 기술 개발의 차원을 넘어 국가 간 주도권 경쟁으로 심화되고 있습니다. 이러한 환경 속에서 AI 기술을 선도하는 국가가 미래 산업과 글로벌 질서를 주도할 것이라는 인식은 점차 확고해지고 있습니다.

글로벌 패권 경쟁이 심화되는 가운데, 미국·중국·EU 등 주요국은 AI를 국가 전략의 핵심 축으로 설정하고 대규모 투자, 산업 생태계 육성, 규범과 표준 선점을 위한 정책을 본격적으로 추진하고 있습니다. 한편, AI 기술과 데이터, 컴퓨팅 인프라가 소수의 글로벌 빅테크 기업에 집중되면서 기술 종속과 디지털 주권 약화에 대한 우려 또한 커지고 있습니다. 이는 중견국인 우리나라에게 AI 경쟁력 제고와 기술주권 확보라는 과제를 더욱 절실하게 제기하고 있습니다.

이러한 문제의식 속에서 본 보고서는 급변하는 글로벌 AI 기술 및 시장 환경을 종합적으로 분석하고, 주요국의 AI 생태계와 정책·규제 동향을 비교·검토함으로써 우리나라가 AI 기술주권을 확보하고 국가경쟁력을 강화하기 위한 정책적 방향을 모색하고자 하였습니다. 또한 글로벌 지표 분석과 전문가 설문 등을 통해 국내 AI 생태계의 강점과 한계를 진단하고, 이를 바탕으로 AI 기술주권과 국가경쟁력 확보를 위한 주요 쟁점과 정책 과제를 제시하였습니다.

AI를 둘러싼 경쟁은 이제 단일 기술이나 특정 기업의 성과로 결정되지 않으며, 국가 차원의 전략과 정책 선택이 그 성패를 좌우하는 단계에 접어들었습니다. 본 보고서가 정책 결정자와 산업계, 학계, 연구자 등 다양한 독자들에게 글로벌 AI

환경을 이해하는 데 유용한 참고 자료가 되고, 우리나라가 AI를 통해 지속 가능한 성장과 국가경쟁력 강화를 실현하는 데 기여할 수 있기를 기대합니다.

본 연구는 정보통신정책연구원의 이경선 연구위원이 총괄 책임을 맡고 김성옥 연구위원, 이경은 부연구위원이 함께 참여하였으며, 성신여자대학교의 오장민 교수와 한양대학교의 윤혜선 교수가 협동 연구진으로 참여하였습니다. 바쁜 일정 속에서도 연구에 참여해 주신 연구진과 자문위원 여러분께 깊이 감사드립니다. 본 보고서를 계기로 AI 기술주권과 국가경쟁력에 대한 논의가 더욱 폭넓고 심도 있게 이어지기를 기대합니다.

2025년 12월

정보통신정책연구원

원장 이 상 규

목 차

서 언	1
요약문	11
제1장 서론	15
1. 연구배경 및 목적	15
2. 연구내용 및 범위	16
제2장 기술 및 시장 동향	18
제1절 기술동향	18
1. 생성형 AI	18
2. 에이젠틱 AI	27
3. 물리적 AI	37
4. AI 모델 경량화	46
제2절 시장동향	53
1. AI 도입·확산 현황	53
2. AI 시장 동향	56
제3장 주요국 AI 생태계 분석	74
제1절 미 국	74
1. AI 산업생태계 구조 및 주요 플레이어	74
2. AI 정책 및 투자 현황	76

3. AI 규제 현황	84
제2절 중국	92
1. AI 산업생태계 구조 및 주요 플레이어	92
2. AI 정책 및 투자 현황	104
3. AI 규제 현황	109
제3절 E U	116
1. AI 산업생태계 구조 및 주요 플레이어	116
2. AI 정책 및 투자 현황	118
3. AI 규제 현황	123
제4절 한국	132
1. AI 산업생태계 구조 및 주요 플레이어	132
2. AI 정책 및 투자 현황	141
3. AI 규제 현황	145
제4장 국내 AI 생태계 진단	156
제1절 국내 AI 기술자립도 및 경쟁력 진단: 전문가 설문결과를 중심으로 ..	156
1. AI 기술자립도 및 경쟁력 진단	158
2. 차세대 AI 시장의 전략적 중요도 및 주도권 확보 가능성	163
제2절 국내 AI 정책 진단: 전문가 설문결과를 중심으로	166
1. 국내 AI 정책의 타당성·정합성·균형성 진단	166
2. 향후 AI 정책방향 및 과제	171
제5장 정책 제언	191
1. 정책방향	191
2. 실행전략	194

참고문헌	204
Abstract	223

표 목 차

〈표 2-1〉	주요 MCP 서버 레지스트리	33
〈표 2-2〉	AI 패러다임 비교	40
〈표 2-3〉	2025년 주요 모바일 NPU 사양 및 특징 비교	50
〈표 2-4〉	AI칩 시장 주요 플레이어 및 수익화 현황(2024-2025)	57
〈표 2-5〉	AI 클라우드(Compute Layer) 주요 기업 및 특징	60
〈표 2-6〉	Foundation Model 개발 기업	62
〈표 2-7〉	물리적 AI(로봇) 시장 규모 및 성장 전망	67
〈표 2-8〉	물리적 AI(자율주행 모빌리티) 시장 규모 및 성장 전망	68
〈표 2-9〉	주요 해외 기업들의 물리적 AI 제품·서비스 출시 및 개발 현황	70
〈표 2-10〉	국내 기업의 물리적 AI 서비스 개발 동향	72
〈표 3-1〉	AI 액션 플랜 이행을 위한 행정명령	78
〈표 3-2〉	반도체과학법을 통한 주요 지원 사항	81
〈표 3-3〉	미국 정부의 2025년 AI R&D 예산	82
〈표 3-4〉	미국 R&D 예산 및 AI 펀딩의 기능별 분류(2022-2025)	83
〈표 3-5〉	화웨이의 AI 생태계 구조	95
〈표 3-6〉	알리바바의 AI 생태계 구조	98
〈표 3-7〉	바이두의 AI 생태계 구조	100
〈표 3-8〉	중국의 주요 AI 기업 특성	101
〈표 3-9〉	중국 AI 생태계의 협력 구조	103
〈표 3-10〉	차세대 AI 발전계획에 제시된 주요 목표와 응용영역	105
〈표 3-11〉	국내 주요 AI 반도체 스타트업	133

〈표 3-12〉 국내 LLM 추론 성능 비교	140
〈표 3-13〉 국정목표별 정책과제	142
〈표 4-1〉 역할/기능별 AI 기술스택	158
〈표 4-2〉 AI 기술스택별 글로벌 경쟁현황	160
〈표 4-3〉 AI 기술스택별 현황에 대한 전문가 의견	161
〈표 4-4〉 차세대 AI 시장의 전략적 중요도 및 주도권 확보 가능성 평가 이유	164
〈표 4-5〉 AI 정책목표별 국정과제	166
〈표 4-6〉 정책진단 지표	168
〈표 4-7〉 국내 AI 정책에 대한 전문가 의견	169
〈표 4-8〉 AI 정책방향: AI 전문가 설문결과	172
〈표 4-9〉 AI 경쟁력 강화를 위한 핵심 과제: AI 전문가 설문결과	175
〈표 4-10〉 AI 경쟁력 강화를 위해 가장 중요하고도 시급한 과제: 전문가 응답결과	176
〈표 4-11〉 기타 정책제언	181
〈표 4-12〉 모두의 AI 실현을 위한 정책과제: AI 전문가 설문결과	185
〈표 4-13〉 AI 인재양성 방안: 전문가 의견	186

그림 목 차

[그림 2-1]	생성형 언어모델 벤치마크 결과	19
[그림 2-2]	GRPO 개념도	25
[그림 2-3]	GLM-4.1V-Thinking 아키텍처	26
[그림 2-4]	MCP의 아키텍처의 구성 요소	31
[그림 2-5]	MCP와 A2A의 차이	35
[그림 2-6]	AI 서비스 경험률	53
[그림 2-7]	AI 서비스 인식	54
[그림 2-8]	정부 및 연구기관의 AI 도입 사례	55
[그림 2-9]	인간과 AI 사이의 역량 차이	63
[그림 2-10]	에이젠텍 AI 시장 전망	64
[그림 2-11]	물리적 AI 시장 전망(2024-2034)	66
[그림 2-12]	산업용 로봇 시장 전망(2024~2033)	67
[그림 3-1]	글로벌 AI 투자액과 미국 AI 투자액	75
[그림 3-2]	미 연방정부의 IT 및 AI R&D 예산(2021-2025)	82
[그림 3-3]	AI 스택 주요 플레이어	93
[그림 3-4]	HyperCLOVA 모델 라인업	135
[그림 3-5]	HyperCLOVA X SEED Think-14B 성능 테스트 결과	136
[그림 3-6]	EXAONE 4.0-32B 성능그래프	137
[그림 3-7]	카카오의 멀티모달 언어모델 라인업	139
[그림 3-8]	국민성장펀드 지원대상	144
[그림 4-1]	전문가 설문 응답자 소속별 구성	156
[그림 4-2]	전문가 설문 응답자 전문영역별 구성	157

[그림 4-3] AI기술스택별 해외의존도 및 공급망위험도: AI 전문가 설문결과	159
[그림 4-4] AI기술스택별 현재 경쟁력 및 향후 경쟁가능성: AI 전문가 설문결과	163
[그림 4-5] 차세대 AI 시장별 전략적 중요도 및 주도권 확보가능성: AI 전문가 설문결과	164
[그림 4-6] 정책목표별 중요도 및 실행난이도: AI 전문가 설문결과	168
[그림 4-7] 정책목표별 타당성·정합성·균형성: AI 전문가 설문결과	169
[그림 4-8] 모두의 AI 실현의 주요 장애물: AI 전문가 설문결과	184
[그림 5-1] AI 기술스택별 공급망 위험도 vs. 향후 경쟁가능성: AI 전문가 설문결과	192

요 약 문

1. 제 목

AI 기술주권과 국가경쟁력 제고 방안 연구

2. 연구 목적 및 필요성

AI 기술은 이제 다양한 산업 영역에서의 생산성 향상 도구, 경제 성장의 견인차를 넘어 사회경제, 국가안보 전반을 좌우할 게임체인저로 여겨진다. AI를 선도하는 국가가 미래 산업과 글로벌 질서를 주도하게 될 것이라는 전망이 우세하며 이에 AI는 이제 단순한 산업혁신의 도구가 아닌 국가경쟁력을 결정짓는 국가 차원의 핵심 전략 자산으로 자리매김하고 있는 것이다.

국가 차원에서의 AI 경쟁력 확보가 중요해짐에 따라 미국, 중국, EU 등 세계 주요국을 둘러싼 AI 기술패권 경쟁은 나날이 격화되고 있다. 미국을 제외한 세계 주요국들은 미국 빅테크 중심 AI 생태계에 종속되지 않기 위해 AI 기술경쟁력 확보에 박차를 가하는 한편 거버넌스, 규제, 표준 등 다양한 방식으로 AI 시장주도권 경쟁에 참여하며 기술 패권이 특정 국가에 집중되지 않도록 견제하고 있다.

미국 빅테크 중심 시장구조가 고착화되고 글로벌 AI 리더십 확보를 위한 국가간 경쟁이 더욱 치열해지고 있으나 아직 AI 생태계가 완성형은 아니며 AI 생태계 및 경쟁구도에 변화를 가져올 다양한 변수들이 등장하고 있다는 점은 긍정적이다. 중국의 기술 굴기는 현재 미국 중심 시장 구도를 흔들 다양한 혁신으로 이어지고 있으며 온디바이스 AI를 가능하게 할 경량화 모델, 실행력을 갖춘 에이전틱 AI, 다양한 하드웨어와 AI와의 융합을 이끌 물리적 AI 등의 부상 또한 향후 AI 생태

계를 다양한 산업으로 연결, 확장시키며 새로운 강자의 등장, 다양한 기업들의 시장 진입을 가능하게 할 잠재력을 지닌 것으로 보인다.

이제 AI 기술은 키워드로 소비되던 시절을 지나 우리 사회경제 전반에 걸쳐 실질적이고도 거대한 변화를 가져올 구조적 전환의 동력으로 부상하고 있다. 거대해진 가능성만큼이나 기업 간, 국가 간 격차가 커지고 있다는 점은 AI 주도권 확보를 위해 국가적 역량, 자원의 집중이 필요하다는 것을 강조하나, 기술생태계가 빠르게 진화하고 있으며 AI 기술력 확보에 국가도 감당하기 힘든 엄청난 자본력이 필요하다는 점에서 매우 전략적인 접근이 필요하다. 이에 본 연구에서는 AI를 둘러싼 기술 및 정책환경 변화, 주요국 대응 현황 등을 면밀히 분석하고 우리 AI 생태계를 진단하여 AI 기술주권 확보와 국가경쟁력 강화를 위한 전략적 방향성을 제시한다.

3. 연구 내용 및 결과

본 연구에서는 AI 기술·시장 동향과 주요국 AI 생태계를 비교분석하고, AI 기술자립도(해외의존도, 공급망 위험도), 우리 AI 산업의 현재 경쟁력 및 향후 경쟁 가능성, 우리 AI 정책의 정합성 등을 진단하여 AI 기술주권과 국가경쟁력 제고를 위한 정책방향과 실행전략을 제시하였다.

주요 연구결과는 다음과 같다. 중국, EU 등 세계 주요국은 풀스택 관점에서 AI 경쟁력 및 자원 간 연결성을 강화·확장하고 공급망 전반에 걸친 통제권을 확보하는 등 각자의 방식으로 AI 기술주권 확립을 위한 노력을 추진 중이다. 주요국 사례에서 볼 수 있듯이 AI 모든 기술 스택에서의 기술 자립만이 AI 기술주권을 확보하는 유일한 방법은 아니다. 게다가 미국이 엄청난 기술격차를 과시하며 사실상 독점하고 있는 현재의 AI 생태계에서 그동안 AI 기술 스택들이 서로 유기적으로 연결, 결합되어 공진화해왔다는 점을 고려하면 풀스택 기술자립은 적어도 단기간에 도달가능한 목표는 아니다. AI 전문가들 또한, 정부는 우리가 경쟁가능한 핵심

기술스택에 자원을 집중하여 AI 경쟁력을 확보해 나가야 한다고 강조한다. 즉, AI 주권확보의 목표는 완전한 자립 생태계라는 이상적 목표보다는, 핵심 기술 분야에서 외부 압력에 흔들리지 않는 자율성을 확보하는 '최소한의 종속과 최대한의 경쟁력'이 더 타당할 것으로 판단된다.

각 기술스택에 대한 정책적 우선순위는 AI 기술자립도를 진단한 결과를 공급망 위험도, 향후 경쟁가능성을 중심으로 재구성해보면 가늠해 볼 수 있다. 첫째, 공급망 위험도가 높으나 향후 우리의 경쟁가능성이 낮은 영역은 기술주권 측면에서 정부 주도로 공급망 위험 대응을 위한 중장기적 노력이 필요한 영역으로 반도체/가속기, 시스템SW가 이에 속한다. 둘째, 공급망 위험도, 향후 경쟁가능성 둘 다 높은 영역은 기술주권뿐 아니라 산업경쟁력 확보 차원에서 정부 주도의 집중 투자가 필요한 영역으로 AI 모델, 클라우드가 이에 속한다. 세 번째, 공급망 위험도는 상대적으로 낮고 향후 경쟁가능성은 높은 영역은 민간 주도로 경쟁력을 확보해 나가되 산업경쟁력 확보를 위한 정부의 지원이 필요한 영역으로 응용SW, 네트워크, 데이터인프라/솔루션이 이에 속한다. 마지막으로 공급망 위험도, 향후 경쟁가능성이 낮은 영역은 정책적 우선순위가 가장 낮은 영역으로 개발플랫폼/프레임워크가 이에 속한다. 그러나 물론 글로벌 AI 생태계 상황은 언제든지 바뀔 수 있다는 점에서 이 영역에 속한 기술스택에 대해서도 공급망 위험대응 측면에서 지속적인 모니터링은 필요하다. 다시 정리해보면, 단기적으로는 AI 모델, 클라우드를 포함하는 데이터·AI모델 학습 생태계에 대한 경쟁력 확보가 필요하며 중장기적으로는 AI 공급망 위험도를 낮추고 우리 AI 생태계가 우리 기술에 기반한 반도체·시스템SW로 연결·확장되어갈 수 있도록 반도체/가속기, 시스템 SW에 대한 지속적인 개발 투자, 그리고 기술스택간 연동·확장을 위한 생태계 설계가 필요하다.

구체적인 실행전략으로는 1) 대규모 AI 모델에서의 글로벌 격차 해소 및 범용-도메인 모델이 상호발전하는 생태계 설계, 2) 독자 AI 모델과 NPU, HW-SW를 잇는 통합 경쟁력 확보, 3) AI 서비스 생태계 활성화로 AI 생태계의 수요기반 자생적 성장 기반 구축, 4) 우리의 전략산업과 연계된 '버티컬 AI', '물리적 AI'로

AI를 진짜 산업으로 육성, 5) 혁신과 안전의 균형있는 성장을 위한 AI 거버넌스 강화, 6) AI 기술한계 극복에 기여하는 도전적 연구생태계 구축을 제안한다.

4. 기대효과

본 연구는 AI 정책수립에 필요한 기초자료를 제공하고 AI 기술주권 및 산업경쟁력 확보를 위한 전략적 우선순위와 구체적인 실행전략을 제시함으로써 현 정부의 핵심 과제인 AI 3강 실현에 기여할 것으로 기대된다.

제1장 서론

1. 연구배경 및 목적

인공지능(Artificial Intelligence: AI) 기술은 산업과 사회 전반에 걸쳐 전례 없는 속도로 확산되고 있다. OpenAI의 GPT-5, Google의 Gemini 2.5 Pro, Anthropic의 Claude 4 등 프론티어 모델들이 인간 전문가 수준의 추론 능력을 선보이고 있으며, 중국발 오픈 소스 모델의 급격한 성장은 글로벌 AI 기술 지형에 새로운 경쟁 구도를 형성하고 있다. 동시에 AI 기술은 디지털 영역을 넘어 로봇, 자율주행, 제조 자동화 등 물리적 세계로 그 적용 범위를 빠르게 확장하고 있다.

AI 기술은 이제 다양한 산업 영역에서의 생산성 향상 도구, 경제 성장의 견인차를 넘어 사회경제, 국가안보 전반을 좌우할 게임체인저로 여겨진다. AI를 선도하는 국가가 미래 산업과 글로벌 질서를 주도하게 될 것이라는 전망이 우세하며 이에 AI는 이제 단순한 산업혁신의 도구가 아닌 국가경쟁력을 결정짓는 국가 차원의 핵심 전략 자산으로 자리매김하고 있는 것이다.

국가 차원에서의 AI 경쟁력 확보가 중요해짐에 따라 미국, 중국, EU 등 세계 주요국을 둘러싼 AI 기술패권 경쟁은 날이 격화되고 있다. AI 기술 확보에 막대한 자금이 요구되며 이에 AI 기술력이 천문학적 자금을 가진 소수 글로벌 빅테크에게 집중되고 있는 현 상황은 많은 나라들에게 산업경쟁력뿐 아니라 데이터 안보, 기술 주권에 대한 위협으로 다가오고 있다. 미국을 제외한 세계 주요국들은 미국 빅테크 중심 AI 생태계에 종속되지 않기 위해 AI 기술경쟁력 확보에 박차를 가하는 한편 거버넌스, 규제, 표준 등 다양한 방식으로 AI 시장주도권 경쟁에 참여하며 기술 패권이 특정 국가에 집중되지 않도록 견제하고 있다.

이러한 노력에 우리나라 또한 예외는 아니다. 특히 파운데이션 모델을 기반으로 한 AI 패러다임 전환 이후 우리 기업들이 글로벌 AI 경쟁에서 고전하고 있다는

점은 우리의 위기의식을 높이며 기술주권 확보와 AI 산업 발전을 위한 우리 정부의 정책적 노력 강화로 이어지고 있다.

미국 빅테크 중심 시장구조가 고착화되고 글로벌 AI 리더십 확보를 위한 국가 간 경쟁이 더욱 치열해지고 있으나 아직 AI 생태계가 완성형은 아니며 또한 가까운 미래에 AI 생태계와 현 AI 경쟁구도에 변화를 가져올 다양한 변수들이 등장하고 있다는 점은 긍정적이다. 미국 우선주위를 내세우는 트럼프의 재집권으로 미국이 자국 기업들의 AI 우위 선점을 위해 더욱 강력한 보호정책, 미국만을 위한 단독 행보를 강화해나갈 가능성이 높아지고 있으나, 이러한 미국의 강력한 견제와 제재는 역설적으로 중국의 기술 굴기를 가속화시키며 현재 미국 중심 시장 구도를 흔들 다양한 혁신으로 이어지고 있다. 온디바이스 AI를 가능하게 할 경량화 모델, 실행력을 갖춘 에이전틱 AI, 다양한 하드웨어와 AI와의 융합을 이끌 물리적 AI 등의 부상 또한 향후 AI 생태계를 다양한 산업으로 연결, 확장시키며 새로운 강자의 등장, 다양한 기업들의 시장 진입을 가능하게 할 잠재력을 지닌 것으로 보인다.

이제 AI 기술은 키워드로 소비되던 시절을 지나 우리 사회경제 전반에 걸쳐 실질적이고도 거대한 변화를 가져올 구조적 전환의 동력으로 부상하고 있다. 거대해진 가능성만큼이나 기업 간, 국가 간 격차가 커지고 있다는 점은 AI 주도권 확보를 위해 국가적 역량, 자원의 집중이 필요하다는 것을 강조하나, 기술생태계가 빠르게 진화하고 있으며 AI 기술력 확보에 국가도 감당하기 힘든 엄청난 자본력이 필요하다는 점에서 매우 전략적인 접근이 필요하다. 이에 본 연구에서는 AI를 둘러싼 기술 및 정책환경 변화, 주요국 대응 현황 등을 면밀히 분석하고 우리 AI 생태계를 진단하여 AI 기술주권 확보와 국가경쟁력 강화를 위한 전략적 방향성을 제시한다.

2. 연구내용 및 범위

본 연구는 총 5장으로 구성되어 있으며 서론을 제외한 각 장의 주요 내용은 다

음과 같다. 제 2장에서는 AI 기술과 시장의 관점에서 국내외 동향을 분석한다. 제 3장에서는 주요국 AI 산업생태계 및 AI 기술패권 시대 글로벌 AI 리더십 확보를 위한 주요국의 위치와 전략을 분석한다. 제 4장에서는 AI 기술전문가 설문 내용을 중심으로 국내 AI 생태계를 진단한다. 마지막 장에서는 주요 연구결과를 종합하여 AI 기술 주권과 국가경쟁력 확보를 위한 정책방향과 실행전략을 제시한다.

제 2 장 기술 및 시장 동향

본 장은 두 개의 절로 구성되어 있으며 각각 기술, 시장의 관점에서 국내외 동향을 서술한다.

제 1 절 기술동향

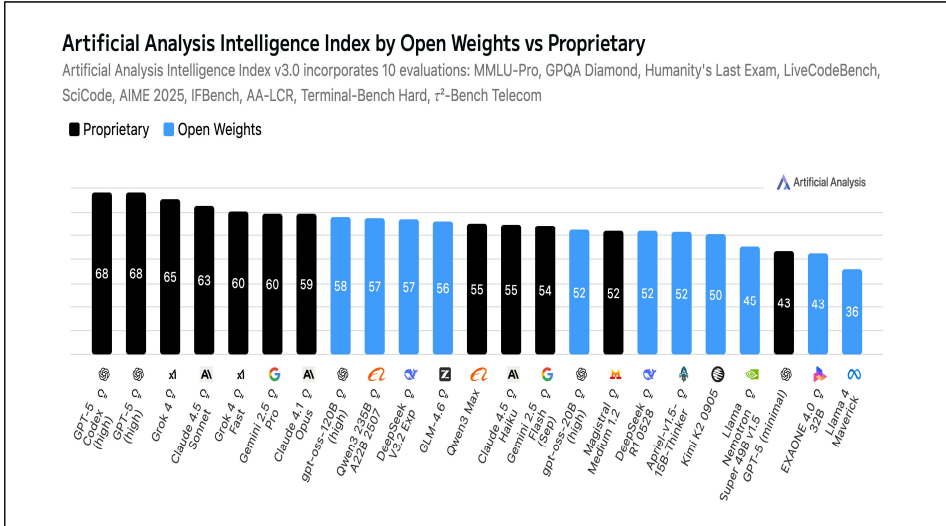
본 절에서는 현재 AI 기술 트렌드 변화를 이끌고 있는 생성형 AI, 에이전틱 AI, 물리적 AI(Physical AI), AI 모델 경량화에 대해 살펴본다.

1. 생성형 AI

가. 전체 현황

2025년 생성형 AI 분야는 글로벌 빅테크 기업들의 치열한 경쟁 속에서 급속한 발전을 거듭하고 있다. ArtificialAnalysis.AI의 2025년 10월 기준 벤치마크 결과에 따르면, OpenAI, Google, Anthropic, xAI 등 주요 빅테크 기업들이 다양한 평가 지표에서 최상위권을 형성하고 있으며, 특히 추론 능력과 멀티모달 처리 기술에서 괄목할 만한 진전을 보이고 있다. AI를 둘러싼 경쟁은 현재 단순한 성능 향상을 넘어 AI 기술의 실질적인 활용 가능성과 산업적 파급력이 확대되는 방향으로 전개되고 있다.

[그림 2-1] 생성형 언어모델 벤치마크 결과



자료: ArtificialAnalysis.ai (2025년 10월 14일 검색)

1) 주요 기업 동향

① OpenAI

OpenAI는 생성형 AI 분야 선도 기업으로서 지속적인 혁신을 추구하고 있다. 2024년 5월에 출시된 GPT-4o는 ‘Omni(전방위)’라는 명칭에 걸맞게 텍스트, 이미지, 오디오를 네이티브로 처리하는 진정한 멀티모달 모델로 평가받고 있다. 이 모델은 128,000 토큰의 컨텍스트 윈도우를 지원하며, 50개 이상의 언어를 처리할 수 있어 글로벌 시장에서의 활용성을 대폭 향상시켰다.

OpenAI의 또 다른 주요 성과는 추론 특화 모델인 o-시리즈의 개발이다. 2024년 9월에 출시된 o1은 Chain-of-thought 추론에 최적화된 최초의 모델로 AI가 복잡한 문제를 단계적으로 분석하고 해결하는 능력을 획기적으로 향상시켰다. 이후 2025년 1월부터 4월 사이에 출시된 o3/o3-mini 모델들은 강화학습을 통해 ‘생각하는 능력’을 더욱 고도화하여, 수학적 추론, 코딩, 과학적 분석 등의 영역에서 인간 전문가 수준에 근접한 성과를 보여주고 있다.

2025년 8월에 출시된 GPT-5는 OpenAI의 기술적 집대성이라 할 수 있다. 이 모델은 기존 GPT 시리즈와 o-시리즈를 단일 시스템으로 통합한 모델로, 실시간 라우터가 사용자의 특성과 질문에 따라 빠른 응답 모드와 깊은 사고 모드를 자동으로 선택하는 혁신적인 아키텍처를 도입하였다. 2025년 10월 GPT-5는 대부분의 벤치마크에서 최고 성능을 기록하며 업계 표준으로 자리매김하고 있다.

② Google

Google은 자사의 방대한 컴퓨팅 인프라와 연구 역량을 바탕으로 Gemini 시리즈를 통해 생성형 AI 시장에서 강력한 경쟁력을 확보하고 있다. 2024년 12월에 출시된 Gemini 2.0 Flash는 네이티브 이미지 및 오디오 출력을 지원하며, 실시간 비디오 및 오디오 상호작용을 위한 Multimodal Live API를 제공한다. 이를 통해 사용자들은 AI와 더욱 자연스럽게 직관적인 방식으로 소통할 수 있게 되었으며, 특히 실시간 응용 프로그램 개발에 있어 새로운 가능성이 열리게 되었다.

2025년 3월에 출시된 Gemini 2.5 Pro는 컨텍스트 윈도우 확장의 새로운 이정표를 세웠다. 이 모델은 1,000,000 토큰 이상의 입력 컨텍스트를 지원하여 업계 최대 규모를 자랑하며, 65,536 토큰에 달하는 출력 용량 역시 업계 최고 수준이다. 특히 'Deep Think' 모드를 통해 복잡한 추론 작업을 수행할 수 있으며, 텍스트, 이미지, 오디오, 비디오 등 다양한 멀티모달 입력을 통합적으로 처리할 수 있다. 이러한 대용량 컨텍스트 처리 능력은 장문의 문서 분석, 코드베이스 전체 이해, 복잡한 대화 맥락 유지 등의 영역에서 획기적인 성능 향상을 가져왔다.

③ Anthropic

Anthropic은 AI 안전성 연구를 기반으로 설립된 기업으로서, Claude 시리즈를 통해 생성형 AI 시장에서 독자적인 위치를 확보하고 있다. 2025년 5월에 출시된 Claude 4는 '듀얼 모드 추론' 아키텍처를 도입하여 간단한 쿼리에는 즉각적으로 응답하고, 복잡한 문제에 대해서는 확장된 사고 모드로 전환하여 심층적 분석을 수행한다. 이러한 유연한 접근 방식은 사용자 경험과 컴퓨팅 효율성을 동시에 최적화하는 데 기여하고 있다.

특히 Claude 4는 코딩 분야에서 최고 수준의 성능을 보여주며, 복잡한 작업 및 에이전트 워크플로우 수행에 강점을 가지고 있다. 2025년 9월과 10월에 각각 출시된 Claude Sonnet 4.5와 Haiku 4.5는 보다 향상된 코딩 능력과 함께 복잡한 에이전트 구축 및 컴퓨터 사용에 최적화된 특성을 갖추고 있다. 또한, AI가 사용자의 컴퓨터를 직접 조작하는 Anthropic의 ‘Computer Use’ 기능은 AI가 단순 답변자가 아닌 행동 주체가 되게 함으로써 에이전트 시대로의 전환을 이끌고 있다.

④ xAI(Grok)

일론 머스크가 설립한 xAI는 Grok 시리즈를 통해 생성형 AI 시장에 도전장을 내밀었다. 2025년 2월 출시된 Grok 3은 대규모 컴퓨팅 인프라를 기반으로 빠른 개발 주기를 확보한 것이 특징이다. xAI는 약 200,000개 GPU 규모의 컴퓨팅 클러스터를 구축하여 경쟁사 대비 빠른 모델 개발과 출시를 가능하게 하였으며, 이는 AI 개발에 있어 컴퓨팅 인프라의 중요성을 다시 한번 입증하는 사례가 되고 있다.

Grok 3은 ‘Think’ 모드와 ‘Big Brain’ 모드를 지원하여 복잡한 문제 해결 시 추가적인 컴퓨팅 리소스를 동적으로 할당할 수 있다. 이러한 접근 방식은 사용자가 필요에 따라 응답 품질과 처리 시간 사이의 균형을 조절할 수 있게 하며, 특히 수학적 추론, 과학적 분석, 복잡한 코딩 작업 등에서 우수한 성능을 발휘한다. xAI의 빠른 성장은 기존 빅테크 기업들의 AI 패권에 새로운 도전이 되고 있으며, 시장 경쟁을 더욱 치열하게 만드는 요인으로 작용하고 있다.

2) 공통 트렌드 및 주요 인사이트

① 강화학습을 이용한 추론 능력 강화

2025년 생성형 AI의 가장 두드러진 트렌드는 사후 훈련(post-training) 단계에서 강화학습을 기반으로 추론의 깊이를 더해 복잡한 문제를 해결하는 능력을 향상시킨 것이다. OpenAI의 Thinking 모드, Google의 Deep Think 모드, Anthropic의 듀얼 모드 추론, xAI의 BigBrain 모드 등 각 기업들이 유사한 접근 방식을 채택하고 있다는 점은 이러한 기술적 방향성이 업계 전반의 합의에 기반하고 있음을 보여준다. 이러한 추론 능력 강화는 AI가 단순한 패턴 매칭을 넘어

진정한 문제 해결 능력을 갖추는 데 핵심적인 역할을 하고 있다.

② Mixture of Expert (MoE) 아키텍처의 부상

OpenAI, Google, xAI 등 주요 기업들은 MoE 아키텍처를 채택하는 것으로 알려져 있다. MoE 아키텍처는 하나의 처리 모듈 대신 여러 개의 전문 모듈 (expert)를 두고, 입력 토큰마다 라우터가 일부 전문 모듈만 선택적으로 활성화하는 조건부 연산 구조이다. 이러한 구조는 모델의 정교함과 효율성을 동시에 추구할 수 있게 해주며, 컴퓨팅 비용을 효과적으로 관리하면서도 높은 성능을 유지할 수 있게 해준다.

③ 컨텍스트 윈도우 확장 경쟁

컨텍스트 윈도우의 확장은 생성형 AI 기업들 사이에서 중요한 경쟁 영역으로 부상하였다. Google Gemini 2.5 Pro의 1백만 토큰, Anthropic Claude Sonnet 4.5의 200,000 토큰, OpenAI GPT-4의 192,000 토큰, xAI Grok 3의 131,072 토큰 등 각 기업들은 장문 처리 능력을 지속적으로 향상시키고 있다. 컨텍스트 윈도우의 확장은 인컨텍스트 학습(In-Context Learning)과 추론 능력 향상 등 전반적인 성능 향상을 견인하는 핵심 요소로 작용하고 있으며, 이를 통해 AI는 더 복잡하고 긴 문서를 이해하고 분석할 수 있게 되었다.

④ 에이전트 시대로의 전환

생성형 AI는 단순한 챗봇 역할을 넘어 자율적으로 작업을 수행하는 에이전트로 진화하고 있다. Google의 Project Mariner와 Anthropic의 Computer Use 등은 AI가 사용자를 대신하여 웹 브라우징, 파일 관리, 애플리케이션 조작 등의 작업을 수행할 수 있음을 보여주는 대표적인 사례이다. 이러한 발전은 AI를 단순한 대화 도구에서 통합 지능 시스템으로 변화시키고 있으며, 향후 기업 운영과 개인의 생산성에 혁신적인 변화를 가져올 것으로 예상된다.

⑤ 멀티모달 고도화 및 모델 통합

생성형 AI 모델들은 텍스트뿐만 아니라 이미지, 오디오, 비디오를 단일 모델에

서 처리하는 진정한 멀티모달 모델로 발전하고 있다. 과거에는 각각의 모달리티를 처리하기 위해 별도의 모델이 필요했으나, 현재는 하나의 통합 모델이 다양한 입출력을 처리할 수 있게 되었다. 이러한 통합은 사용자 경험을 개선하면서도 모델 선택의 부담을 줄이는 방향으로 진화하고 있으며 개발자들에게도 더욱 직관적인 API 환경을 제공하고 있다.

나. 오픈 소스 모델의 약진

2024년 상반기까지만 해도 막대한 자본력을 보유한 빅테크의 폐쇄형 모델과 오픈 소스 모델 사이의 성능 격차는 지속적으로 확대되는 추세였다. 특히 추론 성능의 비약적인 발전에 대해 빅테크 기업들의 공식 논문이나 기술 공개가 이루어지지 않는 상황에서, 오픈 소스 진영은 관련 논문과 엔지니어링 노하우의 공유를 통해 훈련 전략을 추측하고 발전시켜 나갔다.

이러한 상황은 2024년 하반기와 2025년 1월 중국의 DeepSeek에서 공개한 추론 성능 향상 훈련 기술 보고서를 계기로 크게 전환되었다. DeepSeek의 기술 공개로 인해 빅테크가 사용하던 것으로 추정되는 훈련 레시피가 구체적으로 확인되었으며, 이후 오픈 소스 진영의 모델 성능은 급격히 향상되기 시작했다. 2025년 상반기부터 발표되는 오픈 소스 모델들은 일부 벤치마크에서 빅테크의 최상위 모델을 앞서는 성과를 보이며 과열 경쟁 양상을 띠고 있다. 주요 모델을 살펴보면 다음과 같다.

1) DeepSeek-R1

DeepSeek-R1은 중국의 DeepSeek사가 2025년 1월에 MIT 라이선스¹⁾로 공개한 오픈 소스 모델이다(DeepSeek-AI et al. 2025). 이 모델은 617B-A37B의 규모를 가지고 있는데, 이는 전체 6,170억 개 파라미터 규모의 모델 중 추론 시 실제로 활성화되는 파라미터가 370억 개 규모임을 의미한다. DeepSeek-R1은

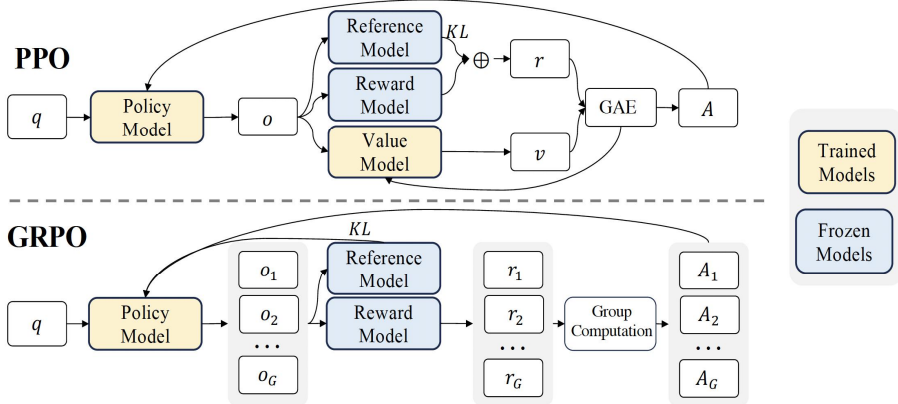
1) 저작권 고지와 라이선스 문구를 함께 포함해야 한다는 조건 하에서 소프트웨어를 자유롭게 사용·수정·배포·상업적으로 활용할 수 있도록 허용하는 오픈소스 라이선스

출시 당시 기준으로 OpenAI-o1과 대등한 추론 성능을 구현하였으며, 무엇보다 중요한 것은 이러한 성능을 달성하기 위한 훈련 레시피를 최초로 공개했다는 점이다.

DeepSeek-R1은 텍스트만을 처리하며 128,000 토큰의 컨텍스트 윈도우를 지원한다. 아키텍처 측면에서 이 모델은 MoE 구조를 채택하여 256개의 전문가(Expert) 중 토큰 생성 시 8개만 활성화되는 효율적인 방식을 사용한다. 또한 Multi-Head Latent Attention(MLA)을 통해 메모리와 계산 비용이 많이 소요되는 어텐션 구조를 단순화하였으며, Multi-Token Prediction(MTP)을 도입하여 다음 스텝의 토큰 예측과 더불어 추가 스텝의 토큰까지 예측하는 새로운 모델링 방식을 제안하였다.

DeepSeek-R1의 핵심적인 기술적 기여는 추론 훈련을 위한 강화학습 레시피인 Grouped Relative Policy Optimization(GRPO)의 공개이다. GRPO는 전통적인 강화학습 방법론인 Proximal Policy Optimization(PPO) 대비 불필요한 모델 사용을 제거하고, 대신 추론 및 정답 생성을 G회 실시하여 그룹을 구성하는 방식을 채택한다. 그룹의 생성 내용을 평가하는 리워드(reward)는 정답이 일치하는지, 추론 과정에서 필요한 포맷이 유지되는지 등을 기계적으로 정량화할 수 있도록 정의된다. 이후 그룹 리워드의 통계를 계산하고, 평균 대비 상대적으로 우수한 정도(advantage)를 강화학습 시그널로 사용한다. 이러한 단순하지만 자동화된 설계로 인해 엔드-투-엔드(end-to-end) 훈련을 통해 최고 수준의 추론 성능에 도달할 수 있는 레시피가 완성되었으며, 이후 공개되는 오픈 소스 모델들은 DeepSeek의 레시피를 채용 및 발전시켜 빅테크의 폐쇄형 모델과의 간극을 좁혀가고 있다.

[그림 2-2] GRPO 개념도



출처: Shao et al. (2024)

2) GLM-4.5

GLM-4.5는 중국의 Zhipu AI가 2025년 7월에 MIT 라이선스로 공개한 오픈 소스 모델이다(GLM-4.5 Team 2025). 모델 규모는 355B-A32B로, 전체 3,550억 개 파라미터 규모 중 활성화되는 모델은 320억 개 규모이다. 이 모델은 에이전트 작업, 추론, 코딩에 특화된 복합 추론에 집중하여 훈련되었으며, 하이브리드 모델로서 Thinking과 Direct 응답 방식으로 동작할 수 있다. 컨텍스트 윈도우는 128,000 토큰을 지원한다.

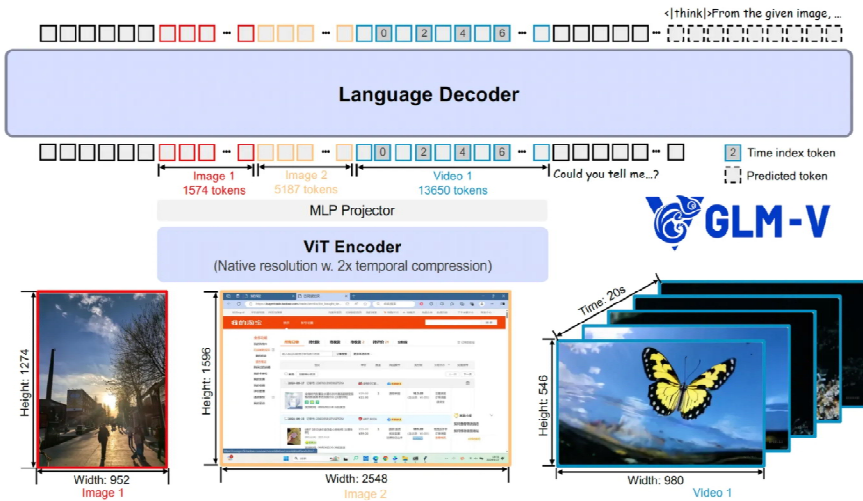
GLM-4.5의 주요 의의는 DeepSeek-R1과 유사한 아키텍처를 채택하면서도 보다 적은 규모의 파라미터 수로 주요 벤치마크에서 개선된 성능을 달성했다는 점이다. 이는 모델 아키텍처와 훈련 방법론의 최적화를 통해 효율성과 성능을 동시에 향상시킬 수 있음을 보여주는 사례로 오픈 소스 생태계의 기술적 성숙도가 빠르게 높아지고 있음을 입증한다.

3) GLM-4.1V-Thinking

GLM-4.1V-Thinking은 Zhipu AI가 2025년 7월에 MIT 라이선스로 공개한 멀티모달 추론 모델이다(GLM-V Team 2025a). GLM-4.1V-9B-Thinking(90

억 개 파라미터)과 GLM-4.5V(1,080억 개 파라미터)의 두 가지 버전이 제공된다. 이 모델은 다목적 멀티모달 추론을 목표로 하여 OCR을 통한 수학 문제 풀이, GUI 스크린 캡처로부터 코드 생성, 차트 이해, 비디오 이해 등 다양한 기능을 지원한다. 컨텍스트 윈도우는 32,000 토큰으로 설정되어 있으며, 이미지는 최대 6,144 토큰, 비디오는 최대 30,000 토큰까지 처리할 수 있다.

[그림 2-3] GLM-4.1V-Thinking 아키텍처



자료: GLM-V Team(2025b)

GLM-4.1V-Thinking의 아키텍처는 자사의 추론 성능을 보유한 GLM 언어 모델과 이미지 인코더, 비디오 인코더를 통합한 구조를 가지고 있다. 이미지와 비디오 토큰을 포함하도록 확장된 텍스트 토큰 공간에서 추론 성능을 최적화하는 방식을 채택하였다. 이 모델의 가장 큰 의의는 기존에 빅테크가 크게 앞서 있던 멀티모달 통합 및 추론 모델 분야에서 오픈 소스 모델의 경쟁가능성을 보여주었다는 점이다. 이는 멀티모달 AI 기술의 민주화에 중요한 이정표가 되고 있다.

4) gpt-oss

gpt-oss는 미국의 OpenAI가 2025년 8월에 아파치 2.0 라이선스²⁾(gpt-oss 사

용 정책이 세분화됨)로 공개한 오픈 소스 모델이다(OpenAI 2025). 모델 규모는 gpt-oss-120B-A5.1B로, 전체 1,200억 개 파라미터 규모 중 활성화되는 모델은 51억 개 규모이다. 이 모델은 추론 능력, 도구 사용, 구조화된 출력에 특화되어 있으며, 128,000 토큰의 컨텍스트 윈도우를 지원한다.

gpt-oss의 아키텍처는 DeepSeek-R1, GLM-4.5와 유사하지만, 그 의미는 빅테크인 OpenAI가 고성능의 모델을 오픈 소스화했다는 데 있다. 이는 오픈 소스 진영이 빅테크에 실질적인 위협이 될 수준까지 성장했음을 의미하며, OpenAI가 오픈 소스 생태계와의 경쟁에서 새로운 전략적 대응이 필요하다고 판단했음을 보여준다. gpt-oss의 공개는 생성형 AI 시장의 경쟁 구도에 새로운 변수를 추가하였으며, 향후 오픈 소스와 폐쇄형 모델 간의 경계가 더욱 모호해질 것임을 예고하고 있다.

2. 에이젠틱 AI

가. 에이젠틱 AI의 정의 및 특징

에이젠틱 AI(Agentic AI)란 AI 시스템이 자율적이고 목표 지향적으로 행동하는 특성이나 능력을 의미한다. 이는 여러 정보 소스로부터 데이터를 처리하고, 해당 데이터에 기반한 의사결정을 내리며, 사람의 개입을 최소화한 상태에서 다단계 워크플로우를 실행하는 자율 시스템을 통칭하는 개념이다. 에이젠틱 AI는 단순히 사용자의 명령에 반응하는 기존의 생성형 AI를 넘어서, 목표 지향적으로 스스로 계획하고 실행하는 자율형 AI 시스템으로의 패러다임 전환을 대표한다.

이러한 자율형 AI 시스템으로의 전환은 기업 운영과 산업 전반에 혁신적 변화를 가져올 것으로 예상된다. 과거의 AI가 주로 특정 질문에 대한 답변이나 단일 작업의 수행에 초점을 맞추었다면, 에이젠틱 AI는 복잡한 비즈니스 프로세스 전

2) 소프트웨어의 자유로운 사용·수정·상업적 이용을 허용하되, 저작권 및 라이선스 고지 유지, 변경 사항 명시, NOTICE 파일 포함, 그리고 기여자의 특허 사용 허용을 조건으로 하는 오픈소스 라이선스

체를 이해하고 자율적으로 수행할 수 있는 능력을 갖추고 있다. 이는 인간 작업자의 역할을 실행에서 감독과 전략적 의사결정으로 전환시키는 근본적인 변화를 의미한다.

에이젠틱 AI의 주요 특징을 살펴보면, 첫째, 사전 구축된 모듈(Pre-built modules)의 활용이다. 이는 워크플로우 템플릿, 데이터 패턴 인식 시스템 등 사전에 정의된 트리거와 액션을 포함하는 구성 요소로 개발 과정을 크게 단순화한다. 두 번째 특징인 메모리 관리(Memory management)는 과거의 상호작용, 사용자 선호도, 이전 결정 등을 저장하여 장기적인 작업 전반에 걸쳐 맥락을 기억하는 능력을 의미한다. 이를 통해 에이전트는 연속적인 작업에서 일관성을 유지하고 사용자의 패턴에 맞춤형 서비스를 제공할 수 있다. 세 번째 특징은 다중 에이전트 협업(Multi-agent collaboration)으로, 여러 전문 에이전트가 협력하여 대규모 프로젝트를 수행하며 오케스트레이터(orchestrator)가 작업을 분배하는 구조이다. 네 번째 특징인 계획 및 추론(Planning and reasoning)은 고정된 지시를 따르는 대신 상황을 분석하고 목표를 설정하며 변화하는 환경에 적응하는 전략을 수립하는 능력을 의미한다. 다섯 번째로, 맞춤형 의사결정(Customized decision-making)은 사용자 선호도와 행동 패턴을 분석하여 개인화된 상호작용을 제공하는 것이다. 여섯 번째 특징인 자가 학습(Self-learning)은 결과와 피드백으로부터 학습하여 지속적으로 성능을 개선하는 능력이며, 일곱 번째 특징인 상황 인식(Contextual awareness)은 여러 정보 흐름을 처리하여 현재 상황과 우선순위를 파악하고 적절하게 대응하는 능력을 의미한다. 이러한 특징들의 조합을 통해 에이젠틱 AI는 단순한 도구를 넘어 진정한 디지털 동료(digital colleague)로서 기능할 수 있게 된다.

나. 기업 채택 현황 및 동인

SuperHuman 블로그에 나온 2025년 조사에 따르면, 96%의 기업이 AI 에이전트 사용을 확대하고 있으며, 경쟁력 유지를 위해 에이젠틱 AI 투자를 필수적으로 인식하고 있다. 주요 채택 동인으로는 복잡한 비즈니스 프로세스의 자동화 요구

증가, 실시간 의사결정 능력 향상, 24시간 7일 지속적 운영 가능성, 그리고 인적 오류 감소 및 정확도 향상 등이 꼽힌다. 기업들은 에이젠틱 AI를 통해 반복적이고 시간 소모적인 작업을 자동화하고, 인간 직원들이 보다 창의적이고 전략적인 업무에 집중할 수 있도록 업무 구조를 재편하고 있다.

그러나 Harvard Business Review(HBR, 2025)에 따르면, 실제 에이젠틱 AI의 채택은 아직 초기 단계에 머물러 있다. 주요 위험 요소로는 예상보다 높은 비용 증가, 불분명한 비즈니스 가치, 부적절한 위험 통제 등이 지적되고 있다. 많은 기업들이 에이젠틱 AI의 잠재력에 대해서는 공감하면서도, 실제 도입 과정에서 발생하는 기술적 복잡성과 조직적 저항, 그리고 투자 대비 수익(ROI)의 불확실성으로 인해 신중한 접근을 취하고 있는 것이 현실이다.

다. 에이전트 개발 프레임워크 생태계

에이젠틱 AI 개발을 위한 프레임워크는 단일 에이전트 구축부터 멀티 에이전트 시스템까지 다양한 수준의 추상화와 기능을 제공하며 빠르게 발전하고 있다. 현재 시장에서는 LangChain/LangGraph, Microsoft AutoGen, CrewAI 등이 대표적인 프레임워크로 자리 잡고 있으며, 각각의 특성과 강점에 따라 다양한 용도로 활용되고 있다.

1) LangChain / LangGraph

LangChain은 현재 가장 널리 채택된 에이전트 프레임워크로서, 모듈식 접근 방식과 광범위한 생태계를 제공한다. 다양한 LLM 제공업체, 벡터 데이터베이스, 도구들과의 통합을 지원하여 개발자들이 빠르게 에이전트 시스템을 구축할 수 있도록 돕는다. LangGraph는 LangChain 위에 구축된 확장 프레임워크로서 그래프 기반 오케스트레이션을 제공하며, 장기 실행 상태 저장 에이전트를 관리하는 저수준 프레임워크이다³⁾.

LangChain/LangGraph의 주요 장점으로서는 대규모 커뮤니티 지원, 풍부한 통

3) <https://github.com/langchain-ai/langgraph>

합 옵션, 그리고 인간 개입(human-in-the-loop) 지원 등이 있다. 이러한 특성으로 인해 복잡한 상태 저장 워크플로우, 다중 에이전트 조율, 프로덕션 배포 등의 활용 사례에 특히 적합하다. 활발한 오픈소스 커뮤니티 덕분에 지속적인 기능 개선과 버그 수정이 이루어지고 있으며, 풍부한 문서와 예제 코드가 학습 곡선을 낮추는 데 기여하고 있다.

2) Microsoft AutoGen

Microsoft의 AutoGen은 엔터프라이즈급 안정성과 대화형 멀티 에이전트 시스템 구축에 중점을 둔 프레임워크이다. 에이전트 간 유연한 대화 상호작용을 강조하며, 코드 실행을 위한 보안 컨테이너 환경을 제공하여 기업 환경에서의 안전한 운영을 보장한다. Microsoft의 기업용 소프트웨어 개발 경험이 반영되어 있어 보안, 확장성, 안정성 측면에서 높은 수준을 유지한다.

AutoGen의 주요 장점으로서는 세밀한 제어 기능, 비동기 지원, 그리고 Microsoft 제품군과의 우수한 통합이 있다. Azure, Microsoft 365, Dynamics 365 등 Microsoft 생태계와의 원활한 연동이 가능하며, 이미 Microsoft 제품을 사용하고 있는 기업들에게 특히 매력적인 선택지가 된다. 복잡한 문제 해결, 연구 중심 작업, 코드 생성 및 디버깅 등의 활용 사례에서 강점을 발휘한다.

3) CrewAI

CrewAI는 역할 기반 워크플로우에 특화된 프레임워크로서, 비즈니스 프로세스 자동화에 적합한 사용자 친화적 구조를 제공한다. '크루(Crew)'라는 개념을 중심으로 각 에이전트에게 특정 역할과 목표를 부여하고, 이들이 협력하여 복잡한 작업을 수행하도록 설계되어 있다. 이러한 직관적인 구조 덕분에 기술적 배경이 부족한 사용자들도 비교적 쉽게 에이전트 시스템을 구축할 수 있다.

CrewAI의 주요 장점으로서는 낮은 진입 장벽, 빠른 배포, 명확한 역할 정의 등이 있다. 비즈니스 프로세스 자동화, 협업 워크플로우, 콘텐츠 생성 등의 활용 사례에서 강점을 발휘하며, 특히 마케팅팀, 고객 서비스 부서, 콘텐츠 제작팀 등 비개발자 조직에서의 AI 도입을 용이하게 한다. 다만, 고도로 복잡한 기술적 워크플로

우에서는 LangGraph나 AutoGen에 비해 유연성이 부족할 수 있다.

라. 프로토콜

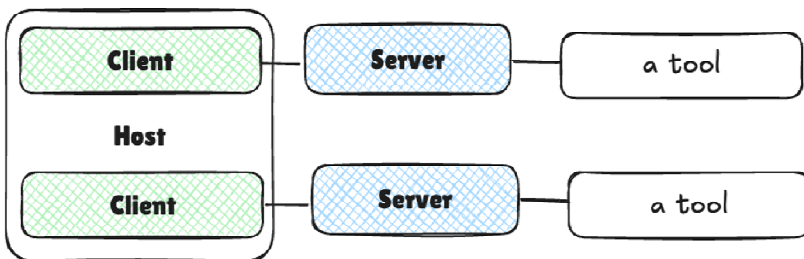
1) Model Context Protocol (MCP)

2024년부터 AI 모델은 외부 도구나 함수를 호출할 수 있는 능력을 갖추기 시작했다. 그러나 도구 조회, 요청, 응답에 대한 표준화가 정립되지 않아 모델과 도구를 대응시키기 위한 개발 복잡도가 크게 증가하는 문제가 발생하였다. 각 AI 제공업체와 도구 개발자들이 서로 다른 방식으로 통합을 구현하면서, 개발자들은 동일한 도구를 여러 AI 모델에 연결하기 위해 중복된 작업을 수행해야 했다.

이러한 문제를 해결하기 위해 Anthropic은 2024년 11월 Model Context Protocol(MCP)이라는 공개 표준 프로토콜을 제안하였다. MCP는 AI 시스템이 외부 도구, 시스템, 데이터 리소스와 통합하는 방식을 표준화함으로써, 개발자들이 한 번의 구현으로 다양한 AI 모델과 호환되는 도구를 만들 수 있도록 지원한다. 이는 마치 USB가 다양한 장치들의 연결 방식을 표준화한 것과 유사한 역할을 AI 도구 통합 영역에서 수행하고자 하는 것이다.

MCP의 아키텍처는 세 가지 핵심 구성 요소로 이루어져 있다. 첫째, 호스트(Host)는 LLM을 실행하고 모든 클라이언트의 연결을 관리하는 중앙 조정자 역할을 수행한다. 호스트는 보안, 권한, 시스템 제어를 담당하며, 전체 MCP 생태계의 핵심 허브로 기능한다.

[그림 2-4] MCP의 아키텍처의 구성 요소



자료: Model Context Protocol 사이트

둘째, 클라이언트(Client)는 호스트와 서버 간 1:1 연결을 유지하는 중개자로서 호스트 안에서 동작하며, 세션 관리와 메시지 라우팅을 수행한다. 각 클라이언트는 특정 서버와의 통신을 담당하며, 연결 상태를 모니터링하고 필요시 재연결을 시도한다. 셋째, 서버(Server)는 도구, 데이터, 리소스를 제공하는 경량 프로그램으로서, MCP 사양을 구현하여 표준화된 방식으로 기능을 노출한다. 서버는 특정 서비스나 데이터 소스에 대한 접근을 제공하며, 예를 들어 데이터베이스 서버, 파일 시스템 서버, 외부 API 서버 등이 이에 해당한다. 이러한 세 구성 요소의 분리는 관심사의 명확한 분리와 독립적인 확장을 가능하게 하여, 복잡한 엔터프라이즈 환경에서도 유연하게 적용될 수 있다.

또한, MCP는 모델과 외부 시스템을 연결하기 위해 네 가지 기능적 구성요소를 가진다. 첫째, 리소스(Resources)는 사용자 또는 모델이 참조하거나 활용할 수 있는 컨텍스트 데이터로, 파일, 데이터베이스 레코드, 실시간 데이터 스트림 등이 이에 해당한다. 둘째, 프롬프트(Prompts)는 템플릿화된 메시지 및 워크플로우로, 반복적인 작업 패턴을 정의하고 재사용할 수 있도록 한다.

셋째, 도구(Tools)는 LLM이 호출할 수 있는 실행 가능한 함수로, 계산 수행, 외부 시스템 호출, 데이터 조작 등의 작업을 처리한다. 넷째, 샘플링(Sampling)은 서버가 모델의 추론을 요청하고 결과를 생성하도록 하는 메커니즘이다.

엔터프라이즈 환경에서 MCP 채택이 확산되고 있는 데에는 여러 이유가 있다. 첫째, 시스템 상호운용성 측면에서 표준화된 프로토콜을 통해 AI 모델이 다양한 엔터프라이즈 시스템과 통신할 수 있게 된다. 둘째, 확장 가능한 지능형 워크플로우 구축이 가능해져 전체 통합 레이어를 재구축하지 않고도 새로운 도구와 데이터 소스를 추가할 수 있다. 셋째, 표준화된 인터페이스를 활용함으로써 맞춤형 연결 구축의 부담이 크게 줄어든다. 넷째, 컴플라이언스 및 접근 제어 측면에서 AI 상호작용이 통제되고 감사 가능한 채널을 통해 이루어지므로 규제 요구사항을 충족하기가 용이해진다.

MCP 서버 생태계는 빠르게 성장하고 있으며, 다양한 레지스트리들이 MCP 서

버의 등록과 조회를 지원하고 있다. Anthropic의 공식 MCP Directory는 검증된 프로덕션급 MCP 서버를 제공하며, PulseMCP는 Tadas Antanavicius와 Mike Coughlin이 운영하는 일일 업데이트 종합 디렉토리이다. Awesome MCP Servers는 Anthropic과 커뮤니티가 함께 관리하는 공식 및 커뮤니티 서버의 종합 컬렉션이며, MCP.so는 독립 개발자와 커뮤니티가 운영하는 사용자 친화적 서버 발견 플랫폼이다. MCP 프로토콜은 지속적으로 업데이트되고 있으며, 보안 강화, 세션 관리, 스트리밍 통신 등의 기능이 추가되고 있다. 향후 MCP 서버 레지스트리는 수요와 공급을 연결하는 마켓플레이스로 진화할 것으로 전망되며, 이를 통해 개발자들은 필요한 도구를 쉽게 찾고, 도구 개발자들은 자신의 솔루션을 널리 배포할 수 있는 생태계가 형성될 것이다.

〈표 2-1〉 주요 MCP 서버 레지스트리

레지스트리	운영주체	유형
Anthropic's MCP Directory	Anthropic (공식)	검증된 프로덕션급 MCP 서버
PulseMCP	Tadas Antanavicius, Mike Coughlin	일일 업데이트 종합 디렉토리
Awesome MCP Servers	Anthropic + 커뮤니티	공식 및 커뮤니티 서버 종합 컬렉션
MCP.so	독립 개발자 + 커뮤니티	사용자 친화적 서버 발견 플랫폼

2) Agent2Agent Protocol(A2A)

에이전트가 개별적으로 도구를 사용하는 것을 넘어서 서로 협력하여 더 복잡한 문제를 해결하기 위해서는 에이전트 간의 표준화된 통신 방법이 필요하다. MCP가 에이전트와 도구 간의 연결을 표준화했다면, 에이전트 간의 협업을 위한 별도의 프로토콜이 요구되었다. 이러한 배경에서 Google은 2025년 4월 Agent2Agent Protocol(A2A)이라는 공개 표준 프로토콜을 제안하였다.

A2A는 서로 다른 에이전트 개발 프레임워크, 벤더, 도메인의 자율 AI 에이전트

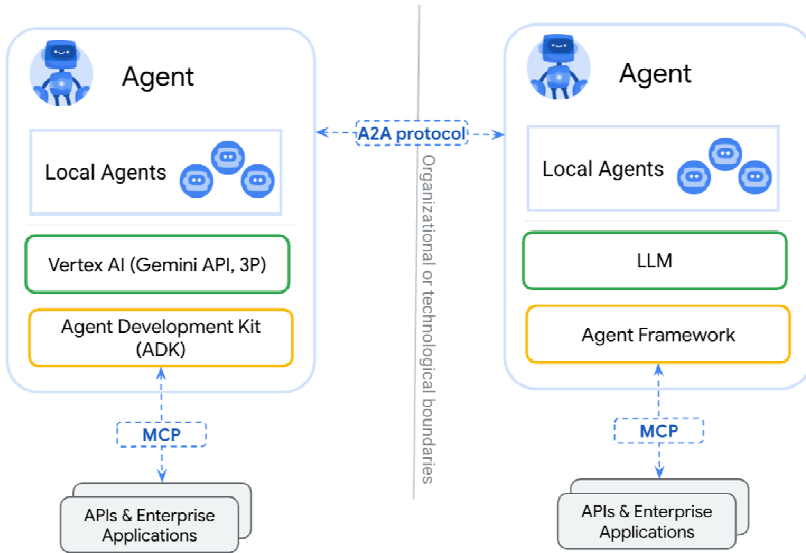
간 안전하고 확장가능한 협업을 지향한다. 특히 주목할 점은 A2A가 에이전트 간 협업에서 다른 에이전트의 내부 상태, 메모리, 도구에 직접 접근하지 않고도 효과적으로 협력할 수 있는 구조를 정의했다는 것이다. 이러한 설계는 보안을 강화하고 지적 재산을 보호하면서도 강력한 멀티 에이전트 시스템 구축을 가능하게 한다.

A2A 핵심 기능을 살펴보면 첫째, 에이전트 디스커버리 및 기능 교환이다. 에이전트는 표준화된 ‘에이전트 카드(Agent Cards)’를 게시하고, 다른 에이전트가 이를 동적으로 발견하고 작업을 협상할 수 있다. 에이전트 카드에는 해당 에이전트가 수행할 수 있는 작업, 필요한 입력, 제공하는 출력 등의 정보가 포함되어 있어 에이전트 간 자동화된 협업이 가능해진다. 두 번째 기능인 구조화된 작업 관리는 작업 위임, 모니터링 및 완료를 위한 명확한 프로토콜을 정의한다. 이를 통해 복잡한 작업이 여러 에이전트에 분배되고 진행 상황이 추적되며 결과가 통합될 수 있다. 세 번째 기능인 표준 기반 통신은 HTTP, JSON-RPC, Server-Sent Events(SSE)와 같은 친숙한 프로토콜을 기반으로 구축되어 기존 인프라와의 호환성을 보장한다. 네 번째 기능인 엔터프라이즈급 보안은 JWT 및 OIDC 인증, 암호화, 세밀한 권한 부여가 내장되어 있어 안전한 에이전트 상호작용을 지원한다. 다섯 번째 기능인 장기 실행 작업 지원은 비동기 푸시 알림을 사용하여 에이전트가 복잡하고 장기적인 워크플로우에서 실시간 상태 업데이트와 피드백 메커니즘으로 협업할 수 있게 한다.

MCP와 A2A는 서로 다른 목적을 가진 상호보완적인 프로토콜이다. MCP는 AI 에이전트와 도구/데이터 소스 연결에 전문화되어 외부 API 등을 도구로서 제공하고 활용하는 데 초점을 맞추고 있다면, A2A는 AI 에이전트 간 통신 및 협업에 전문화되어 에이전트 간 작업 위임 및 워크플로우 조율에 초점을 맞추고 있다. 비유하자면, MCP는 개인이 도구를 사용하는 방식을 표준화한 것이고, A2A는 팀원들이 서로 협력하는 방식을 표준화한 것이라 할 수 있다. 중요한 점은 MCP와 A2A가 서로 독립적이며 상호보완적으로 에이전틱 AI 생태계를 구축한다는 것이다. 하나의 에이전트 시스템은 MCP를 통해 다양한 도구와 데이터 소스에 접근하면서, 동시에 A2A를 통해 다른 에이전트들과 협업할 수 있다. 이러한 이중 구조는

에이젠틱 AI의 능력을 극대화하면서도 각 프로토콜이 자신의 전문 영역에 집중할 수 있게 해준다.

[그림 2-5] MCP와 A2A의 차이



자료: A2A protocol 사이트

A2A의 대표적인 활용 사례로는 국제 여행 계획 작성을 생각해볼 수 있다. 사용자의 여행 요청을 받은 AI 에이전트는 항공편 예약 에이전트, 호텔 예약 에이전트, 현지 액티비티 예약 에이전트, 금융 서비스 에이전트와 조율하여 포괄적인 여행 계획을 수립한다. 각 에이전트는 자신의 전문 영역에서 최적의 옵션을 탐색하고, A2A 프로토콜을 통해 서로의 결과를 공유하며, 최종적으로 사용자에게 통합된 여행 계획을 제시한다. 또 다른 활용 사례로는 구매 컨시어지가 있다. 구매 에이전트는 여러 판매자 에이전트들과 통신하여 주문을 처리한다. 예를 들어, 사용자가 특정 제품을 요청하면 구매 에이전트는 다양한 판매자 에이전트들에게 가격, 재고, 배송 조건 등을 문의하고, 최적의 옵션을 선택하여 주문을 완료한다. 이 과

정에서 각 판매자 에이전트는 자사의 시스템과 정책에 따라 독립적으로 작동하면서 A2A를 통해 일관된 방식으로 협업한다.

3) 시사점

① 상호운용성 및 표준화의 중요성 증대

MCP는 에이전트와 도구 간 통합을 표준화하고, A2A는 에이전트 간 협업을 가능하게 함으로써 에이전트 AI 생태계의 근간을 형성하고 있다. 그러나 이러한 프로토콜들의 통합은 필연적으로 복잡성이 증대되기 마련이다. 실제 배포 환경에서는 인증, MCP 도구 통합, 벡터 데이터베이스 지식 검색, A2A 프로토콜이 함께 작동해야 하며, 이를 원활하게 오케스트레이션하는 것의 중요성이 점점 증가할 것이다. 향후 이러한 복잡한 통합을 단순화하고 관리하는 플랫폼과 도구들이 중요한 비즈니스 기회가 될 것으로 예상된다.

② 멀티 에이전트 오케스트레이션의 복잡성

멀티 에이전트 시스템은 단일 에이전트로는 불가능한 복잡한 문제 해결을 가능하게 하지만, 동시에 새로운 거버넌스 과제를 야기한다. 첫째, 에이전트 분산 (Agent Sprawl) 문제가 있다. 명확한 거버넌스 없이 너무 많은 자율 시스템이 작동하면 누가 어떤 결정에 책임을 지는지 불분명해지는 책임성 부재 문제가 발생할 수 있다. 둘째, 복잡한 통합 문제가 있다. 효과적인 멀티 에이전트 시스템 운영을 위해서는 표준화된 데이터 형식, 명확한 API 계약, 효과적인 오케스트레이션 플랫폼, 모니터링 시스템이 필요하다. 셋째, 신뢰 및 안정성 문제가 있다. 에이전트의 예측 불가능성, 환각(hallucination), 상충되는 결론 등은 지속적인 모니터링과 중재 시스템을 요구한다. 자율 에이전트가 잘못된 결정을 내릴 경우 이를 감지하고 수정하는 메커니즘이 필수적이다. 넷째, 비용 및 ROI 압박이 있다. 대규모 오케스트레이션은 상당한 컴퓨팅 리소스, 통합 노력, 인적 주의를 소비하므로 신속한 가치 입증에 필요하다. 기업들은 에이전트 AI 도입의 비용 대비 효과를 명확히 측정하고 입증해야 하는 과제를 안고 있다.

③ 엔터프라이즈 채택의 현실적 과제

엔터프라이즈 환경에서 에이젠틱 AI의 채택은 여러 현실적 과제에 직면해 있다. 첫째, 거버넌스 및 컴플라이언스 측면에서 자율 시스템이 민감한 워크플로우와 상호작용함에 따라 책임 있는 AI 프레임워크, 편향 감지, 설명 가능성의 표준화가 필요하다. 특히 금융, 의료, 법률 등 규제가 엄격한 산업에서는 AI의 의사결정 과정을 추적하고 설명할 수 있어야 한다. 둘째, 레거시 시스템 통합 문제가 있다. 많은 엔터프라이즈가 10~20년 된 시스템을 운영하고 있으며, 이러한 시스템은 AI API를 수용하도록 설계되지 않았다. 데이터 사일로, 호환되지 않는 형식, 문서화되지 않은 인터페이스 등으로 인해 통합이 복잡해진다. 셋째, 신뢰 격차 문제가 있다. 자율 에이전트의 의사결정 투명성 부족은 비즈니스 이해관계자들의 신뢰를 감소시키고 잠재적 규제 문제를 야기할 수 있다. ‘블랙박스’로 인식되는 AI 시스템에 대해 경영진과 규제 기관은 회의적인 태도를 보일 수 있으며, 이는 에이젠틱 AI의 확산을 저해하는 요인이 된다. 이러한 과제들을 해결하기 위해서는 기술적 발전과 함께 조직 문화, 규제 환경, 인력 역량 등 다양한 측면에서의 변화가 동반되어야 할 것이다.

3. 물리적 AI

가. 주요 개념

물리적 AI란 인공지능이 물리적 시스템과 통합되어 기계가 현실 세계를 직접 인식하고, 추론하며, 상호작용할 수 있도록 하는 기술 패러다임을 의미한다. 이는 디지털 연산과 실체적 행동 사이의 간극을 메우는 혁신적인 기술 영역으로서, 제조업, 물류, 헬스케어, 자율주행 등 수조 달러 규모의 산업 전반에 걸쳐 혁신적인 변화를 예고하고 있다. 물리적 AI의 발전은 파운데이션 모델의 진화, 첨단 로봇틱스 기술의 성숙, 그리고 물리 기반 시뮬레이션 기술의 융합에 의해 주도되고 있다.

물리적 AI가 주목받는 이유는 기존의 디지털 AI가 가진 한계를 극복하고 AI의 적용 범위를 현실 세계로 확장시키기 때문이다. 텍스트와 이미지를 처리하는 생성

형 AI가 디지털 세계에서 혁명을 일으켰다면, 물리적 AI는 공장, 창고, 병원, 도로 등 물리적 공간에서 유사한 혁명을 일으킬 잠재력을 가지고 있다. 이는 단순히 로봇에 AI를 탑재하는 것을 넘어서, AI가 물리적 세계의 법칙을 이해하고 그에 따라 행동할 수 있는 새로운 지능의 형태를 의미한다.

물리적 AI 시스템은 네 가지 핵심 운영 단계로 구성된다. 첫 번째 단계는 구체화(Embodiment)로서, AI가 로봇, 자율주행차, 스마트 기기 등 물리적 형태에 내장되어 실제로 행동할 수 있는 수단을 갖추는 것이다. 이 단계에서 AI는 단순한 소프트웨어에서 벗어나 물리적 세계와 상호작용할 수 있는 '몸'을 얻게 된다. 두 번째 단계는 인식(Perception)으로서, 시스템이 다양한 센서 어레이를 통해 환경에 대한 실시간 데이터를 수집하는 과정이다. 여기에는 컴퓨터 비전용 카메라, 3D 매핑용 LiDAR/Radar, 동작 추적용 IMU(관성측정장치), 그리고 온도와 압력 센서 등이 포함된다. 세 번째 단계는 의사결정(Decision-Making/Reasoning)으로서, AI 알고리즘이 융합된 센서 데이터를 처리하여 맥락을 이해하고, 미래를 예측하며, 적절한 행동 방침을 결정하는 과정이다. 이 단계에서 AI는 단순한 반응을 넘어서 상황을 종합적으로 분석하고 최적의 행동을 선택해야 한다. 네 번째 단계는 행동(Action)으로서, 모터, 그리퍼, 바퀴 등의 액추에이터가 결정된 물리적 과업을 실제로 실행하여 물체를 조작하거나 환경을 탐색하는 것이다. 이 네 단계가 연속적으로 순환하면서 물리적 AI 시스템은 환경과 지속적으로 상호작용하게 된다.

전통적 AI, 특히 대규모 언어 모델(LLM)과 물리적 AI 사이에는 근본적인 차이가 존재한다. 전통적 AI는 화면상의 텍스트와 이미지를 처리하여 디지털 결과물을 생성하는 반면, 물리적 AI는 현실 세계의 센서 데이터를 처리하여 물리적 행동을 생성한다. 작동 환경 측면에서도 전통적 AI는 이산적이고 정제된 디지털 세계에서 작동하지만, 물리적 AI의 환경은 연속적이고 불확실하며 물리 법칙의 지배를 받는다. 디지털 세계에서는 입력과 출력이 명확하게 정의되지만, 물리적 세계에서는 노이즈, 마찰, 중력, 예측 불가능한 상황 등 다양한 변수가 작용한다.

이러한 차이로 인해 물리적 AI는 훨씬 더 복잡한 기술 스택을 필요로 한다. 여

기에는 다양한 센서 데이터를 통합하는 센서 융합(Sensor Fusion) 기술, 지면 시간을 극복하기 위한 엣지 컴퓨팅, 그리고 물리적 속성에 대한 내재적 이해를 갖춘 AI 모델 개발 등이 포함된다. 물리적 AI 시스템은 센서 하드웨어부터 AI 모델의 물리적 기반에 이르는 포괄적 기술 스택을 구축해야 하며, 이는 순수 소프트웨어 기반의 전통적 AI보다 훨씬 높은 기술적 장벽을 형성한다.

물리적 AI는 전통적인 로봇틱스 및 자율주행과도 밀접한 관련이 있으나 명확한 차이점이 존재한다. 전통적인 로봇틱스는 사전에 프로그래밍된 반복 작업을 수행하는 기계 자체 또는 규칙 기반 시스템을 의미한다. 예를 들어, 공장의 조립 라인에서 정해진 동작을 반복하는 로봇팔이 이에 해당한다. 이러한 전통적 로봇틱스는 유연성이 부족하고 환경 변화에 대응하기 어렵다는 한계가 있다. 반면, 물리적 AI는 이러한 로봇에 지능을 부여하여 실시간으로 인식하고, 추론하며, 적응할 수 있는 능력을 제공함으로써 로봇을 자율 에이전트로 전환시키는 역할을 한다.

자율주행은 물리적 AI의 주요 응용 분야 중 하나로 이해할 수 있다. 자율주행 시스템에서 '몸'은 센서와 액추에이터가 장착된 차량이며, '뇌'는 환경을 인식하고 차량을 제어하는 AI 모델로 대응된다. 자율주행차는 카메라, LiDAR, Radar 등을 통해 주변 환경을 인식하고, AI가 이 정보를 바탕으로 주행 경로를 결정하며, 스티어링, 가속, 제동 등의 액추에이터를 통해 실제 주행을 수행한다. 이처럼 물리적 AI는 로봇틱스와 자율주행을 포괄하는 상위 개념으로서, 디지털 AI의 능력을 물리적 세계로 확장하는 기술 패러다임이라 할 수 있다.

〈표 2-2〉 AI 패러다임 비교

특징	전통적 AI (LLM)	전통적 로봇틱스	물리적 AI
주요 영역	디지털	물리	물리
핵심 기능	정보 처리	사전 프로그래밍된 작업 수행	인식, 추론, 행동
데이터 입력	텍스트, 이미지	제한된 센서 신호	다중 모드 실시간 센서 데이터
결과물	텍스트, 이미지	정해진 물리적 동작	적응형 물리적 행동
핵심 과제	환각(Hallucination), 편향	유연성 부족, 환경 변화 대응 불가	안전, 물리 법칙 준수, Sim-to-Real
대표 사례	ChatGPT	조립 라인 로봇팔	자율 창고 로봇, 자율주행차

나. 핵심 기술 동향

1) 월드 파운데이션 모델 (World Foundation Model, WFMs)

LLM이 텍스트와 언어를 이해하는 파운데이션 모델이라면, WFM(World Foundation Model)이란 3D 공간, 물리 법칙, 객체 간 상호작용 등 실세계의 구조를 학습하여 물리적 세계를 이해한 파운데이션 모델이다. WFM의 주요 기능은 크게 세 가지로 구분된다. 첫째, 공간 이해 기능은 3D 환경에서 깊이, 거리, 크기 등을 파악하고 물리 시뮬레이션을 수행하는 능력이다. 둘째, 물리 시뮬레이션 기능은 중력, 충돌, 운동 등 물리 법칙을 예측하는 능력이다. 셋째, 시각-행동 연결 기능은 로봇이 보는 것을 이해하고 적절한 행동 계획을 수립하는 능력이다.

WFM의 대표적인 플랫폼으로는 NVIDIA의 Cosmos 플랫폼이 있다⁴⁾. 2025년 1월 CES에서 공개된 Cosmos는 물리적 AI 전용 WFM으로서 로봇과 자율주행 시스템 개발의 중요 인프라로 부상하고 있다. Cosmos는 2천만 시간 분량의 비디오를 큐레이션하여 훈련한 대규모 생성형 AI 모델을 기반으로 하며, 세 가지 핵심 구성 요소로 이루어져 있다. 첫째, Cosmos Predict는 텍스트, 이미지, 비디오 등

4) <https://www.nvidia.com/en-us/ai/cosmos/>

의 멀티모달 입력을 기반으로 미래의 가상 상태를 생성하는 모델이다. RGB 이미지만이 아니라 깊이, 의미론적 정보, 움직임 벡터를 동시에 예측하는 월드 생성 모델로서, 로봇이나 자율주행 시스템이 행동하기 전에 그 결과를 미리 시뮬레이션할 수 있게 해준다. 둘째, Cosmos Transfer는 세그먼트 맵, 깊이 맵, 라이더 스캔 등 구조화된 시뮬레이션 데이터를 포토리얼리스틱 비디오로 변환하는 멀티컨트롤 넷 모델이다. 이를 통해 시뮬레이션 환경에서 생성된 데이터를 실제 환경과 유사한 형태로 변환하여 AI 훈련에 활용할 수 있다. 셋째, Cosmos Reason은 시공간 인식 능력을 갖춘 멀티모달 추론 모델로서, 비디오 데이터를 분석하고 자연어로 상호작용의 결과를 추론하는 비전 언어 모델(VLM)이다.

2) 거대 행동 모델 (Large Behavior Models, LBM)

거대 행동 모델(Large Behavior Models, LBM)은 물리적 AI의 또 다른 핵심 기술로 부상하고 있다. 2025년 8월 보스턴 다이내믹스와 도요타 연구소(TRI)가 공동으로 공개한 연구 결과에서는 Large Behavior Model을 통해 Atlas 휴머노이드 로봇이 전신 조작과 이동 행동을 자율적으로 수행할 수 있음을 입증하였다. 이는 로봇이 단순히 프로그래밍된 동작을 수행하는 것이 아니라, 다양한 상황에서 스스로 적절한 행동을 선택하고 실행할 수 있음을 보여준 획기적인 성과이다.

LBM에서 로봇은 두 가지 방식으로 학습한다. 첫째, 카메라와 센서로 캡처한 타인의 행동을 관찰하며 학습하는 관찰 학습이다. 이를 통해 로봇은 인간이나 다른 로봇의 동작을 보고 모방할 수 있다. 둘째, 자신의 행동을 통해 경험적으로 학습하는 시행착오 학습이다. 로봇은 다양한 행동을 시도하고 그 결과를 평가하여 점점 더 나은 행동 패턴을 학습해 나간다. 이러한 학습 방식을 통해 LBM은 행동 패턴과 의사결정 과정에 집중하며, 물리적 세계와의 자연스러운 상호작용을 가능하게 한다.

3) 비전 언어 행동 모델 (Vision Language Action Model, VLA)

비전 언어 행동 모델(Vision Language Action Model, VLA)은 시각, 언어, 행동을 통합하여 로봇이 자연어 명령을 이해하고 이를 물리적 행동으로 변환하는

것을 학습하는 모델이다. VLA 모델은 인식(Vision), 지시(Language), 실행(Action)의 세 가지 모달리티를 엔드-투-엔드 프레임워크로 통합함으로써, 전통적인 로봇 제어의 ‘인식 → 계획 → 제어’ 파이프라인을 단일 통합 모델로 변화시켰다. 이러한 통합적 접근은 로봇이 더욱 자연스럽게 유연하게 인간의 명령에 반응할 수 있게 해준다.

VLA 모델의 대표적인 사례로는 2025년 9월 Google DeepMind가 발표한 Gemini Robotics 1.5가 있다(DeepMind 사이트⁵⁾). 이 모델은 현재 최고의 VLA 모델로 평가받고 있으며, 여러 혁신적인 기능을 제공한다. 첫째, 사고 과정 표시(Chain-of-Thought Visualization) 기능을 통해 로봇이 행동 전에 사고 과정을 텍스트로 보여줌으로써 투명성을 향상시킨다. 이는 로봇의 의사결정 과정을 인간이 이해하고 검증할 수 있게 해준다. 둘째, 멀티바디 추론 기능을 통해 휴머노이드, 로봇팔, 모빌 로봇 등 다양한 로봇 형태에서 학습된 기술이 상호 전이될 수 있다. 셋째, 도구 호출 기능을 통해 Google Search와 같은 외부 도구를 자동으로 호출하여 정보를 획득하고 활용할 수 있다.

4) 디지털 트윈 시뮬레이션 (Sim-to-Real 전환)

현실 세계에서 로봇을 훈련하는 것은 느리고 비용이 많이 들며 위험하다. 로봇이 잘못된 행동을 학습하는 과정에서 장비가 손상되거나 주변 환경에 피해를 줄 수 있기 때문이다. 이러한 이유로 대부분의 물리적 AI 훈련은 시뮬레이션 환경에서 이루어진다. 그러나 시뮬레이션에서 훈련된 시스템을 실제 하드웨어에 배포할 때 성능 저하가 발생하는 문제가 있는데, 이를 ‘Sim-to-Real 갭’이라고 한다. 시뮬레이션 환경은 아무리 정교하더라도 현실 세계의 모든 복잡성을 완벽하게 재현할 수 없기 때문에 이러한 격차가 발생한다.

NVIDIA의 Omniverse는 이러한 Sim-to-Real 갭을 메우기 위한 대표적인 기술 플랫폼이다⁶⁾. Omniverse 기반의 훈련 워크플로우는 다음과 같이 진행된다.

5) <https://deepmind.google/models/gemini-robotics/>

6) <https://www.nvidia.com/en-us/omniverse/>

먼저, Omniverse에서 3D 로봇 및 환경 시나리오를 생성한다. 그다음, 물리 시뮬레이션을 통해 CG 영상, 깊이 맵, 세그먼트 맵 등의 구조화된 데이터를 생성한다. 이어서 Cosmos Transfer를 활용하여 이러한 시뮬레이션 데이터를 포토리얼 리스틱 비디오로 변환한다. 변환된 데이터로 VLA 모델을 훈련하고, 마지막으로 Cosmos Predict를 통해 미래 환경을 예측하고 정책을 검증한다. 이러한 파이프라인을 통해 시뮬레이션에서 현실로의 전환 시 발생하는 성능 저하를 최소화할 수 있다.

다. 주요 기업 동향

1) NVIDIA

NVIDIA는 물리적 AI 분야에서 가장 포괄적인 생태계를 구축한 기업으로 평가받고 있다. Cosmos 플랫폼, Isaac 로보틱스 프레임워크, Omniverse 디지털 트윈 운영 시스템, Jetson Thor 로봇 컴퓨터 등을 통해 GPU부터 시뮬레이션 플랫폼, 파운데이션 모델에 이르는 통합 생태계를 운영하고 있다. NVIDIA는 현재 이러한 수직 통합된 물리적 AI 인프라를 제공하는 유일한 플레이어로서, 로봇 개발자들에게 하드웨어부터 소프트웨어까지 일관된 개발 환경을 제공하고 있다.

2) Tesla

Tesla는 '마스터 플랜 4'에서 회사 가치의 80%가 휴머노이드 로봇 Optimus에서 나올 것이라고 선언하며 자율주행 기술을 로보틱스로 확장하고 있다. Tesla는 자율주행차 개발을 통해 축적한 컴퓨터 비전, 센서 융합, AI 훈련 인프라 등의 역량을 휴머노이드 로봇에 적용하고 있으며, 대규모 제조 능력을 바탕으로 저렴한 가격의 범용 로봇 양산을 목표로 하고 있다. Optimus는 공장 작업부터 가정 내 서비스까지 다양한 영역에서 활용될 것으로 기대된다.

3) Boston Dynamics

Boston Dynamics는 현대차 그룹이 2021년 인수한 이후 더욱 활발한 혁신을 보여주고 있다. 2024년 4월에는 완전히 새롭게 설계된 전기 구동 방식의 이족 보

행 휴머노이드 Atlas 신모델을 발표하였다. 2025년 8월에는 도요타 연구소(TRI)와 공동으로 Large Behavior Model(LBM)을 적용한 Atlas의 자율 조작 능력을 입증하는 연구 결과를 발표하여, 휴머노이드 로봇 분야에서 기술적 리더십을 확인하였다.

4) Amazon

Amazon은 물류 자동화 분야에서 물리적 AI를 가장 대규모로 실제 적용하고 있는 기업으로 300개 이상의 물류센터에 100만 대 이상의 로봇을 배치하여 물류 효율성을 25% 이상 향상시켰다. Amazon의 사례는 물리적 AI가 실험실 수준을 넘어 실제 산업 현장에서 대규모로 운영될 수 있음을 보여주는 중요한 증거가 되고 있다.

5) Google DeepMind

Google DeepMind는 Gemini Robotics와 VLA 모델인 RT-2를 개발하여 물리적 AI의 핵심인 ‘두뇌’에 집중하고 있다. DeepMind의 접근 방식은 로봇 하드웨어보다는 AI 모델의 추론, 일반화, 작업 적응성을 향상시키는 데 초점을 맞추고 있으며, 이를 통해 다양한 로봇 플랫폼에 적용될 수 있는 범용 AI 개발을 목표로 한다.

라. 시사점

물리적 AI는 기존의 디지털 AI와 구별되는 몇 가지 독특한 특성을 가지고 있다. 첫째, 체화된 지능(Embodied Intelligence)으로서 AI가 물리적 신체를 통해 세계를 직접 경험하고 학습한다. 이는 추상적인 데이터에서 학습하는 전통적 AI와 달리, 실제 환경과의 상호작용을 통해 더욱 풍부하고 실용적인 지식을 획득할 수 있음을 의미한다. 둘째, 실시간 적응성으로서 예측 불가능한 환경 변화에 즉각적으로 대응하는 능력이 요구된다. 물리적 세계는 끊임없이 변화하며, 로봇은 이러한 변화에 실시간으로 적응해야 한다. 셋째, 멀티모달 통합 능력으로서 시각, 청각, 촉각 등 다양한 센서 데이터를 통합하여 처리해야 한다. 인간이 여러 감각을

종합하여 환경을 인식하듯이, 물리적 AI도 다양한 센서 정보를 융합하여 상황을 파악해야 한다. 넷째, 물리 법칙에 대한 이해가 필수적이다. 마찰, 관성, 충돌, 중력 등 물리적 현상을 학습하고 이에 정밀하게 대응해야 실제 환경에서 안전하고 효과적으로 작동할 수 있다. 이러한 특성들은 물리적 AI가 단순히 기존 AI의 확장이 아닌 새로운 기술 패러다임을 보여준다.

물리적 AI의 발전은 향후 산업과 사회에 광범위한 영향을 미칠 것으로 전망된다. 첫째, 산업 전반의 재편이 예상된다. 제조업, 물류, 헬스케어, 농업, 건설 등 대부분의 산업에서 물리적 AI가 핵심 인프라로 성장할 것이다. 특히 인력 부족이 심화되는 분야에서 물리적 AI는 생산성을 유지하고 향상시키는 핵심 수단이 될 것이다. 둘째, 협업 생태계의 형성이 예상된다. 인간과 로봇이 물리적 공간에서 자연스럽게 협업하는 환경이 구현될 것이다. 로봇이 위험하거나 반복적인 작업을 담당하고, 인간은 창의성과 판단력이 요구되는 업무에 집중하는 협업 모델이 확산될 것이다. 셋째, 일자리 구조의 변화가 예상된다. 반복적이고 위험한 작업은 로봇이 대체하고, 인간은 창의적이고 전략적인 역할에 집중하게 될 것이다. 이는 단기적으로는 일부 직종의 감소를 가져올 수 있으나, 장기적으로는 새로운 형태의 일자리 창출과 업무의 질적 향상으로 이어질 수 있다. 로봇 관리자, AI 훈련 전문가, 인간-로봇 협업 설계자 등 새로운 직종이 등장할 것으로 예상된다.

물리적 AI의 발전에는 여러 기술적 도전 과제가 남아있다. 첫째, 안전성과 신뢰성의 확보가 가장 중요한 과제이다. 물리적 AI의 오류는 디지털 AI와 달리 물리적 피해로 직결될 수 있다. 로봇의 잘못된 동작은 장비 손상, 제품 불량, 심지어 인명 피해까지 초래할 수 있기 때문에 강력한 안전장치와 검증 시스템이 고도화되어야 한다. 둘째, 윤리적 고려사항에 대한 사회적 합의가 필요하다. 로봇의 권리, 개인 정보 보호, 해킹 위험, 책임 소재 등 새로운 윤리적 문제들이 제기되고 있으며, 이에 대한 사회적 논의와 합의가 필요하다.

4. AI 모델 경량화

가. 주요 개념

AI 모델 경량화(Model Lightweighting)란 대규모 AI 모델의 성능 손실을 최소화하면서 연산 효율을 극대화하여, 클라우드에 집중된 컴퓨팅 부하를 엣지(Edge) 디바이스로 분산시키는 전략적 기술을 의미한다.

2025년 현재 생성형 AI 시장은 심각한 병목에 직면해 있다. 최근 시장 조사 결과에 따르면, 기업들이 생성형 AI 도입을 위해 약 300억~400억 달러 규모의 막대한 투자를 단행했음에도 불구하고, 약 95%의 조직이 가시적인 투자 수익(ROI)을 창출하지 못하고 있는 것으로 나타났다(Challapally, et al, 2025.7). 이를 업계에서는 'GenAI Divide(생성형 AI 격차)'라고 부르며, 이는 기술의 부재가 아니라 비용 효율적인 배포와 운영 전략의 부재에서 기인한다. 대다수의 기업(약 3분의 2)은 여전히 파일럿 단계에 머물러 있으며, 실제로 전사적 차원에서 AI를 확장(Scale)하고 있는 기업은 전체의 약 5%에 불과하다(McKinsey 2025. 11). AI 모델 경량화는 추론 비용 절감, 지연 시간 단축, 전력 소비 감소 등 AI 기술의 실용화를 위한 핵심 요소로 향후 AI를 통한 실질적 생산성 향상을 가져올 에이전틱 AI와 물리적 AI의 발전은 경량화 기술 없이는 실현될 수 없다. 에이전틱 AI가 복잡한 워크플로우를 자율적으로 수행하고, 물리적 AI가 로봇이나 자율주행차와 같은 엣지 디바이스에서 실시간으로 인식·추론·행동하기 위해서는 클라우드 의존도를 낮추고 온디바이스에서 고성능 AI를 구동할 수 있는 기술이 필수적이다. 특히 2025년 들어 전 세계 AI 데이터센터의 연간 전력 소비량이 100 테라와트시(TWh) 증가할 것으로 전망되면서, 에너지 효율성은 단순한 비용 절감을 넘어 규제 준수와 기업의 생존을 위한 필수 요건이 되고 있다(WashingtonExaminer 2025. 11. 20).

나. 경량화 학습방법의 진화

경량화 기술은 앞서 간략히 언급한 Depth Up-scaling, Pruning, Distillation 등의 기법 외에도, 양자화(Quantization)와 추측적 디코딩(Speculative Decoding)

등을 통해 새로운 돌파구를 마련해가고 있다. 또한 경량화 기술은 과거의 단순한 파라미터 절삭을 넘어, 수학적 최적화와 하드웨어의 특성을 고려한 ‘공동 설계 (Co-design)’ 방식으로 진화하고 있다.

1) 양자화(Quantization)

양자화는 모델의 가중치(Weight)와 활성화 값(Activation)을 표현하는 정밀도를 낮추어 메모리 사용량과 연산량을 줄이는 기술이다. 2025년에는 기존의 8비트 (INT8)를 넘어 4비트, 심지어 1비트 수준의 극단적 양자화 기술이 상용화 단계에 진입하였다. 특히 2025년 학계와 산업계에 가장 큰 충격을 준 기술 중 하나는 ‘1비트 LLM’의 등장이다. 전통적인 LLM이 16비트(FP16/BF16) 부동소수점 연산에 의존했던 것과 달리, Microsoft Research 등이 제안한 BitNet b1.58과 같은 아키텍처는 가중치를 삼진법(Ternary) 값인 $\{-1, 0, 1\}$ 로 제한한다(Ma et al. 2025. 4. 16). 이 방식은 행렬 곱셈 연산에서 가장 비용이 많이 드는 곱셈 과정을 단순 덧셈으로 대체하여 연산 복잡도를 획기적으로 낮출 뿐만 아니라, 메모리 대역폭 요구량을 줄여 에너지 소비를 급격히 감소시킨다. 놀랍게도 1.58비트 모델은 동일한 크기의 FP16 모델과 비교했을 때 언어 이해 능력에서 성능 저하가 거의 없는 것으로 나타났다. 이는 “모델 성능은 파라미터 수와 훈련 데이터 양에 비례한다”는 기존의 스케일링 법칙(Scaling Law)에 새로운 변수를 추가한 것으로, 향후 AI 하드웨어가 고정밀도 곱셈기(Multiplier) 중심에서 저정밀도 덧셈기 위주로 재편될 가능성을 시사한다.

한편, 거대 모델을 처음부터 다시 학습시키는 비용은 천문학적이기 때문에, 학습이 완료된 모델을 압축하는 ‘사후 학습 양자화(Post-Training Quantization, PTQ)’ 기술이 2025년 추론 시장의 표준으로 자리 잡았다 (Franco et al., 2025. 8. 28). 현재 NVIDIA의 Blackwell 아키텍처 등 최신 GPU는 FP4(4비트 부동소수점) 연산을 네이티브로 지원하며, 연구 결과에 따르면 4비트로 양자화된 Llama 3.1 모델이 16비트 원본 모델과 비교하여 복잡한 벤치마크에서 성능 차이가 거의 없음이 입증되었다. 이는 추론 비용을 4배 이상 절감할 수 있음을 의미하며, 온디

바이스 환경에서의 LLM 구동을 현실화하는 핵심 동력이 될 것으로 기대된다.

2) 추측적 디코딩(Speculative Decoding)

추측적 디코딩은 LLM의 추론 속도를 높이기 위해, 작고 빠른 ‘초안 모델(Draft Model)’이 먼저 토큰을 생성하고 큰 ‘타겟 모델’이 이를 검증하는 기술이다. 이는 메모리 대역폭 병목 현상을 완화하는 데 탁월한 효과가 있다.

2025년 추측적 디코딩은 더욱 지능화되었다. ‘추측적 검증(Speculative Verification, SV)’ 기술은 초안 모델의 정확도를 실시간으로 예측하는 별도의 ‘동반 모델(Companion Model)’을 도입하였다(Kim et al., 2025. 9. 29). 이 동반 모델은 초안 모델과 타겟 모델 간의 정렬 정도를 평가하여 검증 길이를 동적으로 조절함으로써, 잘못된 초안 생성으로 인한 연산 낭비를 줄이고 처리량을 극대화한다. 또한 특정 접미사 패턴을 예측하여 추론 속도를 가속화하는 SuffixDecoding 기술도 반복적인 텍스트 생성 작업에서 효율성을 입증하고 있다.

3) GaLore(Gradient Low-Rank Projection)

추론뿐만 아니라 ‘학습’ 단계에서의 경량화도 비약적으로 발전하였다. ‘GaLore’ 기술은 거대 모델 학습 시 발생하는 그래디언트(Gradient)가 낮은 랭크(Low-Rank) 구조를 가진다는 점에 착안하여, 그래디언트를 저차원 공간으로 투영하여 학습한다(Zhao et al. 2024). 이 기술을 사용하면 70억 파라미터(7B) 규모의 모델을 값비싼 H100 클러스터가 아닌, 24GB 메모리를 가진 일반 소비자용 GPU (예: RTX 4090) 한 장으로도 전체 파라미터 학습(Full Fine-tuning)이 가능하다.

2025년 발표된 후속 연구인 GaLore 2는 SVD(특이값 분해) 연산의 오버헤드를 줄이고 최신 병렬 학습 전략과 통합하여 대규모 사전 학습 시나리오에서도 확장성을 입증하였다(Su et al., 2025). 이는 중소기업이나 학계 연구실에서도 독자적인 고성능 모델을 개발할 수 있는 길을 열어주었으며, AI 기술의 민주화 측면에서 중요한 의의를 갖는다.

4) 하드웨어-소프트웨어 공동 설계 (Co-design)

경량화된 소프트웨어는 이를 물리적으로 구동할 하드웨어와의 결합을 통해 완성된다. 2025년은 모바일 AP(Application Processor) 제조사들과 NPU 스타트업들이 소프트웨어 스택과 하드웨어를 공동 설계하는 ‘Co-design’ 전략이 정점에 달한 해이다.

스마트폰은 가장 보편적인 AI 디바이스로, 주요 칩셋 제조사들은 생성형 AI 구동에 최적화된 NPU 성능 경쟁에 돌입하였다. 퀄컴의 스냅드래곤 8 Gen 5는 온디바이스 ‘에이전트 AI’를 표방하며, 탑재된 헥사곤(Hexagon) NPU는 이전 세대 대비 37% 향상된 속도를 자랑한다. 특히 생성형 AI 모델 구동 시 초당 220 토큰(Tokens per Second)이라는 처리 속도를 기록하였으며, 자체 개발한 Oryon CPU 아키텍처와 결합하여 전력 효율을 35% 개선함으로써 AI 에이전트가 백그라운드에서 지속적으로 사용자의 상황을 인지하고 판단할 수 있는 기반을 제공한다.

삼성전자는 2나노(2nm) GAA(Gate-All-Around) 공정을 적용한 엑시노스 2600을 통해 전력 효율의 극한을 추구하고 있다. 유출된 벤치마크 결과에 따르면, 엑시노스 2600의 NPU는 경쟁작인 애플의 A19 Pro 대비 특정 AI 연산에서 약 6배, 스냅드래곤 8 Gen 5 대비 30% 빠른 성능을 보인 것으로 알려졌다. 삼성은 이 칩셋을 통해 셔터를 누르기 전 AI가 최적의 구도를 예측하고 다중 프레임 합성을 미리 수행하는 ‘예측 이미징(Predictive Imaging)’ 기술을 구현하였으며, 이는 갤럭시 시리즈의 온디바이스 실시간 통번역 및 카메라 기능을 한 차원 높일 것으로 기대된다.

〈표 2-3〉 2025년 주요 모바일 NPU 사양 및 특징 비교

특징	Qualcomm Snapdragon 8 Gen 5 (Qualcomm 2025)	Samsung Exynos 2600 (Samsung 2025)	Apple A19 Pro (Apple 2025)
아키텍처	Oryon CPU + Hexagon NPU	2nm GAA + 차세대 NPU	3nm + 16코어 Neural Engine
AI 처리속도	220 Tokens/sec (생성형 AI)	스냅드래곤 대비 NPU 성능 30% 우위	온디바이스 Siri 및 실시간 추론 최적화
주요 기능	멀티모달 에이전트, 행동형 AI	예측 이미징, 실시간 통번역	하드웨어 가속 레이 트레이싱, 발열 제어
전력 효율	CPU 효율 35% 개선	2nm 공정을 통한 전성비	알루미늄 유니바디 열 분산 설계

자료: Qualcomm, Samsung, Apple 사이트

전력 소모를 근본적으로 해결하기 위해 인간의 뇌 구조를 모방한 뉴로모픽 컴퓨팅 기술도 2025년 가시적인 성과를 내고 있다. 뉴로모픽 칩은 스파이킹 신경망(SNN)을 기반으로 데이터가 발생할 때만 전력을 소비하는 ‘이벤트 구동(Event-driven)’ 방식을 사용한다. 이는 항상 켜져 있어야 하는 IoT 센서나 드론 감시 시스템에서 기존 GPU 대비 수백 배의 에너지 효율을 제공한다. 미국 UT Dallas 연구팀과 한국의 KAIST, ETRI 등은 기존 딥러닝 모델보다 훨씬 적은 학습 연산으로 패턴을 인식하고 예측할 수 있는 뉴로모픽 프로토타입을 개발하였으며, 2025년은 이 기술이 연구실을 벗어나 엣지 AI 디바이스 시장(약 300억 달러 규모)으로 진입하는 원년이 될 전망이다.

하드웨어 성능을 최대한 끌어내기 위한 소프트웨어 엔진 경쟁도 치열하다. 오픈소스 진영에서는 vLLM과 Hugging Face의 TGI(Text Generation Inference)가 양대 산맥을 이루고 있다. vLLM은 ‘PagedAttention’ 기술을 통해 메모리 파편화를 줄여 처리량을 극대화하는 반면, TGI는 엔터프라이즈급 안정성과 다양한 디코딩 전략 지원에 강점이 있다.

국내 스타트업인 프렌들리AI(FriendliAI)는 독자적인 ‘이터레이션 배치(Iteration

Batching)’ 기술을 통해 기존 vLLM 대비 높은 처리량과 낮은 지연 시간을 달성하였다. 또한 ‘Friendly TCache’ 기술로 자주 사용되는 연산 결과를 캐싱하여 중복 연산을 제거하고, 단일 GPU에서 여러 개의 LoRA 어댑터를 동시에 서빙하는 ‘Multi-LoRA’ 기능을 지원하여 SaaS 기업들의 비용 절감 수요를 공략하고 있다. 한국의 AI 인프라 소프트웨어 기업 모레(Moreh)는 짐 켈러가 이끄는 텐스토렌트(Tenstorrent)와 협력하여, NVIDIA GPU가 아닌 하드웨어에서도 대규모 AI 모델을 효율적으로 학습하고 추론할 수 있는 통합 솔루션을 2025년 슈퍼컴퓨팅 컨퍼런스에서 공개하였으며, 이는 특정 하드웨어 종속성을 탈피하려는 ‘AI 독립’ 움직임의 일환으로 주목받고 있다.

라. 시사점

본 보고서에서 살펴본 생성형 AI, 에이젠틱 AI, 물리적 AI의 발전은 모두 경량화 기술과 밀접하게 연계되어 있다. 생성형 AI 분야에서 오픈 소스 진영이 빅테크와의 격차를 좁힐 수 있었던 배경에는 DeepSeek-R1의 GRPO와 같은 효율적인 훈련 기법과 함께 양자화, 증류 등의 경량화 기술이 자리하고 있다. 에이젠틱 AI가 복잡한 워크플로우를 자율적으로 수행하기 위해서는 실시간 추론이 필수적이며, 이는 추측적 디코딩과 NPU 최적화를 통해 달성될 수 있다. 물리적 AI가 로봇이나 자율주행차와 같은 엣지 환경에서 작동하기 위해서는 전력 효율이 높은 경량화된 모델과 뉴로모픽 칩과 같은 저전력 하드웨어가 필수적이다. 소프트웨어 정의 하드웨어로의 전환도 중요하다. 2025년 이후에는 하드웨어의 성능보다 이를 구동하는 컴파일러(TVM, MLIR 등)의 최적화 능력이 칩의 성패를 좌우할 것으로 전망된다. 리벨리온, 퓨리오사AI 등 국내 AI 반도체 기업들이 하드웨어 스펙보다 소프트웨어 호환성과 컴파일러 최적화에 집중하는 것은 이러한 시대적 흐름을 반영한 전략이다. 엔비디아의 CUDA 독점 체제는 다양한 NPU와 개방형 소프트웨어 스택에 의해 서서히 잠식될 것이며, 모레와 텐스토렌트의 협력과 같은 ‘AI 독립’ 움직임은 이러한 변화를 가속화할 것이다.

대한민국은 메모리 반도체(HBM)의 절대 강자로서 AI 하드웨어 공급망의 핵심 위치를 점하고 있으나, 시스템 소프트웨어 역량에서는 글로벌 격차가 존재한다. 특히 반도체 및 AI 엔지니어가 부족할 것으로 전망되며, 하드웨어 설계 인력에 비해 컴파일러, 드라이버 등 시스템 소프트웨어를 다룰 수 있는 고급 인력의 공급이 턱없이 부족하다. 이는 국산 NPU의 상용화를 가로막는 최대 걸림돌로 작용하고 있다. 따라서 정부 R&D 예산의 일부를 하드웨어 제작 지원에서 컴파일러, 드라이버, 최적화 라이브러리 개발로 전환하고, 산학 협력을 통한 융합형 인재 양성이 시급하다. 또한 삼성전자와 SK하이닉스는 메모리 반도체 패권을 활용하여 연산과 저장이 통합된 PIM 기술과 저전력 뉴로모픽 소자의 상용화를 통해 차세대 'AI 메모리' 분야에서 초격차를 확보해야 할 것이다.

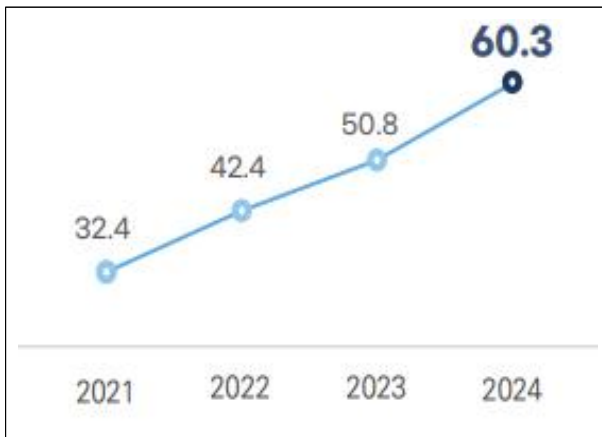
제 2 절 시장동향

본 절에서는 AI 기술의 급진적인 발전이 AI 시장에 가져오고 있는 변화를 다각도로 조명한다.

1. AI 도입·확산 현황

AI 기술이 하루게 다르게 발전하며 빠르게 일상화되고 있다. 과학기술정보통신부의 2024 인터넷이용 실태조사(최문실 외, 2025.4)에 따르면 AI 서비스를 경험한 국민은 2021년 32.4%에서 2024년 60.3%로 3년 새 두 배 증가했으며 AI 서비스 경험자의 50% 이상은 AI 서비스에 대해 긍정적으로 평가하고 있는 것으로 나타났다. 특히 '정부, 기업의 AI 기반 서비스는 사회 전반에 긍정적인 영향을 준다'는 질문에 대해 AI 서비스 경험자의 64.7%, 미경험자의 51.9%가 그렇다 혹은 매우 그렇다라고 답한 것으로 나타나 이제 대부분의 국민들이 AI에 대한 높은 수용도를 가지고 있음을 볼 수 있다.

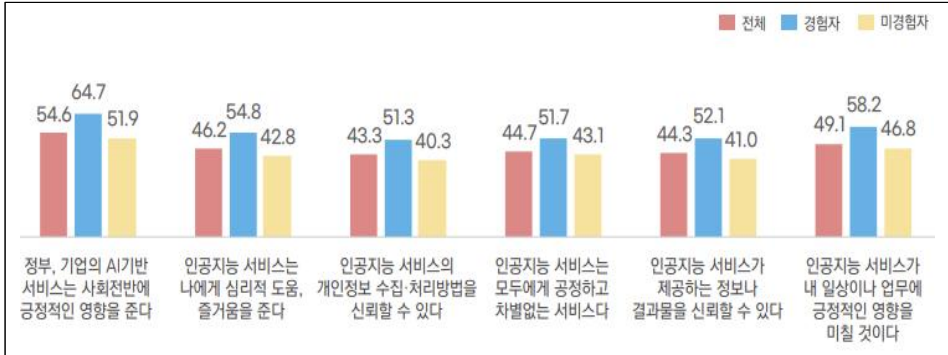
[그림 2-6] AI 서비스 경험률
(만 6세이상, 단위: %)



자료: 2024 인터넷이용 실태조사(최문실 외, 2025. 4)

[그림 2-7] AI 서비스 인식

(만 6세이상, 단위: %)



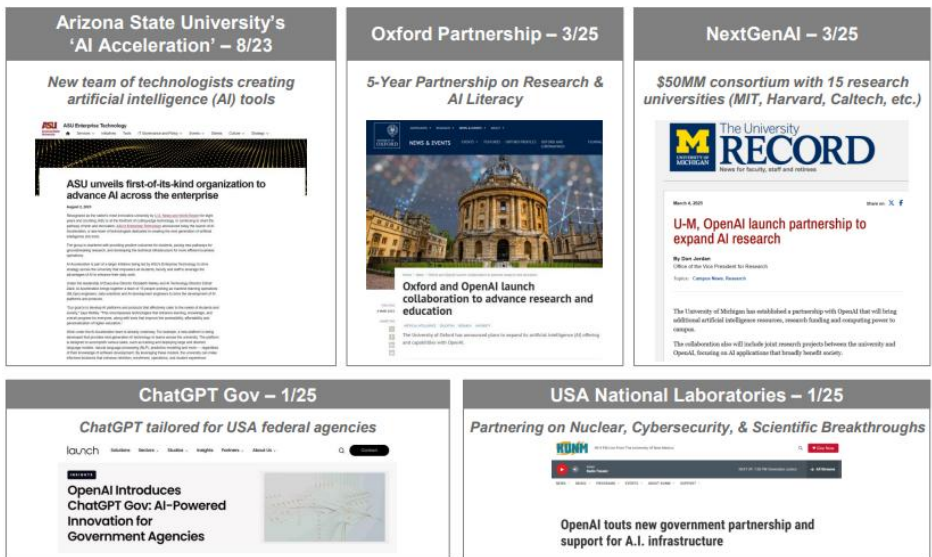
자료: 2024 인터넷이용 실태조사(최문실 외, 2025.4)

2022년 말 ChatGPT의 등장으로 주목받기 시작한 생성형 AI 서비스의 경우 2024년 이용률이 33.3%로 2년 새 약 두 배 가까이 증가했다(최문실 외, 2025. 4). 전 세계적으로 보면 ChatGPT는 출시 17개월 만에 주간 활성 사용자 수가 8배 증가하여 8억 명에 이르렀으며, 특히 비북미 지역에서는 불과 3년만에 사용자 비중이 90%에 도달했다(BOND, 2025). 이는 인터넷이 이 같은 결과를 얻는데 23년이 걸린 것과 뚜렷한 대비를 이루는데 이는 ChatGPT가 인터넷 혁명과 달리, 전 세계적으로 동시 확산되고 있음을 의미한다.

2024 인터넷이용 실태조사(최문실 외, 2025. 4)에 따르면 생성형 AI 이용 분야는 아직 단순 정보검색(81.9%)과 문서작업 보조(44.4%)가 주를 이루고 있으나, 인간과 같은 수준의 언어능력과 추론능력을 갖춘 생성형 AI 기반 서비스는 외국어 번역(40.0%), 창작 및 취미활동 보조(15.2%), 코딩 및 프로그램 개발(6.3%) 등 기존에는 인간 전문가의 영역으로 여겨지던 다양한 분야로 빠르게 확산하고 있다. 미국에서 수행된 여러 조사 결과들은 AI가 단순한 기술적 호기심을 넘어서 실질적인 비즈니스 가치와 사회적 효용을 제공하고 있음을 잘 보여준다. 우선, Pew Research Center(2025. 2. 25)에 따르면 미국 성인 직장인의 72% 이상은 AI 챗봇이 업무를 더 빠르고 효율적으로 수행하는 데 도움을 주었다고 평가했으며, Foutune(2025. 2. 20)에 의하면 미국 내 18~24세 학생들은 주로 연구, 문제해

결, 학습 및 조언 목적으로 ChatGPT를 사용하고 있다고 응답했다. 2024년 4분기 기준 S&P 500 기업 중 절반 이상이 실적 발표에서 'AI'를 언급했으며 (Goldman Sachs, 2025. 2. 27), 2025년 1분기 기준 미국 기업의 AI 채택률은 전 분기 대비 21% 증가한 약 7%로 나타났다(Goldman Sachs Global Investment Research, 2025.3, BOND, 2025. 5.에서 재인용). 오픈AI는 2025년 1월 미국 연방 기관을 위한 ChatGPT Gov를 출시하고 미국 국립 연구소는 핵, 사이버 보안 및 과학적 혁신 분야에서 AI 파트너십을 맺는 등 AI는 국방, 안보, 과학, 행정 분야에서도 적극 활용되고 있다(BOND, 2025. 5).

[그림 2-8] 정부 및 연구기관의 AI 도입 사례



Source: Arizona State University (8/23), Oxford University (3/25), University of Michigan (3/25), Launch Consulting (1/25) via AI Advantage Daily News, NPR (1/25)

자료: BOND(2025. 5)

2. AI 시장 동향

가. AI 칩 제조 및 컴퓨팅 서비스 시장 현황

AI 성능은 고성능 하드웨어, 특히 AI 칩과 이를 기반으로 한 컴퓨팅 서비스에 의해 결정된다. NVIDIA는 AI 칩 시장에서 약 80%의 점유율을 차지하며 지배적 사업자로 군림하고 있다. 2024년 1월 기준 NVIDIA의 분기 매출은 전년 동기 대비 78% 증가하여 390억 달러를 기록했으며, 전 세계 데이터센터 자본지출의 25% 이상을 차지했다(NVIDIA, 2025. 2. 26).

NVIDIA의 독주로 주요 기업들이 공급망에서의 NVIDIA의 의존도를 줄이기 위한 ‘탈엔비디아’ 전략을 가속화함에 따라 Google TPU(Tensor Processing Unit), Amazon Trainium, AMD Gaudi 등 맞춤형 ASIC(Application-Specific Integrated Circuit) 칩의 제작·활용도 증가하고 있다. 맞춤형 ASIC칩은 특정 연산 작업에 최적화되어 있어 높은 성능과 효율성을 동시에 제공하며, 이에 대규모 클라우드 사업자들은 자사 AI 워크로드에 최적화된 칩을 자체 개발함으로써 비용 절감과 성능 극대화를 동시에 추구하고 있다.

Google의 경우, 2013년 자체 AI 전용 ASIC인 TPU를 설계, 개발하기 시작해 2015년 처음 사내에 배포를 시작하였으며 Google Cloud를 통해 제공하고 있다. 2024년 Google TPU의 Google Cloud로의 연간 판매액은 116% 증가해 89억 달러에 이를 것으로 추정된다(BOND, 2025. 5).

Amazon는 2020년 12월 AI/머신러닝 칩인 AWS Trainium 칩을 발표했으며 Trainium 기반 AI 클라우드 매출은 2025년 216% 성장하여 36억 달러에 이를 것으로 전망된다(BOND, 2025. 5).

〈표 2-4〉 AI칩 시장 주요 플레이어 및 수익화 현황(2024-2025)

기업명	시장 점유율 (AI 가속기)	매출 성장률(최신)	주요 제품/전략
NVIDIA	약 80%	분기별 매출 +78% (390억 달러)	<ul style="list-style-type: none"> GPU(Blackwell 등), 데이터센터 CapEx의 25% 이상 차지 GPU 기반 AI 학습·추론 시장 독점. CUDA 생태계로 진입장벽 확보
Google	-	TPU 연간 판매액 +116%(89억 달러 예상)	<ul style="list-style-type: none"> TPU(Tensor Processing Unit), AI 모델 구축 및 실행에 특화된 ASIC 구글 내 부용 중심
Amazon	-	Trainium 연간 판매액 +216%(36억 달러 예상)	<ul style="list-style-type: none"> Trainium 칩, AI 훈련 및 추론에 특화된 ASIC 추론용 칩에서 효율성 높으며, AWS AI 서비스에 통합됨
AMD	-	MI300 AI 가속기 2024년 20억 달러 이상 매출 예상	<ul style="list-style-type: none"> MI300 시리즈, Ryzen AI Pro 300 시리즈, NVIDIA에 대한 경쟁력 있는 대안
Intel	-	Gaudi AI 칩, NVIDIA H100보다 50% 저렴	<ul style="list-style-type: none"> Gaudi 3 가속기, 비용 효율성에 중점, LLM 훈련에 특화
Meta	-	MTIA(Meta Training and Inference Accelerator) 개발에 투자	<ul style="list-style-type: none"> 맞춤형 AI 칩, 딥러닝 속도 향상, 장기적인 AI 인프라 구축 계획

자료: BOND(2025. 5), "Trends - Artificial Intelligence(AI)", patentpc.com

주요 기업들이 탈엔비디아, 풀스택 수직통합 전략 등을 강화해나감에 따라 장기적으로 AI 하드웨어 시장의 경쟁 구도는 지금과는 다른 양상으로 변화해 갈 것으로 예상된다(Morgan Stanley, 2025. 3. 20).

나. 데이터센터 시장 현황

AI의 급격한 확산은 데이터센터 수요를 폭증시키고 있으며, 이는 전력 공급의 구조적 압박으로 이어지고 있다. 2024년에는 글로벌 데이터센터 자본지출(CapEx)이 4550억 달러 규모로 급증했으며, 미국의 데이터센터 용량은 2020년

대비 신규 건설 기준 16배, 기존 대비 5배 확대되었다(CIOdive, 2025. 3. 20). 데이터센터는 2017년 이후 전체 전력 증가율의 약 네 배에 달하는 연평균 약 12% 증가율을 보이고 있으며, 현재 전 세계 전력 소비의 약 1.5%, 특히 미국은 전 세계 데이터센터 전력 소비의 약 45%를 차지하고 있다(International Energy Agency, 2025. 4). IEA(International Energy Agency)는 AI 중심의 데이터센터 전력 수요가 2030년까지 4배 이상 증가하여 약 945TWh에 이를 것으로 전망하고 있다(The Wall Street Journal, 2025. 4).

이러한 상황은 ‘제번스 패러독스(Jevons Paradox)’—즉, 기술 효율 개선에도 불구하고 전체 에너지 소비가 오히려 증가하는 현상—을 명확히 보여준다. 이는 컴퓨팅 능력 부족이 아닌, 안정적인 전력 공급이 새로운 병목이 되었음을 시사한다.

다. AI 클라우드 서비스 시장 현황

AI 클라우드 서비스는 AI 모델의 훈련과 추론을 지원하는 핵심 인프라 계층으로 하이퍼스케일 클라우드 제공업체들이 방대한 자본 지출을 통해 시장을 주도하고 있다. 글로벌 AI as a Service(AIaaS) 시장 규모는 2024년 약 160억 달러였으며, 2025년에는 약 220억 달러로 확대되고 2025~2030년간 연평균 성장률은 36.1%로 전망된다(Grand View Research, 2025a). 기업별 상황을 살펴보면 다음과 같다.

1) Amazon Web Services(AWS)

AWS는 규모의 경제와 폭넓은 서비스 포트폴리오를 기반으로 글로벌 클라우드 시장 점유율 약 30%를 유지하며 1위를 지속하고 있다. 2024년 AWS의 연간 매출은 약 1,076억 달러이며, 전년 대비 약 19% 증가하였다(Amazon IR 사이트). 2025년 2분기 기준 AWS 매출은 약 309억 달러로 분기 기준 최대 규모를 기록했으며, 연환산 매출은 1,240억 달러 수준으로 평가된다(CRN, 2025. 8. 7).

AWS는 AI 인프라 경쟁력 확보를 위해 자체 맞춤형 칩(Trainium·Inferentia)을 적극 도입하고 있다. Trainium은 공개 자료상 동급 GPU 대비 약 30~40%

향상된 가격 대비 성능을 제공한다(Amazon 사이트)⁷⁾.

2) Microsoft Azure

Azure는 AI 기능을 Microsoft 365, Dynamics, Copilot 등 기업용 소프트웨어에 통합함으로써 AI 확장성과 생산성 제고를 동시에 추구하고 있다. 글로벌 클라우드 시장점유율은 약 20%대로 AWS에 이어 2위이며, AIaaS 영역에서도 주요 플레이어로 자리잡고 있다. Microsoft는 Azure 단독 매출을 공식적으로 분리해 공시하지 않으나, 2025 회계연도 기준 Azure 매출이 약 750억 달러를 넘었다는 분석이 존재한다(Constellation, 2025. 6. 30). MS의 공시자료에 따르면 2025년 2분기 기준 'Intelligent Cloud' 사업부문의 매출은 약 255억 달러이며 Azure 및 기타 클라우드 서비스는 전년 대비 31% 성장한 것으로 확인된다(Microsoft IR 사이트).

3) Google Cloud Platform(GCP)

Google은 광범위한 AI 서비스와 TPU 칩에 대한 투자로 AI First 전략을 실현 중이다. GCP의 클라우드 시장 점유율은 약 12% 정도이며, 2025년 2분기 매출은 136억 달러로 전년 대비 32% 성장하였으며, 연환산 기준 약 544억 달러 규모로 추정된다(CRN, 2025. 8. 5).

세 사업자 모두 AI 모델 학습·추론을 위한 전용 칩 개발, AI 컴퓨팅 인프라 확장, 초거대 모델 기반 서비스 통합을 추진하며 AI 클라우드 시장의 구조적 성장을 견인하고 있다.

4) Oracle Cloud Infrastructure(OCI)

Oracle Cloud Infrastructure(OCI)는 전통적으로 데이터베이스 중심의 클라우드 서비스를 제공해 왔으나, 최근 몇 년 사이 급격하게 AI 인프라 강화 전략을 추진하고 있다. Oracle은 'OCI Supercluster'라는 명칭으로, 최대 13만 개 이상의 GPU를 단일 클러스터로 구성할 수 있는 초대형 인프라 아키텍처를 공개하였

7) <https://aws.amazon.com/ai/machine-learning/trainium/>

다. 이 클러스터는 NVIDIA Blackwell과 같은 차세대 GPU 아키텍처를 지원하며, 초고속 RDMA 네트워킹과 고성능 스토리지 시스템을 결합해 대형 모델 학습과 추론 환경을 효율적으로 구현할 수 있도록 설계되었다(Oracle 한국 홈페이지).⁸⁾ Oracle은 기업 고객과 공공기관을 대상으로 AI 모델 학습 전용 클라우드를 제공하고 있으며, OCI의 클라우드 인프라 매출은 FY2024 4분기 기준 전년 동기 대비 42% 성장하여 20억 달러를 기록하였으며, FY2024 4분기에만 OpenAI와의 계약을 포함한 30건 이상의 AI 판매 계약을 체결하여 총 125억 달러 이상의 계약을 확보하였다. 이러한 AI 수요에 힘입어 미이행 계약 잔고(RPO)는 44% 증가한 980억 달러에 달했다(Oracle Corporation, 2024. 6. 11).

〈표 2-5〉 AI 클라우드(Compute Layer) 주요 기업 및 특징

기업	전략 요약	경쟁력	시장점유율 (AI IaaS 기준 추정)
Amazon AWS	SageMaker, Bedrock 등 모델 호스팅 및 API 제공	대규모 고객 기반, 자체 칩 활용	약 40%
Microsoft Azure	OpenAI와 전략적 제휴. Copilot 서비스 통합	ChatGPT와 Azure 통합 효과	약 30%
Google Cloud	Vertex AI, TPU 기반 ML 플랫폼 제공	고급 사용자 중심, Gemini API 연계	약 15%
Oracle	NVIDIA와 협력, OCI 인프라 제공	대기업 ERP+AI 통합 전략	미미하지만 성장 중

자료: BOND(2025. 5)

5) 기타 사업자

Salesforce, IBM, Tencent, Huawei 등이 약 2%의 점유율을 차지하고 있으며 Akamai, Baidu, China Telecom, China Unicom, CoreWeave, Databricks, Fujitsu, NTT, Snowflake, SAP, VMware 등이 각각 약 1% 내외의 점유율을 보유하고 있다(CRN, 2025. 8. 7).

8) <https://www.oracle.com/kr/ai-infrastructure/>

이들 사업자들은 시장점유율 측면에서는 제한적으로 보이지만, 특정 산업(CRM, 금융, 통신, 중국 내수 등)이나 특정 기술영역에서 니치 전략 및 파트너십 중심의 영향력을 발휘하고 있다. 특히 고성능 GPU, 고대역폭 네트워크 및 대규모 병렬처리가 가능한 데이터센터 인프라를 기반으로 한 AI 인프라 특화 클라우드가 새로운 경쟁 축으로 부상하고 있는데 일례로, CoreWeave는 GPU 클라우드 전문 공급자로, 2017년 설립되어 초기에 암호화폐 채굴에 활용되던 GPU 인프라를 기반으로 고성능 컴퓨팅(HPC) 및 인공지능 연산용 클라우드로 사업을 전환하였다. 범용 클라우드와 달리 AI 학습·추론에 특화된 GPU 인프라를 대규모로 제공한다는 점이 핵심 역량이며 NVIDIA와의 긴밀한 전략적 제휴를 통해 최신 GPU를 조기에 도입하고 이를 기반으로 초대형 GPU 클러스터를 구축하였다. OpenAI, Microsoft, Meta 등 세계적인 AI 연구 및 개발 기관에도 인프라를 제공함으로써 안정적인 매출 성장과 장기적 계약 기반의 수익 구조가 가능해졌고, 2024년 매출은 전년 대비 737% 증가하여 19억 1,500만달러에 도달했으며(CoreWeave, 2025. 3. 14), 2025년 1분기 매출은 9억 8,200만 달러, 전년 동기 대비 420%의 높은 성장률을 보이고 있다(CoreWeave, 2025. 5. 14).

라. AI 모델 시장 현황

ChatGPT에 쏠린 세계적 관심은 AI로의 투자, 인력 집중으로 이어지며 멀티모달 처리 기술의 성숙, AI 추론 능력의 비약적 발전 등 AI 모델 성능의 향상, 나아가 AI 모델 시장의 확장을 이끌어내고 있다.

우선, AI 모델은 텍스트 처리에서 출발하여 이미지, 오디오, 비디오 등 다양한 형태의 데이터를 이해하고 생성할 수 있는 ‘멀티모달’ 역량으로 빠르게 확장하고 있다. 이러한 변화는 인간과 유사한 방식으로 세상을 인지하고 상호작용하는 AI 시스템의 등장을 의미한다. 2022년부터 2024년까지 2년간 대규모 멀티모달 AI 모델의 출시는 1,150%, 대규모 언어 모델(LLM)은 420% 증가하였으며, 같은 기간 전년 대비 기준으로 비전 모델은 109%, 음성/오디오 모델은 367%, 비디오 모델은 120% 각각 증가하였다(Epoch AI, 2025.4, Bond, 2025.5.에서 재인용).

주요 기업들은 GPT- 4o, Claude 3, Chameleon 등 완전히 멀티모달화된 시스템을 속속 출시하고 있다.

또한, 하드웨어 효율성 향상(GPU·TPU의 고도화), 알고리즘 최적화, 그리고 모델 압축, 캐싱 기술 등의 발전은 추론 비용 감소로도 이어지고 있는데(NVIDIA, 2025. 3) 언어 모델 실행에 드는 토큰당 비용은 2022년에서 2024년 사이 99.7% 감소하였다(BOND, 2025. 5).

〈표 2-6〉 Foundation Model 개발 기업

기업	대표 모델	전략 및 경쟁력	특징
OpenAI	GPT-4o, GPT-4, GPT-3.5	Microsoft와 독점 계약, Copilot에 내장	세계적 인지도, 응용 확장
Google DeepMind	Gemini 1.5	검색·광고·클라우드 전반에 통합 전략	자체 클라우드+칩 통합
Anthropic	Claude 3	안정성 중심 모델(constitutional AI)	AWS, Google, Salesforce 투자
Meta	LLaMA 3	오픈소스 전략으로 영향력 확대	커뮤니티와 연구용 확산
Mistral	Mixtral 등	고성능, 경량화 모델에 특화	유럽 기반, 오픈 모델 집중
Cohere	Command-R+	엔터프라이즈 NLP 최적화	데이터 보안 중시 기업 대상
xAI(Elon Musk)	Grok	X(구 트위터) 플랫폼 내 통합	Tesla/SpaceX 데이터 활용 예상

자료: BOND(2025. 5), “Trends - Artificial Intelligence(AI)”, 저자 재작성.

파운데이션 모델 기업들은 구독 기반과 API 활용을 통해 빠른 매출 성장세를 보이고 있다. OpenAI의 2024년 매출액은 37억 달러로 2022년 대비 1,050% 성장하였다(Medium, 2025. 6. 8). Anthropic은 18개월 만에 연간 매출이 20배 증가하여 20억 달러에 도달하였고(Anthropic, 2025. 3. 3), Perplexity는 14개월 만에 매출이 7.6배 증가해 1억 2천만 달러를 기록하였다(Bloomberg, 2025. 7.

18). 하지만, AI 모델 훈련 비용은 지속적으로 증가하고 있어 파운데이션 모델 기업들은 구조적 딜레마에 직면해 있다.

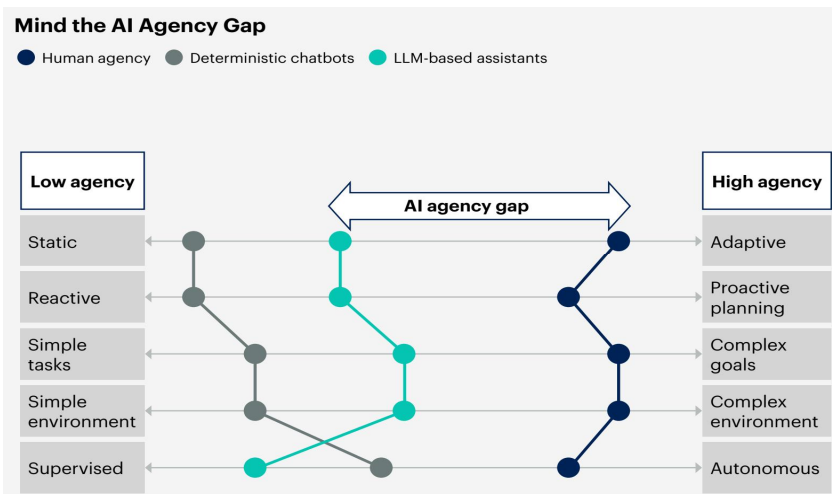
마. AI 서비스 시장 현황

1) 에이전틱 AI(Agentic AI) 시장

AI의 발전으로 이제 AI는 단순한 정보 제공이나 콘텐츠 생성 기능을 넘어 실제 비즈니스 프로세스와 일상 업무를 수행하는 ‘에이전트’ 형태로 진화하고 있다. AI 에이전트는 사용자를 대신하여 추론하고, 실행하며, 다단계 작업을 수행하는 지능형 시스템으로 주어진 목표를 이해하고 이를 하위 작업으로 분해한 뒤, 다양한 주체들과 상호작용하며 작업을 실행하고, 실시간으로 환경 변화에 적응할 수 있다.

Gartner(2024. 10)는 AI 에이전트가 디지털 동료(digital coworker)로 기능하며, 2028년까지 전체 업무 결정의 15%를 수행할 것이라고 전망한다. 이는 RPA(로봇 프로세스 자동화)를 뛰어넘어, 인간 없이도 워크플로우 설계가 가능한 시대가 도래할 것임을 시사한다.

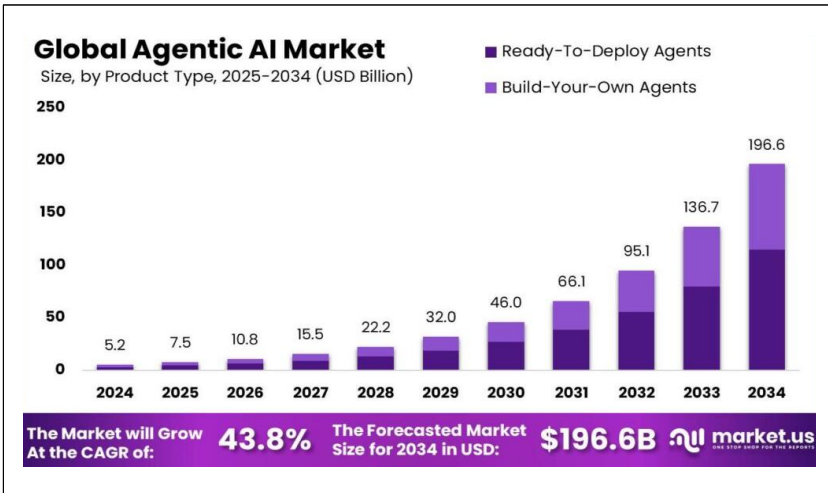
[그림 2-9] 인간과 AI 사이의 역량 차이



자료: Gartner(2024. 10)

복수의 시장 조사 기관에 따르면, 에이전틱 AI 시장은 향후 몇 년간 연평균 40% 이상의 높은 성장률을 기록할 것으로 전망된다. 우선, Market.us⁹⁾에서는 에이전틱 AI 시장이 2024년 52억 달러 규모에서 2034년 1,966억 달러로 확대될 것으로 전망하였다. 이는 연평균성장률 43.8%에 해당하는 폭발적인 성장세로, 에이전틱 AI가 향후 10년간 가장 빠르게 성장하는 기술 분야 중 하나가 될 것임을 의미한다. 지역별로 살펴보면, 2024년 현재 북미 지역이 전체 시장의 38%를 차지하며 선두를 달리고 있으나, 아시아태평양 지역이 향후 가장 빠른 성장세를 보일 것으로 예상되었다.

[그림 2-10] 에이전틱 AI 시장 전망



자료: Market.us(2025)

여러 보고서들 또한 유사한 전망치를 제시하고 있다. Precedence Research (2025. 12)에서는 세계 에이전틱 AI 시장이 2024년 약 52억 5000만 달러에서 연평균 43.84% 성장하여 2034년 190억 500만 달러로 확대될 것으로 전망하였으며, Mordor Intelligence(2025a)는 2025년 69억 6,000만 달러 규모에서 연

9) <https://market.us/report/agentic-ai-market/>

평균 43.61% 성장하여 2030년 425억 6,000만 달러에 이를 것으로 추정하였다. Straits Research(2025)에서는 2025년 83억 1,000만 달러에서 연평균 44.21% 성장하여 2033년 시장 규모를 1548억 4000만 달러 규모가 될 것으로 전망하였다. 해당 보고서는 산업별 채택 속도로는 금융, IT·소프트웨어, 헬스케어 가 가장 빠르며, 특히 금융 분야에서 자동화된 리스크 점검, 고객 응대, 자산 관리 등에서 에이전틱 AI 도입이 가속화될 것이라 분석하였다.

주요 시장 조사기관들의 전망을 종합해 보면 에이전틱 AI가 단순한 생성형 AI 기능 확장 수준을 넘어 기업 운영 자동화, 프로세스 혁신, 자율적 의사결정 등의 수요 증가에 의해 본격적인 상업화 단계로 진입하고 있으며 에이전틱 AI 시장은 향후 10년간 연평균 40~45% 수준의 초고속 성장세를 지속하여 2030년대 중반에는 수천억 달러 규모의 이를 것으로 기대된다.

2) 물리적 AI 시장

물리적 AI(Physical AI)란 센서(카메라, 라이다 등)로 물리 환경을 인지하고, 인지한 정보를 바탕으로 판단하며, 로봇 팔·바퀴 등의 액추에이터로 현실 세계에 행동을 수행하는 AI 시스템을 말한다(SuperbAI, 2025. 6. 25). 로봇, 자율주행차, 드론, 휴머노이드 등 물리적 하드웨어를 수반하면서 AI가 실시간으로 인식·판단·제어를 수행하는 모든 시스템이 물리적 AI 범주에 포함된다. 구현 형태별로는 공장·물류·서비스 현장에서 반복작업을 수행하는 휴머노이드형 로봇, 자율주행차·로보택시·물류배송 로봇 등 자율주행 모빌리티형, 군사·농업·물류·촬영용 드론형, 공장·물류창고 내 AGV/AMR(자율이동로봇)형, 병원·수술실에서 활용되는 수술·의료 로봇형, 호텔·카페·병원 등에서 서비스·안내를 수행하는 서비스 로봇형 등으로 구분할 수 있다(이해수 외, 2025).

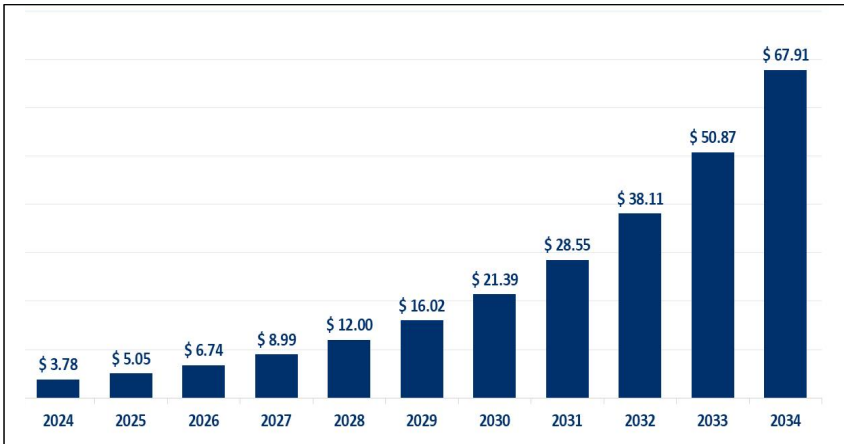
최근 생성형 AI의 발전은 물리적 AI 성능 향상을 가속하며 새로운 전환점을 만들고 있다. 로봇이 자연어 지시를 통해 상황을 이해하고 행동을 계획하며 인간에 가까운 방식으로 작업을 수행할 수 있게 되었으며, 또한 대규모 생성 AI를 활용해 자율주행 학습 데이터를 대량 생성·최적화하거나, 복잡한 물리 시뮬레이션 환경을

정교하게 구축함으로써 물리적 AI의 학습 속도와 범위가 크게 개선되었다 (Kearney, 2025).

엔비디아 CEO 젠슨 황이 2025년 CES 기조연설에서 “Physical AI 시장에 수조 달러 규모의 기회가 있을 것”이라고 하였듯이, 다양한 보고서에서는 산업용 로봇 등 관련 시장의 가파른 성장을 전망하고 있다. Cervicorn Consulting(2025)에서는 물리적 AI 시장이 2024년 37.8억 달러 규모에서 연평균 약 33.5% 성장하여 2034년 679.1억 달러 규모에 이를 것으로 전망한다.

[그림 2-11] 물리적 AI 시장 전망(2024-2034)

(단위: 십억 달러)



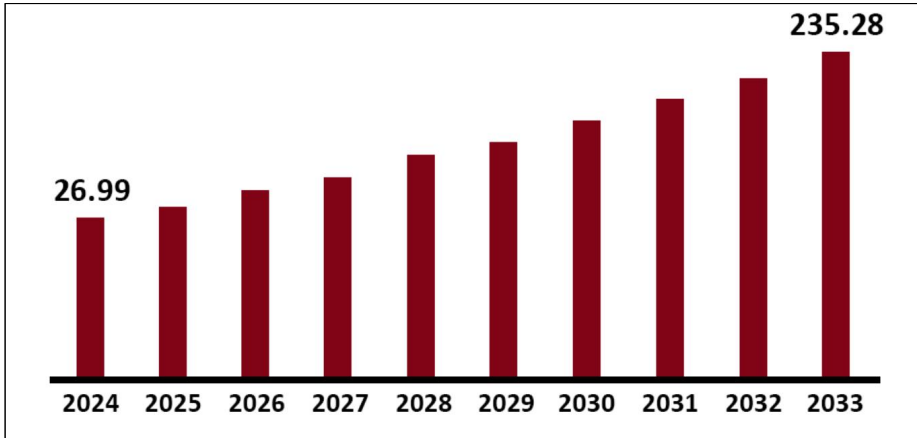
자료: Cervicorn Consulting(2025)

이러한 성장은 AI 기술의 발전, 로봇 하드웨어 비용의 하락, 그리고 자동화에 대한 산업계의 수요 증가가 복합적으로 작용한 결과로 분석되며 특히 코로나19 이후 공급망 안정성과 인력 부족 문제에 대한 인식이 높아지면서 물리적 AI 도입에 대한 관심이 크게 증가하였다.

Astute Analytica(2025)의 전망에 따르면, 글로벌 산업용 로봇 시장 규모는 2024년 약 269억 9천만 달러(약 37조 7천억 원)에서 2033년 2352억 8천만 달러(약 328조 6천억 원)로 9배 가까이 성장할 것으로 예측된다. 물리적 AI의 핵심

응용 분야인 서비스 로봇 시장(물류, 헬스케어, 개인 서비스 등)은 2032년 약 3,890억 달러 규모에 이를 것으로 전망되며, 이는 연평균 약 29%의 성장한 결과이다(Market Research Future, 2024).

[그림 2-12] 산업용 로봇 시장 전망(2024~2033)



자료: Astute Analytica(2025), SuperbAI(2025)에서 재인용

<표 2-7> 물리적 AI(로봇) 시장 규모 및 성장 전망

조사기관	시장 범위	기준연도 시장 규모 (십억달러)	예측연도 시장 규모 (십억달러)	연평균 성장률(CAGR)
Astute Analytica	글로벌 산업용 로봇	2024년 26.99	2033년 235.28	27.1% (2024-2033)
Market Research Future	서비스 로봇 (주요 물리적AI)	2023년 40.5	2032년 388.9	29.02% (2024-2032)
Grand View Research	휴머노이드 로봇	2024년 1.83	2030년 13.8	39.5% (2024-2030)

자료: Astute Analytica(2025), Grand View Research(2025b), Market Research Future (2025)

물리적 AI의 핵심 응용 분야인 자율주행 모빌리티 시장도 빠르게 성장하고 있다. Fortune Business Insights에 따르면, 글로벌 자율주행차 시장은 2023년 약 192억 달러에서 2030년 약 1,363억 달러로 연평균 32.3%의 고속 성장이 기대된다.

<표 2-8> 물리적 AI(자율주행 모빌리티) 시장 규모 및 성장 전망

조사기관	시장 범위	기준연도 시장 규모 (십억달러)	예측연도 시장 규모 (십억달러)	연평균 성장률(CAGR)
Fortune Business Insights	글로벌 자율주행차	2023년 19.21	2030년 136.32	32.3% (2023-2030)
Mordor Intelligence	글로벌 자율주행차	2024년 60.8	2030년 122.04	23.27% (2025-2030)
Business Research Insights	글로벌 자율주행차	2025년 331.62	2034년 6,743.83	39.75% (2025-2034)

주: Business Research Insights는 시장 정의를 타 조사기관에 비해 광범위하게 설정
 자료: Business Research Insights(2025), Fortune Business Insights(2025), Mordor Intelligence(2025b),

Business Research Insights에서는 자율주행차 시장 규모가 2025년 3316억 달러에서 2034년 6조 7천억 달러를 넘어설 것으로 예측하며 연평균 39.75%에 달하는 매우 높은 성장률을 제시하였는데, 이는 자율주행 기술이 단순한 차량 판매를 넘어, MaaS(Mobility as a Service)와 같은 서비스 및 인프라 시장 전체를 포괄하는 거대한 생태계를 형성할 것이라는 기대를 반영한 전망이다.

물리적 AI의 주요 고성장 응용 분야를 살펴보면 첫째, 물류 및 창고 분야이다. 자율 이동 로봇(AMR)과 지능형 지게차는 재고 관리, 주문 처리, 라스트 마일 배송 비용 절감을 위해 적극적으로 투입되고 있다. Amazon은 이미 300개 이상의 물류센터에 100만 대 이상의 로봇을 배치하여 이러한 자동화를 통해 공급망 효율성을 25% 이상 향상시킨 것으로 알려져 있다. 물류 분야에서의 물리적 AI 도입은 24시간 운영, 정확한 재고 관리, 빠른 주문 처리 등의 이점을 제공하며, 전자상거

래의 성장과 함께 지속적인 수요가 예상된다.

두 번째 고성장 분야는 제조 및 자동차 산업이다. 지능형 로봇은 조립, 품질 관리, 예측 유지보수 등에 활용되어 배치 시간을 단축하고 생산 정밀도를 향상시키고 있다. 전통적인 산업용 로봇이 반복적인 단순 작업에 국한되었다면, 물리적 AI가 적용된 지능형 로봇은 다양한 제품과 공정에 유연하게 대응할 수 있다. 특히 다품종 소량 생산이 증가하는 현대 제조 환경에서 이러한 유연성은 매우 중요한 경쟁력이 되고 있다.

세 번째 고성장 분야는 헬스케어로 가장 빠르게 성장하는 응용 분야이다. 로봇 보조 수술 시스템은 외과 의사의 정밀도를 향상시키고 피로도를 감소시키는 데 기여하고 있다. 또한 로봇을 활용한 환자 모니터링, 재활 치료, 노인 돌봄 서비스는 의료 인력 부족과 고령화 사회 문제에 대응하는 중요한 솔루션으로 주목받고 있다. 특히 정밀한 동작이 요구되는 수술 분야와 반복적인 케어가 필요한 재활 분야에서 물리적 AI의 활용 가치가 높게 평가되고 있다.

물리적 AI 시장의 성장에 발맞춰 글로벌 기업들도 적극적인 행보를 보이고 있다. 특히 빅테크와 로봇 스타트업을 중심으로 혁신적인 프로젝트와 투자가 활발하다.

미국 로봇 스타트업 Figure는 BMW와 협력해 휴머노이드 로봇을 미국 스파르탄버그 공장의 생산라인에 투입하였다. 2025년 11월 기준 Figure 02 로봇은 약 11개월 동안 BMW X3 조립라인에서 주 5일, 일 10시간씩 운영되며, 9만 개 이상의 부품 로딩 작업을 수행하여 약 3만 대의 차량 생산에 기여한 것으로 보고된다. Figure는 2024년 오픈AI와 차세대 휴머노이드용 AI 모델 공동 개발을 발표했으나, 2025년에는 자체 대형언어모델(LLM) 개발로 전환하면서 협력을 종결하였다.

아마존은 미국 Agility Robotics의 이족보행 로봇 'Digit'을 물류창고 업무에 테스트 중이며, 2024년부터 미국 일부 물류센터에서 빈 토트(상자) 이동·정리 등 반복 작업을 수행하도록 파일럿을 진행하고 있다.

메르세데스 벤츠는 미국 휴머노이드 로봇 기업 Apptронik에 투자하며 'Apollo' 휴머노이드 로봇을 도입하여 공장에서 부품 운반·품질 검사 등 물류·인트라로지

스틱스 작업에 시험적용하고 있다.

중국은 2024년 기준 글로벌 신규 산업용 로봇 설치의 50% 이상을 차지하고, 전 세계 로봇 관련 특허의 3분의 2를 보유할 정도로 물리적 AI·로봇에서 공격적 투자를 이어가고 있다. 주요 기업으로는 유니트리(Unitree), Deep Robotics, EngineAI, AgiBot 등이 있으며, 정부는 2050년까지 5조 달러 규모의 로봇 시장을 목표로 대규모 정책 지원을 추진 중이다

〈표 2-9〉 주요 해외 기업들의 물리적 AI 제품·서비스 출시 및 개발 현황

기업	제품/서비스명	핵심 전략 및 특징
Figure AI & BMW	Figure 02 휴머노이드 로봇(자동차 공장 부품 로딩·조립 보조)	BMW 스파르탄버그 공장에 11개월간 상시 투입, 주5일 10시간 근무, 9만 개 이상 부품 로딩·약3만 대 차량 생산에 기여. 실제 양산 라인에 휴머노이드를 투입한 대표적 사례
Agility Robotics (Amazon 파트너)	이족보행 휴머노이드 Digit(물류센터에서 빈 토트 상자 이송 등 반복 작업)	Amazon 물류센터 내 파일럿을 통해 사람 대신 반복적·육체적 작업 수행. 기존 AMR·컨베이어와 결합한 “물류 특화 휴머노이드”
Appttronik & Mercedes-Benz	휴머노이드 Apollo(부품 운반, 조립 키트 배송, 품질 검사 등 인프라로지스틱스 용도)	메르세데스-벤츠와 상용화 파트너십 체결 후 독일 Berlin·헝가리 공장에 파일럿 도입. 위험·반복 작업 대체 및 노동력 부족 해소를 목표로 휴머노이드 도입 확대
Unitree Robotics(중국)	휴머노이드 H1, G1, 소비자형R1 등; 4족 보행 로봇 시리즈	중국의 대표적 휴머노이드·4족 보행 로봇 기업. 저가(수천 달러 수준) 휴머노이드 상용화로 대중 시장을 겨냥

자료: Figure AI(2025. 11. 19), Botinfo.ai(2025. 9. 19), Appttronik(2024. 3. 15), Unitree 사이트

국내에서는 대기업이 로봇 전문기업에 전략적 투자를 확대하면서, 휴머노이드·서비스 로봇·모빌리티 로봇 등 물리적 AI 전 영역에서 대기업 중심의 통합 생태계가 빠르게 재편되고 있다. HD현대로보틱스는 한국을 대표하는 산업용 로봇 기업으로, 핸들링·가공·용접용 다관절 로봇과 디스플레이용 FPD 로봇 등 제조용 로

봇 풀 라인업을 보유하고 있으며, 공장 자동화·설계·시운전·A/S를 포함한 토털 솔루션을 제공한다. 최근에는 자율 이동 서비스 로봇 등으로 포트폴리오를 확장하고 있다. 두산로보틱스(Doosan Robotics)는 협동로봇(cobot) 분야 강자로, 2023년 기준 글로벌 협동로봇 시장(중국 제외) 점유율 약 5.4%를 기록하며 세계 4위로 평가된다. 특히 20kg 이상 고하중 협동로봇 시장에서 약 72% 점유율을 보이며 물류·제조 현장의 물리적 AI 도입을 선도하고 있다. 레인보우로보틱스(Rainbow Robotics)는 KAIST 휴머노이드 'HUBO' 계열을 기반으로 한 양발 보행 휴머노이드·협동로봇 기업으로, 2024년 말 삼성전자가 약 2,670억 원을 추가 투자해 최대주주(지분 최대 59% 수준)가 되었다. 삼성전자는 '퓨처 로보틱스 오피스'를 CEO 직속 조직으로 신설하며, 레인보우로보틱스를 축으로 차세대 휴머노이드·서비스 로봇 개발에 착수하였다. LG전자는 호텔·병원·카페·박물관 등을 대상으로 안내·배달·광고·보안 기능을 수행하는 클로이(CLOi) ServeBot·GuideBot 등 다양한 서비스 로봇을 상용화하고 있다. CLOi ServeBot은 자체 자율주행 기능을 탑재한 도어형 배달 로봇으로, 병원·호텔에서 음식·물품 전달 업무를 수행한다. 뉴로메카(Neuromeka)는 중소 제조업 대상 가격 경쟁력 있는 협동로봇 'Indy' 시리즈와 델타 로봇 'LASER' 시리즈를 공급하는 국내 협동로봇 전문기업이다. 스마트팩토리·물류창고·연구용 로봇 등 다양한 환경에서 사용가능하며, 사람-로봇 협업 환경에 최적화된 솔루션을 제공한다. 현대차그룹은 Boston Dynamics 인수를 통해 4족 보행 로봇(Spot), 물류 로봇(Stretch), 전기 구동 휴머노이드 Atlas를 확보하고 있으며, 2025년 이후 그룹 내 공장에 Atlas를 시범 배치하는 계획을 추진하고 있다.

〈표 2-10〉 국내 기업의 물리적 AI 서비스 개발 동향

기업	제품/서비스명	핵심 전략 및 특징
HD현대 로보틱스 ¹⁰⁾	산업용 다관절 로봇 (핸들링·가공·용접 등), FPD 로봇, 협동로봇, 모바일 서비스 로봇	<ul style="list-style-type: none"> • 국내 대표 산업용 로봇·FA(Factory Auto- mation) 기업 • 공장 자동화 전 공정을 아우르는 토털 솔루션 (설계·설치·시운전·A/S) 제공 • 향후 서비스·모바일 로봇까지 확장 계획
두산로보 틱스 ¹¹⁾	협동로봇 M·H·P 시리즈, 고하중 팔레타이징용 P3020 등	<ul style="list-style-type: none"> • 글로벌 협동로봇 시장(중국 제외) 점유율 약 5.4%로 세계4위 • 고하중 팔레타이징 협동로봇 분야에서 약72% 시장점유율
레인보우 로보틱스 ¹²⁾¹³⁾	휴머노이드 HUBO2, DRC-HUBO, 대형 휴머노이드FX-2, 협동로봇RB 시리즈, 4족 로봇RBQ	<ul style="list-style-type: none"> • 삼성전자가 최대주주로 올라서면서 차세대 휴 머노이드·제조용 로봇 공동 개발 추진 • ‘퓨처 로보틱스 오피스’ 신설로 삼성 로봇 전략 의 핵심 파트너
LG 전자 ¹⁴⁾¹⁵⁾	서비스 로봇 브랜드 LG CLOi시리즈 - ServeBot (배달·서빙), GuideBot(안내·광고·보안), BaristaBot 등	<ul style="list-style-type: none"> • 호텔·병원·도서관·쇼핑몰 등에서 안내·배달·광 고·보안 업무 수행 • 최근에는 Google Gemini 기반 생성형AI를 탑재한 GuideBot 공개

10) <https://www.hd.com/en/business/machine/hyundai-robotics/contents>

11) 한국경제(2025. 9. 12.), “Doosan Robotics aims to draw \$256 mn in IPO subscriptions”

12) <https://www.rainbow-robotics.com/>

13) Samsung Newroom(2024. 12. 31.), “Samsung Electronics To Become Largest Shareholder in Rainbow Robotics Accelerating Future Robot Development”
<https://news.samsung.com/global/samsung-electronics-to-become-largest-shareholder-in-rainbow-robotics-accelerating-future-robot-development>

14) 한국경제(2024. 7. 1.), “LG unveils Google’s generative AI- powered robot CLOi”

15) <https://www.lgcorp.com/media/release/27812>

기업	제품/서비스명	핵심 전략 및 특징
뉴로메카 (Neuromeka) ¹⁶⁾	협동로봇 Indy시리즈, 차세대ICoN 협동로봇, 델타 로봇LASER 시리즈, 자율이동로봇 등	<ul style="list-style-type: none"> • 중소 제조업을 주요 타겟으로 “쉽고 경제적인 로봇 자동화” 지향 • 최근 ‘자율지능AI 휴머노이드’ 개발도 추진 중
현대차 그룹 ¹⁷⁾	Boston Dynamics의 Spot, Stretch, 향후 Atlas 계열 휴머노이드, 그룹 자체 서비스 로봇(DAL-e 등)	<ul style="list-style-type: none"> • Boston Dynamics 인수 이후 공장 품질검사·물류·건설 현장에 Spot·Stretch를 시범 적용하고, 미국 내 로봇 생산기지 구축 • 국내에서는 DAL-e Delivery 등 서비스 로봇을 통해 모빌리티·건축·물류·서비스 전반에서 로봇-인간 협업 모델 실증

자료: 각사 홈페이지 및 언론보도

16) <https://en.neuromeka.com/>

17) <https://www.hyundai.com/worldwide/en/brand-journal/mobility-solution/hyundai-boston-dynamics>

제3장 주요국 AI 생태계 분석

AI가 국가 기술패권 확보의 핵심 자산으로 대두됨에 따라 AI 경쟁력 확보를 위한 국가 간 경쟁도 치열해지고 있다. 특히 트럼프 2기 행정부 취임 이후 AI를 둘러싼 국가 간 패권 경쟁은 더욱 심화되는 양상으로 이는 주요국 AI 정책 기조에도 변화를 야기하고 있다. 본 절에서는 AI를 둘러싼 국가 간 경쟁양상을 이해하기 위해 주요국 AI 생태계 및 정책 현황을 분석한다.

제1절 미국

1. AI 산업생태계 구조 및 주요 플레이어

미국 AI 생태계는 빅테크 기업, 풍부한 벤처 캐피털(VC), 세계 최고 수준의 대학, 그리고 정부의 전략적 지원이라는 네 개의 강력한 축이 상호작용하며 전 세계 AI 혁신을 주도하는 독보적인 구조를 갖추고 있다.

첫째, 미국은 NVIDIA, Google, Microsoft, Meta, Apple 등 현재 AI 생태계의 핵심적인 인프라 제공자이자, 원천 기술 연구와 상용화의 선두 주자들을 기반으로 AI 풀스택에 걸친 영향력과 통제력을 강화해 나가고 있다. NVIDIA는 AI 모델 학습에 필수적인 GPU 시장을 사실상 독점하고 있으며, Amazon(AWS), Microsoft(Azure), Google(GCP)은 막대한 클라우드 컴퓨팅 자원을 통해 전 세계 AI 개발의 기반을 제공한다. 또한, OpenAI, Anthropic, Google, Meta, Microsoft(OpenAI와 협업) 등은 최첨단 파운데이션 모델 개발 경쟁을 주도하며 기술적 한계를 넓히고 있으며, AI와 자사 제품·서비스의 통합, AI 서비스의 플랫폼화 등 플랫폼·응용 서비스 스택에서 AI 인프라 비용을 상쇄할 수 있는 AI 수요, 수익모델 창출을 위해 노력하고 있다. 종합해 보면 미국 AI 생태계는 AI 전 기술

스택에 걸쳐 독점적 영향력을 갖춘 거대 빅테크 기업들을 보유하고 있을 뿐 아니라, 이 기업들이 칩-클라우드-모델-서비스가 유기적으로 연결된 수직통합형 생태계를 구축하고 있다는 점에서 기술 표준 설정과 시장지배력 강화, 데이터 축적 및 네트워크 효과의 선순환을 동시에 실현할 수 있는 구조적 경쟁우위를 갖는다.

둘째, 미국은 ‘모험 자본’으로 불리는 벤처 캐피털(VC)을 통해 아이디어가 기술과 만나고, 기술이 기업으로 성장하는 가장 역동적인 시스템을 갖추고 있다. 2024년 2분기 기준, 미국은 전 세계 AI 관련 VC 투자액의 약 66%를 차지하며(CBInsights, 2024. 7. 30) ‘인내 자본(patient capital)’의 특성을 가지고 미국 AI 생태계에 실험적이고 도전적 시도가 가능하도록 지원한다.

[그림 3-1] 글로벌 AI 투자액과 미국 AI 투자액



자료: CBInsights(2024. 7. 30)

이러한 투자생태계를 바탕으로 미국에는 OpenAI, Anthropic, Databricks를 비롯해 1,000개가 넘는 AI 유니콘(기업가치 1조 원 이상 비상장 기업)이 존재하며, 실리콘벨리 외에도 뉴욕 등 새로운 허브가 계속해서 등장하고 있다. 이들은 빅테크가 개척한 영역을 파고들거나 새로운 시장을 창출하며 생태계에 활력을 불

어닝을 역할을 수행한다.

셋째, 스탠퍼드, MIT, 카네기 멜런, UC 버클리 등 세계 최고 수준의 대학들은 미국 AI 생태계의 '두뇌'로 기능한다. 이들 대학은 전 세계에서 가장 뛰어난 인재들을 끌어모아 AI 연구자 및 개발자로 양성하며, 졸업생들은 빅테크나 스타트업의 핵심 인력으로 흡수되거나, 직접 스타트업을 창업하며 선순환 구조를 만든다. 딥러닝을 포함한 수많은 AI의 근간 기술이 이들 대학 연구실에서 탄생했는데, 학계에서 공개한 논문과 오픈소스 연구는 산업계가 새로운 아이디어를 얻고 기술을 발전시키는 중요한 자양분이 되고 있다.

넷째, 미국 정부는 직접적인 시장 개입보다는 민간의 혁신을 촉진하고 국가 안보와 경쟁력을 강화하는 방향으로 AI 생태계를 지원한다. 국립과학재단(NSF), 방위고등연구계획국(DARPA) 등은 장기적인 관점에서 대학과 연구소의 기초 연구 및 고위험·고수익 연구에 자금을 지원하여 미래 기술의 씨앗을 뿌리며, 최근 발표된 '미국의 AI 액션 플랜(America's AI Action Plan)'에서와 같이 정부는 규제 완화, 인프라 투자 촉진, 국제 표준 선도 등 민간이 마음껏 펼 수 있는 운동장을 만드는 데 주력한다. 이는 기업 친화적인 환경을 조성하여 미국의 AI 리더십을 공고히 하려는 전략이다.

결론적으로, 미국 AI 생태계의 가장 큰 강점은 이 네 가지 요소가 유기적으로 결합하여 아이디어 발굴 → 인재 양성 → 기술 개발 → 자본 투자 → 시장 창출로 이어지는 강력한 선순환 구조를 구축했다는 데 있으며 이는 다른 국가들이 쉽게 모방하기 어려운 미국의 핵심 경쟁력이다.

2. AI 정책 및 투자 현황

가. 미국 정부의 AI 정책 방향

미국 정부의 AI 정책은 '글로벌 리더십 유지'라는 일관된 목표 아래 행정부의 변화에 따라 그 방법론을 달리하고 있다. 특히 트럼프 행정부는 시장 주도 혁신과 더불어 AI 안전과 신뢰를 강조했던 바이든 행정부와 달리 AI 기술 개발 속도를

가속화하고, AI 모델 개발·활용에서 미국 기업이 주도권을 유지할 수 있도록 규제 완화에 중점을 둔 정책을 펴고 있다. 이는 미중 기술 패권 경쟁이 심화되는 가운데 AI 기술의 발전 속도를 최대한 끌어올리려는 전략적 판단으로 보인다.

이러한 정책 기조가 구체화된 것이 2025년 7월 발표된 ‘미국의 AI 액션플랜 (America’s AI Action Plan)’(The White House, 2025. 7)이다. 이 계획은 AI 분야 미국의 글로벌 지배력을 유지하고 강화하는 것을 최우선 목표로 삼고 있으며, 다음과 같이 세 가지 핵심축으로 구성되어 있다.

첫째, 혁신 가속화(Accelerating Innovation)로, AI 개발과 도입을 저해하는 연방 정부의 과도한 규제를 식별하고 수정하거나 폐지하는 데 중점을 둔다. 또한, 기업 친화적인 환경을 조성하여 민간 부문의 AI 기술 혁신을 촉진하는 것을 목표로 한다.

둘째, 미국 내 AI 인프라 구축(Building American AI Infrastructure)이다. 데이터센터, 반도체 제조 시설 등 AI 산업의 핵심 인프라에 대한 연방 정부의 허가 절차를 간소화하고 신속하게 처리하여 AI 발전에 필수적인 물리적 기반을 미국¹⁸⁾내에 빠르게 확충한다는 계획이다.

셋째, 국제 외교 및 안보 리더십 확보(Leading in International Diplomacy and Security)이다. 동맹국 및 파트너 국가들과의 협력을 강화하고 미국기업이 구축한 AI 풀스택(하드웨어, 모델, 소프트웨어 등)의 해외 수출을 지원하여 미국이 구축한 AI 기술을 글로벌 표준으로 확산하며, 동시에 미국의 첨단 AI 기술이 적대국으로 확산되는 것을 차단하기 위해 수출 통제를 강화한다.

미정부는 AI경쟁은 속도전이라는 인식하에 AI 액션 플랜의 즉각적 이행을 위해 “데이터센터 인프라 허가 가속화”(The White House, 2025. 7. 23a)¹⁹⁾, “미국 AI 기술 스택 수출 증진”(The White House, 2025. 7. 23b)²⁰⁾, “이념 편향적

18) <https://www.whitehouse.gov/presidential-actions/2025/07/accelerating-federal-permitting-of-data-center-infrastructure/>

19) <https://www.whitehouse.gov/presidential-actions/2025/07/accelerating-federal-permitting-of-data-center-infrastructure/>

AI 방지”(The White House, 2025. 7. 23c)²¹⁾ 등을 담은 구체화된 행정명령도 함께 발표하였다.

〈표 3-1〉 AI 액션 플랜 이행을 위한 행정명령

구분	목적	조치내용
데이터센터 인프라 허가 가속화	AI 인프라 병목 해소 및 국내 생산역량 강화	<ul style="list-style-type: none"> • AI 데이터센터 및 관련 전력 인프라에 대한 연방 인허가 절차 단축 • 환경심사 절차 간소화·신속 처리 • FAST-41 절차(연방 인허가 절차를 신속·조정하기 위한 법적 체계) 적용 확대 • 연방 토지 활용 촉진
미국 AI 기술 스택 수출 증진	글로벌 AI 표준 선점 및 기술 리더십 강화	<ul style="list-style-type: none"> • 반도체·클라우드·AI 모델 등 AI 풀스택의 해외 수출 촉진 프로그램 설계 • 산업계 컨소시엄 구성 지원 • 우방국 대상 협력 및 확산 추진
이념 편향적 AI 방지	연방 AI 활용의 정치적 중립성 확보	<ul style="list-style-type: none"> • 연방정부 조달 AI 시 정치·이념적 중립성 요구 • 조달 계약 시 편향 관련 기준 명시 • 공급업체에 모델의 중립성·사실성 관련 정보 제출 요구

자료: The White House(2025. 7. 23a, 2025. 7. 23b, 2025. 7. 23c)

또한, 미국의 AI 액션 플랜에서 언급된대로 미국은 자국 AI 산업의 압도적 주도권을 지키기 위한 수단으로 강력한 통상 정책을 추진하고 있다. 미국 통상 정책의 가장 강력한 무기는 수출 통제로 이는 중국 AI 기술 발전의 ‘엔진’ 역할을 하는 첨단 반도체 공급망을 직접 차단하는 고강도 전략이다. 주요 타겟은 NVIDIA의 A100·H100·H20, AMD의 MI300 등 AI 모델 학습에 필수적인 고성능 GPU와 이러한 반도체를 만드는 데 필요한 첨단 제조 장비로, 중국이 거대 AI 모델을 개발하고 고도화하는 데 필요한 물리적 인프라 자체를 무력화시켜 기술

20) <https://www.whitehouse.gov/presidential-actions/2025/07/promoting-the-export-of-the-american-ai-technology-stack/>

21) <https://www.whitehouse.gov/presidential-actions/2025/07/preventing-woke-ai-in-the-federal-government/>

발전 속도를 늦추는 것이 주된 목표이다. 이는 중국의 군사적 AI 활용 가능성을 차단하려는 국가 안보적 목적과 직결된다. 주요 집행근거는 미 상무부 산업안보국(BIS)이 관할하는 수출관리규정(EAR: Export Administration Regulations)으로, 이 규정에 따르면 미국 기업이 특정 성능 기준 이상의 반도체나 제조 장비를 중국에 수출하려면 상무부 산업안보국(BIS)의 수출 허가를 받아야 하며, 원칙적으로 거부(presumption of denial) 방침이 적용된다. ‘반도체과학법(CHIPS and Science Act)’ 역시 미국 정부의 보조금을 받는 기업이 중국 내 첨단 반도체 제조시설을 신설, 확장하지 못하도록 하는 ‘가드레일’ 조항을 포함하고 있다.

또 다른 강력한 무기는 ‘관세’ 및 ‘블랙리스트’의 활용이다. 관세는 반도체 수출 통제처럼 직접적인 타격을 주기보다는, 중국의 기술 생태계 전반에 비용 압박을 가하고 자국 산업을 보호하는 역할을 한다. 미국은 무역법 301조에 따라 IT 하드웨어, 통신 장비, 드론 등 광범위한 중국산 기술 제품에 높은 관세를 부과하고 있다. 이는 중국 기업의 가격 경쟁력을 약화시키고, 글로벌 기업들이 공급망을 중국 밖으로 이전하도록 유도하는 효과를 낳는다. 또한, 화웨이, 센스타임, 하이크비전 등 중국의 대표적인 AI 및 기술 기업들을 수출 통제 명단에 올려 미국 기술에 대한 접근을 사실상 차단한다. 이는 특정 기업을 정밀 타격하여 고립시키는 효과적인 수단이다.

그 외에도 미 정부는 미국 자본과 기술이 중국의 AI 발전에 기여하는 것을 막기 위한 정책으로 특정 첨단 기술분야에 대한 해외 투자를 제한하고 있다. 미정부는 2023년 미국 기업이나 투자자가 중국의 AI, 양자컴퓨팅, 반도체 등 첨단 기술 분야에 지분을 투자하는 것을 금지하거나 제한하는 투자 제한 행정명령(The White House, 2023. 8. 9)²²⁾을 도입하였는데, 이는 중국 AI 스타트업들의 자금 조달을 어렵게 만드는 효과가 있다.

22) <https://www.federalregister.gov/documents/2023/08/11/2023-17449/addressing-united-states-investments-in-certain-national-security-technologies-and-products-in>

결론적으로, 미국의 대중국 AI 통상 정책은 수출 통제로 핵심 기술의 숨통을 조이고, 관세와 블랙리스트로 광범위한 압박을 가하며, 투자 제한으로 미래의 자금줄까지 차단하는 다층적이고 종합적인 ‘경제·안보 전략’이라고 할 수 있다.

나. 미국 정부의 AI 투자 현황

미국 정부는 직접적인 R&D 자금 지원과 민간 투자를 유도하는 산업 정책을 결합한 ‘투트랙’ 전략을 통해 AI 기술 패권을 공고히 하고 있다. 미국 정부의 AI 투자는 크게 세 가지 주요 전략적 위에서 수행된다.

1) 국가 AI 이니셔티브 법(United States Congress, 2021)²³⁾

미국이 신뢰할 수 있는 AI 시스템의 개발과 활용에서 세계적인 지위를 확보하기 위해 연방정부 차원의 AI 연구·조정·전략을 법제화한 첫 국가 AI 법률로 2020년 의회를 통과하고 2021년 1월 1일부터 시행되었다. 이 법에서는 연방정부 전체의 AI 활동을 전략적으로 조율하기 위한 프레임워크로서 미국 대통령이 National Artificial Intelligence Initiative를 수립하도록 규정하며, 구체적으로는 초기관 조정위원회 및 조정 사무국 설치, 자문위원회 신설, 국립과학재단(NSF)을 중심으로한 AI 연구소 네트워크 구축 및 공공·민간의 파트너십 촉진, NIST를 통한 AI 평가 및 표준 개발 및 데이터 공유 모범사례 수집, AI 분야 교육·인력양성 프로그램 지원 등의 내용을 담고 있다. 이 법은 미국 내 AI 연구개발을 가속화하고, 자원과 인재를 육성하며, 신뢰할 수 있는 AI 생태계를 구축하기 위한 범정부 차원의 장기적 토대를 제공한다.

2) 반도체과학법(CHIPS and Science Act, 2022)

반도체뿐만 아니라 AI, 양자 컴퓨팅 등 첨단 기술 연구에 향후 10년간 약 2,800억 달러를 투자하는 초대형 산업 육성법으로 이는 AI 연구에 필요한 컴퓨팅 파워와 하드웨어 기반을 국가적으로 강화하는 역할을 수행한다.

23) https://science.house.gov/2020/12/hr-6216-national-artificial-intelligence-initiative-act-2020?utm_source=chatgpt.com

〈표 3-2〉 반도체과학법을 통한 주요 지원 사항

항목	내용
응용과학 지원	<ul style="list-style-type: none"> • 투자액의 4분의 1은 NSF가 올해 설립한 기술혁신국(TIP)에 투입 • 인공지능, 양자 컴퓨팅 등 특정 기술의 시장화 과정을 지원하는 것으로, 스타트업의 죽음의 계곡 극복 지원
기초연구, 과학교육, 연구기회 확대 등 지원	<ul style="list-style-type: none"> • 그 외 예산은 기초연구, 과학교육, 연구 기회 확대 등에 투자 • 주로 과학 소외학생 지원 기관의 보조금으로 운영 • 연구 역량을 강화하고, 연구기금이 상대적으로 부족한 지역 내 학술 기관 지원
과학부 기금	<ul style="list-style-type: none"> • 에너지부(DOE)에 승인된 약 700억 달러 대부분은 과학부 기금으로 활용 • 국립 연구소, 대학 및 민간 기업에서 핵융합 및 핵물리학과 같은 물리 과학 분야 지원

자료: 한국과학창의재단 사이트

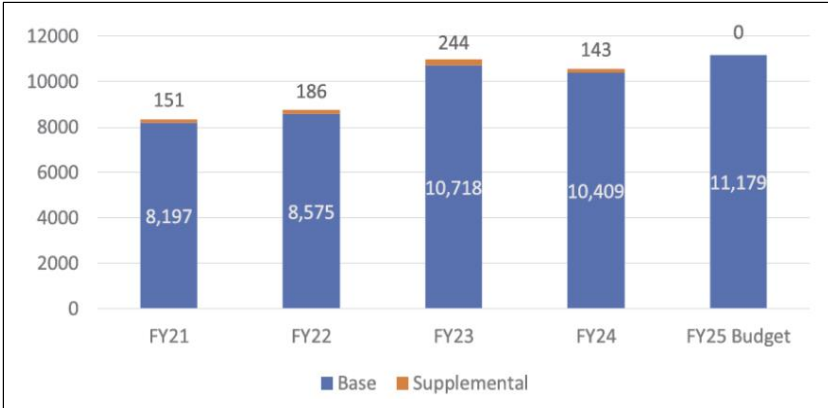
3) 미국의 AI 액션 플랜(America's AI Action Plan, 2025)

앞서 언급한 AI 액션 플랜은 직접적인 자금 지원을 넘어, 규제를 완화하고 데이터센터, 전력망 등 AI 연구·혁신 인프라 구축을 가속화하여 민간 투자를 극대화하는 데 초점을 맞추고 있다. 구체적인 투자계획이 명시되어 있지는 않지만 연방 부처의 예산, 프로그램, 행정명령 등을 통해 AI 혁신 생태계 전반(연구, 인프라, 인력, 통상 전략)의 체계적 지원 및 조정을 뒷받침한다.

미국 2025 회계연도(FY25) AI 관련 예산은 2025년 11월 27일, 백악관 과학기술정책국(OSTP) 산하 국가과학기술위원회(NSTC)가 대통령의 2025 회계연도 예산안에 대해 의회에서 요구하는 보충 자료를 제공하면서 공개되었는데 이에 따르면 인공지능(AI), 네트워크, 슈퍼컴퓨팅 및 빅데이터 분야의 혁신적 발전을 지원하기 위한 연방 정부의 연구개발(R&D) 투자는 기존의 R&D 중심 기관뿐만 아니라 정부 전반으로 계속해서 확대되고 있다. 2021~2025 회계연도 동안 기밀 프로그램을 제외한 모든 연방 정부의 IT 및 AI R&D 예산은 총 28억 달러 증가했으며 이는 연평균 6%씩 꾸준히 성장한 결과이다.

[그림 3-2] 미 연방정부의 IT 및 AI R&D 예산(2021-2025)

(단위: 백만 달러)



자료: Federal budget(2025.1.7)

모든 기관 중 예산증액 규모가 가장 큰 기관은 국립과학재단(NSF)으로 이는 미국 정부가 AI 원천기술 확보에 사활을 걸고 있다는 것을 보여준다. 단일 기관 중 최대 비중을 차지한 기관은 전체 예산의 27%를 확보한 국립보건원(NIH)으로, 이 중 절반에 가까운 14억 달러는 대규모 데이터 관리에 사용할 것으로 알려졌다.

<표 3-3> 미국 정부의 2025년 AI R&D 예산

기관명	예산 규모 및 특징 (FY25 기준)
국립과학재단(NSF)	<ul style="list-style-type: none"> • 20.5억 달러. 전년 대비 가장 큰 폭(+2.85억 달러)으로 증가 • AI 집중 연구 분야(\$4억 → \$4억 9,400만) • ‘차세대 통신 네트워크 및 시스템’과 ‘대규모 데이터 관리 및 분석’을 포함한 7개 카테고리에도 각각 1,200만~2,700만 달러의 예산 증액 계획
국립보건원(NIH)	<ul style="list-style-type: none"> • 30.5억 달러 (전체의 27%) 단일 기관 최대 비중 • 대규모 데이터 관리(\$14억)와 고성능 컴퓨팅 인프라 및 응용(\$4억 2,100만 달러)에 집중
국방부(DOD, DARPA 제외)	<ul style="list-style-type: none"> • 20.3억 달러로 이론적인 연구보다 통신, 로봇처럼 전장에서 즉시 활용가능한 하드웨어와 시스템에 집중 • 차세대 통신 네트워크 및 시스템(\$4억 800만 달러)과 지능형 로봇 및 자율 주행 시스템(\$2억 9,800만 달러)에 중점 투자

기관명	예산 규모 및 특징 (FY25 기준)
국방고등연구계획국 (DARPA)	<ul style="list-style-type: none"> • 14.1억 달러 • 네트워크용 전자 기기 R&D에 집중
에너지부(DOE)	<ul style="list-style-type: none"> • 15.4억 달러 • 고성능 컴퓨팅 인프라 및 응용, 이를 위한 지원 R&D에 예산의 2/3 인 10.3억 달러 투자
국립표준기술연구소 (NIST)	<ul style="list-style-type: none"> • 2억 6,500만 달러
농무부(USDA)	<ul style="list-style-type: none"> • 2억 1,600만 달러
항공우주국(NASA)	<ul style="list-style-type: none"> • 1억 2,100만 달러
해양대기청(NOAA)	<ul style="list-style-type: none"> • 1억 600만 달러
그 외 15개 기관	<ul style="list-style-type: none"> • 4억 4,500만 달러

자료: Federal budget(2025. 1. 7.)

또한, 국가과학기술위원회(NSTC)는 IT 및 AI R&D 지출을 12개의 기능적 범주로 체계화하였는데 그 결과는 다음의 표와 같다.

〈표 3-4〉 미국 R&D 예산 및 AI 펀딩의 기능별 분류(2022-2025)

(단위: 백만 달러)

Category/PCA	FY22	FY23 Actual	FY24 Enacted	FY25 Blden	FY25 +/- FY24	Percent change
Advanced Communications Networks and Systems (ACNS)	569	967	839	792	-47	-6%
Artificial Intelligence (AI) (core funding)	1,838	1,990	1,896	1,955	59	3%
Computing-enabled Human Interaction, Communication and Augmentation (CHUMAN)	637	612	609	655	46	8%
Computing-enabled Networked Physical Systems (CNPS)	150	172	214	230	16	8%
Cyber Security and Privacy (CSP)	722	712	692	778	86	12%
Education and Workforce (EDW)	426	461	458	495	37	8%
Electronics for Networking and IT (ENIT)	527	911	838	991	153	18%
Enabling R&D for High-capability computing systems (EHCS)	575	909	951	943	-8	-1%
High-capability Computing Infrastructure & Applications (HCIA)	1,654	1,526	1,146	1,582	436	38%

Intelligent Robotics and Autonomous Systems (IRAS)	415	507	490	476	-14	-3%
Large-scale Data Management and Analysis (LSDMA)	1,551	1,807	1,826	1,890	64	4%
Software Productivity, Sustainability and Quality (SPSQ)	344	389	383	394	11	3%
Total*	9,409	10,963	10,342	11,181	839	8%

자료: Federal budget(2025. 1. 7)

주요 투자분야는 AI, 대규모 데이터 관리 및 분석(Large-scale Data Management), 고성능 컴퓨팅 인프라(High-Capability Computing Infrastructure and applications) 고성능 컴퓨팅 시스템을 위한 지원 R&D(Enabling R&D for High-capability computing systems), 네트워킹 및 IT용 전자 기기(Electronics for Networking and IT)로 각각 전체 예산의 18%, 17%, 14%, 9%, 9%를 차지하는 것으로 나타났다. 2024 회계연도 대비 가장 큰 폭의 증액은 고성능 컴퓨팅 인프라 분야로 4억 3,600만 달러의 예산이 증가(38% 증가)했으며, 두 번째로 큰 증액은 전자 기기 및 네트워킹분야로, 1억 5,300만 달러가 증가(18% 증가)했다. 이러한 결과는 미국 정부가 AI를 단순한 SW 기술로 보지 않고 데이터, 기초연구, 하드웨어/에너지, 안보를 결합한 국가 전략자산으로 관리하고 있음을 보여준다. 또한 인프라에 대한 막대한 투자는 미국 정부가 단순히 ‘소프트웨어 지능(AI)’ 개발에 만족하지 않고, 이를 뒷받침할 슈퍼컴퓨터와 데이터센터 인프라를 국가 차원에서 압도적으로 구축하겠다는 강력한 신호로 볼 수 있다.

3. AI 규제 현황

가. 기본 철학: AI 패권 경쟁과 혁신 우선주의

미국의 AI 규제 철학은 2025년 트럼프 행정부 출범과 함께 명징한 전환을 맞이하였다. 그 핵심은 “미국이 글로벌 AI 패권을 달성하고 유지하여 인간 번영, 경제 경쟁력, 국가 안보를 촉진한다”는 목표 아래, AI를 지정학적 경쟁의 핵심 수단으로 인식하는 것이다(The White House, 2025). 트럼프 행정부는 AI 규제를 혁

신의 장벽으로 규정하고, 시장 친화적 접근과 기업 자율성 극대화를 통한 글로벌 AI 패권 유지를 일관된 전략 목표로 설정하였다. EU가 기본권 보호를, 중국이 국가 통제를 각각 규제의 중심 가치로 삼는 것과 달리, 미국은 혁신과 경쟁력을 전면에 내세우는 차별화된 철학을 택하고 있다.

그러나 연방 차원의 규제 완화 기조와 달리 주정부 차원에서는 오히려 규제 강화 움직임이 가속화되는 이중 구조가 나타나고 있다. 연방과 주 사이의 이러한 긴장은 2025년 미국 AI 거버넌스의 가장 주목할 만한 특징으로, 미국 내 AI 규제의 향방을 가늠하는 핵심 변수로 부상하고 있다.

나. 핵심 입법 및 정책 체계

1) 트럼프 행정부의 행정명령 체계

① 행정명령 제14179호(2025년 1월 23일)²⁴⁾: AI 리더십 장벽 제거

트럼프 대통령은 취임 3일 후인 2025년 1월 20일 바이든 행정부가 2023년 10월 30일 발표한 「안전하고 신뢰할 수 있는 AI 개발과 사용에 관한 행정명령(Executive Order 14110)」을 포함한 다수의 행정명령을 일괄 폐기하였다(White House, 2025). 이어 1월 23일 서명된 행정명령 제14179호 「미국 AI 리더십 장벽 제거(Removing Barriers to American Leadership in Artificial Intelligence)」는 “이념적 편향이나 조작된 사회적 의제로부터 자유로운” AI 시스템 개발을 강조하며 180일 내에 새로운 AI 액션플랜 개발을 지시하였다(Executive Order 14179). 이 명령은 바이든 행정부의 AI 안전 중심 정책을 완전히 폐기하고 혁신 우선주의로 전환하는 신호탄이 되었다.

② AI 액션플랜 발표(2025년 7월 23일): “AI 경쟁에서 승리하기”

2025년 7월 23일 백악관은 “Winning the AI Race: America’s AI Action Plan”을 발표하였다(The White House, 2025). 10,000건 이상의 공개 의견수

24) Executive Office of the President. (2025). Executive Order 14179: Removing Barriers to American Leadership in Artificial Intelligence. January 23, 2025. 90 Fed. Reg. 8741.

럼 절차를 거쳐 백악관 과학기술정책실(OSTP)이 주도하여 개발한 이 계획은 90개 이상의 연방 정책 조치를 다음 세 가지 핵심 축으로 제시하였다.

첫 번째 축인 혁신 가속화는 규제 완화, 오픈소스 AI 모델 장려, AI 도입 촉진, 차세대 제조업 지원, AI 기반 과학 투자 등 민간 주도 혁신이 번영할 수 있는 조건 조성에 중점을 둔다. 두 번째 축인 미국 AI 인프라 구축은 데이터센터·반도체 제조시설을 위한 간소화된 허가 절차, 전력망 개발, 미국 반도체 제조업 복원, 군사 및 정보기관용 고보안 데이터센터 구축에 초점을 맞춘다. 세 번째 축인 국제 AI 외교 및 안보 리더십은 미국 AI 시스템·컴퓨팅 하드웨어·표준의 전 세계 채택 추진, 동맹국·파트너국에 대한 미국 AI 수출, 국제 거버넌스 기구에서의 중국 영향력 견제, AI 컴퓨팅 수출 통제 강화를 내용으로 한다(The White House, 2025).

같은 날 트럼프 대통령은 3개의 관련 행정명령을 추가로 서명하였다. 행정명령 제14317호는 상무부가 90일 내에 American AI Exports Program을 설립하여 미국의 풀스택 AI 기술 수출 패키지를 지원하도록 지시하였다. 행정명령 제14318호는 데이터센터 건설을 위한 연방 허가 절차 간소화와 환경 규제 완화를 지시하였다. 행정명령 제14319호는 연방 정부가 “이념적 의제를 위해 진실성과 정확성을 희생하는” AI 모델을 조달하는 것을 방지하고, 다양성·형평성·포용성(DEI) 원칙 대신 “진실성”과 “중립성” 기준을 충족하는 모델만 연방 조달에 활용하도록 지시하였다(Executive Orders 14317-14319, 2025)²⁵⁾.

25) Executive Office of the President. (2025). Executive Order 14317: Promoting the Export of the American AI Technology Stack. July 23, 2025.

Executive Office of the President. (2025). Executive Order 14318: Accelerating Federal Permitting of Data Center Infrastructure. July 23, 2025.

Executive Office of the President. (2025). Executive Order 14319: Preventing Woke AI in the Federal Government. July 23, 2025.

③ Genesis Mission 행정명령(2025년 11월 24일)²⁶⁾

2025년 11월 24일 트럼프 대통령은 미국의 AI 연구·개발 및 과학적 응용을 강화하는 새로운 연방 이니셔티브 “Genesis Mission”을 출범시키는 행정명령에 서명하였다. 이 명령은 컴퓨팅 자원 확대, 방대한 연방 데이터셋 접근 확대, 과학 분야에서의 AI 실용화를 목표로 하며, 에너지부 장관에게 연구자들에게 컴퓨팅 파워와 데이터셋을 제공하는 “American Science and Security Platform”의 구축을 지시하였다(NBC News, 2025. 11. 24). 이는 AI 연구를 맨해튼 프로젝트에 비견하는 대규모 연방 연구 투자 이니셔티브로 평가된다.

④ 행정명령 제14365호(2025년 12월 11일)²⁷⁾: 국가 AI 정책 프레임워크 확보

2025년 12월 11일 서명된 행정명령 제14365호 「인공지능에 관한 국가 정책 프레임워크 확보(Ensuring a National Policy Framework for Artificial Intelligence)」는 “최소 부담(minimally burdensome)” 국가 AI 정책 프레임워크를 통해 미국의 글로벌 AI 패권을 유지·강화하겠다는 목표를 제시하며 주정부 AI 규제를 견제하는 내용을 담고 있다. 핵심 내용은 다음과 같다. 법무장관에게 30일 내에 연방 정책과 상충하는 주 AI 법률에 이의를 제기할 AI 소송 태스크포스(AI Litigation Task Force) 설치를 지시하였다. 상무부 장관에게는 주 AI 법률의 평가 보고서 제출을 지시하여 “과도한 주 규제”로 판단되는 법률을 식별하도록 하였으며, 이 과정에서 콜로라도 AI Act를 과도한 규제의 예로 명시적으로 언급하였다. 또한 연방 BEAD 프로그램 자금 지원을 연방 AI 정책 준수 여부에 연동하는 재정 조건화(funding conditionality) 메커니즘을 도입하였다. FTC 위원장에게 FTC Act의 AI 모델 적용에 관한 정책 성명 발표를 지시하고, 미국 특별 AI·암호화폐 자문관 및 과학기술보좌관에게 충돌하는 주 AI 법률을 선점하는 통

26) Executive Office of the President. (2025). Executive Order 14363: Launching the Genesis Mission. November 24, 2025.

27) Executive Office of the President. (2025). Executive Order 14365: Ensuring a National Policy Framework for Artificial Intelligence. December 11, 2025. 90 Fed. Reg. 58499.

일 연방 AI 정책 프레임워크에 관한 입법 권고안 마련을 지시하였다(White & Case, 2025). 다만 이 행정명령 자체는 의회 입법이 아니므로 주 AI 법률을 직접 선점할 법적 효력은 없으며, 향후 법원에서의 연방주의 관련 헌법 논쟁이 예상된다(O'Melveny, 2025).

2) 포괄적 연방 AI 법률의 부재

2025년 12월 기준으로 미국에는 포괄적 연방 AI 법률이 없다. 연방 차원의 AI 규제는 주로 저작권법, 반독점법, 제조물책임법, 차별금지법, 소비자보호법 등 기존 법률의 적용과 FTC, EEOC, CFPB, FDA 등 규제기관들의 법 해석·집행을 통한 사실상의 연성법 방식으로 이루어지고 있다. Stanford HAI에 따르면 연방 AI 관련 규제는 2023년 25개에서 2024년 59개로 두 배 이상 증가하였으며, 규제 제정 기관도 21개에서 42개로 늘어났다(Maslej et al., 2025.4). 트럼프 행정부가 규제 완화를 기치로 내걸고 있음에도 개별 기관 차원의 AI 관련 규제 활동은 오히려 증가하고 있다는 점은 미국 AI 거버넌스의 분산적 성격을 잘 보여준다.

다. 혁신 촉진과 리스크 관리의 균형 방식

미국의 접근법은 규제 완화를 통한 민간 주도 혁신 극대화과 기업 자율 규제를 두 축으로 삼고 있다. 연방 차원에서는 NIST의 AI 위험 관리 체계(AI RMF)를 통한 자발적 표준 설정이 주된 도구로 활용되고 있다. 그러나 트럼프 행정부는 NIST AI RMF에서 다양성·형평성·포용성(DEI) 관련 내용을 제거하도록 지시하여(Brookings, 2025), 기존의 책임 있는 AI 개발 원칙에서 일정 부분 후퇴하는 모습을 보이고 있다. 한편, 2025년 5월 트럼프 대통령이 서명한 TAKE IT DOWN Act는 비동의 친밀한 이미지(NCII)와 딥페이크 성적 콘텐츠의 온라인 유포를 연방 범죄로 규정하는 것으로 2025년 현재 연방 차원에서 제정된 유일한 AI 관련 입법이다(The White House, 2025).

AI 안전, 공정성, 편향 등의 문제에 대한 연방 차원의 체계적 접근은 현저히 약화되었다. 트럼프 행정부는 AI 편향 문제를 규제 대상이 아닌 이념적 문제로 규정하며, 알고리즘 차별 방지를 위한 주정부 규제 조치를 “이념적 편향의 강요”로 간

주하고 있다(Brookings, 2025). 이는 AI 시스템에 내재된 데이터 편향 문제를 방지할 수 있다는 우려를 낳고 있으며, 연방 차원의 AI 안전 거버넌스가 실질적으로 공백 상태에 놓일 수 있다는 비판도 제기되고 있다.

라. 거버넌스 체계와 집행 역량

1) 주정부 규제 강화의 가속화

연방 차원의 규제 완화와 대조적으로 주정부 차원에서는 AI 규제 강화가 급속히 진행되고 있다. 2025년 한 해에만 전국 50개 주에서 1,208개의 AI 관련 법안이 발의되었으며, 38개 주에서 100개 이상의 AI 관련 법률이 채택되었다(NCSL, 2025).

주요 입법 현황을 살펴보면, 캘리포니아는 2025년 9월 29일 뉴섬 주지사가 서명한 「프론티어 AI 투명성법(SB 53)」을 통해 연간 총수입 5억 달러 이상의 대규모 프론티어 모델 개발자에게 안전·보안 프로토콜 수립과 공개를 의무화하였다(California Governor's Office, 2025). 이는 2024년 거부된 SB 1047의 축소 버전으로, 대규모 AI 모델에 대한 투명성 요구사항을 강화한 것이다. 텍사스에서는 2025년 6월 22일 Abbott 주지사가 서명한 「텍사스 책임 AI 거버넌스법(TRAIGA, HB 149)」이 2026년 1월 1일 발효되었다(Governor Abbott, 2025). 이 법률은 자해·폭력 조장, 헌법적 권리 침해, 보호 집단에 대한 불법 차별, 딥페이크 아동 성착취물 생성 등을 목적으로 한 AI 시스템의 개발·배포를 금지하며, 텍사스 법무장관에게 집행 권한을 부여하고 위반 시 건당 최대 20만 달러의 민사 제재를 부과한다. 콜로라도의 경우 2024년 제정된 「콜로라도 AI Act(SB 24-205)」가 미국 최초의 포괄적 주 AI 규제법으로, 고위험 AI 시스템의 알고리즘 차별 방지를 위한 개발자·배포자 의무를 부과하고 있다. 당초 2026년 2월 1일 시행 예정이었으나 2025년 8월 28일 시행일이 2026년 6월 30일로 연장되었으며(SB 25B-004, 2025)²⁸⁾, 트럼프 행정부의 행정명령 제14365호에서 과도한 규제의 예로 명시적으로 언급되기도 하였다(The White House, 2025).

28) State of Colorado. (2025). SB 25B-004: Colorado AI Act — Extension of Implementation Date. August 28, 2025.

2) 연방-주 규제 충돌과 법적 불확실성

연방-주 간 규제 충돌은 2025년 하반기 들어 더욱 표면화되었다. 하원에서 발의된 주정부 AI 규제 10년 모라토리엄 조항이 “One Big Beautiful Bill Act”에 포함될 것으로 예상되었으나 상원에서 초당적으로 거부되어 최종 법안에서 제외되었다(Goodwin Law, 2025). 연방 입법을 통한 주 규제 선점이 무산되자 트럼프 행정부는 행정명령 제14365호를 통해 같은 목표를 추구하는 방향으로 전략을 전환하였다. 그러나 행정명령 자체는 의회 입법이 아니므로 주 AI 법률을 직접 선점할 법적 효력이 없어 향후 법원에서의 헌법적 도전이 예상되는 가운데 법적 불확실성이 지속되고 있다. 2025년 12월 기준으로 상무부의 주 AI 법률 평가 보고서 제출(2026년 3월 11일 기한), 법무장관의 AI 소송 태스크포스 구성, FTC의 AI 정책 성명 발표 등 후속 조치들이 진행 중으로, 이 결과에 따라 미국 내 AI 규제 지형이 상당 부분 재편될 것으로 전망된다.

마. 국제적 영향력 확산 전략

트럼프 행정부의 국제 AI 전략은 “미국 AI 기술 스택(American AI Technology Stack)”의 글로벌 확산을 통한 패권 강화에 방점을 두고 있다. 행정명령 제14317호를 통해 미국산 AI 최적화 하드웨어, 데이터센터 저장소, 클라우드 서비스, 네트워킹을 포함한 풀스택 AI 기술 패키지의 수출을 전략적으로 지원하는 한편, AI 컴퓨팅 수출 통제를 통해 중국의 첨단 AI 기술 접근을 차단하는 투트랙 전략을 취하고 있다(The White House, 2025). 바이든 행정부가 국제 협력과 공통 AI 안전 표준 수립을 강조했던 것과 달리, 트럼프 행정부는 EU AI Act나 국제 AI 안전 네트워크와의 다자적 협력보다 미국 AI 기술의 글로벌 채택을 앞세우는 일방적 접근을 선택하고 있다. 이는 AI 거버넌스를 공동의 규범 형성 문제가 아닌 기술 패권 경쟁의 연장선으로 바라보는 트럼프 행정부의 시각을 잘 보여준다.

바. 평가 및 시사점

미국 모델의 강점은 다음과 같이 네 가지로 정리된다. 첫째, 규제 완화를 통해 민간 주도 혁신이 최대한 빠르게 전개될 수 있는 환경을 조성하고 있다. 둘째, 데

이터센터, 반도체, 컴퓨팅 파워 등 AI 인프라에 대한 대규모 연방 투자를 통해 기술 경쟁력의 물적 기반을 강화하고 있다. 셋째, 시장 경쟁을 통해 AI 기술 생태계가 자연스럽게 발전하도록 유도하는 방식은 혁신의 방향을 정부가 미리 규정하지 않는다는 점에서 유연하다. 넷째, AI에 관한 포괄적 규제 법제 없이도 기존 법률 체계를 AI에 적용하는 방식은 새로운 입법 없이 빠르게 규제 공백을 메울 수 있다는 실용적 장점이 있다.

물론 한계도 적지 않다. 포괄적 연방 AI 법률이 없어 규제 공백이 지속되고 있으며, 연방-주 규제의 분화로 인해 기업들이 부담해야 할 준수 비용이 복잡해지고 있다. AI 편향, 안전, 공정성 문제에 대한 체계적 접근이 약화된 가운데, “이념 중립성” 요건이라는 새로운 규제 부담이 추가되었다는 점도 아이러니이다. 국제 AI 안전 협력의 약화는 글로벌 거버넌스 공백을 확대할 우려가 있으며, 행정명령 제 14365호로 촉발된 연방-주 법적 분쟁이 장기화될 경우 미국 AI 기업들이 오히려 규제 불확실성에 더 크게 노출될 수 있다.

한국에 대한 시사점은 네 가지이다. 첫째, 민간 혁신 촉진을 위한 규제 완화와 전략적 인프라 투자를 병행하는 방식은 AI 기본법 시행과 연계된 산업 육성 정책을 설계하는 데 구체적인 참고가 된다. 둘째, 포괄적 AI 전용 법률 없이 기존 법률 체계를 활용하는 미국의 경험은 한국이 AI 기본법 제정 이후에도 「개인정보 보호법」, 「정보통신망법」 등 기존 법률과의 정합성을 지속적으로 점검해야 함을 상기시킨다. 셋째, 연방-주 규제 충돌에서 드러나는 규제 일관성 문제는 한국이 중앙-지방 차원의 AI 규제 거버넌스를 설계할 때 미리 조정 메커니즘을 갖추어야 한다는 것을 시사한다. 넷째, AI 편향 문제를 이념적 관점에서 접근하는 트럼프 행정부의 방식은 AI 공정성·비차별 원칙이 단순한 규범 선언이 아니라 한국 AI 기본법의 핵심적 실질 가치임을 역설적으로 확인시켜 준다.

제 2 절 중 국

1. AI 산업생태계 구조 및 주요 플레이어

중국 AI 생태계는 AI 기반 빅테크(알리바바, 바이두, 텐센트, 화웨이 등) 기업들의 자체 생태계를 주축으로 각 기업 간 우위를 바탕으로 한 협력을 통해 자국 생태계를 구축하려는 공통의 노력이 동시에 진행되며 빠른 속도로 발전하고 있다. 특히, 딥시크 R1 엔진의 성공에 힘입어 거대 모델과 경량화 모델 혁신을 동시에 추진하며, Gitee 등 중국 내 오픈소스 생태계의 활성화를 추진하는 한편, 글로벌 오픈소스 생태계에도 적극 참여하여 중국산 모델의 글로벌 확산을 촉진하려는 경향도 특징적으로 나타난다.

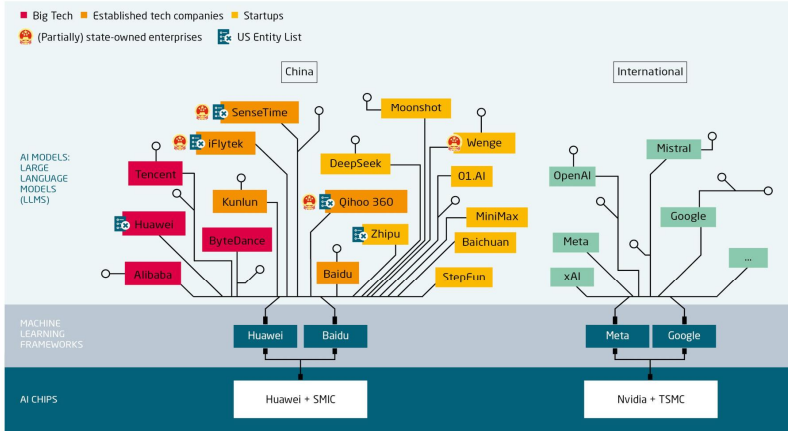
AI 기술 스택별로 보면 반도체 칩은 화웨이, 프레임워크는 화웨이와 바이두, 모델은 알리바바 등의 빅테크와 스타트업들이 혼재되어 생태계를 견인하는 역할을 하고 있다. 반도체 칩의 경우, 최첨단 GPU 및 관련 제조 장비에 대한 미국 및 동맹국의 엄격한 수출 통제로 인해 중국이 자체 개발에 주력하는 분야이다. 자국 반도체 역량에 대한 막대한 국가의 투자가 진행되고 있으며, 이 계층은 화웨이와 중국의 반도체 파운드리인 SMIC가 주축을 담당한다.

머신러닝 프레임워크는 전 세계적으로 대부분 오픈소스에 의존하며, 이 분야에서 중국의 전략은 국내 대안을 육성하는 동시에 글로벌 오픈소스 커뮤니티에 적극적으로 참여하여 빠른 기술개발과 동시에 중국산 소스의 글로벌 확산을 추진함으로써 글로벌 리더십을 확보하는 것이다.

LLM과 다양한 애플리케이션은 자국 소프트웨어 생태계와 글로벌 오픈소스 생태계 발전에 힘입어 중국의 주요 빅테크와 스타트업들이 빠른 속도의 성장세를 보이고 있다. 바이두, 알리바바, 텐센트 등 기존 빅테크들은 자사 전략 뿐 아니라, 중국 내 생태계에서의 빠른 산업 응용을 위한 노력을 전개하며 풀스택의 AI 역량을 완성해 나가는 중이다.

[그림 3-3] AI 스택 주요 플레이어

The AI Stack: Few players at the bottom, many on top



Source: MERICS

자료: MERICS(2025. 7. 22.)

이 과정에서 중국 정부의 지원은 차등적으로 이루어지는데 자본 집약적인 반도체 부문에는 가장 중요한 국가 지원이 제공되는 반면 소프트웨어 프레임워크의 국산화는 주로 주요 국내 기술 기업에 위임된다., AI 모델 및 애플리케이션을 포함한 상위 계층은 GPU, 클라우드, 데이터 활용 등의 지원 환경을 보장받는다(MERICs, 2025. 7. 22).

상기한 중국 정부의 정책 방향에 따라 알리바바, 텐센트, 바이두, 화웨이 등 중국 빅테크들은 인프라부터 애플리케이션, 산업 영역을 아우르는 풀스택 생태계를 구축하기 위한 노력을 전개하고 있다. 고성능 AI 칩과 클라우드 플랫폼은 막대한 연산 자원을 바탕으로 거대 모델의 훈련·배포를 지원하며, 바이두의 패들패들(PaddlePaddle), 화웨이의 마인드스포어(MindSpore) 등 AI 소프트웨어 프레임워크를 통해 개발자 생태계를 조성한다. 나아가 각 기업은 초거대 언어모델(LLM)을 챗봇, 검색, 음성비서, 자율주행 등 자사 서비스 전반에 통합해 나가며, AI 플랫폼을 개방하거나 오픈소스화하여 외부 개발자와 협력 생태계를 구축하는 전략을 공통적으로 가지고 있다.

딥시크, 문샷, 즈푸 AI 등의 스타트업은 혁신적인 모델 아키텍처와 오픈소스 전략으로 승부하며, 대기업들과 전략적 제휴 또는 투자를 통해 성장하면서 생태계의 저변 확산과 혁신 촉진의 주축으로 기능한다. 주요 플레이어별 AI 생태계를 살펴 보면 다음과 같다.

1) 화웨이²⁹⁾

• AI 인프라 구축

화웨이는 미국에 대항하여 중국의 칩 생태계를 주도하는 한편 오픈소스 딥러닝 프레임워크인 MindSpore, 자체 개발한 산업특화 LLM인 판구(Pangu)를 출시하며 사실상 AI 폴스택 기업으로서의 성장을 도모한다. 화웨이가 가장 중추적인 역할을 담당하는 AI 칩 계층에서는 Ascend 시리즈, 특히 올해 출시한 Ascend 910C AI 칩이 엔비디아의 H100 GPU에 필적하도록 설계되었다.

또한, AI 컴퓨팅 파워 수요가 기하급수적으로 증가함에 따라, 시스템 수준 혁신을 특징으로 하는 AI 네이티브 클라우드 인프라인 클라우드 매트릭스(CloudMatrix)를 출시하여 안정적이고 신뢰할 수 있는 AI 컴퓨팅 파워를 제공하는 등 자국의 지정학적 환경에 적합한 AI 인프라 제공을 위한 노력을 지속해 나가고 있다.

• LLM 모델과 산업적용

2024년 출시된 화웨이의 LLM인 판구 모델(Pangu Models) 5.0은 화웨이 클라우드의 핵심 AI 솔루션으로 다양한 산업 시나리오에서 고객의 문제를 해결하고 운영 효율성을 높이는 데에 중점을 둔다. 이 모델은 10억에서 1조 이상의 파라미터를 가진 다양한 크기로 제공되며, 텍스트, 이미지, 비디오, 레이어, 적외선, 원격 감지 등 멀티 모달 기반의 물리적 AI로 발전하고 있으며 2024년 한 해 동안 정부 서비스, 금융, 제조, 제약 R&D, 석탄 채굴, 철강 제조, 철도, 자율주행, 산업 및 건축 설계, 기상학 등 30개 이상의 산업에서 400개 이상의 시나리오에 적용되었다.

29) Hwawei investment&Holding Co.,Ltd 2024 Annual Report에서 발췌, 정리

화웨이는 2022-2024년 3년간 연간 매출의 20% 이상을 R&D에 투자하고 있으며, 2024년 총 R&D 지출은 179.7억 위안으로 총 매출의 20.8%를 차지한다. 화웨이의 2024년 연차보고서에 따르면, 화웨이는 AI 방면에서 AI, 기초 기술, 지능형 자동차 등 미래 지향적인 기초 연구 및 혁신, 하드웨어, 소프트웨어, 칩, 장치, 네트워크 및 클라우드 전반의 전략적 시너지를 강화하고 AI 인프라 강화, AI 도메인 혁신 및 산업 지능화 가속화, LLM 효율화 등에 R&D 투자를 지속적으로 확대해 나갈 예정이다.

〈표 3-5〉 화웨이의 AI 생태계 구조

생태계 계층	추진현황
산업 응용	<ul style="list-style-type: none"> • 산업특화 솔루션: HarmonyOS 등 • 스마트 디바이스: 통신, 제조, 헬스케어, 자율주행 등
AI 모델 및 플랫폼	<ul style="list-style-type: none"> • 판구大모델, MindSpore
클라우드·네트워크	<ul style="list-style-type: none"> • Huawei Cloud, 5.5G
AI 칩셋	<ul style="list-style-type: none"> • Ascend, Kirin

2) 알리바바

중국 최대 전자상거래 기업인 알리바바는 AI 중심 기업으로의 전환을 선언하고 막대한 투자를 감행 중이다. 2023년 말 알리바바의 회장인 마윈은 향후 3년간 3,800억 위안(약 76조 원)을 AI 및 클라우드 인프라에 투자하기로 했고, 2025년에는 AI 칩 구매 예산만 390억 위안으로 전년 대비 50% 이상 늘린 바 있다 (Alibaba, 2024).

알리바바의 AI 전환은 미국의 반도체 수출 규제라는 외부적 압박에 대응하면서도 동시에 AI 시대의 선제적 기회를 포착하려는 전략적 판단에 기반한다. 2025년 AI 칩 구매 예산만 390억 위안으로 전년 대비 50% 이상 증액한 것은 첨단 GPU 수급이 어려워진 상황에서도 AI 인프라 확충에 대한 강한 의지를 보여준다.

이러한 한계를 극복하기 위해 알리바바는 다각적 접근을 취하고 있다. 자회사

핀투오거 반도체(Pingtouge)를 통해 AI 추론용 한광 800 칩과 서버용 Yitian 710 CPU 등을 자체 개발하는 한편, 엔비디아 H800 GPU 등을 대량 확보하여 단기적 수요에 대응하고 있다. 또한 Moonshot AI, MiniMax, Zhipu AI 등 유망 스타트업에 직접 투자하여 최신 AI 기술을 흡수하는 전략을 병행하고 있다.

• 클라우드 인프라의 전략적 우위

알리바바 클라우드는 자사의 거대한 전자상거래 트래픽을 처리하며 성장한 중국 최대 클라우드 서비스로, 현재 중국 IaaS/PaaS 시장에서 약 25%의 점유율을 차지하며 1위를 유지하고 있다. 이 클라우드 인프라 위에 구축된 “Feitian”(飞天) AI 플랫폼은 초거대 모델 개발의 핵심 토대가 되고 있다. 알리바바는 AI 제품이 올해 클라우드 부문 성장의 절반 이상을 견인할 것으로 예상한다고 발표했으며 이는 AI가 단순한 기술적 시도를 넘어 실질적인 수익 창출의 핵심 동력이 되고 있음을 보여준다.

• Qwen 모델군: 오픈소스 전략

알리바바의 대표적인 대규모 언어 모델인 Qwen(통의천문, Tongyi Qianwen)은 2023년 4월 베타 버전으로 처음 출시되어 현재까지 지속적인 발전을 이루고 있다. 2024년 7월 일부 벤치마크에서 최고의 중국어 모델로 선정되었고, 엔트로픽과 오픈AI의 최고 모델에 이어 전 세계적으로 3위를 차지했다.

2024년 9월 알리바바는 새로운 Qwen 2.5 제품군에서 100개 이상의 오픈소스 모델을 출시했으며, 이 중 Qwen 2.5 72B 모델은 메타의 훨씬 큰 규모의 Llama 모델을 능가하는 성능을 보였다. 2025년 2월 기준 상위 10개 오픈소스 LLM 중 최소 7개가 알리바바 클라우드의 Qwen2.5-72b 시리즈를 기반으로 하고 있다는 사실은 Qwen 생태계의 광범위한 영향력을 보여준다. 최신 버전인 Qwen 3는 하이브리드 사고 모드(“Thinking”과 “Non-Thinking”)를 채택하여 추론 성능, 속도, 비용을 유연하게 제어할 수 있도록 했다. 추론 벤치마크에서 AIME25에서 92.3점, 코딩 분야의 LiveCodeBench v6에서 74.1점으로 오픈소스 기록을 세웠다.

이 외에도 알리바바의 전략은 인프라(알리 클라우드)와 접근성 높은 고성능 모

텔(Qwen 오픈소스)을 동시에 제공함으로써 개발자와 기업을 자사 생태계로 유인하는 선순환 효과를 창출하고 있다.

- 전사적 AI 통합과 새로운 서비스 개발

알리바바는 AI를 전자상거래, 금융, 클라우드 등 그룹 전 계열사에 이식하고 있다. 쇼핑앱 타오바오에서는 상품 설명 자동 생성과 추천 고도화, 물류 최적화에 AI를 적용하고 있으며, AI 비서가 탑재된 웹브라우저 Quark, 회의록 요약 도구 Tongyi Tingwu, 기업용 코파일럿 Tongyi Qianwen 지능 비서 등 새로운 AI 기반 서비스를 출시했다.

특히 주목할 만한 것은 도시 뇌(City Brain) 프로젝트를 통한 스마트시티 분야로의 확장이다. 교통 흐름 최적화 등 도시 인프라 관리에 AI 솔루션을 제공함으로써 공공 부문으로의 사업 영역 확장을 도모하고 있다.

- 칩-클라우드-모델-서비스-산업 수직 통합 생태계

알리바바의 AI 생태계는 반도체 칩부터 클라우드 인프라, AI 모델, 서비스, 산업 적용에 이르는 전 단계를 아우르는 수직 통합 구조를 특징으로 한다. 여기에 오픈소스 전략과 파트너십을 결합하여 개발자 커뮤니티를 적극적으로 끌어들이는 것이 핵심이다.

알리바바의 AI 전략은 독자적 기술 개발과 클라우드 플랫폼을 통한 외부 기업 대상 AI 서비스 제공을 포함하며 헬스케어, 제조업, 금융 서비스 등 산업별 특화 AI 솔루션을 출시하는 것이다. 알리바바의 자회사인 클라우드 인텔리전스 그룹(Cloud Intelligence Group)은 AI 혁신의 이점을 기업 고객이 활용할 수 있도록 클라우드 상에서 AI 솔루션을 제공하기 시작하였고, 고객들이 자신들의 비즈니스 시나리오에 맞는 맞춤형 LLM을 개발할 수 있도록 지원하는 한편, 이를 알리바바 클라우드의 매출 상승으로 연계될 수 있는 구조를 형성하고 있다.

이 외에도, 알리바바는 2024년 중국의 유명 AI 스타트업인 문샷AI(Moonshot AI)에 약 8억 달러를 투자하여 약 36%의 지분을 확보하며 생태계 저변 확장에 나섰다. 이 중 3억 달러 이상은 알리바바의 클라우드 컴퓨팅 인프라를 사용할 수

있는 크레딧이다. 이 방식은 미국 내 OpenAI-Microsoft 관계와 유사한 모델이며, 단순 지분 참여를 넘어 운영 인프라 공급자 역할까지를 포괄하며 AI 생태계 전방위 영향력 확보가 가능한 기반으로 작용한다(TECHINASIA, 2025. 7. 12).

즉, 미국의 첨단 칩 수출 제한으로 자원이 부족한 중국 상황에서 스타트업의 AI 모델 훈련에 필요한 컴퓨팅 자원을 제공함으로써 알리바바가 클라우드 기반 AI 스타트업의 파워하우스로서의 입지를 다지고 있음을 의미한다(Alibaba Cloud, 2025. 4. 8).

〈표 3-6〉 알리바바의 AI 생태계 구조

생태계 계층	추진현황
산업 응용	<ul style="list-style-type: none"> • 전자상거래: 타오바오, 티몰 • 물류: Cainiao • 디지털 미디어: Youku 등 • 스마트 시티: City Brain
AI 모델 및 플랫폼	<ul style="list-style-type: none"> • Qwen, Feitian
클라우드	<ul style="list-style-type: none"> • Alibaba Cloud
AI 칩셋	<ul style="list-style-type: none"> • 한광800, Yitian710

3) 바이두

중국 최대의 검색 포털인 바이두(Baidu)는 일찍부터 AI 연구에 매진하여 자체 딥러닝 프레임워크인 패들패들(PaddlePaddle)과 AI 칩 설계 등 풀스택 역량을 갖춘 기업이다. 바이두의 쿤룬(Kunlun) AI 프로세서는 2018년 첫 버전이 공개된 이래 2세대(2021년), 3세대(P800, 2024년)로 거듭 발전하며 중국 내 AI 칩 개발 노선 강화에 기여하고 있다. 바이두의 AI 칩은 검색제공을 통해 쌓아올린 음성, 자연어 처리, 이미지 및 기타 AI 기술에 최적화된 딥러닝 프레임워크를 제공한다(Baidu, 2024).

2025년 4월 바이두는 3세대 P800 칩 3만 개로 구성된 AI 훈련 클러스터를 가동시켰다고 밝혔는데 이 클러스터는 수백억~수천억 매개변수 규모의 딥시크급 모

텔도 훈련가능한 성능을 갖춘 것으로 알려졌다(Reuters, 2025.4.15).

AI 모델 측면에서 바이두 경쟁력의 핵심은 대규모 개발자 커뮤니티를 보유한 패들패들(PaddlePaddle) 프레임워크와 ERNIE(문심, Wenxin) 모델 시리즈이다. 2016년 출시된 패들패들은 477만 명 이상의 개발자와 18만 개 기업이 활용하는 글로벌 플랫폼으로 성장하였고 의미 분석, 이미지 분류, 객체 탐지, 광학 문자 인식, 음성 합성 등 다양한 AI 기능을 지원한다.³⁰⁾ 또한, 중국에서 독자 개발된 최초의 오픈소스 산업급 딥러닝 플랫폼으로, 핵심 프레임워크, 산업급 모델 라이브러리, 개발 키트, 도구 구성 요소, 학습 및 훈련 커뮤니티를 포함한 종합적인 생태계를 제공한다(Baidu Developer Community, 2024).

2023년 최초 공개된 바이두의 AI 모델 어니봇(ERNIE Bot, 文心一言)은 챗GPT에 대응하는 바이두의 대화형 LLM으로, 2023년 10월 ERNIE 4.0 발표를 시작으로 지속적인 모델 업그레이드가 진행되었다. 바이두는 2025년 4월 ERNIE 4.5 Turbo와 고도화된 추론 모델 ERNIE X1을 선보였다. 이 모델들은 코딩과 논리 추론 분야에서 GPT-4.5급 성능을 달성했다고 평가된다.

2025년 6월 바이두는 ERNIE 4.5 멀티모달 모델의 오픈소스 공개를 통해 AI 개발 경제성을 재정의하고 PaddlePaddle을 전 세계 개발자들의 기초 프레임워크로 확산시키려는 전략적 전환을 추진했다. 또한, 바이두는 기존에 쌓아올린 검색 역량과 자사의 경쟁력이 집결된 자율주행이라는 양대 분야에서 강점을 지니고 있으며, 패들패들과 어니봇, 바이두의 산업 애플리케이션은 모두 바이두의 AI 클라우드 인프라를 통해 제공된다.

30) https://www.paddlepaddle.org.cn/documentation/docs/en/guides/index_en.html

〈표 3-7〉 바이두의 AI 생태계 구조

생태계 계층	추진현황
산업 응용	<ul style="list-style-type: none"> • 검색광고 • 자율주행(Apollo) • 스마트 디바이스
AI 모델 및 플랫폼	<ul style="list-style-type: none"> • Ernie Bot, PaddlePaddle
클라우드	<ul style="list-style-type: none"> • Baidu AI Cloud
AI 칩셋	<ul style="list-style-type: none"> • Kunlun

위에서 본 것처럼, 중국의 빅테크들은 모두 AI 기업으로서의 정체성을 쌓아나가며 AI 풀스택 기업으로 입지를 강화해 나가고 있다. 각자가 전통적으로 강점을 가지고 있던 분야를 바탕으로 AI 모델과 솔루션으로의 응용을 확대하고, 이를 자사 클라우드를 통해 제공하고 자체 칩 개발에 주력하며 중국 AI 생태계의 자국화와 그 과정에서의 풀스택 경쟁을 촉진한다.

이 외에도, AI 모델 및 애플리케이션 단에서는 스타트업 중심의 LLM 개발이 활발하며, 특히 DeepSeek-R1 모델이 비용 효율성과 강력한 추론 기능으로 주목을 받은 바 있으며, 문샷 AI, 즈푸 AI, 그리고 오랫동안 중국에서 음성AI 분야의 강자였던 아이플라이테크 등의 중소 스타트업이 중국의 AI 생태계에 각자의 분야에서 혁신 창출자로 지위를 다지는 중이다.

〈표 3-8〉 중국의 주요 AI 기업 특성

기업	AI 칩(반도체)	클라우드 인프라	AI 엔진 (모델/플랫폼)	주요 애플리케이션	산업 적용 분야
화웨이	자체 AI칩 Ascend 910/ 310(데이터센터 가속기)기린 (Kirin) 모바일 AP(NPU 탑재)	화웨이 클라우드 - 中 4위권 IaaS (정부/통신 특화) Atlas AI 컴퓨팅 플랫폼(Ascend 기반)	Pangu(盘古, 판 구)대모델 - NLP /CV/멀티 모달 시리즈(最 大 1000억+ 파라미터) MindSporeAI 프레임워크 (오픈소스)	판구 약물발견플 랫폼(신약개발) 스마트폰 Celia 음성비서 (HarmonyOS) 영상 감시 AI(안 면인식 등 솔루션)	통신망 최적화 제조·농업(맞춤 형 AI모델) 공공 안전/정부 (감시, 교통) 모바일 디바이 스(UX 향상)
알리바바	자체 AI칩 한광 800(추론) · Yitian 710 (서버용 SoC) ※ 반도체 자회 사 핑투오거 (T-Head)	알리바바 클라우 드(Aliyun)- 中 1위 IaaS/ PaaS Feitian AI 플랫폼 (모델 훈련 지원)	Tongyi Qianwen(通義 千問)- 오픈소스 LLM 'Qwen' 시리즈g 모델스코프 (ModelScope) 등 오픈소스 모델 허브	퀵(Qwen) 챗봇 (통역, 문답 등) Quark 브라우저 (AI 비서 탑재) 전자상거래 AI 추천·검색	전자상거래 (리테일) 클라우드 서비스 (기업 DX) 스마트시티 (도시 뇌) 금융(엔트그룹 핀테크)
바이두	자체 AI칩 쿤룬 (Kunlun) 시리즈 - 3세대 P800까지 개발 ※ 3만개 칩 클러스터로 LLM 훈련	바이두 AI 클라 우드- 中 3위권 클라우드 (AI 최적화) 초대규모 완카 (万卡) 클러스터 구축	ERNIE(文心, 원신)시리즈 - Wenxin 4.5 등 LLM 파들패드 (Paddle Paddle) 딥러닝 프레임 워크(오픈소스)	원신일언 (ERNIE Bot)챗 봇 Apollo 자율주행 플랫폼(로보택시) 듀어OS (DuerOS)음성 비서 (스마트기기)	인터넷 검색·광고 자율주행차/ 모빌리티 클라우드 AI 솔루션(제조, 정부) 음성 AI(스마트 스피커 등)
텐센트	자체칩 없음 (GPU 의존) ※ 미리 엔비디아 H800/H100 대량 비축	텐센트 클라우드 (Tencent Cloud) - 中 2위권 IaaS 대규모 GPU 팜운영, AI 슈퍼컴 상용	혼원(混元, Hunyuan)- 자체 LLM (Hundreds of billions param) 엔젤(Angel) 등 머신러닝 프레임워크	위챗(WeChat) AI 조력 (예: 지능형 비서) 광고 생성 AI(이 미지/영상 생성) 게임 AI 생성 플 랫폼(혼원 Game)	소셜미디어/ 메신저 게임/엔터테인 먼트 온라인 광고(콘텐츠 생성) 클라우드/ 핀테크 (위챗페이 등)

기업	AI 칩(반도체)	클라우드 인프라	AI 엔진 (모델/플랫폼)	주요 애플리케이션	산업 적용 분야
딥시크 (Deep Seek)	자체칩 없음 - 엔비디아 H800 GPU로 대규모 모델 훈련 ※ 저사양 GPU로 고효율 알고리즘 구현	(별도 클라우드 없음, 타 클라우드 활용) ※ API 서비스로 LLM 제공	DeepSeek R1 - GPT-4급 성능 LLM(2025. 1 공개) DeepSeek V3 - 2024년 공개 LLM(오픈소스)	딥시크 챗봇 앱- iOS 미국 다운로드 1위(25. 1) LLM API 서비스- 초저가 과금 정책	(주로 B2B API 제공) 일반 소비자용 챗봇 앱 ※ 텐센트 등 타 기업 서비스에 모델 기술 제공
문샷 AI (Moon shot)	자체칩 없음 - 클라우드 상 GPU로 모델 훈련 ※ 혼합전문가 (MoE)기술로 GPU 비용 절감	(별도 클라우드 없음, 타 인프라 활용) ※ 모델 오픈소스로 배포, 협업 유도	Kimi 시리즈 Kimi-1(2024), Kimi-2(2025 공개) LLMm.dongascience.com ※ Kimi K2: 1조 매개변수 MoE 모델, 오픈소스	Kimi Chat- 멀티모달 챗봇(텍스트+코딩 능력) Kimi API 서비스- 클라우드 등 대비 저렴한 사용료	(주로 모델 제공) 코딩 보조 AI (SW개발률 통합) 창작물 생성(문장·이미지 등)
지푸 AI (Zhipu)	자체칩 없음 - 협력사 GPU/TPU 활용 ※ 정부·빅테크 투자로 인프라 확보	(클라우드 없음) ※ 산학 협력: 칭화대 연구진 주축 R&D	GLM 시리즈- ChatGLM 2/3 등 다국어 LLM(130억~1300억+파라미터) CogGLM등 지식 그래프 통합 모델 연구	ChatGLM챗봇 - 中 첫 양방향 영-중 챗봇(오픈소스 공개) 기업 맞춤형 지니 (ChatGPT류) 솔루션 제공	(B2B 솔루션 위주) 클라우드 상 LLM API AI 에이전트 플랫폼(실험 단계) ※ VC 펀드 조성으로 스타트업 투자
아이플라 이텍 (iFLYTEK)	자체칩 없음 - AI 가속 칩 설계 경험 제한 ※ 화웨이 등과 협력하여 HPC 인프라 구축	iFlytek Open Platform- 음성 AI API 클라우드(기존 서비스) Spark Desk 서비스 - LLM 엔진 제공	Spark 대모델 讯飞星火LLM v3.0+(중영 다국어 지원) 멀티모달 SparkDesk- 코드, 이미지 생성 등 강화	Spark Chat- 지식 Q&A 및 코딩 비서(챗 GPT 대항) AI 교육 비서- 학생 맞춤형 튜터 기기 음성 AI 서비스- 음성인식, 합성 (기존 강점)	교육(스마트 학습) 음성 서비스(통역기 등 대바이스) 사무 생산성(오피스 업무 비서) 헬스케어(의료 상담봇)

2025년 7월 26일 개최된 상하이 AI 컨퍼런스(WAIC)에서 화웨이 등 주요 기업들이 모델-칩 생태계 혁신 연합(模芯生态创新联盟, Model-Chip Ecosystem Innovation Alliance)을 출범하였고, 화웨이·Biren·StepFun 등 기업이 참여해 칩-모델-인프라 통합 표준 구축을 추진 중이다. 이 연합은 마지막 애플리케이션 단의 산업 적용과 활용을 염두에 두고, 미국의 반도체 칩 수출통제 등과 중국의 현재 기술 수준을 감안하여 가장 효율적으로 AI 생태계를 구축할 수 있도록 애플리케이션단과 인프라단의 기업들이 협력을 진행하기 위해 건립되었다. 이 연합은 중국의 유명 AI 기업 중 하나인 즈푸AI(智谱AI)의 모회사인 StepFun이 주도하여 모델-칩 매칭 플랫폼을 제공하며 반도체 칩 설계 기업인 Biren(璧仞科技), LLM 개발 기업인 즈푸AI, 효율형 LLM을 개발하는 딥시크, 이 외에 바이두, 알리바바, 텐센트 등이 참여한다.

〈표 3-9〉 중국 AI 생태계의 협력 구조

App Layer / Inference Engine / Public Use		
AI Chips	Frameworks	LLM Models
Huawei 등	MindSpore, Paddle 등	(Zhipu, DeepSeek 등)

기업 간 협력을 통해 고성능 칩 없이도 LLM을 위한 학습-운영이 가능하도록 칩 벤더와 모델 개발사가 직접 협력하여 추론 효율을 높이고, 칩-프레임워크-모델-애플리케이션까지의 풀스택 생태계를 구축하며, 자국산 칩 활성화를 촉진하여 자국 현실에 맞는 생태계를 구축하려는 노력은, 중국의 AI 발전 방향성을 가장 잘 드러내는 사례로 볼 수 있다.

결론적으로, 중국의 AI 정책은 중앙 정부의 강력한 의지, 막대한 자원 투입, 그리고 산업 전반에 걸친 전략적 통합을 특징으로 하는 포괄적이고 역동적인 접근 방식을 보여준다. 그 과정에서 중국은 AI 분야의 글로벌 리더로 거듭날 수 있는 자국의 AI 풀스택 생태계를 갖추어나가기 위해 빅테크 기반의 인프라와 프레임워

크를 중심으로 혁신적인 모델들이 탄생할 수 있는 환경을 조성하고 정부는 그에 적합한 지원을 해나가고 있다. 특히, 기업들이 미국의 수출통제에 대응하기 위해 중국 내 실정에 적합한 인프라와 모델 구축 및 효율화의 길을 모색하며 기존 미국 중심 글로벌 AI 생태계와는 또 다른 양상의 궤적이 형성되고 있어 AI 생태계 진영에서는 향후 이원화 경향이 강해질 가능성이 있다.

그러나, 글로벌 생태계의 주도권을 다지기 위해서는 중국 정부와 기업들이 자국 내에서 그리고 글로벌 시장에서 고려해야 할 기술적, 경제적, 윤리적, 지정학적 도전 과제가 남아있다.

중국 주요 AI 기업들은 이 과정에서 공통적으로 아래와 같은 점들을 위협요인으로 지목하고 있다. 먼저, 화웨이는 지정학적 갈등, 탈세계화, 세계 경제 침체로 인한 불확실성 증가를 인식하고 있으며, 이에 대해 국제 무역 규칙을 옹호하고 사이버 보안 및 개인정보 보호 거버넌스 시스템을 지속적으로 강화해 나가는 중이다. 알리바바는 AI 기술 개발 및 사용이 수반하는 편향, 오류, 시스템 장애, 사이버 보안, 개인정보 보호, 지적재산권 및 윤리적 문제 등의 문제와 중국의 국내 AI 관련 법제도가 강화됨에 따라 야기될 규제 준수 비용의 증가, 기술 및 비즈니스 관행 변경 등에 우려를 표시하고 있다. 또한, 미국 등의 첨단 컴퓨팅 칩 및 반도체 제조 장비에 대한 수출 통제 조치는 클라우드 인텔리전스 그룹의 제품 및 서비스 제공 능력에 부정적인 영향을 미칠 수 있음을 지적한다. 바이두 역시 생성형 AI 기술과 관련된 규제 및 법적 프레임워크 강화가 규제 준수 비용을 증가시킬 수 있으며 미국의 반도체 수출 통제 조치가 바이두의 AI 기술 개발 및 역량에 제한을 가하여 미래 사업 및 재무 성과에 부정적인 영향을 미칠 수 있음을 인지하고 있다.

2. AI 정책 및 투자 현황

가. 중국 정부의 AI 정책 방향

중국 정부의 AI 정책은 2015년 발표한 중국제조 2025와 인터넷플러스 전략부터 시작된다. 중국제조 2025는 제조업의 지능화, 디지털화, 네트워크화를 통해

제조업 전반의 산업구조 고도화를 추진하는 전략이며, 인터넷 플러스 전략은 인터넷, 빅데이터 클라우드 컴퓨팅, 인공지능 등 정보 기술을 전통 산업과 융합하여 새로운 성장동력을 창출한다는 방향성을 제시하고 있다. 이 두 정책은 중국의 디지털 전환과 그 근간인 통신 네트워크, 반도체, 클라우드 및 디지털 기술의 자체 개발과 생태계 혁신을 통해 중국의 글로벌 리더십을 확보하려는 목적을 갖고 있다. 또한, 중국 국무원의 “차세대 인공지능 발전계획(2017)”을 통해 2020년까지 세계적 수준의 AI 역량을 달성하고, 2025년까지 선도국이 될 것이며, 2030년까지 글로벌 AI 리더십을 확보할 것을 천명한 이후, 중국은 AI 역량 제고 및 기술자립, 그리고 미중 기술패권경쟁에 대응하기 위한 전방위적인 자국 생태계 구축을 위해 매진하고 있다.

〈표 3-10〉 차세대 AI 발전계획에 제시된 주요 목표와 응용영역

구분	전략 목표	응용 영역
1단계 (~'20)	<ul style="list-style-type: none"> AI 전체적인 기술 수준을 글로벌 선진국 수준으로 향상(핵심 산업 규모 1,500억 위안, 관련 산업 규모 1조 위안) 고급 인재 유치·육성정책 AI 관련 법률·윤리·규제적 프레임워크 확립 	<ul style="list-style-type: none"> 중점 영역의 혁신적 응용 확대 추진
2단계 (~'25)	<ul style="list-style-type: none"> AI가 산업을 업그레이드하고, 경제 및 전반의 엔진으로 작동하는 사회 구현 기존 이론 중심 돌파를 만들어 일부 기술 응용이 글로벌 최고 수준 도달(핵심 산업 규모 4,000억 위안, 관련 산업 규모 5조 위안) AI 이론·기술·응용 모두 세계 최고 수준 도달, 중국을 글로벌 AI 혁신 선도국 조성 	<ul style="list-style-type: none"> AI 산업 경쟁력을 최고로 높이고, 응용 분야 확대(스마트제조·스마트의료·스마트도시·스마트농업 등) 분야에서 폭넓은 응용 추진
3단계 (~'30)	<ul style="list-style-type: none"> AI 이론·기술·응용 모두 세계 최고 수준에 도달, 중국을 글로벌 AI 혁신 선도국 조성 AI 산업경쟁력 리더십 확보(핵심 산업 규모 1조 위안, 관련 산업 규모 10조 위안) AI 인재 양성 기지 구현 AI 관련 법률·윤리·규제체계 개선 	<ul style="list-style-type: none"> 생산·생활·사회 거버넌스 등 각 분야에서 응용의 폭과 깊이 확장

자료: 이정아(2025. 3.)

AI 이론, 기술, 응용을 모두 세계 최고 수준으로 끌어올리고, AI를 통해 산업구조의 고도화를 추진하는 방향은 현재 중국 정부의 AI 정책으로까지 이어지고 있다. 2024년 3월에 개최된 정부 업무보고에서는 인공지능+ 이니셔티브가 제기되었으며, 제조업 디지털 전환과 적극적인 데이터 개발, 개방, 유통으로 생산성의 질적 제고를 목표로 하고, 독자적으로 AI 산업을 적극적으로 육성하며, 향후 AI 컴퓨팅 플랫폼의 국산화, 운용 효율 향상, 중국 자체 GPU 생태계의 산업경쟁력 강화, 산업현장에서 AI의 빠른 확산 등의 내용을 골자로 하고 있다(유재홍, 2024. 7. 30.). 더 자세히 보자면, 대규모 모델의 광범위한 적용 가속화와 새로운 서비스와 제품의 등장 촉진, 신에너지 자동차, AI 기반 휴대폰 및 컴퓨터, 스마트 로봇, 지능형 제조 장비 등 차세대 스마트 단말기 기반에 중점을 두어 AI가 단순한 소프트웨어 솔루션을 넘어 물리적 제품과 인프라에 내재화되는 방향을 고려하고 있다. 뿐만 아니라 이를 위해 중소기업을 포함한 더 많은 기업들의 AI 접근가능성, 실행가능성에 중점을 두고 저비용, 고성능 AI 등의 개발과 확산을 통한 비용효율성을 강조한다(新华社, 2025. 3. 6.).

2025년 8월에는 인공지능+ 행동계획의 심층 이행이 국무원을 통과했다(新华社, 2025. 8. 1). 대대적인 산업 응용을 강조하는 이 정책은 방대한 국내 시장과 다양한 산업 분야가 AI 모델 훈련을 위한 대규모 데이터셋을 제공한다면 기술이 빠르게 확산될 수 있는 환경적 우위로 작용할 것이라고 인식하고 있다. 이 정책 문건에서는 다음과 같은 내용을 강조하고 있다. 첫째, 실용주의적 통합이다. AI를 단순한 첨단 기술이 아닌 제조, 농업, 의료 등 '실물 경제' 전반에 걸쳐 생산성을 높이고 효율성을 개선하는 범용 기술로 활용하려는 의지가 드러나며 이는 '신질 생산력'이라는 새로운 경제 성장 모델의 핵심 동력으로 AI를 자리매김하는 것을 의미한다. 둘째, 자급자족 및 회복력 강화 노력이다. 미국의 기술 제재에 대한 대응으로 국내 컴퓨팅 인프라(동수서산) 및 반도체 자급자족 노력을 지속하고 오픈소스 생태계를 통해 외부 의존도를 줄이면서도 혁신을 가속화하려는 전략이 강조된다. 셋째, 균형잡힌 거버넌스 구축이 강조되어 혁신을 장려하면서도 콘텐츠 통제, 데이터 보안, 윤리적 원칙을 강조하는 '발전과 안보의 균형'을 추구하고, 특히

AI 생성 콘텐츠에 대한 라벨링 의무화와 같은 구체적인 규제는 정보 환경에 대한 국가의 통제를 강화하려는 의도를 보여준다. 넷째, 인재 및 데이터의 전략적 활용을 중시하여 ‘역두뇌 유출’을 활용한 인재 유치 및 양성과 데이터의 자산화를 통한 국가적 활용 등을 통해 AI 발전을 위한 핵심 자원을 확보하려는 노력이 강화될 것으로 보인다. 마지막으로, 국내 발전에 집중하면서도 글로벌 AI 거버넌스 논의에 적극적으로 참여하고 ‘중국 솔루션’을 제시하며 국제적 영향력을 확대하려는 의지를 표명하고 있다.

나. 중국 정부의 AI 투자현황

중국의 AI 전략은 자체 생태계 구축과 산업 응용으로 집약될 수 있다. 반도체 칩부터 LLM(대형언어모델)에 이르기까지 AI 전체 기술 생태계의 자립이 전략적 목표이며 미국의 수출통제에 대응하는 국가적 전략이다.

주요 AI 인프라를 구성하며, 미국의 수출통제에 가장 큰 영향을 받는 반도체 분야는 자본집약적으로 국가의 중점 지원 대상이며 소프트웨어 프레임워크 개발은 기존 대형 민간 기술기업이 담당하고 AI 모델과 응용은 비교적 자유로운 민간에서 추진하는 등 AI 생태계 구축이 빠른 속도로 진행되는 중이다.

칩 제조 기술 지원을 위한 중국 정부의 대규모 공공 자금 지원 프로그램으로는 “국가 집적회로 산업 투자 기금”(“빅 펀드”)가 있다. 2014년 9월에 설립된 빅 펀드는 국무원의 “국가 집적회로 산업 발전 지침”에 대한 직접적인 대응으로 집적회로 산업 발전을 위한 집중적인 지원을 제공하는 것을 목표로 한다. 이 기금은 산업정보화부(MIIT)와 재정부 산하의 법인으로 운영되는데, 이사회는 전략을 수립하고 주요 프로젝트를 승인하며, 시노 IC 캐피탈(Sino IC Capital)은 투자를 실행하고 자금을 관리하는 2단계 관리 구조로 운용하고 있다. 펀드는 칩 제조 촉진, 인수합병 장려, 2030년까지 반도체 공급망의 모든 부문에서 글로벌 리더십 달성을 포함하고 있는데, 미국 수출통제에 대응하며 기금의 투자 방식도 변화해 왔으며 후기 단계에서는 특히 공급망의 취약점을 보완하는 데에 집중하고 있다. 이 펀드는 2014년 설립 이후 5년을 주기로 한 기씩 운용되고 있는데, 1기(2014-

2019)에는 칩 제조, 2기(2019-2024)에는 웨이퍼 제조를 중심으로 장비, 재료, 칩 설계, 패키징 및 테스트를 포함한 집적회로 산업 체인 강화를 목표로 하였다면, 3기(2024-2029)는 고대역폭 메모리(HBM) 산업 및 AI 반도체의 다른 핵심 영역에 중점을 두고 있다. 3기의 등록 자본금은 3,440억 위안(약 475억 달러)으로 이는 이전 두 단계의 총액을 합친 것보다 훨씬 많으며, ICBC, 중국농업은행, 중국건설은행, 중국은행, 교통은행, 중국 우정저축은행을 포함한 주요 국영 은행들이 총 1,140억 위안의 자본을 출자하였다(央视网, 2024. 5. 28).

또 다른 펀드로는 국가급 인공지능산업투자펀드(国家人工智能产业投资基金)가 있다. 2025년 1월 설립된 이 펀드는 총 규모 600.6억 위안(한화 약 11조 6천 억 원)으로 중국 공업정보화부(MIIT)와 재정부가 주도하여 설립되었다. 이는 국가 차원에서 AI 산업을 직접 지원하고 육성하겠다는 강력한 신호이다. 펀드의 주요 출자자들은 중국의 핵심 국영 기업 및 투자 기관들로 구성되어, 펀드의 안정성과 전략적 중요성을 더한다(第一财经, 2025. 1. 21).

‘국가대기금 3기’로 불리는 국가집적회로산업투자펀드 3기 주식회사(国家集成电路产业投资基金三期股份有限公司)는 총 등록 자본금 3,440억 위안(약 66조 원) 규모로 중국 역사상 최대 규모의 반도체 투자 펀드이다. 이 펀드가 국가 AI 펀드에 출자함으로써, AI 관련 칩, 컴퓨팅 칩, HBM 칩 등 AI 구동의 핵심 영역에 대한 투자를 확대할 것으로 예상된다. 이는 AI 산업의 하드웨어 인프라 강화에 중점을 두겠다는 의미로 해석될 수 있다. 이 펀드는 AI 산업의 전반적인 생태계를 포괄하는 광범위한 투자를 목표로 하는데 인공지능 전체 산업 체인에 걸쳐 투자를 진행하며, 컴퓨팅 파워, 알고리즘, 데이터 및 응용 프로그램 등 AI 산업의 다양한 핵심 링크를 포괄한다. 특히 ‘구현 지능(具身智能)’ 방향을 매우 중요하게 여기며, 관련 첨단 신기술, 신제품, 산업화 및 상업화 진행 상황을 면밀히 주시하며 투자 배치를 진행할 예정이다(第一财经, 2025. 1. 3.). ‘구현 지능’은 로봇공학, 자율주행 등 물리적 세계와 상호작용하는 AI 기술을 의미하며, 이는 중국이 실물 경제와의 AI 융합을 강조함을 시사한다. 이 펀드는 산업 투자 속성뿐만 아니라 일정 부분 벤처 투자 속성도 가지고 있어 초기 단계의 유망한 AI 스타트업에도 투자하여

혁신을 촉진할 것으로 보인다. 운영 원칙은 공업정보화부와 재정부의 감독 및 지도 하에 제한된 자금을 효율적으로 활용하고, 핵심 링크에 집중하며, ‘인내 자본(patient capital)’³¹⁾의 역할을 발휘하여 중국 인공지능 산업의 혁신적인 발전을 위한 장기적이고 안정적인 자금 요소를 보장한다. 이러한 정책과 펀드 운용 계획은 중국 정부가 AI 산업, 특히 LLM과 관련된 핵심 기술 및 응용 분야에 대한 강력한 통제와 지원을 통해 국가 주도형 AI 생태계를 구축하려는 전략을 명확히 보여준다.

이 외에도, 지방정부들 역시 다양한 지원책을 통해 AI 생태계의 발전을 위한 노력을 경주 중이다. 베이징시의 경우, 지방 기업이 국산 GPU를 구매할 경우, “지능형 컴퓨팅 서비스 구축 투자 규모의 일정 비율”에 대해 보조금을 지급하며 (Reuters, 2024. 4. 26), 상하이, 항저우, 선전, 청두 등 주요 지방 정부들도 비슷한 지원정책을 마련하여 AI 클라우드 인프라 보조금, 데이터 수집 및 모델 개발 지원을 포함한 종합 보조금 패키지를 제공 중이다. 상하이는 컴퓨팅 보조금 6억 위안, AI 모델 할인 3억 위안, 데이터·학습 보조 1억 위안 규모의 프로그램을 운영한다(연합뉴스, 2025. 7. 29).

3. AI 규제 현황

가. 기본 철학: 디지털 주권과 기술경쟁력의 이중 목표

중국의 AI 규제는 국가 안보와 사회 안정을 위한 포괄적 AI 통제라는 뚜렷한 목표 아래 구축되어 있다. 이는 단순한 기술 규제가 아니라 ‘디지털 주권’ 확보를 통한 종합적 사회 거버넌스의 핵심 수단으로 AI를 활용한다는 전략적 인식을 반영한다. 중국의 AI 규제 체계는 “입법 우선, 윤리적 지침 및 분류된 거버넌스”라는 접근법을 취하고 있으며(IAPP, 2025), 데이터 컴플라이언스, 알고리즘 컴플라이언스, 사이버 보안 및 윤리를 포괄하는 다층적 법령 체계에 기반을 두고 있다.

31) 단기 수익보다는 장기적인 성장과 전략적 목표 달성을 중시하는 투자

그러나 중국 AI 거버넌스의 보다 중요한 특징은 강력한 규제적 통제와 적극적 기술 진흥이 서로를 강화하는 방식으로 설계되어 있다는 점이다. 중국은 2017년 국무원이 발표한 ‘신세대 인공지능 발전계획’을 통해 2030년까지 AI 분야에서 글로벌 리더가 되겠다는 목표를 설정하고 이를 3단계 로드맵으로 추진해 왔다(State Council, 2017). 2025년 현재 중국의 AI 관련 특허 보유량은 전 세계의 약 60%를 차지하며(Global Times, 2025), 2025년 6월 기준 생성형 AI 사용자 기반은 5억 1,500만 명에 달해 반년 만에 두 배로 증가하였다(CNNIC, 2025). 이러한 기술 경쟁력을 바탕으로 중국의 AI 규제는 뒤따라가는 규제가 아니라 산업 발전을 주도하고 형성하는 선제적 성격을 띠고 있다.

나. 핵심 입법 및 정책 체계

1) 단계적·기술별 규제 확대 전략

중국 AI 규제의 가장 독창적인 특징은 포괄적 단일 법률 대신 기술 유형별로 규제를 순차적으로 도입하는 단계적 확대 전략이다. 2022년부터 2025년에 걸쳐 이 전략은 다음과 같이 전개되었다.

2022년 3월 시행된 「인터넷 정보서비스 알고리즘 추천 관리규정」은 여론 형성이나 사회적 동원 능력을 갖춘 알고리즘 추천 서비스 제공자에게 서비스 제공 후 10일 이내에 국가인터넷정보판공실(CAC)에 알고리즘을 등록하도록 의무화하였다. 2024년 10월 기준 수천 건 이상의 알고리즘이 등록되어 있으며, 이 제도는 국가가 AI 시스템의 작동 방식을 직접 파악하고 통제하는 기반으로 기능하고 있다(IAPP, 2025).

2023년 1월에는 「인터넷 정보서비스 심층합성 관리규정」(CAC, 2023a)이 시행되어 딥페이크를 포함한 합성 콘텐츠의 제작·배포를 규율하기 시작하였다. 서비스 제공자는 불법 정보의 생성·배포를 금지하고 이용자 등록, 알고리즘 심사, 윤리 심사, 콘텐츠 모니터링 체계를 갖추어야 한다(White & Case, 2025).

같은 해 8월에는 세계 최초의 구속력 있는 생성형 AI 규제인 「생성형 인공지능 서비스 관리 잠정방법」이 시행되었다. AI가 생성하는 콘텐츠가 사회주의 핵심 가

치를 견지하도록 규정하고, 국가 권력 전복이나 국가 안보를 위태롭게 하는 내용을 금지하였다(CAC, 2023b). 서비스 제공자는 CAC에 서비스를 등록해야 하며, 2024년 말 기준 302개 생성형 AI 서비스가 승인을 완료하였고 등록 사용자 수는 6억 명을 넘어섰다(중국국제방송국, 2025).

2025년 3월 14일에는 CAC, 공업정보화부,公安부, 국가방송TV총국의 4개 부처가 공동으로 「인공지능생성형내용표식방법」을 발표하고 같은 해 9월 1일부터 시행하였다(CAC et al., 2025). 이에 따라 AI가 생성한 텍스트, 이미지, 오디오, 비디오, 가상 콘텐츠 등 모든 디지털 결과물에는 이용자가 쉽게 인지할 수 있는 명시적 라벨과 파일 메타데이터에 내장되는 암시적 라벨을 함께 표시하여야 한다.

2025년에는 규제가 더욱 체계적으로 강화되었다. 2025년 4월 25일 국가시장감독관리총국과 국가표준화관리위원회가 공동 발표한 3개 국가 표준이 2025년 11월 1일 발효되었다(White & Case, 2025). 이 표준들은 생성형 AI 데이터 주석 보안 사양, 생성형 AI 사전훈련 및 미세조정 데이터 보안 사양, 생성형 AI 서비스 기본 보안 요구사항으로 구성되어 AI 개발 전 과정에 걸친 보안 체계를 규율한다.

같은 해 3월 21일 CAC와公安부가 공동 발표한 「안면인식 기술 응용 안전 관리방법」은 2025년 6월 1일 발효되었다. 이 조치는 안면인식의 목적 명확성·필요성·최소화 원칙, 미성년자 권익 보호, 데이터 로컬 저장, 주요 용도 사전 신고 의무 등을 규정하고 있다(36氪, 2025).

국가 사이버보안 표준화 기술위원회(TC260)는 2025년 9월 9일 AI 거버넌스 프레임워크 2.0을 발표하였다. 2024년 9월 발표된 1.0 버전을 업그레이드한 이 문서는 AI 위협을 내재적 위협과 응용 위협으로 분류하고, 거대언어모델 관련 문제, 윤리 문제, 편향·차별, 훈련 데이터셋 오염, 데이터 유출 등에 대한 구체적 대응 방안을 제시한다(IAPP, 2025). 다만 구속력 없는 모범사례 가이드라인의 성격을 갖는다.

2025년 10월 28일 전국인민대표대회 상무위원회를 통과한 사이버보안법(CSL) 개정안은 2026년 1월 1일 발효되었다(NPC Standing Committee, 2025).

2017년 제정 이후 최초의 주요 개정으로, 국가 사이버보안법에 AI 관련 조항을 처음으로 도입하였다. AI 기초 연구 및 핵심 알고리즘 혁신 지원, AI 관련 인프라 개발, AI 윤리 기준 강화, 보안 위협 모니터링 강화 등이 주요 내용이며, 구체적인 이행 규정과 기술 표준은 추후 별도 하위 법령으로 제정될 예정이다(Global Policy Watch, 2025).

2) 포괄적 AI 법안 추진 현황

단계적 접근이 지속되는 가운데, 중국은 포괄적 AI 법률 체계의 구축도 함께 모색하고 있다. 2024년 5월 AI 법안 초안을 발의하고 국무원 입법 의제에 포함시켰으나 2025년 12월 기준으로 포괄적 AI법의 공식 제정·공포는 이루어지지 않았다. 이는 중국이 포괄적 단일 입법보다 기술 유형별 단계적 규제와 사이버보안법, 개인정보보호법, 데이터보안법 등 기존 법률 체계를 활용하는 방식을 당분간 지속할 것임을 시사한다.

다. 혁신 촉진과 리스크 관리의 균형 방식

1) AI Plus 행동계획 — 혁신과 규제의 병행

2025년 8월 26일 국무원은 “인공지능+” 이니셔티브 심화 추진에 관한 의견(国发[2025]11号)(이하 ‘AI Plus 행동계획’)을 발표하였다(State Council of China, 2025). 이 계획은 AI를 경제, 사회 전 분야에 광범위하고 심층적으로 통합하여 ‘지능 경제(智能经济)’와 ‘지능 사회(智能社会)’를 형성하겠다는 중국의 장기 전략을 집약한 문서로, 단순한 산업 진흥책을 넘어 국가 발전 모델 자체를 AI 중심으로 재편하겠다는 의지를 담고 있다.

AI Plus 행동계획은 2025년부터 2035년까지 3단계 목표를 제시한다. 2027년까지는 과학기술 발전, 산업 활용, 소비 서비스, 공공복지, 거버넌스·안보, 국제협력의 6개 핵심 분야에서 AI의 광범위한 통합을 달성하고 차세대 스마트 단말기 및 AI 에이전트 보급률을 70% 이상으로 확대한다. 2030년까지는 AI가 고품질 발전을 전면적으로 지원하는 체계를 갖추고 보급률 90% 이상을 달성하며, 2035년까지는 완전한 지능 경제, 지능 사회로의 전환을 완성한다는 것이 이 계획의 골자

이다(State Council of China, 2025). 아울러 AI Plus 행동계획은 AI를 기존 기업에 접목하는 수준을 넘어, 처음부터 AI를 근간으로 설계되는 ‘AI 네이티브(AI-native)’ 기업의 육성을 별도로 강조하고 있다.

2) AI 글로벌 거버넌스 행동계획 — 국제 표준 선도 전략

2025년 7월 세계 인공지능대회(WAIC)를 계기로 중국은 ‘인공지능 글로벌 거버넌스 행동계획’을 발표하였다. 13개 항목으로 구성된 이 행동계획은 각국 정부, 기업, 개인의 AI 참여를 독려하고, 기술 장벽을 완화하며, AI를 다양한 분야에 적용하고, 글로벌 사우스 국가들의 AI 발전을 지원하는 내용을 담고 있다(USSC, 2025). 리창 총리는 이를 계기로 ‘세계 AI협력기구’ 설립을 제안하면서 “더 많은 중국의 방안을 국제 사회에 제공하고 세계 AI 거버넌스에 더 많은 중국의 지혜를 공헌할 것”이라고 강조하였다. 미국 중심의 AI 거버넌스 질서에 대한 대안으로 중국 모델을 국제 무대에 적극 제시하려는 의도가 담겨 있다.

라. 거버넌스 체계와 집행 역량

1) 7개 부처 공동의 다부처 통합형 구조

중국의 AI 거버넌스는 7개 정부 부처가 공동으로 규제를 마련하고 집행하는 다부처 통합형 구조로 운영된다. CAC가 주도하고 국가발전개혁위원회(NDRC), 교육부, 과학기술부, 공업정보화부(MIIT), 공안부, 국가방송TV총국 등이 협력한다. NDRC는 AI Plus 행동계획의 총괄 조정 기관으로, 각 지역과 부서가 실정에 맞는 조치를 취하도록 감독한다(State Council of China, 2025). 규제 권한이 분산되어 있으면서도 국가 차원의 전략적 목표 아래 부처 간 협력이 긴밀하게 이루어진다는 점이 이 구조의 핵심 특징이다.

2) 강력한 집행 체계

중국 AI 규제는 강력한 사법 집행력을 바탕으로 실질적인 구속력을 갖추고 있다. 2025년 4월 CAC는 “밝고 깨끗한: AI 기술 남용 정화”라는 3개월 전국 캠페인을 시행하여 3,500개 이상의 AI 관련 제품을 제거하고, 96만 개 이상의 불법

정보를 삭제하였으며, 3,700개 이상의 계정을 처벌하였다(HSF Kramer, 2025). 2024년 8월에는 베이징 인터넷법원이 AI 음성 저작권 침해 소송에서 피고에게 25만 위안의 손해배상을 명령하였는데, 이는 중국에서 처음으로 AI 저작권 침해를 인정한 판결로 평가된다(Xinhua, 2024. 8. 21). 행정적 단속과 사법적 집행을 결합하는 방식으로 중국은 규제를 선언적 수준에 머물지 않게 하는 집행 문화를 만들어가고 있다.

3) 과학기술 윤리 심사 제도

2023년 12월부터 시행된 과학기술 윤리 심사 제도는 AI 및 특정 기술 활동에 종사하는 기업에게 의무적 윤리 심사를 요구한다. 이 제도는 기술 개발 단계에서부터 국가 가치와 정책 방향에 부합하는지를 검증함으로써, 중국이 AI 기술 발전의 전 주기에 걸쳐 실질적인 감독 권한을 행사할 수 있는 법적 기반을 마련한 것으로 평가된다(Chambers and Partners, 2025a).

4) 국제적 영향력 확산 전략

중국은 Sovereign AI 정책을 통해 자국의 AI 기술 개발·배포·사용에 대한 완전한 통제권을 확보하는 한편, AI 글로벌 거버넌스 행동계획을 통해 국제 무대에서의 영향력을 넓혀가고 있다. 특히 개발도상국을 대상으로 한 AI 발전 지원과 ‘중국 모델’의 글로벌 사우스 국가들로의 확산은 미국과 EU 중심의 글로벌 AI 표준에 맞서는 대안적 거버넌스 전략으로 읽힌다(USSC, 2025). 2025년 초 Deep Seek의 부상은 이러한 전략의 기술적 기반이 실제로 형성되고 있음을 보여주는 사례로, ChatGPT를 제치고 중국과 미국 모두에서 애플 앱스토어 다운로드 1위를 기록하며 글로벌 AI 시장에서 중국의 기술 경쟁력을 입증하였다(IAPP, 2025).

국제 표준화 측면에서도 중국은 적극적인 행보를 이어가고 있다. ISO/IEC 기술 위원회 참여를 통해 AI 국제 표준 논의에서 주도권을 확보하고, 상하이협력기구(SCO), BRICS 등 다자 프레임워크를 통해 중국 AI 규범을 확산시키는 한편, ‘세계 AI협력기구’ 창설을 제안하며 미국, EU와는 다른 새로운 국제 AI 거버넌스 플랫폼을 구축하려 하고 있다.

바. 평가 및 시사점

중국 모델의 강점은 크게 네 가지로 정리된다. 첫째, 포괄적 단일법 대신 기술 유형별로 규제를 순차적으로 도입하는 방식은 기술 발전 속도에 맞춰 적절한 제도를 마련할 수 있다는 점에서 합리적인 접근이라고 할 수 있다. 둘째, 강력한 집행 체계를 통해 규제가 실제로 작동하도록 만드는 능력은 다른 국가들이 갖추기 어려운 구조적 강점이다. 셋째, AI Plus 행동계획으로 대표되듯이 규제와 진흥을 전략적으로 병행하여 산업 발전과 통제를 동시에 추구한다. 넷째, 알고리즘 등록제, AI 생성 콘텐츠 라벨링 등 구체적이고 집행 가능한 규제 도구를 설계한다는 점에서 규제의 실무적 완성도가 높다.

그러나 한계도 분명하다. 국가 통제 중심의 접근은 민간 혁신 생태계의 자율성을 제약할 우려를 낳고, 사회주의 핵심 가치 준수 등 정치적 의무는 중국 AI 기술의 국제적 신뢰성에 부정적 영향을 미친다. 포괄적 AI법이 부재한 상황에서 규제가 여러 부처에 분산되어 있어 일관성 확보가 쉽지 않으며, 역외 적용의 한계로 인해 글로벌 기업들이 중국 규제 환경에 적응하는 데 상당한 부담을 안고 있다.

한국에 대한 시사점은 크게 세 가지다. 첫째, 기술 유형별 단계적 규제 도입 전략은 기술 발전 속도와 제도적 대응 속도 간의 격차를 줄이는 데 효과적이다. 한국의 경우 AI 기본법에서 고영향 AI의 범위를 처음부터 고정하기보다는 기술 발전에 따라 지속적으로 확장·조정할 수 있는 동태적 분류 체계를 갖추는 것이 중요하다. 둘째, 알고리즘 등록제와 AI 생성 콘텐츠 라벨링 등 집행 가능한 구체적 도구의 설계 방식은 AI 기본법 하위 규범을 만들 때 직접 참고할 수 있다. 셋째, AI Plus 행동계획처럼 규제 체계와 산업 진흥 전략을 긴밀하게 연동하는 방식은 AI 기본법 시행과 산업 육성 정책 사이의 연계를 강화하는 데 시사하는 바가 크다. 다만 국가 통제 중심의 거버넌스 모델은 민주적 법치국가인 한국에 그대로 이식하기 어려운 부분이 있으므로 중국 모델의 집행력이라는 강점은 취하되 기업의 자율성과 우리의 공적 가치를 존중하는 방식으로 재설계하는 것이 필요하다.

제 3 절 E U

1. AI 산업생태계 구조 및 주요 플레이어

EU의 AI 생태계는 중요한 구성 요소에 대해 미국 기업에 크게 의존하고 있다. 예를 들어, 유럽 칩 제조업체는 AI 가속기 분야에서 NVIDIA와 같은 시장 선두주자와 직접 경쟁할 만한 경쟁 우위를 가지고 있지 않으며 구글, 아마존, MS 등에 대한 클라우드 의존도는 점차 높아지는 상황이다. 최첨단 대규모 언어 모델(LLM) 개발의 상당 부분은 여전히 OpenAI와 같은 미국 모델에 의존하거나 비유럽 출처로부터 상당한 투자를 유치하고 있다. EU의 클라우드 시장은 AWS, 구글, MS가 70% 이상의 압도적인 점유율을 차지하고 있으며, 유럽 공급업체 전체의 합산 점유율은 15%에 불과한 실정이다(ITdaily, 2025. 7. 3).

EU는 이러한 의존성을 줄이고 자국 AI 풀스택을 구축하기 위한 다각적인 노력을 전개하고 있는데 EU 내 언어다양성과 문화다양성 등을 존중한 다국어 LLM 모델과, 프랑스를 중심으로 하는 자국산 LLM 개발 노력이 가장 주축이 된다. 이 LLM 개발 과정에서 역시 EU 내 클라우드 활용을 통해 대미 인프라 의존도를 낮추려는 노력도 동시에 진행된다. 중국에서 이와 같은 노력이 중국 내 기업들을 통해 주로 이루어진다면 EU 내에서는 정부 산하 연구기관의 프로젝트 형태가 주를 이룬다. OpenEuroLLM 프로젝트는 유럽 언어 및 기타 사회경제적으로 중요한 언어를 위한 강력한 다국어 기초 모델 개발을 진행하고, EU 회원국은 아니지만 스위스에서도 ETH 취리히와 EPFL이 협력하여 공공 인프라에서 개발된 다국어 LLM을 출시할 예정이다. 대표적인 AI 기업인 프랑스의 Mistral AI(프랑스), 유럽 기업가들이 설립했지만 뉴욕에 본사를 둔 허깅페이스 등 유럽 기업들이 오픈소스 LLM과 프레임워크를 개발하며 인정을 받고 있으며 이들의 노력은 유럽 내에서 AI 풀스택 기업 생태계를 구축하기 위한 중요한 축으로 작용한다.

AI 칩 및 하드웨어 측면에서 유럽은 ASML, Infineon, STMicroelectronics와 같은 세계적인 반도체 전문 기업을 보유하고 있지만 AI 가속기 분야에서 엔비디

아와 같은 시장 선두 주자와 직접 경쟁하지는 않는다. 그러나 SiPearl(프랑스), Fractile(영국), Axelera AI(네덜란드), Semron(독일), GreenWaves Technologies(프랑스) 등 AI 칩 개발에 적극적으로 참여하는 유럽 스타트업들이 점차 늘고 있다. 프랑스 파리 근교의 대규모 AI 캠퍼스 건설은 이러한 하드웨어 역량을 유럽 내에 구축하려는 핵심적인 노력으로, NVIDIA의 GPU 기술과 유럽의 인프라 및 자금 지원이 결합된 형태이다.

• 딥러닝 프레임워크

허깅페이스는 PyTorch, TensorFlow와 같은 주요 딥러닝 프레임워크와의 통합을 통해 개발자들이 AI 모델을 쉽게 개발, 훈련, 배포할 수 있도록 지원한다. 또한, OpenEuroLLM 프로젝트는 유럽 연구 기관, 기업, 고성능 컴퓨팅(HPC) 센터의 컨소시엄을 통해 유럽의 가치와 규제에 부합하는 오픈소스 다국어 LLM을 개발하며, EuroHPC 인프라를 활용한다. 스위스의 ETH 취리히와 EPFL도 ‘알프스(Alps)’ 슈퍼컴퓨터에서 훈련된 오픈소스 다국어 LLM을 출시할 예정이며 이는 AI 주권과 언어적 포괄성을 강조한다. OpenDR과 Flower와 같은 오픈소스 딥러닝 프레임워크도 유럽에서 개발되고 있다.

• 클라우드 및 인프라

미스트랄 AI와 스케일웨이의 협력은 유럽 내에서 주권적인 클라우드 컴퓨팅 역량을 구축하는 중요한 사례이다. EU는 ‘AI 기가팩토리’ 이니셔티브를 통해 차세대 AI 인프라를 구축하고 주권 컴퓨팅 역량을 제공하며, ‘AI 온디맨드 플랫폼(AIoD)’은 유럽의 AI 도구, 데이터셋 및 서비스를 통합하는 중앙 디지털 접근 지점 역할을 하여 분절된 AI 자원을 통합된 인프라로 전환한다.

결론적으로, EU는 EU 내 AI 풀스택을 구축하고 미국에 대한 의존도를 줄이기 위해 정부 차원의 막대한 투자와 다양한 협력 이니셔티브를 추진하고 있다. 칩, 프레임워크, LLM 등 AI 스택의 모든 계층에서 역량을 강화하려는 전략적 움직임이 두드러지며 여전히 높은 대미 의존도를 완화하기 위해 ‘기술 주권’을 확보하고 윤리적이고 인간 중심적인 AI 개발의 글로벌 표준을 제시하겠다는 비전을 가지고

있다. 중국의 자체 생태계 구축 노력과는 또 다른 축의 AI 표준을 만들어나가기 위해 노력으로 평가할 수 있다.

2. AI 정책 및 투자 현황

가. EU의 AI 정책

AI 기술이 국가경쟁력의 핵심 요소로 부상하면서 EU는 AI 분야에서의 기술 주권 확보를 주요 정책 목표로 설정하였다. 2024년 AI Act의 시행과 함께, EU는 단순한 규제적 접근을 넘어서 자립적이고 경쟁력 있는 AI 생태계 구축을 위한 전략을 종합적으로 추진하고 있는데 특히 현재 EU의 AI 생태계가 미국 기업에 크게 의존하고 있는 현실을 타개하고자 하는 노력이 각 기술스택에서 진행되고 있다.

EU의 인공지능에 대한 전반적인 비전은 인간 중심적이고 신뢰할 수 있는 AI 개발과 동시에 세계적인 AI 허브가 되겠다는 약속으로 명확히 표현된다. 기술적 리더십과 윤리적 거버넌스를 동시에 강조하는 전략은 EU를 다른 글로벌 AI 강국과 차별화하는 특징이다. 이러한 전략적 접근 방식은 2021년 4월 발표된 AI 패키지를 기반으로 구체적인 정책 조치 및 행동으로 나타났다. 이 패키지에는 유럽의 AI 접근 방식을 설명하는 커뮤니케이션, EU 회원국과 공동으로 개발된 AI 조정 계획에 대한 포괄적인 검토, 그리고 관련 영향평가와 함께 AI에 대한 획기적인 규제 프레임워크 제안이 포함되었다.

이러한 비전을 더욱 강화하기 위해 2025년 4월에 발표된 “AI 대륙 행동 계획”은 유럽을 AI 분야의 글로벌 리더로 자리매김 하는 것을 목표로 한다. 이 계획의 핵심은 유럽의 전반적인 경쟁력을 향상시키는 동시에 민주적 가치를 적극적으로 보호하고 발전시키기 위한 신뢰할 수 있는 AI 기술을 개발하는 것이다.

EU의 AI 법안과 같은 규제에 대한 선도적이고 포괄적인 채택은 단순한 규제를 넘어서는 의도적인 전략적 포지셔닝으로 볼 수 있다. 중국의 AI 개발이 국가 주도의 자국 생태계 구축 노력으로 특징지어지고 미국이 민간 부문 혁신에 크게 의존하는 반면, EU의 입장은 윤리적이고 신뢰할 수 있는 AI에 대한 글로벌 표준 설정자

가 되는 것이다. 이러한 움직임은 책임감 있는 AI 관행을 우선시하는 기업과 개발자를 유치하여, 신뢰, 투명성, 윤리적 고려가 광범위한 채택 및 대중의 수용을 위해 점점 더 중요해지는 미래 글로벌 시장에서 뚜렷한 경쟁 우위를 창출할 수 있다. 이는 “책임감 있는 AI”가 강력한 시장 차별화 요소로 발전할 것이라는 전략적 계산을 나타낸다. 이러한 선제적인 규제 리더십은 데이터 프라이버시에 대한 GDPR의 전 세계적 영향력과 유사하게 EU 표준이 사실상 글로벌 규범이 되는 브뤼셀 효과를 발휘할 잠재력을 가지고 있다. 그러나 이러한 전략은 관련 규제 준수 비용이 신생 스타트업 및 소규모 기업에 과도한 부담이 될 경우 혁신을 저해할 수 있는 내재된 위험도 내포하고 있다.

EU AI법은 규제뿐 아니라 AI 혁신을 지원하기 위한 여러 조항들을 담고 있으며, 이는 크게 AI 규제 샌드박스, 실제 환경에서의 테스트 허용, 중소기업 및 스타트업 지원 세 가지로 요약된다.

• 인공지능 규제 샌드박스

EU AI법은 각 회원국의 감독기관이 적어도 하나 이상의 ‘AI 규제 샌드박스’를 구축·운영하도록 의무화하고 있다. AI 규제 샌드박스란 시장 출시 전에 AI 시스템을 한정된 기간 동안 통제된 환경에서 개발·실험·테스트·검증할 수 있도록 지원하는 제도이다. 규제기관의 감독 하에 기업들이 혁신적 AI를 자발적으로 시험해 볼 수 있으며, 이 과정에서 EU법상의 요구사항 준수 여부를 점검하게 된다. 특히 샌드박스 내에서는 필요한 경우 현실 환경에서의 시험 운영도 일정 범위 내에서 허용되어, 새로운 AI 기술을 실제 환경에서 검증할 수 있다(정재욱, 2024. 7). 이러한 규정은 AI 기업들이 안전하고 합법적인 방법으로 혁신을 시도하도록 독려함으로써, 기술 성숙도를 높이고 시장 출시 소요 시간을 단축하는 효과를 가져올 것으로 기대된다. 한편 법령에는 샌드박스 운영 시 개인정보 보호 등 타 감독기관과의 협력, 위험 식별 시 즉각적 대응 및 테스트 중단 등 안전장치에 관한 세부 규정도 포함되어 있다.³²⁾ 이러한 조항은 EU 전체에 샌드박스 기반의 AI 혁신 인프라를 마련함으로써, 규제와 혁신의 동반 향상을 도모하는 핵심 장치라 할 수 있다.

• 실제 환경 테스트 허용

앞서 언급한 대로, AI 법은 일정 조건 하에서 샌드박스 외부의 현실 환경에서도 AI 시스템 테스트를 허용하고 있다. 이는 고위험 AI라 하더라도 안전조치가 갖춰진 경우 제한된 실제 조건에서 시범운영을 할 수 있도록 한 규정이다. 예컨대 해당 AI 시스템이 규제 샌드박스가 아닌 실제 환경에서 테스트되어야 할 기술적 필요성이 인정되는 경우 등 특정 요건을 충족하면 규제당국의 승인 하에 현장 실증(pilot)이 가능하다(박강민 외, 2024). 이 조치는 혁신 기술의 현실 적합성 검증을 촉진하여 규제 샌드박스에서 나온 기술을 실사용 환경으로 원활히 이행할 수 있게 돕는다. 단, 테스트 중 식별된 중대한 위험은 즉시 완화하거나 테스트를 중지해야 하며, 감독당국의 상시 모니터링을 받아야 한다는 조건이 붙어 있어 안전성과 공익 보호를 함께 도모한다

• 중소기업·스타트업 지원

EU AI 법은 혁신 주체인 중소기업 AI 기업에 대한 특별 지원 조항도 마련하고 있다. 구체적으로 스타트업·중소기업에 샌드박스 참여 기회 우선 부여, 법 준수에 관한 맞춤형 안내와 질의 창구 제공, 적합성 평가 수수료 감경 등의 내용을 담고 있어 규모가 작은 혁신기업들이 규제에 의한 진입장벽을 최소화할 수 있도록 배려하고 있다. 예를 들어 제71조(최종안 기준)에서는 회원국은 중소기업에게 규제 샌드박스에 대한 우선 접근권을 부여하고 이들에게 AI법 이행과 관련된 인식제고 활동과 소통채널을 제공한다”고 규정하고 있다. 또한 적합성 평가 비용도 기업 규모에 따라 차등 감면하도록 하여 스타트업의 규제준수 비용 부담을 덜어 주는 조치를 명시하였다(박강민 외, 2024). 이는 곧 AI 기술기업의 혁신 활동을 지원하면서도 윤리·안전 기준을 준수하도록 유도하여, 유럽산 AI의 경쟁력을 높

32) EU AI법 최종안 제69조에서는 “회원국은 2026년 8월 2일까지 최소 1개의 AI 규제 샌드박스를 국가 수준에서 운영하도록 보장해야 한다. 이 샌드박스들은 혁신을 촉진하고 AI 시스템의 개발, 훈련, 테스트 및 검증을 일정 기간 동안 지원하는 통제된 환경을 제공해야 한다. 필요시 현실 조건에서의 시험도 해당 샌드박스 감독 하에 포함될 수 있다.”

이고 EU 디지털 주권을 확보하려는 전략으로 볼 수 있다.

EU AI 법에 따라 2026년까지 전 회원국에 최소 하나 이상의 AI 샌드박스가 마련될 예정이며 이에 앞서 몇몇 국가는 이미 자국 주도의 AI 샌드박스 프로그램을 시범 운용하거나 준비 중이다. 예를 들어 핀란드는 AI 법 이행을 위해 2024년 작업반을 구성하여 국내 AI 샌드박스의 법적 근거와 운영 계획을 수립하고 있으며, 2025년 8월까지 관련 국내법을 정비한 뒤 2026년 8월까지 샌드박스를 가동한다는 로드맵을 발표했다(Chambers and partners, 2025b). 각국의 샌드박스는 국경 간 협력도 가능하며, 여러 회원국이 공동 운영하는 샌드박스나 기존 타 분야 규제 샌드박스와의 연계도 허용된다. 궁극적으로 EU는 샌드박스 네트워크를 통해 회원국 감독당국 간 베스트 프랙티스의 공유와 기업 교차 참여를 촉진하고 있다.

나. EU의 AI 투자 현황

유럽 연합의 AI 전략은 AI에 대한 전략적 투자, 연구기관과 민간 기업 간의 강력한 협력 촉진, 숙련된 인력 확보 및 유지 조치 시행, 데이터 보호 및 공유 명확화 및 촉진, AI 개발을 위한 명확한 윤리적 지침 수립, AI 관련 보안 문제 해결 등 7가지 핵심 영역을 명시하는 포괄적인 프레임워크를 중심으로 구성된다. 이러한 투자를 주도하는 핵심 재정 수단은 Horizon Europe 및 Digital Europe 등 기존의 R&D 프로그램이다. Digital Europe 프로그램(2021-2027)은 슈퍼컴퓨팅, 인공지능, 사이버 보안 및 고급 디지털 기술과 같은 핵심 역량 영역을 전략적으로 지원하기 위해 81억 유로가 넘는 예산을 투입한다. 이 프로그램은 2021년부터 2024년까지 AI 개발 및 배포에 10억 유로 이상을 투입하였으며 AI 혁신 패키지, 복구 및 복원력 기금, GenAI4EU 이니셔티브 등도 AI 개발 및 배포에 상당한 자금을 지원한다.

EU의 주요 연구 자금 지원 프로그램인 Horizon Europe은 2021-2022년 기간 동안 AI 연구 및 개발에 26억 유로를 할당했다. 또한 2024년 4월에 시작된 공모에서는 대규모 AI 모델 발전에 5천만 유로를 포함하여 AI 및 양자 기술 프로젝트에 1억 1,200만 유로 이상이 배정되었다(European Commission).

2024년 1월에 도입된 AI 혁신 패키지는 2024-2027년 기간 동안 최대 40억 유로 상당의 이니셔티브를 통해 AI 투자를 더욱 강화했으며, 특히 생성형 AI 기술의 개발 및 배포를 목표로 한다(KERC, 2025.6).

“AI 팩토리” 및 “기가팩토리” 설립에 대한 명시적인 강조는 컴퓨팅 능력이 AI 개발의 근본적인 병목 현상이라는 인식이 EU 내에 자리잡고 있음을 보여준다. AI 기가팩토리는 유럽의 AI 역량을 강화하고 기술 주권을 확보하기 위해, AI 개발에 필수적인 컴퓨팅 인프라 부족 문제를 해결하고, 유럽 내에서의 AI 모델 훈련, 개발 및 배포를 위한 기반을 마련하기 위한 이니셔티브이다. 이 이니셔티브는 미국 등 비유럽 클라우드 인프라 및 하드웨어에 대한 의존도를 낮추고 유럽산 컴퓨팅 역량을 구축하여 디지털 자율성을 확보하는 것을 가장 큰 목적으로 하며 슈퍼컴퓨팅 자원 제공, 대규모 AI 모델의 알고리즘 개발, 테스트, 평가 등을 위한 통합인프라, 저탄소 하이퍼스케일 데이터 센터를 통한 개방형 플랫폼 제공, 데이터 주권 보장, 산업 전반의 AI 응용 지원 등을 진행한다(EuroHPC 사이트).

EU의 AI 투자 전략은 단순히 이론적인 연구 프로젝트에 자금을 지원하는 것을 넘어 대규모 AI 개발 및 훈련에 필요한 실질적인 물리적 및 디지털 인프라를 적극적으로 구축하는 데 중점을 둔다. 또한, 중국 대비 뒤쳐져있는 것으로 평가받아왔던 EU의 데이터 인프라를 확충하여 경쟁력을 확보하고 EU의 혁신 환경을 특징짓는 “분절화와 불충분한 규모”라는 만연한 문제에 대한 직접적으로 대응하기 위해 노력 중이다. 이러한 공공 투자의 증가는 AI 팩토리 등 최첨단 AI 인프라의 구축을 통해 유럽 스타트업 및 연구자에게 컴퓨팅 능력을 개선하고 더 쉽게 접근할 수 있도록 함으로써 EU AI 부문의 전반적인 경쟁력을 강화하고 중요한 컴퓨팅 자원에 대한 비EU 하이퍼스케일러에 대한 의존도를 줄이는 데 기여할 것으로 예상된다.

EU의 전략은 “연구실에서 시장으로의 전환을 원활하게 하는” 이니셔티브를 명시적으로 설정하고, 공공-민간 파트너십에 대한 적극적으로 장려를 통해 실행된다. 이는 유럽의 기초 연구 강점과 혁신 기술의 상업화 및 확산에 대한 약점이라는 인식을 분명히 보여준다. 이러한 전환에 집중함으로써 EU는 보다 통합되고 효율적인 혁신 파이프라인을 구축하고자 하며 이러한 접근 방식은 공공 자금으로

지원되는 연구 및 과학적 돌파구가 EU 내에서 실질적인 경제적 및 사회적 이점으로 효과적으로 전환되도록 보장하는 것을 목표로 한다.

3. AI 규제 현황

가. 기본 철학: 기술주권 및 국가경쟁력과 AI 규제의 관계 설정

EU의 AI 규제는 ‘인간 중심적 AI’와 ‘실행할 수 있는 AI 생태계 구축’이라는 두 가지 목표를 동시에 추구한다. EU AI Act(Regulation (EU) 2024/1689)³³⁾는 “인공지능 시스템이 안전하고 기본권과 EU 가치를 존중하며, EU의 혁신을 촉진하고 강화”하는 것을 목표로 명시하고 있는데(Article 1) 이는 단순한 산업 규제를 넘어 EU의 ‘디지털 헌법주의(digital constitutionalism)’ 지향을 구현한 것이다. 특정 부문에 국한되지 않고 모든 경제, 사회 영역에 적용되는 수평적 접근 방식을 채택함으로써, EU 역내 시장에서 AI 시스템의 일관된 작동을 보장하는 동시에 회원국 간 규제 장벽을 최소화하는 것이 핵심 설계 원리이다.

그러나 2025년 들어 이 철학은 중요한 긴장 국면에 접어들었다. Stanford HAI의 『2025 AI Index Report』에 따르면 2024년 미국이 40개의 주목할 만한 AI 모델을 개발하는 동안 EU는 단 3개를 개발하는 데 그쳤다(Maslej et al., 2025. 4). 이러한 현실을 정면으로 다룬 것이 2024년 9월 Mario Draghi가 발표한 『유럽 경쟁력의 미래(The Future of European Competitiveness)』 보고서이다. Draghi는 EU가 “낮은 산업 역동성, 낮은 혁신, 낮은 투자, 낮은 생산성 성장”의 악순환에 빠져 있으며, 미국과 중국에 뒤처진 원인으로 혁신 금융에 이용 가능한 자본의 제한, 기술과 인적자본의 부족, 규제에 의한 단일시장 접근성 제한을 꼽았다(Draghi, 2024). 이를 타개하기 위해 연간 7,500~8,000억 유로(EU GDP의 4~5%)의 추가 투자가 필요하다는 처방을 제시한 이 보고서는, 이후 EU AI 규제

33) European Commission(2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council on Artificial Intelligence (EU AI Act). Official Journal of the European Union. August 1, 2024.

의 방향을 ‘안전 우선’에서 ‘혁신과 경쟁력 강화’로 재조정하는 정치적 동력이 되었다. 2025년 1월 29일 발표된 “EU 경쟁력 나침반(EU Competitiveness Compass)”은 이러한 권고사항을 정책으로 구현하는 로드맵으로, 폰데어라이엔 위원장은 이를 향후 5년간 EU 집행위원회의 “북극성”이라고 설명하였다(European Commission, 2025a). 이로써 EU의 AI 거버넌스는 기본권 보호의 규범적 토대를 유지하면서도 혁신 경쟁력 회복이라는 현실적 압박을 동시에 수용해야 하는 이중적 과제 앞에 서게 되었다.

나. 핵심 입법 및 정책 체계

1) EU AI Act의 위험 기반 규제 체계(European Commission, 2024)

2024년 8월 1일 발효된 EU AI Act은 세계 최초의 포괄적 AI 규제법으로, AI 시스템을 위험 수준에 따라 4개 유형으로 분류하여 차등 규제를 적용한다.

① 허용불가 위험(Unacceptable Risk) — 금지(Article 5)

Article 5에서 규정하는 금지행위들은 2025년 2월 2일부터 적용되었으며, 2025년 8월 2일부터 제재 집행이 가능해졌다. 금지 행위의 목록에는 개인의 행동을 조작하거나 취약계층을 착취하는 시스템, 공공기관의 사회적 점수제, 직장·교육기관에서의 감정 인식, 인터넷이나 CCTV에서의 무작위 안면 이미지 수집, 민감한 속성(인종, 정치적 견해, 종교 등)에 기반한 생체인식 분류, 공공장소에서 실시간 원격 생체인식 등이 포함되어 있다. 위반 시 전 세계 연간 매출액의 7% 또는 3,500만 유로 중 높은 금액이 과징금으로 부과된다. EU 집행위원회는 2025년 2월 4일 금지행위에 관한 가이드라인을 발표하여 집행의 일관성을 지원하고 있다.

② 고위험(High Risk) — 엄격한 준수 의무(Annex III)

Annex III에 열거된 8개 분야의 AI 시스템이 고위험으로 분류되며, 해당 시스템이 자연인에 대한 프로파일링을 수행하는 경우에는 예외 없이 고위험으로 간주된다(Article 6(3)). 생체인식 및 분류, 중요 인프라 관리, 교육 및 직업훈련, 고용

및 근로자 관리, 필수 민간·공공 서비스 접근, 법집행, 이민·망명·국경통제 관리, 사법행정 및 민주적 절차가 Annex III에 열거된 8개 분야에 해당한다. 고위험 AI 시스템을 제공하는 사업자는 위험관리 체계 구축(Article 9), 데이터 거버넌스(Article 10), 기술 문서 유지(Article 11), 투명성 및 정보 제공(Article 13), 인간 감독(Article 14), 정확성·견고성·사이버보안 확보(Article 15), 적합성 평가(Article 43) 등의 의무를 수명주기 전반에 걸쳐 이행해야 한다. 이러한 의무들의 적용 시점은 원칙적으로 2026년 8월 2일이나, 아래에서 살펴볼 Digital Omnibus Package 법률개정안에 따라 조정될 수 있다.

③ 제한적 위험(Limited Risk) - 투명성 의무(Article 52)

챗봇과 같이 인간과 상호작용하는 AI 시스템, AI가 생성한 콘텐츠, 감정 인식이나 생체 분류 시스템 등에 대해서는 이용자가 AI와 상호작용하고 있음을 인지할 수 있도록 하는 투명성 의무가 부과된다. 특히 딥페이크나 AI 생성 콘텐츠에는 명확한 라벨링이 요구된다.

④ 최소 위험(Minimal Risk) - 자발적 행동규범(Article 69)

대부분의 AI 시스템이 이 범주에 속하며, EU AI Act상 별도의 의무나 요건은 없다. 다만 Article 69에 따라 집행위원회, AI 오피스(AI Office), 회원국은 고위험에 해당하지 않는 AI 시스템 제공자에게 위험 관리, 데이터 거버넌스, 인간 감독 등에 관한 자발적 행동규범 준수를 권장하고 있다.

2) 범용 AI(GPAI) 모델 규제 체계(European Commission, 2025b)

GPAI 모델은 “다양한 고유 작업을 유능하게 수행할 수 있는 상당한 일반성을 보이는” 모델을 말한다(Article 3(63)). 2025년 8월 2일부터 GPAI 모델 제공자들은 Chapter V의 규정에 따른 의무를 이행하여야 한다.

의무의 내용은 모델의 위험 수준에 따라 달라진다. 일반 GPAI 모델 제공자는 Article 53에 따라 훈련 방법론, 데이터셋, 성능 지표 등을 담은 기술 문서를 작성하고 유지해야 하며, 다운스트림 제공자에게 상세 정보를 제공하고, EU 저작권

법 준수 정책을 수립하며, 훈련 데이터 출처 요약은 공개해야 한다. 여기에 더하여 시스템적 위험을 갖는 GPAI 모델 제공자는 Article 55에 따라 표준화된 프로토콜을 통한 정기적 모델 평가, 위험 요소에 대한 완화 조치 설계 및 실행, 사이버 보안 보호 체계 구축, 에너지 소모 및 환경적 영향의 측정·보고 의무를 추가로 부담한다. 현재 훈련에 사용된 컴퓨팅 파워가 10^{25} FLOPs를 초과하는 모델이 시스템적 위험 모델로 분류된다.

이러한 의무의 구체적 이행을 지원하기 위해 EU AI 사무국은 2025년 7월 10일 약 1,000명의 이해관계자와 함께 개발된 GPAI 실행규범(Code of Practice) 최종본을 발표하였다. 투명성, 저작권, 안전·보안의 세 챕터로 구성된 이 실행규범은 GPAI 모델 제공자가 AI Act의 의무를 어떻게 구체적으로 이행해야 하는지에 대한 실무적 지침을 담고 있다(European Commission AI Office, 2025a). 2025년 8월 1일 집행위원회와 AI 위원회가 적합성을 확인하였다. 이 실행규범에 서명한 제공자는 이를 AI Act 의무 준수를 입증하는 주요 수단으로 활용할 수 있다. 준수는 자발적이지만, 서명하지 않은 제공자는 다른 방법으로 AI Act 준수를 별도로 입증하여야 한다. 한편 2025년 7월 18일에는 GPAI 모델 의무 범위에 관한 가이드라인이, 2025년 7월 24일에는 훈련 데이터 공개 요약 템플릿이 각각 발표되어 실무적으로 활용할 수 있는 구체적 도구들이 갖추어지고 있다(European Commission AI Office, 2025b; European Commission AI Office, 2025c).

3) 2025년 11월 Digital Omnibus on AI — 혁신 친화적 재조정

2025년 11월 19일 EU 집행위원회는 ‘Digital Omnibus on AI’(EU AI Act 개정안, 이하 ‘디지털 옴니버스 개정안’)를 발표하였다. 이 개정안은 AI Act 이행을 간소화하고 준수 부담을 완화하며, 2026년 8월 2일 AI Act의 전면 시행일자를 조정하는 것을 목적으로 한다(European Commission, 2025b). 디지털 옴니버스 개정안의 주요 내용은 다음과 같다.

첫째, 고위험 AI 시스템 의무가 적용되는 시행일자를 조정하는 메커니즘을 도입하였다. 법정 시행일 대신 관련 조화 표준, 공통 사양, 가이드라인 등 지원 수단의

가용성을 확인하는 집행위원회 결정에 시행일자를 연동하되, Annex III에 따른 고위험 시스템의 최종 시행일은 2027년 12월 2일, Annex I에 따른 고위험 시스템은 2028년 8월 2일로 설정하였다. 현행 AI Act는 관련 표준이나 공통 사양이 마련되지 않더라도 법정 일자가 도래하면 해당 규정이 시행되는 구조인데, 이번 개정안은 가이드라인과 표준이 준비되었음을 집행위원회가 확인한 시점부터 의무가 적용되도록 유연성을 부여한 것이다. 다만 최종 기한 자체는 명시적으로 설정하여 무기한 연기를 방지하였다는 점이 핵심이다. 둘째, 중소기업(SME) 뿐만 아니라 중견기업(SMC)에 대한 규제 준수 부담 완화하여 2029년까지 SME의 규제 부담을 35% 감축하는 것을 목표로 하고 있다. 셋째, AI 사무국을 중앙 집행 허브로 강화하여 대규모 온라인 플랫폼에 통합된 AI 시스템에 대한 감독 권한을 AI 사무국에 부여하였다. 디지털 옴니버스 개정안은 2025년 12월 기준 유럽의회 내 부시장 및 소비자보호 위원회(IMCO) 위원회와 산업·연구·에너지위원회(ITRE)에서 심의 중이다. 최종 채택을 위해서는 유럽의회, 유럽이사회와의 삼자 협상을 거쳐야 한다.

한편, AI 손해배상책임지침(AI Liability Directive)은 이해관계자 간 합의 부족과 혁신 저해 우려를 이유로 2025년 2월 철회되었다. 대신 2024년 12월 8일 발효된 개정 「제조물 책임 지침」(Product Liability Directive)이 AI 시스템으로 인한 손해에 대한 책임 체계를 보완하는 역할을 담당하게 되었다(IIEA, 2025).

다. 혁신 촉진과 리스크 관리의 균형 방식

EU는 규제 체계를 정비하는 한편, 혁신 진흥을 위한 구체적 정책들을 함께 추진하고 있다. 2025년 2월 11일 발표된 'InvestAI 이니셔티브'는 2,000억 유로 규모의 AI 투자 동원 계획으로, 200억 유로 신규 펀드를 조성하여 연구, 혁신 및 AI 상업화를 위해 지역 허브 역할을 수행할 AI 기가팩토리(AI Gigafactories) 구축을 지원한다(European Commission, 2025c). 2025년 4월 9일 발표된 'AI 대륙 실행 계획(AI Continent Action Plan)'은 대규모 AI 컴퓨팅 인프라 구축, 고품질 데이터 접근성 증대, 전략 부문에서의 AI 활용 촉진, AI 기술 및 인재 강

화, EU AI Act 이행 간소화의 5대 핵심 분야를 제시하며, 73억 유로 규모의 Horizon Europe 프로그램과 연계하여 추진한다(European Commission, 2025d). 2025년 10월 채택된 ‘Apply AI Strategy’는 AI를 11개 핵심 부문에 적용하는 전략으로, 조직이 전략적·정책적 결정을 내릴 때 AI를 우선적으로 고려하는 ‘AI-first 접근’을 촉진하고 있다(European Commission, 2025e).

그러나 혁신 촉진과 리스크 관리 사이의 긴장은 여전히 해소되지 않고 있다. DIGITALEUROPE 등 주요 기술 기업들과 유럽 산업계는 복잡한 규제 요건의 간소화를 지속적으로 요구하고 있으며, 폴란드는 EU 순환의장국 기간 중 GPAI 모델 실행규범이 제때 완성되지 않을 경우 관련 규칙 적용을 연기하자고 제안하기도 하였다(Bird & Bird, 2025). 반면 시민사회단체들은 규제 완화가 기본권 보호를 약화시킬 수 있다고 경계하고 있으며, 이미 새로운 규정 준수에 투자한 기업들 사이에서도 규제 방향의 잦은 변화가 오히려 불확실성을 키운다는 우려가 제기되고 있다.

라. 거버넌스 체계와 집행 역량

1) 다층적 감독 구조

EU AI Act의 거버넌스는 EU 차원과 국가 차원이 역할을 나누는 다층적 구조로 설계되어 있다. EU 차원에서는 2025년 8월 2일 공식 운영을 시작한 AI 사무국이 GPAI 모델에 대한 감독을 전담하며, 회원국 대표들로 구성된 EU AI 위원회(European AI Board)는 AI Act의 일관된 적용 보장과 정책 지침 개발을 맡고 있다(Articles 64-67). 과학자문단(Scientific Panel)은 GPAI 모델의 시스템적 위험에 대한 독립적인 기술 자문을 제공한다(Articles 68-70).

국가 차원에서는 각 회원국이 2025년 8월 2일까지 국가 감독기관을 지정하도록 하였다. 그러나 집행 방식은 국가별로 상당한 차이를 보이고 있다. 스페인은 전담 AI 감독청(Spanish AI Supervisory Agency)을 설립하는 중앙집중형 접근을 택한 반면, 다른 많은 국가들은 기존 부문별 규제기관들에 권한을 분산하는 방식을 검토하고 있다. 이러한 접근법의 다양성은 EU 전체 차원에서 규제 일관성을

확보하는 데 쉽지 않은 과제를 남기고 있다(Wilson Sonsini, 2025).

2) 집행 체계의 현실적 한계

EU AI Act의 집행은 아직 본격적인 궤도에 오르지 못하고 있다. 금지행위(Article 5)에 대한 제재는 2025년 8월 2일부터 집행이 가능해졌으나, 2025년 12월 기준으로 금지행위에 대한 공식적인 집행 조치가 발표된 사례는 없다(FPF, 2025). 고위험 AI 시스템에 대한 규정의 시행은 2026년 8월 2일 이후에야 가능하며, 디지털 옴니버스 개정안이 채택될 경우 2027년 12월 이후로 더 늦춰질 수 있다. AI 기술의 급속한 발전과 복잡성은 규제당국의 감독 역량을 지속적으로 압박하고 있어, 전문 인력과 자원을 어떻게 확보하느냐가 EU AI 거버넌스의 실효성을 가르는 핵심 변수로 부상하고 있다.

3) 보완적 법률 체계와의 시너지

EU AI Act는 EU의 광범위한 디지털 규제 생태계의 일부로서 기능한다. 「일반 데이터보호규정(GDPR)」, 「데이터법(Data Act)」, 「데이터 거버넌스법(Data Governance Act)」 등이 AI 시스템의 데이터 처리 활동을 규율하고, 「디지털서비스법(DSA)」과 「디지털시장법(DMA)」은 AI를 활용하는 대형 플랫폼 기업들의 의무를 규정한다. NIS2 지침과 「사이버복원력법(Cyber Resilience Act)」은 AI 시스템의 보안에 관한 규제 사항을 다루고 있다. 이처럼 EU는 AI 거버넌스를 단일 법률로 완결하려 하지 않고, 데이터·플랫폼·보안 규제가 서로 맞물려 작동하는 디지털 규제 생태계 전반에서 구현한다는 전략을 취하고 있다. 이는 규제의 빈틈을 줄이는 동시에 규제 중복과 충돌이라는 새로운 과제를 낳기도 한다.

마. 국제적 영향력 확산 전략

EU AI Act는 역외 적용(Article 2) 조항을 통해 EU 경계를 넘는 광범위한 영향력을 행사하고 있다. EU 시장에 AI 시스템을 출시하거나 EU 내에서 사용하는 모든 기업에 적용되기 때문에, 미국의 OpenAI, 중국의 ByteDance 등 주요 글로벌 AI 기업들은 사실상 EU 기준에 맞춰 제품을 개발할 수밖에 없는 상황이다. 이

는 Anu Bradford가 제시한 ‘브뤼셀 효과(Brussels Effect)’의 전형적 사례로, 캐나다, 영국, 브라질 등 주요국들이 EU의 위험 기반 접근법을 참조하여 자국 AI 규제를 설계하고 있다(artificialintelligenceact.eu, 2025).

EU는 G7 AI 프로세스, OECD AI 정책 관측소, Global Partnership on AI 등 주요 국제기구에서 AI 거버넌스 논의를 주도하는 한편, 개발도상국을 대상으로 한 AI 거버넌스 역량 지원 프로그램을 통해 EU AI Act를 글로벌 표준으로 확산시키는 전략을 추진하고 있다. 국제 표준화 측면에서는 CEN, CENELEC, ETSI 등 유럽 표준화 기구들이 ISO/IEC 42001 등 국제 표준과의 조화를 추구하며 글로벌 상호운용성 확보에 적극 참여하고 있다. 이러한 다차원적 접근을 통해 EU는 AI 거버넌스 분야에서 규범 형성자로서의 위상을 공고히 하려 하고 있다.

바. 평가 및 시사점

EU AI Act는 세계 최초의 포괄적 수평적 AI 규제 체계로서 여러 측면에서 의미 있는 성과를 거두고 있다. 위험 수준에 따른 차등 규제를 통해 혁신과 안전의 균형을 제도적으로 구현하였고, 다층적 거버넌스 체계를 구축하여 EU 차원의 일관된 집행 기반을 마련하였다. GPAI 실행규범과 같이 법적 구속력과 자율 준수를 결합하는 유연한 준수 메커니즘을 개발한 것도 주목할 만하며, 브뤼셀 효과를 통해 글로벌 AI 규제 논의를 선도하고 있다는 점에서 규범 형성자로서의 위상을 확보하고 있다.

그러나 한계도 분명하다. 규제의 복잡성으로 인한 높은 준수 비용, AI 혁신 경쟁에서 미국, 중국 대비 열세, 2025년 12월 기준 금지 행위에 대한 집행 조치가 전무한 집행 공백, 회원국 간 규제 일관성 확보의 어려움, 디지털 옴니버스 개정 안으로 예시되는 규제 후퇴 압력 등이 지속적으로 제기되고 있다. 특히 규제 설계 단계에서 집행 인프라 구축을 병행하지 못한 점은 법률을 먼저 제정하고 하위 규범과 집행 체계를 뒤따라 정비하는 방식이 갖는 구조적 위험을 보여준다.

한국에 대한 시사점은 크게 네 가지로 정리된다. 첫째, 위험 수준에 따라 규제 강도를 달리하는 차등 규제 방식은 AI 기본법의 고영향·고성능·생성형 AI 구분과

유사한 설계 원리를 공유하고 있어, EU의 구체적 기준 설정 방식을 참조할 수 있다. 둘째, GPAI 실행규범과 같이 법적 의무와 자율 준수를 결합하는 메커니즘은 AI 기본법이 강행 규정만으로는 채울 수 없는 실무적 이행의 공백을 메우는 데 유효한 방법이 될 수 있다. 셋째, InvestAI와 AI 대륙 실행 계획처럼 규제와 진흥 정책을 함께 패키지로 설계하는 방식은 AI 기본법 시행과 연계된 혁신 지원 정책을 구성하는 데 직접적인 참고가 된다. 넷째, 디지털 옴니버스 개정안을 통해 기존 규제를 유연하게 재조정하는 과정은 AI 기본법의 하위 규범과 고영향 AI 분류 기준을 설계할 때 처음부터 조정 메커니즘을 내재화하는 것이 왜 중요한지를 잘 보여준다.

제 4 절 한 국

1. AI 산업생태계 구조 및 주요 플레이어

한국은 AI 메모리 등 일부 분야에서 세계 최고 수준의 경쟁력을 보유하고 있으며 하드웨어와 소프트웨어 전반에 걸쳐 기술력을 갖춘 기업을 가지고 있다는 점에서 하드웨어-소프트웨어 간 시너지 창출에 유리한 구조를 가지고 있다. 생태계 핵심 플레이어로는 하드웨어 분야 삼성전자(AI 반도체, 온디바이스 AI), SK하이닉스(HBM 등 AI 메모리), AI 모델 분야 네이버(하이퍼클로바X), LG(LG 엑사원), 업스테이지(솔라), KT(믿음), SKT(에이닷 엑스) 등이 있으며 이들은 국내 AI 기술 생태계의 핵심 축으로서 AI 인프라와 모델 경쟁력 확보에 중요한 역할을 수행하고 있다. 또한 리벨리온(사피온 통합법인), 퓨리오사AI, 딥엑스와 같은 AI 반도체(NPU) 스타트업들이 원천기술을 바탕으로 글로벌 시장에서 두각을 나타내며, 대기업 중심의 생태계에 새로운 기술적 활력을 불어넣고 있다.

그러나 국내 AI 생태계는 대기업이 기술 개발, 인재 확보, 스타트업 투자 등 생태계 전반을 주도하는 구조로 빠른 기술 도입과 대규모 프로젝트 추진에는 유리하지만 대기업 의존도가 높고 스타트업 중심의 중간층이 상대적으로 약하다는 한계가 존재한다. 비록 AI 반도체 분야를 필두로 딥테크 스타트업들이 체급을 키우며 ‘허리’ 역할을 강화하고 있으나, 글로벌 선도국과의 원천기술 격차를 해소하고 지속가능한 자생적 생태계를 구축하는 것은 여전히 중요한 당면 과제로 지적되고 있다.

가. 하드웨어

한국은 AI 연산에 필수적인 고대역폭 메모리(HBM) 시장에서 압도적인 경쟁력으로 가지고 글로벌 AI 인프라의 중심축 역할을 수행하고 있다. 카운터포인트리서치에 따르면 2025년 2분기 HBM 출하량을 기준으로한 글로벌 시장 점유율에서 SK하이닉스 62%, 마이크론 21%, 삼성전자 삼성전자 17%로 SK하이닉스와

삼성전자가 전 세계 HBM 시장 점유율의 약 80%를 차지한 것으로 나타났다(서울와이어, 2025. 10. 8). HBM은 2013년 SK하이닉스가 최초 개발하며 한국이 주도하고 있는 기술로 AI 가속기의 연산 성능을 좌우하는 핵심 부품이다. 글로벌 빅테크의 AI 데이터센터 투자 확대와 맞물려 그 전략적 중요성이 더욱 커지고 있으며, HBM3E를 넘어, 6세대 HBM4에서는 SK하이닉스와 삼성전자의 시장지배력이 더욱 공고해질 것으로 전망되고 있다(서울와이어, 2025. 10. 8).

한편 반도체 가속기 부문에서는 엔비디아 GPU에 절대적으로 의존하고 있으나, 리벨리온, 퓨리오사AI, 딥엑스, 에이직랜드 등 팹리스 스타트업들이 고성능 NPU를 양산하며 기술 자립 기반을 확대하고 있다. 이들 기업은 데이터센터용 AI 가속기뿐 아니라 온디바이스 AI와 엣지 AI 시장을 겨냥한 전용 칩 개발을 추진하며 글로벌 AI 반도체 공급망 내에서 새로운 경쟁 축을 형성하고 있다.

〈표 3-11〉 국내 주요 AI 반도체 스타트업

기업	전문분야	기업가치 (추산)	매출액 및 실적	주요 제품 포트폴리오
리벨리온 (사피온 합병)	데이터 센터용 AI 추론 NPU	약 2조 원	<ul style="list-style-type: none"> • 2024년 합병 후 156 억원 • 국내외 데이터센터 공급 확대 	<ul style="list-style-type: none"> • ATOM: 국내 최초 데이터 센터용 NPU, KT클라우드 NPU 서비스에 적용 • REBEL: 초거대 LLM과 멀티모달 모델 가속 특화칩
퓨리오사AI	데이터 센터용 AI 추론 NPU	약 1조 5000억 원	<ul style="list-style-type: none"> • 시리즈 D 투자 유치 중 (4000억 규모) • 글로벌 양산 계약 추진 	<ul style="list-style-type: none"> • WARBOY: 1세대 비전특화 NPU (양산중) • RNGD: 2세대 고성능LLM 가속칩, 2025년 7월 LG연 구원 역사원에 공급
딥엑스 (DEEPIX)	온디바이스/ 엣지용 NPU	약 1조 3000억 원	<ul style="list-style-type: none"> • 2024년 매출 3억 원 • CES 2026 혁신상 2개 부문(컴퓨팅 하드웨어와 임베디드 기술) 수상 	<ul style="list-style-type: none"> • DX-M1: 소형가전/드론용, 바이두의 패들패들과 파트 너십 • DX-M2: 차세대2 나노기반 피지컬 AI 칩, 2027년 양산 계획

기업	전문분야	기업가치 (추산)	매출액 및 실적	주요 제품 포트폴리오
모빌린트	온디바이스/ 엣지용 NPU	누적 300억 원 투자 유치	• 2025년 목표 매출 50~100억 원	<ul style="list-style-type: none"> • ARIES: 온프레미스 AI용, 싱글칩 완료, 2025년 하반기부터 본격적 대량양산 계획 • REGULUS: 온디바이스 AI용, PoC 진행 중
에이직랜드 (ASICLAND)	글로벌 ASIC(주문형 반도체) 디자인 솔루션	시가총액 약 3,000억 원대	• 2024년 매출 940억 원 기록	<ul style="list-style-type: none"> • 고객 맞춤형 SoC 설계, 칩 설계 및 검증, 파운더리 생산 및 패키징 지원 • TSMC 등 다양한 IP 업체와 협력

자료: 나무위키 사이트, THEELEC(2025. 3. 4), Byline Network(2025. 6. 4.), 매일경제(2025. 6. 6.), the bell(2025. 9. 17.), 아시아경제(2025. 9. 4.), Platum(2025. 11. 6.), 에이직랜드 사이트

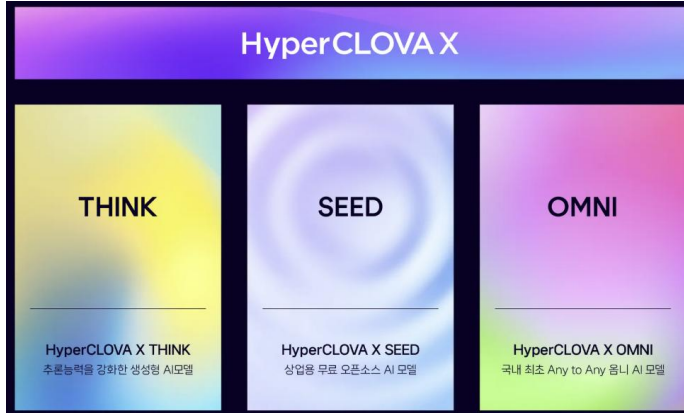
나. AI 모델

중국 오픈 소스 모델이 추론 성능 측면에서 빅테크와 견줄 수 있는 수준에 도달하고 파운데이션 모델의 훈련 레시피가 공개됨에 따라 국내 주요 IT 기업들은 이러한 오픈 소스 레시피를 채용하고 발전시켜 자체적인 파운데이션 모델을 개발하는 사례가 늘어나고 있으며 의미있는 진전을 보이고 있다. 한국의 주요 AI 모델들을 살펴보면 다음과 같다.

1) 네이버

네이버의 AI 모델은 HyperCLOVA X를 중심으로 추론, 멀티모달, 경량 모델로 확장되고 있다. HyperCLOVA X는 한국어와 국내 데이터를 기반으로 학습된 초거대 언어모델로 네이버 AI 서비스의 기반 모델 역할을 하며, 이를 바탕으로 복잡한 질의를 단계적으로 분석하고, 텍스트와 이미지를 함께 이해해 추론 결과를 생성하는 추론 특화 모델 HyperCLOVA X Think, 텍스트·이미지·오디오 등 다양한 모달리티 데이터를 동시에 이해하고 처리하는 멀티모달 모델 HyperCLOVA X Omni, 그리고 개발자 생태계 확산을 위해 공개된 경량 모델 계열인 HyperCLOVA X SEED 등이 이러한 확장의 사례이다.

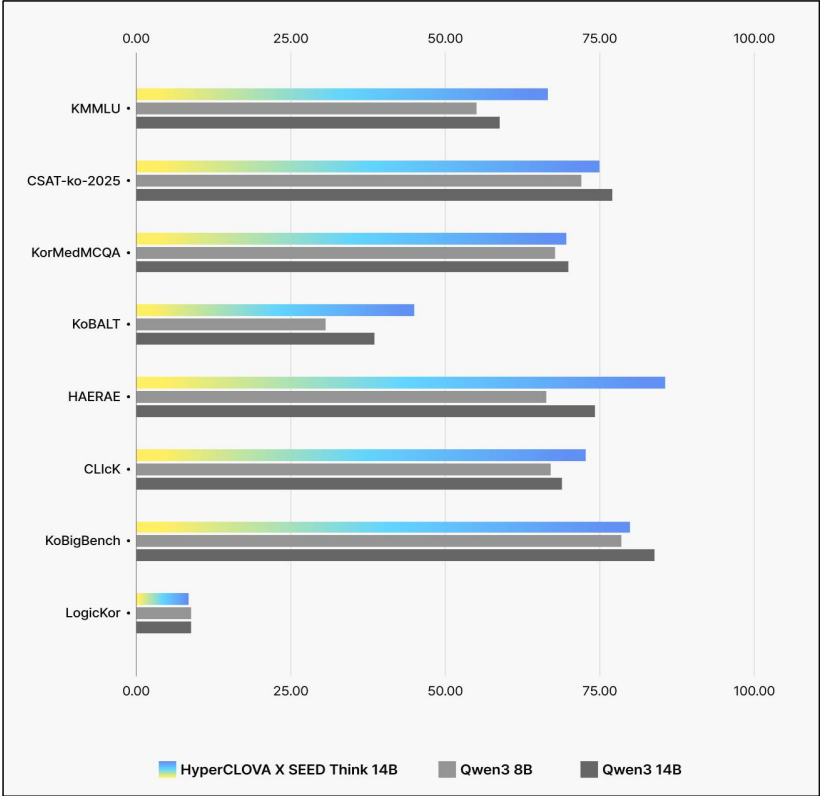
[그림 3-4] HyperCLOVA 모델 라인업



자료: <https://clova.ai/hyperclova>

이러한 모델 확장은 대형 기초 모델을 기반으로 추론 및 에이전트 역량을 강화하고 다양한 서비스로 확장하려는 네이버의 AI 플랫폼 전략을 반영한다. 이러한 흐름 속에서 2025년 7월 공개된 HyperCLOVA X SEED Think-14B는 복잡한 문제 해결을 위한 추론 능력과 도구 호출 기능을 지원하는 추론 특화 모델로, 에이전트 기반 작업 수행을 염두에 두고 설계되었다. 이 모델은 약 140억 파라미터 규모의 중형 모델로 유사한 규모의 Qwen3-14B 대비 낮은 훈련 비용으로 학습되었으며 한국어 벤치마크에서 경쟁력 있는 성능을 보였다.

[그림 3-5] HyperCLOVA X SEED Think-14B 성능 테스트 결과



자료: CLOVA(2025. 7. 22.)

2) LG AI연구원³⁴⁾

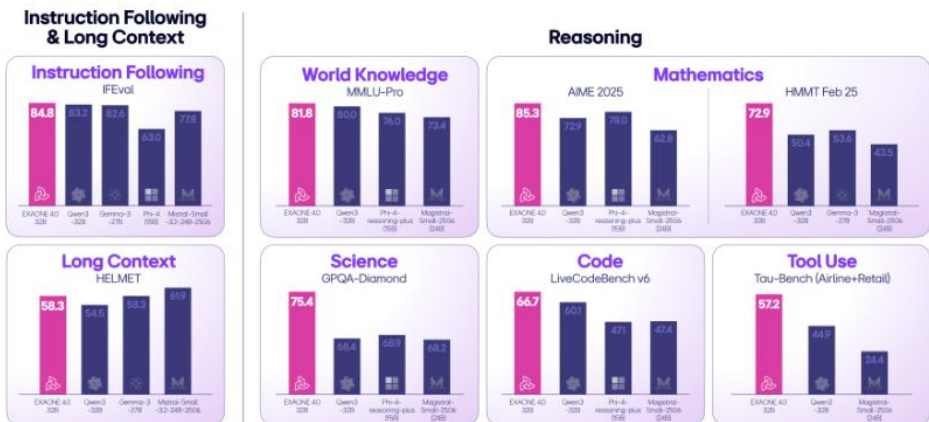
LG AI연구원의 AI 모델은 ‘EXAONE(엑사원)’ 계열 파운데이션 모델을 중심으로 다양한 기능과 산업 적용 모델로 확장되는 구조이다. LG는 2021년 국내 최초 독자 개발한 초거대 멀티모달 모델 EXAONE 1.0을 공개한 이후 지속적으로 성능을 개선하며 모델을 발전시켜 왔는데 2024년에는 국내 최초 연구용 오픈소스 모델 EXAONE 3.0과 스탠퍼드 AI 보고서에 소개된 EXAONE 3.5를, 2025년에는 국내 최초 추론기반 모델 EXAONE Deep과 EXAONE 4.0을 발표했다.

34) <https://www.lgresearch.ai/news/view?seq=577&utm>

EXAONE-4.0는 추론능력과 자연어 이해 역량을 결합한 하이브리드 모델로 32B 규모 대형모델과 온디바이스용 1.2B 경량모델로 설계되었다. 특히 EXAONE-4.0-32B는 국내 오픈 소스 모델로는 유일하게 ArtificialAnalysis.AI의 글로벌 벤치마크에서 상위권에 이름을 올리며 국제적인 경쟁력을 입증하였다. 그동안 국내 AI 모델들이 한국어 성능에서는 경쟁력을 보였으나 영어 성능에서는 다소 부족한 것으로 평가되었는데 LG AI 연구원의 EXAONE 4.0 32B는 MMLU-Pro(지식·추론) 81.8점, LiveCodeBench v6(코딩) 66.7점, GPQA-Diamond(과학) 75.4점, AIME 2025(수학) 85.3점을 기록하며 글로벌 빅테크의 플래그십 모델들에 상당 부분 근접한 성능을 보여주었다³⁵⁾.

또한 Work Agent 서비스인 ChatEXAONE, 의료 AI 모델(조직병리 이미지 분석에 특화된 EXAONE Path 2.0) 등 산업에 적용가능한 서비스, 분야별 특화 모델 등도 개발하며 AI 생태계를 확장하고 있다. 이러한 흐름은 대형 파운데이션 모델을 기반으로 추론·멀티모달·산업 특화 모델을 단계적으로 발전시키고, 이를 다양한 산업 현장에 적용하려는 LG의 산업 중심 AI 플랫폼 전략을 반영한다.

[그림 3-6] EXAONE 4.0-32B 성능그래프



자료: LGRsearch 블로그

35) <https://www.lgresearch.ai/exaone/>

3) SKT

SKT는 국내 데이터에 최적화된 LLM을 자체 개발하며 ‘소버린 AI(주권형 AI)’ 전략을 추진해 왔다. SKT의 핵심 모델 계열은 A.X 시리즈 LLM으로 A.X 3 시리즈는 자체 개발 모델을 중심으로 한 “주권형 AI 모델”에 해당한다. 2025년 7월에는 모델 설계와 학습 전 과정을 자체적으로 수행한 340억 파라미터 규모의 A.X 3.1 모델을 공개하였는데 이 모델은 한국어 대화 성능에 집중했던 A.X 3.0과는 달리 코딩·수학 능력 등 추론 성능이 강화된 모델로, 고급 추론 작업과 복잡한 문제 해결을 지원하도록 설계되었다(SK telecom Newsroom, 2025. 7. 24).

A.X 4 시리즈는 오픈소스 모델을 기반으로 한국어 데이터를 추가 학습해 기업 환경에 최적화한 모델이다. 공개된 A.X 4.0(72B 표준 모델과 7B 경량 모델) 중 72B는 국내 오픈 소스 LLM으로서는 가장 큰 규모이며 알리바바의 Qwen2.5 모델을 시작점으로 하여 지속 학습을 통해 대용량 한국어 말뭉치를 학습시켜 한국어 벤치마크에서 우수한 성능을 달성하였다(SK telecom Newsroom, 2025. 7. 3).

2025년 12월에는 5000억 개 규모의 초대형 모델 A.X K1을 독자 AI 파운데이션 모델 프로젝트 1차 발표회에서 공개할 것이라 언급하며, 이를 소형 모델을 학습시키는 ‘교사 모델’ 역할의 기반 모델로 활용하는 전략도 제시했다(뉴시스, 2025. 12. 28). SKT는 이러한 모델들을 기반으로 통신 서비스(에이닷 등)와 기업용 AI, 에이전트 플랫폼으로 확장하는 AI 플랫폼 전략을 추진하고 있다.

4) KT

KT는 자체 파운데이션 모델 ‘Mi:dm’ 계열 LLM을 중심으로 기업·공공 환경에 적용가능한 한국어 특화 생성형 AI 플랫폼을 구축하고 있다. KT는 2025년 7월 기존 모델을 고도화하여 Mi:dm 2.0 라인업을 발표하였다. 이 중 Mi:dm 2.0 Pro는 기업용 차상위 모델로 가장 큰 파라미터 규모를 가지며 복잡한 추론, 지식기반 판단, 장문 분석에 최적화된 모델로 알려졌다. 115억 파라미터를 가진 Mi:dm 2.0 Base는 누구나 AI를 개발할 수 있는 생태계 확장을 위해 오픈소스로 공개되었는데 이는 국내에서 11B(110억 개 파라미터)이상 LLM이 상업적 오픈소스로

제공되는 첫 사례이다. 또한, KT는 이를 Distillation한 2.3B(23억 개 파라미터)의 경량화 모델인 Mi:dm 2.0 Mini도 함께 공개하였다. Mi:dm 모델은 한국어 AI 역량 평가지표인 Ko-Sovereign, 한국 관련 전문지식 이해도를 측정하는 KMMLU, 한국어 언어모델 평가지표인 HAERAE 등에서 국내외 주요 오픈소스 모델보다 우수한 성능을 기록한 것으로 알려졌으며, 국산 AI 반도체 기업 리벨리온과 협력해 국산 AI 반도체에서 최적화를 진행했으며, 프렌들리AI와 함께 사용자가 무료로 체험할 수 있는 환경도 제공한다(ZDNET Korea, 2025. 7. 3).

5) 카카오

카카오는 자체 LLM 개발과 외부 AI 협력을 병행하는 하이브리드 전략을 취하고 있다. 우선, 카카오는 자체 AI 모델 개발을 지속하고 있는데 한국어 특화 대형 언어모델 KoGPT를 공개하며 AI 모델 개발을 시작했고, 이후 새로운 LLM 계열 Kanana를 개발해 성능과 효율을 개선했다. Kanana는 한국어와 영어를 처리하는 LLM 시리즈로, 문맥 이해와 코드·수학 문제 해결 능력 등을 강화한 모델이며 텍스트뿐 아니라 이미지까지 처리하는 멀티모달 모델도 공개되었다.

[그림 3-7] 카카오의 멀티모달 언어모델 라인업



자료: kakao tech 사이트

자체 모델 개발과 함께 외부 AI 모델을 결합하는 전략도 추진하고 있다. 카카오 브레인 조직을 본사 AI 조직으로 통합하며 AI 전략을 재정비했고, 2025년에는 OpenAI와 전략적 협력을 맺어 카카오톡 등 주요 서비스에 생성형 AI 기능을 도입하는 계획을 발표했다(kakao, 2025. 2. 4). 이를 통해 카카오는 자체 모델을 기반으로 하되 글로벌 AI 기술을 결합해 서비스 경쟁력을 높이는 방향을 추진하고 있다.

〈표 3-12〉 국내 LLM 추론 성능 비교

모델 이름	모델 크기	GPQA(영어)	CLick(한국어)
DeepSeek-R1-0528	37B	81	81.5
Qwen3-14B	14B	49.8	68.8
Mi:dm 2.0	11.5B	33.5	X
HyperCLOVAX-SEED-Think-14B	14B	X	72.8
A.X.4.0	72B	67.6	79.4
kanana-1.5-8B	8B	X	X
EXAONE-4.0-32B	32B	75.4	X

주: 각 모델 공개시 발표한 추론형 벤치마크 결과를 인용함

현재 한국어 특화 성능에서의 강점을 유지하면서도 영어를 비롯한 다국어 성능을 향상시키는 것이 국내 AI 생태계의 글로벌 경쟁력 확보를 위한 핵심 과제로 부각되고 있다. 빅테크와 중국의 오픈 소스 주도 기업들은 초거대 규모의 파운데이션 모델을 보유하고 있으며, 이를 32B 규모의 모델 완성도를 높이는 데 활용하고 있다. 국내에서도 글로벌 추세에 발맞추어 32B 규모 또는 그 이상의 모델에서 격차를 좁히는 방향으로 투자와 연구개발이 집중될 필요가 있다.

다. 물리적 AI

삼성전자는 2024년 12월 레인보우로보틱스의 최대주주가 되면서 물리적 AI 분야의 본격 진출을 알렸다. 레인보우로보틱스는 협동로봇, 양팔로봇, 자율이동

로봇 등 다양한 로봇 제품군을 보유한 기업으로, 삼성전자는 이를 통해 제조 및 물류 자동화에 로봇을 투입할 계획이다. 또한 2025년 8월에는 INNOEX Lab을 신설하여 물리적 AI 기반 제조 자동화를 본격적으로 추진하고 있다. 삼성전자의 반도체, 디스플레이, 가전 제조 역량과 레인보우로보틱스의 로봇 기술이 결합되면 강력한 시너지가 창출될 것으로 기대된다.

현대자동차그룹은 2021년 Boston Dynamics 인수를 통해 글로벌 로보틱스 분야에서 독보적인 위치를 확보하였다. 현대차그룹의 목표는 완전한 로보틱스 가치 사슬을 구축하는 것으로, 로봇 부품 제조에서부터 스마트 물류 및 공장 자동화 솔루션에 이르는 종합 생태계를 만들어가고 있다. 자동차 제조를 통해 축적된 대규모 생산 역량과 품질 관리 노하우는 로봇의 대량 생산과 신뢰성 확보에 중요한 자산이 되고 있으며, 자율주행 기술과의 시너지도 기대된다.

국내 빅테크 외에도 마음 AI 등 국내 유망 AI 기업들도 물리적 AI 관련 다양한 기술 스택에서 활동 중이며 2025년 10월 1일에는 마음AI 주도로 국내 물리적 AI 산업발전을 이끌 다양한 기업들이 참여하는 한국피지컬AI협회가 출범하였다.

2. AI 정책 및 투자 현황

우리 정부는 AI를 국가 성장의 핵심 동력으로 활용한다는 강력한 의지를 바탕으로 'AI G3 강국'을 목표로 AI 산업경쟁력 확보 및 AI 대전환을 위한 전방위적 정책을 추진 중이다. 특히 이러한 정책 기조는 2025년 9월에 발표한 국정과제에 잘 담겨있는데 국정과제³⁶⁾에서 제시된 AI 관련 주요 정책 목표는 크게 기술주권 확보, 기술주도 성장, 모두의 AI로 구분되며 각 정책 목표별 주요 주요 과제를 정리해보면 다음과 같다.

36) <https://www.korea.kr/govVision/>

〈표 3-13〉 국정목표별 정책과제

목표	내용
기술 주권 확보	① AI 3대 강국 도약을 위한 AI 고속도로 구축 <ul style="list-style-type: none"> • (컴퓨팅 인프라) GPU 확보 및 민간 AIDC 생태계 육성 • (네트워크) 6G 상용화와 실시간·초정밀 AI 서비스 위한 지능형 기지국 확산 • (데이터) 공공·민간 AI 학습용데이터 통합플랫폼/데이터 스페이스/통계 메타 데이터 구축 • (컨트롤타워) 실질적 컨트롤 타워로서 AI 국가위원회 역할 강화 ② 독자 AI 모델 개발 지원 <ul style="list-style-type: none"> • 글로벌 수준의 독자 AI모델 개발을 위해 AI 정예팀 GPU·데이터·인재 집중 지원 ③ AI 반도체 산업 생태계 확립 <ul style="list-style-type: none"> • NPU·PIM R&D 지원, 국산 NPU+AI 서비스 패키지 테스트베드 확대, 사업화 적시 지원
기술 주도 성장	① 핵심·융합인재 양성 <ul style="list-style-type: none"> • AI 중심대학, 과기원 AI 단과대학 설립, 대학·대학원 간 패스트트랙 운영 및 국내외 우수인재 확보 • 다양한 산업수요 기반 AX 인재양성을 위해 이노베이션 아카데미 확대, AX 대학원 설립 ② 미래 AI 판도를 흔들 차세대 AI 원천기술 확보 <ul style="list-style-type: none"> • 혁신 AI 연구에 도전할 수 있는 국가 AI 연구소 육성 ③ AI 반도체 산업 생태계 확립 <ul style="list-style-type: none"> • NPU·PIM R&D 지원, 국산 NPU+AI 서비스 패키지 테스트베드 확대, 사업화 적시 지원 ④ 피지컬 AI 적용 가속화 <ul style="list-style-type: none"> • 핵심기술 확보 및 공장·로봇·자율주행·자율운행선박 등 적용 가속화 ⑤ 벤처 투자시장 확대 및 넥스트 유니콘 프로젝트 <ul style="list-style-type: none"> • AI혁신펀드, 성장사다리, AI 창업패키지, 해외진출 등 AI 스타트업 창업·스케일업 지원 • 넥스트 유니콘 프로젝트 추진
모두의 AI	① 모두의 AI 기반 구축 <ul style="list-style-type: none"> • AI 파운데이션 모델 기반 국민 AI 접근성 증진 및 공공·경제·사회 AX 도모 • 누구나 AI 서비스를 개발·체험할 수 있도록 AI 캠페인·배움터 조성 ② 산업AI 전환 촉진 <ul style="list-style-type: none"> • 제조기업의 AI 팩토리 전환, 중소기업 대상 AI 바우처 지원, 서비스업의 AI 전환 지원 ③ 지역AI 전환 촉진 <ul style="list-style-type: none"> • 지역특화산업의 AI 전환거점인 광역별 AI혁신거점을 구축, 첨단 산단을 AI로 전환하는 AX 실증산단 조성

목표	내용
모두의 AI	④ 공공 AI / K-AI 시티 / AI 기본사회 실현 <ul style="list-style-type: none"> • AI 정부실현을 위한 30대 핵심과제, 사회시스템 전반의 개혁을 주도할 AI 프로젝트 추진 • 범정부 AI 공통기반 구현, 전 공무원 AI 역량 강화 • 공공 AI 시장 창출, 공공데이터 개방, 공공 AI 활용에 대한 안전·신뢰성 확보 • 거주하며 실증하는 AI 특화 시범도시 조성·확산 • 다수 국가가 참여하는 글로벌 AI 컨소시엄, 글로벌 AI 공동기금 및 AI 특화지구 조성 ⑤ 안전·신뢰기반 확보 <ul style="list-style-type: none"> • AI 오남용 대응 핵심 기술 개발, 디지털 보안·안전 확보, 개인정보보호 체계 확립

자료: 정부의 국정과제를 저자가 분류하여 정리

관련하여 AI 예산도 대폭 확대되었는데 특히 R&D, 대규모 인프라 구축, 산업 및 인재 육성에 투자가 집중되고 있다. 우선 우리 정부의 AI 관련 투자는 연구개발(R&D) 예산 증액을 통해 가장 두드러지게 나타나고 있다. 과학기술정보통신부(2025. 8. 22)에서 발표한 자료에 따르면 2026년도 국가 R&D 예산은 전년 대비 약 19.3% 증가한 35.3조 원으로 역대 최대 규모이며 이 중 AI 분야 예산은 전년 대비 106.1% 증가한 총 2.3조 원이다. R&D 예산은 기술주도 성장, 모두의 성장을 양대축으로 수립되었으며 중점 투자분야는 범용인공지능(AGI), 경량·저전력 AI, 피지컬 AI 원천기술 및 파운데이션 모델 확보 확보와 실증 등으로 AI 생태계 전반에 걸친 독자적 역량 강화에 집중될 예정이다.

다음으로 AI 인프라 구축 및 산업 활용 촉진을 위한 대규모 투자 프로젝트 추진이다. 정부는 대형·중소 데이터센터 간 고성능 네트워크 연동 기술개발로 'AI 고속도로'를 실현하고 GPU 자원의 집적 및 공동활용 체계를 구축하여 급증하는 고성능 GPU 기반 연구수요에 대응한다는 계획이다. 이를 위해 'AI 컴퓨팅자원활용 기반강화' 사업에 2조 1,087억 원을 편성(GPU 서버 확충에 1조 3,998억 원), 통합운영환경 구축에 6,832억 원)했으며 GPU는 2025년 1.3만 장(B200, H200 등), 2026년 B200 1.5만 장(서버 1,875대)를 구입하여 총 2.8만장을 확보하게 되며 이는 네이버클라우드, NHN, 카카오 등의 CSP 데이터센터에 설치되어 산업

계, 학계, 연구기관 등에 제공될 예정이다(IT DAILY, 2025.11.4). 또한 NPU, PIM 등 AI 반도체 기반 클라우드 핵심기술의 국산화를 지원할 예정이다.

산업 AI에 대한 투자는 산업통상자원부가 주도하며 2025년 5월 28일 보도자료에 따르면 산업통상자원부는 2025년 445개 과제에 4,787억 원을 투자하여 산업 전반의 AI 전환(AI)을 추진한다는 계획이다. 주요 투자 분야는 AI 팩토리(44개 과제, 627억원), 인공지능 반도체(20개 과제, 216억 원), 자율주행차(82개 과제, 1,206억원), 첨단바이오(80개 과제, 682억 원), 지능형로봇(31개 과제, 296억 원), 디스플레이(14개 과제, 138 억 원), 핵심소재(17개 과제, 277억 원), 에너지 신산업(10개 과제, 74억 원) 등이며 투자 방식은 첫째, 다수 업종·기업이 공통으로 활용가능한 산업 AI모델 개발, 둘째, 바이오, 이차전지 등 개별업종에 특화된 산업 AI 모델 개발, 적용, 확산, 셋째, 산업 AI 인프라 조성이다.

그 밖에도 정부는 인공지능을 포함한 첨단 전략산업 및 관련 기업 육성을 위해 국민성장펀드를 조성하여 향후 5년간 150조 원 이상을 지원할 예정이다. 국민성장펀드는 특히 산업 파급효과가 큰 범부처 메가 프로젝트에 중점 투자되며 장기 인내자본 투자가 필요한 벤처생태계에도 적극적으로 투입된다.

[그림 3-8] 국민성장펀드 지원대상

대한민국의 미래 = 첨단전략산업 + 벤처·스케일업 + 지역성장

산업 파급효과가 큰 범부처 “메가프로젝트” 중심으로 지원
(대규모 자금 + 규제·재정·세제·지자체·인력 토탈솔루션)

AI 데이터센터 구축

에너지 고속도로

바이오·반도체 투·융자

- ① 미래 20년 新성장동력이 될 “첨단전략산업”(대기업 / 중견기업 / 중소기업)
- ② 미래 산업발전의 씨앗인 “R&D, 벤처·스케일업”
- ③ 지역 소멸 위기를 타개할 수 있는 “지역성장 프로젝트”

자료: 금융위원회(2025. 9. 10)

AI 인재양성에는 과기정통부는 1조 3,000억 원, 교육부는 1조 4,000억 원을 투자한다. 과기정통부는 특히 ‘브레인 투 코리아’ 사업 등을 통해 석박사 연구자 처우를 개선하고 초격차 해외 인재 유치를 위해 노력할 계획이다(조선경제, 2025. 8. 25). 교육부는 AI 인재양성을 위해 2026년 초등등 교육에 9,000억 원, 고등교육에 5,000억 원의 예산을 투입하며, 5.5년 만에 학사·석사·박사 과정을 모두 마칠 수 있는 ‘패스트트랙’도 신설할 계획이다(시사저널, 2025. 11. 10.). 또한 7개 시도에 AI 교육지원센터 설치, AI 거점대학을 지정하여 지역 허브 구축, 최고 수준 석학이 정년 제한 없이 연구할 수 있도록 국가석좌교수제도 도입할 계획이다(시사저널, 2025. 11. 10.).

3. AI 규제 현황

가. 기본 철학: 기술주권 및 국가경쟁력과 AI 규제의 관계 설정

한국의 AI 거버넌스는 “인공지능의 건전한 발전과 신뢰 기반 조성에 필요한 기본적인 사항을 규정함으로써 국민의 권익과 존엄성을 보호하고 국민의 삶의 질 향상과 국가경쟁력을 강화” 한다는 이중 목표를 바탕으로 구축되었다(AI 기본법 제1조)³⁷⁾. 이는 법률의 명칭에서도 드러나듯이 혁신과 신뢰를 동시에 추구하는 균형적 접근이다.

한국 모델의 핵심 철학은 세 가지 측면에서 구체화된다. 첫째, 법 제3조의 기본 원칙은 국가가 “인공지능사업자의 창의정신을 존중”하도록 명시하면서도 “안전한 인공지능 이용환경의 조성을 위하여 노력하여야 한다”고 규정함으로써 진흥과 안전의 균형을 법 원칙으로 선언하고 있다. 둘째, 규제 대상을 포괄적으로 설정하지 않고 사람의 생명·신체·기본권에 중대한 영향을 미치는 ‘고영향 인공지능’과 ‘생성형 인공지능’에 대한 선별적 규제를 채택하였다. 셋째, 형사처벌 없이 과태료 최대 3,000만 원의 경미한 제재 체계를 설계하여 기업 부담을 최소화하면서도 필

37) 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」. 법률 제20676호. 2025. 1. 21. 제정.

요한 법적 집행력을 확보하였다.

이러한 철학적 토대는 4년 이상에 걸친 국회 논의의 산물이다. EU의 포괄적 사전 규제 모델과 미국, 일본의 자율 규제 모델 사이에서 한국만의 독자적 균형점을 모색한 결과물로, 세계 두 번째, 아시아 최초의 포괄적 AI 기본법이라는 역사적 의미를 갖는다.

나. 핵심 입법 및 정책 체계

1) AI 기본법의 제정과 시행³⁸⁾

2025년 1월 21일 제정된 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」(법률 제20676호, 이하 ‘AI 기본법’)과 동법 시행령(대통령령 제36053호)³⁹⁾은 2026년 1월 22일 시행된다. 법 제4조는 역외 적용 조항을 두어 국외에서 이루어진 행위라도 국내 시장 또는 이용자에게 영향을 미치는 경우 적용하도록 규정하고 있으며, 국방·국가안보 목적의 AI는 적용 대상에서 제외하였다(AI 기본법 제4조).

2025년 12월 30일에는 국가인공지능전략위원회 개편 사항의 법제화, 인공지능 연구소 설립·운영 근거 마련, AI 분야 창업 활성화 지원, 공공분야 AI 수요 창출, 공공데이터의 학습용 데이터 제공 근거 마련 등을 내용으로 하는 개정안이 국회 본회의를 통과하여 시행 초기부터 강화된 정책 추진력을 확보하게 되었다.

2) 핵심 개념 체계

AI 기본법은 규제 대상과 의무 주체를 다음과 같이 체계적으로 정의하고 있다.

• 인공지능사업자

법 제2조제7호는 인공지능사업자를 “인공지능산업과 관련된 사업을 하는 자로서 법인, 단체, 개인 및 국가기관 등”으로 정의하면서, ‘인공지능개발사업자’(AI를 개발하여 제공하는 자)와 ‘인공지능이용사업자’(개발사업자가 제공한 AI를 이용하

38) 과학기술정보통신부. (2025). 인공지능 기본법 시행령 주요 내용. 과학기술정보통신부 보도자료.

39) 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 시행령」. 대통령령 제36053호. 2026. 1. 22. 제정.

여 인공지능제품 또는 인공지능서비스를 제공하는 자)로 구분한다. 규율 대상에 국가기관 등이 명시적으로 포함된다는 점이 중요한 특징으로, 정부, 지방자치단체, 공공기관이 AI 서비스를 제공하는 경우에도 법상 의무를 부담하게 된다. 한편 시행령 제27조제3항은 이용사업자가 개발사업자로부터 AI 시스템을 제공받아 중대한 기능 변경 없이 활용하는 경우 개발사업자의 의무 이행이 이용사업자에게도 이전된다고 규정하여 중복 의무를 완화하고 있다.

• 고영향 인공지능

법 제2조제4호는 에너지 공급, 보건의료, 의료기기, 범죄 수사·체포를 위한 생체인식, 채용·대출 심사 등 개인 권리·의무에 중대한 영향을 미치는 판단, 교통수단·시설·체계의 주요 작동, 국가 기관등의 의사결정, 유아·초등·중등교육에서의 학생 평가 등 10개 영역을 고영향 AI로 열거하고, 대통령령으로 추가 지정할 수 있도록 하는 개방적 열거 방식을 채택하였다.

• 생성형 인공지능

법 제2조제5호는 입력 데이터의 구조와 특성을 모방하여 글·소리·그림·영상 등 다양한 결과물을 생성하는 AI 시스템으로 정의한다.

• 고성능 인공지능

AI 기본법은 ‘고성능 인공지능’이라는 별도 용어를 두지 않는 대신, 인공지능 안전성 확보 의무를 규정한 법 제32조에서 "학습에 사용된 누적 연산량이 대통령령으로 정하는 기준 이상인 인공지능시스템"에 대해 안전성 확보 의무를 부과하고 있다. 시행령 제24조는 이를 ① 학습 누적 연산량 10^{26} FLOPs 이상, ② 현재 최첨단 AI 기술 적용, ③ 사람의 생명·신체·기본권에 광범위하고 중대한 영향을 미칠 우려의 세 가지 요건을 모두 충족하는 경우로 한정하였다. EU AI Act의 GPAI 시스템 규제나 미국의 프론티어 AI 개념에 상응하는 것으로, 현재의 연산량 기준상 전 세계적으로 소수의 최대 규모 AI 모델만이 해당된다.

3) 인공지능사업자의 주요 의무 체계

AI 기본법은 AI의 유형에 따라 사업자의 의무를 달리 정하고 있다. 고영향 인공지능, 생성형 인공지능, 고성능 인공지능 순으로 각각의 의무를 살펴보고, 외국 사업자에게 적용되는 의무를 함께 검토한다.

① 고영향 인공지능 사업자의 의무

고영향 AI 사업자의 의무는 AI의 생애주기에 따라 순차적으로 부과된다.

첫째, 고영향 AI 또는 이를 이용한 제품·서비스를 제공하려는 사업자는 먼저 해당 AI가 고영향 AI에 해당하는지를 스스로 검토하여야 한다(법 제33조). 판단이 불확실한 경우에는 과기정통부장관에게 확인을 요청할 수 있으며, 장관은 30일 이내에 결과를 회신하여야 한다(1회에 한해 30일 연장 가능). 결과에 이의가 있는 사업자는 회신을 받은 날부터 10일 이내에 재확인요청을 할 수 있다. 전문위원회(50명 이상)가 자문 기구로 기능한다(시행령 제25조 및 제26조). 이 조항은 제40조의 사실조사 대상에 포함되지 않으며 별도의 제재 규정도 없어, 사업자의 자율적 준수에 의존하는 구조이다.

둘째, 고영향 AI를 이용한 제품·서비스를 출시하기 전에 사람의 기본권에 미치는 영향을 평가하도록 노력하여야 한다(법 제35조제1항). 이는 강행 의무가 아닌 노력 의무이므로 별도의 제재 규정이 없다. 다만 국가 기관 등이 고영향 AI를 이용하려는 경우에는 영향평가를 실시한 제품·서비스를 우선적으로 고려하여야 한다는 점에서(법 제35조제2항), 공공 부문에서는 사실상 강한 유인으로 작동한다. 시행령 제28조는 영향평가에 영향받는 개인·집단의 식별, 기본권 유형의 식별, 영향의 내용과 범위, 위험의 예방·완화·복구 방안, 개선이 필요한 경우 이행계획 등이 포함되어야 함을 규정하고 있다.

셋째, 서비스를 제공할 때에는 고영향 AI에 기반하여 운용된다는 사실을 이용자에게 미리 알려야 한다(법 제31조제1항). 시행령 제23조는 제품에 직접 기재하거나 계약서, 이용약관에 명시하는 방법, 이용자 화면에 표시하는 방법, 서비스 제공 장소에 게시하는 방법 등을 허용하고 있다. 다만 고영향 AI를 활용한 사실이

제품명이나 화면 표시 등에서 이미 명백한 경우, 또는 사업자 내부 업무 용도로만 사용되는 경우에는 고지 의무가 면제된다. 이 의무는 제40조의 사실조사 대상에 포함되지 않으며, 위반 시 3,000만 원 이하의 과태료가 직접 부과된다(법 제43조 제1항제1호).

넷째, 서비스 운영 중에는 안전성과 신뢰성을 확보하기 위한 조치를 지속적으로 이행하여야 한다(법 제34조제1항). 구체적으로는 위험관리방안을 수립하여 운영하고, 기술적으로 가능한 범위에서 AI가 도출한 최종 결과와 그 주요 기준을 이용자에게 설명할 수 있는 방안을 마련하여야 한다. 이용자를 보호하기 위한 방안도 갖추어야 하며, 고영향 AI에 대한 사람의 관리·감독 체계를 유지하고, 이러한 조치의 내용을 확인할 수 있는 문서를 5년간 보관하여야 한다. 시행령 제27조는 위험관리방안의 주요 내용, 설명 방안, 이용자 보호 방안, 관리·감독자의 성명 및 연락처를 사업장 또는 홈페이지에 공개하도록 규정하되, 영업비밀에 해당하는 사항은 제외할 수 있도록 하였다. 이 의무는 제40조에 따른 사실조사 대상에 포함되며, 위반이 확인되는 경우 과기정통부장관은 중지·시정명령을 내릴 수 있고, 이를 이행하지 않으면 3,000만 원 이하의 과태료가 부과된다(법 제40조제3항, 제43조 제1항제3호).

② 생성형 인공지능 사업자의 의무

생성형 AI를 이용한 제품·서비스를 제공하는 사업자는 그 결과물이 생성형 AI에 의해 생성되었다는 사실을 표시하여야 한다(법 제31조제2항). 특히 딥페이크 처럼 실제와 구분하기 어려운 음향·이미지·영상을 제공하는 경우에는 이용자가 명확하게 인식할 수 있는 방식으로 고지하거나 표시하여야 한다(법 제31조제3항). 다만 해당 결과물이 예술적·창의적 표현물에 해당하는 경우에는 전시나 향유를 저해하지 않는 방식으로 고지·표시할 수 있다. 시행령 제23조는 사람이 인식할 수 있는 방법과 워터마크 등 기계가 판독할 수 있는 방법을 모두 허용하되, 딥페이크의 경우 이용자가 명확하게 인식할 수 있는 방법만을 인정하여 비가시적 워터마크만으로는 의무를 이행한 것으로 볼 수 없도록 하였다. 이 의무는 제40조에

따른 사실조사 대상에 포함되며, 위반이 확인되는 경우 중지·시정명령을 거쳐 이를 이행하지 않으면 3,000만 원 이하의 과태료가 부과된다(법 제40조제3항·제43조제1항제3호).

③ 고성능 인공지능 사업자의 의무

고성능 인공지능에 해당하는 시스템을 제공하는 사업자는 수명주기 전반에 걸쳐 위험을 식별하고 평가하며 완화하여야 하고, AI 관련 안전사고를 모니터링하고 대응하는 위험관리체계를 구축하여야 한다(법 제32조제1항). 이행 결과는 과기정통부장관에게 제출하여야 하며(법 제32조제2항), 이 의무 역시 제40조에 따른 사실조사 대상에 포함된다. 위반이 확인되는 경우 중지·시정명령을 거쳐 이를 이행하지 않으면 3,000만 원 이하의 과태료가 부과된다(법 제40조제3항 및 제43조제1항제3호).

④ 국내대리인 지정 의무

국내에 주소나 영업소가 없는 외국 AI 사업자로서 ① 전년도 매출액 1조 원 이상, ② AI 서비스 부문 전년도 매출액 100억 원 이상, ③ 직전 3개월간 국내 이용자 수 1일 평균 100만 명 이상, ④ 중지·시정명령 불이행으로 과태료를 부과받은 전력이 있는 경우 중 어느 하나에 해당하는 자는 국내대리인을 서면으로 지정하고 과기정통부장관에게 신고하여야 한다(법 제36조제1항 및 시행령 제29조). 국내대리인은 ① 고성능 AI 안전성 확보 의무 이행 결과의 제출, ② 고영향 AI 해당 여부 확인 요청, ③ 고영향 AI 사업자의 안전성·신뢰성 확보 조치 이행 지원의 세 가지 사항을 대리한다. 국내대리인이 이 법을 위반한 경우에는 해당 대리인을 지정한 외국 사업자가 그 행위를 한 것으로 본다(법 제36조제3항). 국내대리인을 지정하지 않은 경우에는 3,000만 원 이하의 과태료가 부과된다(법 제43조제1항제2호).

다. 혁신 촉진과 리스크 관리의 균형 방식

1) AI 기본법의 균형 구조

AI 기본법은 혁신 촉진과 위험 관리를 동시에 추구하는 이중적 구조를 취하고

있다. 법 제1조는 입법 목적을 "인공지능의 경쟁력 강화와 안전하고 신뢰할 수 있는 인공지능 기반 조성"으로 명시하여 두 가치를 병렬적으로 제시하고 있다.

혁신 촉진 측면에서 AI 기본법 제3장은 기술 개발과 산업 육성을 위한 광범위한 국가 지원 체계를 규정하고 있다. 기술 개발 분야에서는 연구개발 사업 지원(법 제13조), 기술 표준화(법 제14조), 학습용 데이터 관련 시책 수립(법 제15조)을 규정하고 있다. 산업 활성화 분야에서는 AI 기술의 도입·활용 지원(법 제16조), 중소기업에 대한 특별 지원(법 제17조), 창업 활성화(법 제18조), AI 융합 촉진(법 제19조), 관련 법·제도 개선(법 제20조)을 규정하고 있다. 인프라 측면에서는 전문인력 양성(법 제21조), 국제협력 및 해외 진출 지원(법 제22조), AI 집적단지 지정(법 제23조), 실증기반 조성(법 제24조), 데이터센터 관련 시책 추진(법 제25조)을 규정하며, AI 산업계의 자율 협력 기반으로 한국인공지능진흥협회 설립 근거도 마련하였다(법 제26조). 아울러 새로운 AI 기술·서비스에 대해 기존 규제의 적용을 한시적으로 유예하는 규제 샌드박스 제도를 도입하여 혁신 공간을 확보하였다(법 제28조~제30조).

위험 관리 측면에서는 고영향·생성형·고성능 AI에 대한 법적 의무와 제재를 규정하는 한편, 그 외의 영역에서는 연성법과 자율규제, 인센티브 장치를 조합하는 방식을 채택하였다. 정부는 인공지능 윤리원칙을 제정·공표하고(법 제27조), 민간 자율인공지능윤리위원회 설치를 권장하며(법 제28조), 안전성·신뢰성 검·인증 활동을 지원한다(법 제30조). 이는 규제 의무의 직접 부과보다 자율적 준수 문화의 형성을 우선하는 접근이다.

사업자의 이행 부담 완화를 위한 장치도 마련되어 있다. 정부는 법 시행 초기 1년 이상의 과태료 제도 기간을 운영하고, 고영향 AI 가이드라인 마련, 컨설팅 및 비용 지원을 통해 이행 부담을 최소화할 계획이다(과학기술정보통신부, 2025). 또한 법 제34조제3항은 다른 법령에 따라 이에 준하는 조치를 이미 이행한 경우 고영향 AI 사업자 책무를 이행한 것으로 간주하여 중복 규제를 방지하고 있다.

2) 국가인공지능전략위원회의 AI 행동계획⁴⁰⁾

혁신 촉진을 위한 국가 차원의 종합 전략도 병행하여 마련되고 있다. 국가인공지능전략위원회는 2025년 12월 15일 출범 100일 기념 기자간담회에서 “대한민국 인공지능행동계획(안)”을 발표하고 각계 의견수렴에 착수하였다(2025년 12월 31일 현재 안(案) 단계). 행동계획(안)은 ‘세계 3대 인공지능 강국 도약’을 비전으로, ① 인공지능 혁신 생태계 조성, ② 범국가 인공지능 기반 대전환, ③ 국제 인공지능 기본사회 기여의 3대 정책축 아래 98개의 세부 과제를 담고 있다. 인프라 측면에서는 민관 합작 특수목적법인(SPC)을 통해 2028년까지 GPU 1.5만 장, 2030년까지 5만 장 확보를 목표로 하는 국가 AI 컴퓨팅 센터 구축을 핵심 과제로 제시하였다. 인재 양성 측면에서는 초·중·고 전 학년 AI 필수 교육체계 구축을 추진하며, 산업 측면에서는 2030년 피지컬 AI 세계 1위 달성과 제조업 AI 대전환(AI) 가속화를 목표로 설정하였다.

라. 거버넌스 체계와 집행 역량

1) 거버넌스 체계

AI 기본법은 정책 수립, 조정, 집행의 각 기능을 담당하는 기관을 층위별로 구분하여 규정하고 있다. 정책 최상위 기구로는 대통령 소속 국가인공지능전략위원회를 두고 있다(법 제7조). 위원장은 대통령이 맡으며, 60명 이내의 위원으로 구성되며 민간 위촉위원이 전체의 과반수를 차지하도록 하여 민관 협력 구조를 제도화하였다. 위원회는 AI 기본계획 및 주요 정책을 심의·의결하고 부처 간 정책을 조율하는 역할을 수행한다. 2025년 9월 8일 대통령 주재로 출범한 위원회는 기술혁신·인프라, 산업·공공 AX, 데이터, 사회, 국방·안보 등 8개 분과위원회와 3개 태스크포스를 구성하여 운영하고 있으며, AI 기본법 시행에 따라 법정 기구로 전환되었다.

정책 수립 및 집행의 주무부처는 과학기술정보통신부이다. 과기정통부장관은 3년마다 AI 기본계획을 수립하고(법 제6조), 고영향 AI 확인 요청에 대한 회신, 사

40) 국가인공지능전략위원회. (2025). 대한민국 인공지능 행동계획(안).

실조사, 중지·시정명령, 과태료 부과·징수 등 법 집행 전반을 담당한다(법 제33조, 제40조 및 제43조). 전문적 자문을 위해 고영향 AI 전문위원회(50명 이상)를 운영하며(시행령 제26조), AI 관련 정책 개발과 국제규범 정립·확산 업무를 위한 인공지능정책센터를 별도로 지정할 수 있다(법 제11조). 시행령 제9조는 한국지능정보사회진흥원(NIA)을 포함하여 대학 부설연구소, 공공기관, 비영리법인, 정부출연연구기관 등 중 과기정통부장관이 인정하는 기관을 정책센터로 지정할 수 있도록 규정하였다.

AI 안전 전담기관으로는 인공지능안전연구소를 운영한다(법 제12조). 안전연구소는 ① AI 안전 관련 위험 정의 및 분석, ② AI 안전 정책 연구, ③ AI 안전 평가 기준·방법 연구, ④ AI 안전 기술 및 표준화 연구, ⑤ 국제교류·협력, ⑥ AI 시스템 안전성 확보 지원 등의 사업을 수행하며, 시행령 제10조는 안전 관련 자문·교육, 평가 시스템 구축·운영, 데이터 확보·공개, 통계·정보 분석·제공을 추가 사업으로 규정하였다. 2025년 12월 30일 국회를 통과한 AI 기본법 개정안에서는 안전연구소의 설립 근거가 더욱 구체화되었다.

2) 집행 역량

AI 기본법의 집행은 사실조사권과 중지·시정명령권을 핵심 수단으로 삼고 있다. 과기정통부장관은 생성형 AI 표시의무(법 제31조제2항 및 제3항), 고성능 AI 안전성 확보 의무(법 제32조), 고영향 AI 안전성·신뢰성 확보 조치(법 제34조제1항) 위반이 발견되거나 신고·민원이 접수된 경우 자료 제출을 요구하거나 사업장에 직접 출입하여 조사할 수 있으며(법 제40조제1항 및 제2항), 위반 사실이 확인되면 중지·시정에 필요한 조치를 명할 수 있다(법 제40조제3항). 과태료는 최대 3,000만 원으로(법 제43조), EU AI Act의 최대 3,500만 유로나 중국의 형사책임 수준에 비해 낮게 설정되어 있어 실질적 억지력보다는 준수 유도과 협력적 집행에 방점을 두는 구조이다. 법 시행 초기 1년 이상의 제도 기간 운영 계획도 이러한 집행 기초를 잘 보여준다.

마. 국제적 영향력 확산 전략

한국의 국제 AI 거버넌스 참여는 법적 근거, 다자 외교, 양자 협력의 세 차원에서 이루어지고 있다. 법적 차원에서 AI 기본법은 국제 협력을 명시적 정책 과제로 규정하고 있다. 법 제22조는 AI 기술 및 산업의 국제 협력과 해외 진출 지원을 정부의 책무로 정하고 있으며, 인공지능안전연구소의 사업 범위에도 국제교류·협력이 명시적으로 포함되어 있다(법 제12조). 국가인공지능전략위원회 역시 글로벌 협력 분과위원회를 별도로 두어 국제 AI 거버넌스 논의에 대한 전략적 대응 체계를 갖추었다.

다자 외교 차원에서 한국은 2024년 5월 AI 서울 정상회의를 개최하여 AI 안전·혁신·포용의 세 가치를 글로벌 AI 거버넌스의 핵심 의제로 제시하였다. 서울 정상회의는 2023년 영국 블레츨리 AI 안전성 정상회의의 후속으로 개최된 것으로, 한국은 이를 통해 AI 거버넌스 국제 논의에서 중견국으로서의 조정자 역할을 수행하였다. OECD AI 원칙, GPAI(AI에 관한 글로벌 파트너십), UNESCO AI 윤리 권고 등 주요 다자 규범 형성 과정에도 적극적으로 참여하고 있으며, 2025년 9월 OECD가 발표한 한국 AI 정책 현황 분석에서 한국은 통신 인프라, 디지털 정부, 공공데이터 개방 분야에서 최상위권을 기록한 것으로 평가되었다(OECD, 2025).

양자 협력 차원에서는 AI 기본계획 수립 시 국제 협력 방향을 포함하도록 법으로 규정하고 있으며, 미국·일본·EU 등 주요국과의 AI 정책 대화와 기술 협력을 지속적으로 추진하고 있다. 한국 AI 기본법 자체가 아시아 최초의 포괄적 AI 입법으로서 아시아·태평양 지역 국가들의 입법 모델로 주목받고 있다는 점에서, 한국의 규범 수출 잠재력도 점차 부각되고 있다.

다만 OECD는 한국이 다양한 채널을 통해 국제 협력을 양적으로는 충분히 추진하고 있으나, 보다 정교한 AI 외교 전략을 마련하여 질적 수준을 높여야 한다고 권고하고 있다(OECD, 2025). 이는 한국이 국제 AI 규범 형성 과정에서 소비자를 넘어 실질적인 형성자로 자리매김하기 위해 해결해야 할 과제이다.

바. 평가 및 시사점

한국 AI 기본법이 제시하는 균형 모델은 다음과 같은 강점을 지닌다. 첫째, 한국은 세계에서 두 번째로, 아시아에서는 최초로 포괄적 AI 기본법을 제정함으로써 국제 AI 규범 형성 논의에서 선도적 위치를 점하게 되었다. 둘째, 고영향 AI 개념을 통해 규제 대상을 선별함으로써 혁신에 대한 부담을 최소화하면서도 핵심적인 위험 영역을 포괄하는 균형 잡힌 설계를 실현하였다. 셋째, 대통령 직속 최상위 거버넌스 기구를 법으로 명문화하여 부처 간 AI 정책을 실질적으로 조율할 수 있는 제도적 기반을 갖추었다. 다섯째, 역외 적용 조항과 국내대리인 제도를 통해 국내에 거점을 두지 않은 글로벌 AI 기업에 대해서도 규제를 실효성 있게 적용할 수 있는 장치를 마련하였다.

반면 이 모델이 안고 있는 한계도 분명하다. 첫째, 법 시행 초기 계도 기간이 운영됨에 따라 실질적인 규제 집행까지는 상당한 시간이 걸릴 것으로 예상되어 규제가 실제로 작동하는 데 시간이 필요하다. 둘째, 핵심 사항들이 대통령령, 고시, 가이드라인으로 위임되어 있어 하위 규범을 제때 정비하고 구체화하는 것이 법을 제대로 운용하기 위한 핵심 과제로 남아 있다. 셋째, 안전성 확보 의무(법 제32조)의 적용 요건이 10^{26} FLOPs 이상으로 매우 엄격하게 설정되어 있어, 현실적으로 적용 대상이 지극히 제한적이라는 비판이 있다. 넷째, AI 기본법과 개인정보 보호법, 지능정보화기본법 등 기존 법률 사이의 정합성을 확보하는 것이 지속적인 과제로 남아 있다.

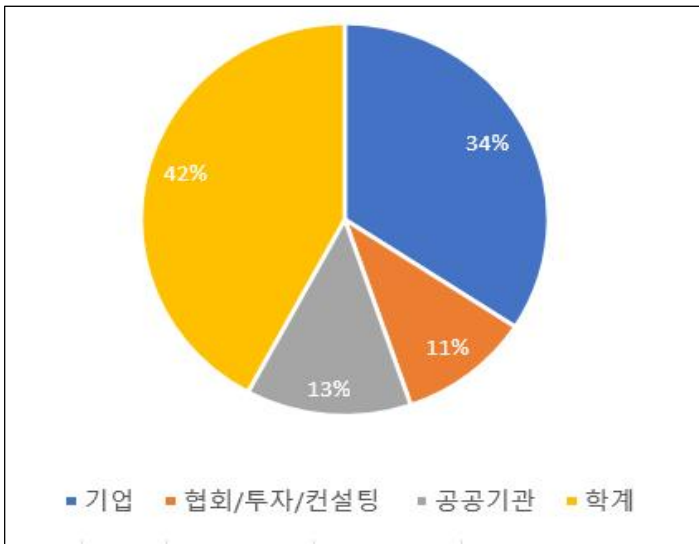
이러한 강점과 한계를 종합할 때, 한국의 균형 모델은 EU의 포괄적 위험 기반 규제와 미국과 일본의 자율규제 사이에서 독자적인 위치를 점하고 있으며, 특히 아시아·태평양 지역 국가들과 글로벌 사우스 국가들에게 현실적인 참조 모델이 될 수 있다. 제재는 가볍게 설계하되 거버넌스 구조는 강력하게 구축하고, 혁신을 최대한 보장하되 핵심 위험은 관리하는 이 접근법은 AI 기술 발전 단계와 규제 역량이 서로 다른 다양한 국가들이 채택하기에 유연한 모델로 주목받고 있다.

제 4 장 국내 AI 생태계 진단

제 1 절 국내 AI 기술자립도 및 경쟁력 진단: 전문가 설문결과를 중심으로

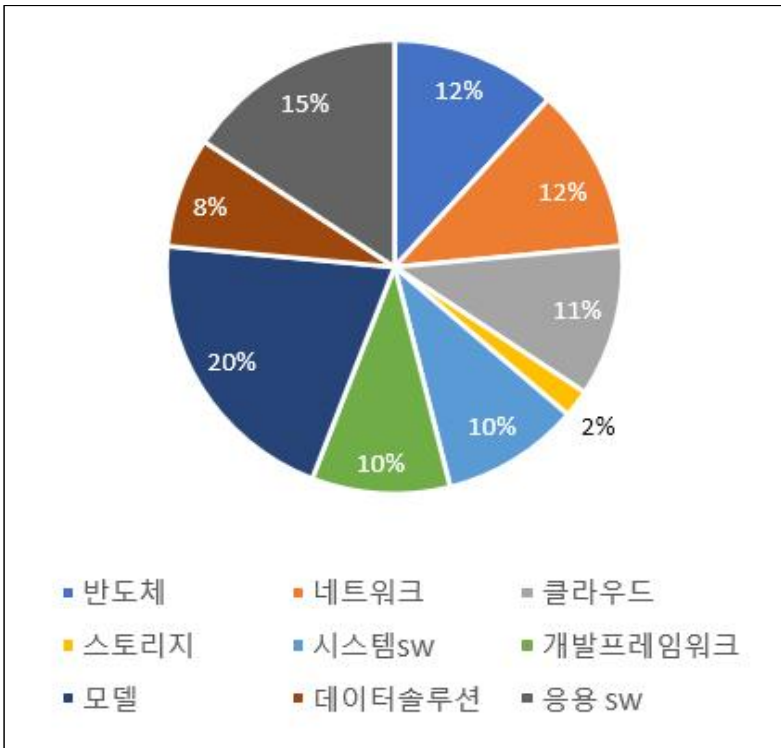
본 절에서는 AI 기술 전문가 설문 결과를 바탕으로 우리나라 AI 기술 자립도, 경쟁력, 차세대 AI 시장에서의 우리의 주도권 확보 가능성, AI 대전환의 주요 장애물 등을 진단한다. AI 전문가 설문 응답자는 총 38명이며, 소속별 응답자 구성은 학계 16명(42%), 기업 13명(34%), 공공기관 5명(13%), 협단체 및 투자·컨설팅 4명(11%)이다.

[그림 4-1] 전문가 설문 응답자 소속별 구성



또한, AI 기술스택을 반도체, 네트워크, 클라우드, 스토리지, 시스템SW, 개발 프레임워크, 모델, 데이터 솔루션, 응용SW 등 총 9개의 영역으로 구분하여 설문 응답자들의 전문영역을 기술스택별로 살펴보면 다음의 그림과 같다.

[그림 4-2] 전문가 설문 응답자 전문영역별 구성



주: 중복선택 허용

그림에서 볼 수 있듯이 응답자 구성은 AI 모델이 21명(20%)으로 가장 높은 비중을 차지했으며, 가장 적은 영역은 스토리지로 2명(2%)으로 스토리지를 제외하면 응답자들의 전문성은 AI 기술스택 전반에 걸쳐 고르게 분포되어 있다.

1. AI 기술자립도 및 경쟁력 진단

AI 기술자립도 및 경쟁력은 앞서 언급한 9개의 기술스택을 중심으로 진단한다. 우선 9개의 기술스택의 주요 역할과 기능은 다음과 같다.

〈표 4-1〉 역할/기능별 AI 기술스택

기술스택	역할/기능
반도체(GPU/메모리/패키징/인터커넥트)	AI 모델 학습·추론에 필요한 고성능 연산을 수행하는 핵심 하드웨어
네트워크(스위치/패브릭/광모듈)	여러 GPU/노드를 고속으로 연결, 병렬 학습·분산 처리 지원
클라우드/하이퍼스케일러	컴퓨팅·스토리지·네트워크 자원을 유연하게 관리·확장
스토리지 (병렬 I/O)	GPU와 스토리지 간 초고속 데이터 읽기/쓰기 파이프라인 제공
시스템 SW(분산학습/서빙)	하드웨어와 프레임워크를 연결하고 자원 관리·통신 최적화 담당
개발플랫폼/프레임워크	모델 학습·추론을 위한 고수준 프로그래밍 인터페이스 제공
모델·SW(LLM/틀체인)	데이터를 기반으로 지능을 구현해, 상위 계층(플랫폼/서비스, 응용)을 가능하게 만드는 코어
데이터 솔루션 (정제/거래/보호)	데이터를 수집·정제·가공·관리·교환해 모델 학습과 서비스 운영이 가능하도록 지원
응용SW	모델을 실제 서비스·제품에 통합, 사용자가 활용할 수 있는 기능을 구현·제공

기술스택별 기술자립도 및 경쟁력 진단에는 ① 해외의존도, ② 공급망위험도, ③ 현재 우리나라의 경쟁력, ④ 향후 경쟁가능성 4개의 지표를 활용하였으며 AI 기술전문가 각자에게 자신이 평가가능한 스택을 선택하여 5점 척도로 평가(1점 매우 낮음, 2점 낮은 편, 3점 보통, 4점 높은 편, 5점 매우 높음)하도록 하였다. 참고로 스택별 평가 인원은 최대 23명(AI모델), 최소 8인(응용 SW)이며, 평가 인원이 3인으로 집계된 스토리지 부문은 평가 결과의 신뢰성을 담보하기 어려워 정량화된 진단 결과에서는 제외하였다.

가. 해외의존도 및 공급망 위험도

AI 기술전문가 설문을 통해 진단한 국내 AI기술스택별 해외의존도 및 공급망 위험도 수준은 다음과 같다.

[그림 4-3] AI기술스택별 해외의존도 및 공급망위험도: AI 전문가 설문결과



주: 5점척도(1 매우 낮음, 2 낮은 편, 3 보통, 4 높은 편, 5 매우 높음) 평가결과

우선, AI 기술스택별 해외의존도 평가결과를 살펴보면 반도체/가속기(4.93), 개발플랫폼/프레임워크(4.81), 시스템SW(4.50), 네트워크(4.43), 클라우드(4.33), AI모델(4.00), 데이터인프라/솔루션(3.00), 응용SW(2.88) 순으로 데이터인프라/솔루션, 응용SW를 제외한 모든 기술스택에서 우리의 해외의존도가 보통(3)보다 높은 수준으로 나타났다. 실제로 현재 거의 모든 AI 기술스택에서 미국 기업들이 독점적인 영향력을 행사하고 있어 이러한 전문가 설문 결과는 AI 기술생태계의 글로벌 경쟁현황에서 도출가능한 결론과 일맥상통한다.

〈표 4-2〉 AI 기술스택별 글로벌 경쟁현황

스택	글로벌 경쟁현황	공급망 리스크 요인
반도체(GPU/메모리/패키징/인터커넥트)	• GPU: NVIDIA(독점)	• GPU 단일공급
	• 메모리: SK hynix·삼성(과점)	
	• 패키징: TSMC(과점)	• TSMC 패키징 종속
네트워크 (스위치/패브릭/광모듈)	• Broadcom(ASIC 과점)	• Broadcom ASIC 종속
	• NVIDIA(InfiniBand 준독점)	• InfiniBand 종속
	• 광모듈: 美·중국·대만·일본 (경쟁적)	
클라우드 (AI 클라우드/ 하이퍼스케일러)	• AWS·MS·Google(과점)	• GPU 구매력 Hyperscaler 집중
스토리지(병렬 I/O)	• DDN·WekaIO·NetApp(과점)	• 병렬 SW 해외 종속
시스템 SW (분산학습/서빙)	• NVIDIA CUDA·NCCL(독점) • AMD ROCm • Intel OneAPI • OpenAI의 Triton	• NVIDIA 절대독점 • PyTorch-CUDA-NCCL
개발 플랫폼 (프레임워크·툴)	• PyTorch(논문 70~80%, 무료 오픈소스) • TensorFlow·JAX	• 사실상 PyTorch 독점 • 대체재 진입 불가
모델	• OpenAI·Anthropic·Google (과점) • Meta·Mistral(경쟁)	• Big Tech 종속
데이터 솔루션 (정체/거래/보호)	• Databricks, Snowflake, Palantir(과점)	• 데이터 마켓·클린룸해외 의존 • EU 표준 리스크

그러나, 기술의 해외의존도가 높은 것이 반드시 공급망 위험도가 높은 것을 의미하지는 않다. 실제로 AI 전문가 설문 결과, 네트워크, 개발플랫폼/프레임워크의 공급망 위험도는 각각 2.71, 2.63으로 보통보다 낮은 것으로 나타났다. 네트워크의 경우 사실상 해외 기술에 100% 의존하고 있으며, 원천 경쟁력이 취약하지만 광소자, 패키징, CPO, 실리콘 포토닉스 등에서 중장기적으로 기회는 있을 것이라는 의견, 개발 플랫폼의 경우 국내 플랫폼은 없으나 오픈소스라는 점에서 공급망

위험도는 낮다는 의견 등이 제시되었다. 물론 전문가들의 의견이 모두 일치하지는 않았는데 각 기술스택과 관련하여 전문가들의 주요 의견을 정리해보면 다음의 표와 같다.

〈표 4-3〉 AI 기술스택별 현황에 대한 전문가 의견

AI 기술스택	주요 의견
반도체/가속기	<ul style="list-style-type: none"> • GPU 설계 및 생산 미국/대만이 사실상 독점, 패키징도 마찬가지 • HBM을 제외한 나머지 AI 반도체 역량 미흡 • 메모리(HBM), 첨단패키징에서는 경쟁력 有 • 메모리영역에서 중국의 성장세
시스템SW	<ul style="list-style-type: none"> • 대부분 NVIDIA에 의존하나, 오픈 소스 생태계이므로 위험도는 낮음 • 해외 주요 업체/주요 연구그룹의 주도, 중국 업체/주요 연구그룹의 주도, 투자 부족/난이도높음
AI 모델	<ul style="list-style-type: none"> • 프론티어 모델 성능은 해외 빅테크가 압도적, 한국어 특화, 특정 도메인, 산업별 모델은 가능성 있음, • 기술 공유, 공개가 적어, 다수 참여 프로젝트 부재, 대기업 주도 AI 개발도 기초 모델에 대한 투자는 미미 • 공급망 위험도는 현재 OpenAI, 구글 등이 경쟁 중이기에 분산되어 있으나 잠재적으로 매우 높음
클라우드/하이퍼스케일러	<ul style="list-style-type: none"> • 네이버 등 국산 클라우드 경쟁력 향상 중 vs. 최신 AI 기술 활용, 확산 관점에서 특히 경쟁력 낮음 • GPU 종속 및 규모 한계, 기술 및 규모로 선도업체 전무 • 공공 등 규제요건
응용SW	<ul style="list-style-type: none"> • AI 활용 SW는 모두에게 열려있음, B2B는 확장 및 글로벌 경쟁 가능한 SW 존재, 그러나 Big Tech 기업의 기술 수준과 격차 • B2C 한국사용자 대상 서비스는 다수 존재하나 글로벌 경쟁력은 미지수
네트워크	<ul style="list-style-type: none"> • 네트워크 모듈 및 장비 시장 공략 가능 vs. 네트워크 칩 및 제조업체 부족 • SW 등 핵심 기술이 거의 전량 해외 의존(Broadcom·NVIDIA·Marvell 등 해외 기술에 사실상 100% 의존)
데이터솔루션	<ul style="list-style-type: none"> • AI 신뢰성 및 데이터 모니터링 분야는 글로벌 신생 시장이라 아직 주도권이 정해지지 않아 경쟁가능 • 데이터 저작권 이슈, 정부의 규제 및 관리 • 데이터의 수집, 정제, 가공에 투자하는 기업 부족, 데이터 소유권자들의 비용 청구 과도

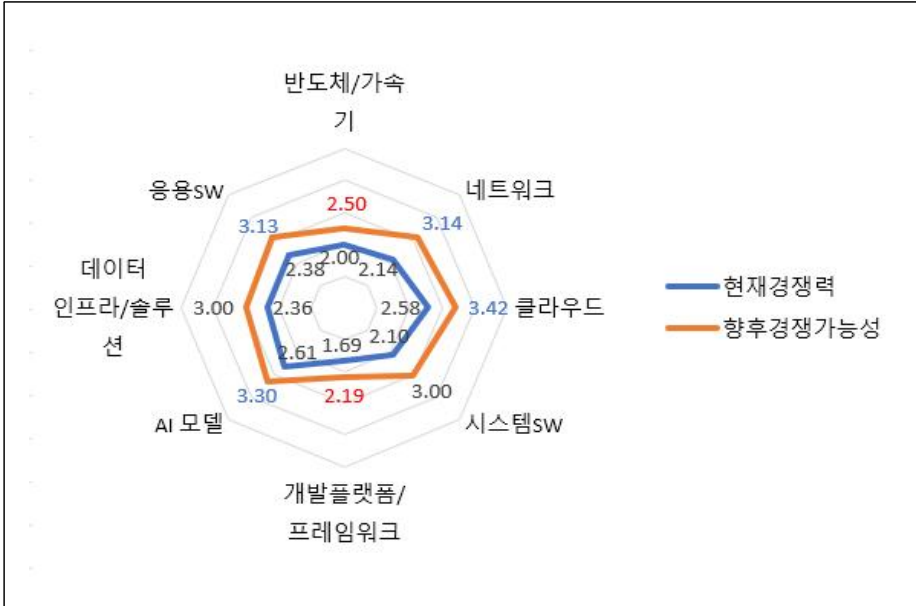
AI 기술스택	주요 의견
개발플랫폼/ 프레임워크	<ul style="list-style-type: none"> • 오픈소스 활성화 가능 vs. 한국 내 오픈소스, 프레임워크 진영을 주도하는 기업, 단체 부재 • 글로벌 AI 개발 스택은 CUDA-PyTorch/TensorFlow-Langchain 기반 미국 주도 오픈소스 프레임워크에 절대적으로 의존하며 이 기술 체인은 향후 거부할 수 없는 영향을 미칠 것 • 퓨리오사AI 같은 국내 NPU가 개발되어도 결국 이 해외 생태계에 종속되어야 하는, 장기적인 관점 생태계 종속 위험 매우 큼

종합해보면 해외의존도와 공급망 위험도 모두 보통보다 높게 나타난 기술 스택은 반도체/가속기(4.15), 시스템SW(3.80), 클라우드(3.50), AI 모델(3.36) 등이다.

나. 현재 경쟁력 및 향후 경쟁가능성

다음으로 경쟁력 측면에서 살펴보면, 기술스택별 현재 우리의 경쟁력은 AI모델(2.61), 클라우드(2.58), 응용SW(2.38), 데이터인프라/솔루션(2.36), 네트워크(2.14), 시스템SW(2.10), 반도체/가속기(2.00), 개발플랫폼/프레임워크(1.69) 등의 순으로 모든 기술스택에서 보통보다 낮은 수준으로 평가되었다. 그러나, 향후 경쟁가능성은 현재 경쟁력보다는 모두 높은 수준으로 평가되었으며 클라우드(3.42), AI모델(3.30), 네트워크(3.14), 응용SW(3.13)는 보통보다 다소 높게, 데이터인프라/솔루션(3.00), 시스템SW(3.00)은 보통 수준으로 나타났다. 반도체/가속기의 경우 2.50, 개발플랫폼/프레임워크는 2.19로 향후 경쟁가능성 측면에서도 보통보다 낮은 수준으로 평가되었다.

[그림 4-4] 시기술스택별 현재 경쟁력 및 향후 경쟁가능성: AI 전문가 설문결과

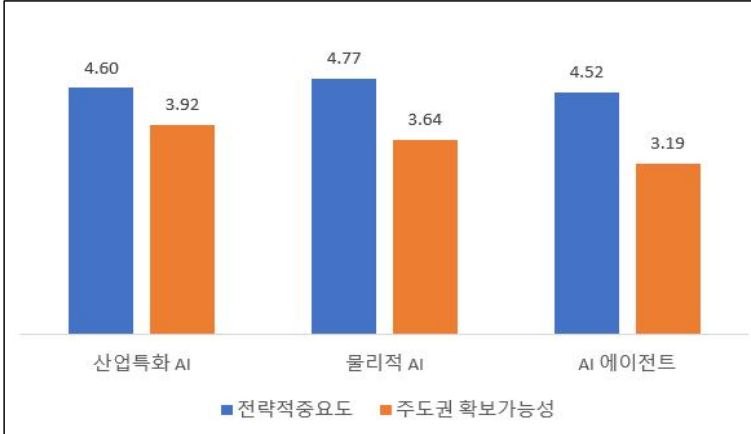


주: 5점척도(1 매우 낮음, 2 낮은 편, 3 보통, 4 높은 편, 5 매우 높음) 평가결과

2. 차세대 AI 시장의 전략적 중요도 및 주도권 확보 가능성

기술주도 성장의 측면에서 가까운 미래에 우리가 주목해야할 AI 시장을 찾아보고 차세대 AI 시장별(잠재 시장규모+전략적 파급력을 고려한) 전략적 중요도, 우리의 주도권 확보가능성을 평가한 결과, AI 전문가들이 가장 많이 언급한 시장은 산업특화 AI(25명), 물리적 AI(22명), AI 에이전트(21명)로 조사되었다. 전략적 중요도는 물리적 AI 4.77, 산업특화 AI 4.60, AI 에이전트 4.52로 모두 매우 높게 나타났으며, 주도권 확보가능성은 산업특화 AI 3.92, 물리적 AI 3.64, AI 에이전트 3.19 순으로 나타났다.

[그림 4-5] 차세대 AI 시장별 전략적 중요도 및 주도권 확보가능성: AI 전문가 설문결과



주: 5점척도(1 매우 낮음, 2 낮은 편, 3 보통, 4 높은 편, 5 매우 높음) 평가 결과

그 밖에도 로봇틱스, 휴머노이드, 자율주행, 온디바이스 AI(HW-SW 통합솔루션, 온디바이스 LLM), AI 인프라(반도체, 네트워크, 스토리지, 클라우드, 모델 등), (사이버보안, AI 모니터링 및 신뢰성(AgentOps/Trustworthy AI) 등) 목적 특화 AI, AI-RAN for 6G, AI 시티 등이 차세대 유망 AI 시장으로 언급되었다.

<표 4-4> 차세대 AI 시장의 전략적 중요도 및 주도권 확보 가능성 평가 이유

분야	점수	이유
산업특화 AI	5	<ul style="list-style-type: none"> 제조업 강국이기에 산학협력이 잘 이루어진다면 데이터 확보나 가치있는 문제 발굴이 수월해 발전 가능성이 클 것으로 예상 한국 기업은 결국 적은 인원으로 최대의 성과를 내야함. 이를 위해 다양한 관점에서 AI를 기반으로한 업무 프로세스 간소화, 자동화 필요 제조, 자동차, 반도체, 의료기기 분야 세계적 경쟁력 보유. 규제-로컬 지식 기반의 산업별 특화는 한국 기업에 비교우위 HW 내재 솔루션 강점
	4	<ul style="list-style-type: none"> 구현 및 요구 시점이 가장 가까우며 반도체, 자동차, 조선 등 국내 주력 산업 중심으로 주도 가능성 존재
	3	<ul style="list-style-type: none"> 제조 등 일부 국내 시장은 경쟁력이 있으나, 법률, 메디컬 시장 등은 상대적으로 규제 및 기술도입 지연

분야	점수	이유
산업특화 AI	2	<ul style="list-style-type: none"> • 기반은 갖추었으나 기반의 온전한 활용 여부는 의문. 데이터 활용 관점에서 규제와 개별 기업별 자체 개발/활용 등의 추세라 파편화, 중복투자 많음 • 현대·포스코 등 제조업 적용 사례 있으나, 시물레이션 SW 역량 미흡 • 결국 범용AI에서 특화AI로 발전해 각 산업 섹터에서 선도업체의 차별성을 부여할 것으로 예상되나 국내 산업 섹터별 특화AI에 대응 인력 부족
물리적 AI	5	<ul style="list-style-type: none"> • 산업 특화 AI에서 해소하고자 하는 것은 사무직군을 중심으로 관리직을 포괄하는 형태 자동화, 현장직무에서 필요한 AI는 결정이 아닌 행동 로봇
	4	<ul style="list-style-type: none"> • 대한민국은 하드웨어, 소프트웨어 가치사슬을 모두 갖춘 몇 안되는 나라. 로봇 HW 제조 우위 (중국 제외) • 자동차, 조선, 헬스케어 등에서 실증 기술을 보유한 국가로서, 이 분야 발전에 주도권 강화 가능
	3	<ul style="list-style-type: none"> • 로봇 및 자율주행 분야는 미국/중국 글로벌 선도 기업 주도. 다만 반도체/부품 생태계 연계를 통한 강점 발휘 가능성 • 제조업 중심의 혁신으로 시장 주도를 기대하나, 기진출 기업들의 기술력 추격 필요, 그래픽스 분야나 효율적 AI 분야 등 관련 분야가 약함
	2	<ul style="list-style-type: none"> • 로봇틱스 및 제조 디바이스 등 피지컬 시장 자체가 경쟁열위로, 물리적 AI도 영향 예상
AI 에이전트	4	<ul style="list-style-type: none"> • 응용 도메인 별로 국내 기술 선도 가능 • 사회가 고도의 자동화에 익숙 • 근 미래에 폭발적 성장 예상, 다양한 국내 플랫폼이 존재하며, 이들의 디지털 전환 및 멀티 에이전트 전환과정에서 큰 가능성 존재
	3	<ul style="list-style-type: none"> • AI 생태계 기반은 글로벌 상위권이나 속도면에서 우려. 대기업 등 AI 활용 사례는 유럽보다 느림, 비전, 목표, 조직적 추진이 미흡하며 파편적 시도 • 한국어 특화 에이전트, 정부/공공 도메인 최적화 등 틈새시장, 일부 국내 시장/사용자 특화 서비스 등에 가능성 존재, 그 이외 경우 경쟁력은 미지수 • 산업 도메인에서 활용되기 위해서는 AI 플랫폼, SW보안, 안전, 윤리 등 과제가 여전히 존재해 한동안 제한적 확산 전망 • SW생태계 열위, 글로벌 Big Tech 주도 예상, 삼성/LG의 가전 기업 역량이 있으나 구글/아마존과의 경쟁

제 2 절 국내 AI 정책 진단: 전문가 설문결과를 중심으로

1. 국내 AI 정책의 타당성·정합성·균형성 진단

국내 AI 정책의 목표는 크게 3가지로 구분된다. 첫째, 기술주권 확보로 AI 핵심 기술에 대한 자립 혹은 통제 역량 확보를 의미한다. 둘째, 기술주도 성장으로 AI 산업을 국가 성장엔진으로 육성하여 경제성장을 견인하는 것으로 목적으로 한다. 셋째, 모두의 AI로 중소기업·개발자·시민·공공기관 누구나 AI를 쉽게 활용할 수 있는 환경을 구축함으로써 국가 전체의 혁신 역량을 제고하여 AI 대전환을 이끌어내는 전략이다. 이번 정부가 발표한 국정과제를 AI정책목표별로 구분하여 보면 다음 표와 같다.

〈표 4-5〉 AI 정책목표별 국정과제

목표	내용
기술 주권 확보	① AI 3대 강국 도약을 위한 AI 고속도로 구축 <ul style="list-style-type: none"> • (컴퓨팅 인프라) GPU 확보 및 민간 AIDC 생태계 육성 • (네트워크) 6G 상용화와 실시간·초정밀 AI 서비스 위한 지능형 기지국 확산 • (데이터) 공공·민간 AI 학습용 데이터 통합플랫폼/데이터 스페이스/통계 메타 데이터 구축 • (컨트롤타워) 실질적 컨트롤타워로서 AI 국가위원회 역할 강화 ② 독자 AI 모델 개발 지원 <ul style="list-style-type: none"> • 글로벌 수준의 독자 AI모델 개발을 위해 AI 정예팀 GPU·데이터·인재 집중 지원 ③ AI 반도체 산업 생태계 확립 <ul style="list-style-type: none"> • NPU·PIM R&D 지원, 국산 NPU+AI 서비스 패키지 테스트베드 확대, 사업화 적시 지원
기술 주도 성장	① 핵심·융합인재 양성 <ul style="list-style-type: none"> • AI 중심대학, 과기원 AI 단과대학 설립, 대학·대학원 간 패스트트랙 운영 및 국내외 우수인재 확보 • 다양한 산업수요 기반 AX 인재양성을 위해 이노베이션 아카데미 확대, AX 대학원 설립 ② 미래 AI 판도를 흔들 차세대 AI 원천기술 확보 <ul style="list-style-type: none"> • 혁신 AI 연구에 도전할 수 있는 국가 AI연구소 육성

목표	내용
기술 주도 성장	③ AI 반도체 산업 생태계 확립 • NPU·PIM R&D 지원, 국산 NPU+AI 서비스 패키지 테스트베드 확대, 사업화 적시 지원 ④ 피지컬 AI 적용 가속화 • 핵심기술 확보 및 공장·로봇·자율주행·자율운항선박 등 적용 가속화 ⑤ 벤처 투자시장 확대 및 넥스트 유니콘 프로젝트 • AI 혁신펀드, 성장사다리, AI 창업패키지, 해외진출 등 AI 스타트업 창업·스케일업 지원 • 넥스트 유니콘 프로젝트 추진
모두의 AI	① 모두의 AI 기반 구축 • AI 파운데이션 모델 기반 국민 AI 접근성 증진 및 공공·경제·사회 AX 도모 • 누구나 AI 서비스를 개발·체험할 수 있도록 AI 캠페인·배움터 조성 ② 산업AI 전환 촉진 • 제조기업의 AI 팩토리 전환, 중소기업 대상 AI 바우처 지원, 서비스업의 AI 전환 지원 ③ 지역AI 전환 촉진 • 지역특화산업의 AI 전환거점인 광역별 AI혁신거점을 구축, 첨단산단을 AI로 전환하는 AX 실증산단 조성 ④ 공공 AI / K-AI 시티 / AI 기본사회 실현 • AI 정부실현을 위한 30대 핵심과제, 사회시스템 전반의 개혁을 주도할 AI 프로젝트 추진 • 범정부 AI 공통기반 구현, 전 공무원 AI 역량 강화 • 공공 AI 시장 창출, 공공데이터 개방, 공공 AI 활용에 대한 안전·신뢰성 확보 • 거주하며 실증하는 AI 특화 시범도시 조성·확산 • 다수 국가가 참여하는 글로벌 AI 컨소시엄, 글로벌 AI 공동기금 및 AI 특화지구 조성 ⑤ 안전·신뢰기반 확보 • AI 오남용 대응 핵심 기술 개발, 디지털 보안·안전 확보, 개인정보보호 체계 확립

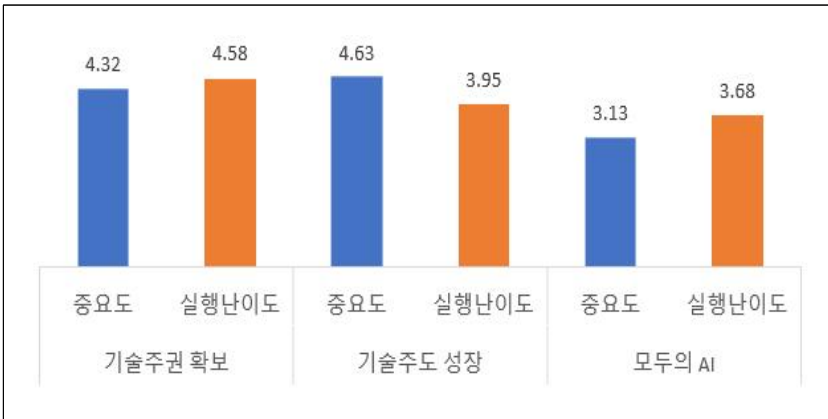
자료: 국정과제에 기반하여 저자가 분류

가. 정책목표별 중요도 및 실행난이도

AI 전문가들을 대상으로 각 정책목표별 중요도 및 실행난이도에 대해 조사한 결과, 기술주도 성장, 기술주권 확보의 중요도는 각각 4.63, 4.32로 매우 높게 평가되었으며 모두의 AI는 3.13으로 상대적인 중요도는 기술주도 성장, 기술주권 확보 대비 낮게 평가되었다. 실행난이도 측면에서는 기술주권이 4.58로 매우 높

게 나타났으며 기술주도 성장 3.95, 모두의 AI 3.68로 나타났다.

[그림 4-6] 정책목표별 중요도 및 실행난이도: AI 전문가 설문결과



주: 5점척도(1 매우 낮음, 2 낮은 편, 3 보통, 4 높은 편, 5 매우 높음) 평가결과

나. AI 정책의 타당성·정합성·균형성 진단

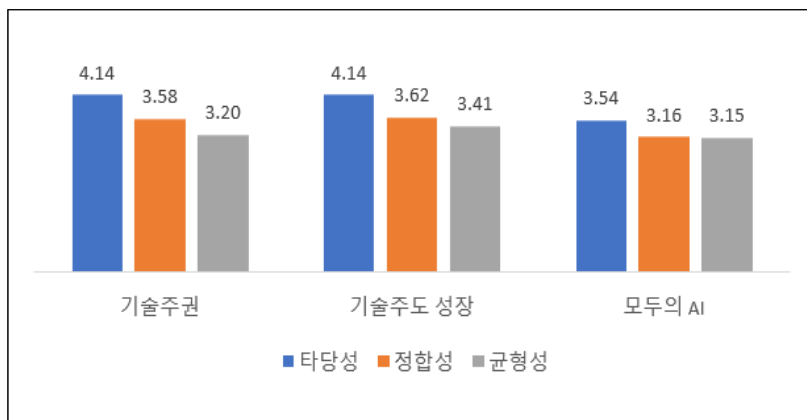
다음으로 AI 정책을 타당성, 정합성, 균형성의 측면에서 평가하였으며 각 지표가 의미하는 바는 다음과 같다.

<표 4-6> 정책진단 지표

지표	내용
타당성	<ul style="list-style-type: none"> 정책 목표 및 설계 방향이 타당한가? 개별 사업·과제가 정책 목표 달성에 기여하는 방향으로 설정되었는가? (ex. AI 고속도로 구축, 독자 AI 모델 개발 등은 기술주권 확보에 실질적으로 기여할 수 있는가, GPU 및 6G 중심의 전략은 정책목표 달성에 타당한가)
정합성	<ul style="list-style-type: none"> 개별 사업들이 큰 전략적 그림 안에서 유기적으로 연결되어 있는가?
균형성	<ul style="list-style-type: none"> 특정 영역에 자원이 과도하게 집중되지는 않았는가?(ex. GPU 확보 등) 중요한 영역이 빠져 있거나 방치되어 있지는 않은가?

AI 전문가들은 기술주권, 기술주도성장 측면의 국정과제에 대해 타당성은 높지만 상대적으로 정합성, 특히 균형성이 낮은 것으로 평가했다. 모두의 AI는 기술주권, 기술주도성장 대비 타당성이 낮은 것으로 평가되었는데 이는 앞서 정책목표별 중요도 평가에서 모두의 AI가 다른 정책목표 대비 중요도가 낮다고 평가된 결과와 동일선상에서 이해할 수 있다.

[그림 4-7] 정책목표별 타당성·정합성·균형성: AI 전문가 설문결과



주: 5점척도(1 매우 낮음, 2 낮은 편, 3 보통, 4 높은 편, 5 매우 높음) 평가결과

<표 4-7> 국내 AI 정책에 대한 전문가 의견

정책목표	주요 의견
기술주권	<ul style="list-style-type: none"> • GPU 편중 → 공급망 다변화(NVIDIA·AMD·HPC·국산 NPU)로 수정 필요 • GPU 품귀현상과 투자는 대규모 서비스(ChatGPT, Claude)의 등장과 초거대 동시 AI 서비스를 글로벌 단위로 영위하게 때문에 필요, NPU는 학습에 사용할 수 없을 뿐만 아니라, 현재 엔비디아, AMD를 비롯한 기존 GPU가 지원하는 산업군에서도 아주 특정한 산업 혹은 목적에서만 사용 가능하기 때문에 태생적으로 한계 명확, 그럼에도 대한민국의 성장이 반도체 기반이기 때문에 앞으로의 성장을 과대평가하여 일부 소수 기업에게 지나친 기대를 하고 있다 생각 • 독자 AI 모델의 활용처에 대한 논의가 더 필요 • 독자AI 모델 개발이나 특화 모델로 이어지는 구조를 국가 과제로 출범했다는 점에서 의미가 있으나, 후속 과제들과의 정합성이 떨어지고 지원규모가 크지 않다는 점은 한계

정책목표	주요 의견
기술주권	<ul style="list-style-type: none"> • K-AI 사업은 대규모 언어모델(LLM) 개발에 집중하는 것으로 보여 이미지 생성이나 영상 기반 멀티모달 모델, 온디바이스 환경 등을 고려한 경량화 전략 등 기술전략의 다양성 강화 등이 필요 • AI모델보다 훨씬 더 주권을 확보하기 어려운 분야는 AI 시스템SW/클라우드 및 커넥티드AI인프라(6G 포함) 기술인데, 이에 대한 투자 비전은 불명확 • AI 3대 강국 도약을 위한 AI 고속도로 구축 <네트워크>에서 6G 상용화보다는 AI 데이터센터 내부 CXL·광모듈·이더넷·InfiniBand 등의 내용이 들어가야 함 • 시스템SW에 대한 부분 추가 필요(분산학습, 모델 최적화 등) • 전력 인프라 확보에 대한 내용 추가 필요
기술주도 성장	<ul style="list-style-type: none"> • 벤처 투자의 경우 현재의 분위기나 정책적 드라이브에 대한 기대감은 충분히 긍정적. 해외 진출 및 Open Innovation 등에 대해서는 지금도 지원이 더 많아지는 추세긴 하나, 더욱 적극적인 지원이 있을 경우 큰 성과를 기대할 수 있을 것 • 인공지능 분야별 예산의 쏠림 현상. 인공지능은 LLM뿐만 아니라 각 산업별 및 분야별 AI 도입 필요 • 기술주도성장은 명확한 방향성 없이 개별 기관, 학계에 맡기고 지원해주는 구조가 최선이라고 보기 어려울 듯함 • 산업계의 뒷받침 없이 무분별한 인력 양성 사업으로 인한 인력 공급 chain 열화 가능성 • PIM이라는 특정 기술 및 방법론이 왜 언급되어있는지 모르겠음. 시장성이 검증되지 않았을 뿐 아니라 대안 기술 중 하나인데 특정한 기술을 목표로 하기보다 달성하려는 목적과 방향을 정책에 담아야하지 않나 생각
모두의 AI	<ul style="list-style-type: none"> • 실행 측면에서, 너무 지방, 지역, 특정 거점 등으로 나누어 진행되거나, 유사/동일 사업을 여기저기서 따로 실행하는 것은 리스크, 국가 전체의 역량을 집중하면 경쟁가능하나 우왕좌왕하는 사이에 타이밍을 놓치고 있다는 점이 염려. • 지역 균형 발전 등이 무리하게 엮이는 경향이 있음. 국가 경쟁력 관점과 기술 트렌드를 고려하여 시급한 과제부터 하나씩 진행해야 함 • 모두를 위한다는 것은 모두하지 않겠다는 것과 같은 역설적 효과 창출. “모두를 위하여”라는 프레임처럼 얇고 넓게 지원하는 정책은 지양 필요. 선택과 집중을 통하여, 주력산업과, 사회 필수적인 산업을 집중육성 필요, 국가의 정책은 마중물 역할이지, A to Z까지 전부 지원하는 것은 바람직하지 않음 • 모두가 사용할만한 서비스를 개발하는데 투자해야하고, 모두를 위한 AI 교육이나 홍보 등은 예산 낭비라 생각. • 모두의 AI는 모든 사람이 AI를 잘 알필요는 없다 생각. 그것보다는 AI를 잘 접목시킨 시스템 하나를 잘 개발하고, 이를 사용하는 법을 잘 교육하는 방법으로 가는 것이 좀더 낫다고 생각 • 모두의 AI에 공공에 대한 내용은 자세한 반면, 산업 AI 전환에 대한 내용은 상대적으로 구체성이 떨어져보임 • 실제 교육과 적응의 대상들이 수용력이 있는지와 어떤 형태로 역량을 상향평준화할 것인지에 대한 방향성 미흡

정책목표	주요 의견
기타	<ul style="list-style-type: none"> • 기술주권 및 기술주도성 확보를 위해 AI 모든 영역의 역량 확보 시도는 비현실적 • 선택과 집중이 필요한데, 차세대 응용 기술을 선점하는 전략이 더 효율적. 특정 분야(응용)에서 기술을 선점하고 새 경쟁 시장에서 주도권을 잡는 전략이 비용 측면에서 합리적 • 인프라·하드웨어 중심 강하고, 서비스·응용·지역분산·포용 측면이 상대적으로 약한 것으로 판단 • AI 적용의 전체 레이어를 고민해보고 각각에 대한 지원도 고민 필요 • 모든 정책은 단기적이고 응용 투자에 집중. 국내 AI 생태계 마련을 위해서는 기초연구-R&D-실증-공공도입-민간확산 등 선순환 구조가 되어야 하지만 현재는 연계성이 매우 낮음 • 우리나라가 집중해야 하는 것은 1) 미국, 중국 등의 AI 기술에 이은 fast follower 정책을 취하고, 2) 미국, 중국의 AI 기술을 우리의 온디바이스 및 클라우드 시스템에 잘 적용하여 다양한 AI 서비스에 활용하는 것, 3) 사람들이 AI 를 어떻게 활용할 것인지를 너무 깊이 고민하지 않는 것이라고 생각 • AI정책실행단계에서는 부처별 사업·과제가 서로 중복 또는 단절 현상도 나타나고 있으며, 정책 조정 거버넌스 측면에서도 아직 “컨트롤타워” 역할과 부처 간 집행 연계에 대한 개선 필요

2. 향후 AI 정책방향 및 과제

가. 향후 AI 정책방향

우리 정부의 AI 정책방향에 대해서는 ‘정부는 우리가 경쟁가능한 핵심 기술 스택에 자원을 집중하고, AI 인프라 경쟁력 확보를 위한 민간 차원의 투자 촉진·지원’해야한다는 의견에 가장 높은 동의수준을 보였다(5점 척도에서 4.32). 우리가 집중해야할 핵심 기술스택으로는 GPU/NPU, 인공지능 반도체, 시스템SW 등 반도체 관련 항목이 가장 많이 언급되었고, 다음으로 물리적 AI, 제조 및 로보틱스, 온디바이스 LLM 등 물리적 AI 관련 항목이 많았으며 그 밖에도 산업특화 AI, 응용SW, 에이전트, 데이터, AI 모델, 인프라, 소비자 AI 등이 언급되었다. 이외에도 우리 정부의 AI 정책방향으로 ‘AI 병목 해소(경량화 등), 미래 AI 생태계 재편 등에 기여할 핵심 기술개발을 위한 도전적 R&D, 핵심 인력 양성 등에 정부 자원을 집중’(4.08), ‘GPU·NPU 등 핵심 자원의 공급망 다변화 및 리스크 대응 전략 수립’(4.03), ‘기술주권 확보는 기술추격에 중점을 두고, 아직 선점되지 않은 차세대

AI 시장에서의 경쟁력 확보, 핵심 자원 기술력 확보에 국가 자원을 집중'(4.0) 등이 높은 수준의 동의를 얻은 것으로 나타났다.

〈표 4-8〉 AI 정책방향: AI 전문가 설문결과

정책방향	동의 수준
정부는 우리가 경쟁가능한 핵심 기술 스택에 자원을 집중하고, AI 인프라 경쟁력 확보를 위한 민간 차원의 투자 촉진·지원	4.32
AI 병목 해소(경량화 등), 미래 AI 생태계 재편 등에 기여할 핵심 기술개발을 위한 도전적 R&D, 핵심 인력 양성 등에 정부 자원을 집중	4.08
GPU·NPU 등 핵심 자원의 공급망 다변화 및 리스크 대응 전략 수립	4.03
기술주권 확보는 기술추격에 중점을 두고, 아직 선점되지 않은 차세대 AI 시장에서의 경쟁력 확보, 핵심 자원 기술력 확보에 국가 자원을 집중	4.00
미국과의 협력을 강화하여 미국 중심 공급망에서의 우리의 입지를 강화·확대	3.89
우리 기술력에 기반한 AI 생태계를 연결·확장해 나가기 위해 기술 스택간 수직·수평적 상호운용성 확보를 위한 국내외적 노력 강화	3.82
중장기적 관점에서 중국과 같이 AI 풀스택에 걸친 기술자립·독자생태계 구축	3.55
기술주권 확보는 기술추격에 중점을 두고, 사회경제 전반의 AI 대전환을 촉진함으로써 국가성장을 견인	3.32
EU처럼 공공 영역을 중심으로 국내 클라우드 + 오픈소스 기반 SW 스택을 활용한 개방형 생태계를 조성하여 풀스택에 걸친 대체 공급망 구축	3.16
글로벌 정합성을 갖춘 규범을 통해 미국 빅테크에 대한 통제권 강화	2.53

주: 5점척도(1 매우 낮음, 2 낮은 편, 3 보통, 4 높은 편, 5 매우 높음) 평가결과

기타 의견으로는 ‘최소한의 통제, 준수를 위한 규정 정비 및 미국 등 해외 빅테크에 대한 견제 장치 마련이 필요하나 폐쇄적 방향이 아니라 오히려 빅테크들이 한국에서 AI 사업하고 서비스 공급하기 좋다는 인식을 갖도록 진흥 여건을 열어주는 방향 필요. 빅테크와의 연계, 협업, 경쟁 등을 거쳐 국내 AI 기술 및 서비스 수준도 높아지고, 글로벌 진출도 호환성, 레퍼런스, 사용성, 기술력 등 고려 시 더 원활해질 것으로 기대, 국내 클라우드를 고집하는 것은 EU처럼 실패한 정책이 될 것’, ‘AI 전담부처 설립 추진’, ‘국가 AI 서비스 인프라 벤치마크(인프라 경쟁력 표

준) 개발을 통한 민간투자/경쟁 강화' 등이 제시되었다.

나. AI 경쟁력 강화를 위한 정부과제

AI 경쟁력 강화를 위한 정부 과제로는 'AI 핵심 인재 양성', '기술발전 속도에 맞춘 법제도 설계', '산업특화 AI 개발·활용 지원'이 중요하고도 시급한 과제로 선정되었으며 뒤를 이어 'AI 인프라 혁신, AI 기술한계 극복에 기여하는 도전적 R&D 지원', '차세대 AI 반도체, 고효율·저전력 컴퓨팅 기술 개발 지원', '다양한 산업에서의 AI 전환을 이끌 테크 스타트업의 등장, 성장 지원', '규제 샌드박스 등 혁신 촉진형 제도설계' 등이 선정되었다.

인재양성과 관련된 의견으로는 '산업계가 마련되지 않은 상태에서 대규모 인력 양성 사업 중단(실업자 양산 및 이공계 기피 심화)', 'AI 시스템 핵심 인재 양성', '숫자보단 정말 똑똑한 사람들의 유입 필요', '민간 기업 등의 AI 전문가 등이 대학에서 강의·연구를 병행할 수 있도록 겸직 규제 완화 및 대학 내 AI 전문가에 대한 우수한 연구 환경 제공', '응용 및 산업 적용도 인재 양성 및 재직자 교육 등이 필요' 등의 의견이 제시되었다.

'기술발전 속도에 맞춘 법제도 설계'와 관련하여서는 민간의 자율활동 보장에서 혁신이 출현한다는 점을 강조하였으며, '산업특화 AI 개발·활용 지원' 관련하여서는 '생성형 AI에서 에이전틱 AI로 진화함에 따라 학습데이터도 인터넷 공개 데이터에서 기업 내 업무 데이터로 변화. 산업별 AI 적용을 통해 산업 특화 Agentic AI를 개발하고 이를 확산하려는 추세로 국내 AI 모델, 인프라, 인력에 기반하여 산업 적용 사례를 확보하고 이를 해외로 수출하는 전략이 필요하며, 이를 가속화하기 위해 국산 AI 모델 활용도를 높이는 정책 필요(API 가격 보조금 지급 등)', '스타트업 중심 산업 특화 AI(AIaaS 포함) 개발 지원', 이를 위해 고용 유연화 및 정부 지원 필요, '한국형 온톨로지를 통해 산업별 데이터와 지식을 표준화하고, 국내 스타트업이 빠르게 도메인 특화 AI를 개발할 수 있도록 Playground 제공' 등의 의견이 제시되었다.

'AI 인프라 혁신, AI 기술한계 극복에 기여하는 도전적 R&D 지원'은 'LLM 및

딥러닝뿐 아니라 다양한 AI 기술에 대한 학문적 접근 촉진 및 꾸준한 육성', 'AI 기술은 이미 미국, 중국 등에서 고도화되고 있어 AI 기술 자체를 고도화하는 R&D 정책 지양(중복투자 방지차원), R&D정책은 AI 기술의 산업 영역 적용에 투자하는 것이 바람직', '일회성이 아닌 연속 지원해서 스킬업 필요', '독자 파운데이션 모델 개발이나 모두의 AI보다는 새로운 핵심 기술 분야 개척과 우수 기술인력 확보가 더 중요하고 시급' 등의 의견이 제시되었다.

'공공시장 개방'은 '현재까지의 수준, 속도를 넘어 확실한 개방 필요. 국내 클라우드, 모델에 한정되지 말고 글로벌 진출 고려. 국내 인프라와 규제에 한정되면 개방이 아니라 갈라파고스일 뿐, 대규모 공공 프로젝트 시, 스타트업 등 새로운 플레이어가 확실한 롤을 가져갈 수 있는 구조로 진행하거나 대규모 사업과 동시에 소규모, PoC 등 많은 기회를 함께 제공' 등의 의견이 있었다. 그 밖에도 '따라가기보다 오히려, 온디바이스용 칩에서 AI모델 최적 지원을 위한 자원 최적화 중요. 삼성전자 등 엔드단말 회사들과 함께 온디바이스에서 AI 모델을 잘 활용하여 좋은 서비스를 구현하는 방향으로 집중 투자. 이를 위해 다양한 폼팩터 AI 서비스의 사용자 경험성 향상 지원 및 다양한 온디바이스AI 서비스용 데이터 수집 등 중요', '규제에 구애받지 않고, 국제 최고 수준의 AI 서비스/제품/인프라 기술을 plug and play 할 수 있는 실증 편의성을 갖춘 실증 도시(단지보다 대규모) 필요' 등의 의견이 있었다.

〈표 4-9〉 AI 경쟁력 강화를 위한 핵심 과제: AI 전문가 설문결과

평가항목	중요도	시급성	실행 난이도
AI 핵심 인재 양성	4.79	4.58	4.18
기술발전 속도에 맞춘 법제도 설계	4.50	4.34	3.55
산업특화 AI 개발·활용 지원	4.45	4.29	3.50
AI 인프라 혁신, AI 기술한계 극복에 기여하는 도전적 R&D 지원	4.37	4.03	3.66
차세대 AI 반도체, 고효율·저전력 컴퓨팅 기술 개발 지원	4.32	4.03	4.16
다양한 산업에서의 AI 전환을 이끌 테크 스타트업의 등장, 성장 지원	4.11	4.00	3.53
규제 샌드박스 등 혁신 촉진형 제도설계	4.08	4.03	3.11
GPU 중심의 학습 인프라와 NPU 중심의 추론·엡지 인프라의 병행 육성	3.89	3.74	3.71
부처 간 AI 정책 조정체계 강화	3.87	3.79	3.53
독자 파운데이션 모델 시장성 확보 방안 및 후속 지원체계 수립	3.87	3.61	4.26
컴퓨팅자원, 공공데이터 등 기반 인프라를 확충하고 학계, 중소·벤처의 활용 지원	3.84	3.84	2.61
청정 에너지 및 데이터센터 인프라 확보	3.82	3.74	3.66
차세대 AI 시장 선점을 위한 엡지, 온디바이스용 칩, 모델 최적화 지원	3.82	3.47	3.71
공공 시장을 적극 개방하고 기업들이 end-to-end 서비스 설계, 실증, 협업 경험 등을 축적할 수 있도록 대규모 공공 프로젝트 기획·추진	3.63	3.45	3.16
주요 기술변곡점 대응을 위한 기업들의 풀스택 협력모델을 선정, 공동 R&D, 표준대응 등을 지원	3.55	3.11	3.50
다양한 수준의 이용자를 대상으로한 AI 도구, 서비스 산업(AIaaS) 육성	3.53	3.39	3.39
산업현장 재교육 및 업스킬링(upskillng) 지원	3.50	3.32	3.00
AI 윤리·신뢰·보안 기반 강화	3.42	3.03	3.22
기술, API, 거버넌스 등의 표준 수립	3.39	3.18	3.47
고속 네트워크·패브릭 자립화	3.30	3.05	3.59

주: 5점척도(1 매우 낮음, 2 낮은 편, 3 보통, 4 높은 편, 5 매우 높음) 평가결과

또한, 객관식 문항이 아닌 주관식 문항으로 가장 시급한 과제가 무엇인지를 질

문하여 얻은 결과는 다음과 같다.

〈표 4-10〉 AI 경쟁력 강화를 위해 가장 중요하고도 시급한 과제: 전문가 응답결과

구분	응답
킬러 솔루션 /수요창출 /활용 중심 전략	<ul style="list-style-type: none"> • 인프라가 아닌 솔루션, 모델, 데이터 중심 성장어어야함, 한국 기업들(특히 IT)은 원천 기술 개발/고도화에 특화된 사업자가 아닌, 기존 기술을 이용하여 돈을 버는 것에 치중되어 있어 원천 기술(모델, NPU 등)에 투자를 하더라도 실질적으로 성과가 나기는 어려울 것. 오히려 원천 기술을 활용하여 다양한 산업계에서 활용할 수 있는 솔루션 개발, 산업계에서 활용할 수 있는 데이터 구축에 더 많은 투자가 이루어질 필요
	<ul style="list-style-type: none"> • AI를 활용하여, 어떤 서비스를 구체적으로 어떻게 만들지, 혹은 AI를 활용한 서비스를 만드는 생태계 구축을 어떻게 가속화할 수 있을지에 관련된 논의가 더 되어야 된다고 생각
	<ul style="list-style-type: none"> • AI 시장에 대한 수요를 키워야 함. 민간에서 AI 시장 수요를 키운다면 정부는 이에 대해 집중 지원, 다른 산업에서 다른 방식으로 AI를 적용하는 시장을 발견, 확대하는 조직에게 적극적으로 인센티브를 제공하면서 AI 수요 자체를 키워나가고 민간도 다양한 시장에 도전할 수 있도록 지원 필요
	<ul style="list-style-type: none"> • 단순한 기술 우월성을 넘어, 사람들의 삶을 근본적으로 변화시키는 ‘꼭 필요한 솔루션(Killer Application)’ 이 등장해야만 기술은 역사의 전환점으로 기록, 국내 스타트업 기업들이 AI를 활용한 ‘꼭 필요한 솔루션(Killer Application) 또는 꼭 필요한 AI Agent’를 개발할 수 있는 토대 확보 필요
	<ul style="list-style-type: none"> • 국민 체감형 AI서비스를 통한 전 국민의 실질적 AI체감
	<ul style="list-style-type: none"> • AI 반도체, 인프라, 서비스까지 모든 영역에서 실제 돈이 되는 사업과 기업을 발굴하고 키우는 것이 가장 중요
	<ul style="list-style-type: none"> • 옛지AI 관점에서 수요 창출이 가장 중요. 아직도 옛지 시장에서 AI는 비전을 넘어서지 못하고 있고, 수요 창출이 더딘 상황. 이 부분에 대한 고민 필요. 수요가 늘어야 공급이 늘어난다고 생각하기에, 공급에 대해서 많이 고민하는 것보다 수요에 대해 고민하고 이에 대한 경험을 축적하게 되면 자연스럽게 경쟁력 확보도 따라올 것
제조AI/AIX	<ul style="list-style-type: none"> • AI 소프트웨어 및 서비스 생태계 구축, • 제조업은 전체 GDP의 30%이상 차지하고, 세계적 경쟁력을 갖춘 산업이므로, AI기술의 산업적용을 빠르게 지원. AI자체의 기술향상지원, AI인력양성도 중요하나, 노동기능인구 부족으로 무너지가는 제조업을 살리는 것이 가장 시급. 미국중국 등 선진국의 AI기술을 도입하되, 국산화, 현지화를 지원함으로써, 국내제조업이 자생적으로 활용할 수 있는 생태계 조성. R&D를 적극 지원하면서, 관련 법제도 보완

구분	응답
제조AI/AIX	<ul style="list-style-type: none"> • 우리나라는 제조업이 강한 나라로, 이를 AI와 결합하는 것이 AI 3대 강국 도약의 가장 현실적이고 빠른 길. 경쟁력있는 스타트업/중소기업도 쉽게 참여할 수 있는 실증공장과 데이터 공유 생태계를 만들고, 현장 중심 AI 인재를 키운다면 산업 경쟁력과 혁신이 함께 성장할 것
	<ul style="list-style-type: none"> • AI가 우리 사회에 체감가능한 실질적 이득을 주려면, 수많은 제조업 섹터와 AI의 유기적 결합이 필요하고, 일부 그룹이 개발하는 독자AI모델 등이 이러한 광범위한 효과와 연결될 지는 미지수. 결국, 다양한 제조업 섹터 지원, 이를 뒷받침할 수 있는 대규모 AI연구개발인력양성이 필요
	<ul style="list-style-type: none"> • AI는 우리나라가 기술력 및 테스트베드 환경에 있어서 유리한 고지를 선점할 수 있는 것이 크게 있지 않으므로 fast follower 전략이 맞다고 보며, 주요 AI 모델개발 및 알고리즘이 아니라 이를 활용한 온디바이스 LLM 혹은 클라우드 LLM 등의 기술을 실제 제조 현장에 적용하는 기술이 필요, 물론 이를 위해서는 네트워크 및 데이터 인프라도 함께 받쳐줘야 함. 굳이 미국과 같이 대량 학습을 위해 값비싼 GPU를 엄청나게 많이 사들이는 것 보다 오히려 온디바이스의 엣지 방향으로 가서 네트워크와 함께 기술력을 향상시키는 방향이 더 좋은 방향이고, 시급하게 적용해야 하는 방향이라 생각
	<ul style="list-style-type: none"> • AI는 시장이 있어야하고 수요가 있어야하는 산업. 한국의 제조업 역량을 기반으로 제조·물류·에너지·의료 등 AI활용 전방산업의 연계 필요 • AI 독자 모델 및 AI 반도체 확보를 위한 지속적이고 장기적인 투자가 이루어져야 하며 의료, 자율자동차, 법률 등 AX 확산을 위한 법제도의 정비가 이루어져야 함. 특히대규모 데이터 확보를 위한 개인정보 보호 및 사이버 보안에 대한 제도 개선 시급
국산AI 모델 개발/활용/확산	<ul style="list-style-type: none"> • 독자 AI 모델 개발이 가장 시급한 과제로, 국내 AI 활용에 있어 해외 의존도를 근본적으로 개선시킬 것, 이를 바탕으로 기술주도성장, 모두의AI가 실질적으로 구현될 것, 다만 독자 AI 모델이 국내 소수의 기업에게만 혜택이 돌아가는 것을 지양하는 방안 마련 필요
	<ul style="list-style-type: none"> • 정책 목표가 유의미한 수준의 소비러 AI 확립과 이를 기반으로 한 국가 전반에 AI 응용 및 확산이라면, 지금 당장은 모두의 AI라는 국정과제 목표에 자원을 낭비하기 보다는 미국, 중국의 최신 LLM 모델 성능에 버금가는 국산 LLM 확보를 최우선 과제로 생각하고 모든 자원을 집중, 이를 위한 합리적 보상 체계를 토대로 인재 영입이 진행되는 것이 우선. catch up 방식은 중국의 방식을 모사하는 것이 합리적
	<ul style="list-style-type: none"> • 국산 AI 모델의 활용/확산 중요, 중국이 딥시크 모델을 만들었다고 주목을 받았지만, 더 중요한 것은 이를 활용하는 자국 기업, 기관에게 많은 지원을 하였음. 현재 해외 모델의 API 가격이 매우 낮음. 이는 규모의 경제에 기반한 것이기도 한데, 해외 모델이 성능도 좋고 가격도 낮아서 모두가 선택할 수밖에 없음, 국산 모델의 성능은 점점 따라가고 있는데, 초기 API 가격을 낮춰주어 수요가 창출되면 규모의 경제가 실현되어 점점 더 많이 쓰게 될 것임. 우

구분	응답
국산AI 모델 개발/활용/확산	<p>리나라에서 먼저 국내 모델, 반도체, 데이터센터 등을 활용해야 해외에서도 사용하게 될 것</p> <ul style="list-style-type: none"> • 현재 시점의 아젠다인 독자 AI 모델 등에 대해서는 더 열린 지원 필요
인재양성	<ul style="list-style-type: none"> • AI전문인력의 육성 및 해외탈출 막을 수 있는 정책과 전략 • 글로벌 최상위 핵심 인재 확보 + 이들이 생존하고 활약할 수 있는 혁신 생태계 조성 <ul style="list-style-type: none"> - AI 기술전장은 최상위 인재들이 주도 - AI 시장의혁신이나 글로벌표준경쟁은 민첩한 테크 스타트업에서 등장. 시장에서 가장 먼저, 가장 잘 작동하는 기술을 개발하고 이를 오픈소스로 풀어버리면 표준이 됨 - 뛰어난 인재가 있어도,이들이 경제적 보상이 안정적인 대기업으로 대부분 취직하여 민첩함을 잃어버리는 사례가 많음. 스타트업으로는 기대할 수 있는 보상안정성이 낮고 극소수 인원을 제외하면 한계가 뚜렷 - 이들을 위한 컴퓨팅 자원과 불확실성 해소가 뒷받침 되지 않으면, 어렵게 확보한 인재와 아이디어가 글로벌 시장에서 경쟁할 좋은 시기를 놓치게될 수밖에 없음 • 인력양성. 특히 전산학 자체에 대한 근본적 이해가 있는 인력 필요 • 대학/대학원 지원을 통해 우수 인력이 인공지능 분야를 전공할 수 있도록 유도 • 우수 인재가 걸 맞는 대우를 받을 수 있는 보상현실화와 사회적 공감대 필요 • AI 모델 및 서비스 개발 및 운용할 핵심 인력 양성
소버린AI/AI 시스템SW 기술/AI 인프라/혁신적 연구	<ul style="list-style-type: none"> • 소버린 AI를 위한 AI 컨소시엄이 제일 시급. 개인, 스타트업이 할 수 없는 분야이고, 국가적으로 AI 경쟁력 및 경제 효과를 일으키는데 제일 중요하고 시급, 잘 짜여진 거대 모델은 이후 파생, 적용시킬 수 있는 부분들이 무궁무진. 탄탄한 제조업, 반도체, 헬스케어 등 여러 분야에 우리 AI를 적용하여, 작업 효율성을 올리고 신규 기술개발에까지 AI를 사용하게 되면, 앞으로도 경쟁력을 충분히 유지할 수 있을 것 • 시스템 S/W, 추론/학습 프레임워크 분야의 경쟁력 확보, AI반도체, AI모델 일변도 지원이 아니라 전체 에코시스템을 고려한 시스템 S/W, 추론/학습 프레임워크 등의 지원 필요. HPC를 위한 AI데이터센터 네트워크 분야의 투자 필요. 네트워크 지원이 이동통신에 쏠려있음. • 컴퓨터 시스템 SW에 대한 핵심 기술 개발과 인재 개발 • 양질의 데이터 수집 및 관리, 사용 및 적용 • AI G3를 목표로 한다면, 현재 완전히 부재한 “AI 시스템” 기술 확보 필요. “오픈소스 시스템 SW를 활용하여 국내 AI 반도체 및 AI 네트워크 기술과 연동하여 고성능 AI 시스템”을 구성할 수 있다면 AI 시스템 기술을 확보한 것으로 인정할 수 있음.

구분	응답
소버린AI/AI 시스템SW 기술/ AI 인프라/ 혁신적 연구	<ul style="list-style-type: none"> • 특정 회사의 장비와 SW를 활용하여 AI 시스템을 “구동하는 것” 또는 “일부 수정하여 구동하는 것”은 기술주권 확보와 어떠한 관계도 없음. “널리 알려진 AI 파운데이션 모델을 가져와 알려진 방법론에 기반하여 AI 모델을 학습시킨 것” 또한 이와 마찬가지로 기술주권 확보와 관계가 없음. • AI 기술주권 확보는 본질적으로 들어가면 대규모 GPU 확보 이후 비로소 시도해 볼 수 있음. (현재는 large model에 연구 인프라 부재로, novel 아이디어를 검증해보기 어려워, 소규모 모델에 적용되는 소규모 아이디어들 만들 주로 연구) • AI 서비스 경쟁력은 같은 세대(유사수준지능)내에서는 반응 속도 ,tokens/s 등 체감 응답 성능에서 발생하는 경우가 많아 ,AI 시스템/AI 인프라 국제경쟁력이 AI 서비스 경쟁력과 직결. <hr/> <ul style="list-style-type: none"> • 최근 GPU 26만장 공급 뉴스가 있었으나, GPU·전력·데이터센터·대용량 데이터 인프라가 빠른 시일 내에 구축되어야함. 따라서, 관련 정책 내용도 보완 필요 • AI 고속도로 등 인프라 구축 <hr/> <ul style="list-style-type: none"> • AI 3강 실현을 위해서는 GPU 등의 계산자원 확보가 가장 시급. • 다만, 국가에서 투자를 충분히 하더라도 미국, 중국과 비교해 절대적으로 HW 인프라 부족. • 인프라 확보와 더불어 격차를 일거에 해소할 수 있는 혁신적 연구를 장려하는 정책, 제도, 문화가 정착되어야 목표를 달성하고 또한 그 이상을 바라볼 수 있음. 많은 정부 과제들이 단순히 기존 기술 대비 성능 향상 폭, 발표 논문 개수를 목표로 하는데, 이를 통해서도 혁신적 연구 수행 불가능. 평가 어려움 때문에 어쩔 수 없는 부분도 있지만 바뀌지 않는다면 연구 현장에서 혁신을 만들기는 요원
차세대 AI 원천기술 확보	<ul style="list-style-type: none"> • 미래 AI 판도를 흔들 차세대 AI 원천기술 확보 • 아직 AI의 혁명은 제대로 오지 않았다고 생각하며, 지금 LLM을 만든 구조가 아닌 다른 AI 모델이 추론, 멀티모달, 퍼지컬AI 측면에서 월등한 성능을 보일 것이기 때문
GPU/에너지 인프라 확충/배분	<ul style="list-style-type: none"> • 객관적 평가체계에 따라 효율성 및 경쟁력을 기준으로 국가 공통 AI 데이터 센터 자원을 배분해, 혁신성과 산업 파급력이 높은 주체가 우선 활용 (공정성 보다는 효율성)
국가자원의 통합 관리/ 성과공유 및 확산	<ul style="list-style-type: none"> • 국가적 차원에서 큰 전략적 그림을 잘 안배하고, 이 그림 안에서 각 기업, 기관, 지역이 담당할 부분에서는 또 각각의 특성을 고려한 최적의 사업을 추진할 수 있도록 자율성은 부여해야 됨 • 어디까지 자율적 영역이고, 어디까지 국가적 공통의 틀과 호환을 갖춰야 하는지를 구체적으로 정의 필요

구분	응답
<p>국가자원의 통합 관리/성과공유 및 확산</p>	<ul style="list-style-type: none"> • 공통의 가이드라인을 제시하고, 특정기관, 지역의 성과가 반드시 다른기관, 지역, 기업 나아가 글로벌 시장에도 확산될 수 있는지를 고려해사업을 추진 • AI를 잘 모르는 담당자들이 독자적으로 사업을 발주, 추진, 관리하는 것은 문제, 실패 사례에 대해서도 문책이 아니라 진지한 고찰과 반성을 바탕으로 다음에는 실패하지 않을 방법에 대해 공유할 기회를 주고, 성공 사례에 대해서는 결과적 성과만 보는 것이 아니라 그 성공이 있기까지에 과정과 실무자들의 노력이 어떤 것이 있었는 지를 실질적으로 공유, 전파할 수 있는 기회가 늘어나야 함. • 정부/민간의 다양한 노력으로 GPU, 전력, 네트워크, 데이터센터 인프라는 물리적으로 빠르게 확충되어 가고 있으나 이러한 자원들이 하나의 생태계로 효율적으로 작동하지 않음. 부처/기관/과제별 클러스터가 쪼개져 있고 기업·연구기관·지자체가 동일한 GPU/데이터를 중복 구축하고 있으며 데이터 이동·접속·보안·성능 모니터링이 표준화되어 있지 않음. 즉, 자원은 늘어나고 있는데 경쟁력은 분산되어가는 양상. 개별 산업 정책은 별개로 진행되더라도, 어렵게 확보된 AI인프라 자원의 관리만이라도 통합할 수 있는 컨트롤 타워의 구성이 필요 • 거액의 정부 예산 투입 과제의 경우, 그 과정과 결과를 투명하게 공개하여 국민이 활용할 수 있어야 함. 예를 들어서, foundation model 학습의 경우, 학습에 사용된 데이터와 학습 방법을 공개하여 국가의 구성원들이 이를 무상으로 활용할 수 있어야 함
<p>분야선별/자원 집중/속도감 있는 정책</p>	<ul style="list-style-type: none"> • 한정된 자원 재고 및 효율적 자원 배분 필요. 0~100까지 전부 자주권을 확립하려는 시도는 매우 한정된 자원(ex. GPU 26만장은 미국 대비 크게 부족, 예산 등 가용 자본 한계)으로 비효율적. 우방국(미국) 공급망 의존을 현실적으로 인정하되, 해당 파운데이션 환경을 기반으로 우리 실정에 맞는 응용, 차세대 경쟁산업에 집중해서 일부 분야라도 선점하는 전략이 합리적 • 미국, 중국 등에 비해서 기술적으로 후발주자임에도 검증의 검증을 거쳐 진행되는 지원 프로세스는 속도 측면에서 뒤쳐질 수밖에 없어, 국가 거버넌스 관점에서의 컨센서스와 함께 밀집되고 집중된 지원 필요 • 에이젠틱 AI, 피지컬 AI 등 근 미래 고성장이 예상되지만 과열되기 직전인 글로벌 핵심 AI 분야를 정의하고, 이 부분에 정부/대학/기업의 역량을 집중하는 선별적 기술주도 투자 필요
<p>end-to-end</p>	<ul style="list-style-type: none"> • 파운데이션 모델 개발에만 그치지 않고, 어떤 End-product로 상용화할 것인지에 대한 명확한 그림이 없는 상태에서는 실질적인 성과로 이어지기 어려움. 특히 피지컬 AI 시대에는 수요기업에서부터 로보틱스, 소재부품, ICT, AI 모델, 데이터 기업 등 가치사슬이 복잡해지기 때문에 이들을 원팀으로 묶어 구체적인 성과를 창출할 수 있도록 판을 깔아줘야 함.

구분	응답
현실판단 지표	<ul style="list-style-type: none"> • 실제 우리의 AI 경쟁력이 어느 정도인지 객관적으로 볼 수 있는 현실적인 판이 더 필요. 국가적 지원이 지속적으로 이루어지고 있다고 하나, 현실적으로 일부 GPU나 데이터 등을 지원하는 것 외에는 민간 영역에 기대는 것이 현실.
창업/투자 환경 개선	<ul style="list-style-type: none"> • 창업 촉진을 위해 실패시 법적 책임 완화 및 실패 후 재기 할 수 있는 방안 마련 (배임/횡령 등의 포괄적 적용 지양) • 스타트업 세제혜택 등 실질적인 사업 환경개선 • 국제적 기업유치를 위한 노동법 개정
기타	<ul style="list-style-type: none"> • AI기업을 위한 규제를 완화하고 세제 혜택을 주는 등 기업 투자 유치를 장려. 산업이 커지면 일자리도 늘어나고, 기업의 투자도 늘어날 것 • 물리적 AI나 산업별 특화 모델, NPU 등에 대해 국가 전략 자산관점에서의 검토와 지원, 이를 기반으로 한국 내 대기업과의 협업 등이 정착되어야 자생적인 AI생태계를 이루어나갈 수 있음 • 기존 정부출연 연구소의 연구능력 정상화(국가AI연구소 신설 재고 필요)

그 밖에도 AI 정책과 관련하여 전문가들이 제시한 기타 의견들은 다음과 같다.

〈표 4-11〉 기타 정책제언

구분	내용
원천기술 개발에 집중/원천성에 대해 정부 조직내 전문성 제고	<ul style="list-style-type: none"> • 코어AI, AI시스템, AI인프라 각각에 대한 “원천기술” 개발에 투자가 집중되어야 하며, 이를 실증할 수 있는 규제 샌드박스 실증 단지 운영 (인프라 메인テナンス + 지원인력)에 많은 투자가 집중되어야 함 • 컴퓨터 분야의 “원천성”은 세계 최초, 세계 최고성능 등 기계적인 기준이 아니라 해당 분야의 현존 “병목 (난제)” 중 얼마나 결정적인 것을 해결했는지에 집중해서 평가해야함. 컴퓨터 분야 “원천성”에 대한 정확한 이해없이 특정 분야/기술에 집중 투자하는 것은 매우 부적절하며, 현재의 대규모 AI 연구개발 투자 또한 비효율이 산재해있고, 균형성 부족 • 이는 기재부/혁신본부/국회 예산심의 과정의 “시스템 분야” 전문성/이해도 부재와도 연결되어 있는 문제로 따로 심도있게 다룰 필요.
반도체 산업 선도	<ul style="list-style-type: none"> • NPU 활용 생태계 구축을 위한 방안 마련: 하드웨어 칩 개발만으로는 CUDA 생태계 종속을 벗어나기 매우 어려움, 국내 NPU 활성화를 위해 AI 바우처 지원사업에서 따로 트렉을 만들어두긴 했지만 스타트업 입장에서는 NPU를 사용할 동기가 약함. 단순 GPU 지원과 병행하여,

구분	내용
반도체 산업 선도	<p>국산 NPU(퓨리오사, 리벨리온 등)를 기반으로 한 'NPU 클라우드 테스트베드 구축 필요. 국가 AI R&D 과제 및 AI 인재양성 사업에서, 국산 NPU 활용 및 관련 SW 스택(컴파일러, 커널, 프레임워크 최적화) 개발할 경우 강력한 경제적 인센티브를 주는 방안.</p> <ul style="list-style-type: none"> • 칩 개발 예산의 50% 이상을 이 SW 생태계 구축에 투입하여, 개발자들이 '실제로 쓸 수 있는' 환경을 마련, 다만 결국 NPU의 경우 인센티브가 없는 글로벌에서 경쟁이 필요하다보니 적절한 수준을 찾는 것은 필수적 • 미래 반도체 산업 선도를 위해서는 “신소재 기반 AI 반도체(뉴로모픽, 광기반, spintronics 등) 개발” 사업 신설이나 장기적인 추진전략 필요 • 조달청 '우수물품지정제도' 고시 개정을 통한 AI반도체(부품) 조달 정책 추진: 제4조, 제4조의 2 • HBM·패키징·NPU가모텔·시스템SW와 함께 'Tunkey프로젝트'로 최적화
전력	<ul style="list-style-type: none"> • AI 데이터센터 구축에 필수 인프라인 전력/입지/냉각/계통의 병목현상 해결정책 (예)AI데이터센터 특구를 연계한 전력계통선 확보 정책추진(지역선정등)
피지컬 AI	<ul style="list-style-type: none"> • 중국의 피지컬 AI 수준을 정확하게 파악하고 벤치마킹할 부분을 빠르게 받아들여야 함
신뢰기술의 전략자산화	<ul style="list-style-type: none"> • 'AI 신뢰(TRiSM)' 기술의 전략 자산화 및 공공 선도 도입: 신뢰가능한 AI Agent 및 모니터링 분야는 AI시대 데이터 주권과 기술 안보가 걸린 핵심 전략 분야로 해외 프레임워크/모니터링 솔루션이 결합하여 국내에 유입되면, 기업의 핵심 프롬프트, 데이터, Agent, 톨체인 로그가 해외 솔루션에 종속되거나 유출될 수 있음 • AI신뢰·위험·보안관리분야에 대한 집중지원: 단순한 프롬프트 공격과 같은 위협보다 AI가 가진 내재적 오류 모니터링 • 공공부문 및 금융/의료등 민감 분야에서 국내 솔루션 도입을 선도 필요
데이터 표준화/개방·유통체계 구축	<ul style="list-style-type: none"> • 데이터 개방·유통 체계의 실질적 운영 전략 수립/실행 필요, 공공기관이 보유 중인 대규모 데이터(의료, 교통, 환경, 행정 등)를 익명화·정형화하여 개방하고, 이를 AI 학습용 데이터셋으로 활용할 수 있도록 라이선스 체계 및 포맷 표준을 마련 • 민간 데이터 공급자를 위한 인센티브 제도를 도입하여, 중소 데이터 기업이나 스타트업이 데이터를 수집·가공해 유통 플랫폼에 올릴 수 있도록 해야할 것 • 데이터 유통 플랫폼에 대해서는 품질평가·신뢰성 검증체계를 운영하여, 학습데이터의 신뢰도를 확보해야 함 • '산업 데이터'의 전략적 표준화 및 주권형 플랫폼 구축: 한국이 강점을 가진 제조업(반도체, 자동차, 조선 등)은 막대한 현장 데이터를 보유하

구분	내용
데이터 표준화/ 개방·유통체계 구축	<p>고 있으나, 표준화되어 있지 않고 사일로에 갇혀있어 AI 학습에 활용하기 어려움. 잘 정제해서 버티컬 솔루션 형태로 만들어낼 수 있다면 충분한 경쟁력이 있을 것</p> <ul style="list-style-type: none"> • 산업분야(예: 배터리, 반도체장비, 로보틱스)별 AI학습용 표준 데이터 셋 구축 • 해당서비스에 모니터링 플랫폼 번들링하여 글로벌 솔루션 및 데이터 표준 주도
모두의AI	<ul style="list-style-type: none"> • AI 활용 확산을 위해 산업별/영역별 AI 모범사례 발굴/홍보/교육 • 공공AI 영역에서는 공공서비스를 대상으로 공공AI플랫폼을 구축하고 마켓플레이스를 개방해 누구나 참여할 수 있는 공공서비스 AI化 촉진 방안 마련
스타트업 생태계	<ul style="list-style-type: none"> • 국가적으로 AI를 성공시키기 위해선 퓨리오사AI 같은 성공적인 스타트업이 AI 스택별로 여러개가 나와야함 • 초기 스타트업이 대표의 지분을 보호하며 성장할 수 있도록 정부 지원
정부는 큰 방향성만 제시	<ul style="list-style-type: none"> • 다른 분야도 마찬가지지만, 특히 빠르게 변화하는 AI 분야는 정책 설계와 적용보다 더 빠르고 다양한 방향으로 발전. 정부에서 큰 방향성을 제시하고 인프라 구축에 힘써주고, 연구자들이 세부적인 연구 방향을 주체적으로 그리고 혁신적으로 계획하고 수행할 수 있으면 더욱 바람직
장기적이고 꾸준한 육성 전략 필요	<ul style="list-style-type: none"> • R&D 일관성 및 예측성 확보: 유행만 따르지 말고 꾸준한 지원(작은 규모라도 오랜 기간 연구자를 신뢰하며 한결 같은 지원) • AI는 기술이라기보다는 학문. 물론 LLM 및 CV는 최근 산업화 수준까지 진행되었으므로 산업 측면에서 다룰 필요는 있음. 하지만 LLM 및 CV 기술의 최근 발전이 AGI를 가능케 할 것인지에 대해서는 아직 consensus가 없음. 학문적 측면에서 장기적이고 꾸준한 육성 전략 필요.
지원방향/방식	<ul style="list-style-type: none"> • 미국, 중국과 똑같은 트랙으로 경쟁하기 보다는, 보다 실질적인 제조업/사회적 이득을 얻는 방향으로 정책을 세우고 세밀하게 지원 필요 • 정부 사업들에 대한 성과를 잘 평가하여, 효과가 좋은 사업들은 더욱 확장하는 순환 프레임워크 구성
거버넌스	<ul style="list-style-type: none"> • 자율주행, AI반도체, 피지컬AI 등 핵심 기술로 부상하면 과기부, 산업부, 중기부, 교육부 등다부처에서 중복적이고 파편적으로 R&D 및 인력양성 사업이 진행 • 적어도 R&D는 부처 중립적으로 추진할 수 있도록 기획 및 관리할 수 있는 체제 개편이 주체가 필요

구분	내용
거버넌스	<ul style="list-style-type: none"> • 정부 조직법 상 산업통상부는 제조 산업의 AX를 위하여 선도적인 정책 추진 중. 그러나, 과기부 중기부 모두 제조 AX 정책을 중복지원(피지컬AI 정책, 제조AI 정책)에 따른 기업간 혼란이 야기 • 산업통상부 내 주력 산업 업종과에서 관련 제조 역량을 키울 수 있도록 지원하는 것은, 정부 조직법 내 부처의 역할을 충실하고 있음. 특히 피지컬AI는 과기부가 아니라, 산업부가 하는게 맞고, 제조데이터 육성 관련 정책 및 법 제정 등을 하는 것은 중기부의 역할이 아니라, 과기부나 산업부 역할

현재 시점에서 기술주도 성장, 기술주권 대비 정책적 중요도는 다소 낮지만, 향후 AI 기술이 진화함에 따라 모두의 AI 정책의 중요도는 점차 올라갈 것이라는 점에서 모두의 AI 실현을 위한 주요 장애물과 실현 전략도 살펴보았다. 우선, 모두의 AI를 가로막는 주요 장애물은 ‘AI 전문인력 부족’, ‘고품질 데이터 확보의 어려움’, ‘AI 내재화 역량 부족’ 등으로 조사되었다.

[그림 4-8] 모두의 AI 실현의 주요 장애물: AI 전문가 설문결과
(단위: 명)



모두의 AI 실현을 위한 전략으로는 ‘공급역량 강화 및 생태계 육성’, ‘인프라 및 공공기반 확충’, ‘수요창출 및 확산 촉진’ 등이 중요한 것으로 평가되었다. ‘공급역량 강화 및 생태계 육성’과 ‘수요창출 및 확산 촉진’은 사실 수요-공급이 서로 맞물려 돌아간다는 점에서 어느 한쪽이 더 중요하다고 말하기 어려우며 이에 평가결과 상 큰 차이는 없었다. 다만, 공급 측면이 중요도에서 약간 더 높게 나타났으며 수요 측면이 시급성 측면에서 약간 더 높게 나타났으며, 산업 분야 응답자들은 ‘수요창출 및 확산 촉진’을 가장 중요하고 시급한 전략으로 평가하였다.

〈표 4-12〉 모두의 AI 실현을 위한 정책과제: AI 전문가 설문결과

	중요도	시급성	실행 난이도
공급 역량 강화 및 생태계 육성	4.05	3.76	3.92
인프라 및 공공기반 확충	4.03	3.79	3.61
수요 창출 및 확산 촉진	4.00	3.89	3.61
신뢰·표준 및 안전 기반 강화	3.71	3.45	3.45
인재 및 역량 확산	3.71	3.37	3.29

주: 5점척도(1 매우 낮음, 2 낮은 편, 3 보통, 4 높은 편, 5 매우 높음) 평가결과

다. 인재양성 방안

AI 인재양성은 중요하지만 매우 어려운 과제이다. AI 인재양성과 관련된 전문가들의 주요 의견을 정리해보면 다음과 같다. 첫째, AI 전문가들은 AI 인재 정책에서 글로벌 AI 생태계를 주도할 수 있는 최상위 인재를 양성하는 것이 필수적이며 이에 이들이 성장할 수 있는 환경을 조성하는 것이 가장 중요하고도 시급한 과제라고 강조하였다. 둘째, 빅테크 등 산업에서 AI 기술발전을 주도하고 있다는 점에서 최상위 인재육성은 반드시 대학에서만 이루어지지 않으며 이에 산업 전문가가 대학에서 인재육성에 참여할 수 있는 방안, 산업과 대학이 연계되어 인재를 육성할 수 있는 방안 등에 대해서도 고민이 필요하다고 언급하였다. 셋째, 소수의 최상위 핵심 인재 육성, 산업현장에서 AI 도입을 주도할 산업 인재 육성, 일반 국민들을 대상으로한 AI 리터러시 교육 등 AI 교육은 그 목적과 수요에 따라 철저

히 구분되어 추진될 필요가 있다고 강조하였다. 특히 최상위 핵심 AI 인력은 매우 소수의 인력이며, 산업인력 또한 AI로 기존 직무가 변화하고 있으며 업무 생산성이 높아지고 있어 산업수요가 뒷받침되지 않은 대규모 인력양성 사업은 지양되어야 한다고 언급하였다. AI 인력양성과 관련된 구체적인 전문가 의견은 아래 표와 같다.

〈표 4-13〉 AI 인재양성 방안: 전문가 의견

구분	내용
핵심 고급 인재가 성장할 수 있는 환경 조성	<ul style="list-style-type: none"> • 해외 우수인재에 대한 유치와 그를 중심으로 한 연구/실증 거버넌스 고도화 필요. 현재 AI 인력은 다수의 80점짜리보다 소수의 99점짜리가 밸류를 높게 가지는 구조이며, 산업계의 핵심 인재풀이 학계와 연구계의 실증 성과를 넘어서기 시작. 이에, 글로벌에서의 경쟁력을 지닌 인재의 회귀를 촉진하는 관점과 그를 기반으로 한 인재육성 관점에서 핵심 인재에 대한 국내 밀도를 올리는 것이 중요한 키가 될 것 • 인재는 성장할 수 있는 환경에서 근무하고자 하며, 성장할 수 있는 환경은 좋은 과제와 좋은 데이터, 좋은 동료가 있는 기업 및 기관 • Hero 과제와 Hero 기업을 많이 만들어야 함. 도전적인 연구를 하고 성과를 내는 곳에 자연스럽게 인재가 모이게 되어 있음. • 대기업, 중소기업, 스타트업 나누지 말고 잘 하는 곳에 집중 지원 • 폭넓은 지원이 필요한 경우도 있지만, 선택과 집중을 해야될 영역 존재. 즉, 많은 수의 인재양성이 어느 정도 필요하기는 하지만, 정말 수준 높은 연구/개발을 할 수 있는 인재 양성이 필수적. • 정부는 국내 출신 AI 전문가가 해외로 나가지 않도록 혹은 나가더라도 국내와 연계할 수 있도록 귀환 인센티브, 국내 스타트업 창업 지원, 공동연구 기회 제공 등을 강화해야 함, • 또한, 고급인재를 양성할 때 “세계수준의 연구-프로젝트 경험”을 제공해야할 것, AI는 빠르게 변화하는 영역이므로 국민대상 평생학습 및 재교육 시스템 강화 필요
핵심 고급 인재가 성장할 수 있는 환경 조성	<ul style="list-style-type: none"> • 글로벌 AI 기술 표준을 선점하고 생태계를 주도할 최상위 인재를 확보하고 이들이 활약할 환경 구성이 필요 • 숫자가 아닌 질적으로 최상위 인재 확보, SOTA 모델과 핵심 프레임워크는 소수 천재들이 주도, 논문이나 특허 같은 정량적 실적보다는 보상, GPU 등 자원에 대한 접근 확대 등이 필요

구분	내용
핵심 고급 인재가 성장할 수 있는 환경 조성	<ul style="list-style-type: none"> • 글로벌 오픈소스 생태계 기여를 핵심 성과로 인정, 다만 어느정도 표준화 된 분야에서의 실효성은 낮다고 생각, 주요 글로벌 AI 오픈소스 프로젝트의 활동을 SCI급 논문에 준하는 핵심 연구 실적으로 인정, R&D 자금 지원 • 혁신 스타트업 집중지원, 스톡옵션이 많이 유명무실해졌는데, 활성화 필요, 단순 자금 투자보다는 기업들과의 협업 기회 활성화 • AI 핵심 인력을 위한 AI 병역 특례 제도 신설
처우개선/우대정책/고용유연화	<ul style="list-style-type: none"> • AI 인재뿐만 아니라 첨단산업 전체 발전을 위해서 한국의 의대 선호 문화를 개선해야 함. 과학 기술자들이 의사와 같은 수준의 대우와 보상을 받을 수 있게 해야만 AI를 비롯한 첨단산업에 우수한 인재가 모여들 수 있을 것 • 기업 간 인재 유치를 위한 자율경쟁을 권고하고 공공영역에서도 직무/직함/자격인증 등으로 고위 발탁함으로써 경쟁력 있는 AI인재가 인정 받는 사회적 인식 제고가 필요, • 국내에서도 AI인재가 안정적으로 연구개발할 수 있도록 연구환경 조성 필요 (고용 안정성, 연구과제 지속성, 연구 자율성 등 보장), 즉, AI인재를 생산하는 것에서 그치는 것이 아닌 사후관리(처우개선) 환경 마련 • 학교는 이론과 실습, 논문 작성 등으로 인재를 양성하지만, 이러한 인재는 산업에서 필요한 인재와는 차이가 있음. 산업형 인재를 키우려면 회사원들이 무한 경쟁을 해야하고, 이를 위해선 적절한 보상 시스템과 고용 유연화 필수 • 고용 경직성 완화 및 능력에 맞는 보상 체계 확보 • 장기적으로 (비단 AI 뿐만 아니라) 연구자들이 국내에서 충분한 대우를 받고 자부심을 가지며 근무할 수 있는 환경 조성 필요. 현재 AI 인재양성 정책을 아무리 잘 세워서 수행한다 해도 결국 많은 수가 혹은 가장 우수한 인재들은 해외로 빠져나갈 것. • 급여도 문제지만 단순히 그것에만 국한된 문제는 아님. 이공계 인재들의 자긍심을 고취시킬 수 있는 문화, 우대 정책 등이 필요 • 고급 AI인력의 외국 유출을 막기 위해서는 한국에서 AI를 연구하거나 서비스를 해도 보상이 충분한 환경의 조성 필요. 이를 위해 다양한 스타트업 육성을 지원하고, 산업계와 일체화된 인재 육성 프로그램도 계획. • 대학에 우수한 AI전문가가 교수로 활동할 수 있도록 “AI 교수”의 처우 대폭 개선 (최소 대기업 임원급 급여 보장)
인재유치	<ul style="list-style-type: none"> • AI 인재 양성에 앞서 현재 풀에서 유의미한 인재 유치가 선행되어야 후행하는 인재양성이 실질적인 의미를 가질 것. 피상적인 인재보다는 실무에 즉각 투입될 수 있는 인재를 글로벌 AI 인재 시장이 자유경쟁시장이라는 점을 인지하고 합리적인 보상(한국의 인재 유출의 핵심 원인을 비합리적인 명목적/사회적 보상 체계에 있다고 봄)을 토대로 섭외

구분	내용
인재유치	<ul style="list-style-type: none"> • 민간기업 등의 AI 전문가 등이 대학에서 강의 및 연구를 병행할 수 있도록 겸직 규제 완화하고 대학 내 AI 전문가 등에 대한 우수한 연구 환경(컴퓨팅 인프라 등) 제공 • 국제 공동연구프로젝트 확대 • 해외 우수인재 유치 지원
목적/수준에 부합하도록 인재양성 체계 정비	<ul style="list-style-type: none"> • 1) AI 리터러시 교육, 2) 산업AI 인재 양성, 3) AI 고급인재 양성은 철저히 분리해야함. • AI 대학/대학원도 AI 최고급 인재 양성에 해당하지 않는 경우는 축소/폐지해야함. AI 대학/대학원 출신 학생들은 AI최고급 인재로 성장하지 않는 경우, 각 산업 도메인지식을 쌓고 AI를 추가로 배운 전통적 대학/대학원 학생들 대비 산업 AI 인재 역할에서 크게 뒤떨어짐. • 코어 AI 및 AI시스템 고급인재로 성장할 수 있는 학생은 대한민국 전체로 봤을 때 기껏해야 1년에 200명 내외일 것, 그 외의 학생들에 대해서는 도메인 지식 교육을 반드시 병행해야함. AI에 대한 지식만 가진 보통 수준의 학생/석박사들은 곧 공급 과잉임.
수요 창출/수요기반 교육	<ul style="list-style-type: none"> • 핵심 AI 인재와 현장 AI 인재를 분리해서 정책 설계 • 핵심 AI Scientist·AI Engineer·Chip/Model Co-design 인재 : 석·박사·전문대학원 • 현장(제조·물류·서비스·공공) : 단기 직업훈련, 실습 중심, 현장 문제 해결 기반
	<ul style="list-style-type: none"> • 많은 정책은 공급 중심. 인재 양성이라는 말처럼 공급에 초점을 맞추어서 AI에 맞는 인재를 길러낸다는 것인데, 이 또한 중요하지만 수요에 집중해야 한다 생각 • AI 시장 규모 자체를 키우고 그 수요가 증가한다면 자연스럽게 AI 시장에 참여하고자 하는 인재가 늘어나고 그에 따른 경쟁이 가속화되면서 AI 인재가 양성되고 그에 맞는 프로그램이 설립된다 생각. • 이는 마치 의과 대학에 대한 별다른 인재 양성 정책을 하지 않아도 그 수요가 분명하고 이득이 되기에 많은 인재가 몰리고, 그에 맞추어서 인재 양성 프로그램이 사교육에서 등장하는 것과 같다 생각
	<ul style="list-style-type: none"> • 산업계가 마련되지 않은 상태에서 대규모 인력 양성 사업 중단 (실업자 양산 및 이공계 기피 심화) • 단순히 특정 기간 교육을 거쳐 자격증이 부여된 실무자 대량 배출 지양
	<ul style="list-style-type: none"> • AX의 베스트 성과 창출(성공의 경험)이 중요. • 사례전파를 통한 기업 니즈기반의 교육 중요. • AI 기술 자체의 교육도 중요하나, AI기술의 활용 관점에서의 지원정책 중요

구분	내용
산업별 인재 육성	<ul style="list-style-type: none"> 인재양성에서 대학의 역할을 배제할 수는 없으나, 이미 빅테크들 비롯하여 대학의 중요성이 떨어진 것은 관찰되고 있는 상황. 산업별 인재 양성을 위한 아카데미 등의 경우 그런 관점에서 대학보다 더 앞으로 중요한 역할이 기대되는 만큼 더욱 체계적인 정책 설계와 실행이 필요.
산업 수요 기반 실전형 교육	<ul style="list-style-type: none"> 대학 등 정규 교육기관의 역량과 네트워크를 충분히 활용하되, 이러한 교육이 꼭 학사, 석사 등 학위과정에 국한되지 않아야 함. 반대로 8주 온라인 교육 등 너무 가벼운 수준의 교육에 예산이나 지원이 중복적으로 제공되어서도 안됨 또한 교육과정에 반드시 실제 기업, 산업 현장에 적용된 AI 실전 사례에 대한 부분도 중요한 축으로 포함되어야 함. 실제 AI 프로젝트를 수행한 담당자나 개발자, 혹은 AI 프로젝트의 결과물을 업무에 사용하는 사용자 등과 멘토링, 인터뷰, 프로젝트 리뷰 등이 교육에 필수적으로 일정 부분 이상 편성되어야 실질적 교육 효과가 있을 것. 대학생들이 학부 시절에 실질적인 AI 기업 업무를 경험해볼 수 있는 제도적 지원 단순 이론 교육에서 벗어나 국가 공통 AI 데이터센터를 이용하여 산업 현장과 연계된 실습 중심의 프로젝트형 교육을 통해 실무형 인재 양성 필요 지역별 공공립 대학중심의 AI실습 인프라지원 필요 대학에 의존한 단순 이론 교육과 정량적 확대만으로는 한계, 한국의 AI 인재 정책도 대학에 의존하기 보다 “산업 수요 기반 실전형 교육” 과 “생태계 연계형 인재 양성”으로 전환 필요. 파리의 ‘에콜 42(École 42)’와 같은 선진 모델을 벤치마킹해서, 국내 기업과의 긴밀한 협업을 통해 한국형 실용주의 AI 인재양성 시스템을 구축 검토 AI 관련 산업에 대해 산학 연계 활성화 방안이 필요하고 기업입장에서 신입사원에 대한 실무능력 향상에 시간과 경비를 절감 가능 AI관련 인프라를 활용하여 실습할 수 있는 교육프로그램 필요 (실전 프로젝트 기반 교육 과정 운영)
중소기업 인재 육성	<ul style="list-style-type: none"> 일부 과기원 등에 집중 지원하여 양성한 인력은 모두 대기업으로 흡수. 하지만, 대기업은 자본과 역량이 풍부하므로 오래 전부터 AI내재화에 많은 투자와 노력을 기울이고 있음. 정부가 실질적으로 지원해야 할 인력은 중소기업을 위한 인력으로 중소기업을 지원할 AI연구개발인력 양성을 위해 보다 폭넓은 영역, 레벨의 인력 양성이 필수적 AI를 직접 개발하는 기업만이 아니라, AI 도입을 고려하고 있으나 자체 개발 비용이 부족한 기업들을 위한 인력 양성도 충분히 이루어져야 함

구분	내용
융합 교육 /관련 프로그램 다양화	<ul style="list-style-type: none"> • 미국 주요 대학은 AI 학과가 아니고 기존 데이터 과학, 전자공학 등에서 기초 학문을 배우면서 융합하는 방식이나 국내는 트렌드한 기술이 있으면 우후죽순 격으로 특성화 학과가 생겨나 파편화되고 결국에 융합에 방해가 되는 결과가 나타남 • AI 반도체 설계와 AI SW의 학제간 분리 교육은 추격형에 머무르고 있는 국내의 관련 산업은 영원한 추격형에 머무를 수 밖에 없음, 예를 들어, AI알고리즘에 대해 제대로 파악해야 관련 알고리즘의 효율적인 계산을 위해 AI반도체 설계에 어떻게 반영할 것인가 고민이 이루어지고 이를 바탕으로 추격형에서 선도형으로 전환될 수 있음 • 꼭 AI 학과, 전공이 아니더라도 관련 컴퓨터 및 데이터사이언스 등 관련 대학원 학생들의 인력 양성 프로그램 다양화 및 확대 필요
인프라 자원의 효율적 배분	<ul style="list-style-type: none"> • AI 고속도로 구축으로 구성된 인프라 자원의 분배/활용 및 관리 효율화 필요. 희망자 모두에게 GPU/NPU 사용 혜택을 제공하되, 사용량과 성과를 평가하여 자원 회수/확대하는 최적 운영 전략 수립 필요.
민간자율성 부여	<ul style="list-style-type: none"> • 비단 AI뿐 아니라 인재양성은 자율성 중요. 정부는 많은 것을 민간에 맡기고 장기적 관점에서 산업적 생태계가 작동할 수 있도록 하는 등 민간은 하기 어려운 일을 해야 함
과도한 편중 지양	<ul style="list-style-type: none"> • 특정 학교 및 특정 세부 분야에 리소스가 과도하게 편중되는 것을 방지해야 함. 국가의 리소스에 한계가 있는데, 특정 학교가 과도하게 많은 교원을 충원하고, 학생을 선발하고, 과도하게 정부과제를 수주하는 것은 정부의 방향성과도 부합하지 않음.
통합적 거버넌스 필요	<ul style="list-style-type: none"> • 분절된 부처별 추진 체계의 통합 필요. 과기정통부(AI, SW대학원 중심), 산업부(산업별 전문인력), 중기부(SMB, 스타트업 인력), 고용부(전환교육), 교육부(기초·교원 양성) 등으로 나뉘어 구조적 중복·비연계가 발생

제 5 장 정책 제언

본 장에서는 앞서 제시한 연구결과를 종합하여 우리나라의 AI 기술주권과 국가경쟁력 제고를 위한 정책방향과 실행전략을 제언한다.

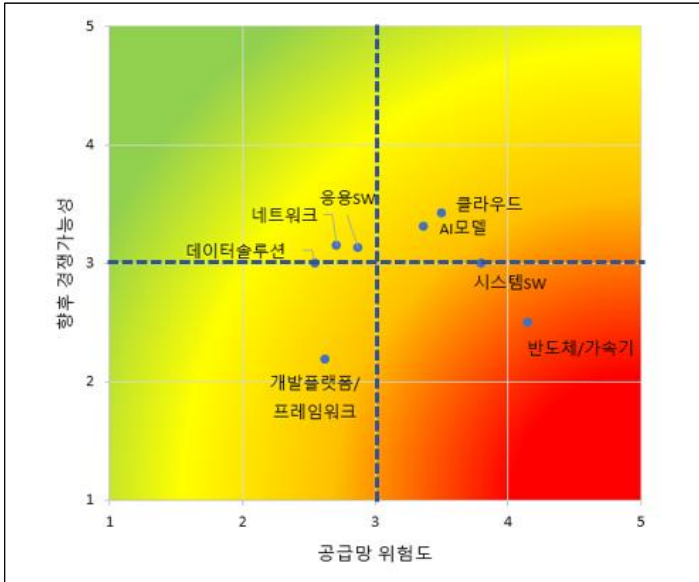
1. 정책방향

중국, EU 등 세계 주요국은 풀스택 관점에서 AI 경쟁력 및 자원 간 연결성을 강화·확장하고 공급망 전반에 걸친 통제권을 확보하는 등 각자의 방식으로 AI 기술주권 확립을 위한 노력을 추진 중이다. 이러한 사례에서 보듯이 AI 모든 기술스택에서의 기술 자립만이 AI 기술주권을 확보하는 유일한 방법은 아니다. 게다가 미국이 엄청난 기술격차를 과시하며 사실상 독점하고 있는 현재의 AI 생태계에서 그동안 AI 기술 스택들이 서로 유기적으로 연결, 결합되어 공진화해왔다는 점을 고려하면 풀스택 기술자립은 적어도 단기간에 도달가능한 목표는 아니다. 또한 강한 기술 종속성이 존재하는 상태에서 AI 풀스택 구축은 매우 조심스럽게 접근해야 할 필요가 있다. 만약 글로벌 시장과 동떨어진 독자 기술 스택을 무리하게 채택할 경우, 이는 결국 기술 활용의 효율성 저하와 시장 수요 제한이라는 문제로 귀결될 위험이 있기 때문이다. 특히 GPU 기반의 클라우드 AI와 범용 LLM의 학습 및 추론 영역에서 엔비디아 생태계 참여 없이는 경쟁력 확보가 어렵거나 뒤처질 위험이 크다는 점을 직시해야 한다. 이에 AI 전문가들 또한, 정부는 우리가 경쟁가능한 핵심 기술스택에 자원을 집중하여 AI 경쟁력을 확보해 나가야 한다고 강조한다. 즉, AI 주권확보의 목표는 완전한 자립 생태계라는 이상적 목표보다는, 핵심 기술 분야에서 외부 압력에 흔들리지 않는 자율성을 확보하는 ‘최소한의 종속과 최대한의 경쟁력’이 더 타당하다고 판단된다.

그렇다면 ‘최소한의 종속과 최대한의 경쟁력’을 위한 전략은 무엇인가? 우선,

AI 기술자립도를 진단한 결과를 공급망 위험도, 향후 경쟁가능성을 중심으로 재구성해보면 각 기술스택에 대한 정책적 우선순위를 가늠해 볼 수 있다.

[그림 5-1] AI 기술스택별 공급망 위험도 vs. 향후 경쟁가능성: AI 전문가 설문결과



주: 5점척도(1 매우 낮음, 2 낮은 편, 3 보통, 4 높은 편, 5 매우 높음) 평가결과

첫째, 공급망 위험도가 높으나 향후 우리의 경쟁가능성이 낮은 영역은 기술주권 측면에서 정부 주도로 공급망 위험 대응을 위한 중장기적 노력이 필요한 영역으로 반도체/가속기, 시스템SW가 이에 속한다. 둘째, 공급망 위험도, 향후 경쟁가능성 둘 다 높은 영역은 기술주권뿐 아니라 산업경쟁력 확보 차원에서 정부 주도의 집중 투자가 필요한 영역으로 AI 모델, 클라우드가 이에 속한다. 세 번째, 공급망 위험도는 상대적으로 낮고 향후 경쟁가능성은 높은 영역은 민간 주도로 경쟁력을 확보해 나가되 산업경쟁력 확보를 위한 정부의 지원이 필요한 영역으로 응용SW, 네트워크, 데이터인프라/솔루션이 이에 속한다. 마지막으로 공급망 위험도, 향후 경쟁가능성이 낮은 영역은 정책적 우선순위가 가장 낮은 영역으로 개발

플랫폼/프레임워크가 이에 속한다. 그러나 물론 글로벌 AI 생태계 상황은 언제든 바뀔 수 있다는 점에서 이 영역에 속한 기술스택에 대해서도 공급망 위험대응 측면에서 지속적인 모니터링은 필요하다. 다시 정리해보면, 단기적으로는 AI 모델, 클라우드를 포함하는 데이터·AI모델 학습 생태계에 대한 경쟁력 확보에 집중할 필요가 있으며 중장기적으로는 AI 공급망 위험도를 낮추고 우리 AI 생태계가 우리 기술에 기반한 반도체+시스템SW로 연결·확장되어갈 수 있도록 반도체/가속기, 시스템 SW에 대한 지속적인 개발 투자, 그리고 기술스택간 연동·확장을 위한 생태계 설계가 필요하다.

이러한 방향성은 현재 한국 정부가 취하고 있는 전략과 일면 맞닿아 있는 것처럼 보인다. 우리 정부는 GPU의 대량 확보를 통해 AI 인프라 성능을 높이고 글로벌 수준의 LLM 모델 개발을 지원하는 한편, NPU 개발과 국산 반도체 비중 확대 등 점진적 전략을 취하고 있다. 지난 APEC 회의 이후, 26만장 규모의 GPU를 확보하고, 국산 NPU를 포함한 국가 AI 컴퓨팅 센터 구축과 독자 파운데이션 모델 확보의 마중물을 마련하였다. 그러나, 아직까지 독자 파운데이션 모델 개발 사업을 통해 정부가 달성하고자 하는 목표가 무엇인지, 향후 어디까지를 레거시(Legacy) 기술 스택으로 활용하고 어디서부터를 우리가 독자적으로 가져갈 것인지에 대한 명확한 방향성, 로드맵은 잘 보이지 않는다. GPU 확보 및 독자 파운데이션 모델 개발 사업에 국가 차원에서의 막대한 투자가 이루어지고 있는 만큼 이러한 정부의 노력이 지속가능한 기술 생태계 구축의 마중물이 될 수 있도록 이어지는 내용에서는 제시한 정책방향을 구체화하는 실행전략에 대해 논의하도록 하겠다.

2. 실행전략⁴¹⁾

정부가 확보한 GPU가 AI 생태계의 기술적 자립과 혁신을 유도하고, 독자 파운데이션 모델이 우리 산업 전반에 지능을 공급하는 두뇌이자 우리 AI 산업의 점진적 성장을 이끄는 핵심 기반으로 기능할 수 있도록 본 연구에서는 1) 대규모 AI 모델에서의 글로벌 격차 해소 및 범용-도메인 모델이 상호발전하는 생태계 설계, 2) 독자 AI 모델과 NPU, HW-SW를 잇는 통합 경쟁력 확보, 3) AI 서비스 생태계 활성화로 AI 생태계의 수요기반 자생적 성장 기반 구축, 4) 우리의 전략산업과 연계된 ‘버티컬 AI’, ‘물리적 AI’로 AI를 진짜 산업으로 육성, 5) 혁신과 안전의 균형있는 성장을 위한 AI 거버넌스 강화, 6) AI 기술한계 극복에 기여하는 도전적 연구생태계 구축을 실행전략으로 제안한다.

1) 대규모 AI 모델에서의 글로벌 격차 해소 및 범용-도메인 모델이 상호발전하는 생태계 설계

오픈소스 기술의 발전, 훈련 레시피의 공유 등으로 AI 모델 성능은 빠르게 상향 평준화되고 있으며 이미 글로벌 시장은 AI 모델 경쟁에서 나아가 에이전틱 AI 경쟁으로 이동하고 있다. 이는 우리에게 기회이자 위협으로 이에 우리는 선도국 AI 모델 경쟁력과 격차를 빠르게 해소하고 우리 독자 AI 모델이 실제 산업에서 활용·확산될 수 있도록 기반을 구축할 필요가 있다.

이에 우선, 오픈소스 전략을 적극 강화할 필요가 있다. 국내 기업들이 개발한 고성능 모델의 오픈소스화를 장려하여 국내 AI 개발자 생태계를 활성화하고 기술 혁신을 가속화해야 한다. 이미 NAVER, SKT, KT, 카카오, LG AI연구원 등 국내 주요 기업들이 오픈소스 모델을 공개하고 있으나, 이러한 노력을 더욱 확대하고 체계화할 필요가 있다. 아울러 이들 모델과 데이터셋, 그리고 경량화 기법이 자유

41) 실행전략은 김성욱·이경선(2025), [특집 1회] 글로벌 AI 경쟁전략 및 대응방안: 풀스택(Full-stack)의 관점에서, KISDI Premium Report에 먼저 게재되었으며 보고서의 내용은 이를 추가적으로 수정·보완하여 작성됨

롭게 공유되고 발전될 수 있는 국가 AI 허브 플랫폼을 고도화하여, 개방형 혁신의 구심점 역할을 수행하도록 해야 한다. 또한, 빅테크와 중국의 오픈소스 주도 기업들은 초거대 규모의 파운데이션 모델을 보유하고 있으며, 이를 32B 규모의 모델 완성도를 높이는 데 활용하고 있다는 점에서 현재 32B 이상 규모 모델에서 존재하는 글로벌 격차를 해소하기 위해 대규모 GPU 클러스터 등 국가 전략적 컴퓨팅 인프라에 대한 과감한 투자, 그리고 국내 기업과 학계가 이러한 인프라에 원활하게 접근할 수 있도록 보장하는 정책적 지원이 필요하다.

둘째, 범용 모델과 도메인 모델이 상호작용하는 생태계 설계로 범용 모델과 도메인 모델이 상호 작용하며 발전하는 순환 구조의 생태계를 설계해야 한다. 이는 거대한 지식 기반과 강력한 추론 능력을 제공하는 기반 모델(Foundation Model)로서의 범용 모델을 도메인 특화 모델이 탄생할 수 있는 토대로 활용하는 구조를 의미하며, 더 나아가 도메인 모델이 현장에서 확보한 데이터와 피드백이 다시 범용 LLM을 개선하는데 활용될 수 있는 상호작용 구조를 의미한다. 이러한 이중화 설계가 성공하려면 도메인 해당 분야 기업이 가진 고유한 데이터와 비즈니스 요구사항을 반영해야 하므로 개발 초기부터 도메인 기업의 참여가 필수적이다. 이를 위해 독자 파운데이션 모델 선정 이후 초기부터 도메인, 서비스 영역 기업들의 참여에 수요가 확실한 특화 모델을 만들고 이를 실제 산업 현장에 적용할 수 있는 후속 프로젝트를 설계해야 한다. 또한 모델 개발에서 멈추지 않고 기업들이 기존 시스템에 AI 모델을 쉽게 통합할 수 있도록 API, 개발 툴킷 등 사용하기 편리한 도구를 포함하는 산업 현장용 풀스택 플랫폼을 구축해야 한다. 즉, 우리 범용 모델의 시장성과 실효성, 고도화된 모델 및 산업 적용 확산을 동시에 추진할 수 있는 형태로 지원 구조가 마련되어야 한다.

셋째, 글로벌 시장성과 리더십 확보를 통한 지역적 전문화 달성이다. AI 기술이 글로벌 리더십 확보로 이어질 수 있도록 시장 창출과 레퍼런스 구축에 나서야 한다. 월드베스트 LLM 프로젝트 등에 대한 후속 지원으로 정부의 공공 시장 창출을 통한 시장 레퍼런스 구축을 고민할 필요가 있다. 참여 기업의 자사 및 민간 활용 뿐만 아니라 행정용, 교육용, 스마트시티용 등으로 조달 시장을 개방하여 트랙 레

코드를 쌓을 수 있도록 지원해야 한다. 이를 바탕으로 우리의 강점을 결합한 차별화된 풀스택 AI 패키지를 산업용으로 상품화하여 글로벌 시장을 공략해야 한다. 이는 미국, 중국 등과 차별화된 새로운 AI 생태계 표준을 제시하며, 한국이 글로벌 AI 시장의 리더십을 확보하는 핵심 동력이 될 수 있을 것이다.

2) 독자 AI 모델과 NPU, HW-SW를 잇는 통합 경쟁력 확보

한국은 삼성과 SK하이닉스의 HBM, 스타트업들의 NPU 등 하드웨어 경쟁력을 확보하고 있으며, 이는 도메인 특화 영역 및 엣지(Edge), 온디바이스(On-device) 등 차세대 기술 변곡점 대응에 유리한 조건이다. 그러나, 글로벌 수준의 LLM 개발을 통해 달성하고자 하는 생태계의 청사진이 명확하지 않아, 기술 스택 간 최적화된 공급이 실제 수요와 미스매치될 우려가 있으며 AI 경쟁력 확보에 필수적인 하드웨어(HW), 소프트웨어(SW)의 유기적 연결성에 대한 고민 또한 부족한 실정이다. 국내 AI 기업 중 대부분이 NPU 도입 계획이 없으며, 도입 의향이 있는 기업들조차 ‘호환성 등 SW 지원’이 중요하다고 언급한다. 이러한 상황을 해소하기 위해서는 우선 엔비디아와 CUDA 중심 기존 AI 패러다임 하에서 메모리와 NPU를 중심으로 새로운 생태계를 구축하고, 독자 파운데이션 모델이 그 생태계의 기반 역할을 할 수 있는 구조 설계를 고민해 볼 필요가 있다. 예를 들어 학습과 모델 개발 단계는 GPU 기반 생태계와 협력하여 빠르게 경쟁력을 갖추는 한편 NPU와 엣지 중심의 생태계를 활성화하여 서비스단의 점진적 자립화를 추구하는 방안이다. 이 과정은 GPU 생태계에서 학습된 글로벌 경쟁력을 갖춘 LLM을 한국형 NPU-메모리 인프라에 맞게 최적화하여 엣지와 온디바이스 AI 시장에서 기술 주도권을 확장해나가는 전략으로 연결될 수 있는데 이를 위해서는 독자 파운데이션 모델의 활용도, 시장성, 글로벌 호환성, 차세대 인프라 및 모델과의 시너지 등을 초기 단계부터 함께 설계할 필요가 있으며, 구체적인 정책과제들은 다음과 같다.

첫째, GPU와 NPU가 같이 가는 하이브리드 생태계 구축으로, GPU 중심에서 NPU 생태계로의 확장과 전환을 촉진하는 정책 수립이 필요하다. 현재 한국의 AI 모델 대다수는 GPU 기반이며, 엔비디아 GPU-CUDA 기반 기술 스택의 대체재

가 없는 상황에서 학습 단계에서의 GPU 활용은 불가피하다. 따라서 학습은 GPU 기반으로 수행하되, 추론 및 서비스 단계는 NPU 기반으로 전환하여 빠른 시장 경쟁력 확보와 비용 효율성, 그리고 기술 주권 확보의 기반을 마련해야 한다. 도메인 특화 서비스에서는 점진적으로 NPU 중심 생태계로 전환하여 독립성을 확보하고, NPU의 강점에 기반한 엣지 및 물리적 AI 시대의 새로운 스택을 구성해야 한다.

둘째, 국산 NPU가 시장성을 가지려면, PyTorch나 TensorFlow 같은 기존 프레임워크와 CUDA 기반 환경에 익숙한 개발자들이 기존의 학습된 모델을 NPU에서 쉽게 구동시킬 수 있도록 호환성을 제공하는 소프트웨어 툴체인이 전제되어야 한다. NPU의 저전력·고효율 추론 능력을 최대한 활용하려면 하드웨어에 최적화된 소프트웨어가 필요하며, 단일 기업이 AI 풀스택을 모두 개발하는 것은 현실적으로 불가능하므로 정부의 지원은 NPU에 최적화된 오픈소스 소프트웨어 생태계 구축에도 집중될 필요가 있다.

셋째, 최적화와 효율화를 위한 기술 스택 간 협업 생태계 조성으로, 모델 개발사와 칩 설계 및 제조사가 참여하는 협의체를 구성하여, 칩과 모델이 함께 최적화되며 개발될 수 있도록 지원해야 한다. AI 생태계의 성공을 위해서는 칩을 만드는 하드웨어 기업과 모델을 개발하는 소프트웨어 기업이 초기 단계부터 긴밀하게 협력하여 서로의 성능을 최대로 이끌어 낼 필요가 있다. 중국의 모델-칩 연합 사례를 참조하여 민간의 수요와 공급에 따른 자율적인 협의체 구성을 장려하되, 정부는 그 성과물의 표준화 및 조달 시장 창출 등의 지원을 고민해야 한다.

또한, 이 과정에서 확보된 GPU 제공자들의 역할 또한 재정립될 필요가 있다. 특히, 네이버 클라우드와 같은 CSP들은 국가 AI 인프라와 모델의 상용화를 연결하는 오케스트레이터로서 역할을 재정립할 필요가 있다. 국가 AI 컴퓨팅 센터의 운영 주체로서, 물리적인 GPU 자원을 가상화하여 기업들에게 제공하는 관문을 담당하는 CSP 기업들이 클라우드 상에서 범용 모델에 파인튜닝이 가능한 원스톱 개발 환경을 제공하고, 추론 요청에도 대응가능한 국산 NPU를 자동으로 할당하

도록 클라우드 단에서 처리할 수 있다면, 특화 모델의 응용이나 국산 NPU에 대한 도입 저항감을 낮출 수 있을 것이다.

마지막으로, 글로벌 수준의 시스템 SW 역량을 갖추기 위한 고급인재 육성이 중요하다. 시스템 SW는 AI 모델보다 기술주권 확보가 어려운 분야로 현재 하드웨어 설계 인력에 비해 컴파일러, 드라이버 등 시스템 소프트웨어를 다룰 수 있는 고급 인력의 공급은 턱없이 부족하다. 이는 국산 NPU의 상용화를 가로막는 최대 걸림돌로 작용하고 있다. 따라서 정부 R&D 예산의 일부를 하드웨어 제작 지원에서 컴파일러, 드라이버, 최적화 라이브러리 개발로 전환하는 한편, AI 시스템 SW 부문 최고급 인재 육성에 투자할 필요가 있다. 단 코어AI 및 AI 시스템 고급인재로 성장할 수 있는 학생은 매우 소수라는 점에서 양적 성과에 치중하는 인재육성 정책은 지양되어야 할 것이다.

3) AI 서비스 생태계 활성화로 AI 생태계의 수요기반 자생적 성장 기반 구축
 지금까지 AI 산업은 AI 잠재력에 대한 기대와 초거대 AI 모델 경쟁에 기반한 인프라 수요에 기반하여 성장해왔다. 그러나 실질적인 활용수요 없이는 AI 생태계의 성장은 지속가능하지 않다. AI가 일상생활과 산업 현장에서의 실질적인 가치창출로 이어지지 않는다면 AI 모델 및 인프라에 대한 투자 비용은 상쇄될 수 없으며 데이터 피드백을 제공하는 AI 서비스 없이는 AI 모델의 지속적 고도화는 기대하기 힘들다. AI로 투자가 집중됨에도 불구하고 AI 거품론이 지속적으로 제기되는 이유이다. 이에 AI 서비스 생태계 활성화를 통해 AI의 실질적인 수요를 창출하여 AI 생태계의 자생적 성장기반을 마련할 필요가 있다.

이를 위해 첫째, 누구나 쉽게 AI를 서비스로 이용할 수 있도록 AIaaS(AI as a Service) 생태계를 육성할 필요가 있다. AIaaS 전략의 핵심은 복잡한 AI 기술을 쉽게 사용할 수 있도록 돕는 '중간 역할' 기업들의 육성이다. AI 서비스 기업은 AI의 대중적 활용을 촉진하고 초기 시장 수요를 창출하는 키 플레이어로, AI의 도입 장벽을 낮춰 AI의 빠른 확산을 돕고 AI를 활용한 새로운 서비스, 가치의 창출을 이끌 수 있다. 나아가 산업 특화 서비스의 신속한 적용과 시장화를 통해 물리

적 AI 등 기술 변곡점 선점도 앞당길 수 있다. AI 서비스 생태계 활성화를 위해 정부는 AI 스타트업 기업들의 AI 도입, 활용, 솔루션 평가, AI 신뢰성 테스트까지 한 번에 해결할 수 있는 원스톱 지원 체계를 고려할 필요가 있는데 관련하여 모델 개발자, 애플리케이션 배포자, 제3자 테스터 등이 참여하여 생성형 AI 솔루션을 선정하고 중소기업들에게 활용 기회를 제공하는 싱가포르의 ‘생성형 AI 평가 샌드박스’ 사례를 참고할 만하다.

또한, AI 모델, 인프라에 대한 정부 투자가 성공하기 위해서는 정부의 투자성공물이 실제 비즈니스 현장의 문제를 해결하는 서비스로 연결될 필요가 있다. 투자의 결과물이 실질적인 수익모델로 완성될 때 그 수익이 모델 고도화로 재투자되고 고도화된 모델이 다시 수요 창출로 이어지며 자생적인 성장궤도로 올라갈 수 있다.

4) 전략산업과 연계된 버티컬 AI, 물리적 AI로 AI를 진짜 산업으로 육성

범용 모델 영역에서 기업 간 기술격차가 빠르게 좁혀지고는 있으나 전환비용으로 인해 AI 활용기업들은 AI 선도기업들의 모델을 채택하고 이에 락인될 가능성이 높다. 이에 우리는 한국이 강점을 가진 제조업, 조선, 의료 분야에 특화된 경량화 모델을 개발하여 차별화된 시장을 개척해야 한다. 예를 들면 네이버와 카카오의 경량화 모델 기술을 산업 현장에 접목하여 스마트 팩토리의 디지털 트윈이나 선박 자율 운항 시스템 등에 특화된 ‘산업용 AI’ 시장을 개발하는 것이다. 이는 데이터 보안이 중요한 수요 기업들에게 클라우드 기반 범용 모델의 확실한 대안이 될 수 있으며, 온디바이스 AI의 강점인 프라이버시 보호와 저지연 특성을 극대화할 수 있다. 또한 에이전틱 AI로의 진화로 이제 기업 내 업무 데이터를 활용한 AI 모델 개발이 중요해지고 있어 버티컬 AI 시장에서의 데이터 확보 및 서비스 개발 기회는 우리 기업들이 글로벌 시장에서 차별화된 경쟁력을 가질 수 있는 기회를 제공할 수 있다. 이에 국내 AI 모델, 인프라, 인력에 기반하여 산업 적용 사례를 확보하고 이를 해외로 수출하는 전략이 필요하며, 이를 위해 AI 관련 기술, 인재, 법제도 정책을 주력산업 육성 정책과 연계할 필요가 있으며 API 가격 보조금 지

급 등 국산 AI 모델의 활용도를 높이는 지원 정책도 필요하다. 데이터와 지식의 연계 체계화가 부족해 많은 기업들이 아직 AI 적용에 어려움을 겪고 있다는 점에서 한국형 온톨로지를 통해 산업별 데이터와 지식을 표준화하여 국내 스타트업이 빠르게 도메인 특화 AI를 개발할 수 있는 기반을 제공할 필요도 있으며, AI 기본법의 신뢰 기반 진흥 방침과 규제 샌드박스 제도를 적극 활용하여, 자율주행 에이전트나 뉴로모픽 기반의 스마트 시티 기술을 자유롭게 테스트할 수 있는 과감한 규제 특례 지역을 지정하고, 이를 통해 글로벌 AI 스타트업들을 한국으로 유인하는 ‘AI 테스트베드 코리아’ 전략도 고려해볼 수 있다.

또한, 가까운 미래에 AI가 진짜 산업이 되는 ‘아이폰 모먼트’가 열릴 것으로 예상됨에 따라 AI가 진짜 산업이 되는 차세대 AI 시장에서의 우리의 자리를 확보하기 위한 노력이 필요하다. AI 선도 기업들이 차세대 AI 시장선점을 위해 막대한 자금력과 노력을 투입하고 있지만 아직 생태계가 형성되는 시점이며 물리적 AI 생태계는 다양한 HW, SW 기업들을 포괄하는 형태로 확장되어 갈 것이라는 점에서 아직 다양한 국가, 기업들에게 많은 기회가 존재한다. 특히, 우리나라는 HW, SW에 걸친 역량을 보유하고 있다는 점에서 물리적 AI 분야 선도를 위한 국가 R&D 역량의 집중이 필요하다. 우리나라가 보유한 가장 큰 강점 중 하나는 제조, 로봇, 자율주행 분야의 하드웨어 역량이다. 삼성전자와 현대차그룹으로 대표되는 국내 기업들은 글로벌 수준의 로봇틱스, 제조, 자율주행 하드웨어 기술과 운영 역량을 보유하고 있다. 따라서 이러한 하드웨어 강점과 AI 기술을 결합하여 시너지를 창출할 수 있도록 HW-SW 협력모델을 구축, 지원할 필요가 있다. 현재 AI 생태계는 개별 기업단위를 경쟁력 확보, 글로벌 가치사슬 진입이 어렵다는 점에서 더욱 우리 기업들이 글로벌 가치사슬에서 핵심 노드로 진입할 수 있도록 기업간 연합체, 산학연 컨소시엄 등을 구성하여 이들의 공동 대응을 전방위적인 측면에서의 지원할 필요가 있다.

단순한 LLM 경쟁을 넘어서, 3D 공간, 물리 법칙, 객체 상호작용을 학습하는 ‘월드 파운데이션 모델(World Foundation Model, WFM)’과 타인의 행동 관찰 및 자신의 경험을 통해 물리적 세계와의 상호작용을 학습하는 ‘거대 행동 모델

(Large Behavior Model, LBM)' 등 특화된 파운데이션 모델 개발에 대한 선제적 지원도 필요하다. 이러한 특화 모델들은 향후 물리적 AI 시대의 핵심 인프라가 될 것이며, 국내 제조업과 로봇틱스 산업의 경쟁력을 한 단계 도약시키는 기반이 될 것이다. 또한 물리적 AI가 엷지 환경에서 효과적으로 작동하기 위해서는 경량화 기술이 필수적이므로, VLA(Vision Language Action) 모델의 경량화 및 NPU 최적화 연구에 대한 지원도 병행되어야 한다. NVIDIA의 Omniverse와 Cosmos 플랫폼이 보여주듯이, 고정밀 3D 디지털 트윈 환경에서 물리 시뮬레이션을 수행하고, 이를 통해 데이터를 생성하며, AI 모델을 안전하게 훈련하고 검증할 수 있는 통합 시뮬레이션 플랫폼이 물리적 AI 개발의 핵심 인프라로 부상하고 있어 국가 차원에서 이러한 통합 시뮬레이션 플랫폼을 공공 인프라로 구축하는 것도 고려해볼 필요가 있다. 이를 통해 대기업뿐만 아니라 자본과 하드웨어가 부족한 중소기업과 스타트업도 저비용으로 로봇 AI를 개발할 수 있는 산업 기반을 마련함으로써, 물리적 AI 분야에서 국내 생태계 전반의 경쟁력을 강화할 수 있을 것이다.

엷지, 온디바이스용 칩에서의 AI모델 지원을 위한 자원 최적화도 중요하다. 삼성전자와 같이 세계를 리드하는 엔드단말 제조기업이 있으므로 이런 회사들과 함께 온디바이스에서 AI 모델을 잘 활용하여 좋은 서비스를 구현하는 방향으로 투자가 많이 이루어져야 한다. 정부는 다양한 온디바이스 AI 서비스에서 데이터 수집이 용이하도록 기술·제도적 지원, 다양한 폼팩터 AI 서비스 사용자의 경험성 향상을 위한 지원 등을 통해 이를 촉진시킬 수 있을 것이다.

마지막으로 특정 기업, 특정 국가가 차세대 AI 생태계를 독점하지 않도록 국제 표준, 규범 수립과정에 적극 참여하고 개방형 생태계 조성을 위한 국가적 차원에서의 글로벌 협력 노력이 필요하다.

5) 혁신과 안전의 균형있는 성장을 위한 AI 거버넌스 강화

AI 기술 확산을 위해서는 기술적 역량과 제도적 역량을 동시에 강화해야 한다. 이를 위한 구체적 과제로는 첫째, AI 기본법 하위 규범의 신속한 정비와 이행 지

원 체계 구축이 필요하다. AI 기본법의 실효성은 법률 조항 자체보다 이를 구체화하는 하위 규범의 품질과 구체성에 달려 있다. 고영향 AI 가이드라인, 생성형 AI 표시 방법 기준, 안전성 확보 의무 이행 고시 등을 조속히 정비하는 한편, 기업들이 실제로 준수할 수 있도록 컨설팅과 비용 지원 체계를 함께 갖추어야 한다. 아울러 AI 기술이 빠르게 발전하는 현실을 고려하면, 특정 시점의 기술 수준을 전제로 한 고정적 기준보다는 기술 발전에 따라 동태적으로 조정되는 메커니즘을 처음부터 설계에 내재화하는 것이 중요하다.

둘째, AI 안전연구소와 주무 부처의 규제 집행 역량을 실질적으로 강화해야 한다. AI 안전연구소가 형식적 기구에 머물지 않고 실질적인 연구 역량과 국제 협력 능력을 갖춘 기관으로 발전하려면, 세계 수준의 연구진 확보, 첨단 연구 인프라 구축, AI 안전 국제 공동연구 참여 확대가 뒷받침되어야 한다. 과기정통부의 AI 전문 집행 인력도 확충이 필요하며, 의료, 금융, 교육, 미디어, 공공 부문 등 각 분야별 AI 응용에 전문적으로 대응할 수 있는 범부처 전문가 네트워크도 함께 구축될 필요가 있다.

셋째, 글로벌 AI 거버넌스 논의에 전략적으로 참여하는 노력을 강화해야 한다. AI 서울 정상회의의 성과를 발판으로 OECD, ISO/IEC, AI 안전 연구소 네트워크, APEC 등 다양한 다자 채널에서 한국의 균형 모델을 구체적 정책 제안으로 발전시켜 나가야 한다. 캐나다, 영국, 호주, 싱가포르, 브라질 등 유사한 위치의 중견국들과의 연대를 통해 EU·미국 중심의 글로벌 AI 거버넌스 논의에 아시아·태평양 지역의 시각을 반영하는 노력도 병행해야 한다. 개도국을 대상으로 한 AI 거버넌스 역량 지원 프로그램은 한국 모델의 글로벌 확산과 소프트파워 제고를 동시에 도모하는 유효한 수단이 될 수 있다.

넷째, AI 기본법과 연관 법률 간 정합성을 확보해야 한다. AI 기본법과 「개인정보 보호법」, 「정보통신망법」 등 유관 법령 사이의 규범 충돌과 공백을 체계적으로 정비하는 작업이 필요하다. 특히 개인정보 처리와 AI 학습 데이터 활용, 고영향 AI 영향평가와 개인정보 영향평가의 관계 등에 대한 명확한 기준 마련이 시급하다.

다섯째, 적응형 거버넌스 체계를 실질적으로 구현해야 한다. 이 개념은 이론적으로 자주 언급되어 왔으나 한국에서는 아직 제도적으로 구현된 사례가 충분하지 않다. AI 기술 동향과 시장 변화를 지속적으로 추적하는 모니터링 체계, 정책 효과를 정기적으로 평가하는 시스템, 필요 시 신속하게 정책을 조정할 수 있는 의사 결정 체계를 갖추어야 한다. 인공지능정책센터가 이러한 역할을 실질적으로 지원할 수 있도록 기능과 권한을 명확히 하는 것이 중요하다. AI 기본법에 도입된 규제 샌드박스 제도(법 제28조~제30조)도 적극 활용하여 새로운 AI 기술과 서비스에 대한 실증 기반을 넓히고, 그 과정에서 도출되는 정보가 거버넌스 체계에 실시간으로 반영되는 선순환 구조를 만들어 나가야 한다.

6) AI 기술한계 극복에 기여하는 도전적 연구생태계 구축

AI 기술 발전은 점진적 개선의 누적이라기보다, 트랜스포머와 같은 아키텍처 전환, 생성형 AI의 등장처럼 기존 질서를 바꾸는 불연속적 혁신을 통해 도약해 왔다. 이러한 특성은 후발주자에게 단순 추격만으로는 선도국과의 격차를 좁히기 어렵다는 점을 시사한다. 따라서 격차해소를 위한 투자 뿐 아니라 새로운 패러다임을 만들어낼 수 있는 도전적이고 실험적인 연구에도 전략적인 투자가 필요하다. 비록 단기적 성과 창출 가능성은 낮지만, 이러한 연구는 기술의 방향 자체를 재정 의하고 경쟁 구도를 전환할 수 있는 잠재력을 지니며, 동시에 세계적 수준의 문제에 도전할 수 있는 연구환경을 통해 최고 수준의 인재를 유치·배출하는 기반을 제공할 수 있다.

참 고 문 헌

[국내 문헌]

과학기술정보통신부(2025. 8. 22.), “내년 R&D 예산 파격 투자…연구생태계 완전 복원, ‘진짜 성장’ 실현”, 대한민국 정책브리핑.

<https://www.korea.kr/news/policyNewsView.do?newsId=148948043>.

과학기술정보통신부(2025), “인공지능 기본법 시행령 주요 내용”, 과학기술정보통신부 보도자료.

국가인공지능전략위원회(2025), “대한민국 인공지능 행동계획(안)”.

금융위원회(2025. 9. 10.), “대한민국 경제 재도약150조원 국민성장펀드가 함께합니다.”, 보도자료.

김성욱·이경선(2025), “[특집 1호] 글로벌 AI 경쟁전략 및 대응방안: 풀스택(Full-stack)의 관점에서”, KISDI Premium Report, 2025-06.

뉴스시스(2025. 12. 28.), “SKT, 5000억 매개변수 초거대 AI 모델 30일 공개…‘모두의 AI 만들 것’”.

매일경제(2025. 6. 6.), “AI 반도체 설계 딥엑스 ‘유니콘 기업’ 눈앞”.

박강민, 장진철, 안성원(2024), “유럽연합 인공지능법의 주요내용 및 시사점”, SW 중심사회, 2024년 10월호. 소프트웨어정책연구소.

산업통상자원부(2025. 5. 28.), “’25년 산업AI 개발·확산에 4,800억원 투자”, 보도자료.

서울와이어(2025. 10. 8.), “삼성·SK HBM 점유율 80%…한국의 저력 어디서 나오나”.

시사저널(2025. 11. 10.), “AI 인재 양성에 1조4000억원 투입…‘패스트트랙’도 도입한다”.

아시아경제(2025. 9. 4.), “세계가 눈독 들인다…‘가치 1조 훌쩍’ 한국 기업들, AI강

자 사이서도 빛나”.

<https://www.asiae.co.kr/article/2025090408121378075>.

연합뉴스(2025. 7. 29.). “中상하이, AI산업 육성 위해 2천억원 규모 보조금 지원”.

<https://www.yna.co.kr/view/AKR20250729144300009>.

유재홍(2024. 7. 30.). 세계인공지능대회를 통해 살펴 본 중국의「AI+ 행동」계획과 산업계 동향. CSF 전문가오피니언, 대외경제정책연구원.

https://csf.kiep.go.kr/issueInfoView.es?article_id=54576&mid=a20200000000&board_id=4.

이정아(2025. 3.), “중국의 인공지능(AI) 정책 전략 현황과 변화 방향: AI 대국으로 도약하기 위해 중국은 어떻게 변화하고 있는가?”, The AI Report 2025-1, 한국지능정보사회진흥원(NIA).

이해수, 유재홍, 안성원(2025), “퍼지컬 AI의 현황과 시사점“, 소프트웨어정책연구소. 「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」. 법률 제20676호. 2025. 1. 21. 제정.

정재욱(2024. 7.). “세계 최초로 통과된 EU 「AI법」, 우리 기업의 대응 방향은?“ 외교부.

「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법 시행령」. 대통령령 제36053호. 2026. 1. 22. 제정.

조선경제(2025. 8. 25.), “R&D 예산 역대 최대… 인재양성은 여전히 쥐꼬리”.

최문실, 정기호, 최형인(2025. 4), “2024 인터넷이용 실태조사”, NIA.

한국경제(2024. 7. 1.), “LG unveils Google’s generative AI-powered robot CLOi”
<https://www.kedglobal.com/artificial-intelligence/newsView/ked202406270007>.

_____ (2025. 9. 12.), “Doosan Robotics aims to draw \$256 mn in IPO subscriptions”

<https://www.kedglobal.com/robotics/newsView/ked202309110020>.

혁신의숲(2025. 3. 26.), “글로벌 AI 반도체 스타트업 성장 분석”.

byline Network(2025. 6. 4.), “[AI 반도체 스타트업은 지금 ②] 사피온 품은 리벨리온, 사업 본궤도에 한 발짝”.

CLOVA(2025. 7. 22.), “AI 생태계에 새로운 싹을 틔우다: HyperCLOVA X SEED Think”.

IT DAILY(2025. 11. 4.), “李 정부 2026년 AI 예산안 10조 원 돌파…‘AI 3대’ 강국 도약 첫걸음 내딛나”.

<http://www.itdaily.kr/news/articleView.html?idxno=236161>.

kakao(2025. 2. 4.), “카카오, 국내 최초로 오픈AI와 전략적 제휴 체결”.

Kearney(2025. 4. 22.), “Physical AI 시대의 도래와 대비 방안”.

KERC(2025. 6), “2024 유럽 분야별 연구 및 정책 동향”, 한-EU연구협력센터 (KERC).

NVIDIA blog(2024. 3), “소버린 AI란?”

<https://blogs.nvidia.co.kr/blog/what-is-sovereign-ai/>.

Platum(2025. 11. 6.), “딥엑스, CES 2026 혁신상 2개 부문 수상”.

Samsung Newroom(2024. 12. 31.), “Samsung Electronics To Become Largest Shareholder in Rainbow Robotics Accelerating Future Robot Development”.

<https://news.samsung.com/global/samsung-electronics-to-become-largest-shareholder-in-rainbow-robotics-accelerating-future-robot-development>.

SK telecom Newsroom(2025. 7. 24.), “SKT, 자체 LLM 기술 바탕으로 초거대 AI 모델 도전”.

_____ (2025. 7. 3.), “SK텔레콤, 에이닷 엑스 4.0 지식형 모델 오픈소스로 공개”.

SuperbAI(2025. 6. 25.), “젠슨 황이 선언한 AI의 다음 물결, 피지컬 AI란 무엇인가? 개념 및 역사 정리”.

<https://blog-ko.superb-ai.com/what-is-physical-ai-next-wave/>
 the bell(2025. 9. 18.), “퓨리오사AI, 4000억 시리즈D 라운드 시동건다”.
 THEELEC(2025. 3. 4.), “[Y인사이트] 온디바이스AI NPU 스타트업 모빌린트, 올해 매출 가시화”.
 ZDNET Korea(2025. 7. 3.), “KT, 한국적AI ‘믿음2.0’ 오픈소스 공개...국내 AI 대중화”.

[해외문헌]

36氪 (36Kr). (2025). 「人脸识别技术应用安全管理办法」解读 [Analysis of the Measures for the Safety Management of Facial Recognition Technology Applications]. 2025년 3월.

Alibaba Cloud.(2025. 4. 8.). The Best Practice of Moonshot AI in Massive Data Preprocessing for the Kimi Large Model.

https://www.alibabacloud.com/blog/the-best-practice-of-moonshot-ai-in-massive-data-preprocessing-for-the-kimi-large-model_602119.

Alibaba(2024). Fiscal Year 2024 Annual Report.

Anthropic(2025. 3. 3.), “Anthropic raises Series E at \$61.5B post-money valuation”.

Appttronik(2024. 3. 15.), “Appttronik and Mercedes-Benz Enter Commercial Agreement”

<https://appttronik.com/news-collection/appttronik-and-mercedes-benz-enter-commercial-agreement>.

ArtificiaAnalysis.ai, <https://artificialanalysis.ai/>, 2025.

artificialintelligenceact.eu.(2025). Brussels Effect and Global AI Regulation.

<https://artificialintelligenceact.eu>(최종방문일 2026. 1. 15.).

Astute Analytica(2025), “Industrial Robotics Market Revenue To Attain USD 235.38 Billion By 2033”.

Baidu Developer Community.(2024). PaddlePaddle 3.0 Release Notes.
<https://github.com/PaddlePaddle/Paddle/releases/tag/v3.0.0>.

Baidu(2024). Form 20-F for Baidu Inc.

Bird & Bird(2025). AI Regulatory Horizon Tracker — EU: Japan. Bird & Bird LLP.

Bloomberg(2025. 7. 18). “AI Startup Perplexity Valued at \$18 Billion With New Funding”.

Boston Dynamics(2025. 4. 3.), “Boston Dynamics & Hyundai Motor Group Expand Collaboration to Drive Mobility Manufacturing & Innovation”

<https://bostondynamics.com/news/boston-dynamics-hyundai-motor-group-expand-collaboration-drive-mobility-manufacturing-innovation/>.

BOND(2025. 5), “Trends - Artificial Intelligence(AI)”.

Botinfo.ai(2025. 9. 19.), “Agility Robotics Digit: The Logistics-First Humanoid Built for Work”

<https://botinfo.ai/articles/agility-robotics-digit>.

Brookings Institution.(2025). AI Policy Under Trump: Regulatory Rollback and State Divergence. Washington, D.C.: Brookings Institution.

Business Research Insights(2025), “Autonomous Vehicles Market Overview”

<https://www.businessresearchinsights.com/market-reports/autonomous-vehicles-market-106395>.

CAC (Cyberspace Administration of China). (2023a). 「互联网信息服务深度合成管理规定」 [Provisions on the Administration of Deep Synthesis Internet Information Services]. 2023년 1월 10일 시행.

- CAC (Cyberspace Administration of China). (2023b). 「生成式人工智能服务管理暂行办法」 [Interim Measures for the Management of Generative Artificial Intelligence Services]. 2023년 8월 15일 시행.
- CAC, 공업정보화부, 공안부, 국가방송TV총국 (2025). 「人工智能生成合成内容标识办法」 [Measures for the Labeling of AI-Generated Synthetic Content]. 2025년 3월 14일 발표.
- California Governor's Office. (2025). Senate Bill 53: Frontier AI Transparency Requirements. September 29, 2025.
- CBInsights(2024. 7. 30.), "State of AI Q2'24 Report".
https://www.cbinsights.com/research/report/ai-trends-q2-2024/?utm_source=chatgpt.com.
- Challapally, A., Pease, C., Raskar, R., Chari, P.(2025. 7), "The GenAI Divide STATE OF AI IN BUSINESS 2025", MIT NANDA.
- Chambers and Partners(2025a), Artificial Intelligence 2025: China; Japan; South Korea Sections. Global Practice Guides.
- _____ (2025b), Artificial Intelligence 2025: Finland Trends and Developments.
<https://practiceguides.chambers.com/practice-guides/artificial-intelligence-2025/finland/trends-and-developments>.
- Cervicorn Consulting(2025), <https://www.cervicornconsulting.com/physical-ai-market>.
- CIOdive(2025. 3. 20.), "Data center investments surged to \$455B last year: report".
- Constellation(2025. 6. 30.), "Microsoft delivers strong Q4, Azure delivers \$75 billion in annual revenue".
<https://www.constellationr.com/blog-news/insights/microsoft-delivers-strong-q4-azure-delivers-75-billion-annual-revenue>.

CoreWeave, Inc.(2025. 3. 14.), “Form S-1/A (Amendment No.1).” U.S. Securities and Exchange Commission.

<https://www.sec.gov/Archives/edgar/data/1769628/000119312525058309/d899798ds1a.htm>.

CoreWeave(2025. 5. 14.), Q1 2025 Earnings Press Release.

<https://www.sec.gov/Archives/edgar/data/1769628/000176962825000010/coreweave1q25earningspress.htm>.

CRN(2025. 8. 5.), “AWS Vs. Microsoft Vs. Google Cloud Earnings Q2 2025 Face-Off”.

<https://www.crn.com/news/cloud/2025/aws-vs-microsoft-vs-google-cloud-earnings-q2-2025-faceoff>.

CRN(2025. 8. 7.), “Cloud Market Share Q2 2025: Microsoft Dips, AWS Still Kingpin”.

<https://www.crn.com/news/cloud/2025/cloud-market-share-q2-2025-microsoft-dips-aws-still-kingpin?page=>.

DeepSeek-AI et al.(2025), “DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning”.

Draghi, M. (2024). The Future of European Competitiveness: A Competitiveness Strategy for Europe. Brussels: European Commission. September 2024.

Epoch AI(2025. 4). “Large-Scale AI Model Releases by Modality.”

<https://ourworldindata.org/artificial-intelligence>

European Commission, European research development and deployment of AI,

<https://digital-strategy.ec.europa.eu/en/policies/european-ai-research>.

European Commission(2024). Regulation (EU) 2024/1689 of the European

- Parliament and of the Council on Artificial Intelligence (EU AI Act). Official Journal of the European Union. August 1, 2024.
- European Commission(2025a). EU Competitiveness Compass. Brussels: European Commission. January 29, 2025.
- _____ (2025b). Proposal for a Regulation Amending Regulation (EU) 2024/1689 (Digital Omnibus on AI), COM(2025) 836. Brussels: European Commission. November 19, 2025.
- _____ (2025c). InvestAI Initiative: €200 Billion AI Investment Plan. Brussels: European Commission. February 11, 2025.
- _____ (2025d). AI Continent Action Plan. Brussels: European Commission. April 9, 2025.
- _____ (2025). Apply AI Strategy. Brussels: European Commission. October 2025.
- European Commission AI Office(2025a), General-Purpose AI Code of Practice: Final Version. Brussels: EU AI Office. July 10, 2025.
- _____ (2025b). Guidelines on the Scope of the Obligations for Providers of General-Purpose AI Models. Brussels: EU AI Office. July 18, 2025.
- _____ (2025c). Template for the Summary of Training Data. Brussels: EU AI Office. July 24, 2025.
- Executive Office of the President.(2025). Executive Order 14179: Removing Barriers to American Leadership in Artificial Intelligence. January 23, 2025. 90 Fed. Reg. 8741.
- _____ (2025). Executive Order 14317: Promoting the Export of the American AI Technology Stack. July 23, 2025.
- Executive Office of the President. (2025). Executive Order 14318:

Accelerating Federal Permitting of Data Center Infrastructure. July 23, 2025.

Executive Office of the President.(2025). Executive Order 14319: Preventing Woke AI in the Federal Government. July 23, 2025.

_____(2025). Executive Order 14363: Launching the Genesis Mission. November 24, 2025.

_____(2025). Executive Order 14365: Ensuring a National Policy Framework for Artificial Intelligence. December 11, 2025. 90 Fed. Reg. 58499.

Federal budget(2025. 1. 7.), “Federal AI and IT Research and Development Spending Analysis”.

<https://federalbudgetiq.com/insights/federal-ai-and-it-research-and-development-spending-analysis/>.

Figure AI(2025. 11. 19.), “F.02 Contributed to the Production of 30,000 Cars at BMW.”

<https://www.figure.ai/news/production-at-bmw>.

FPF(Future of Privacy Forum)(2025). Understanding Japan’s AI Promotion Act: An “Innovation-First” Blueprint for AI Regulation. Washington, D.C.: FPF.

Fortune(2025. 2. 20.), “ChatGPT is used by 1 in 3 college students—but variability across states is cause for concern, say OpenAI”

<https://fortune.com/education/articles/openai-report-more-than-one-third-college-aged-adults-use-chatgpt/>.

Fortune Business Insights(2025),

<https://www.fortunebusinessinsights.com/autonomous-vehicle-market-109045>.

- Franco, G., Monteagudo-Lago, P., Colbert, I., Fraser, N., Blott, M(2025. 8. 28), “Improving Quantization with Post-Training Model Expansion, AMD. <https://arxiv.org/html/2503.17513v2>.
- Gartner(2024. 10), Top 10 Strategic Technology Trend for 2025.
- GLM-4.5 Team(2025a), “GLM-4.5: Agentic, Reasoning, and Coding (ARC) Foundation Models”.
- GLM-V Team(2025b), “GLM-4.5V and GLM-4.1V-Thinking: Towards Versatile Multimodal Reasoning with Scalable Reinforcement Learning”.
- Global Policy Watch. (2025). AI Office Publishes Final Version of the Code of Practice for General-Purpose AI Models. Covington & Burling LLP. July 30, 2025.
- Global Times.(2025). China AI Patent Holdings: 60% of Global Share. Beijing: Global Times.
- Goodwin Law.(2025). One Big Beautiful Bill Act: AI Moratorium Provisions Analysis. Goodwin Procter LLP. July 2025.
- Goldman Sachs(2025. 2. 27), “Earnings Takeaways: Q4 2024.” Marcus by Goldman Sachs.
- Governor Abbott, State of Texas. (2025). Texas Responsible AI Governance Act (TRAIGA, HB 149). June 22, 2025.
- Grand View Research(2025a), “Artificial Intelligence As A Service Market Summary”
<https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-as-a-service-market-report>.
- _____ (2025b), “Humanoid Robot Market (2025-2030) Summary”
<https://www.grandviewresearch.com/industry-analysis/humanoid-robot-market-report>.

HBR(2025), <https://hbr.org/2025/10/why-agentic-ai-projects-fail-and-how-to-set-yours-up-for-success>.

HSF Kramer(Herbert Smith Freehills Kramer)(2025). AI Tracker: Mainland China — CAC Enforcement Campaign 2025.

Huawei Investment & Holding Co., Ltd.(2024). 2024 Annual Report.

IAPP(International Association of Privacy Professionals). (2025). China AI Regulatory Update 2025; Global AI Regulatory Tracker 2025. Portsmouth: IAPP.

IIEA(Institute of International and European Affairs)(2025). The Transition to a New Digital Policy Agenda: EU Digital Policy 2025-2026. Dublin: IIEA.

International Energy Agency(2025. 4), “Energy and AI”.

ITdaily.(2025. 7. 3.). European Cloud Providers See Revenue Growth, but not Market Share.

<https://itdaily.com/news/cloud/market-share-european-providers-no-growth/>.

Kim, S., Kim, J., Yoon, D., Shin, J., Lee, J., Seo., J.(2025. 9. 29.), “Speculative Verification: Exploiting Information Gain to Refine Speculative Decoding.”<https://arxiv.org/html/2509.24328v1>.

Ma, S., Wang, H., Huang, S., Zhang, X., Hu, Y., Song, T., Xia, Y., Wei, F.(2025. 4. 16.), “BitNet b1.58 2B4T Technical Report”, Microsoft Research,. <https://arxiv.org/html/2504.12285v1>, 2025.

Market Research Future(2025), “Service Robotics Market Summary”.

<https://www.marketresearchfuture.com/reports/service-robotics-market-2437>.

Nestor Maslej, Loredana Fattorini, Raymond Perrault, Yolanda Gil,

Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, Toby Walsh, Armin Hamrah, Lapo Santarlaschi, Julia Betts Lotufo, Alexandra Rome, Andrew Shi, Sukrut Oak, (2025.4), "The AI Index 2025 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA.

McKinsey (2025. 11), "The state of AI in 2025: Agents, innovation and transformation".

<https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>.

MERICs (2025. 7. 22.), China's drive toward self-reliance: Artificial intelligence chips & large language models.

<https://meric.org/en/report/chinas-drive-toward-self-reliance-artificial-intelligence-chips-large-language-models>.

Medium (2025. 6. 8.), "OpenAI's Explosive Growth: A Revenue Breakdown and Industry Comparison".

Mordor Intelligence (2025a), "Agentic AI Market Size & Share Analysis - Growth Trends & Forecasts (2025-2030)

<https://www.mordorintelligence.com/industry-reports/agentic-ai-market>".

_____ (2025b), "Autonomous (Driverless) Cars Market Size & Share Analysis - Growth Trends And Forecast"

<https://www.mordorintelligence.com/industry-reports/autonomous-driverless-cars-market-potential-estimation>.

Morgan Stanley (2025. 3. 20.), "AI's Next Leap: 5 Trends Shaping Innovation and ROI"

- <https://www.morganstanley.com/insights/articles/ai-trends-reasoning-frontier-models-2025-tmt>.
- NBC News(2025. 11. 24.), “Trump signs executive order launching Genesis Mission AI project”.
- NCSL(National Conference of State Legislatures). (2025). 2025 State AI Legislation Tracker. Denver: NCSL.
- NPC Standing Committee (전국인민대표대회 상무위원회). (2025). 网络安全法修正案[Cybersecurity Law (CSL) Amendment].
- NVIDIA Newsroom(2025. 2. 26.), “NVIDIA Announces Financial Results for Fourth Quarter and Fiscal 2025”.
- <https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-fourth-quarter-and-fiscal-2025>.
- NVIDIA(2025. 3), GTC Keynote
- https://www.youtube.com/watch?v=_waPvOwL9Z8.
- OECD(2025). OECD AI Policy Review of Korea 2025. OECD Publishing, Paris.
- O’Melveny. (2025). Executive Order 14365: Constitutional Questions and State AI Law Implications. O’Melveny & Myers LLP. December 2025.
- OpenAI(2025), <https://openai.com/ko-KR/index/introducing-gpt-oss/>.
- Oracle Corporation(2024. 6. 11.). “Oracle Announces Fiscal 2024 Fourth Quarter and Fiscal Full Year Financial Results.”
- <https://investor.oracle.com/investor-news/news-details/2024/Oracle-Announces-Fiscal-2024-Fourth-Quarter-and-Fiscal-Full-Year-Financial-Results/default.aspx>.
- Pew Research Center(2025. 2. 25.), “U.S. Workers Are More Worried Than Hopeful About Future AI Use in the Workplace”.

<https://www.pewresearch.org/social-trends/2025/02/25/u-s-workers-are-more-worried-than-hopeful-about-future-ai-use-in-the-workplace/>.

Precedence research(2025. 12), “Agentic AI Market Size, Share and Trends 2025 to 2034”.

<https://www.precedenceresearch.com/agentic-ai-market>.

Reuters(2024. 4. 26.), Beijing city to subsidise domestic AI chips, targets self-reliance by 2027.

<https://www.reuters.com/technology/beijing-city-subsidise-domestic-ai-chips-targets-self-reliance-by-2027-2024-04-26/>.

_____(2025. 4. 15.), Baidu Announces Third-Generation P800 Chip Cluster Deployment.

<https://www.reuters.com/technology/baidu-announces-third-generation-p800-chip-cluster-deployment-2025-04-15/>.

Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y.K., Wu, Y., Guo, D.(2024), “DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models”.

State of Colorado. (2025). SB 25B-004: Colorado AI Act — Extension of Implementation Date. August 28, 2025.

State Council of China. (2017). 「新一代人工智能发展规划」[New Generation Artificial Intelligence Development Plan]. 国发[2017]35号. 2017년 7월 20일.

State Council of China. (2025). 「关于深入实施“人工智能+”行动的意见」[Opinions on Deepening the Implementation of the “AI+” Initiative]. 国发[2025]11号. 2025년 8월 26일.

Straits Research(2025), “Agentic AI Market Size & Outlook, 2025-2033”

<https://straitsresearch.com/report/agentic-ai-market>.

Su, D., Gu, A., Xu, J., Tian, Y., Zha, J.(2025), “GaLore 2: Large-Scale LLM Pre-Training by Gradient Low-Rank Projection”.

<https://arxiv.org/abs/2504.20437>, 2025.

TECHINASIA(2025. 7. 12.). Alibaba-backed AI startup Moonshot AI launches open-source model.

The Wall Street Journal(2025. 4), “AI Boom to Fuel Surge in Data Center Energy Needs, IEA Says”.

The White House(2023. 8. 9.), “Addressing United States Investments in Certain National Security Technologies and Products in Countries of Concern”.

_____ (2025. 7), “Winning the RaceAMERICA’S AI ACTION PLAN”.

_____ (2025. 7. 23a), “ACCELERATING FEDERAL PERMITTING OF DATA CENTER INFRASTRUCTURE”.

_____ (2025. 7. 23b), “PROMOTING THE EXPORT OF THE AMERICAN AI TECHNOLOGY STACK”.

_____ (2025. 7. 23c), “PREVENTING WOKE AI IN THE FEDERAL GOVERNMENT”.

United States Congress(2021). National Artificial Intelligence Initiative Act of 2020.

USSC (United States Studies Centre). (2025). China’s AI Global Governance Action Plan. Sydney: USSC.

WashingtonExaminer(2025. 11. 20.), “AI is redefining national power and sovereignty”.

<https://www.washingtonexaminer.com/restoring-america/courage-strength-optimism/3893352/ai-redefining-national-power-sovereignty/>.

White & Case LLP. (2025). AI Watch: Global Regulatory Tracker — United States; China. White & Case LLP.

Wilson Sonsini(2025). EU Releases Final Code of Practice for General-Purpose AI Models. Wilson Sonsini Goodrich & Rosati. July 10, 2025.

Xinhua(2024. 8. 21). “목소리도 저작권” 中, AI 음성권 침해 첫 배상 판결. 신화망.

Zhao, J., Zhang, Z., Chen, B., Wang, Z., Anandkumar, A., Tian, Y.(2024) GaLore: Memory-Efficient LLM Training by Gradient Low-Rank Projection, <https://arxiv.org/abs/2403.03507>.

第一财经(2025. 1. 21.), 国家AI产业投资基金成立 600.6亿元撬动中国人工智能市场再升级, <https://www.yicai.com/video/102450232.html>

_____(2025. 1. 3.), 大基金三期出手1640亿元, 2025半导体行业动向如何?, <https://www.yicai.com/news/102427078.html>

新华社(2025. 3. 6.), “人工智能+”: 硬核科技加速走进产业竞技场, https://www.gov.cn/yaowen/liebiao/202503/content_7011011.htm.

_____(2025. 8. 1.), 国务院常务会议解读|我国部署深入实施“人工智能+”行动, <http://www.news.cn/government/20250801/11ed8b53dfcb4e5a98c1ea47736b4af8/c.html>.

央视网(2024. 5. 28.), 注册资本3440亿, 国家大基金三期来了! 或将投资这些重点项目, <https://finance.cctv.com/2024/05/28/ARTIhM6N831KrZyhRYMQziH8240528.shtml>.

중국어제방송국(2025). “중국 생성형 AI 서비스 승인 현황”.

[웹사이트]

나무위키 사이트,

<https://namu.wiki/w/%EB%82%98%EB%AC%B4%EC%9C%84%ED%82>

%A4.

에이직랜드 사이트, <https://eng.asicland.com/business-scope>.

이재명정부 123대 국정과제 사이트, <https://www.korea.kr/govVision/>.

한국과학창의재단 사이트,

<https://www.kosac.re.kr/menus/248/boards/459/posts/37656?brdType=R&thisPage=1&bbIdx=37758&brdCodeValue=&searchField=&searchText=>.

현대차 사이트,

<https://www.hyundai.com/worldwide/en/brand-journal/mobility-solution/hyundai-boston-dynamics>.

A2A protocol 사이트, <https://a2a-protocol.org/>.

Apple 사이트,

<https://www.apple.com/newsroom/2025/09/apple-unveils-iphone-17-pro-and-iphone-17-pro-max/>.

AWS 사이트. <https://aws.amazon.com/ai/machine-learning/trainium/>(최종 검색일: 2025. 10. 1.).

Amazon IR 사이트,

<https://ir.aboutamazon.com/quarterly-results/default.aspx>.

CLOVA 사이트, <https://clova.ai/hyperclova>.

DeepMind 사이트, <https://deepmind.google/models/gemini-robotics/>.

EuroHPC 사이트,

https://www.eurohpc-ju.europa.eu/ai-factories_en#:~:text=Expanding%20and%20scaling%20business%20innovation%20for%20SMEs,13%20AI%20Factory%20Antennas%20that%20offer%20free%2C.

github 사이트, <https://github.com/langchain-ai/langgraph>.

HD현대로보틱스 사이트,

- <https://www.hd.com/en/business/machine/hyundai-robotics/contents>.
- kakao 사이트, <https://www.kakaocorp.com/page/detail/11334>.
- kakao tech 사이트, <https://tech.kakao.com/posts/702>
- KT 사이트, https://enterprise.kt.com/pd/P_PD_NE_00_316.do.
- LG전자 사이트, <https://www.lgcorp.com/media/release/27812>.
- LG AI Research 사이트, <https://www.lgresearch.ai/exaone/>.
- <https://www.lgresearch.ai/news/view?seq=577&utm>.
- LGRsearch 블로그, <https://www.lgresearch.ai/blog/view?seq=575>.
- Market.us 사이트, <https://market.us/report/agentic-ai-market/>.
- Microsoft IR 사이트,
<https://www.microsoft.com/en-us/investor/earnings/fy-2026-q2/press-release-webcast>.
- Model Context Protocol 사이트,
<https://modelcontextprotocol.io/docs/learn/architecture>.
- NAVER 블로그, <https://clova.ai/tech-blog>.
- neuromeka 사이트, <https://en.neuromeka.com/>.
- NVIDIA 사이트, <https://www.nvidia.com/en-us/ai/cosmos/>.
- NVIDIA 사이트, <https://www.nvidia.com/en-us/omniverse/>,
- Oracle 한국 웹페이지.
<https://www.oracle.com/kr/ai-infrastructure/> (최종검색일: 2025. 11. 5.)
- paddlepaddle 사이트,
https://www.paddlepaddle.org.cn/documentation/docs/en/guides/index_en.html.
- Qualcomm 사이트,
<https://www.qualcomm.com/smartphones/products/8-series/snapd>

ragon-8-gen-5-mobile-platform.

Rainbow robotics 사이트, <https://www.rainbow-robotics.com/>.

Samsung 사이트,

<https://semiconductor.samsung.com/processor/mobile-processor/exynos-2600/>.

SuperHuman 블로그,

<https://blog.superhuman.com/enterprise-agentic-ai-adoption/>.

Unitree 사이트, <https://www.unitree.com/h1/>.

Abstract

Kyoungsun LEE, Sungok KIM, Kyungeun LEE, Jangmin Oh,
Hyesun Yoon

■ Title

Research on Enhancing AI Technological Sovereignty and National Competitiveness

■ Purpose of Research

Artificial Intelligence(AI) has emerged as a strategic asset that extends beyond supporting industrial innovation to exert a significant influence on national economic growth, industrial competitiveness, security, and society as a whole. This study aims to analyze the global AI technological and market environment, as well as policy trends in major countries, and to propose policy directions for Korea to secure AI technological sovereignty and utilize AI as a core driver of national competitiveness.

■ Main Outcomes of Research and Policy Implications

This study analyzes global trends in AI technology and markets and

conducts a comparative assessment of AI ecosystems across major countries. It further evaluates Korea's AI technological self-reliance—including external dependency and supply chain risks—along with the current competitiveness and future potential of the domestic AI industry, as well as the coherence of existing AI policies. Based on these analyses, the study proposes policy directions and implementation strategies to strengthen AI technological sovereignty and enhance national competitiveness.

The main findings are as follows. Major economies such as China and the EU are pursuing AI technological sovereignty in their own ways by strengthening and expanding connectivity across resources from a full-stack perspective and by securing control over the entire AI supply chain. As evidenced by these cases, achieving technological independence across all layers of the AI stack is not the only path to AI sovereignty. Moreover, given that the current AI ecosystem—largely dominated by the United States—has evolved through the organic integration and co-evolution of interconnected technology stacks, full-stack technological self-reliance is not a feasible short-term objective. AI experts also emphasize that governments should focus resources on core technology stacks where competitive advantages can realistically be achieved. In this regard, the goal of AI sovereignty should not be an idealized fully self-sufficient ecosystem, but rather “minimum dependency with maximum competitiveness,” ensuring autonomy in critical technological domains without excessive vulnerability to external pressures.

Policy priorities across different technology stacks can be derived by

restructuring the assessment of AI technological self-reliance around two key dimensions: supply chain risk and future competitiveness. First, areas with high supply chain risk but low future competitiveness—such as semiconductors/accelerators and system software—require government-led, long-term efforts to mitigate risks from a technological sovereignty perspective. Second, areas with both high supply chain risk and high future competitiveness—such as AI models and cloud—require concentrated government investment to secure both technological sovereignty and industrial competitiveness. Third, areas with relatively low supply chain risk but high future competitiveness—such as application software, networks, and data infrastructure/solutions—should be driven primarily by the private sector, with government support to enhance industrial competitiveness. Finally, areas with both low supply chain risk and low future competitiveness—such as development platforms and frameworks—should be assigned the lowest policy priority. However, given the rapidly evolving nature of the global AI ecosystem, continuous monitoring remains necessary even for these areas. In summary, in the short term, strengthening competitiveness in the data and AI model training ecosystem—including AI models and cloud—is essential. In the mid-to-long term, reducing supply chain risks and enabling the expansion of Korea’s AI ecosystem based on domestic technologies will require sustained investment in semiconductors/accelerators and system software, as well as ecosystem design that facilitates integration and expansion across technology stacks.

정보통신정책연구원 기본연구 안내

■ 2023 기본연구

- 기본연구 23-01 데이터 경제 활성화를 위한 민관 역할분담과 정책 개입영역 연구
(윤성욱, 박소연)
- 기본연구 23-02 해외 주요국 신규 사업자 이동통신시장 진입 효과 분석(김민희)
- 기본연구 23-03 유사국 사례를 통해 본 북한 통신시장 발전 방안 연구(임동민, 서소영,
이종화, 조대근, 서흥수)
- 기본연구 23-04 메타버스 시대 기본권 보호에 관한 연구: 인격권 논의를 중심으로
(권은정, 한혜지, 계인국, 김법연, 이승민)
- 기본연구 23-05 디지털화폐 생태계 변화에 대응한 중앙은행 디지털화폐(CBDC) 도입
정책 연구(박동욱)
- 기본연구 23-06 ICT 확산에 따른 노동시장 임금격차(최지은, 이은영, 최세림, 이현옥)
- 기본연구 23-07 데이터 생산·관리 역량 평가를 위한 성숙도 모형 개발 연구(노희용,
장신재, 박지원)
- 기본연구 23-08 인터넷 생태계 환경 변화에 따른 네트워크 인프라 비용 분담에 관한
연구(염수현, 강인규, 이상규, 김태오)
- 기본연구 23-09 미디어 이용 조사의 모드 효과 비교(신지형)
- 기본연구 23-10 방송미디어분야 자율규제 제도화 방안 연구(이종원)
- 기본연구 23-11 이동통신 네트워크 인프라 산업 생태계의 전환 방향 연구(여재현,
박지현, 윤도원, 장희선, 김선우, 정인준, 박의환)

■ 2024 기본연구

- 기본연구 24-01 AI 반도체 생태계 조사(윤성욱, 김민식, 민대홍)
- 기본연구 24-02 인공지능 감시에 의한 권력의 확대와 그 규범적 대응방안 연구
(문광진, 허진주, 전민경)
- 기본연구 24-03 시청각미디어 콘텐츠의 공정한 거래를 위한 정책 방안 연구: 수익
배분 관련 이슈를 중심으로(강준석)
- 기본연구 24-04 온라인 광고 시장의 경매 경쟁에 대한 연구(백소성)
- 기본연구 24-05 기술진보에 따른 산업별 업무 변화 연구(서영선)
- 기본연구 24-06 5G 요금제 특성이 이용자 편익에 미치는 영향: 단말기와 요금제의
결속을 중심으로(박진환)

- 기본연구 24-07 방송 프로그램 장르 구분의 타당성 검토: 방송법과 방송사업자 및 수용자의 장르 분류 기준 차이를 중심으로(황현정)
- 기본연구 24-08 생성형 AI가 미디어 분야에 미칠 영향에 대한 탐색적 연구(곽동균, 김남두, 주성희, 황현정, 강하연, 김예빈)
- 기본연구 24-09 디지털 경제 활성화를 위한 실물자산토큰(Real World Asset) 시장의 탈중앙화 금융 알고리즘 적용 가능성 연구(김 찬)
- 기본연구 24-10 AI 반도체 정책의 효과성 제고 방안 연구(정현준, 김민식, 오정숙)

■ 2025 기본연구

- 기본연구 25-01 통신시장의 경쟁활성화를 위한 공동지배력 평가 방법론 연구 (라성현)
- 기본연구 25-02 디지털 심화에 따른 몰입형 서비스 생태계 연구 (이재영, 심홍진, 성윤희, 이승환, 구자근)
- 기본연구 25-03 중국 유통플랫폼의 글로벌 확장과 대응방안 (김성욱, 장신재, 손가녕, 전민경)
- 기본연구 25-04 디지털 융합 시대 정보통신법제의 통합과 재편에 관한 연구 (권은정, 김법연, 김지훈, 선지원, 박상미)
- 기본연구 25-05 디지털 환경 변화를 반영한 데이터 트래픽 예측 연구(김민희, 정광재)
- 기본연구 25-06 글로벌 미디어 플랫폼 확산 대응을 위한 국내 미디어 서비스 경쟁력 분석 및 강화 전략 연구(강준석, 권용재)
- 기본연구 25-07 월드뱅크 등 국제기구와의 협력 연구 (황준호, 양종민, 김지원, 문용일, 이효원)
- 기본연구 25-08 생성형 AI의 생산성 분석(민대홍)
- 기본연구 25-09 디지털 전환에 따른 소매패턴 변화와 정책 방향(장신재, 이선희)
- 기본연구 25-10 인공지능·자율시스템 기반 도시의 윤리적 설계방안 연구 (이현경, 문광진, 전민경)
- 기본연구 25-11 디지털 기술 활용이 기업 성과에 미치는 영향 분석과 정책 시사점 연구(손녕선)
- 기본연구 25-12 생성형 인공지능 서비스 채택 선행요인에 관한 탐색적 연구 (주성희, 임연수, 김예빈)
- 기본연구 25-13 주파수 경매에서 네트워크 구축조건 도입 방식에 대한 연구(백소성)
- 기본연구 25-14 AI 기술주권과 국가경쟁력 제고 방안 연구

(이경선, 김성옥, 이경은, 오장민, 윤혜선)

기본연구 25 - 15 플랫폼 확산의 인플레이션 영향에 대한 연구(김경은, 조수진, 심명규)

기본연구 25 - 16 AI 시대의 포용과 상생을 위한 사회적 의제 연구

(문아람, 문정옥, 조성은, 연소라, 김휘홍, 이으뜸, 전민경, 신진호)

기본연구 25 - 17 ICT 혁신이 사회·경제에 미치는 영향(최지은, 서영선, 노희용)

기본연구 25 - 18 차세대 통신기술이 통신시장 경쟁구조 및 네트워크 투자에 미치는 영향
분석(김민희, 김경모)



● 저 자 소 개 ●

이 경 선

- Purdue Univ. 경영학 박사
- 현 정보통신정책연구원 연구위원

이 경 은

- 서울대학교 경영학 석사
- 현 정보통신정책연구원 부연구위원

윤 혜 선

- 현 한양대학교 법학전문대학원/
인공지능대학원 교수

김 성 옥

- 중국사회과학원 경영학과 박사
- 현 정보통신정책연구원 연구위원

오 장 민

- 현 성신여자대학교 AI융합학부 교수

기본연구 25-14

AI 기술주권과 국가경쟁력 제고 방안 연구

2025년 12월 일 인쇄

2025년 12월 일 발행

발행인 이 상 규

발행처 정보통신정책연구원

충청북도 진천군 덕산읍 정통로 18

TEL: 043-531-4114 FAX: 043-535-4695~6

인쇄 인성문화

ISBN 979-11-7000-437-0 93320

〈비매품〉