

지능 공장의 시대, 데이터센터 진화에 따른 재정 정책 과제

AI FACTORY



‘저장’ 공간에서 ‘생산’ 공간으로 패러다임 전환

인터넷 보급 이후 데이터센터는 대표적인 디지털 인프라(Infrastructure)로 자리 잡았다. 과거 데이터센터의 역할은 기업 데이터의 안전한 저장, 웹 트래픽 중개, 서버의 중단 없는 가동 등 ‘데이터 저장소(Data Storage)’에 국한되었다. 정적 데이터를 보관하고 사용자 요청에 따라 이를 제공하는 수동적 공간에 가까웠다. 그러나 생성형 AI(Generative AI)의 등장은 데이터센터의 패러다임을 근본적으로 전환하였다. 현재의 데이터센터는 단순한 보관 창고를 넘어, 데이터를 원료로 고부가가치 산출물을 창출하는 ‘지능형 제조 공장(AI Factory)’으로 진화하고 있다. 엔비디아(NVIDIA)의 최고경영자 젠슨 황(Jensen Huang)을 비롯한 글로벌 기술 리더들은 미

래 데이터센터를 토큰(Token)을 생산하는 새로운 시대의 공장으로 규정한다.¹ 원료인 데이터와 동력원인 전력을 투입하여 비즈니스에 즉시 활용 가능한 ‘지능(Intelligence)’을 실시간으로 생산하는 제조 기지로 전환되었다는 의미다.

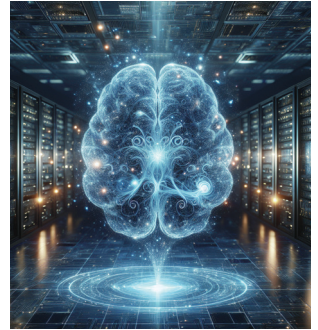
이러한 전환은 우리 재정에도 직접적인 함의를 던진다. 지능 생산 기지의 구축에는 기가와트(GW)급 전력망과 수조 원대의 컴퓨팅 자원이 요구되며, 이는 곧 대규모 공공투자와 재정 배분의 우선순위 문제로 귀결되기 때문이다. 본 고에서는 데이터센터가 지능 생산 공장으로 변모하는 메커니즘을 분석하고, 이로 촉발된 물리적 인프라(Infrastructure)의 격변과 기업의 과제를 살펴본 뒤, 이를 뒷받침할 재정 정책의 방향성을 논하고자 한다.

1 NVIDIA, “AI is a 5-Layer Cake”, 엔비디아 공식 블로그(2026.3). 최초 발표는 2026년 1월 세계경제포럼 연설

생성형 AI의 확산으로 데이터센터는 단순한 데이터 저장시설을 넘어 지능을 생산하는 핵심 산업 인프라로 진화하고 있다. AI 경쟁력 확보를 위해서는 전력망과 컴퓨팅 자원, 첨단 반도체 등 대규모 인프라 구축이 필수적인 만큼, 효율적인 재정 투자와 정책적 지원이 중요한 과제로 부각되고 있다.

AI 팩토리 시대에 필요한 재정 정책의 방향과 국가의 역할을 살펴본다.

이상현 산업연구원 디지털전환연구실 선임연구위원



<표 1> 전통적 데이터센터와 AI 팩토리의 핵심 패러다임 비교

구분	전통적 데이터센터(디지털 창고)	AI 팩토리(지능 제조 공장)
핵심 역할	데이터 저장·호스팅·트래픽 중개	데이터 학습 및 모델 추론, 지능 생산
아키텍처	개별 서버의 병렬적 나열독립 작동	초고속 네트워크로 묶인 단일 슈퍼컴
입·출력물	(입력) 사용자 요청 → (출력) 정적 파일	(입력) 전력+데이터 → (출력) 지능형 토큰
전력 규모	수십 MW 수준의 범용 전력망	GW 수준의 대규모 분산형 전력망

자료: 본문 내용 정리

AI 팩토리의 가동 메커니즘 :

‘5단 케이크(5-Layer Cake)’ 모델

인공지능 생태계를 고도로 연계된 단일 공정으로 이해하기 위해 글로벌 기술 업계에서는 젠슨 황 엔비디아 최고경영자가 제시한 ‘5단 케이크(5-Layer Cake)’ 개념을 활용한다. 이 개념은 2026년 1월 세계 경제포럼(다보스)에서 처음 제시된 뒤 엔비디아 공식 블로그를 통해 정식화되었으며, 에너지를 기저 동력으로 삼아 최상단의 비즈니스 애플리케이션에

이르는 수직 계열화 구조를 설명한다. 데이터센터는 이 다섯 계층이 유기적으로 맞물려 지능을 생산하는 물리적 공장의 역할을 담당한다.

1) 에너지(Energy)와 칩(Chips): 지능 생산의 물리적 토대
실시간으로 지능형 연산을 처리하는 데이터센터는 상시 안정적인 대규모 전력이 요구된다. 생산되는 모든 산출물(토큰)이 전력 투입과 전자의 물리적 이동을 거쳐 완성되므로, 저렴하고 지속 가능한

에너지의 확보는 공장 가동의 최우선 전제 조건이다. 에너지 계층 위에는 고성능 연산 가속 프로세서(GPU·NPU)가 위치한다. 이 반도체 칩은 전력을 연산 작업으로 변환하는 장치로서, 대규모 병렬처리 능력과 고대역폭 메모리(HBM) 기술을 결합하여 전력당 연산 성능(Performance per Watt)을 극대화하는 역량이 이 단계의 경쟁력을 좌우한다.

2) 인프라: 전통적 데이터센터에서 AI 팩토리로 진화

5단 케이크에서 인프라는 전통적 데이터센터가 AI 팩토리로 진화하는 물리적·논리적 전환점을 상징한다. 과거의 인프라는 독립된 서버들을 병렬로 나열하여 정적 데이터를 저장·인출하는 ‘디지털 창고’에 불과했다. 반면, AI 시대의 인프라는 개별 칩을 고성능 네트워킹(Network Fabric)과 초고속 인터커넥트로 묶어 단일한 거대 슈퍼컴퓨터이자 초정밀 가공

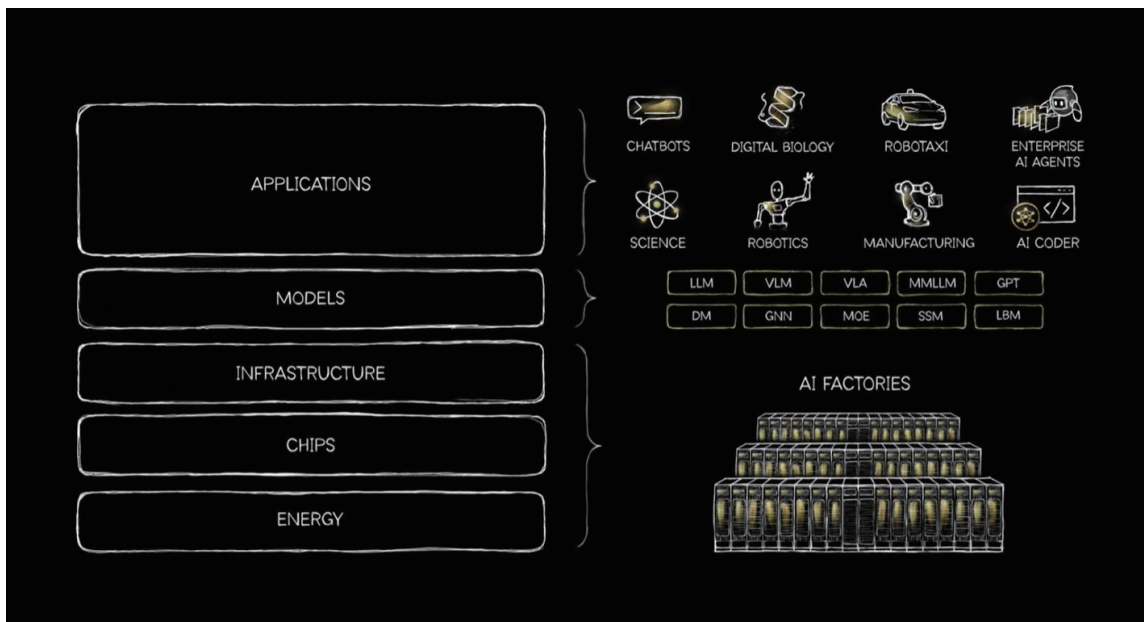
설비로 작동시킨다. 즉, 인프라 계층은 하단의 자원을 결합하여 지능을 양산하는 물리적 실체를 제공한다.

3) AI 모델 및 애플리케이션(Application):

범용 지능에서 경제적 가치로

통합된 하드웨어 인프라 위에서 데이터를 학습하고 고도화되는 소프트웨어 엔진이 AI 모델이다. 흔히 대규모 언어 모델(LLM)이 주목받지만, 이 계층은 언어에 국한되지 않는다. 생물·화학·물리 현상과 로봇 제어를 이해하는 피지컬 AI(Physical AI)와 산업 도메인 특화 모델이 오히려 제조 현장에서 결정적 부가가치를 창출하며, 이는 후술할 우리나라의 강점과 직결된다. 최상단 애플리케이션은 이렇게 산출된 지능이 의료·금융·제조 등 실제 도메인과 결합해 자본 회수와 경제적 가치 실현으로 이어지는 단계다.

<그림 1> 엔비디아의 5단 케이크(5-LAYER CAKE) 모델



자료: <https://blogs.nvidia.co.kr/en/blog/ai-5-layer-cake/>(검색일: 2026.7.15.)

인프라의 물리적 격변 :

전력·냉각·부지 그리고 하드웨어 경쟁력

데이터센터가 지능의 생산 공장으로 진화함에 따라, 하드웨어 인프라 영역에서도 전통 제조업의 공장 건설에 준하는 물리적 변화와 제약이 발생하고 있다.

1) 기가와트(GW)급 전력 공급의 한계와 분산화

초거대 AI 모델이 구동되는 데이터센터는 GW 단위의 전력망 구축을 요구한다. 국제에너지기구(IEA)는 전 세계 데이터센터 전력 소비가 2024년 415TWh에서 2030년 945 TWh로 두 배 이상 증가할 것으로 전망했다.² 도심 인근 부지의 확보가 사실상 불가능해지면서, 데이터센터 입지는 발전 설비 인근이나 송전 인프라가 안정적인 지역으로 분산되는 추세다. 이에 따라 재생에너지 단지와의 연계, 소형 가스터빈 등 분산형 전원의 내재화도 대안으로 시도되고 있다.

2) 열 제어 솔루션: 액체 냉각(Liquid Cooling)의 도입

서버 랙의 전력 밀도 급상승으로 발생하는 고열의 제어는 공장 가동률을 결정하는 핵심 공정이다. 공

랭식의 한계를 넘기 위해, 비전도성 액체에 서버를 직접 담그는 액침 냉각(Immersion Cooling)이나 냉각수가 칩의 바닥면을 순환하는 수랭식 공법이 보편화되고 있다.³

3) 노후 공장 부지의 전환과 재활용

물리적 입지 측면에서도 구조적 변화가 관찰된다. 전통 제조업 쇠퇴로 유휴화된 철강 공장이나 화력 발전소 부지를 최첨단 AI 데이터센터로 전환하는 사례가 늘고 있다. 해당 부지는 대용량 전력 공급을 위한 고압 송전선로와 냉각수용 수자원을 이미 갖추고 있어, 건설 비용 절감과 인허가 기간 단축을 위한 실효성 있는 대안으로 평가받는다.

4) 고대역폭 메모리(HBM)와 고효율 가속 인프라의 경쟁력

전력·냉각과 함께 AI 팩토리의 성능을 좌우하는 또 하나의 축은 하드웨어 자체의 효율이다. AI 가속기의 연산 성능은 곧 메모리 대역폭에 의해 제약되며, 이 병목을 해소하는 핵심 부품이 고대역폭 메모리(HBM)다. HBM은 우리나라 반도체 기업이 세계 시

<표 2> AI 팩토리 전환에 따른 인프라 물리적 격변과 대응 방향

격변 이슈	핵심 내용	대응방향 및 시사점
GW급 전력 공급·계통	데이터센터 전력소비 급증, 도심 입지 곤란	발전설비 인근 분산 배치, 분산형 전원 내재화
열제어(냉각)	고밀도 랙 발열 → 공랭식 한계, 액침·수랭식 보편화	전력효율(PUE) 개선, 랙 밀도 소형화 유리
부지확보 및 전환	유휴 철강·화력부지 전환 (송전선로·수자원 기 확보)	건설비 절감, 인허가 단축의 실효적 대안
HBM·하드웨어	연산 성능의 메모리 대역폭 병목 → HBM이 핵심(우리나라 주도)	고효율 AI 가속 인프라 패키지 공급으로 도약

자료: 본문 내용 정리

2 국제에너지기구(IEA), “Energy and AI”,(2025.4)

3 MarketsandMarkets, “Data Center Liquid Cooling Market-Global Forecast to 2029”

장을 사실상 주도하는 소수 영역으로, AI 팩토리 시대에 우리가 하드웨어 계층에서 확보한 가장 강력한 교두보다. 여기에 고효율 전력 반도체, 정밀 냉각 설비, 저전력 추론 가속기까지 결합하면, 우리나라는 단순 부품 공급을 넘어 ‘고효율 AI 가속 인프라’를 하나의 패키지로 제공하는 공급자로 도약할 수 있다.

기업의 생존 체력: 데이터 보안과 산업 특화형 버티컬 AI

데이터센터의 공장화는 개별 기업의 경쟁력과 직결된다. 범용 AI 모델의 활용 단계를 넘어, 기업이 자체 특화 인프라를 구축하고 차별화된 지능을 확보해야 하는 현실적 이유가 여기에 있다.

1) 데이터 보안과 지적재산권(IP) 보호

생성형 AI 도입 시 기업이 직면하는 핵심 장벽은 ‘보안’이다. 퍼블릭 AI 서비스의 경우, 민감 정보의 유출 위험이 존재한다. 이에 따라 기업들은 독립된 전용 데이터센터나 격리된 프라이빗 클라우드(Private Cloud)에 독자 AI 인프라를 구축하려 한다. 보안이 보장된 자체 인프라인 경우에만 핵심 데이터의 안전한 처리와 생산이 가능하기 때문이다.

2) 범용 지능의 한계와 버티컬 AI에서의 우리의 강점

빅테크의 범용 AI는 고도의 정밀성이 요구되는 산업

현장 적용에 명확한 한계를 지닌다. 반도체 미세 공정 최적화, 조선 공정의 정밀 제어 등은 일반적 정보 학습만으로 불가능하다. 이 지점에서 우리나라 산업계가 지닌 풍부한 제조 현장 데이터와 깊이 있는 도메인 역량은 차세대 버티컬 AI(Vertical AI) 시장을 선점할 강력한 무기다.⁴ 공정 엔지니어링 지식과 설비 제어 노하우(OT)를 특화 모델에 주입하고 전용 AI 팩토리 연계할 때, 우리나라는 범용 기술의 격차를 넘어 특정 산업 영역에서 압도적 지배력을 확보할 수 있다.

3) 기업 전용 AI 인프라의 구축

자사 사업에 정확히 부합하는 고부가가치 산출물(토론)을 얻으려면, 기업 특화 프로세스를 학습한 독자 소형언어모델(sLLM)과 이를 구동할 맞춤형 인프라가 필수적이다. 이때 ‘전용 인프라’는 국가 차원의 소버린 AI(Sovereign AI)와는 차별화된, 기업 단위의 프라이빗 AI를 뜻한다. 독자 인프라 확보 여부는 디지털 전환기 기업의 중장기 생산성 격차를 규정하는 핵심 척도가 된다.

4) 기업이 직면한 그림자: 하단 스택 종속과 과투자 리스크

다만 이러한 전환에는 하방 위험이 병존한다.

첫째, 5단 케이크의 하단(칩·인프라)이 특정 해외 플랫폼에 강하게 종속되어 있어, 애플리케이션 계층으

<표 3> 기업 특화 AI 인프라의 3대 기회와 위험 요인

기업 특화 AI 인프라 구축 - 3대 기회 요인		
① 데이터 보안·IP 전용·프라이빗 환경에서 민감 정보 안전 처리	② 버티컬 AI 제조 도메인(OT) 데이터로 특정 산업 시장 선점	③ 기업 전용 AI 독자 sLLM·맞춤 인프라로 중장기 생산성 격차 결정
↓		
그림자 - 병존하는 하방 위험		
① 종속의 역설 하단 스택(칩·인프라)의 해외 플랫폼 의존 심화	② 과투자 위험 수익화 지연 시 좌초자산·재정 부담	③ 계통·인허가 병목 전력 과부하·인허가로 투자 일정 지연

자료: 본문 내용 정리

로 올라갈수록 외국 생태계 의존이 심화되는 ‘중속의 역설’이 나타날 수 있다. 둘째, GW급 전력과 수조 원대 컴퓨팅에 대한 투자는 애플리케이션 계층의 수익화가 지연될 경우, 대규모 좌초 자산과 재정 부담으로 되돌아올 수 있다. 셋째, 전력 계통의 국지적 과부하와 인허가 병목은 투자 일정 자체를 지연시키는 상시적 위험 요인이다. 이러한 위험을 관리하는 일이 AI 전환 시대 재정 정책의 출발점에 해당한다.

AI 팩토리 시대의 재정 정책 과제

지능 생산 기반의 구축은 본질적으로 대규모 자본 투입을 전제하며, 그 상당 부분이 공공 재정과 정책 금융의 몫이다. AI 팩토리 시대의 재정 정책은 ‘공급 부문 투자’와 ‘수요 부문 지원’을 이원적으로 설계하되, 재정 부담과 민간 유인 사이의 정교한 균형 조율이 필요하다.

1) 공급 부문의 투자:

국가 컴퓨팅·전력망의 재정 소요와 배분 원칙

정부는 2026년 AI 관련 예산을 총 9.9조 원 규모로 편성해 전년 본예산(3.3조 원)의 약 세 배로 확대했고, 이 가운데 GPU 등 컴퓨팅 자원 확보에 2조 원대 사업(AI 컴퓨팅 자원 활용 기반 강화 사업 등)을 배정하여 2028년까지 첨단 GPU 5만 장 이상 확보를 추진하고 있다.⁴ 총사업비 2.5조 원 규모의 국가 AI 컴퓨팅 센터도 민관합작 특수목적법인(SPC) 방식으로 추진된다. 재정 관점의 핵심 원칙은 세 가지다. 첫째, 개별 기업이 산발적으로 GPU를 조달하기보다 국가가 대량 구매로 협상력과 수급 안정성을 확보하는 ‘규모의 경제’가 재정 효율에 유리하다. 둘째, 자산의 소유·운영 구조를 명확히 하여 재정 투입 자산이 저활용되지 않도록

록 가동률과 성과 지표(SLA)를 관리해야 한다. 셋째, 전력망·계통 보강 투자는 데이터센터 개별 사업이 아니라 국가 기간 인프라 관점에서 장기 재정 계획에 편입되어야 한다.

2) 수요 부문의 지원:

컴퓨팅 바우처와 중소·중견기업 격차 해소

공급 기반을 확충하더라도, 이를 활용할 수요 기반이 취약하면 투자 효과는 반감된다. 특히 자체 GPU 확보가 어려운 중소·중견 제조기업의 컴퓨팅 격차 해소가 관건이다. 국가가 확보한 컴퓨팅 자원을 학계·연구계에는 무상, 산업계에는 시장가 대비 낮은 요율로 제공하는 방식과, AI·클라우드 바우처를 통합한 ‘윈스톱 바우처’ 같은 수요 지원 수단이 이미 가동되고 있다. 재정 지원은 단순 요금 할인을 넘어, 도메인 데이터의 표준화와 특화 모델 개발까지 연계하는 ‘패키지형 바우처’로 고도화될 필요가 있다.

3) 균형점: 재정 부담과 민간 유인의 조율

모든 소요를 재정으로 충당하는 것은 불가능하며 바람직하지도 않다. 정부는 AI를 조세특례제한법상 국가전략기술로 지정해 투자 세액공제를 강화하고, 2026년 제정된 AI 데이터센터 특별법을 통해 전력·용수·인허가 등 비재정적 병목을 완화하는 방향으로 정책 조합을 설계하고 있다. 재정의 역할은 직접 투자와 세제·보증을 통한 민간 유인, 그리고 민간 위험 부담(Risk-Sharing)을 정교하게 배합하여, 최소한의 재정으로 최대한의 민간 투자를 유발하는 마중물이 되는 데 있다. 국가 컴퓨팅 센터의 민간 지분 확대 개편은 이러한 균형 모색의 한 사례에 해당한다.

4 한국지능정보사회진흥원(NIA), “2025-2026 글로벌 AI 인프라 전략 및 소버린 AI 트렌드 분석”

5 과학기술정보통신부, “2026년 AI컴퓨팅 자원 활용기반 강화사업”(2026.3)