



AI 심장, 데이터센터 : 정책·산업·기술 동향과 우리의 과제¹⁾

1 AI 시대, 데이터센터가 국가 핵심 인프라로 부상

가. 생성형 AI 확산과 컴퓨팅 수요의 폭증

➔ 데이터 폭발과 연산 수요의 패러다임 전환

- 초거대 언어모델(LLM)과 생성형 AI의 발전으로 글로벌 컴퓨팅 연산 수요가 기하급수적으로 폭증하고 있음
- 과거의 연산이 정형 데이터의 단순 처리와 저장에 집중되었다면, 현재는 수천억 개의 파라미터를 갖춘 모델의 학습(Training)과 실시간 추론(Inference)을 위한 대규모 병렬 연산을 요구함

➔ 컴퓨팅 자원의 희소성 심화

- 대규모 언어모델이 컨텍스트 길이와 동시 사용자 수를 확장함에 따라, 그래픽 처리장치(GPU) 및 고대역폭메모리(HBM) 수요가 공급을 압도하고 있음
- 이는 단순한 칩 확보 경쟁을 넘어 칩을 가동할 전력, 냉각, 통신망이 집적된 ‘물리적 인프라’의 공급 병목 현상으로 이어짐

나. 전통 데이터센터에서 AI 데이터센터로의 전환

➔ 아키텍처의 근본적 변화

- 기존 클라우드 데이터센터가 중앙처리장치(CPU) 중심의 범용 워크로드를 위해 설계되었다면, AI 데이터센터는 GPU, 신경망처리장치(NPU) 등 고밀도·고전력 연산에 특화된 가속기 기반 구조로 진화함

➔ 전력 밀도와 물리적 스트레스 급증

- 전통적 데이터센터의 랙(Rack)당 전력 밀도는 평균 3~10kW 수준이었으나, AI 데이터센터는 40kW에서 120kW 이상의 초고밀도 전력을 요구함
- 이러한 부하 패턴은 전력망에 극심한 과도 상태(Transient) 스트레스를 유발하며, 인프라의 전면적인 재설계가 불가피함²⁾

1) 정보통신기획평가원 디지털인프라단 디지털인프라전략팀 백선경 팀장(sunkyoung100@iitp.kr)
본 고는 저자의 개인적인 견해이며 과학기술정보통신부와 KISTEP의 공식적인 의견이 아닙니다.

다. AI 데이터센터를 둘러싼 글로벌 투자 경쟁

⇒ 하이퍼스케일러 및 자본의 대규모 유입

- 하이퍼스케일러 기업들은 글로벌 AI 주도권을 확보하기 위해 연간 수천억 달러를 AI 인프라에 투자하고 있으며, 투자 규모는 지속 확대 중
 - ※ '25년 대비 '26년 투자 규모(달러): (마이크로소프트) 800억 → 1900억, (아마존) 1,000억 → 2,000억, (구글) 750억 → 1,850억 수준
- 전 세계 AI 데이터센터 인프라 투자 규모는 '30년까지 약 5조 5천억 달러 (약 8,030조 원)에 이를 것으로 전망됨³⁾
- 미국·EU·UAE 등 주요국은 국가 주도로 대규모 AI 인프라에 투자 중임

〈 주요국 AI 인프라 투자 현황 〉

미국	EU	UAE	싱가폴	대만	사우디
스타게이트(Stargate) 프로젝트	AI대륙 액션플랜(AI Continent Action Plan)	국부펀드 전용 투자사 MGX	친환경 데이터센터 로드맵 (Green DC Roadmap)	TSMC 중심 반도체 및 AI 파운드리 인프라 구축	트랜센더스 프로젝트 & 국부펀드
최대 5,000억 달러 (약 730조원)	총 2,000억 유로 (약 310조원)	최대 1,000억 달러 (약 146조원)	기존 1.4GW + 추가 300MW~1.5GW 전력 공급 목표	연간 600억 달러 ↑ (약 88조원)	최대 1,000억 달러 (약 146조원)

출처 : 저자 작성

⇒ '전력-부지-냉각' 선점 전쟁

- 최신 AI 데이터센터 경쟁의 핵심은 칩 구매력이 아니라 '전력 및 수자원'의 안정적 확보로 이동함
- 이에 따라 대형 원자력 발전 및 소형모듈원전(SMR)과 결합된 오프그리드 (Off-grid) 데이터센터 투자가 폭발적으로 증가하고 있음

라. AI 데이터센터의 국가 전략적 의미

⇒ AI 주권(Sovereign AI)과 컴퓨팅 주권

- 자체적인 기초 모델과 이를 구동할 컴퓨팅 인프라를 보유하지 못한 국가는 기술적·경제적·문화적 종속을 피할 수 없음
- AI 데이터센터는 민간 상업 시설을 넘어 국가의 '디지털 영토'를 결정짓는 핵심 안보 자산으로 격상됨

⇒ 산업 경쟁력과 AI 생태계

- 대규모 컴퓨팅 자원 접근성은 자국 스타트업, 학계, 연구기관의 혁신 속도를 좌우함

2) 다우 IDC(<https://blog.daoudc.com/>), AI 데이터센터란 무엇인가?(25.11)

3) JP모건, Barron's·Dow Jones·Investor's Business Daily(26.6)

- 고성능 인프라가 구축된 국가와 지역으로 인재와 자본이 집중되는 ‘클러스터 효과’가 극대화되고 있음

⇒ 반도체·전력·통신·클라우드 산업과의 연계

- AI 데이터센터는 첨단 반도체(초고속 메모리, NPU), 차세대 전력망, 초저지연 통신 패브릭, 액체 냉각 설비 등 가치사슬 전반의 기술이 융합된 최종 집결지로, 인프라 구축 경쟁력이 곧 유관 산업의 글로벌 패권으로 직결됨

2 AI 데이터센터의 개념과 산업 생태계

가. AI 데이터센터의 정의와 특징

⇒ (정의) 방대한 비정형 데이터의 고속 처리 및 초저대 AI 모델의 분산 학습·추론 워크로드에 최적화되도록 설계된 특수 목적의 컴퓨팅 인프라

⇒ (특징) ▲수천~수만 개의 AI 가속기(GPU/NPU/TPU/DPU 등) 병렬 연결, ▲초저지연 고대역폭 네트워크(InfiniBand, Ultra Ethernet 등) 적용, ▲랙당 40~120kW 이상의 초고전력 밀도, ▲고발열·고열밀도 문제를 해소하기 위한 액체 냉각(Liquid Cooling) 기반의 열 관리 시스템이 필수적임

〈 기존 클라우드 데이터센터와 AI 데이터센터의 차이 〉

구분	기존 클라우드 데이터센터	AI 데이터센터
주요 워크로드	트랜잭션, 웹 서비스, 범용 IT (비동기적 처리)	AI 모델 학습(Training) 및 대규모 추론 (동기화된 집단 연산)
핵심 프로세서	CPU 기반	GPU, NPU 등 AI 가속기 기반
랙당 전력 밀도	평균 3 ~ 10kW	40 ~ 120kW 이상 (초고밀도)
네트워크 구조	이더넷(Ethernet) 중심, 트래픽 지연 내성 존재	InfiniBand, NVLink, Ultra Ethernet 등 초저지연 무손실 패브릭 필수
주요 냉각 방식	공랭식 (CRAC/CRAH)	직접액체냉각(D2C), 액침냉각(Immersion)
인프라 부하 특성	무작위 분포(Random Distribution) 형태의 부하	칩들이 동시에 연산하는 동기화된 집단행동(High di/dt) 발생

출처 : 저자 작성

나. AI 데이터센터의 주요 유형

〈 AI 데이터센터 유형별 특징 〉

유형	특징
AI 학습(Training) 중심	• 수만 개의 GPU를 거대한 단일 클러스터로 묶어 파운데이션 모델을 훈련하는 시설 - 지연시간보다 무중단 대규모 전력 공급이 중요하여 부지와 전기료가 저렴한 외곽 지역에 하이퍼스케일 규모로 구축됨
AI 추론(Inference) 중심	• 완성된 모델을 바탕으로 사용자 질의에 실시간 응답하는 시설 - 낮은 지연시간(Latency) 확보가 필수적이므로 사용자 밀집 지역이나 도심 근교에 엣지(Edge) 형태로 구축되는 경향이 큼
Hyperscale / Colocation / Sovereign AI	• 대형 테크 기업의 자가 사용형(하이퍼스케일), 공간 및 설비를 임대하는 상업용(코로케이션), 국가 데이터 주권 보호를 위해 정부 주도로 구축하는 폐쇄형 인프라(소버린 AI)로 구분됨
Neocloud (네오클라우드)	• AI 연산에만 100% 특화하여 GPU를 서비스형(GPUaaS)으로 제공하는 신흥 클라우드 - 범용 클라우드의 가상화 오버헤드를 제거한 베어메탈 인프라를 제공하며, 기존 하이퍼스케일러 대비 동급 인프라 사용료를 약 65% 절감하는 압도적 비용 경쟁력으로 급성장 중임⁴⁾ - 코어위브(CoreWeave), 람다랩스(Lambda Labs), 크루소(Crusoe), 네비우스(Nebius) 등이 대표적임

출처 : 저자 작성

다. AI 데이터센터 산업의 가치사슬

〈 AI 데이터센터 산업 가치사슬 〉

분류	기업 형태
AI 반도체	• AI 가속기*, 데이터 프로세싱 유닛(DPU)**, 고대역폭메모리(HBM)*** 등 반도체 설계 및 제조 기업 * 대규모 AI 학습·추론 연산을 수행하는 GPU·NPU·TPU 등 ** 데이터 이동·네트워크·스토리지 처리를 가속하여 연산 효율을 높이는 장치 *** AI 가속기의 메모리 병목을 완화하는 메모리 장치
AI 반도체 솔루션	• CPU·가속기·메모리 간 고속 연결 및 메모리 확장·공유를 지원하는 CXL(Compute Express Link) 기반 컨트롤러·스위치 및 메모리 솔루션 등을 설계·제조하는 반도체 및 관련 솔루션 기업
서버	• GPU·NPU·TPU 등 고성능 AI 가속기와 CPU, HBM 및 고속 인터커넥트를 집적한 고밀도 AI 서버·랙 시스템을 설계·통합하는 서버 제조사 및 시스템 통합 업체
스토리지	• AI 학습·추론 과정에서 발생하는 대규모 데이터셋과 모델 파라미터를 신속하게 공급·저장하기 위한 NVMe SSD, NVMe-oF, 병렬 파일시스템 및 분산 스토리지 기반의 고성능·대용량 스토리지 시스템 공급 업체
네트워크	• 수천~수만 개의 AI 가속기 및 서버 노드를 하나의 대규모 클러스터로 연결하고, 분산 학습 과정에서 발생하는 대용량 데이터 교환과 집단통신(Collective Communication)의 병목을 최소화하기 위한 초저지연·고대역폭 네트워크 인프라 공급 업체
전력·냉각	• 랙당 수십~100kW 이상으로 증가하는 AI 데이터센터의 초고전력 밀도와 이에 따른 고발열·고열밀도 문제를 안정적으로 처리하기 위한 전력 공급·분배 및 열관리 인프라 공급 업체

4) 디지털서비스(<https://www.digitalmarket.kr/>), 네오클라우드- AI 시대의 새로운 인프라 패러다임(26.1)



분류	기업 형태
데이터센터 설계·건설	AI 서버의 초고중량·고전력·고발열 특성을 반영해 데이터센터의 부지 선정부터 전력 인입, 건축, 기계·전기·배관(MEP), 냉각 및 네트워크 인프라까지 통합 설계·시공하는 엔지니어링·EPC(Engineering-Procurement -Construction)·건설업체
클라우드·AI 플랫폼	- 대규모 GPU·AI 가속기 인프라를 구축하고 이를 기업·개발자에게 서비스 형태로 제공하는 하이퍼스케일러 - 클라우드 서비스 사업자(CSP) - GPU 클라우드 및 전문 AI 인프라 서비스 사업자
운영·관리 소프트웨어	AI 데이터센터 내 서버·네트워크·전력·냉각 설비를 실시간으로 모니터링하고 자원 활용률과 에너지 효율을 최적화하는 DCIM*, BMS**, EMS** 및 AI 기반 운영 소프트웨어 공급 업체 * Data Center Infrastructure Management ** Building Management System *** Energy Management System

출처 : 저자 작성

라. 주요 사업자 및 산업구조 변화

⇒ 기존 클라우드 인프라는 AWS, MS, Google 등 소수 하이퍼스케일러가 지배했으나, AI 특화 네오클라우드(코어위브, 람다 등)가 엔비디아의 전략적 투자와 최신 칩 우선 할당 지원을 받으며 새로운 강력한 축으로 부상함

- H100, GB200 NVL72, Vera Rubin NVL72 등 최신 GPU 플랫폼을 대형 클라우드 사업자와 동등하거나 일부 세대에서는 더 빠르게 도입·상용화하며 시장 영향력을 확대

⇒ 빅테크 기업들 역시 엔비디아 의존도를 낮추기 위해 자체 맞춤형 AI 반도체(Custom Silicon) 개발을 강화하는 한편, 전력·냉각 설비부터 네트워크 인프라까지 데이터센터 아키텍처 전반을 내재화(수직계열화)하려는 시도를 가속화하고 있음

- Google, AWS, Microsoft, Meta 등 글로벌 빅테크 기업들은 TPU, Trainium, Maia, MTIA 등 자체 맞춤형 AI 반도체(Custom Silicon) 개발을 확대
- 나아가 자체 칩 간 인터커넥트와 AI 클러스터 네트워크, 서버·랙 시스템, 전력·액체 냉각 설비 및 소프트웨어 스택까지 공동 설계(Co-design)하는 추세임

3 AI 데이터센터 핵심 기술 동향

가. AI 컴퓨팅 기술

⇒ 대규모 모델 분산 처리 및 3D 병렬화(3D Parallelism) 고도화

- 최신 가속기들은 단일 칩 내 연산 능력의 극대화를 넘어, 거대한 모델을 여러 GPU로 쪼개어 연산하는 텐서 병렬화 파이프라인 기술을 필수적으로 적용함

- 데이터, 텐서, 파이프라인을 결합한 3D 병렬화(DeepSpeed, Megatron-LLM 등)를 통해 수만 개 규모의 GPU 클러스터 컴퓨팅 효율을 극대화할 수 있음⁵⁾

⇒ 초고성능 가속기 고도화

- 대규모 AI 연산에 특화된 범용 GPU 및 커스텀 ASIC/NPU 도입이 가속화됨
※ NVIDIA Blackwell/Rubin, Google TPU, AWS Trainium/Inferentia 등
- 저정밀도 연산 포맷(FP8, FP4, MXFP8/Microscaling formats) 표준화로 컴퓨팅 성능 극대화 및 메모리 사용량 절감의 이점이 있음

⇒ 초고대역폭 메모리 및 3D 패키징

- HBM3E 상용화 및 HBM4(2048-bit 인터페이스, 로직 다이 통합) 전환으로 가속기 메모리 대역폭 한계를 돌파함
- CoWoS(Chip-on-Wafer-on-Substrate) 및 2.5D/3D 이종 집적 패키징 기술을 통한 칩렛(Chiplet) 구조의 컴퓨팅 밀도를 향상시킴

⇒ CXL(Compute Express Link) 기반 메모리 확장 및 풀링

- CXL 2.0/3.0 인터페이스를 활용한 메모리 확장 및 풀링 기술로 AI 서버 내 메모리 단편화 해소 및 동적 리소스 할당을 구현함
※ CXL 4.0 규격은 PCIe 7.0을 기반으로 128GT/s 전송 속도를 내며, 여러 물리 포트를 묶어 1.5TB/s 대역폭을 제공하는 ‘번들 포트’를 도입함⁶⁾

나. AI 서버 및 시스템 아키텍처

⇒ 초고밀도 랙 스케일 아키텍처(Rack-Scale Architecture)로의 진화

- 개별 서버 단위 설계에서 탈피하여, 엔비디아의 NVL 72, OCP Open Rack v3 등과 같이 랙 전체를 하나의 거대한 단일 컴퓨팅 노드로 통합하는 추세임
 - 초거대 AI 모델을 랙 내부에서 단일 시스템처럼 처리하여 지연을 최소화하고, 랙당 100kW~130kW 이상의 초고밀도 전력 및 냉각을 통합 관리하는 구조를 채택하고 있음

⇒ 서버 내 이종 프로세서 간 호환성 및 개방형 폼팩터 표준화

- 이종 프로세서 간의 메모리 공유 및 캐시 일관성 확보를 위해 CXL 및 PCIe 6.0/7.0 기반의 범용 인터커넥트 기술 도입이 가속화됨
- OCP 주도의 OAM(OCP Accelerator Module) 규격 및 UCIe(Universal Chiplet Interconnect Express) 칩렛 표준을 통해 이종 가속기를 동일 메인보드나 패키지 내에서 혼용 및 교체할 수 있는 상호운용성이 강화됨⁷⁾

5) NVIDIA, Megatron-LM & 3D Parallelism, Blackwell Architecture Technical Brief(‘24)

6) Intel Blog(<https://intel.com/>), CXL 4.0 인프라 계획 가이드: AI 스케일을 위한 메모리 풀링(‘26.3)

➔ 시스템 확장성(Scalability) 이슈 극복 아키텍처 도입

- 스케일업/아웃의 물리적 한계를 극복하기 위해 연산, 메모리, 스토리지를 물리적으로 분리한 뒤, 고속 패브릭망을 통해 워크로드 요구량에 맞춰 동적으로 자원을 할당 및 조합하는 기술을 적용함

다. 초고속 네트워크 기술

➔ 초고속 가속기 간(Scale-up) 인터커넥트: NVLink 및 UALink

- NVIDIA 독자 규격인 고대역폭 NVLink 기반 랙 단위 통합 기술에 대항하는 개방형 가속기 연결 표준인 UALink 도입 경쟁이 가속화됨

➔ 대규모 분산 클러스터(Scale-Out) 백본망 및 UEC(Ultra Ethernet Consortium) 이더넷

- 대규모 통신에 독보적인 Infiniband(Quantum-X800 등) 환경에서, AI 워크로드 처리에 특화된 UEC기반 차세대 이더넷(RoCE: RDMA over Converged Ethernet) 기술 등장으로 주도권 경쟁 중

➔ 광 인터커넥트(Optical Interconnect) 및 CPO(Co-Packaged Optics) 기술 적용

- 고대역폭·초저지연 백본망의 800Gbps 도입 확산 및 1.6Tbps 스위치/트랜시버 전환 가속화 추세에 따라 구리선의 신호 감쇄와 전력 소모 한계 극복이 필요함
- 이를 극복하기 위해 스위치 칩(ASIC) 패키지 내부에 광학 소자를 직접 통합하는 CPO 및 LPO(Linear-drive) 기술 도입을 시작함

➔ AI 최적화 패브릭 아키텍처

- 다대다(All-to-All) 통신 병목 해소를 위한 레일 최적화 Leaf-Spine, 초거대 AI 클러스터의 Dragonfly+, Leaf-Spine 다계층 스위칭 구조인 Non-blocking Fat-Tree 토폴로지 등을 채택하여 설계 중임

라. 스토리지 및 데이터 인프라 기술

➔ 초고속 병렬 파일 시스템 및 AI 전용 스토리지

- 수만 개 GPU에 데이터 공급이 지연되지 않도록 분산 병렬 파일 시스템(General Parallel File System) 필수화되는 추세임
- 텍스트, 이미지 등 페타바이트급 비정형 데이터를 AI 노드에 병목 없이 공급하기 위한 NVMe 기반의 올-플래시 스토리지 기술이 적용됨

7) OCP, OCP Open Rack v3(ORV3) Specification/UCI Consortium, Universal Chiplet Interconnect Express Specification

➔ GPU Direct Storage(GDS) 아키텍처

- CPU 메모리를 거치지 않고 NVMe 스토리지와 GPU 메모리 간 직접 DMA (Direct Memory Access) 전송을 통해 I/O 지연 시간 단축 및 CPU 부하 경감 효과를 극대화할 수 있는 구조임

마. 전력 인프라 기술⁸⁾

➔ 고전압 직류(HVDC) 및 랙 레벨 배전 효율화

- 교류-직류(AC-DC) 변환 단계를 최소화하여 전력 손실을 방지하고, 배전 설비가 차지하는 물리적 공간을 절감하여 컴퓨팅 인프라 밀집도 최적화를 위한 기존 48V DC 및 400V DC 버스바 랙 직결 구조 기술이 확산되고 있음

➔ 에너지 저장 시스템(BESS) 및 고효율 UPS(Uninterruptible Power Supply)

- BESS 결합과 수십 MW급 대용량 모듈러 UPS 수요가 증가함
 - 대규모 연산에 따른 순간적인 전력 스파이크 발생 시 즉각적으로 부하를 흡수하여 핵심 자원의 무중단 운영 및 설비 생존성 보장이 목적

➔ 전력망 병목 해소를 위한 온사이트 전력 생산

- AI 데이터센터는 지속적이고 안정적인 기저 부하를 요구하므로, 외부 전력망에 의존하지 않는 자체 마이크로그리드 시스템이 대두됨
- 대규모 전력 인입 지연 대응을 위해 SMR(소형 모듈 원자로), 천연가스 터빈, 수소 연료전지 기반 자가발전 연계 검토가 증가하는 추세임

바. 냉각 기술

➔ 액체 냉각(Liquid Cooling)의 필수화⁹⁾

- 공랭(Air Cooling)의 한계에 따라 D2C(Direct-to-Chip/Cold Plate) 냉각이 주류로 정착하고 있음
 - 발열이 심한 CPU 및 GPU 표면에 쿨드플레이트를 부착해 직접 냉각수를 순환시킴
 - ※ 공랭식 대비 전력 소비를 50% 이상 절감하며, 랙당 100kW 이상의 열 부하를 처리할 수 있어 최신 엔비디아 GB200 서버 등의 주력 냉각 방식으로 채택됨

8) Uptime Institute, Global Data Center Survey “Power and Resiliency Trends”/OCP, Open Rack v3 Power/DC Busbar Specification

9) ASHRAE TC 9.9, Liquid Cooling Guidelines for Datacom Equipment Centers

➔ 액침 냉각(Immersion Cooling)

- 서버 전체를 비전도성 절연 냉각액에 담가 열을 원천적으로 흡수하는 기술로 단상 및 2상 액침 냉각 방식이 있음
 - 팬이 필요 없어 소음이 없고, 냉각에 소모되는 전력을 90% 이상 획기적으로 줄여 데이터센터 전력사용효율(PUE)을 이론적 한계인 1.05 수준까지 낮출 수 있는 차세대 게임 체인저로 평가¹⁰⁾

➔ CDU(Cooling Distribution Unit) 다중화 및 폐열 회수

- 시스템 장애 시에도 냉각수가 지속 공급되도록 배관 및 펌프를 다중화(N+1)하고, 정밀 유량·온도 제어가 가능한 차세대 CDU 기술이 부상하고 있음
- 데이터센터 발생 온수(폐열)를 인근 지역난방/스마트팜으로 공급하는 열 회수 시스템 연계로 폐열의 재활용 기술도 등장함

사. 친환경·에너지 기술

➔ 핵심 지속가능성 지표 관리

- 막대한 수자원 낭비를 줄이기 위해 외부 보충수 공급이 필요 없는 ‘폐쇄형 순환 무수 냉각(Closed-loop Zero Water Cooling)’ 시스템 전환이 시도되고 있음
 - PUE(전력효율지수) 1.1 이하 목표, WUE(물사용효율) 0에 수렴하는 무수 냉각 시스템 및 CUE(탄소사용효율) 등 관리를 강화하고 있음¹¹⁾

➔ 24/7 무탄소 에너지(CFE, Carbon-Free Energy)

- 기존 RE100의 한계를 탈피하여, 시간 단위로 재생에너지와 전력 소비를 일치시키는 차세대 조달 표준 24/7 CFE 전력구매계약(PPA)으로 차별화를 추진 중¹²⁾

아. AI 데이터센터 운영·최적화 기술

➔ 디지털 트윈 및 지능형 DCIM

- AI 데이터센터와 동일한 가상 환경에서 AI 컴퓨팅 자율 운영, 열유체 및 전력 부하를 사전 예측, 대규모 장비 교체 시 물리적 위험 요소 사전 방지가 가능한 기술 도입
 - 센서 데이터를 기반으로 인프라의 디지털 트윈을 구축하고, AI 알고리즘을 통해 핫스팟(발열 집중 구간)을 사전에 감지하여 유량을 조절하는 예측 냉각 기술이 도입되어 물리적 장애를 예방함

10) OCP, Cooling Environments Project: Immersion Cooling and Cold Plate Requirements

11) The Green Grid, PUE, WUE, CUE Metrics and Guidelines

12) UN Sustainable Energy for All(<https://www.seforall.org/>), 24/7 Carbon-Free Energy Compact

⇒ AI 기반 전력·냉각 자동화 제어

- 시설 외부 환경 변수와 데이터 트래픽 폭증 양상을 실시간 학습하여, 개입 없이 최적의 냉각 인프라 가동률을 스스로 산출하는 자율형 제어가 가능함

⇒ 지능형 워크로드 오케스트레이션

- 개별 노드의 상태를 초 단위로 감시하여 연산 자원의 병목을 해소하고, 유휴 자원을 중요 임무에 즉각 재할당 처리를 위한 소프트웨어(Kubernetes/Slurm 등) 기반 분산 오케스트레이션 고도화는 필수적임

자. 보안 및 신뢰성 기술

⇒ 기밀 컴퓨팅(Confidential Computing) 및 RoT(Root of Trust)

- 하드웨어 기반 신뢰 실행 환경을 GPU 메모리로 확장하여 데이터 학습·추론 중 민감 데이터 유출 전면 통제 가능한 보호 기술
- 칩 부팅의 가장 첫 단계부터 하드웨어(RoT 칩)적으로 펌웨어를 검증하여, 외부로부터의 악의적인 백도어 삽입 및 변조 시도를 원천 봉쇄함

⇒ 네트워크 실시간 암호화(MACsec) 및 양자 내성 암호(PQC) 도입

- AI 데이터센터 간 트래픽 전송 시 하드웨어 가속을 통해 지연 없는 실시간 암호화를 수행하고, 미래 양자 컴퓨터의 해킹 위협까지 방어 가능한 최상위 보안망 구축을 지원함

⇒ 장애 예측 및 자율 복구

- 하드웨어 파손 전 발생하는 미세한 이상 징후를 선제 감지하여, GPU 통신 오류 발생 시 정상 노드로 워크로드를 자동 이관하는 무중단 방어 체계 확립 고려도 필수적임

4 글로벌 AI 데이터센터 산업 및 정책 동향

가. 글로벌 시장 및 산업 동향

⇒ 하이퍼스케일러 설비투자(CAPEX*) 폭발적 증가¹³⁾

* Capital Expenditures: 기업이 공장, 기계장치, 건물, IT 인프라, 소프트웨어 등 고정자산을 구매하거나 업그레이드하기 위한 지출

13) Synergy Research Group, Hyperscale Spending Spree ('25.12) / CBRE Investment Management, Data Centers: Ain't No Mountain High Enough ('26.2)

- 클라우드 선도 4사(Amazon, Microsoft, Google, Meta)를 중심으로 대규모 언어 모델(LLM) 학습 및 추론 인프라 확보 경쟁이 심화되고 있음
 - 글로벌 하이퍼스케일러들의 분기별 설비투자 규모가 2025년 1,400억 달러를 돌파하며 역대 최대치 경신
- 기존 일반 클라우드 워크로드용 데이터센터 비중은 축소되고, AI 전용 초고밀도 인프라(서버, 네트워크, 공간)를 선점하기 위한 10년 이상의 장기 임대 계약 및 포트폴리오 전환이 글로벌 시장 성장을 주도하고 있음

➔ GPUaaS(GPU-as-a-Service) 신흥 강자 부상 및 인프라 공급망 양극화

- 대규모 자본을 바탕으로 NVIDIA 최신 가속기(Blackwell 등)를 선제 확보하여 클라우드 형태로 전용 임대하는 코어위브, 람다 등 AI 특화 GPUaaS 기업들의 시장 점유율이 급속히 확대되고 있음
- 빅테크 기업들은 범용 GPU 확보 경쟁과 더불어, 엔비디아 의존도 탈피 및 인프라 구축 비용 절감을 위해 자체 설계한 맞춤형 AI 반도체 중심의 전용 인프라 확충을 병행하며 공급망을 재편 중임

➔ 전력망 제약 심화에 따른 입지 분산 및 신흥 거점 부상¹⁴⁾

- 북미 노던 버지니아(공실률 0.3%), 유럽 프랑크푸르트 등 기존 글로벌 Tier-1 데이터센터 허브의 극심한 송전망 포화로 인해, 신규 입지 선정 시 ‘네트워크 지연’보다 ‘전력 확보’가 최우선 기준으로 격상됨
- 부지 확보 후 전력을 끌어오던 방식에서 탈피하여, 풍부한 전력(원전, 재생에너지 등)과 확장 가능한 부지를 갖춘 중남미(케레타로 등), 아태지역(조호르, 바탐) 및 북미 외곽 지역(텍사스 서부 등)으로 대규모 자본이 선제 이동하며 신흥 허브를 형성 중임

➔ 인프라 후방 산업(전력 및 냉각 설비)의 메가 사이클 도래

- 랙당 전력 밀도의 급증(고집적화)으로 인해 글로벌 데이터센터 물리적 인프라(DCPI) 시장은 2030년까지 1,200억 달러 규모로 성장 전망(연평균 22% 성장)¹⁵⁾
- 단일 서버 단위를 넘어선 기반 인프라 수요 폭증으로 열 관리 부문이 가장 빠르게 성장 중이며, 액체 냉각 기술이 선택적 옵션을 넘어 신규 AI 인프라 구축의 필수 전제조건으로 자리매김

14) CBRE Research, Global Data Center Trends 2026('26.6)

15) Dell'Oro Group, Data Center Physical Infrastructure Forecast('26.8)

➔ 사모펀드(PE) 및 대체 자본의 AI 데이터센터 메가 펀드 투자 집중¹⁶⁾

- AI 인프라가 국가 필수 인프라이자 장기 수익이 보장되는 최우량 상업용 부동산 자산으로 부상하면서, 글로벌 사모펀드(Blackstone, KKR 등) 및 국부펀드 주도의 메가 펀드 조성을 통한 설비 직접 투자가 가속화되고 있음
- 전통적 부동산 자본의 AI 전력/공간 인프라 포트폴리오 재배치가 본격화 중임
※ 세계 최대 독립 데이터센터 운영사 QTS를 100억 달러에 인수한 Blackstone 社가 펜실베이니아 지역 데이터센터 및 전력망에 250억 달러를 투입

나. 주요국 정책 동향

➔ 미국

- (전력망 접속 규제 대폭 철폐) 미 연방에너지규제위원회(FERC)는 수년씩 걸리던 심사 절차를 최대 60~90일 이내로 단축하는 강력한 행정 조치를 단행¹⁷⁾
※ AI 팩토리의 전력망 연계 지연을 해소하기 위해 6대 권역 전력망 운영기관에 연방전력법 206조 기반 ‘소명 명령(Show Cause Order)’을 발동
- (세계 최대 원전 확장과 빅테크 동맹) 도널드 트럼프 정부의 원전 활성화 정책 (행정명령)과 맞물려, 빅테크와 원전 산업 간 대규모 자본 제휴가 본격화¹⁸⁾
 - 마이크로소프트는 스리마일 아일랜드 원전 1호기 재가동에 16억 달러를 투자해 20년 전력구매계약(PPA)을 체결함
 - 아마존(AWS)은 도미니언 에너지 인근 및 X-에너지 SMR에 5억 달러 이상을 투입하는 등 데이터센터의 민간 자체 전력 조달 생태계가 정립됨

➔ EU¹⁹⁾

- (AI Continent Plan 가동) 미국·중국 주도의 AI 패권에 대응하기 위해 2,000억 유로 규모의 단일 시장 최대 기술 프로젝트를 추진함
- (AI 팩토리 구축) 슈퍼컴퓨팅(EuroHPC) 인프라에 2만 개 이상의 GPU를 결합한 ‘AI 팩토리’를 역내 13개 지역에 거점화함
 - 이탈리아의 ‘IT4LIA’ 등이 대표적이며, 역내 스타트업과 연구기관에 컴퓨팅 자원과 데이터를 개방하여 유럽만의 독자적인 ‘소버린 AI’ 기반을 육성함

16) Blackstone Press Release, Digital and Energy Infrastructure Investment(‘25.7) / PitchBook, Private Equity Infrastructure Investment Report(‘26)

17) 임팩트온(<http://www.impacton.net/news/>), 美 FERC, AI 데이터센터 전력망 대란에 정책 전환(‘26.6)

18) 한국데이터경제신문(<https://www.dataeconomy.co.kr/news/>), AI 데이터센터 환경 오염 논란(‘26.5)

19) AI-BOOST meets AI factories(<https://aiboost-project.eu/>), The engine behind Europe’s AI-Continent strategy(‘25.8)

→ 영국²⁰⁾

- **(국가 중요 인프라(CNI) 격상)** 2024년부터 데이터센터를 전력·수도 등과 동일한 CNI로 지정하여 사이버 공격, 테러, 기후 재난으로부터 국가 차원의 보호망을 가동함
- **(AI Growth Zones 지정)** 데이터센터 신규 유치를 위해 간결한 인허가 패스트 트랙과 친환경 전력 우선 접속 혜택을 제공하는 ‘AI 성장 구역’을 신설하고, 국가 AI 연구 자원(AIRR) 용량을 ’30년까지 20배 이상 대폭 확장함

→ 중국²¹⁾

- **(동수서산(东数西算) 대규모 국책 사업)** 베이징, 상하이 등 동부 지역의 폭증하는 데이터를 에너지 리소스가 풍부하고 서늘한 서부 지역(구이저우, 닝샤, 간쑤, 네이멍구 등)으로 전송해 처리하는 인프라 분산 프로젝트를 가동함
 - 10개의 국가급 빅데이터 클러스터를 구축하여 전력 수급 불균형과 지연 문제를 동시 타개함
- **(자립형 칩 생태계와 컴퓨팅 바우처)** 미국의 첨단 칩 수출 통제에 맞서, 화웨이(Huawei), 캄브리콘 등 자국산 AI 반도체의 공공 조달을 의무화함
 - 또한 ‘컴퓨팅 바우처(算力券)’ 제도를 통해 기업의 자국 인프라 수요를 촉발시키며 폐쇄형 자체 생태계 고도화에 총력을 기울임

→ 일본²²⁾

- **(지방 분산 및 복원력 강화)** 도쿄·오사카 등 대도시에 집중된 인프라를 지진 등 자연재해 리스크로부터 분산시키기 위해 홋카이도 및 규슈 지역 거점화를 국가 차원에서 지원함
- **(소프트뱅크의 초거대 프로젝트)** 소프트뱅크는 최대 300억 엔의 정부 보조금을 바탕으로 홋카이도 도마코마이시에 내년 50MW를 시작으로 향후 최대 1GW까지 확장되는 일본 최대 규모의 자체 LLM 구동용 AI 데이터센터를 착공함

→ 싱가포르·중동 등 신흥 AI 데이터센터 허브

- 싱가포르는 협소한 국토와 전력난에도 불구하고, 2024년 예산에서 10억 달러(\$\$) 이상을 AI 컴퓨팅 지원에 투입²³⁾
 - 안정적인 규제 환경과 금융 허브 지위를 활용해 AWS(90억 달러), 구글(50억 달러) 등 빅테크로부터 수백억 달러 규모의 데이터센터 투자를 유치하며 아시아 최고 허브 지위를 공고히 함

20) 국회도서관 포털, Data centres : planning policy, sustainability, and resilience(’25.8)

21) 한국통신학회 추계종합학술발표회, 중국 AI 반도체 발전요인과 정책적 시사점(IITP 이현재, ’25.11)

22) 요코하마 한국기업인 연합회(<https://yokokoba.org/>), 소프트뱅크, 홋카이도에 일본 최대 데이터센터 착공(’25.8)

23) Introl Blog(<https://introl.com/>), 싱가포르의 270억 달러 AI 혁명이 2025년 동남아시아를 견인한다(’25.8)

- UAE·사우디아라비아 등 중동 국가들은 풍부한 오일머니와 저렴한 전력을 무기로 데이터센터 및 AI 칩 구매에 약 3,000억 달러 규모의 투자를 기획함²⁴⁾
- 그러나 최근 이스라엘-이란 분쟁 등 중동 지역의 지정학적 무력 충돌로 데이터센터가 피격(바레인, UAE 등) 대상이 되면서 인프라 안보 리스크가 부각됨

〈 주요국 정책의 공통점과 차이점 〉

국가/지역	핵심 정책 및 전략 방향	주요 차별점
미국	민간 주도 확장, 전력망 규제 타파, SMR-데이터센터 융합	인프라 규제를 연방 차원에서 제거하며 시장 논리와 에너지 자본(원전)을 결합
EU	AI Continent Plan, 소버린 AI, 공공 'AI 팩토리' 구축	대규모 공공 자금 투입을 통한 역내 스타트업 보호 및 빅테크 종속 탈피 (하향식+상향식)
중국	'동수서산', 자립형 AI 반도체 생태계 강제 육성	철저한 국가 통제 및 지역(동/서) 분산 정책, 미국 제재 우회를 위한 강제 국산화
일본	재난 복원력(Resilience) 중심의 핫카이드/규슈 분산	정부 보조금 지급을 통한 지정학적 리스크 회피 및 로컬 LLM 생태계 구축 집중

출처 : 저자 작성

5 국내 AI 데이터센터 산업 및 정책 동향

가. 국내 데이터센터 산업 현황과 AI 전환

➡ 국내 데이터센터 폭발적 성장 및 상업용 시장 확대

- 국내 데이터센터는 '00년 52개소에서 '25년 182개소로 약 3.5배 급증하였으며, 현재 신규 구축이 구체적으로 추진 중인 부지만 72곳에 달함²⁵⁾
- 국내 상업용 민간 데이터센터 매출은 '24년 기준 6조 2,200억 원 규모이며, 대규모 AI 데이터센터 클러스터 사업이 본격화됨에 따라 '29년까지 연평균 12.4%의 고성장이 전망됨

➡ 초고밀도 AI 데이터센터 패러다임 전환 및 IT 전력 부하량 급증

- 기존 일반 클라우드 중심에서 대규모 언어 모델(LLM) 학습 및 추론에 특화된 하이퍼스케일 AI 전용 데이터센터로 수요가 급속히 재편됨
- 전체 IT 부하 용량은 '30년 6,320MW 규모로 폭증할 것으로 예측되며, 랙당 전력 요구량 역시 기존 10kW 이하에서 40~100kW 이상으로 고집적화되는 추세임

24) 매일경제(<https://www.mk.co.kr/news/>), "AI 허브의 꿈 날아갔다"...400조 투입 계획, 전쟁 포화에 묻힌 중동(26.3)

25) 한국데이터센터연합회(KDCC), KOREA DATA CENTER MARKET 2026~2029(26.8)

➔ 민간 AI 데이터센터의 수도권 극단적 편중 및 전력 수급 한계

- 네트워크 지연시간 최소화, 고객사 및 IT 인력 수급의 용이성 등으로 인해 현재 민간 데이터센터의 74.8%가 수도권에 밀집되어 있음
- 한국전력 송전망 포화로 수도권 신규 전력 연계가 대거 보류되면서, 전력망 쪼टा운 위험 해소 및 지방 분산이 업계 최대 화두로 대두됨

➔ 국내 IT장비·SW·솔루션의 외산 대비 큰 기술 격차와 낮은 국산화율

- 데이터센터의 두뇌인 IT 코어 부문의 평균 기술 수준은 2.0~2.5점(5점 만점)에 불과하며, 독자 제조 메인보드가 개발되어도 코어 칩셋 등 라이선스를 외산에 대부분 의존하고 있음²⁶⁾
 - 랙을 제외한 서버, 스토리지, 네트워크 장비의 국산화율은 1%~11.1%, DCIM/EMS·클러스터 SW·솔루션 등은 10~25% 수준임

나. 국가 AI 컴퓨팅 인프라 확충 정책

➔ 'AI G3' 글로벌 중추국가 도약 비전

- 대통령 주재 '국가인공지능위원회'를 중심으로 대한민국 AI 혁신 청사진인 「국가 AI전략 정책방향」을 수립함

➔ 범국가적 4대 AI 플래그십 프로젝트의 일환으로 국가 AI 컴퓨팅 인프라 대폭 확충

- AI 경쟁력의 근간이 되는 대규모 국가 AI 컴퓨팅 센터 구축, 민관 합작 투자를 통한 고성능 컴퓨팅 자원 대형화 및 집적화 지원을 제시함
- '27년까지 민간 부문 AI 투자 대폭 확대 유도를 위한 정부의 투자 활성화 환경 조성 및 산업계 전반의 자발적 AI 기술 및 인프라 투자 촉진 방안을 제시함

다. 국가 AI 컴퓨팅센터 추진 현황

➔ 민관 합작 운영 법인 '코아크(KOAK)' 설립 후 해남 솔라시도 인프라 착공²⁷⁾

- 공공 마중물과 민간 자본을 결합해 과기정통부, 국민성장펀드, 삼성SDS 컨소시엄 출자로 특수목적법인 출범
- 총사업비 약 2조 5천억 원이 투입되며, '28년까지 엔비디아 GPU 등 15,000장 규모를 집적해 1엑사플롭스(1EF) 이상의 연산 능력 확보를 목표로 함

26) 정보통신기획평가원(IITP), 국내 데이터센터 장비·SW 기술 경쟁력 분석('26.7)

27) 대한민국 정책브리핑, '국가 AI 컴퓨팅센터' 첫 삽...1만 5000장 규모·2028년 완공('26.08)

➔ 국산 반도체 생태계 조성

- 초기에는 첨단 글로벌 GPU 중심으로 서비스하되, 국산 AI 반도체의 개발·검증 과정을 거쳐 '30년까지 국산 칩 비중을 50% 수준으로 확대하여 자생적인 AI 반도체 생태계를 조성하고자 함

라. 대한민국 대도약, 메가프로젝트 'AI 데이터센터'

➔ 1,500조 원 규모의 초대형 메가프로젝트 및 AI 데이터센터 집중 육성²⁸⁾

- '26년 6월 정부는 '반도체, 퍼지컬 AI, AI 데이터센터'를 3대 축으로 하는 국가 전략 산업화를 위한 초대형 메가프로젝트를 발표
- 10년간 민간 투자 및 정부 인프라(전력·용수·인허가) 지원을 결합하여, AI 데이터센터 분야에만 550조 원의 대규모 자본을 집중할 계획임
 - SK, GS, 네이버 등 민간 3개 기업이 1단계 8.4GW, 장기적으로 '35년까지 총 18.4GW의 전례 없는 초대형 AI 데이터센터 클러스터를 권역별로 구축함

➔ 국내 AI 데이터센터 인프라 자체를 수출로 산업화

- 국내 기업의 고효율 냉각 기술, 전력 관리 솔루션, 국산 NPU 등 IT 장비 솔루션까지 모두 패키지화하여 글로벌 시장에 'AI 솔루션 및 데이터센터' 자체를 수출 동력으로 육성하는 청사진을 제시함

마. 국내 기업의 AI 데이터센터 투자 동향

➔ 통신사(SKT, KT, LGU+ 등)와 IT 플랫폼(네이버, 카카오 등) 국내 기업들은 GPUaaS (서비스형 GPU) 확보 및 맞춤형 AI 모델 훈련을 위한 수천억 규모의 선제 투자를 집행

- (네이버) 글로벌 대체 투자사 Brookfield 등과 손잡고 약 10조 원 이상의 초대형 글로벌 AI 데이터센터 프로젝트 추진을 구상함²⁹⁾
- (SK에코플랜트 및 SKT) 엔비디아 등과 협력하여 전력 소모가 극심한 AI 워크로드를 감당할 대규모 AI 데이터센터 확장 및 울산 AI 데이터센터 프로젝트 등 인프라 고도화 추진³⁰⁾
- (삼성SDS) 엔트로픽과 기업용 AI 사업 협력을 확대하며, 국가 AI 컴퓨팅센터 구축 및 운영 역량 내재화를 추진 중

28) 삼일PwC, Q&A로 보는 '3대 메가프로젝트': AI, 기술을 넘어 국가성장 전략으로('26.7)

29) 연합뉴스, AI 데이터센터 판 키운다...네이버 14조·SKT 2GW 추진('26.7)

30) SK ecoplant Newsroom, AI 데이터센터 특별법 통과... AI 인프라 경쟁 본격화('26.5)

바. 지역별 AI 데이터센터 구축 및 입지 변화

→ 수도권 전력 포화에 따른 탈수도권 가속화

- 수도권의 송전망 부족으로 인해 신규 AI 데이터센터의 핵심 입지 기준이 '통신 지연시간'에서 '전력 수급 및 냉각 용수 확보'로 변모

→ 에너지 거점 중심의 신흥 AI 데이터센터 허브로 비수도권 지역의 부상

- 풍부한 재생에너지와 대규모 부지를 갖춘 서남권, 수열 에너지 활용 가능 및 기저 발전원 여유가 있는 강원권 등 에너지 접근성이 높은 비수도권으로 입지가 재편되고 있음

사. 전력·통신·에너지 정책과의 연계

→ 「인공지능 데이터센터 산업 진흥에 관한 특별법」 제정(2026.5)

- '27년 1분기 시행을 앞둔 AI 데이터센터 특별법은 AI 인프라를 국가 전략 자산으로 공식 인정하고 전력·입지 규제를 대폭 완화하는 내용을 포괄하고 있음

→ 비수도권 전력계통 영향평가 면제 특례

- 기존 「분산에너지 활성화 특별법」에 따라 대규모 전기사용시설(10MW 이상)에 의무화되어 1년 이상 지연을 초래하던 '전력계통 영향평가'를 비수도권 AI 데이터센터 신축·확장 시 전면 면제하여 지방 분산을 촉진함

→ 인허가 타임아웃제 및 일괄 처리 도입

- 과기정통부 장관으로 인허가 창구를 일원화하고, 관계기관이 일정 기간(40~150일) 내에 거부 통지하지 않으면 자동으로 인허가가 완료된 것으로 간주하는 '타임아웃제'를 신설함

→ 재생에너지 직접 전력구매계약(PPA) 특례 및 특구 지정

- 막대한 전력을 24시간 소모하는 AI 데이터센터의 특성을 고려하여, 한국전력을 거치지 않고 재생에너지로 생산한 전기를 사업자가 직접 공급받을 수 있는 PPA 제도 활성화의 법적 근거를 마련함
- 국가인공지능전략위원회의 심의를 거쳐 비수도권에 'AI 데이터센터 특구'를 지정하고, 기반시설(전력·용수·신재생설비)을 국가가 우선 지원할 수 있도록 제도화함

6 AI 데이터센터 확산을 둘러싼 주요 쟁점

가. 전력 확보와 전력망 병목(Grid Bottleneck)

⇒ 대규모 AI 데이터센터는 수십~수백 MW 이상의 상시 기저 전력을 요구하나, 지역 이기주의(NIMBY), 지중화 요구 등으로 인해 초고압 송배전망 건설의 극심한 지연이 계속될 수 있음³¹⁾

- 전기를 생산해도 송전 제약으로 전송하지 못하는 병목 현상은 AI 인프라 확충의 가장 치명적인 약점으로 작용함

나. 데이터센터 수도권 집중과 지역 분산의 한계

⇒ 분산에너지법 시행 이후 수도권 전력 계통 허가율이 2.1%에 불과한 상황에서, 규제(10MW)를 회피하기 위해 9.8MW 이하로 쪼개기 수전을 신청하는 등 제도적 사각지대와 편법이 등장하며 정책 실효성을 저해할 수 있음

⇒ 분산을 유도하여 비수도권 지역에 입지하더라도 해당 지역의 송전망이 부실하여 생산된 전력을 정상적으로 공급받지 못하거나, 핵심 IT 인력의 지방 근무 기피 현상이 겹쳐 실수요자들의 불만이 고조될 수 있음

다. 전기요금 및 계통비용 분담

⇒ 특정 지역에 초대형 부하가 신설됨에 따라 발생하는 막대한 송전선로 및 변전소 신설 비용(계통 보강 비용)을 원인 제공자인 데이터센터 사업자가 부담할 것인지, 아니면 한국전력과 일반 소비자가 분담할 것인지에 대한 사회적 비용 분담 갈등이 발생할 수 있음³²⁾

라. 용수·냉각·환경 문제

⇒ AI 데이터센터가 냉각탑을 통해 직접 증발시키는 물보다 해당 데이터센터의 전력을 생산하기 위해 발전소에서 소모하는 '간접 물 사용량'의 문제가 부각되고 있음³³⁾

⇒ SMR 1기(680MW) 가동에 매일 5,680만 리터(약 2만 3천 명의 일일 생활용수)의 물이 소모되는 등 수자원 고갈 및 농업 용수에 대한 갈등이 전망됨¹⁸⁾

31) 한국경제(<https://www.hankyung.com/>), AI 인프라 데이터센터, 수도권에 짓기 더 어려워진다('26.6)

32) 산업통상자원부, 데이터센터 수도권 집중 완화 방안('23.9)

33) 파이낸셜뉴스(<https://www.fnnews.com/news/>), AI 데이터센터, 수자원 고갈시키나('26.7)

- 세계적 기후변화로 식수가 고갈되는 가운데 데이터센터의 기하급수적 용수 소비는 향후 지역 농업 및 생활용수 배분 갈등의 뇌관으로 작용할 전망이다

마. 탄소중립과 AI 컴퓨팅 확대의 딜레마

- ⇒ 국내 기업의 RE100 달성 및 글로벌 ESG 규제 준수 목표가 연중무휴 전력을 소비하는 AI 인프라 확산과 정면으로 충돌하고 있음
- 간헐적인 태양광/풍력만으로는 데이터센터 가동이 불가능하므로 원전 기반의 무탄소 에너지(CFE) 도입 비중을 둘러싼 사회적 합의가 지연되고 있음

바. GPU 등 핵심 부품 공급망 병목 및 해외 기술·플랫폼 의존도

- ⇒ 엔비디아 GPU 및 고대역폭메모리(HBM) 등 필수 인프라 장비를 소수의 글로벌 독점 벤더에 의존하고 있어 비상 상황* 발생 시 국가 AI 연산망 전체가 마비될 수 있는 공급망 리스크를 안고 있음
- * 제품 조달 지연, 천문학적인 구축 단가 급등, 지정학적 무역 제재 발동 등
- ⇒ 하드웨어뿐 아니라 CUDA 플랫폼, 인피니밴드 네트워크 등 소프트웨어와 통신 프로토콜 전반이 해외 생태계에 종속되어 있어 국내 기업이 글로벌 경쟁에 주도권을 갖기 힘든 진입 장벽으로 작용하고 있음
- ⇒ 군사, 금융, 공공 데이터 등 민감 정보를 다루는 파운데이션 모델을 훈련할 때 외산 인프라에 의존할 경우 데이터 유출 및 주권 침해 소지가 존재함
- 폐쇄된 국경 내에서 구동되는 완벽한 ‘소버린 AI 인프라’ 확보 역량이 안보와 직결되어 있음

아. 투자 과열 및 사업성 문제

- ⇒ 천문학적 설비투자(CAPEX) 대비 수익화 지연 및 수요 격차
- 글로벌 빅테크의 AI 인프라 투자 규모가 수천억 달러로 폭증했으나, 실제 AI 서비스로 회수되는 누적 매출은 투자액의 일부에 불과하여 ‘수요-수익 격차’가 심화될 수 있음
- 차세대 AI 주도권 경쟁에서 도태될 것을 우려한 기업 간 ‘죄수의 딜레마’ 식 방어적 과잉 투자가 지속됨에 따라, 단기 투자 대비 수익률(ROI)에 대한 회의론이 확산되고 있음³⁴⁾

34) International Banker, AI's Capital-Spending Boom: Sustainable Investment or Emerging Bubble?(26.8)

➔ 초고가 GPU의 짧은 수명주기 및 감가상각에 따른 재무 리스크

- 데이터센터는 통상 7~15년의 장기 투자 회수 기간을 전제로 대규모 부채 (Private Credit 등)를 조달하나, 핵심 자산인 AI 가속기(GPU/NPU)의 유효 경제 수명은 1~2년에 불과함
- 차세대 칩셋의 빠른 출시로 기존 하드웨어의 자산 가치가 급격히 하락(기술적 진부화)함에 따라, 담보 가치 하락 및 금융 연계 리스크로 전이될 우려가 있음³⁵⁾

➔ 전력·냉각 중심의 막대한 운영비(OPEX) 부담 및 좌초 자산(Stranded Asset) 위험

- 초고밀도 전력 소모와 특수 냉각(액체/액침 냉각) 설비 도입으로 인해 초기 건축비뿐만 아니라 지속적인 유지·운영비가 기하급수적으로 증가함
- 장기 입주사나 전력구매계약(PPA)을 사전에 확정 짓지 않은 ‘투기적 데이터 센터’의 경우, 향후 투자 열기 둔화 시 전력망만 잠식하는 적자 자산으로 전락할 위험이 있음³⁶⁾

7 결론 및 정책적 시사점

가. AI 데이터센터, AI 인프라 생태계 경쟁으로의 전환

➔ 컴퓨팅 주권 및 ‘클러스터(집적) 효과’의 핵심 기반 시설

- AI 데이터센터는 우수 인재와 자본, 혁신 기업을 집중시키는 구심점 역할을 하는 핵심 안보 자산이자, 국가의 디지털 영토를 결정짓는 중추 시설임

➔ 범정부 차원의 거시적 접근 필요

- 주요국의 대규모 AI 인프라 투자에 대응하여, 단순 물리적 설비 확충을 넘어 전력·통신·데이터 안보를 포괄하는 인프라 생태계 전반의 경쟁력 확보가 필수적임

나. 반도체-IT장비-전력/냉각-AI를 망라한 통합 국가 전략 추진

➔ 첨단 기술 융합의 집결지

- 국산 NPU, 고대역폭메모리(HBM), DPU, 초저지연 네트워크 인프라, 고효율 전력·액체 냉각 설비 등 가치사슬 전반의 융합 경쟁력을 선점해야 함

35) Man Group Insights, The AI Bubble: Hidden Risks and Opportunities('26.2)

36) KKR Global Macro & Asset Allocation, Beyond the Bubble: Why AI Infrastructure Will Compound('25.11)/Digital Infrastructure, Data centre infrastructure investment - AI bubble, boom or bust?('26)

➔ 통합 컨트롤타워 및 기술 실증 지원

- 현재 외산 의존도가 높은 국내 IT 장비 및 운영 소프트웨어(DCIM 등)의 자립도를 높이기 위해, 부처 간 협력 체계 구축이 필요
- 메가프로젝트의 대한민국 AI 데이터센터가 인프라 기술의 성능 검증과 레퍼런스 구축 사이트로 적극 활용되어야 함

다. 국내 AI 데이터센터 산업 생태계 육성 및 수출 산업화

➔ 국내 기술력 기반의 칩·장비 생태계 조성

- AI 데이터센터 필수 인프라 장비의 소수 글로벌 벤더 독점에 따른 공급망 리스크(조달 지연, 단가 급등 등) 완화를 위해, 국가 컴퓨팅 시설 내 국산 AI 칩 및 솔루션 도입 비중을 점진적으로 확대해야 함

➔ “AI 데이터센터 코리아 패키지” 수출 동력화

- 국내 우위 기술인 첨단 메모리, 고효율 냉각 기술, 전력 관리 솔루션에 NPU, DPU, CXL 등의 국산 기술력을 접목한 IT 장비·SW 솔루션까지 풀 패키지화 하여 신흥 글로벌 거점을 겨냥한 신규 수출 모델로 육성할 필요가 있음

라. 전력망 제약 극복을 위한 지역 분산형 인프라 구축

➔ 수도권 전력 포화 대응 및 입지 분산

- 민간 데이터센터의 70% 이상이 수도권에 편중된 바, 전력계통 영향평가 면제 등 특별법의 제도적 특례를 활용해 재생에너지 및 기저 발전 여유가 있는 비수도권으로 입지를 신속히 재편해야 함

➔ 지역 특화 AI 클러스터 연계 및 인센티브 강화

- 단순한 시설 이전을 넘어 지역 핵심 산업과 연계된 특화 거점을 조성하고, 입주 기업 및 필수 IT 인재의 자발적 이전을 유도할 수 있는 실효적 혜택(세제, 정주 여건 등)을 병행해야 함

마. 지속가능한 AI 데이터센터 정책 이행 방안 마련

➔ 친환경·고효율 에너지 체계 확립

- 막대한 수자원 고갈 우려를 불식시키기 위해 폐쇄형 무수 냉각 시스템 전환을 독려하고, 무탄소 에너지(CFE) 중심의 전력 공급 체계 연계를 강화해야 함

➔ 지역 상생 모델 및 인프라 비용 분담

- 데이터센터 발생 폐열의 지역사회(스마트팜, 난방 등) 재활용을 장려하여 주민 수용성을 높이고, 대규모 송배전망 신설에 따른 계통 보강 비용의 합리적 분담 기준을 확립하여 사회적 갈등을 최소화해야 함

마. 데이터 주권 보호 및 과잉 투자 리스크 선제 관리

➔ 소버린 AI(Sovereign AI) 인프라 확보

- 공공·민감 데이터의 해외 플랫폼 종속 및 외부 유출 리스크를 차단하기 위해, 독자적인 데이터 주권을 보장하는 폐쇄형 인프라 환경 조성이 선결되어야 함

➔ 재무적 지속가능성 검증 (좌초 자산 방지)

- 초고가 GPU의 짧은 교체 주기(1~2년)와 인프라 장기 회수 기간(7~15년) 간의 괴리로 인한 재무 리스크를 방지하고자, 철저한 수요 예측과 유연한 인프라 확장 모델을 채택하여 투기성 적자 자산으로의 전락을 방지해야 함