ICT 진흥 및 혁신기반 조성사업 RS-2025-02173110

정책연구 25-28-11

디지털 전환기 AGI 기술 경쟁력 확보를 위한 중장기 전략

Mid- to Long-Term Strategies for Securing AGI Technological Competitiveness in the Era of Digital Transformation

2025. 12

주관연구기관 : 정보통신정책연구원

연구기관 : (사)한국정보과학회

과학기술정보통신부 정보통신기획평가원

ICT 진흥 및 혁신기반 조성사업 RS-2025-02173110
정책연구 25-28-11

디지털 전환기 AGI 기술 경쟁력 확보를 위한 중장기 전략

(Mid- to Long-Term Strategies for Securing AGI Technological Competitiveness in the Era of Digital Transformation)

송길태 외

2025. 12

주관연구기관 : **정보통신정책연구원**
연구기관 : **(사)한국정보과학회**

과학기술정보통신부 정보통신기획평가원

이 보고서는 2025년도 과학기술정보통신부 방송통신발전기금 ICT 진흥 및 혁신기반 조성사업의 연구결과로서 보고서 내용은 연구자의 견해이며, 과학기술정보통신부의 공식입장과 다를 수 있습니다.

제 출 문

과학기술정보통신부 장관 귀하

본 보고서를 『**디지털 대전환 메가트렌드 연구: 디지털 전환기 AGI 기술 경쟁력 확보를 위한 중장기 전략**』의 연구결과보고서로 제출합니다.

2025년 12월

연 구 기 관 : (사)한국정보과학회
총괄책임자 : 송길태 교 수
참여연구원 : 조성배 교 수
최용석 교 수
김지희 교 수
신현정 교 수
이지형 교 수
최정훈 박사과정
김기범 박사과정
문은정 연 구 원
오다름 연 구 원

목 차

표 목 차

그 림 목 차

요 약 문

1. 제 목

디지털 전환기 AGI 기술 경쟁력 확보를 위한 중장기 전략

2. 연구 배경 및 목적

범용 인공지능(Artificial General Intelligence, AGI)은 인간 수준의 사고·학습·추론·창의성을 통합적으로 수행할 수 있는 지능체계를 의미한다. 기존의 협의적 인공지능(Artificial Narrow Intelligence, ANI)이 특정 과제에 최적화된 기능적 지능이었다면, AGI는 다양한 문제를 자율적으로 해결하고 새로운 지식을 생성하는 일반지능의 단계로 진화한다. 최근 전 세계적으로 대규모 언어모델(LLM), 멀티모달 학습, 시뮬레이션 기반 학습기술이 빠르게 고도화되면서 AGI의 실현 가능성이 가시화되고 있다. 그러나 이러한 기술적 진전에도 불구하고, AGI의 개념과 범위는 여전히 불명확하며, 기술 발전과 사회적 수용 사이의 간극이 존재한다.

대한민국은 인공지능 분야에서 일정 수준의 기술 경쟁력을 확보했지만, AGI와 같이 초대형·범용 지능 체계를 구현하기 위해 필요한 연산 인프라, 데이터 자원, 제도적 기반은 아직 제한적이다. 또한 AGI의 잠재적 사회적 영향력—노동 구조의 변화, 정책결정의 자동화, 신뢰성 및 안전성 확보 문제—를 고려할 때, 기술의 단순한 발전을 넘어 국가 차원의 전략적 대응이 필요하다. 이에 본 연구는 AGI를 기술적·사회적·정책적 관점에서 통합적으로 분석하고, 그 개념적 정의, 발전 경로, 촉진 및 제약 요인을 규명하여 국가 전략 수립의 근거를 제시하는 것을 목적으로 수행되었다. 연구는 AGI의 기술 구조와 산업 확산 가능성을 실증적으로 검토하고, 이를 기반으로 향후 정책·연구개발(R&D)·거버넌스 체계의 정렬 방향을 도출하였다.

3. 연구의 구성 및 범위

본 연구는 AGI의 개념적 정의에서부터 기술 요소, 발전 단계, 산업별 확산 전략, 정책 시사점까지의 전 과정을 포괄적으로 다루었다. 연구는 총 7개 장으로 구성되며, 1장은 연구의 필요성과 목적을 제시하고, 2장은 AGI의 개념적 정의 및 범위를 다루었다. 3장은 AGI 구현을 위한 핵심 기술 요소를 분석하였으며, 4장은 기술적·사회적 발전 단계를 단계별 시나리오로 제시하였다. 5장은 AGI 발전을 촉진하는 기술·산업·사회적 요인을 규명하였고, 6장은 발전을 저해하는 병목 요인을 분석하였다. 7장은 AGI의 산업별 적용과 확산 전략을 다루어, 기술의 실질적 파급효과를 평가하였다.

연구의 범위는 기술적 분석에 국한되지 않고, AGI가 사회·경제·정책 전반에 미칠 영향을 포함하는 다층적 구조로 설정되었다. 기술적 범위에서는 기억, 추론, 학습, 메타인지, 창의성, 추상화, 세계모델 등 인간 인지 구조를 모사하는 일곱 인지 기능을 중심으로 분석하였다. 산업적 범위에서는 반도체·제조, 바이오·의료, 금융·경제, 교육·법률, 해양·물류, 문화콘텐츠 등 여섯 개 주요 산업군을 대상으로 AGI의 도입 가능성과 확산 전략을 검토하였다. 정책적 범위에서는 기술 성숙도(TRL), 사회적 수용도, 정렬성(Alignment) 평가를 포괄하는 국가 차원의 관리 프레임 구축을 목표로 하였다. 연구의 방법론적 접근은 문헌분석, 기술동향 평가, 산업 적용 사례 검토, 전문가 자문 등 다단계 절차를 병행하였다. 이를 통해 AGI의 개념에서 산업·정책적 적용까지 연계되는 종합적 프레임을 구축하였다.

4. 연구 내용 및 결과

연구의 핵심 성과는 AGI를 “기술-산업-정책의 상호작용적 체계”로 재정의하고, 각 축별 구체적 발전 경로를 도식화한 데 있다. 우선 AGI의 기술적 구조 분석 결과, 인간의 인지 기능을 구현하기 위한 핵심 요소는 기억, 추론, 학습, 메타인지, 창의성, 추상화, 세계모델의 일곱 기능으로 정리되었다. 본 연구는 이들 기능을 실현하기 위한 접근을 발전형·전환형·도전형의 세 기술군으로 분류하였다. 발전형 기술은 LLM을 기반으로 한 효율성 및 안정성 향상을 목표로 하고, 전환형 기술은 멀티모달 통합 및 시뮬레이션 기반 학습을 통해

환경 적응능력을 강화하며, 도전형 기술은 새로운 인지 구조를 창출해 기존 AI의 한계를 넘어서는 방향으로 전개된다. 이러한 세 기술군의 균형적 발전이 AGI의 질적 전환을 촉진하는 핵심 요인으로 제시되었다.

AGI의 발전 단계를 다룬 분석에서는, AGI가 단일한 기술의 확장이 아니라 사회적 신뢰와 제도적 기반이 병행되는 공진화적 성장임을 확인하였다. 연구는 AGI 발전을 Emerging-Competent-Expert-Virtuoso-Superhuman의 다섯 단계로 구분하고, 각 단계별로 기술적 성숙도, 안전성 검증 수준, 사회적 수용도를 병렬적으로 평가하였다. 특히 Competent 단계 이후에는 기술의 불확실성이 급격히 확대되기 때문에, 기술적 검증 체계와 정렬성 평가 기준의 조기 구축이 필수적임을 강조하였다.

AGI의 발전을 촉진하는 요인으로는 고성능 컴퓨팅 자원 확보, 효율적 알고리즘 개발, 데이터 품질 향상, 멀티모달 융합, 자기지도학습의 확립이 기술적 측면에서 중요하게 제시되었다. 산업적 요인으로는 공공 조달, 민관 공동투자, 오픈소스 생태계 조성, 표준화 정책의 추진이, 사회적 요인으로는 윤리적 신뢰와 제도적 안정성 확보가 핵심으로 도출되었다. 반면 병목 요인은 컴퓨팅 인프라 부족, 전문 인력의 한계, 규제 불확실성, 데이터 편향, 사회적 신뢰 결핍 등으로 확인되었다. 이러한 분석은 AGI가 기술·산업·사회 인프라의 동시적 개선 없이는 실질적으로 진전하기 어렵다는 점을 보여준다.

산업별 확산 분석에서는 AGI의 파급효과가 전 산업 영역으로 확장될 수 있음을 구체적으로 제시하였다. 반도체 및 제조 산업에서는 공정 자동화와 예지 정비, 생산 최적화가 핵심 응용 영역으로, 생산 효율성과 품질 안정성을 동시에 높일 수 있다. 바이오·의료 산업에서는 신약개발, 유전자 분석, 진단 지원 등 연구개발 속도와 정확성을 향상시킬 수 있으며, 금융·경제 분야에서는 리스크 예측과 규제 자동화를 통해 산업 효율성을 높일 수 있다. 교육·법률 분야에서는 맞춤형 학습과 법률 자문 자동화가 가능해지고, 해양·물류에서는 운항 최적화와 공급망 예측이 실현 가능하다. 문화콘텐츠 산업에서는 창의적 생성 모델을 활용한 콘텐츠 생산의 혁신이 가능하다. 이러한 산업별 분석을 통해 본 연구는 AGI 확산이 단일 산업에 국한되지 않고, 기술 성숙도·데이터 인프라·규제 환경·사회적 수용성의 조합에 따라 상이한 속도로 진행된다는 점을 확인하였다. 따라서 산업별 특성을 반영한 단계적 확산 전략과 우선순위 설정 체계의 필요성이 제기되었다. AGI의 산업적 확산은 기술의 준비도(TRL)뿐 아니라 사회적 신뢰와 정책 정합성을 함께 고려해야 하며, 정부

와 민간이 이를 조정할 통합 프레임워크를 마련하는 것이 중요하다.

5. 기대효과

본 연구의 결과는 AGI를 기술혁신 중심의 연구대상에서 국가 전략기술의 범주로 격상시키는 개념적 전환을 제공한다. 첫째, AGI 발전의 기술·산업·사회 통합모델을 구축함으로써 향후 국가 R&D 투자와 정책 설계의 근거를 마련하였다. 특히 발전형-전환형-도전형 기술 구조와 단계별 발전 시나리오는 중장기 기술 로드맵 수립의 기초 자료로 활용될 수 있다. 둘째, 산업별 AGI 확산 매트릭스는 정부와 기업이 산업별 우선 적용 영역을 선정하고, 자원 배분의 효율성을 제고하는 정책 설계 도구로 기능한다. 셋째, AGI 발전의 촉진·병목 요인 분석은 기술 투자뿐 아니라 제도 개선과 사회적 합의 형성의 필요성을 입증함으로써, 균형 잡힌 정책 수립에 기여할 것이다.

또한 본 연구는 AGI를 단순히 인공지능 기술의 확장으로 보지 않고, 기술·산업·정책이 상호작용하는 국가 혁신체계로 재해석하였다. 이를 통해 대한민국이 향후 AGI 선도국가로 도약하기 위해 갖추어야 할 전략적 기반—컴퓨팅 인프라, 데이터 거버넌스, 인재양성 체계, 안전성 검증 제도—를 구체적으로 제시하였다. AGI의 발전은 기술적 성취만으로 달성되지 않으며, 사회적 신뢰와 정책적 정렬이 함께 이루어질 때 실질적인 경쟁우위로 전환될 수 있다. 따라서 본 연구의 결과는 국가 차원의 AGI 추진 기본계획, 기술표준화 정책, 산업별 R&D 로드맵 수립 등 후속 전략 수립에 직접적인 참고자료로 활용될 수 있을 것이다.

SUMMARY

1. Title

Mid- to Long-Term Strategies for Securing AGI Technological Competitiveness in the Era of Digital Transformation

2. Research Objectives

Artificial General Intelligence(AGI) refers to an intelligence system capable of performing human-level cognition, learning, reasoning, and creativity in an integrated manner. Unlike Artificial Narrow Intelligence(ANI), which is optimized for specific, task-oriented functions, AGI represents a higher stage of evolution—an intelligent system capable of autonomously solving diverse problems and generating new knowledge. With the rapid advancement of large language models(LLMs), multimodal learning, and simulation-based training technologies worldwide, the realization of AGI has become increasingly tangible. However, despite these technological developments, the conceptual boundaries and scope of AGI remain ambiguous, and a considerable gap persists between technological progress and social acceptance.

While the Republic of Korea has achieved a competitive level of expertise in artificial intelligence, the nation still lacks sufficient computational infrastructure, data resources, and institutional frameworks necessary to realize large-scale, general-purpose intelligence systems such as AGI. Furthermore, given AGI's potential societal impact—including labor market restructuring, policy decision automation, and challenges related to reliability and safety—there is a growing need for a strategic national response that extends beyond

mere technological advancement.

Accordingly, this study aims to analyze AGI from technological, societal, and policy perspectives, to define its conceptual boundaries, development trajectory, and enabling and constraining factors, and to present the empirical foundation for establishing a national AGI strategy. The study also examines the technological architecture and industrial diffusion potential of AGI and, based on this analysis, derives future directions for the alignment of policy, R&D investment, and governance systems.

3. Contents and Scope of the Research

This study comprehensively examines AGI—from its conceptual definition and technological components to its development stages, industrial diffusion strategies, and policy implications. The report consists of seven chapters. Chapter 1 introduces the research necessity and objectives. Chapter 2 defines the conceptual scope of AGI. Chapter 3 analyzes its core technological elements, while Chapter 4 presents stage-based scenarios for its technical and societal evolution. Chapter 5 identifies the key technological, industrial, and social drivers that promote AGI development, and Chapter 6 investigates the major bottlenecks that hinder progress. Finally, Chapter 7 assesses the industrial applications and diffusion strategies of AGI, focusing on its practical and economic implications.

The scope of this research extends beyond a purely technical analysis to encompass AGI's anticipated influence on social, economic, and policy systems. From the technical perspective, the analysis centers on seven cognitive functions that emulate human cognition—memory, reasoning, learning, metacognition, creativity, abstraction, and world modeling. From the industrial perspective, the research focuses on six key sectors—semiconductors and manufacturing, biotechnology and healthcare, finance and economics, education and law, maritime logistics, and cultural content—to examine AGI's potential

adoption pathways and diffusion strategies. From the policy perspective, the study aims to establish a national management framework that integrates technology readiness(TRL), social acceptance, and alignment evaluation. The methodology combines literature review, technology trend analysis, case studies of industrial applications, and expert consultation, thereby constructing an integrated framework that connects AGI's conceptual understanding with its industrial and policy implementation dimensions.

4. Research Results

The central achievement of this research lies in redefining AGI as an "interactive system among technology, industry, and policy" and outlining concrete development trajectories for each dimension.

The analysis of AGI's technical architecture identifies seven core elements essential to replicating human cognition: memory, reasoning, learning, metacognition, creativity, abstraction, and world modeling. The study classifies the technical approaches to realizing these functions into three categories—evolutionary, transitional, and pioneering.

The evolutionary approach focuses on enhancing efficiency and stability through optimization of existing LLM-based technologies. The transitional approach integrates multimodal learning and simulation-based training to strengthen contextual understanding and environmental adaptability. The pioneering approach seeks to design entirely new cognitive structures that transcend the inherent limitations of conventional AI. Balanced advancement among these three approaches is presented as the key driver of AGI's qualitative transformation.

The analysis of AGI's developmental stages confirms that AGI's evolution is not a linear extension of existing technologies but rather a process of co-evolution between technological innovation and societal readiness. The study categorizes AGI's evolution into five stages—Emerging, Competent, Expert, Virtuoso, and Superhuman—and evaluates each

stage in parallel across three dimensions: technological maturity, safety verification, and social acceptance. In particular, from the Competent stage onward, technological uncertainty expands rapidly, underscoring the need for early establishment of technical verification systems and alignment evaluation standards to ensure responsible development.

The study identifies several key drivers that accelerate AGI advancement. From a technological standpoint, securing high-performance computing infrastructure, developing efficient algorithms, improving data quality, advancing multimodal integration, and institutionalizing self-supervised learning are essential. Industrial drivers include public procurement initiatives, public-private co-investment, open-source ecosystem development, and the establishment of standardization policies. Social drivers center on ensuring ethical trust and institutional stability. Conversely, major bottlenecks are found in limited computing infrastructure, workforce shortages, regulatory uncertainty, data bias, and lack of public trust. The findings demonstrate that AGI cannot progress meaningfully without the concurrent advancement of technological, industrial, and social infrastructures.

The industrial diffusion analysis highlights AGI's potential to transform all major economic sectors. In the semiconductor and manufacturing industries, AGI enables process automation, predictive maintenance, and production optimization, improving both efficiency and quality control. In biotechnology and healthcare, it accelerates research and development through enhanced drug discovery, genomic analysis, and diagnostic support. In the financial and economic sectors, AGI enhances predictive modeling, risk management, and regulatory automation. In education and law, it enables personalized learning and automated legal consulting, while in maritime and logistics, it facilitates route optimization and real-time supply chain prediction. Finally, in the cultural content industry, AGI-driven creative generation technologies redefine production and localization paradigms. These analyses confirm that AGI's industrial diffusion will not proceed uniformly across sectors but will instead vary according to each industry's technological

readiness, data infrastructure, regulatory environment, and level of social acceptance. Consequently, the study underscores the need for a phased diffusion strategy and prioritization framework that reflect the characteristics of each industry. AGI's industrial expansion must consider not only technological readiness(TRL) but also social trust and policy alignment, requiring an integrated coordination framework between government and industry.

5. Expectations

The findings of this study provide a conceptual shift that elevates AGI from a subject of technological innovation to a domain of national strategic technology.

First, by establishing an integrated model linking AGI's technological, industrial, and social dimensions, the research offers an empirical foundation for guiding future national R&D investments and policy design. In particular, the tri-level technical structure—evolutionary, transitional, and pioneering—and the stage-based development scenario can serve as foundational references for formulating mid- to long-term technology roadmaps.

Second, the industrial diffusion framework enables government and private stakeholders to identify priority sectors for AGI deployment and to enhance resource allocation efficiency in technology policy.

Third, by analyzing AGI's enabling and constraining factors, the study contributes to balanced policymaking by evidencing the need for not only technological investment but also institutional reform and social consensus-building.

Furthermore, this research reinterprets AGI not merely as an extension of artificial intelligence but as a national innovation system in which technology, industry, and policy co-evolve. Through this perspective, the study identifies the strategic foundations necessary for Korea to become a leading AGI nation—computing infrastructure, data governance, talent cultivation, and safety verification systems. AGI development cannot

be achieved through technical advancement alone; it requires the concurrent maturation of social trust and policy alignment to translate technological potential into national competitiveness. Accordingly, the outcomes of this study are expected to directly inform the establishment of Korea's National AGI Initiative, the development of AGI-related technical standards, and the formulation of sector-specific R&D roadmaps, ultimately supporting the nation's transition toward sustainable and trustworthy AGI leadership.

CONTENTS

제1장 서 론

제1절 연구의 목적 및 필요성

ChatGPT, Gemini 등 이미지, 텍스트, 음성 등 다양한 형태(Modal) 데이터를 입력받아 인간이 원하는 결과물을 자유자재로 반환해주는 멀티모달(Multi-modal) 인공지능들이 보급되면서, 다음 단계 '인공 일반지능(Artificial General Intelligence, 이하 AGI)'에 대한 논의가 학계, 산업계, 그리고 세계 각국 정부 부처 등에서 활발하게 논의되기 시작했다.

AGI 의 대중적 정의는 '모든 영역에서 인간 수준의 범용적 사고, 추론, 그리고 자율적 학습 및 개선 능력을 갖춘 강(強) 인공지능' 시스템으로 정의되며 특정 영역 지식만을 가지고 수동적으로 작동하는 특화 인공지능(Aritificial Narrow Intelligence, 이하 ANI)와 대비되는 개념으로 인식된다. 하지만 '인공지능의 인간 수준 사고, 추론, 자율학습 및 자기 개선 능력'에 대한 엄밀한 과학적, 학술적 정의는 여전히 학자마다 시각이 다르며 논쟁의 영역이다. 이에 따라 AGI의 학술적 정의는 여전히 정립되지 않은 상태이다.

그럼에도 불구하고 최근 상용 언어모델(Large Language Model, 이하 LLM)들이 거의 모든 영역에서 '범용성에 가깝다고 여겨지는' 사고 및 추론 능력을 획득했고, 에이전틱 인공지능(Agentic AI) 시스템들이 주어진 과업(Task) 해결을 위해 자율적인 계획 수립, 행동 수정, 그리고 탁월한 자기 개선 능력을 획득하며 AGI의 대중적 정의에 성큼 다가선 모습을 보여주고 있다. 이에 따라 시민사회 및 관련 전문가 커뮤니티에서 AGI 개념, 능력, 그리고 가능성에 대한 관심이 날로 높아지고 있다. 본 보고서는 이 같은 뜨거운 관심에 대한 학술적 응답이자 안내서이며, AGI에 대한 다양한 학술적 관점을 집대성해 선도적으로 일관된 학술적 정의를 내리고자 한다.

AGI 개념 정립 및 기술 연구는 글로벌 기술 경쟁 구도 속에서의 기술 선점, AI 관련 기술 융합의 가속, 그리고 AGI로 인한 위험 관리와 윤리적 선도 세 가지 측면에서 분명한 필요성을 갖는다. AGI는 '범용 지능' 확보를 위한 기술 경쟁이자 디지털 주권, 안보, 산업 패

권이 복합적으로 결합된 글로벌 단위 국가 전략 기술 확보 생존 레이스이다. AGI 개념을 가장 먼저 정밀하게 정립하고, 기술 규범 및 표준을 선점하는 국가가 시장 규칙을 설계하는 패권국이 될 것이다.

AI 안전 규범을 선점하기 위한 글로벌 경쟁은 이미 활발하게 이루어지고 있다. EU는 2024년 8월 1일 EU AI Act를 발효하여 위험 기반 규제 체계를 세계 최초로 단계적 시행하였으며 글로벌 AI 시장의 규제 기준점을 선점하고자 했다(European Commission, 2024). 미국의 경우 2023년 10월 EO 14110으로 미 연방 차원의 안전 및 보안 프레임을 제시해 AI 규제 기준점을 선점하려 시도했다(The White House, 2023). 대한민국은 2024년 5월 영국과 함께 AI Seoul Summit을 개최해 '프런티어 AI 안전 Commitment'등 국제 공조 장치를 마련함으로써 글로벌 AI 규범 선점 경쟁에 대응하고 있다(UK Department for Science, Innovation and Technology, 2024). 하지만 AGI 기술에 대한 개념 정립 및 기술 개발은 아직 확고한 경쟁우위를 확보한 시장 내 선도적 플레이어가 없기 때문에, 시장 진입 장벽이 상대적으로 낮고 경쟁우위 확보가 수월한 것으로 평가할 수 있다. 따라서 대한민국은 AGI 개념을 선도적으로 정의하고, 해당 기술 개발에 자원을 집중하여 기술 규범 및 표준을 선점하는 것이 필요하다.

AI 관련 기술 융합의 가속 측면에서, 최근 LLM, 멀티모달, 로보틱스, 지식 그래프, 시뮬레이션, 도구 사용이 하나의 에이전틱 스택(Agentic stack) 으로 수렴하고 있다. 예를 들어, GPT-4o 는 멀티모달 LLM으로써 텍스트, 음성, 영상의 실시간 통합 추론을 수행하며 멀티모달 기술의 경계를 실질적으로 낮추었다(OpenAI, 2024). RT-2 등 비전-언어-행동 모델은 웹 및 로봇 데이터를 결합해 AI의 지식이 행동으로 전이 가능함을 입증하며 로보틱스와 LLM을 결합하였다(Google DeepMind, 2023). NVIDIA는 최근 Blackwell 등의 차세대 가속기와 AI 팩토리 아키텍처를 출시하며 대규모 멀티모달 및 에이전틱 스택 컴퓨팅 집적화를 가속화하고 있다(NVIDIA Corporation, 2024). 이 같은 최신 흐름은 AI의 실행력이 텍스트 자동화를 넘어 현실(물리) 세계로 점차 확대되고 있음을 의미한다. AI가 사용할 수 있는 자원이 풍부해지고 고도화되는 현실 속에서, AGI 기술 개발은 물리 세계에 영향력을 행사하는 인터페이스를 직접 통제하는 지능(Intelligence)의 고도화를 의미한다.

AGI를 기반으로 보다 고도화된 지능형 에이전트(Intelligent agent)는 다양한 산업 분야에 적용되어 생산성을 획기적으로 증대할 수 있다. 제조 현장의 '라인 변경, 신규 품목 투입

(Changeover) 자동화'를 그 예시로 들 수 있다. 지능형 에이전트가 기업 내 PLM/MES, 정비 이력, SOP, BOM 규정 등을 지식그래프로 통합해 토크 규격, 작업 순서, 안전 기준 등을 근거와 함께 조회한다(Hedberg, T.D., et al., 2016). 이후 멀티모달 LLM이 이를 바탕으로 툴 교체 목록, 로봇 동작 시퀀스와 같은 공정 전환 계획을 자동 설계한다. 디지털 트윈에서 시나리오를 사전 검증하고, 안전성 및 효과성이 승인되면 로보틱스/PLC에 배포되어 작업 전 주기를 수행한다(Brohan, A., et al., 2023). 현장 작업자의 경우 코파일럿을 통해 단계별 지시, 리스크 알림 등을 확인 받는다. 결과적으로 작업 준비와 승인까지 걸리는 시간을 획기적으로 단축해 Changeover time 단축, 초기 불량 및 재작업 감소, 예지 정비 및 부품 자동 제안 등을 통해 공정 가동률을 상승시킬 수 있다(McKinsey & Company, 2024). 에이전트의 모든 의사결정 및 행동은 근거 인용 및 감사 로그로 기록되어 규정 준수 및 확산 재사용성을 확보할 수 있다.

AGI로 인한 위험 관리와 윤리적 선도 측면에서, AGI 기술 연구는 AGI의 산업 전 분야 통합 및 응용에서 안전성, 보안, 책임성에 대한 선제적 AI 거버넌스 확립을 통한 시장 내 경쟁우위 구축에 있어 중요하다. 글로벌 시장에서, 현재 많은 국가들이 선제적 AI 거버넌스 확립을 위해 경쟁하고 있다. 예컨대 미국은 2023년 NIST AI RMF 1.0을 발표해 AI의 산업 통합으로 발생하는 위험을 식별하고 측정하며, 이를 완화하는 실무 프레임을 제시했다(National Institute of Standards and Technology, 2023). 또한 영국과 함께 AI Safety Institute를 설립해 인공지능 모델을 평가하고 유사시에 대비하는 레드팀 구축을 공공 주도로 추진하고 있으며, 민간 기업과의 안전 연구 협력도 확대하고 있다(UK Department for Science, Innovation and Technology, 2023). AGI의 경우 고도화된 AI 시스템 이므로, 기존에 구축된 AI 거버넌스 체계로는 안전성, 보안, 책임성 측면에서 대응이 다소 부족할 수 있다. 선제적이고 공격적인 AGI 개념 정립 및 기술 개발을 통해 AGI 시스템에 걸맞는 거버넌스를 타국보다 선제적으로 구축함으로써 시장 내 표준을 최초로 정립하고 선발주자로 자리매김하는 것이 필요한 시점이다.

한편 AGI 개념 정립 및 기술 연구는 국가 전략 기술 확보, 경제 및 산업 혁신, 사회 문제 해결, 그리고 과학적 탐구 확장 네 가지 측면의 분명한 목적을 갖는다. 국가 전략 기술 확보의 측면에서, AGI는 범용 추론, 학습, 도구 사용 능력을 결합해 사회 모든 분야 문제 해결에 적용될 수 있기 때문에 기존 ANI 대비 파급력과 대체효과가 매우 큰 기술이다. 더불

어, AGI 기술 구현을 위해서는 반도체 제조-컴퓨팅 인프라-데이터 확보-알고리즘 구축-시스템 응용 및 보급에 이르는 사슬적(Chain) 생태계가 동시적, 유기적으로 결합 되어야 하므로, 선도권을 확보한 소수 국가가 장기간 초과이익을 누리는 전략 기술이다. 장차 AGI 자립 수준에 따라 각 국가의 경쟁우위, 디지털 전환 속도, 범위, 안전성, 사회 인프라 안정성 등이 좌우될 것이기 때문에 AGI를 국가 전략 기술로서 선도적으로 확보하는 것이 중요하다.

국가 전략 기술로서 AGI 기술은 주권형(Sovereign) AGI 아키텍쳐, 국가형 컴퓨팅 클러스터, 핵심 원천 IP, 그리고 표준 및 규제 샌드박스 네 가지로 좀 더 세분화 할 수 있다. 주권형 AGI 아키텍쳐는 한국어, 다국어, 그리고 도메인 융합을 전제로 한 범용 인공지능 알고리즘과 로컬 및 에지 환경에서도 사용할 수 있는 경량 파생 모델 인공지능 알고리즘 체계를 동시 구축하는 것을 의미한다. 국가형 컴퓨팅 클러스터는 GPU 및 NPU와 같은 AI 알고리즘 계산 가속기, 고대역 통신 체계, 그리고 탄소 및 전력 최적화를 완료한 계산 자원을 의미한다. 핵심 원천 IP는 메모리, 장기학습, 도구(Tool) 사용, 그리고 안전성(거버넌스 확보 및 레드팀 운용을 통한 사이버 보안 확보)을 통합적으로 확보한 범용 인공지능 알고리즘 핵심 모듈의 국내 IP 확보를 의미한다. 마지막으로 표준 및 규제 샌드박스는 개발 시스템의 안전성 평가, 데이터 책임성, 그리고 기록 및 감사(Logging & Audit) 프레임을 마련하는 과정을 의미한다.

경제 및 산업 혁신의 관점에서, AGI는 모든 영역에서의 복합 문제 프로세스를 스스로 분해, 계획, 도구 호출하는 능력을 통해 업무 자동화의 범위를 RPA 수준에서 지식 및 창의 업무로까지 확장할 수 있다. 이를 통해 AGI는 제조-의료-바이오-금융-교육-컨텐츠-공공서비스 전반에서 생산성, 품질, 그리고 안전성을 동시 개선할 수 있다.

경제 및 산업 혁신 기술로서 AGI는 산업별 AGI 레퍼런스 스택, 에이전틱(Agentic) 업무 코파일럿(Copilot), 현장 데이터 연계, 그리고 안전 및 보안 내재화 네 가지로 세분화 할 수 있다. 산업별 AGI 레퍼런스 스택은 제조(공정 및 품질관리), 바이오 연구(R&D), 금융(리스크 관리 및 컴플라이언스), 교육(개인화, 맞춤화 학습 플래닝) 등 도메인 특화 파이프라인 기술을 의미한다. 에이전틱 업무 코파일럿은 계획-실행-평가 루프로 이루어진 에이전틱 인공지능 시스템으로서, 기업 내부 시스템(ERP/PLM/LIMS) 등 도구 호출을 통합한 패키지를 말한다. 현장 데이터 연계는 OT/IoT/로봇, 영상, 텍스트, 음성 등 멀티모달 현장 데이터

취득 후 전처리하고 관리하며 AGI 시스템에 올바르게 학습시키는 데이터 거버넌스 전 과정을 일컫는다. 안전 및 보안 내재화는 AGI 시스템에 입력하기 위한 안전한 최적 프롬프트를 설계하고 주입하며, AGI 시스템에 대한 적대적 공격으로 인한 데이터 유출 방지책을 수립하고, 무엇보다 AGI 시스템의 의사결정 오류 대응 가드레일을 수립하는 절차를 의미한다.

사회 문제 해결 관점에서, AGI는 기후, 에너지, 보건, 고령화, 재난, 그리고 교통 등 복잡계 문제 해결에 대해 데이터, 모형, 정책의 상호작용을 통합적으로 다루는 범용 추론을 제공함으로써 시나리오 생성, What-if 기반 정책 실험, 다목표(Multi-purpose) 최적화 지원 등으로 정책 설계의 품질과 속도를 높일 수 있다.

사회 문제 해결 도구로서 AGI는 복잡계 모델을 결합하는 복합 시스템, 정책 시뮬레이터, 데이터 거버넌스, 그리고 AGI 윤리 및 책임 체계로 세분화 할 수 있다. 복잡계 문제 해결을 위해 AGI는 물리 시뮬레이션, 경제 모형, 지식 그래프 등을 결합한 하나의 복합 시스템으로 기능해야 한다. 또한 AGI는 정책 시뮬레이터로서 기능할 수 있으며 보건, 환경, 교통, 재난 분야의 정책 대안을 자동 생성하고 평가하며, 직접 대중들에게 설명을 제공할 수 있다. 데이터 거버넌스의 측면에서, AGI 학습을 위해 개인정보 및 공공데이터를 안전하게 결합하고, 지역 및 부처 간 데이터 연계 허브를 구축하는 것이 필요하다. 무엇보다 AGI에 대한 고위험 응용 가이드라인을 설정하고, 설명가능성 및 감사(Auditing) 가능성을 확보하는 것이 중요하다.

과학적 탐구 확장의 관점에서, AGI는 문헌, 데이터 수집, 시뮬레이션 및 실험 전 과정을 유기적으로 연결해 가설 생성-설계-분석-해석의 연구 전 주기를 가속한다. 특히 AGI는 바이오, 재료, 우주, 뇌과학 등 고난도 과학 연구 분야에서 탐색 비용을 극적으로 낮추고 실패율을 줄일 수 있다.

과학적 탐구 확장 도구로서 AGI는 과학 코파일럿, 멀티모달 연구 스택, 자율실험(Closed-loop), 그리고 개방형 리소스 네 가지 항목으로 세분화 할 수 있다. AGI는 과학 코파일럿으로서 기능할 수 있으며, 해당 시스템은 문헌, 코드, 실험노트를 통합하고 해당 지식을 갖춘 뒤 최적 가설을 자동으로 생성하고, 실험 계획을 자율 수립하고, 결과를 해석할 수 있다. 또한 AGI는 오믹스/영상/시계열/그래프 데이터 등을 동시에 처리해 통합적인 정제된 정보를 생산할 수 있으며, 지식 그래프를 AGI에 결합해 신뢰가능한 추론 체인을 구

축할 수 있다. AGI는 로봇 및 기타 장비 제어와 연동되어 실험 최적화 루프를 설정하고 자율 실험을 반복하여 최적의 결과 값을 찾아낼 수 있다. 이 같은 AGI 시스템은 연산 자원 및 환경이 변하더라도 동일하며 신뢰가능한 결과값을 적기에 제공할 수 있어야 하는데, 이를 위해 AGI의 사전학습 가중치, 훈련 데이터셋 등을 오픈소스로 공개함으로써 재현성을 확보하는 것이 필요하다.

상기 서술된 AGI 기술 개념 정립 및 R&D의 필요성 및 목적을 다음과 같이 종합 요약할 수 있다. 첫째, AGI는 국가의 미래를 결정짓는 전략기술이자 디지털 주권의 핵심 인프라다. AGI는 특정 분야의 개별 과제 자동화를 넘어서, 범용 추론, 학습, 도구 사용, 자기 개선을 통합한 차세대 지능으로 기능할 것이며, 이는 시장을 넘어 인간 사회 전반의 표준, 규범, 질서를 파괴적(Destructive)으로 재편할 것이다. 한국은 주권형(Soverign) AGI 아키텍쳐와 평가 체계를 글로벌 경쟁 속에서 가장 먼저 선점해 기술과 거버넌스 양 축 모두에 대한 수출 기반을 마련해야 한다. 궁극적으로는 공동체 내의 집단적 디지털 지능에 대한 대외 의존도를 낮추고, 안전한 디지털 전환의 기준을 우리 손으로 정립하는 것이 필요하다. 둘째, AGI는 산업 전반의 생산성, 품질, 안전성 동시 개선을 견인하는 실용 엔진이자 핵심(Core) 기술이다. 예컨대 LLM, 로보틱스, 지식그래프, 그리고 디지털 트윈 기술을 결합한 AGI 기반의 지능형 에이전트는 계획하고 실행하며 그 결과를 평가하는 제조 공정 전 주기를 자동화할 것이다. 그리고 이는 Changeover time 단축, First-Pass Yield 개선, 그리고 OEE 상승으로 직결된다. 셋째, '안전이 곧 경쟁력'인 생태계를 선도적으로 구축해야 한다. AGI의 산업 전 영역 통합에 있어 안전 가드레일-감사기능 로그-레드팀 운용-상호운용 평가에 이르는 AGI 거버넌스를 선도적으로 구축함으로써 업계 안전 기준을 수립해야 한다.

한편, AGI 기술 연구 및 개발은 제조업 생산성 강화, 인구 구조 변화 대응, 글로벌 기술 의존 탈피, 그리고 학계와 산업계 간 연계 강화 네 가지 측면에서 우리나라에 매우 필요하다. 먼저 제조업 생산성 강화 측면에서, AGI 기술은 대한민국의 반도체, 바이오, 의료, 로보틱스 산업에 긴밀하게 결합 되어 생산성을 획기적으로 증대할 수 있다. 대한민국은 메모리 및 HBM 반도체 분야에서 세계적 경쟁 우위를 보유하고 있으며 전세계 AI 인프라 구축의 병목을 좌우하는 핵심 반도체 공급국이다. 또한 네이버의 HyperCLOVA X와 같은 국산 대규모 언어모델도 한국어 및 다국어 성능을 축적해 주권형(Soverign) AI의 토대를 갖추었다. 대한민국의 이러한 산업 기반은 AGI기술 연구 및 개발에 있어 최적의 환경이라

할 수 있으며, 이를 기반으로 AGI의 제조업 통합을 촉진할 수 있다. 예컨대 주권형 AGI 아키텍쳐와 도메인 특화 에이전트를 동시 구축하거나, 디지털 트윈 및 시뮬레이션 결과가 현장 로봇/PLC 로 이어지는 Closed-Loop(계획-검증-배포)를 표준화 할 수 있을 것이다. 이를 기반으로 각 제조업에서 Changeover time 단축, FPY 개선, 그리고 OEE 상승 등의 효과를 거둘 수 있다.

한편 인구 구조 변화 대응의 측면에서, AGI 기술은 초고령 사회에 진입하고 있는 대한민국의 생산성 향상에 매우 중요한 핵심 엔진이 될 것이다. 현재 한국의 합계 출산율은 2023년 기준 0.72로 세계 최저 수준이었고, 2024년에도 0.75로 세계에서 가장 낮은 편에 속했다(OECD, 2024). 또한 2025년에는 전체 인구 대비 65세 이상 비중이 20%를 넘겨 '초고령 사회'에 진입했다. 결과로, 한국사회의 노동공급 축소와 복지지출 증가는 불가피하며, 생산성 중심의 성장모델로의 전환이 필수이다. AGI와 산업의 통합이 이 같은 한국의 인구 절벽을 극복하고 생산성을 다시 끌어올릴 수 있다. 예컨대 산업 현장에 AGI 에이전트를 도입해 제조, 물류, 에너지, 공공 서비스 자동화 범위를 넓히고, 더 나아가 지식 및 창의 업무까지 확대함으로써 부족한 노동력을 대체할 수 있을 것이다. 또한 어르신들을 위한 의료 및 돌봄 AGI 에이전트를 도입해 병원 의무기록을 요약하거나 어르신들이 알기 쉽게 설명해주고, 복지 자원을 더욱 효과적으로 배분할 수 있다.

글로벌 기술 의존 탈피의 측면에서, AGI 기술 개발은 주권형 AGI 구축 및 국제 규범 선도 효과를 가져올 것이다. 이미 대한민국 정부는 AI G3 도약을 국가 목표로 설정한 바 있으며, 공격적인 국가 AI전략 수립, 컴퓨팅 인프라 확충, 안전성 프레임웍(Framework) 구축을 병행하고 있다. 이는 기술 및 표준을 동시 선점하려는 전략으로, 보다 고도화된 AI 시스템인 AGI 기술개발에도 같은 전략이 적용될 수 있다. 특히 미국과 중국에서 기술적 우위를 선점하여 이미 기술 외부 의존도가 높은 LLM 시스템과는 달리, AGI 기술은 아직 시장 내 선두주자가 없는 상황이다. 선제적이고 공격적인 AGI 기술 R&D 투자는 대한민국이 진정한 의미에서의 AI 기술 글로벌 선두주자로 발돋움할 수 있는 전략이다.

학계-산업계 연계 강화의 측면에서 AGI 기술 개발은 기초 기술 개발-응용-사업화의 선순환 구조를 강화할 수 있다. 대한민국 정부는 현재 AI G3 청사진 제시와 함께 국가 AI 전략위원회 출범, AI 연구허브 공모 등 AI 거버넌스 및 투자 기반을 공격적으로 확대하고 있다. AGI 기술 개발은 성공할 경우 투자금액 대비 생산 부가가치가 압도적으로 큰 기술로

산업계의 관심이 높은 사안이며, 인공지능 공학계에서도 AGI 개발의 높은 학술적 가치로 해당 기술에 대한 관심이 나날이 높아지고 있다. 정부 주도로 구축한 AI 거버넌스 및 투자 기반 위에서, 학계와 산업계의 집약적 협력 촉진이 기대되는 상황이며, 이 같은 선순환이 학계의 재현성 높은 연구 결과가 빠른 산업화로 이어지도록 할 수 있다.

상기 서술된 한국적 맥락에서의 AGI 개념 정립 및 연구의 필요성을 다음과 같이 정리할 수 있다. 첫째, 한국의 주력 산업인 제조업(반도체, 바이오, 로보틱스 등)과 AGI를 결합해 공급망-인프라-인공지능-응용이 맞물리는 차세대 성장엔진을 구축할 수 있다. 해당 엔진은 초저출산 및 초고령 사회에 접어든 대한민국의 노동 생산성 악화 문제를 효과적으로 해결할 수 있을 것이다. 둘째, 초고령 및 저출산 구조에 대응하는 의료 및 돌봄 지능형 에이전트를 도입해 복지 서비스의 생산성 및 효과를 끌어올리고, 공공 의료 및 돌봄 서비스의 질을 획기적으로 개선할 수 있다. 셋째, 국가 AI거버넌스, 글로벌 AI규범 선도, 산학 연계를 삼각축으로, 안전성과 상호운용 표준을 선도하여 '안전이 곧 경쟁력'인 AGI 시장을 설계하고 시장 내 선도적 플레이어가 될 수 있다.

AGI 대응 전략은 이제 단순한 기술 트렌드가 아니라, 국가 미래가 달린 생존 전략이다. 범용 추론, 학습, 도구 사용, 자기 개선 능력을 함께 갖춘 AGI는 산업 구조, 노동시장, 국가 안보, 사회 및 시장 규범을 동시에 재편할 것이다. 선점국은 기술과 표준을 함께 수출하며 장기간에 걸친 초과이익을 확보하고, 후발국은 데이터, 컴퓨팅 인프라, 규제 준수 비용에서 장기간에 걸친 구조적 열위를 감수해야 한다. 한국은 반도체, 자동차, 의료, 로보틱스 등 다방면에 걸친 제조업 경쟁력을 지렛대로 삼아, 주권형(Soverign) AGI 역량을 확보하고, 더 나아가 국제 안전 및 상호 운용 표준 형성에 적극 참여해야 한다. 따라서 현 시점에서 AGI에 대한 장기적, 체계적 연구 전략을 수립하는 것은 매우 시의적절하며 필요한 일이다. 본 보고서는 다음의 4대 기술 축과 3대 기반을 국가 전략으로 정립할 것을 제안한다.

4대 기술 축

1. 주권형 AGI 코어(한국어, 다국어, 멀티모달, 도구 사용, 장기기억) 기술 개발
2. 산업 별 AGI 코파일럿/에이전트(제조, 물류, 의료, 금융, 공공) 연구
3. AGI 기반 정책 시뮬레이터(기후, 보건, 재난 등 복잡계 의사결정) 연구
4. 과학 연구 사이클을 가속할 수 있는 코파일럿/자율실험 연구

3대 기반

1. AGI 개발 및 운용 위한 국가형 컴퓨팅, 데이터 신뢰 인프라(보안 구역·감사/로깅·권한화·그린 컴퓨팅) 구축
2. AGI 안전성·거버넌스(고위험 가드레일, 레드팀, 사고보고·사후학습, 상호운용 평가) 구축
3. AGI 연구 개발을 위한 인재, 생태계(학제 간 트랙, 산학 공동랩, 책임 있는 오픈소스 등) 구축

추진 원칙은 안전 우선, 개방 및 상호 운용, 효율 및 지속 가능성, 사람 중심, 그리고 성과 계량화 다섯 가지로 정립할 수 있다. 먼저 안전 우선의 경우, '안전이 곧 경쟁력' 이 되도록 고위험 도메인 가드레일과 감사 가능성을 내재화 한다. 개방 및 상호 운용을 위해 데이터, 모델, API, DB 스키마 표준화로 재사용성을 극대화 한다. 효율 및 지속 가능성을 위해 효율 KPI를 예산 및 프로젝트 결과 평가에 연동한다. 또한 휴먼 인더 루프와 책임 배분을 사전에 명확히 함으로써, 현장 수용성을 높여 '사람 중심'의 AGI 개발을 추진한다. 마지막으로 모든 산업 통합 AGI 시스템에 FPY, OEE 등의 정량 KPI를 부착해 확산 기준을 정립한다.

AGI 개념 정립 및 기술 개발의 기대효과는 명확하다. 산업 측면에서 AGI는 생산성, 품질, 안전성을 동시에, 그리고 획기적으로 개선할 수 있다. 사회 측면에서 AGI는 고령화, 지역 격차 등 구조 문제에 데이터 및 모형 기반으로 효과적 해법을 제시할 수 있다. 학계-산업계의 측면에서 AGI 연구 개발은 재현성 높은 지식을 생산하고, 이를 빠르게 사업화 하는 선순환을 이룰 수 있다.

제2장 AGI의 개념적 정의와 범위 설정

제1절 AGI의 개념 관련 자료 분석 및 정리

인공지능의 발전과 함께 AGI(Artificial General Intelligence, 범용 인공지능)를 정의하려는 시도는 연구 동향에 따라 지속적으로 변화해 왔다. 초기 연구에서는 AGI를 단일 단계로 규정하는 경향이 두드러졌으나, 최근 연구들은 이를 여러 단계로 세분화하고, 다차원적 속성을 포함하는 방향으로 개념을 확장하고 있다.

1. 단일 단계의 AGI 정의

가장 대표적인 초기 시도는 1950년 앨런 튜링(Alan Turing)이 제안한 튜링 테스트 이다. 튜링은 "기계가 생각할 수 있는가?"라는 질문에 답하기 위해, 기계가 인간처럼 대화할 수 있는지를 판별하는 모방 게임을 제안했다. 이 접근은 역사적으로 큰 의의를 지니고, 구현이 단순하다는 장점이 있다. 그러나 기계의 지능 그 자체를 측정하기보다 사람을 속일 수 있는지 여부에 초점이 맞춰져 있다. 특히 대규모 언어모델(Large Language Model, LLM) 중 일부는 일부 태스크에서 이 테스트를 통과할 수 있으므로, 튜링 테스트만으로 AGI를 정의하거나 벤치마킹하기에는 한계가 있다.

이후 Strong AI 개념은 Searle(1980)에 의해 제시되었다. 그는 적절히 프로그래밍된 컴퓨터가 실제로 '마음'을 지니고, 이해와 인지 상태를 보유할 수 있다고 주장하였다. 이는 AGI 논의를 의식과 주관적 경험의 차원까지 확장했다는 점에서 의의가 있다. 그러나'마음'의 존재 여부를 과학적으로 판별할 합의된 방법이 없다는 점에서 실용성이 제한적이다.

인간 뇌 유사성을 기반으로 한 정의는 Mark Gubrud(1997)에 의해 제안되었다. 그는 인간 뇌의 복잡성과 속도를 능가하거나 이에 필적하며, 일반 지식을 습득하고 추론할 수 있는 인공지능을 AGI로 정의했다. 이 정의는 직관적으로 이해하기 쉽지만, 현대 AI의 핵심 구조인 트랜스포머 아키텍처(Vaswani et al., 2023)는 인간 뇌와 유사한 방식이 아니더라도

높은 성능을 발휘한다. 따라서 뇌 유사성을 AGI의 필수 조건으로 간주할 필요성은 점차 약화되고 있다.

Goertzel(2014)은 AGI를 "인간이 일반적으로 수행할 수 있는 인지 과제를 수행할 수 있는 기계"로 정의했다. 이 정의는 로봇과 같은 물리적 구현 없이도 범용 인지 능력에 집중할 수 있다는 장점이 있다. 그러나 어떤 과제를 기준으로 할지, 그리고 인간의 범위를 어떻게 설정할지에 대한 명확한 합의가 없어, 여전히 개념적 모호성이 남아있다.

Shanahan(2015)은 AGI를 특정 과제에 특화되지 않고 폭넓은 범위의 과제를 학습·수행할 수 있는 지능으로 정의하며, 메타인지적 학습 능력을 필수 조건으로 제시하였다. 이는 단순히 주어진 능력을 발휘하는 것을 넘어, 새로운 환경에서 적응하고 지식을 습득하는 유연성을 AGI의 핵심 요소로 강조한다.

이에 비해 OpenAI는 AGI를 "대부분의 경제적으로 가치 있는 작업에서 인간을 능가하는 고도로 자율적인 시스템"으로 정의하였다.[1] 이 접근은 성능을 정량적으로 평가할 수 있는 지표를 제공한다는 장점이 있다. 그러나 창의성이나 감성지능처럼 경제적 가치로 환산하기 어려운 영역은 포괄하지 못한다.

Marcus(2022)는 AGI의 핵심 요소를 유연성과 범용성으로 보고, 영화/소설 이해, 임의의 주방에서 요리하기, 대규모 프로그래밍, 자연어 수학 증명 변환 등 구체적인 벤치마크 과제를 제시하였다. 이러한 접근은 AGI의 역량을 평가하기 위한 실질적 예시를 제공한다는 장점이 있다. 그러나 일부 과제를 통과했다고 해서 해당 시스템을 곧바로 AGI로 인정할 수 있는지에 대해서는 여전히 논쟁이 지속되고 있다.

Artificial Capable Intelligence(ACI)는 개방된 환경에서 복잡하고 다단계 작업을 수행할 수 있는 충분한 성능과 범용성을 갖춘 AI 시스템을 의미한다(Mustafa Suleyman and Michael Bhaskar, 2023). 현대판 튜링 테스트로 불리는 ACI는 AI에게 10만 달러를 주고 이를 수개월 내에 100만 달러로 증대시키는 경제 기반 평가 접근법이다. 이러한 접근법은 실세계 과제를 중시하고 생태적 타당성을 확보하려는 장점이 있지만, 재정적 이익만을 목표로 함으로써 정렬 위험을 초래할 수 있으며, 범위가 제한적이라는 한계를 가진다.

마지막으로 최신 LLM 발전에 따라 일부 연구자들은 이미 LLM이 AGI에 해당한다고 주

1) https://openai.com/charter/

장한다(Agüera y Arcas & Norvig, 2023). AGI의 핵심 특성이 범용성이라고 강조하며, 최신 LLM이 다양한 주제를 논의하고, 여러 작업을 수행하며, 다중 모달 입력과 출력을 처리하고, 다국어 환경에서 작동하며, zero-shot 또는 few-shot 학습을 통해 새로운 작업에 적응할 수 있다고 설명한다. 그러나 범용성만으로는 충분하지 않으며, 성능 평가가 반드시 병행 되어야 한다. 예를 들어 LLM이 코드를 작성하거나 수학 문제를 풀 수 있더라도, 그 결과가 신뢰할 만하지 않다면 아직 충분히 성숙한 범용성이라 보기 어렵다.

2. 단계별 AGI 정의

최근 연구에서는 AGI의 발전 경로를 체계적으로 설정하기 위해, 단일 종착점이 아닌 AGI 정의를 여러 단계로 구분하고 각 단계의 특징과 필요 조건을 구체적으로 규정하는 접근이 등장하고 있다(Morris et al., 2024.; Feng et al., 2024).

가. 6단계 AGI

〈표 2－1〉 성능과 범용성에 기반해 단계별로 정의한 AGI

성능(행) × 범용성(열)	좁은 범위(Narrow)	범용(General)
레벨 0: No AI	Narrow Non-AI AI계산기 소프트웨어; 컴파일러	General Non-AI 인간 개입 기반 컴퓨팅 (예: 아마존 Mechanical Turk)
레벨 1: Emerging (비숙련 인간과 비슷하거나 약간 우수)	Emerging Narrow AI 단순 규칙 기반 시스템(예: SHRDLU)	Emerging AGI ChatGPT, Bard, Llama 2, Gemini
레벨 2: Competent (숙련 성인의 하위 50% 이상)	Competent Narrow AI AI유해 콘텐츠 탐지기(Jigsaw), 스마트 스피커(Siri, Alexa), VQA 시스템(PaLI), Watson, 일부 과제(SOTA LLMs)	Competent AGI 아직 달성되지 않음
레벨 3: Expert (숙련 성인의 상위 10% 이상)	Expert Narrow AI 맞춤법·문법 검사기(Grammarly), 생성형 이미지 모델 (Imagen, DALLE 2)	Expert AGI 아직 달성되지 않음

성능(행) × 범용성(열)	좁은 범위(Narrow)	범용(General)
레벨 4: Virtuoso (숙련 성인의 상위 1% 이상)	Virtuoso Narrow AI Deep Blue, AlphaGo	Virtuoso AGI 아직 달성되지 않음
레벨 5: Superhuman (모든 인간 능력 초월)	Superhuman Narrow AI AlphaFold, AlphaZero, StockFish	Artificial Superintelligence 아직 달성되지 않음

최근 인공지능 연구에서는 AGI를 단일한 완성 지점으로 규정하기 보다, 기술 발전의 각 단계의 특성을 명확히 구분하는 단계별 정의가 중요하게 논의되고 있다. Morris et al.(2024)은 이러한 필요성에 따라 성과와 범용성이라는 두 가지 축을 중심으로 AGI를 6단계로 구분하는 분류 체계를 제안하였다. 이 접근은 자율주행 분야에서 Society of Automotive Engineers(SAE)[2] 자동화 단계(Levels of Driving Automation)가 정책 논의와 기술 진척 상황의 명확한 전달을 가능하게 했던 것과 유사하게, AGI 발전 경로를 사회·기술적으로 이해하고 관리하기 위한 표준 프레임워크 역할을 한다.

이 프레임워크는 각 단계별로 달성해야 할 명확한 성능 지표와 평가 기준을 설정하는 동시에, 해당 수준에서 새롭게 등장하는 위험 요소와 인간-인공지능 상호작용 방식의 변화를 함께 고려한다. 이를 통해 여러 기존 AGI 정의가 단일 틀 안에서 공존할 수 있게 된다. 예컨대 Agüera y Arcas & Norvig(2023)이 제시한 정의는 본 체계에서 "Emerging AGI"에 해당하고, OpenAI가 제시한 '노동 대체 가능성'을 기준으로 한 정의는 "Virtuoso AGI"에 가깝게 위치한다. 또한, Goertzel(2014), Shanahan(2015) 등이 제시한 다수의 기존 AGI 개념들은 "Competent AGI" 범주로 분류될 수 있다.

단계별 성과는 특정 단계에서의 인간 대비 수행 수준을 나타내며, 일반성은 다양한 인지 과제에서 해당 성과를 안정적으로 유지하는 범위를 의미한다. 특히, 단계별 성과 기준은 "대부분의 과제에서 달성해야 하는 최소 성능"을 기준으로 하되, 일부 과제에서는 더 높은 성능을 보일 수 있다는 것이다. 또한, 성과와 범용성의 발달은 반드시 선형적으로 진행되지 않는다. 새로운 과제를 스스로 학습하는 메타인지 능력이 확보되는 시점에서 급격한 성능 도약이 발생할 수 있다.

0단계 General Non-AI는 인공지능이 아닌 단순 계산이나 규칙 처리 수준을 의미한다. 1

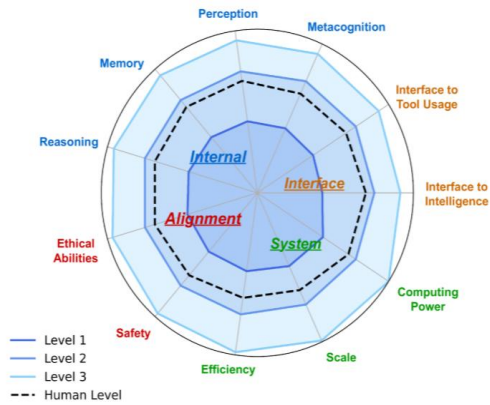
2) 미국 자동차학회

단계 Emerging AGI는 일부 인지 과제에서 초보적이거나 숙련되지 않은 인간과 비슷하거나 약간 나은 성능을 보이지만, 전반적인 수행 능력은 제한적이며, ChatGPT, Llama 2, Gemini 등이 이에 해당한다. 2단계 Competent AGI는 대부분의 인지 과제에서 숙련된 성인의 하위 50% 이상 성능을 보이며, 일부 영역에서는 전문가 수준을 달성할 수 있지만, 현재 공개된 범용 AI 중 이 수준에 도달한 사례는 없다. 3단계 Expert AGI는 숙련된 성인의 상위 10% 이상의 성과를 보이며, 창의적 문제 해결에서도 높은 성취를 보인다. 4단계 Virtuoso AGI는 대부분의 과제에서 상위 1% 수준을 유지하며, 최고 수준의 인간 전문가와 동등하거나 그 이상의 창의성을 발휘한다. 5단계 ASI(Artificial Superintelligence)는 모든 인지 과제에서 인간을 능가하며, 인간이 본질적으로 수행할 수 없는 질적으로 새로운 과제까지 해결할 수 있는 능력을 포함한다. AlphaFold가 단백질 구조 예측에서 세계 최고 과학자를 넘어선 것이 단일 과제에서의 대표적인 사례다.

이와 같은 단계별 정의는 단순한 기술적 성취의 평가를 넘어, 각 단계에서 요구되는 사회적 수용성, 법·윤리적 기준, 그리고 안전성 검증 체계 마련에 중요한 근거를 제공한다. 특히 AGI를 단일 종착점이 아닌 다층적·연속적인 발전 경로로 바라볼 경우, 각 단계별로 잠재적 위험 요소를 조기에 식별하고 이에 대응할 수 있는 체계적 관리가 가능해진다.

나. 3단계 AGI

〔그림 2－1〕 Internal, Interface, Sytem, Alignment에 따른 3단계 AGI



자료: Feng et al., 2024. Internal, Interface, Sytem, Alignment에 따른 3단계 AGI

〈표 2－2〉 AGI 수준 및 특성

카테고리	특성	L1	L2	L3
일반(General)	특정 분야에서 인간 성능을 능가	✓	✓	✓
	실제 환경에서 인간 성능을 능가	✗	✓	✓
	인간 개입 없이 자체 진화	✗	✗	✓
내부(Internal)	최소한의 인간 개입으로 새로운 상황에 적응	✗	✓	✓
	도메인 간 지식 일반화	✗	✓	✓
	창의성과 혁신 발휘	✗	✗	✓
	복잡한 의사결정 과정 수행	✗	✗	✓
인터페이스(Interface)	인간 및 다른 AI 시스템과 원활히 협업	✗	✓	✓
	새로운 도구를 자율적으로 생성	✗	✗	✓
	자기 학습과 적응을 통한 지속적 향상	✗	✓	✓
	공감, 감성 지능, 사회적 지능 발휘	✗	✗	✓
시스템(System)	초안정성, 저지연, 고처리량 제공	✓	✓	✓
	데이터, 전력, 컴퓨팅 효율성을 갖춘 설계	✗	✓	✓
	자동 학습, 조정, 협업, 배포 지원	✗	✗	✓
정렬(Alignment)	인간 지시를 정확히 따름	✓	✓	✓
	주어진 사용자의 선호를 정확히 반영	✓	✓	✓
	사용자·사회 수준의 가치와 목표에 강하게 정렬	✗	✓	✓

자료: Feng et al., 2024. AGI 수준 및 특성 비교

3단계 Ultimate AGI는 AGI 발전의 정점에 해당하는 개념으로, 인간이 제공하는 세부적인 지시 없이도 고수준의 모호한 목표를 이해하고, 스스로 발전하며, 모든 학습·추론·의사결정을 자율적으로 수행할 수 있는 단계이다. 이 단계의 AGI는 인간의 능력을 전 범위에서 초월하며, 기술 개발 과정에서조차 인간 개입이 전혀 필요하지 않다. 궁극적 AGI는 인간의 가치와 목표에 깊이 정렬되어야 하며, 그렇지 않을 경우 막대한 사회적·윤리적 위험을 초래할 수 있다. 또한, 이 단계의 AGI는 공감(empathy), 사회적 인식(social awareness), 심지어는 자기 인식(self-consciousness)에 가까운 특성을 보여줄 가능성이 있으며, 이를 통해 인간 및 다른 AI 시스템과의 원활하고 장기적인 협력이 가능해진다. 그러나 궁극적 AGI는 여전히 이론적 개념으로, 그 실현 가능성과 시점은 불확실하며, 현재도 활발한 연구와

논쟁의 대상이다.

이와 같은 3단계 AGI 정의는 기술의 현재 위치를 평가하고 향후 발전 목표를 명확히 설정하는 동시에, 각 단계에서의 위험 요소를 식별하고 대비책을 마련하는 데 중요한 기준이 된다. 특히, Embryonic 단계에서 Superhuman 단계를 거쳐 Ultimate AGI로 향하는 여정은 단순한 성능 향상을 넘어, 인간과 인공지능의 관계, 사회 구조, 그리고 인류의 미래 방향성을 재정립하는 과정이기도 하다.

상기에 서술된 AGI 개념 정립 및 논의의 역사를 다시 한번 정리하면 다음과 같다. 초기에는 AGI를 단일 단계로 규정하려는 시도가 주류였다. 대표적으로 튜링 테스트는 “기계가 인간처럼 대화할 수 있는지”를 기준으로 AGI를 판단하려는 최초의 접근으로 제시되었으나, 이는 기계의 실제 지능보다 사람을 속일 수 있는지에 초점을 둔 한계를 지닌 것으로 평가되었다. 이후 Strong AI 논의는 AGI를 ‘마음을 지닌 인공지능’으로 확장하며 의식과 이해 여부를 포함하는 정의를 제시하였으나, 실증적 판별 방법 부재로 실용성이 부족한 것으로 보고했다. 또 다른 흐름에서는 인간 뇌 유사성 기반 정의가 등장했으나, 현대 트랜스포머 기반 모델이 뇌 유사 구조 없이도 높은 성능을 발휘함에 따라 필수 조건으로서의 타당성이 약화된 것으로 평가했다. 이후 연구 경향은 단일 정의에서 벗어나 AGI를 단계별로 구분하고 범용성 및 성능과 같은 다차원 기준을 포함하는 방향으로 확장되었으며, 최근 연구들은 이러한 단계적 접근의 중요성을 강조한다.

본 보고서는 이러한 흐름을 종합한 결과, AGI 정의의 최종 형태로 Morris et al.(2024)의 6단계 정의를 채택한다. Morris et al.은 성능과 범용성 두 축을 기준으로 AGI를 Level 0부터 Level 5까지 단계별로 구분하였다. 해당 분류에서 Level 1 Emerging, Level 2 Competent, Level 3 Expert, Level 4 Virtuoso, Level 5 Superhuman은 각각 범용성과 성능의 결합 수준에 따라 정의되며, 현재 ChatGPT, Gemini 등은 Emerging AGI 단계에 위치한 것으로 구분된다. 보고서는 이러한 단계 기반 접근이 AGI를 단일 종착점이 아닌 경로로 이해하게 해주며, 각 단계별 목표와 위험 요인, Human-AI Interaction 변화를 명확히 제시한다는 점에서 가장 타당한 정의라고 판단하고, 해당 6단계 모델을 본 연구의 AGI 최종 정의로 채택한다.

제 2 절 최신 LLM의 주요 특징 및 향후 발전 방향 분석

인공지능의 발전과 함께 AGI(Artificial General Intelligence, 범용 인공지능)를 정의하려는 시도는 연구 동향에 따라 지속적으로 변화해 왔다. 초기 연구에서는 AGI를 단일 단계로 규정하는 경향이 두드러졌으나, 최근 연구들은 이를 여러 단계로 세분화하고, 다차원적 속성을 포함하는 방향으로 개념을 확장하고 있다.

1. 최신 LLM의 주요 특징

최근의 LLM과 생성형 AI 모델은 단순히 주어진 입력 문장을 다음 단어로 예측하는 확률적 텍스트 생성기를 넘어, 다양한 모달리티를 처리하고 복잡한 추론 및 계획을 수행하는 지능형 시스템으로 진화하고 있다. 특히 GPT-5, GPT-4o, Gemini Ultra, LLaMA 3와 같은 최신 모델들은 단순 언어 이해를 넘어서 멀티모달 데이터까지 처리 가능한 Perception(지각), 복잡한 논리 및 인과관계를 다루는 Reasoning(추론), 장·단기 정보의 관리와 활용을 포함하는 Memory(기억), 그리고 자기 인식·전략 수정·사회적 추론 능력을 아우르는 Metacognition(메타인지)로 발전 구조를 세분화하고 있다. 이러한 네 가지 축은 단순한 기술적 분류를 넘어, 인공지능이 AGI(Artificial General Intelligence)로 향하는 필수적인 인지 구조의 기초 틀로 간주된다.

가. Perception

Perception 측면에서, 초기 LLM은 텍스트라는 단일 모달리티에 한정되어 있어 실세계의 복잡한 정보를 충분히 반영하지 못했다. 그러나 최근에는 이미지, 음성, 영상, 심지어 그래프나 센서 데이터까지 처리 가능한 멀티모달 LLM(He et al., 2024; Laurençon et al., 2024)이 등장하여 이러한 제약을 극복하고 있다. 이들은 서로 다른 모달리티의 데이터를 하나의 통합 표현 공간으로 매핑하여 모델이 일관된 방식으로 이해·추론할 수 있도록 한다. 멀티모달 통합 방식은 크게 외부 결합(External connection)과 내부 결합(Internal connection)으로 구분된다. 외부 결합은 LLM 외부에 별도의 인식 모듈을 두고 이를 LLM과 연결하는 방식이다. Flamingo(Alayrac et al., 2022) 계열 모델은 Cross-attention 기법을 활용해 언어 모델 내부에 시각적 토큰을 주입함으로써 텍스트와 이미지 간의 상호작용을 가능하게 한

다. 내부 결합은 LLM 내부 구조에 직접 다른 모달리티를 주입하는 방식이다. Fuyu(Bavishi et al., 2023)와 같은 모델은 시각 정보를 언어 토큰과 동일하게 처리해 완전한 autoregressive 구조에서 멀티모달 학습을 수행한다. 이러한 발전은 단순히 모달리티 확장을 넘어, 텍스트·이미지·비디오·음성 간 상호 변환(generation)까지 가능하게 하여 LLM의 적용 범위를 넓히고 있다.

나. Reasoning

Reasoning은 논리적이고 체계적으로 생각하는 과정으로, 증거와 과거 경험을 바탕으로 결론을 도출하거나 결정을 내리는 것이다(Huang and Chang, 2022). 먼저, 사고 경로 탐색 기법은 모델이 복잡한 문제를 단계적으로 분해하여 해결하도록 돕는다. Chain of Thought (CoT)는 다단계 문제를 중간 추론 단계로 분해하여 해석 가능성과 오류 추적성을 높이며, Tree of Thoughts(ToT)는 트리 기반 탐색을 통해 여러 경로를 비교·선택하고 필요시 되돌아가는 유연성을 제공한다. Graph of Thoughts(GoT)는 추론 단계를 그래프 구조로 조직해 복잡한 경로와 의존관계를 처리하며, 피드백 메커니즘을 통해 더욱 정교한 추론을 가능하게 한다. 다음으로, 자기 일관성 기법은 결과의 안정성과 신뢰성을 강화한다. Self-consistency는 다양한 추론 경로 중 일관된 결과를 선택하고, Maieutic Prompting은 설명과 추론을 순환적으로 연결해 결론을 도출하며, Progressive-Hint Prompting은 이전 답변을 힌트로 활용해 점진적으로 정답에 접근한다. 마지막으로, 언어·세계·에이전트 모델 통합 접근은 단일 LLM의 한계를 보완해 보다 인간과 유사한 계획·추론을 구현한다. LAW 프레임워크는 언어모델, 세계모델, 에이전트모델을 결합해 인지적으로 완결된 추론 체계를 구성합니다. 이러한 Reasoning 기법들은 AI가 복잡하고 변화하는 환경에서도 일관성 있고 신뢰성 있는 문제 해결과 의사결정을 수행하는 데 중요한 역할을 한다.

다. Memory

Memory 능력 역시 현재 LLM 발전의 핵심 축이다. LLM의 기억은 단기 기억과 장기 기억 두 범주로 나눌 수 있다. 단기 기억은 현재 의사결정 과정에 필요한 정보를 일시적으로 저장하고 활용한다. 대표적으로 in-context prompting이 있으며, 이는 모델이 대화나 작업의 컨텍스트를 사용해 추가 정보나 예시를 제공(Wang et al., 2020)하거나 중간 추론을 수행(Wei et al., 2022)하게 한다. 단기 기억에는 (1) 지각 모듈에서 수집·처리한 실시간 데이

터, (2) 추론·계획·자기 진화 모듈의 즉시 출력, (3) 장기 기억에서 신속히 검색한 정보가 포함된다. 이 정보들은 종합되어 이후 행동이나 의사결정에 직접 반영된다. 장기 기억은 경험과 지식으로 구성되며, 의사결정과 추론의 기반이 된다. 경험은 과거의 관찰, 행동, 피드백 등을 이벤트 흐름, 경로, 사고 기록, 입출력 쌍 등의 형태로 저장하여 관련 사례 검색과 일반화된 판단을 가능하게 한다. 지식은 세계와 자기 자신에 대한 이해로, 경험에서 습득하거나 외부 지식원을 통해 보완된다. 예를 들어, Voyager(Wang et al., 2023)는 실행 가능한 코드를 저장하고, ReAct(Yao et al., 2022)는 Wikipedia API, ChatGPT Browse with Bing은 인터넷 검색을 통해 부족한 정보를 보완한다. 또한Retrieval-augmented methods (Borgeaud et al., 2022) 기법은 비정형 텍스트 기반 지식을 활용한다.

라. Metacognition

Metacognition는 자기 인식, 정서적 회복력, 그리고 리더십과 혁신에 대한 동기 등 고차원적 인지와 정서 능력을 가리키며, 조직과 AGI 모두에 필수적이다(Choudrie & Selamat, 2006). AGI에서는 이러한 능력이 자기 인식, 의식, 자기 진화, 그리고 마음 이론(Theory of Mind)으로 구체화되며, 이는 사회적 상호작용을 원활하게 하고 윤리적 의사결정을 가능하게 하며 자율적 학습과 진화를 촉진한다(Huang et al., 2023c). 자기 진화는 반복 학습, 코드 실행, 시뮬레이션 피드백, 프롬프트 최적화 등을 통해 구현되며, 최근에는 과거 경험을 체계적으로 활용하는 방법론이 주목받고 있다(Qian et al., 2024).

2. 향후 발전 방향

현재의 대형언어모델과 생성형 AI는 이미 텍스트, 이미지, 음성 등 다양한 형태의 데이터를 처리하고 복잡한 언어 추론을 수행하는 능력을 보여주고 있지만, 여전히 AGI (Artificial General Intelligence) 수준에 도달하기 위해서는 인지 구조 전반의 질적 도약이 필요하다. 특히, Perception, Reasoning, Memory, Metacognition의 네 가지 핵심 영역에서 고도화가 필수적이며, 이들 각각은 독립적으로 발전하는 것뿐만 아니라 상호 긴밀히 연결되어야 한다.

가. Perception

AGI 수준의 Perception은 단순히 멀티모달 입력을 병렬로 처리하는 수준을 넘어, 실제

세계의 물리적·인과적 구조를 이해하고 실시간으로 감각 정보를 통합하는 방향으로 발전해야 한다. 예를 들어, 로봇이 카메라를 통해 얻은 시각 데이터와 마이크로 수집한 음성 신호, 그리고 촉각 센서를 통해 감지한 질감을 하나의 통합된 장면 인식으로 결합하여 즉각적으로 상황을 판단하는 능력이 필요하다. 이를 위해서는 대규모 멀티모달 데이터 학습과 더불어, 환경 변화에 따른 맥락 파악과 미래 상태 예측까지 가능하게 하는 '세계 모델(World Model)'의 구현이 핵심이 될 것이다.

나. Reasoning

Reasoning 영역에서는 현재의 Chain of Thought(CoT) 기반 다단계 추론을 넘어서, 복합적인 인과 관계 분석, 불완전한 정보 속에서의 의사결정, 그리고 장기적인 계획 수립 능력까지 확대되어야 한다. 이러한 고차원 추론은 단순히 대규모 데이터 학습만으로는 달성되기 어렵고, 심볼릭 추론과 신경망 기반 추론을 결합하는 하이브리드 구조를 통해 구현될 가능성이 크다. 또한, AGI급 Reasoning은 환경과의 지속적인 상호작용 속에서 가설을 세우고 실험·검증을 반복하며 스스로 추론 전략을 수정하는 능동적 특성을 가져야 한다.

다. Memory

Memory 시스템의 발전 방향은 현재의 제한된 컨텍스트 윈도우나 외부 검색 시스템을 통한 일시적 정보 참조를 넘어, 장기적으로 중요한 정보를 영구적으로 보관하고 의미에 따라 선택적으로 갱신하는 자율적 기억 관리 체계로 향해야 한다. 이는 불필요하거나 오래된 정보를 정리하고, 새로운 지식을 구조적으로 통합하며, 경험과 기억을 재구성해 점점 더 정교한 판단을 가능하게 한다. 결국 AGI급 메모리는 '기억의 양'이 아니라 '기억의 질'과 '지속적 진화'에 중점을 두게 될 것이다.

라. Metacognition

마지막으로, Metacognition의 고도화는 AGI가 단순히 문제를 푸는 도구를 넘어 자기 성찰과 자기 개선이 가능한 '자율 학습자'로 변화하는 핵심 동력이 될 것이다. AGI급 메타인지는 자신이 무엇을 알고 있고 무엇을 모르는지 인식하며, 부족한 부분을 보완하기 위해 능동적으로 학습 전략을 수정한다. 더 나아가, 타인의 지식 상태와 의도를 파악하고 이에 맞추어 상호작용 방식을 조정하는 사회적 메타인지 능력까지 확보해야 한다. 이러한 기능

이 결합될 때, AGI는 스스로 학습 목표를 설정하고 지식 구조를 재편하며, 환경 변화와 새로운 문제 상황에 유연하고 창의적으로 대응할 수 있는 진정한 '지능'에 가까워질 것이다.

결론적으로, 향후 LLM과 생성형 AI의 발전은 Perception, Reasoning, Memory, Metacognition의 각 영역을 개별적으로 강화하는 동시에, 이들 간의 상호작용을 극대화하여 통합적인 인지 구조를 형성하는 방향으로 진행될 것이다. 이러한 통합적 도약이 이루어질 때, 현재의 생성형 AI는 단순한 정보 처리 엔진을 넘어, 인간과 유사한 수준의 일반 지능을 갖춘 AGI로 한 걸음 더 나아갈 수 있을 것이다.

제3장 AGI 구현을 위한 핵심 기술 요소 분석 및 분류

제1절 AGI를 위한 기존 연구들의 핵심 기술 요소

1. AGI 핵심기술 요소 도출을 위한 기존 AGI 정의들

본 절에서는 핵심기술요소를 도출하기 위한 관점에서 AGI의 개념적 정의를 간단히 개관해본다. 앞장에서도 논의되었다시피, Morris et al.(2024) 등에서는 AGI를 범용성을 충분히 갖추면서 각 과업에서 인간과 비슷한 성능을 달성하는 시스템으로 정의하고 있으며, 범용성과 성능(또는 Speciality)를 점차적으로 고도화하는 방향으로 단계별로 AGI를 정의하였다. 추가로, Zhang et al.(2024)에서는 범용성을 특정 도메인에 제한하면서 응용성을 높이는 방향으로의 Specialist Generalists Intelligence(SGI) 같은 AGI 중간 단계로 제안하기도 했다.

먼저, Morris et al.(2024)는 AGI 정의를 위해 다음 6원칙을 제시하였다.

〈표 3－1〉 AGI 정의를 위한 6원칙

원칙	내용
과정보다는 능력(Capabilities)에 초점	• 핵심은 무엇을 달성할 수 있는가이지, 어떻게 수행하는가(Process)가 아님. • 의식(consciousness), 감각(sentience) 등은 과정 중심 개념으로 AGI의 필수 조건이 아님.
범용성(Generality)과 성능(Performance)에 초점	• AGI 정의에는 두 요소 모두 필요. • 범용성만, 혹은 성능만 강조하는 정의는 불완전함.
인지·메타인지(Cognitive & Metacognitive) 과제에 초점, 물리적(Physical) 과제는 아님	• 로봇적 구현(embodiment)은 선택적 요소일 뿐 필수는 아님. • 다만 메타인지 능력(새로운 과제 학습, 인간에게 도움 요청 등)은 범용성 달성에 핵심적.
배치(Deployment)가 아닌 잠재력(Potential)에 초점)	• 실제 세계에 배치하지 않아도 일정 수준의 과제를 수행할 능력을 보이면 AGI라 할 수 있음.

원칙	내용
배치(Deployment)가 아닌 잠재력(Potential)에 초점	배치를 조건으로 하면 비기술적(법·사회·윤리) 문제로 정의가 불필요하게 복잡해짐.
생태적 타당성(Ecological Validity)에 초점	• AGI 평가는 사람들이 실제로 가치를 두는 과제 기반이어야 함. • 경제적 가치뿐 아니라 사회적·예술적 가치까지 포함해야 하며, 단순·형식적 벤치마크는 피해야 함.
단일 종착점이 아닌 경로(Path)에 초점	• AGI는 하나의 최종점이 아니라 단계별(Level-based) 경로로 정의해야 함. • 각 단계마다 지표, 위험 요인, Human-AI Interaction 변화가 명확히 제시되어야 함.

자료: Morri et al., 2024. AGI 정의를 위한 6원칙

Morris et al.(2024)에서 제시한 2번째 원칙인 **범용성과 성능**, 6번째 원칙인 **경로로의 목표**에 따라, 기존 연구자료들 대부분 AGI를 단계별 목표로 제시하고 있으며, 먼저 저자들은 앞 장에서 소개되었듯이 Morris et al.(2024)에서는 Emerging(Level 1), Competent(Level 2), Expert(Level 3), Virtuoso(Level 4), Superhuman(Level 5)로 구분하였다. Feng et al.(2025)에서는 Embroyonic AGI(Level 1), Superhuman AGI(Level2), Ultimate AGI(Level3)의 3단계로 나누었다. 두 논문 모두 ChatGPT, Gemini등 SOTA LLM, LVM은 모두 가장 낮은단계의 AGI에 속한다고 구분하였으며, 다시 말해 모두 특정한 범주의 태스크에서만 인간과 비슷하거나 그 이상의 성능을 보여주는 수준으로 간주하고 있다. Morris et al.(2024)에서는 Level 1에서 Level 5까지 레벨이 올라갈때, **범용성과 성능 모두를 고려하여 두 지표의 "곱"**이 증가되는 것으로 분류하고 있다. Feng et al.(2025)의 Level 2의 Superhuman AGI는 Morris et al.(2024)의 Level 4에 해당되는 것으로 분석되며, Level 3의 Ultimate AGI는 인간의 개입이 없이 완전히 자기 진화(Self-evolve)를 통해 스스로 개선할 수 있는 수준으로, Morris et al.(2024)의 Level 5를 달성하기 위한 보다 세부적인 접근법을 제시한 것으로 이해된다.

또한, Zhang et al.(2024)은 AGI를 달성하기 위한 경로상에서 중간 단계로 **Specialist Generalists Intelligence(SGI) 개념**을 제시하였는데, SGI는 최소한 한 가지 이상에서 인간 전문가를 뛰어넘는 전문성(specialized ability)을 가지면서도, 동시에 다양한 과제를 처리할 수 있는 범용성도 유지하는 AI 시스템을 말한다. Zhang et al.(2024)은 LLM은 범용성은 있으나, 전문성, 혁신, 실제응용에서 한계가 있음을 지적하면서, SGI가 실제 응용에서 더욱 적합한 접근으로 AGI 중간단계로 가치를 강조하였다. 예를 들면, 의료 등 특정 도메인에

유용한 대부분의 복잡합 태스크들은 인간 전문가 수준으로 전문적으로 수행하는 시스템이나, 해당 도메인내 "다양한" 태스크 수행에 제한되면서 타 도메인이나 새로운 태스크까지는 확장이 아직 안되어서 완전한 AGI로 볼 수 없는 시스템이더라도, 도메인내에서는 충분히 범용성이 높으므로 시장 경제적으로 효용 가치가 있다고 할 수 있다. 예시로 Saab et al.(2024)의 Med-Gemini를 1단계 SGI인 Emerging SGI로 구분하고 있다.

2. 기존연구에서의 AGI 구현을 위한 핵심기술 요소들

Feng et al.(2025)은 AGI구현을 위한 핵심기술 요소들을 Internal, Interface, System, Alignment의 4개 범주로 나누어서 정리하였다. 다음 테이블은 Feng et al.(2025)에서 소개한 4개 범주와 핵심 요소, 그리고 도전과제들을 일부 선택하여 제시한다.

〈표 3－2〉 AGI 구현을 위한 범주별 핵심 요소 기술

분류	핵심 요소	AGI를 위한 도전과제
Internal	○ Perception ○ Reasoning ○ Memory ○ Metacognition	• Percepton: 모달리티 다양화, 강건하고 신뢰할 수 있는 멀티모달 시스템, 설명 가능하고 투명한 멀티모달 모델 등 • Reasoning: 단순 상관관계가 아닌 인과 메커니즘 학습을 통한 인과학습 강화. prompting 없이도 내재적으로 복잡한 reasoning 수행 등 복합 추론 고도화, 불확실성 표현과 모호성 처리 능력 개선. 사회적 추론 능력 확보, 다중 도메인상 추론, 시뮬레이션 기반 복잡한 실세계상 추론 등 • Memory: 다양하고 계층적인 메모리 관리, 메모리와 고도화 추론의 통합, 새로운 환경 대응을 위해 메모리 Retrieval 과정 자체를 학습 및 적응 가능, RAG와 달리 스스로 새로운 지식 생성·평가·통합, 지속적 학습(Self-evolve), 과거 경험·통찰로 지식 기반 확장. • Metacognition: 호기심·내적 동기를 기반으로 스스로 학습 목표를 설정하고 지속적으로 진화하는 자율적 자기 진화(Self-evolution)와 개방형 학습, 타인의 정신상태를 깊이 이해하도록 사회적 추론과 Theory of Mind 강화, 깊은 자기 인식과 성찰 그리고 내적 Reflection능력 고도화을 통한 자기 인식(Self-awareness)과 의식(Consciousness)능력 강화 등

분류	핵심 요소	AGI를 위한 도전과제
Interface	○ 디지털 월드 및 물리적(Physical) 월드에 대한 인터페이스 ○ 다른 에이전트와의 인터페이스 ○ Human-AGI 인터페이스	• 디지털 월드: 자율적 도구 발명(Autonomous Tool Creation), VR/XR·웨어러블·스마트 환경 등 차세대 인터페이스와 원활히 통합 • 물리적 월드: 인간과 유사한 수준의 환경 해석 및 적응 능력을 확보, 시각·청각·언어·촉각 데이터를 종합적으로 처리하는 능력, 엣지 컴퓨팅 기반 실시간 대응 강화 등 • 다른 에이전트와의 인터페이스: 협력적 학습을 통해 서로의 성공과 실패에서 학습 그리고 지식·전략 공유, 상황 인식 기반(context-aware) 통신을 통해 에이전트들이 맥락에 맞춰 상호작용, 에이전트 간 안전한 행동과 소통 등 • Human-AGI 인터페이스: 다양한 환경에서의 사용자의 실질적 이익 보장, AGI 통제 강화, 개인·사회적 리스크 관리 등
System	○ 스케일 가능한 아키텍처 ○ 대규모 학습 기법 ○ 추론 최적화 ○ 비용 효율화 ○ 컴퓨팅 플랫폼	• 과학적 발견, 월드 시뮬레이션과 같은 복잡한 과업 지원을 위한 데이터 센터형 모델 • 장문맥·대규모 데이터 처리, 사용자 로컬 특화 모델과 중앙모델간 협업, 다양한 도구와 외부자원을 효과적으로 사용 • 정교한 메모리 저장을 위한 Lifelong Learning 고도화 • 다중 에이전트간 통신, 자원 공유, 모듈화, 에이전트 증가와 같은 멀티에이전트 오케스트라 방법 등
Alignment	○ 인간 가치 및 목표에 맞춘 안전한 AGI 정렬 기술	• 도구 및 다른 에이전트, 인간과의 인터페이스에 기반한 통합 정렬, 일관적이고 효율적인 정렬 기술, 모델의 생성 의도를 파악할 수 있는 투명한 정렬 방법 등

자료: Feng et al. 2025. 저자 번역 및 정리

한편, Zhang et al.(2024)에서는 SGI에서의 프레임워크로 System1과 System2 각각 개선시키고 이들간의 협력, 그리고 Dual system을 통한 자가 진화로 구성된 3개 레이어의 요소들을 제안하고 있다.

〈표 3－3〉 SGI 구현을 위한 System 1과 System 2 융합기반 3계층 구조

레이어	내용	내용
Layer 1	○ System 1 능력 강화 (Enhancing System 1 Abilities)	• System 1은 world knowledge base에 의해 구축되는데, 이는 system2에서 유도된 추상화된 규칙에 따라 확장되고, system2상 경험이 누적되면 system 1 다시강화
	○ System 2 능력 강화 (Enhancing System 2 Abilities)	• System 2는 전문가 수준의 기술과 구조적 일반화에 필수적이며, 반복 탐색, 출력 피드백, 자기 수정 과정을 통해 전문가 수준의 복잡한 문제 해결 능력 강화
Layer 2	○ System 1-System 2 협력적 융합 (Collaborative Fusion of System 1 & 2)	• System 1은 신속한 실행을 담당하고, System 2는 그 결과를 교정·검증하는 역할로, 두 시스템 간의 협업과 피드백 루프를 통해 전체 성능 상승
Layer 3	○ 자기 진화(Self-Evolving) 메커니즘	• System 1과 2는 듀얼-시스템 인터랙션 및 동적이고 unseen 태스크상 자기 평가를 통해서 자기 진화, 환경과 상호작용하며 스스로 탐험하고 학습, 피드백 기반으로 점진적 향상, 장기적으로 지속적 자기 개선(Self-improvement) 과 자율적 학습(Self-learning)의 기반

자료: Shang et al. 2024. 저자 재정리

또한, Shang et al.(2024)에서는 한계없는 Context Length에 기반한 LLM이 AGI의 답은 아니라고 주장하며, RAG와 같은 외부지식 활용은 메모리기반으로 볼 수 있으나 Memory는 단순히 raw data라고 하면서, AI-Native Memory의 도입을 주장하였다. 다시 말해, LLM을 컴퓨터의 CPU(프로세서), LLM의 컨텍스트를 RAM, 그리고 장기 기억으로 기능할 수 있는 'AI-Naive Memory'을 **Lifelong Personal Model(LPM)**로 구성시켜, AGI의 핵심 인프라로서 다음과 같이 L0-L1-L2 구조를 제안하고 있다.

〈표 3－4〉 AI-Naive Memory에 기반한 L0-L1-L2 구조

단계	컴포넌트	내용
L0	Raw Data Memory	기존 RAG 방식처럼 원시 데이터 자체를 메모리로 활용
L1	Natural-Language Memory	추론 결과를 자연어 요약 형태로 저장 (예: 사용자 프로필, 주요 사실, 요약)
L2	AI-Native Memory(Lifelong Personal Model, LPM)	뉴럴 네트워크 기반으로, 언어로 설명하기 어려운 기억까지 파악하고 압축하여 저장하는 개인화된 메모리 모델

자료: Shang et al. 2024. 저자 재정리

그리고, 메모리와 관련하여, AGI를 목표로 명시적으로 밝힌 것은 아니지만, Wu et.(2025)에서는 Human 메모리와 비교하여, 기존 AI시스템의 메모리 방법들을 대상(personal과 system memory), 형태(non-parametric 과 parametric memory), 그리고 시간(short-term과 long-term memory)의 3개 차원으로 구분하여 정리하였다. 향후 메모리 구조 연구방향으로 멀티모달메모리, 정적 메모리와 실시간 메모리 통합, 단기 도메인 트고하 메모리를 넘어 다양한 메모리 유형을 통합한 범용성을 높인 포괄적(Comprehensive) 메모리, 개별 메모리에서 공유 메모리, 자율진화형 메모리 유형을 강조하였는데, 정리하자면 통합형 **범용(General) 메모리 구조**를 추구하는 방향이라고 할 수 있다.

기타, Raman et al.(2025)에서는 AGI 경로로 사회적 경로, 기술적 경로, 설명가능성 경로, 인지/윤리 경로, 뇌 영감(Brain-inspired) 경로를 언급하면서 이들이 서로 연계되어 함을 주장하고 있다. Kurshan(2025)에서는 AGI가 단순한 모델 규모 확장에서 비롯되는 것이 아니라, 시스템 아키텍처 설계를 통한 구조적 전환이 본질적이라는 점을 강조하고, 시스템 전반의 설계를 위한 Systemic AI 접근으로 인간 뇌를 모사하는 시스템 설계, 에너지 효율성 확보, 도덕적 판단을 위한 통합 Alignment 구조 설계 등을 제안하였다.

이들 내용에서 보다시피, 기존연구에서는 일부 AGI를 위한 핵심 요소로서 “메타인지”(Megacognition) 를 꼽고 있다는 점을 들 수 있고, 또한 Reasoning이나 System 2요소를 별도로 분리해서 다루고 있다는 점, 환경과 상호작용 강화가 강조되고 있다는 점, 또한 메모리 구조를 보다 범용성을 높이고 인간의 메모리에 보다 가깝게 구성하도록 하는 점, 인간 뇌를 모사하는 시스템 설계, 설명가능성과 윤리적 책임을 고려한 설계를 지향하도록 하는 점들이 특징이다.

제 2 절 AGI의 핵심요소기술 분석 및 분류

1. LLM으로부터 AGI로의 방향

AGI에서는 범용성을 지향하면서 인간지능에 보다 가까운 레벨로 도약하는 것이 목표이므로, 현재 대규모 데이터에 기반을 둔 LLM 프레임워크나 방향성도 기술적으로 재검토가 필요할 수 있다. 우선 AGI에 대한 연구는 아니지만, Mahowald et al.(2024)에서는 사람의 인지능력과 대조하면서, 언어사용에 대한 형식적 능력(formal competence)와 실세계 추론을 포함하는 기능적 능력(functional competance)영역에서 LLM에서 보여지는 능력들을 조사한 결과, LLM은 형식적 능력에는 매우 강하지만 사람의 고차인지능력에 해당되는 기능적 능력들을 실현하는데는 한계가 있다는 점을 지적하였다.

한편, AGI를 목표로 현재 ChatGPT와 같은 SOTA LLM을 기반을 둘 수 있는 가에 대해서는, 연구자들은 최소한 데이터 효율적인 학습패러다임이 설계되어가 하거나 또는 더 나아가 완전히 새로운 인지아키텍처가 필요하다고도 주장하기도 한다. 비교적 LLM 아키텍처에 대해서 우호적인 입장을 보이는 Bubeck et al.(2023)에서는 LLM이 다양한 도메인에서 인간 수준에 가까운 성능을 보인다는 점을 강조하면서, 특히 GPT-4가 수학, 코딩, 비전, 의학, 법률, 심리학 등 여러 영역에서 특별한 프롬프트 없이도 문제를 해결할 수 있는 능력을 실험적으로 보여주었다. 저자들은 GPT-4가 여전히 불완전하지만 일종의 초기 AGI 시스템(Early yet incomplete version of an AGI system)으로 간주될 수 있다고 평가하면서, AGI의 첫 사례라는 점을 강조하였다. 동시에 GPT-4의 한계를 지적하면서, AGI를 위해서 단순한 모델 크기 확대나 데이터 양 증가만으로는 충분하지 않으며, **다음 단어 예측을 넘어선 새로운 패러다임이 필요**하다는 점을 추가로 제안하였다.

반면, Goertzel(2023)에서는 LLM이 인지적 측면에서 보여주는 능력을 평가하면서도, **AGI로 나아가기 위해서는 근본적으로 다른 인지 아키텍처 설계가 필요**하다고 강조하고 있고, 점진적 모델 확장은 한계가 명확하며, LLM을 AGI의 일부 구성 요소로 활용하되 **보다 포괄적인 인지 구조 및 시스템 설계가 병행**되어야 함을 주장하고 있다.

이들 연구들을 정리하자면, LLM에 기반한 GPT-4류는 AGI의 초기 형태로 간주도리 가능성이 있지만, 그러나 AGI까지 나아가려면, 새로운 구조적 개념적 접근이 필수적이라는

점을 공통적으로 제시하고 있다.

2. AGI의 핵심요소기술 도출: 새로운 혁신 기술요소 및 LLM 개선 및 고도화 병행

LLM류의 대규모 데이터 기반 모델은 현재까지는 성공적으로 적용되어 AGI의 가장 낮은 레벨은 달성된 상황이며, AGI를 위한 새로운 접근이 필요하지만, LLM류가 AGI를 향해서는 기술 최전선에 있는 구조이며, 실제로 이러한 방향으로서 고도화를 진행했을 때 AGI의 실현 한계가 어느정도인지는 가늠하지 어렵기 때문에, AGI를 위해 현재의 LLM류의 프레임워크의 학습 효율성을 개선하면서 더욱 고도화 하는 기술적 방향도 동시에 진행될 필요가 있을 것이다. 따라서, 본 절에서 제시하고자 하는 AGI 핵심요소기술은 LLM류의 개선 및 확장, 스케일을 포함하면서 또한 프레임워크 자체를 개선하는 새로운 방향 두 가지 모두를 아우르고자 한다.

종합적으로 본 절에서는 AGI 핵심요소기술을 **모델구조, 인식, 학습, 추론, 체화, 해석, 시스템, 메타인지, 윤리** 이렇게 8개로 나누어서 정리한다. 언급했듯이, 현재 SOTA LLM에서 다루는 토픽들도 다수 포함되어 있으며, 메모리는 별도의 범주에 다루지 않고, 모델 구조의 범주에 편성시켰다. 다음 테이블에 각 범주별 요소기술과 목표 예시를 제시하였다.

〈표 3－5〉 본 절에서 제안하는 AGI 핵심 기술 범주, 요소기술 및 목표 예시

기술 범주	요소 기술 예시	요소기술 목표 예시
모델구조	○ 트랜스포머 대안구조 ○ Human-like General memory ○ Brain-spired Model ○ Diffusion model 확장 및 대안 구조	• 대규모 데이터·연산 자원 의존도를 줄이면서도 장기 문맥 및 추론 능력을 확보 • 인간의 단기/장기/작업기억을 모사한 구조적 메모리 통합, 범용 메모리 구조 • 신경과학 원리를 반영한 계층적·모듈형 아키텍처 설계 • Diffusion model 확장 및 대안구조 제시
인식 (Perception)	○ 멀티모달리티(텍스트, 이미지, 영상, 시계열, 사운드, 기타 센싱정보) ○ 멀티모달리티 통합 표현 ○ Generalist 인코더/디코더 확장	• 텍스트·음성·이미지·영상·센서 데이터를 공통 의미 공간에서 처리 • 서로 다른 모달리티 간 zero-shot 변환 및 통합 이해 능력 확보

기술 범주	요소 기술 예시	요소기술 목표 예시
학습	○ 다음단어 예측을 넘어선 General 학습 ○ 데이터 효율적 파운데이션 모델 학습 ○ Lifelong 연속 학습 ○ Long-term 지식 마스터링 ○ 지식 편집 및 점진적 확장 ○ 도메인 특화 학습 ○ Test-Time Adaptation ○ Human-like 메모리기반 Long-term 학습 ○ Deep Alignment ○ Multi-agent Collaborative 학습	• 행동, 계획, 인과적 추론까지 포함한 General 학습 패러다임 구축 • 적은 데이터로도 일반화 가능한 학습 • catastrophic forgetting 없는 지속적 지식 축적 • 개별 지식 요소 간 연결·추상화를 통한 고수준 이해 • 기존 모델 파라미터를 유지하면서 지식 삽입/수정 가능 • 경험 기반 장기 기억 저장 및 재활용 • 인간 피드백을 고도화하여, 가치, 윤리의 정교한 개념을 내재화한 학습 • 여러 AI 간 상호작용을 통한 집단지성 학습
Reasoning	○ 일반 추론(General reasoning) ○ 이해 기반 Robust Reasoning ○ 문제 복잡도에 따른 Reasoning Scaling ○ Retrieval과 Reasoning 통합 ○ Efficient reasoning ○ Unseen 환경에서 creative reasoning ○ System 2 Reasoning 스케일링 ○ System 2 Reasoning을 System 1 능력으로 내재화	• 수학, 코딩을 위한 추론을 넘어, 다양한 도메인의 추론으로 범용성 확장 • 문제 구조와 의도를 이해할 수 있는 유연한 추론 메커니즘 • 문제의 복잡성이 높아지더라도 붕괴하지 않고 견고하게 Long-term Reasoning을 수행 • Retrieval과 Reasoning통합 메카니즘을 통한 RAG 고도화 • Overthinking극복을 위한 효율적 Reasoning • 학습되지 않은 문제에서 새로운 해결법 창출 • AI의 인간 유사 추론 능력 강화, MCTS 기반 reasoning 제어, neural-symbolic 통합 아키텍처 • System 2의 Reasoning경험을 내재화
체화 (Embodiment)	○ 디지털 월드 ○ 물리 월드 ○ 내적 월드 ○ 사회적 월드 ○ Theory of Mind ○ 월드 모델	• 디지털 및 물리 월드와의 상호작용 강화 및 세계 자가 탐색 및 확장 • 내적 월드에 대한 상호작용은 Thinking 액션을 고도화하여 에이전트 인지구조를 세계 탐색 대상으로 간주 • 다중 에이전트 환경 속 사회적 규범, 행동 학습, 다른 에이전트에 대한 이해 • 환경의 동적, 인과적 구조를 월드모델화 하여 LLM통합

기술 범주	요소 기술 예시	요소기술 목표 예시
해석 (Interpretation)	○ In-context learning 분석 ○ 기계론적 해석(Mechanistic interpretation) ○ Causality 모델링 ○ Uncertainty 모델링 ○ 설명가능한 AGI ○ AGI Scaling Law	• 프롬프트 기반 학습 메커니즘 창발의 구조적 및 근본적 이해 • 내부 뉴런, 회로 수준에서 모델 동작 원리 해석 • 인과관계 학습 및 추론을 통한 설명 가능성 확보 • 불확실성 추정 및 추론결과에 다양한 내적 상태로 연결 • AGI의 범용성과 성능 곱과, 모델의 파라미터·데이터·연산량 간 관계 규명
시스템 (System)	○ Energy-efficient AGI ○ 범용성/성능을 모두 고려한 모델 경량화 ○ 온 디바이스 AGI를 위한 Efficient finetuning ○ Long context 문맥 처리 ○ 멀티모델 병합 및 라우팅 ○ Federated Learning	• 에너지 효율성을 염두한 AGI 모델 • AGI 범용성과 성능을 모두 고려한 모델 경량화 • AGI 모델의 온 디바이스 탑재를 위한 최적화된 파라미터 효율적 학습 기법 • 수백만~수천만 토큰 이상의 in-context 입력 또는 출력을 효율적으로 처리, 사용자-AGI 대화 히스토리를 처리하거나 메모리화 • AGI 모델에서 파생된 다양한 특화 모델들을 효율적으로 병합, 라우팅을 통한 고도화 • 분산 환경에서 프라이버시 보장 학습
메타인지 (Metacognition)	○ 메타인지 Emergent Learning ○ 메타인지기반 학습 및 추론 ○ 내적상태 모델링 ○ 학습과 추론 통합	• 학습과 추론 등 영역에서의 메타인지 능력을 창발시키는 방법 • 메타인지에 기반한 새로운 학습과 추론 메커니즘 • 동기, 목표, 다양한 내적상태 모델링, 그리고 메타인지에 기반한 내적 월드에 대한 상호작용 고도화 • 학습과 추론을 통합
윤리 (Ethics)	○ 인간가치와 충돌 해결 ○ 고도의 인간가치 이해 및 행동 제어	• 인간 가치 충돌 회피를 최우선 취하는 행동 메카니즘 • 문화 및 사회 맥락을 정교하게 이해하여 행동하도록 제약

여기서 주목할만하게, 메타인지는 Feng et al.(2025)에서도 AGI 내부구조에서 다룬 범주이기도 한데, Feng et al.(2025)을 포함하여 AGI 기존 연구들 몇몇에서 필요성을 언급하였음과 동시에 기존 AI연구에는 없는 새로운 요소기술 범주라고 볼 수 있다. 본 절에서는, AGI에서 범용성을 갖추기 위해서는 메타인지(또는 그에 상응하는 능력)가 출현되는 것이 중요한 목표 중 하나가 될 수 있을 것이라 가정하였다. 다시말해, 현재 LLM에서의 인지능력들을 단순히 Scaling을 한다고 AGI가 달성되는 것이 아니라, LLM에서 대규모 데이터로부터 In-context learning과 같은 능력들이 창발되었던 것처럼, 현재 수준의 in-context learning과 같은 창발적 능력을 한단계 뛰어넘은 AGI 실현에 핵심에 해당되는 능력인 메타인지가 창발되거나 도출되는 방향으로서 연구가 필요하다고 가정한 것이다. 인지과학적으로 메타인지는 '생각에 대한 생각'을 지칭하지만, 본 절에서 메타인지는 AGI의 핵심이 되는 AI와의 차별화된 능력으로, 학습을 위한 동기 부여에 대한 인지, 자기 지식상태에 대한 인지, 추론패턴에 대한 인지, 귀납적으로 무의식적으로 파악하던 내용에 대한 개념적 인지(개념화), 내적 상태에 대한 신념적 인지, 더 나아가서 의식과 주관적 인식체계 등, 보다 포괄적이고 차별화된 의미를 갖는다고 볼 수 있다. 결국, 메타인지 능력이 도출될 때, AGI에서 범용성과 성능이 동시에 비약적인 breakthough를 이루게 되는 것으로 기대할 수 있다.

메타인지와 같은 기술 범주를 별개로 설정하는 이유는, AGI는 우리가 가보지 못한 미지의 영역이고 기술적으로 어떤 요소들이 필요할지를 가늠하기는 매우 어렵지만, 몇가지 Breaththough가 분명히 달성되어야 도달될 수 있는 성질임을 명확히 하여, 결과적으로, 이를 통해 "발견적으로" 기술을 탐구하는 방향으로서의 접근법이 필요하다는 것을 분명히 하고자 하는데 있다.

제 4 장 AGI 기술 발전 단계별 시나리오 구축

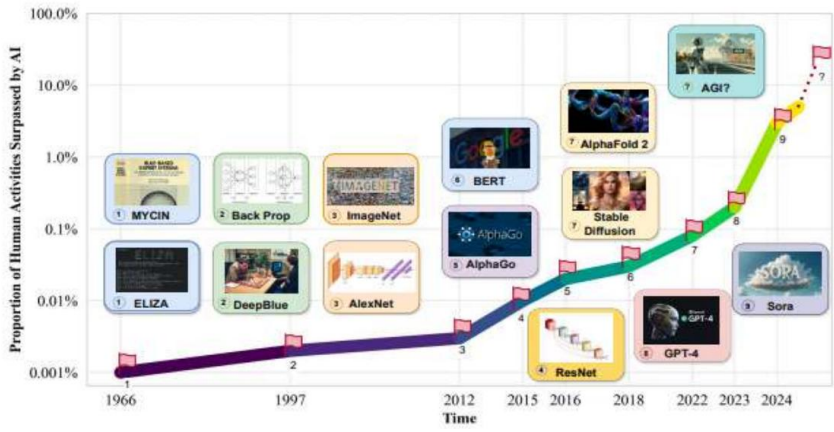
제 1 절 기술 성숙도를 고려한 AGI 발전 단계

1. 인공지능 기술의 진화와 AGI 기술 발전 트렌드 분석

가. 인공지능 기술 발전 트렌드

인공지능 기술은 그 용어와 개념이 처음 정립된 이래로 수많은 연구자에 의해 인간의 수준을 능가하기 위한 기술적 도전이 이루어지고 있다(Feng et al., 2024). 초기 인공은 전문가 시스템과 같은 낮은 수준의 인공지능에서 시작하여 인공신경망과 CNN/RNN의 심층 신경망을 거쳐 관련 기술 수준이 눈부시게 발전하여 Chat-GPT로 대표되는 LLM이 등장한 이후 사람에 필적하는 수준의 AGI에 도전하고 있는 단계에 이르고 있다.

〔그림 4-1〕 인간의 수준을 능가하기 위한 인공지능 기술 발전 트렌드



자료: Feng et al.(2024)

나. AGI 달성을 위한 기술 발전 트렌드

1) 하드웨어 가속화

GPU/TPU 클러스터 확산에 따라 딥러닝 연산 병렬화를 통해 대규모 모델을 훈련하는데 요구되는 시간을 급격히 단축할 수 있게 되었다. 이들 중 대표적인 사례는 NVIDIA의 H100 및 H200 GPU에 적용된 호퍼(Hopper) 아키텍처(NVIDIA, 2023)를 고려할 수 있다. 이 아키텍처는 트랜스포머 연산에 최적화된 FP8/FP16/FP32 혼합 정밀도를 지원하여 기존 대비 6배 정도 성능을 향상시킬 수 있다. 구글에서는 4,096개의 TPU v4 칩으로 구성되는 TPU v4 Pod를 통해 2023년 기준으로 1.1 엑사플롭스(Exa FLOPs) 연산 능력을 구현하여 대규모 언어모델 훈련 시간을 기존 대비 최대 70% 단축 가능하다고 제시하고 있다(Jouppi et al., 2023). 그밖에 NCCL(NVIDIA Collective Communications Library)과 GPUDirect RDMA (Remote Direct Memory Access) 기술 개발을 통해 다중 GPU 간 초고속 데이터 전송 체계를 구축(Hu et al., 2025)하는 등 분산 컴퓨팅 표준화 작업도 추진하고 있다.

하드웨어 가속화를 위한 또 다른 접근으로 뉴로모픽 칩의 등장을 고려할 수 있다. 기존의 CPU 아키텍처와는 달리 인간 뇌의 시냅스 구조를 모방한 저전력 고속 처리 장치를 개발하고 있다. 이러한 뉴로모픽 칩의 사례로 IBM의 TrueNorth(Akopyan et al., 2015)와 NorthPole(Modha et al., 2023), 그리고 Intel의 Loihi(Davies et al., 2018, Intel, 2024)를 대표적인 사례로 고려할 수 있다.

〈표 4－1〉 뉴로모픽 칩의 등장과 특징 분석

	기술 특징	현재 개발 현황
IBM TrueNorth/ NorthPole	TrueNorth는 10^6개의 뉴런으로 2억 5,600만 개의 시냅스를 구현	2023년 NorthPole은 12nm 공정을 기반으로 칩셋 제작
Intel Loihi/ Hala Point	스파이크 신경망 기반 실시간 적응형 학습 기능 구현	Hala Point는 로이히 2 칩 1,152개를 통합, 최대 1.15 B 개 뉴런 지원

그 밖에 메모리 내 컴퓨팅(In Memory Computing, IMC)을 통해 데이터를 메모리에 저장하고 연산까지 수행하는 기술도 개발(Morita et al., 2024)하고 있다. 기존에는 CPU가 데이터를 메모리에서 읽어와 연산 후 다시 메모리에 저장하는 방식을 사용했지만, IMC는 메모

리 내에서 직접 연산을 수행하여 데이터 이동을 최소화하고 처리 속도를 비약적으로 향상시킬 수 있다. 특히, 대용량 데이터 처리나 인공지능 분야, 그 중에서도 딥러닝 분야에서 효율적인 컴퓨팅을 가능하게 한다.

AGI를 위한 기술 발전 트렌드 중 하드웨어 가속화 분야에 대해 고려할 또 다른 기술은 양자 컴퓨팅 기술(Beer et al., 2020)이라 할 수 있다. IBM에서는 양자 신경망에 대한 이론화 연구를 통해 VQE(Vector Quantum Eigensolver) 알고리즘을 최적화 문제 해결에 활용할 수 있음을 보인 바 있다(Liao et al., 2024). 또한 Quantinuum H-Series 시스템 사례에서 보여주는 바와 같이 양자 프로세서가 특정 계산 작업만 분담하는 방식으로 기존 HPC와 결합하는 연구도 진행되고 있다. 하지만 실제 활용을 위해서는 1,500 큐비트 이상의 오류 정정 양자 컴퓨터가 필요하지만, 2023년 기준 최대 1,121개의 큐비트 머신인 IBM Condor(Gambetta et al., 2023)) 수준에 머무르고 있다. 이러한 한계에도 불구하고 양자 알고리즘 기반 최적화 문제 해결 연구가 AGI의 추론 속도 향상에 기여할 수 있을 것으로 기대되므로 양자 컴퓨팅의 활용 가능성을 탐구할 필요가 있다.

마지막으로 차세대 메모리 아키텍처 기술을 살펴볼 필요가 있다. SK 하이닉스에서는 2024년 기준 대역폭 9.6Gbps로 기존 대비 50% 증가한 성능으로 처리가 가능한 HBM3E 메모리를 공급하고 있다(SK Hynix, 2024). 또한 Xilinx Versal ACAP 사례 분석에서 살펴보듯이 프로그래머블 CMOS를 통해 FPGA 기반 재구성 아키텍처가 AGI의 동적 컴퓨팅 요구를 충족하도록 하는 연구도 진행하고 있다.

결론적으로 AGI 개발을 위해 하드웨어 기술은 에너지 효율성 개선이 필수적이다. 현재의 인공지능 아키텍처를 기반으로 하는 AGI 시스템은 대략 20W를 소모한다고 알려진 인간 두뇌 대비 최소 50,000배 이상 전력을 요구한다. 이 문제 해결 없이는 실용화가 어렵다는 점에서 저전력 설계가 최우선 과제라 할 수 있다.

2) 딥러닝 아키텍처의 진화

딥러닝 아키텍처는 2006년 제프리 힌튼의 심층 신뢰 신경망(Deep Belief Networks. DBN) 논문(Hinton et al., 2006)과 2012년 ImageNet 챌린지에서의 AlexNet(Krizhevsky et al., 2012)을 통해 그 가능성이 제시된 이래 급격한 발전을 이루어내고 있다. 초기 CNN과 RNN 중심의 구조와 대량의 학습 데이터를 이용한 지도 학습 중심의 2세대 모델에서 출발

하여 AGI에 적용이 가능한 3세대 모델을 개발하는 방향으로 발전하고 있다. 3세대 모델은 트랜스포머 기반의 초대규모 모델(Vaswani et al., 2017)로서 few-shot 학습 및 상식 전이 능력을 갖추고 있어 데이터 효율성을 개선하고 있다. 이를 위한 학습 방식으로 자기 지도 학습과 강화 학습을 결합한 형태로 기술 개발이 이루어지고 있다.

〈표 4－2〉 딥러닝 아키텍처의 진화에 대한 특징 분석

구분	2세대 모델	3세대 모델(AGI 접근형)
모델의 구조	CNN/RNN	트랜스포머 기반 초대규모 모델(GPT-4, PaLM 등)
학습 방식	지도 학습 중심	자기지도학습 + 강화학습 결합
데이터 효율성	대량 데이터 의존	Few-shot 학습 및 상식 전이 능력 확보

이상의 내용을 항목별로 구분하여 살펴보면 다음과 같다. 먼저 모델의 구조에서 트랜스포머 아키텍처의 부상을 고려할 수 있다. 트랜스포머 모델의 Self-Attention 메커니즘은 장거리 의존성 문제를 해결하여 기존 RNN/CNN 대비 20배 이상의 장기 기억 능력을 구현하는 것이 가능하다. 또한 모델의 크기, 데이터 양, 그리고 계산량 간 상관관계 입증으로 초대규모 훈련 전략을 수립하기 위한 기반을 마련하기 위한 규모의 법칙이 제시되고 있다(Kaplan et al., 2020). GPT-3가 공개된 2020년 이래로 모델의 크기는 급격히 증가하고 있고 이에 따라 훈련 데이터의 용량과 계산 요구량 역시 급격히 증가하고 있다.

〈표 4－3〉 트랜스포머 아키텍처와 규모 법칙

모델 특성	GPT-3(2020)	PaLM(2022)	Llama 4(2025)[3]
파라미터 규모	175B	540B	> 1T
훈련 데이터 용량	570GB	3.12TB	> 140TB
계산 요구량(FLOPS)	3.12e23	2.56e24	>4.2e26

자료: Brown et al. 2020, Chowdhery et al. 2022

3) Llama4의 값은 Meta에서 공식적인 발표가 없어 스케일링의 법칙, 기술적 추세 등을 고려한 추정치임

다음으로 학습 방식은 자기 지도 학습 방식이 확대될 것으로 기대된다. 사전 훈련 전략을 살펴보면 기존 BERT의 경우 Masked Language Model을 한정적으로 적용(Devlin et al., 2019))하고 있으나 GPT 계열의 모델에서는 전체 시퀀스를 예측하는 방향으로 확장하고 있다. 모델의 다중 모달리티 대응을 위해서 이미지-텍스트 매칭을 지원하는 CLIP(Radford et al., 2021)에서 시각-언어-음성을 통합 처리하는 FLAVA 등을 통해 크로스 모달리티의 이해 능력을 향상시키는 방향으로 진화하고 있다(Singh et al., 2022). 또한 few-shot learning을 통해 3-5개의 예제만으로 새로운 작업을 습득 가능한 기술(Vinyals et al., 2016)이 개발되고 있으며 이를 통해 GPT-4에서는 약 10%의 성능 향상이 가능하다고 제시되고 있다(OpenAI, 2023).

대규모 신경망의 효율화를 위한 연구로서 Mixtral 8x7B 모델에서는 Mixture of Experts (MoE) 기술을 활용하여 활성화된 전문가 유닛만 계산하여 약 130억 개의 활성 파라미터로 약 700억 개의 파라미터를 가지는 Llama 2 모델과 비슷하거나 더 뛰어난 성능을 보인다고 발표한 바 있다(Mistral AI, 2023). LLM 모델의 양자화 학습에 있어서도 FP32에서 INT8로 변환하여 사용함으로써 모델의 크기를 75% 축소하고 에너지 소모도 현저하게 감소시키는 것이 가능하다고 알려져 있다(Gholami et al., 2022). 또한 신경망 가지치기(pruning)을 통해 불필요한 가중치를 제거함으로써 BERT-large 대비 98% 정도의 성능을 유지하면서 4배 경량화를 실현하는 것이 가능하다는 연구도 발표되어 있다(Sanh et al., 2020).

현재 트랜스포머 모델은 우수한 성능을 보이지만 AGI를 위해서는 여전히 한계점을 가지고 있다고 할 수 있다. 최근 로봇 등에 적용하는 피지컬 AI 기술이 제시되고 있으나 여전히 상당 부분은 물리적 작업 수행 능력에서 한계를 가진다고 할 수 있다. 이를 극복하는 방안으로 신경 과학 기반의 하이브리드 아키텍처 개발 등이 고려될 수 있다.

3) 신경과학 기반 접근법

AGI를 위해서는 기존 모델이 가지는 한계점을 해결하기 위한 한 가지 방법으로 신경과학 기반의 접근을 고려할 수 있다. 뇌-컴퓨터 인터페이스(BCI)에서는 Neuralink 등이 개발 중인 직접적 신경 신호 처리 기술로 AGI의 생물학적 구현 가능성을 탐구하고 있다(Lavazza et al., 2025). 심층 신경망 기반 디코딩 기술과 생체 역학 모델링을 통한 새로운 접근법이 지속적으로 연구되고 있다.

BCI 방식이 bottom-up 방식이라고 한다면 인지 아키텍처 모델링은 top-down 방식이라 할 수 있다. ACT-R(Anderson, 2007)과 SOAR(Laird, 2012) 같은 이론들이 인간 사고 프로세스를 계산 모델화하는 데 적용되고 있다. ACT-R은 생명체 두뇌의 해마와 전두엽에 해당하는 작업 기억 및 장기 기억을 구현하여 인지 과학적 두뇌 모델을 개발하고 있으며SOAR 역시 ACT-R과 유사한 인지과학 기반의 모델로 계층적 주의 메커니즘과 지속적 학습 시스템을 구현하고 있다. 이러한 인지 아키텍처 모델을 LLM과 결합하여 LLM의 한계를 보완하고 AGI를 개발하는 방향으로의 시도가 이루어지고 있다(Wu et al., 2024). 이 과정에서 LLM은 방대한 텍스트나 이미지를 이해하고, 인간과 자연스럽게 상호작용하는 창구 역할을 하고 인지 모델은 LLM이 이해한 정보를 바탕으로, 복잡한 문제 해결을 위한 논리적 추론, 계획, 의사 결정과 같은 고차원적인 인지 기능을 수행한다. 이를 통해 설명 가능성(explainability)을 확보하고 견고성(robustness)과 데이터 효율성(data efficency)을 제고할 수 있다고 알려져 있다. 하지만 보다 강력한 AGI 구현을 위해서는 이러한 접근에 더해 신경 가소성을 모방하여 동적 네트워크를 재구성하는 등 뉴로 컴퓨팅 기술과 결합한 하이브리드 솔루션 개발이 필요하다.

4) 데이터 및 지식 표현의 혁신

데이터 및 지식 표현에 있어 현재 기술과 AGI 요구 수준을 살펴보면 〈표 4-4〉에서 보여주는 바와 같다. 먼저, 지식베이스의 경우 전통적으로 구조화된 데이터베이스를 통한 계층적 온톨로지를 주로 사용해왔다. 이 경우 지식에서의 의미 관계 추론은 논리 연산을 통해 대부분 이루어진다. 하지만 최근에는 지식베이스를 신경망에 임베딩시키는 형태로 발전하고 있으며 의미 관계 추론도 그래프 신경망(Graph Neural Networks, GNN)을 이용하고 있다(Kipf et al., 2017).

다음으로 멀티모달 데이터 처리에 대해 살펴보면 전통적으로 이미지 데이터와 텍스트 데이터는 각각 별도로 분리하여 처리해왔다. 하지만 최근에는 이들을 통합하여 처리하는 방식으로 발전하고 있다. LLaMA3에서 이들을 통합하여 처리한 경우 의미적 일관성 지수(Semantic Consistency Index, SCI)가 향상하는 것을 확인할 수 있다(AI@Meta, 2024).

AGI를 달성하기 위해서는 이들 각각의 기술 요소들이 더욱 발전할 필요가 있다. 지식베이스의 경우에는 상시 주론이 가능한 의미론적 네트워크가 필요하며 멀티모달 데이터의

경우 시청각 정보를 통합하여 이해하고 생성하는 능력이 요구된다.

〈표 4－4〉 데이터 및 지식 표현에 대한 AGI 요구 수준 분석

구분	현재 기술	AGI 요구 수준
지식베이스	구조화된 데이터베이스 및 비구조 데이터	상식 추론 가능한 의미론적 네트워크 필요
멀티모달 데이터	이미지/텍스트 분리 처리 또는 단순 통합 처리	시청각 정보 통합 이해 및 생성 능력

5) 윤리적 기술적 과제 해결

AGI를 위한 윤리적 과제로서 먼저 유네스코(UNESCO 2024) 등 국제기구의 윤리 프레임워크에 대한 이행 수준을 제시하고 있다. 데이터 수집 단계에서의 동의 여부에 대한 관리율도 지역별로 크게 차이가 나고 있다. 실제 데이터에 대해 익명화 기술을 적용한다고 하더라도 재식별이 가능한 경우도 편차가 매우 크다. 또한, 편향성의 문제를 해결하기 위한 방법으로 편향 탐지 알고리즘 및 공정성 보정 메커니즘 등의 마련이 필요하다. 이를 통해 AGI 모델이 OECD AI 원칙(OECD, 2019) 및 EU AI 법(EU, 2024) 등에서 제시하는 안전성 프레임워크에서 책임성 기준을 충족하도록 할 필요가 있다. OECD의 AI 원칙은 AI 시스템을 위한 5가지 원칙을 규정하고 있다. 그 내용은 포용적 성장, 인간 중심의 가치와 공정성, 투명성과 설명 가능성, 견고성과 보안 및 안전, 그리고 책임성으로 구성된다. EU의 AI 법은 2024년 최종 통과되어 현재 시행 과정에 있다. 이 법은 AI 기술을 위험 수준에 따라 4단계로 나누고 각 단계 별로 다른 수준의 규제를 적용하도록 규정하고 있다. 가장 놈은 단계가 용납할 수 없는 위험이고 이 수준에서는 인간의 안전과 권리에 명백히 위배되어 AI 시스템의 사용이 전면 금지된다. 그 아래로 고위험, 제한된 위험, 그리고 최소 위험 단계로 위험 수준이 낮아진다. 이와 같은 규정을 통해 혁신을 저해하지 않으면서 시민의 안전과 기본권을 보호하는 것을 목표로 한다.

기술적 과제로서 중요한 것 중 하나가 앞에서 언급한 바 있는 에너지 효율성 개선이다. GPT-3 훈련에 소모된 1,287MWh의 전력 문제 해결을 위한 녹색 AI 연구가 활발히 진행 중이다(Patterson et al., 2021). 이를 위한 접근 방법으로 실리콘 포토닉스 기반 칩셋을 이용

하는 광자 컴퓨팅 아키텍처, 양자 컴퓨팅과 기존 알고리즘의 하이브리드 시스템 개발을 고려할 수 있다.

2. 기존 AGI 기술 발전 단계에 대한 조사 분석

가. 구글 딥마인드(Google DeepMind)의 5단계 로드맵

구글 딥마인드는 인공지능 시스템의 성능과 일관성이라는 두 가지 축을 기준으로 AGI 발전 단계를 5 단계로 구분하는 프레임워크를 제안하고 있다. 이들이 제안하는 모델에서 Level 0는 특정 규칙에 따라 작동하는 규칙 기반의 단계로서 이 단계의 시스템은 지능이 없는 시스템으로 AGI 발전 단계에 포함하지 않고 있다. 따라서 초기 AI 단계인 Level 1에서 마지막 초지능에 해당하는 Level 5까지 5단계로 규정하고 있다. 이러한 프레임워크는 "AGI가 존재하는가?"와 같은 이분법적 질문에서 벗어나 AGI로 발전해 나가는 과정을 구체적으로 측정하고 평가하자는 목적으로 제안하고 있다.

구글 딥마인드의 접근 방식은 자율주행 기술의 단계를 레벨 0부터 5까지 나누는 것과 유사하며, AI의 능력을 더욱 객관적으로 평가하고 안전성 문제를 단계별로 대비하는 데 중점을 두고 있다.

나. OpenAI의 AGI 5단계

OpenAI는 AGI를 향한 과정을 능력의 심화 관점에서 5단계로 나누어 제시하고 있다. 이 분류는 내부적으로 AGI 개발의 진척도를 가늠하고, 각 단계별로 필요한 안전 조치와 사회적 준비를 논의하기 위해 만들어졌다. OpenAI의 발전 단계 구분에서는 ChatGPT의 초기 버전에서와 같이 질의 응답을 통해 기본적인 대화가 가능한 수준을 제일 낮은 Level 1으로 고려하고 있다. OpenAI는 자사의 모델이 Level 1을 지나 Level 2에 근접하고 있다고 자체적으로 평가하고 있다. Level 3이 가지는 주요 특징으로는 오류가 발생하더라도 스스로 문제를 해결하는 능력으로 고려하고 있다. Level 4의 시스템이 가져야 하는 주요 특징으로는 창의성을 고려하고 있다. 따라서 이 단계에서는 새로운 아이디어나 기술을 스스로 만들어 낼 수 있는 수준을 의미한다. 마지막 Level 5에서는 인간의 개입이 전혀 없어도 통합적으로 동작하는 초지능의 단계를 의미한다.

〈표 4－5〉 OpenAI의 AGI 5단계 분류

단계(Level)	명칭	설명
Level 1	챗봇(Chatbot) / 대화형 AI	ChatGPT-3처럼 간단한 질문에 답하고 기본적인 대화가 가능한 수준
Level 2	추론가(Reasoner)	별도의 도구 없이 박사급 인간과 유사한 수준의 추론 및 문제 해결 능력을 갖춘 단계입
Level 3	에이전트(Agent)	사용자를 대신하여 며칠에 걸쳐 자율적으로 복잡한 작업을 계획하고 수행할 수 있는 시스템
Level 4	혁신가(Innovator)	특정 분야에서 세계 최고 수준의 인간 전문가를 능가하는 능력을 가지는 단계(예: 노벨상 수준의 과학적 발견을 하는 AI)
Level 5	조직(Organization)	거의 모든 분야에서 인간의 능력을 압도적으로 뛰어넘는, 가상의 초지능(ASI) 단계

자료: Forbes, 2024

다. 마이크로소프트의 AGI 발전에 대한 관점

마이크로소프트는 구글 딥마인드나 OpenAI처럼 공식적인 단계별 프레임워크를 발표하지는 않고 있다. 하지만 AGI에 대해 신중한 입장을 보이면서도, AI 에이전트의 발전을 중요한 과정으로 고려하고 있다. 이를 위해 마이크로소프트는 자사의 AI 비서인 코파일럿(Copilot)의 확장을 통해 코파일럿을 단순한 챗봇을 넘어, 사용자의 업무와 일상에 깊숙이 통합되어 복잡한 작업을 대신 처리해주는 AI 에이전트로 발전시키는 데 집중하고 있다.

마이크로소프트의 일부 고위 임원은 OpenAI의 샘 알트먼과는 다르게 AGI가 단기간에 구현되기는 어려울 것이라는 신중한 전망을 내놓기도 하고 있다. 이는 AGI로 가는 길이 점진적이고 많은 기술적 난관이 남아있음을 시사하고 있다.

라. 기타 학계에서 바라보는 AGI 발전에 대한 관점

글로벌 빅테크 기업 이외의 학계와 연구계에서도 AGI의 발전 단계와 수준을 정의하고 평가하기 위한 다양한 프레임워크를 제시하고 있다. 이는 기업의 관점과는 조금 다른 보다 분석적이고 근본적인 접근을 취하는 경우가 많다. 본 절에서는 이러한 연구 결과를 몇 가지 소개하고 분석하고자 한다.

1) 호세 에르난데스-오라요(José Hernández-Orallo)의 능력 중심 접근법

스페인의 발렌시아 공과대학 교수이자 케임브리지 대학 연구원인 호세 에르난데스-오라요(2017)는 AI 평가 분야에 대한 연구를 진행하고 있다. 이들의 연구에서는 특정 과업(task) 중심의 평가에서 벗어나, 보다 일반적이고 근본적인 능력(ability)을 측정해야 한다고 주장하고 있다.

이들은 "The Measure of All Minds"에서 인간을 포함한 모든 지능체를 동일한 척도로 평가할 수 있는 보편적 지능 테스트를 제안하고 있다. 이는 특정 과업 수행 능력뿐만 아니라 학습, 추론, 적응, 그리고 지식 활용 등의 근본적인 인지 능력을 평가하는 데 중점을 두고 있다.

AI 모델의 단계 구분에 있어서 명시적인 단계를 제시하기보다는 AI의 능력을 다차원적인 인지 프로필(cognitive profile)로 표현하고자 한다. 예를 들어, 어떤 AI는 기억력은 뛰어나지만 추론 능력이 부족할 수 있고, 다른 AI는 그 반대일 수 있다. 이 프로필이 모든 영역에서 인간 수준 이상으로 균형 잡힌 상태를 AGI로 고려할 수 있다.

이러한 접근법은 "AGI가 출현했는가?"라는 단순한 질문을 넘어, "어떤 종류의 지능이, 어느 수준까지 발전했는가?"라는 더 정교한 분석을 가능하게 한다. 이를 통해 AI의 강점과 약점을 명확히 파악하여 안전성 연구와 정책 수립에 기여할 수 있다.

2) 괴르첼과 왕펑(Goertzel & Wang, 2007)의 AGI 발달 심리학 모델

AI 연구자 벤 괴르첼(Ben Goertzel)과 왕펑(Pei Wang) 등은 인간 아기의 인지 발달 과정에 영감을 받아 AGI의 발전 단계를 설명하는 모델을 제시한 바가 있다. 이는 AI가 지능을 획득하는 과정을 인간의 학습 및 성장 과정에 비유한 것이라 할 수 있다. 이들의 연구에서 AGI는 처음부터 모든 것을 아는 상태로 만들어지는 것이 아니라, 경험을 통해 점진적으로 학습하고 성장하며 더 높은 수준의 지능으로 발달한다는 관점이다.

이들이 예시적으로 제시하고 있는 AGI 발전 단계는 사람의 성장 과정과 동일하게 유아기, 아동기, 그리고 성인기의 3가지 단계로 구성되어 있다. 이 모델은 AGI의 지식뿐만 아니라 가치관과 사회성이 어떻게 형성될 수 있는지에 대한 중요한 질문을 던지고 있다. AGI가 인간 사회에 긍정적으로 통합되기 위해서는 올바른 데이터와 상호작용을 통해 사람과 마찬가지로 교육과 사회화 과정이 필수적이라는 점을 강조하고 있다.

〈표 4－6〉 AGI 발달의 심리학적 모델

단계(Level)	명칭	설명
Level 1	유아기 (Infant-like) AGI	기본적인 감각 정보를 처리하고, 간단한 패턴을 인식하며, 원초적인 반응을 보이는 단계
Level 2	아동기 (Child-like) AGI	언어를 배우고, 기본적인 상식과 개념을 습득하며, 간단한 문제 해결과 사회적 상호작용을 하는 단계
Level 3	성인기 (Adult-like) AGI	복잡하고 추상적인 추론, 장기적인 계획 수립, 전문 지식 학습 및 창의적인 작업 수행이 가능한 단계

제 2 절 AGI 발전 단계별 기술의 주요 특징 조사 분석

1. AGI 발전 단계별 주요 기술 수준과 요구 수준

AGI로의 발전은 특정한 한 두 개의 기술이 아니라 여러 개의 핵심 기술이 상호 작용하면서 복합적으로 성능이 향상되는 과정이라 할 수 있다. 본 절에서는 AGI 발전 단계별 주요 기술 분야를 6개로 나누고 현재 기술 수준을 AGI를 향한 초기 단계로 보고 향후 인간 수준의 AGI 구현을 위한 요구 수준을 비교 분석하고자 한다.

가. 추론 및 문제 해결(Reasoning and Problem Solving)

추론 및 문제 해결 기술 분야에서의 현재 기술 수준을 살펴보면 다단계 복합 추론, 귀납적 추론, 그리고 일부 계획 수립 기술이 개발되고 있다. 복잡한 문제를 해결하기 위해서 필요한 세부 단계를 논리적으로 분해하고 각각을 해결하여 주어진 문제를 해결하는 방식으로 처리한다(Wei et al., 2022). 또한 주어진 정보를 바탕으로 일반 원리나 새로운 가설을 생성하는 귀납적 추론도 가능하다. 또한, 주어진 목표 달성을 위한 구체적인 실행 계획을 수립하는 능력을 보이고 있다(Yao et al., 2023). 특정 유형의 문제 해결, 수학 및 코딩 문제 해결 능력은 우수하나 여러 도메인이 섞인 복합 문제나 상식 기반의 추론에는 취약하여 그럴듯한 오답을 생성하는 경우가 자주 발생한다.

AGI를 위해서는 기존에 없는 새로운 방식으로 문제에 접근하고 해결하는 창의적이고 전략적 사고의 개발이 필요하다. 그리고 자신의 사고 과정을 스스로 평가하고 수정하는 초인지(metacognition)도 요구된다. 그밖에 정보가 불완전한 상황에서도 최적의 판단을 내리는 것이 가능해야 한다. 초기 단계에서 창의적 아이디어의 생성은 가능할 수 있으나 그 품질을 스스로 평가하고 발전하도록 하는 초인지 능력은 매우 부족하다고 할 수 있다.

나. 멀티모달리티(Multimodality)

멀티모달 데이터의 처리 분야의 현재 기술 수준은 텍스트, 영상, 음성, 그리고 동영상 등을 동시에 이해하고 상호 간의 맥락 관계를 파악하는 데이터의 통합 이해가 어느 정도 가능하다고 할 수 있다(Radford et al., 2021). 예를 들어 영상의 시각적 내용과 배경 음악의 분위기를 함께 이해하는 것이 일부 가능하다. 이와 관련된 기술 수준은 빠르게 발전하고

있어 텍스트와 영상 사이의 상호 변환이나 영상 내용을 요약하는 기술 수준은 상당히 발전해 있다고 할 수 있다(Alayrac et al., 2022). 하지만 현재 기술 수준은 각 모달리티를 단순히 연결하는 수준이며 진정한 수준에서의 융합적 이해는 할 수 없다. 예를 들어 풍자적 밈(meme)에 있어서 시각적 정보와 내포하고 있는 문화적 맥락을 동시에 이해하는 것은 매우 어려운 수준이다.

멀티모달리티 기술 분야에서 AGI에 이르기 위해서는 기본적으로 완전한 감각의 융합이 필요하다. 인간과 마찬가지로 모든 감각 정보를 하나의 통합된 경험으로 인식하고 이를 바탕으로 세상과 상호작용 하는 것이 가능해야 한다. 하지만 이와 관련된 기술은 아직 개념적 연구 수준에 머물러 있으며 특히 로보틱스와 결합된 물리적 세계에서의 감각 융합은 극히 초보적인 단계라고 할 수 있다.

다. 자율성 및 에이전트(Autonomy and Agent)

자율성 및 에이전트 기술 분야의 현재 수준은 목표 지향적 행동 수준이라 할 수 있다. 예를 들어 "여름 휴가 계획을 세워줘"와 같은 구체적 목표를 받아서 스스로 웹을 검색하고 예약 및 결제를 하는 하위 작업을 자율적으로 수행하는 것이 가능하다. 이와 관련해 Chat-GPT 또는 Gemini와 같은 LLM 서비스가 제공되고 있으며 LLaMA나 DeepSeek 등 많은 오픈소스 프로젝트를 통해 다양한 수준에서 가능성을 제시하고 있다. 하지만 오류 발생시 대처 능력이 부족하고 장기적인 작업을 수행하는 경우 목표를 잃어 버리는 Goal Drift 경향이 발생하기 쉽다. 따라서 낮은 신뢰성이 가장 부족한 부분이라 할 수 있다.

AGI 수준에 이르기 위해 자율성 및 에이전트 기술을 장기적이고 개방적인 목표를 수행하는 수준으로 향상시켜야 한다. 즉, "내가 지금 하고 있는 사업의 시장 점유율을 높여줘"와 같은 추상적이고 장기적인 목표를 이해하고 수 개월 이상의 장기적이고 독립적인 전략을 수행할 수 있어야 한다. 관련하여 현재 기술 수준으로 구현이 불가능한 자기 개선, 장기 기억, 그리고 동기 부여 등의 복합적인 기술 개발이 요구된다.

라. 월드 모델(World Models)과 상식(Common Sense)

상식 처리 기술 분야의 현재 수준을 살펴보면 "A를 하면 B가 일어난다."와 같은 명확한 인과 관계를 학습하고 예측하는 것이 가능하다. 또한 현실 세계를 지배하는 기본적인 물리 법칙을 이해하고 시뮬레이션할 수 있는 구글 딥마인드의 지니(Ginie)와 같은 월드 모델

도 등장하기 시작하고 있다(Bruce et al., 2024). 지니는 텍스트, 이미지, 스케치 등 간단한 프롬프트만으로 사용자가 직접 조작하고 상호작용할 수 있는 가상 세계(playable world)를 즉석에서 만들어 제공하는 혁신적인 AI입니다. 기존의 동영상 생성 AI(예: Sora)가 정해진 영상을 만들어내는 데 그쳤다면, 지니는 한 걸음 더 나아가 실시간으로 사용자의 입력에 반응하는 2.5D 게임과 같은 환경을 생성합니다.

하지만, 인간 사회의 복잡한 규칙, 관계, 감정, 그리고 문화적 뉘앙스를 깊이 이해하고 행동을 결정하는 것은 매우 어렵고 이 기술 수준에 이르러야 AGI 수준이라고 할 수 있다(LeCun, 2022). Stanford 대학교의 Smallville와 같은 사회적 상호작용을 시뮬레이션하는 생성 에이전트 연구(Park et al., 2023)가 시도되고 있다. 그럼에도 불구하고 실제 인간 사회와 비교하면 그 복잡도와 일반화 측면에서 매우 큰 격차가 있다(Bender et al., 2021).

마. 학습 능력(Learning Ability)

학습 능력 기술 분야의 현재 수준을 살펴보면 적은 데이터로 새로운 기술을 빠르게 배우는 퓨샷(few-shot) 또는 원샷(one-shot) 학습 능력을 구현하기 위한 연구(Brown et al., 2020)가 진행되고 있으며 전이 학습을 통해 한 분야에서 배운 지식을 완전히 다른 분야에 효과적으로 적용하기 위한 연구도 진행 중이다. 이러한 기술은 LLM의 핵심 능력 중 하나로 활용되고 있다. 하지만, 전이 학습의 효율성이 아직도 많이 부족하여 완전히 새로운 도메인으로의 지식 전이는 매우 어려운 실정이다.

학습 능력을 AGI 수준으로 끌어 올리기 위해서는 지속적 평생 학습 기술의 개발이 요구된다. 새로운 정보를 학습하더라도 기존의 정보를 잊지 않고 지속적으로 자신을 업데이트하고 성장해야 한다. 또한 치명적 망각(catastrophic forgetting)의 문제를 해결(Kirkpatrick et al., 2017)하고 인간과 같은 유연한 학습 능력이 평생에 걸쳐 지속적으로 이루어지도록 하는 것은 AGI의 학습 모델로서 요구되는 수준이라 알 수 있다.

바. 안정성 및 가치 정렬(Safety and Alignment)

인공지능이 유해하거나 편향된 결과물을 생성하지 않도록 하는 강력한 통제 기술을 개발하고 있다. 또한 해석 가능성을 확보하기 위해 인공지능 모델의 판단 근거를 이해하기 위한 연구를 진행하고 있다. 이러한 기술로 유해 콘텐츠 필터링이나 인간의 피드백을 통한 강화 학습(Reinforcement Learning with Human Feedback, RLHF) 등을 사용(Ouyang et

al., 2022)하고 있으나 새로운 방식의 공격에는 취약하다는 문제가 있다. 또한 인공지능이 대답한 내용에 대한 이유를 근본적으로 설명하는 데 있어서 요구되는 해석 가능성 기술(Olah et al., 2020)은 아직도 초기 단계라 할 수 있다.

AGI 수준에 이르기 위해서는 기본적으로 가치의 내재화가 요구된다(Bai et al., 2022). 인간은 복잡하고 때로는 모순적인 가치를 깊이 내재화하여 어떠한 상황에서도 이익에 부합하는 판단을 내린다. 인간과 동등하거나 또는 초월하는 AGI 수준에 이르기 위해서는 이러한 가치의 내재화가 필수적이라 할 수 있다(Christian, 2020). 물론 이때의 이익은 특정 개인이나 집단의 이익이 아니라 공공의 이익에 해당해야 할 것이다. 이 과정에서 인류의 보편적 가치가 무엇인지 정의하는 문제에 제일 먼저 직면하며 이는 기술의 문제라기 보다는 철학의 문제라 할 수 있어 기술로 해결이 불가능한 매우 어려운 문제라 할 수 있다.

2. AGI 기술 수준 분석

현재 최고 수준의 인공지능(LLM)은 약 인공지능(weak AI)이라고 규정하기에는 그 수준이 조금 더 높다고 할 수 있으므로 그 한계를 벗어나 초기 AGI로 진입하는 문턱에 있다고 할 수 있다. 더 나아가 일부 능력, 특히 자연어 처리에서 해당 단계의 요구 수준에 근접하고 있다고 할 수 있다. 앞에서 살펴본 결과에 따르면 AGI로의 도약을 가로막는 가장 큰 기술적 장벽은 상식의 문제와 자율성의 문제라 할 수 있다. 단순히 데이터의 패턴을 학습하는 것을 넘어 세상의 인과 관계를 이해하고 스스로 목표를 설정하고 행동하는 능력을 구현하는 것은 여전히 미지의 영역이라 할 수 있다.

이러한 기술이 개발되고 목표로 하는 수준이 달성된다고 하더라도 안전성 및 가치 정렬 연구의 중요성은 기하급수적으로 증가한다고 할 수 있다. 만약 기술 개발과 안전성의 연구가 불균형을 이루는 경우 통제 불가능한 위험을 초래할 수 있다. 따라서 인공지능 기술 수준의 고도화와 안전성을 확보하기 위한 기술은 반드시 함께 발전해야 한다.

본 절에서는 앞에서 살펴본 6개 기술 분야에 대한 기술 수준 현황과 핵심 격차에 대해 심층 분석을 수행하고자 한다.

가. 규모의 힘에 기반한 현재의 발전과 그늘

현재 AI 기술 발전의 핵심 동력은 규모에 기인한다고 할 수 있다. 수십억 개의 파라미터

를 가진 거대 모델을 수 페타바이트(PB)의 데이터로 학습시키는 방식은 자연어 처리(NLP)와 이미지-텍스트 통합 분석 및 생성 분야에서 괄목할 만한 성과를 이어 내고 있다. 이는 인간 수준에 근접한 유창한 언어 구사 능력, 코딩 보조, 이미지 생성 등에서 높은 성능을 보이며 생산성 향상에 기여하고 있다는 측면에서 살펴본다면 초기 AGI 단계에서 요구하는 수준의 일부를 만족시킨다고 할 수 있다.

이러한 우수한 능력을 다른 측면에서 살펴본다면 데이터에 내재된 패턴을 모방하고 보간(interpolation)하는 것에 가깝다고 할 수 있다. 주어진 문제가 데이터에 없거나 매우 드문 상황, 즉 외삽(extrapolation)이 필요한 새로운 문제에 직면했을 때 급격히 성능이 저하된다. 그럴듯한 거짓말을 하는 환각(hallucination) 현상은 이러한 통계적 학습의 근본적 한계를 명확히 보여주는 증거라고 할 수 있다.

나. 가장 큰 장벽: 추론과 상식의 부재

인간 아기는 몇 번의 경험만으로도 세상의 기본적인 물리 법칙(예: 물체는 중력에 의해 아래로 떨어진다)과 인과 관계를 학습한다. 하지만 현재의 AI는 수억 개의 문서를 읽고도 이러한 직관적 이해, 즉, 세계 모델(world model)과 상식(common sense)을 내재화하는 데 실패하고 있다.

AI는 "달걀을 깨면 노른자가 나온다"는 것을 텍스트적으로 알지만, '깨다'라는 행위의 물리적 결과와 그 불가역성을 진정으로 이해하지 못한다. 이로 인해 복합적인 문제 해결에 필수적인 다단계 추론, 귀납적 가설 설정, 창의적 문제 해결 능력이 현저히 떨어진다. 이는 AGI로 가는 길에 놓인 가장 높고 두꺼운 장벽이라 할 수 있다.

이 문제의 해결은 단순히 데이터 양을 늘리는 것으로는 불가능하며, 인과 추론(causal inference), 기호 주의와 연결 주의를 결합한 뉴로-심볼릭(neuro-symbolic) AI, 더 나아가 새로운 모델 아키텍처 등 근본적인 R&D 패러다임 전환이 요구된다.

다. 미완의 혁명: 자율적 에이전트로의 도약 과제

AGI의 궁극적인 모습 중 하나는 인간의 지시를 받아 자율적으로 목표를 달성하는 AI 에이전트이다. 최근 Auto-GPT, LLaMA, DeepSeek 등 오픈 소스 형태의 LLM과 여러 글로벌 빅테크의 상용 서비스 등 여러 시도가 진행되고 있으나, 신뢰성과 안전성 측면에서 아직도 부족한 점이 많다.

현재의 에이전트는 명확하고 단순한 단기 과업에서는 가능성을 보이지만, 복잡하고 장기적인 목표 수행 시 쉽게 길을 잃는다. 우리는 이를 goal drift라고도 부른다. 이러한 문제가 발생하는 것은 앞서 언급한 추론 및 상식 능력의 부재에 기인한다. 외부 환경의 예기치 않은 변화에 대응하고, 실패했을 때 스스로 원인을 분석하여 계획을 수정하는 능력이 없기 때문이다.

신뢰 가능한 AI 에이전트 기술은 향후 디지털 경제의 게임 체인저가 될 것으로 기대된다. 이 기술의 확보는 단순히 API를 연결하는 수준을 넘어, 견고한 장기 기억(long-term memory), 자기 성찰(self-reflection), 그리고 강건한 계획 수립 능력의 확보에 달려있다고 할 수 있다.

라. 영원한 숙제: 안전성 기술의 더딘 발전

AI의 능력이 강력해져 인간의 능력에 필적하거나 또는 더 나아가 초월하는 상황에 이를수록 인간이 AI를 통제하지 못할 때 발생하는 위험은 기하급수적으로 커진다. 하지만 안타깝게도 AI의 능력(capability) 발전 속도는 안전성 및 통제(safety & control) 기술의 발전 속도를 크게 앞지르고 있다.

현재의 AI 안전 기술이라 할 수 있는 유해성 필터나 RLHF 등은 이미 알려진 위험을 막는 사후적 또는 수동적 대응에 가깝다. 또한 AI가 왜 그런 결정을 내렸는지 근본적으로 설명하는 해석 가능성(interpretability) 기술은 아직 초기 단계이며, AI가 인간의 복잡한 가치를 내재화하여 스스로 올바른 판단을 내리게 하는 가치 정렬(value alignment) 문제는 여전히 해결되지 않은 난제이다.

AGI 단계로 발전할수록 이 기술 격차는 심각한 존재론적 위험으로 이어질 수 있다. 따라서 안전성 연구는 더 이상 부가적인 요소가 아닌, 기술 개발의 핵심 전제 조건이 되어야 한다.

제 3 절 AGI 발전 단계별 기술 수준 평가 지표 및 시각화 방법

AGI는 특정 작업에 국한되지 않고 인간처럼 광범위한 영역에서 학습, 이해, 적용, 그리고 문제 해결 등을 수행할 수 있는 지능을 의미한다. 초기 인공지능 기술에서 AGI에 이르기까지 발전 단계를 학계와 산업계의 여러 기관에서 제시하고 있고 이에 대한 조사 분석 결과를 이전 절에서 제시하고 있다. 본 절에서는 먼저 여러 기관에서 제시하고 있는 발전 단계를 종합 분석한 결과를 바탕으로 공통의 발전 단계를 제시하고 이 발전 단계를 기반으로 기술 수준을 평가하기 위한 지표를 제시하고 시각화하기 위한 방법을 살펴본다.

1. AGI 발전 단계별 주요 기술 수준 평가 지표

가. AGI 발전 단계의 정의

본 보고서에서는 구글 딥마인드와 OpenAI를 비롯하여 여러 학계와 산업계에서 제안한 AGI 기술 로드맵 및 발전 단계를 바탕으로 공통적인 부분을 모으고 단순화하여 AGI 발전을 4개 수준(Level 0-3)으로 단순화하여 정의한다. 본 문서에서 정의하고 있는 AGI 발전 단계는 〈표 4-7〉에서 보여주고 있다.

〈표 4-7〉 본 문서에서 고려하는 AGI 발전 단계

단계(Level)	특징	설명
Level 0	특화된 영역의 좁은 AI	특정 영역의 작업에서 높은 성능을 보이지만 영역 전이나 새로운 문제 해결에 대한 능력은 제한적
Level 1	멀티모달 멀티태스크 AI	다수의 서로 다른 작업을 동일 모델로 수행할 수 있으며 적은 수의 샘플로 제한적 맥락 전이가 가능
Level 2	인간과 대등한 수준의 AGI	넓은 범위의 인지·추론·계획·사회적 상호작용에서 평균적 인간과 대등한 수준으로 새로운 환경에 스스로 학습
Level 3	초지능 (Superintelligence)	거의 모든 지적 과제에서 인간 최고 전문가를 상당히 능가하는 수준

먼저 가장 낮은 단계인 Level 0는 딥러닝의 등장과 함께 오늘날 인공지능이 가장 성공적인 결과를 보여주고 있는 단계라 할 수 있다. 이 단계를 좁은 AI(Narrow AI) 또는 특화 모델이라 부를 수 있다. 이 단계에서는 특정한 주어진 영역의 문제 해결에서는 아주 우수한 능력을 보이고 상당한 문제 영역에서 사람의 능력을 뛰어넘는 수준의 결과를 보여주고 있다. 이 단계의 응용 사례로는 영상을 분류(Krizhevsky et al., 2012)하거나 바둑 문제 해결(Silver et al., 2017)하거나 아니면 특정한 영역에 특화된 언어 모델(Radford et al., 2019) 등을 고려할 수 있다. 하지만 특정한 문제 영역에서 매우 우수한 성능을 보이는 모델을 다른 문제 영역으로 전이하거나 새로운 문제 해결에 적용했을 때의 능력은 매우 제한적이라 할 수 있다.

두 번째 단계인 Level 1은 멀티 모달 데이터를 기반으로 멀티 태스크를 처리할 수 있는 수준으로 이전 단계에 비해 범용성이 강화된 단계라 할 수 있다. 이 단계에서는 언어, 비전, 프로그램밍 등을 하나의 모델로 수행할 수 있다. 또한 제한적인 맥락 전이를 적은 수의 샘플로 빠르게 적용할 수 있다. 이 단계는 ARC-AGI(Abstraction and Reasoning Corpus for AGI)(Collet, 2019; ARC-AGI, 2019)등과 같은 고난도 일반화 테스트에서 의미있는 성과를 보이는 수준이라 할 수 있다. ARC-AGI는 기계가 학습된 데이터나 패턴을 단순히 반복하지 않고 완전히 새로운 상황에서도 인간과 유사하게 추론과 추상화를 통해 문제를 해결할 수 있는지를 평가하기 위한 목적으로 개발된 벤치마크이다.

세 번째 단계인 Level 2는 인간 수준과 거의 대등한 범용성을 인공지능이 가지는 단계로서 Human-level AGI 단계라 할 수 있다. 이 단계에서는 넓은 범위의 인지, 추론, 계획, 그리고 사회적 상호작용에서 평균적인 인간과 대등하거나 일부 상회하는 성과를 보인다. 또한 새로운 환경에 대해 스스로 탐색하고 학습하면서 능력을 확장할 수 있다. 이 단계에 이르면 인공지능의 안전 문제 또는 정책적 리스크가 실질적으로 대두되는 단계이다(Morris et al., 2023)

마지막 네 번째 단계인 Level 3은 초지능 단계로 거의 모든 지적 과제에서 인간 최고 전문가의 수준을 상당히 능가하는 단계라 할 수 있다. 이 수준의 인공지능은 속도·효율·창의성 면에서 질적 도약을 보여 사회적·안전적 영향이 매우 크다. 닉 보스트롬(Bostrom 2016)은 자신의 저서 초지능(Superintelligence)에서 이 단계의 도래로 인해 존재론적 위험을 수반할 수 있음을 경고하고 있다.

나. AGI 발전 단계별 평가 지표

앞에서 정의한 AGI 발전 단계별로 해당 단계에 기술 수준이 이르렀는지 객관적으로 평가하기 위한 평가 지표에 대한 고찰이 필요하다. 여기서는 몇 가지 평가 지표와 이를 적용하기 위한 방법론을 살펴보고자 한다.

본 보고서에서는 AGI 발전 단계 평가를 위한 핵심 지표로 〈표 4-8〉에서 보여주고 있는 성능 지표, 전이 지표, 일반화 지표, 메타 학습 지표, 그리고 인지 및 자기 인식 지표를 포함하는 5가지를 제안하고 있다. 이들 지표에 대한 상세한 설명은 아래와 같다.

〈표 4-8〉 AGI 발전 단계별 평가 지표

구분	주요 평가 내용	대표 벤치마크	AGI 단계별 중요도
성능 지표	단일 과제 정확도, 효율성	ImageNet, GLUE, MMLU	Level 0-1
전이 지표	학습된 지식의 재활용	CLIP, ARC, XNLI	Level 1-2
일반화 지표	새로운 규칙 또는 개념에 대한 적응	ARC-AGI, SCAN, CLEVR	Level 2-3
메타학습 지표	학습 속도 및 효율성	MAML, Meta-Dataset	Level 3-4
인지 및 자기 인식 지표	불확실성 판단 및 자기 검증	Calibration, ToM Test	Level 4+

먼저, 성능 지표(Performance Metrics)는 특정 과제(Task)에 대해 인공지능 모델이 달성하는 정확도, 효율성, 그리고 일관성 등을 정량적으로 측정하는 가장 기본적인 지표이다. 이 지표는 특히 AGI 발전 단계에서 Level 0에 해당하는 특화된 영역의 좁은 인공지능 모델의 발전 수준을 평가하는 핵심 지표라 할 수 있다. AGI 발전에 있어 상위 단계로 갈수록 단일 과제의 점수보다는 다중 과제 전반에서의 평균 성능이나 성능 편차의 안정성이 더욱 중요해진다. 성능 지표의 대표적인 예시로 다음과 같은 4가지를 고려할 수 있다.

- ImageNet Top-1 Accuracy: 시각 모델(Deng et al., 2009; Krizhevsky et al., 2012)
- GLUE/SuperGLUE Score: 언어 이해(Wang et al., 2019; 2020)
- MMLU(Massive Multitask Language Understanding): 57개 학문 분야에 걸친 다중 과제 평가(Hendrycks et al., 2021)

- Big-Bench(Beyond the Imitation Game Benchmark): 인간적 추론 능력 포함(Srivastava et al., 2022)

이상과 같은 성능 지표는 모델의 정확성을 가장 잘 보여주는 지표이지만 모델의 이해, 추론, 그리고 일반화 능력과 같은 높은 차원의 지능은 반영하지 못하는 한계를 가진다.

두 번째로 전이 지표(Transfer Metrics)는 모델이 새로운 문제 영역이나 환경으로 전이될 때 성능의 저하를 얼마나 최소화하는가를 평가하는 지표이다. 이 지표는 학습된 지식을 재활용하는 능력을 측정한다는 점에서 AGI를 구분하는 필수 지표라 할 수 있다. 전이 지표는 하나의 데이터셋으로 학습한 모델을 유사하나 다른 데이터셋이나 언어 또는 환경에서 테스트하고 성능이 얼마나 잘 유지되는지를 다음(식 4-1)과 같은 전이 효율(Te; Transfer Efficiency)로 정의한다.

$$T_e = \frac{P_{new}}{P_{prev}} \times 100\% \tag{4-1}$$

예를 들어 원래 주어진 문제에 대한 모델의 정확도가 90%인데 새로운 문제 영역에 대해서는 정확도가 72%라면 전이 효율은 80%로 주어진다. 전이 지표의 대표적인 예시로 다음과 같은 3가지를 고려할 수 있다.

- CLIP(OpenAI): 이미지-텍스트 전이 평가(Radford et al., 2021; Brown et al., 2020)
- Zero-shot MMLU, Cross-lingual NLI(XNLI): 다언어 전이 테스트 Cross-lingual Evaluation(Conneau et al., 2018 – XNLI)
- ARC(Abstraction and Reasoning Corpus): 추상 규칙을 새로운 문제에 전이(Chollet, 2019)

전이 지표는 모델의 적응력과 지식 재사용 능력을 반영하는 지표로서 단일 과제 중심의 평가만으로는 측정이 어려운 AGI가 가지는 특성의 핵심 요소라 할 수 있다.

다음 3번째 지표는 일반화 지표(Generalization Metrics)이다. 이 지표는 학습 데이터에 존재하지 않는 완전히 새로운 상황이나 규칙에 대해 올바른 판단을 내릴 수 있는지를 평가한다. 이를 통해 단순한 데이터 전이가 아니라 개념적이고 논리적인 일반화 능력을 측

정한다. 일반화 지표의 대표 사례로는 다음 3가지를 고려할 수 있다.

- ARC-AGI(Abstraction and Reasoning Corpus for AGI): 다양한 패턴과 추상적 규칙을 학습해야 하는 고난도 테스트로 인간은 대체로 80~90% 수준이고 현존하는 AI 모델은 30% 이하의 성능을 보임(Chollet, 2019)
- CLEVR / ShapeWorld: 시각적 추론 및 조합적 일반화 평가(Johnson et al., 2017)
- SCAN / COGS(Compositional Generalization): 문법적 조합의 일반화 능력 측정(Lake & Baroni, 2018; Kim & Linzen, 2020)

메타학습 지표를 네 번째 평가 지표로 고려할 수 있다. 메타학습 지표는 모델이 새로운 과제를 학습하는 속도와 효율성을 평가한다. 인간은 새로운 환경에 대해 학습하는 법을 배우는 능력이 있지만, 대부분의 인공지능 모델은 이 기능이 제한적이다. 메타학습 지표는 Few-shot 또는 One-shot learning 실험을 통해 일반화 가능 여부를 평가하고 Online Adaptation Speed를 통해 추가 데이터를 주었을 때 학습효율이 얼마나 향상되는지를 측정한다. 메타학습 지표의 대표 사례로 다음 3가지를 고려할 수 있다.

- Meta-Dataset(Triantafillou et al., 2019)
- LEAF Benchmark: Federated + Meta-learning 평가(Caldas et al., 2019)
- Model-Agnostic Meta-Learning(MAML) 기반 테스트(Finn et al., 2017)

마지막 5번째 지표는 인지 및 자기 인식 지표(Cognitive & Self-awareness Metrics)이다. 이 지표를 통해 모델이 스스로의 지식 상태, 불확실성, 그리고 한계를 조정할 수 있는지를 평가한다. 이 단계에 이르면 AGI 모델에서 의식 또는 자기 참조적 사고의 초기 형태가 가능하다고 고려할 수 있다. 이 지표의 측정 요소로는 답변에 대한 신뢰도 예측 정확도를 통해 모델의 자기 평가(self-consistency)를 할 수 있으며, 모델의 확신(confidence)과 실제 정답률 간의 일치 정도를 통해 불확실성에 대한 보정(uncertainty calibration)을 할 수 있다. 또한 타인의 믿음이나 의도를 추론할 수 있는지에 대한 심리학 기반의 평가(Theory of Mins Test)도 수행한다. 인지 및 자기 인식 지표의 대표 연구 사례로 다음을 고려할 수 있다.

- Self-consistency / Calibration Error(Jiang et al., 2021)

– Theory of Mind Test(Kosinski, 2023)

– Reflection-based Reasoning(OpenAI, 2023)

인공지능 모델이 아직 이 단계에 완전히 이르지 못했지만 GPT-4부터는 스스로의 답변 불확실성을 추정하거나 잘못된 판단을 교정하려는 시도를 통해 부분적인 자기 인식 능력을 보인다. 하지만, 이 능력은 아직 완전한 자각(self-awareness)에는 이르지 못하고 있다.

2. AGI 평가 지표를 위한 시각화 방법

가. AGI 평가 지표를 위한 시각화 목적

앞에서 살펴본 AGI 평가를 위한 정량적 지표(성능, 전이, 일반화, 메타학습, 그리고 인지/자기 인식)는 다차원적이며 서로 다른 스케일과 해석을 요구한다. 따라서 효과적인 시각화는 다중 지표를 한눈에 비교할 수 있게 해주고, 지표 간 트레이드 오프와 분포(불확실성 및 편차)를 드러내며, 시계열 학습 곡선과 자원(비용) 요소를 동시에 표현할 수 있어야 한다. 본 보고서는 각각의 시각화 기법의 원리와 적용 가능성, 그리고 장단점 등을 정리한다.

나. AGI 평가 지표별 시각화 방법

성능 지표에 대한 시각화는 주로 혼동 행렬(Confusion Matrix), ROC 및 PR 곡선, 그리고 Top-k accuracy를 보여주는 Bar chart 등을 고려할 수 있다. 성능 지표를 시각화하는데 있어서 고려해야 할 점은 단일 점수(Top-1)만 시각화하는 것보다 클래스별 오류나 임계값 민감도, 그리고 표본 간 분산을 함께 제시하는 것이 바람직할 수 있다.

다음으로 전이 지표의 시각화에는 전이 매트릭스(Transfer Matrix)의 히트맵(heatmap)을 통한 시각화와 domain-to-domain success rate의 flow를 함께 고려할 수 있다. 전이 지표를 적절히 시각화하기 위해서는 어떤 소스 영역이 어떤 대상에는 잘 전이되고 어떤 대상에는 잘 전이되지 않는지를 방향성과 비율로 보여주는 것이 적절하다. 그밖에 문제 인스턴스 전반에 걸친 일관성을 나타내는 Performance Profile을 통해 다양한 인스턴스에 대해 어느 정도 비율로 좋은 성능을 얻을 수 있는지 시각화하는 것도 고려할 필요가 있다.

세 번째 지표인 일반화 지표는 ARC-스타일 문제별 성공률 히트맵과 문제 특성별 성능 패턴을 통해 시각화하는 것이 적절하다. 그밖에 요약 프로파일을 제시하는 Radar chart와

성능 분포를 나타내는 박스(Box) 또는 바이올린(Violin) 플롯(plot)을 사용한다. 일반화 지표의 시각화에는 추상화나 조합성과 같은 문제 유형별로 성능이 어떻게 변화하는지를 시각적으로 분리하고 비교할 필요가 있다. 특히 실패 사례는 샘플의 시각화를 통해 보강하면 더욱 효과적인 시각화가 가능하다.

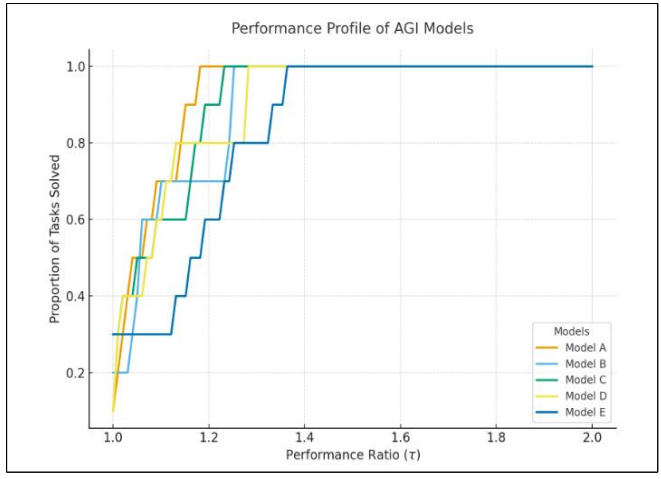
다음 메타학습 지표는 학습 곡선(Learning curve)과 여러가지 few-shot 조건에 대해 시각화하는 것이 적절하다. 또한 박스 플롯이나 바이올린 플롯을 통해 적응 속도의 분포를 제시할 수 있다. 그밖에 초기 샘플 수, 최종 성능, 수행 시간 등을 다차원 공간에서 표현할 수도 있다. 메타학습이 적절하게 이루어지면 샘플 수에 대비하여 성능 향상의 기울기를 통해 학습 효율을 나타낼 수 있다. 특히 초기 몇 개의 샘플에서 성능 개선을 명확하게 제시할 수 있다.

마지막으로 인지 및 자기 학습 지표는 Reliability diagram과 Calibration histogram으로 시각화할 수 있다. 그밖에 confidence와 correctness를 나타내는 Scatter plot으로 시각화할 수도 있고 혼동 행렬을 사용할 수도 있다. 이 지표의 시각화에는 모델이 얼마나 신뢰성이 있는지를 적절히 보여주어야 하며 적절한 보정 작업이 이루어졌는지 전후 비교를 병행하여 개선 효과를 명확하게 제시할 필요가 있다.

다. AGI 시각화 예시

Performance profile의 시각화 예시는 다음과 같다. 이 그래프에서 x-축은 performance ratio로 각 모델의 성능이 최고 성능 대비 몇 배까지 허용되는지 나타낸다. y-축은 주어진 performance ratio 이하의 성능비를 보이는 태스크의 비율을 나타낸다. 이 곡선이 왼쪽 위로 올라갈수록 더 많은 태스크에서 최고 수준의 성능을 유지하는 모델이라 할 수 있고 곡선이 오른쪽 아래에 위치할수록 성능 편차가 큰 모델임을 나타낸다. 이 시각화 방법은 여러 모델이나 AGI 단계의 상대적 효율성이나 일관성을 비교하는데 효과적이다.

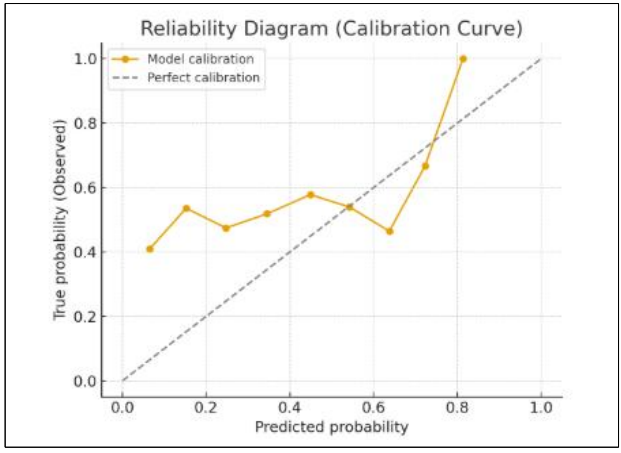
〔그림 4－2〕 AGI 모델별 Performance Profile의 시각화 예시



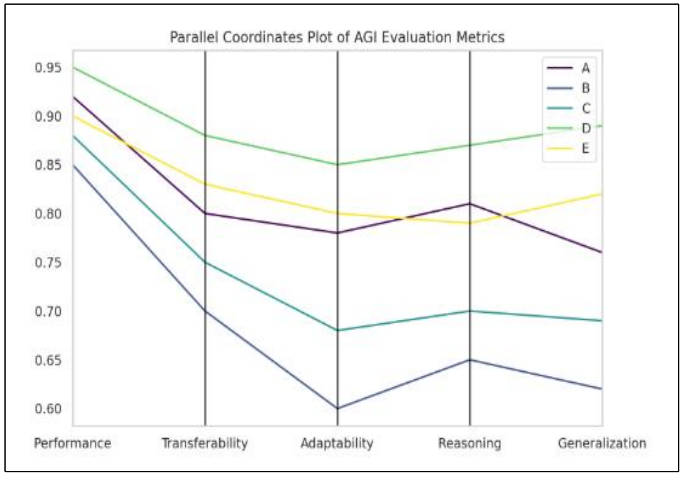
Reliability diagram의 시각화 예시는 〔그림 4－3〕과 같다. 이 그래프는 캘리브레이션 곡선이라고 부르며 예측 확률과 실제 정답 간의 정합도를 보여준다. 그림에서 점선으로 표시된 대각선은 완벽한 보정을 나타내며 실선으로 표시된 선이 실제 정답 확률을 보여준다. 실선이 점선에 가까울수록 모델의 확률 예측을 신뢰할 수 있다고 생각할 수 있다.

다음으로 평행 좌표계 시각화로 불리는 Parallel Coordinates Plot은 AGI에 대한 여러가지 평가 지표를 한눈에 비교할 수 있도록 제공하는 다차원 시각화 방법이다. 여기서 각각의 선은 개별 모델을 의미하며 각각의 축마다 해당 모델의 상대적 지표값을 연결하여 균형적 강점과 약점 패턴을 확인하도록 한다. 우수한 성능의 모델은 플롯의 상단에 위치하며 상대적으로 덜 우수한 모델은 하단에 위치한다. 〔그림 4－4〕에서 Parallel Coordinates Plot의 예시를 보여주고 있다.

〔그림 4－3〕 AGI 모델의 Reliability diagram 시각화 예시



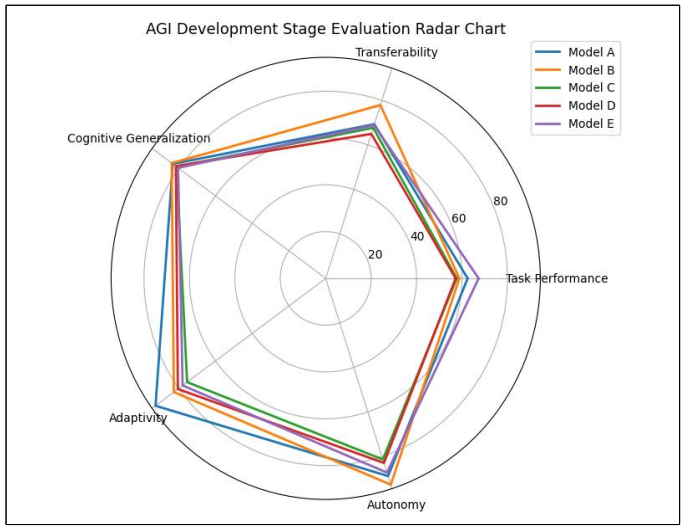
〔그림 4－4〕 AGI 모델에 대한 Parallel Coordinates Plot 시각화 예시



마지막으로 Radar chart의 예시를 〔그림 4－5〕에서 보여주고 있다. 이 방법을 통해 여러 AGI 모델에 대해 성능 지표별로 어떤 차이를 보이는지 확인할 수 있다. 예를 들어 어떤 모델은 특정한 성능 지표에서는 아주 우수한 결과를 보이지만 다른 지표에 대해서는 저조하다는 것을 한눈에 알 수 있다. 반면에 어떤 모델은 개별 성능 지표 중 아주 우수한 결과를

보이지는 않지만 전반적으로 모든 영역의 성능 지표에 대해 우수한 결과를 가진다는 것을 한눈에 확인할 수 있다.

〔그림 4－5〕 AGI 모델에 대한 Radar Chart 시각화 예시



제5장 AGI 기술 발전 단계별 기술적·산업적 촉진 요인 식별

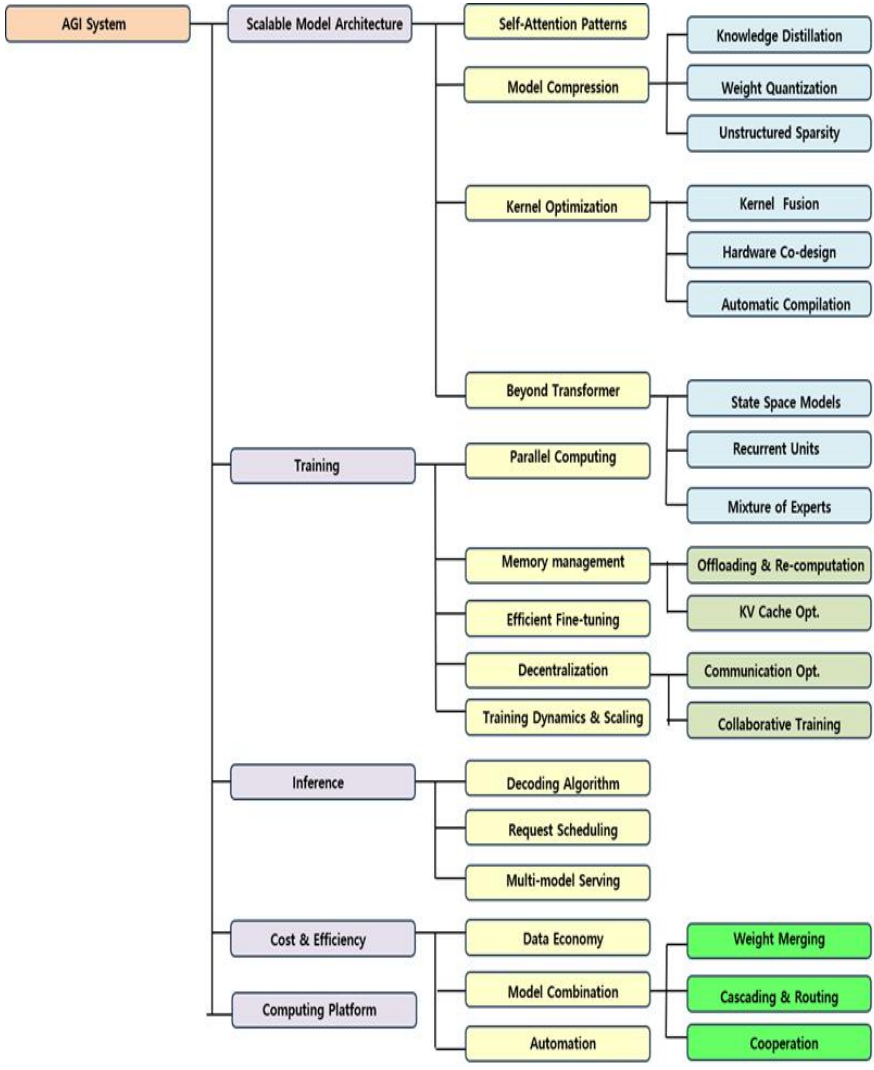
Llama 2(Touvron et al., 2023), GPT-4(OpenAI, 2023a), Gemini(Team et al., 2023), Claude 3(AI, 2024), Mistral(Jiang et al. 2023d)와 같은 여러 대형 모델에서 나타나는 새로운 창발적인 현상(Wei et al., 2022a)은 모델의 매개변수 수가 일정 수준까지 증가하면서 나타나며, 다양한 시스템에서 LLM의 효율성을 유지하며 확대되는 기본적인 요소들은 1) 계산 및 모델링을 근본적으로 그리고 알고리즘적으로 정의하는 확장 가능한 모델 아키텍처, 2) 대규모 학습 기술은 지리적으로 분산될 수 있는 더 많은 하드웨어 가속기의 활용을 최적화, 3) 여러 모델에 대한 안정적이고 높은 처리량을 보장하는 추론 인프라, 4) 데이터, 모델 조합 및 자동화 프로세스를 훨씬 더 효율적으로 만드는 다양한 방법론의 비용 및 효율성, 5) 소프트한 물리적 제약을 극복하고 차세대 계산 기능과 미래 알고리즘 혁신을 위한 하드웨어 기반을 제공하는 하드웨어 컴퓨팅 플랫폼 측면 등이다.

시스템 연구의 발전은 이러한 요소들의 확장성을 촉진하는 데 필수적이며, 이러한 추세는 인공 일반 지능으로 나아가는 과정에서 여전히 중추적인 역할을 할 것으로 예상된다.

AI 시스템 연구 및 엔지니어링의 지속적인 발전을 통해, 클라우드 전반에 걸쳐 1만 개가 넘는 이기종 가속기에서 모델이 학습되는 것을 상상할 수 있다. 이를 통해 여러 에이전트를 하나의 다기능 융합 시스템으로 연결할 뿐만 아니라, 일상생활에서 사람들에게 즉각적이고 개인적인 지원을 제공할 수 있다.

이 장에서는 AGI 시스템이 해결해야 할 주요 과제부터 시작하여 모델 아키텍처 혁신, 학습, 추론, 비용 절감 및 컴퓨팅 플랫폼에 대한 노력을 소개한다.

〔그림 5－1〕 Taxonomy of Current AGI Systems



제 1 절 AGI System의 속성과 과제

AGI 시스템이 갖추어야 할 주요 속성과 과제는 아래와 같이 7가지 개념으로 정리할 수 있다.

1. 대용량 학습 데이터(The Large Amount of Training Data)

대형 모델은 친칠라 최적성(Chinchilla optimality)을 달성하기 위해 방대한 학습 데이터가 필요하다(Hoffmann et al., 2022). 동시에, 생성 모델과 사용자 콘텐츠 제작의 엄청난 성공으로 인해 인터넷에서 이용 가능한 데이터의 양은 급증할 가능성이 높지만, 평균적인 품질과 진위성은 그만큼 빠르게 개선되지 않을 수 있다. 따라서 효율적인 학습을 위해 다양한 소스에서 데이터를 선택, 구조화, 정리 및 혼합할 수 있는 더욱 정교하고 자동화된 데이터 처리 파이프라인이 필요하다.

2. 반복의 속도와 비용(The Speed and Cost of Iteration)

대형 모델 학습의 각 반복은 프로토타입 제작 및 실험 중에 막대한 리소스와 시간을 소모할 수 있다. 실제로는 인적 오류 및 시스템 오류와 같은 여러 가지 다른 간섭이 학습 선점을 초래할 수 있다. 자동화된(하이퍼파라미터 및 아키텍처) 검색 파이프라인과 잘 설계된 학습 인프라(Shoeybi et al., 2020; Aminabadi et al., 2022)는 반복 비용을 획기적으로 줄이고 모델 개발 속도를 암묵적으로 향상시킬 수 있다.

3. 개인정보 보호가 중요하고 자원이 제한된 환경(Privacy-sensitive and Resource-constraint Settings)

현재 가장 성공적인 대규모 모델은 사용자 요청이 중앙 집중식으로 처리되는 데이터 센터에 배포되어 있지만(Achiam et al., 2023; AI, 2024), 데이터와 쿼리가 로컬에서 사용되고 처리되는 에지에서 모델을 제공해야 할 필요성은 개인정보 보호나 지연 시간에 더 민감한 상황에서 더욱 중요해질 것이다. 그러나 에지 디바이스는 일반적으로 컴퓨팅 및 메모리

측면에서 성능이 떨어지므로, 자원 제약 환경에서 활용도를 최적화하면서도 모델 성능을 저하시키지 않는 기술 개발이 필요해진다.

4. 미세 조정 및 적응을 위한 효율적인 방법(Efficient Methods for Fine-tuning and Adaptation)

사전 훈련된 모델을 작업별 데이터에 대하여 미세 조정하는 것이 가장 널리 사용되는 패러다임이다. 그럼에도 불구하고 전체 모델 가중치 업데이트에 필요한 컴퓨팅 자원과 시간은 여전히 많은 사용자에게 부담스러운 수준이다. 효율적인 미세 조정 방법은 도메인 적응, 에이전트 학습, 작업별 최적화에 대한 진입 장벽을 낮추는 데 도움이 된다.

5. 서비스 지연 시간 및 처리량(Serving Latency and Throughput)

AGI 시스템은 원활한 사용자 경험과 참여를 위해 낮은 지연 시간과 높은 처리량을 지원해야 한다. 그러나 현재 시스템은 배치 처리 최적화, 첫 번째 토큰까지의 시간 또는 단일 쿼리 완료 시간과 같은 다른 요소들 사이에서 균형을 이루는 경우가 많다(Yu et al., 2022; Aminabadi et al., 2022). 이러한 모든 지표 간의 균형을 맞추는 것은 어려운 문제이다.

6. 메모리 사용량(Memory Footprint)

대규모 모델을 배포할 때 중요한 과제 중 하나는 메모리 사용량인데, 이는 셀프 어텐션의 이차적 특성으로 인해 긴 컨텍스트 및 다중 모드 입력의 경우 더욱 심각해진다. KV 캐시는 더 빠른 추론을 위해 메모리를 희생하는 일반적인 기법이지만(Pope et al., 2022), 적절히 처리하지 않으면 상당한 메모리 부담을 초래할 수 있다.

7. 하드웨어 호환성 및 가속(Hardware Compatibility and Acceleration)

모델 제공 성능은 엔지니어가 하드웨어의 기능을 얼마나 잘 활용할 수 있는지에 크게 좌우된다. 다양한 가속기 아키텍처에 맞게 설계된 특수 커널과 알고리즘은 추론 속도를 크게 향상시킬 수 있다. 이기종 장치와의 호환성을 확보하고 통일된 소프트웨어 추상화를 구축함으로써 대규모 모델의 잠재력을 완전히 발휘할 수 있다.

제 2 절 확장 가능한 모델 아키텍처(Scalable Model Architectures)

시스템 연구자와 엔지니어가 끊임없이 고민하는 주제 중 하나는 더 크고 강력한 모델을 만드는 방법이다. 하지만 확장을 위해서는 여러 가지 요소를 고려해야 한다. 매개변수 수와 학습 데이터 양(친칠라 스케일링 법칙(Hoffmann et al., 2022), 그리고 유효 컨텍스트 길이와 서빙 용량(컨텍스트 스케일링 법칙(Xiong et al., 2023)이 고려 사항이다. 이를 위해, 모델 아키텍처 수준에서 기존 연구의 최적화 기법을 도입하는 것부터 시작하여, 모델에 더 독립적인 학습 및 추론 인프라 구축이 필요하다.

1. 셀프 어텐션 패턴(Self-attention Patterns)

트랜스포머 아키텍처의 핵심 메커니즘인 기본 셀프 어텐션(Vaswani et al., 2017)은 입력 시퀀스 길이에 따라 계산량이 이차적으로 증가하여 긴 컨텍스트에 대한 하드웨어 메모리를 제한한다. 이를 극복하기 위해 계산 부하와 메모리 요구량을 줄이기 위한 다양한 주의 마스크 패턴이 개발되었다. 국소적 어텐션으로 제한하거나 해상도를 낮추는 방식으로, 성능 저하를 감수하면서 효율성을 개선하기 위한 슬라이딩 윈도우(Beltagy et al., 2020; Jiang et al., 2023d), 효율성과 넓은 범위 커버리지를 결합한 확장 패턴(Ding et al., 2023b) 등이 있다.

2. 모델 압축(Model Compression)

모델 압축의 목표는 대규모 모델의 메모리 사용량, 계산 복잡도 및 배포 비용을 줄이는 것이다. 여러 접근법 중 지식 증류(Knowledge Distillation, KD)(Hinton et al., 2015; Hsieh et al., 2023)는 더 큰 교사 모델이 학습한 지식을 추출하여 더 작은 학생 모델로 이전하는 과정으로, 종종 더 효율적으로 서비스를 제공할 수 있다. 더 크고 성능이 좋은 '교사(teacher)' 모델의 지식을 더 작고 효율적인 '학생(student)' 모델로 전달하는 방법으로는 교사 모델의 API를 호출하여 입력/출력 쌍을 수집하고 이를 학생 모델 훈련에 사용하는 블랙 박스 KDAlpaca(Taori et al., 2023), Vicuna(Peng et al., 2023b), 교사 모델의 가중치나 손실 값 같은 더 많은 정보에 접근하여 더 나은 지식 전달을 유도하는 화이트 박스 KDMiniLLM(Gu

et al., 2023), 학생 모델이 자체 생성한 출력 시퀀스를 활용하고 교사 모델로부터 피드백을 받아 전통적인 지도 학습 KD의 한계를 극복하는 일반화된 KD(Generalized KD, GKD) (Agarwal et al., 2024) 등이 있다. 또한 모델의 희소성(sparsity)을 활용하여 추론 시 모델 가중치의 일부를 제거하는 가지치기(Pruning)와 역방향 학습(reverse-learning) 또는 초정렬(super-alignment)의 가능성을 탐구하여 더 약한 모델들로부터 지식을 추출하고(예: 분석, 병합, 적응적 업데이트를 통해) 현재 모델을 개선하는 방법도 있다.

3. 커널 최적화(Kernel Optimization)

커널은 대규모 네트워크에서 일반 행렬 곱셈과 같은 기본 연산을 가속화하기 위해 개발된다. 커널 융합은 두 개 이상의 커널을 하나로 병합하는 기술로, 커널 실행 오버헤드와 중복 메모리 접근을 줄일 수 있다. 이는 LightSeq(Wang et al., 2021), FasterTransformer (NVIDIA, 2023a), ByteTransformer(Zhai et al., 2023)와 같은 많은 추론 엔진에서 광범위하게 구현되었으며, 인스턴스화는 주로 그룹화된 GEMM, 레이어 정규화, 활성화 계산 및 자기 주의로 구성된다, Dao et al.(2022); Dao(2023). 이 방향을 따르는 대부분의 연구는 GPU 메모리 계층 구조를 활용하여 메모리 지연 시간을 숨기고 스레드 점유율을 극대화하려 한다. 다양한 계산 패턴에 대한 고도로 최적화된 커널 구현에 대한 수요가 증가함에 따라, 자동 컴파일은 매우 활발한 연구 분야가 되었다. 이는 고도로 최적화된 커널 구현(예: CUTLASS, cuBLAS, cuDNN)을 위한 소스 라이브러리보다 더 큰 유연성을 제공하는 접근 방식이다. 이 범주 내에서 TVM(Chen et al., 2018), Triton(Tillet et al., 2019), JAX(Bradbury et al., 2018)는 하드웨어 가속기의 잠재력을 활용하여 고효율 저수준 코드로 컴파일하는 동시에 빠르고 쉬운 프로토타이핑을 위한 파이썬 인터페이스를 제공한다. 이는 사용자 정의 커널 작성의 학습 곡선을 크게 낮출 뿐만 아니라 다른 컴퓨팅 환경에 대한 코드 적응을 위한 추상화를 제공한다.

4. 트랜스포머를 넘어(Beyond Transformers)

이차 스케일링 한계를 극복하기 위해 Transformer 아키텍처의 대안을 연구하는 데에는 전문가 혼합(MoE: Shazeer et al., 2017; Roy et al., 2020), 상태 공간 모델(SSM: Gu et al.,

2022a; Gupta et al., 2022; Smith et al., 2023), 그리고 순환 신경망(RNN)의 재등장이 포함된다. 이러한 접근 방식은 토큰을 특수 네트워크로 라우팅하거나, 선형 복잡도를 갖는 시퀀스 변환을 모델링하거나, 긴 컨텍스트 처리를 위해 순환 관계를 사용하는 등 다양한 메커니즘을 제공한다. 또한, 아키텍처 하이브리드화, 다양한 모델의 요소 결합, 그리고 더욱 효율적인 구조를 생성하기 위한 자동화된 검색 프로세스가 포함된다.

제 3 절 대규모 학습(Large-scale Training)

최신 하드웨어를 사용하는 대규모 모델 학습은 여러 가지 과제에 직면하고 있다. 예를 들어, 메모리 요구량 증가로 인해 단일 GPU에 모델을 모두 담을 수 없다는 점, 최소한의 오버헤드를 발생시키면서 더 많은 컴퓨팅 유닛을 사용하여 학습 속도를 높이는 것(선형 확장), 그리고 분산된 리소스를 활용하는 것 등에 직면하고 있다.

대규모 사전 학습과 다운스트림 작업 적응을 위한 효율적인 미세 조정을 가능하게 한 여러 연구와 분산 학습을 통해 다양한 가능성을 모색하는 방법으로 1) 하드웨어 및 네트워크 이질성을 포함하는 많은 경우 수동 설계와 동등하거나 더 나은 성능을 보이는 자동 병렬 스케줄링 방법(Jiang et al., 2023e), 2) 대규모 모델을 학습하고 제공하는 메모리 관리, 3) 비용(구현 난이도, 데이터 요구 사항, 학습 예산 등)과 지속적인 사전 학습으로 인한 성능 격차 간의 균형을 찾을 수 있는 효율적인 미세 조정(Efficient Fine-tuning), 4) 클라우드를 통해 분산 및 하드웨어 이기종 컴퓨팅 장치를 활용하여 모델 학습 및 추론을 수행하는 분산화(Decentralization), 5) 모델을 다양한 크기로 확장할 때 훈련 역학에 대한 추론에 도움이 될 수 있도록 데이터 및 모델 가중치 측면 뿐만 아니라 중간 체크포인트 및 아티팩트 로깅과 같은 다른 측면에서도 완전히 “개방적”일 수 있는 훈련 역학 및 확장(Training Dynamics & Scaling) 등이 요구된다.

1. 병렬 컴퓨팅(Parallel Computing)

대규모 언어 모델의 병렬 처리는 4D 병렬 처리로 불리는 네 가지 주요 모드를 사용하여 여러 컴퓨팅 유닛에 걸쳐 효율성을 높인다. 이는 모델 및 데이터를 분산하여 처리하며, 최신 프레임워크와 자동화된 스케줄링 기술을 통해 최적의 성능을 달성한다. 주요 병렬 처리 모드로는, 모델을 여러 유닛에 복제하고, 데이터를 분할하여 각 모델에 공급하는 분산 데이터 병렬(DDP: Aminabadi et al., 2022),(FairScale: FairScale, 2021),(Megatron-LM: Shoeybi et al., 2020), 모델 가중치를 여러 GPU에 분할하고 데이터를 병렬로 처리한 후 결과를 집계하는 텐서 병렬(TP: Shoeybi et al., 2020), 모델 계층을 여러 GPU에 수직으로 분할하고, 데이터가 여러 유닛을 통해 단계별로 이동하는 파이프라인 병렬(PP: Huang et al.,

2019), 긴 컨텍스트 작업을 위해 시퀀스 차원을 따라 연산 및 저장 부하를 분할하는 시퀀스 병렬(SP: Liu et al., 2023j, Shoeybi et al., 2020) 등이 있다. 최적의 병렬 처리 전략을 자동으로 찾는 기술은 하드웨어 활용을 극대화하는 자동 병렬 스케줄링의 사례인 HexGen(Jiang et al., 2023e) 등은 검색 기반 알고리즘을 사용해 클러스터 구성의 다양한 요소를 고려하여 병렬화 프로세스를 자동화한다.

2. 메모리 관리(Memory Management)

대규모 언어 모델의 학습 및 서비스에서 메모리 관리는 매우 중요하며, 특히 긴 컨텍스트 처리 시 KV 캐시가 모델 가중치와 활성화 메모리를 능가할 수 있다. 메모리 단편화 문제를 해결하고 GPU 활용을 극대화하기 위해 페이지 기반 관리, KV 캐시 압축, 메모리 계층 간 오프로딩 등의 기술이 사용된다. 주요 메모리 관리 기술로는 기존 OS의 페이지 기법을 활용하여 KV 캐시를 불연속적인 메모리 블록으로 분할하는 페이지 어텐션(Paged Attention)(vLLM: Kwon et al., 2023a), 구조 프로파일링을 통해 중요하지 않은 토큰을 동적으로 제거하여 메모리 사용량을 줄이는 적응형 KV 캐시 압축(FastGen, Ge et al., 2024), 경험적 관찰을 바탕으로 핵심 토큰만 유지하여 메모리를 절약하고 성능을 유지하는 핵심 토큰 기반 메모리 관리(Liu et al., 2023a; H2O, Zhang et al., 2023j), 어텐션 계산을 서브루틴으로 분할하고, GPU 및 CPU 메모리를 동적으로 관리하는 전용 서버를 통해 효율적으로 분산 처리하는 분산 메모리 관리(LLM: Lin et al., 2024a) 등이 있다. 이외에 모델 가중치나 KV 캐시를 메모리가 더 많은 CPU로 이동시켜 GPU의 메모리 제약을 완화하는 CPU 오프로딩(FlexGen: Sheng et al., 2023b), 역전파 시 계산 그래프의 일부를 재계산하여 최대 메모리 사용량을 줄이는 그래디언트 체크포인팅(Chen et al., 2016)이 있다. 결론적으로 효율적인 메모리 관리는 확장 가능한 시스템 구축과 대규모 병렬 처리를 위한 필수적인 요소이다.

3. 효율적인 미세 조정(Efficient Fine-Tuning)

사전 학습된 대규모 모델을 특정 작업에 맞게 조정하는 미세 조정은 매우 비용이 많이 들고 비효율적일 수 있다. 매개변수 효율적 미세 조정(parameter-efficient fine-tuning

PEFT) 기법은 학습해야 할 매개변수의 수를 최소화하여 비용과 성능 간의 균형을 맞추는 다양한 기법을 의미한다. 이러한 방법들은 적은 노력으로도 우수한 성능을 달성하는 것을 목표로 한다. 주요 PEET 기법으로는 델타 가중치 행렬의 저랭크 분해인 학습 가능한 행렬을 삽입하거나(PiSSA: Meng et al.(2024)), 매우 적은 계산 예산으로 LLaMA를 명령어 추종 모델로 효율적으로 미세 조정하는 기법(LLaMA-Adapter: Zhang et al., 2023d) 등이 있다. 또한 모델 내부에 학습 가능한 요소를 추가하는 다양한 어댑터 및 프롬프트 기법도 있다.

4. 탈중앙화(Decentralization)

탈중앙화는 여러 지역에 분산된 이기종 컴퓨팅 장치(클러스터 또는 개인 기기)를 활용하여 대규모 모델의 학습 및 추론을 수행하는 것이다. 주요 과제는 통신 오버헤드를 극복하고 효율성을 유지하는 것이며, 이를 위해 데이터 압축, 최적화된 연합 학습, 그리고 개인 장치를 활용하는 협업 학습 등의 기술이 사용된다. 주요 탈중앙화 기술 및 사례로는 인코더를 사용하여 KV 캐시를 압축함으로써 컨텍스트 처리 지연 시간을 줄이거나(CacheGen: Liu et al., 2023c), 희소화 및 양자화 기술을 결합하여 느린 네트워크 환경에서도 통신량을 최소화하며 LLM을 미세 조정하는 통신 오버헤드 최적화 기법(CocktailSGD: Wang et al., 2023g)과 연결 상태가 좋지 않은 장치들로 구성된 네트워크에서 효율적인 연합 평균화 알고리즘을 사용여 통신 속도를 크게 낮추면서도 동기식 최적화와 비슷한 성능을 제공하는 연합학습 기법(DiLoCo: Douillard et al., 2023), 여러 사용자의 GPU를 크라우드소싱하여 BLOOM-176B 및 OPT-175B와 같은 대형 모델을 제공하고 미세 조정하는 협업 학습(Petals: Borzunov et al., 2022등이 있다. 결론적으로 탈중앙화는 전 세계의 컴퓨팅 자원을 활용하여 확장성, 내결함성, 그리고 프라이버시 보호를 강화하는 차세대 AI 시스템의 핵심 방향이다.

5. 학습 역학 및 확장(Training Dynamics & Scaling)

대규모 언어 모델의 학습 역학을 이해하는 것은 AI 개발에 중요하지만, 많은 성공적인 LLM이 학습 과정의 세부 사항을 완전히 공개하지 않아 어려움이 있다. 최근에는 중간 체크포인트와 학습 데이터를 공개하는 노력을 통해 확장성 및 학습 역학에 대한 분석이 활

발해지고 있다.(Xia et al., 2023a)는 다양한 하위 작업에서 OPT 모델(Zhang et al., 2022a)의 중간 체크포인트를 분석하고, 모델 크기보다 복잡도를 모델 성능의 예측 지표로 강조하며, 더 큰 모델일수록 환각 현상이 덜하고 학습 초기 단계에서 모델이 최소한의 성능을 보이는 경향이 있음을 보여준다. 이와 더불어(Tirumala et al., 2022)는 모델 크기, 데이터셋 크기, 학습률 전반에 걸쳐 다양한 기억 능력을 연구하는 데 중점을 두고 있으며, 개별 학습 사례를 기억하기 위한 고유 식별자로서 명사와 숫자의 중요성에 대한 흥미로운 가설을 제시한다. 이외에 Pythia(Biderman et al., 2023)는 7천만 개에서 120억 개에 이르는 매개변수 크기의 공개 데이터를 기반으로 학습된 16개의 LLM(학습 모델)을 소개한다. 이러한 중간 체크포인트가 더 광범위한 커뮤니티에 공개됨에 따라, 연구자들은 개별 저장된 가중치와 손실을 검토하고 벤치마킹함으로써 학습 역학 관련 질문에 대한 답을 훨씬 더 쉽고 효율적으로 찾을 수 있게 된다. 마지막으로, OLMo(Groeneveld et al., 2024)는 학습 데이터와 학습 및 평가 코드를 포함한 전체 프레임워크를 기꺼이 공개하여 LLM의 과학적 연구를 더욱 용이하게 한다.

제 4 절 추론 기법(Inference Techniques)

AGI 추론 시스템은 사용자 응답성, 가용성 및 효율성을 보장해야 하며, 이는 학습 단계에서 대규모 모델의 궁극적인 잠재력을 최대한 발휘하고 사용자와 시스템의 상호 작용 방식을 혁신하는 데 도움이 된다. 따라서 자동 회귀 디코딩을 가속화하고, 요청 스케줄링의 균형을 맞추고, 클러스터에서 다양한 기능을 가진 수많은 모델을 처리하는 여러 기법으로 1) 정확한 디코딩 가속에 중점을 두고, 정확도를 떨어뜨리지 않으면서 원래 모델에 충실하면서 성능을 극대화하는 디코딩 알고리즘(Miao et al., 2023), 2) 주어진 입력에 대한 컨텍스트(사용자 정보, 이전 KV 캐시, 모델 어댑터 등)의 효율적인 프리페칭과 최대 GPU 활용도를 위해 가변 시퀀스 길이의 예제 처리 및 첫 번째 토큰까지의 시간(TTFT), 작업 완료 시간(JCT), 배치 토큰 처리량, 추론 지연 시간과 같은 다양한 요청 수준 지표 간의 균형 유지 Orca(Yu et al., 2022) 등이 효율적으로 실행되는 스케줄링(Request Scheduling), 3) 여러 애플리케이션 시나리오(LLM 에이전트, 페르소나 챗봇, 개인 정보 보호 지원 등)에서 동일 모델의 여러 복제본을 제공하고, 여러 작업 특화 모델을 효율적으로 배포할 수 있는 다중 모델 서비스(Multi-model Serving) 등이 있다.

1. 디코딩 알고리즘(Decoding Algorithm)

정확한 디코딩 가속화는 대규모 언어 모델의 성능을 극대화하면서도 원래 모델의 정확도를 유지하는 것을 목표로 한다. 이를 위해 추측 디코딩 및 하드웨어 인식 알고리즘과 같은 다양한 기술들이 활용된다. 가벼운 초안(draft) 모델이 여러 토큰을 예측하고, 더 강력한 대상(target) 모델이 이를 병렬로 검증하여 디코딩 속도를 높이는 추측 디코딩(Speculative Decoding)의 사례로는 초안 모델 대신 기존 데이터베이스의 문자열 매칭을 사용해 후보 토큰을 생성하는 Promptlookup Decoding(Saxena, 2023), 초안 모델의 출력을 토큰 트리로 구성하여 병렬 검증을 용이하게 하는 SpecInfer(Miao et al., 2024), 특수 마스크 패턴을 사용한 트리 어텐션으로 여러 메두사 헤드의 토큰을 동시에 확인하는 Medusa(Cai et al., 2024a), 초안 모델 없이 중간 계층을 건너뛰어 후보 시퀀스를 생성하는 Self-speculative decoding(Zhang et al., 2023k) 등이 있고, 하드웨어의 특성을 고려하여 메모리

접근 및 계산을 최적화하는 기법으 사례로는 시퀀스를 블록으로 나누고 Flash-Attention으로 병렬 처리한 후 결과를 집계하여 효율성을 높이는 Flash-Decoding(Dao et al., 2022; Dao, 2023)과 Flash-Decoding의 한계를 보완하여 비동기 소프트맥스, 최적화된 GEMM, 하드웨어 적응형 데이터 흐름 등을 통해 HuggingFace 대비 4배 이상 빠른 속도를 달성한 FlashDecoding++(Hong et al., 2023a)이 있다.

2. 요청 스케줄링(Request Scheduling)

대규모 언어 모델의 요청 스케줄링은 가변적인 시퀀스 길이와 자기회귀적 특성 때문에 복잡하다. 핵심은 GPU 활용을 극대화하고, 지연 시간을 최소화하며, 처리량을 높이기 위해 컨텍스트 프리페칭, 동적 배칭, 그리고 요청의 특성을 고려한 예측적 스케줄링 기법을 활용하는 것이다. 요청 스케줄링 전략의 몇 가지 중요한 특징은 1) 주어진 입력에 대한 컨텍스트(사용자 정보, 과거 KV 캐시, 모델 어댑터 등)의 효율적인 프리페칭, 2) 최대 GPU 활용을 위한 가변 시퀀스 길이의 예제 처리, 3) 첫 번째 토큰까지의 시간(TTFT), 작업 완료 시간(JCT), 배치 토큰 처리량, 추론 지연 시간과 같은 다양한 요청 수준 지표 간의 균형을 맞추는 것이다. 요청을 단계별로 처리하고 동적으로 배치를 구성하여 GPU 활용도를 극대화하고 서로 다른 길이의 요청으로 인한 지연 시간을 줄이는 반복 수준 및 동적 배칭 사례로는 반복 수준 스케줄링과 선택적 배칭을 결합하여 요청의 자기회귀적 특성을 더 잘 처리함으로써 이전 엔진보다 우수한 성능을 발휘한 Orca(Yu et al., 2022), 시퀀스가 완료됨에 따라 배치가 동적으로 변경되어 GPU 리소스를 계속 사용할 수 있도록 하는 연속 배칭을 구현한 vLLM(Kwon et al., 2023a), “인플라이트 배칭”을 사용하여 대기 시간을 제거하고 컨텍스트 및 생성 단계를 함께 처리한 TensorRT-LLM(NVIDIA, 2023b) 등이 있다. 또한 입력 또는 잠재적 출력 길이에 대한 정보를 사용하여 스케줄링 결정을 최적화하는 예측 및 적응형 스케줄링 접근 방식으로는 JCT를 줄이기 위해 입력 길이를 고려하는 다중 레벨 피드백 큐 스케줄러를 사용하여 토큰 수준에서 선점을 가능하게 한 FastServe(Wu et al., 2023c), 각 요청에 대한 잠재적 응답 길이를 예측하여 더 많은 예제를 메모리에 저장함으로써 짧은 요청과 긴 요청 간의 성능 격차를 최소화한 S3(Jin et al., 2023b)이 있다. 토큰을 조작하여 더욱 균일한 배치를 생성하여 효율성과 응답성을 향상시킨 방법으로는 프롬프트

와 생성에서 토큰을 일관된 전달 크기로 구성하고, 긴 프롬프트는 분할하고 짧은 프롬프트는 결합하여 요청 분산을 줄인 Dynamic SplitFuse(Holmes et al., 2024), DeepSpeed-FastGen 방법이 있다.

3. 멀티 모델 서빙(Multi-model serving)

멀티 모델 서빙은 단일 기본 모델에 PEFT(매개변수 효율적 미세 조정) 어댑터를 동적으로 로드하여 여러 작업 특화 모델을 효율적으로 제공하는 방식이다. 이는 여러 개의 완전한 모델을 배포하는 것보다 계산적으로 훨씬 효율적이고 자원 낭비가 적다. 주요 과제는 다양한 요청에 맞춰 올바른 어댑터를 신속하게 로드하는 것이다. 주요 기술 사례로는 단일 사전 학습 모델을 공유하는 이기종 LoRA 어댑터들의 연산을 배치 처리하는 전용 CUDA 커널을 사용하고, 이를 통해 지연 시간을 약간만 추가하면서도 처리량을 최대 12배까지 높일 수 있는 Punica(Chen et al., 2023h), 동적 어댑터 관리를 위한 통합 메모리 풀을 사용하는 통합 페이징(Unified Paging) 기술을 도입하고, 이와 함께 최적화된 CUDA 커널을 사용하여 단일 GPU에서 수천 개의 LoRA 어댑터를 효율적으로 서빙할 수 있게 한 S-LoRA(Sheng et al., 2023a), 어댑터 교환 스케줄링을 구현하여 GPU와 CPU 메모리 간에 어댑터를 비동기적으로 프리페치 및 오프로드하고, 배칭을 스케줄링하여 전체 처리량을 최적화한 LoRAX(Predibase, 2023)이 있다. 이러한 기술들은 단일 GPU로도 수많은 LoRA 헤드를 제공할 수 있게 하여, 모델 협업, 작업 일반화, 모델 병합 등 광범위한 가능성을 열어준다.

제 5 절 컴퓨팅 플랫폼(Computing Platforms)

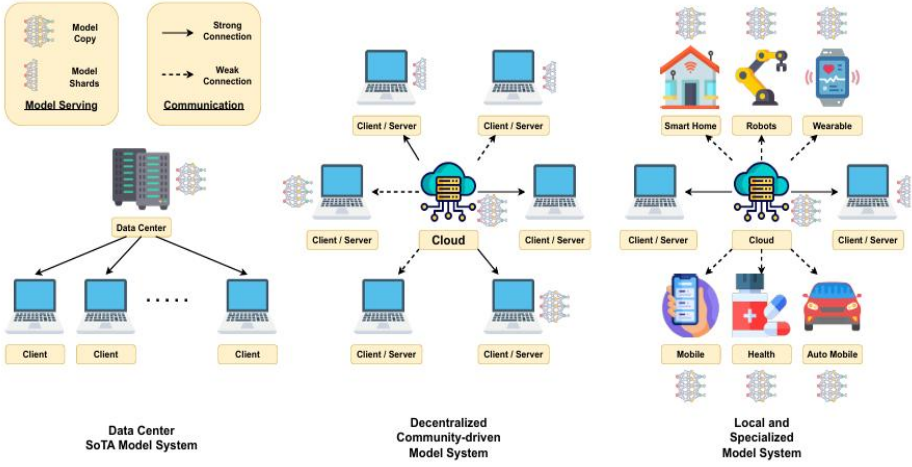
대규모 언어 모델의 발전과 실용성을 결정하는 주요 요인은 끊임없이 진화하는 하드웨어 가속기이다. GPU는 가장 널리 사용되는 선택으로, 빠른 스레드 공유 메모리를 통해 병렬 계산을 최적화한다. GPU는 풍부한 벡터 및 행렬 곱셈을 사용하는 최신 딥 러닝에 적합하다. NVIDIA의 Ampere 및 Hopper GPU 아키텍처는 향상된 메모리 용량, 액세스 속도 및 컴퓨팅 성능(텐서 코어 증가) 덕분에 많은 최첨단 모델의 초석이 되었다. 이러한 GPU는 다양한 산술 정밀도(32비트 및 16비트 부동 소수점)와 형식(텐 서 부동 소수점 및 브레인 부동 소수점)을 지원하며, 이는 수치 정밀도와 효율성의 균형을 이룬다. NVIDIA 외에도 다른 제조업체들도 TPU(Jouppi et al., 2023), FPGA(Yemme and Garani, 2023), AWS Inferential 그리고 Groq의 LPU 6과 같은 딥 러닝 애플리케이션을 위한 특수 가속기에 투자하고 있으며, 각각 고유한 장점을 가지고 있다.

대형 모델은 학습 및 추론을 지원하기 위해 막대한 메모리 용량이 필요하다(추가 최적화 없이 네이티브 Llama-70B를 처리하려면 80GB VRAM을 갖춘 A100 8개가 필요하다). 그러나 효율적인 알고리즘을 개발하려면 기본 하드웨어 사양(모델 병렬 처리, 메모리 계층 구조, 네트워크 구성 등)에 대한 깊은 이해가 필요하다. 모델을 1조 개 이상의 규모로 확장함에 따라 더 복잡한 병렬 처리 기술이 필수적이며, 이는 개념화, 구현 및 유지 관리가 어려울 수 있다. NVIDIA DGX GH200 7은 상호 연결된 Grace Hopper Superchip(Grace CPU와 Grace GPU)에서 대규모 공유 메모리 공간(최대 144TB)을 제공하여 프로그래밍 모델을 간소화한다. Qualcomm Cloud AI 100 Ultra 8은 단일 150와트 카드에서 1,000억 개의 매개변수 모델을 처리할 수 있다(LED 전구와 동일한 전력 소비량).

가속기의 강력한 성능과 효율성은 유연성을 제공하며, 이는 NVIDIA의 CUDA 및 AMD의 ROCm과 같은 특수 설계된 프로그래밍 언어를 통해 스레드 사용 및 계산 로직을 더욱 세밀하게 제어할 수 있도록 한다. TVM(Chen et al., 2018) 및 MLC-LLM(MLC team, 2023)과 같은 여러 연구들은 컴파일러 가속을 통해 모든 기기에 머신 러닝 및 딥 러닝 모델을 배포하는 것을 보편화하고자 하며, 이는 다양한 가속기의 잠재력을 극대화하는 것을 목표로 한다. AI 하드웨어 연구 및 엔지니어링은 차세대 AI 진화의 출현을 주도할 것으로 예상되

며, AGI 시스템은 현재의 한계를 극복하고 계산 및 전력 효율성의 경계를 한 단계 더 높일 수 있는 차세대 하드웨어 플랫폼이 필요할 것으로 예상된다.

〔그림 5－2〕 The Future Forms of AGI Systems



제 6 절 비용 및 효율성(Cost and Efficiency)

모델 훈련 및 추론과 관련된 비용은 쉽게 간과될 수 있지만, 실제로는 특히 산업 환경에서 이러한 요소들이 종종 모델 아키텍처 설계, 데이터 믹스 선택, 서비스 가격 책정과 같은 많은 의사 결정에 영향을 미칠 수 있다. 개발 주기를 가속화하고 모델의 경제적 성능을 향상시키는 몇 가지 대표적인 선행 노력으로는 1) 성능 향상을 위해 기존 데이터 믹스에 어떤 데이터를 추가해야 하는지와 모델을 더 견고하게 만들기 위해 비필수 데이터(outliers)를 제거할 수 있는지 그리고 데이터 제공업체에 지불할 합리적인 비용 등 데이터 재무와 관련된 사항, 2) 간단한 평균화(Wortsman et al., 2022), 작업 산술(Ilharco et al., 2023), 다중 모드(인코더) 병합(Wu et al., 2023b; Sung et al., 2023), 학습된 라우팅 함수를 기반으로 한 병합(Lu et al., 2023), SLERP, 가중(복합 경사 하강법(Tam et al., 2023), 확률 및 인구 기반 최적화 알고리즘(Huang et al., 2024) 병합 등 일련의(전문화된) 대형 모델을 조정하거나 병합하여 전체 시스템의 성능을 향상시키기 위한 모델 조합(Model Combination), 3) 대규모 기반 모델 구축의 복잡성이 증가함에 따라 민주화와 민첩한 개발을 위해 보다 성숙한 자동화 프로세스가 있다.

제6장 AGI 기술 발전 단계별 기술적·산업적 병목 요인 식별

제1절 기술적 병목 요인 식별

AGI 기술 발전에서 컴퓨팅 자원 한계, 학습 데이터 한계 및 품질, 알고리즘 복잡도와 학습 효율, 신뢰성 및 강건성, AGI 안전성 및 목표 정렬(Alignment), 지식 전이 및 일반화 실패, 해석 가능성 및 투명성 등 AGI 실현을 저해하는 핵심 기술적 병목 요인을 식별하고 분석해 보고자 한다.

1. 컴퓨팅 인프라 및 자원 한계 분석

AGI 개발의 가장 핵심적인 기술 병목 중 하나는 고성능 연산 자원의 확보와 효율적 활용이다. 예를 들어, GPT-4의 학습에는 약 25,000대의 NVIDIA A100 GPU가 90~100일간 투입되었으며, 총 연산량은 약 2.15×10^{25} FLOPs에 달하는 것으로 추정된다. 이는 단일 국가 연구기관이 자체 보유하기에는 지나치게 높은 수준의 자원 요구이며, Google Gemini Ultra 역시 TPU v4 클러스터 기반으로 수천 대의 장비를 병렬로 운용해야만 학습이 가능하다. TPUv4는 장치당 275 TFLOPS, 32GB HBM 메모리를 제공하며, NVIDIA A100 대비 1.2~1.7배 빠른 성능, 1.3~1.9배 높은 에너지 효율을 달성하며, AGI급 모델 학습에는 이를 뒷받침할 대규모 네트워크·전력 인프라가 병행되어야 한다.

가. 에너지·탄소 비용

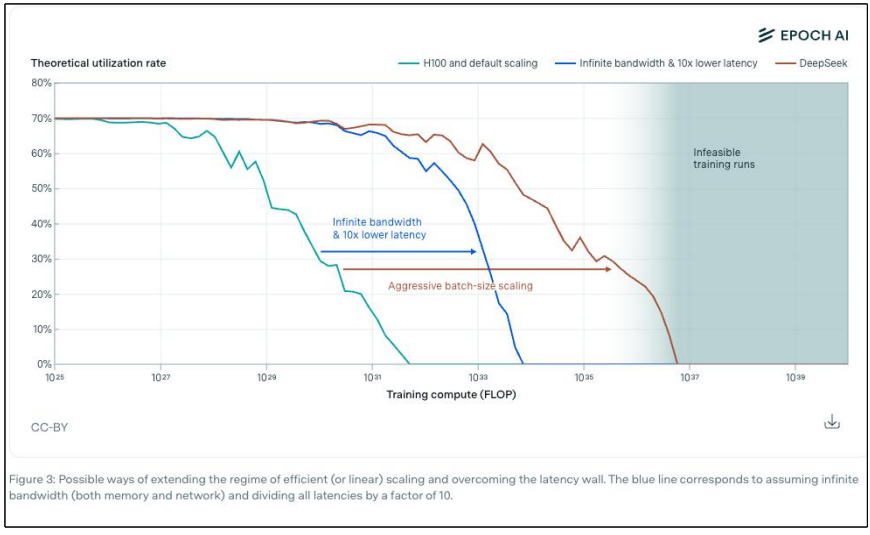
최근 분석에 따르면 GPT-4 학습에는 약 5,177만 kWh의 전력이 소모된 것으로 추정되며, 미국 캘리포니아 West US Azure 리전의 평균 탄소집약도(240.7 gCO2e/kWh) 기준 약 15,000톤 CO2e에 해당한다. 이는 미국 평균 가정 1,000세대가 평균적으로 5~6년간 사용하는 전력량에 상응한다. 이러한 전력·탄소 비용은 환경 문제를 넘어, AGI 연구의 지속가능

성과 인프라 운영비용에 직결되는 병목 요소가 된다.

나. 네트워크·메모리 병목

모델 규모가 커질수록 네트워크 전송 지연(latency)과 대역폭 제약이 분산 학습의 주요 병목으로 작용한다. 특히, 대규모 노드를 활용하는 InfiniBand HDR(200 Gbps) 환경에서도 통신 지연이 학습 속도 한계(지연 장벽, latency wall)의 주요 요인으로 작용할 수 있음이 지적되고 있다. Megatron-LM, DeepSpeed-ZeRO 등 최신 분산 학습 프레임워크는 트릴리언(Trillion)-scale 모델 학습을 가능하게 하고, 대규모 메모리 절감을 통해 단일 클러스터에서도 초대형 모델을 처리할 수 있으며, CPU/NVMe 오프로드를 활용하면 훨씬 큰 모델 학습까지 지원한다. 그러나 네트워크 대역폭 제약과 I/O 병목 문제는 여전히 시스템 성능 한계로 존재한다. 이는 대량의 파라미터와 활성화 정보가 노드별로 빈번히 교환되어야 하는 분산 환경의 특성 때문이다.

〔그림 6－1〕 효율적 스케일링 한계와 지연(latency) 장벽 극복 방안 비교



자료: E Erdil et al. 2024

다. 공급망·지정학 리스크

고성능 AI 하드웨어 공급은 NVIDIA, Google, AWS 등 소수 글로벌 기업이 장악하고 있다. 특히 NVIDIA H100은 A100 대비 TOPS/Watt 효율이 약 3배 높지만, 공급량이 제한적이며, 미국의 수출 규제 강화로 중국은 H100·A100 도입에 큰 제약을 받고 있어 현지 초거대 모델 개발 속도가 둔화된 사례가 있다. 이러한 공급 집중 구조는 반도체 공급망 교란, 지정학 리스크, 무역 규제에 취약하다.

라. 시사점

AGI 개발에서 연산 자원과 전력 인프라, 통신 구조, 메모리 아키텍처는 상호 연동된 복합 요소로 작용하며, 단일 요소 개선만으로는 한계를 극복하기 어렵다. 따라서 국가 차원의 AGI 연구 역량 강화를 위해서는 1) 고성능 연산 자원의 안정적 공급망 확보, 2) 저탄소·재생에너지 기반 대규모 데이터센터 설계, 3) 통신·I/O 병목 완화를 위한 차세대 분산 아키텍처 연구가 필요하다.

2. 데이터 부족 및 품질 분석

AGI 모델의 성능은 데이터의 양과 품질 모두에 크게 의존한다. Common Crawl은 300억 페이지 이상의 웹 데이터를 제공하지만, 최근 KISTEP 분석(2024)에 따르면 전체 데이터의 약 40%가 중복 콘텐츠이며, 문단 단위로 분석시 중복률이 70%까지 도달한다. 또한 광고, 메뉴바, 저작권 문구, 깨진 문자 등 의미 없는 노이즈 데이터의 비율도 매우 높다(Baack, Stefan., 2024)(Morgan Stanley, 2025). The Pile 데이터셋 역시 학술논문, 코드, 뉴스 등 22개의 다양한 도메인을 포함하지만, 여전히 몇 가지 구조적 한계를 안고 있다. 첫째, Common Crawl 기반으로 구축된 Pile-CC는 여전히 웹 크롤링 특유의 노이즈와 품질 저하 문제를 포함하고 있으며, GitHub·ArXiv·수학 데이터처럼 자연어와 다른 형식은 일반 언어 모델링에 적합하지 않다. 둘째, Books3, Enron Emails 등 일부 서브셋은 저자 동의나 저작권 문제가 있어 법적·윤리적 논란 소지가 있다. 셋째, 성별·인종·종교와 관련된 편향과 유해 콘텐츠가 포함되어 있으며, 전체 데이터의 97% 이상이 영어로 구성되어 다국어 일반화에도 제약이 크다. 마지막으로, 데이터셋 분할 간 중복 가능성과 특정 도메인의 성능 저하가 보고되어 AGI 모델 학습용으로는 여전히 구조적 한계가 존재한다.

가. 데이터 품질 저하와 모델 성능 영향

데이터 품질 저하는 모델 환각(hallucination)과 편향(bias) 발생률을 높인다. 특히, LAION-5B와 같은 초대규모 웹 크롤링 기반 데이터셋은 중복된 이미지-텍스트 쌍, 잘못 매칭된 캡션, 품질이 낮은 웹 콘텐츠가 다수 포함되어 있다. 이러한 저품질 데이터는 학습 과정에서 모델이 특정 패턴을 과도하게 암기하게 하여 일반화 성능을 저해하며, 결과적으로 이미지-텍스트 검색이나 분류와 같은 다운스트림 태스크에서 오류율을 증가시킨다. 또한 데이터셋 내부의 사회적 편향과 유해 콘텐츠는 모델이 이를 그대로 재생산하거나 강화할 위험이 있으며, 이는 AGI 실현 과정에서 모델의 성능 뿐 아니라 안전성과 신뢰성을 저해하는 주요 병목으로 작용한다.

나. 고품질 데이터 고갈과 합성 데이터 의존성

Villalobos, Pablo, et al. 에 따르면, 공개된 인간이 생산하는 고품질 텍스트 데이터는 빠르면 2026년, 늦어도 2032년 사이에 대부분 소진될 것으로 예상된다. 이는 곧 AGI 학습이 합성 데이터에 상당 부분 의존하게 될 것임을 시사한다. 그러나 합성 데이터만을 반복적으로 활용할 경우, 모델이 원래의 인간 데이터 분포에서 점차 벗어나 출력이 동질화되고 비현실적으로 변질되는 위험이 지적된다. 실제로 선행 연구들은 모델이 자기 생성 데이터를 다시 학습할 때 편향과 오류가 누적되며 성능이 저하되는 현상(model collapse)이 발생한다고 보고한다. 이러한 경향은 합성 데이터 의존도가 커질수록 더욱 심화될 수 있다.

다. 데이터 접근성과 법적·제도적 장벽

AGI 학습을 위한 데이터 확보 과정에서는 프라이버시 보호와 저작권 규제가 중요한 현실적 장벽으로 작용한다. 먼저 개인정보보호법(GDPR, CCPA 등)은 의료기록·금융 거래, 교육 데이터와 같은 고가치 도메인의 대규모 활용을 엄격히 제한한다. 이로 인해 AGI 학습에 필수적인 고품질·고정밀 데이터 접근성이 본질적으로 제약된다. 저작권 문제 역시 주요 장벽인데, 학술 논문과 상업 출판물은 무단 크롤링 시 법적 분쟁에 직면할 수 있으며, 실제로 Getty Images와 New York Times가 AI 기업들을 상대로 제기한 소송은 저작권 데이터 활용이 초래할 수 있는 법적 리스크를 단적으로 보여준다. 이러한 제약은 단순한 윤리적 쟁점을 넘어 학습 데이터 구축 자체의 기술적 한계로 이어진다. 즉, 실제 활용 가능한 데이터 풀(pool)이 점점 축소되고, 모델이 특정 도메인에서 충분히 학습이 이뤄지지

못해 모델의 일반화 성능에 불균형을 초래한다. 나아가 라이선스 제약으로 인해 일부 데이터셋은 공개적으로 배포되지 않고 특정 기업·기관에 독점적으로 귀속되면서, AGI 연구 생태계 전반에 데이터 독점과 불균형이라는 구조적 병목이 심화되는 것이다.

라. 멀티모달 데이터의 품질 문제

멀티모달 학습에서는 시각·언어 간 의미적 불일치뿐만 아니라, 헤테로지니어스(heterogeneous) 데이터 소스 간의 구조적 차이와 메타데이터 오류가 모델의 일반화 성능을 저해할 수 있다. 예를 들어, 이미지-텍스트 페어링 과정에서 캡션이 실제 시각 정보를 온전히 반영하지 못하거나, 도메인별 데이터 분포가 심각하게 불균형할 경우, 모델은 특정 도메인에 과적합(overfitting)되거나 반대로 핵심 의미를 놓치는 문제가 발생한다. P. Villalobos, L. Bing, S. Xu, et al.에서는 이러한 문제를 방지하기 위해 도메인별 커버리지 평가와 품질 균형 조정이 필요하다고 강조한다. 더 나아가, Y. Jiang, et al.은 멀티모달 시계열 데이터가 텍스트, 이미지, 표, 그래프 등 이질적 모달리티로부터 수집될 때, 데이터 이질성(heterogeneity), 잡음(noise), 그리고 시간적·의미적 불일치(misalignment)가 누적되어 분석과 예측 성능을 저해할 수 있음을 지적한다. 이는 멀티모달 맥락을 효과적으로 정합(alignment)하지 못할 경우, AGI 모델이 요구하는 범용성(generality) 확보를 제약하는 핵심 한계로 작용한다. 따라서 대규모 멀티모달 데이터셋 구축 시 단순한 규모 확장보다는 커버리지 다양성, 정합성, 이질적 출처 간 조화를 보장하는 것이 성능 유지와 AGI 실현에 더 중요하다.

마. 시사점

학습 데이터 한계 및 품질 문제는 AGI 연구의 정확도와 효율성에 직결된다. 따라서 국가 및 글로벌 차원에서 다음과 같은 노력이 필요하다. 1) 고품질·저편향 데이터 정제 및 검증 체계 구축, 2) 멀티모달 데이터 커버리지 평가 및 품질 관리, 3) 합성·실세계 데이터 혼합 비율 최적화, 4) 데이터 접근성의 법적·제도적 장벽 완화가 필수적이다. 특히 데이터 필터링 및 정제 파이프라인에서 품질 우선 기준을 적용하고, 노이즈 제거 및 샘플 가중치 재조정을 통해 데이터 효율성을 보장해야 한다.

3. 알고리즘 복잡도 및 학습 효율성 한계 분석

AGI 모델 개발의 또 다른 핵심 기술적 병목은 알고리즘 구조의 복잡성과 이에 따른 학습 효율성 문제이다. 현재 Transformer 구조는 AGI 모델의 핵심이지만, 입력 길이에 따라 시간·메모리 복잡도가 $O(n^2)$로 증가한다. 예를 들어 32k 토큰 이상의 긴 문맥을 처리할 경우, 단일 GPU에서 메모리 한계로 학습이 중단되거나 처리 속도가 급격히 저하될 수 있다. 이를 개선하기 위해 제안된 FlashAttention은 불필요한 메모리 접근을 줄여 처리 효율을 높였으며, 실험적으로 최대 2.7배의 속도 향상을 달성하였다. 그러나 긴 문맥에서 의미 일관성 저하에 관한 가능성이 보고 되고 있다.

가. Scaling Law와 데이터 효율성

Chinchilla scaling law에 따르면, 동일한 연산량에서 데이터와 파라미터 수의 균형이 최적 성능을 낸다(J. Hoffmann et al.). 그러나 GPT-4, Gemini 등 초대형 모델은 메모리·연산 제약으로 인해 최적 비율보다 파라미터가 많은 구조를 채택하는 경우가 많으며, 이는 데이터 효율성을 저하시킨다(OpenAI, 2023),(R. Anil, 2023). 즉, 단순히 파라미터 수 확장 만으로는 AGI급 모델의 성능을 극대화하는 데 한계가 존재한다. 따라서 데이터 효율성을 높이는 방향의 연구가 병행되어야 한다.

나. KV 캐시 관리 및 통신 병목

긴 시퀀스 학습의 병목은 단순히 메모리 문제 뿐 아니라 key-value(KV) 캐시 관리와 GPU 간 통신 대역폭 제약에서도 발생한다(T. Dao, 2022),(D. Li, 2024). 수만 토큰 이상 컨텍스트를 유지하려면 GPU 간 KV 캐시 공유가 필요하며, 이 과정에서 통신 지연(latency)과 대역폭 한계가 모델 전체 처리 속도를 제한하는 주요 원인으로 작용한다. 특히 대규모 분산 학습 환경에서는 이러한 제약이 두드러진다.

다. MoE(Mixture-of-Experts) 구조의 도전 과제

MoE(Mixture-of-Experts) 구조는 일부 전문가 네트워크만 활성화하여 연산량을 줄이는 방식으로 학습 효율성을 개선한다. 하지만 전문가 선택의 편향, 학습 불안정성, 그리고 전문가 간 로드 밸런싱 문제로 인해 성능 편차를 키울 수 있다는 점이 지적된다. 이를 완화하기 위해 토큰 라우팅 최적화, 전문가 수 및 활성화 비율의 동적 조정과 같은 추가 기법

이 필요하다(D. Harvey, 2025).

라. 시사점

알고리즘 복잡도와 학습 효율성의 병목을 해결하기 위해서는 Sparse Attention, Longformer, Hyena 등 차세대 모델 구조 도입이 필요하며, 메모리-통신 통합 최적화 및 KV 캐시 압축·분산 전략 구축 및 데이터-파라미터 균형 설계 기반 모델 개발, MoE 라우팅 편향 완화 기법 등이 병행되어야 한다. 이러한 노력을 통해 장문 입력 처리 효율성과 모델 확장성을 동시에 확보하는 것이 AGI 실현의 필수 조건이 된다.

4. 지식 전이 · 일반화 실패 문제 분석

AGI 모델은 본질적으로 여러 과제와 도메인 전반에서 지식을 전이하고, 새로운 상황에서도 안정적으로 일반화할 수 있어야 한다. 그러나 현행 LLM 및 Foundation Model은 여전히 이러한 요구를 충족하지 못하고 있다. 특히 파국적 망각(Catastrophic Forgetting) 문제와 표현 편향(Representation Bias)으로 인해 새로운 과제를 학습하는 과정에서 기존 지식이 급격히 손실되거나, 특정 도메인에 과도하게 최적화된 표현 구조 때문에 전이 효율이 크게 제한된다(N. Haque, 2025),(U. Peters and B. Chin-Yee, 2025).

이와 함께, 합성 데이터 기반 전이 역시 현실 세계의 다양성과 변동성을 충분히 반영하지 못해 일반화 실패 가능성을 내포한다. 결국, 망각-편향-전이 실패의 삼중 병목은 AGI가 요구하는 범용성(generality)과 적응성(adaptability) 확보를 가로막는 근본적 제약으로 작용한다.

가. 파국적 망각과 일반화 실패

최근 연구에 따르면 LLM(GPT-4, Claude 등)은 DomainNet, WILDS와 같은 OOD (Out-of-Distribution) 벤치마크에서 학습 도메인과의 유사도가 낮아질수록 정확도가 최대 20~40%p까지 하락하는 것으로 보고되었다(D. Roussinov et al., 2025). 이는 새로운 도메인 특성이나 편향이 주어졌을 때 LLM의 일반화 및 전이 능력이 근본적으로 제한됨을 보여준다. 또한 continual learning 환경에서는 새로운 과제를 학습하는 과정에서 이전 과제 성능이 30~50%까지 저하되는 파국적 망각(catastrophic forgetting) 현상이 빈번하게 발생한다(N. Haque, 2025). 이러한 결과는 AGI 모델이 장기적이고 지속적인 학습 환경에서 안정적으로 지식을 유지 및 확장하기 어렵다는 점을 실증적으로 입증한다.

나. 지식 전이 효율성의 한계

AGI를 지향하는 LLM은 다양한 도메인 간 지식 전이를 수행해야 하지만, 최근 연구들은 도메인 전이 효율이 근본적으로 제한적임을 보여준다. 전문 분야(법률, 의학 등)로 미세 조정된 모델은 다른 도메인으로 전이될 때 성능 개선이 뚜렷하지 않으며, 경우에 따라서는 성능 저하까지 관찰된다(Z. Ke et al., 2025). 이러한 한계는 모델 내부의 표현 공간이 특정 도메인에 과도하게 최적화되는 표현 편향(representation bias)에서 기인한다는 분석이 제시된다. 또한 합성 데이터 기반 전이 학습 역시 현실 세계의 변동성과 도메인 고유 특성을 충분히 반영하지 못해, 일반화 실패 및 편향 누적 위험을 내포한다(A. Setlur et al., 2024).

다. 대응 접근법과 제약

지식 전이 한계를 극복하기 위해 다양한 방법론이 제안되고 있다. 대표적으로 모델 모듈화(modular architecture)는 특정 도메인 지식을 분리·재활용하여 새로운 영역으로의 적응을 용이하게 하고, 메타러닝(meta-learning)은 적은 샘플로도 빠르게 도메인 적응이 가능하도록 설계된다. 또한 프롬프트 기반 적응(prompt-tuning, in-context learning)은 기존 모델을 재훈련하지 않고도 새로운 도메인에 제한적 적응을 가능하게 한다. 그러나 초거대 LLM에 이러한 기법들을 적용하는 데에는 확장성(scalability) 문제와 안정성 검증 부족이라는 제약이 따른다. 실제로 대규모 벤치마크에서 모듈화·메타러닝 기법은 제한된 태스크에서는 효과적이지만, 범용적 상황에서는 여전히 불안정한 결과를 보인다. 프롬프트 기반 접근 역시 OOD 환경에서는 성능 편차가 크고, 합성 데이터나 편향된 도메인 지식에 의존할 경우 전이 효과가 제한적이다(A. Setlur et al., 2024). 따라서 현 단계에서는 범용성·신뢰성을 보장하는 수준의 전이 학습 기법은 확립되지 않았다는 점이 주요 한계로 지적된다.

라. 시사점

AGI 실현을 위해서는 단순히 특정 도메인에서의 성능 향상이나 파인튜닝을 넘어, 범용 표현 학습(universal representation learning)을 확립하여 지식 전이와 일반화의 효율성을 극대화할 필요가 이를 위해 다음과 같은 과제가 시사된다. 1) 전이 성능 평가를 위한 표준화된 벤치마크 개발, 2) 다양한 도메인을 아우르는 범용 표현 학습 기법 연구, 3) 지속적 학습 환경에서의 안정성 보장 메커니즘 확립이 요구된다. 이를 통해 AGI가 미지의 상황에서도 안정적이고 효율적으로 지식을 적용할 수 있도록 설계 해야 한다.

5. 신뢰성/강건성/안전성 및 목표 정렬 문제 분석

AGI 수준 모델의 신뢰성(reliability)은 단순한 평균 성능이 아니라 환경 변화, 적대적 공격, 목표 정렬 불완전성, 그리고 물리적 환경 전이 상황에서도 일관된 결과를 내는 능력에 의해 결정된다. 그러나 현재 LLM은 훈련 데이터 분포에서 벗어난 Out-of-Distribution (OOD) 입력과 적대적 예제(adversarial examples)에 취약하며, 목표 정렬(alignment)의 불완전성으로 인해 사회적 위험을 수반한다. 더 나아가 시뮬레이션-현실 간 불일치로 인한 물리적 안정성 문제는 AGI의 실세계 적용을 제약하는 또 다른 핵심 병목으로 지적된다.

가. 환경 변화 및 적대적 공격 취약성

언어 모델(Language Model, LM)은 훈련 데이터의 분포에서 벗어난 입력, 즉 환경 변화(distribution shift)에 대해 성능이 급격히 저하되는 강건성(robustness) 부족을 보인다. 이러한 강건성 부족은 모델의 예측 안전성을 저하시켜, AGI의 안전성(safety) 문제로 직결될 수 있다.

이와 관련하여 R. Bommasani et al. 에서는 HELM(Holistic Evaluation of Laguage Models) 평가 체계를 통해 두 가지 핵심 강건성 개념에 주목한다. 첫번째는 불면성(invariacne)으로, 의미를 보존하는 미세한 교란(예: 오타, 문장 부호 변화 등)에 대해 모델의 출력이 얼마나 안정적인지를 측정하는 것이고, 두번째는 등변성(equivariance)으로 의미를 변경하는 교란(예: 문장 내 논리 반전, 대조적 표현 등)에 대해 모델의 출력이 적절하게 변하는지를 평가하며, 모델이 입력의 비관련 요소에 과도하게 의존하지 않는지를 검증한다. HELM 평가 결과에 따르면, 대부분의 대규모 언어 모델은 표준 조건 대비 강건성 및 공정성 교란(robustness/fairness perturbations)이 존재할 때 일관적으로 5~10% 수준의 성능 저하를 보이며, 특히 의미 변화 교란을 사용하는 Contrast Sets 평가에서는 성능 하락 폭이 훨씬 더 크게 나타난다. 예를 들어, Contrast Sets가 적용된 BoolQ와 IMDB 데이터셋에서 모든 모델의 성능이 표준 정확도 대비 크게 감소하는 것으로 나타났는데, IMDB 실험에서 고성능 모델 중 하나인 GLM(130B)은 표준 정확도 95.6%에서 의미 변화 교란 시 86.9%로 하락(약 8% 감소)하였다. 또한 TNLG v2(530B) 모델은 NarrativeQA에서 표준 정확도 72.6%를 기록했으나, 강건성 교란(robustness perturbations)이 추가되었을 때에는 38.9%로 급격히 감소하는 결과를 보였다. 이러한 결과는 대규모 언어 모델이 현실 세계의 미세한 입력 변화나 의미

적 변형에 여전히 취약함을 보여준다.

나. 목표 정렬(Alignment) 방법론의 한계: RLHF와 헌법적 AI(Constitutional AI)

언어 모델은 규모가 커질수록 표현력과 추론 능력은 향상되지만, 사용자의 의도나 가치 체계에 대한 정렬(alignment)가 자동적으로 개선되지는 않는다. 모델은 여전히 허위 정보, 편향된 발화, 유해한 응답을 생성할 수 있으며, 이는 AGI 의 신뢰성과 안전성을 저해하는 주요 병목 요인 중 하나로 지적된다. 이러한 문제를 완화하기 위해 대표적 접근으로 인간 피드백 강화 학습(Reinforcement Learning from Human Feedback, RLHF)과 헌법적 AI(Constitutional AI, CAI)가 있지만, 두 방법 모두 구조적 한계를 내포한다.

인간 피드백 강화학습(RLHF)는 인간 평가자의 피드백을 통해 모델 출력을 사용자 의도에 맞게 조정하는 대표적인 정렬 방법이다(D. Krueger, 2023). 실제로 OpenAI의 연구 결과에 따르면, RLHF로 미세 조정된 InstructGPT(1.3B) 모델은 GPT-3(175B)보다 인간 평가에서 더 높은 선호도를 기록했으며, 진실성(truthfulness) 측면에서도 개선을 보였다. 하지만 RLHF는 인간 피드백을 활용하는 특성상 몇가지 한계를 가진다. 먼저, 피드백의 품질과 다양성에 대한 의존성 문제가 있다. 인간 평가자의 가치관, 문화, 편향에 따라 정렬 방향이 달라질 수 있다. 또한 단기적 순응(short-term alignment) 문제가 있다. 짧은 대화나 단일 작업에서는 효과적이지만, 장기 대화나 다단계 추론 상황에서는 모델의 의도 정렬이 점진적으로 붕괴된다는 것이다. 마지막으로 비용과 확장성의 제약이 있다. 고품질 피드백 확보에는 상당한 인적 자원과 비용이 소요되므로, AGI 수준의 광범위한 도메인 커버리지를 달성하는 것은 사실상 불가능하다.

헌법적 AI(CAI)는 RLHF의 인간 의존도를 줄이기 위해 제안된 접근으로, 인간의 개입을 '헌법(constitution)'이라 불리는 원칙 목록 제공으로 제한한다(Bai, Y et al., 2022). 모델은 이 헌법을 바탕으로 자기 비판(self-critiques)과 자체 수정(revisions)을 수행하며, 인간 대신 AI 스스로 피드백을 생성하는 AI 기반 정렬(self-alignment) 과정을 거친다. 이 방법은 정렬 비용을 줄이고, 규범적 일관적을 유지한다는 장점이 있으나, 여전히 한계가 존재한다. 우선, 모델에 제공하는 헌법 또한 인간이 사전에 정의한 규칙에 기반하기 때문에 복잡한 윤리적 상황이나 맥락 변화에 충분히 대응하지 못하는 불완전성이 남아 있다. 또한 모델이 스스로 생성한 피드백을 반복 학습함으로써 오류나 편향이 강화되는 자기 증폭 위험(Self-reinforcement risk)이 내제되어 있다.

따라서 RLHF와 CAI는 현재 AGI 정렬 문제 해결의 주요 경로이지만, 장기적이고 맥락적인 정렬은 여전히 미해결 상태라는 점에서 AGI 실현에 본질적 병목 요인으로 남아 있다.

다. 물리적 환경에서의 안정성 문제

AGI 수준의 인공지능이 로봇 제어, 자율 시스템 등 물리적 환경에 적용될 때는 시뮬레이션 단계에서의 성능이 실제 환경으로 전이(sim-to-real)되는 과정에서 급격히 저하될 수 있다(W. Zhao et al., 2020). 실제 로봇 데이터 수집은 샘플 비효율성, 높은 비용, 안전 위험 등의 제약으로 시뮬레이션에 의존하지만, 시뮬레이션과 현실 간의 동역학 및 감지 불일치(mismatch)는 정책 전이 시 성능 저하와 제어 실패를 초래한다. 특히 감지(sensing) 단계의 불일치는 시각 센서의 미세한 노이즈나 적대적 공격(adversarial perturbation)에도 모델의 판단이 왜곡될 수 있음을 보여준다. 이러한 입력 분포 취약성은 현실 환경에서의 안정적 동작을 보장하기 어렵게 만든다.

이를 완화하기 위해 대표적 접근으로 도메인 무작위화(domain randomization)와 도메인 적응(domain adaptation)이 활용되지만, 극단적 환경 변화나 복합 상호작용 상황에서는 여전히 한계가 있다. 따라서 AGI의 물리적 적용에서는 도메인 전이 강건성 테스트와 안정성 시뮬레이션 기반 검증, 더 나아가 메타 학습(meta learning) 및 지속 학습(continual learning)을 통한 환경 적응 능력 강화가 필수적이다. Sim-to-Real 전이의 불안정성과 감지 단계의 취약성은 AGI의 실세계 신뢰성 확보를 가로막는 핵심 병목 요인으로 평가된다.

라. 시사점

AGI의 신뢰성, 강건성, 안전성 및 목표 정렬 문제를 해결하기 위해서는 국가적 차원, 연구적 차원의 노력이 모두 필요하다. 국가 차원에서는 모델 강건성 및 정렬 평가 기준을 법제화하는 것이 필요하며, 도메인 전이 안전성 검증 절차의 의무화를 추진해야 한다. 동시에 연구 차원에서는 환경 변화와 적대적 교란, 시뮬레이션-현실 간 전이 불일치 등 다양한 변동 요인을 고려한 표준화된 강건성 및 안정성 벤치마크 개발이 필요하다. 또한 RLHF와 CAI의 구조적 한계를 보완하기 위해 비용 효율적이고 문화적 다양성을 반영한 피드백 수집 체계 및 지속 학습, 메타 학습 기반의 자가 정렬(self-alignment) 연구가 병행되어야 한다. 궁극적으로 이러한 기술은 AGI가 환경 변화와 사회적 가치 충돌 속에서도 예측 가능하고 신뢰 가능한 행동을 유지하는 체계적 안전장치로 기능해야 할 것이다.

6. 해석 가능성 및 투명성 부족 문제 분석

AGI의 해석 가능성(Explainability)과 투명성(Transparency)은 신뢰성, 안전성, 법적 책임성 확보를 위해 필수적이다. 그러나 현재 초거대 모델의 내부 의사결정 과정은 '블랙박스'로 남아 있으며, 특정 답변이 어떤 경로 및 패턴을 통해 생성되었는지, 어떤 학습 데이터가 근거가 되었는지를 명확히 추적하기 어렵다.

가. 기존 XAI 기법의 한계

SHAP, LIME, Integrated Gradients 등 기존 설명 가능 인공지능(XAI) 기법은 단일 입력에 대한 부분적 설명을 제공할 수 있지만, 본래 단순 예측 모델이나 지역(local) 수준 설명을 목표로 설계되었기 때문에 AGI 수준의 장기·다단계 추론을 설명하기에는 근본적 한계가 있다. 특히 LIME은 무작위 표본화 과정에서 설명 결과가 변동될 수 있어 일관성이 낮으며, 고차원 데이터에서는 설명력이 저하되는 문제가 있다(M. T. Ribeiro et al., 2016). SHAP은 이론적으로 안정적인 해석을 제공하지만, 계산 복잡성이 높아 초거대 모델 적용 시 막대한 연산 자원 부담을 요구한다(S. M. Lundberg and S.-I. Lee, 2017). Integrated Gradients는 입력 변화 경로에 민감하며, 복잡한 피처 상호작용을 완전하게 포착하기 못한다(M. Sundararajan, 2017). 더불어, 초거대 모델에서는 해석 가능성 알고리즘 자체가 막대한 연산 자원을 요구하고, 복잡한 내부 상호작용을 단순화하여 설명하는 과정에서 정보 손실이 불가피하다.

나. 투명성 부족과 책임성 문제

해석 가능성의 한계는 곧 투명성 부족으로 이어진다. 의료·법률·금융 등 고위험 분야에서 AGI가 잘못된 결정을 내렸을 때, 원인 분석과 책임 규명을 위해서는 모델 구조, 학습 데이터, 결정 경로가 공개되어야 한다. 그러나 OpenAI, Google DeepMind 등 주요 개발사는 상업적 비밀·보안상의 이유로 학습 데이터, 모델 파라미터, 설계 세부사항을 공개하지 않고 있다(C. Roscher et al., 2021).(M. Chen, 2024),(C. Roscher et al., 2021)에서는 이러한 비공개 관행이 모델 오남용 위험 평가와 안전성 검증을 어렵게 만든다고 지적하며, 규제기관과 연구자 간의 협력 필요성을 강조한다.

다. 시사점

AGI 신뢰성과 투명성 확보를 위해서는 기존 XAI 기법의 한계를 넘어서는 통합형 해석 가능성 프레임워크 구축이 필요하다. 특히 다단계 추론, 동적 의사결정, 멀티모달 입력을 포괄하는 설명 구조를 설계하고, 설명 가능성을 단순 부가 기능이 아닌 모델 설계의 내재적 목표로 포함해야 한다. 이를 위해 모델 내부 상태를 단계적으로 추적하는 계층적 추적(hierarchical tracing), 체인 오브 소트(chain-of-thought) 정렬 기반 해석 기법, 모델 내부 피처 간 상호작용 가시화 기술 등이 유망한 접근으로 제시된다. 이러한 기술적 방향은 정보 손실을 최소화하면서 모델의 의사결정 과정을 인간이 이해할 수 있는 형태로 재구성하는 데 중점을 둔다.

제2절 산업적 병목 요인 식별

AGI 산업화 과정에서의 인적 자원 부족, 글로벌 기술/인프라 독점, 법·제도 미비 등 AGI 실현에 영향을 미치는 주요 산업적 병목 요인을 국내외 현장 사례, 정책자료, 통계분석 등을 통해 다각적으로 식별하고 분석함

1. 인적 자원 및 교육 체계 병목 분석

AGI 실현의 핵심은 고성능 컴퓨팅과 대규모 데이터뿐 아니라, 이를 설계하고 운영 및 확장할 수 있는 고급 인적 자원과 지속 가능한 교육 및 연구 생태계 확보에 있다. 한국은 AI 인프라 및 반도체 제조 분야에서 높은 경쟁력을 보유하고 있음에도 불구하고, 인재 양성 및 교육 체계 전반에서 구조적 병목 현상을 보인다. 이는 장기적으로 AGI 시대의 기술 자립과 산업 경쟁력 확보에 제약 요인으로 작용할 가능성이 있다.

가. 고급 AI 연구 인력의 부족 및 수급 불균형 심화

The Global AI Index 2024에 따르면, 한국은 AI 종합 순위 6위를 기록하며 기술 인프라 부문(6위)에서는 선진국 수준의 경쟁력을 확보했으나, 인재(Talent) 부문은 13위로 상대적으로 낮은 평가를 받았다(한국과학기술기획평가원(KISTEP), 2024). 세부적으로는 과학자 11위(표준점수 0.87), 개발자 16위(0.51), 전문가 13위(0.34)로 모두 10위권 밖에 머물렀다. 이는 AI 연구·개발·응용 단계에서 주도적 역할을 할 수 있는 고급 AI 기술 인력의 양적·질적 부족을 의미한다. 특히 AGI 핵심 기술 영역인 대규모 언어 모델(LLM), 멀티모달 AI, 자기지도학습(Self-supervised learning), 강화학습(Reinforcement learning) 등 고난도 분야에서 석·박사급 연구자층의 협소함과 전문성 부족이 뚜렷하다.

한국과학기술한림원의 제236회 한림원탁토론회(2024)에서는 AI 고급 인력의 해외 유출 심화와 연구처우 격차 확대가 국가 차원의 중대한 리스크로 지적되었다(한국과학기술한림원, 2025). 북미·유럽 주요 AI 연구기관은 막대한 연구비, 대규모 데이터 접근성, 고성능 연산 자원, 높은 보상 체계를 기반으로 전 세계 인재를 흡수하고 있다. 반면 한국은 이에 상응하는 연구환경을 제공하지 못해 박사급 연구자의 해외 이탈이 지속적으로 증가하고

있으며, 국내 연구자들은 "보상·자율성·연구지원의 부족"을 가장 큰 이탈 요인으로 꼽고 있다. 이러한 추세는 단순한 두뇌 유출(Brain Drain)을 넘어, AGI 연구를 선도할 지식 핵심층의 축소로 이어질 가능성이 크다.

KISTEP의 '주요국의 AI 연구기반 구축 현황과 시사점' 보고서에 따르면, 한국은 연구기관 간 AI 전문인력 순환 구조가 미비하며, 교육기관-산업-연구소 간의 인력 교류도 매우 제한적이다(박혜영, 2025). 미국의 NSF AI Institute나 영국의 Turing Institute, 캐나다의 CIFAR AI Program 등은 국가 차원의 AI 허브를 통해 연구-교육-산업 인력이 순환하는 구조를 구축하고 있다. 반면 대한민국은 기관·과제 단위의 단절적 인력 구조로 인해 장기적 연구 역량 축적이 어려운 구조적 제약을 겪고 있다.

나. 교육 및 연구 생태계의 한계

주요 선진국은 대학-국가연구소-산업체가 연계된 AI 교육·연구 복합거점(AI Hub) 체계를 운영하고 있다(박혜영, 2025). 예를 들어 미국은 AI Institutes를 통해 매년 20여 개 기관이 공동으로 핵심 인재를 양성하고, 영국의 Alan Turing Institute는 대학 연합체 기반의 AI 석·박사 공동과정을 운영한다. 반면 한국은 단기 사업 중심의 분절적 인재 양성 체계에 머물러 있으며, 대학 중심·단기 교육 위주의 구조로 인해 고급 연구자의 연속적 육성과 장기 연구 인프라 구축이 어렵다.

국가 차원의 고성능 컴퓨팅 자원 및 오픈 데이터 플랫폼이 존재하더라도, 교육기관과의 연계성이 낮아 학생·연구자가 실제로 모델 학습·실험에 접근하기 어려운 구조가 지속되고 있다. 이로 인해 AGI 관련 핵심 연구—예를 들어 대규모 모델 학습, 시뮬레이션-현실 전이 검증, 안전성 연구 등—이 대학이나 공공기관 차원에서 원활히 수행되지 못하고 있다.

다. 시사점

AGI 실현을 위해서는 핵심 AI 인재의 질적 강화, 교육·연구·산업 간 연계, 기초 연구역량 확충이 병행되어야 한다. 우선 박사급 중심의 장기 인재 양성 및 안정적 연구직 제도를 마련하고, 보상과 자율성 제고, 국제 공동연구 지원 확대를 통해 해외 유출을 방지해야 한다. 또한 대학-연구소-산업체 간 순환형 인력 교류 구조와 융합대학원·산학 공동과정을 확충하여 교육과 연구가 산업 현장과 지속적으로 연결되는 생태계를 구축해야 한다. 마지

막으로 기초 AI 및 안전성, 지속학습 분야의 장기 도전형 R&D 투자를 강화하고, CIFAR, ELLIS, Turing Institute 등과의 국제 협력 네트워크를 통해 글로벌 연구 생태계를 확대할 필요가 있다.

2. 기술 독점 및 글로벌 공급망 병목 분석

AGI의 발전은 기술적 성취뿐 아니라, 이를 지속 가능하게 뒷받침하는 산업 생태계의 구조적 안정성에 의존한다. 최근 반도체 및 AI 산업은 생성형 AI 확산에 힘입어 고성능 메모리와 AI 가속기 중심으로 빠르게 성장하고 있으나, 그 이면에서는 기술 및 자본의 집중과 공급망 불균형이 심화되고 있다. 특히, 초거대 AI 모델 학습에 필요한 컴퓨팅 자원과 첨단 반도체 공정이 소수 글로벌 기업 및 특정 지역에 편중되면서, AGI 연구개발의 접근성이 제한되고 시스템적 리스크가 확대되는 양상이 나타난다. GPU 및 병렬 컴퓨팅 소프트웨어 생태계의 사실상 독점화, 대만·네덜란드 등 특정 지역에 집중된 반도체 제조·장비 공급 구조는 기술 발전의 속도를 높이는 동시에 불안정성과 종속성을 내재화하고 있다.

가. 기술 및 자본 집중에 따른 기업 독점

최근 AGI 개발의 주요 자원인 컴퓨팅 인프라, 인재, 자본이 소수 글로벌 빅테크 기업, 이른바 '하이퍼스케일러(Hyperscaler)' 기업에 고도로 집중되는 현상이 나타나고 있다. 2024년 기준 마이크로소프트, 알파벳, 메타, 아마존 등 주요 빅테크 기업들은 AI 인프라에 총 1,900억 달러 이상을 투자한 것으로 추정되며, 이는 대부분의 학술기관이나 신흥기업이 감당하기 어려운 수준이다(A. Mocle, 2025). 이러한 자본 집중은 사실상 AI 연구개발의 진입 장벽을 높이고, 대규모 모델 학습을 수행할 수 있는 주체의 범위를 제한하는 구조적 요인으로 작용한다.

또한 이들 기업은 AI 스타트업에 투자하면서 동시에 클라우드 자원을 공급하는 '투자-공급'의 이중 구조를 형성하고 있다(A. Mocle, 2025),(J. Vipra and S. M. West, 2023). 즉, 스타트업에 대한 투자금이 다시 컴퓨팅 자원 이용료 형태로 하이퍼스케일러의 클라우드 사업으로 환류되는 순환 구조가 존재하는 것이다. 이러한 순환 구조는 산업 전반의 지대 추구(rent-seeking) 성향을 강화하며, 결과적으로 소수 기업이 AGI 연구 생태계 방향성을 간접적으로 통제할 수 있는 구조적 영향력을 확보하게 만든다(J. Vipra and S. M. West, 2023).

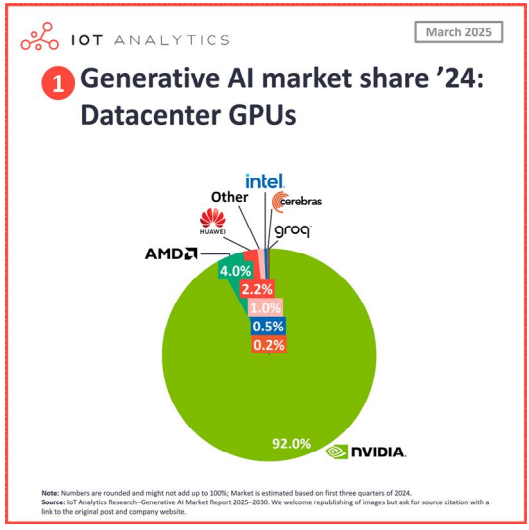
이는 연구 자원의 효율적 배분보다는 시장 지배적 사업자의 경제적 이해관계에 부합하는 연구 방향이 강화되는 결과로 이어질 수 있다.

나. GPU·소프트웨어 생태계의 집중과 컴퓨팅 접근성 격차

AGI 개발의 주요 자원인 고성능 GPU와 병렬 컴퓨팅 소프트웨어 생태계는 현재 소수의 글로벌 기업을 중심으로 집중되어 있다. 글로벌 데이터센터용 GPU 시장은 엔비디아(NVIDIA)가 약 92% 이상을 점유하고 있으며, 그 성능과 CUDA 기반 생태계의 높은 호환성을 바탕으로 AI 연구의 사실상 표준 인프라로 자리 잡았다(J. Fernandez, 2025). 이러한 시장 구조는 하드웨어 경쟁력뿐 아니라, 20년 이상 축적된 CUDA·cuDNN·TensorRT 등 독점적 병렬처리 플랫폼과 개발 툴체인을 중심으로 구축된 소프트웨어 생태계에 의해 더욱 강화되고 있다. 이 플랫폼은 방대한 라이브러리와 개발자 커뮤니티를 보유하고 있어, AI 모델 학습 및 추론의 표준 인터페이스로 작동한다. 이에 따라 AMD(Instict 시리즈)나 인텔(Gaudi, Ponte Vecchio 시리즈) 등 주요 경쟁사 및 신규 업체가 자체 하드웨어를 개발하더라도 호환성 확보에 막대한 시간과 비용이 소요되며, 결과적으로 기술 생태계가 단일 구조로 수렴되는 경향을 보인다.

또한 GPU 공급은 TSMC 중심의 첨단 반도체 제조 병목과 결합되어 수요 대비 공급이 지속적으로 부족한 상황이다. 2024년 기준, H100·B200 등 고성능 AI GPU는 평균 3~6개월의 리드타임을 보이며, 개당 가격은 수만 달러 수준으로 형성되어 있다(A. Thadani and G. C. Allen, 2023). 이러한 희소성은 연구기관·중소기업·공공부문의 컴퓨팅 자원 접근성 불균형을 심화시키며, 산업 전반의 기술 주권 확보 및 다양성 유지에도 영향을 미친다.

〔그림 6－2〕 2024년 생성형 AI용 데이터센터 GPU 시장 점유율



자료: IoT Analytics Research Brief(2025)

다. 첨단 반도체 공급망의 지정학적 취약성

AI 반도체는 AGI 연구의 물리적 기반이며, 특히 고성능 메모리(HBM)와 첨단 패키징(CoWoS 등), EUV(극자외선) 노광 기술이 필수적이다. 그러나 이러한 공정과 장비는 극도로 제한된 지역과 기업에 집중되어 있어, 공급망 전반이 지정학적 불안정성에 취약한 구조를 갖고 있다.

2024년 기준, 개만의 TSMC는 전 세계 첨단(10nm 이하) 반도체의 약 90% 이상을 생산하고 있으며, AI용 칩의 핵심 공정 대부분을 담당하고 있다(A. F. Gomes and J. van den Wijngaard, 2025). 네덜란드 ASML은 EUV 노광 장비 시장을 독점하고 있어, 해당 장비의 생산 차질이나 수출 규제가 곧 글로벌 AI 반도체 생산에 직접적인 영향을 미친다(A. F. Gomes and J. van den Wijngaard, 2025). 이러한 공급망 집중은 기술적 효율성 측면에서는 유리하지만, 동시에 단일 실패 지점(Single Point of Failure)을 만들어낸다. 지정학적 분쟁, 자연재해, 혹은 특정 장비 공급 중단만으로도 글로벌 AI 산업의 생산 라인이 정지될 가능성이 존재한다. 특히 대만 해협의 지정학적 긴장은 AGI 연구의 연속성과 예측 가능성을 제약하는 중요한 리스크 요인으로 평가된다.

한편, AI 반도체 수요 급증으로 HBM(High Bandwidth Memory) 시장은 2025년 전체 DRAM 매출의 5% 이상을, 2028년까지 30.6% 차지할 것으로 전망된다(GO OVER AI, 2025). SK하이닉스, 삼성전자, 마이크론 등 주요 기업들은 생산 능력 확대와 기술 고도화 경쟁을 가속화하고 있으나, 여전히 일부 소재, 공정, 장비는 특정 해외 기업 의존도가 높다. 이러한 구조는 국가 단위의 기술 자립성 확보를 어렵게 만드는 요인으로 작용한다.

라. 시사점

기술 독점과 공급망 집중은 AGI 발전을 가속화하는 한편, 접근성 불균형과 지정학적 리스크를 동반한다. 이러한 구조적 병목을 완화하기 위해 다음과 같은 정책적 대응이 필요하다. 첫째, 공공 컴퓨팅 인프라 확충을 통해 연구기관과 중소기업의 GPU 접근성을 제고하고, 오픈소스 기반 병렬처리 프레임워크를 지원하여 특정 생태계 의존도를 완화해야 한다. 둘째, 공급망 다변화 및 라질리언스 강화를 추진해야 한다. 첨단 패키징, 소재, 테스트 공정 등 국내 강점을 중심으로 기술 자립을 확대하고, 동맹국 간 공급 안정화 협력체계를 구축해야 한다. 셋째, 대안적 기술 패러다임 연구를 지원해야 한다. 초대규모 컴퓨팅 투입에 의존하지 않는 알고리즘 효율화, 온디바이스 AI, 경량 모델 등 자원 절약형 AI 기술 개발을 장기적으로 육성해야 한다. 마지막으로, 국제 협력 차원에서 글로벌 기술 표준 및 윤리 프레임워크를 마련하여 기술 독점이 공공 이익을 저해하지 않도록 관리할 필요가 있다. AGI 실현을 위한 기술적 진보는 산업 구조의 균형과 공급망의 안정성을 전제로 해야 한다. 산업적 집중과 지정학적 의존을 완화하는 전략은 단순한 경제 경쟁력을 넘어, 국가 차원의 AI 주권과 기술안보를 지키는 수단이 될 것이다.

제 7 장 산업별 AGI 기술 활용 현황 및 중기(5년후), 장기(10년후) 산업별 AGI 전망 및 경쟁력 확보를 위한 대응 전략

제 1 절 반도체/제조

AGI 시대를 대비한 반도체 제조 및 메모리 산업의 경쟁력 확보 전략은 기술 혁신뿐 아니라 생태계, 인재, 공급망, 정책 등 다층적인 접근이 필요하다. 기업들은 고도로 전문화된 하드웨어로 전환하고, AGI를 활용해 연구개발과 제조를 가속화하며, 혁신적인 메모리 및 패키징 기술을 개발해야 한다.

1. AGI 시대를 대비한 반도체 제조 및 메모리 산업의 경쟁력 확보 전략

1.1. 중기(5년 후) 반도체/제조 분야 AGI 활용 전망

가. 특수 목적 칩을 위한 AI 기반 설계 자동화 구현

생성적 적대 신경망(GAN) 및 변이형 자동 인코더(VAE)와 같은 고급 머신러닝 모델을 활용하는 생성적 AI 도구는 아키텍처 탐색, 회로 최적화, 레이아웃 생성을 포함한 다양한 설계 단계를 자동화하고 최적화할 수 있다. 또한, 데이터 가용성, 모델 해석 가능성, 그리고 AI가 생성한 설계를 기존 검증 워크플로에 통합할 수 있다. 즉 생성적 AI가 아키텍처 탐색, 회로 최적화, 레이아웃 생성 등 복잡한 작업을 자동화하므로써 설계 역량을 향상시키고, 개발 비용을 절감하며, 반도체 기술 혁신을 가속화할 수 있다.(Prashis Raghuwanshi, 2025)

나. 예측 제조 및 수율 최적화를 위한 AI 도입

AI로 제조 장비의 방대한 데이터셋을 분석하여 패턴을 식별하고 고장을 예측하므로써 제조 최적화에 핵심적인 공정-구조-특성-성능(PSPP) 관계에 대한 종합적 이해를 촉진할 수 있다. 결론적으로 머신러닝 모델과 AI는 반도체 제조 분야에서 예측 유지보수 및 결함

탐지에 활용되고 있고, 가동 중단 시간을 최소화하고 수율을 극대화하는 완전 자율적·자가 최적화형 '스마트 팹'을 구현할 수 있고, 더불어 데이터 세트 품질, 모델 일반화, 자율 실험과 관련된 과제를 해결할 수 있다.(Yiqiang Zheng et al. 2025)

다. AGI 워크로드를 위한 HBM 개발 가속화

생성형 AI 애플리케이션의 확대와 폭증으로 인해 고성능 메모리, 특히 고대역폭 메모리(HBM)에 대한 수요가 급증하고 있다. 그러므로 AGI시대에 컴퓨팅 서비스 수요를 충족시키기 위해 HBM 기술 혁신에 집중해야 한다. 여기에는 더 높은 용량의 스택 개발과 성능 한계 돌파가 포함된다. 차세대 HBM을 시작으로 가치 사슬의 변화는 요소 기술 공급업체의 범위를 확대하고 새로운 비즈니스 모델을 창출해야 한다.(Mitsui & Co Report, 2025)

라. 첨단 패키징 기술 고도화

칩릿(chiplet) 및 3D 적층과 같은 첨단 패키징 기술은 AGI 시대의 핵심 차별화 요소가 될 것이다. 2030년까지 AI 수요에 힘입어 첨단 패키징 시장은 약 800억 달러 규모로 성장할 전망이다. 팬아웃 패널 레벨 패키징(FOPLP) 및 유리 코어 기판과 같은 기술에 투자하면 더욱 강력하고 통합된 시스템을 구현할 수 있다. AI가 고밀도·이종 통합을 위한 혁신적 기술을 요구하며 고급 패키징 전환을 가속화하고 있으며, 이종 통합을 통한 실리콘과 패키지의 공동 설계는 HPC 및 AI의 성능 요구사항 달성에 핵심이다.(Yole Group Report, 2025)

마. AGI 협업을 위한 인력 역량 강화 우선순위 설정

AGI의 부상은 반도체 인력의 전환을 요구할 것이며, AI 도구와 효과적으로 협업할 수 있는 엔지니어 및 연구원에 대한 수요가 증가할 것이다. 중기 전략에는 복잡한 조사 및 전략적 과제와 같이 AGI가 전투력 증강 요인으로 작용하는 분야에서 인간의 전문성을 개발하기 위한 맞춤형 교육 프로그램이 포함되어야 한다. 2025년 세계경제포럼(WEF)은 AGI가 인간 분석가가 고차원적 업무에 집중할 수 있도록 하는 강력한 증폭 장치 역할을 할 수 있음을 논의했다. 맥킨지 앤 컴퍼니가 강조한 바와 같이, 이러한 기술의 잠재력을 최대한 활용하기 위해서는 인공지능 및 기계학습(AI/ML) 교육이 필수적이다. 특히 AGI 시대 사이버 보안과 위협 환경에 대한 대응 및 혁신 방안은 우수한 인력의 사전 양성을 시작으로 준비

해야한다.(World Economy Forum, Annual Meetings of the Global Future Councils and Cybersecurity, 2025)

1.2. 장기 전략(10년 후) 반도체/제조 분야 AGI 활용 전망

가. AI 네이티브 메모리 연구 추진

장기적으로 기존 메모리 아키텍처를 넘어서는 것이 핵심이 될 것이다. AGI를 위해 근본적으로 설계된 AI 네이티브 메모리에 집중이 요구된다. 이는 모든 유형의 메모리를 매개변수화하고 압축하는 심층 신경망 모델을 구축하는 것을 포함하며, 컨텍스트 윈도우의 대폭적 확장과 "거의 무한한 메모리 용량"으로 이어질 수 있다. 최근 문헌에 따르면 거의 무제한적인 맥락 길이를 가진 LLM이 AGI를 실현할 수 있다고 주장하면서, LLM의 유효 맥락 길이는 주장된 맥락 길이보다 상당히 짧으며 추론 실험으로 부터 긴 맥락에서 관련 정보를 찾는(단순) 추론을 동시에 수행하는 것이 거의 불가능하다고 보고하고 있다. 궁극적으로 모든 에이전트/개인은 자신만의 대규모 개인 모델, 즉 자연어로 설명할 수 없는 메모리를 포함하여 모든 유형의 메모리를 매개변수화하고 압축하는 심층 신경망 모델(따라서 AI 네이티브)을 가져야 한다고 주장한다.(Xiaofei Sun, Xiaokang Wu, 2024).

나. 양자역학 기반 신소재 개발

신소재 발굴을 위한 기존의 시행착오 방식은 시간 소모적이고 비효율적이다. 장기 전략으로는 양자역학 기반 AGI를 목표로 양자 암호화 및 센싱 분야를 포함한 새로운 반도체 소재와 양자 장치를 발견하고 혁신하는 것이 필요하다. 최근 한 반도체 산업 보고서에 따르면 AI 기반 방법이 특히 반도체 스크리닝 및 장치 성능 최적화와 같은 분야에서 재료 발견 파이프라인을 가속화할 수 있음을 피력하고 있다.(New General Group, Quantum Mechanics-Informed AI for Semiconductor, 2025) 이 보고서에서는 또한 반도체 응용 분야를 위한 양자역학 기반 AI 시스템은 양자역학의 기본 원리에서 도출된 물리적 제약 조건을 통합하는 특수 신경망 아키텍처를 사용하며, 이러한 시스템은 반도체 소자 작동을 지배하는 기본 물리 법칙을 반영하는 도메인별 유도 편향을 내장한다는 점에서 기존의 머신러닝 접근 방식과 근본적으로 다름을 보고한다. 결론적으로 양자 파동 함수의 간결한 표현 학습, 2차 계산 모듈은 소자의 기하학적 구조, 재료 조성, 도핑 프로파일 및 그에 따른 전류-전압 특성 간의 매핑을 학습하여 전자 전달 문제를 해결 등의 새로운 기술 프레임워

크 및 구현 아키텍처 등 양자역학 기반 AI 시스템을 위한 전략이 필요하다.

다. 완전 자율적 반도체 생산 시설 구축

장기적으로 반도체 산업은 "라이트아웃" 팹으로 알려진 완전 자율 생산 시설로 전환될 것입니다. 이러한 시설은 AGI를 활용해 재료 취급부터 공정 미세 조정까지 제조의 모든 측면을 최소한의 인간 개입으로 제어하고 최적화할 것이며, 이것은 자율적 자재 처리 및 실시간 장비 조정 등 반도체 생산 시설의 완전 자동화와 자급자족적인 제조 공정으로 운영 비용과 인적 오류를 줄이는 전략이 될 수 있다. AI 기반 기술로 혁신할 완전 자율 생산 항목으로는 AI 기반 웨이퍼 결함 감지 및 분류, 리소그래피 및 에칭 장비의 예측 유지보수, 클린룸에서의 자율 자재 취급, AI 기반 R&D 실험 및 실시간 장비 튜닝을 위한 엣지 AI 등이 있다.(ICT-Strypes, 2025 and Beyond, 2025)

라. 공동 패키징 광학(CPO) 및 실리콘 포토닉스 통합

기존 전기적 상호 연결 방식의 한계는 AGI 시스템의 병목 현상으로 작용할 것이다. 장기적 전략은 공동 패키징 광학(CPO) 및 실리콘 포토닉스를 반도체 패키지에 직접 통합하는 것이다. 이를 통해 칩과 메모리 간 초고속 데이터 전송이 가능해져 AGI에 필요한 방대한 데이터 흐름을 지원할 수 있다. 최근 한 민간 연구소에서는 반도체 기업들이 첨단 실리콘 포토닉스 통합과 같은 최첨단 제조 공정에 투자할 것을 권고하며, 첨단 패키징 보고서에서 공동 패키징 광학 기술의 통합을 업계의 핵심 차세대 솔루션으로 강조했다.(Capgemini Research Institute Report, 2025)

마. 지속 가능하고 탄력적인 글로벌 공급망 구축

AGI가 막대한 수요를 주도하게되면 반도체 산업은 더욱 견고하고 지속 가능한 공급망이 필요할 것이다. 이것은 자원 배분을 최적화하고, 공급망 리스크를 관리하며, 환경적 영향을 줄이는 방안이 포함되어야 함을 의미한다. 즉 공급망 차질을 예측하고 물류를 자동화하며, 효율성과 지속 가능성 모두를 위해 원자재 조달을 최적화할 것을 요구한다. 산업계가 다양한 소재 통합의 복잡성을 해결하면서 환경적 영향을 완화해야 하는 압박에 직면할 수도 있으며, AGI 시대에 공급망 회복탄력성과 지속가능성에 대한 집중이 강조된다(Capgemini Research Institute Report, 2025).

제 2 절 바이오/의료

본 절에서는 AGI 기술이 바이오 및 의료 산업에서 현재 어떻게 활용되고 있으며 중장기적으로 발전할지 살펴보고자 한다. 또한, 중장기적 발전 전망에 따라 경쟁력 확보를 위한 전략도 함께 고려해보고자 한다. 이를 위해 먼저 바이오 및 의료 산업 분야의 기술을 분류하고 각 세부 분야별로 AGI 기술의 활용에 대해 분석한다.

1. 바이오 및 의료 산업 분류

바이오 및 의료 산업은 생명 과학과 공학 기술이 융합된 광범위한 분야라 할 수 있다. 이에 대한 활용 현황 및 전망을 살펴보기 전에 먼저 바이오 및 의료 산업의 세부 분류 체계를 살펴보고자 한다.

바이오 및 의료 산업을 분류하는 방법은 여러 가지가 있으나 본 보고서에서는 OECD와 EU 바이오협회에서 제안한 분류 체계를 고려한다. 이 분류 체계에 따르면 각 세부 분야가 상징하는 대표적인 이미지를 색상으로 각각 표현하고 있다. 먼저 붉은색의 혈액을 상징하는 레드 바이오(Red Bio)는 인간의 건강과 생명, 그리고 의료 분야를 대표한다. 다음으로 식물의 녹색에서 유래한 그린 바이오(Green Bio)는 바이오 농업, 식량 및 자원과 관련된 분야를 의미한다. 마지막으로 미생물을 이용한 청결하고 효율적인 산업 공정을 상징하는 화이트 바이오(White Bio)는 산업 생산, 에너지, 그리고 환경 정화 분야를 나타낸다. 이 분류 체계는 전세계적으로 통용되는 바이오 및 의료 산업의 표준 분류법처럼 사용되고 있다(기술과 혁신, 2020).

여기서는 바이오 및 의료 산업 분류 체계를 레드, 그린, 바이오 세 가지로 구분하여 각 분야를 구성하는 세부 제품 또는 산업군을 상세하게 살펴본다.

가. 레드 바이오(Red Bio): 보건 및 의료 분야

레드 바이오는 인체의 건강과 질병 치료를 목적으로 하는 바이오 및 의료 기술 분야이다. 일반적으로 3가지 분야 중 가장 큰 비중을 차지한다고 알려져 있으며, 의약품. 의료 기기, 헬스 케어 및 의료정보시스템 등을 포함하고 있다. 여기서는 이들 3가지 세부 분류에 대해 자세히 살펴본다.

1) 바이오 의약품(Biopharmaceuticals)

바이오 의약품은 생물체의 유전자나 단백질 등을 이용하여 만드는 의약품을 의미한다. 바이오 의약품 분야는 〈표 7-1〉에서 보여주고 있는 바와 같이 세포 치료제 및 유전자 치료제, 항체 의약품, 백신, 단백질 의약품, 그리고 혈액 제제 등을 포함하고 있다.

〈표 7-1〉 바이오 의약품 분야

세부 분야	설명
세포 치료제 및 유전자 치료제	살아있는 세포나 유전물질을 인체에 투여하여 질병을 치료하는 의약품
항체 의약품	특정 항원(질병 원인 물질)에만 반응하는 항체를 이용한 치료제
백신	병원체의 항원을 이용하여 면역 반응을 유도, 질병을 예방하는 의약품
단백질 의약품	치료 효과가 있는 단백질을 재조합 기술로 대량 생산한 의약품
혈액 제제	혈액 성분을 분리, 정제하여 만든 의약품

2) 의료 기기(Medical Devices)

의료 기기 분야에는 질병의 진단이나 치료 및 예방을 위해 사용되는 기기나 장치 또는 재료 등을 포함한다. 세부적으로는 크게 진단 기기와 치료 기기로 구분할 수 있으며 먼저 진단 기기와 치료 기기의 세부적인 내용은 〈표 7-2〉에서 보여주는 바와 같다.

의료 기기에는 인체 검체를 분석하여 질병을 진단하는 체외 진단 기기, 인체 내부를 영상으로 시각화하여 진단하는 의료 영상 기기, 그리고 각종 생체 신호를 계측하는 생체신호 측정기기 등으로 구성된다. 치료 기기에는 질병을 치료하고 회복을 보조하는 수술 및 지료 기기, 방사선 치료기, 그리고 재활 및 보조기기 등을 포함하고 있다.

〈표 7－2〉 의료 기기-진단 및 치료 기기 분야

세부 분야		설명
진단 기기	체외진단기기	혈액, 소변 등 인체 검체를 이용해 질병을 진단하는 기기
	의료영상기기	인체 내부를 시각화하여 진단하는 장비(X-ray, CT, MRI, 초음파 등)
	생체신호 측정기기	심전도(ECG), 뇌파(EEG) 등 생체 신호를 측정하는 기기
치료 기기	수술 및 치료기기	레이저 수술기, 수술용 로봇, 인공호흡기 등
	방사선 치료기	암 치료에 사용되는 선형가속기 등
	재활 및 보조기기	인공관절, 임플란트, 보청기 등

3) 디지털 헬스 케어 및 의료정보 시스템

디지털 헬스 케어 및 의료정보시스템 분야는 정보통신 기술을 이용하여 개인의 건강과 질병을 관리하는 산업 및 서비스 분야를 의미한다. 이 분야는 크게 원격의료 및 재택 의료 분야와 의료정보 시스템 분야로 나누어 고려할 수 있다. 먼저, 원격의료 및 재택 의료 분야는 정보통신망을 통해 원격으로 진료하거나 가정에서 건강을 관리하는 기기 또는 시스템을 의미한다. 다음으로 의료정보 시스템은 전자 의무 기록(EMR), 병원 의료정보 시스템(HIS) 등 의료 데이터 관리 시스템을 포함하고 있다.

나. 그린 바이오(Green Bio): 농업, 식품 및 자원 분야

그린 바이오는 농업, 식품, 그리고 자원 등과 같이 생물 자원을 활용하여 유용한 물질을 생산하는 기술 분야이다. 유전자 재조합 식품으로 알려진 개량종자, 건강 기능 식품, 그리고 친환경 농약 및 사료 첨가제 등이 이 분야에 포함된다. 본 보고서에서는 바이오 농업 및 종자와 바이오 식품과 같은 두 가지 세부 분야에 대해 살펴본다.

1) 바이오 농업 및 종자

바이오 농업 및 종자는 유전자 변형을 통해 병충해에 대한 저항성을 개선하거나 생산성을 향상시키는 분야를 의미한다. 또한, 미생물을 활용하는 친환경 농약 및 비료도 포함하고 있다. 〈표 7－3〉에서 바이오 농업 및 종자 분야를 정리해서 보여주고 있다.

〈표 7−3〉 바이오 농업 및 종자 분야

세부 분야	설명
유전자 변형 작물	생산성, 특정 영양소, 그리고 병충해 저항성 등을 개선한 작물
미생물 농약 및 비료	병충해 방제나 작물 생장에 도움이 되는 미생물을 활용한 친환경 제품

2) 바이오 식품

바이오 식품 분야에는 기능성 식품, 식품 첨가물, 그리고 사료 첨가제 등이 포함되어 있다. 〈표 7−4〉에서 바이오 식품 분야를 세부적으로 정리해서 보여주고 있다.

〈표 7−4〉 바이오 식품 분야

세부 분야	설명
기능성 식품	건강 증진에 도움이 되는 성분을 강화한 식품
식품 첨가물	미생물 발효 등을 통해 생산된 효소 및 감미료
사료 첨가제	동물의 성장을 돕거나 건강을 개선하는 바이오 물질

다. 화이트 바이오(White Bio): 산업, 에너지, 환경 분야

화이트 바이오는 기존 화학 산업의 소재를 미생물이나 효소를 이용하여 바이오 기반으로 대체하고 환경 문제를 해결하는 기술 분야라 할 수 있다. 이 분야를 세부적으로 살펴보면 바이오 소재 및 화학 분야와 바이오 에너지 및 환경의 두 가지 분야로 크게 나누어 고려할 수 있다.

1) 바이오 소재 및 화학

바이오 소재 및 화학은 친환경 소재나 효소 또는 화학물질 등을 포함하는 분야를 의미한다. 〈표 7−5〉에서 바이오 소재 및 화학 분야를 세부적으로 정리해서 보여주고 있다. 대표적인 친환경 소재로는 생분해성 플라스틱을 고려할 수 있고 그밖에 산업 공정에 사용되는 산업용 효소와 미생물 공정을 통해 화학물질을 생산하는 분야를 포함하고 있다.

〈표 7－5〉 바이오 소재 및 화학 분야

세부 분야	설명
바이오 플라스틱	옥수수 등 식물 자원을 원료로 만든 생분해성 플라스틱
산업용 효소	세제, 섬유 가공 등 산업 공정에 사용되는 효소
바이오 화학 소재	미생물 공정을 통해 생산하는 화학물질

2) 바이오 에너지 및 환경

바이오 에너지 및 환경은 바이오 연료 또는 바이오 환경 등을 포함하는 분야를 의미한다. 〈표 7－6〉에서 바이오 에너지 및 환경 분야를 정리해서 보여주고 있다. 바이오 연료는 기존 화석 연료와 다르게 생물 유기체를 변환하여 만든 연료를 의미하고 바이오 환경은 미생물을 이용하여 오염된 환경을 정화하는 기술 분야를 포함한다.

〈표 7－6〉 바이오 에너지 및 환경 분야

세부 분야	설명
바이오 연료	바이오 매스(생물 유기체)를 변환하여 만든 연료 (예: 바이오에탄올, 바이오디젤)
바이오 환경	미생물을 이용하여 폐수 및 폐기물을 처리하고 오염된 환경을 정화하는 기술

2. 바이오 및 의료 산업 분야 AGI 기술 활용 현황

바이오 및 의료 산업은 인류의 건강 증진과 삶의 질 향상에 직접적으로 이바지하는 핵심 분야로서 앞에서 살펴본 바와 같이 레드(보건·의료), 그린(농업·식품), 그리고 화이트(산업·환경) 바이오로 분류된다. 최근 AGI 또는 그에 준하는 인공지능 기술이 급속도로 발전하면서, 이들 기술이 바이오 및 의료 산업의 전통적인 연구개발 및 생산 패러다임을 근본적으로 바꾸고 있다. 여기서는 오늘날 바이오 및 의료 산업에서의 세부 분야별 AGI 기술 활용 현황을 분석하고자 한다. AGI는 방대한 바이오 데이터를 분석하여 새로운 패턴과 인사이트를 도출하고 복잡한 생명 현상의 시뮬레이션을 통해 신약 개발, 맞춤형 의료, 지

속 가능한 농업 및 친환경 소재 개발 등의 난제를 해결할 잠재력을 지니고 있다고 할 수 있다(Fleming, 2018).

가. 레드 바이오 분야 AGI 활용 현황

레드 바이오 분야는 보건 및 의료 분야를 지칭하며 바이오 및 의료 산업 전 분야에서 AGI 기술의 도입이 가장 활발하게 이루어질 수 있는 영역이라 할 수 있다.

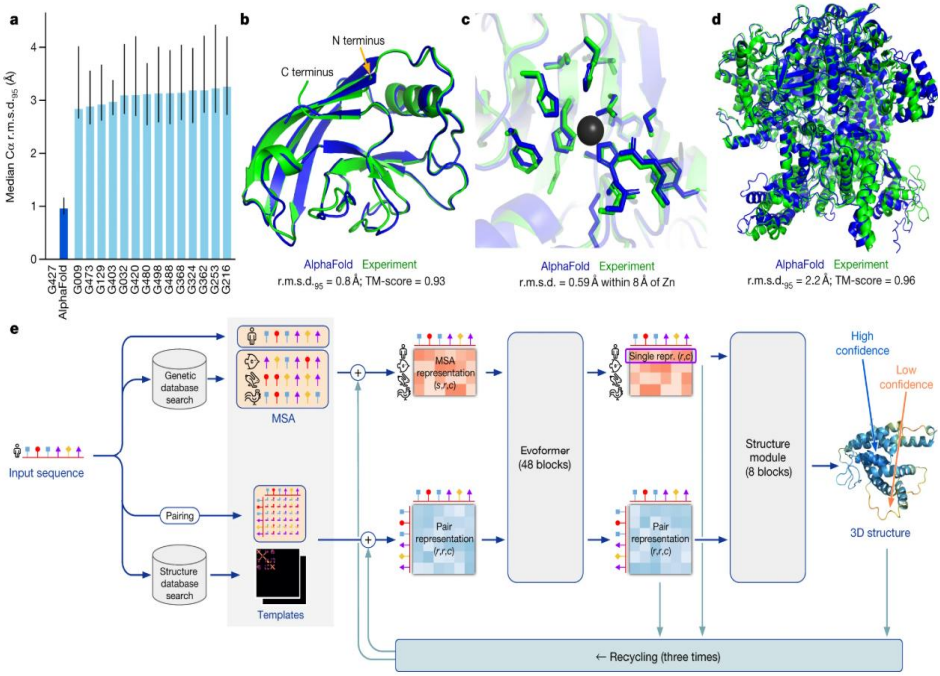
1) 바이오 의약품(Biopharmaceuticals)

오늘날 바이오 의약품 분야에서는 AGI 기술까지는 아니더라도 인공지능 기술의 도입을 통해 신약 개발의 전 주기에 걸쳐 시간과 비용을 획기적으로 단축시키고 있다. 최신 인공지능 기술은 과거 수년에서 십수 년이 걸리던 신약 후보 물질 발굴 과정을 수개월 단위로 줄이는 것을 가능하게 하고 있다. 여기서는 신약 후보 물질 발굴과 임상시험 최적화의 두 가지 영역으로 구분하여 살펴보고자 한다.

구글 딥마인드의 알파폴드(AlphaFold)와 같은 인공지능 모델은 단백질 3차원 구조를 높은 정확도로 예측하여, 질병의 원인이 되는 단백질의 작용 기전을 이해하고 이에 맞는 약물 구조를 설계하는 데 결정적인 역할을 한다(Jumper et al., 2021).

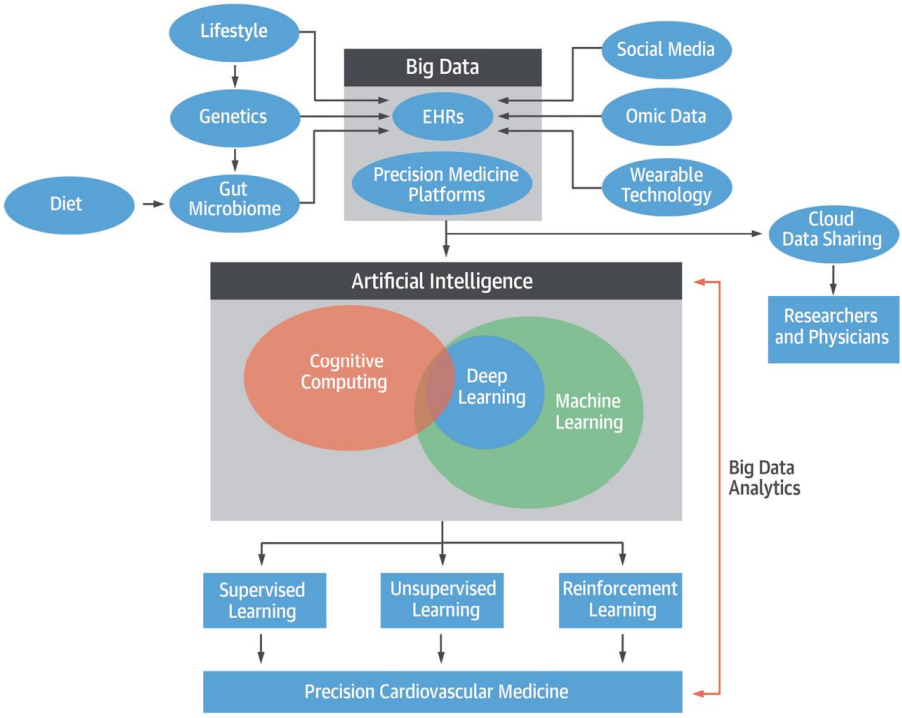
〔그림 7-1〕에서 알파 폴드를 이용하여 정밀한 단백질 구조를 생성하는 과정을 보여주고 있다. 알파 폴드의 인공지능 기술은 수십억 개의 화합물 라이브러리에서 특정 단백질에 가장 효과적으로 결합할 수 있는 후보물질을 가상으로 탐색(Virtual Screening)하여 R&D 효율을 극대화한다.

〔그림 7-1〕 알파 폴드를 이용해 정밀한 단백질 구조를 생성하는 과정



또한, 인공지능 기술은 환자의 유전체 정보와 생활 습관 데이터 등을 분석하여 임상시험에 가장 적합한 환자군을 선별하고, 약물 반응성을 예측하여 임상 성공률을 높이는 데 활용된다. 또한, 임상 데이터 분석을 통해 부작용을 사전에 예측하고 시험 프로토콜을 동적으로 최적화하는 데 활용된다(Krittanawong et al., 2021). 〔그림 7-2〕에서 이러한 인공지능 기반 환자 유전체 정보 및 생활 습관 등의 복합 분석 모델의 구조를 보여주고 있다.

〔그림 7−2〕 인공지능 기반 환자 유전체 정보 및 생활 습관 등의 복합 분석 모델



2) 의료 기기 및 진단

의료 영상 진단과 스마트 의료 기기를 포함하는 의료 영상 데이터 분석은 오늘날 인공지능이 가장 괄목할만한 결과를 보이는 분야라 할 수 있다. CT, MRI, X-ray 등 의료 영상을 딥러닝 기반으로 분석하여 인간 의사보다 빠르고 정확하게 암, 뇌졸중, 그리고 안과 질환 등의 병변을 찾아내는 기술이 상용화되고 있다. 이는 조기 진단율을 높이고 진단 오류를 줄이는 데 크게 기여하고 있다.(Topol, 2019)에서는 이러한 다양한 사례를 소개하고 있으며 〈표 7−7〉에서 FDA에서 승인한 인공지능 모델의 사례를 보여주고 있다. 2019년 기준 14개의 인공지능 모델이 FDA 승인을 통과해 실제 진단에 활용되고 있으며 이러한 모델은 점차 획지적으로 증가 할 것으로 기대된다.

〈표 7－7〉 FDA에서 승인한 인공지능 의료 모델 사례

기업	승인 일자	의료 모델
Apple	September 2018	Atrial fibrillation detection
Aidoc	August 2018	CT brain bleed diagnosis
iCAD	August 2018	Breast density via mammography
Zebra Medical	July 2018	Coronary calcium scoring
Bay Labs	June 2018	Echocardiogram EF determination
Neural Analytics	May 2018	Device for paramedic stroke diagnosis
IDx	April 2018	Diabetic retinopathy diagnosis
Icometrix	April 2018	MRI brain interpretation
Imagen	March 2018	X-ray wrist fracture diagnosis
Viz.ai	February 2018	CT stroke diagnosis
Arterys	February 2018	Liver and lung cancer(MRI, CT) diagnosis
MaxQ-AI	January 2018	CT brain bleed diagnosis
Alivector	November 2017	Atrial fibrillation detection via Apple Watch
Arterys	January 2017	MRI heart interpretation

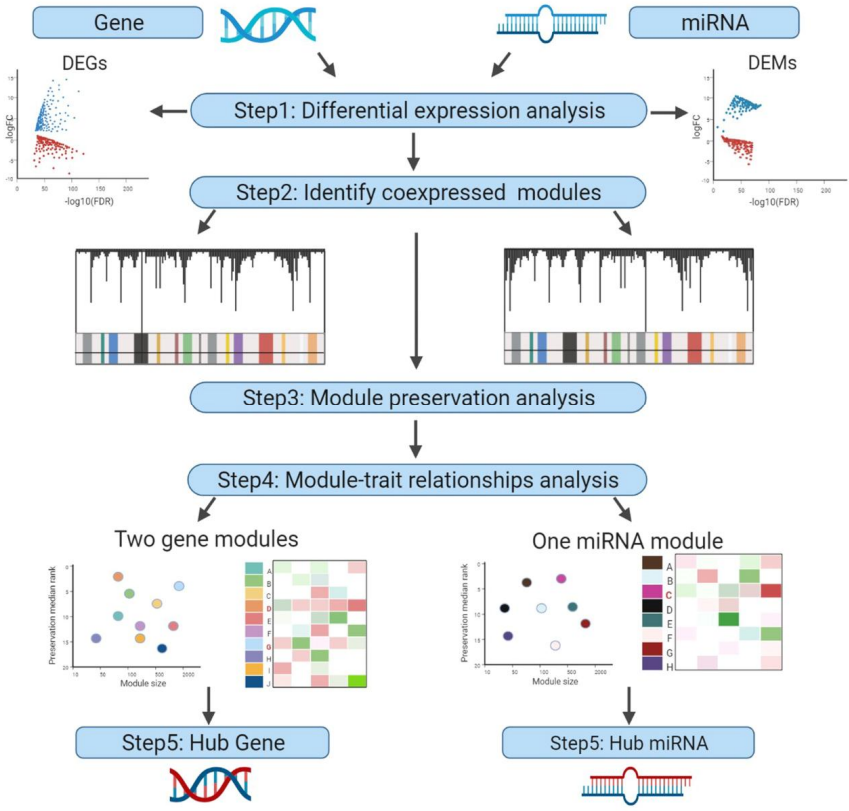
스마트 의료 기기 분야에서는 수술용 로봇에 인공지능 기술을 접목하여 의사의 정밀한 수술을 보조하고, 수술 과정 중에서 발생할 수 있는 예기치 못한 상황에 실시간으로 대응하는 시스템이 개발되고 있다. 또한, 웨어러블 기기에서 수집된 생체신호를 실시간으로 분석하여 심장마비나 뇌졸중 같은 응급 질환의 발생을 사전에 예측하고 경고하는 서비스도 등장하고 있다.

3) 디지털 헬스케어 및 재생의료

오늘날 인공지능 기술은 개인 맞춤형 의료 시대를 여는 핵심 기술로 활용되고 있다. 환자의 유전 정보, 의료 기록, 그리고 실시간 생체 데이터를 종합적으로 분석하여 개인에게 최적화된 치료법과 약물을 추천하는 것이 가능해지고 있다. 특히 항암 치료에서 환자 개별 암세포의 특성을 분석해 가장 효과적인 항암제를 추천하는 서비스가 활발히 연구되고 있다(Fleming, 2018; You et al., 2022). 그밖에 세포 치료제 개발을 위해 줄기세포의 분화 과정을 시뮬레이션하고 제어하여 특정 질병 치료에 가장 효과적인 세포를 생산하는 공정을 최적화하는 데 인공지능 기술이 활용되고 있다.

〔그림 7-3〕에서 네트워크 기반으로 우수한 항암체를 구분하는 흐름을 보여주고 있다 (You et al., 2022).

〔그림 7-3〕 네트워크 기반의 우수 항암체 구분 모델 흐름도

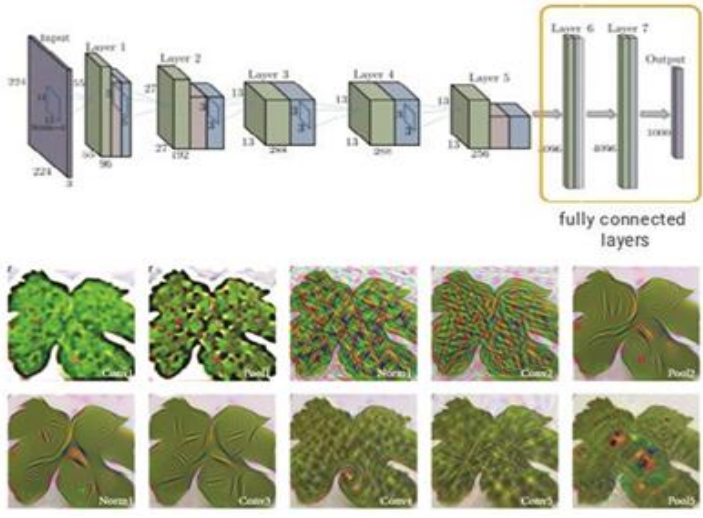


나. 그린 바이오 분야 AGI 활용 현황

오늘날 그린 바이오 분야에서는 기후 변화와 식량 위기 문제에 대응하기 위해 인공지능 기술을 적극적으로 활용하고 있다. 이 분야에서도 현재 상황에서는 AGI라고 할 수는 없지만 최신 인공지능 기술이 적용되고 있는 단계라 할 수 있다. 여기서는 그린 바이오 분야에서의 인공지능 활용 현황을 스마트 팜 및 정밀 농업, 종자 개발 및 육종, 그리고 대체 식품 개발의 3가지로 구분하여 살펴본다.

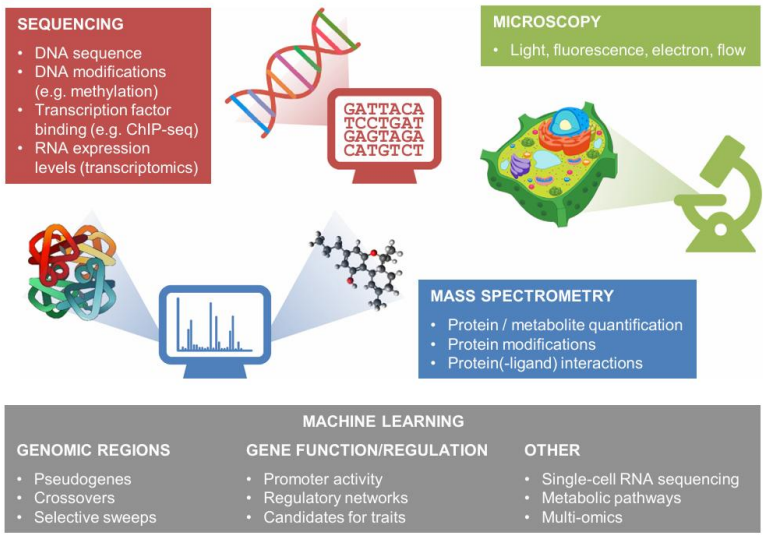
먼저 스마트 팜 및 정밀 농업 분야에서는 드론과 센서가 수집한 작물의 생육 데이터, 토양 상태, 그리고 기상 정보를 바탕으로 최신 인공지능 기술을 적극적으로 활용하여 물, 비료, 그리고 농약을 필요한 곳에 필요한 만큼만 정밀하게 공급하도록 제어하고 있다. 이를 통해 작물의 생산성을 극대화하고 환경 오염은 최소화하도록 한다(Kamilaris et al., 2018). 〔그림 7-4〕에서는 CaffeNet이라는 식물의 질병을 분류하는 CNN 기반의 네트워크를 보여주고 있다(Sladojevic et al., 2016).

〔그림 7-4〕 CaffeNet의 구조와 식물 질병 분류 예시



다음으로 종자 개발 및 육종을 위해 인공지능 기술을 활용하여 식물의 유전체 정보를 분석하여 특정한 기후 또는 병충해에 강인한 유전자를 발굴하는 것이 가능해지고 있다. 이를 통해 원하는 형질을 가진 품종을 개발하는데 요구되는 시간을 단축할 수도 있다. 그 밖에 유전자 가위 기술(CRISPR)과 인공지능 기술을 결합하여 육종 효율을 높이는 연구도 활발히 진행되고 있다(Dijk et al., 2020). 〔그림 7-5〕에서 이러한 연구의 사례로서 기계 학습을 이용한 유전체 분석의 예시를 보여주고 있다.

〔그림 7－5〕 기계 학습을 이용한 유전체 분석 예시

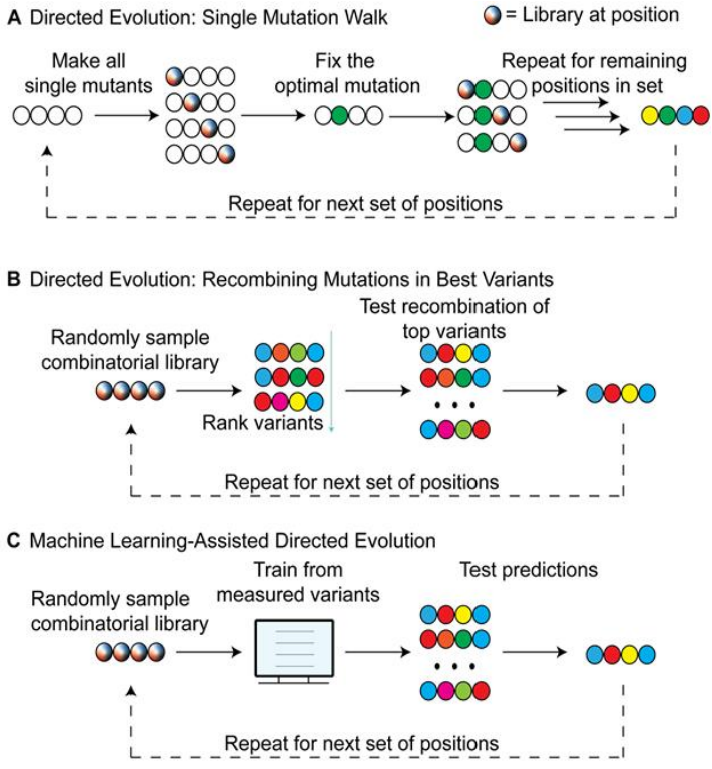


마지막으로 대체 식품 개발 분야에서는 식물성 고기나 배양육의 맛과 식감을 실제 고기와 유사하게 만들기 위해 다양한 단백질 조합을 시뮬레이션하고 최적의 배율을 찾는데 인공지능 기술을 활용하는 것이 가능하다.

다. 화이트 바이오 분야 AGI 활용 현황

화이트 바이오 분야는 기존의 화석 연료에 의존하는 석유화학 산업을 대체하고 지속 가능한 생산 체계를 구축하는 것을 주된 목표의 하나로 하고 있다. 이를 위해 인공지능 기술은 미생물을 설계하거나 공정 최적화를 하는데 핵심적 역할을 수행하고 있다. 이 분야의 기술 수준 역시 AGI라고 할 수는 없고 인공지능이라고 보는 것이 타당하다. 여기서는 여러 세부 기술 분야 중에서 가장 많은 주목을 받고 있는 바이오 소재 및 화학 분야와 바이오 에너지 및 환경 정화의 두 가지만 살펴보고자 한다.

〔그림 7−6〕 기계 학습 기반 미생물 진화 시뮬레이션 과정



먼저, 바이오 소재 및 화학 분야에서 인공지능 기술은 특정 화학 물질이나 플라스틱을 분해하는 능력이 뛰어난 미생물 균주나 효소를 설계하는 데 사용된다. 원하는 기능을 가진 단백질(효소) 구조를 처음부터 설계하고, 미생물의 대사 경로를 최적화하여 바이오 플라스틱과 바이오 연료 등 유용한 물질의 생산 효율을 높인다(Wu et al., 2019). 〔그림 7−6〕에서 기계 학습을 기반으로 미생물의 진화를 시뮬레이션하는 과정을 보여주고 있다.

다음으로 바이오 에너지 및 환경 정화 분야에서는 인공지능 기술을 활용하여 바이오매스로부터 바이오에탄올, 바이오디젤을 생산하는 발효 공정을 최적화하고 있다. 또한, 폐수나 토양의 오염 물질을 가장 효율적으로 분해할 수 있는 최적의 미생물 조합을 찾아내는 데도 인공지능 기반 시뮬레이션을 적극적으로 활용하고 있다.

라. 바이오 및 의료 분야 AGI 활용 현황 분석 및 제언

앞에서 살펴본 바와 같이 인공지능 기술이 아직 완전한 AGI 수준에 이르지는 않았기에 현재 상황에서는 최신 인공지능 기술을 활용하고 있는 단계라 할 수 있다. 인공지능 기술은 바이오 및 의료 분야 전반에 걸쳐 연구개발의 속도를 가속화하고, 생산 효율을 극대화하며, 과거에는 해결 불가능했던 문제에 대한 새로운 해법을 제시하고 있다. 특히 질병의 정복, 식량 안보 확보, 그리고 지속 가능한 산업 생태계 구축이라는 인류의 목표 달성에 결정적인 기여를 할 것으로 기대된다.

그러나 AGI 기술의 성공적인 도입과 확산을 위해서는 해결해야 할 과제도 존재한다. 고품질의 방대한 데이터 확보, 데이터 표준화, 기술의 신뢰성 및 안전성 검증, 그리고 기술 활용에 따른 윤리적·법적 문제에 대한 사회적 합의가 필요하다. 따라서 정부와 산업계, 학계는 긴밀한 협력을 통해 데이터 인프라를 구축하고 관련 규제를 정비할 필요가 있다.

이와 함께 오늘날 인공지능 기술이 바이오 및 의료 분야뿐 아니라 산업 전반에 걸쳐 확산과 활용이 이루어짐에 따라 융합형 인재를 양성하고 확보하기 위한 노력도 함께 지속해야 할 것이다.

3. 중기(5년 후) 바이오/의료 분야 AGI 활용 전망 및 경쟁력 확보 전략

AGI 기술은 바이오 및 의료 산업의 연구개발, 생산, 그리고 산업화 전반에 걸쳐 패러다임 전환을 가속화할 것으로 기대된다. 특히 방대한 생명과학 데이터로부터 새로운 지식을 창출하고 복잡한 생물학적 시스템을 예측 및 제어하는 AGI의 능력은 이전에 해결 불가능했던 난제들을 해결할 열쇠로 주목받게 될 것이다. 여기서는 향후 5년 후의 중기적 관점에서 레드, 그린, 그리고 화이트 바이오 및 의료 분야별 AGI 기술의 발전 양상을 전망하고, 이에 대응하여 국내 기업 및 연구 기관이 글로벌 경쟁력을 확보하기 위한 핵심 전략을 제시하고자 한다. 이는 단순한 기술 도입을 넘어, AGI를 핵심 역량으로 내재화하고 새로운 비즈니스 모델을 창출하는 선제적 대응의 필요성을 강조할 필요가 있다. 향후 5년간 AGI 기술은 특정 작업을 자동화하는 단계를 넘어, 가설 수립부터 실험 설계, 결과 해석까지 과학적 발견의 전 과정에 깊숙이 관여하는 AI 과학자(AI Scientist) 모델로 진화할 것으로 기대된다.

가. 레드 바이오 분야

1) 전반적 전망

의료, 제약 및 진단 등을 포함하는 레드 바이오 분야에서는 향후 5년 이후 초개인화 및 예측 의료가 본격화될 것으로 예상된다. 특히 디지털 트윈 기술과 결합하여 질병의 예방과 치료를 혁신할 것으로 예상된다. 환자 개인의 유전체, 생활 습관, 그리고 의료 기록을 통합한 가상 인체를 구축하여 특정 약물에 대한 반응이나 질병의 진행 과정을 필요한만큼 반복적으로 시뮬레이션하여 최적의 치료법을 부작용없이 찾아내는 시대가 될 것이다.

2) 주요 세부 기술 전망

신약 개발 분야에서는 알파폴드(AlphaFold)를 넘어, 특정 질병 타겟에 최적화된 새로운 단백질과 항체 구조를 처음부터 설계하는 생성형 AI 기술이 신약 개발의 표준으로 자리잡을 것이다(Jumper et al., 2018). 이를 통해 후보물질 발굴부터 임상 전 단계까지의 기간이 현재의 절반 이하로 단축될 수 있다. 특히, 항생제·항미생물 펩타이드 등에서 대규모 후보를 발굴하는 사례가 실험적 검증 단계까지 전개되어 초기 전임상 파이프라인을 채울 것으로 기대된다(Torres et al., 2024; Wan et al., 2024). 하지만 임상과 승인까지의 과정에서는 약효, 독성, 그리고 임상 설계 등의 병목 현상이 남아있어 완전한 자동 신약 개발은 5년 후에도 어려울 것으로 예상된다.

그 밖에 예측 진단 분야에서는 혈액 내 미량의 DNA나 단백질 마커를 분석하여 암이나 치매와 같은 중증 질환을 증상 발현 수년 전에 예측하는 AI 진단 알고리즘이 상용화될 수도 있을 것이다. 이는 치료 중심의 의료 패러다임을 예방 및 관리 중심으로 전환시키는 기폭제가 될 것이다(Topol, 2019). 다만, 진단, 영상, 그리고 임상 의사 결정 등과 같은 예측 진단 인공지능 모델은 규제 및 임상 검증 기준 강화로 상용화 속도가 조절될 것으로 생각된다. 그럼에도 불구하고 FDA와 EU의 규제 동향에 따라 이 분야에서 인공지능을 넘어 AGI 기술의 활용 범위는 더욱 넓어질 것으로 전망된다(Aboy et al., 2024; FDA, 2025).

3) 경쟁력 확보를 위한 핵심 전략

레드 바이오 분야의 경쟁력 확보를 위해서는 데이터와 실험 인프라에 대한 동시 투자가 필요하다. 인공지능 모델의 고도화를 위해 대규모의 고품질 실험 및 임상 데이터와 자동화 실험을 위한 로봇랩과 같은 것이 필요하다. 이를 위해 우선적으로 데이터 파이프라인

과 실험 정보 관리 시스템을 구축할 필요가 있다(Jumper, 2021). 또한, 제품 개발의 초기 단계부터 FDA와 EU 등의 규제 요구를 반영한 검증 및 문서화 프로세스를 설계하여 상용화 리스크를 낮출 수 있다(Aboy et al., 2024; FDA, 2025).

모델의 상용화와 산업화 적용 확장을 위해 제약사 등과 연계한 협력 네트워크 구축이 필요하다. 인공지능 모델과 데이터셋은 파트너십을 통해 확장하고 초기 완성도가 낮은 부분을 외부 CRO/CPO와 연계하여 빠르게 검증하는 모듈형 파이프라인 구축 및 운영이 요구된다(Torres et al., 2024).

이와 더불어 생성형 모델이 유해 물질을 설계할 가능성이 상존하므로 내부 거버넌스 및 외부 심사가 필수적이다. 이를 통해 바이오 시큐리티와 윤리 이슈를 극복할 필요가 있다. 그리고 인공지능 모델이 제시한 후보가 임상에서 실패할 가능성에 대응하여 사업화가 가능한 후보로 전환하는 역량을 제고할 필요가 있다.

나. 그린 바이오 분야

1) 전반적 전망

지속 가능한 자율형 농업 시스템을 구축하고자 하는 그린 바이오 분야에서는 기후 변화와 식량 안보 위기에 대응하는 기술 개발이 요구된다. 이 분야에서 AGI는 최소한의 자원으로 최대의 수확을 달성하는 자율 운영 스마트 팜(Autonomous Smart Farm)을 구현할 것으로 기대된다. 이는 단순히 데이터를 분석하는 수준을 넘어, 농장 전체의 생육 환경과 에너지 효율을 스스로 판단하고 최적화하는 단계로 발전하는 것을 의미한다.

2) 주요 세부 기술 전망

그린 바이오 분야의 세부 기술로 디지털 육종과 정밀 농업 제어 두 분야를 중심으로 살펴보고자 한다. 먼저 디지털 육종 분야에서 AGI는 유전체 정보와 실제 생육 데이터를 결합하여 특정한 기후(고온, 가뭄 등)나 토양 환경에 최적화된 맞춤형 품종을 가상 환경에서 시뮬레이션하고 개발 기간을 획기적으로 단축할 것이다. 이를 위해 드론, 위성, 그리고 센서 데이터와 유전체를 기반으로 인공지능 기술을 활용한 예측을 결합하는 정밀 육종이 확산되어 품종 개발 주기가 단축되고 현장 적응성이 향상될 전망이다(Fu et al., 2024; Aijaz et al., 2025).

정밀 농업 제어 분야에서는 드론과 로봇이 수집한 초분광 이미지와 토양 센서 데이터

등을 실시간으로 분석하여, 개별 식물의 상태에 맞춰 물, 영양분, 그리고 생장 호르몬을 마이크로 단위로 정밀하게 공급하는 자율형 시스템이 보편화될 것이다. 물과 비료의 절감 등을 포함하는 지속 가능성을 지원하고 기후 회복력을 가진 품종을 개발하는 것도 가능할 것으로 기대된다. 그밖에 병해충의 조기 탐지 서비스 산업도 빠르게 성장해 나갈 것으로 예측된다.

3) 경쟁력 확보를 위한 핵심 전략

그린 바이오 분야의 경쟁력 확보를 위해서는 농업 생산 현장과 유전체 정보를 연동하는 데이터 플랫폼 구축이 우선 필요할 것으로 생각된다. 원격 탐지 영상과 토양 및 기상, 그리고 유전체 데이터의 통합 저장 및 분석 환경을 마련하고 데이터 파이프라인 구축과 표준화 작업도 우선 추진할 필요가 있다(Aijaz et al., 2025). 또한 모델과 하드웨어를 패키지로 결합한 서비스형 솔루션의 개발도 요구된다. 예를 들어 농가에서 쉽게 도입할 수 있도록 센서와 인공지능 분석 모델을 결합하여 패키지화하고 초기 도입 비용을 경감할 수 있도록 구독형 서비스로 시장에 진입할 필요가 있다. 기후와 토양이 지역별로 다르며 이러한 점이 그린 바이오 산업에 크게 영향을 미친다는 점을 고려하여 서로 다른 지역별 모델을 마련하고 농기계 및 유통사와의 협업으로 실수익으로 연결되도록 하는 것이 중요하다.

그린 바이오 분야에서의 경쟁력 확보를 위해 데이터 편향성 문제와 현장 전이성 문제로 인한 실사용 환경에서 성능 저하가 발생할 가능성이 크므로 현장 데이터 확보 및 지속적 재학습을 효과적으로 수행할 수 있는 기술 개발이 요구된다. 또한 유전자 편집 등의 기술에 대한 각종 규제와 사회 및 소비자 수용성도 함께 고려하고 대응할 필요가 있다.

다. 화이트 바이오 분야

1) 전반적 전망

화이트 바이오 분야에서 AGI는 인공지능 기반 바이오 파운드리의 핵심 엔진으로 작동할 것으로 기대된다. 바이오 파운드리는 미생물 균주 개발, 공정 최적화, 그리고 시제품 생산까지의 전 과정을 자동화한 시설을 의미한다. AGI는 이 모든 과정을 지능적으로 설계하고 제어하여 바이오 화학, 바이오 플라스틱 등 고부가가치 소재의 개발 속도와 생산성을 극대화할 것으로 예상된다.

2) 주요 세부 기술 전망

화이트 바이오 분야에서는 합성 생물학 회로 설계를 대표적으로 고려할 수 있다. 이는 인간이 원하는 특정 화학 물질을 생산하도록 미생물의 유전자 회로(Metabolic pathway)를 AGI가 자동으로 설계하고, 수만 가지의 설계안을 가상 세포 안에서 시뮬레이션하여 최적의 균주를 예측하는 기술이 핵심이 될 것이다. 또한 바이오 공정의 디지털 트윈을 구축하고 인공지능 기술을 기반으로 공정 최적화를 수행하여 원가 절감 및 수율 향상에 직접 기여할 수 있다. 모니터링 및 기계 학습 기술을 이용하여 유지 보수를 예측하고 실시간 품질 제어를 함으로써 조속한 생산 안정성 개선을 기대할 수 있을 것이다. 그 밖에 효소 및 미생물 기반 공정에서 인공지능 기술을 이용하여 촉매와 경로 설계를 가속화함으로써 기존 화학 공정의 대체 가능성도 기대할 수 있다(Bähner et al., 2021).

3) 경쟁력 확보를 위한 핵심 전략

화이트 바이오 분야의 경쟁력 확보를 위해서 산업 생산 현장인 공장에 디지털 트윈 기술을 적용하여 스마트 공장을 단계적으로 구축해 나갈 필요가 있다. 먼저 파일럿 라인에서 센서 데이터를 수집하고 이후 단계에서 모델을 구축하며 마지막 단계에서 스케일 확장을 해나가는 순차적 접근을 할 필요가 있다(Bähner et al., 2021). 또한 모듈별 공정 설계 역량을 확보하여 여러 제품에 재사용 가능한 공정 모듈을 개발하고 각 모듈의 재사용 체계를 마련해 가야 할 것이다. 그밖에 데이터와 바이오를 융합한 인력 확보를 위한 인력 재교육도 요구된다. 이를 위해 공정 개발자에게 데이터 사이언스 역량을 보강하여 모델과 공정의 연계를 강화할 필요가 있다.

화이트 바이오 분야에 AGI를 적용해나가는 과정에서 운영 데이터 부족과 데이터의 품질 문제를 극복하기 위해 공정 데이터의 표준화와 고품질의 라벨링을 위한 투자가 요구된다. 또한 화이트 바이오에 활용되는 AGI 기술에 대한 산업 공정의 안전 확보와 환경 규제 준수 여부도 필수적이라 할 수 있다.

4. 장기(10년 후) 바이오/의료 분야 AGI 활용 전망 및 경쟁력 확보 전략

향후 10년 후, AGI는 바이오 및 의료 산업에서 단순한 분석 도구를 넘어, 생명 현상의 근본 원리를 이해하고 새로운 생명 시스템을 창조하는 공동 창조자(Co-creator)로 자리매김할 것으로 기대된다. AGI는 유전자를 읽고, 쓰고, 설계하는 전 과정에 관여하며 데이터의 분석 수준을 넘어 자연에 존재하지 않는 새로운 기능의 단백질과 미생물을 설계하고, 복잡한 생태계의 상호작용을 시뮬레이션하여 미래를 예측하고 제어하는 단계에 이를 것이다. 더불어 인류가 지금까지 경험하지 못했던 속도와 규모로 질병, 식량, 그리고 환경 문제를 해결할 잠재력을 가지게 될 것이다. 여기서는 향후 10년의 장기적 관점에서 AGI가 가져올 바이오 및 의료 산업의 근본적 변화를 전망하고, 미래의 기술 경쟁력을 제고하기 위해 우리가 반드시 확보해야 할 핵심 선도 전략을 제시하고자 한다. 이는 기존의 추격자(Fast Follower) 전략을 넘어, 새로운 규칙을 만드는 퍼스트 무버(First Mover)로의 전환을 촉구하는 로드맵이어야 할 것이다.

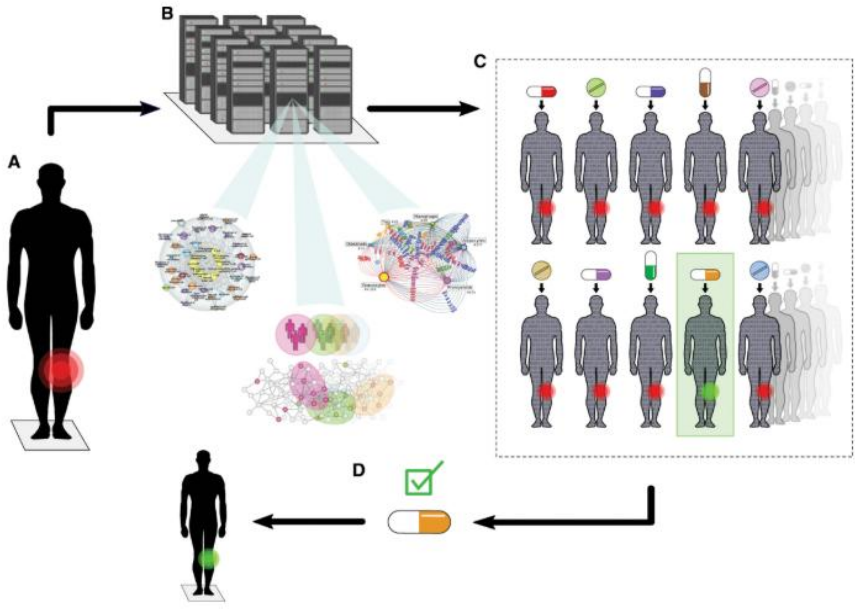
가. 레드 바이오 분야

1) 전반적 전망

향후 10년 후 AGI는 개인의 유전체와 생체 데이터를 기반으로 구축된 개인용 디지털 트윈(Personal Digital Twin)을 통해 질병을 평생 관리하고 노화를 제어하는 시대를 열 수도 있을 것이다(Benson, 2023). 암이나 치매와 같은 질병은 발병 전 예측을 넘어, AGI가 설계한 세포 유전자 치료제(Cellular Gene Therapeutics)를 통해 원인 유전자를 정상으로 복원하거나, 노화 세포를 제거하여 건강 수명을 연장하는 것이 가능해질 것이다.

〔그림 7-7〕에서 이러한 방향의 연구 사례로 개인별로 디지털 트윈을 구축하여 다양한 약물의 반응을 시뮬레이션하고 개인에 맞추어 최적의 처방을 하는 연구를 보여주고 있다.

〔그림 7−7〕 개인 디지털 트윈을 이용한 맞춤형 처방 과정

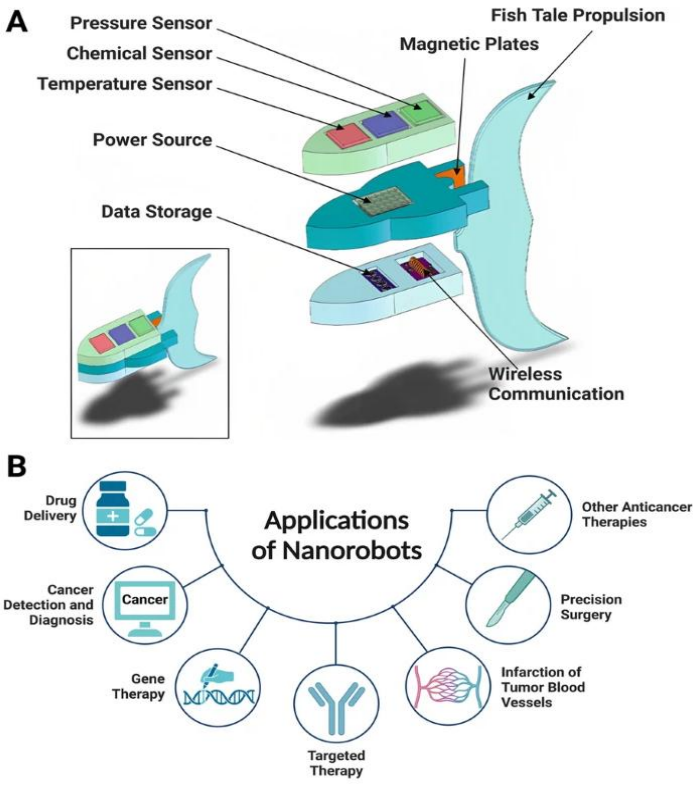


2) 주요 세부 기술 전망

레드 바이오 분야의 대표적인 세부 기술로는 장기적으로 먼저 자율형 치료 로봇을 고려할 수 있을 것이다. AGI는 체내에서 특정 타겟에 약물을 전달하거나, 특정 암세포만 찾아내 공격하고 손상된 조직을 재생시키는 나노미터 크기의 생체 내 로봇(In-vivo Nanobot) 또는 프로그래밍된 박테리아를 설계할 것으로 기대된다(Yang et al, 2025; Kong et al., 2023; Cooper et al., 2023). 이는 약물 전달의 부작용을 최소화하고 치료 효과를 극대화하는 궁극의 정밀 의료를 실현할 수 있을 것이다. 〔그림 7−8〕에서 바이오 및 의료에 적용되는 나노 로봇의 연구 사례를 보여주고 있다.

또한, 온디맨드(on-demand) 신약 및 백신 플랫폼으로서 새로운 감염병이 발생하면, AGI가 바이러스의 유전 정보를 분석하여 수 시간 내에 최적의 mRNA 백신이나 항체 치료제 구조를 설계하고, 분산된 바이오 프린팅 시설로 전송하여 즉시 생산하는 전염병 즉시 대응 시스템이 구축될 것이다.

〔그림 7-8〕 나노 미터 크기의 생체 로봇 연구 예시



그 밖에 AGI는 단백질-리간드(Ligand) 설계, 분자 생성, 독성 예측, 그리고 제형 설계 등을 통합한 자동 설계 루프를 운영할 것이다. 예를 들어, AGI 모델이 스스로 후보물질을 설계하고 시뮬레이션을 실행하여 실험실 자동화 시스템에 지시하는 사이클이 가능해질 것이다. 즉, 스스로 실험하는 인공지능 모델이 구현될 것이고 이 흐름은 연구 초기 단계의 속도를 엄청난 규모로 높일 것이다(Groff-Vindman et al., 2025). 궁극적으로는 AGI를 기반으로 하는 의료 에이전트가 진단, 처방, 그리고 모니터링을 통합 제어하는 시스템이 출현할 수 있다.

3) 경쟁력 확보를 위한 핵심 전략

레드 바이오 분야의 경쟁력 확보를 위해서는 임상시험에서의 윤리 규정 준수와 규제 강

화 및 인간 대상 검증에 대한 부담을 극복할 필요가 있다. 또한, AGI 모델의 개발 및 활용을 확산하기 위해서는 AGI 블랙박스의 해석 가능성을 확보할 필요가 있다. 그 밖에도 대규모 환자 데이터 확보, 개인정보 보호, 그리고 데이터 표준화도 극복해야 할 주요 이슈라 할 수 있다.

나. 그린 바이오 분야

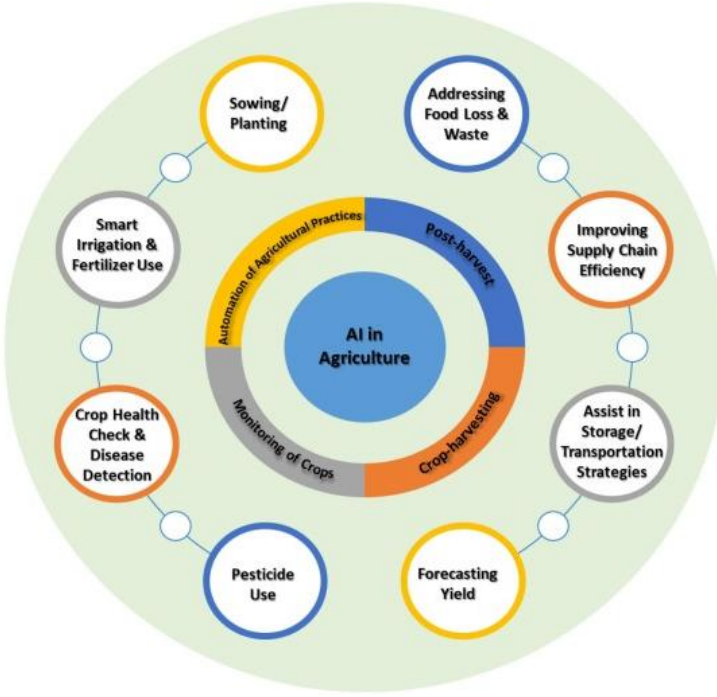
1) 전반적 전망

그린 바이오 분야에서 AGI는 개별 농장의 한계를 극복하고 특정 지역의 기후, 토양, 그리고 생태계를 총체적으로 분석하여 식량 생산과 환경 보전을 동시에 달성하는 지능형 생태계 농업(Intelligent Ecosystem Farming)을 구현할 것이다. 이는 인류가 지구 환경의 적극적인 설계자이자 관리자가 되는 것을 의미한다.

2) 주요 세부 기술 전망

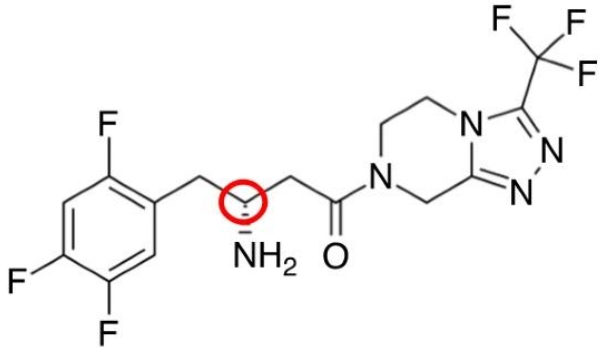
그린 바이오 분야에서의 세부 기술을 전망해보면 장기적으로 기후 변화에 대응하는 작물을 창조하는 단계에 이를 수 있을 것이다. AGI는 고온, 사막, 또는 염분 토양 등 극한 환경에서도 생존하고 높은 생산성을 내는 새로운 식물 종을 유전자 코드 설계(Genetic Code Design)를 통해 창조할 수도 있을 것이다. 이는 기존의 품종 개량을 넘어, 원하는 기능을 가진 유전자를 조합하여 완전히 새로운 생명체를 만드는 단계라 할 수 있다. 그린 바이오 분야에서 인공지능의 중요성에 대해 〔그림 7-9〕에서 보여주고 있다(Pandey et al., 2024).

〔그림 7－9〕 그린 바이오 분야에서의 인공지능의 중요성



이 밖에도 개인 맞춤형 영양 식품이 개발될 수도 있다. 개인의 장내 미생물(마이크로바이옴)과 유전 정보에 맞춰 영양소 흡수를 최적화한 맞춤형 식품이 AGI에 의해 설계될 수 있다. 새롭게 설계된 식품은 수직농장이나 식품 프린터를 통해 즉시 생산되어 가정으로 배송되는 서비스가 나타날 수도 있다. 〔그림 7－10〕에서는 합성을 통해 새로운 유형의 식품 또는 분자 구조를 생산하는 연구의 사례를 보여주고 있다(Voigt, 2020).

〔그림 7－10〕 새로운 유형의 식품 또는 분자 구조 생산 연구 사례



이 밖에 자율 농업 에이전트 및 로보틱스의 융합으로 드론, 로봇, 센서, 토양 및 기상 시스템을 AGI가 통합 제어할 수 있을 것이다. 이러한 스마트 농장 에이전트를 통해 실시간 작물 관리와 시비가 가능할 수 있고 관개 및 병해충 대응을 자동화할 것이다.

또한, 생태 복원 및 탄소 순환 바이오 시스템에서는 식물, 미생물, 그리고 토양 생태계를 대상으로 AGI 기반 복원 전략 모델이 기후 대응, 탄소 저감, 그리고 지속 가능성 설계를 가능하게 할 것이다.

3) 경쟁력 확보를 위한 핵심 전략

그린 바이오 분야의 경쟁력 확보를 위해서는 현장 환경 다양성과 예측 불확실성에 대한 극복이 필수적이라 할 수 있다. 또한 그린 바이오 산업 특성상 요구되는 막대한 인프라 비

용과 농업 데이터의 지역 편향성도 극복해야 한다. 그 밖에 유전자 편집과 생태 리스크에 대해 필수적으로 동반되는 사회적 수용성 이슈도 해결해야 할 문제라 할 수 있다.

다. 화이트 바이오 분야

1) 전반적 전망

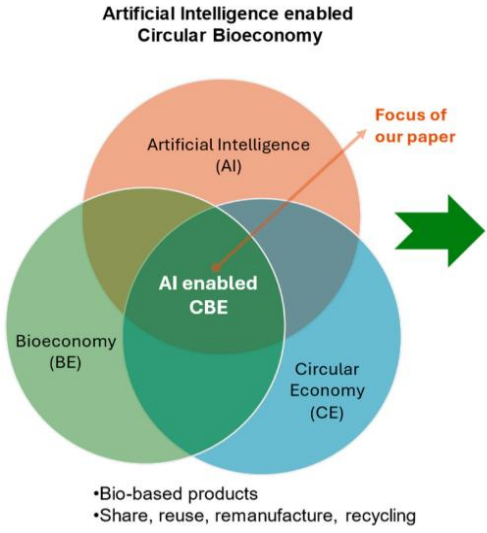
향후 10년 후 화이트 바이오 분야는 AGI가 도시와 산업에서 발생하는 모든 폐기물과 탄소를 자원으로 전환하는 완전 순환 바이오 경제(CBE; Circular Bioeconomy)를 완성할 수도 있을 것이다. AGI는 자연의 물질 순환 원리를 모방하여, 인간의 산업 활동이 지구 환경에 부하를 주지 않는 지속 가능한 시스템을 설계하고 운영하는 것이 가능할 수도 있다.

2) 주요 세부 기술 전망

화이트 바이오 분야에서의 세부 기술을 전망해보면 장기적으로 환경 정화용 합성 미생물을 고려할 수 있다. AGI는 해양의 미세 플라스틱을 분해하거나 대기 중의 이산화탄소를 흡수하여 유용한 바이오 연료나 플라스틱 원료로 전환하는 맞춤형 합성 미생물(Designer Synthetic Microbes)을 설계하는 단계에 이를 수 있다. 이는 환경 문제를 해결하는 동시에 새로운 자원을 창출하는 혁신을 이끌 수 있다(Shah et al., 2025). 〔그림 7-11〕에서 인공지능을 활용하는 순환 바이오 경제의 개념 모델을 보여주고 있다. 이를 위해 AGI는 발효, 정제, 그리고 효소 반응과 같은 공정 전체를 디지털 트윈화하고 실시간 적응형 제어를 수행할 것이다.

또한 AGI는 자가 치유 및 적응형 소재 개발도 주도할 것으로 보인다. 외부 환경 변화에 스스로 반응하여 형태나 기능이 변하고, 손상 시 자가 치유 능력을 갖춘 살아있는 소재(Living Materials)를 AGI가 설계하고, 이를 건축, 섬유, 전자제품 등에 적용하여 제품의 수명을 연장하고 폐기물을 원천적으로 줄일 것으로 기대된다. 이를 위해 AGI가 분자 수준 생합성 경로를 탐색하고, 미생물 세포 공장으로 자동 배포하고 최적화하는 방식이 가능해질 것이다.

〔그림 7－11〕 인공지능을 활용하는 CBE 개념 모델



	Intended Outcomes
Environment	Reduced GHG emissions, waste reduction, ecosystem regeneration
Resources	Efficient use of biomass, cascading value chains, nutrient cycling
Economy	Green jobs, rural development, bio-based innovation
Innovation	New materials, circular business models, biotechnologies
Food & Agriculture	Less food waste, better soil health, more resilient food systems
Urban Systems	Biowaste valorisation, sustainable packaging

3) 경쟁력 확보를 위한 핵심 전략

화이트 바이오 분야의 경쟁력 확보를 위해서는 공정 안정성과 안전성, 그리고 스케일업의 불확실성을 극복해야 한다. 또한 센서 기반 실시간 정보 부족 문제와 데이터 노이즈 문제를 해결하여 고품질의 대규모 데이터 확보가 전제되어야 한다. 그 밖에 AGI 시스템의 활용에 있어 비용 구조에서의 경쟁력 확보도 필수적이라 할 수 있다.

제 3 절 금융/경제

금융·경제 부문은 인공지능 기술이 산업 전반에 가장 빠르게 확산되고 있는 분야 중 하나로, 데이터 기반 의사결정과 위험 관리가 산업 구조의 핵심을 이루고 있다. 2025년 현재, 글로벌 금융권은 규칙 기반의 특정 태스크를 수행하는 좁은 인공지능(Narrow AI)을 넘어 범용 인공지능(AGI)으로의 이행이 진행되고 있으며, 이는 금융의 효율성 및 안정성, 투명성을 근본적으로 재편하는 변곡점이 되고 있다. 대한민국 역시 이러한 흐름에 발맞추어 AGI 기술을 중심으로 금융 서비스의 자동화와 지능화를 추진하고 있다. 이 절에서는 금융·경제 분야의 AGI 기술 활용 현황 및 중기(5년 후), 장기(10년 후) AGI 활용 전망 및 경쟁력 확보를 위한 대응 전략에 대해서 다룬다.

1. 금융 · 경제 분야 AGI 기술 활용 현황

가. 글로벌 금융 산업의 AGI 활용 현황

2025년 현재 글로벌 금융 산업은 '협의의 인공지능(Narrow AI)를 넘어, 범용 인공지능(General-Purpose AI) 및 파운데이션 모델 단계로 진입하였다. 과거에는 이상거래탐지(FDS)나 규칙 기반 챗봇 등 제한된 업무 자동화 중심이었다면, 최근에는 대규모 언어 모델(LLM)과 에이전트형 AI 시스템(Agentic AI)이 결합되어 금융의 핵심 기능(리스크 관리·자산 운용·투자 의사결정 등)을 구조적으로 변화시키고 있다. 세계 주요 금융기관들은 AI를 단순히 '기술 활용 수단'이 아니라 핵심 경쟁력 확보의 인프라로 인식하며, 자체 연구개발(R&D)과 데이터 생태계 구축에 대규모 투자를 하고 있다.

JPMorgan Chase는 전 세계 은행 중 AI 관련 논문 비중 37% 이상을 차지하며 독자적 금융 특화 LLM을 운영중이고, 이로 인해 연간 20억 달러 이상의 비용 절감 효과를 달성하였음을 발표하였다. Goldman Sachs는 AI 기반 파생상품 구조 설계 도구 'Visual Structuring'을 상용화해 트레이딩 효율성과 정확성을 향상시켰다. 또한 DBS Bank는 370개 이상의 활용 사례에 걸쳐 1,500개 이상 AI 모델을 운영하며, 'PURE(Principled, Unbiased, Relevant, Explainable)' 윤리 프레임워크를 통해 신뢰 가능한 AI 거버넌스를 실현하고 있다. HSBC는 'HSBC AI Markets'와 'Wealth Intelligence'를 통해 자연어 처리와 대규모 언어 모델을 활용

한 트레이딩 분석, 리스크 관리, 맞춤형 투자 인사이트 제공 등 고도화된 AI 자문 서비스를 운영하고 있다. 또한 금융 정보 단말기 시장의 강자인 Bloomberg는 500억 파라미터 규모의 BloombergGPT를 개발하였는데, 이는 일반적인 NLP 벤치마크에서도 동등하거나 더 나은 성능을 보였으며 비슷한 규모의 기존 오픈 모델보다 금융 업무에서 매우 우수한 성능을 보였다.

〔그림 7－12〕 BloombergGPT의 금융 및 범용 태스크 성능 비교 결과

Finance-Specific	**BloombergGPT**	**GPT-NeoX**	**OPT-66B**	**BLOOM-176B**
Financial Tasks	**62.51**	51.90	53.01	54.35
Bloomberg Tasks (Sentiment Analysis)	**62.47**	29.23	35.76	33.39

General-Purpose	**BloombergGPT**	**GPT-NeoX**	**OPT-66B**	**BLOOM-176B**	**GPT-3**
MMLU	**39.18**	35.95	35.99	39.13	**43.9**
Reading Comprehension	**61.22**	42.81	50.21	49.37	**67.0**
Linguistic Scenarios	**60.63**	57.18	58.59	58.26	**63.4**

이러한 흐름은 금융 산업이 JPMorgan Chase나 Bloomberg 처럼 자체적인 AI 연구개발 역량을 바탕으로 기술 생태계를 주도하는 '인프라 구축자(Infrastructure Builder)' 그룹과 대한민국 금융권을 포함한 대다수의 '응용기술 채택자(Application Adopter)' 그룹으로 양극화되고 있음을 시사한다. 전자는 독자 모델과 데이터 생태계를 확보하여 기술 종속을 탈피하고 장기 경쟁력을 확보하는 반면, 후자는 외부 기술 의존도를 높이며 근본적인 기술 격차를 심화시키고 장기적으로는 기술 종속의 위험을 내포한다.

나. 대한민국 금융권의 AGI 도입 현황

국내 금융권의 인공지능 도입은 기존 머신러닝 기반 분석 중심에서 생성형 AI 및 대규모 언어 모델(LLM)을 활용한 지능형 자동화 체계로 전환되는 단계에 있다. 2025년 현재, 주요 시중은행들은 내부망 보안체계를 유지하면서도 생성형 AI의 잠재력을 검증하기 위한 자체 모델 구축, 내부 비서형 에이전트 도입, 고객 서비스형 AI 행원 운영을 추진하고 있다. 이러한 흐름은 금융산업이 기존의 정형데이터 분석 중심에서 자율 추론과 자연어 기

반 대화형 지능 시스템으로 진화하고 있음을 보여준다. 다만 금융 데이터의 보안성 등의 규제로 인해 현재는 도메인 특화된 생성형 AI 또는 준(準)에이전트 형태이며, 완전한 범용 AGI 수준을 활용한 사례는 아직 드물다.

1) 내부 운영 효율화: 생성형 AI 업무 비서 및 내부망 기반 모델 구축

신한은행은 'AI ONE' 업무 비서 플랫폼을 통해 직원들이 금융 지식 Q&A, 주요 시장지표 확인, 마케팅 타겟리스트 작성, 대출업무에서의 서류 발송 등 약 40여 종의 업무를 생성형 AI 기반으로 수행하도록 지원하고 있다. AI ONE 플랫폼에는 약 10만 건의 사내 문서와 업무 지식을 학습시킨 GPT 모델이 활용되었으며, 2025년 3월 금융보안원 보안 평가를 최종 통과했다. KB국민은행은 2023년 4월부터 12월까지 'KB-GPT'를 운영하며 금융 서비스 내 검색, 채팅, 요약, 문서 작성, 코딩 기능을 모두 GPT로 처리할 수 있도록 지원하고, 총 9개의 GPT 기반 앱 기능을 데모로 사용할 수 있도록 제공했다. 또한 'KB AI Translator' 서비스를 함께 개발하여 금융 분야 특화 번역 및 외국인 고객 응대, 문서 번역과 같은 업무를 자동화하였다. 케이뱅크는 2025년 2월 KT, 업스테이지와 함께 MOU를 체결하여 케이뱅크 앱에 KT의 믿음과 업스테이지의 솔라 모델을 베이스로 금융 특화 LLM을 도입했다. 그 외 여러 은행들도 직원 내부 업무를 중심으로 생성형 AI 기술 검증(PoC)을 수행하고 있는 추세이다.

2) 고객 서비스 증강: 대화형 에이전트 기반 AI 행원 확산

고객 접점에서는 대화형 LLM 에이전트가 확산되고 있다. 신한은행은 전국 영업점에 약 150대의 '디지털 데스크 AI행원'을 배치하여 자연어 기반 금융상담과 외국어 번역 서비스를 제공하고 있으며, 무인 점포 형태의 'AI 브랜치'를 함께 운영하며 예적금 신규, 계좌 입출금, 증명서 발급 등 창구에서 주로 발생하는 60여 개의 업무를 셀프뱅킹(Self banking) 형태로 처리할 수 있다. NH농협은행은 국내 시중은행 중 최초로 전국 1,100여 개 영업점에 AI 행원을 배치했다. 상품 설명 및 표준 상담 업무를 수행하고 있으며, 고령층과 외국인을 위한 음성 및 텍스트 기반 생성형 상담 서비스도 병행하고 있다. 마지막으로 KB국민은행은 개인 맞춤 'AI 금융비서'를 개발하여 영상합성 엔진, STT(Speech-to-Text), TTS(Text-to-Speech), 챗봇, 딥러닝 등의 기술을 바탕으로 입출금 내역 조회, 계좌 이체, 금융 관련

질의 응답, 금융 상품 소개 및 필요 서류 안내 등 개인화된 맞춤형 서비스 상용화를 추진하고 있다. 이러한 서비스들은 단순 질의응답 수준을 넘어, 고객 데이터와 실시간 연계된 반응형 AI 에이전트 구조이다.

3) AI 인프라 및 생태계 고도화: 자체 모델·플랫폼·데이터 구축

국내 은행들은 생성형 AI를 안정적으로 운영하기 위해 전용 인프라 및 데이터 생태계 조성에도 투자하고 있다. 카카오뱅크는 2024년 AI 전용 데이터센터를 개소하여, 내부 AI 디비전과 AI 프로덕트를 전반을 통합 관리하는 'AI 에코시스템' 전략을 수립하였다. 해당 센터는 모델 학습·검증·배포의 전 과정을 내부망에서 안전하게 처리하도록 설계되어 있으며, 이를 통해 고객 정보 보호와 규제 준수를 동시에 달성하고 있다. 또한 이 인프라는 향후 다중 에이전트 기반 금융 서비스 실험 및 AI 의사결정 자동화를 위한 핵심 기반으로 활용될 수 있다.

금융당국 또한 2025년 상반기에 '금융권 AI 플랫폼'을 구축하여 전문가 그룹이 선정한 오픈소스 AI 모델을 내부망에 바로 설치할 수 있도록 지원하고, 내부 업무에 적용하기 전에 성능을 점검할 수 있는 환경을 제공함으로써 AI 인프라 구축 비용을 절감할 수 있도록 지원한다. 이와 함께 금융 특화 한글 말뭉치 및 보안 처리된 공통 데이터셋을 제공함으로써, 금융기관이 RAG(Retrieval-Augmented Generation) 기반 서비스나 내부 특화 LLM 학습에 활용할 수 있도록 환경을 조성하고 있다.

4) 정책 및 제도 현황

국내 은행 전산망은 기본적으로 인터넷망과 분리된 환경에서 운영되므로 인터넷을 기반으로 한 서비스들을 업무에 직접 활용하기엔 제약이 많다. 이러한 환경에서 정부는 금융권의 생성형 AI 활용을 촉진하기 위해 '혁신금융서비스 지정 제도(규제 샌드박스)'를 운영하고 있다. 이 제도는 망분리 규제로 제한받던 금융사들이 상용 인터넷 기반 AI 서비스를 제한적으로 활용할 수 있도록 허용하는 제도적 장치이다. 2024년 12월 기준 141건, 2025년 3월 기준 총 199건의 신청이 접수되었으며, 사례로는 신한은행의 생성형 AI 기반 AI 은행원 및 투자·금융지식 Q&A 서비스가 있으며, KB국민은행의 생성형 AI 금융상담 Agent, NH농협은행의 생성형 AI 플랫폼 기반 금융서비스, 카카오뱅크의 대화형 금융계산기가 있

다. 이러한 서비스들은 고객 상담 및 금융지식 제공 등 고객 경험 개선형 서비스에 집중되어 있다.

정부는 동시에 AI 리스크 관리 및 거버넌스 체계를 강화하고 있다. 금융위원회와 금융보안원은 정확성, 보안, 개인정보 보호를 중심으로 한 가이드라인을 마련 중이며, 혁신금융서비스 지정 기업에도 보안성 검증과 개인정보 관리 요건을 의무화하고 있다. 이러한 접근은 혁신 촉진과 리스크 통제의 병행을 목표로 하며, 금융권의 안전한 생성형 AI, 더 나아가 안전한 AGI의 도입 기반을 구축하는 방향으로 진행되고 있다.

2. 중기(5년 후) 금융 · 경제 분야 AGI 활용 전망

가. 금융 특화 파운데이션 모델의 확산과 활용 생태계 전망

2025년 이후 5년간은 금융 산업이 단순한 LLM 활용 단계를 넘어서, 금융 도메인에 최적화된 파운데이션 모델(Financial Foundation Model, FFM) 생태계를 구축하는 시기가 될 것으로 전망된다. 금융 특화 파운데이션 모델(FFM)은 단순히 금융 언어를 이해하는 모델이 아니라, 기존의 LLM이 가진 일반 언어 처리 성능을 유지하면서 금융 도메인에서의 정확성과 신뢰성을 함께 확보하는 것을 목표로 한다.

현재 Bloomberg의 500억 개 파라미터 규모 모델인 BloombergGPT는 금융 언어 태스크에서 기존 LLM을 능가할 수 있음을 보여준 사례이다. 이러한 흐름을 바탕으로 향후 5년 내에 다음 세 축을 중심으로 금융 특화 파운데이션 모델이 본격화될 가능성이 높다:

1) 금융 언어 파운데이션 모델(Financial Language Foundation Model, FinLFM)
2) 금융 시계열 파운데이션 모델(Financial Time-Series Foundation Model, FinTSFM)
3) 금융 비전-언어 파운데이션 모델(Financial Vision-Language Foundation Model, FinVLFM)

〈표 7－8〉 금융 특화 파운데이션 모델(FFM) 유형별 비교 및 대표 모델

FFM 유형	설명	주요 입력 데이터	핵심 활용 분야	대표 연구/모델
FinLFM	금융 텍스트와 지식 중심의 모델로, 문서 이해, 요약, 감성 분석 등에 특화	• 금융 보고서 • 뉴스 기사 • 계약서 • 공시 자료 등	• 시장 동향 예측 • 투자 보고서 자동 생성 • 규제 준수 자동화	BloombergGPT(M. Chen, D et al., 2024), FinGPT (OpenAI, 2023)
FinTSFM	고차원적인 금융 시계열 데이터를 중심으로 패턴 학습, 예측, 변동성 모델링, 이상 탐지 등에 중점	• 주가 및 거래량 • 금리, 경제 지표 등	• 알고리즘 트레이딩 • 실시간 리스크 평가 • 포트폴리오 관리 최적화	TradeFM (J. Hoffmann et al., 2024)
FinVLFM	금융 텍스트와 차트, 표, 이미지 등을 동시에 처리하는 멀티모달 모델	• 차트 및 이미지, 표 등을 포함한 재무 보고서	• 투자 분석 고도화 • 기업 실사(Due Diligence) 자동화 • 정보 검색 및 추출	FinLLaMA (R. Anil et al, 2023), FinTral (D. Harvey, et el., 2025)

이러한 금융 특화 파운데이션 모델들은 재무 문서 요약, 감성 분석, 주가 변동성 예측, 리스크 평가 등 금융의 핵심 기능을 자동화함으로써, 금융기관 내부의 AI 스택(AI Stack) 구축을 가속화하고 산업 전반의 업무 효율성을 높일 것이다.

한편, 글로벌 클라우드 사업자(예: AWS, Microsoft Azure, Google Cloud 등)는 금융기관이 안전한 환경에서 대규모 모델을 독자적으로 개발하거나 미세조정(Fine-tuning)할 수 있도록, 보안 강화형 데이터 환경, 전용 AI 컴퓨팅 인스턴스, 파이프라인 자동화 도구 등을 포함한 인프라와 서비스를 지속 확대하고 있다. 이러한 인프라는 향후 금융기관이 고비용의 독자 개발 없이도 첨단 기술을 활용하여 금융 특화 파운데이션 모델(FFM)을 미세조정하고 자사 서비스에 통합할 수 있는 실질적인 기반이 될 것으로 전망된다.

국내에서도 금융데이터거래소(KDX), 금융보안원, 한국금융연구원 등을 중심으로 금융 데이터의 안전한 유통 및 활용을 위한 데이터 허브 고도화가 본격 추진될 것으로 보이며, 이에 따라 금융 도메인 특화 모델 개발을 위한 협력 연구 및 시범 사업이 확산될 것으로 예상된다. 이처럼 중기적으로는 대형 파운데이션 모델을 완전히 독자적으로 개발하기보다

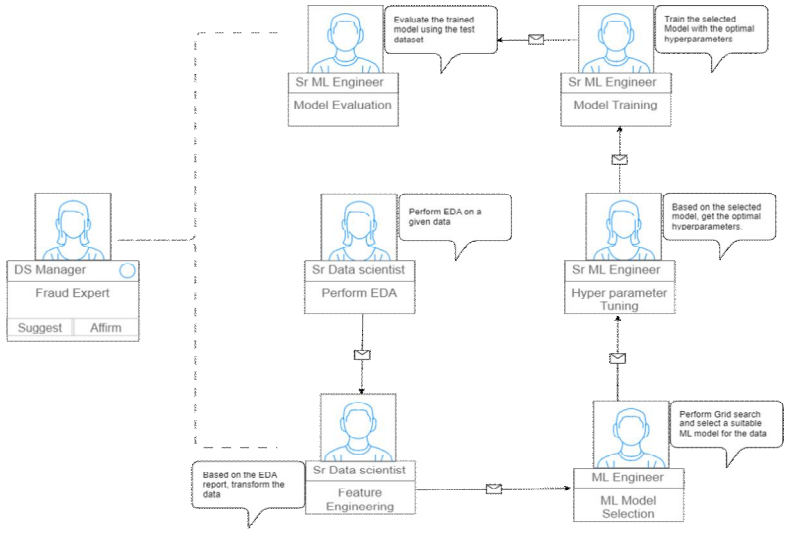
는, 글로벌 금융 특화 파운데이션 모델을 국내 환경과 규제 체계에 맞게 미세조정(fine-tuning)하는 방향이 주류 전략으로 자리잡을 가능성이 높다.

나. 금융 지능형 에이전트의 부상과 서비스 자동화 전환 전망

중기(2025-2030)에는 AI가 자율적 의사결정과 실행을 수행하는 '에이전트형 AI(Agentic AI)'로 확산될 것이다. Morgan Stanley는 2025년을 "에이전트 AI의 해"로 규정하며, 금융산업의 주요 변곡점을 예고하였다.

에이전트형 AI는 '관리자 에이전트', '데이터 분석 에이전트', '모델 검증 에이전트' 등으로 구성된 AI 팀(Crew)을 통해 신용 리스크 모델링이나 모델 리스크 관리와 같은 고도화된 금융 업무 전체를 인간의 감독 하에(human-in-the-loop) 자율적으로 수행한다. 이러한 에이전트 시스템은 초과 수익률 발굴, 포트폴리오 최적화, 거시 경제 시뮬레이션, 규제 준수 자동화 등 금융의 핵심 영역에 적용되어, 금융 운영의 속도와 규모를 기하급수적으로 확장시킬 잠재력을 가지고 있다. 이러한 구조는 단순 업무 자동화를 넘어, 인간이 AI 도구를 '활용'하던 단계에서 인간이 AI 팀을 '관리·감독'하는 단계로의 패러다임 전환을 의미하기도 한다.

〔그림 7－13〕 사기 탐지 모델링을 위한 AI 크루의 역할 수행 및 협업 구조



최근 글로벌 금융기관들은 대규모 언어 모델(LLM)을 기반으로 한 금융 지능형 에이전트(Financial Intelligent Agents)를 핵심 경쟁력으로 내세우며, 기존의 디지털뱅킹 서비스에서 한 단계 진화한 자율형 금융 운영 체계(AI-Driven Finance)로 전환하고 있다. 이러한 흐름은 단순한 챗봇이나 자동 응답 수준을 넘어, 고객 맞춤 자산관리, 리스크 평가, 결제 관리 등 복합적 의사결정 업무를 수행하는 에이전트형 AI의 상용화를 통해 가속화되고 있다. 대표적으로 미국 뱅크오브아메리카(Bank of America)는 2018년 출시한 가상 금융 비서 '에리카(Erica)'를 통해 금융 지능형 에이전트의 대규모 상용화를 선도하고 있다(C. Schuhmann et al., 2022). 에리카는 자연어 처리와 예측 분석 기술을 결합해 고객의 금융 상황을 실시간으로 분석하고 안내하며, 출시 이후 6년 만에 누적 20억 회 이상의 상호작용을 기록하였다. 현재 4,200만 명 이상의 고객이 에리카를 통해 계좌 관리, 송금, 소비 분석, 구독 모니터링, 자산 리스크 안내 등 30여 종 이상의 기능을 일상적으로 활용하고 있다. 이용 고객의 98% 가 평균 44초 내에 원하는 답변을 획득했다고 평가하며 금융권 내에서 가장 성공적인 에이전트 고객 서비스 사례로 평가받는다. 이와 함께, JPMorgan Chase의 COiN(Document Intelligence), Morgan Stanley의 AI 리서치 어시스턴트인 AskResearchGPT 등 주요 글로벌 기관들도 AI 에이전트를 자산관리 및 리스크 관리, 기업 금융 전반에 확대 적용하고 있다.

반면, 우리나라의 금융권은 AI 에이전트 서비스를 실험적으로 도입하고 있다. 국내 은행 및 증권사는 주로 내부망 규제로 인해 일부 은행사에서만 'AI 행원', '업무 비서형 AI' 등을 시범 운영하고 있으나, 향후 5년 내(2025-2030년)에는 정부의 AI 혁신금융서비스 확대와 금융보안원 주도의 기술 검증 체계 정립, 국내 금융기관의 자체 모델 및 인프라 구축이 병행됨에 따라 지능형 에이전트의 본격적 상용화 단계로 진입할 것으로 전망된다. 특히 고도화된 금융 파운데이션 모델(FFM)과 멀티 에이전트 운영 프레임워크가 결합되면, 향후 금융기관 내 AI가 단순한 조언자를 넘어 자율적 판단과 실행을 수행하는 준(準)AGI 수준의 금융 운영체계로 발전할 것으로 예상된다.

다. AGI 도입에 따른 제도적·경제적 전환 전망

1) 제도 및 규제 환경의 변화

향후 5년 내 대한민국 금융 및 경제 부문은 AGI 도입을 위한 제도적 실험기에서 제도화

의 초기 단계로 이행할 것으로 전망된다. 현재 금융당국은 혁신금융서비스(규제 샌드박스)를 통해 AI 기반 금융서비스의 실증과 검증을 병행하는 체계를 운영 중이며, 중기적으로는 이러한 실증 중심의 제도가 AGI 응용 기술의 상용화 촉진 플랫폼으로 기능하게 될 것이다.

정부는 '혁신 촉진'과 '신뢰 구축'이라는 이중 트랙 규제 모델을 유지하며, 한편으로는 혁신금융서비스를 통한 시장 진입 완화를 지속하면서도, 다른 한편으로는 AI 모델의 정확성·보안·설명가능성에 관한 인증제 도입을 추진할 가능성이 높다. 이 과정에서 금융위원회와 금융보안원은 공동으로 모델 신뢰성 평가 지침을 표준화하고, 데이터 결합 제한 규제를 완화하여 금융-비금융 데이터 융합 기반 AGI 학습 허용 범위를 확대할 것으로 예상된다. 또한, 이미 금융 및 경제 도메인에서 AGI 서비스 상용화를 선도하고 있는 글로벌 주요국의 규제 방향이 국내 정책 설계의 주요 준거점으로 작용할 것으로 보인다.

2) AGI 도입의 금융·경제적 파급

AGI 수준의 지능형 시스템이 금융 부문에 확산되면, 5년 내 한국 경제 전반에 다음과 같은 거시경제적 변화가 예상된다. 첫째, 금융 효율성의 극대화이다. AI 기반 위험 평가, 포트폴리오 관리, 자금세탁방지(AML) 등 복잡 업무의 자동화는 금융기관의 생산성 향상과 거래비용 절감을 동시에 실현한다. 한국은행 시뮬레이션에 따르면, AI 기술 확산은 GDP 최대 12.6%, 총요소생산성(TFP) 최대 3.2% 상승 효과를 유발할 수 있다. 둘째, AGI의 파급력이 실물경제로 확장된다. 금융·제조·유통 등 산업 전반에서 자금조달, 신용평가, 위험예측이 AGI 기반으로 자동화되며, 이는 투자 효율성 향상과 산업간 자본 흐름 최적화를 유도한다. 특히 중소기업의 자금 접근성이 개선되고, 산업 내 자본배분의 투명성이 향상될 것으로 기대된다. 셋째, 노동시장 및 고용구조의 이중화이다. AGI 도입으로 데이터 과학, 리스크 관리, 투자 전략 등 고숙련 직무는 생산성과 임금이 상승하는 반면, 단순 사무직이나 반복적 서비스 직무는 AI 대체 위험이 높아질 것으로 예측된다. 이로 인해 산업 내 인력 재교육(Re-skilling)과 AGI 활용 역량 강화가 국가 차원의 핵심 과제로 부상할 전망이다. 마지막으로, 경제 구조의 자동화·플랫폼화 심화이다. AI 및 데이터 중심 경제로의 전환이 가속되며, 금융기관은 단순 서비스 제공자에서 데이터 기반 경제 인프라 운영자로 전환된다. 이는 AGI가 금융시장의 안정성을 관리하고, 정책결정에 필요한 거시지표를 실시간 산출하는 등 AI 기반 경제 운영체계로 발전하는 과도기적 단계로 해석된다.

3) AGI 규제 및 경제 정책의 향후 방향

중기적으로 대한민국은 AGI의 금융·경제적 활용 확대에 따라 정책의 초점을 '통제'에서 '조율'로 전환될 것이다. 현재의 데이터 보안 중심 접근은 유지하되, 금융·산업 전반에서 AGI를 활용하기 위해 법제·윤리·경제 3축 프레임워크가 병행될 필요가 있다. 먼저 법제적 측면에서는 AI 모델에 대한 인증제 및 사전검증제 도입이 이루어질 것이며, AI 행위 책임 범위를 명확히 하는 'AGI 신뢰성법' 제정 논의가 진행될 가능성이 높다. 경제적책 측면으로는 AI 기반 산업 자동화에 따른 노동력 대체 문제를 완화하기 위해 디지털 전환 인력 재교육 프로그램과 혜택 제도화가 필요할 것이다. 마지막으로 금융 안정성 측면에서 금융보안원은 AI 리스크 대응을 위해 모델 보안성 평가 및 사후 모니터링 체계를 확립할 것이다. 이러한 대응은 AGI 기반 시스템 리스크를 완화하고, 글로벌 규제기관과의 협력 기반을 구축하는 방향으로 전개될 것이다.

요약하면, 2025~2030년의 대한민국 금융·경제 분야는 AGI 상용화의 제도화 초기 단계로 진입하게 될 것이다. 금융권은 Narrow AI 중심 효율화에서 벗어나 FFM 및 Agentic AI를 활용한 자율적 의사결정 체계로 전환하며, 정부는 이에 대응해 규제 완화와 윤리 검증 병행 정책을 본격화할 것으로 예상된다. 이러한 변화는 단기적으로는 금융 효율성 및 생산성 향상을, 중기적으로는 산업 구조 재편과 거시경제 안정화 기여를 가져올 것이다. 반면, 대형 금융사 중심의 기술 집중, 노동시장 불균형, 데이터 의존성 증대 등의 부작용이 나타날 수 있어 AI 거버넌스 및 경제정책적 대응의 통합 관리체계 구축이 필요하다.

3. 장기(10년 후) 금융 · 경제 분야 AGI 활용 전망

가. 개요: 중기 이후의 과제와 장기적 전환 방향

2030년대 초반, 대한민국 금융·경제 부문은 AGI 응용의 제도화 단계를 넘어 지속 가능한 생태계 확립기로 진입할 것으로 예상된다. 이는 중기 단계에서 드러난 문제점—언어모델 의존, 데이터 한계, 규제 불균형—을 체계적으로 보완하는 시기로 보는 것이 타당하다.

중기(2025-2030)에 이르는 동안, 금융기관들은 글로벌 범용 LLM을 파인튜닝한 금융 특화 파운데이션 모델(FFM)을 적극 도입하고, 이에 기반한 에이전틱(Agentic) AI 서비스를 내부 업무 및 고객 상담 영역으로 확신시킬 것이다. 또한, 정부는 모델 인증·검증·감독

체계를 제도화하며 금융 AI의 제도권 편입을 추진할 것으로 보인다.

그러나 이러한 전환 과정에서 다음과 같은 구조적 문제가 불가피하게 등장할 가능성이 있다. 첫째, 해외 범용 모델에 대한 과도한 의존은 금융 데이터의 맥락적 해석 오류와 규제 해석의 불일치 문제를 야기할 수 있다. 둘째, 기관 간 데이터 품질 및 접근 격차로 인해 AI 성능과 신뢰성의 양극화가 심화될 수 있으며, 특히 비표준 데이터(서식·용어·법률 문체 등)에 대한 모델의 일반화 실패가 금융 리스크로 전이될 우려가 있다. 셋째, AI 의사결정의 불투명성(Black-box problem)과 책임 주체 불분명성이 확대되면서, 감독당국이 제시한 설명가능성 및 공정성 기준에 대한 실증적 검증 체계가 부족하다는 한계를 맞닥뜨릴 수 있다. 마지막으로, 에이전트형 AI의 자율 실행 범위와 법적 책임 경계에 대한 명확한 기준이 부재하여, 금융소비자 보호 및 금융사고 책임소재 등 제도적 리스크가 병존할 가능성도 있다. 따라서 2030년대 초반은 이러한 중기 단계의 기술적·제도적 불균형을 교정하고, 대한민국의 금융 AI 생태계의 독립성과 지속 가능성을 강화하는 시기로 전환될 것으로 전망된다.

특히 중기까지 국내 금융기관들은 글로벌 FFM을 미세조정한 모델에 의존할 것이나, 장기적으로는 한국어 중심의 금융 특화 모델(Ko-FFM) 구축이 핵심 경쟁력으로 부상할 가능성이 높다. 이는 단순히 번역 정확도를 개선하는 수준이 아니라, 국내 금융 문서의 문체 및 계약 용어, 규제 언어 등 고유한 금융 언어 체계 전반을 정밀하게 반영하는 도메인 적합도 확보의 문제이기 때문이다. 따라서 중기 이후에는 해외 모델의 단순 활용을 넘어, 한국 금융 환경에 최적화된 파운데이션 모델(Ko-FFM, Ko-FinVLFM 등)을 독자적으로 개발 및 운영하는 방향으로 전환될 가능성이 있다. 이러한 국산 금융 AI의 자립화는 단순한 기술 독립을 넘어, 모델의 신뢰성·설명가능성·규제 적합성·안전성까지 국내 기준에 맞춘 통합 생태계 확립으로 이어질 것으로 예상된다.

나. 기술적 보완 과제: 언어 모델, 데이터, 해석 가능성

1) 언어 모델: 한국어 기반 금융 언어 모델의 필요성

현재 국내 금융 기관들은 GPT-4, Claude, Gemini 등 영문 중심의 범용 대규모 언어 모델(LLM)을 주로 활용하고 있으며, 일부 기관은 LLaMA, Mistral, Qwen 등 공개형 모델을 파인튜닝하여 내부 서비스에 적용하고 있다. 중기적으로는 이러한 범용 모델을 기반으로

한 외국어 기반의 금융 파운데이션 모델(FFM)이 활용될 가능성이 높지만, 장기적으로는 한국 금융 환경에 최적화된 한국어 기반 금융 언어모델의 구축이 필수적이다.

그 이유는 단순히 언어 번역의 문제가 아니라, 금융 및 경제 영역의 특성을 고려한 정확성과 안전성을 보장하기 위한 구조적 필요성 때문이다. 금융 대화나 문서는 일반 언어와 달리 금리, 환율, 지수, 단위, 수치 등의 정량 정보가 빈번하게 등장하며, 계약서, 회계문서, 감사보고서 등에는 문맥적 미묘함과 규제 언어의 해석 정확도와 직결된다. 따라서 일반적인 LLM이나 영문 데이터로 구성된 FFM을 그대로 활용할 경우, 수치 단위 오류 및 용어 오역, 정책적 문맥 왜곡 등으로 인한 할루시네이션(Hallucination) 및 보안, 신뢰성 리스크가 누적될 가능성이 크다.

이러한 한계를 극복하기 위해, 장기(2030-2035년)에는 금융 도메인에 특화된 한국어 데이터셋 구축 및 파운데이션 모델 학습 체계를 구축하고자 하는 시도가 있을 것으로 보인다. 이러한 한국어 금융 파운데이션 모델 구축은 국내 금융권에서 사용하는 약식 용어, 기관별 표현 차이, 회계/재무 보고서의 어미 구조 등 한국어 특유의 문법적 변형을 반영가능한 언어적 일관성 확보를 가능하게 하며, 원화 단위(KRW), 금리 및 세율 계산식 등 수치적 문맥을 정밀하게 해석하여 오답 가능성을 줄이고 수치와 단위의 정합성을 강화할 수 있다. 또한 국내 금융 감독 기관이 사용하는 규제 용어체계(K-IFRS 등)를 학습함으로써 정책 및 규제 문맥에 맞는 응답을 생성할 수 있을 것이며, 국내 자체 모델들을 개발함으로써 해외 클라우드 기반 API의 의존도를 낮추고, 민감한 금융 데이터의 국내 처리 및 검증 체계를 확립함으로써 AI 신뢰성(Trustworthiness)을 강화할 수 있을 것이다.

현재도 국내 연구기관과 산업계에서는 한국어 중심 대형 언어 모델(KoGPT, SOLAR, HyperCLOVA X 등)이 활발히 개발되고 있다. 이러한 모델들은 한국어 문맥 처리력, 맞춤형 파인튜닝 용이성, 고해상도 형태소 분석 성능 등에서 강점을 보이고 있다. 향후 이들 모델을 기반으로 금융 특화 데이터셋을 추가 학습하여, 한국어 금융 파운데이션 모델로 진화시킬 경우, 이는 데이터 주권 확보와 금융 AI의 안정성 및 신뢰성 제고를 동시에 달성할 수 있는 전략적 경로가 될 것이다.

2) 한국어 금융 도메인 데이터셋 구축 및 정합성 강화 전망

앞서 지적된 기관 간 데이터 품질 및 접근 격차 문제는 AI 성능의 양극화와 금융 리스크 전이로 이어질 가능성이 높다. 특히 금융권 데이터는 서식, 용어, 법률 문체 등이 기관별로

상이하고, 비표준 문서 구조가 많아 모델의 일반화 성능을 저하시킬 수 있다. 이러한 한계를 해소하기 위해, 국가 차원의 한국어 금융 도메인 데이터셋 구축 및 정합성(Consistency) 강화 전략이 장기적으로 추진될 것으로 전망된다.

우선, 데이터 표준화와 품질 인증 체계가 핵심 과제로 부상할 것이다. 금융위원회, 금융보안원, 한국은행, 금융감독원 등 주요 기관이 공동으로 1) 금융 데이터의 수집 및 가공, 활용 기준을 통일하고, 2) 문서의 포맷과 용어, 단위, 라벨링 체계를 표준화하며, 3) 데이터 품질을 인증하는 '금융 AI 학습데이터 품질지수(Financial Data Quality index, FDQI)'를 마련해야 할 것이다. 이를 통해 각 기관이 구축하는 데이터셋 간의 구조적 정합성을 확보하고, 모델 학습 단계에서 발생하는 데이터 편향 및 중복 문제를 최소화할 수 있을 것이다. 둘째, 금융 데이터의 특성상 민감정보 보호와 활용 간 균형이 중요하므로, 비식별화·합성데이터·연합학습(Federated Learning) 기반 데이터 활용 모델이 병행이 필요하다. 이를 통해 실제 고객정보를 직접 노출하지 않으면서도 고신뢰 학습 데이터를 지속적으로 확보할 수 있다. 셋째, 이러한 정합성 확보 노력은 앞서 제시한 한국어 금융 파운데이션 모델(Ko-FFM) 구축과 긴밀히 연계될 것이다. 표준화된 한국어 금융 데이터셋을 기반으로 Ko-FFM을 학습하면 금융 문서의 고유 구조적 문맥을 이해하고, 수치 및 단위, 법률 언어의 정밀 해석이 가능해지며 기관별 규제 문체 차이의 통합적 반영이 가능해진다. 이는 단순한 모델 성능 향상을 넘어, 국가 단위의 금융 AI 신뢰성 프레임워크 확립으로 이어질 것이다.

3) 설명가능성 및 투명성 강화

앞서 제기된 AI 의사결정의 불투명성과 책임 주체의 불명확성 문제는 금융 AI의 제도권 편입 과정에서 가장 시급히 해결해야 할 과제로 꼽힌다. 이에 따라 2030년대 초반에는 설명가능 인공지능(XAI) 기반의 모델 구조 개편과 책임 주체 명확화를 위한 제도적 틀의 본격화가 병행될 것이다.

우선, 금융당국은 모델의 내부 판단 과정을 외부 검증이 가능하도록 하는 설명가능 인공지능 표준(XAI-FIN Framework)을 마련해야 하며, 금융기관은 AI가 산출한 결과에 대해 입력 변수의 기여도, 판단 근거, 불확실성 수준을 명시적으로 기록 및 제출하도록 요구받는다. 이 과정은 기존 모델 리스크 관리(Model Risk Management, MRM) 체계와 연계되어 상시 점검되는 형태로 발전할 수 있다. 또한 AI 의사결정으로 발생한 결과에 대한 책임 주

체를 모델 개발 및 운영, 검증 단계별로 명확히 구분하는 제도 정비가 추진되어야 한다. 특히 AI가 자율적으로 의사결정을 수행하는 경우에도 결정 로그(Decision Log)를 통해 책임 추적이 가능한 AI 감사(Audit) 체계를 구축이 필요할 것이다. 이로써 AI의 판단 오류나 오남용이 발생하더라도, 책임 소재를 명확히 규정하고 제재 및 보상, 개선 절차를 일관되게 수행할 수 있는 기반이 마련될 것이다. 결국, 금융권의 AI 신뢰성 확보는 '모델의 투명성 확보'와 '책임 구조의 제도화'라는 두 축을 중심으로 진화할 것이며, 이는 금융 및 경제 산업에 AGI를 안정적으로 도입하기 위한 필수 전제 조건이 될 것이다.

다. 산업·경제 구조의 점진적 시스템화

금융·경제 분야에 AGI 기술 활용의 안정기가 찾아오면 장기적으로 금융 AGI는 산업과 경제 전반의 자율적 조정 시스템(Self-Adjusting System)을 향해 발전할 것이다. 이는 인간-AGI 협업 기반의 준시스템화로 접근될 가능성이 크다. 즉, 정책·금융·산업 각 부문이 AGI의 예측 및 판단 결과를 실시간 반영해 의사결정을 보완하는 협업적 시스템 운영 구조로 발전되는 것이다. 이 시기에는 다음과 같은 산업 구조 변화가 동반된다.

- **금융 부문**: 금융사별 독립적 모델에서, 금융권 공동 FFM 플랫폼으로 통합하여 데이터, 리스크 관리 효율 극대화
- **경제 부문**: AGI 기반 경제지표 예측·금리정책 시뮬레이션을 통한 정책 결정의 자동 보조 시스템화
- **산업 부문**: AGI를 통한 자본 배분 및 공급망 위험 예측을 통한 제조·유통·서비스 간 상호 최적화

4. 경쟁력 확보를 위한 대응 전략

2035년까지 금융·경제 분야의 AGI 도입이 안정기에 진입하기 위해서는, 기술적 자립 기반 확립과 함께 신뢰 가능한 제도적 관리체계 구축이 병행되어야 한다. 중기(2025-2030년) 단계에서는 금융 파운데이션 모델의 확산과 에이전트형 AI 서비스를 도입하기 위해 데이터 활용 인프라를 확장하고 규제를 완화하는 방향으로, 장기(2030-2035년) 단계에서는 국내 금융 AI 생태계의 독립성과 지속 가능성을 강화하는 방향으로 대응 전략이 전개되어야 한다.

가. 한국어 금융 파운데이션 모델 개발 생태계 구축

한국어 금융 파운데이션 모델 개발 생태계 구축은 금융·경제 분야의 핵심 기술 자립과 데이터 주권 확보를 목표로 한다. 이를 위해 국내 주요 연구기관, 금융사, 그리고 클라우드 사업자가 협력하여 '한국어 금융 파운데이션 모델(Ko-FFM)' 공동 개발 컨소시엄을 구성한다. 이 컨소시엄은 금융감독원과 금융보안원을 중심으로 규제 언어, 금융 보고서, 회계 및 감사 데이터 등 고품질 한국어 금융 코퍼스를 구축하고, 이를 기반으로 한 학습 데이터 생태계를 조성한다. 또한 HyperCLOVA X, KoGPT 등 기존 한국어 대규모 언어 모델을 토대로, 금융 용어 체계 및 계량 정보, 법률 문체를 반영한 도메인 특화 파인튜닝 프레임워크를 개발함으로써 모델의 정밀성과 전문성을 강화한다. 마지막으로, 한국어 모델을 활용한 AI 보안 검증 및 성능 인증 제도화를 추진하여, 금융권 내 상용 글로벌 모델 대비 신뢰성과 투명성 측면에서 우위를 확보하는 것을 전략적 방향으로 삼는다.

나. 금융 데이터 표준화 및 품질 인증 체계 확립

금융 데이터 표준화 및 품질 인증 체계 확립은 기관 간 데이터 품질 및 접근 격차를 해소하고, 모델 학습의 정합성을 확보하는 것을 핵심 목표로 한다. 이를 위해 금융 AI 데이터 표준(Financial AI Data Standard, FADS)를 제정하고, 금융 데이터의 형식 및 단위, 라벨링 체계를 일원화된 기준으로 통합한다. 또한, 데이터 신뢰도와 활용도를 정량적으로 평가하기 위한 '금융 AI 학습데이터 품질지수(FDQI)를 도입하여, 데이터셋의 품질 인증을 제도적으로 시행한다. 개인정보 보호 원칙을 전제로, 연합학습과 합성데이터 생성 허용체계를 병행 구축함으로써 데이터 활용과 보안 간 균형을 달성한다. 이러한 기반 위에서 학습-검증-피드백의 지속적 품질 관리 체계를 확립함으로써 금융 AI 생태계 전반의 데이터 신뢰성과 효율성을 제고한다.

다. 설명가능 인공지능(XAI) 기반 책임 구조 제도화

설명가능 인공지능(XAI) 기반 책임 구조 제도화는 AI 의사결정의 투명성을 확보하고 금융사고 발생 시 책임소재를 명확히 규정하는 것을 목표로 한다. 이를 위해 금융당국은 'XAI-FIN Framework'를 제정하여, AI 모델이 수행하는 모든 의사결정 과정에서 내부 판단 근거, 입력 변수의 기여도, 불확실성 정보를 의무적으로 기록하도록 규정한다. 금융기관은 기존의 모델 리스크 관리(MRM) 체계에 XAI 검증 프로세스를 통합하여, AI 의사 결정 로그

를 체계적으로 저장하고 외부 감사(Audit) 절차를 통해 투명한 검증이 가능하도록 운영한다. 또한 모델의 개발 및 운영, 검증 단계별로 책임 주체를 명확히 분리함으로써, 법적·제도적 책임 구조를 명문화한다. 아울러 감독기관은 AI 시스템의 오류나 오작동이 발생할 경우 제재 및 보상 프로세스를 자동화하는 체계를 구축하여, 금융 AI의 신뢰성과 공정성을 제도적으로 보장한다.

라. AI 인프라 및 연합 데이터 허브 고도화

AI 인프라 및 연합 데이터 허브 고도화는 안전한 AI 개발 및 운영 환경을 조성하고 산업 간 데이터 융합을 촉진하는 것을 목표로 한다. 이를 위해 금융보안원은 '금융권 전용 클라우드 샌드박스'를 구축하여, 외부망과의 연결 없이도 AI 모델의 학습과 검증, 배포가 가능한 안전한 폐쇄형 환경을 제공한다. 또한 금융데이터거래소(KDX)를 중심으로 산업별 데이터 허브 통합 체계를 구축함으로써, 금융 데이터와 비금융 데이터의 융합 활동을 촉진하고 새로운 데이터 기반 비즈니스 모델을 창출할 수 있는 기반을 마련한다. 정부는 GPU 및 AI 컴퓨팅 자원 확보를 위해 '금융 AI 컴퓨팅 풀(FinAI Compute Pool)'을 운영하여 공공 지원을 확대하고, 민간 클라우드 사업자와의 협력을 통해 AI 인프라 공동 활용을 위한 표준 계약 체계를 마련함으로써 산업 전반의 AI 접근성과 자원 효율성을 제고한다.

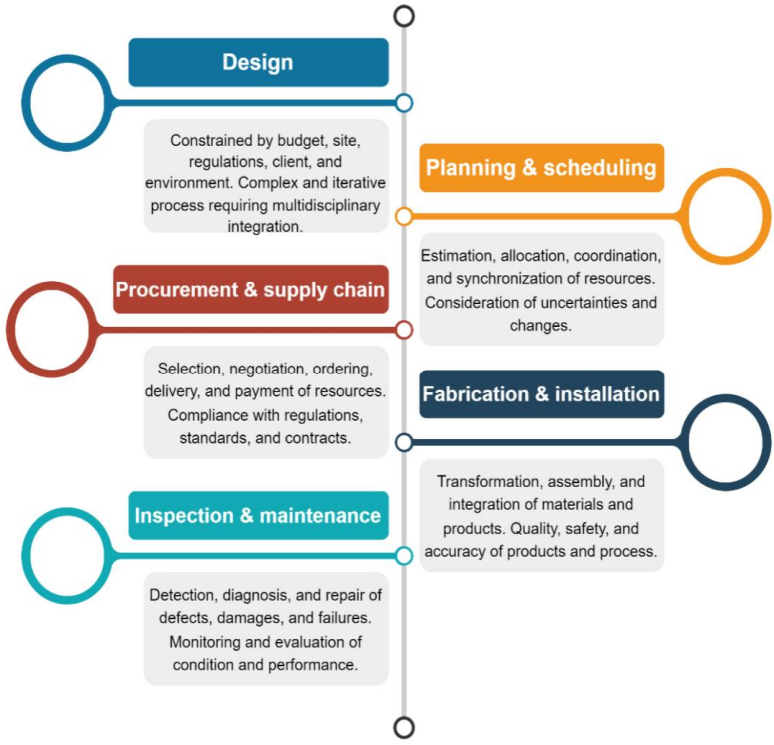
마. 금융 AI 거버넌스 및 윤리

AI 인프라 및 연합 데이터 허브 고도화는 안전한 AI 개발 및 운영 환경을 조성하고 산업 간 데이터 융합을 촉진하는 것을 목표로 한다. 이를 위해 금융보안원은 '금융권 전용 클라우드 샌드박스'를 구축하여, 외부망과의 연결 없이도 AI 모델의 학습과 검증, 배포가 가능한 안전한 폐쇄형 환경을 제공한다. 또한 금융데이터거래소(KDX)를 중심으로 산업별 데이터 허브 통합 체계를 구축함으로써, 금융 데이터와 비금융 데이터의 융합 활동을 촉진하고 새로운 데이터 기반 비즈니스 모델을 창출할 수 있는 기반을 마련한다. 정부는 GPU 및 AI 컴퓨팅 자원 확보를 위해 '금융 AI 컴퓨팅 풀(FinAI Compute Pool)'을 운영하여 공공 지원을 확대하고, 민간 클라우드 사업자와의 협력을 통해 AI 인프라 공동 활용을 위한 표준 계약 체계를 마련함으로써 산업 전반의 AI 접근성과 자원 효율성을 제고한다.

제 4 절 건설/문화콘텐츠

1. 건설

〔그림 7－14〕 건설 분야의 Lifecycle

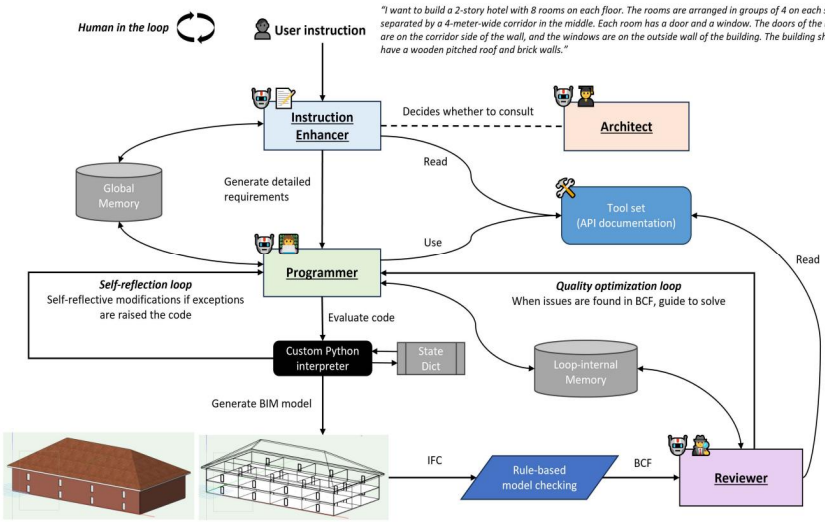


자료: Taiwo et al.(2024)

건설 산업은 설계(Design), 시공(Construction), 조달 및 공급망 관리(Procurement and Supply Chain Management), 제작 및 설치(Fabrication and Installation), 그리고 검사 및 유지보수(Inspection and Maintenance) 등 전 주기(Lifecycle) 단계 전반에 걸쳐있다(Parsamehr et al., 2023). 각 단계는 다양한 자원, 전문 분야, 이해관계자의 조정과 통합을 필요로 하는 복잡한 과정으로 구성되며, 동시에 여러 제약 조건, 불확실성, 그리고 변화 요인을 고려하

고 관리해야 한다(Taiwo et al., 2024). 이러한 산업적 특성은 AGI(Artificial General Intelligence)의 핵심 구성 요소인 추론(Reasoning), 자율성(Autonomy), 학습(Learning), 안전성 및 가치 정렬(Safety & Alignment) 기술을 실제 환경에서 검증할 수 있는 응용 분야로 평가된다. 그러나 최근 연구들은 여전히 단일 과업 중심의 Artificial Narrow Intelligence (ANI)이나 초기 AGI 수준에 머물러 있다.

〔그림 7－15〕 Text2BIM 프레임워크



자료: Taiwo et al.(2024)

설계 단계에서는 언어 기반 추론(Language-to-Action Reasoning) 기술이 발전하고 있다. Du et al.(2024)의 Text2BIM 연구는 여러 LLM 에이전트가 사용자의 자연어 지시를 해석하고 이를 Vectorworks API 명령어로 변환하여, 사람이 직접 편집 가능한 BIM(Building Information Model)을 자동 생성하는 프레임워크를 제안하였다. 이후 룰 기반 모델 검사기(Rule-based model checker)를 통해 생성된 BIM의 품질을 검증함으로써, 언어 이해와 설계 행위가 하나의 연속된 과정으로 작동할 수 있음을 실험적으로 입증하였다.

시공 단계에서는 자율적 의사결정과 적응형 제어 기술이 주목받고 있다. Yao et al.(2024)은 건설 일정 최적화 문제를 마르코프 의사결정 과정(Markov Decision Process,

MDP)으로 모델링하고, 그래프 신경망(GCN)과 딥 강화학습(DRL)을 결합한 알고리즘을 제안하였다. 이 연구는 복잡한 제약 조건을 고려하면서도 효율적인 일정을 자동으로 도출할 수 있음을 보여주었으며, 건설 데이터의 구조적 관계를 학습하는 지능형 스케줄링 모델의 가능성을 제시하였다. 한편 Wang et al.(2023)은 BIM-Driven Human-Robot Collaborative Construction Workflows with Closed-Loop Digital Twins 연구에서 BIM 데이터를 로봇 행동 계획에 통합하고, 현장 피드백을 반영하여 계획을 실시간으로 조정하는 폐쇄 루프(closed-loop) 협업 구조를 구현하였다. 로봇은 센서 피드백을 기반으로 환경 변화를 감지하고 작업 계획을 수정함으로써, 자율성과 적응성을 갖춘 협업형 로봇 시스템의 가능성을 보여주었다.

운영 및 유지관리 단계에서는 학습과 안전을 중심으로 한 연구가 확대되고 있다. Lee et al.(2022)은 디지털 트윈(Digital Twin) 환경에서 딥 강화학습(DRL)을 적용해 작업 분배 정책을 학습하고, 환경 변화에 따라 동적으로 제어 전략을 조정하는 자율 제어 모델을 제시하였다. 이 연구는 시뮬레이션을 통해 경험 데이터를 반복적으로 축적하고 정책을 개선하는 형태의 적응 학습을 구현했다는 점에서 의의가 있다. Al-Azani et al.(2024)은 개인 보호장비(PPE, Personal Protective Equipment) 착용 여부를 인식하기 위해 대규모 데이터셋을 구축하고, Vision Transformer와 YOLOv7 등 다양한 딥러닝 모델의 성능을 비교 및 평가하였다. 그 결과, 제안된 접근법은 건설 현장의 작업자 안전 준수 상태를 실시간으로 감지할 수 있음을 입증하였으며, 시각적 감시를 통한 안전 규범 모니터링의 가능성을 제시하였다.

가. 중기(5년후), 장기(10년후) 건설 산업 분야 AGI 전망

건설 산업은 설계, 시공, 운영 전 과정에서 발주자, 설계자, 엔지니어, 시공사, 하도급업체 등 다양한 이해관계자가 참여하며, 복잡한 의사결정 구조와 물리적 제약, 안전 및 규제 요건이 얽혀 있는 산업이다(Taiwo et al., 2024). 이러한 특성은 인공지능(AI)이 단순한 자동화 수준을 넘어, 고차원적 추론, 지각, 자율성 및 에이전트, 학습, 안전성 및 가치 정렬을 결합한 통합형 지능 구조로 발전해야 함을 시사한다. 최근 연구들은 건설 산업의 AI 적용이 여전히 특정 과업에 한정된 ANI 단계나 초기 AGI에 머물러 있으며(Du et al., 2024), 향후 10년 내에는 이러한 개별 지능 요소들이 상호 연계되고 통합되는 방향으로 진화할 것으로 전망된다.

1) 중기 AGI 전망

향후 5년은 AGI를 구성하는 핵심 기술들이 개별 영역에서 일정 수준의 발전을 이룬 이후, 산업 현장에서 통합적으로 실증되는 과도기적 단계(Transitional Phase)가 될 것으로 보인다. 이 시기의 연구개발은 언어-행동 추론(Language-to-Action Reasoning), 자율성 및 에이전트, 학습 능력, 안전성 및 가치 정렬을 중심으로 고도화될 것으로 보인다.

언어-행동 추론(Language-to-Action Reasoning)영역에서는 LLM이 인간의 언어적 지시를 설계나 시공 절차로 변환하고, 이를 구조화된 지식 체계와 연계하는 방향으로 발전할 것으로 예상된다. 예를 들어 “지하 2층 방수공정의 자재 규격을 변경하라”는 명령을 이해하고, BIM(Building Information Modeling) 및 디지털 트윈 시스템을 통해 시공 계획, 인력 배치, 규정 검증까지 자동으로 수행하는 수준으로 확장될 것이다. 단순한 명령 수행을 넘어, 규제 조건이나 물리적 제약을 고려한 인과적 추론이 가능해지면서 설계 변경의 결과를 미리 시뮬레이션하고 대안을 제시하는 능력을 갖출 것으로 보인다.

자율성 및 에이전트 영역에서는 인공지능이 단일 공정 자동화를 넘어, 여러 목표를 동시에 이해하고 환경 변화에 따라 계획을 스스로 조정하는 반자율형 구조로 발전할 것이다. 예기치 못한 기상 변화나 자재 공급 차질, 장비 고장 등 다양한 현장 변수를 실시간으로 인식하고 일정 조정이나 자원 재배분을 수행할 수 있는 수준이 될 것으로 예상된다. 이와 같은 시스템은 클라우드 기반 건설 관리 플랫폼과 연계되어, 인간 관리자와 협력적인 의사결정을 수행하는 방향으로 진화할 것이다.

학습 능력 영역에서는 단발적인 재학습을 넘어 지속적인 피드백을 통해 경험적으로 자기 갱신이 가능한 형태로 발전할 전망이다. 디지털 트윈 기반의 시뮬레이션, 센서 데이터, 드론 영상, 품질 검사 로그 등이 통합되면서, 인공지능은 실제 환경의 데이터를 바탕으로 스스로 학습 효율을 향상시킬 수 있다. 이러한 구조는 반복되는 시공 작업의 품질을 개선하고, 예측 가능한 유지보수를 지원하는 데 기여하게 된다.

안정성 및 가치 정렬영역에서는 산업 규범과 안전 기준을 인공지능이 스스로 해석하고 반영하는 학습 방식이 중요해질 것이다. AI가 규정의 문맥과 의도를 이해하여 학습이 실현되면, 안전사고 예방이나 품질 점검을 인간 감독에 의존하지 않고 자동으로 수행할 수 있게 된다. 이 시기의 AI는 인간의 감독을 여전히 필요로 하지만, 제한된 범위 내에서 목표 유지, 계획 조정, 규범 반영이 가능한 수준의 인지 일관성과 자율성을 확보할 것으로 예상

된다. 이러한 단계는 AGI로의 진입을 위한 실험적 기반이 될 것이다.

2) 장기 AGI 전망

10년 후의 건설 산업은 AGI의 완전한 통합 단계에 진입하며, 추론과 문제 해결, 자율성, 학습 능력, 가치 정렬이 고도화된 형태로 작동할 것으로 전망된다. AGI는 단순히 365일 작동하는 자동화 시스템을 넘어(AI 2027 Project, 2025), 설계적 사고, 공정 관리, 환경적 판단을 종합적으로 이해하고 스스로 최적의 의사결정을 수행하는 지능형 협력자(Intelligent Collaborator)로 진화할 것으로 전망된다.

추론 및 문제 해결 영역에서 AGI는 초인지(Metacognitive) 기반의 자기 검증 구조를 통해 불확실성을 평가하고, 판단 근거를 명시적으로 해석하며, 결과를 사후적으로 점검할 수 있는 능력을 갖추게 된다. 건설 분야에서는 이러한 기능이 구조적 안정성, 공간 효율성, 에너지 절감을 동시에 고려하는 다목적 최적화(Multi-objective Optimization) 설계로 확장될 것이다. 예를 들어, 건물의 일사량, 통풍, 하중 분포 데이터를 실시간 분석해 설계안을 자동으로 조정하거나, 구조물의 균열 가능성을 사전에 예측하고 보수 계획을 자동 생성하는 시스템이 구현될 것으로 예상된다. 이는 고위험 건설 프로젝트에서 신뢰성과 투명성을 확보하기 위한 핵심 요건이자, 인간-기계 협력적 설계의 토대를 형성할 것이다.

자율성 및 에이전트 영역에서는 단기적 작업 수행을 넘어 장기 목표(Long-term Goal)를 스스로 설정하고, 환경 변화에 따라 계획을 수정하며, 상호 협력적 의사결정을 수행하는 방향으로 발전할 것이다. 건축과 건설 전반에서 AGI는 설계, 시공, 운영, 유지관리, 건축물 해체 등 전 생애주기(Lifecycle) 하나의 지능형 체계로 통합 관리하는 역할을 수행하게 된다. 예를 들어, 건물의 실시간 사용 데이터를 바탕으로 공간 재배치나 에너지 효율 개선을 스스로 제안하고, 시공 단계에서는 로봇 시공 시스템과 협력하여 계획을 즉시 수정·반영할 수 있다. 또한, 도시 단위의 스마트시티 설계에서는 다수의 AGI 시스템이 교통, 환경, 주거 데이터를 통합 분석하여, 장기적인 도시 계획(Urban Planning)의 시뮬레이션과 검증을 수행하게 될 것이다.

다. 중기(5년후), 장기(10년후) 건설 산업 AGI 대응 전략

AGI의 발전은 건설 산업의 생산성, 안전성, 그리고 지속가능성을 근본적으로 변화시킬 잠재력을 지니고 있다. 그러나 이러한 전환이 실현되기 위해서는 기술 개발뿐 아니라 데

이터, 제도, 인력, 표준 등 산업 전반의 구조적 대응이 병행되어야 한다. 따라서 건설 산업의 AGI 경쟁력 확보는 기술 중심의 혁신을 넘어, 산업 생태계 전체를 지능화하는 방향으로 추진되어야 한다.

1) 중기 AGI 대응 전략

향후 5년은 AGI 기술의 개별 요소를 산업 현장에 통합하고 실증하는 단계가 될 것이다. 이 시기의 목표는 기술의 완성보다 산업 적응력과 현장 실효성 확보에 있다. 정부와 산업계는 BIM, 디지털 트윈, IoT, 대규모 언어모델 등 다양한 기술을 연계한 AGI 실증 플랫폼을 구축하고, 공공 인프라 프로젝트를 중심으로 시범 사업을 추진해야 한다. 이를 통해 데이터 기반의 설계 자동화, 공정 최적화, 안전 관리 등 실제 업무 환경에서 AGI의 활용성을 검증할 수 있다. 또한, 데이터의 표준화와 개방을 추진하여 AGI 학습 생태계를 조성해야 한다. 이를 위해 BIM, 시공 로그, 센서 데이터 등 이기종 데이터를 통합 관리하는 건설 데이터 허브를 구축하고, 국제 표준(IFC, ISO 등)에 부합하는 데이터 거버넌스 체계를 확립해야 한다. 아울러, 건설 엔지니어/관리자 대상의 AGI 활용 교육을 통해 인적 역량을 강화해야 한다. 기술적 확산과 함께 법/제도적 기반 정비도 필수적이다. AGI의 판단 결과에 대한 책임 기준, 안전성 평가 절차, 윤리적 검증 체계 등이 선제적으로 마련되어야 한다. 결국 중기 전략의 핵심은 “AGI 기술의 산업 통합과 실증”이며, 이는 향후 자율적 산업 지능 생태계로 발전하기 위한 기초 단계가 될 것이다

2) 장기 AGI 대응 전략

10년 후 건설 산업의 AGI는 인간의 감독을 넘어, 스스로 판단, 학습, 및 의사결정을 수행하는 자율 지능형 산업 구조로 진화할 것이다. 이 시기에는 기술 고도화보다 지속 가능한 운영 체계와 신뢰 기반 거버넌스 구축이 중심이 되어야 한다. AGI의 자율적 판단과 의사결정을 제어하기 위해, 판단 근거의 투명성·안전성·책임성을 확보하는 제도적 관리 체계를 마련해야 한다. 산업 전반의 데이터를 통합 관리하고, 지속적인 학습·검증이 가능한 데이터 순환형 생태계를 구축해 AGI의 신뢰성과 적응력을 강화할 필요가 있다. 또한, AGI와 협업할 수 있는 전문가와 관리자를 양성하여, 인간이 단순 감독자가 아닌 지능 협력자로서 역할을 수행하도록 인력 구조를 재편해야 한다. 더불어, AGI의 안전성, 데이터 구조, 책임 기준을 국제 표준으로 발전시켜 한국형 AGI 모델의 글로벌 확산을 추진할 필요가 있다.

결국 장기 대응 전략의 핵심은 기술 통합에서 산업 운영과 제도의 통합으로의 확장이다. AGI가 산업과 사회의 신뢰 가능한 파트너로 작동하기 위해서는, 기술적 진보와 함께 윤리/법/교육이 조화된 지속 가능한 생태계 조성이 필수적이다.

2. 문화 콘텐츠

가. 문화 콘텐츠 산업 AGI 기술 활용 현황

현재 문화 콘텐츠 산업에서는 영상, 음악, 게임, 출판, 마케팅 등 다양한 영역에서 AI의 활용이 두드러진다(State of AI Report, 2025). 예를 들어, 영상 산업에서는 OpenAI의 Sora는 텍스트 명령으로부터 최대 1분 길이의 물리적으로 일관된 영상을 생성하며, 이는 영화·애니메이션 분야의 프리비즈(previsualization)나 콘셉트 시안 제작 과정에서 실험적으로 활용되고 있다. Google의 DreamFusion(Poole et al., 2022)은 텍스트 기반의 3D 자산을 자동 생성하여, 기존의 모델링-렌더링 파이프라인을 효율화하는 역할을 하고 있다. 이들은 비록 자율적 의사결정이나 지속학습 능력을 갖추지는 못했지만, 초기 단계의 추론과 멀티모달리티 능력을 구현한 사례로 평가된다.

〔그림 7－16〕 사회적 시뮬레이션 환경에서의 생성형 에이전트

자료: Park et al.(2023)

게임 산업에서는 LLM을 기반으로 한 지능형 NPC(Non-Player Character) 연구가 활발하다. 대표적으로 Stanford의 Generative Agents(Park et al., 2023)는 캐릭터가 대화를 통해 사회적 관계를 형성하고, 기억을 기반으로 행동을 계획하는 프레임워크를 제시하였다. 이러한 기술은 완전한 AGI는 아니지만, 제한적 자율성 과 상황 기반 추론을 실험적으로 구현한 사례로 볼 수 있다. 이는 기존의 스크립트 기반 게임 설계에서 벗어나, 플레이어와의 상호작용을 통해 내러티브가 유동적으로 전개되는 새로운 창작 패러다임을 제시하고 있다.

음악과 오디오 산업에서도 생성형 모델이 멀티모달리티영역에서 중요한 진전을 보이고 있다. Meta의 MusicGen(Copet et al., 2023)은 텍스트나 멜로디 입력으로 음악을 생성하며, 이는 AGI 구성요소 중 멀티모달 이해와 표현 생성 능력을 제한적으로 보여준다. 또한 Meta의 SeamlessM4T(Barrault et al., 2023)는 end-to-end 방식으로 처리하여 감정 보존형 음성 합성 및 자동 더빙에 활용되어 글로벌 현지화(Localization) 과정의 효율을 높이고 있다.

출판과 미디어, 광고 분야에서는 언어모델을 기반으로 한 LLM-강화 추천 시스템(LLMERS)이 확산되고 있다(Liu et al., 2024). LLMERS는 단순한 시청 패턴 분석을 넘어, 이용자의 심리적 상태나 맥락을 추론해 콘텐츠를 추천하는 방식으로 발전하고 있다. 이는 데이터 기반 개인화를 넘어서 언어적 추론과 맥락 이해를 결합한 초기 AGI적 접근으로 볼 수 있다.

이처럼 문화 콘텐츠 산업에서의 인공지능 활용은 완전한 AGI 단계에 도달하지 않았으나, LLM 기술이 발전함에 따라 AGI의 일부 핵심 역량인 언어적 추론, 멀티모달 이해, 제한적 자율성 등이 개별 영역에서 실험적으로 구현되고 있다. 이러한 기술적 진전은 단순한 생산 자동화를 넘어, 창작자와 인공지능이 상호 보완적으로 아이디어를 탐색하고 결과물을 완성하는 새로운 협업 구조를 형성할 수 있을것으로 예상된다(State of AI Report, 2025; AI 2027 Project, 2025).

나. 중기(5년후), 장기(10년후) 문화 콘텐츠 산업 AGI 전망

문화 콘텐츠 산업은 언어, 시각, 음향, 감정 등 다양한 모달리티를 다루며 인간의 감수성과 창의적 판단이 핵심인 산업이다. 이러한 특성은 AI가 단순 생성 단계를 넘어, 추론, 자율성, 멀티모달리티, 학습, 가치 정렬이 결합된 통합형 지능 구조로 발전해야 함을 시사한다. 향후 10년은 AI가 단순 보조 도구를 넘어, 인간과 협력하는 창의적 파트너로 진화하는 시기가 될 것으로 전망된다.

1) 중기 AGI 전망

향후 5년은 인공지능이 단순한 창작 보조 도구를 넘어, 인간과 함께 아이디어를 구상하고 결과물을 만들어내는 협업적 창작 단계로 발전하는 시기가 될 것이다(Yu et al., 2025, Kadenhe et al., 2025). 이 시기의 기술 발전은 크게 네 가지 방향에서 뚜렷한 변화를 보일 것으로 예상된다. 현재 문화 콘텐츠 산업에서 AI는 대본 작성, 영상 편집, 음원 제작 등 특정 과업을 자동화하는 수준에 머물러 있지만, 가까운 미래에는 이들이 하나의 통합된 지능 구조로 작동하며 지능형 협업 환경(Intelligent Co-Creation Environment) 이 본격적으로 자리 잡을 것으로 전망된다. 이 시기의 기술 발전은 주로 네 가지 축에서 전개될 것으로 보인다.

첫째, 멀티모달 추론(Multimodal Reasoning) 기술의 진전으로 텍스트, 영상, 음향, 3D 데이터가 유기적으로 연결되는 통합형 창작이 가능해질 것이다. AI는 시나리오의 감정선이나 스토리 전개를 분석해 카메라 구도, 조명, 배경음악을 자동 설계할 수 있으며, 음악이나 게임에서는 사용자의 감정 상태를 반영한 적응형 생성이 실현될 것으로 예상된다.

둘째, 에이전트형 창작 시스템(Agent-based Creation System) 의 확산으로, 각기 다른 역할을 수행하는 다중 AI가 하나의 프로젝트 안에서 협력하게 된다. 예를 들어, 대본을 담당하는 에이전트와 콘셉트 아트를 설계하는 에이전트, 그리고 홍보 전략을 기획하는 에이전트가 실시간으로 상호작용하며 하나의 작품을 만들어내는 구조가 가능해진다. 특히 게임 산업에서는 이러한 시스템이 대화형 스토리텔링이나 지능형 NPC(Non-Player Character) 의 형태로 구현되며, 플레이어와 AI가 함께 서사를 만들어가는 새로운 창작 경험을 제공할 것이다.

셋째, 적응형 학습(Adaptive Learning) 능력이 강화되면서 AI는 관객의 반응과 플랫폼 데이터를 분석해 콘텐츠를 실시간으로 보완하거나 다음 회차의 방향을 스스로 조정하는 수준으로 발전할 것이다. 이러한 능동적 피드백 구조는 AI가 인간의 평가를 학습하고, 그 결과를 창작 과정에 반영하는 초기형 지능 순환 구조(quasi-intelligent feedback loop) 로 이어질 것으로 보인다.

넷째, 안전성 및 가치 정렬의 중요성이 커질 것이다. AGI로 진화하는 과정에서 AI가 생성하는 콘텐츠가 사회적/윤리적 기준을 벗어나지 않도록 하기 위해, 모델 수준의 가치 정렬 메커니즘과 생성 제어 기술(safety-constrained generation) 이 함께 발전할 것으로 예상

된다. 이와 더불어, 출처 인증(C2PA), 디지털 워터마킹, 블록체인 기반 진위 검증 기술이 상용화되어, 생성 콘텐츠의 안전성을 보장하는 체계가 구축될 전망이다.

2) 장기 AGI 전망

2035년 전후의 문화 콘텐츠 산업은 협업 중심의 지능형 창작 환경(Intelligent Co-Creation Environment)을 넘어, AGI가 인간 수준의 자율적 창의 주체(Autonomous Creative Entity)로 진화할 것으로 전망된다(Wu et al., 2025). 이 시기의 AGI는 단순히 인간의 지시를 따르는 보조 도구가 아니라, 기획자, 감독, 작가 등의 역할을 일정 부분 수행하며, 창작의 기획부터 결과 검증까지 일부 과정을 스스로 관리할 수 있는 수준에 도달할 것으로 예상된다.

첫째, 지속학습(Continual Learning)이 고도화되면서 AGI는 사회적 트렌드, 대중의 정서, 문화적 변화를 학습해 창작 방향을 조정할 수 있게 된다. 콘텐츠는 고정된 결과물이 아닌, 관객의 반응과 사회적 흐름에 따라 개선되는 창작 구조로 발전할 가능성이 높다.

둘째, 통합형 생성 구조(Unified Generative Architecture)가 완성되며, 언어/시각/음향/감정 표현이 단일 지능 프레임 내에서 처리된다. AGI는 시나리오, 연출, 음악, 시각적 요소를 유기적으로 결합해 일관된 예술적 세계를 구현하고, 관객 데이터를 분석해 장면, 톤, 스토리 등을 스스로 보완하는 Self-Evolving Agent(Gao et al., 2025)의 초기 단계로 작동할 것으로 예상된다.

셋째, 문화적 감수성과 사회적 책임성이 AGI 창작의 핵심 기준으로 자리 잡을 것이다. AGI는 단순히 미적 완성도를 추구하는 것을 넘어, 작품이 사회적 맥락 속에서 어떤 의미로 해석되는지를 이해하고 조정할 수 있게 된다. 예를 들어, 특정 문화나 세대, 성별에 대한 편향적 서사를 줄이고, 다양한 배경의 인물과 정서를 자연스럽게 반영하는 균형 잡힌 창작 방식이 확산될 것이다. 이러한 변화는 AI가 인간의 감수성을 대체하는 것이 아니라, 사회적으로 공감 가능한 예술적 표현을 확장하는 방향 으로 발전함을 의미한다.

다. 중기(5년후), 장기(10년후) 문화 콘텐츠 산업 AGI 대응 전략

AGI의 진화는 문화 콘텐츠 산업의 창작 방식을 근본적으로 재편할 것으로 예상되어 진다. 그러나 이러한 기술적 전환이 산업 경쟁력 강화로 이어지기 위해서는 단순한 기술 도입을 넘어, 데이터, 제도, 인력, 및 생태계 전반에 걸친 구조적 대응이 병행되어야 한다. 특

히 문화 콘텐츠 산업은 감정과 창의가 중심인 산업이기 때문에, 기술만으로는 창작의 본질적 가치를 구현하기 어렵다. 따라서 AI와 인간이 협업하는 구조를 마련하는 것이 중요하다. 향후 10년은 이러한 방향성을 바탕으로, AGI 기술 발전 단계를 고려한 중기와 장기 대응 전략을 체계적으로 추진해야 한다.

1) 중기 AGI 대응 전략

향후 5년은 AGI를 문화 콘텐츠 제작 현장에 실제로 적용하며 산업 적응력을 높이는 시기이다. 우선, 창작자와 AI가 함께 작업하는 구조가 확산되기 위해서는 공정한 협업 생태계가 뒷받침되어야 한다. AI 학습에 사용되는 창작물의 원저작자에게 합리적인 보상이 돌아가는 보상 체계를 구축해야 하며, 창작 과정에서 발생하는 아이디어나 결과물의 소유권을 명확히 규정하는 AI-창작자 공동 저작 원칙을 제도화해야 한다.

또한, AI가 창작을 위해 학습할 수 있는 데이터의 질과 다양성을 확보하기 위해 멀티모달 공공데이터 허브를 조성할 필요가 있다. 저작권이 정리된 텍스트, 음원, 영상, 3D 모델 데이터를 체계적으로 수집하고 공개함으로써, 국내 기업이 합법적으로 활용할 수 있도록 지원해야 한다.

산업 경쟁력의 핵심은 결국 사람에 있다. 예술과 기술의 경계를 넘나드는 AI-크리에이티브 전문 인력을 양성하기 위해 대학과 산업계가 협력하는 융합 교육 프로그램을 마련하고, 현장 중심의 실습형 연구소(AI Creative Lab)를 통해 창작자들이 직접 AI와 협업하는 경험을 축적할 수 있도록 해야 한다.

마지막으로, 생성형 AI의 확산에 따라 사회적·윤리적 논란이 커질 가능성에 대비해 AI 창작 윤리 기준을 확립해야 한다. AI가 생산한 콘텐츠의 진위, 출처, 편향성 문제를 점검하기 위한 감수성 평가 시스템을 구축하고, 정부와 민간이 함께 운영하는 윤리 검증 위원회를 통해 책임 있는 창작 환경을 조성할 필요가 있다.

2) 장기 AGI 대응 전략

2030년대 중반 이후 문화 콘텐츠 산업은 인간과 AI가 협업하는 단계를 넘어, AGI가 독립적인 창의 주체로 활동하는 시대로 전환될 것으로 전망된다. 이 시기의 핵심은 기술적 효율이 아니라 인간과 AGI가 공존하며 창의적 가치를 공동으로 생산하는 지능형 창작 생태계의 완성에 있다.

AGI는 인간 창작자의 역할은 단순한 지시자(director)가 아닌, 창의적 큐레이터(creative curator)로 변화하게 된다. 인간은 AGI가 제안하는 수많은 창작 가능성 중에서 의미와 방향을 선택하고, 사회적 감수성이나 미학적 판단을 통해 결과물을 조정하는 역할을 담당하게 된다. 이는 창작 주체의 개념을 인간에서 인간-AGI로 확장시키며, 창작 행위의 본질을 '지능 간 협업'으로 재정의하게 될 것이다. 이와 함께, AGI 윤리 및 가치 정렬의 중요성은 더욱 커질 전망이다. AGI가 스스로 서사를 구성하고 메시지를 선택하는 단계에 이르면, 사회적 편향이나 문화적 왜곡이 작품 속에 반영될 가능성이 높아진다. 따라서 세대/성별/문화별 감수성을 반영할 수 있는 Cultural Alignment Model(Alkhamissi et al., 2024)과, 창작물의 사회적 영향을 검증할 수 있는 평가 시스템을 정립해야 한다. 이를 통해 AGI가 단순히 효율적인 생산자가 아닌, 윤리적·사회적 판단이 가능한 창의적 주체로 기능하도록 관리해야 한다.

장기적으로는 이러한 기술적·제도적 발전을 기반으로, 한국형 AGI 창작 생태계가 글로벌 표준으로 자리매김할 가능성이 있다. AGI 기반의 콘텐츠 제작 인프라, 저작권·데이터 활용 규범, 협업형 창작 모델 등이 체계화되면, 한국은 '인간 중심의 창의 AI' 라는 새로운 문화 기술 브랜드를 구축할 수 있다. 결국 AGI 시대의 문화 콘텐츠 산업은 단순한 생산 효율의 경쟁을 넘어, 감성/가치/창의성을 결합한 고차원적 문화 지능 산업으로 진화하게 될 것이다.

제 5 절 교육/법률

1. 교육분야 AGI기술 활용 현황

Feng et al.(2025)이 제시한 초기 AGI 단계에 해당하는 ChatGPT의 등장 이후, LLM 및 에이전틱(Agentic) AI 기술이 교육 현장 전반에 걸쳐 교수·학습 지원, 행정 효율화, 맞춤형 학습환경 구현 등을 중심으로 본격적으로 활용되면서, 교육 분야의 AI 시장은 가속적인 성장세를 보이고 있다. Grand View Research(2024)의 시장 분석에 따르면, 2024년 기준 전 세계 교육 분야 AI 시장 규모는 약 58.8억 달러(약 7조 9,400억 원)로 추정되고, 또한 2025년부터 2030년까지의 연평균 복합성장률(CAGR)은 약 31.2%에 이를 것으로 예상, 2030년에는 약 322.7억 달러(약 43조 6,600억 원) 규모로 성장할 것으로 전망된다.

〈표 7－9〉 교육 분야 대표 AI 기업 및 제공 서비스

기업	기업소개 요약	대표서비스
튜터러스랩스 (국내) http://tutoruslabs.io/	LLM을 기반으로 한 대화형 AI 튜터링 솔루션을 개발하는 에듀테크 기업. 디지털 교과서와 온라인 학습 환경을 연계해 학생 맞춤형 학습 대화 지원.	• learnCoach: AI 기반 튜터링 솔루션 • learnClip: 영상강의 기반 대화 솔루션 • learnSpeak, 한영 발음평가 및 어학학습 솔루션
클래스팅(국내) https://www.classting.com/	학생의 학습 데이터를 기반으로 맞춤형 문제 출제·채점·피드백을 자동화. 교사 업무를 지원하는 AI 기반 학습·관리 통합 서비스를 제공	• AI 기반 학습관리 및 맞춤형 교육 플랫폼: 학생의 학습 수준 진단 → 맞춤 학습 경로 추천 → 자동 문제 출제·채점 → 학습 리포트 제공 등을 통해 개인화된 맞춤 학습
맥스 AI(국내) https://maxai.co.kr/	1: 1 영어 학습을 중심으로, 원어민 튜터 및 AI 코칭 시스템을 결합하여 학생들의 영어 스피킹 능력 향상을 지원하는 국내 에듀테크 서비스	• AI 튜터와 원어민 영상 기반의 1:1 영어 스피킹 학습 플랫폼: 사용자의 영어 수준과 학습 상황에 맞춰 실시간 발음·문법 피드백을 제공하며 프리토킹 및 맞춤 커리큘럼을 운영

기업	기업소개 요약	대표서비스
매직스쿨(해외) https://www.magicschool.ai/	교사용 생성형 AI 플랫폼으로, 50여 가지 AI 도구를 통해 수업 계획·평가문항·학습자료를 자동 생성하고 LMS와 연동해 수업 준비와 평가를 효율화. 최근 학생용 지원 도구도 출시	• 교사용 주요 도구: 표준에 맞춘 수업계획 생성, 평가문항 자동 작성, 개별화 학습자료 자동 추천 등 교육현장의 반복 업무를 AI로 경감
Revision.ai(해외) https://www.revision.ai/	학생들이 강의 노트·PDF·슬라이드를 업로드하면 자동으로 플래시카드, 퀴즈 등 AI 생성 학습도구로 바꿔주는 학습 플랫폼	• AI 퀴즈/플래시카드 자동생성: PDF·PPT·노트 파일을 업로드하면 자동으로 플래시카드 및 퀴즈 형태의 학습도구로 변환

교육분야 AI는 ChatGPT이전에도 오래전부터 관심을 받아온 영역이며, 교사와 튜터를 대상으로 AI 기반 교육 서비스를 제공하는 국내외 스타트업 및 기업들은 〈표 7−9〉와 같다.

한편, 유네스코 한국위원회(2021) 보고서에 따르면, 교육 분야에서의 AI 활용은 크게 ① **학습·평가 영역**과 ② **교사 역량 강화 및 교수법 향상** 영역으로 구분된다. 먼저, 학습과 평가 영역에서는 AI가 1) 지능형 튜터링 시스템(ITS), 2) 대화 기반 학습시스템(DBTS), 3) 자동 작문 평가(AWE), 4) 탐구 학습시스템(ELE) 등의 형태로 적용되고 있다. ITS와 DBTS는 학습자 중심의 상호작용과 자기주도 학습을 촉진하며, AWE와 ELE는 평가의 자동화 및 탐구 기반 사고력 향상을 지원한다. 다음으로, 교사 역량 강화 및 교수법 향상 영역에서의 AI 활용은 교사의 역할을 대체하기보다는 증강(augmentation)의 관점에서 교사의 역량을 확장하고 교수법 혁신을 지원하는 것을 강조하면서, 대표적으로 1) 출석·채점·과제 관리 등을 자동화하는 AI 학습 조교(AI Teaching Assistant), 2) 학습자 상호작용을 분석하여 교사에게 학습진단 정보를 제공하는 AI 기반 토론 포럼 모니터링, 3) 원격 지역에서도 교사-학생 간 지도를 지원하는 AI-인간 듀얼 교사모델(Dual Teacher Model) 등을 언급하고 있다.

LLM Agents기반 교육 응용 연구 현황으로, , Chu et al.(2025)는 교사와 학습자를 지원하기 위한 교수·학습 지원 에이전트(Pedagogical Agents)를 (1) 교사 지원 에이전트(Teaching Assistance Agents)와 (2) 학습자 지원 에이전트(Student Support Agents)로 구분하여 개관하였으며, 그 주요 내용은 다음 표에 정리하였다.

〈표 7－10〉 교수・학습 지원 에이전트의 구성 요소

Agents 유형	요소범위	내용
교사 지원 Agents (Teaching Assistance)	교실 시뮬레이션 (Classroom Simulation)	• 교사-학습자 간 대화, 협력 활동, 문제 해결 과정을 자동으로 모델링. • 교실 시뮬레이션은 실제 수업 이전에 교수방법을 테스트하고 학생 반응을 예측하며, 교사의 업무 부담을 경감
	피드백 코멘트 생성 (Feedback Comment Generation)	• 과제, 퀴즈, 프로젝트 등에 대한 자동 평가 피드백을 생성 • 향후 과제의 난이도가 높아질수록 외부 도구 연동(tool-use) 및 메모리 모듈 강화 필요
	학습자료 추천 (Learning Resource Recommendation)	• 학습자에게 적합한 콘텐츠를 추천하기 위해 기존 자료 검색(retrieval-based) 및 맞춤형 생성(generation-based) 방식을 활용 • 향후 실시간 학습자 성과 반영한 동적 추천, 멀티모달 콘텐츠 포함, 생성+검색 혼합 방식 등
학습자 지원 Agents (Student Support)	적응형 학습 (Adaptive Learning)	• LLM 에이전트는 학습자 프로파일을 구성하고 과거 상호작용·성향·감정상태 등을 반영하여 학습경로를 자동 설계하고 수정 • 정서모델(affective state)·성격(traits)을 도입해 몰입도 및 학습효율을 높이는 방향이 제시
	지식 추적 (Knowledge Tracing)	• 학습자의 지식 습득·변화 과정을 추적해 미래 성과를 예측하고 맞춤 설명 제공 • 기존방법 대비 LLM 에이전트는 자연어 상호작용을 통해 개념마스터리(conceptual mastery)를 보다 동적이고 세밀하게 파악 가능
	오류검출 및 수정 (Error Correction & Detection)	• 학습자의 응답에서 오류 또는 오개념을 실시간으로 탐지하고, 적절한 피드백 및 수정 경로 제안. 최근 손필기 또는 디지털 초안을 자연어로 전환 후 분석하는 멀티모달 접근 도입 • 향후 복잡한 도메인(프로그래밍, 수학 논리)에서의 신뢰성 확보 및 멀티모달 오류 탐지 강화

자료: Chu et al.(2025). 교육분야에서의 LLM Agents - Pedagogical Agents

추가로, Chu et al.(2025)에서는 LLM 기반 **분야 특화 교육 에이전트(Domain-specific Educational Agents)**를 과학 학습(Science learning), 언어 학습(language learning), 전문직

업영역(professional development)으로 구분하여 제시하였다. 과학 분야에서는 수학·물리·화학·생물 등에서 문제해결·시뮬레이션·실험 지원·과학적 발견이 가능한 에이전트들을, 언어 분야에서는 읽기·쓰기·말하기·번역 등 기능별로 문법 교정·난이도 조정·대화 시뮬레이션·문화적 맥락 번역을 지원하는 시스템들을 소개하였으며, 또한 전문직업 교육에서는 의료·법률·컴퓨터과학 등에서 진단·피드백·모의훈련을 제공하는 에이전트를 언급하고 있따. 전반적으로 LLM기반 교육 에이전트의 경우 학습의 실증성과 몰입도를 높이며, 교수·학습의 효율성과 개인화를 강화하는 방향으로 발전하고 있는 추세이다.

2. 교육분야 AGI기술 발전 전망 및 경쟁력 확보 대응방안

가. 교육분야 AGI기술 활용 중기(5년후), 장기(10년후) 전망

현재 교육환경은 학교 중심의 정규교육과 비정규교육(사교육 및 온라인교육 등)으로 구분된다. 정규교육의 경우 교사 1인과 다수의 학생 간 상호작용을 기반으로 운영되고 있으나, 개별 학생의 수준과 학습 속도에 맞춘 맞춤형 교육이 체계적으로 구현되기 어려운 한계가 존재한다. 그러나 AGI 기술의 발전은 이러한 제약을 해소하고, 학생 개별 수준에 따른 맞춤형 교육을 전 학제와 교육과정에 걸쳐 실현할 수 있는 기반을 제공할 것으로 전망된다.

우선, 단기적으로 분아별 **교육특화 파운데이션 모델이 구축**될 것이다. 해당 모델은 초·중·고 및 대학 교육과정을 포괄하는 구조로, 파운데이션 모델 자체가 "교육학적(educationally informed)" 특성을 내재하여 **교과목 텍스트북 제작, 강의자료 개발, 학생 수준 트래킹, 강의계획서 작성, 시험·과제 출제 및 채점 등 교육 전주기 과업을 자동화·고도화할 수 있는 기능을 수행**하게 될 것이다. 이러한 교육특화 모델은 과목 및 수준별 교과내용을 정교하게 학습하고, 교사의 강의 및 1:1 튜터링 데이터를 대규모로 정렬·학습(Alignment)함으로써, 1:1 맞춤형 튜터링 서비스가 파운데이션 모델 기반에서 구현되는 환경이 조성될 것으로 보인다.

AGI의 범용성이 증가함에 따라, 이러한 교육 특화 파운데이션 모델은 초기에는 일부 과목 또는 특정 과업에 특화된 형태로 개발될 수 있겠지만, **점차 적용 범위가 확장되면서 교육 전 영역을 포괄하는 통합형 모델로 발전**할 것이다. 이에 따라 **기존의 개별 LLM 에이전트들은 점차 파운데이션 모델 기반의 통합 에이전트 서비스 형태로 재편**되며, 교육 생태

계 전반의 지능화와 자동화 수준을 한 단계 끌어올리는 핵심 플랫폼으로 자리 잡을 것으로 전망된다.

이와 함께, 교사의 업무를 지원함과 동시에 학생 맞춤형 AGI 교사 서비스가 파운데이션 모델을 기반으로 본격화될 전망이다. AGI는 학생과의 대화형 상호작용을 통해 수준·이해도·학습 속도를 실시간으로 파악하고, 그 결과에 따라 맞춤형 교육자료 및 강의 콘텐츠를 생성한다. 또한 기존 강의 동영상을 자동 편집하거나, 새로운 교육영상을 생성형 AI로 제작함으로써, 학생은 자신의 수준에 최적화된 속도와 난이도의 학습 콘텐츠를 제공받게 된다. 나아가 시험 문제 출제·평가·피드백이 모두 자동화되며, 교사 부재 환경에서도 고품질의 1:1 튜터링이 가능한 구독형 교육서비스 모델이 등장할 것으로 예측된다.

정규수업 보조를 위한 AGI 학습보조 서비스 또한 다양하게 확산될 것이다. 예를 들어, 강의 내용을 자동 요약·추적하고, 부족한 부분을 실시간으로 설명하거나, 보충자료를 추천함으로써 학생이 정규교육을 보다 심화적으로 이해할 수 있도록 지원하는 기능이 발전할 것으로 전망된다.

특히 언어교육 분야는 맞춤형 학습의 효과가 가장 두드러지게 나타날 영역이다. 문법이나 독해 중심의 초기 교육은 앞서 언급한 교육특화 AGI 모델을 활용하여 자동화될 수 있으며, 말하기·듣기 중심의 학습에서는 학습자의 발화 수준·억양·이해도를 실시간 분석하여 적절한 학습 콘텐츠를 자동 생성·추천하는 AI 언어코치형 AGI가 도입될 것으로 보인다. 이를 통해 학습자의 언어능력 향상뿐 아니라 학습 몰입도와 지속성이 크게 개선될 것이다.

상기 내용을 요약하면, 향후 5년 및 10년 단위의 AGI 기반 교육 분야 발전 전망은 다음 〈표 7-11〉과 같이 정리할 수 있다.

〈표 7－11〉 교육분야 AGI의 중기(5년 후) 및 장기(10년 후) 전망

구분	중기(5년 후, 2030년 내외)	장기(10년 후, 2035년 내외)
AGI 기술 수준	– AGI 레벨 1(Embryonic AGI)에서 레벨 2(Superhuman AGI) 초기 단계로 진입. LLM 기반 Agent의 기능 다변화 및 실용화 확대.	– AGI 레벨 2(Superhuman AGI)로의 기술 고도화를 통하여 범용성 지속적 증가. 일부 교육 응용에서 고도화된 추론·학습·창의 능력을 갖춘 자율형 교육 AGI Agent 등장
교육특화 파운데이션 모델	– 교과별(국어, 수학, 과학 등) 데이터와 교수법 데이터를 기반으로 한 분야별 교육 특화 파운데이션 모델 시범 구축 – 교사 강의자료, 교과서, 시험문항 자동생성 등 교육 행정·콘텐츠 업무 중심의 초기 적용 단계	– 초·중·고·대학 전 교육단계를 포괄하는 통합형 교육 파운데이션 모델로 발전 – 정규·비정규 교육을 통합 지원하는 범용형 교육 AGI 플랫폼 – 학습자 수준·흥미·인지 스타일을 반영하여 자동 교과 설계 및 개인화 학습경로 제시 가능
교육특화 생성형 AGI	– 기존 MOOC 콘텐츠를 학습한 생성형 요약·보충 설명 AGI 등장 – 강의 자막·노트 자동 생성, 핵심 개념 요약, 퀴즈 자동출제 등 학습 보조 중심 기능 확립 – 학습자의 이력·선호 기반으로 추천 강좌·학습 경로 자동 제안 시스템 활성화 – 강의평가 및 피드백 자동 수집·요약으로 운영 효율성 개선	– MOOC 플랫폼 전반이 AGI 기반 자동 콘텐츠 생성·운영 체계로 전환 – 강사가 입력한 핵심 개요만으로 전체 강의 스크립트·슬라이드·영상 자동 생성 가능 – 학습자별 수준·언어·속도에 맞춘 개인화형 생성 강의(Adaptive MOOC) 보편화 – 전 세계 학습데이터를 활용한 자가학습형 글로벌 AGI MOOC 생태계 형성
교사 지원 에이전트	– 강의자료 제작, 학습 평가, 과제 관리 등 반복 업무 자동화 실현. AI 조교·강의지원형 Agent 도입 확대. – 교사의 수업계획, 시험문항 출제, 채점, 피드백 등 반복 업무를 자동화하는 보조교사형 AGI 서비스 확산 – 수업자료 추천, 학습 진단 보고서 자동화 등 실무 중심 활용	– AGI 기반으로 교수 지원 Agents 통합 고도화. AGI Agent형태로 수업 설계, 학습 진단, 피드백 자동화 등 교사의 의사결정 보조 획기적 강화. – 교사와 공동으로 수업을 설계·진행하는 Co-Teaching형 AGI 파트너 등장 – 학급 운영, 학생관리, 정서지도까지 지원하는 자율수업 관리형 AGI 교사로 발전

구분	중기(5년 후, 2030년 내외)	장기(10년 후, 2035년 내외)
학생 지원 에이전트	– 과목별 대화형 AI 튜터 서비스 보급 – 학습 수준·속도에 따른 맞춤형 학습자료 제공 및 기초형 개인화 서비스 정착. 시험 자동출제·평가 기능 제공	– 자율형 AGI 튜터링 서비스 확산. 실시간 진단·콘텐츠 생성·피드백을 통합한 완전 맞춤형 학습 – 학습자의 장기목표·진로 기반으로 학습설계·평가를 수행하는 자율학습 파트너형 AGI 확산 – 정서인식·메타인지 기능을 통합한 인지-정서 통합형 학습 동반자로 진화
STEM (과학·기술·공학·수학) 분야	– 수학·물리·화학·생물 등 주요 교과목에 도메인 특화 튜터 AGI Agent – 수학문제 풀이, 과학 실험 시뮬레이션, 코딩 교육 지원, 개념 학습 자동화 기능 확산. – 실습형 콘텐츠 자동 생성 및 문제 풀이 설명기능 확립	– 멀티모달 STEM AGI Agent로 발전. 실험 설계·시각화·탐구학습을 통합한 가상 실험 기반 학습 환경 구현. – 가상실험·데이터 분석·연구보고서 작성까지 수행하는 준-AGI급 STEM 학습환경 구축 – 복합적 문제 해결을 수행하는 탐구형 AGI 연구조교 시스템 등장
언어 (문해·소통) 분야	– 언어 이해 및 문법·쓰기 교육 중심의 LLM 튜터링 서비스 확산. – 발음, 억양, 문법 피드백이 가능한 대화형 언어튜터 AGI 상용화 – 학습자 수준별 회화 시나리오 생성·추천 기능 확립	– 말하기·듣기 중심의 대화형 AGI - 언어교사(Conversational AGI Tutor) 고도화 – 실시간 음성·제스처·상황인식 기반의 몰입형 언어교육 환경 정착. – 학습자의 문화·정서까지 반영하는 인지·감성 통합형 언어교육 AGI 실현
교육 생태계 변화	– 학교·온라인 교육 간 AI 학습보조 서비스 본격 도입. 교사·학생 간 상호작용 지원 중심.	– 정규·비정규 교육이 AGI 기반 통합 학습 생태계로 재편. 학습 효율·접근성·몰입도가 획기적으로 향상

이상의 내용은 AGI 발전에 따라 교육분야에서 구현될 대표적인 변화 전망을 요약한 것이다. 이러한 기술적 발전이 심화되면, 교육특화 파운데이션 모델을 중심으로 한 통합적 교육서비스 생태계가 형성될 것으로 예상된다. AGI의 범용성 특성에 따라, 현재 다수의 중소형 에듀테크 기업이 분산적으로 제공하던 기능들이 통합·고도화되면서, 일부 거대 교육 AI 기업 중심의 시장 재편 및 집중화 현상이 나타날 가능성이 높다. 특히 대규모 데이터와

모델 자원을 보유한 글로벌 사업자들이 교육 파운데이션 모델의 표준을 선도하게 되면, 독점적 구독형(Subscription-based) AI 교육서비스 모델이 시장의 주류로 자리잡을 것으로 예상된다.

나. 교육분야 AGI기술 경쟁력 확보를 위한 대응 전략

범용성이 높은 AGI기술이 플랫폼화될 경우, 사용자의 흡수력은 기존 플랫폼의 영향력을 초월하는 수준으로 확장될 것이다. 따라서, 만약 미국이나 중국에서 강력한 교육 특화 AGI 서비스가 출시된다면, 국내 대비가 미흡한 상황에서 넷플릭스와 유사한 현상처럼 국내 사용자들이 빠르게 글로벌 대기업의 구독형 서비스에 편입될 가능성이 높다. 특히 우리나라는 세계적으로 교육열이 높고 교육시장 규모 또한 큰 편이므로, 해외 기업이 국내 교육 AI 시장을 장악할 경우 막대한 국부 유출이 우려된다.

따라서 이러한 위험을 오히려 기회로 전환하여, 높은 교육열을 기반으로 한 교육분야 AGI 기술 확보 전략이 필요하다. 즉, 미국과 중국을 위협할 수 있는 교육 버티컬(Vertical) AGI를 확보하기 위해 정책적·연구적 투자가 병행되어야 하며, 정부·민간기업·대학이 협업하는 국가 차원의 AGI R&D 네트워크를 구축해야 한다.

향후 1~2년간은 **AGI 기반 교육 특화 파운데이션 모델 구축을 위한 데이터 확보 및 인프라 조성**이 최우선 과제가 되어야 한다. 현재 과기정통부 주도로 진행 중인 범용·도메인 특화 파운데이션 모델 사업은 일부 선정된 기관 중심으로 운영되어 참여 확장성의 한계가 존재한다. 범용 파운데이션 모델의 경우 ChatGPT나 Gemini와 같은 글로벌 선례가 명확해 국가대표급 사업으로 추진하기 용이하나, 분야별 특화 파운데이션 모델은 독점 구독형 사례가 아직 희소하므로 다양한 기업과 대학이 자유롭게 시도·참여할 수 있는 개방형 연구 환경 구축이 절대적으로 필요하다.

특히 **AGI 혁신을 위한 고급 인재 양성**이 무엇보다 중요하다. 현재 국내 대학원 단계에서조차 파운데이션 모델 구축을 직접 실험해볼 수 있는 환경이 부족한 점은 큰 제약이다. 따라서 대학이 GPU 인프라를 '기초 연구장비'로 간주하고 적극 투자해야 하며, 이를 통해 **더 많은 연구자가 직접 파운데이션 모델을 구축·실험**할 수 있도록 해야 한다. 이러한 기반이 마련된다면 국내 연구진이 AGI를 향한 다양한 아이디어를 생산적으로 검증하고, 교육 AGI 혁신의 내재적 역량을 확보할 수 있을 것이다.

또한 교육 AGI 개발에는 데이터 가공 및 변환 인프라 구축이 필수적이다. 동영상 강의, 학습 콘텐츠뿐 아니라 교사-학생 간 피드백 데이터를 수집·정제하여 데이터화해야 하며, 이를 위해 자연스러운 학습·상호작용 데이터를 확보할 수 있는 하드웨어·인터페이스 환경을 마련해야 한다. 수집된 데이터는 국가 AI 데이터 허브를 통해 공유하되, 해외 유출을 방지하기 위한 **'소버린 데이터(Sovereign Data)' 정책**이 병행되어야 한다. 데이터 등급에 따른 접근·보안 정책을 강화하고, 필요시 관련 법령을 제정하여 보호 체계를 확립해야 한다.

아울러, **교육특화 파운데이션 모델 개발사업과 병행하여 교육 AGI 에이전트 기술 개발사업**을 정부 차원에서 추진해야 한다. 각 기업과 대학이 개발한 파운데이션 모델을 기반으로, 다양한 교육 AI 기업이 에이전트형 서비스 출시를 통해 경쟁할 수 있는 환경을 조성해야 하며, 혁신형 유니콘 기업 발굴 정책을 병행하여 글로벌 대형 AI 기업과의 경쟁력을 확보해야 한다.

결국, 파운데이션 모델 인프라 구축 → 대학 인프라 확장 → AGI 고급 인재 양성 → AGI 기술사업화로 이어지는 전 주기적 전략이 필요하다. 이는 단순한 기술개발을 넘어, 국내 교육 AGI 대표기업을 육성하기 위한 시스템적 기반 조성으로 볼 수 있다. 우리나라의 **세계적 수준의 교육열과 풍부한 학습시장 규모는 교육 AGI 유니콘 기업이 성장할 수 있는 강력한 토대가 될 것**이며, 이를 위해 **파운데이션 모델·인프라·인재·사업의 네 축이 유기적으로 선행 구축**된다면, 국내에서도 전세계적으로 경쟁력있는 독점적 구독형(Subscription-based) AI 교육서비스 모델을 운영하는 교육 AGI 기업이 탄생할 수 있으리라 기대한다.

3. 법률분야 AGI기술 활용 현황

법률산업 또한 LLM과 생성형 AI의 도입으로 자동화 및 지식기반 의사결정 지원이 빠르게 확산되고 있다. Grand View Research(2024)에 따르면, 글로벌 법률 AI 시장 규모는 2024년 약 14.45 억 달러(약 1조 9,700억 원)에서 2030년 39.18 억 달러(약 5조 3,500억 원)로 성장하여, 연평균 17.3 %(CAGR)의 높은 성장세를 지속할 것으로 전망된다. 국내외 법률 AI 분야 기업들은 다음 〈표 7-12〉에 정리되어 있다.

〈표 7－12〉 법률 분야 대표 AI 기업 및 제공 서비스

기업	기업소개 요약	대표서비스
넥서스에이아이(국내) https://www.nexusai.kr/	초거대 언어모델(LLM) 기반으로 일반인을 위한 법률상담 챗봇과 변호사 대상 업무보조 솔루션을 개발	• 법률상담: 판결문, 유사 상담사례, 법령 등 공개 데이터를 기반으로 답변을 생성하며, 답변 내용에 출처를 명시해 신뢰성을 확보한 실시간 법률상담 솔루션 • AI사건진단: 변호사 또는 로펌이 서비스의 주체가 되는 SaaS 형태의 솔루션으로, 의뢰인의 사건정보를 AI가 분석하여, 예상형량, 집행유예가능성, 죄명, 공소방식 등을 리포트로 제공
엘박스(LBox)(국내) https://lbox.kr/	총 250만 개 가량의 판례가 등록된 국내 최대 판례 검색서비스 제공	• 판례 검색·AI 리서치 플랫폼: 400만 건 이상의 판례 데이터를 바탕으로 법률 전문가들이 자연어 질의를 통해 관련 판례 검색, 문서 분석 및 초안 작성 기능을 활용할 수 있도록 지원하는 법률 AI 플랫폼
로폼(LawForm)(국내) https://www.lawtalk.is/	국내법에 최적화된 AI 기반 법률문서 자동작성 플랫폼으로, 대화형 채팅을 통해 법률요건을 분석하고 계약서 - 서명 - 관리까지 한 번에 처리	• 계약서·법률문서의 자동생성, 채팅기반 법률요건 분석, 그리고 전자서명 및 클라우드 보관까지 통합 지원하는 문서관리 플랫폼
로앤컴퍼니(국내) https://www.lawtalk.is/	이용자가 변호사와 연결될 수 있는 플랫폼인 법률 상담 앱 'LawTalk' 운영	• LawTalk: 변호사를 찾는 일반 사용자와 변호사를 연결해주는 온라인 법률 상담 플랫폼
AI Lawyer(해외) https://ailawyer.pro/	계약서·소송 서류·클레임 등 법률문서를 자동 생성·요약하고, 실시간 질의응답을 제공하는 다기능 법률 AI 어시스턴트 플랫폼으로 비전문가도 손쉽게 법률 업무를 처리할 수 있게 하는 서비스 제공	• 문서 자동화: 계약서·소장 등 법률 문서를 AI가 자동 생성하거나 요약 • 실시간 질의응답 및 연구 보조: 사용자 질문에 대해 판례·법령 검색, 법률리서치 지원을 제공하는 가상 법률 어시스턴트

기업	기업소개 요약	대표서비스
vLex(해외) https://vlex.com/	전 세계 100여 개국의 판례·법령·논문 등 10억 건 이상의 법률 자료를 통합 제공하는 글로벌 AI 법률 리서치 플랫폼으로, 자연어 검색과 인공지능 요약 기능을 통해 법률가의 판례·계약·소송 업무를 자동화하고 지원	• Vincent: 자연어 질문으로 법률 자료(판례·법령 등을 포함한 10억 건 이상)를 검색하고, 정확한 인용과 함께 답변을 제공하는 AI 기반 법률 리서치 플랫폼
Darrow AI(해외) https://www.darrow.ai/	법률 리스크 탐지 및 문서자동화 기능을 갖춘 AI 플랫폼으로, 기업이 계약서·법률문서 내 잠재적 리스크나 불리한 조항을 자동으로 식별하고 대응할 수 있도록 지원	• AI 기반 계약서·문서 리스크 탐지 및 조항 분석 플랫폼: 기업이 보유한 법률 문서를 자동으로 스캔하여 불리하거나 비표준적인 조항을 식별하고, 이를 쿼리 가능하게 표시하여 대응을 권고하는 기능

한편, 법률 분야의 AI 기술과 관련하여, 스탠퍼드 코덱스(Stanford CodeX)[4] 분류체계에서는 전 세계 리걸테크(LegalTech) 기업들이 제공하는 서비스를 9개 범주로 구분하고 있으며, 2025년 10월 현재, 각 범주별 기업 수 및 주요 특징은 다음 〈표 7－13〉과 같다.

〈표 7－13〉 법률 AI 서비스의 9개 범주와 대표 기업

범주	주요내용 및 정의	대표기업	등록 기업수
분석 및 인사이트 (Analytics & Insights)	예측 모델, 데이터 시각화, 벤치마킹 등을 통해 법률 의사결정을 지원하는 정량적·정성적 분석 도구를 포함	Lex Machina, KenRi	148
준법 및 리스크 관리 (Compliance & Risk)	규제 준수, 데이터 프라이버시, 사이버 보안, 부패방지, ESG(환경·사회·지배구조) 리스크 관리를 위한 솔루션	Accurex, AI Compliance	431
계약 관리 (Contract Management)	계약서의 작성, 검토, 협상, 관리 전 과정을 지원하는 플랫폼	Ironclad, DraftPilot, SpotDraft	245

4) CodeX TechIndex. https://techindex.law.stanford.edu/

범주	주요내용 및 정의	대표기업	등록 기업수
지식 및 법률 리서치 (Knowledge & Research)	법률 데이터베이스 접근, 조직 내 지식 관리, 실무 가이드 및 참고자료 제공 서비스	vLex, Casetext	258
마켓플레이스 및 대체법률서비스 (Marketplace & ALSPs)	대체법률서비스 제공업체(ALSP), 법률 인재 매칭 플랫폼, 프로젝트 기반 법률 자문 중개 마켓플레이스 등	UpCounsel, Axiom	370
소송 및 분쟁 해결 (Litigation & Dispute Resolution)	소송관리, 전자증거개시(eDiscovery), 중재·조정(ADR) 등을 지원	Relativity, Everlaw	326
법률 실무 관리 (Practice Management)	의뢰인 관리, 일정관리, 청구·회계, 사건 진행 추적 등 로펌 운영 전반의 효율성을 높이는 플랫폼	Clio, PracticePanther	395
문서 관리 및 자동화 (Document Management & Automation)	법률 문서의 작성, 구성, 저장, 자동화 기능을 제공하는 시스템	GoFirmex, Litera	561
지식재산 관리 (IP Management)	특허, 상표, 저작권 등 지식재산권 포트폴리오의 등록·관리·보호를 지원하는 도구	Anaqua, PatSnap	197

Hou et al.(2025)에 따르면, 학계에서 연구 중인 법률 AI의 대표 과업은 데이터셋 중심으로 **법률문서 분류, 판결 예측, 법률 질의응답, 판례 검색, 법률 요약** 등 다섯 가지 핵심 영역으로 구분된다. 또한, **법률 도메인 특화 대형언어모델(Legal LLM)**은 2023년 이후 급격한 발전세를 보이며, 모델 규모가 수십억(6B) 수준에서 2025년에는 100B 이상으로 확대되는 추세를 보인다. 초기에는 중국어권의 ChatLaw-33B, HanFei, SaulLM 등이 시장을 주도하였으며, 이후 영어권의 InLegalLLaMA, Lawyer-LLaMA, 다국어 모델 MindsLegal 등이 잇따라 개발되었다. 이러한 모델들은 주로 LLaMA, BLOOM 등 오픈소스 기반 프레임워크를 활용하여 구축되었으며, 법률문서 생성, 판례검색, 질의응답, 조항추론 등 실무 중심 응용 분야로 빠르게 확산되고 있다. Hou et al.(2025)는 향후 법률 LLM의 주요 현안으로 1) 할루시네이션(비사실적 정보 생성), 2) 멀티모달리티(비텍스트 데이터 미활용), 3) 다국어성(법체계

별 차이로 인한 성능 저하), 4) 해석가능성의 도전과제를 제시하였으며, 향후 연구 방향으로 지침형 학습데이터 기반의 고품질 합성데이터셋 구축(Synthetic Dataset Synthesis), RAG(Retrieval-Augmented Generation)을 활용한 해석가능성 강화, 음성·영상 등 멀티모달 데이터 융합, 다국어 법률 LLM 개발 및 관할권별 정렬(Alignment), 복합 사건 처리(다수 피고·다중 혐의), 저비용 배포 가능한 소형 언어모델 개발 등을 제안하였다.

4. 법률분야 AGI기술 발전 전망 및 경쟁력 확보 대응방안

가. 법률분야 AGI기술 활용 중기(5년후), 장기(10년후) 전망

법률 분야 또한 단기적으로는 **파운데이션 모델과 LLM 에이전트를 활용한 서비스 확장과 품질 개선**에 초점이 맞춰질 것으로 보인다. 그러나 AGI 기술이 발전하고 시간이 지남에 따라, **개별적으로 운영되던 법률 서비스들이 법률 특화 파운데이션 모델의 범용성이 높아짐에 따라 통합**되고, **초거대 법률 특화 파운데이션 모델**을 기반으로 하는 서비스로 전개될 가능성이 높다.

변호사의 법률 자문 및 행정 업무 보조의 경우, AGI의 발전에 따라 **기존의 변호사가 AI에 하나하나 지시하던 수작업형 인터페이스가 고도화된 자동화 방식으로 전환**될 전망이다. 변호사가 사건을 맡을 때 수행하던 법률 조사 및 사전 검토 과정은 장기적 법률 추론(long-term legal reasoning)과 사무 기능이 결합된 형태로 발전하여, "AI 보조변호사" 수준의 AGI 서비스가 구현될 것으로 예상된다. 즉, 사건 관련 고객 정보를 문서 형태로 입력하면, "AI 보조변호사"가 변호사의 업무 스타일과 법률 규정을 학습하여 장시간 대화 없이도 일정 시간이 지나면 다양한 변호 대안을 제시할 수 있게 될 것이다.

또한 사무 행정 분야에서도 **"AI 법률행정 보조"가 소송에 필요한 서류 작성, 일정 관리, 고객 응대 및 관리 등 변호 행정 업무를 자동화**함으로써, 기존에 분산되어 있던 개별 서비스들이 통합된 형태로 등장할 가능성이 높다. 중기적으로는 이러한 변호·행정 업무 전반이 통합되어, "AI 변호사무소" 형태의 일원화된 서비스 구조가 정착될 것으로 예상된다.

한편, 고객 측면에서는 변호사 선임 없이 AI 변호 상담이 점차 고도화되면서, 변호사 도움 없이도 일정 부분의 법률 자문이 가능해질 전망이다. 단순 사건의 경우 단기적으로 상당 부분 대체가 가능할 것으로 보이며, 복잡한 추론과 논리가 필요한 사건은 중기적으로

일부 대체가 가능할 것으로 판단된다. 장기적으로는 난이도가 높은 사건을 제외한 대부분의 사안에서, AI 변호 상담만으로도 고품질의 법률 자문 서비스를 제공받을 수 있을 것으로 예상된다.

결국, 현재는 분야별로 특화된 법률 서비스가 독립적으로 운영되고 있으나, 향후에는 유사 도메인 간 통합이 가속화되면서 AGI 기반의 범용적 법률 파운데이션 모델로 발전할 것으로 전망된다. 이에 따라 법률 서비스의 범위가 확대되고, 서비스 품질과 효율성 또한 비약적으로 향상되는 방향으로 진화할 것으로 예상된다.

상기 내용을 요약하면, 향후 5년 및 10년 단위의 AGI 기반 법률 분야 발전 전망은 다음 〈표 7－14〉과 같이 정리할 수 있다.

〈표 7－14〉 법률분야 AGI 활용 중기(5년 후), 및 장기(10년 후) 전망

구분	중기(5년 후, 2030년 내외)	장기(10년 후, 2035년 내외)
법률 특화 파운데이션 모델	– 법률문서, 판례, 행정데이터를 학습한 법률 특화 파운데이션 모델 구축이 확산될 전망. – 각 로펌·법률기관별로 개별 운영되는 전문 LLM 모델이 등장하여, 업무 자동화 및 맞춤형 법률지원 기능 강화.	– 법률 영역 전반을 포괄하는 초거대 통합형 법률 파운데이션 모델이 등장. – 다양한 서비스가 통합되어 범용적 법률 AGI 생태계로 발전하며, 법률 AI 시장 구조의 대형화 및 집중화 가속.
변호업무 지원 AGI 에이전트	– 변호사 업무를 지원하는 보조형 LLM Agent가 확산. – 사건 분석, 문서작성, 리스크 예측 등 반복적·표준화된 업무 중심의 자동화	– 개별 Agent들이 상호 연결되어 자율적 판단 및 법률추론을 수행하는 “AI 보조변호사” 수준으로 진화 – 변호사의 스타일과 논리 패턴을 반영한 장기적 리걸 추론(Long-term Legal Reasoning) 탑재형 고등 변호 보조업무 수행
법률행정 보조 AGI 에이전트	– 소송 서류작성, 일정관리, 고객상담 등 사무행정 기능을 자동화하는 AI 법률비서 서비스가 본격 확산.	– 변호사 행정·고객관리·소송절차 관리가 완전히 통합된 “AI 법률행정 플랫폼”이 등장. 행정과 변호 업무가 일원화된 디지털 “AI 변호사무소” 형태로 정착.

구분	중기(5년 후, 2030년 내외)	장기(10년 후, 2035년 내외)
AI 법률 상담 및 변호 대체	– 단순 사건 중심의 AI 법률상담 기능이 보편화되어, 기초 자문·문서 작성 등에서 변호사 의존도 점진적 감소.	– 복잡한 사건의 논리적 변호까지 수행 가능한 고도화된 AI 변호상담 서비스 등장. – 변호사 없이도 일정 수준의 변호 품질을 제공하는 준(準) AI 변호 서비스 체계 확립.
AGI형 법률 통합 서비스 생태계	– LLM 기반 개별 법률 서비스들이 독립적으로 운영되며, 도메인별 전문성 중심의 분산형 구조 유지.	– 유사 도메인 간 융합 및 통합이 가속화되어, 법률·행정·상담 기능이 하나의 AGI 허브로 통합. – 글로벌 수준의 법률 AGI 네트워크가 형성되어 초연결형 법률 생태계로 발전.

법률분야의 AGI는 단기적으로는 파운데이션 모델과 LLM 에이전트를 통한 업무 효율화 및 서비스 확장에 집중될 것으로 보이며, 중기 이후에는 이들 기술이 통합되어 “AI 보조변호사–AI 법률행정비서–AI 변호사무소” 형태로 발전할 가능성이 크다. 궁극적으로는 법률 특화 파운데이션 모델을 중심으로 AGI형 법률 생태계가 구축되며, 주요 로펌 및 기술기업을 중심으로 한 구독형 법률 AGI 서비스 시장의 독점화 및 플랫폼화가 본격화될 것으로 예상된다.

다. 법률 AGI기술 경쟁력 확보를 위한 대응 전략

법률 분야에서도 도메인 특화 파운데이션 모델을 조속히 추진해야 하며, 이를 위해 법률 도메인 데이터 구축과 관련된 현안을 면밀히 조사할 필요가 있다. 교육 분야는 교육 콘텐츠가 공개되어 있고, 교육과정 자체도 공유가 가능한 상황이며, 학생-교사 간 1:1 튜터링 조차 공개에 큰 문제가 없는 데이터 유형이다. 반면, 법률 분야는 의뢰인과 변호사 간 상담 등에서 개인정보 이슈가 존재하므로, 의료 분야와 마찬가지로 AGI 확장을 위한 데이터 품질 향상에 있어 데이터 구축 문제가 핵심 과제가 될 수 있다. 따라서 공개 가능한 데이터와 개인정보가 포함된 비공개 데이터를 구분하여, 당장은 공개 가능한 데이터의 확보 및 구축을 통해 대규모 법률데이터를 구성하고, 이를 기반으로 파운데이션 모델 개발에 집중해야 한다. 이후에는 비공개 데이터를 활용한 추가 확장 방안으로서, 변호사나 로펌 단위로 의뢰인과의 소통 내용, 사건 정보 등을 기밀 데이터로 활용하되, 변호 업무가 완료

된 이후에는 해당 데이터를 모델 학습에 반영하여 해당 변호사의 **AGI 기반 개별 모델**을 지속적으로 개선할 수 있도록 하는 구조가 필요하다. 이 과정에서는 높은 수준의 개인정보 보호를 유지하면서도 법률 특화 파운데이션 모델을 추가 학습할 수 있는 기밀성 유지형 Alignment 기술이 필수적이다. 또한 공개 가능한 경우에도 로펌 간, 변호사 간 데이터 공유가 쉽지 않으므로, 국가 차원에서 데이터 공유 체계와 인센티브를 마련해야 한다.

또한, 법률 분야에서 변호 업무 등을 대체하기 위해서는 방대한 법률 지식과 법리적 추론 능력을 법률 도메인에 특화하여 강화할 필요가 있다. 따라서 1차적으로 법률 특화 파운데이션 모델이 구축된 이후에는, 이를 기반으로 **법률 특화 대규모 법률추론모델(Large Legal Reasoning Model)을 추가 개발**하여, **복잡한 법리 해석과 논리적 판단을 수행할 수 있는 수준으로 고도화**해야 할 것이다.

교육 분야와 달리, 법률 AI 서비스는 상시 수요가 아닌 사건 발생 시점에 한시적으로 이용되는 특성이 있으므로, 일반 사용자보다는 변호사나 법률 관계자를 위한 구독형 AGI 서비스가 자리 잡을 가능성이 높다. 변호사는 업무 효율성을 높이기 위해 AGI를 활용할 것이며, 이러한 기술이 발전함에 따라 변호사를 대체할 수 있는 기능이 점차 강화될 가능성도 있다. 변호 업무 또한 매우 다양하므로, 파운데이션 모델 개발 이후에는 여러 AI 개발자와 사업자가 다양한 법률 에이전트를 개발할 수 있는 환경이 조성되어야 한다. 그런데, **변호 업무는 전문성이 높아 외부에서 접근하기 어려우므로, 변호 프로세스를 구체적으로 데이터화하여 공개하고, AGI 개발자가 이를 쉽게 활용할 수 있도록 접근성을 높이는 제도적 장치가 필요**하다. 교육 분야와 마찬가지로, **다양한 법률 AGI 사업자가 파운데이션 모델을 기반으로 각종 에이전트 서비스를 출시해 경쟁할 수 있는 개방형 생태계를 조성**해야 한다. 법률 도메인은 의료 분야처럼 전문성이 높아 진입장벽이 존재하므로, **변호 경험이 없더라도 기술 역량만으로 참여할 수 있도록 하는 접근 용이성 확보**가 마련되어야 할 것이다.

궁극적으로, 법률 분야에서도 **다양한 AGI 개발자들이 법률 도메인에 참여하여, 미국과 중국의 거대 기업에 대응할 수 있는 소버린(Sovereign) AGI 생태계를 구축**해야 한다. 이를 통해 국내 법률 AI 시장을 보호하고 동시에 글로벌 경쟁력을 확보할 수 있는 대표적 법률 AI 기업의 성장 기반을 마련하는 것이 중요하다.

제 6 절 해양/물류

해양·물류 산업은 글로벌 공급망의 핵심축으로, AI·로보틱스·자율운항 등 기술이 빠르게 융합되고 있는 분야이다. 본 절에서는 AGI 기술이 해양 운항, 항만 물류, 해상 안전 및 효율성 향상에 미치는 영향을 분석하고, 중기(5년)·장기(10년) 관점에서 산업 전반의 구조적 변화를 전망하며 AGI 기반 경쟁력 확보 전략을 제시하고자 한다.

1. AGI 기술 활용 현황

해양·물류 산업은 세계 교역의 80% 이상(World Trade Organization, 2023)을 담당하는 핵심 산업으로, 글로벌 공급망 안정성과 국가 경제안보에 직결된다. 전통적으로 이 산업은 해운-항만-물류 의 세 축으로 정의되어 왔으며(Caliskan, A., and Y. Ozturkoglu, 2018), 각 영역이 상호 의존적으로 작동해 왔다. 최근 10년간 인공지능의 급격한 발전은 이러한 구조에 변화를 가져왔다. 자율운항, 디지털트윈 기반 관리 등 세 축간 기능적 통합이 가속화(Markopoulos, E., et al., 2023)되고 있으며, 이는 향후 AGI로의 진화를 준비하는 과도기로 평가된다. 그러나 현재 기술은 대부분 Narrow AI 수준에 머물러 있거나 일부 응용 영역에서는 레벨 1: Emerging AGI이 관찰되고 있다. 본 절에서는 해운, 항만, 물류의 세 부문별로 AGI 기술의 적용 현황과 기술적 성숙 단계를 분석하고, 각 부문의 융합 가능성을 조망한다.

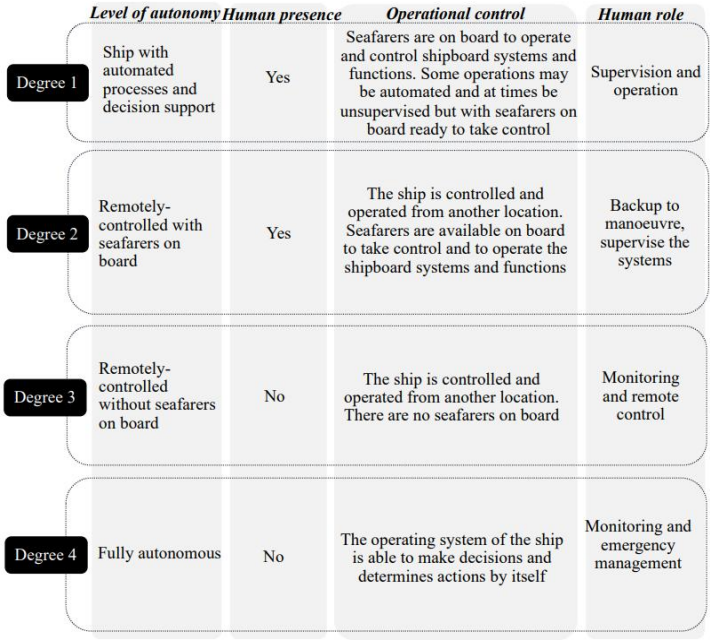
가. 해운 분야: 자율운항 및 운항 의사결정 지원

해운 분야는 AGI 기술이 가장 빠르게 확산되는 영역으로, 자율운항선박(MASS: Maritime Autonomous Surface Ship) 이 대표적이다. 국제해사기구(IMO)는 2024년 “MASS Code”를 통해 자율운항 수준을 자동화 지원(Degree 1) 에서 완전 자율운항(Degreel 4) 까지 구분하고, AI 기반 운항체계의 법적 책임과 안전기준을 제시하였다(International Maritime Organization, 2024).

국외 사례로는 노르웨이의 Yara Birkeland호가 있다. 항로 설정·충돌 회피·접안을 AI가 수행하는 세계 최초의 완전 자율운항 화물선으로, 이는 해운·항만·물류 전 주기 데이터를 통합한 대표 사례로 평가된다(Yara International & Kongsberg Gruppen, 2021). 국내에서도 해양수산부가 주관하는 Korea Autonomous Surface Ship(KASS) 프로젝트가 추진 중이며,

2025년도 기준 IMO 기준의 Degree 3 자율운항 기술 확보를 목표로 하고 있다(Korea Autonomous Surface Ship Project Office, 2020),(Ministry of Oceans and Fisheries, 2020). 이러한 국내외 사례들은 선박 데이터·기상정보·주변 환경을 결합하여 실시간 자율 의사결정을 수행함으로써, 인간 개입 없이 운항을 최적화하는 레벨1: Emerging AGI 단계의 의사결정 시스템으로 발전하고 있다. 다만 자율운항의 확산에는 안전 및 사이버보안 위험 등의 과제가 남아 있다. 예를들어 20,000 TEU(20피트 컨테이너 1개) 컨테이너 선박의 경우 US$54,500/TEU 수준이라는 추정을 선박이 나르는 물류의 가치가 약 10억 달러(약1.4조원)이 될 것으로 예상된다. 즉, 자율운항을 위해서는 그 자율운항 AI의 안전성과 신뢰성 및 보안대 대한 강건성이 매우 중요한 특징이 될 것으로 예상뇐다.

〔그림 7－17〕 IMO에 의해 정의된 해운 분야 자율운항 수준



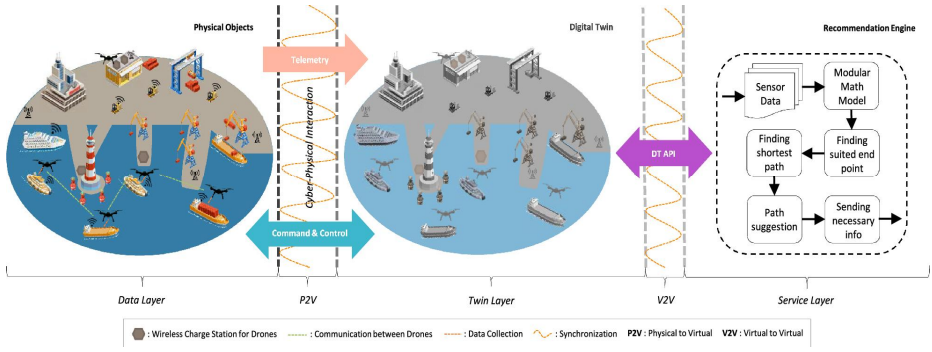
자료: Sharma, Amit, et al. 2019

나. 항만 분야: 디지털트윈 및 예지정비 자동화

항만 분야는 해양·물류 산업의 핵심 거점으로, 선박 입출항 및 하역 장비 운영, 육상 운송이 복합적으로 연계된 지능형 산업 생태계이다. 최근 AGI 핵심 요소인 지각, 추론, 기억 기술이 항만 운영에 도입되면서, 물리적 항만을 가상 공간에 복제해 의사결정을 지원하는 디지털트윈(Digital Twin) 기술이 빠르게 확산되고 있다. 디지털트윈은 선박 위치, 하역 장비 상태, 기상 및 물류 흐름 데이터를 통합 시뮬레이션함으로써 예지정비, 운영 최적화, 에너지 효율 향상 등을 지원한다. 이는 향후 AGI 수준의 지능형 항만 운영체계로 발전하기 위한 과도기로 평가된다.

국외에서는 네덜란드 로테르담항(Port of Rotterdam Authority, 2023)이 대표적이다. 이 항만은 인프라·기상·수위·선박 이동 데이터를 통합 관리하는 디지털트윈 플랫폼을 운영하며, 2030년까지 자율운항선박을 수용할 수 있는 스마트 인프라 전환을 추진하고 있다(World Trade Organization, 2023). 또한 중국 상하이 양산항(Ding, Y., et al., 2023)은 5G·AI·IoT 기술을 결합해 자동화 터미널의 실시간 모니터링과 자원배치 최적화를 수행하는 디지털트윈 시스템을 적용함으로써 장비 운영 효율을 높이고 비용을 절감하고 있다. 또한, 최근 연구(Yigit, Y. et al., 2023)에서는 드론과 5G 네트워크를 활용하여 항만 전역의 실시간 데이터를 디지털트윈 환경과 연계하는 자동화 방식도 제안되었다. 국내에서는 부산 신항(Eom, J.-O., Yoon et al., 2023)이 디지털트윈 활용의 핵심 사례로 꼽힌다. 부산항만공사와 국내 연구기관은 항만 운영 데이터를 활용해 선박 도착예측, 배치 최적화, 탄소배출 절감 전략을 통합한 지능형 항만 운영 플랫폼을 개발하고 있으며, 이는 항만 운영 자동화 및 탄소중립화의 실질적 기초로 평가된다. 이러한 사례들은 항만 분야가 데이터 기반 지능형 의사결정으로 이행하고 있음을 보여주지만, 여전히 인간의 감독 및 규칙 기반 제어에 의존하고 있다는 점에서 완전한 AGI 수준에는 도달하지 못한 레벨 1: Emerging AGI 수준의 초기 단계로 평가된다.

〔그림 7－18〕 디지털트윈과 연계한 항만 분야 자동화 사례



자료: Yigit, Yagmur, et al. 2023

다. 물류 분야: AI 기반 예측 및 최적화

물류 산업은 상품의 입고, 저장, 출고, 운송으로 이어지는 복합 흐름을 지니며, 이들이 서로 유기적으로 연결된 거대한 사이버-물리시스템이다. 최근 들어 인공지능 기술이 물류 경로 예측, 재고 최적화, 운송수단 스케줄링 등에 폭넓게 적용되면서, 이러한 흐름을 통합적으로 제어할 수 있는 예측 및 최적화 기능이 부각되고 있다. 특히, 시뮬레이션 기반 디지털트윈 구조와 결합되며, 물류환경의 실시간 가상복제 및 최적의 운영방법 도출이 가능해지는 방향으로 진화하고 있다.

국외에서는 세계적 물류기업인 DHL(DHL Trend Research, 2023)이 디지털트윈을 활용한 물류센터 설계 및 물류흐름 시뮬레이션 사례를 공개했다. 이들은 물류창고 내부의 이동 동선, 재고 흐름, 자동화차량 및 드론 운용을 가상모델에 반영하고, AI 기반 시나리오 분석을 통해 공간 활용과 작업 효율을 개선하였다. 국내에서는 CJ대한통운(CJ Logistics, 2022)의 “아폴로(APOLO)”프로젝트가 대표적이다. 이 기업은 군포 스마트 풀필먼트센터에 디지털트윈을 적용하여 실제 창고와 동일한 조건의 가상물류센터를 구축하였으며, 물류 설비의 위치·작업자의 동선·물량 변화 등을 가상환경에서 시뮬레이션하고 장비 고장 및 병목 지점을 실시간으로 파악할 수 있도록 설계 했다. 그럼에도 불구하고 현재 물류 분야의 기술은 여전히 부분적 자동화와 제한적 최적화 수준에 머물러 있다. 전 구간을 아우르는 통합적이고 자율적인 운영체계, 즉 AGI 수준의 완전한 의사결정형 플랫폼으로 발전하기 위해서는 데이터 통합, 실시간 변수 반영, 인간 감독 최소화 등 다수의 과제가 남아 있다.

〔그림 7－19〕 디지털트윈과 연계한 물류 분야 최적화 사례

2. 중기(5년 후) AGI 발전 전망

해양·물류 산업은 향후 5년 내 Emerging-Competent AGI(레벨 1-2) 단계로의 이행기에 진입할 것으로 전망된다. 현재 Narrow AI가 주로 반복적 판단과 단일 기능 수행에 한정되어 있다면, 5년 후에는 복합적 상황 판단, 예측, 자율 최적화 기능을 갖춘 초기 범용지능의 구현이 예상된다. 이 시기에는 해운, 항만, 물류의 각 기능이 독립적으로 운영되는 것을 넘어, 데이터·시뮬레이션·의사결정이 통합된 복합 지능형 생태계로 재편될 것이다.

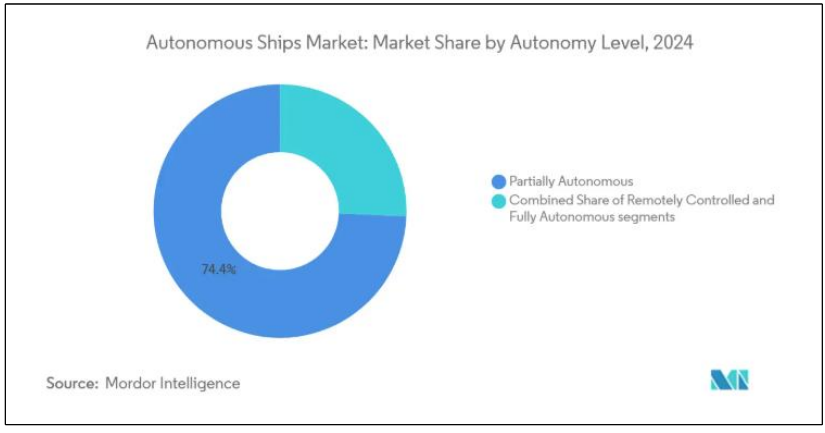
가. 해운 분야: 설명 가능성 기반 자율운항 고도화

해운 산업은 중기 시점에서 AGI 기술의 직접적 적용이 가장 빠르게 진행될 분야로 평가된다. IMO가 2024년 발표한 MASS Code는 현재 전 세계적으로 실증 운항 중인 비승선 원격제어형 선박(Degree 3)을 주요 적용 대상으로 하며, 2025년까지 비의무적 초안을 완성하고 2028년부터 의무화할 예정이다(International Maritime Organization, 2024). 이는 2030년 이전에 글로벌 해운 시장 전반에서 Emerging-Competent AGI(레벨 1~레벨 2) 수준의 운항 지능 시스템이 상용화될 가능성이 높음을 시사한다.

시장조사기관 Mordor Intelligence의 보고서(Mordor Intelligence, 2025)에 따르면, 2024년 기준으로 부분 자율형(Partially Autonomous) 선박이 전체 시장의 약 74.4%를 차지하며, 완

전 자율형 및 원격 제어형(Remotely Controlled & Fully Autonomous) 선박이 점진적으로 증가하는 추세를 보이고 있다(그림 7-20). 또한, 자율운항선박 시장 규모는 2025년 약 6.9억 달러에서 2030년 11억 달러 수준으로 성장할 것으로 전망된다. 이러한 추세는 단순한 시장 팽창이 아니라, 자율운항 기술의 상용화 확대에 따라 운항 안전성과 책임성 확보에 대한 요구가 급격히 높아지고 있음을 의미한다. 특히 자율운항선박의 판단이 인간의 감독 없이 이뤄지는 비율이 증가함에 따라, 각국 해사당국과 보험기관은 AI 판단의 근거, 불확실성, 재현 가능성을 핵심 평가 요소로 보고 있다. 이에 따라 최근 연구들은 조건부 자율형(Degree 2~3) 중심의 제어 시스템을 넘어, 판단·추론·학습 기능과 설명가능성(Explainability)을 동시에 구현하려는 방향으로 진화하고 있다.

〔그림 7-20〕 자율형 선박의 시장 점유율

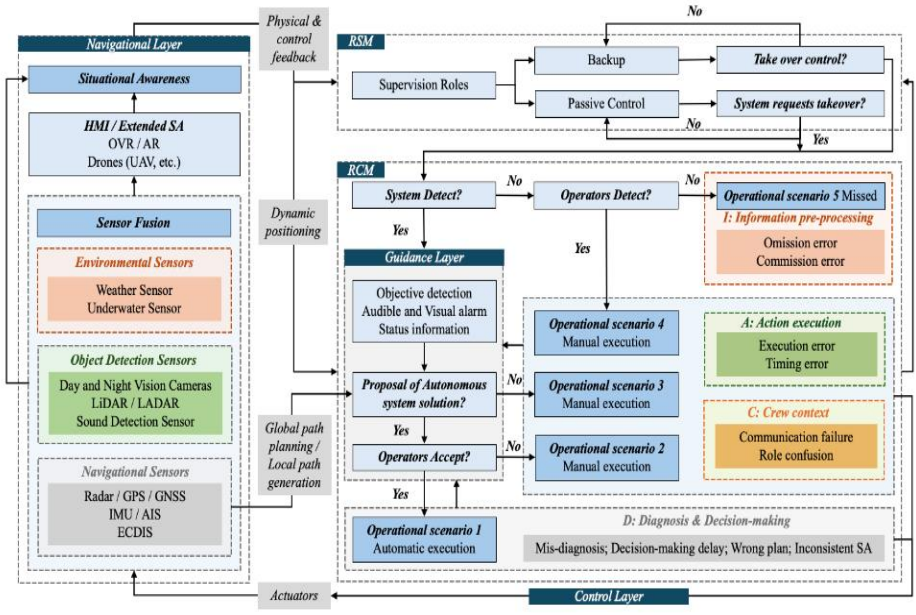


예컨대 Zhang et al.는 자율운항선박의 운항 의사결정 과정에서 AI의 판단 근거를 실시간으로 시각화하고, 인간 운항자와의 신뢰 기반 협업을 지원하는 설명가능(그림 7-21)를 제안하였다. 이는 아직 초기 연구 단계에 머물러 있지만, 설명가능성 연구는 자율 판단의 신뢰성과 책임성을 확보하기 위한 핵심 전이 단계로서, Emerging-Competent AGI 수준으로 진화하기 위해 반드시 거쳐야 하는 기술적 기반을 의미한다.

더불어 자율운항 선박이 위성 통신을 통해서 스스로 각국의 관련 관제 시스템과 소통하여 태풍 등 기상 상황을 고려하고 운항 경로에 있는 다른 선박들과 소통을 하면서 최적 경

로를 조정하고 경로상의 전체 시스템을 최적화 할 것으로 예상된다. 이 과정에서 AI는 주변 환경에 대한 정보를 수집하여 주요 정보를 전달하는 것, 상황에 따라 연료와 경로를 최적으로 바꾸는 것 등을 수행할 것으로 예상된다.

〔그림 7－21〕 자율운항선박의 의사결정에 대한 설명가능 인터페이스 도식



나. 항만 분야: 자율학습형 디지털트윈 기반 항만 운영 고도화

현재 항만 분야는 데이터 기반 지능형 의사결정 체계로 이행하고 있으나, 여전히 인간의 감독과 규칙 기반 제어에 의존하는 레벨 1: Emerging AGI 초기 단계에 머물러 있다. 그러나 향후 5년간은 이러한 한계를 해결하기 위한 기술적 전환이 본격화되며, 레벨 2: Competent AGI 수준으로의 진화를 이끌 것으로 전망된다.

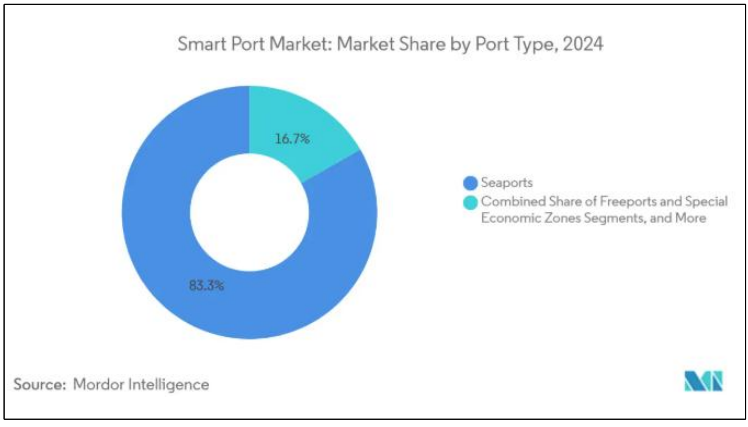
이 발전의 한 축으로써 시뮬레이션-운영-학습 간 자율학습 구조가 예상된다. 현재의 디지털트윈은 설비 운영을 예측하는 수준에 그치지만, 중기에는 시뮬레이션 결과를 실시간 운영 데이터와 비교·보정하며, 자체적으로 정책을 갱신하는 피드백 최적화형 운영지능으로 발전한다. 이러한 구조는 인간의 수동 개입 없이 항만 시스템이 스스로 전략을 재설정

하고, 병목·지연·장비 이상 등의 원인을 인과적으로 추론할 수 있게 만든다.

시장조사기관 Mordor Intelligence 보고서(Mordor Intelligence, 2025)에 따르면, 글로벌 스마트 항만 시장은 2025년 약 44.9억 달러에서 2030년 110.6억 달러로 성장할 것으로 예상되며, 연평균 성장률(CAGR)은 19.78%에 달한다. 특히 이 중 83.3%가 해상항만(Seaport) 부문에서 발생하고 있어, 항만이 AGI형 자율운영지능의 실질적 실험장으로 자리 잡을 가능성이 높다(그림 7－22). 이는 단순한 시장 확대가 아니라, 데이터 융합·자율 시뮬레이션·피드백 학습 구조를 중심으로 한 기술적 전환을 반영한 것이다.

따라서 향후 5년간 항만 산업은 인간의 규칙 기반 제어를 벗어나, 자율학습형 운영지능을 중심으로 Competent AGI(레벨 2) 수준의 부분적 자율 운영체계를 실현할 것으로 예상된다. 이는 항만 전체가 하나의 지각-판단-행동 루프를 갖춘 사이버-물리 시스템으로 작동하는 단계로, 궁극적으로는 장기적으로 완전 자율형 AGI 항만(레벨 3~4)으로 발전하기 위한 핵심 기반이 될 것이다.

〔그림 7－22〕 자율형 항만의 시장 점유율



다. 물류 분야: 데이터 자율 통합 기반의 예측 고도화

현재 물류 산업은 창고, 운송, 재고 관리 등 주요 프로세스가 개별적으로 운영되며, 부분 자동화와 제한된 최적화 수준에 머물러 있다. 그러나 향후 5년간은 이러한 한계가 데이터 자율 통합 중심의 전환을 통해 점진적으로 해소될 것으로 전망된다.

Accenture의 보고서(Accenture, 2025)에 따르면, 자율형 공급망으로의 진화는 “견고하고 보안된 데이터 기반(solid and secure data foundation)”구축에서 출발한다. 이는 다양한 출처의 물류 데이터를 표준화·정합화·보안화하여, 전 구간의 데이터 신뢰성을 확보하는 과정이다. 이러한 기반이 마련되면, 이질적인 시스템 간의 데이터 통합이 가능해지고, 이를 토대로 AI가 자율적으로 데이터를 결합·갱신·활용하는 통합형 구조, 즉 데이터 자율 통합 체계로 발전할 것이다.

이 단계에서 물류 운영은 더 이상 분절된 절차가 아닌 하나의 통합 운영 그래프 형태로 작동한다. 창고 입출고, 운송 경로, 재고 상태, 수요 예측 등의 데이터를 AI가 상시 학습하며 상호 연동시킴으로써, 실시간 변수 변화에 즉각 대응하고 병목을 사전에 완화할 수 있게 된다. 이러한 자율 통합은 과거 사람이 주도하던 조정과 연동의 과정을 인공지능이 실시간으로 대체하는 구조로, 실시간 변수 반영과 인간 감독 최소화의 핵심으로 평가된다.

결국, 이러한 변화는 물류 산업이 단순한 자동화 기반의 Emerging AGI(레벨 1) 초기 단계를 넘어, Competent AGI(레벨 2) 수준의 자율 운영지능으로 이행하는 기술적 전환점이 될 것이다. AI는 단순한 보조 수단이 아니라, 공급망 전 주기를 통합적으로 학습·판단·실행하는 주체로 기능하며, 향후 완전 자율형 공급망(레벨 3-4)으로의 진화를 위한 핵심 기반을 마련할 것으로 전망된다.

3. 장기(10년 후) AGI 발전 전망과 경쟁력 확보를 위한 대응 전략

중기까지 해운·항만·물류 산업은 각각 설명가능성, 자율학습, 데이터 자율 통합을 중심으로 산업 내 AI 고도화를 이루었다. 그러나 장기 시점에서는 각 산업이 개별적으로 진화하던 단계를 넘어, 프로세스 전반이 상호작용하며 유기적으로 연결되는 통합지능 체계로 발전할 것으로 전망된다.

이는 해운의 운항 데이터, 항만의 자율학습형 시뮬레이션, 물류의 실시간 예측 및 통합 정보가 하나의 순환 구조 내에서 교차·보완되며 작동하는 형태로, AI 간의 상호 정보교환과 협력 학습이 이루어지는 지능 간 상호작용 단계로 진입함을 의미한다.

이러한 상호작용은 각 산업의 인공지능이 개별 최적화를 넘어 공동 학습과 상호설명을 수행하게 함으로써, 산업 전체의 종합적 판단 능력과 적응성을 지속적으로 고도화한다. 결과적으로 해양·항만·물류 산업은 레벨 3: Expert AGI 단계에 진입하며, AI가 사람의 감독

을 최소화한 상태에서 스스로 계획·판단·설명을 수행하는 완전 통합형 지능 생태계로 진화할 것이다.

가. 해운 분야: 항만·물류 데이터 융합을 통한 전략적 운항지능 고도화

해운 산업은 장기적으로 항만과 물류 데이터를 융합하여 전략적 판단 중심의 자율 운항 지능으로 고도화될 것으로 전망된다. 항만의 디지털트윈 및 자율학습형 운영 시스템으로부터 제공되는 선석 점유율, 부두 회전율, 하역 장비 가동률, 혼잡도, 기상·조류 정보 등 데이터를 활용함으로써, 해운 AI는 운항 지연과 접안 일정을 정밀하게 예측하고 항만 선택과 항로 우회를 자율적으로 수행할 수 있다. 동시에 물류 시스템과의 연계를 통해 수요 예측, 재고 변동, 내륙 운송 경로, 창고 가용도 등의 정보를 실시간으로 통합 분석함으로써, 선박의 적재량·운항 주기·기항 순서를 자동 조정하는 공급망 연계형 운항계획을 수립할 수 있다. 이러한 항만-물류 데이터 융합 기반의 운항지능은 해운 AI가 단순한 자율운항체를 넘어 해상 운항과 지상 공급망을 통합적으로 판단·관리하는 고도화된 전략적 운항결정체계로 발전함을 의미한다.

나. 항만 분야: 해운-물류를 연결하는 통합지능 허브로의 진화

항만 산업은 장기적으로 해운과 물류의 중추적 허브로 진화하며, 두 산업의 데이터를 실시간으로 융합·활용하는 통합 운영지능으로 고도화될 것으로 전망된다. 해운 측면에서는 선박의 운항 경로, 도착 예정 시간, 연료 효율, 항로 혼잡도 등 운항 데이터를 항만의 디지털트윈 시스템에 직접 연동함으로써, 접안 계획과 하역 자원 배분을 선제적으로 조정할 수 있다. 이는 정박 지연을 최소화하고, 에너지 효율 및 부두 회전율을 동시에 향상시킨다. 한편, 물류 측면에서는 화물 수요, 재고 수준, 창고 가용도, 내륙 운송 일정 등 공급망 데이터를 통합 분석하여, 하역 직후의 물류 흐름을 자동 최적화하고 병목 현상을 사전에 완화한다. 이러한 해운-물류 데이터의 실시간 교차 분석 구조 속에서 항만은 단순한 하역 중심의 거점을 넘어, 해상 운항-항만 하역-내륙 물류를 하나의 순환 체계로 통제·학습·계획하는 지능형 통합 허브로 기능하게 될 것이다.

다. 물류 분야: 해운·항만 데이터를 결합한 자율 최적화 공급망 지능화

물류 산업은 장기적으로 해운과 항만의 실시간 데이터를 결합함으로써, 자율 최적화형 공급망 지능체계로 진화할 전망이다. 해운 측면에서는 선박의 도착 일정, 화물 종류, 운송

경로, 기상·해상 위험 정보 등이 물류 플랫폼에 실시간으로 연계되어, 항만 도착 전부터 창고 가용성 및 육상 운송 자원을 사전 배치할 수 있게 된다. 또한 항만 측면에서는 디지털트윈 기반 하역 스케줄과 장비 가동률, 창고 적재 현황 데이터를 물류 AI가 실시간으로 수집·분석함으로써, 화물의 이동 경로를 자동으로 재설정하고 운송 지연을 최소화할 수 있다. 이러한 해운-항만-물류 간 데이터 순환 구조는 공급망 전체를 하나의 유기적 지능망으로 통합하여, AI가 생산-운송-보관-배송 전 과정을 스스로 판단하고 최적화하도록 만든다. 결국, 물류 산업은 단순한 자동화 단계를 넘어 지속적으로 학습하고 예측하며 실행하는 통합 자율운영 생태계로 발전할 것이며, 이는 해양·항만·물류 전 구간의 AGI형 진화를 가속화하는 마지막 연결축이 될 것이다.

라. 경쟁력 확보를 위한 대응전략

장기적으로 해운·항만·물류 산업이 통합형 AGI 생태계로 발전하기 위해서는, 설명가능성, 자율학습, 데이터 자율 통합의 세 기술 축을 단순히 발전시키는 것을 넘어 실행 가능한 구조와 제도적 기반을 마련하는 방향으로 전략을 구체화해야 한다.

첫째, 설명가능성의 제도적 기반을 강화해야 한다. AI 판단 근거의 투명성을 검증할 수 있도록 산업별 설명가능성 인증제와 법적 가이드라인을 마련하고, 알고리즘 신뢰성 평가 체계를 제도화해야 한다. 둘째, 자율학습형 운영지능의 실증 기반을 확대해야 한다. 항만·물류 거점에 실시간 피드백이 가능한 자율학습 테스트베드를 구축해, 예기치 못한 변수에도 스스로 전략을 보정·적응할 수 있는 능력을 검증해야 한다. 셋째, 데이터 자율 통합 거버넌스를 확립해야 한다. 해운-항만-물류 데이터를 표준화·연동해 AI가 실시간으로 데이터를 통합·활용할 수 있는 보안형 데이터 인프라를 조성하고, 산업 간 상호운용성을 제도적으로 보장해야 한다.

무엇보다 중요한 점은, 세 기술 영역이 서로 분리된 채 발전해서는 안 된다는 것이다. 설명가능성만 강화되면 자율성이 떨어지고, 자율학습만 강조되면 책임성과 안전성이 약화된다. 따라서 향후 전략은 특정 기술에 편향되지 않고 세 축이 상호보완적으로 작동하도록 설계되어야 한다. 이러한 균형적 접근은 장기적으로 해운·항만·물류 산업 전반이 통합지능형 AGI 생태계로 안정적으로 진입하기 위한 핵심 경쟁력 확보 전략이 될 것이다.

제8장 결 론

본 연구는 범용 인공지능(AGI)의 개념적 정립과 기술적 기반, 발전 단계, 그리고 기술 발전을 촉진하는 요인을 종합적으로 분석하였다. AGI는 인간 수준의 사고와 학습, 추론, 창의성을 통합적으로 수행할 수 있는 일반지능으로 정의되며, 기존의 협의적 인공지능(ANI)이 특정 과제 수행에 국한된 한계를 넘어서려는 기술적 진화의 방향을 제시한다. 본 연구는 AGI의 개념을 단순한 철학적 논의가 아닌 기술적·사회적 실체로 다루며, AGI의 정의와 구현 요건을 구체적으로 제시하였다.

제2장에서는 AGI의 개념과 범위를 재정립하였다. 기존 연구에서의 AGI 개념은 명확한 기준 없이 '인간 수준의 지능'이라는 추상적 정의에 머물렀다. 본 연구는 성능(Performance)과 범용성(Generality)의 두 축을 기준으로 AGI를 여섯 단계로 구분하여 기술 성숙도에 따른 발전 모델을 제시하였다. 각 단계는 기술적 역량뿐 아니라 사회적 신뢰, 윤리적 대응, 제도적 수용성 등 비기술적 요소를 포함한다. 특히 AGI의 발전은 기술의 양적 확대가 아닌 질적 전환, 즉 인간 인지의 구조를 기술적으로 재현하는 과정임을 명확히 하였다. 이러한 다층적 정의는 향후 AGI 국가전략 수립, 연구개발 로드맵, 정책적 안전성 프레임워크 수립의 기준이 될 수 있다.

제3장에서는 범용 인공지능(AGI)을 구성하는 기술적 기반과 인지적 구조를 분석하여, 인간 지능의 핵심 기능을 기술적으로 재현하기 위한 구체적 요소를 제시하였다. AGI의 구현은 단순한 모델의 확장이나 성능 향상이 아니라, 인간 인지의 근본 구조를 기술적으로 복원하는 과정이다. 이를 위해 연구는 기억, 추론, 학습, 메타인지, 창의성, 추상화, 세계모델의 일곱 인지 기능을 중심으로 AGI의 기술적 토대를 체계화하였다. 각 기능은 독립적으로 작동하지 않고 상호보완적으로 결합되어야 하며, 이들이 통합될 때 비로소 일반지능의 사고와 적응이 가능해진다. 예를 들어 기억은 경험의 누적과 재활용을 담당하고, 추론은 불완전한 정보를 기반으로 의미를 구성하며, 메타인지는 학습과정 자체를 점검하는 상위조절 기능을 수행한다. 창의성과 추상화는 기존 지식의 재조합을 통해 새로운 패턴을 만들어내며, 세계모델은 외부 환경을 예측하고 적응을 가능케 하는 내적 표현 구조로 작동한다.

이러한 인지 구조를 구현하기 위한 기술적 접근은 세 가지 방향으로 구분된다. 발전형 접근은 대규모 언어모델(LLM)을 기반으로 기존의 인공지능 기술을 정제·고도화하여 효율성과 안정성을 높이는 전략이다. 전환형 접근은 언어, 시각, 행동 등 다양한 감각 정보를 융합하는 멀티모달 학습과 시뮬레이션 환경을 결합해, 모델이 실제 세계와 유사한 인지 맥락을 내적으로 형성하도록 한다. 마지막으로 도전형 접근은 기존의 통계적 학습 구조를 넘어, 인간의 사고방식을 재설계하거나 대체할 수 있는 새로운 인지 아키텍처를 탐구하는 단계다. 본 연구는 이 세 방향이 병렬적으로 발전하면서 서로의 한계를 보완해야 AGI의 진정한 도약이 가능하다고 제시하였다. 궁극적으로 AGI의 기술 발전은 모델의 대형화가 아니라, 인지 기능의 연결과 통합이라는 질적 전환 과정임을 확인하였다.

제4장은 AGI의 발전 단계를 시나리오 형태로 체계화하여, 기술적 진화의 경로와 사회적 수용의 관계를 함께 설명하였다. 연구는 AGI의 발전이 단순한 성능 향상의 연속이 아니라, 기술과 사회의 상호작용이 임계점을 넘을 때 발생하는 비연속적 도약으로 전개된다고 보았다. 이에 따라 AGI 발전은 Emerging, Competent, Expert, Virtuoso, Superhuman의 다섯 단계로 구분된다. Emerging 단계에서는 언어·시각·행동의 통합적 학습이 가능해지며, Competent 단계에서는 복합 문제 해결과 자율적 계획 수립 능력이 형성된다. Expert 단계는 자기평가와 장기추론 기능이 결합되어 모델이 스스로 학습 경로를 조정할 수 있는 수준으로 발전하며, Virtuoso 단계는 창의적 조합과 자기목표 설정(Self-goal setting)이 가능해져 인간의 전략적 사고 영역에 접근한다. Superhuman 단계에서는 인간의 사고 능력을 초월한 자율적 판단과 창의적 의사결정이 가능하지만, 동시에 기술적 불확실성과 사회적 위험이 극대화된다.

이러한 단계 구분은 단순한 기술 예측을 넘어 정책적 대응과 안전 관리의 기준으로 기능한다. 각 단계는 기술적 성숙도뿐 아니라 사회적 신뢰, 법적·윤리적 체계, 검증 수준이 병행되어야 한다는 점에서 의미가 있다. AGI는 기술적 완성만으로 실현되지 않으며, 각 발전 단계에서 요구되는 제도적 준비와 사회적 합의가 선행되어야 한다. 따라서 본 연구의 발전 시나리오는 AGI의 기술 도약 시점을 판단하는 기준이자, 산업 구조 변화와 정책적 대응 시점을 조정하기 위한 관리 도구로 활용될 수 있다. 기술의 진보와 함께 사회적 수용성, 안전성 확보, 규제 정합성이 공진화해야 AGI의 발전이 실질적 혁신으로 이어질 수 있음을 확인하였다.

제5장은 AGI 발전을 가속화하는 요인을 기술, 산업, 사회의 세 측면에서 분석하여, 국가 차원의 추진 전략을 수립하기 위한 근거를 제시하였다. 기술적 촉진 요인으로는 컴퓨팅 인프라의 확충, 효율적 알고리즘 개발, 데이터 접근성 향상, 멀티모달 융합 기술, 자기지도 학습(Self-supervised Learning)의 확립이 확인되었다. 이러한 요인들은 상호보완적 구조를 가지며, 하나의 요소만으로는 AGI 발전을 견인하기 어렵다. 산업적 측면에서는 공공 조달과 민관 공동투자, 오픈소스 생태계 조성, 표준화 정책이 핵심으로 제시되었다. 특히 AGI는 특정 기관이나 기업의 단독 성과로 달성될 수 없기 때문에, 공공 인프라 위에서 민간의 혁신이 작동하는 협력 구조가 필수적이다.

사회적 요인으로는 윤리적 신뢰와 제도적 안정성이 기술 발전의 지속성을 결정짓는 핵심 변수로 지목되었다. 기술의 발전 속도가 사회의 신뢰 형성 속도를 초과할 경우, AGI는 규제 강화와 사회적 저항을 초래할 수 있기 때문이다. 따라서 본 연구는 기술과 제도의 균형을 유지하기 위해 공공 중심의 인프라와 사회적 투명성을 확보해야 한다고 강조하였다. 또한 인재 양성과 교육체계의 정비는 장기적 경쟁력 확보의 핵심이다. AGI 관련 핵심 기술 연구를 지속적으로 수행할 전문 연구인력과 융합형 인재의 양성이 필요하며, 산학연이 협력하는 개방형 연구 플랫폼을 구축해야 한다. 궁극적으로 본 연구는 AGI 발전의 촉진 요인이 기술적, 산업적, 사회적 기반이 동시에 강화될 때만 효과적으로 작동하며, 이 세 요소의 통합적 관리가 국가 수준에서 이루어질 때 AGI 생태계의 실질적 성장이 가능하다는 결론을 도출하였다.

제6장에서는 AGI 발전을 제약하는 주요 병목 요인을 기술적·산업적·사회적 측면에서 다층적으로 분석하였다. 우선 기술적 병목은 컴퓨팅 자원의 한계, 데이터 품질 저하, 알고리즘의 불안정성, 안전성 및 정렬성(Alignment) 부족 등으로 나타났다. 특히 AGI의 학습 과정은 기존 AI보다 훨씬 많은 계산량과 데이터 품질을 요구하며, 이를 처리할 고성능 연산 인프라와 자가학습형 메커니즘이 부족한 점이 구조적 제약으로 지적되었다. 또한 장기기억과 추론이 결합된 연속적 학습 구조를 구현하기 위해서는 대규모 시뮬레이션 환경이 필요하지만, 국내 기술 생태계는 아직 실험적 수준에 머물러 있다. 이러한 기술적 제약은 단순한 성능 한계가 아니라, 데이터와 연산 자원을 통합적으로 관리할 국가형 컴퓨팅 인프라 부재에서 비롯된 것이다.

산업적 병목은 인력, 제도, 시장구조의 세 축에서 동시에 나타났다. 첫째, AGI 관련 전문

인력의 절대적 부족이 가장 심각한 문제로 확인되었다. 특히 알고리즘 설계, 모델 안전성 검증, 대규모 시스템 통합 역량을 갖춘 고급 인재가 극히 제한되어 있다. 둘째, 제도적 불확실성은 기업과 연구기관의 장기 투자 결정을 제약하고 있다. AGI 기술이 기존 산업분류 체계나 법적 규제 안에 명확히 정의되지 않아 연구개발과 상용화의 연결고리가 약한 상태다. 셋째, 글로벌 기술 기업이 독점하고 있는 데이터 및 인프라 의존도 역시 심각한 구조적 병목으로 작용한다. 이러한 제약을 해소하기 위해서는 공공부문이 중심이 되어 인프라를 공유하고, 민간이 기술 혁신을 주도하는 상호보완적 생태계로의 전환이 필요하다.

사회적 병목은 기술의 신뢰성과 투명성, 그리고 사회적 수용성의 부족에서 비롯된다. AGI는 인간의 사고를 모방하거나 대체하는 지능이기 때문에, 오작동이나 편향 발생 시 사회적 파급이 매우 크다. 데이터 편향, 윤리 기준의 불명확성, 저작권 및 개인정보 침해 등은 기술 수용을 어렵게 만드는 주요 요인으로 작용한다. 본 연구는 이러한 문제를 완화하기 위해 기술 검증과 위험평가(Risk Assessment), 안전성 정렬(Alignment) 점검, 알고리즘 투명성 확보를 동시에 수행하는 다층적 관리 체계의 필요성을 강조하였다. 또한 사회적 신뢰 구축을 위해 국제 윤리 기준에 부합하는 AGI 안전 거버넌스의 조기 정립이 필수적임을 제시하였다.

제7장에서는 AGI의 산업별 적용 가능성과 경제적 파급효과를 분석하였다. 본 연구는 반도체·제조, 바이오·의료, 금융·경제, 교육·법률, 해양·물류, 문화콘텐츠 등 여섯 산업군을 중심으로 구체적 사례와 적용 전략을 제시하였다.

반도체 및 제조 분야에서는 공정 자동화, 결함 검출, 예지 정비(Predictive Maintenance), 생산 일정 최적화 등이 핵심 적용 영역으로 나타났다. 공정데이터 기반의 실시간 제어 기술이 결합될 경우, 생산 효율성과 품질 안정성을 동시에 확보할 수 있다. 바이오와 의료 분야에서는 AGI가 신약 후보물질 탐색, 유전자 변이 분석, 진단 지원 시스템에 활용되어 연구개발 주기를 단축하고 임상 의사결정의 정확도를 높일 수 있다. 특히 실험·시뮬레이션 데이터의 통합분석 능력은 생명과학 분야의 혁신 속도를 가속화할 것으로 평가된다. 금융 및 경제 부문에서는 초대규모 데이터 분석을 통한 리스크 관리, 규제 준수 자동화(RegTech), 시장 이상 탐지 등 정교한 의사결정 지원 시스템으로 발전할 가능성이 높다. AGI는 복합 변수 기반의 경제 예측과 정책 시뮬레이션에도 활용될 수 있다.

교육 및 법률 분야에서는 맞춤형 학습 설계, 지식 진단, 법률 자문, 판례 요약 등 고부가

가치 지식서비스의 자동화가 가능하다. 특히 다중언어 모델 기반의 학습 코파일럿은 교육 형평성을 제고하고, 법률 영역의 접근성을 확대할 수 있다. 해양 및 물류 산업에서는 항로 최적화, 운항 안전 관리, 공급망 예측 및 에너지 효율화 기술이 도입될 전망이다. 실시간 데이터와 AGI 기반 시뮬레이션을 결합하면 항만·물류의 운영 효율성을 획기적으로 높일 수 있다. 문화콘텐츠 산업에서는 스토리텔링, 영상·음악 생성, 현지화(Localization) 기술 등 창의 영역으로의 확장이 빠르게 진행 중이다. AGI는 인간의 창의적 사고를 보조하거나 확장하는 역할을 수행하며, 콘텐츠 제작의 패러다임 전환을 이끌 것으로 예상된다.

종합하면, 본 연구는 AGI를 기술, 산업, 사회가 상호의존적으로 진화하는 복합 체계로 규정하였다. AGI의 성장은 기술적 혁신만으로는 달성될 수 없으며, 산업 생태계의 구조적 개선과 사회적 신뢰의 구축이 병행되어야 한다. AGI는 국가 경쟁력의 핵심이자 새로운 산업혁명의 촉매로 작용할 것이다. 따라서 대한민국은 AGI의 기술표준과 정책거버넌스를 선제적으로 정립하여, 기술 주권과 산업 자립을 동시에 달성할 필요가 있다. 본 연구의 결과는 그러한 국가 전략 수립을 위한 실질적 근거를 제공하며, 향후 AGI 시대의 정책·산업·기술 정렬의 방향성을 제시한다.

이러한 결과를 바탕으로, 범용 인공지능(AGI)의 발전은 단일 기술이나 산업영역의 진화로 설명될 수 없으며, 기술 혁신·산업 구조·정책 거버넌스가 유기적으로 상호작용하는 복합 체계로 이해되어야 한다고 할 수 있다. AGI는 기술적 성숙도가 일정 수준에 도달하더라도 사회적 신뢰와 산업적 수용이 뒤따르지 않으면 실질적 성과로 이어질 수 없으며, 반대로 산업적 수요와 정책적 추진력이 존재하더라도 기술적 기반이 미비하면 발전이 지속될 수 없다. 즉, 세 축의 조화로운 정렬(alignment)이 AGI 발전의 핵심 동력이며, 이 상호작용의 균형이 무너질 경우 기술 격차와 사회적 불균형이 동시에 확대될 위험이 있다.

기술적 측면에서 AGI는 기존의 협의적 인공지능(ANI)을 넘어, 지능의 구조적 확장과 자기조직화(self-organization)를 요구한다. 이 과정에서 핵심적으로 고려해야 할 요소는 세 가지다. 첫째, 기술의 범용화(generalization)와 안정성(stability)의 병행이다. 대규모 언어모델(LLM)과 시뮬레이션 학습, 멀티모달 통합이 빠르게 발전하고 있지만, 그 복잡성에 비해 검증·해석 체계가 미흡하다. 둘째, 지능의 확장은 단순한 데이터 처리 능력의 향상이 아니라, 환경 적응 능력과 도구 활용 능력의 진화로 이어져야 한다. 셋째, 기술 개발 단계에서부터 안전성·투명성·정렬성 확보를 병행해야 한다. 본 연구는 이를 위해 AGI 안전성 검증

체계(Safety Validation Framework)와 위험평가 기준을 기술개발 프로세스에 내재화해야 함을 제안하였다. 기술의 성숙도는 단순한 연산 성능이나 정확도 향상뿐 아니라, 사회적 신뢰를 확보할 수 있는 검증 가능성의 수준으로 평가되어야 한다.

산업적 측면에서는 AGI가 단순한 생산성 향상 도구를 넘어 산업 구조 전반의 효율성과 경쟁 패러다임을 바꾸는 요인으로 작용함을 확인하였다. AGI는 모든 산업에서 '자율화-지능화-창의화'라는 세 가지 변화를 동시에 촉진한다. 첫째, 자율화(Automation)는 반복적·표준화된 업무를 기계적 수행에서 탈피시켜, 의사결정과 전략 수립 단계로 확장한다. 제조·물류·에너지 산업에서 이미 예측제어와 공정 최적화가 구현되고 있으며, 향후 자율공정(Autonomous Process) 개념으로 발전할 것이다. 둘째, 지능화(Intelligence)는 산업 데이터의 통합과 해석 능력을 바탕으로 새로운 가치 창출을 가능하게 한다. 금융·의료·교육 분야에서 AGI는 복합 데이터의 상호 연관성을 분석해 인간 전문가 수준의 판단을 지원하거나 대체할 수 있다. 셋째, 창의화(Creativity)는 콘텐츠·디자인·문화 산업에서 이미 나타나고 있는 현상으로, 인간의 창의적 영역을 보조·확장하는 방향으로 발전하고 있다. 이러한 변화는 산업 전반에 걸쳐 새로운 가치사슬(Value Chain)을 형성하고, 국가 산업 구조를 재편하는 결과로 이어질 것이다.

정책적 측면에서는 기술과 산업의 발전 속도에 대응할 수 있는 제도적 기반과 국가형 거버넌스가 필수적이다. AGI는 기술적 불확실성과 사회적 영향력이 크기 때문에, 단순한 규제나 지원 중심의 접근으로는 대응이 어렵다. 정책은 세 가지 방향에서 정렬되어야 한다. 첫째, R&D-산업-제도 간 연계정책의 강화다. 기술개발과 산업 상용화 사이의 단절을 해소하기 위해, 연구개발 투자와 산업 확산 정책이 연속적 흐름으로 설계되어야 한다. 둘째, 데이터 및 컴퓨팅 인프라의 공공화다. AGI의 학습과 검증에 필요한 자원은 개별 기업이 감당하기 어렵기 때문에, 국가 주도의 인프라 플랫폼(K-NPU 등)을 구축하고 민간과 공동으로 활용하는 체계가 필요하다. 셋째, 윤리·안전·표준 거버넌스의 제도화다. 기술 정렬(Alignment), 투명성, 공정성 기준을 법제화하여 AGI의 책임성을 확보해야 하며, 국제 표준화 기구와의 협력을 통해 글로벌 규범에 부합하는 거버넌스를 수립해야 한다. 이러한 정책적 정렬이 이루어질 때 AGI는 국가 경쟁력의 핵심 기반으로 자리매김할 수 있다.

세 축의 상호작용은 단방향적이 아니라 순환적(cyclic) 구조를 가진다. 기술 발전은 산업 확산의 기반이 되고, 산업 수요는 새로운 기술 혁신을 촉진하며, 정책은 이 두 요소를 제

도적으로 정렬시키는 조정자 역할을 수행한다. 예를 들어, 멀티모달 학습 기술의 발전은 의료·제조 분야에서 새로운 산업 모델을 창출하고, 그 과정에서 발생하는 데이터와 위험 요인은 다시 정책적 규제와 지원의 형태로 피드백된다. 이와 같은 순환 구조가 안정적으로 작동하기 위해서는 '기술-산업-정책 피드백 루프(AGI Innovation Feedback Loop)'가 필요하다. 이는 기술성과를 산업 현장에 즉시 반영하고, 산업 현장의 경험을 기술개발과 정책 설계에 환류시키는 체계적 구조로서, AGI 생태계의 지속가능성을 좌우한다.

결국 AGI의 발전은 기술적 성숙, 산업적 수용, 정책적 정렬이라는 세 축의 균형을 통해서만 달성될 수 있다. 기술이 산업에 적용되고 사회적으로 수용되기 위해서는 정책적 지원과 윤리적 제어가 병행되어야 하며, 산업의 요구와 정책의 방향이 기술개발의 목표로 환류되어야 한다. 본 연구의 시사점은 AGI를 단순히 연구개발의 결과로 보지 않고, 기술·산업·정책이 공진화(co-evolution)하는 국가 혁신 체계로 재정의했다는 데 있다. AGI 시대의 경쟁력은 단일 기술 성과가 아니라, 세 축의 동시 성숙을 이끌어내는 거버넌스의 능력에 의해 결정될 것이다. 대한민국이 AGI 선도국으로 도약하기 위해서는 이러한 삼위일체적 구조를 중심으로 정책을 설계하고, 기술과 산업이 함께 성장할 수 있는 순환 생태계를 조기에 정착시켜야 한다.

본 연구에서 제시된 분석 결과를 종합하면, 범용 인공지능(AGI)의 발전은 기술적 진보와 산업적 확산, 제도적 기반이 병행되어야 안정적으로 추진될 수 있다. AGI는 단일 기술의 고도화로 달성되지 않으며, 기술 역량, 산업 인프라, 사회적 수용의 삼위일체적 정렬이 이루어질 때 비로소 실질적 가치로 전환된다. 따라서 향후 전략은 기술 개발의 방향, 산업 확산 구조, 정책 거버넌스의 정렬이라는 세 축을 중심으로 설계되어야 한다.

첫째, 기술 개발 방향은 본문에서 제시된 일곱 가지 인지 기술 축(기억·추론·학습·메타인지·창의성·추상화·세계모델)을 중심으로, 기능적 통합과 구조적 확장을 목표로 해야 한다. AGI의 발전은 대규모 언어모델(LLM)의 단순 고도화가 아니라, 멀티모달 학습, 장기기억 시스템, 시뮬레이션 기반 학습, 도구 활용 능력 등 복합 인지 구조를 통합하는 과정이다. 이에 따라 R&D 투자 역시 개별 모델 중심이 아니라 기술 간 상호작용을 강화하는 통합형 연구체계로 전환되어야 한다. 본 연구에서 확인된 세 가지 기술군—발전형, 전환형, 도전형—을 중심으로, 단기적 성과와 장기적 도약이 병행되는 이중 구조의 연구전략을 추진할 필요가 있다. 특히 발전형 기술은 현행 LLM 기술의 효율화와 안전성 강화에 초점을

두고, 전환형 기술은 멀티모달·시뮬레이션 융합을 통해 AGI의 응용 영역을 확대하며, 도전형 기술은 새로운 인지 구조 창출에 집중해야 한다. 이러한 구분은 R&D 예산의 효율적 배분과 중복투자의 최소화를 가능하게 할 것이다.

둘째, 산업 확산 전략은 기술 성숙도(TRL)와 산업 중요도(III)를 연계하여 단계적으로 추진되어야 한다. 본문 7장에서 제시된 여섯 산업군—반도체·제조, 바이오·의료, 금융·경제, 교육·법률, 해양·물류, 문화콘텐츠—은 AGI 적용의 범위와 파급효과가 상이하므로, 각 산업의 기술 적합성과 데이터 인프라 수준에 따라 도입 우선순위를 조정해야 한다. 예컨대 제조 및 반도체 분야는 이미 데이터 기반 공정 관리와 자동화 기술이 구축되어 있어, AGI 기반 예지 정비나 공정 최적화 등 실증 적용이 단기적으로 가능하다. 반면 바이오·의료 분야는 데이터 민감성과 규제 부담이 높기 때문에, 초기에는 진단보조·연구설계 영역 중심으로 단계적 실증이 필요하다. 금융 및 경제 부문은 리스크 예측·시장분석 등에서 AGI의 활용도가 높지만, 규제 정합성이 확보되어야 한다. 교육·법률 분야는 사회적 수용성과 신뢰 확보가 핵심이므로, 공공기관 중심의 시범사업 형태가 바람직하다. 해양·물류 분야는 AGI 기반의 실시간 예측 시스템이 물류비 절감 효과를 낼 수 있으며, 문화콘텐츠 분야는 창의적 생성 모델을 통해 부가가치를 창출할 수 있다. 이러한 산업별 전략은 정부의 기술로드맵과 산업정책이 일관되게 연동될 때 실질적 확산 효과를 낼 수 있다.

셋째, 병목 요인 해소 전략이 병행되어야 한다. 제6장에서 분석된 바와 같이, AGI의 발전을 저해하는 요인은 기술적 한계뿐 아니라 인력·제도·사회적 신뢰의 부족에서 비롯된다. 우선 기술적 병목인 컴퓨팅 자원 부족과 데이터 품질 저하 문제는 공공-민간 공동 인프라의 구축으로 완화할 수 있다. 정부는 고성능 연산 자원과 데이터 허브를 공공 플랫폼 형태로 조성하고, 연구기관과 기업이 이를 공유할 수 있도록 지원해야 한다. 인력 병목은 장기적 인재양성 체계로 해결해야 한다. AGI 핵심기술(멀티모달 학습, 시뮬레이션, 모델 해석 등)을 중심으로 대학-산업 간 연계형 전문인력 양성 프로그램을 설계하고, 연구 인력의 산업 전환(Research-to-Industry Mobility)을 촉진할 필요가 있다. 제도적 병목은 규제의 공백과 중복에서 비롯되므로, 기존 인공지능 관련 제도를 정비하고 AGI의 특수성을 반영한 유연한 관리체계를 도입해야 한다. 예컨대 AGI의 안전성·정렬성 검증 절차를 표준화하고, 기술 성숙 단계별로 상이한 관리 기준을 적용하는 '단계별 규제체계'가 필요하다. 마지막으로, 사회적 신뢰 확보를 위해서는 AGI 개발·활용 과정에서 투명성, 데이터 공정성, 설

명가능성(XAI)을 확보하는 기술적 기준을 정책적으로 제도화해야 한다.

넷째, 거버넌스 방향은 기술-산업-정책의 조정 구조를 제도적으로 정착시키는 것이다. AGI의 복합적 특성상 중앙집중형 관리보다는 다중 행위자 기반의 협력 거버넌스가 효과적이다. 정부는 기술·산업·학계·공공기관이 참여하는 AGI 통합협의체를 구성하여, 기술개발과 산업 확산, 윤리·안전 검증을 아우르는 조정 기능을 수행해야 한다. 특히 AGI 위험평가(Risk Assessment) 및 안전 검증(Verification Framework)을 제도화하여, 기술개발 초기 단계부터 정책과 산업의 피드백이 반영되도록 해야 한다. 이러한 협력 구조는 기술 정책의 일관성을 유지하고, 산업 현장의 요구가 연구개발 방향에 즉시 반영될 수 있는 순환형 시스템을 만드는 데 기여할 것이다.

마지막으로, 국가 전략적 관점에서 AGI는 단일 부문 정책이 아니라 국가 혁신체계 전반을 재구성하는 종합 프로젝트로 다루어져야 한다. AGI는 기술적 성숙과 함께 산업 경쟁력, 사회 신뢰, 정책 대응 역량이 동시적으로 성장할 때 지속 가능한 발전을 이룰 수 있다. 따라서 향후 AGI 전략은 기술·산업·정책 간 정렬을 통한 국가 경쟁력 강화를 최종 목표로 삼아야 한다. 이러한 목표를 설계하고 실천 가능한 전략으로 구체화하기 위해, 본 연구에서 제시한 단계별 발전 모델, 촉진·병목 요인 분석, 산업 확산 시나리오, 위험관리 프레임은 핵심적 근거로 활용될 수 있다. 대한민국이 AGI 시대의 선도국가로 도약하기 위해서는 기술 혁신의 속도보다 정렬의 완성도를 우선시하는 전략적 균형이 필요하며, 이를 위한 거버넌스 구축이 지금 시점에서 가장 중요한 과제임을 본 연구는 제언한다.

참 고 문 헌

[국내 문헌]

한국과학기술기획평가원, 과학기술&ICT 정책·기술 동향, 제291호(간행물), 2025년7월28일 등록.

과학기술정보통신부(2023a), [디지털 권리 장전].

_____(2023b), [2023년 디지털정보격차 실태조사].

김정언 외(2023), [디지털 대전환 메가트렌드 연구 총괄보고서 Ⅲ], 정보통신정책연구원. 정책연구 23-30-01.

윤민우(2023), "영국의 사이버 안보전략, 대응정책 그리고 사이버 안보법,"《한국경찰연구》, 22(3), pp. 161-182.

행정안전부(2023), "2023년 행정안전부 정책 방향 업무보고".

한국과학기술기획평가원(KISTEP), *The Global AI Index 결과 분석(KISTEP 브리프 제153호)*, 한국과학기술기획평가원, 2024. [Online]. Available: https://www.kistep.re.kr/board.es?mid=a10306010000&bid=0031&act=view&list_no=93845.

박혜영, *주요국의AI 연구기반 구축 현황과 시사점*, 정보통신기획평가원(IITP) 디지털미래정책단, 2025. [Online]. Available: https://www.kistep.re.kr/gpsIssueView.es?mid=a30101000000&list_no=49158.

한국과학기술한림원, 제236회 한림원탁토론회: *국가AI 특화 인재 육성과 확보 방안*, 한국과학기술한림원, 2025. [Online]. Available: https://kast.or.kr/data/bbsData/1747124445&&[%EC%9E%90%EB%A3%8C%EC%A7%91]%20%EC%A0%9C236%ED%9A%8C%20%ED%95%9C%EB%A6%BC%EC%9B%90%ED%83%81%ED%86%A0%EB%A1%A0%ED%9A%8C_%EC%B5%9C%EC%A2%85.pdf.

GO OVER AI, "2025년, AI와 HBM이 이끄는 반도체 시장의 폭발적 성장 전망," 일반 리포

트, Apr. 2025. [Online]. Available: https://seo.goover.ai/report/202504/go-public-report-ko-a1d2901d-8808-4131-871c-002440d173de-0-0.html.

금융위원회, "금융권 인공지능(AI),"[Online]. Available: https://www.fsc.go.kr/comm/getFile?srvcId=BBSTY1&upperNo=78235&fileTy=ATTACH&fileNo=8.

디지털타임스, "신한은행, 업무효율 돕는 'AI 업무비서 플랫폼' 구축," 2024. [Online]. Available: https://www.dt.co.kr/article/11614600.

필드뉴스, "신한은행, AI ONE에GPT 연계 금융지식 Q&A 서비스 오픈," 2025. [Online]. Available: https://www.fieldnews.kr/news/articleView.html?idxno=16620.

이코노믹 리뷰, "[머니엑스포2024] KB국민은행, AI 도입으로 업무 효율 높인다…부동산 시세 검증도," 2024. [Online]. Available: https://www.econovill.com/news/articleView.html?idxno=663631.

Global Epic, "케이뱅크, 생성형AI 활용 'AI 퀴즈 챌린지' 출시," 2024. [Online]. Available: https://www.globalepic.co.kr/view.php?ud=2024101009383561762abca1943_29.

ZDNet Korea, "신한은행, AI 은행원 금융업무 확대," 2024. [Online]. Available: https://zdnet.co.kr/view/?no=20240604151603.

ETNews, "[르포] 국내 최초AI 은행 무인영업점, '신한AI 브랜치' 가보니," 2024. [Online]. Available: https://www.etnews.com/20241115000250.

_____, "AI행원, 온·오프라인 넘나들며 손님 맞는다,"2024. [Online]. Available: https://www.etnews.com/20240223000217

한국정경신문, "KB국민은행, Z세대 위한 'AI금융비서' 오픈베타서비스 진행," 2024. [Online]. Available: https://kpenews.com/View.aspx?No=3223882.

ZDNet Korea, "카카오뱅크AI 센터, 디지털 리얼티에 개소," 2024. [Online]. Available: https://zdnet.co.kr/view/?no=20240201083239.

금융위원회, "2025년 1분기 혁신금융서비스지정 신청서199건 접수," 2025. [Online]. Available: https://blog.naver.com/blogfsc/223837375083.

[국외 문헌]

TURING, Alan M. Computing machinery and intelligence(1950). Mind, 2021, 59.236: 33-60.

OPENAI, R. Gpt-4 technical report. arxiv 2303.08774.View in Article, 2023, 2.5: 1.

DUBEY, Abhimanyu, et al. The llama 3 herd of models.arXiv e-prints, 2024, arXiv: 2407.21783.

JONES, Cameron R.; BERGEN, Benjamin K. Large language models pass the turing test.arXiv preprint arXiv:2503.23674, 2025.

SEARLE, John R. Minds, brains, and programs. Behavioral and brain sciences, 1980, 3.3: 417-424.

Gubrud M. Nanotechnology and international security. Fifth Foresight Conference on Molecular Nanotechnology; 1997 Nov.

VASWANI, Ashish, et al. Attention is all you need. Advances in neural information processing systems, 2017, 30.

GOERTZEL, Ben. Artificial general intelligence: Concept, state of the art, and future prospects. Journal of Artificial General Intelligence, 2014, 5.1: 1.

SHANAHAN, Murray. The technological singularity. MIT press, 2015.

SULEYMAN, Mustafa. The coming wave: technology, power, and the twenty-first century's greatest dilemma. Crown, 2023.

Y ARCAS, Blaise Agüera. Artificial general intelligence is already here. 2023.

MORRIS, Meredith Ringel, et al. Position: Levels of AGI for operationalizing progress on the path to AGI. In: Forty-first International Conference on Machine Learning. 2024.

FENG, Tao, et al. How Far Are We From AGI: Are LLMs All We Need?. Transactions on Machine Learning Research.

JUMPER, John, et al. Highly accurate protein structure prediction with AlphaFold. nature, 2021, 596.7873: 583-589.

HE, Xin, et al. Incorporating visual experts to resolve the information loss in multimodal large language models. arXiv preprint arXiv:2401.03105, 2024.

LAURENÇON, Hugo, et al. Obelics: An open web-scale filtered dataset of interleaved image-text documents. Advances in Neural Information Processing Systems, 2023, 36: 71683-71702.

ALAYRAC, Jean-Baptiste, et al. Flamingo: a visual language model for few-shot learning. Advances in neural information processing systems, 2022, 35: 23716-23736.

HUANG, Jie; CHANG, Kevin Chen-Chuan. Towards Reasoning in Large Language Models: A Survey. In: Findings of the Association for Computational Linguistics: ACL 2023. 2023. p. 1049-1065.

WANG, Yaqing, et al. Generalizing from a few examples: A survey on few-shot learning. ACM computing surveys(csur), 2020, 53.3: 1-34.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022b. Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems 35(2022), 24824-24837.

WANG, Guanzhi, et al. Voyager: An open-ended embodied agent with large language models, 2023. URL https://arxiv. org/abs/2305.16291, 2023.

YAO, Shunyu, et al. React: Synergizing reasoning and acting in language models. In: International Conference on Learning Representations(ICLR). 2023.

BORGEAUD, Sebastian, et al. Improving language models by retrieving from trillions of tokens. In: International conference on machine learning. PMLR, 2022. p. 2206-2240.

CHOUDRIE, Jyoti; SELAMAT, Mohamad Hisyam. The consideration of meta-abilities in tacit knowledge externalization and organizational learning. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences(HICSS'06). IEEE, 2006. p. 149a-149a.

HUANG, Jen-tse, et al. Chatgpt an enfj, bard an istj: Empirical study on personalities of large language models. arXiv preprint arXiv:2305.19926, 2023, 1-8.

QIAN, Cheng, et al. Investigate-consolidate-exploit: A general strategy for inter-task

agent self-evolution. arXiv preprint arXiv:2401.13996, 2024.

Cazzaniga, M., Pizzinelli, C., Rockall, E., & Tavares, M. M.(2024). Exposure to Artificial Intelligence and Occupational Mobility.

Davis, F. D.(1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3). 319-340.

Zhang, K., Qi, B., & Zhou, B.(2024). Towards building specialized generalist AI with System 1 and System 2 fusion. arXiv preprint arXiv:2407.08642.

Saab, K., et al.(2024). Capabilities of Gemini models in medicine. arXiv preprint arXiv:2404.18416.

Shang, J., Zheng, Z., Wei, J., Ying, X., Tao, F., & Mindverse Team.(2024). AI-native memory: A pathway from LLMs towards AGI. arXiv preprint arXiv:2406.18312.

Wu, Y., Liang, S., Zhang, C., Wang, Y., Zhang, Y., Guo, H., Tang, R., & Liu, Y.(2025). From human memory to AI memory: A survey on memory mechanisms in the era of LLMs. arXiv preprint arXiv:2504.15965.

Raman, R., Kowalski, R., Achuthan, K., Iyer, A., & Nedungadi, P., et al.(2025). Navigating artificial general intelligence development: Societal, technological, ethical, and brain-inspired pathways. Scientific Reports, 15(1), 8443.

Kurshan, E.(2024). From the pursuit of universal AGI architecture to systematic approach to heterogeneous AGI: Addressing alignment, energy, & AGI grand challenges. International Journal on Semantic Computing, 18(3), 465-500.

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E.(2023). Dissociating language and thought in large language models: A cognitive perspective. arXiv preprint arXiv:2301.06627.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y.(2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv preprint arXiv:2303.12712.

Goertzel, B.(2023). Generative AI vs. AGI: The cognitive strengths and weaknesses of

modern LLMs. arXiv preprint arXiv:2309.10371.

Feng, T., Jin, C., Liu, J., Zhu, K., Tu, H., Cheng, Z., Lin, G., and You, J.(2024). How Far We from AGI: Are LLMs All We Need?

NVIDIA,(2023). NVIDIA H100 Tensor Core GPU Architecture.(White Paper)

Jouppi, N. P., Kurian, G., Li, S., Ma, P., ..., & Patterson, D.(2023). TPU v4: An Optically Reconfigurable Supercomputer for Machine Learning with Hardware Support for Embeddings. arXiv preprint arXiv:2304.01433.

Hu, Z., Shen, S., Bonato, T., Jeaugey, S., Alexander, C., Spada, E., Dinan, J., Hammond, J., & Hoefler, T.(2025). Demystifying NCCL: An In-depth Analysis of GPU Communication Protocols and Algorithms. arXiv preprint arXiv:2507.04786.

Akopyan, F., Sawada, J., Cassidy, A., Alvarez-Icaza, R., Arthur, J., ..., & Modha, D.(2015). TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 34, no. 10, Oct. 2015.

Modha, D. S., Akopyan, F., Andreopoulos, A., Appuswamy, R., Arthur, J. V., Cassidy, A. S., ... & Taba, B.(2023). Neural inference at the frontier of energy, space, and time. Science, 382(6668), 329-335.

Davies, M., Srinivasa, N., Lin, T., Chinya, G., Cao, Y., ..., & Wang, H.(2018). Loihi: a Neuromorphic Manycore Processor with On-Chip Learning, IEEE Micro, Jan. 2018.

Morita, A., Bhattacharjee, A., Li, Y., Kim, Y., & Panda, P.,(2024). When In-memory Computing Meets Spiking Neural Networks. arXiv preprint arXiv:2408.12767.

Beer, K., Bondarenko, D., Farrelly, T., Osborne, T. J., Salzmann, R., Scheiermann, D., & Wolf, R.(2020). Training deep quantum neural networks. Nature communications. 11(1), 808.

Liao, H., Wang, D. S., Sitdikov, I., Salcedo, C., Seif, A., & Minev, Z. K.(2024). Machine learning for practical quantum error mitigation. Nature Machine Intelligence, 1-9.

Hinton, G., Osindero, S., & The., Y.(2006). A Fast Learning Algorithm for Deep Belief Nets. Neural Computations, vol. 18, issue 7, 2006.

Krizhevsky, A., Sutskever, I., & Hinton, G.(2012). ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Hallacy, C., Kalliam, A., Klimov, A., Krueger, G., Neelakantan, A., Ryder, N., Ramesh, A., Xiao, C., & Amodei, D.(2020). Scaling laws for neural language models. arXiv preprint arXiv:2001.08361.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K.(2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Volume 1(Long and Short Papers)(pp. 4171-4186).

Singh, A., Hu, R., Goswami, V., Couairon, G., Galuba, W., Rohrbach, M., & Kiela, D.(2022). FLAVA: A foundational language and vision alignment model. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(pp. 10037-10047).

Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., & Wierstra, D.(2016). Matching networks for one shot learning. In Advances in Neural Information Processing Systems, 29(NIPS 2016).

Gholami, A., Kim, S., Dong, Z., Yao, Z., Mahoney, M. W., & Keutzer, K.(2022). A survey of quantization methods for efficient neural network inference. In Low-Power Computer Vision(pp. 31-82). CRC Press.

Sanh, V., Wolf, T., & Rush, A. M.(2020). Movement pruning: Adaptive sparsity by fine-tuning. In Advances in Neural Information Processing Systems, 33(NeurIPS 2020).

Lavazza, A., & Pizzetti, F. G.(2025). Neuralink's brain-computer interfaces: medical innovations and ethical challenges. Frontiers in Human Dynamics, 7.

Anderson, J. R.(2007). How can the human mind occur in the physical universe? Oxford University Press.

Laird, J. E.(2012). The Soar cognitive architecture. The MIT Press.

Wu, S., Oltramari, A., Francis, J., Giles, C. L., & Ritter, F. E.(2024). Cognitive LLMs: Towards integrating cognitive architectures and large language models for manufacturing decision-making. arXiv preprint arXiv:2408.09176.

Kipf, T. N., & Welling, M.(2017). Semi-supervised classification with graph convolutional networks. In International Conference on Learning Representations(ICLR).

UNESCO.(2024). AI ethics and governance: A global perspective on readiness. UNESCO Publishing.

OECD.(2019). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments.

European Parliament, & Council of the European Union.(2024). Regulation(EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations(EC) No 300/2008,(EU) No 167/2013,(EU) No 168/2013,(EU) 2018/858,(EU) 2018/1139 and(EU) 2019/2144 and Directives 2006/42/EC, 2014/30/EU, 2014/35/EU and 2014/53/EU (Artificial Intelligence Act).

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., ... & Dean, J.(2021). Carbon emissions and large neural network training. arXiv preprint arXiv:2104.10350.

Hernández-Orallo, J.(2017). The Measure of All Minds: Evaluating Natural and Artificial Intelligence. Cambridge University Press.

Goertzel, B., & Wang, P.(Eds.).(2007). Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms. IOS Press.

Yao, S., Yu, D., Zhao, J., Sha, D., Butler, N., & Narasimhan, K.(2023). Tree of Thoughts: Deliberate Problem Solving with Large Language Models. arXiv preprint arXiv:2305.10601.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I.(2021). Learning Transferable Visual Models From Natural Language Supervision.

Proceedings of the 38th International Conference on Machine Learning, 8748-8763.

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S.(2023). Generative Agents: Interactive Simulacra of Human Behavior. Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology.

Bruce, J., Dennis, M., Parker-Holder, J., Wang, J., Guo, R., Greaves, J., ... & de Freitas, N.(2024). Genie: Generative Interactive Environments. arXiv preprint arXiv: 2402.15391.

LeCun, Y.(2022). A Path Towards Autonomous Machine Intelligence. OpenReview.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S.(2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D.(2020). Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems, 33, 1877-1901.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... & Hadsell, R.(2017). Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences, 114(13), 3521-3526.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., ... & Lowe, R.(2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35, 27730-27744.

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., ... & Amodei, D.(2022). Constitutional AI: Harmlessness from AI Feedback. arXiv preprint arXiv:2212.08073.

Christian, B.(2020). The Alignment Problem: Machine Learning and Human Values. W. W. Norton & Company.

Olah, C., Cammarata, N., Schubert, L., Goh, G., Petrov, M., & Carter, S.(2020). Zoom In: An Introduction to Circuits. Distill.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023.

Llama 2: Open foundation and f ine-tuned chat models. arXiv preprint arXiv:2307.09288(2023).

Gemini Team, Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, et al. 2023. Gemini: a family of highly capable multimodal models. arXiv preprint arXiv: 2312.11805(2023).

Anthropic AI. 2024. Claude 3. https://www.anthropic.com/news/claude-3-family.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023d. Mistral 7B. arXiv:2310.06825.

Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W. Rae, Oriol Vinyals, and Laurent Sifre. 2022. Training Compute-Optimal Large Language Models. arXiv:2203.15556.

Mohammad Shoeybi, Mostofa Patwary, Raul Puri, Patrick LeGresley, Jared Casper, and Bryan Catanzaro. 2020. Megatron-LM: Training Multi-illionParameter Language Models Using Model Parallelism. arXiv:1909.08053.

Reza Yazdani Aminabadi, Samyam Rajbhandari, Minjia Zhang, Ammar Ahmad Awan, Cheng Li, Du Li, Elton Zheng, Jeff Rasley, Shaden Smith, Olatunji Ruwase, and Yuxiong He. 2022. DeepSpeed Inference: Enabling Efficient Inference of Transformer Models at Unprecedented Scale. arXiv:2207.00032.

Reiner Pope, Sholto Douglas, Aakanksha Chowdhery, Jacob Devlin, James Bradbury, Anselm Levskaya, Jonathan Heek, Kefan Xiao, Shivani Agrawal, and Jeff Dean. 2022. Efficiently Scaling Transformer Inference. arXiv:2211.05102.

Wenhan Xiong, Jingyu Liu, Igor Molybog, Hejia Zhang, Prajjwal Bhargava, Rui Hou, Louis Martin, Rashi Rungta, Karthik Abinav Sankararaman, Barlas Oguz, Madian Khabsa, Han Fang, Yashar Mehdad, Sharan Narang, Kshitiz Malik, Angela Fan, Shruti Bhosale, Sergey Edunov, Mike Lewis, Sinong Wang, and Hao Ma. 2023. Effective Long-Context Scaling of Foundation Models. arXiv:2309.16039.

Youhe Jiang, Ran Yan, Xiaozhe Yao, Beidi Chen, and Binhang Yuan. 2023e. HexGen: Generative Inference of Foundation Model over Heterogeneous Decentralized Environment. arXiv:2311.11514.

Xupeng Miao, Gabriele Oliaro, Zhihao Zhang, Xinhao Cheng, Hongyi Jin, Tianqi Chen, and Zhihao Jia. 2023. Towards Efficient Generative Large Language Model Serving: A Survey from Algorithms to Systems. arXiv:2312.15234.

Gyeong-In Yu, Joo Seong Jeong, Geon-Woo Kim, Soojeong Kim, and Byung-Gon Chun. 2022. Orca: A Distributed Serving System for Transformer-Based Generative Models. In 16th USENIX Symposium on Operating Systems Design and Implementation(OSDI 22). USENIX Association, Carlsbad, CA, 521–538.

Anil Yemme and Shayan Srinivasa Garani. 2023. A Scalable GPT-2 Inference Hardware Architecture on FPGA. In 2023 International Joint Conference on Neural Networks(IJCNN). 1–8.

Mitchell Wortsman, Gabriel Ilharco, Samir Yitzhak Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S. Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmon, Simon Kornblith, and Ludwig Schmidt. 2022. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. arXiv:2203.05482.

Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. 2023. Editing Models with Task Arithmetic. arXiv:2212.04089.

Chengyue Wu, Teng Wang, Yixiao Ge, Zeyu Lu, Ruisong Zhou, Ying Shan, and Ping Luo. 2023b. π-Tuning: Transferring Multimodal Foundation Models with Optimal

Multi-task Interpolation. arXiv:2304.14381.

Yi-Lin Sung, Linjie Li, Kevin Lin, Zhe Gan, Mohit Bansal, and Lijuan Wang. 2023. An Empirical Study of Multimodal Model Merging. arXiv:2304.14933.

Zhenyi Lu, Chenghao Fan, Wei Wei, Xiaoye Qu, Dangyang Chen, and Yu Cheng. 2024. Twin-merging: Dynamic integration of modular expertise in model merging. arXiv preprint arXiv:2406.15479(2024).

Derek Tam, Mohit Bansal, and Colin Raffel. 2023. Merging by Matching Models in Task Subspaces. arXiv:2312.04339.

Chengsong Huang, Qian Liu, Bill Yuchen Lin, Tianyu Pang, Chao Du, and Min Lin. 2024. LoraHub: Efficient Cross-Task Generalization via Dynamic LoRA Composition. arXiv:2307.13269.

P. Schmid, "The carbon footprint of GPT-4," Medium, Mar. 2023. [Online]. Available: https://medium.com/data-science/the-carbon-footprint-of-gpt-4-d6c676eb21ae.

X. Wei, Y. Tian, and H. Wang, "How Far Are We From AGI?"arXiv preprintarXiv: 2405.10313, 2024.

S. Birhane, V. Prabhu, H. Suresh, et al., "A Critical Analysis of the Largest Source for Generative AI Training Data: Common Crawl," in Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency(FAccT), 2024.

P. Villalobos, L. Bing, S. Xu, et al., "Will we run out of data? Limits of LLM scaling based on human-generated data," arXiv preprintarXiv:2211.04325, 2022.

A. Goyal, M. Shoeybi, M. Rajbhandari, et al., "The Efficiency Spectrum of Large Language Models: An Algorithmic Perspective," arXiv preprintarXiv:2312.00678, 2023.

M. Chen, D. Guestrin, and K. P. Murphy, "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME," Advanced Intelligent Systems, vol. 6, no. 4, 2024.

C. Roscher, J. Lorenz, and S. Regneri, "Notions of explainability and evaluation approaches for explainable AI," Artificial Intelligence in Medicine, vol. 117, 2021.

Chowdhery, A., et al.(2022). PaLM: Scaling Language Modeling with Pathways. Journal of

Machine Learning Research, 24(112), 1-13.

M. Shoeybi, M. Patwary, R. Puri, P. LeGresley, M. Casper, and B. Catanzaro. "Megatron-LM: Training Multi-Billion Parameter Language Models Using Model Parallelism." arXiv preprint arXiv:1909.08053, 2019.

T. Dao, D. Y. Fu, S. Ermon, A. Rudra, and C. Ré, "FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness," in Advances in Neural Information Processing Systems(NeurIPS), 2022.

D. Li, R. Shao, A. Xie, E. P. Xing, X. Ma, I. Stoica, J. E. Gonzalez, and H. Zhang, "DistFlashAttn: Distributed Memory-efficient Attention for Long-context LLMs Training,"in Proc. Conf. on Language Modeling(COLM), 2024,

J. Hoffmann, C. Schuhmann, S. Borgeaud, and T. Henighan et al., "Beyond Chinchilla-Optimal: Accounting for Inference in Language Model Scaling Laws," in Proc. 41st Int. Conf. on Machine Learning(ICML), 2024.

D. Harvey, et el., "Optimizing MoE Routers: Design, Implementation, and Evaluation in Transformer Models," arXiv preprint arXiv:2506.16419, 2025.

L. Gao, S. Biderman, S. Black, L. Golding, T. Hoppe, C. Foster, J. Phang, H. He, A. Thite, N. Nabeshima, S. Presser, and C. Leahy, "The Pile: An 800 GB Dataset of Diverse Text for Language Modeling," arXiv preprint arXiv:2101.00027, 2020.

C. Schuhmann, R. Beaumont, R. Vencu, C. Gordon, R. Wightman, M. Cherti, T. Coombes, A. Katta, C. Mullis, M. Wortsman, P. Schramowski, S. Kundurthy, K. Crowson, L. Schmidt, R. Kaczmarczyk, and J. Jitsev, "LAION-5B: An open large-scale dataset for training next generation image-text models," Proc. Advances in Neural Information Processing Systems(NeurIPS), 2022.

N. Haque, "Catastrophic Forgetting in LLMs: A Comparative Analysis Across Language Tasks," arXiv preprint arXiv:2504.01241, Apr. 2025.

U. Peters and B. Chin-Yee, "Generalization Bias in Large Language Model Summarization of Scientific Research," Royal Society Open Science, 2025.

D. Roussinov, S. Sharoff, and N. Puchnina, "Controlling Out-of-Domain Gaps in LLMs for

Genre Classification and Generated Text Detection," in Proc. 31st Int. Conf. on Computational Linguistics(COLING), Abu Dhabi, UAE, Jan. 2025, pp. 3329-3344.

Z. Ke, Y. Ming, X.-P. Nguyen, C. Xiong, and S. Joty, "Demystifying Domain-adaptive Post-training for Financial LLMs," arXiv preprint arXiv:2501.04961, 2025.

A. Setlur, S. Garg, X. Geng, N. Garg, V. Smith, and A. Kumar, "RL on Incorrect Synthetic Data Scales the Efficiency of LLM Math Reasoning by Eight-Fold," arXiv preprint arXiv:2406.14532, 2024.

N. Sharma, T. Wolfers, and C. Y ı ld ı z, "Beyond Benchmarks: A Novel Framework for Domain-Specific LLM Evaluation and Knowledge Mapping," arXiv preprint arXiv:2506.07658, June 2025.

M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.

S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Advances in Neural Information Processing Systems(NeurIPS), vol. 30, pp. 4768-4777, 2017.

M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," Proceedings of the 34th International Conference on Machine Learning(ICML), vol. 70, pp. 3319-3328, 2017.

Y. Jiang, K. Ning, Z. Pan, X. Shen, J. Ni, W. Yu, A. Schneider, H. Chen, Y. Nevmyvaka, and D. Song, "Multi-modal Time Series Analysis: A Tutorial and Survey," arXiv preprint arXiv:2503.13709, Mar. 2025.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Hassabis, D.(2017). Mastering the game of Go without human knowledge. Nature, 550(7676), 354-359.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Hassabis, D.(2017). Mastering the game of Go without human knowledge. Nature, 550(7676), 354-359.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I.(2019). Language models are unsupervised multitask learners. OpenAI Technical Report.

Chollet, F.(2019). On the Measure of Intelligence. arXiv:1911.01547.

Bostrom, N.(2016), Superintelligence-Paths, Dangers, Strategies, Oxford University Press, USA.

Deng, J. et al.(2009). ImageNet: A large-scale hierarchical image database. CVPR.

Wang, A. et al.(2019). GLUE: A multi-task benchmark and analysis platform for natural language understanding. ICLR.

Srivastava, A. et al.(2022). Beyond the Imitation Game Benchmark(BIG-bench). arXiv: 2206.04615.

Conneau, A. et al.(2018). XNLI: Evaluating cross-lingual sentence representations. EMNLP.

Lake, B. M., & Baroni, M.(2018). Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. ICML.

Kim, N., & Linzen, T.(2020). COGS: A compositional generalization challenge based on semantic interpretation. EMNLP.

Johnson, J. et al.(2017). CLEVR: A diagnostic dataset for compositional language and elementary

Triantafillou, E. et al.(2019). Meta-Dataset: A dataset of datasets for learning to learn from few examples. ICLR.

Caldas, S. et al.(2019). LEAF: A benchmark for federated settings. arXiv:1812.01097.

Finn, C., Abbeel, P., & Levine, S.(2017). Model-Agnostic Meta-Learning for fast adaptation of deep networks. ICML.

Jiang, H. et al.(2021). How can we know when language models know? TACL.

Kosinski, M.(2023). Theory of Mind may have spontaneously emerged in large language models. arXiv:2302.02083.

Fleming, N.(2018). How artificial intelligence is changing drug discovery. Nature

Jumper, J., et al.(2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589.

Krittanawong, C., et al.(2021). Artificial intelligence in precision cardiovascular medicine. Journal of the American College of Cardiology, 78(3), 265-277.

Topol, E. J.(2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56.

You, Y., Lai, X., Pan, Y. et al.(2022). Artificial intelligence in cancer target identification and drug discovery. Sig Transduct Target Ther 7, 156.

Kamilaris, A., & Prenafeta-Boldú, F. X.(2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D.(2016). Deep neural networks based recognition of plant diseases by leaf image classification. Computational intelligence and neuroscience, 2016.

van Dijk, A. D. J., et al.(2020). Machine learning in plant science and plant breeding. iScience

Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J., & Arnold, F. H.(2019). Machine learning-assisted directed protein evolution with combinatorial libraries.(PNAS)

Torres, M.D.T. et al.,(2024). Mining human microbiomes reveals an untapped source of antimicrobial peptides, Cell.

Wan, F. et al.,(2024). Deep-learning-enabled antibiotic discovery, Nature Biomedical Engineering.

Aboy, M. et al.,(2024). Navigating the EU AI Act: implications for regulated digital products,(PMC review).

Fu, J. et al.,(2024). Breeding 5.0: Artificial intelligence-decoded breeding strategies, JIPB/Wiley.

Aijaz, N., et al.,(2025). Artificial intelligence in agriculture: Advancing crop productivity and sustainability, Journal of Agriculture and Food Research.

Bähner, F.D. et al.,(2021). Challenges in optimization and control of biobased processes, Industrial & Engineering Chemistry Research.

Benson, M.(2023). Digital Twins for Predictive, Preventive Personalized, and Participatory

Treatment of Immune-Mediated Diseases, Arterioscler Thromb Vasc Biol. 2023 Mar; 43(3): 410-416.

Yang, Y., et al.(2025). Beyond the pill: How nanorobots are transforming drug delivery, Journal of Drug Delivery Science and Technology.

Kong, X., et al.(2023). Advances of medical nanorobots for future cancer treatments, Journal of Hematology & Oncology.

Cooper, R. et al.(2023). Engineered bacteria detect tumor DNA, 381(6658), Science.

Groff-Vindman, C., et al.(2025). The convergence of AI and synthetic biology: the looming deluge, Nature npj biomedical innovations.

Pandey, D., et al.,(2024). Towards sustainable agriculture: Harnessing AI for global food security, Artificial Intelligence in Agriculture.

Voigt, C.,(2020). Synthetic biology 2020–2030: six commercially-available products that are changing our world, Nature Communications.

Shah, M., et al.(2025). A Framework for Assessing the Potential of Artificial Intelligence in the Circular Bioeconomy, Sustainability, 17(8).

Chu, Z., Wang, S., Xie, J., Zhu, T., Yan, Y., Ye, J., Zhong, A., Hu, X., Liang, J., Yu, P. S., & Wen, Q.(2025). LLM Agents for Education: Advances and Applications. arXiv preprint arXiv:2503.11733.

Latif, E., Mai, G., Nyaaba, M., Wu, X., Liu, N., Lu, G., Li, S., Liu, T., & Zhai, X.(2023). AGI: Artificial General Intelligence for Education. arXiv preprint arXiv:2304.12479.

Wang, S., Xu, T., Li, H., Zhang, C., Liang, J., Tang, J., Yu, P.S., & Wen, Q.(2024). Large Language Models for Education: A Survey and Outlook. arXiv preprint arXiv:2403.18105.

Hou, Z., Ye, Z., Zeng, N., Hao, T., & Zeng, K.(2025). Large Language Models Meet Legal Artificial Intelligence: A Survey. arXiv preprint arXiv:2509.09969.

Guha, N., Nyarko, J., Ho, D. E., Ré, C., Chilton, A., Narayana, A., Chohlas-Wood, A., Peters, A., Waldon, B., Rockmore, D. N., … Li, Z.(2023). LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language

Models. arXiv preprint arXiv:2308.11462.

Li, H., Chen, J., Yang, J., Ai, Q., Jia, W., Liu, Y., Lin, K., Wu, Y., Yuan, G., Hu, Y., Wang, W., Liu, Y., & Huang, M.(2024). LegalAgentBench: Evaluating LLM Agents in Legal Domain. arXiv preprint arXiv:2412.17259.

Cai, H., Zhao, S., Zhang, L., Shen, X., Xu, Q., Shen, W., Wen, Z., & Ban, T.(2025). Unilaw-R1: A Large Language Model for Legal Reasoning with Reinforcement Learning and Iterative Inference. arXiv preprint arXiv:2510.10072.

Taiwo, R., Bello, I. T., Abdulai, S. F., Yussif, A. M., Salami, B. A., Saka, A., & Zayed, T.(2024). Generative AI in the Construction Industry: A State-of-the-art Analysis. arXiv preprint arXiv:2402.09939.

Parsamehr, M., Perera, U. S., Dodanwala, T. C., Perera, P., & Ruparathna, R.(2023). A review of construction management challenges and BIM-based solutions: perspectives from the schedule, cost, quality, and safety management. Asian Journal of Civil Engineering, 24(1), 353-389.

Du, C., Esser, S., Nousias, S., & Borrmann, A.(2024). Text2BIM: Generating building models using a large language model-based multi-agent framework. arXiv preprint arXiv:2408.08054.

Yao, Y., Tam, V. W., Wang, J., Le, K. N., & Butera, A.(2024). Automated construction scheduling using deep reinforcement learning with valid action sampling. Automation in Construction, 166, 105622.

Wang, Xi, et al. "Enabling Building Information Model-driven human-robot collaborative construction workflows with closed-loop digital twins." Computers in Industry 161(2024): 104112.

Lee, D., Lee, S., Masoud, N., Krishnan, M. S., & Li, V. C.(2022). Digital twin-driven deep reinforcement learning for adaptive task allocation in robotic construction. Advanced Engineering Informatics, 53, 101710.

Al-Azani, S., Luqman, H., Alfarraj, M., Sidig, A. A. I., Khan, A. H., & Al-Hamed, D.(2024). Real-time monitoring of personal protective equipment compliance in surveillance

cameras. IEEE Access, 12, 121882-121895.

Poole, B., Jain, A., Barron, J. T., & Mildenhall, B.(2022). Dreamfusion: Text-to-3d using 2d diffusion. arXiv preprint arXiv:2209.14988.

Park, Joon Sung, et al. "Generative agents: Interactive simulacra of human behavior." Proceedings of the 36th annual acm symposium on user interface software and technology. 2023.

Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., ... & Défossez, A.(2023). Simple and controllable music generation. Advances in Neural Information Processing Systems, 36, 47704-47720.

Barrault, L., Chung, Y. A., Meglioli, M. C., Dale, D., Dong, N., Duquenne, P. A., ... & Wang, S.(2023). SeamlessM4T: massively multilingual & multimodal machine translation. arXiv preprint arXiv:2308.11596.

Liu, Q., Zhao, X., Wang, Y., Wang, Y., Zhang, Z., Sun, Y., ... & Tian, F.(2024). Large Language Model Enhanced Recommender Systems: A Survey. arXiv preprint arXiv:2412.13432.

Yu, W. F.(2025). AI as a co-creator and a design material: Transforming the design process. Design Studies, 97, 101303.

Kadenhe, N., Al Musleh, M., & Lompot, A.(2025, August). Human-AI Co-Design and Co-Creation: A Review of Emerging Approaches, Challenges, and Future Directions. In Proceedings of the AAAI Symposium Series(Vol. 6, No. 1, pp. 265-270).

Wu, W., Zhu, Z., & Shou, M. Z.(2025). Automated movie generation via multi-agent cot planning. arXiv preprint arXiv:2503.07314.

Gao, H. A., Geng, J., Hua, W., Hu, M., Juan, X., Liu, H., ... & Wang, M.(2025). A survey of self-evolving agents: On path to artificial super intelligence. arXiv preprint arXiv:2507.21046.

Alkhamissi, B., ElNokrashy, M., Alkhamissi, M., & Diab, M.(2024, August). Investigating Cultural Alignment of Large Language Models. In Proceedings of the 62nd Annual

Meeting of the Association for Computational Linguistics(Volume 1: Long Papers) (pp. 12404-12422).

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The Long-Document Transformer. arXiv:2004.05150 [cs.CL].

Yiran Ding, Li Lyna Zhang, Chengruidong Zhang, Yuanyuan Xu, Ning Shang, Jiahang Xu, Fan Yang, and Mao Yang. 2024b. Longrope: Extending llm context window beyond 2 million tokens. arXiv preprint arXiv:2402.13753(2024).

Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. 2023. Stanford Alpaca: An Instruction-following LLaMA model. https: //github.com/tatsu-lab/stanford_alpaca.

Albert Gu and Tri Dao. 2023. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv:2312.00752 [cs.LG].

Kaiwen Wang, Rahul Kidambi, Ryan Sullivan, Alekh Agarwal, Christoph Dann, Andrea Michi, Marco Gelmi, Yunxuan Li, Raghav Gupta, Avinava Dubey, et al. 2024a. Conditioned Language Policy: A General Framework for Steerable Multi-Objective Finetuning. arXiv preprint arXiv:2407.15762(2024).

Xiaohui Wang, Ying Xiong, Yang Wei, Mingxuan Wang, and Lei Li. 2021. LightSeq: A High Performance Inference Library for Transformers. arXiv:2010.13887 [cs.MS].

Xiaohui Wang, Ying Xiong, Yang Wei, Mingxuan Wang, and Lei Li. 2021. LightSeq: A High Performance Inference Library for Transformers. arXiv:2010.13887 [cs.MS].

NVIDIA. 2023a. FasterTransformer. https://github.com/NVIDIA/FasterTransformer.

Yujia Zhai, Chengquan Jiang, Leyuan Wang, Xiaoying Jia, Shang Zhang, Zizhong Chen, Xin Liu, and Yibo Zhu. 2023. ByteTransformer: A High-Performance Transformer Boosted for Variable-Length Inputs. arXiv:2210.03052 [cs.LG].

Binhang Yuan, Yongjun He, Jared Quincy Davis, Tianyi Zhang, Tri Dao, Beidi Chen, Percy Liang, Christopher Re, and Ce Zhang. 2023. Decentralized Training of Foundation Models in Heterogeneous Environments. arXiv:2206.01288 [cs.DC].

Tianqi Chen, Thierry Moreau, Ziheng Jiang, Lianmin Zheng, Eddie Yan, Meghan Cowan,

Haichen Shen, Leyuan Wang, Yuwei Hu, Luis Ceze, Carlos Guestrin, and Arvind Krishnamurthy. 2018. TVM: An Automated End-to-End Optimizing Compiler for Deep Learning. arXiv:1802.04799 [cs.LG].

Philippe Tillet, H. T. Kung, and David Cox. 2019. Triton: an intermediate language and compiler for tiled neural network computations. In Proceedings of the 3rd ACM SIGPLAN International Workshop on Machine Learning and Programming Languages(Phoenix, AZ, USA)(MAPL 2019). Association for Computing Machinery, New York, NY, USA, 10–19. https://doi.org/10.1145/3315508.3329973.

James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. 2018. JAX: composable transformations of Python+NumPy programs. http://github.com/google/jax..

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. 2017. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. arXiv preprint arXiv:1701.06538(2017).

Aurko Roy, Mohammad Saffar, Ashish Vaswani, and David Grangier. 2020. Efficient Content-Based Sparse Attention with Routing Transformers. arXiv:2003.05997 [cs.LG].

Ankit Gupta, Albert Gu, and Jonathan Berant. 2022. Diagonal State Spaces are as Effective as Structured State Spaces. arXiv:2203.14343 [cs.LG].

Alisa Liu, Zhaofeng Wu, Julian Michael, Alane Suhr, Peter West, Alexander Koller, Swabha Swayamdipta, Noah A Smith, and Yejin Choi. 2023h. We're Afraid Language Models Aren't Modeling Ambiguity. arXiv preprint arXiv:2304.14399(2023).

Yanping Huang, Youlong Cheng, Ankur Bapna, Orhan Firat, Mia Xu Chen, Dehao Chen, HyoukJoong Lee, Jiquan Ngiam, Quoc V. Le, Yonghui Wu, and Zhifeng Chen. 2019. GPipe: Efficient Training of Giant Neural Networks using Pipeline Parallelism. arXiv:1811.06965 [cs.CV].

Rongjie Huang, Mingze Li, Dongchao Yang, Jiatong Shi, Xuankai Chang, Zhenhui Ye,

Yuning Wu, Zhiqing Hong, Jiawei Huang, Jinglin Liu, et al. 2023b. Audiogpt: Understanding and generating speech, music, sound, and talking head. arXiv preprint arXiv:2304.12995(2023).

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. 2023a. Efficient Memory Management for Large Language Model Serving with PagedAttention. arXiv: 2309.06180 [cs.LG].

Lin Gui, Cristina Gârbacea, and Victor Veitch. 2024. BoNBoN Alignment for Large Language Models and the Sweetness of Best-of-n Sampling. https://doi.org/10.48550/arXiv.2406.00832 arXiv:2406.00832 [cs].

Ying Sheng, Lianmin Zheng, Binhang Yuan, Zhuohan Li, Max Ryabinin, Daniel Y. Fu, Zhiqiang Xie, Beidi Chen, Clark Barrett, Joseph E. Gonzalez, Percy Liang, Christopher Ré, Ion Stoica, and Ce Zhang. 2023b. FlexGen: High-Throughput Generative Inference of Large Language Models with a Single GPU. arXiv:2303.06865 [cs.LG].

Fanxu Meng, Zhaohui Wang, and Muhan Zhang. 2024. PiSSA: Principal Singular Values and Singular Vectors Adaptation of Large Language Models. arXiv preprint arXiv:2404.02948(2024).

Renrui Zhang, Jiaming Han, Chris Liu, Peng Gao, Aojun Zhou, Xiangfei Hu, Shilin Yan, Pan Lu, Hongsheng Li, and Yu Qiao. 2023d. LLaMA-Adapter: Efficient Fine-tuning of Language Models with Zero-init Attention. arXiv:2303.16199 [cs.CV].

Yuhan Liu, Hanchen Li, Kuntai Du, Jiayi Yao, Yihua Cheng, Yuyang Huang, Shan Lu, Michael Maire, Henry Hoffmann, Ari Holtzman, Ganesh Ananthanarayanan, and Junchen Jiang. 2023c. CacheGen: Fast Context Loading for Language Model Applications. arXiv:2310.07240 [cs.NI].

Jue Wang, Yucheng Lu, Binhang Yuan, Beidi Chen, Percy Liang, Christopher De Sa, Christopher Re, and Ce Zhang. 2023g. CocktailSGD: Fine-tuning Foundation Models over 500Mbps Networks. In Proceedings of the 40th International Conference on

Machine Learning(Proceedings of Machine Learning Research, Vol. 202), Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett(Eds.). PMLR, 36058-36076. https://proceedings.mlr.press/v202/wang23t.html.

Mengzhou Xia, Mikel Artetxe, Chunting Zhou, Xi Victoria Lin, Ramakanth Pasunuru, Danqi Chen, Luke Zettlemoyer, and Ves Stoyanov. 2023a. Training Trajectories of Language Models Across Scales. arXiv:2212.09803 [cs.CL].

Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, MonaDiab, Xian Li, Xi Victoria Lin, Todor Mihaylov, Myle Ott, Sam Shleifer, Kurt Shuster, Daniel Simig, Punit Singh Koura, Anjali Sridhar, Tianlu Wang, and Luke Zettlemoyer. 2022a. OPT: Open Pre-trained Transformer Language Models. arXiv:2205.01068 [cs.CL].

Kushal Tirumala, Aram H. Markosyan, Luke Zettlemoyer, and Armen Aghajanyan. 2022. Memorization Without Overfitting: Analyzing the Training Dynamics of Large Language Models. arXiv:2205.10770 [cs.CL].

Stella Biderman, Hailey Schoelkopf, Quentin Anthony, Herbie Bradley, Kyle O'Brien, Eric Hallahan, Mohammad Aflah Khan, Shivanshu Purohit, USVSN Sai Prashanth, Edward Raff, Aviya Skowron, Lintang Sutawika, and Oskar van der Wal. 2023. Pythia: A Suite for Analyzing Large Language Models Across Training and Scaling. arXiv:2304.01373 [cs.CL].

Dirk Groeneveld, Iz Beltagy, Pete Walsh, Akshita Bhagia, Rodney Kinney, Oyvind Tafjord, Ananya Harsh Jha, Hamish Ivison, Ian Magnusson, Yizhong Wang, Shane Arora, David Atkinson, Russell Authur, Khyathi Raghavi Chandu, Arman Cohan, Jennifer Dumas, Yanai Elazar, Yuling Gu, Jack Hessel, Tushar Khot, William Merrill, Jacob Morrison, Niklas Muennighoff, Aakanksha Naik, Crystal Nam, Matthew E. Peters, Valentina Pyatkin, Abhilasha Ravichander, Dustin Schwenk, Saurabh Shah, Will Smith, Emma Strubell, Nishant Subramani, Mitchell Wortsman, Pradeep Dasigi, Nathan Lambert, Kyle Richardson, Luke Zettlemoyer, Jesse Dodge, Kyle Lo, Luca

Soldaini, Noah A. Smith, and Hannaneh Hajishirzi. 2024. OLMo: Accelerating the Science of Language Models. arXiv:2402.00838 [cs.CL].

Apoorv Saxena. 2023. Prompt Lookup Decoding. prompt-lookup-decoding/.

Xupeng Miao, Gabriele Oliaro, Zhihao Zhang, Xinhao Cheng, Zeyu Wang, Zhengxin Zhang, Rae Ying Yee Wong, Alan Zhu, Lijie Yang, Xiaoxiang Shi, Chunan Shi, Zhuoming Chen, Daiyaan Arfeen, Reyna Abhyankar, and Zhihao Jia. 2024. SpecInfer: Accelerating Generative Large Language Model Serving with Tree-based Speculative Inference and Verification. arXiv:2305.09781 [cs.CL].

Jun Zhang, Jue Wang, Huan Li, Lidan Shou, Ke Chen, Gang Chen, and Sharad Mehrotra. 2023k. Draft & Verify: Lossless Large Language Model Acceleration via Self-Speculative Decoding. arXiv:2309.08168 [cs.CL].

Tri Dao. 2023. FlashAttention-2: Faster Attention with Better Parallelism and Work Partitioning.(2023).

Ke Hong, Guohao Dai, Jiaming Xu, Qiuli Mao, Xiuhong Li, Jun Liu, Kangdi Chen, Yuhan Dong, and Yu Wang. 2023a. FlashDecoding++: Faster Large Language Model Inference on GPUs. arXiv:2311.01282 [cs.LG].

Yunho Jin, Chun-Feng Wu, David Brooks, and Gu-Yeon Wei. 2023b. S3: Increasing GPU Utilization during Generative Inference for Higher Throughput. arXiv:2306.06000 [cs.AR].

Connor Holmes, Masahiro Tanaka, Michael Wyatt, Ammar Ahmad Awan, Jeff Rasley, Samyam Rajbhandari, Reza Yazdani Aminabadi, Heyang Qin, Arash Bakhtiari, Lev Kurilenko, and Yuxiong He. 2024. DeepSpeed-FastGen: High-throughput Text Generation for LLMs via MII and DeepSpeed-Inference. arXiv:2401.08671 [cs.PF].

Lequn Chen, Zihao Ye, Yongji Wu, Danyang Zhuo, Luis Ceze, and Arvind Krishnamurthy. 2023h. Punica: Multi-Tenant LoRA Serving. arXiv:2310.18547 [cs.DC].

Ying Sheng, Shiyi Cao, Dacheng Li, Coleman Hooper, Nicholas Lee, Shuo Yang, Christopher Chou, Banghua Zhu, Lianmin Zheng, Kurt Keutzer, Joseph E. Gonzalez, and Ion Stoica. 2023a. S-LoRA: Serving Thousands of Concurrent LoRA Adapters.

arXiv:2311.03285 [cs.LG].

Predibase. 2023. Multi-LoRA inference server that scales to 1000s of fine-tuned LLMs. https://github. com/predibase/lorax.

MLC team. 2023. MLC-LLM. https://github.com/mlc-ai/mlc-llm.

Quantum Mechanics-Informed AI for Semiconductor: A Comprehensive Investment and Technology Analysis, Oct. 2025, New York General Group, https://www.newyorkgeneralgroup.com/qmiagi-for-semiconductor#:~:text=AGI%20systems%2C%20however%2C%20can%20create%20even%20greater,that%20humans%20may%20not%20be%20able%20to.

Capgemini Research Institute Report, 2025, The semiconductor industry in the AI era Innovating for tomorrow's demands, https://www.capgemini.com/wp-content/uploads/2025/01/Semiconductors-report.pdf.

D. Roussinov, S. Sharoff, and N. Puchnina, "Controlling Out-of-Domain Gaps in LLMs for Genre Classification and Generated Text Detection," in *Proc. 31st Int. Conf. on Computational Linguistics(COLING)*, Abu Dhabi, UAE, Jan. 2025, pp. 3329–3344.

R. Bommasani, T. Lee, and P. Liang, "Holistic Evaluation of Language Models (HELM),"Stanford CRFM, 2022.

W. Zhao, J. Peña Queralta, and T. Westerlund, "Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: A Survey," *in Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence(SSCI)*, IEEE, 2020.

A. Mocle, "Artificial General Intelligence: What Are We Investing In?" *Tech Policy Press,* 2025.

J. Vipra and S. M. West, "Computational Power and AI," *AI Now Institute Report*, 2023.

J. Fernandez, "The Leading Generative AI Companies" *IoT Analytics Research Brief*, 2025.

A. Thadani and G. C. Allen, "Mapping the Semiconductor Supply Chain: The Critical Role of the Indo-Pacific Region," *Center for Strategic and International Studies(CSIS),* 2023.

A. F. Gomes and J. van den Wijngaard, "How AI Chips Became a Tool of Global Power,"

IE Insights, 2025.

FinTech Global, "Five Banks Dominate Global AI Research in Finance," 2025.

OpenReview, "TRADEFM: A Generative Foundation Model for Trade-Flow and Market Microstructure," *under reviewat the Int. Conf. on Learning Representations(ICLR),* 2026.

L. Chen, S. Liu, J. Yan, X. Wang, H. Liu, C. Li, et al., "Advancing Financial Engineering with Foundation Models: Progress, Applications, and Challenges," *arXiv preprint* arXiv:2507.18577, 2025.

N. Wang, H. Yang, and C. D. Wang, "FinGPT: Instruction Tuning Benchmark for Open-Source Large Language Models in Financial Datasets," in *Proc. 37th Conf. on Neural Information Processing Systems(NeurIPS)*, New Orleans, LA, USA, Dec. 2023.

. Huang, M. Xiao, D. Li et al., "Open-FinLLMs: Open Multimodal Large Language Models for Financial Applications," *arXiv preprint* arXiv:2408.11878, 2024.

G. Bhatia, E. M. Billah Nagoudi, H. Cavusoglu, and M. Abdul-Mageed, "Fintral: A Family of GPT-4-Level Multimodal Financial Large Language Models," *arXiv preprint* arXiv:2402.10986, 2024.

Morgan Stanley, "Thematics: Uncovering Alpha in AI's Rate of Change," *Morgan Stanley Research Report*, 2025.

I. Okpala*, A. Golgoon*, and A. R. Kannan*, "Agentic AI Systems Applied to Tasks in Financial Services: Modeling and Model Risk Management Crews," *arXiv preprint* arXiv:2502.05439, 2025.

Bank of America Newsroom, "BOFA's Erica Surpasses 2 Billion Interactions—Helping 42 Million," Apr. 2024.

CTOMagazine, "AI in Morgan Stanley: Shaping the Future of Financial Services,", 2025,

GoBeyond AI, "JPMorgan COiN AI Contract Analysis Legal Docs,", 2025.

Morgan Stanley, "Morgan Stanley Research Announces AskResearchGPT," 2024.

EY, "The EU AI Act: What It Means for Your Business," *EY Insights*, 2024.

AI.gov, "President Trump's AI Strategy and Action Plan,"

D. Kim, C. Park, et al., "SOLAR 10.7B: Scaling Large Language Models with Simple yet Effective Depth Up-Scaling," in *Proc. 2024 Conf. of the North American Chapter of the Association for Computational Linguistics(NAACL)*, Mexico City, Mexico, June 2024.

K. M. Yoo et al., "HyperCLOVA X: Scaling Korean Large Language Models for Industry and Research," *arXiv preprint* arXiv:2404.01954, Naver Cloud Corp., 2024.

World Trade Organization, Maritime Transport Services: Overview and Key Statistics., WTO Publications, 2023.

Caliskan, A., and Y. Ozturkoglu, "Maritime Logistics," Intelligent Transportation and Planning: Breakthroughs in Research and Practice, IGI Global, 2018, pp. 822-845.

Markopoulos, E., et al., The Impact of Digital Twins Technology in Maritime Fleet and Safety Management, UCL Discovery, 2023.

International Maritime Organization, Autonomous Shipping, IMO Media Centre, 2024.

Yara International & Kongsberg Gruppen, Yara Birkeland: The First Ever Fully Electric and Autonomous Container Ship, Yara International Press Release, 2021.

Korea Autonomous Surface Ship Project Office, Korea Autonomous Surface Ship(KASS) Project Outline, KASS Project, 2020.

Ministry of Oceans and Fisheries(Republic of Korea), Korea Autonomous Surface Ship Project, News Release, 27 November 2020.

Sharma, A., et al., "Catching Up with Time? Examining the STCW Competence Framework for Autonomous Shipping," Proceedings of the Ergoship Conference, Haugesund, Norway, 2019.

Port of Rotterdam Authority, Smart Infrastructure ¦ Port of Rotterdam, 2023.

Ding, Y., et al., "Real-Time Monitoring and Optimal Resource Allocation for Automated Container Terminals: A Digital Twin Application at the Yangshan Port," Journal of Advanced Transportation, 2023.

Yigit, Y., et al., "TwinPort: 5G Drone-Assisted Data Collection with Digital Twin for Smart Seaports," Scientific Reports, vol. 13, no. 1, p. 12310, 2023.

Eom, J.-O., Yoon, J.-H., Yeon, J.-H., and Kim, S.-W., "Port Digital Twin Development for Decarbonization: A Case Study Using the Pusan Newport International Terminal," Journal of Marine Science and Engineering, vol. 11, no. 9, p. 1777, 2023.

DHL Trend Research, Digital Twins in Logistics, DHL Innovation Center Report, 2023.

CJ Logistics, "Digital Twin-Based Smart Logistics Center," CJ Logistics Newsroom, 2022.

Mordor Intelligence, Autonomous Ships Market Size & Share Analysis - Growth Trends & Forecasts(2025-2030).

Zhang, Z., and H. Xu, "Explainable AI for Maritime Autonomous Surface Ships(MASS): Adaptive Interfaces and Trustworthy Human-AI Collaboration," arXiv preprint arXiv:2509.15959, 2025.

Mordor Intelligence, Smart Port Market Size & Share Analysis - Growth Trends & Forecasts(2025-2030).

Accenture, Making Autonomous Supply Chains Real, 2025.

[기타 자료]

BAVISHI, Rohan, et al. Introducing our multimodal models, 2023. URL https://www. adept. ai/blog/fuyu-8b, 2023, 2.

Intel.(2024, April 17). Intel Builds World's Largest Neuromorphic System to Enable More Sustainable AI. Intel Newsroom. Retrieved from https://www.intel.com/content/www/us/en/newsroom/news/intel-builds-worlds-largest-neuromorphic-system.html.

Gambetta, J., & Gil, D.(2023, December 4). The hardware and software for the era of quantum utility is here. IBM Research Blog. Retrieved from https://www.ibm.com/quantum/blog/quantum-roadmap-2033.

SK Hynix.(2024. 9. 26). SK하이닉스, 세계 최초 '12단 적층HBM3E' 양산 돌입. SK hynix Newsroom. Retrieved from https://news.skhynix.co.kr/12hi-hbm3e-production/.

Mistral AI.(2023, December 8). Mixtral of experts. https://mistral.ai/news/mixtral-of-experts/.

Forbes(2024), "OpenAI's 5 Levels Of 'Super AI' (AGI To Outperform Human Capability)", https://www.forbes.com/sites/jodiecook/2024/07/16/openais-5-levels-of-super-ai-agi-to-outperform-human-capability/.

MLC team. 2023. MLC-LLM. https://github.com/mlc-ai/mlc-llm.

S. Shukla, "Comprehensive Analysis of FLOP Calculations in Large Language Models," Medium, 2023. [Online]. Available: https://medium.com/my-aiml/comprehensive-analysis-of-flop-calculations-in-large-language-models-ae023647f66a.

D. Krueger, "Continuous Adversarial Quality Assurance: Extending RLHF and Beyond," Alignment Forum, 2023, https://www.alignmentforum.org/posts/QGaioedKBJE39YJeD/continuous-adversarial-quality-assurance-extending-rlhf-and.

"AI alignment," Wikipedia, 2024, https://en.wikipedia.org/wiki/AI_alignment.

K. Semba, "Artificial Intelligence, Real Consequences: Confronting AI's Growing Energy Appetite" Extreme networks, August. 2024, https://www.extremenetworks.com/resources/blogs/confronting-ai-growing-energy-appetite-part-1.

NVIDIA, "Introducing 200G HDR InfiniBand Solutions," White Paper, 2018, https://network.nvidia.com/files/doc-2020/wp-introducing-200g-hdr-infiniband-solutions.pdf.

Microsoft, "ZeRO: Zero Redundancy Optimizer," DeepSpeed Tutorials, 2025. https://www.deepspeed.ai/tutorials/zero/

NVIDIA, "Megatron-LM," GitHub repository. https://github.com/NVIDIA/Megatron-LM.

E. Erdil, "Data Movement Bottlenecks to Large-Scale Model Training: Scaling Past 1e28 FLOP," Epoch AI Blog, Nov. 2, 2024, https://epoch.ai/blog/data-movement-bottlenecks-scaling-past-1e28-flop.

Gcore, "NVIDIA GPUs: H100 vs. A100 ¦ a detailed comparison," Gcore Blog, 2025, https://gcore.com/blog/nvidia-h100-a100.

Signity Solutions, “Understanding GDPR and CCPA in the Context of AI Systems”, Signity Solutions Blog. [Online]. Available: https://www.signitysolutions.com/blog/understanding-gdpr-and-ccpa.

AP News, “Getty Images drops copyright infringement allegations in high-profile UK lawsuit against Stability AI,” AP News, Jun. 25, 2025. https://apnews.com/article/getty-images-stability-ai-copyright-trial-stable-diffusion-7208c729fb10c1f133cb49da2065d72a.

The New York Times, “NYT case against OpenAI and Microsoft can advance,” Axios(summarizing NYT v. OpenAI/Microsoft), Apr. 1, 2025. https://www.axios.com/2025/04/01/nyt-openai-microsoft-lawsuit-advances.

ARC-AGI(2019). ARC-AGI(The General Intelligence Benchmark), https://arcprize.org/arc-agi?utm_source=chatgpt.com.

기술과 혁신(2020). 미래의 먹거리 바이오 산업, 바로 알자, http://webzine.koita.or.kr/201604-specialissue/INTRO-%EB%AF%B8%EB%9E%98%EC%9D%98-%EB%A8%B9%EA%B1%B0%EB%A6%AC-%EB%B0%94%EC%9D%B4%EC%98%A4%EC%82%B0%EC%97%85-%EB%B0%94%EB%A1%9C-%EC%95%8C%EC%9E%90.

FDA(2025). Artificial Intelligence in Software as a Medical Device / AI-enabled guidance pages(FDA website, 2024-2025). https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device.

AI 2027 Project.(2025). AI 2027: Scenarios for Artificial Intelligence Development. https://ai-2027.com.

Benaich, N., & Hogarth, I.(2025). State of AI Report 2025. Air Street Capital. https://www.stateof.ai.

REVOLUTIONIZING SEMICONDUCTOR DESIGN AND MANUFACTURING WITH AI, Prashis Raghuwanshi, Vol. 3 No. 3(2024): Exploring Ethical Dilemmas in Artificial Intelligence and Machine Learning, pp.272-277, https://doi.org/10.60087/jklst.vol3.n3. pp. 272-277.

Artificial Intelligence-Driven Approaches in Semiconductor Research, Yiqiang Zheng, Hao Xu, Zhexin Li, Linlin Li, Yongchao Yu, Pengfei Jiang, Yanmeng Shi, Jing Zhang, Yuqing Huang, Qing Luo, Zheng Lou, Lili Wang, PMID: 40534303 DOI: 10.1002/adma.202504378, https://pubmed.ncbi.nlm.nih.gov/40534303/.

SEMICONDUCTOR MEMORY AT A TURNING POINT, Mitsui & Co. Global Strategic Studies Institute Monthly Report June 2025, https://www.mitsui.com/mgssi/en/report/detail/__icsFiles/afieldfile/2025/07/22/2507_m_ogawa_ex_e_1.pdf.

Artificial Intelligence(AI) is accelerating the shift to advanced packaging with FOPLP, glass cores, and Co-Packaged Optics(CPO) integration., https://www.yolegroup.com/strategy-insights/ai-fuels-the-future-of-advanced-packaging/.

Annual Meetings of the Global Future Councils and Cybersecurity, How we enhance cybersecurity defences before the attackers in an AGI world, https://www.weforum.org/stories/2025/10/how-we-enhance-cybersecurity-defences-before-the-attackers-in-an-agi-world/.

Times of India, "JPMorgan Chase CEO Jamie Dimon Doesn't Sugarcoat AI Impact on Jobs; Says: Yes AI Will Lead to Layoffs, but This Does Not Mean…," 2025. [Online]. Available: https://timesofindia.indiatimes.com/technology/tech-news/jpmorgan-chase-ceo-jamie-dimon-doesnt-sugarcoat-ai-impact-on-jobs-says-yes-ai-will-lead-to-layoffs-but-this-does-not-mean-/articleshow/124472661.cms.

Goldman Sachs, "Insights, Analytics and Trading Platform ¦ Goldman Sachs Marquee," [Online]. Available: https://marquee.gs.com/welcome/our-platform/overview.

FinTech News Singapore, "DBS Recognised as World's Best AI Bank in Global Finance's Inaugural Awards," 2025. [Online]. Available: https://fintechnews.sg/120178/ai/dbs-ai-bank/.

HSBC, "HSBC AI Markets ¦ Products,"[Online]. Available:

https://www.business.hsbc.com/en-gb/products/hsbc-ai-markets.

Private Banker International, "HSBC Private Bank Unveils New AI-Driven Platform," 2025. [Online]. Available: https://www.privatebankerinternational.com/news/hsbc-private-bank-unveils/.

GPT-3 Demo / BloombergGPT, "BloombergGPT ¦ Discover AI Use Cases," [Online]. Available: https://gpt3demo.com/apps/bloomberggpt.

S. Wu, O. Irsoy, et al., "BloombergGPT: A Large Language Model for Finance," arXiv preprintarXiv:2303.17564, 2023. [Online]. Available: https://arxiv.org/abs/2303.17564.

Samsung SDS, "AI in Banking in 2025," Samsung SDS Insights, 2025. [Online]. Available: https://www.samsungsds.com/kr/insights/ai-in-banking-in-2025.html.

AI Times, "AI Times News Article," 2021. [Online]. Available: https://www.aitimes.com/news/articleView.html?idxno=141620.

European Commission, "AI Act enters into force," European Commission Newsroom, 2024. [Online]. Available: https://commission.europa.eu/news-and-media/news/ai-act-enters-force-2024-08-01_en.

The White House, "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," Federal Register, Doc. 2023-24283, Nov. 1, 2023. [Online]. Available: https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.

Department for Science, Innovation and Technology, "Frontier AI Safety Commitments, AI Seoul Summit 2024," GOV.UK, May 21 2024. [Online]. Available: https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024.

OpenAI, "Hello GPT-4o," OpenAI, May 13 2024. [Online]. Available: https://openai.com/ko-KR/index/hello-gpt-4o/?utm_source=chatgpt.com.

Google DeepMind, "RT-2: New model translates vision and language into action," DeepMind Blog, July 28 2023. [Online]. Available:

https://deepmind.google/discover/blog/rt-2-new-model-translates-vision-and-language-into-action/?utm_source=chatgpt.com.

NVIDIA Corporation, "NVIDIA Blackwell Platform Arrives to Power a New Era of Computing," NVIDIA Newsroom, Mar. 18 2024. [Online]. Available: https://nvidianews.nvidia.com/news/nvidia-blackwell-platform-arrives-to-power-a-new-era-of-computing?utm_source=chatgpt.com.

National Institute of Standards and Technology(NIST). Artificial Intelligence Risk Management Framework(AI RMF 1.0) [NIST AI 100-1]. Gaithersburg, MD: U.S. Dept. of Commerce, Jan. 2023. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf.

UK Department for Science, Innovation and Technology, "Introducing the AI Safety Institute," GOV.UK, Nov. 2023. [Online]. Available: https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute?utm_source=chatgpt.com.

OECD, Korea's Unborn Future: Understanding Low-Fertility Trends, OECD Publishing, Paris, Mar. 2025. [Online]. Available: https://www.oecd.org/en/publications/korea-s-unborn-future_75aa749c-en.html.

Hedberg, T.D., et al. "Testing the Digital Thread in Support of Model-Based Manufacturing and Inspection," NIST, 2016. [Online]. Available: nist.gov/publications/testing-digital-thread-support-model-based-manufacturing-and-inspection.

Brohan, A., et al. "RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control," arXiv:2307.15818, 2023. [Online]. Available: arxiv.org/abs/2307.15818.

McKinsey & Company, "Digital twins: The next frontier of factory optimization," McKinsey Operations Insight, Jan. 10, 2024. [Online]. Available: mckinsey.com/capabilities/operations/our-insights/digital-twins-the-next-frontier-of-factory-optimization.

디지털 대전환 메가트렌드 연구 시리즈

25-28-01 디지털 대전환 메가트렌드 연구 총괄보고서Ⅴ: 2035 AI·디지털 대전환 메가트렌드 (문아람 외, 정보통신정책연구원)

25-28-02 2035 AI · 디지털 대전환 미래전략 (문아람 외, 정보통신정책연구원)

25-28-03 디지털 전환 기반 구축을 위한 차세대 통신기술 및 양자기술 활용 방안 (이인규 외, 한국통신학회)

25-28-04 디지털 자산 시장 발전에 따른 혁신과 공정의 기업 생태계 조성 (양성병 외, 한국경영학회)

25-28-05 디지털 전환기 기업 동태 분석 기반 혁신성장 정책 방향 (이경원 외, 정보통신정책학회)

25-28-06 AI·디지털 전환기 노동 구조의 재편과 노동권 보호를 위한 안전망 설계 (임운택 외, 한국사회학회)

25-28-07 디지털 전환기 민주주의 회복력 강화를 위한 정치제도의 변화 (김범수 외, 한국정치학회)

25-28-08 디지털 전환에 대응하는 AI·데이터 기반 디지털 복지서비스 개발 연구 (윤건 외, 한국행정학회)

25-28-09 AI·디지털 전환에 따른 공공부문 데이터 거버넌스 전략 (남태우 외, 한국정책학회)

25-28-10 디지털 저탄소 전환을 위한 기술적·제도적 방안 (김현 외, 대한전자공학회)

25-28-11 디지털 전환기 AGI 기술 경쟁력 확보를 위한 중장기 전략 (송길태 외, 한국정보과학회)

● 저 자 소 개 ●

송 길 태

- 펜실베니아주립대학교 컴퓨터공학 박사
- 현 부산대학교 정보컴퓨터공학부 교수
- 현 부산대학교 인공지능융합혁신대학원장
- 현 한국정보과학회 부회장

박 진 영

- 한국과학기술원 전산학 박사
- 현 성균관대학교 소프트웨어학과 부교수

이 상 국

- 프랑스국립루앙대학교 전기전자공학 박사
- 현 가톨릭대학교 미디어기술콘텐츠학과 정교수
- 현 국가과학기술인력개발원 KIRD 석좌교수
- 현 한국정보과학회 인공지능소사이어티 부회장

최 재 식

- 일리노이대학교 컴퓨터공학 박사
- 현 한국과학기술원 AI대학원 교수
- 현 ㈜인이지 대표이사

김 기 범

- 부산대학교 대학원 AI전공 박사과정

나 승 훈

- 포항공과대학교 컴퓨터공학 박사
- 현 울산과학기술원 AI대학원 부교수

배 창 석

- 연세대학교 전기전자공학 박사
- 현 대전대학교 정보통신공학과 교수
- 현 한국정보과학회 인공지능소사이어티 부회장

최 영 준

- 서울대학교 전기컴퓨터공학 박사
- 현 아주대학교 소프트웨어학과 교수
- 현 아주대학교 인공지능융합혁신대학원장
- 현 더에이아이랩(주) 대표

최 정 훈

- 부산대학교 대학원 AI전공 박사과정

ICT 진흥 및 혁신기반 조성사업 RS-2025-02173110
정책연구 25-28-11

디지털 전환기 AGI 기술 경쟁력 확보를 위한 중장기 전략

(Mid- to Long-Term Strategies for Securing AGI Technological Competitiveness in the Era of Digital Transformation)

2025년 12월 일 인쇄
2025년 12월 일 발행

발행인 과학기술정보통신부 장관
발행처 과학기술정보통신부
세종 갈매로 477 정부세종청사 4동
Homepage: www.msit.go.kr
인 쇄 인 성 문 화