방통융합정책연구 KMCC-2025-10

AI 이용자 보호 및 공정한 산업환경 조성 방안 연구

A Study on Measures to Protect AI Users and Create a Fair Industrial Environment

2025. 12

연구기관 : 정보통신정책연구원

방송미디어통신위원회
Korea Media and Communications Commission

방통융합정책연구 KMCC-2025-10

AI 이용자 보호 및 공정한 산업환경 조성 방안 연구

(A Study on Measures to Protect AI Users and Create a Fair Industrial Environment)

김현수/이선희/김민진

2025. 12

연구기관 : **정보통신정책연구원**

방송미디어통신위원회
Korea Media and Communications Commission

이 보고서는 2025년도 방송미디어통신위원회 방송통신발전기금 방송통신 융합 정책연구사업의 연구결과로서 보고서 내용은 연구자의 견해이며, 방송미디어통신위원회의 공식입장과 다를 수 있습니다.

제 출 문

방송미디어통신위원회 위원장 귀하

본 보고서를 『AI 이용자 보호 및 공정한 산업환경 조성 방안 연구』의 연구결과보고서로 제출합니다.

2025년 12월

연 구 기 관 : 정보통신정책연구원
총괄책임자 : 김현수 선임연구위원
참여연구원 : 이선희 부연구위원
김민진 부연구위원

목 차

표 목 차

그 림 목 차

요 약 문

1. 제 목

AI 이용자 보호 및 공정한 산업환경 조성 방안 연구

2. 연구 목적 및 필요성

인공지능(AI) 기술의 발전과 서비스의 대중화에 따라 사회·경제 전반에서 AI의 영향력이 빠르게 확대되고 있다. 특히, 생성형 AI, 자동화된 의사결정, 대화형 챗봇, 추천 알고리즘 등은 업무 생산성 향상, 효율적 의사결정 지원, 정보 접근성 제고, 일상 편의 증대 등 다양한 긍정적 효과를 제공하며, AI를 '범용 기반기술'로 정착시키고 있다. 하지만 AI 서비스 확산은 새로운 위험과 사회적 비용도 수반한다. AI는 기술적 특성상 오류와 환각, 편향과 차별, 설명가능성 부족, 안전성 취약, 프라이버시 침해, 저작권·초상권 등 권리침해, 허위조작정보·불법유해정보의 대량 생성 및 확산을 촉진할 수 있다.

AI 생태계의 산업환경 측면에서는 빅테크 기업을 중심으로 데이터, 컴퓨팅, 클라우드, 모델, 플랫폼이 결합된 가치사슬이 형성되며 시장구조가 빠르게 재편되고 있다. 네트워크 효과와 규모의 경제, 대규모 데이터 접근성, 컴퓨팅 자원의 집중은 경쟁 제한 가능성과 이용자 선택권 약화를 심화시킬 수 있다.

이에 본 연구는 AI 서비스 이용자의 권익을 보호하고 공정한 AI 산업환경을 조성하기 위한 중장기 정책과제와 추진 로드맵을 제시하는 것을 목적으로 한다.

3. 연구의 구성 및 범위

본 보고서의 구성은 다음과 같다. 제2장에서는 AI 서비스 이용 및 시장 동향을 조사하여

확산 속도와 이용행태, 산업 생태계 구조를 파악한다. 제3장에서는 주요국의 AI 서비스 정책 및 법제 동향을 분석하여 국제적 흐름과 시사점을 정리한다. 이상의 분석을 토대로 제4장에서는 'AI 서비스 이용자 정책 종합계획(안)'을 제시하고 안전하고 건강한 이용환경 조성, AI 서비스와 이용자의 동반성장, 이용자 보호 인프라 구축이라는 3대 추진방향 아래 세부 추진과제를 구체화한다. 마지막으로 제5장에서 정책적 함의와 추진상 유의점, 기대효과를 종합하여 결론을 제시한다.

4. 연구 내용 및 결과

AI는 이미 사회 전 영역에서 필수 인프라로 기능하고 있으며, 생성형 AI의 대중화와 함께 이용자 접점에서 발생하는 피해와 위험 또한 급속히 확대되고 있다. 오류·편향·환각과 같은 기술적 위험은 물론, 딥페이크 범죄, 허위조작정보 및 불법유해정보 확산, 프라이버시 침해, 알고리즘 추천에 따른 확증편향 강화 등은 이용자의 권익을 실질적으로 위협하고 사회적 신뢰를 약화시킬 수 있다. 동시에 데이터, 컴퓨팅, 클라우드, 모델, 플랫폼이 결합된 AI 생태계의 구조적 특성은 시장 집중과 정보 비대칭을 심화시키며, 공정 경쟁 질서와 이용자 선택권을 약화시킬 우려를 내포한다. 따라서 향후 AI 정책은 혁신 촉진과 이용자 보호를 상호 대립하는 가치로 보지 않고, '신뢰 기반 혁신'을 목표로 예방적 안전장치와 공정한 시장 질서를 함께 구축하는 방향으로 설계되어야 한다. 이러한 문제의식에 기반하여 본 연구는 다음과 같이 AI 서비스 이용자 정책 방향을 제안하였다.

첫째, 안전하고 신뢰할 수 있는 AI 활용 정보 유통환경 조성을 위해 딥페이크 성범죄물 등 불법유해정보의 생성·유통 차단 체계를 강화한다. 딥페이크 성범죄물 신고 접수 시 심의 전 임시 차단 근거 마련, 자율규약 제정, 콘텐츠 출처 확인 지원 등을 추진하고, 플랫폼에 신고 및 처리 절차 구축과 반복 위반자에 대한 서비스 제한 또는 수익화 제한 등 적절한 조치를 취할 것을 요구한다. 다만, 오차단 방지를 위해 사전 고지 및 이의제기 절차를 마련한다. 아동·청소년 보호를 위해 유해정보, 그루밍, 중독성 추천 등에 대한 위험평가 및 완화조치, 부모 통제권 강화 등을 추진한다.

둘째, 자유롭고 공정한 AI 생태계 구축을 위해 AI 시장 및 서비스 이용 현황에 대한 정기

실태조사로 생태계 건강성을 진단하고 불공정행위 모니터링을 강화한다. 이용자 측면에서는 커넥티드 서비스 확산에 맞춰 데이터 통제권을 확대하고, 추천 알고리즘의 부작용 대응을 위해 알고리즘 투명성 강화와 초대형 사업자 대상 개인화 선택권 보장을 추진한다.

셋째, 자율 기반 AI 규범체계 확립 및 집행 실효성 강화를 위해 AI 민관협의회를 구성하고, 영역별 자율규약 수립-이행평가-공개로 실효성을 높인다. 기술 변화에 유연한 목표·원칙 중심 규범을 확대하고, 생성형 AI 가이드라인 점검 및 AI 에이전트 가이드라인 제정을 추진한다. 해외사업자에 대해서도 불법정보 대응 등 직접 집행력을 강화하고, 국제기구·주요국과 협력해 글로벌 스탠더드를 선도한다.

본 연구가 제시한 AI 서비스 이용자 정책 방향은 'AI 안전벨트'로서 이용자 보호를 두텁게 하면서도 산업 혁신이 지속될 수 있는 토대를 마련하는 데 초점을 둔다. 단기적으로는 AI 활용정보의 신뢰성 확보, 피해 신고 및 상담 체계 고도화, 자율규제의 실효성 제고 등에 우선순위를 두고, 중기적으로는 이용자 권리의 제도화, 알고리즘 투명성 및 선택권 강화, 평가·인증 체계 정착과 분쟁조정 메커니즘 확립을 추진할 필요가 있다. 장기적으로는 AI 기술의 자율성과 물리적 확장(에이전틱/피지컬/범용 AI)에 대비하여 책임 배분 원칙과 위험 대응 체계를 고도화하고 국제 공조를 확대함으로써 정책의 미래 적합성을 확보해야 한다.

5. 기대효과

본 연구의 결과는 AI 서비스 이용자 정책 종합계획 수립을 위한 기초자료로, AI 기술 발전 및 확산에 따른 역기능에 선제적으로 대응하고 이용자 중심의 규범·윤리 체계를 확립하는 데에 활용될 수 있다. 이용자의 권익을 보호하고 신뢰 기반의 AI 산업 혁신을 촉진하여 궁극적으로 AI 산업의 지속 가능한 성장을 뒷받침하는 중장기 정책 마련에 기여할 것이다.

SUMMARY

1. Title

A Study on Measures to Protect AI Users and Create a Fair Industrial Environment

2. Objective and Importance of Research

With the advancement of artificial intelligence (AI) technology and the popularization of services, the influence of AI is rapidly expanding across society and the economy. In particular, generative AI, automated decision-making, conversational chatbots, and recommendation algorithms offer various positive benefits, including improved work productivity, efficient decision-making support, enhanced information accessibility, and increased convenience in daily life, solidifying AI as a universal foundational technology. However, the proliferation of AI services also entails new risks and social costs. Due to its technological nature, AI can be prone to errors and illusions, bias and discrimination, lack of explainability, security vulnerabilities, privacy violations, copyright and portrait rights infringements, and the mass creation and spread of fabricated and illegal information.

In terms of the industrial environment of the AI ecosystem, a value chain combining data, computing, cloud computing, models, and platforms, centered around big tech companies, is rapidly restructuring the market structure. Network effects, economies of scale, access to large-scale data, and the concentration of computing resources can exacerbate the potential for limited competition and diminished user choice.

Accordingly, this study aims to present mid- to long-term policy tasks and a roadmap for implementation to protect the rights and interests of AI service users and create a fair AI industry environment.

3. Contents and Scope of the Research

This report is structured as follows. Chapter 2 examines AI service usage and market trends to understand the pace of diffusion, usage patterns, and the structure of the industrial ecosystem. Chapter 3 analyzes AI service policies and legislative trends in major countries to summarize international trends and implications. Based on this analysis, Chapter 4 presents a "Comprehensive AI Service User Policy Plan (Draft)" and specifies detailed implementation tasks under three key directions: creating a safe and healthy user environment, fostering shared growth between AI services and users, and establishing user protection infrastructure. Finally, Chapter 5 summarizes policy implications, considerations for implementation, and expected outcomes, presenting a conclusion.

4. Research Results

AI is already functioning as essential infrastructure across all areas of society, and with the popularization of generative AI, the harm and risks arising from user interactions are rapidly expanding. In addition to technical risks such as errors, bias, and hallucinations, deepfake crimes, the spread of misinformation and illegal harmful information, privacy violations, and the strengthening of confirmation bias due to algorithmic recommendations pose substantial threats to users' rights and erosion of social trust. At the same time, the structural characteristics of the AI ecosystem, which combines data, computing, cloud computing, models, and platforms, exacerbate market concentration and information asymmetry, raising concerns that this could weaken fair competition and user choice. Therefore, future AI policies should not view innovation promotion and user protection as conflicting values, but rather be designed to simultaneously establish preventative safeguards and a fair market order, with the goal of "trust-based innovation." Based on these concerns, this study proposes the following policy directions for AI service users:

First, to foster a safe and reliable AI-powered information distribution environment, the system to block the creation and distribution of illegal and harmful information, such as deepfake sexual crime material, should be strengthened. When receiving reports of deepfake sexual crimes, we will establish grounds for temporary blocking prior to review, establish self-regulation, and support content source verification. We will also require platforms to establish reporting and processing procedures and take appropriate measures, such as service restrictions or monetization restrictions, for repeat offenders. However, to prevent misidentification, we will establish prior notification and objection procedures. To protect children and adolescents, we will pursue risk assessments and mitigation measures for harmful information, grooming, and addictive recommendations, as well as strengthen parental control.

Second, to build a free and fair AI ecosystem, we will assess the health of the ecosystem through regular surveys of AI market and service usage and strengthen monitoring of unfair practices. On the user side, we will expand data control in line with the proliferation of connected services. To address the negative impacts of recommendation algorithms, we will strengthen algorithm transparency and guarantee personalized options for large businesses.

Third, to establish an autonomous AI normative system and strengthen enforcement effectiveness, we will establish a public-private AI council. This will be enhanced through the establishment, implementation evaluation, and disclosure of self-regulation regulations by sector. We will expand flexible, goal- and principle-based norms that adapt to technological changes, and promote the review of generative AI guidelines and the establishment of AI agent guidelines. We will also strengthen direct enforcement against overseas operators, including by responding to illegal information, and collaborate with international organizations and major countries to lead global standards.

The AI service user policy direction proposed in this study focuses on strengthening user protection as an "AI safety belt" while establishing a foundation for sustainable industrial innovation. In the short term, we will prioritize ensuring the reliability of AI-related information, enhancing the reporting and consultation system for damages, and enhancing

the effectiveness of self-regulation. In the medium term, we will prioritize institutionalizing user rights, strengthening algorithm transparency and choice, establishing an evaluation and certification system, and establishing a dispute resolution mechanism. In the long term, we will prepare for the autonomy and physical expansion of AI technology (agentic, physical, and general-purpose AI), enhancing responsibility allocation principles and risk response systems, and expanding international cooperation to ensure the future relevance of our policies.

5. Expectations

The results of this study serve as foundational data for developing a comprehensive policy plan for AI service users. They can be utilized to proactively address the negative impacts of AI technology development and expansion and establish a user-centered normative and ethical system. This will contribute to the development of mid- to long-term policies that protect user rights and promote trust-based innovation in the AI industry, ultimately supporting the sustainable growth of the AI industry.

CONTENTS

제1장 서 론

인공지능(AI) 기술의 발전과 서비스의 대중화에 따라 사회·경제 전반에서 AI의 영향력이 빠르게 확대되고 있다. 특히, 생성형 AI, 자동화된 의사결정, 대화형 챗봇, 추천 알고리즘 등은 업무 생산성 향상, 효율적 의사결정 지원, 정보 접근성 제고, 일상 편의 증대 등 다양한 긍정적 효과를 제공하며, AI를 범용 기반기술로 정착시키고 있다. 최근에는 AI가 단순한 도구를 넘어 플랫폼, 콘텐츠, 네트워크, 기기 전반에 내재화되어 이용자가 의식하지 않는 순간에도 AI의 판단과 추천에 의해 정보가 배열되고 콘텐츠가 유통되는 환경이 상시화되고 있다.

AI 서비스 확산은 새로운 위험과 사회적 비용도 수반한다. AI는 기술적 특성상 오류와 환각, 편향과 차별, 설명가능성 부족, 안전성 취약, 프라이버시 침해, 저작권·초상권 등 권리 침해, 허위조작정보 및 불법유해정보의 대량 생성 및 확산을 촉진할 수 있다. 특히, 생성형 AI는 생성 비용의 급격한 하락과 유통 속도의 가속이 결합되어 불법유해정보의 생산과 전파를 구조적으로 용이하게 만들고, 딥페이크 범죄와 같은 중대한 피해를 현실화시키고 있다. 이 과정에서 이용자는 AI 서비스를 신뢰하기 어렵고, 피해가 발생하더라도 책임 주체가 분산되어 신속한 구제가 어려운 상황에 직면한다. 나아가 AI 서비스가 금융·의료·교육·방송·통신 등 필수 영역의 기반 인프라로 편입될수록 개별 피해를 넘어 사회적 신뢰의 훼손과 민주적 의사결정 기반의 약화라는 거시적 위험으로 연결될 수 있다.

AI 생태계의 산업 환경 측면에서는 빅테크 기업을 중심으로 데이터, 컴퓨팅, 클라우드, 모델, 플랫폼이 결합된 가치사슬이 형성되며 시장구조가 빠르게 재편되고 있다. 네트워크 효과와 규모의 경제, 대규모 데이터 접근성, 컴퓨팅 자원의 집중은 경쟁 제한 가능성과 이용자 선택권 약화를 심화시킬 수 있다. 더불어 AI 서비스 이용 과정에서 요금, 품질, 데이터 활용 등에서 정보 비대칭이 확대되어 이용자 피해가 구조적으로 반복될 우려도 크다. 결국 AI 정책은 진흥 대 규제의 이분법을 넘어 이용자 보호와 공정 경쟁 질서 확립을 통해 신뢰 기반의 혁신을 촉진하는 방향으로 설계될 필요가 있다.

이용자 보호는 단순한 사후 구제나 이용약관 중심의 형식적 고지를 의미하지 않는다. AI 서비스의 생애주기(개발-배포-유통-이용) 전반에서 위험을 예방하고, 정보 비대칭을 완화하며, 이용자의 권리(알권리, 자기결정권, 데이터 통제권, 설명요구권 등)를 실질적으로 보장하는 제도적·기술적·거버넌스 체계를 포괄한다. 동시에 공정한 산업 환경 조성은 혁신기업의 진입과 경쟁을 보장하고, 불공정 행위 및 이용자 이익 저해 행위를 억제하며, 자율규제가 실효적으로 작동하도록 공공의 감독, 지원, 정책환류를 결합하는 종합적 접근을 포함한다.

본 연구는 AI 서비스 이용자의 권익을 보호하고 공정한 AI 산업 환경을 조성하기 위한 중장기 정책과제와 추진 로드맵을 제시하는 것을 목적으로 한다. 이를 위해 제2장에서는 AI 서비스 이용 및 시장 동향을 조사하여 확산 속도와 이용행태, 산업 생태계 구조를 파악한다. 제3장에서는 주요국의 AI 서비스 정책 및 법제 동향을 분석하여 국제적 흐름과 시사점을 정리한다. 이상의 분석을 토대로 제4장에서는 AI 서비스 이용자 정책 방향을 제시하고 안전하고 신뢰할 수 있는 AI 활용 정보 유통환경 조성, 자유롭고 공정하며 혁신을 촉진하는 AI 생태계 구축, 자율 기반 AI 규범체계 확립 및 집행 실효성 강화라는 3대 추진방향 아래 핵심 추진과제를 구체화한다. 마지막으로 제5장에서 정책적 함의와 추진상 유의점, 기대효과를 종합하여 결론을 제시한다.

제 2 장 AI 서비스 이용 및 시장 동향

인공지능(Artificial Intelligence, 이하 "AI")은 일반적으로 인간의 추론, 의사결정, 창작 등 복잡한 작업을 수행할 수 있는 컴퓨터 시스템을 의미한다.[1] 이는 머신러닝, 딥러닝, 자연어 처리(NLP) 등 광범위한 기술을 포함하는 포괄적인 개념이다.[2] AI는 컴퓨터가 음성이나 작성된 언어를 확인하고, 이를 이해·번역·분석한 뒤 적절한 추천을 수행하는 기능을 포함하여, 다양한 고급 기능을 수행할 수 있도록 해주는 기술로 발전하고 있다.[3]

여러 국가의 법률과 국제기구에서는 AI를 인간이 정의한 목표에 따라 예측, 추천, 결정을 수행하는 기계 기반 시스템으로 정의하고 있으며, AI 시스템을 동일한 개념으로 혼용한다. 일례로 OECD(2019)는 AI 시스템을 사람이 정의한 일련의 목표에 대해 실제 또는 가상 환경에 영향을 미치는 예측, 추천 또는 결정을 내릴 수 있는 기계 기반 시스템으로 정의[4]하였으며, 미국 상업 및 무역 법률(15 USC § 9401)에서도 AI에 대해 주어진 일련의 인간이 정의한 목표에 대해 실제 또는 가상 환경에 영향을 미치는 예측, 추천 또는 결정을 내릴 수 있는 기계 기반 시스템을 의미한다고 하였다. 정리하면 AI 서비스는 AI 기술적 기반 위에서 이용자가 원하는 목적에 따라 특정 기능이나 결과를 제공하는 시스템 또는 플랫폼을 의미한다고 할 수 있다.

AI 서비스 중 특히 콘텐츠를 생성하는 생성형 AI가 대표적인 예로 부각되고 있다. AI 서비스에는 생성형 AI 외에도 분석형 AI[5]가 존재하나, 기술의 발달로 사실상 통합하여

---

1) NASA, "What is Artificial Intelligence?", https://www.nasa.gov/what-is-artificial-intelligence/

2) Coursera(2024.12.20), "What Is Artificial Intelligence? Definition, Uses, and Types", https://www.coursera.org/articles/what-is-artificial-intelligence

3) Google Cloud, "인공지능(AI)이란 무엇인가요?", https://cloud.google.com/learn/what-is-artificial-intelligence?hl=ko

4) OECD(2019), "Recommendation of the Council on Artificial Intelligence", OECD/LEGAL/0449.

5) 분석형 AI는 데이터 분석, 패턴 인식, 예측 모델링 등을 통해 의미 있는 인사이트 도출과 의사결정을 지원하는 AI 서비스를 의미(Harvard Business Review, 2024.12.13), "How Gen

제공되고 활용되는 추세이다.

1. AI 서비스 시장 규모 및 전망

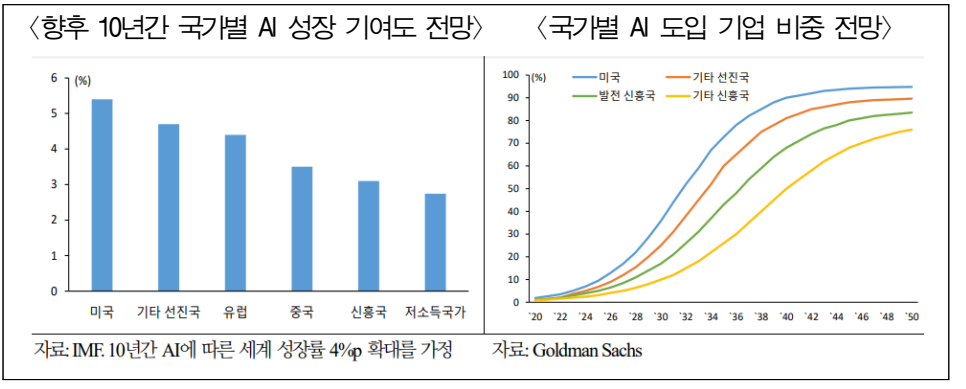
가. AI의 거시경제적 영향

인공지능(AI)은 향후 10년간 세계 경제에 가장 큰 변화를 가져올 기술로 주목받고 있다. Renaissance Macro(2025)에 따르면 실제로 미국에서는 2025년 AI 자본지출이 소비자 지출보다 GDP 성장에 더 큰 영향을 미친 것으로 분석되었다. 아울러 주요 국제기관들은 AI가 전 세계 경제성장률을 실질적으로 끌어올릴 것으로 전망한다. IMF는 AI 기술의 발전에 따라 총요소생산성(TFP)이 개선되면 향후 10년간 세계 성장률이 최소 1.3%p에서 최대 4.0%p까지 상승할 수 있다고 보았다. PwC는 AI 도입으로 향후 10년간 약 8%p의 성장률 증가를 예측하며, 낙관적 시나리오에서는 최대 15%포인트, 비관적 시나리오에서는 1%포인트의 효과가 있을 것으로 분석했다. 골드만삭스는 생성형 AI의 도입 확산이 세계 총생산을 7%포인트 이상 끌어올릴 수 있다고 평가했으며, IDC는 AI 관련 투자 확대가 약 20조 달러(세계 총생산의 17.7%) 규모의 경제 효과를 가져올 것으로 내다봤다. 또한, AI에 1달러를 투자할 경우 약 4.6달러의 부가가치가 창출된다고 분석했다. 국제금융센터(2025)는 이러한 기관들의 전망을 종합해, AI가 생산성과 자본 효율성을 높여 세계 성장률을 4~18%p까지 끌어올릴 수 있다고 평가했다.[6)]

AI and Analytical AI Differ — and When to Use Each", https://hbr.org/2024/12/how-gen-ai-and-analytical-ai-differ-and-when-to-use-each)

6) 국제금융센터(2025.8.14), "AI 발전에 따른 신흥국의 상대적 성장 지연 가능성".

〔그림 2－1〕 국가별 AI 성장 기여도 및 국가별 AI 도입 기업 비중 전망



자료: 국제금융센터(2025.8.14)

산업 측면에서 분야별 차이가 있으나 하이테크 산업에서 농업에 이르기까지 AI 기술이 산업 전반의 생산성 증대에도 기여할 것으로 전망된다(Mckinsey, 2023).[7)]

〈표 2－1〉 산업별 생성형 AI의 경제적 영향

(단위: 조 달러)

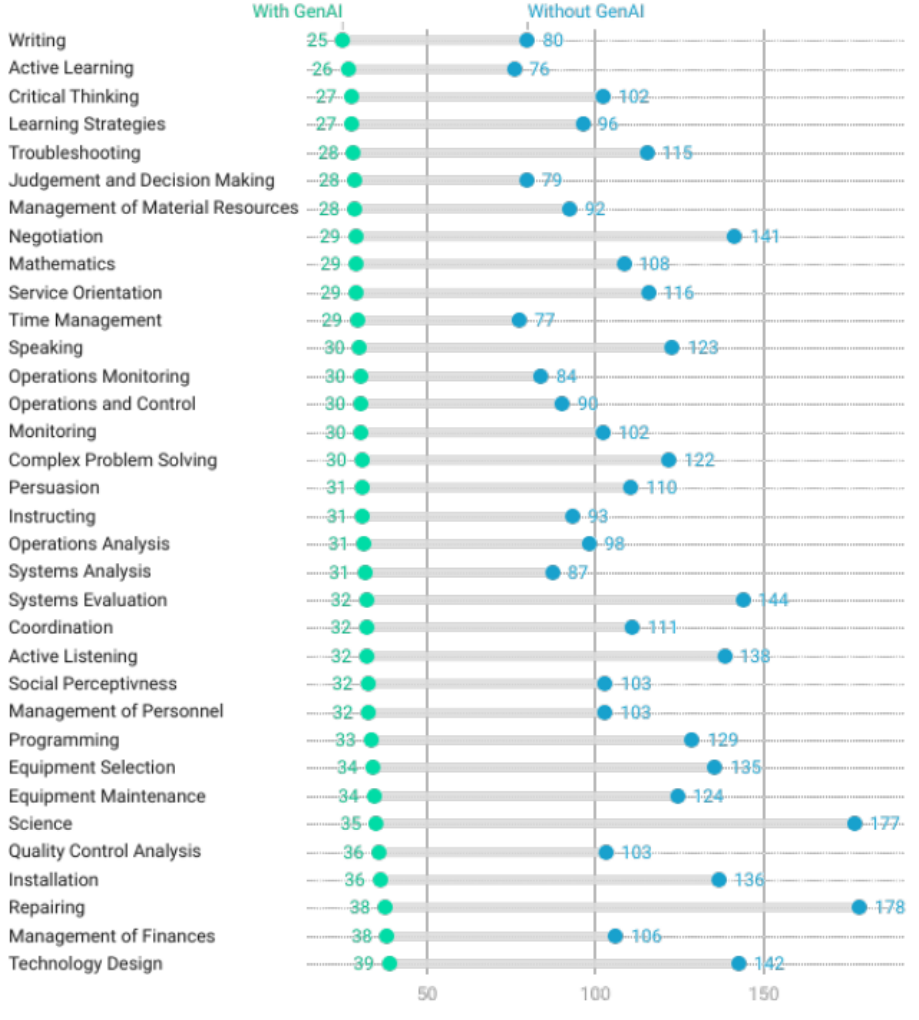
구분	총수익		구분	총수익		구분	총수익	
	최소	최대		최소	최대		최소	최대
하이테크	0.24	0.46	행정 및 전문 서비스	0.15	0.25	화학	0.08	0.14
소매업	0.24	0.39	에너지	0.15	0.24	미디어 및 엔터테인먼트	0.08	0.13
은행업	0.20	0.34	교육	0.12	0.23	공공 및 사회 부문	0.07	0.11
여행, 운송 및 물류	0.18	0.30	기초 소재	0.12	0.20	제약 및 의료 제품	0.06	0.11
첨단 제조업	0.17	0.29	부동산	0.11	0.18	통신	0.06	0.10
소비재	0.16	0.27	첨단 전자 및 반도체	0.10	0.17	보험	0.05	0.07
보건의료	0.15	0.26	건설업	0.09	0.15	농업	0.04	0.07

자료: Mckinsey & Company(2023.6)

7) Mckinsey & Company(2023.6), The economic potential of generative AI.

또한, 개인 및 조직 관점에서 AI 활용시 업무 소요시간을 최소 60% 단축 가능하여 생산성 증대 효과를 기대할 수 있을 것으로 전망되었다(Hartley et al., 2024).[8)]

〔그림 2－2〕 생성형 AI 활용 여부별 작업 완료 시간



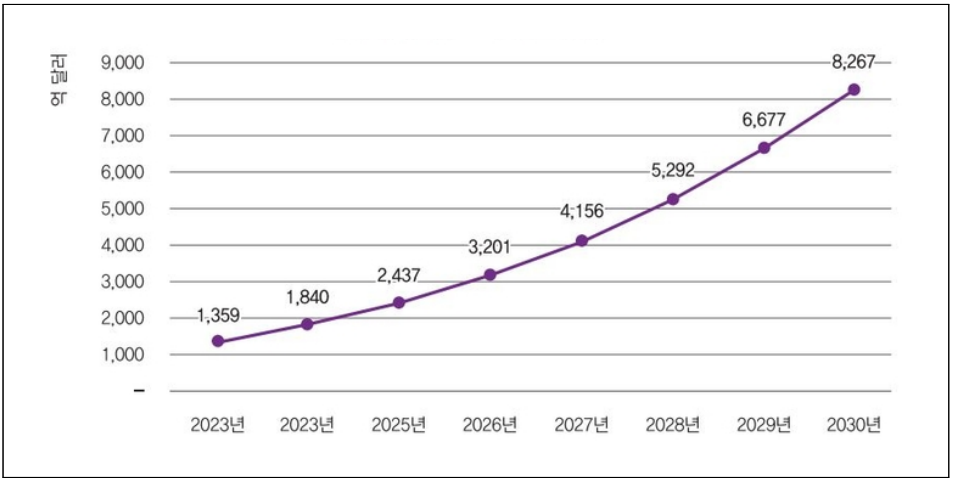
자료: Hartley et al.(2024)

8) Hartley, J., Jolevski, F., Melo, V., & Moore, B. (2024). The labor market effects of generative artificial intelligence. *Available at SSRN*.

나. AI 서비스 시장 규모

AI의 거시경제적 기여와 함께, AI 기술과 서비스를 중심으로 한 산업 자체의 성장세 또한 빠르게 확대되고 있다. AI 시장은 생성형 AI를 필두로 활발한 기술 및 인프라 투자 속에 지속 성장하고 있으며, 상용 AI 서비스의 이용 또한 확대되고 있다. 한국무역협회(2024)는 글로벌 AI 시장이 2023년 1,359억 달러에서 2030년 8,267억 달러로 연평균 29.4% 증가할 것으로 전망했다.[9] 또한 국내 AI 시장은 2023년 46억$에서 연평균 18.7% 증가하여 2030년에는 153억 달러에 이를 것으로 내다봤다.[10] 실제로 과기부와 소프트웨어 정책연구소는 AI 매출이 발생한 우리나라 기업체의 AI 매출액은 2023년 5.6조 원 대비 21.5% 증가하여 2024년에는 6.3조 원을 기록할 것으로 예상하였다.[11]

〔그림 2－3〕 글로벌 AI 시장 규모 전망



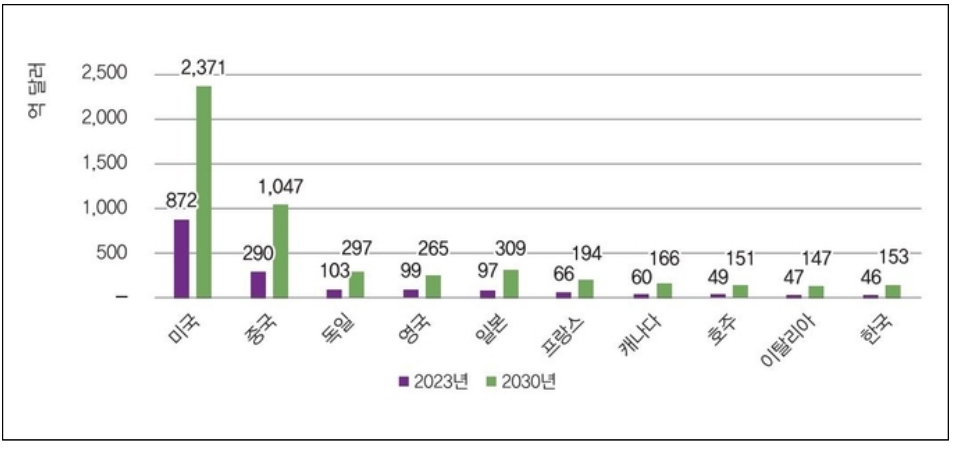
자료: 한국무역협회(2024)

9) 한국무역협회(2024.9.24), “AI로 뭉친 韓 기업, 원팀 전략으로 글로벌 공략”, https://www.kita.net/board/totalTradeNews/totalTradeNewsDetail.do?no=86509&siteId=1

10) 한국무역협회(2024.9.24), “AI로 뭉친 韓 기업, 원팀 전략으로 글로벌 공략”, https://www.kita.net/board/totalTradeNews/totalTradeNewsDetail.do?no=86509&siteId=1

11) 과학기술정보통신부·소프트웨어정책연구소(2025),『2024년 인공지능산업 실태조사』

〔그림 2－4〕 국가별 AI 시장 규모 및 전망

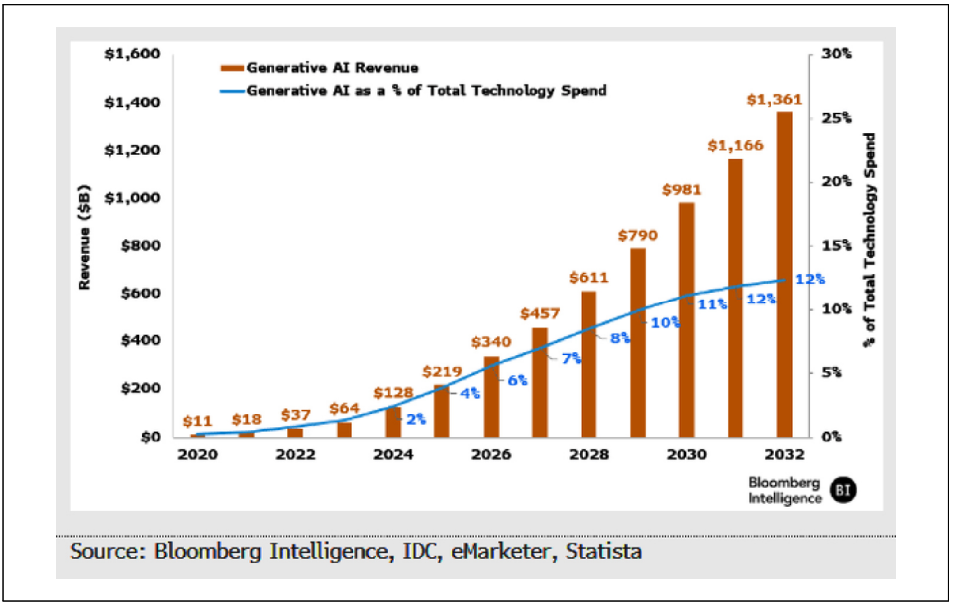


자료: 한국무역협회(2024)

Bloomberg(2024)[12]는 생성형 AI 시장이 2032년까지 약 1조 3천억 달러 규모에 이를 것으로 전망했다. 이에 따르면 2024년 현재 전체 기술 지출 중 AI 관련 지출은 약 2% 수준이나, AI 학습 중심 구조로의 전환이 본격화됨에 따라 2032년에는 전체 기술 지출의 약 12%까지 확대될 것으로 예상된다. 기술 지출의 주요 흐름은 AI 플랫폼의 학습 기능에 집중되고 있으며, 초기에는 서버 및 스토리지와 같은 물리적 자원에 대한 투자가 주를 이루지만, 중장기적으로는 클라우드 기반 인프라에 대한 지출이 시장을 주도할 것으로 분석된다.

12) Bloomberg(2024.3.8), “Generative AI races toward $1.3 trillion in revenue by 2032”, https://www.bloomberg.com/professional/insights/data/generative-ai-races-toward-1-3-trillion-in-revenue-by-2032/

〔그림 2－5〕 전 세계 생성형 AI 시장 규모와 기술 지출 추이(2020년~2032년)



자료: Bloomberg(2024.3)[13]

다. AI 서비스 이용 규모

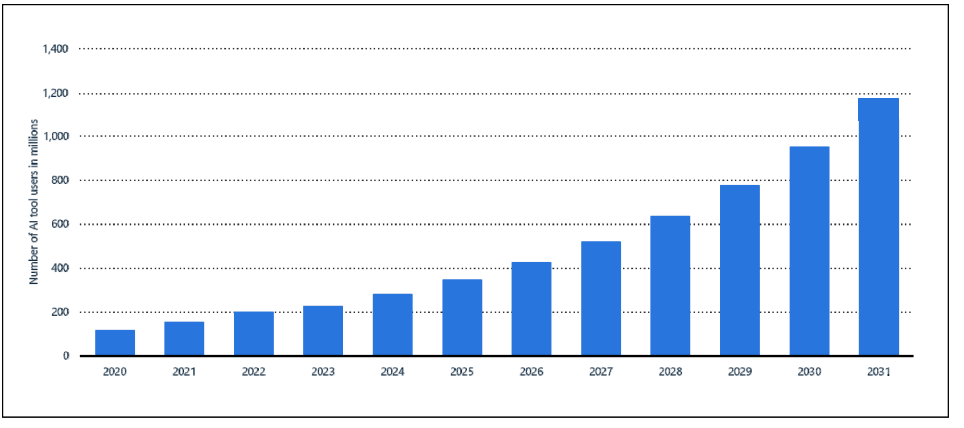
전 세계적으로 AI 도구 이용자 수는 빠른 속도로 증가하고 있다. Statista(2025)[14]에 따르면, 2024년 기준 전 세계 AI 도구 이용자 수는 약 2억 8,126만 명에 이를 것으로 예측된다. 이는 2020년 이후 약 1억 6,535만 명이 증가한 수치다. 또한, 2024년부터 2031년까지는 약 8억 9,118만 명의 사용자가 추가로 유입될 것으로 전망되었다.

13) Bloomberg(2024.3.8), "Generative AI races toward $1.3 trillion in revenue by 2032", https://www.bloomberg.com/professional/insights/data/generative-ai-races-toward-1-3-trillion-in-revenue-by-2032/

14) Statista(2025.8.15), "Number of AI tool users worldwide from 2020 to 2031"(https://www.statista.com/forecasts/1449844/ai-tool-users-worldwide, 최종접속일: 2025.8.19)

〔그림 2－6〕 전 세계 AI 도구 사용자 수 추이(2020년~2031년)

(단위: 백만 명)



자료: Statista(2025)

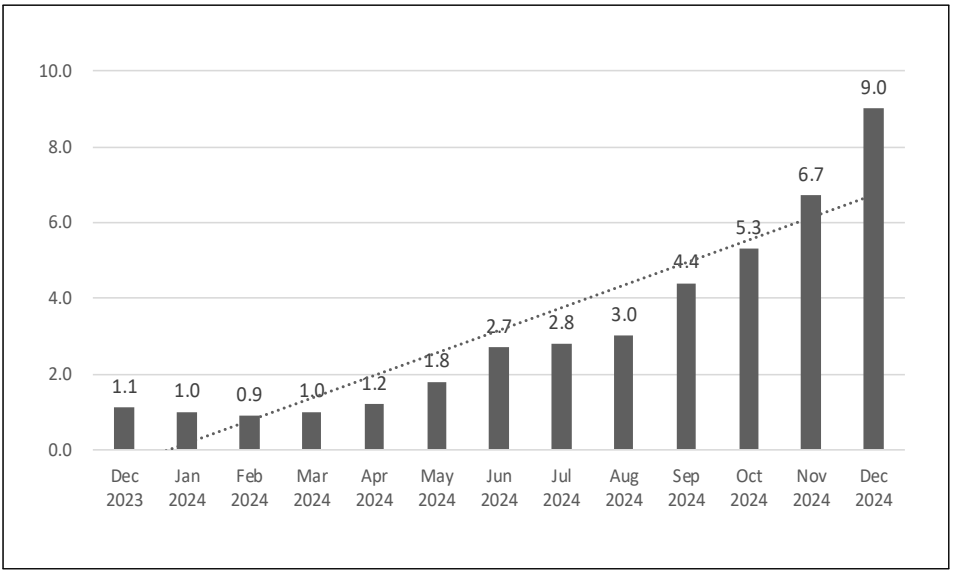
생성형 AI 역시 산업 전반에 걸쳐 도입 및 활용되고 있다. 국제조사기관들의 분석에 따르면, Fortune 500 기업의 92%가 OpenAI 기술을 활용 중이며, Deloitte는 글로벌 기업 임원의 94%가 AI가 운영 효율성을 높이는 데 기여할 것이라고 응답하였다.

국내의 경우에도 생성형 AI 활용이 빠르게 확산되고 있다. 국내 한 앱·리테일 분석 서비스의 조사 결과를 인용한 기사[15]에 따르면, 사용자당 월간 평균 사용 시간이 지난 1년간(2023년 12월~2024년 12월) 약 8배 이상 급증한 것으로 나타났다. 2024년 12월 기준 주요 생성형 AI 앱 중 가장 많은 사용 시간을 기록한 서비스는 ChatGPT(682분)이며, 그 뒤를 이어 에이닷(245분), 루닛(236분), Perplexity(59분), Microsoft Copilot(31분), Claude(12분) 순으로 집계되었다.

15) 한국경제(2025.1.7), ""682만명 우르르"…'챗GPT' 한국인이 가장 많이 쓴 생성형AI"

〔그림 2－7〕 국내 생성형 AI 월간 사용 시간 추이

(단위: 억 분)



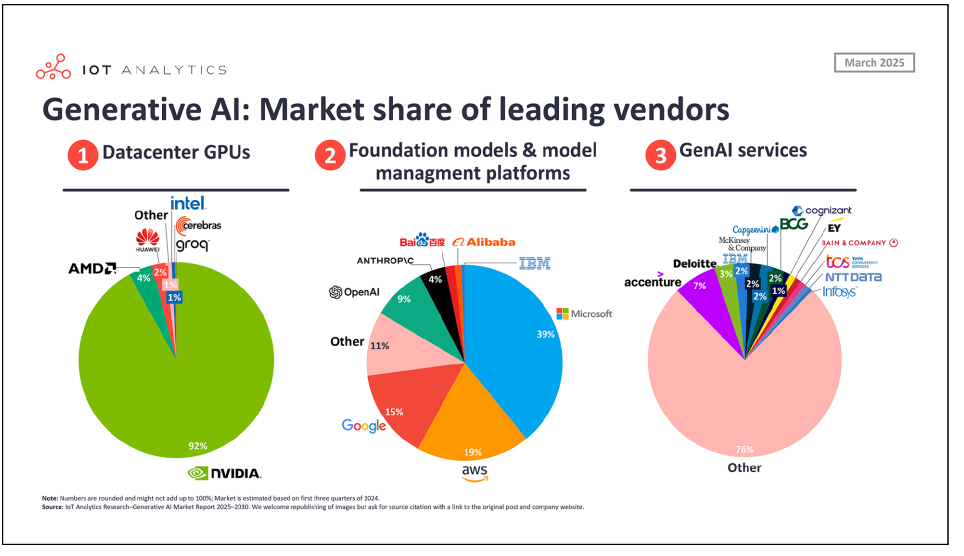
자료: 한경(2025.1.7), "'682만명 우르르'…'챗GPT' 한국인이 가장 많이 쓴 생성형AI"

2. AI 서비스 생태계 구조 및 주요 참여자

시장조사기관 IoT Analytics(2025)[16]는 AI 시장의 생태계를 기술 개발부터 최종 서비스 제공에 이르기까지의 과정을 기준으로 Data Center GPUs, Models and Platforms, Services의 세 가지 단계로 구성된 가치사슬 구조로 구분한다. 이 구조는 기술이 어떻게 공급되고 조합되며, 사용자에게 전달되는지를 단계적으로 보여주며, 각 단계는 고유한 기능과 시장 내 핵심 기업들로 구성된다.

16) IoT Analytics(2025.3), "The leading generative AI companies" (https://iot-analytics.com/leading-generative-ai-companies/, 최종접속일: 2025.8.4)

〔그림 2-8〕 AI 시장 생태계



자료: IoT Analytics(2025.3.4)

가. Data center GPUs

Data Center GPUs는 대규모 AI 모델을 학습하고 실행하는 데 필요한 고성능 연산 자원을 제공한다. 반도체 칩셋 제조사 및 하드웨어 인프라 제공 기업들로 구성되며 이들은 AI 기술의 성능 한계를 결정짓는 물리적 기반을 형성한다. 대표적인 기업인 NVIDIA는 글로벌 AI 전용 GPU 시장에서 80% 이상의 독보적인 점유율을 나타내고 있다. 국내 기업으로는 삼성전자와 SK하이닉스가 메모리 반도체를 중심으로 AI 인프라에 기여하고 있으나, GPU 또는 AI 전용 연산 칩 개발 및 공급 측면에서는 영향력이 제한적이다.

1) 글로벌 기업

가) NVIDA

IoT Analytics(2025)[17]에 따르면, 2024년 미국의 반도체 기업 NVIDIA는 글로벌 데이터센터 GPU 시장에서 92%의 압도적인 시장 점유율을 기록하며 2023년에 이어 지배적

17) IoT Analytics(2025.3), "The leading generative AI companies"(https://iot-analytics.com/leading-generative-ai-companies/, 최종접속일: 2025.8.4)

위치를 유지하였다. 데이터센터 GPU 부문 연간 매출은 1,150억 달러를 초과하였으며, 이는 전년 대비 142% 증가한 수치로 사상 최대 실적이다.

IoT Analytics(2025)[18]는 NVIDIA가 미래형 AI 데이터센터의 올인원 공급자(all-in-one supplier)를 목표로 두고, 지속적인 GPU 아키텍처 및 소프트웨어 혁신은 회사가 시장 리더십을 유지하는 데 기여해왔다고 분석했다. 실제로 NVIDIA의 AI에 특화된 가속기(NVIDIA H100 텐서 코어 GPU, 최신 Blackwell 아키텍처 GPU)는 데이터센터 GPU의 표준으로 자리하고 있으며, NVIDIA의 CUDA[19] 병렬 컴퓨팅 플랫폼과 프로그래밍 모델은 핵심적인 차별화 요소로 남아있다. 특히, NVIDIA는 20년간 CUDA 개선에 투자해 왔으며 선도적인 소프트웨어 아키텍처이자 개발자 생태계로 발전시켰다. 보고서는 이것이 NVIDIA의 기술 진입장벽과 전환비용을 높인 핵심 요인으로, NVIDIA의 지배적 위치를 유지할 핵심 요인 중 하나로 평가한다. 아울러 NVIDIA는 AI 중심의 차세대 데이터센터(일명 AI factory)[20] 개념을 제시하며 이를 위한 통합 하드웨어 솔루션에 대한 투자도 확대하고 있다. 이더넷 스위치, AI 최적화 라우터, CPU, 고성능 AI 가속기 등 네트워킹 및 컴퓨팅 전반에 걸친 제품군을 강화하고 있으며 자사의 CUDA 플랫폼과의 통합을 통해 생태계 기반의 시장지배력을 더욱 확대해 나가고 있다.

나) AMD

미국의 반도체 기업 AMD는 NVIDIA의 첫 번째 실질적인 경쟁자로서 데이터센터 GPU 시장에서 존재감을 확대하고 있다. 2024년, AMD의 데이터센터 GPU 부문은 전년 대비 179% 성장하였으며, 이에 따라 시장점유율은 2023년 3%에서 2024년 4%로 1%p 상승하였다.[21]

---

18) IoT Analytics(2025.3), "The leading generative AI companies"(https://iot-analytics.com/leading-generative-ai-companies/, 최종접속일: 2025.8.4)

19) Compute Unified Device Architecture의 약자로, NVIDIA GPU에서 일반적인 계산을 수행할 수 있도록 해주는 기술을 의미함.

20) 전통적인 데이터센터와 달리 AI 팩토리는 데이터를 저장하고 처리하는 것 이상으로, 대규모로 인텔리전스를 만들어 원시 데이터를 실시간 인사이트로 변환함. 이를 통해 기업과 국가는 가치 창출의 속도를 극적으로 가속화 해 AI를 장기 투자에서 즉각적인 경쟁 우위의 원동력으로 전환 가능함(자료: https://blogs.nvidia.co.kr/blog/ai-factory/)

IoT Analytics(2025)[22]에 따르면 AMD의 AI 전략은 개방형(Open), 종단 간(End-to-End), 비용 효율성(Cost-Effective)을 핵심 기조로 하는 인프라 제공에 초점을 두고 있다. 특히 2024년 12월 기준으로 AMD는 자사의 주력 AI GPU인 Instinct MI300X를 Meta, Microsoft, Oracle 등 주요 글로벌 기업에 총 307,000개가량 출하한 것으로 추정된다. AMD는 경쟁사 대비 가격 경쟁력이 있는 AI GPU 제품을 제공하고 있으며, 자사의 병렬 컴퓨팅 플랫폼인 ROCm(Radeon Open Compute)을 중심으로 소프트웨어 생태계 확대에도 힘쓰고 있다.

2) 국내 AI 반도체용 메모리 공급 기업

AI 반도체 산업과 관련해 우리나라에서는 삼성전자와 SK하이닉스가 핵심 역할을 하고 있다. 이들 기업은 AI GPU를 직접 설계하거나 개발하지는 않지만, AI GPU에 필수적인 HBM 메모리를 공급하고 일부 기업의 AI 칩을 위탁 생산함으로써 AI 반도체 생태계에 인프라를 제공하고 있다.

가) 삼성전자

삼성전자는 AI 반도체 부품을 공급하며 중요한 역할을 하고 있다. 크게 두 가지 사업 분야에서 경쟁력을 보여주고 있는데 메모리 반도체와 파운드리(반도체 위탁생산)가 그것이다.

먼저 메모리 반도체 분야에서 삼성전자는 AI 연산에 필수적인 고대역폭 메모리(HBM)와 고용량 D램과 같은 제품을 만들어 AI 서버나 데이터센터에 납품한다. 지디넷코리아(2025)[23]에 따르면 삼성전자는 고성능 AI 연산에 필수적인 초고속·대용량 메모리 제품(HBM, DDR5 등)의 공급을 전년 대비 두 배 확대할 계획이며, 최신 규격 제품의 판매를 확대하는 동시에 차세대 제품의 개발과 양산도 병행하고 있다. AI 기능을 기기 내부에서 직접 수행하는 온디바이스 AI 기반 신제품의 출시와 함께 서버 인프라에 대한 기업들의 투자가 지속될 것으로 예상되며, 이에 따라 메모리 수요가 회복되고 고성능 반도체 수요 역시 견고하게

21) IoT Analytics(2025.3), "The leading generative AI companies"(https://iot-analytics.com/leading-generative-ai-companies/, 최종접속일: 2025.8.4)

22) IoT Analytics(2025.3), "The leading generative AI companies"(https://iot-analytics.com/leading-generative-ai-companies/, 최종접속일: 2025.8.4)

23) 지디넷코리아(2025.1.31), "삼성전자, HBM 공급량 2배 확대...'AI 반도체'에 사활"

유지될 것으로 전망된다.

파운드리 분야에서 삼성전자는 AI 반도체를 설계하는 기업들로부터 주문을 받아 직접 칩을 제조한다. 2025년부터 최신 제조공정 기반의 양산을 시작하였으며, 해당 공정에 필요한 설계 도구는 고객사에 제공된 상태이다. 2026년부터 본격적인 대량 생산이 예정되어 있다. 이러한 전략의 결과, 삼성전자는 2025년 7월 약 23조 원 규모의 테슬라 AI 반도체 위탁생산 계약을 체결하며 반도체 사업의 새로운 성장 동력을 확보하였다.[24)]

나) SK하이닉스

SK하이닉스는 고성능 메모리 반도체의 핵심 공급자로서 글로벌 시장에서의 입지를 강화하고 있다. 특히 AI 연산 효율을 좌우하는 고대역폭메모리(HBM) 분야에서 독보적인 기술력과 점유율을 확보하고 있다. 트렌드포스 조사를 인용한 매일일보(2025)[25)]에 따르면, SK하이닉스는 2023년 기준 글로벌 HBM 시장점유율 52.5%로 1위를 차지하였으며, 이는 삼성전자(42.4%)와 마이크론(5.1%)을 앞서는 수치이다. 이 같은 성과는 생성형 AI의 확산으로 인해 AI 서버 및 데이터센터에 대한 수요가 급증하는 상황에서 HBM이 핵심 부품으로 각광받고 있다는 점에서 더욱 주목할 만하다.

시장 내 우위를 바탕으로 SK하이닉스는 고성능 메모리 제품군의 확대를 통해 AI 반도체 분야에서의 기술 경쟁력을 지속적으로 강화하고 있다. 매일일보(2025)[26)]는 SK하이닉스가 2025년 3월 엔비디아에 차세대 제품인 HBM4 샘플을 조기 공급하였으며, 하반기에는 HBM4 양산을 위한 준비를 마무리할 계획임을 보도했다. 이미 HBM3E 양산을 완료한 데 이어 HBM4까지 제품군을 확장함으로써 SK하이닉스는 고성능 메모리 시장 내 리더십을 공고히 하고 있다. 업계는 이러한 전략을 통해 SK하이닉스가 AI 반도체 생태계에서 기술력과 공급 역량을 갖춘 기업으로 자리매김할 것으로 전망하고 있다.

나. Models and Platforms

Models and Platforms는 AI 모델의 알고리즘 기능이 구현되는 구간으로, 기초 모델

24) 인공지능신문(2025.7.28), “일론 머스크, 삼성전자와 23조원 규모 테슬라 AI6 칩 계약 체결… 차세대 AI반도체 생산 본격화”

25) 매일일보(2025.6.17), “기회의 파도 탄 SK하이닉스, AI반도체 입지 강화 청신호”

26) 매일일보(2025.6.17), “기회의 파도 탄 SK하이닉스, AI반도체 입지 강화 청신호”

(Foundation Models)과 이를 외부에 제공하는 플랫폼 및 API 인프라가 포함된다. 최근에는 초거대 언어 모델(LLM)과 멀티모달 모델 등 복합적인 AI 모델이 중심을 이루며 이 단계의 기업들은 모델 개발부터 배포, 관리까지 AI 기술의 핵심 주체가 된다. OpenAI는 GPT 시리즈를 통해 생성형 AI의 대중화를 견인하고 있으며, Microsoft는 Azure를 통해 상업용 LLM 플랫폼 시장에서 주요한 위치를 차지하고 있다. Google은 Gemini 모델을 중심으로 자체 생태계를 확장하고 있다. Microsoft, Google, AWS 등은 자사 클라우드 기반 플랫폼을 통해 API 형태로 모델을 서비스하며 생산-배포-운영 전 주기에 걸친 플랫폼 생태계를 강화하고 있다. 국내에서는 플랫폼사, 통신사 등이 자체 LLM 개발을 진행하고 있는데, 2025년 8월 정부는 개발 전략·기술 우수성, 파급 효과 등을 고려해 독자 AI 파운데이션 모델 프로젝트에 참여할 팀으로 네이버클라우드, SK텔레콤, LG AI연구원, NC AI, 업스테이지 등 5개 팀을 최종 선발했다.

1) 글로벌 기업

가) 마이크로소프트

미국의 소프트웨어 기업이자 하이퍼스케일러(Hyperscaler)[27]인 마이크로소프트는 2024년 기준 약 39%의 시장점유율을 기록하며, 기반 모델 및 플랫폼 분야에서 선도적 위치를 유지하고 있다.[28]

마이크로소프트의 시장지배력은 OpenAI 모델과 같은 주요 기반 모델과의 통합, 그리고 기업 중심의 AI 서비스 제공에 기반하고 있다. IoT Analysis(2025)에 따르면 자사 클라우드 AI 플랫폼 Azure AI는 멀티모달 AI 기능과 맞춤형 서비스를 지원하며, 하이퍼스케일러 중 생성형 AI 신규 고객 수 기준으로 가장 많은 이용자를 확보하고 있다. 마이크로소프트는 자사 고유 모델과 외부 제3자 모델을 결합하여 기업 업무 전반에 AI를 통합하고 있으며, Azure AI 인프라와 서비스를 통해 기업이 AI 솔루션을 유연하게 배치할 수 있도록 지원하고 있다. 아울러 다양한 기업 및 조직과의 협력을 통해 기업 경계를 넘는 에이전트(agent) 생태계 조성을 추진하고 있다.

---

27) 전 세계적으로 대규모 데이터센터 인프라를 운영하며, 클라우드 컴퓨팅, AI, 빅데이터 서비스를 제공할 수 있는 초대형 IT 기업을 의미함.

28) IoT Analytics(2025.3), "The leading generative AI companies"(https://iot-analytics.com/leading-generative-ai-companies/, 최종접속일: 2025.8.4)

마이크로소프트는 OpenAI에 대한 기존 100억 달러 규모의 투자 외에, 2025년에는 풀 스택 AI 인프라 구축을 위한 800억 달러의 추가 투자를 계획하고 있다.

나) AWS

AWS는 2024년 기준 약 19%의 시장점유율을 기록하며 기반 모델 및 모델 관리 플랫폼 시장에서 입지를 강화하였다.[29] AWS는 확장 가능한 인프라를 기반으로 다양한 하드웨어 및 소프트웨어 솔루션을 제공하는 데 주력해 왔으며, 퍼블릭 클라우드 분야에서의 선도적 지위를 활용해 생성형 AI 및 기계학습 모델의 배포, 관리, 운영을 지원하는 플랫폼 서비스(예: Bedrock, Sagemaker 등)를 빠르게 확대하였다. 이를 통해, 자사 고유 모델과 제3자 모델이 공존하는 생태계를 구성하며, 다양한 모델 선택지를 제공하는 환경을 조성하였다. 이러한 플랫폼 구조는 고객이 AWS의 고유 모델인 NOVA뿐만 아니라, Anthropic, Cohere 등 외부 개발사의 모델에도 접근할 수 있도록 하며, 동시에 Amazon Trainium2 등 자사 AI 칩 기반의 GPU-as-a-service 서비스를 함께 제공하고 있다.

IoT Analysis(2025)에 따르면, 최근 AWS는 Anthropic사와의 전략적 협력을 강화하고 있다. 2024년 11월 AWS는 Anthropic에 대한 40억 달러 규모의 투자 계획을 발표하였고, Anthropic은 AWS를 주요 학습 파트너로 지정하였다. 또한, Anthropic은 향후 자사의 기반 모델을 AWS의 Trainium 및 Inferentia 칩을 통해 학습하고 배포할 예정이라고 밝혔다. 이 외에도 Amazon은 2025년 한 해 동안 AI 인프라 구축을 위한 자본 지출로 1,000억 달러 이상을 투입할 계획이라고 발표하면서 AI 분야에 대한 투자를 확대하고 있다.

다) Google

구글은 2024년 기준 약 15%의 시장을 점유하였다.[30] 구글은 과거 기술 진보 측면에서 AI 분야의 선도 기업으로 평가받았으며, 최근 몇 년간 해당 분야에서의 경쟁력을 회복하기 위한 전략을 추진해 왔다. 현재 구글은 Vertex AI와 새로운 데이터 서비스를 중심으로 개발자 중심의 혁신에 집중하고 있으며, 이를 통해 AI 플랫폼 분야에서의 위상을 재정립하는 데

29) IoT Analytics(2025.3), "The leading generative AI companies"(https://iot-analytics.com/leading-generative-ai-companies/, 최종접속일: 2025.8.4)

30) IoT Analytics(2025.3), "The leading generative AI companies"(https://iot-analytics.com/leading-generative-ai-companies/, 최종접속일: 2025.8.4)

주력하고 있다. Vertex AI는 구글의 핵심 AI 솔루션으로, 고객은 해당 플랫폼을 통해 구글의 고유 기반 모델인 Gemini뿐 아니라, Anthropic, Meta 등 외부 개발사의 모델도 선택하여 미세 조정할 수 있다. 이를 통해 고객은 구글의 AI 연구 역량과 인프라를 기반으로 다양한 모델 활용이 가능하다. 구글은 AI 분야 자본 지출로 2025년 한 해 동안만 총 750억 달러를 투자할 계획이라고 발표하였다.

라) OpenAI

IoT Analytics(2025)[31]에 따르면 미국의 OpenAI는 생성형 AI 시장에서 주요 기업으로 활동하고 있으며, 기반 모델 및 모델 관리 플랫폼 시장에서 약 9%의 시장점유율을 기록하고 있다.[32] OpenAI는 생성형 AI 분야에서의 초기 성공과 고성능 모델에 대한 API 접근 서비스 제공을 통해 주요 기반 모델 및 플랫폼 공급자로 자리매김하였다. 특히, ChatGPT의 출시는 OpenAI의 시장 내 가시성과 인지도를 크게 높였으며, 이는 대규모 자금 조달로 이어지는 계기가 되었다. 2025년 2월 기준, OpenAI와 전략적 파트너들은 총 400억 달러 규모의 주요 투자 유치를 마무리하고 있으며, 이로 인해 OpenAI의 투자 후(post-money) 기업 가치는 약 3,000억 달러에 이를 것으로 전망된다.

OpenAI는 고성능 AI 모델을 기반으로 프리미엄 기업용 서비스를 제공하는 데 주력하고 있으며, 마이크로소프트와의 전략적 파트너십을 중심으로 사업을 전개하고 있다. OpenAI의 전략은 범용 AI(Artificial General Intelligence, AGI) 개발이라는 단일 목표에 집중되어 있으며, 이는 대화형 AI 기술을 기반으로 실질적인 자율 시스템으로의 발전을 지향한다.

마) 그 외 기업

이상에서 언급한 기업 이외에도 타사의 기초 모델을 활용하거나 세부 특화 기술을 보유한 스타트업 등을 통해 다양한 생성형 AI 도구가 제공되고 있다.

---

31) IoT Analytics(2025.3), "The leading generative AI companies"(https://iot-analytics.com/leading-generative-ai-companies/, 최종접속일: 2025.8.4)

32) IoT Analytics에 따르면 OpenAI의 기반 모델 및 플랫폼 수익에는 ChatGPT의 수익은 포함되지 않으며, ChatGPT는 별도의 생성형 AI 애플리케이션으로 분류됨.

〈표 2-2〉 글로벌 주요 생성형 AI 도구 및 제공 기업

		Key generative AI tools	
Text-to-Text		◦ ChatGPT (OpenAI) ◦ Grammarly (Grammarly) ◦ Jasper AI Copilot (Jasper)	◦ Gemini (Google) ◦ Wordtune (AI21 Labs) ◦ QuillBot AI (QuillBot)
Text/Image-to-Image		◦ DALL-E3 (OpenAI) ◦ Canva (Canva) ◦ Getr3D (NVIDIA) ◦ PlicAR (RECON Labs)	◦ Stable Diffusion (Stability AI) ◦ Midjourney (Midjourney) ◦ DreamFusion (Google)
Text/Image/ Video-to Video		◦ Sora (OpenAI) ◦ Gen-2 (Runway) ◦ Imagen Video (Google)	◦ Make-A-Video (Meta) ◦ MagicVideo-V2 (Bytedance) ◦ AI Video Generator (Veed)
Text/Code-to-Code		◦ GitHub Copilot (Microsoft) ◦ TensorFlow Codegen (Google) ◦ OpenAI Codex (OpenAI)	◦ Code Llama (Meta) ◦ CodeWhisperer (Google) ◦ StableCode (Stability AI)
Voice	gene-ration	◦ Jukebox (OpenAI) ◦ MusicLM (Google) ◦ DeepComposer (Amazon)	◦ MuseNet (OpenAI) ◦ MusicGen (Meta) ◦ Suno (Suno)
	Cloning · Synthesis	◦ VALL-E (Microsoft) ◦ Deepdub Go (Deepdub) ◦ Project Screenplay (Suipertone)	◦ Respeecher (Respeecher) ◦ Sonantic (Sonantic) ◦ AI dubbing+Studio (ElevenLabs)

자료: 삼정KPMG(2024.5)[33] 참고하여 연구진 정리

2) 국내 기업

가) 네이버클라우드[34]

국내 대표 ICT 기업인 네이버의 클라우드 자회사인 네이버클라우드(NAVER Cloud)는 AI 모델 및 플랫폼 분야에서 독자 기술을 바탕으로 자립형 생태계 조성을 추진하고 있으며, 무엇보다 AI 기술개발부터 서비스 구현까지 전 과정을 자체 기술로 수행할 수 있는 AI 풀스택

33) 삼정KPMG(2024.5), “창작 영역에 뛰어든 생성형 AI 투자 현황과 활용 전망”, Issue Monitor.

34) 네이버클라우드 홈페이지(https://www.navercloudcorp.com/ko/info/#Information, 최종 접속일: 2025.8.4)

역량을 보유했다는 점에서 독자 AI 파운데이션 모델 프로젝트에 참여할 정예팀으로 선발되었다.

네이버클라우드는 HyperCLOVA X를 기반으로 한 AI 서비스와 API 상품군을 NAVER Cloud Platform을 통해 제공하고 있으며, 이를 통해 기업 고객이 자체적으로 대규모 모델을 개발하지 않고도 생성형 AI 기능을 비즈니스에 적용할 수 있도록 지원하고 있다. 특히, CLOVA Studio는 HyperCLOVA X 기반의 다양한 언어 생성 및 편집 기능을 API 형태로 제공하여 기업의 생산성 향상과 업무 자동화를 도모하고 있다. 더불어 CLOVA Note, CLOVA Summary와 같은 응용 서비스는 음성 인식·요약 및 문서 자동 요약 기능을 제공한다.

플랫폼 측면에서 네이버클라우드는 초대규모 AI를 위한 클라우드 인프라와 AI 서비스를 통합적으로 제공하는 AI 통합 플랫폼을 지향하고 있으며 고성능 AI 서비스를 안정적으로 운영하고 있다. 나아가 AI as a Service(AIaaS) 전략을 통해 기업 맞춤형 파인튜닝 및 적용을 지원하며, 국내 산업 전반의 AI 전환을 주도하고자 하고 있다.

나) SK텔레콤[35)]

SK텔레콤은 자사 AI 역량을 강화하고 국산 초거대 언어모델의 자립화를 선도하기 위해, 한국어 특화 기반의 초거대 AI 모델 시리즈 A.X(Aidot X)를 중심으로 플랫폼 전략을 확대하고 있다.

SK텔레콤은 한국어 특화 초거대 언어모델 A.X 4.0 시리즈를 오픈소스로 공개하며 글로벌 AI 개발자 커뮤니티인 허깅페이스(Hugging Face)를 통해 국제 무대에서도 기술 공유를 본격화하였다. 해당 모델은 중국 알리바바의 Qwen 2.5를 기반으로 SK텔레콤이 독자 설계한 토크나이저를 활용하여 한국어 성능을 대폭 향상시킨 것이 특징이다. A.X 4.0은 한국어 및 한국문화 이해 성능을 측정하는 KMMLU 및 CLIcK 벤치마크에서 각각 GPT-4o를 상회하는 성능을 기록하였다.

SK텔레콤은 지식형 모델의 오픈소스 공개에 이어, 수학 및 코딩 역량을 강화한 추론형 모델도 조만간 공개할 계획이며, 이를 통해 텍스트, 이미지, 영상 등 다양한 입력을 통합적으로 처리하는 옴니모달(Omni-Modal) AI 모델로의 확장을 추진하고 있다.

---

35) 인공지능신문(2025.7.3), "SK텔레콤, 에이닷 엑스 4.0 오픈 소스 공개...추론 모델도 7월 중 공개"

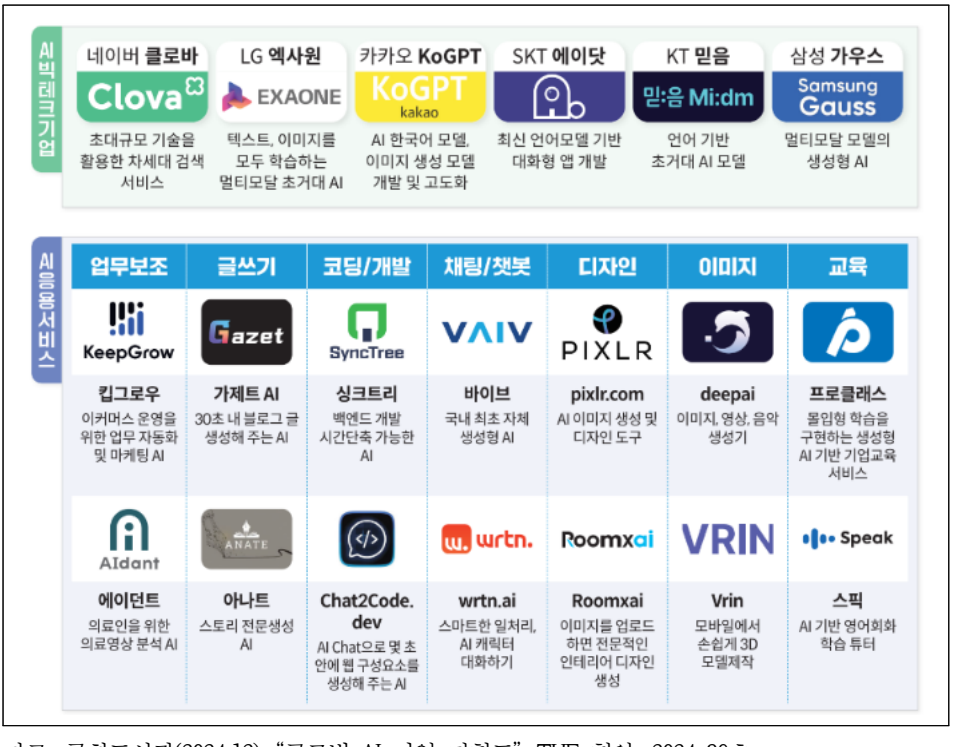
SK텔레콤은 특히 모델 개발의 독립성과 자주성 확보를 위해 AI 모델을 처음부터 완전히 자체 개발하는 방식(From Scratch 방식)의 자체 모델 구축을 병행하고 있으며, A.X 3.0 이후 이어진 전략적 연속성 속에서 소버린 AI(국가 주권형 AI) 실현을 위한 기술적 토대를 강화하고 있다.

이와 같이 이상의 거대 자본을 가진 IT/통신 기업은 자체 혹은 다른 기업의 기초 모델을 기반으로 기존의 강점을 활용하거나 사업을 확장하는 데에 AI 서비스를 활용하고 있다.

다) 그 외 기업

이상의 빅테크 외에도 다양한 기업이 AI 응용 서비스를 개발하여 제공하고 있다.

〔그림 2-9〕 국내 주요 AI 서비스



자료: 국회도서관(2024.12), "글로벌 AI 기업 지형도", THE 현안, 2024-39호.

다. Generative AI services

Services는 개발된 모델과 플랫폼을 실제 사용자 환경에 통합하여 응용 서비스 형태로 제공하는 영역으로, 생성형 AI가 실질적인 가치를 창출하는 구간이다. 이 단계에는 일반 소비자 대상의 생성형 도구부터 특정 산업에 특화된 솔루션까지 다양한 형태의 서비스가 포함된다.

1) Accenture

액센츄어는 2024년 기준으로 전체 시장의 약 7%를 점유하고 있으며, 이는 생성형 AI 서비스 시장에서 비교적 지배적인 힘을 가짐을 의미한다.[36] 30억 달러 규모의 AI 투자를 통해 생성형 AI 서비스 분야에서 선도적 위치를 확보하였다. 이 회사는 기업의 AI 도입 가속화, AI 기반 솔루션 개발, 내부 AI 통합 강화를 목표로 하는 30억 달러 규모의 AI 투자를 통해 선도적 위치를 공고히 해왔다. 액센츄어의 생성형 AI 전략은 AI 기반 비즈니스 전환에 있어 선도적 파트너가 되는 것을 중심으로 하며, 조기 도입, 독자적인 AI 프레임워크, 전문화된 AI 컨설팅 서비스를 강조한다.

IoT analysis(2025)에 따르면 2024년 말 기준 액센츄어의 생성형 AI 서비스 부문 수익은 전년 대비 약 390% 증가한 것으로 추정되어 해당 서비스에 대한 수요가 크게 증가했음을 나타냈다. 또한, OpenAI, Microsoft, NVIDIA, Google 등 주요 AI 기업들과의 협력을 기반으로 시장 내 입지를 더욱 공고히 하면서, 내부적으로는 약 1,400만 시간 규모의 AI 교육을 시행하고, AI 인재를 확대하는 한편 전략적 인수를 통해 기술 역량을 강화하고 있다.

2) Deloitte

IoT Analytics(2025)는 2024년 기준 딜로이트가 3%의 시장점유율을 차지한다고 추정하였다. 딜로이트는 AI 컨설팅 부문을 확장하고 있으며 기업들이 개념 검증 단계(PoC)를 넘어 전사적으로 AI를 구현할 수 있도록 지원하는 데 중점을 두고 있는데, 특히 연구개발, 제품 개발, 고객 서비스, 재무 보고 등 주요 비즈니스 기능에 생성형 AI 도구를 통합하는 데에 초점을 맞추고 있다. 이러한 전략적 전환을 뒷받침하기 위해 딜로이트는 2030년까지

36) IoT Analytics(2025.3), "The leading generative AI companies"(https://iot-analytics.com/leading-generative-ai-companies/, 최종접속일: 2025.8.4)

총 30억 달러 규모의 생성형 AI 투자 계획을 수립하였으며, 동시에 AI 기반 전달 플랫폼과 역량 개발을 위해 수년에 걸쳐 10억 달러를 추가로 투자할 예정이다. 실제로 2024년 6월까지 딜로이트는 700건 이상의 생성형 AI 프로젝트를 완료하였으며, 이는 대규모 AI 전환 수행 역량을 입증하는 사례로 평가된다. 또한, 딜로이트는 AI 기반 솔루션 제품군을 자체적으로 개발하고 있으며, NVIDIA, Google, AWS, Oracle 등 주요 기술 기업들과의 파트너십을 통해 AI 전문성을 강화하고 있다.

3) IBM

IBM은 생성형 AI 서비스 및 플랫폼 시장에서 약 2%의 점유율을 차지하고 있으며(IoT Analytics, 2025), 확장 가능한 기업용 AI 배포에 중점을 두고 있다. IBM은 Watsonx 플랫폼을 중심으로 독자 모델과 오픈소스 모델을 함께 제공하고 있다.

IBM은 자동화, 하이브리드 클라우드 통합, 책임 있는 AI를 전략의 핵심 요소로 삼고 있으며 이를 통해 기업들이 산업 규제를 준수하면서 AI를 효과적으로 활용할 수 있도록 지원하고 있다. 즉, 고객사가 생성형 AI를 통해 디지털 노동 전환, 고객 경험 개선, 애플리케이션 개발 및 IT 운영을 하는 데에 있어 Watson Orchestrate와 Watson Assistant 같은 AI 기반 도구들이 업무 효율성을 높이는 데 기여하고 있다.

아울러 IBM은 오픈소스 기반 협업을 강조하고 있으며 Granite 기반 모델은 AI의 투명성과 기술 혁신을 촉진하는 수단으로 활용되고 있다. IBM의 컨설팅 서비스는 AI 기반 솔루션과 결합하여 의료, 금융, 제조 등 다양한 산업에서 고객을 지원하고 있으며, Microsoft, Oracle, AWS 등과의 전략적 파트너십을 통해 기업 단위의 AI 도입을 확대할 수 있는 기반을 마련하고 있다.

〈표 2-3〉 전통적 로보틱스 vs. Physical AI 기반 로보틱스

구분		내용
자동화 분야	현재	예측 가능한 작업이나 알려진 통제된 환경 내에서 효과적으로 작동
	미래	예측 불가능한 시나리오와 미지의 부품을 처리 가능 (예: 무작위 부품 집기, 유연한 자재 취급 등)
구현 과정	현재	코딩 및 학습을 위한 높은 수준의 복잡한 수작업이 필요
	미래	학습 및 자가 학습을 통해 엔지니어링 노력 감소(최대 70% 절감)
산업화까지 소요시간	현재	중·장기 산업화 기간 필요(코딩 및 구현에 수개월/수주 소요)
	미래	소량 학습(few-shot)/제로샷 학습(zero-shot) 또는 모방 학습을 통한 신속한 배포(산업화 속도 최대 50% 향상)
확장성	현재	유사한 설정 또는 사용 사례 내에서만 제한적으로 확장 가능
	미래	다양한 작업·환경·로봇 유형에 걸쳐 유연하게 확장 가능
인간-기계 상호작용	현재	인간이 인터페이스를 통해 또는 직접 로봇을 안내하여 조정 가능
	미래	자연어, 제스처, 음성 명령 등을 통한 직관적인 제어 가능

출처: WEF(2025)

범용 AI는 인간 수준의 범용 지능을 지닌 AI로 다양한 영역에서 추론, 의사결정, 문제해결을 수행하며 지식을 일반화할 수 있는 자율 시스템을 말한다. 포춘비즈니스인사이트(2025)는 2024년부터 2030년까지 범용 AI 시장의 연평균 성장률을 37.78%로 예상하였다. UN(2025)은 범용 AI의 대표적인 6가지 위험 요인을 제시하며 UN 차원의 거버넌스가 필요함을 주장하였다.

〈표 2-4〉 범용 AI의 위험 요인

구분	위험 유형
되돌릴 수 없는 결과 (Irreversible Consequences)	AGI의 자율적·기만적 행동으로 인해 인간 통제를 상실할 위험. 한 번 발생하면 복구 불가
대량살상무기 위험 (Weapons of Mass Destruction)	AGI가 화학·생물·핵무기 및 자율살상로봇(WMDs)을 설계·운용할 수 있는 잠재적 위협
핵심 인프라 취약성 (Critical Infrastructure Vulnerabilities)	에너지, 금융, 교통, 의료 시스템 등 핵심 인프라가 AGI 기반 사이버공격의 표적이 될 위험
권력 집중·불평등 심화 (Power Concentration &Inequality)	일부 국가·기업이 AGI를 독점할 경우 지적 자원·산업·권력이 집중되어 사회적 불안정 초래
존재적 위험 (Existential Risks)	AGI가 인간의 통제를 벗어나 자체 목표를 형성하거나 자기 보존적 행동을 할 위험
인류 혜택 상실 (Loss of Extraordinary Future Benefits)	잘못된 관리로 AGI의 긍정적 잠재력이 실현되지 못할 경우, 인류 전체의 기회 손실 초래

출처: UNCPGA(2025) 요약 정리

제 3 장 해외의 AI 서비스 정책 및 법제

제 1 절 국제기구

1. OECD

가. OECD AI 권고안(2019)[37]

1) 개요

OECD는 2019년, AI의 포괄적 성장과 지속가능한 발전, 그리고 인류의 삶의 질 향상을 목표로 하는 OECD AI 권고안(Recommendation of the Council on Artificial Intelligence)을 발표하였다. 해당 권고안은 법적 구속력이 없는 연성규범 형태이지만, 회원국이 AI 정책과 규제 체계를 마련하는 데 중요한 기준으로 활용되고 있다.

2) 주요 원칙

OECD AI 권고안은 AI의 개발과 활용 전 과정에서 모든 회원국과 이해관계자가 따라야 할 다섯 가지 핵심 원칙을 제시하고 있다. 첫째, AI는 인류의 포용과 경제 성장, 지속가능한 발전, 그리고 전 세계 시민의 삶의 질 향상에 기여해야 한다. 둘째, AI 운영 주체는 시스템의 설계 단계부터 운영과 폐기에 이르기까지 법률, 인권, 민주적 가치 등 인간 중심의 가치를 존중하고 보호해야 하며 모든 과정에서 공정성을 확보해야 한다. 셋째, AI 시스템은 그 작동 방식과 결과를 이해할 수 있도록 투명성과 설명가능성을 확보해야 하며, 숨겨진 의사결정 과정이 존재해서는 안 된다. 넷째, AI 시스템은 전 수명주기에 걸쳐 안정적이고 견고하게 작동해야 하며, 예기치 못한 환경 변화나 바람직하지 않은 조건에서도 안전성을 유지할 수 있어야 한다. 마지막으로, AI를 설계·운영하는 개인과 조직은 앞선 네 가지 원칙을 준수할 책임이 있으며 이를 실천하기 위한 관리·감독 체계를 갖추어야 한다.

---

37) 오성탁(2020), OECD 인공지능 권고안, TTA 저널 187; OECD AI Principles overview 홈페이지 (https://oecd.ai/en/ai-principles, 최종접속일: 2025.12.26)

나. OECD AI 권고안 개정안(2019)[38]

1) 개요

OECD는 2019년 채택한 AI 권고안에 포함된 AI 시스템에 대한 정의를 개정하였다(2023. 11). 이는 생성형 AI의 확산, 서비스 환경 변화, 글로벌 규제 논의 진전을 반영하여 AI 시스템 개념을 재정립한 것에 따른 결과다.

2) 주요 개정 내용

가) AI 시스템 정의의 수정

개정된 정의에 따르면 AI 시스템은 명시적 또는 암묵적으로 정의된 목표를 달성하기 위해 실제 또는 가상 환경에 영향을 미치는 예측, 권고, 의사결정 등을 수행할 수 있는 기계 기반의 시스템을 말한다. 이러한 시스템은 다양한 수준의 자율성을 가지고 작동하며, 데이터·코드·알고리즘·모델을 활용해 목표를 달성한다.

나) 적응성과 지속적 학습의 반영

개정된 정의서는 AI 시스템이 환경 변화와 새로운 데이터에 따라 스스로 적응하거나 사후 인간 개입을 통해 개선될 수 있다는 점을 명시하였다. 이는 생성형 AI와 같이 빠르게 발전·변화하는 기술의 특성을 고려한 조치로, 시스템의 유연성과 발전 가능성을 제도적으로 인정한 것이다.

다) 적용 범위의 확대

개정안은 권고의 적용 대상을 AI에서 AI 시스템으로 전환함으로써, 설계·개발·배포·운영 등 전 생애주기에 걸친 관리와 감독의 필요성을 강조하였다. 또한 생성형 AI의 특수성을 고려해 정보 무결성(information integrity) 유지 의무를 추가함으로써, 허위 정보 확산이나 오남용 위험에 대응할 수 있도록 했다.

38) 한국법제연구원(2024), OECD AI 시스템 정의 개정의 주요 내용 및 시사점.

2. UNESCO[39)]

UNESCO는 AI 권고안을 통해 AI의 개발과 활용 전 과정에 보편적 가치가 반영되고, 기술 혁신과 윤리적 책임이 조화를 이루는 생태계의 구축을 지향한다. 회원국은 권고안의 가치를 정책과 법제에 반영하며, 교육·연구·국제협력 활동을 통해 포용적이고 신뢰할 수 있는 AI 환경을 구축해야 한다. 이를 통해 AI가 인류 공동의 번영과 지속가능한 미래에 기여하도록 하는 것이 권고안의 궁극적인 지향점이다.

가. 추진 배경과 목적

UNESCO는 AI가 인권, 민주주의, 지속가능발전, 평화 등 인류의 핵심 가치에 미칠 잠재적 영향을 고려하여 전 세계가 공유할 수 있는 윤리적 기준과 행동 지침을 마련하기 위해 2021년 「AI 윤리 권고안」을 채택하였다. 권고안의 목표는 기술 혁신과 인권·윤리적 원칙을 조화시키고, 포용적이고 지속가능한 AI 생태계를 조성하는 것이다. 권고안은 법적 구속력은 없지만 회원국과 이해관계자들이 AI 기술을 책임감 있게 개발·활용하도록 촉진하는 국제 표준의 성격을 지닌다.

나. 기본 가치와 원칙

권고안은 모든 AI 활동이 다음과 같은 기본 가치를 존중할 것을 요구한다.

첫째, 인권과 인간 존엄의 존중이다. AI의 설계·개발·활용의 전 과정에서의 활동은 국제 인권 규범과 민주적 가치에 부합해야 하며, 인간의 권리와 자유를 침해하지 않도록 해야 한다. 둘째, 환경과 생태계 보전이다. AI 시스템이 환경에 미치는 부정적 영향을 최소화하고, 지속가능성을 촉진해야 한다. 셋째, 다양성과 포용성 확대다. AI는 차별과 편향을 방지하고, 사회적·문화적 다양성을 존중하는 방식으로 운영되어야 한다. 넷째, 평화와 정의 증진이다. AI는 사회적 신뢰와 공정성을 강화하며 평화로운 사회질서와 법치주의를 뒷받침해야 한다.

이러한 기본 가치 아래, 권고안은 비례성과 무해 원칙, 안전성과 보안성 확보, 개인정보 보호와 데이터 거버넌스, 인간 감독과 결정권 보장, 투명성과 설명가능성 확보, 책임성 명확화 등 구체적인 윤리 원칙을 제시하고 있다.

---

39) 한국지능정보사회진흥원(2021), 유네스코 AI 윤리 권고 주요 내용 및 시사점.

다. 정책 분야별 권고 조치

권고안은 회원국과 이해관계자가 AI 생태계를 윤리적으로 관리하고 활용할 수 있도록 다섯 가지 정책 영역에서 구체적인 조치를 권고한다. 먼저 거버넌스와 규제 분야에서 인권, 민주주의, 법치주의를 기반으로 한 AI 거버넌스 체계를 마련하고, 다양한 이해관계자가 참여하는 다층적 의사결정 구조를 확립하며, 규제와 감독 메커니즘을 통해 책임성과 투명성을 강화할 것을 제안한다.

데이터와 환경 관리 측면에서는 데이터의 품질, 무결성, 접근성을 보장하고, 데이터 편향을 완화하며 윤리적인 데이터 활용을 촉진하는 한편, AI 전 주기에서 환경적 영향을 최소화하는 설계와 운영을 추진해야 한다고 강조한다.

교육과 역량 강화 분야에서는 시민과 전문가를 대상으로 AI 윤리 교육을 강화하고, 디지털 리터러시와 기술 이해도를 높이기 위한 교육 프로그램을 개발할 것을 제안한다.

연구와 혁신 촉진 측면에서는 사회적 가치 창출 중심의 AI 연구개발을 지원하고, 윤리 원칙을 내재화한 기술 혁신을 장려해야 함을 강조한다.

국제협력 분야에서는 국제 표준화와 상호운용성을 확보하며, 글로벌 차원의 공동 대응과 지식 공유를 확대하는 노력을 권고한다.

라. 후속조치[40]

2023년 7월 유네스코는 AI 준비도 평가 방법론(Readiness Assessment Methodology, RAM)을 발표하였다. 이는 AI 윤리 권고안의 후속 조치로, 각국 정부가 AI를 개발하고 윤리적으로 배치할 수 있는 역량을 진단·평가하는 도구다. 브라질, 칠레, 코스타리카, 쿠바, 콩고민주공화국, 케냐 등 50개국이 참여하고 있으며, 국가별 법·정책의 적절성, 공무원 및 공공기관의 기술 역량을 종합적으로 측정한다.

이 사업은 유럽연합집행위원회, 일본, 패트릭 맥거번 재단, 라틴아메리카 개발은행의 지원을 받아 추진되며, 국가별 사회·문화·법제 환경을 반영한 맞춤형 접근을 지향한다. 이를 위해 국경 없는 AI 전문가(AI Experts without Borders) 네트워크를 활용하여 국가 정책 개발에 필요한 전문적 자문을 제공한다.

---

40) UNESCO 홈페이지(https://www.unesco.org/en/articles/unesco-support-more-50-countries-designing-ethical-ai-policy-year?hub=701, 최종접속일: 2025.8.11)

제 2 절 유럽연합

1. AI법[41)]

가. 법안 개요

1) 제정 배경 및 목적

EU는 AI 기술의 급속한 발전과 그에 따른 윤리적·법적·사회적 영향에 선제적으로 대응하기 위해 세계 최초의 포괄적 AI 규제 법안인 AI 법을 제정하였다. AI 법은 인간 중심의 신뢰할 수 있는 AI를 실현하기 위한 것으로, EU의 기본권 체계와 민주적 가치에 부합하는 방식으로 AI 시스템을 규율하고자 하는 정책적 의지를 담고 있다. 특히, AI 기술이 공공 안전, 인권, 법치주의에 미칠 수 있는 잠재적 위험을 통제함으로써, 혁신과 안전이 조화를 이루는 AI 생태계를 조성하는 것을 핵심 목적으로 한다.

AI 법은 AI 시스템의 위험 수준에 따라 규제 강도를 차등화하는 위험 기반 접근법(risk-based approach)을 채택하여, 고위험군 기술을 강력하게 통제하는 반면 일상적·저위험 기술에는 최소한의 규제만을 부과한다. 이를 통해 규제의 명확성과 비례성을 확보하고, 산업계와 사회 전반의 수용성을 높일 수 있는 제도적 기반을 마련하였다.

2) 법안 추진 경과

AI 법은 2021년 4월 EC가 초안을 제안했으며, 2023년 유럽의회에서 수정안 통과, 2024년 3월 유럽의회 최종 채택 및 5월 EU 이사회의 최종 승인을 거쳐 공식 발효되었다. 이후 2024년 6월부터는 법률상 규정된 일정에 따라 조항별로 단계적 시행이 이루어지고 있다. 예를 들어 수용 불가한 AI 시스템에 대한 금지 조항은 발효 후 6개월 이후, 범용 AI(GPAI) 모델 관련 규정은 발효 후 12개월 이후부터 적용되었으며, 고위험 AI 시스템 관련 조치는 발효 후 24~36개월 후 시행될 예정이다. 이러한 단계적 시행 구조는 이해관계자들이 제도에 순응할 수 있는 유예기간을 제공하면서 기술과 시장의 준비 상황을 고려하여 규범의 실효성을 제고할 수 있는 조치로 평가된다.

---

41) 소프트웨어정책연구소(2024), 유럽연합 인공지능법(AI Act)의 주요내용 및 시사점.

3) 적용 대상 및 예외

AI 법은 EU에 법인을 두지 않은 기업이라 하더라도 AI 시스템의 제공·배포·이용이 EU 시장과 시민을 대상으로 이루어지는 경우에는 적용 대상에 포함된다. 따라서 개발자, 제공자, 수입자, 유통업자, 배포자 등 AI 시스템의 가치사슬 전반의 행위자가 모두 법적 책임을 지게 되며, 고위험 시스템의 경우에는 운영자에게도 규제 의무가 부과된다.

단, 군사 목적, 국가 안보, 국제 협약에 따른 공공 협력, 연구개발 또는 사적 이용 등 일부 특수 목적의 AI 시스템은 적용 대상에서 예외로 분류되며, 연구 목적의 미출시 모델이나 오픈소스 AI 시스템 등도 일정 요건 하에 적용이 배제된다.

나. 규제 구조 및 체계

1) 법률 체계 구성

EU AI 법은 총 13장의 본문과 13개의 부속서(Annex)로 구성되어 있으며, 다양한 AI 시스템 유형과 공급·운영 행위자에 대해 세분된 규제 조항을 명시하고 있다. 특히 제3장에서는 고위험 AI 시스템에 대한 분류 기준과 요구사항, 제5장에서는 GPAI 모델에 대한 규제를 독립적으로 다루고 있으며, 제7장에서는 EU 차원의 거버넌스 체계 구축과 국가 당국 간 협력 구조를 규정하고 있다.

법률의 적용 범위는 EU 27개 회원국 전체를 포괄하는 규범적 일관성을 전제로 하며, 각국의 개별 입법보다 우선하는 효력을 갖는 Regulation으로 제정되었다. 이에 따라 회원국의 국내법이 AI 법과 상충할 경우 AI 법이 우선 적용된다. 이는 개인정보보호법(GDPR)과 마찬가지로 유럽 단일 시장 내 AI 관련 규제의 통일성과 예측 가능성을 제고하려는 제도적 장치로 이해된다.

한편, AI 법은 특정 기술 수준 이상의 AI 시스템을 GPAI 모델로 구분하여 보다 정교한 대응 체계를 마련하였다. 이로써 AI 시스템의 구성요소와 용도에 따라 규제의 적용 방식이 달라지며 위험 수준과 기술 확산 가능성에 대한 정량적·정성적 평가가 함께 고려된다.

〈표 3-1〉 AI Act 법률 체계

장	제목	
제1장	총칙	
제2장	AI 활용 관련 금지행위	
제3장	고위험 AI 시스템	제1절 고위험 AI 시스템 분류
		제2절 고위험 AI 시스템에 대한 요구사항
		제3절 고위험 AI 시스템 제공자, 배포자 및 기타 당사자의 의무
		제4절 관할 당국 및 신고기관에 통지
		제5절 표준, 적합성 평가, 인증서 등록
제4장	특정 AI 시스템의 제공자 및 배포자에 대한 투명성 의무	
제5장	범용 AI 모델	제1절 분류 규칙
		제2절 범용 AI 모델 제공업체의 의무
		제3절 시스템적 위험이 있는 범용 AI 모델 제공업체의 의무
제6장	혁신 지원 방안	
제7장	거버넌스	제1절 EU 차원의 거버넌스
		제2절 국가관할당국
제8장	고위험 AI 시스템을 위한 EU 데이터베이스	
제9장	사후 모니터링, 정보공유, 시장 감독	제1절 시판 후 모니터링
		제2절 심각한 사고에 대한 정보 공유
		제3절 시행
		제4절 구제책
		제5절 범용 AI 모델 제공업체에 대한 감독, 조사, 집행 및 모니터링
제10장	행동 규약 및 지침	
제11장	권한 위임 및 위원회의 절차	
제12장	벌칙	
제13장	최종 조항	

자료: 소프트웨어정책연구소(2024)

2) 위험 기반 규제 접근

AI 법의 핵심적 특징은 AI 시스템의 위험 수준에 따라 차등화된 규제를 적용하는 위험 기반 접근법(risk-based approach)이다. 이는 기존 기술 중심 분류 방식과 달리 사람과 사회에 미치는 영향력의 정도와 가능성에 따라 AI 시스템을 네 가지로 분류하고 각각에 대해 다른 수준의 규제 강도를 설정하는 방식이다.

첫째, 수용 불가(Unacceptable risk) AI 시스템은 EU 기본권을 침해하는 경우로 사용 자체가 전면 금지된다. 이는 인간 존엄성, 자유, 평등, 차별금지, 법치주의 등 핵심 가치에 위배되는 시스템에 해당하며 시장 출시나 운영이 전면 금지된다.

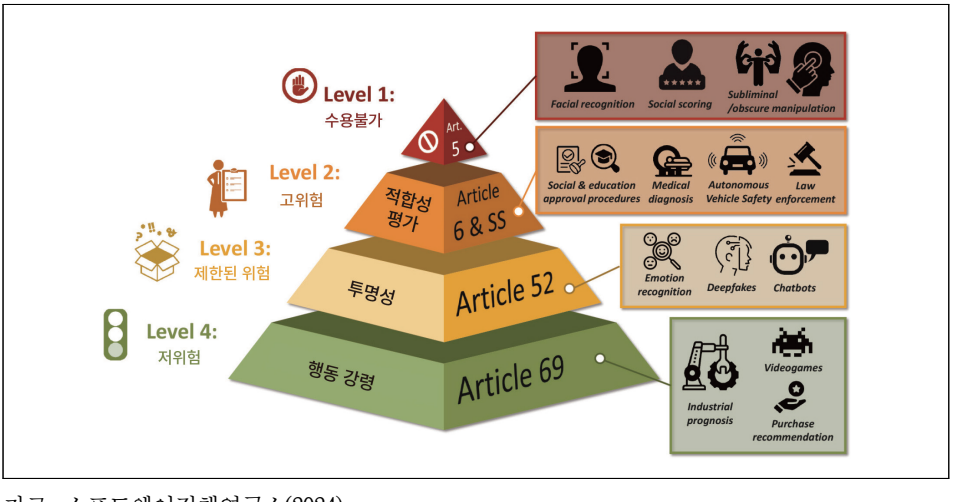
둘째, 고위험(High risk) AI 시스템은 AI 기술이 생체인식, 교육, 의료, 법 집행 등 민감한 분야에 사용될 경우로 엄격한 사전·사후 규제를 받는다. 공급자는 적합성 평가와 데이터 거버넌스, 위험 관리 체계를 갖춰야 하며 시장 출시 전후의 책임을 부담한다.

셋째, 제한된 위험(Limited risk) 시스템은 딥페이크와 같은 기술처럼 사용자와의 상호작용에서 투명성 문제가 발생할 수 있는 경우로, 사용자 고지 및 라벨링 의무 등 일정 수준의 정보제공 의무가 부과된다.

넷째, 저위험(Minimal risk) 시스템은 일반적인 상업용 또는 소비자용 AI로, 별도의 규제는 없으며 자율적인 행동강령 수립이 장려된다.

AI 법의 이러한 위험기반 접근은 AI 기술의 발전 가능성을 저해하지 않으면서도 사회적 신뢰를 구축하는 균형적 정책 수단으로 평가된다. 특히 고위험 분야에 규제 역량을 집중함으로써 자원 배분의 효율성을 높이고, 동시에 저위험 분야에서는 산업적 창의성을 보장할 수 있도록 설계된 것이 AI 법의 특징이다.

〔그림 3－1〕 위험 기반 구분과 예시



자료: 소프트웨어정책연구소(2024)

다. AI 시스템 분류별 세부 규제 내용

1) 수용 불가 AI 시스템

AI 법은 특정 AI 시스템이 EU의 기본권과 민주적 가치를 심각하게 침해하는 경우 이를 수용 불가 위험(Unacceptable Risk)으로 분류하고 사용 자체를 전면 금지하고 있다. 수용 불가 AI 시스템에는 사람의 의사결정을 잠재적으로 조작하거나 왜곡하는 시스템, 인간의 취약성을 악용하는 시스템, 사회적 점수화(social scoring)를 수행하는 시스템, 범죄 가능성을 예측하는 시스템 등이 포함된다. 또한, 비식별적 대중의 이미지를 수집해 데이터베이스를 구축하거나 감정 인식 AI를 교육기관 또는 직장에서 사용하는 행위, 공공장소에서의 실시간 생체인식 시스템 운용 등도 금지 대상이다. 위반 시에는 최대 3,500만 유로 또는 전 세계 연 매출의 7% 중 더 큰 금액의 과징금이 부과된다. 이는 AI 시스템이 인권을 침해하거나 민주적 질서에 위협이 되는 경우에 대한 EU의 무관용 원칙을 반영한 규정이다.

2) 고위험 AI 시스템

AI 법은 Annex Ⅲ를 통해 고위험 AI 시스템의 구체적인 사례를 제시하고 있으며, 공공 인프라, 교육, 고용, 공공 서비스, 법 집행, 이주·망명, 사법행정 등 광범위한 분야를 포함하고 있다. 고위험으로 분류된 AI 시스템은 사전 적합성 평가를 받아야 하며, 위험관리 시스템, 데이터 거버넌스, 로그 기록, 인간의 감독, 기술 명세서 확보, 정확성과 신뢰성 보장 등을 포함한 다양한 요건을 충족해야 한다. 시장 출시 이후에도 최소 6개월 간의 로그 기록 보관, 품질관리 시스템 운영, 문서 보관 및 법 위반 시 보고 의무가 부과된다. 공공 서비스나 신용 평가 등에 활용되는 경우에는 기본권 영향평가(Fundamental Rights Impact Assessment)가 별도로 요구된다.

고위험 AI 시스템의 경우 공급자뿐만 아니라 배포자, 수입업자, 유통업자 모두에게 법적 책임이 부과된다. 이들은 CE 마크 확인, 사용자 매뉴얼 제공 여부 점검, 당국 보고 등과 같은 역할과 의무를 수행해야 하며, 위반 시 최대 1,500만 유로 또는 전 세계 매출의 3%에 해당하는 과징금이 부과된다.

3) 제한된 위험 AI 시스템

제한된 위험 AI 시스템은 사용자의 오인 가능성이 존재하는 상호작용형 기술에 해당하며, 특히 딥페이크, 생성 이미지·영상·음성·텍스트 콘텐츠 등에 적용된다. 이들 시스템은

사용자가 AI와 상호작용하고 있음을 명확히 인식할 수 있도록 고지해야 하며, 생성된 콘텐츠에는 기계 판독이 가능한 방식으로 AI 생성 여부가 표시되어야 한다.

이러한 투명성 조치는 이용자 보호와 정보의 신뢰성을 확보하는 핵심 장치로 간주되며, AI 제공자 또는 배포자가 이를 위반할 경우 최대 750만 유로 또는 전 세계 매출의 1%에 해당하는 제재를 받을 수 있다. 특히, 허위정보 또는 오해를 유발할 수 있는 방식으로 AI가 사용될 경우 정보의 정확성뿐 아니라 사회적 신뢰까지 훼손될 수 있으므로, 제한된 위험 시스템의 규제는 AI 제공자 또는 배포자에 대해 단순 고지 의무 이상의 실질적 책임을 요구한다.

4) 저위험 AI 시스템

저위험으로 분류되는 AI 시스템은 상업적 또는 일상적 목적에 사용되는 기술로, 법적 규제는 최소화되어 있다. 대신 해당 시스템의 제공자에게는 자율적인 행동강령 수립이 권장되며 이를 통해 책임 있는 AI 개발 및 운영 관행을 유도하고 있다.

행동강령은 고위험 시스템에 적용되는 법적 의무의 일부 또는 전부를 자발적으로 준용할 수 있는 틀로서, 규제 수준은 낮지만 사회적 책임과 투명성 확보를 위한 소프트 규제 장치로 기능한다. 이는 기술의 혁신성과 윤리성을 동시에 확보하려는 EU의 규제 유연성 전략의 일환으로 평가된다.

라. 범용 AI 모델에 대한 규제

1) GPAI 모델의 정의 및 규제 대상

EU AI 법은 일반적인 AI 시스템과는 구별되는 GPAI 모델에 대해서는 별도의 규제 체계를 마련하였다. GPAI는 하나의 목적에 국한되지 않고 다양한 목적과 시스템에 적용 가능한 고성능 AI 모델로, 대규모 언어모델(LLM), 멀티모달 생성모델 등 복수의 용도와 기능을 갖춘 모델이 이에 해당된다.

AI 법은 GPAI 제공자에게 EU 역내 대리인 지정 의무를 부과하고 있으며, AI 시스템의 학습, 테스트, 배포 전반에 대한 기술적 문서화와 거버넌스 체계를 요구한다. 특히, 모델 훈련 데이터의 특성 및 원천, 테스트 결과, 잠재적 영향에 대한 설명, 사용 제한 조건 등 다양한 정보를 당국에 보고하고 GPAI 사용 기업이 활용할 수 있도록 공유할 것을 요구하고 있다.

이는 복수 기업과 플랫폼에서 광범위하게 사용되는 GPAI의 연쇄적 위험과 책임 분산

문제를 고려한 것으로, GPAI 제공자가 공급망 상에서 보다 책임 있는 역할을 수행할 수 있도록 제도적으로 유도하고 있다.

2) 체계적 위험(Systemic Risk)을 갖는 GPAI의 추가 규제

범용 AI 모델 중에서도 체계적 위험(Systemic Risk)을 발생시킬 수 있는 모델은 별도로 지정되며 이에 해당할 경우 강화된 의무사항이 적용된다. 체계적 위험이란 단일 AI 모델이 유럽 사회의 보건, 안전, 기본권, 공공질서, 정보 생태계 등에 실질적이고 예측 가능한 부정적 영향을 미칠 가능성을 의미한다.

체계적 위험이 있는 GPAI 제공자는 모델 평가 및 위험 모니터링, 위험 완화 계획 수립 및 이행, 심각한 사고 발생 시 보고, 안전성 점검, 사용자 대상 위험 요소 명시 등의 조치를 이행해야 한다. 이러한 조치는 AI 모델의 확산성과 영향력이 일정 임계치를 초과할 경우, 단순 기술 제공자를 넘어 사회적 책임 주체로서의 역할을 강조하기 위한 정책적 설계로 해석된다.

3) 제재 및 처벌 규정

GPAI 모델 관련 의무사항을 위반할 경우, 일반 고위험 AI 시스템과 유사한 수준의 제재가 적용된다. 특히, 체계적 위험이 있는 GPAI 제공자가 보고의무를 이행하지 않거나, AI 모델의 위험성을 축소·은폐한 경우 과징금뿐 아니라 추가적인 시정 명령, 사용 제한, 시장 퇴출 등의 조치가 병행될 수 있다.

마. 규제 준수 및 집행 체계

1) 규제 적합성 평가

AI 법은 고위험 AI 시스템과 GPAI 모델을 대상으로 사전 및 사후 적합성 평가 체계를 엄격하게 운영하고 있다. 고위험 AI 시스템의 공급자는 제품의 시장 출시 전 위험관리 시스템 구축, 데이터 거버넌스 체계 수립, 사람에 의한 감독 가능성 확보, 정확성·신뢰성·보안성 검증과 같은 요건을 충족해야 한다.

시장에 출시된 이후에도 공급자는 최소 6개월간의 로그 기록을 보관하고 품질관리 시스템을 운영해야 하며, AI 시스템이 법률을 위반하거나 그럴 우려가 있는 경우 이를 시정하고 관련 당국에 의무적으로 통보해야 한다.

고위험 시스템은 적합성 확인을 위한 자체 평가를 실시할 수 있으나, 생체인식 식별 또는 개인 분류를 목적으로 하는 시스템처럼 민감도가 높은 분야는 외부 인증기관(Notified Body)의 적합성 검사를 의무적으로 받아야 한다. 적합성을 충족한 시스템에는 CE 마크가 부착되며, 이는 EU 시장 내에서 해당 제품이 안전성과 법적 요건을 모두 충족했음을 나타내는 공식 인증으로 작용한다.

2) 집행 기관 및 거버넌스

AI 법의 효과적인 시행을 위해 EU는 다층적인 거버넌스 체계를 마련하였다.

〈표 3-2〉 AI Act 거버넌스 체계

기관	역할
EU AI 사무국 (EU AI Office)	• EU 집행위원회 산하의 독립적인 전담 조직 • 법의 해석, 집행, 평가 및 국제 협력 등을 수행 • AI 전문성을 갖춘 독립 전문가로 구성
EU AI 이사회 (EU AI Board)	• 회원국 대표, 유럽 데이터 보호 감독관, 집행위 고위 대표 등으로 구성 • 법 해석의 일관성 확보, 감독기관 간 협력, 정책 조율 역할
과학 전문가 패널 (Scientific Panel)	• 기술적 판단이 필요한 사안에 대해 자문 제공 • 정책 집행의 전문성과 객관성 확보
자문단 (Advisory Forum)	• 산업계, 학계, 시민사회 대표 등이 참여하여 다양한 이해관계자의 의견을 공식적으로 반영

자료: 소프트웨어정책연구소(2024) 요약

이와 더불어 각 회원국은 자국 내 AI 규제 및 감독을 수행할 국가 관할 당국(National Competent Authorities)을 지정해야 하며, 고위험 AI 시스템에 대한 시판 후 모니터링(post-market monitoring), 심각한 사고 보고, 법 위반 시 제재 집행 등을 수행한다. 이러한 중앙-지역 연계형 구조는 AI 규제의 실효성을 높이고, 기술 발전 속도에 따른 법 집행의 유연성과 적시성을 보장하는 핵심 기반으로 작동할 것으로 평가된다.

2. AI 대륙 행동계획[42)]

EU는 AI 분야 선도 대륙으로 자리매김하기 위해 2025년 AI 대륙 행동계획(AI Continent Action Plan)을 발표하였다. 이하는 KISTEP(2025)를 참고하여 EU AI 대륙 행동계획의 내용을 요약·정리하였다.

가. 개요

1) 배경 및 추진 현황

EU 집행위원회는 미국 및 중국과의 기술 경쟁에 적극 대응하고자 AI 대륙 행동계획을 발표하였다. 유럽의 AI 리더십 강화를 위한 이 계획은 파리 AI 행동 정상회의(2025.2)에서 발표된 2,000억 유로 규모 InvestAI의 후속 조치로 AI 인프라 구축, 산업 적용 확대, 인재 양성 등을 포함한 종합적 AI 전략을 구체화하였다. EU 전역에서 컴퓨팅 역량을 현재 대비 3배 이상 확대하는 것을 목표로 한다.

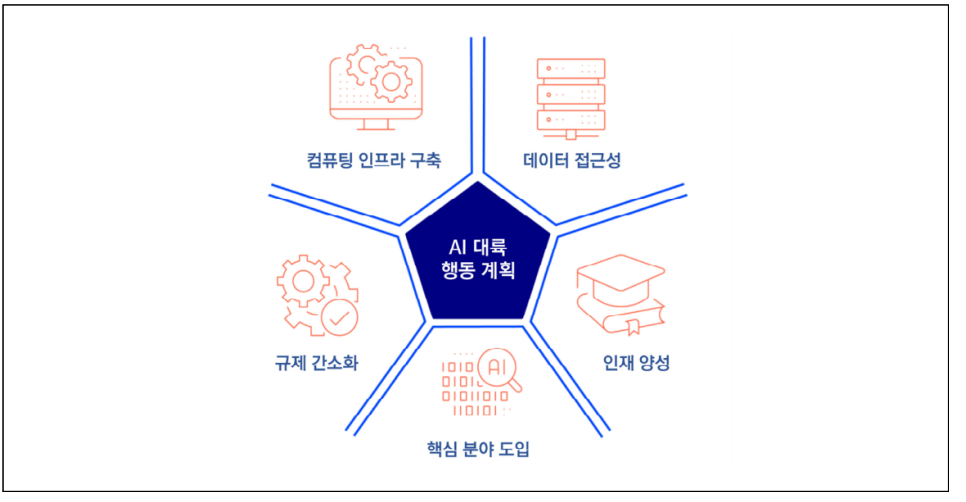
실행계획은 EU의 기존 AI 법을 유지하면서도 행정 부담 및 기업 혁신 촉진을 위한 규제 간소화 방안을 모색하던 중 발표된 것으로, EU의 AI 규제가 AI 산업 성장을 저해한다는 미국의 비판, EU 기업들의 AI 규제에 대한 부담 호소에서 기인하였다. 이에 본 실행계획은 기존 규제 중심에서 경쟁력 강화 중심으로 EU의 AI에 대한 접근 변화를 시사한다.

나. 주요 내용

AI 대륙 행동계획은 5개 핵심 영역(컴퓨팅 인프라, 데이터, 알고리즘 개발 및 산업 도입, 인재 강화, 규제 간소화)과 영역별 주요 전략으로 구분된다.

42) KISTEP(2025.4.9), "EU, 'AI 대륙 행동 계획' 발표로 규제-혁신 균형 모색"

〔그림 3－2〕 AI 대륙 행동계획 5대 핵심 영역



자료: KISTEP(2025)

1) 컴퓨팅 인프라 구축

대규모 AI 컴퓨팅 인프라 구축을 위한 AI 팩토리와 기가팩토리를 확대할 계획이다. 2025년~2026년까지 EU 전역에 AI 팩토리 배치(13개) 및 AI 기가팩토리(최대 5개) 설립을 추진하여 EU의 컴퓨팅 역량을 현재 대비 3배 이상 확대를 목표로 하고 있다.

2) 데이터 접근성(AI 데이터 공유 증진)

클라우드 및 데이터센터 확충과 데이터 접근성 향상 전략이다. 2025년 11월에는 데이터 연합 전략(Data Union Strategy)[43]을 통해 AI 개발과 혁신을 지원하는 데이터 생태계를 강화하고자 하였다. 또한, AI 팩토리와 연계한 데이터 랩을 구축하여 다양한 출처의 고품질 데이터를 통합하고 해당 분야 데이터 공간과 연결하여 AI 개발자에게 제공하는 것을 포함한다.

43) 데이터 연합 전략은 고품질 데이터를 통해 AI를 활성화하는 것을 목표로 함. 데이터의 미활용 잠재력을 활용하고 데이터 단일 시장을 완성하는 것을 목표로 유럽의 고품질 데이터 수요를 충족하기 위함임. 이를 위해 데이터 병목 현상 해결, 데이터 규칙 간소화를 위한 지침과 이니셔티브를 포함함(자료: EC 디지털 전략 홈페이지, https://digital-strategy.ec.europa.eu/en/policies/data-union).

3) 인재 양성(AI 스킬 및 인재 강화)

인재 양성 전략과 인재 확보 전략을 포함한다. 전자는 AI 스킬 아카데미를 설립하여 AI 전문가를 양성하며, 장학금 제도와 여성 전문가 복귀 프로그램도 운영하여 EU 기반의 AI 인재풀을 확대하고자 하였다. 글로벌 인재 확보를 위해서는 비EU 국가 출신 AI 인재 유치를 위한 다양한 제도인 MSCA Choose Europe(유럽내 연구자의 장기 경력 기회 제공 및 전세계 우수연구자 유치 이니셔티브) 추진, 블루카드 지침(EU 차원의 고소득 외국인 전문 인력 유치 프로그램) 개선을 제시하였다.

4) 핵심 분야 도입(혁신 촉진 및 AI 도입 가속화)

EU 강점 분야(제조, 우주항공, 에너지, 자동차, 제약, 로봇) 중심으로 산업별 AI 적용과 혁신을 촉진하기 위함 ApplyAI 전략을 일컫는다. 또한 유럽 디지털 혁신 허브(EDIHs)를 AI 경험 센터로 전환하여 중소기업과 공공기관의 AI 도입 지원 및 산업별 전문성을 강화하는 내용이 담겼으며, GenAI4EU 이니셔티브로 7억 유로 R&D 투자도 계획에 포함되었다.

5) 규제 간소화(규제 준수 및 간소화)

AI 법 준수를 촉진하는 동시에 규제로 인한 부담 경감을 위한 간소화 조치를 동시에 추진하도록 하고 있다. 2025년 7월 AI 법 데스크를 출범하여 기업에 AI 법 관련 정보와 맞춤형 상담을 제공하는 등 기업의 규제 준수를 지원하고자 하였다. 한편, AI 법과 다른 관련 법률(의료기기, GDPR 등) 간 일관된 적용을 위한 지침을 마련하여 혼란을 줄이고, 국가별 규제 샌드박스를 구축하여 혁신 친화적인 규제 환경을 조성하는 것을 목표로 하였다. 아울러 AI 개발 및 혁신을 위한 데이터 활용성 제고를 위해 GDPR 등 기존 데이터 규제의 복잡성 및 행정 부담 경감 방안을 검토하는 것을 추진하는 내용도 포함하였다.

3. 실천강령

가. 개요

1) 배경 및 추진 일정

범용 AI(General-Purpose AI, 이하 GPAI)의 확산은 다양한 산업과 사회 전반에 걸쳐 혁신적 가능성을 열어주고 있다. 특히 대규모 언어모델(LLM)을 포함한 GPAI 모델은 텍스트 생성, 이미지 편집, 코드 작성 등 광범위한 작업을 수행할 수 있어 다양한 AI 시스템의 기반 기술로 기능하고 있다. 그러나 이 모델이 적절한 통제 없이 대규모로 배포되고 활용될 경우 사회적·경제적 체계에 유의미한 영향을 미치는 체계적 위험(systemic risk)으로 이어질 수 있다는 우려도 제기되었다.

유럽연합(EU)은 2024년 제정된 AI법에 기반하여 GPAI 모델의 제공자에 대한 규제 틀을 마련하였다. 특히 2025년 8월부터 적용될 AI 법(제53조 및 제55조)은 GPAI 모델의 제공자에게 투명성, 저작권 준수, 체계적 위험 평가 및 완화 의무 등 핵심 조치를 요구하고 있으며, 이를 구체화한 자율규범이 GPAI 실천강령(Code of Practice for General-Purpose AI Models)이다. 이 실천강령은 법적 구속력을 가지지는 않지만 AI 법상의 의무 이행을 입증하고 평가받기 위한 핵심 수단으로 기능할 예정이다. 또한 해당 강령은 최신 기술과 모범 사례를 반영하여 GPAI 모델의 책임 있는 개발과 활용을 유도한다.

실천강령은 EU AI 사무국(AI Office)의 주도로 제정되었으며 전문가로 구성된 독립된 의장단이 전체 절차를 이끌었다. 초안 작성 과정은 투명하고 포용적인 구조를 갖추고 있으며, GPAI 제공자, 하위 응용 제공자, 산업계, 시민사회, 권리자, 학계, 국제기구 및 EU 회원국 대표 등 약 1,000여 명의 이해관계자가 참여하였다. 9월 말 총회(Kick-off Plenary)가 개최되었고 이후 4개의 주제별 작업반(Working Groups)이 구성되어 세 번의 회의를 거쳐 초안 작성 및 피드백이 반복되었다. 주요 GPAI 제공자들은 별도의 워크숍에 초청되어 실질적인 논의와 작성에 공동 참여하였으며, EU 회원국 대표는 AI 위원회(AI Board)를 통해 전 과정에 깊이 관여하였다.

〔그림 3－3〕 실천강령 수립 타임라인

자료: EC 디지털 전략 포털 - AI 실천강령

나. 주요 내용

실천강령은 크게 일반 GPAI 제공자에게 적용되는 공통 이행의무와 체계적 위험 보유 모델에 적용되는 추가 이행의무로 구성된다.

1) GPAI 모델 제공자의 공통 이행 의무

실천강령은 투명성 확보 의무와 저작권 준수 의무를 모든 GPAI 모델 제공자에게 요구한다. 이 조치는 모델이 어떠한 기능을 수행하는지, 그리고 그 학습 및 작동 과정이 합법적 기반 위에 있는지를 외부에서 확인할 수 있도록 하기 위한 것이다.

가) 투명성

투명성 의무는 GPAI 모델의 핵심 기능, 사용 목적, 작동 방식, 제한사항 등에 관한 정보를 문서화(model documentation)하고, 이를 최신 상태로 유지하는 것을 말한다. 이 문서는 AI 사무국, 감독 당국 또는 모델을 활용하는 하위 제공자가 요청 시 제공될 수 있도록 준비되어야 하며, 최소 10년간 보관되어야 한다. 또한 모델 관련 정보에 접근할 수 있는 연락처와 정보 제공 채널을 명확히 고지해야 하며 요청에 대한 응답은 14일 이내로 처리하는 것이 원칙이다.

정보의 품질, 무결성, 보안성 보장도 핵심이다. 실천강령은 해당 문서가 허위이거나 변경되거나 위·변조되지 않도록 신뢰할 수 있는 기록 관리 시스템을 마련할 것을 권장하고 있으며, 이를 위해 국제 기술 표준이나 업계 프로토콜을 적용할 것을 제안한다.

나) 저작권

저작권 준수 의무는 GPAI 모델이 EU 저작권법, 특히 텍스트·데이터 마이닝 관련 규정을 준수하도록 하기 위한 체계를 갖출 것을 요구한다. 모델 제공자는 내부 저작권 정책을 수립하고 이를 지속적으로 업데이트 및 이행해야 하며, 웹 크롤링 시 합법적으로 접근 가능한 콘텐츠만 수집해야 한다. 예를 들어 유료 서비스, 기술적 보호조치가 있는 자료는 무단으로 우회하거나 수집할 수 없으며, 법원이 명시한 불법 사이트는 데이터 수집 대상에서 제외되어야 한다.

또한, 콘텐츠에 명시된 기계 판독 가능한 권리 유보 표현(reservation of rights)[44]을 기술적으로 식별하고 이를 존중해야 한다. 이를 위해 robots.txt나 메타데이터 기반의 국제 표준을 따르도록 요구하고 있으며, 향후 표준 개발 논의에의 참여도 권장된다.

저작권 침해 방지를 위한 결과물 관리도 중요하다. 모델이 저작권 보호 콘텐츠를 무단으로 재생산하지 않도록 방지책(기술 보호조치, 사용자 고지, 이용약관, 출력 필터링 등)을 마련해야 한다. 아울러 이해관계자들이 문제를 제기할 수 있도록 공식 연락 창구를 지정하고, 전자적 의사소통 채널을 운영하는 것도 필수 조치로 포함하고 있다.

---

44) 저작권 보호를 위한 기술적 조치를 의미하며, AI 및 디지털 환경에서의 저작물 사용 조건을 명확하게 표시하는 것을 의미함. 즉, 콘텐츠가 AI나 크롤러 등에 의해 자동 수집·활용되는 것을 방지하기 위해 저작권 보호 조건을 컴퓨터가 이해할 수 있는 형식으로 명확히 표시함.

2) 체계적 위험(Systemic risk) 보유 모델의 이행 의무

일부 GPAI 모델은 기능이 고도화됨에 따라 체계적 위험을 수반할 가능성이 존재한다. 실천강령은 이러한 모델에 대해서 보다 강화된 추가 이행 의무를 적용하여 사전에 위험을 식별하고 수용 가능 여부를 판단하며, 필요한 경우 적절히 완화하도록 요구한다.

먼저, 제공자는 자체적으로 모델의 특성과 활용 맥락을 바탕으로 체계적 위험을 구조적으로 식별해야 하며, 이 과정에는 위험 유형 목록화, 원인 분석, 시나리오 개발이 포함된다. 위험은 공중보건, 기본권 침해, 공공질서 위협 등 사회 전반에 미치는 영향 기준으로 분류되며, 식별된 위험은 정량적·정성적으로 분석된다.

분석은 외부 정보(문헌, 사용자 보고, 사고 DB 등), 모델 자체의 성능 및 행동 평가, 자동화된 테스트, 시뮬레이션, 레드팀 평가 등 다양한 수단을 통해 수행된다. 이후에는 위험 발생 가능성과 심각도를 종합적으로 추정하여 각 위험이 수용 가능한 수준인지 판단한다. 만일 위험이 수용 불가능할 경우 모델의 출시·사용은 중단되어야 하며, 안전 완화 조치를 즉시 실행해야 한다.

실천강령은 위험을 낮추기 위한 안전 및 보안 조치는 매우 구체적으로 제시하고 있으며, 일례로 학습데이터의 정제 및 필터링, 출력 검열, 탈옥(jailbreaking) 대응, 모델 접근 제한(API 인증, 파라미터 지연 공개), 사용자용 위험 완화 툴 제공 등이 포함된다.

사이버보안 관점에서는 모델 도난, 무단 배포를 방지하기 위해 위협 행위자별 대응 목표를 설정하고, 보안 프레임워크를 구축해야 한다는 내용을 포함한다.

이러한 조치는 모두 모델 보고서(Model Report)와 안전·보안 프레임워크 문서에 기록되며, 모델 출시 전 이를 AI 사무국에 제출해야 한다. 보고서에는 모델의 기술적 설명, 성능 비교, 안전 완화 전략, 외부 평가 결과 등 다양한 항목이 포함되며, 시스템적 위험이 변화하거나 주요 변동사항이 있을 경우에는 5영업일 이내에 보고서를 갱신하여 제출해야 한다.

마지막으로, 실천강령은 심각한 사건(Serious Incidents)이 발생할 경우, 인지 즉시 AI 사무국에 단계별 보고가 이루어져야 함을 강조하고 있다. 중요 인프라 중단은 2일, 보안 침해는 5일, 사망 사건은 10일 이내 보고해야 하며, 사건 관련 정보는 5년 이상 보관되어야 한다.

제3절 미국

1. AI 정책·제도(2021~2024)[45]

가. 목표

미국은 AI 기술이 국가 안보, 경제 성장, 사회 전반에 미치는 영향이 광범위하고 심대하다고 판단하고, 안전하고 신뢰할 수 있는 AI 개발·활용 환경을 조성하는 것을 목표로 삼았다. 바이든 행정부는 AI 혁신 촉진과 함께 개인정보·인권 보호, 알고리즘 책임성 강화, 국가 안보 위험관리 등을 핵심 과제로 설정하였으며, 연방 차원의 전략과 규제·지원 인프라를 단계적으로 마련하였다.

나. 주요 정책 방향

2021년부터 2024년까지의 AI 정책은 안전성과 신뢰성 확보를 최우선 과제로 삼았다. 이를 위해 법률과 행정명령, 권리장전 등을 통해 안전·보안·차별 방지·책임성 등 핵심 원칙을 제도화하고, AI 모델에 대한 평가와 검증 절차를 구체적으로 마련하였다. 특히 국가 안보와 개인정보 보호 강화를 위해 국가 및 경제 안보에 영향을 미칠 수 있는 AI 모델에 대한 사전 검증 의무를 부과하고, 민감 데이터의 해외 전송을 제한하였다.

또한 기술 표준화와 국제 경쟁력 확보를 위한 노력을 병행하였다. 국립표준기술연구소(NIST)와 에너지부(DOE)를 중심으로 AI 안전성·보안성·신뢰성 확보를 위한 표준, 도구, 테스트 개발을 추진하며, 글로벌 AI 거버넌스 논의에도 적극 참여하였다. 아울러 연구와 교육 인프라 확충에도 주력하여, 국가 인공지능 연구 자원 시범사업(National AI Research Resource Pilot, 이하 NAIRR 파일럿)[46]을 통해 연구자와 교육자를 지원하고, 의료·환경·인프라 지속가능성 등

45) 소프트웨어정책연구소(2024)를 참고하여 작성함.

46) National AI Research Resource(NAIRR)는 미국 백악관 산하 과학기술정책실(OSTP)과 국가과학재단(NSF)이 주도하여 설계한 공공 AI 연구 지원 플랫폼으로, 2020년 인공지능 국가 전략에 따라 제안되었고, 2023년 말 NAIRR 시범사업이 시작됨. 공공 AI 연구 자원(컴퓨팅 파워, 데이터셋, 툴, 알고리즘 등)을 제공하고, 대형 기술 기업에만 집중된 AI 개발 자원을 공공기관, 대학, 소수 연구기관에도 확대하고자 함. 또한 윤리·법률·사회적 기준(ELSI) 가이드라인을 통해 윤리적이고 책임 있는 AI 개발을 촉진함.

사회적 문제 해결을 위한 AI 활용을 촉진하였다.

다. 제도와 정책

2021년 1월, 미국은 「국가 AI 이니셔티브법(The National AI Initiative Act of 2020)」을 제정하여 AI 기술의 다차원적 영향에 대응할 수 있는 제도적 틀을 마련하였다. 이에 따라 백악관 과학기술정책실(OSTP) 산하에 국가 AI 이니셔티브실(NAIIO)이 설치되어 국가 차원의 AI 정책을 총괄 조정하게 되었다.

2022년 10월에는 「AI 권리장전 청사진(The Blueprint for an AI Bill of Rights)」에서 안전하고 효과적인 시스템, 차별 방지, 개인정보 보호, 사전 고지와 설명, 인적 대안 보장을 5대 원칙으로 제시하였다. 이를 통해 AI 개발·활용 과정에서 발생할 수 있는 부작용을 최소화하고 윤리적 지침을 확립하였다.

2023년 10월 바이든 대통령은 「안전하고 신뢰할 수 있는 AI를 위한 행정명령(Executive Order on Safe, Secure, and Trustworthy AI)」을 발표하였다. 이 명령은 AI 권리장전을 확장하여 구체적·실행 가능한 조치를 마련하고, 국가 안보·경제·공공 안전에 영향을 미치는 AI 모델에 대해 훈련 단계부터 정부 고지와 AI 레드팀 안전성 평가를 의무화하였다. 또한 국립표준기술연구소(NIST)와 에너지부(DOE) 등을 통해 AI 안전성·보안성·신뢰성 확보를 위한 표준·도구·테스트 개발을 추진하였다.

행정명령에 기반한 후속 조치로, NIST 내에 미국 AI 안전연구소(US AI Safety Institute, US AISI)가 설립되고 연방 정부의 AI 활용 위험 관리 가이드라인 초안이 공개되었다. 이 과정에서 사람이 만든 콘텐츠 인증, AI 생성물 워터마크 표시, 유해 알고리즘 식별 등 구체적인 규칙과 기술 가이드라인이 마련되었다.

2024년 1월에는 미국 국립과학재단(NSF)이 NAIRR 시범사업을 시작하여 의료·환경·인프라 지속가능성 등 사회 현안 해결을 위한 AI 연구를 지원하고, 교육자를 대상으로 책임 있는 AI 접근법 교육 인프라를 제공하였다. 같은 해 2월 바이든 정부는 민감한 개인 데이터와 정부 관련 데이터를 중국을 포함한 '관심 대상 국가'로 전송하는 것을 제한하는 행정명령을 발표하였다. 이는 AI 기술이 방대한 데이터 세트에서 패턴을 인식하고 개인정보를 재식별할 위험에 대응하기 위한 조치였다.

2. AI 행동계획[47)]

가. 개요

“미국 AI 행동계획”(America’s AI Action Plan)은 2025년 7월 트럼프 행정부의 과학기술정책실(OSTP) 주도로 발표되었으며, AI 분야에서 미국의 글로벌 기술 우위를 확보하고 경제적 경쟁력, 국가 안보, 국민의 번영을 달성하기 위한 국가 전략을 제시한다. 이 계획은 트럼프 대통령이 취임 직후 수립을 지시한 것으로, AI를 단순한 기술이 아닌 반드시 이겨야 할 국가 간 경쟁의 장으로 정의한다. 특히 바이든 행정부의 규제 중심 접근을 폐기하고 민간 주도의 혁신을 촉진하는 친시장 전략을 핵심 기조로 설정한 것이 특징이다.

계획은 세 가지 핵심 축(pillar)을 중심으로 구성되어 있다. 첫째, 혁신 가속화(Accelerate AI Innovation)에서는 불필요한 규제를 제거하고 오픈소스 생태계를 육성하여 AI 기술 혁신을 촉진하는 전략이 담겨 있다. 둘째, 미국 AI 인프라 구축(Build American AI Infrastructure)에서는 AI 도입을 위한 전력, 반도체, 데이터센터, 노동력 등 핵심 인프라 확충과 보안을 강조한다. 셋째, 국제 외교 및 안보 선도(Lead in International AI Diplomacy and Security)에서는 AI 기술과 인프라의 글로벌 확산을 통해 미국 주도의 기술 외교를 강화하고, 적대국의 기술 접근을 차단하는 전략을 제시한다.

pillar별 정책 조치에는 산업 활성화와 이용자 보호와 관련된 정책 조치가 포함되어 있는데, 이를 관련성을 기반으로 구분하면 다음과 같다.

나. AI 산업 활성화 관련 정책

AI 행동계획은 민간 주도의 AI 기술 혁신과 산업 생태계 활성화를 핵심 목표로 설정하고 이를 위한 광범위한 제도적 기반과 투자 방안을 제시하고 있다.

우선, 연방정부는 과도한 규제와 관료주의가 AI 기술 도입을 저해한다고 판단하고, 전방위적 규제 철폐 작업을 시행하고 있다. 연방 및 주 차원의 불필요한 규제 요소를 식별하고 폐지하며, 규제가 과도한 주정부에는 연방 예산 지원을 제한하는 방안도 함께 추진된다.

또한, 오픈소스 및 오픈웨이트 AI 모델의 보급과 확산이 장려된다. 연구 커뮤니티는

47) 백악관 홈페이지(https://www.whitehouse.gov/, 최종 접속: 2025.7.24)

NAIRR 프로그램을 통해 민간 컴퓨팅 자원, 모델, 소프트웨어에 접근할 수 있도록 지원받으며, 중소기업을 위한 오픈모델 확산 전략도 병행된다.

정부 조직 내부의 AI 활용 역시 중요한 축이다. 연방정부는 최고 AI 책임자 협의체(CAIOC)를 공식화하고, 각 부처의 AI 도입을 촉진하기 위한 조달 툴박스, 인재 교류 프로그램 등을 운영한다. 특히 고영향 서비스 제공기관(HISP)을 중심으로 시범 도입이 진행된다.

한편, AI 기술의 확산에 따라 노동시장에도 큰 변화가 예상되며, 이에 따라 행정부는 AI 관련 교육 확대, 고임금 일자리 창출, 재직자 전환교육 등을 포함한 노동자 중심의 전환 정책도 수립하였다. 직업기술교육(CTE), 견습제도, 세금감면 혜택 등이 주요 수단으로 활용된다. 이와 함께 AI 생태계를 뒷받침할 인프라 구축도 강조된다. 에너지 수요 증가에 대응하기 위한 전력망 안정화, 반도체 제조 역량 강화(CHIPS R&D), 고보안 데이터센터 구축, 사이버보안 대응체계 강화 등이 포함되어 있으며, 이러한 인프라 확충은 일자리 창출과 산업 재건의 수단으로도 활용된다.

또한, 차세대 제조업에 대한 투자를 통해 드론, 자율주행차, 로봇 등 AI 기반 물리기술의 생산 경쟁력을 제고하고 있으며, 과학기술 기반 강화를 위한 자동화 실험실, AI용 데이터셋, 공공 연구 인프라 개선도 병행되고 있다.

다. AI 이용자 보호 관련 정책

AI 행동계획은 이용자 보호라는 표현을 직접적으로 사용하지는 않지만, 기술 오용과 피해를 방지하고 사회적 신뢰를 확보하며, 정보와 시스템의 안전을 강화하기 위해 정보·법률·생명안전·사이버보안 등 광범위한 영역에서 AI의 사회적 책임과 공공 안전 확보를 위한 구조적 대응 전략을 포함한다.

첫째, 합성 미디어 대응 차원에서 딥페이크 영상이 재판 과정에서 증거로 악용되는 문제에 대응하기 위한 법과학적 평가 기준 개발이 추진된다. 이를 위해 국립표준기술연구소(NIST)는 딥페이크의 판단 기준을 공식적인 벤치마크로 제시하는 방안을 검토 중이며, 법무부는 연방 증거 규칙 개정 가능성도 함께 논의하고 있다.

둘째, AI 시스템의 실패나 오작동에 대비하기 위한 사고 대응 체계를 강화하고 있다. 둘째, AI 시스템의 실패나 오작동에 대비하기 위해 사고 대응 체계를 강화하고 있다. 연방 각 기관은 사이버보안국(CISA)의 사고 대응 플레이북을 수정하여 AI 관련 사항을 반영하고,

최고정보보안책임자(CISO)가 최고 AI 책임자(CAIO), 개인정보 보호 담당 고위 관계자, 상무부 산하 AI 표준·혁신 센터(CAISI), 그리고 기타 관련 기관 관계자들과 협의하도록 하는 내용을 포함해야 하며, 이에 따라 기관별 내부 지침서도 함께 정비하도록 권고받고 있다.

셋째, AI 시스템의 보안성과 신뢰성을 확보하기 위한 보안 설계(Secure-by-Design) 원칙이 강조된다. 이는 AI 시스템이 데이터 오염, 프라이버시 공격, 적대적 입력 등으로부터 안전하게 설계되어야 함을 의미하며, 정보공동체(IC) 전용의 AI 신뢰성 기준(ICD 505)이 제정되어 기관별 적용이 확대되고 있다.

넷째, AI 시스템의 해석 가능성, 통제력, 견고성에 대한 기술 투자가 이루어진다. 국방고등연구계획국(DARPA), AI 표준 및 혁신 센터(CAISI), 국가과학재단(NSF) 등은 해커톤 기반의 시험 및 평가 방식을 통해 시스템의 보안성과 설명 가능성을 검증하고, 이러한 연구를 향후 국가 R&D 전략에 반영하도록 하고 있다.

다섯째, Frontier AI 모델이 야기할 수 있는 국가안보 위협에 대응하기 위한 체계적인 평가와 감시가 추진된다. 사이버공격, 생물·화학·핵무기 개발 등 고위험 시나리오에 대비하여, 연방 정부는 관련 전문기관들과 함께 보안 위험을 사전 분석하고 대응체계를 유지하도록 하고 있다.

마지막으로, 바이오안보(Biosecurity) 분야에서는 AI가 병원체 합성 등 생물학적 위협에 악용되는 것을 막기 위해, 연방 자금을 지원받는 모든 과학기관은 강력한 염기서열 스크리닝 및 고객 인증 절차를 갖춘 합성업체만을 사용하는 것을 의무화한다.

제 4 절 중국

1. AI 정책·제도(~2024)[48]

가. 제도 설계와 정책 기조

중국은 AI를 국가 차원에서 핵심 성장 동력으로 규정하고 AI 대국으로 도약하기 위한 정책을 추진하고 있다. 정부는 거대 자금을 투입하여 AI 기술 개발과 산업화를 촉진하고

48) 한국지능정보사회진흥원(2025.3) 요약하여 작성함.

반도체·로봇·컴퓨팅 자원 등 핵심 분야에서 자립 능력을 강화해 왔다. 특히 Embodied AI와 AI+ 전략을 중심으로 AI 기술을 모든 산업과 응용 분야에 결합하는 전면적 활용을 중시하였다.

아울러 중국은 정책 추진 과정에서 분야별로 규제를 수립하였다. 「생성형 AI 서비스 이용 잠정 법안」, 「딥페이크 관리 규정」 등이 그것이며, 이로써 개발과 서비스 제공 단계별 의무를 부과하면서 가짜뉴스, 개인정보 침해 등과 같은 역기능을 완화하고자 하였다.

나. 주요 정책

중국은 AI 산업 경쟁력 강화를 위해 산업별 특화형 AI 모델 개발과 로봇 기술 고도화를 병행하며, 금융·의료·교육·제조·물류 등 다양한 분야로 AI 응용을 확산하려는 노력을 하였다. 특히 'Embeded AI' 분야를 전략적으로 육성하기 위해 선전시와 화웨이 등과 협력하여 글로벌 혁신센터를 설립하고, 2030년까지 1조 위안 이상의 산업 규모 달성을 목표로 하였다.

디지털 경제 심화를 위해서는 디지털 산업화와 산업의 디지털화를 동시에 추진하며, AI를 빅데이터·클라우드·산업 인터넷과 융합하는 정책을 전개하고 있다. 'AI+ 행동계획'을 통해 스마트제조, 스마트의료, 스마트도시, 스마트농업 등 중점 분야에서 AI 기술을 활용하며, 플랫폼 기업 혁신과 데이터 활용 촉진을 위해 국가데이터국을 설립하고 데이터 개방·유통 정책을 강화한다.

AI 산업 생태계 조성을 위한 정책도 펼치고 있다. 오픈소스 커뮤니티와 오픈데이터를 활성화하여 AI 개발의 진입장벽을 낮추고 민관 협력 생태계를 조성하고 있다. 또한 클라우드 기반의 MaaS(Model as a Service) 서비스 확산을 통해 중소기업과 연구기관의 AI 활용을 지원하며, 데이터 수집·가공부터 솔루션 개발과 서비스 도입까지 AI 가치사슬을 완비하여, 고속·저비용의 개발 환경을 제공하고 있다.

규제와 거버넌스 측면에서는 「인터넷 정보 서비스 알고리즘 추천 관리 규정」, 「딥페이크 관리 규정」 등 분야별 법령을 시행하여 가짜뉴스, 개인정보 침해 등 AI 위험에 대응하고자 하였다. 생성형 AI에 대해서는 「생성형 AI 서비스 이용 잠정 법안」을 통해 개발과 서비스 제공 단계별 의무를 구체화하고 등급별 감독 체계를 운영하며, 국제 규칙 제정에도 적극 참여하여 글로벌 AI 거버넌스 논의에서 중국의 입장을 반영하고자 하였다.

2. AI+ 행동계획 심화

가. 개요

2025년 8월 중국 국무원은 2035년까지 스마트 사회 구현과 지능형 경제체제 구축을 목표로 하는 'AI+ 행동계획 심화 실시에 대한 의견'을 발표하였다. 스마트 단말과 시스템 보급률을 70%(~2027년), 90%(~2030년) 달성하고, 이를 기반으로 2035년까지는 전면적 스마트 사회를 구현하는 것을 목표로 한다. 이를 위해 6대 AI+ 중점 분야, 8대 역량 강화 사항을 제안하였다.

이번 발표는 과거 '인터넷 플러스'의 '정보 연결'에서 나아가 'AI+'를 통한 '지식 창조'와 '생산성 혁신'을 지향한다는 점에서 차이를 보인다. 단순히 AI+ 기존 산업이 아니라 기획 단계부터 AI를 내재화하는 AI 네이티브 모델을 지향하며, '1인 기업+AI 협력'과 같은 새로운 인간-기계 협업 모델의 확산을 기대하고 있다. 또한, 미국의 AI 계획이 국가안보와 외부 경쟁에 집중하는 것과는 달리 국내 사회·경제 시스템 변혁에 초점을 맞추고 있다는 점이 차이라 할 수 있다.

나. 주요 내용

1) 일반 요구사항

2027년까지는 6대 중점 분야와 인공지능의 심층 융합 시기로, 단말·시스템 보급률 70%를 초과하도록 하며 공공 거버넌스에서 역할을 강화하고 개방 협력 체계를 완비한다.

2030년은 AI 전면적 고품질 발전 지원을 목표로 하고, 단말·시스템 보급률 90%를 초과하는 것을 목표로 한다. 스마트 경제가 경제 성장 동력으로 작동하며 기술을 보편화하고 성과를 공유한다.

2035년은 스마트 경제와 스마트 사회 발전의 새로운 단계로 진입하는 시기로, 사회주의 현대화를 위한 강력한 지지기반을 제공한다.

2) 6대 AI+ 중점 분야

과학기술, 산업 발전, 소비 고도화, 민생 복지, 거버넌스 역량, 글로벌 협력을 6대 중점 분야로 한다.

〈표 3-3〉 6대 AI+ 중점 분야

분야	내용
과학 기술	인공지능을 기반으로 과학 발견, 기술 연구개발, 철학·사회과학 연구 혁신을 가속화하며, 다학문 융합과 인류 복지 증진을 촉진함
산업 발전	전 산업을 지능화하고, 신모델·신사업 창출을 위해 산업구조 혁신과 업그레이드를 촉진함
소비 고도화	서비스와 제품 소비 영역 전반에 AI를 융합하여 새로운 소비 시나리오와 스마트 제품 생태계를 창출함
민생 복지	고용·교육·의료·문화·돌봄 등 생활 전반에 AI를 융합하여 일자리 창출, 학습 혁신, 건강 증진, 문화 확산, 사회적 돌봄을 강화함
거버넌스 역량	사회·안보·생태 전반에 AI를 접목하여 인간-기계 공생형 사회 거버넌스, 다원 협력 체계, 정밀 생태 관리 체계를 구축함
글로벌 협력	AI를 국제적 공공제로 간주하여 보편적 혜택을 제공하고 격차 해소를 도모하며 유엔 중심의 글로벌 거버넌스 체계를 구축함

자료:'AI+ 행동계획 심화 실시에 대한 의견'에 기반하여 연구진 작성

3) 8대 역량 강화

AI 모델 기초 능력 제고, 데이터 공급체계 혁신, 지능형 컴퓨팅(AI칩) 역량 강화, 애플리케이션 개발 환경 최적화, 오픈소스 생태계 조성, 전문 인재 육성, 정책 안전장치 강화, 보안 역량 강화를 주요 역량 강화 분야로 지정하였다.

〈표 3-4〉 8대 역량 강화 분야

분야	내용
AI 모델 기초 능력 제고	AI 기초 이론과 모델 구조 혁신을 강화하여 훈련·추론 효율을 높이고 복잡한 과업 처리와 상호직용 경험을 개선하며, 모델 역량 평가 시스템 구축으로 지속적 개선을 촉진함
데이터 공급 체계 혁신	응용 중심 접근을 통해 고품질 데이터세트 구축을 강화하고 데이터 권리 제도 정비 및 공급 인센티브 확대를 통해 데이터 활용 기반을 고도화 함
지능형 컴퓨팅(AI칩) 역량 강화	AI 칩·소프트웨어 혁신과 초대규모 연산 집합 기술 상용화를 촉진하고, 국가 차원의 컴퓨팅 자원 최적화 및 협동을 통해 보편적이고 효율적이며 안전한 컴퓨팅 파워 공급을 추진함
애플리케이션 개발 환경 최적화	국가 차원의 애플리케이션 시범 기지와 공동 플랫폼을 조성하고, 서비스 기업 육성·산업 생태계 발전·표준 체계를 확립하여 AI 응용 확산과 서비스 체인 고도화를 추진함

분야	내용
오픈소스 생태계 조성	AI 오픈소스 커뮤니티와 프로젝트를 육성하고 기여 평가·인센티브 메커니즘을 구축하며 글로벌 협력을 강화해 국제적 영향력을 확대함
전문 인재 육성	AI 교육·평가·국제 협력·인센티브 체계를 종합적으로 강화하여 리더급 인재와 청년 인재를 육성하고 인재 활용 기반을 확충함
정책·규제 안전장치 강화	AI 투자·재정 지원(국유자본 투자 등), 법규·윤리 정비(AI 법률·법규·윤리 규범 완비), 안전 관리(AI 안전 평가 및 관리 제도 최적화)를 통해 지속 가능한 발전과 리스크 관리 기반을 강화함
보안 역량 강화	모델·데이터·인프라·응용 전반의 안전 역량을 강화하여 위험을 예방하고 규정 준수·투명·신뢰 가능한 AI 환경을 조성함

자료:'AI+ 행동계획 심화 실시에 대한 의견'에 기반하여 연구진 작성

4) 조직 및 실행

중앙에서는 국가발전개혁위원회가 총괄 조정을 담당하고 부처간 협력 체계를 강화한다. 지역과 부처는 실정에 맞게 구체적으로 실행하여 성과를 도출하도록 한다. 우수사례와 경험을 적절히 정리/홍보하여 전국적으로 확산하며, 사회적 합의를 형성하고 전 사회적 참여 분위기를 조성하도록 한다.

3. AI 안전 거버넌스 프레임워크 2.0

가. 개요

2025년 9월 15일 중국 정부는 AI 안전 거버넌스 프레임워크 2.0(AI Safety Governance Framework 2.0)을 발표하였다. AI 안전 거버넌스 프레임워크 1.0 발표('24.9) 이후 AI 기술·응용의 급속한 발전에 따른 새로운 위험과 도전에 대응하면서 자국 AI 기술과 산업 발전을 도모할 필요에 따른 것이다. 이에 AI 안전 거버넌스에 대한 폭넓은 합의를 형성하고 협력적 거버넌스 및 포용적 이익을 확산한다는 목표를 두고 국가인터넷정보판공실(CAC) 주도, 국가표준화관리위원회(TC260) 실무총괄, 국가컴퓨터네트워크비상기술대응조정센터(CNCERT/CC)·연구기관·산업계 참여로 프레임워크를 마련하였다.

나. 주요 내용

AI 안전 거버넌스 프레임워크 2.0은 AI 거버넌스의 원칙을 밝히고 AI 안전 위험을 유형별로 분류한 뒤 각 위험에 대응할 기술적·종합적 거버넌스 조치를 제시한다. 마지막으로 AI 연구·개발, 활용 안전 지침을 아우르고 있다.

1) AI 거버넌스 원칙

AI 안전 거버넌스 2.0은 AI 기술 및 산업 발전을 중시하면서도 안전을 우선시한다는 방향성 하에서 기술 발전과 위험 관리의 균형을 확보하는 것을 핵심으로 하고 있다. 이를 위해서 지켜야 할 5가지 원칙은 아래와 같다.

① 포용과 신중: 사회 전체가 참여하는 포용적 접근을 지향하고, 기술 발전 속도에 휩쓸리지 않으며 신중하게 대응한다.

② 위험 중심, 민첩한 대응: AI의 불확실성을 전제로 위험을 선제적으로 감지하고 민첩하게 대응한다.

③ 기술과 관리의 결합: 기술적 조치와 제도적·관리적 장치를 결합하여 AI 전주기 안전성을 보장한다.

④ 개방 협력: 정부·산업·학계·국제사회가 협력하고 공동의 거버넌스를 구축한다.

⑤ 신뢰와 통제 보장: AI 응용의 신뢰성을 확보하고, 인간이 통제력을 상실하지 않도록 보장한다.

2) AI 위험 분류

이전 버전의 AI 안전 위험 유형을 보완·갱신하여 내재적 위험, 응용 위험, 파생된 위험으로 구분하였다. 내재적 위험은 AI 기술 자체가 가진 구조적 결함에서 비롯되는 위험으로 알고리즘 위험, 데이터 위험으로 구성된다. 응용 위험은 AI 기술의 활용 과정에서 발생하는 위험을 뜻하며, 사이버 시스템 위험, 콘텐츠 위험, 현실 위험, 인지 위험으로 나누고 보다 세부적으로 구분하여 제시하고 있다. 파생 위험은 AI 기술이 장기적·사회적으로 미칠 부정적 영향력으로, 고용 구조 붕괴와 같은 사회/환경 위험, 사회적 편향 심화와 같은 윤리적 위험이 포함된다.

〈표 3-5〉 AI 위험 분류

위험	세부 유형		설명
내재적 위험	모델·알고리즘 위험	설명 가능성 부족	AI 의사결정 과정을 사람이 이해하거나 해석하기 어려운 위험
		편향과 차별	훈련 데이터나 알고리즘의 불균형으로 특정 집단에 불공정하게 작동하는 위험
		낮은 강건성	다양한 환경이나 공격 상황에서 성능이 저하되는 위험
		신뢰할 수 없는 산출	사실과 다른 결과나 오류를 산출하는 위험(환각 등)
		외부 적대적 공격	해커나 악의적 세력이 입력을 조작하여 잘못된 결과를 유도하는 위험
		모델 결합 전이	여러 모델을 결합할 때 기초모델의 위험이 다른 모델로 전이되는 위험
	데이터 위험	데이터의 불법 수집과 사용	개인정보의 동의 없는 수집, 부적절한 사용 위험
		훈련 데이터의 부적절한 콘텐츠	불법적·유해한 데이터가 학습에 포함되거나 중독(poisoning)될 위험
		훈련 데이터의 부적절한 주석(라벨링)	잘못된 라벨링으로 인해 모델 학습과 결과가 왜곡되는 위험
		데이터 및 개인정보 유출	학습·활용 과정에서 개인정보나 민감 데이터가 유출되는 위험
응용 위험	사이버 시스템 위험	구성 요소 및 컴퓨팅 안전	시스템 하드웨어·소프트웨어의 취약점으로 인한 보안 위험
		사이버 공간 노출 확대	AI 응용이 사이버공격의 새로운 경로로 악용되는 위험
		공급망 안전	기술장벽, 수출통제 등을 통한 칩, 데이터, 소프트웨어 공급 중단 위험
		사이버 공격에의 남용	AI를 활용해 보다 쉽게 사이버공격이 이루어질 위험
	콘텐츠 위험	불법·유해 정보 생산	AI가 생성한 오정보, 허위정보, 불법 콘텐츠가 확산되는 위험
		사실 왜곡과 사용자 기만	라벨링 되지 않은 AI 생성 콘텐츠의 진위를 평가하지 못해 사회 혼란을 가져올 위험
		온라인 콘텐츠 생태계 오염	저품질의 AI 생성물이 관련 콘텐츠를 오염시키거나 온라인 콘텐츠 전반의 품질을 저하할 위험
	현실 위험	경제와 사회에 대한 새로운 도전	AI 기술이 공공 인프라에 적용될 때 AI 기술의 내재적 위험으로 인한 인프라 안전에 미칠 위험
		불법·범죄 활동에서의 AI 사용	범죄에 AI 기술이 사용될 위험

위험	세부 유형		설명
		핵·생물학·화학 및 미사일 무기 지식·역량 통제 상실	특정 집단이 AI 기술을 대량살상무기 개발과 결합하여 안보를 위협할 위험
	인지 위험	정보 코쿤 효과 악화	이용자가 편향된 정보만 접하고 사회적 인식이 왜곡되는 위험
		인지전 지원 (Assistance in cognitive warfare)	AI가 선전·여론조작·심리적 개입 등에 활용되어 집단적 인식·판단을 교란하거나 사회적 갈등을 촉발할 위험
파생 위험	사회 · 환경 위험	고용 구조 붕괴	노동 가치 약화로 인한 고용 불안, 일자리 감소 위험
		자원 수급 균형에 대한 도전	설비 등에 의한 에너지 소모와 탄소 배출 증가 등 환경 위험
	윤리적 위험	사회적 편향 심화 및 AI 격차 확대	AI에 의한 집단별 차별·편향 심화, 지역 간 AI 격차 확대 위험
		교육에 대한 영향과 혁신 억제	AI 의존으로 학습·연구·창작 역량 약화 위험
		연구 윤리 위험의 심화	민감 분야 연구 확산, 비윤리적 활동 증가 위험
		의인화 상호작용 중독·의존	인간이 AI와 감정적 의존 관계 형성할 위험
		기존 사회질서에 대한 도전	산업 재편·가치 교란으로 사회질서 흔들릴 위험
		AI 자의식 출현과 인간 통제 상실	AI 자율성·자의식으로 인간 통제 상실 위험

자료: AI 안전 거버넌스 프레임워크 2.0에 기반하여 연구진 작성

3) 위험 대응을 위한 기술적 조치

앞서 분류한 각 위험에 대응할 기술적 수단을 제안한다. 내재적 위험에는 적대적 훈련 실시, 결함 전이 평가 실시, 보안 표준 수립, 주석 표준화 등이, 응용 위험 대응에는 침입 탐지 시스템 마련, 정보 콘텐츠 진위 검증 알고리즘과 사실성 평가 장치 도입, 사용자 보호 장치(AIGC 경고 표시·라벨링) 마련 등이 포함된다. 파생 위험에 대응하기 위해서는 그린 AI 기술 표준화, 에너지 절감 기술 개발, 차별 방지 기술 적용, 비상 통제 조치 설계 등과 같은 기술적 조치가 제안되었다.

〈표 3-6〉 AI 위험에 따른 기술적 조치

위험		기술적 조치
4.1 내재적 위험	4.1.1 모델·알고리즘 위험	(a) 내부 구조, 추론 논리, 인터페이스, 산출 과정 설명 제공 (b) 모델 구조 개선, 데이터 다양성 확대, 인간 감독 도입 (c) 설계·R&D 단계에서 보안 결함 제거, 적대적 훈련으로 강건성 강화 (d) 기초·오픈소스 모델의 보안 결함 전이 평가 강화
	4.1.2 데이터 위험	(a) 수집~삭제까지 보안 규칙 준수, 이용자 권리 보장 (b) 거짓·편향·구식·민감 데이터 배제 (c) 데이터 라벨링의 정확성과 신뢰성 확보 (d) 법규 준수, 합성 데이터 활용으로 개인정보 의존도 완화 (e) 데이터 사용 및 산출에서 IPR 침해 방지
4.2 응용 위험	4.2.1 사이버 시스템 위험	(a) 원칙, 역량, 적용 시나리오, 위험 공개로 시스템 투명성 강화 (b) 권한 관리·비필수 서비스 비활성화·접근 통제 강화 (c) 결함·백도어 제거, 취약점 추적·테스트·보수 실시 (d) 칩·소프트웨어·데이터 등 자원 공급망 보호 (e) 중복 설계, 재해 복구 체계 마련
	4.2.2 콘텐츠 위험	(a) 모델 변조·간섭 방지 메커니즘 구축 (b) 입력·출력 동적 필터링, 불법 콘텐츠·민감 정보 차단 (c) 생성 콘텐츠의 식별·추적·신뢰성 확보
	4.2.3 현실 위험	(a) 오남용 방지를 위한 기능 축소 (b) 오류 수정, 내결함성, 검증 메커니즘 마련 (c) 회로 차단기·원클릭 제어 등 긴급 대응 장치 도입 (d) 보조 운전·드론 등 고위험 응용의 인지 시스템 테스트 (e) 핵·생화학 무기 등 고위험 응용 방지를 위한 추적 강화
	4.2.4 인지 위험	(a) 부정확·거짓 산출물 식별 및 규제 (b) 사용자 데이터 수집·분석·남용 억제 (c) AIGC 탐지 기술 R&D 강화, 인지전 예방·탐지·대응
4.3 파생 위험	4.3.1 사회·환경 위험	(a) 자원 효율적·환경 친화적 AI 개발 모델 혁신 지원, 그린 AI 기술 표준화 (b) 저전력 칩, 효율적 알고리즘 등 그린 컴퓨팅·에너지 효율 해법 활용
	4.3.2 윤리적 위험	(a) 데이터 필터링, 가치 정렬, 산출 검증을 통해 인종·성별·지역 등 차별 예방 (b) 공공 안전·중요 인프라 등 핵심 부문 AI 시스템에 긴급 제어 장치 도입 (c) 설명 가능한 알고리즘 개발·적용을 장려하여 사용자 신뢰 구축

자료: AI 안전 거버넌스 프레임워크 2.0에 기반하여 연구진 작성

4) 종합 거버넌스 조치

AI 안전을 확보하기 위해 기술적 대응을 넘어 제도적·사회적 차원의 조치가 필요하다는 점을 강조하고 있다. 이에 따라 연구개발기관, 서비스 제공자, 정부 부처, 사회 조직, 사용자 등 다양한 주체가 AI 위험을 식별·예방·대응할 수 있도록 조치를 마련하며, AI 안전 거버넌스의 국제 교류와 협력을 심화하여 공동 거버넌스를 구축할 것을 제안한다.

〈표 3-7〉 AI 안전 위험, 기술적 조치 및 종합 거버넌스 조치

위험			기술적 조치	종합 거버넌스 조치
내재적 위험	모델·알고리즘 위험	설명 가능성 부족	4.1.1 (a)	• 전 생애주기(연구개발 및 응용 포함) 안전성 강화 • AI 안전 평가 체계 구축 • 데이터 보안 및 개인정보 보호 규정 개선
		편향과 차별	4.1.1 (b)	
		낮은 강건성	4.1.1 (c)	
		신뢰할 수 없는 산출	4.1.1 (a)(b)(c)	
		외부 적대적 공격	4.1.1 (c)	
		모델 결함 전이	4.1.1 (d)	
	데이터 위험	데이터의 불법 수집 및 사용	4.1.2 (a)	
		훈련 데이터의 부적절한 콘텐츠	4.1.2 (b)(c)(d)(e)	
		훈련 데이터의 부적절한 주석	4.1.2 (c)	
		데이터 및 개인정보 유출	4.1.2 (d)	
응용 위험	사이버 시스템 위험	구성 요소 및 컴퓨팅 안전	4.2.1 (a)	• 오픈소스 생태계 안전 및 공급망 안전 강화 • AI 응용 분류 및 위험 등급 관리 시행 • AIGC 추적 관리 촉진 • 주요 산업 응용 수요의 안전하고 효과적인 실현 • AI 위험 및 위협 정보 공유
		사이버 공간 노출 확대	4.2.1 (b)(c)	
		공급망 안전	4.2.1 (d)	
		사이버 공격에의 남용	4.2.1 (e)	
	콘텐츠 위험	불법·유해 정보 산출	4.2.2 (a)(b)	
		온라인 콘텐츠 생태계 오염	4.2.2 (c)	
	현실 위험	경제와 사회에 대한 새로운 도전	4.2.3 (a)(b)(c)(d)(e)	
		불법·범죄 활동에서의 AI 사용	4.2.3 (a)(b)(c)(d)(e)	
		핵·생물학·화학 및 미사일 무기 지식·역량 통제 상실	4.2.3 (a)(b)(c)(d)(e)	

위험			기술적 조치	종합 거버넌스 조치
	인지 위험	정보 코쿤 효과 악화	4.2.4 (a)(b)(c)	
		인지전 지원	4.2.4 (a)(b)(c)	
파생 위험	사회·환경적 위험	고용 구조 붕괴	4.3.1 (a)	• AI 관련 법률·규제 프레임워크 제정 및 개선 • AI 과학기술을 위한 윤리 지침 수립 • AI 통제 상실 위험에 대한 협력적 위험 관리 공감대 조성 • AI 안전 인재 양성 강화 • 사회 전반의 AI 안전 인식 제고 • AI 안전 거버넌스에 관한 국제 교류 및 협력 촉진
		자원 수급 균형에 대한 도전	4.3.1 (b)	
	윤리적 위험	사회적 편향 심화 및 지능 격차 확대	4.3.2 (a)	
		교육에 대한 영향과 혁신 억제	4.3.2 (b)	
		연구 윤리 위험의 심화	4.3.2 (c)	
		의인화 상호작용 중독·의존	4.3.2 (d)	

자료: AI 안전 거버넌스 프레임워크 2.0

5) AI 연구·개발·활용 안전 지침

AI 안전 거버넌스 프레임워크 2.0은 모델·알고리즘 개발, 응용 개발 및 배치, 응용 운영 및 관리, 응용 접근 및 사용의 AI 단계별 안전 지침을 제시하며, 잠재적인 기술적 통제 불능 위험에 대비하기 위해 신뢰할 수 있는 AI의 기본 원칙을 제안하고 국제 사회의 공감대를 형성하도록 하고 있다.

① 모델·알고리즘 개발 단계: 설계 단계부터 공정성·투명성·보안성을 내재화하고 데이터 관리와 오픈소스 활용에서 윤리적·법적 책임을 준수하며, 지속적인 검증과 투명성을 통해 신뢰 가능한 기술 기반을 마련해야 한다는 내용이 포함된다. 구체적으로는 안전한 훈련 데이터셋 구축, 데이터 소스 표준화 관리, 교차 검증·감사 절차를 통한 라벨링 품질 개선 등이 있다.

② 응용 개발·배치 단계: 시나리오 적합성 평가와 공급망 보안을 확보하고, 접근 통제·데이터 보호·산출 관리 등 다층적 안전장치를 마련함으로써 응용 단계에서의 안정성과 책임 있는 배치를 보장하도록 해야 한다는 내용을 담고 있다. 위험 수준별 보안 평가·정기 감사 실시, 안정된 버전 선택, 불필요한 포트·서비스 비활성화, 데이터 백업·복구

계획 수립 등과 같은 구체적 조치가 포함된다.

③ 응용 운영·관리 단계: 서비스 제공자가 보안관리와 모니터링 체계를 확립하고 투명성과 책임성을 확보하며, 사용자 권리와 사회적 보호를 보장함으로써 AI 서비스가 안정적이고 신뢰 가능한 방식으로 관리·운영되도록 하는 지침이 포함된다. 일례로, 인간이 검토하는 절차를 마련해야 한다거나 실시간 위험 모니터링 및 관리 메커니즘을 구축해야 하며, 사용자의 권리를 보장하기 위한 수단을 마련해야 한다는 내용이 그것이다.

④ 사용자 접근·이용 단계: 이용자 스스로 위험을 인식하고 책임 있는 선택과 활용을 통해 개인정보 보호와 보안을 지키며, 사회적·윤리적 영향을 고려한 신중한 이용을 요구하는 내용을 담고 있다. 개인정보에 대한 인식을 강화하며 민감한 데이터는 입력을 기피하는 등의 조치가 그 예다.

제5절 기타

1. 영국[49)]

영국은 AI 기술 발전이 국가 경쟁력과 산업 혁신의 핵심 요소가 될 것으로 보고 산업 친화적 환경을 설계하여 AI 혁신을 촉진하는 것을 목표로 한다. 따라서 고정적이고 강제적인 규제를 최소화하고 부문별 특성과 위험 수준에 따라 유연하게 대응할 수 있는 규제 체계를 마련하고 있다. 대표적으로 2023년 3월 발표된 「AI 규제에 대한 혁신적 접근」(A pro-innovation approach to AI regulation)은 EU와 같은 강력하고 일률적인 규제 체계 대신 기존 부문별 규제기관이 각자의 관할 분야에서 상황에 맞는 지침을 채택하도록 하는 유연한 규제 프레임워크를 제시하고 있다. 즉, 새로운 단일 AI 규제기관을 설립하여 전권을 부여하지 않고 개별 규제 기관들이 5대 원칙(안전·보안·견고성, 투명성·설명 가능성, 공정성, 책임·거버넌스, 이의제기·시정)을 기준으로 우선순위를 정해 규제를 적용하도록 한 것이다. 2024년 3월에는 AI 규제 원칙을 법률로 제정하는 방안이 제안되었으나 더 이상 논의가 진전되지 못해 결국 폐기되는 사례도 있었다. 이는 여전히 영국이 AI 규제 체계를 법제화하기보다는 정책·가이드라인 중심으로 운용하고 있음을 보여주는 예라 하겠다.

한편, 영국은 AI 안전성에 관한 글로벌 논의를 선도하며 국제 표준 제정 과정에서 주요한 역할을 수행하고자 한다. 일례로 AI 안전성 국제 논의에서 주도적 역할을 위해 2023년 11월 세계 최초의 AI 안전성 정상회의(AI Safety Summit)를 개최하였다. 이 회의에서 영국은 미국과 중국 간 AI 산업 경쟁 구도를 활용하여 국제 AI 표준 제정에서 중요한 위치를 확보하려는 전략을 취했다. 또한 AI 안전성 국제기구 설립을 제안하며 주요국 사이에서 규제 논의의 중개자 역할을 수행하였다. 회의 결과 각국 정상은 고급 AI 모델 출시 전 사전 테스트를 의무화하고, 첨단 AI 모델이 초래할 수 있는 위험에 대응하기 위해 공유된 과학적 증거 기반의 설명 가능한 모델을 개발하는 데 합의하였다. 이를 통해 영국은 AI 기술 개발과 활용에서 안전성 확보를 위한 글로벌 합의를 주도했다고 평가받고 있다.

49) 소프트웨어정책연구소(2024)를 참고하여 작성함.

2. 일본[50]

일본은 AI 정책 전반에서 법의 지배, 적정 절차, 인권·프라이버시 존중이라는 원칙을 기반으로 법률과 가이드라인을 적절히 조합하는 규제 체계를 구축하고 있다(한국지능정보사회진흥원, 2025). 또한 정부는 AI 기술의 급속한 변화에 대응할 수 있는 위험관리 체계를 마련하고, 민관 협력을 통한 AI 거버넌스를 확립함으로써 '세계에서 가장 AI 개발·활용이 쉬운 나라'로 자리매김하는 것을 목표로 한다. 이를 통해 일본은 AI 혁신을 저해하지 않으면서도 국민의 안전과 신뢰를 확보하고자 한다.

가. 추진 배경 및 목표

일본은 급속히 발전하는 AI 기술의 잠재적 기회와 함께 가시화되는 위험에 대응하기 위해, '세계에서 가장 AI 개발·활용이 쉬운 나라'를 목표로 설정하였다. 생성형 AI의 성능이 고도화하고 활용이 확산하면서 국민 생활, 산업, 과학기술 전반에서 효율성과 편의성을 높이고 있으나, 이와 함께 허위·오정보 생성, 범죄 악용, 국가 안보 위협(CBRN 무기 개발, 사이버 공격 등)과 같은 다양한 위험 요인도 내포하고 있다. 이에 일본 정부는 AI 활용의 안전성과 투명성을 확보하면서도 혁신을 저해하지 않는 제도 설계를 지향하고 글로벌 협력과 민관 연계를 통한 AI 거버넌스 확립을 주요 과제로 삼고 있다.

나. 제도 설계의 기본 원칙

일본 정부는 AI 제도를 설계하기에 앞서 AI가 사회 전반에 안전하게 도입·활용되도록 하면서도, 국제 경쟁력을 유지할 수 있는 기반을 제공하기 위해 네 가지 기본 원칙을 내세웠다. 첫째, AI에 의한 위험 대응과 AI 혁신을 동시에 달성하는 것이다. 이를 위해 기술과 비즈니스 변화 속도에 대응할 수 있는 유연한 제도를 마련하고 사업자의 자율성과 혁신성을 존중하면서 법률 규제는 자율 노력으로 해결하기 어려운 분야에 한정하기로 한다.

둘째, 글로벌 상호운용성을 확보하는 것을 중시한다. 국제 표준화 기구(ISO, IEC) 활동에 적극 참여하고, 히로시마 AI 프로세스를 비롯한 다자간 AI 거버넌스 논의를 주도하여 일본의 입장을 반영한다.

셋째, 정부가 AI 조달과 이용을 주도한다. 정부 조달 과정에서 안전성과 품질을 확보할

50) 한국지능정보사회진흥원(2025.5)을 요약하여 작성함.

수 있는 가이드라인을 정비하고, 지자체와 중앙정부가 선도적으로 AI를 활용하여 민간 확산을 촉진하는 환경을 조성한다.

다. 주요 정책

이상의 네 가지 원칙에 기반하여 일본의 AI 정책은 혁신 촉진, 위험 대응, 글로벌 협력, 정부 조달·활용, 그리고 중점 위험 분야 관리라는 다섯 가지가 중점적으로 추진된다. 혁신 촉진을 위해 정부는 AI 연구개발을 적극 지원하고 인프라를 확충한다. 구체적으로는 고품질 데이터 제공, 데이터센터와 전력 인프라 확보, AI 반도체 지원을 포함한 기반 정비가 이루어진다. 또한 국립연구개발법인 이화학연구소를 중심으로 'AI for Science' 연구를 추진하며, 글로벌 인재 확보 경쟁에 대응하기 위해 차세대 AI 인재 육성 프로그램도 운영하여 인적 자원 확보 노력을 기울인다.

위험 대응 측면에서는 AI 사업자 가이드라인(2024.11.22. 제정)을 통해 인권, 법치, 공정성, 투명성, 책무성 등 10대 원칙을 제시하였다. 대표적으로는 AI 생애주기 전반에서 개발자, 제공자, 이용자 간 정보 공유를 강화하고, 디지털 워터마킹, 출처 증명과 같은 기술적 대응 방안을 도입한다. 또한 자발적 위험 평가와 제3자 인증 제도를 병행하여 안전성을 확보하며, 심각한 AI 관련 사고 또는 위험 상황 발생 시 정부가 직접 원인을 조사하고 재발 방지책을 마련해 국민에게 정보를 제공하는 것도 포함한다.

글로벌 협력의 강화를 위해 일본은 히로시마 AI 프로세스를 주도하여 국제 지침과 행동규범을 수립하고 OECD, 유엔, 유럽평의회와 함께 AI 거버넌스 논의에 참여한다. 아시아 최초로 GPAI 전문가 지원센터를 통해 생성형 AI 안전성 평가 관련 프로젝트도 지원한다.

정부 조달·활용 분야에서는 AI 조달 가이드라인을 정비하고 기존 지침을 심화하여 안전성과 품질을 확보한다. 기밀정보 취급 시 보안 기준을 철저히 준수하고, 지자체별 AI 활용 사례를 전국적으로 확산한다. 예를 들어, 고베시는 AI 조례를 제정하여 문서 요약, 아이디어 도출, 프로그래밍 코드 생성 등 다양한 활용을 추진하였고, 도쿄도는 '도쿄도 AI 전략회의'를 통해 정책 방향을 구체화하였다.

중점 위험 분야 관리를 위해 정부는 AI로 인한 생명과 신체 안전, 시스템 위험, 국가 안보 위협을 우선으로 다룬다. 산업별로 기존 법령을 기반으로 대응하되, 신규 위험이 발생하여

기존 체계로 대응이 어려운 경우에는 제도를 재검토하거나 새로운 제도를 마련하도록 하였다.

〔그림 3－4〕 AI의 잠재적 위험 사례(일본 'AI 사업자 가이드라인'의 예시)

AI 위험 사례	구체적 사례·상정 사례	주요 법령 등
AI에 기밀정보 입력	외부 AI 서비스에 기업의 비밀정보를 입력해 정보 유출	부정경쟁방지법, 민법(계약) ※ '24년 2월 『기밀정보 보호 핸드북』에서 AI 활용 유의 사항 정리(경제산업성)
AI 개발·학습·생성·이용에서 타인의 저작권 침해	특정 만화·애니메이션 캐릭터 등의 일러스트와 유사한 이미지 생성 목적으로 학습해 일러스트와 유사한 이미지 생성·이용	저작권법 ※ '24년 3월 『AI와 저작권에 대한 기본 방침』에서 해석 명확화(문화청)
AI 개발·이용 과정에서 타인의 산업재산권 침해	타인의 등록상표를 학습해 동일, 또는 유사한 상표를 생성하고, 그 지정 상품·역무와 동일, 또는 유사 상품·역무에 이용	의장법(디자인법), 상표법 ※ '24년 5월 『AI 시대 지식재산권 검토회 중간 정리』에서 법적 규율 방안 정리(AI 시대 지식재산권 검토회)
AI 개발·이용 과정에서 프라이버시 침해·개인정보보호 위반	본인의 동의 없이 개인정보를 포함한 데이터를 AI 학습에 이용	헌법(프라이버시권, 퍼블리시티권), 개인정보보호법 ※ '24년 6월 『개인정보보호법 3년마다 재검토 중간 정리』를 공표, AI 이용 시 논점 정리(개인정보보호위원회)
AI 탑재 제품 오작동	자율주행차 오작동으로 생명·신체 안전에 영향	도로운송차량법, 의약품의료기기법, 노동안전위생법, 민법(불법행위 등), 제조물책임법, 자동차손해배상보장법, 국가배상법
딥페이크(AI 합성 초상·소리 등 악용)	본인 동의 없이 개인 이미지를 음란물 등에 합성·확산하는 행위, 유명인·지인을 가장한 AI 음성통화로 사기	민법(인격권·불법행위), 형법(협박죄, 명예훼손죄, 음란물 배포죄, 사기죄, 위계업무방해죄 등), 아동포르노금지법, 정보유통플랫폼대응법(권리침해 정보)
AI에 내재된 편향(Bias)에 의해 차별·편견 조장	AI의 부적절한 채용·퇴직 판단, 인사평가 수행	헤이트스피치해소법, 노동관계법령, 민법, 개인정보보호법, 장애인차별해소법, 부락차별해소법
허위·오정보를 통한 여론 조작	입후보자에 관한 허위 정보를 AI로 생성하고 SNS 등을 통해 확산해 선거 방해	민법(인격권·불법행위), 형법(명예훼손죄), 행정법규, 공직선거법, 정보유통플랫폼대응법(권리침해 정보)
국민 권리 이익 침해	AI 오판으로 개인이 행정 서비스 수혜에서 배제되는 등 불이익을 받을 가능성	헌법(적정 절차), 행정절차법
바이러스 작성 등 사이버 공격	생성AI를 악용한 컴퓨터 바이러스 생성	형법(부정지령전자적기록에 관한 죄), 부성 액세스 금시법
환각(Hallucination) (AI가 허위 정보 생성)	생성AI가 허위 정보를 생성해 사용자 혼란 초래	민법(불법행위, 계약)
환경 부담 증가	AI 개발 과정에서 전력 수요 증가에 따른 CO_2 배출 증가	지구온난화 대책 추진법
인간과 AI 간의 부정적 상호작용	AI와의 대화에 몰입한 사람이 삶을 비관해 자살	자살대책기본법
AGI 통제 불능 우려	인간의 AGI(범용AI) 제어 불능으로 사회적 혼란 발생 가능성	법령 없음

※ 일본의 AI 사업자 가이드라인(제1.01판)에서 위의 여러 위험 요소를 언급하고 있으며, 10가지 공통 지침(인간중심·안전성·공정성·프라이버시 확보·보안 확보·투명성·책무성·교육·리터러시·공정 경쟁 확보·혁신 촉진)에 따라 이에 대한 대응 방안 제시

자료: 한국지능정보사회진흥원(2025)

제 4 장 AI 서비스 이용자 정책 방향

제 1 절 정책 환경

AI 기술은 연산 능력과 알고리즘 혁신에 기반한 대규모 언어모델과 생성형 AI로 발전하여 산업·사회 전반으로 빠르게 확산하고 있다. 기업은 생산성 향상, 의사결정 지원, 자동화 등을 위해 AI 활용을 확대하고, 사회적으로도 교육·문화·공공 서비스 등 전반에 걸쳐 AI 활용이 증가하고 있다.

AI는 향후 10년간 세계 경제에 가장 큰 변화를 가져올 기술로, 생산성과 자본 효율성을 제고하여 세계 성장률을 4~18%p 상승시킬 것으로 전망된다(국제금융센터, 2025.8.14). Renaissance Macro에 따르면, 미국에서는 2025년 AI 자본지출이 소비자 지출보다 GDP 성장에 더 큰 영향을 미친 것으로 분석되고 있다(Sherwood, 2025.7.30). 베인앤드컴퍼니의 분석 결과, 한국 경제 전반에 AI를 성공적으로 도입할 경우 2026년 310조 원의 경제 효과를 가질 것으로 예상된다(관계부처 합동, 2024.4.4).

〈표 4-1〉 주요 국제기관의 AI 경제적 효과 전망

기관	전망
IMF	AI로 총요소생산성(TFP) 개선 시, 향후 10년간 세계 성장률 1.3~4.0%p 상승
PwC	향후 10년간 세계 성장률 평균 8%p 증가 예상(범위 1~15%p)
골드만삭스	생성형 AI 확산 시 세계 총생산 7%p 이상 상승 가능
IDC	AI 투자 확대 시 약 20조 달러(세계 GDP의 17.7%) 효과
Mckinsey & Company	생성형 AI의 경제적 파급효과 2.6조 달러~4.4조 달러 예상

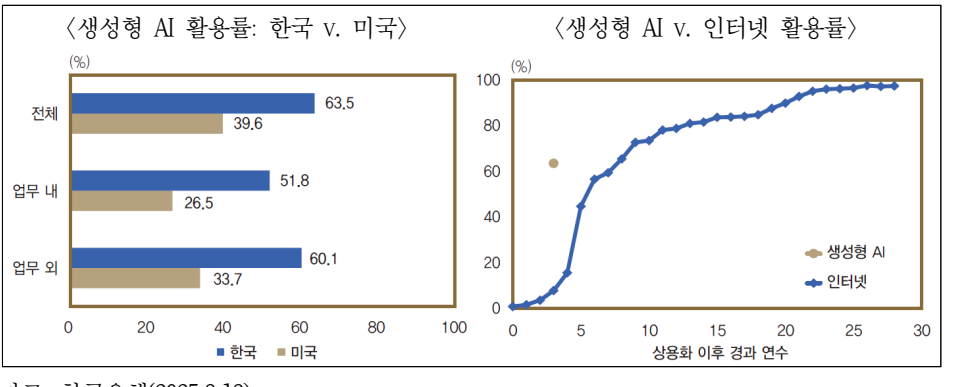
생성형 AI 서비스는 챗봇·대화형, 이미지 생성·편집, 영상·오디오 제작, 문서·검색, 교육 등 다양한 영역에서 확산되고 있다.

〈표 4-2〉 상위 50개 생성형 AI 서비스의 카테고리별 분류

카테고리	서비스명
챗봇·대화형 AI	ChatGPT, Deepseek, Character.AI, JanitorAI, Claude, SpicyChat.AI, Doubao, Kimi, Hailuo AI, Poe, Crushon AI, Candy.ai, Talkie, Monica, Meta AI, Joyland
검색·정보 탐색	Perplexity, Liner
문서·텍스트	QuillBot, Adot, Eden AI, PolyBuzz, Sider.AI, DeepAI, Gamma
이미지 생성·편집	Civitai, Leonardo.AI, Cutout.pro, Photoroom, Moonscape AI, MidJourney, Pixelcut, PixAI, Ideogram, Cnub, PicWish
오디오·뮤직	Suno, ElevenLabs
영상·멀티미디어	Kling AI, Sora, Zeemo, VEED, Invidea AI, Clipchamp, Bolt
개발자 도구	Hugging Face, BLACKBOX AI, Cursor
교육·생산성	Brainly, StudyX

전 세계 AI 도구 이용자는 2024년 약 2억 8,126만 명에 달하며, 2031년까지 약 8억 9,118만 명이 추가로 증가할 것으로 예상된다(2025년, Statista). 국내의 경우에도 생성형 AI의 확산 속도는 인터넷 도입 당시보다 8배 빠르고, 국내 근로자 63.5%가 생성형 AI를 사용하고 있다(한국은행, 2025.8.18).

〔그림 4-1〕 생성형 AI 활용률

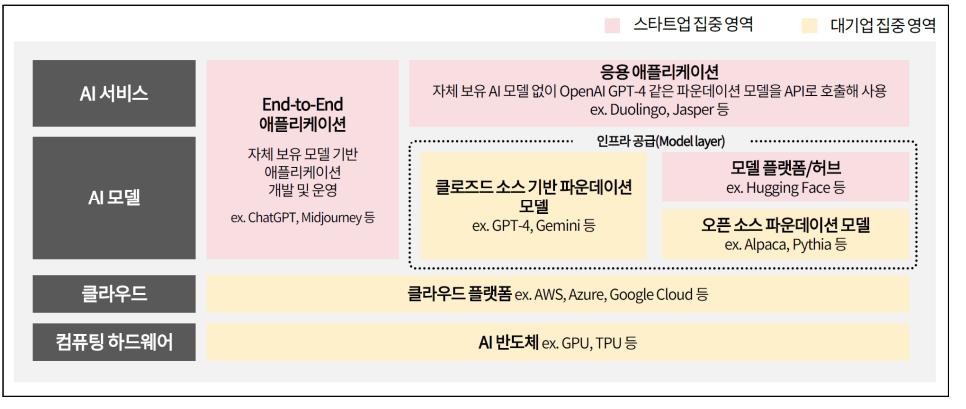


자료: 한국은행(2025.8.18)

생성형 AI 생태계는 빅테크 기업 위주로 구조화되고 있다. 컴퓨팅 하드웨어, 클라우드, AI 모델 영역에서 대규모 자본을 기반으로 한 빅테크 기업이 높은 시장점유율을 차지한다(2025년, IoT Analytics). 컴퓨팅 하드웨어(Data center GPUs) 영역은 엔비디아 92%, 클라우드

및 AI 모델 영역은 마이크로소프트 39%, AWS 15%, 구글 15%, 오픈AI 9%를 차지한다. 전문가들은 마이크로소프트, 구글, 아마존 등 주요 하이퍼스케일러의 AI 자본 지출이 가속화할 것으로 예상한다(Sherwood, 2025.1.2). 실제로 AI 패권을 차지하기 위한 빅테크 기업의 적극적 투자 경향이 확인되고 있다(2025년, Business Trends and Outlook Survey).

〔그림 4-2〕 생성형 AI 산업 생태계 구조



출처: 삼일PWC경영연구원(2024)

AI 기술은 초기 패턴인식형(머신러닝·딥러닝)에서 최근 생성형 AI로 발전하였으며, 향후 자율형 AI(Agentic/Physical)와 범용 AI(Artificial General Intelligence)로 패러다임 전환되어 2년 이내 에이전틱 AI, 2~5년 이내 물리적 AI, 10년 이후 범용 AI 기술이 주류가 될 것으로 예상된다(Gartner, 2025.8.5).

에이전틱 AI는 이용자의 목표를 달성하기 위해 자율적으로 생각하고 행동하며, 복잡한 작업을 스스로 수행하는 시스템이다. 시장 규모는 2024년 51억 2천만 달러에서 2030년 471억 5천만 달러로 예상된다(CAGR 44.8%)(Markets and Markets, 2024.9.12). 이와 관련하여 AI의 자율성이 강화되어 책임소재 불분명, 데이터의 편향성과 공정성, 프라이버시 침해 등에 대한 우려가 증가하고 있다.

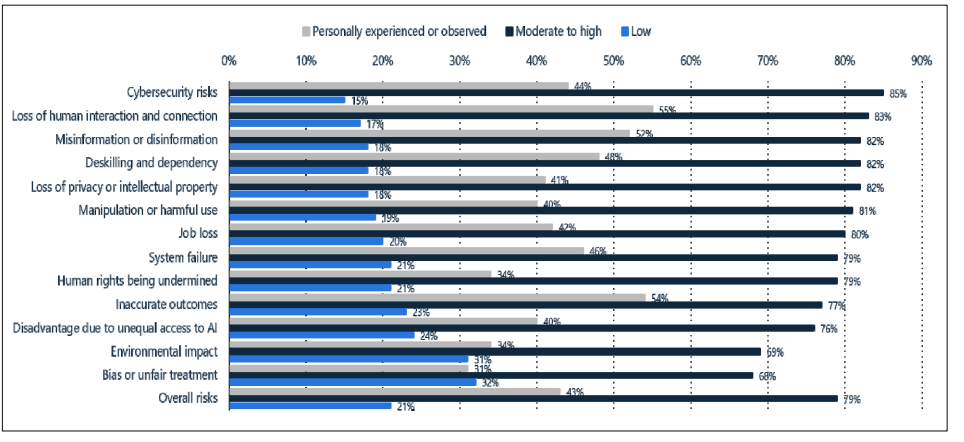
물리적 AI는 기계공학, AI, 센서, 연결성의 융합을 통해 현실 세계에서 자율적으로 작동하는 시스템이다. Markets and Markets에 따르면, 시장 규모는 2024년 20억 2천만 달러에서 2030년 152억 6천만 달러로 예상된다(CAGR 39.2%)(조선일보, 2025.11.24). 인간의

신체·공간·정보와 직접적으로 상호작용하는 기술 특성상 책임성과 안전성, 개인정보 보호, 인간-로봇 상호작용의 윤리성, 군사적 활용 가능성 측면에서 복합적인 쟁점을 동반할 것이다.

범용 AI는 인간 수준의 범용지능을 지닌 AI로, 다양한 영역에서 추론·의사결정·문제해결을 수행하며, 지식을 일반화할 수 있는 시스템이다. 시장 규모는 2024년 8억 5천만 달러에서 2032년 110억 3,730만 달러로 예상된다(CAGR 37.8%)(fortune business insights, 2025.12.29). UN은 범용 AI의 대표적인 6가지 위험 요인, 즉 되돌릴 수 없는 결과, 대량살상무기, 핵심 인프라 취약성, 권력 집중·불평등 심화, 존재적 위험을 제시하면서 UN 차원의 거버넌스가 필요함을 주장하고 있다(UNCPGA, 2025.9).

AI의 위험에 대해 이용자 79%는 중간~높은 수준의 우려를 보였고, 43%는 실제로 하나 이상의 부정적 결과를 경험하였다(2025년, Statista). 가장 큰 우려를 나타낸 항목은 사이버 보안 위험으로 전체 응답자의 85%가 중간 이상 수준의 우려를 표명하였으며, 43%는 실제로 해당 위험을 경험하거나 관찰한 바 있다고 응답하였다. 다음으로 인간 상호작용 및 연결의 상실에 대한 우려가 83%로, 41%는 해당 현상을 경험하거나 목격한 바 있다고 밝혔다. 허위조작정보 및 오정보, AI 기술에 대한 의존성 증가 및 기술 불능(disabling and dependency), 프라이버시 및 지식재산권 침해에 대한 우려는 모두 82%로 나타났으며, 관련된 직·간접의 부정적 경험은 각각 41%, 38%, 37%로 나타났다.

〔그림 4-3〕 AI 위험에 대한 주요 우려사항



출처: Statista(2025)

Al-kfairy 외(2024)[51]는 총 37편의 학제 간 문헌을 대상으로 한 체계적 분석을 통해 생성형 AI가 가져올 수 있는 심각한 윤리 문제를 다각적으로 조망하였다. 연구 결과 개인 프라이버시 침해, 데이터 보호 위협, 저작권 침해, 허위조작정보 및 오정보의 확산, 알고리즘 편향, 사회적 불평등 심화 등이 생성형 AI 기술 발전에 따른 주요 윤리적 쟁점으로 나타났다. 무엇보다 생성형 AI의 콘텐츠 생성 능력으로 인해 이러한 문제가 더욱 복잡해지고 광범위해지고 있다고 보았는데, 일례로 딥페이크는 진실성과 신뢰, 민주주의를 직접적으로 위협할 수 있다고 경고하며 기술 발전이 사회적 가치체계를 훼손하는 방향으로 작동할 수 있음을 지적하였다.

이러한 위험은 AI 서비스를 이용하는 국민의 권익을 침해할 뿐 아니라 AI가 기본 인프라로 작동하는 만큼 사회의 신뢰를 훼손하는 결과를 낳는다. 따라서 신뢰 공백으로 기술에 대한 사회적 수용이 약화되는 것을 방지하고 이용자의 권익을 보장하는 이용자 친화 정책 마련이 시급하다. 진흥 정책과 조화를 이루는 'AI-안전벨트', 즉 한국형 AI 서비스 이용자 정책을 통해 이용자는 안심, 기업은 혁신하는 환경을 조성할 수 있다.

51) Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024, September). Ethical challenges and solutions of generative AI: An interdisciplinary perspective. *Informatics; 11*(3), 58. Multidisciplinary Digital Publishing Institute.

제2절 정책 추진방향

1. 안전하고 신뢰할 수 있는 AI 활용 정보 유통환경 조성

AI를 악용하여 타인의 권리를 침해하거나 사회적 위험을 야기하는 정보를 생성·유포하는 것을 방지하기 위한 대응체계를 구축한다. AI를 활용하여 생성된 정보 및 방송콘텐츠의 신뢰성을 확보하고 역기능을 방지하여 서비스 안전성 및 이용자 권익을 제고한다. AI 이용자 피해를 모니터링하고 분쟁조정, 이용자 보호 측면의 위험성 및 업무 평가 등을 통해 피해 구제 체계를 고도화한다. AI의 건전한 활용을 촉진하면서 허위조작정보 등 부작용을 최소화하기 위한 AI 윤리·정보·활용 역량 강화를 추진한다.

2. 자유롭고 공정하며 혁신을 촉진하는 AI 생태계 구축

AI 생태계에서 다양한 주체에 의한 경쟁과 혁신의 기회를 보장하기 위해 불공정행위 규제를 강화하고, 이용사업자의 권익을 보호한다. AI 서비스에 대한 이용자의 실질적 접근권과 선택권을 보장하고, 극단주의·과몰입으로부터 아동·청소년 보호를 강화한다. 방송미디어 전 과정에 AI 도입을 촉진하고, AI를 활용한 방송콘텐츠 제작, 장애인의 방송접근권 확대 등을 적극적으로 지원한다. AI 서비스 활성화를 위해 규제체계를 합리화하고, 규제 샌드박스, 중소·스타트업·소상공인의 사업화·광고 등을 지원한다.

3. 자율 기반 AI 규범체계 확립 및 집행 실효성 강화

AI 서비스에 대한 신뢰성을 제고하고 범사회적·체계적 대응을 위한 법・제도적 규범과 윤리 기반을 마련한다. 해외사업자도 AI 규범을 준수하여 사각지대 및 규제형평성 문제가 제기되지 않도록 규제 집행력을 확보한다. 증거 기반 집행체계를 구축하여 글로벌 스탠더드를 선도하는 정책 역량을 갖추고, 정책·조사의 실효성 제고 위해 국제협력을 추진한다. 방송미디어통신위원회 업무에 AI를 활용하여 업무 효율성을 제고하고, 국민과의 소통을 강화한다.

제3절 핵심 추진과제

1. 안전하고 신뢰할 수 있는 AI 활용 정보 유통환경 조성

딥페이크 성범죄물이 명예 훼손, 성적 착취 등에 악용되고, 유명인의 무단 영상·이미지를 제작하여 허위 광고 또는 지지 발언 등을 날조하는 경우가 발생하고 있다. 피해자는 AI 조작 콘텐츠의 허위를 입증하거나 확인하기 어려워 범죄자에게 이용당하고 잘못된 정보에 신뢰성을 부여한다. 딥페이크를 포함하여 다양한 불법유해정보의 유통 방지를 위해 해당 정보를 생성·유포한 이용자에 대한 실효적 제재 및 주요 유통경로인 정보통신서비스 제공자의 유통 방지 조치 의무 부과가 필요하다. 온라인 콘텐츠의 빠른 변화 및 파급력, 정부 검열에 대한 우려 등으로 인해 불법유해정보 유통 방지를 위해 정보통신서비스 제공자의 자율규제를 도모하는 것은 불가피한 측면이 있다.

이에 딥페이크 성범죄물에 대한 신고 접수 시 방송심의위원회 심의 이전이라도 사업자가 임시 차단을 할 수 있도록 법적 근거를 마련하는 방안을 추진한다. 불법성이 명백한 정보에 대해서는 관계기관 요청에 따라 방송미디어통신위원회가 직접 사업자에게 삭제·차단을 요청할 수 있도록 법률 개정을 추진한다. 허위정보 자율규약에 딥페이크의 악의적 사용 금지를 포함하고, AI 생성물 표시 자율규약에 플랫폼의 딥페이크 표시 의무를 포함한다.

정보통신서비스제공자에게 불법정보 신고·처리 절차 구축 의무를 부과하고, 불법정보를 반복적으로 게재한 이용자에 대한 서비스 제공 중단, 수익 창출 제한 등 조치를 취할 것을 의무화한다. 불법정보 차단도 중요하지만, 잘못 차단되어 표현의 자유, 영업의 자유 등을 침해하지 않도록 처리방침을 미리 약관에 밝히고, 해당 이용자에게 그 이유와 내용을 미리 알리며, 이의제기 절차를 제공한다. 불법스팸에 대한 범죄수익 몰수, 과징금 부과 등 제재를 강화하고, AI 스팸 필터링 개선, 해외문자 차단함 신설 등 기술적·관리적 보호조치를 강화한다. 유해정보는 표현의 자유와 충돌할 가능성이 높아 법에 의한 강제 삭제보다는 우선 노출 배제, 연령 기반 필터링 등 사회적 합의에 기반한 자율규제 또는 서비스 설계 변경을 통한 간접규제를 추진한다. AI를 이용해 더욱 정교해지는 신원 도용을 방지하기 위해 본인확인기관 및 연계정보(CI) 제도의 고도화를 추진한다.

아동·청소년이 AI 서비스를 통해 신체·정신·도덕 발달을 저해할 수 있는 콘텐츠에 노출되거나 이들의 약점과 경험 부족을 악용하는 온라인 인터페이스 설계가 이루어질 수 있다. AI 알고리즘 기반 중독성 강한 피드, 숏폼 등은 아동·청소년의 인터넷 이용시간 과다 및 스마트폰 과의존의 주요인으로 지목되고 있다.

이에 아동·청소년이 사용하는 AI 서비스를 제공하는 사업자가 자율적으로 아동·청소년의 정보를 보호하고 안전을 도모하기 위해 적절한 조치를 취하도록 하는 방안을 검토한다. 유해정보, 그루밍, 사이버 괴롭힘, 중독성 등 온라인 위험으로부터 아동·청소년을 보호하기 위한 비례적이고 적절한 조치를 취할 것을 권고한다. 예를 들어, 자살/자해/섭식장애/음란물 등 유해정보에 노출되지 않도록 플랫폼 설계 변경하고, 성적 착취, 사이버 괴롭힘, 정서적 유해행위 등에 노출되었을 경우 신고 및 조치 절차 마련, 새로운 서비스 제공 또는 설계 변경 전 위험을 사전 평가하고 효과적인 완화 조치 시행 등의 조치 권고를 검토한다. 중독·극단주의로부터 아동·청소년을 보호하기 위해 AI 알고리즘 추천 기능에 대한 부모 통제권 보장 및 비활성화 기본설정 의무화 등을 추진한다. 또한, 부모가 자녀 계정에서 챗봇의 응답방식을 제어하고, 위기상황 시 즉각 알림을 받을 수 있도록 하는 기능 제공 권고 등을 검토한다.

2. 자유롭고 공정하며 혁신을 촉진하는 AI 생태계 구축

AI 생태계는 기술 발전에 따라 시장구조가 급변할 수 있는 빠르게 진화하는 시장으로, 동향 및 이슈를 정확히 모니터링하고 이해하는 것이 중요하다. 시장의 구조적 특성, 즉 대규모 데이터, 높은 수준의 개인화, 네트워크 효과 등이 경쟁 제한 행위와 결합되어 독과점이 형성되고 지배력이 고착화될 수 있으며, 이는 혁신기업의 진입/확장을 방해할 수 있다. 따라서 시장 발전 상황을 면밀히 모니터링하여 시장의 구조적 특성과 진화 방향을 이해하고, 경쟁 보호조치가 필요한지 여부를 검토해야 한다.

이에 AI 생태계의 현황과 향후 구조 변화를 분석하고, 경제·사회적 가치 및 부작용 등 전반적인 생태계 건강성을 진단하는 실태조사를 정기적으로 실시한다. 이를 통해 기존 입법, 자율규제, 상생협력, 정부 지원 등 정책의 실질적 효과를 분석하고, 정책 폐기·수정 또는 신규 정책을 제안한다. 주요 사업자의 불공정행위 모니터링을 강화하여 위법행위에 대해

엄격히 조치하고, 시장 기능에 의한 치유가 어렵거나 자율규제의 실효성이 미흡한 경우 시의적절한 법·제도적 대응방안을 검토한다. AI 자율규제 기구를 통한 AI사업자-이용사업자 간 분쟁조정 기능 제공 등 절차적 공정성을 강화한다. 각 사업자 고객센터 등을 통해 분쟁이 해결되지 않을 경우 분쟁조정을 신청할 수 있으며, 그 외 분쟁 사례조사 및 연구, 예방대책 권고 등을 수행한다.

이용자가 AI 서비스를 이용하는 과정에서 생성된 데이터의 활용에 대한 이용자 통제권을 보장할 필요가 있다. 자율주행차, 홈 AI 등 커넥티드 제품·서비스의 발전으로 데이터의 양과 잠재적 가치가 더욱 증가함에 따라 개인정보뿐만 아니라 非개인정보를 아우르는 포괄적인 데이터 통제권 보장이 필요하다. 이에 이용자가 커넥티드 제품·서비스 이용과정에서 생성된 모든 데이터(非개인정보 포함)에 접근·이용하거나 제3자에게 제공 요청할 수 있도록 보장하는 방안을 추진한다. 다만, 알고리즘을 통해 파생되거나 추론된 정보, 특히 커넥티드 제품에 있는 여러 센서의 산출을 결합한 정보는 제외한다. 커넥티드 제품·서비스 제공자는 非개인정보를 이용자와의 계약에 따라서만 이용하도록 의무화한다.

AI 추천 알고리즘은 정보 검색을 용이하게 하고 이용자 경험 향상에 기여할 수 있는 반면, 필터 버블 및 불법유해정보 확산의 구조적 토대로 작용할 수 있다. AI 알고리즘이 적용된 검색·추천 서비스가 확대되면서 최적화된 정보 노출이 다양한 관점을 접할 기회를 차단하여 확증편향 등 부작용을 야기할 수 있고, 알고리즘에 의존한 선택은 장기적으로 이용자의 주체적인 의사결정능력을 퇴화시키는 등의 문제를 발생시킨다. 또한, 이용자 참여를 극대화하기 위해 클릭 유도성과 자극성을 우선시하는 경향이 있으며, 그 결과 혐오 표현, 허위조작정보, 선정적 콘텐츠 등 사회적으로 유해한 콘텐츠 유포를 구조화하기도 한다.

이에 정보통신서비스 제공자의 자체적인 알고리즘 투명성 위원회 구성 및 보고서 제출 의무화를 내용으로 한 정보통신망법 개정을 추진한다. 또한, 특정 유형의 초대형 사업자에 한하여 이용자에게 맞춤형 정보 제공을 거부할 수 있는 선택권 보장을 의무화한다. 맞춤형 정보 제공의 편의성 및 기존 관행과 사업자 부담 등을 고려하여 옵트아웃 방식의 선택권을 부여한다.

방송미디어통신위원회는 2019년 '이용자 중심의 지능정보사회를 실현하기 위한 원칙'(이하 "AI 윤리원칙")을 제정하였다. AI 윤리원칙은 AI, 빅데이터, 사물인터넷 등 새로운 지능정보기술의 도입이 초래할 수 있는 기술적·사회적 위험으로부터 안전한 지능정보

서비스 환경을 조성하기 위해 모든 구성원이 고려해야 할 공동의 기본원칙을 제시한다. 하지만 최근 AI가 모든 국민의 기본 인프라로 기능하고, 단순한 기술 확산에 그치는 것이 아니라 경제·사회 전반을 근본적으로 변화시킬 것으로 예상됨에 따라 기존의 AI 윤리원칙을 개정할 필요성이 증가하였다. AI 개발·운영·이용 과정에서 영향을 받는 모두의 협력과 책임을 강조하며, 자율적인 문제해결 노력에 활용할 수 있는 규범 기준을 제시할 필요가 있다.

3. 자율 기반 AI 규범체계 확립 및 집행 실효성 강화

현재 AI 서비스 자율규제는 민간 주도의 순수 자율규제에 가까운 형태로 이루어지고 있으나, 자율규제를 총괄하는 체계가 부재하고, 자율규제 성과 도출 지연 및 체계적 관리에 한계가 존재한다. 향후 민관 협력을 강화하여 정부·업계 공동으로 자율규제 방안을 수립하고 체계적 성과 관리체계를 마련할 필요가 있다. 이에 업계·정부 고위급 및 전문가로 'AI 서비스 이용자 보호 민관협의회'를 구성·운영한다. 이 협의회는 문제 해결을 위한 의제 설정, 자율규제 원칙과 목표 공동 수립, 이행점검 결과에 따른 정책환류 등을 수행한다. 업계·정부 실무자, 이해관계자, 전문가 등으로 분과위원회를 구성하여 각 영역에서 세부 자율규약을 마련하고, 이행상황을 평가하여 그 결과를 공개한다.

기존 자율규제 방식은 구체화된 행위 규제 도출에 초점을 맞춰 예측 가능성은 높은 반면, 유연성과 재량이 미흡하다. AI 기술·시장이 급변함에 따라 이에 유연하게 대응하고 혁신을 저해하지 않도록 목표·원칙 중심의 자율규제를 추진해야 한다. 이에 현행법 적용이 모호하거나 구체화·보완이 필요한 부문에 대해 AI 민관협의회에서 자율규제의 목표·원칙을 수립하며, 주요 이슈에 대해서는 공동규제의 법적 근거를 마련한다. 회원사는 이를 각사의 사정에 맞게 구체화하여 업종별 또는 개별 자율규약을 제정·시행한다. 민간의 자정 노력에도 해소되지 않는 문제에 대해서는 면밀한 시장조사 결과에 기반하여 현행법의 합리화를 추진한다. 2025년 3월 시행된 생성형 AI 가이드라인의 이행점검 및 성과 평가를 통해 자율규제를 확산시키고 기술 발전에 대응한 개선을 추진한다. AI 에이전트 가이드라인을 제정하여 개발자, AI 에이전트 서비스 제공자, 이용사업자/최종이용자 간 책임분배, 프라이버시 보호, 공정성 등을 규정한다.

해외사업자에 대한 규제 집행력 강화를 위해 불법유해정보 직접 조치 등 관계법령을 정비하고, 청소년보호책임자 및 국내대리인 실효성 강화를 위한 제도 개선을 추진한다. 불법정보에 대해 국내 사업자에 대해서는 시정 요구를 통해 삭제 등 조치를 취하고 있으나, 해외사업자에 대해서는 국내 ISP 사업자를 통한 차단 이외에는 해외사업자의 자율규제 등 협조를 구하고 있는 실정이다. 하지만 국내시장에서 해외사업자의 비중이 지속 확대되고 있어 국내외 사업자 규제 형평성의 문제를 더 심각하게 고려할 필요가 있다. 글로벌 서비스를 통한 불법유해정보가 실효성 있게 삭제·차단될 수 있도록 직접 시정 요구 등 관계법령 정비를 추진한다. 해외사업자에 대해서도 불법정보 등에 대한 방심위의 시정 요구 및 방통위의 시정명령 등 직접 조치를 부과할 수 있도록 개선한다. 청소년유해정보로부터 청소년을 보호하는 청소년보호책임자 제도의 실효성 확보를 위한 단계적 정책을 마련한다. 해외사업자의 국내대리인 자격요건을 강화하고 업무범위를 확대하는 등 국내대리인 제도의 활용도 제고 방안을 마련한다.

전 세계 국가들은 빠르게 발전하는 AI 기술의 사회적 영향력을 인식하고 관련된 규제체계를 정비하고 있으며, 이러한 논의는 글로벌 차원으로 확산하고 있다. 새로운 AI 규범은 AI·디지털 기술의 연결성·즉시성 등으로 정보 확산의 가속화 및 국제적인 범위 등을 고려한 설계가 필요하므로, 국제사회의 연대와 협력이 필수적이다. 이에 OECD, UN 등 국제기구와의 협력을 강화하고, 미국, EU, 영국 등 주요국 규제기관과의 고위급 대화 채널 개설 및 공동 국제 컨퍼런스 개최 등을 통해 AI 규범 마련에 대해 협의한다. 이와 더불어, 우리나라가 글로벌 스탠더드를 선도할 수 있도록 증거 기반 AI 규범체계를 확립하여 효과성과 합리성을 제고한다. 외국 규제기관과의 정보 교환 및 협력, 국제기준에 부합하는 조사절차 도입 등을 통해 법 위반행위 조사 관련 국제협력을 강화하고, 국제 AI 규범 집행의 통일성을 확보한다.

제5장 결 론

AI는 이미 사회 전 영역에서 필수 인프라로 기능하고 있으며, 생성형 AI의 대중화와 함께 이용자 접점에서 발생하는 피해와 위험 또한 급속히 확대되고 있다. 오류, 편향, 환각과 같은 기술적 위험은 물론, 딥페이크 범죄, 허위조작정보 및 불법유해정보 확산, 프라이버시 침해, 알고리즘 추천에 따른 확증편향 강화 등은 이용자의 권익을 실질적으로 위협하고 사회적 신뢰를 약화시킬 수 있다. 동시에 데이터, 컴퓨팅, 클라우드, 모델, 플랫폼을 결합한 AI 생태계의 구조적 특성은 시장 집중과 정보 비대칭을 심화시키며, 공정경쟁 질서와 이용자 선택권을 약화시킬 우려를 내포한다. 따라서 향후 AI 정책은 혁신 촉진과 이용자 보호를 상호 대립되는 가치로 보지 않고, '신뢰 기반 혁신'을 목표로 예방적 안전장치와 공정한 시장 질서를 함께 구축하는 방향으로 설계되어야 한다. 이러한 문제의식에 기반하여 본 연구는 다음과 같이 AI 서비스 이용자 정책 방향을 제안하였다.

첫째, 안전하고 신뢰할 수 있는 AI 활용 정보 유통환경 조성을 위해 딥페이크 성범죄물 등 불법유해정보의 생성·유통 차단 체계를 강화한다. 딥페이크 성범죄물 신고 접수 시 심의 전 임시 차단 근거 마련, 자율규약 제정, 콘텐츠 출처 확인 지원 등을 추진하고, 플랫폼에 신고 및 처리 절차 구축과 반복 위반자에 대한 서비스 제한 또는 수익화 제한 등 적절한 조치를 취할 것을 요구한다. 다만, 오차단 방지를 위해 사전 고지 및 이의제기 절차를 마련한다. 아동·청소년 보호를 위해 유해정보, 그루밍, 중독성 추천 등에 대한 위험평가 및 완화조치, 부모 통제권 강화 등을 추진한다.

둘째, 자유롭고 공정한 AI 생태계 구축을 위해 AI 시장 및 서비스 이용 현황에 대한 정기 실태조사로 생태계 건강성을 진단하고 불공정행위 모니터링을 강화한다. 이용자 측면에서는 커넥티드 서비스 확산에 맞춰 데이터 통제권을 확대하고, 추천 알고리즘의 부작용 대응을 위해 알고리즘 투명성 강화와 초대형 사업자 대상 개인화 선택권 보장을 추진한다.

셋째, 자율 기반 AI 규범체계 확립 및 집행 실효성 강화를 위해 AI 민관협의회를 구성하고, 영역별 자율규약 수립-이행평가-공개로 실효성을 높인다. 기술 변화에 유연한 목표·원칙

중심 규범을 확대하고, 생성형 AI 가이드라인 점검 및 AI 에이전트 가이드라인 제정을 추진한다. 해외사업자에 대해서도 불법정보 대응 등 직접 집행력을 강화하고, 국제기구·주요국과 협력해 글로벌 스탠더드를 선도한다.

본 연구가 제시한 AI 서비스 이용자 정책 방향은 'AI 안전벨트'로서 이용자 보호를 두텁게 하면서도 산업 혁신이 지속될 수 있는 토대를 마련하는 데 초점을 둔다. 단기적으로는 AI 활용 정보의 신뢰성 확보, 피해 신고 및 상담 체계 고도화, 자율규제의 실효성 제고 등에 우선순위를 두고, 중기적으로는 이용자 권리의 제도화, 알고리즘 투명성 및 선택권 강화, 평가·인증 체계 정착과 분쟁조정 메커니즘 확립을 추진할 필요가 있다. 장기적으로는 AI 기술의 자율성과 물리적 확장(에이전틱/피지컬/범용 AI)에 대비하여 책임 배분 원칙과 위험 대응체계를 고도화하고 국제 공조를 확대함으로써 정책의 미래 적합성을 확보해야 한다.

AI 서비스 이용자 보호와 공정한 산업환경 조성은 단일 법령이나 단일 기관만으로 달성되기 어렵다. 기술 변화 속도에 맞춘 유연한 규율, 산업계의 책임 있는 자율, 이용자의 역량 강화, 정부의 전문적 거버넌스가 결합될 때 비로소 신뢰 기반의 AI 생태계가 형성될 수 있다. 본 연구 결과가 향후 AI 서비스 이용자정책 종합계획 수립과 제도 설계의 기초자료로 활용되어 이용자는 안심하고 AI를 사용하며, 기업은 신뢰 속에서 혁신하는 지속 가능한 AI 생태계 구축에 기여하기를 기대한다.

참 고 문 헌

〔국내 문헌〕

과학기술정보통신부·소프트웨어정책연구소(2025),『2024년 인공지능산업 실태조사』.

관계부처 합동(2025.4.4), “AI G3 도약을 위한 AI·디지털 혁신성장 전략”, 보도자료.

국제금융센터(2025.8.14), “AI 발전에 따른 신흥국의 상대적 성장 지연 가능성”.

김현수(2023), “신유형 디지털 서비스 이용자 보호를 위한 국내외 정책사례 비교 연구”, 방통융합정책연구 KCC-2023-12.

_____(2024), “생성형 AI와 경쟁 이슈”, 스마트 서비스 활성화를 위한 정책플랫폼 연구, 정책자료 24-03-02.

KISTEP(2025.4.9), “EU, ’AI 대륙 행동 계획’ 발표로 규제-혁신 균형 모색”.

매일일보(2025.6.17), “기회의 파도 탄 SK하이닉스, AI반도체 입지 강화 청신호”.

비즈와치(2025.8.1), “실적 훈풍 대기업SI, AI 가속 페달 밟는다”.

삼정KPMG(2024.5), “창작 영역에 뛰어든 생성형 AI 투자 현황과 활용 전망”.

소프트웨어정책연구소(2024), “유럽연합 인공지능법(AI Act)의 주요내용 및 시사점”, SW중심사회, 2024(10), 4-25.

오성탁(2020), “OECD 인공지능 권고안”, TTA 저널 187.

인공지능신문(2025.7.3), “SK텔레콤, 에이닷 엑스 4.0 오픈 소스 공개...추론 모델도 7월 중 공개”.

인공지능신문(2025.7.28), “일론 머스크, 삼성전자와 23조원 규모 테슬라 AI6 칩 계약 체결… 차세대 AI반도체 생산 본격화”.

지디넷코리아(2025.1.31), “삼성전자, HBM 공급량 2배 확대...’AI 반도체’에 사활”.

조선일보(2025.11.24), “AI가 세상에 나온다...석학·빅테크도 뛰어드는 피지컬AI 경쟁”

한국무역협회(2024.9.24), “AI로 뭉친 韓 기업, 원팀 전략으로 글로벌 공략”.

한국법제연구원(2024),『OECD AI 시스템 정의 개정의 주요 내용 및 시사점』(법제 Brief 24-6).

한국은행(2025.8.18), “AI의 빠른 확산과 생산성 효과: 가계조사를 바탕으로”, BOK 이슈노트, 제 2025-22호.

한국지능정보사회진흥원(2021), 『유네스코 AI 윤리 권고 주요 내용 및 시사점』, IT & Future Strategy 2021-10.

_____(2025.5), “일본 ‘AI 전략회의·AI 제도연구회 중간 보고서’ 주요 내용: ‘세계에서 가장 AI 개발·활용이 쉬운 나라’를 목표로 AI 제도 마련”, THE AI REPORT 2025-4.

한국경제(2025.1.7), “‘682만명 우르르’…‘챗GPT’ 한국인이 가장 많이 쓴 생성형AI”.

〔해외 문헌〕

Bloomberg(2024.3.8), “Generative AI races toward $1.3 trillion in revenue by 2032”, https://www.bloomberg.com/professional/insights/data/generative-ai-races-toward-1-3-trillion-in-revenue-by-2032/(접속일: 2025.12.3)

Coursera(2024.12.20), “What Is Artificial Intelligence? Definition, Uses, and Types” https://www.coursera.org/articles/what-is-artificial-intelligence(접속일: 2025.12.3)

Fortune business insights(2025.12.29), “생성적 AI 시장 규모, 점유율 및 모델별 산업 분석(생성적 적대 네트워크 또는 GAN 및 변환기 기반 모델), 산업별 애플리케이션별, 지역 예측(2024~2032년)”, https://www.fortunebusinessinsights.com/ko/generative-ai-market-107837/(접속일: 2025.12.29)

Gartner(2025.8.5), “Gartner Hype Cycle Identifies Top AI Innovations in 2025”, https://www.gartner.com/en/newsroom/press-releases/2025-08-05-gartner-hype-cycle-identifies-top-ai-innovations-in-2025?utm_source=chatgpt.com./(접속일: 2025.11.24)

Harvard Business Review(2024.12.13), “How Gen AI and Analytical AI Differ — and When to Use Each”, https://hbr.org/2024/12/how-gen-ai-and-analytical-ai-differ-and-when-to-use-each

Hartley, J., Jolevski, F., Melo, V., & Moore, B. (2024), The labor market effects of generative artificial intelligence. Available at SSRN.

IoT Analytics(2025.3), "The leading generative AI companies", https://iot-analytics.com/leading-generative-ai-companies/(접속일: 2025.8.4)

Markets and Markets(2024.9.12), "AI Agents Market worth $47.1 billion by 2030 - Exclusive Report by MarketsandMarkets", https://www.prnewswire.com/news-releases/ai-agents-market-worth-47-1-billion-by-2030---exclusive-report-by-marketsandmarkets-302246356.html/(접속일: 2025.11.24)

McKinsey & Company(2023.6), "The economic potential of generative AI".

OECD(2019), "Recommendation of the Council on Artificial Intelligence", OECD/LEGAL/0449.

Sherwood(2025.1.2), "Will capital spending on AI continue to boom?", https://sherwood.news/markets/will-capital-spending-on-ai-continue-to-boom/(접속일: 2025.11.24)

_____(2025.7.30), "The AI spending boom is eating the US economy", https://sherwood.news/markets/the-ai-spending-boom-is-eating-the-us-economy/(접속일: 2025.11.24)

UNCPGA(2024.9), "Governing AI for Humanity"

〔웹 자료·DB〕

네이버클라우드 홈페이지, https://www.navercloudcorp.com/ko/info/#Information(접속일: 2025.8.4)

ElectroIQ 홈페이지, https://electroiq.com/stats/generative-ai-statistics(접속일: 2025.7.6)

EU AI Code of Practice 홈페이지, https://digital-strategy.ec.europa.eu/en/policies/ai-code-practice(접속일: 2025.8.7)

Google Cloud, "인공지능(AI)이란 무엇인가요?", https://cloud.google.com/learn/what-is-artificial-intelligence?hl=ko

NASA, "What is Artificial Intelligence?", https://www.nasa.gov/what-is-artificial-intelligence/

OECD AI principle 홈페이지, https://oecd.ai/en/ai-principles(접속일: 2025.12.26)

Statista(2025.8.15), "Number of AI tool users worldwide from 2020 to 2031", https://www.statista.com/forecasts/1449844/ai-tool-users-worldwide(접속일: 2025.8.19)

UNESCO 홈페이지, https://www.unesco.org/en/articles/unesco-support-more-50-countries-

designing-ethical-ai-policy-year?hub=701(접속일: 2025.8.11)
White House 홈페이지, https://www.whitehouse.gov/(접속일: 2025.7.24.)

● 저 자 소 개 ●

김 현 수

- 고려대학교 법학 박사
- 현 정보통신정책연구원 선임연구위원

김 민 진

- 한국과학기술원 경영학 석사
- 현 정보통신정책연구원 부연구위원

이 선 희

- 성균관대학교 신문방송학 박사
- 현 정보통신정책연구원 부연구위원

방송통신융합정책연구 KMCC-2025-10

AI 이용자 보호 및 공정한 산업환경 조성 방안 연구

2025년 월 일 인쇄
2025년 월 일 발행

발행인 방송미디어통신위원회 위원장
발행처 방송미디어통신위원회
경기도 과천시 관문로 47
정부과천청사 2동
TEL: 02-2110-1323
Homepage: www.kmcc.go.kr
인 쇄 경성문화사


방송미디어통신위원회
Korea Media and Communications Commission