


경제관계장관회의 겸
산업경쟁력강화 관계장관회의
25-16-7
(서면, 공개)

 역동경제로
서민·중산층 시대 구현!

첨단 GPU 확보 추진방안

2025. 5. 14.

관 계 부 처 합 동

첨단 GPU 확보 추진방안 (요약)

1 추진 배경

- AI 컴퓨팅 인프라가 AI 생태계 혁신성장의 관건으로 부상하며, 전 세계가 AI 경제안보 관점에서 AI컴퓨팅 인프라 확충에 **사활**

※ (美) 트럼프 정부, 5천억 달러 규모의 민관 협력 '스타게이트 프로젝트' 발표('25.1)

- 우리는 절대적 AI컴퓨팅 인프라 부족으로 글로벌 AI 경쟁에 실질적 한계 직면 → 지속적 경각심과 **기민한 정책 대응 시급**

- 이에 정부는 민관 합작 **국가 AI컴퓨팅 센터** 구축(GPU 1.5만장급)과 **슈퍼컴 6호기** 구축(GPU 0.85만장급) 등을 **가속화**하는 한편,

- 보다 과감한 대응 위해 GPU 연내 1만장 확보를 공식화하고*, **1.46조원** 규모의 **첨단 GPU 확보 추경 편성**('25. 5월)

* AI컴퓨팅 인프라 확충을 통한 국가 AI역량 강화방안('25. 2월), 후속조치(4월)

⇒ 조속한 첨단 GPU 확보 → 신속한 국내 산학연 등 지원을 위해, **「첨단 GPU 확보 추진방안」**을 마련하고 **속도감 있게 추진**

2 추진 방향

추진 방향

연내 GPU 지원 착수를 목표로, 민·관 협력 등을 통해 GPU ①구매·②구축·③사용 절차에 거쳐 **속도감 있게 추진**

구매

클라우드 기업^{CSP}을 공모·선정하여 GPU 구매 등 대행

구축

CSP의 기존 데이터센터에, 확보된 GPU 순차 구축

사용

국가 AI컴퓨팅 센터 중심의 GPU 사용, 전략적 배분

3 추진 방안

1 구매

① (공모·선정) 첨단·대규모 GPU(H200·B200)를 신속히 확보·구축·운용할 수 있는 클라우드 기업(CSP) 공모·선정 추진

※ 공모에 참여할 클라우드 기업(CSP)이 가용예산 內 GPU 확보·구축 계획 제시

- 국가 AI컴퓨팅 센터^{SPC} 참여한 CSP 중심으로 우선 선정 추진
- 별도 전문가 위원회를 통한 평가 진행 거쳐 현장실사 병행

※ (평가) △첨단·대규모 GPU 확보·구축 역량(재정투입 대비 컴퓨팅성능), △서비스 실행력, △CSP의 GPU 자원 자체활용 비중과 계획 등을 중점적으로 평가

- CSP 우선순위 선정 이후, 예산, 선호 GPU 기종 등을 고려, CSP들의 GPU 구매 기종·수량 직권조정 추진(공모서 內 명기)

구분	주요원칙
CSP 선정대상	<p>▶ <u>국가 AI컴퓨팅 센터(SPC, 5월말까지 사업참여계획서 접수예정)에 참여한 클라우드 기업(CSP)을 우선적으로 선정</u> 원칙</p> <p>※ 필요시 그 외 클라우드 기업(CSP) 선정 검토</p>
GPU 활용방향	<p>▶ <u>확보한 첨단 GPU 모두, 국내 산학연 대상 지원</u> 추진 원칙</p> <p>※ 단, 선정된 클라우드 기업(CSP)의 데이터센터 운영비 성격도 고려하여 CSP의 GPU 자원 자체활용도 일부 허용</p>

② (협약) 선정된 CSP와 추후 권리 등에 관한 협약(5년 검토) 체결

- ¹⁾NIPA의 GPU 소유권 확보, ²⁾국가 AI컴퓨팅 센터^{SPC}의 GPU 사용, ³⁾선정된 CSP의 일부 GPU 자체활용 허용* 등 중점 협약

* CSP의 데이터센터 운영비 등 총당 차원(구체적 비중은 공모·선정과정에서 결정)

③ (구매) 협약 후, 선정된 CSP로 하여금 GPU 구매발주 신속 추진

2 구축

- ① **(준비)** 선정된 CSP의 GPU 구매·확보 과정에서 CSP가 보유한 기존 데이터센터 인프라(전력·냉각·네트워크 등) 사전정비
- ② **(구축·안정화)** GPU 순차적 확보와 맞물려, 데이터센터 內 순차 구축 후 GPU 서비스(GPUaaS) 조속 안정화 병행

구분	목표일정
`25. 5월	▶ 첨단·대규모 GPU 구매·구축 대행 클라우드 기업(CSP) 공모
`25. 6월	▶ CSP 평가·심사(현장실사 등) 및 선정·협약(직권조정 등)
`25. 7월	<div>▶ GPU 구매 발주 (CSP, GPU 전량 확보(~10월))</div> <div>▶ 데이터센터 제반 인프라 (전력, 설계 등) 사전 준비(CSP)</div>
`25. 8월	
`25. 9월	
`25. 10월	▶ GPU 구축, 서비스 안정화 이후 GPU 서비스 순차 개시
`25. 11월	
`25. 12월	

※ GPU 기종에 따라 순차 확보될 경우 일부 변동 가능

3 사용

- ① **(지원)** 국가 AI컴퓨팅 센터(SPC) 중심으로 국내 산학연, 국가적 프로젝트 등에 전략적 GPU 배분 추진(상시 지원체계 마련)

※ GPU 구매·구축 과정 中, 국내 GPU 수요·용처에 근거한 맞춤형 GPU 지원 방안 수립
- ② **(조정)** 시장수요(총조사·분석), 정책 우선순위(WBL 총수요 파악 등), 실적·성과 등에 따라 GPU 지원 비중 탄력 조정
- ③ **(산학연 매칭)** 산학연 등의 GPU 사용 부담은 적정 수준 설정

※ 기관 규모, 他 GPU 지원사업 자부담 수준(학연 : 무상 & 중소기업 : 10% 미만) 고려
- 국가 AI컴퓨팅 센터^{SPC}에서의 GPU 사용 전까지 구체적 부담 수준 마련(~25.9월)

⇒ **민·관 협력** 등 통해 최대한 **GPU 연내 확보·서비스** 추진에 **총력**






순 서

I. 추진 배경	1
II. 추진 방향	4
III. 추진 방안	5
[붙임] 국가 시컴퓨팅 센터(SPC) 추진 현황 및 계획	12
IV. 추진 일정	14

I. 추진 배경

① 전세계가 사활, AI컴퓨팅 인프라는 AI 생태계 경쟁력의 관건

- 전세계가 AI 경제안보 관점에서 AI컴퓨팅 인프라 확충에 사활, 美 등 주요국은 국가적 전략 대응체계 가동에 총력

	▶ 트럼프 정부, 5천억 달러 규모의 민관 협력 '스타게이트 프로젝트' 발표('25.1)
	▶ 최대 20억 캐나다달러 규모의 '소버린 AI 컴퓨팅 전략' 발표('24.12)
	▶ 유럽 전역에 대규모 AI 인프라 'AI 팩토리' 설립, 15억 유로 투자 추진('24.12)
	▶ '30년까지 공공 AI 컴퓨팅 클러스터 AI 연구자원(AIRR) 20배 확대 추진('25.1)
	▶ AI 주도권 확보 차원, 민간 AI 슈퍼컴 개발 지원('24년까지 1조원 지원상회)

- AI컴퓨팅 인프라는 단순한 HW를 넘어 AI 생태계 혁신성장의 필수 기반이자 국가 AI 경쟁력의 핵심요소로 부상

② 글로벌 AI 전쟁 속 우리의 한계

- 우리는 절대적 AI컴퓨팅 인프라 부족으로, '더 작은' AI 개발함에도 '더 많은' 시간 소요 → 당장의 실질적 경쟁이 곤란한 상황

※ AI 인프라가 충분할 경우, 여러개로 나눠 동시에 다양한 학습방법 시도 가능
↔ 반대로 부족할 경우, 1가지 학습방법에 집중해야 하고 실패시 기회비용 급증

구분	AI 모델(발표시기)	모델 크기	투입 H100	개발기간
LG AI연구원	엑사원 3.5 32B('24.12.)	32B	512장	4개월 ¹⁾
Meta	라마 3.1-405B('24.7.)	405B	16,384장	54일 ²⁾

※ 출처 : 1) LG AI연구원장 발언('25.2월), 2) AI타임스('24.7월) / 동일 비교에 한계

- 글로벌 AI 경쟁 양상이 단일 기업 규모를 넘어 생태계 전체를 아우르는 국가 총력전으로 전개되는 현 상황 속,

- 우리나라가 新 AI 경쟁 패러다임에서 주도적 역할을 수행하기 위해서는 지속적인 경각심과 기민한 정책 대응 시급

[참고] 주요국의 AI컴퓨팅 인프라 확충 · 발전 전략

- **(분석)** 전 세계는 중요성이 증대하는 AI컴퓨팅 인프라 확보 위해, 국가적 차원에서 AI컴퓨팅 인프라 지원 전략 전개에 사활

 미국	<ul style="list-style-type: none"> • 바이든 정부는 대규모 데이터센터 구축 등 국가 AI 인프라 확충을 위한 'AI 인프라 리더십 강화'를 위한 행정명령* 발표('25.1.14) * Executive Order on Advancing United States Leadership in Artificial Intelligence Infrastructure • 트럼프 2기 행정부는 민간과 협력해 미국 내 새로운 거대 AI 인프라를 구축하는 5천억 달러 규모의 '스타게이트 프로젝트' 계획 발표('25.1.21) ※ 정부·민간 협력 합작법인 설립, 1,000억\$ 시작으로 향후 4년간 최대 5,000억\$ 투자 계획
 캐나다	<ul style="list-style-type: none"> • 최대 20억 캐나다달러 규모 '소버린 AI 컴퓨팅 전략' 발표('24.12) • 대규모 공공 소버린 슈퍼컴퓨팅 구축에 10억 캐나다달러 + 기존 공공 컴퓨팅 인프라 확충 2억 캐나다 달러 투자 국가 차원 전개 • 정부가 7억 캐나다 달러를 지원하며 민간 AI 인프라 투자 유인 + 연구기관·기업 등의 자원 구매 지원 위한 3억 캐나다달러 기금 조성
 EU	<ul style="list-style-type: none"> • EU집행위는 'AI 혁신 패키지'의 일환으로 유럽 전역의 AI 컴퓨팅 인프라 구축을 위한 'AI 팩토리 이니셔티브' 추진('24.12) • EU와 AI 팩토리 선정국^(1차 7개국) 간 매칭방식으로 인프라 구축에 15억 유로 투자 계획 발표('24.12) + 활용기업 육성에 1~10억 유로 투자 계획 • AI 팩토리 구축 신속화(~'26), 가칭 AI 기가팩토리 구축법 등 예고(EC, '25.1)
 영국	<ul style="list-style-type: none"> • 英 과학혁신기술부(DSIT)는 대규모 AI 인프라 중심 클러스터 구축 등을 포함한 'AI 기회 액션 플랜' 마련('25.1.) • '30년까지 영국의 공공 AI 컴퓨팅 클러스터인 'AI 연구자원(AIRR)'을 최소 20배 확대 추진(인프라 장기계획 수립 및 10년간 투자지원) • AI 데이터센터 구축 가속화 위한 AI 성장구역 설정(전력 원할공급)
 일본	<ul style="list-style-type: none"> • AI 주도권 확보 차원에서, 민간기업의 AI 슈퍼컴퓨터 개발을 정부 보조금(예산) 형태로 지원('23.下~, '24년까지 총 1조원 상회) ※ Ensuring a stable supply of cloud programs • 기업별 AI 슈퍼컴퓨터 구축(GPU 조달 등) 소요 비용의 1/3 ~ 1/2 보조금 지원 + 지원받은 기업들의 스타트업 지원* 의무화 * 최소 3년 이상 시세보다 저렴한 가격으로 AI 인프라 제공 의무 부과

- **(시사점)** 글로벌 AI 패권전쟁 속, 민·관 원팀으로 국가적 AI컴퓨팅 인프라 확충을 전개하여, AI 생태계 혁신성장 기반 조성 절실

③ 그간 정부의 AI컴퓨팅 인프라 확충 정책 대응

- 정부는 AI컴퓨팅 인프라의 중요성 인식 下에, 민관 합작 '국가 AI 컴퓨팅 센터' 구축(GPU 1.5만장급) 정책 추진('25년~)

※ 국가 AI전략 정책방향('24.9.), 국가 AI컴퓨팅 센터 구축 실행계획('25.1.) 등으로 공식화

- 슈퍼컴 6호기 구축(GPU 0.85만장급)도 가속화('26.上 서비스)

- 그러나, 현재 글로벌 AI 경쟁 속도와 규모를 고려할 때, 보다 신속하고 과감한 대응이 시급한 시점

- 이에, ^{AI컴퓨팅 인프라 확충을 통한} 汎부처 '국가 AI역량 강화 방안'('25.2월), '후속조치'('25.4월) 등에서 국가전략자산 차원의 GPU 연내 1만장 확보 공식화

'25년 내 1만장 확보

국가 AI컴퓨팅 센터 중심으로, 민관 협력을 통해 확보

'26.上 0.8만장 확보

슈퍼컴(국가초고성능컴퓨터) 6호기 구축을 통해 확보

④ 추경을 통한 첨단 GPU 확보 추진

- 정부는 AI컴퓨팅 인프라 수요, 민간시장 여건 등을 종합 고려, 1.46조원 규모의 첨단 GPU 확보 추경 편성('25.5월)

※ 첨단 GPU 서버 확보 外 필수적으로 요구되는 대규모 GPU 클러스터링 포함

- 속도감 있는 집행으로, ①국내 AI컴퓨팅 인프라 즉각 확충, ②생태계 활력 제고, ③글로벌 AI 경쟁 기반 확대 추진

▶ (추경 규모) 총 1.46조원

▶ (구성) GPU 서버(H200 6,400장 + B200 3,600장) ①확충(0.86조원) + ②통합운영환경 구축(0.60조원, 대규모 클러스터링 HW·SW 등)

※ H200, B200 비중은 시장선호, 기술발전 속도 등을 고려하여 조정 가능

※ 효율적 예산활용 위해 ①(서버확충) + ②(통합운영환경구축) 내역 간 통합·증감 등 가능

※ CSP를 통해 첨단 GPU 1만장 + @도 확보 가능

⇒ 조속한 첨단 GPU 확보 → 신속한 국내 산학연 등 지원을 위해, '첨단 GPU 확보 추진방안'을 마련하고 속도감 있게 추진

Ⅱ. 추진 방향

기본계획

- ① **(확보)** 민·관 협력 下 첨단·대규모 GPU 확보·구축 → 서비스화
- ② **(소유)** 국가전략자산 차원 소유권 확보 → 정책적 활용 뒷받침
- ③ **(사용)** 국가 AI컴퓨팅 센터에서 GPU 사용 → 통합지원 시스템

구매·구축	소유	사용
클라우드 기업 (CSP)	정보통신산업진흥원 (NIPA)	국가 AI컴퓨팅 센터 (SPC)

추진방향

◇ **(방안)** 연내 GPU 지원 착수 목표로, 민·관 협력 등을 통해 GPU ①구매·②구축·③사용 절차에 거쳐 속도감 있게 추진

- ① **(구매)** 첨단·대규모 GPU를 신속 확보·구축·운용할 수 있는 클라우드 기업(CSP)을 공모·선정하여 GPU 구매 등을 대행

※ (CSP와의 협약) ¹NIPA가 GPU 소유권 확보 + ²국가 AI컴퓨팅 센터가 사용 + ³선정된 CSP의 운영비 등 고려, 일부 GPU 자원은 자체활용 허용

- ② **(구축)** 공모로 선정된 클라우드 기업(CSP)의 기존 데이터센터 상면 내에, 확보된 GPU를 순차적으로 구축(클러스터링 기반)

- ③ **(사용)** 국가 AI컴퓨팅 센터(SPC) 중심으로 GPU를 사용하며, 국내 산학연 + 국가적 프로젝트 지원 등에 전략적 GPU 배분

※ 국가 AI컴퓨팅 센터^{SPC}로 GPU 이관 추진(단, 비용 등 고려하여 시기·방안은 별도 검토)

◇ **(추진일정)** 클라우드 기업^{CSP} 공모·선정(5~6월) 및 협약(7월) → GPU 구매발주(7월) → GPU* 확보·구축 및 순차서비스(10월~)

* 비교적 최신 모델인 B200의 경우 일정 등 순연 가능

○ 대내외 여건에 따른 변동 가능성과 리스크 등 최소화 만전

Ⅲ. 추진 방안

1 구매

기본 방향

첨단·대규모 GPU(H200·B200)를 신속히 확보·구축·운용할 수 있는 클라우드 기업(CSP)을 공모·선정하여 GPU 구매 등을 대행

① (공모) 첨단·대규모 GPU 구매·구축 대행할 클라우드 기업 공모

※ 정부 차원의 조달 등 절차는 장기간 소요 → 민간 통한 GPU 구매로 기간 단축

① (대상) 국내 클라우드 기업(복수개 선정 가능)

② (평가항목) ¹⁾GPU^{첨단·대규모} 확보, ²⁾구축·가동, ³⁾연내서비스 등 역량

1) 첨단·대규모 GPU 확보 역량(재정투입 대비 컴퓨팅성능) 평가

2) GPU 구축에 적합한 데이터센터 상면 확보, 제반 사양, GPU 구축·가동, 클라우드 서비스(GPUaaS) 전문성·보안성 등 진단

3) 연내 GPU 서비스(GPUaaS) 실행력, CSP의 GPU 자원 자체활용 비중(운영비 등 고려), 방안·계획 등을 종합적으로 분석

③ (선정원칙) 국가 AI컴퓨팅 센터(SPC, 5월말까지 사업참여계획서 접수 예정)에 참여한 클라우드 기업(CSP)을 우선 선정 추진

GPU 확보 기종(^① H200 vs B200 / ^② HGX vs DGX 관련)

※ ^① B200이 H200 대비 △고가, △고성능이나, △기술적 리스크 존재

^② HGX : OEM社 제공하는 커스텀 필요 제품 ↔ DGX : NVIDIA 제공 완성형 기성품

① (H200 vs B200) GPU △수급전망, △비용효율성, △민간여력 등 고려하여, 국내업체 논의, CSP 공모 등을 거쳐 최적의 기종·조합으로 확보 추진

② (HGX vs DGX) 공모 과정에서 기업 대상 제시 요청(구축일정 등 감안)

⇒ CSP 우선순위^{1~3순위} 등 선정 이후, 예산, 선호 GPU 기종 등 고려, CSP들의 GPU 구매 기종·수량 직권조정 추진(공모서 內 명기)

[참고] NVIDIA H200 vs B200 성능 비교

구분	H200	B200
출시	▶ '24. 2분기	▶ '25. 1분기 ⇒ 보다 최신 모델에 해당
아키텍처	▶ Hopper 기반	▶ Blackwell 기반 ⇒ 최신 설계로 빠르고 효율적
성능	▶ FP8(학습) 기준 4PF(Tensor Core) - 8 GPU 서버 : 32PF(Tensor Core)	▶ FP8(학습) 기준 9PF(Tensor Core) - 8 GPU 서버 : 72PF(Tensor Core) ⇒ 연산성능 2배 이상 높음
메모리 용량	▶ 1,128GB GPU 메모리 ▶ 최대 2TB 시스템 메모리	▶ 1,440GB GPU 메모리 ▶ 최대 4TB 시스템 메모리 ⇒ 대용량 메모리 구성으로 대규모 AI 모델의 고속 처리, 병렬 연산 가능
메모리 대역폭	▶ HBM3e 메모리 대역폭 141GB/s ▶ 양방향 GPU-to-GPU 대역폭 7.2TB/s	▶ HBM3e 메모리 대역폭 64TB/s ▶ all-to-all 메모리 대역폭 14.4TB/s ⇒ GPU 간 통신 대역폭 증가로 데이터 전송 속도 및 병렬 처리 효율 향상
소비 전력	▶ 전체 시스템 최대 소비 전력 10.2kW ▶ GPU당 최대 TDP* 700W * Thermal Design Power	▶ 전체 시스템 최대 소비 전력 14.3kW ▶ GPU당 최대 TDP 1,200W ⇒ 전력소모는 증가했으나, 단위작업 처리 시간 단축으로 총 에너지 효율 개선
확장성	▶ 4세대 NVLink, NVSwitch 사용 ▶ 확장성 측면에서 B200에 비해 제한적	▶ 5세대 NVLink, NVSwitch 사용 ⇒ 다수 GPU-서버 간 유기적 연결 가능 (고확장성), 대규모 DC 구축에 적합
열관리 방식	▶ 작동온도 : 5~30도 ▶ 관리 : 공랭식 시스템 지원	▶ 작동온도 : 5~30도 ▶ 관리 : 공랭·수냉식 시스템 모두 지원 ⇒ 고발열·고성능 연산 환경을 고려한 서버 설계·냉각 기술 적용
사용 사례	▶ 생성형 AI 모델 학습, 자연어 처리, 딥러닝 등 연구소, 대학 등에서 활용	▶ 초대규모 AI 모델 학습, 실시간 추론, 고성능 컴퓨팅 등 대규모 DC에 활용 ⇒ 최신 모델에 해당하여 GPU 서비스 (GPUaaS) 개시 지연 등 불확실성 존재

② **(선정)** ① 평가(서류 + 발표) + ② 현장실사 병행 통해 CSP 선정

- ① **(평가)** 전문가 위원회 구성 통해 서류 + 발표 평가 진행하여, 제출 내용·계획의 구체성, 이행 가능성 등을 종합 평가

※ 산학연 전문가 7인 내외로 전문가 위원회 구성 (필요시 NIPA 참여)

- ② **(실사)** 첨단·대규모 GPU 구축할 데이터센터 현장실사, 인프라 시설 준비상황, 가동 가능성 등을 종합 점검하여 CSP 선정

③ **(협약)** 선정된 CSP와 추후 권리 등에 관한 협약(5년 검토) 체결

- ① **(소유)** 국가전략자산 확보·활용 차원에서 CSP가 구매한 GPU 소유권은 정보통신산업진흥원(NIPA)에 귀속

- ② **(사용)** GPU 사용(CSP는 GPU 서비스 제공) 권리는 국가 AI컴퓨팅 센터(SPC)로 귀속하여 AI컴퓨팅 인프라 통합지원체계 뒷받침

※ 과기정통부-NIPA 협약 내 '향후 과기정통부가 정하는 기관이 GPU 사용권을 가질 수 있음'을 명시 → SPC 출범 후 정책지정하여 NIPA-SPC 사용권 협약

- ③ **(지원)** 확보한 첨단 GPU 모두, 국내 산학연 대상 지원 추진을 원칙으로 하되, 선정된 CSP에게 일부 GPU 자체활용 허용

※ CSP의 데이터센터 상면비, 운영비 등 지원 차원

(구체적 비중은 공모 과정에서 기업 측이 제시 → 선정과정에서 결정)

구분	주요원칙
CSP 선정대상	<p>▶ 국가 AI컴퓨팅 센터(SPC, 5월말까지 사업참여계획서 접수예정)에 참여한 클라우드 기업(CSP)을 우선적으로 선정 원칙</p> <p>※ 필요시 그 외 클라우드 기업(CSP) 선정 검토</p>
GPU 활용방향	<p>▶ 확보한 첨단 GPU 모두, 국내 산학연 대상 지원 추진 원칙</p> <p>※ 클라우드 기업(CSP)의 데이터센터 운영비 성격도 고려하여 CSP의 GPU 자원 자체활용 지원도 일부 허용</p>

⇒ 협약 체결 후, 선정된 CSP로 하여금 **GPU 구매발주 신속 추진**

2 구축

기본 방향

공모로 선정된 클라우드 기업(CSP)의 기존 데이터센터 상면 내에 확보된 GPU(H200·B200)를 순차적으로 구축(대규모 클러스터링 기반)

- ① **(사전준비)** GPU 구매·확보 과정에서, GPU 구축할 데이터센터 內 전력·냉각·네트워크 등 제반 인프라 정비 등 사전 준비

※ 특히, 대규모 GPU 클러스터링에 필요한 인프라 준비에 주안

- ② **(구축)** GPU 순차적 확보와 맞물려 CSP에서 사전 정비한 기존 데이터센터 상면 內 GPU 순차 구축

※ GPU 기종에 따라 확보·구축 일정 등 일부 변동 가능성 존재

- ③ **(안정화)** GPU 구축 이후 GPU 서비스(GPUaaS) 안정화 등 필요 최소한의 절차 거쳐 신속한 GPU 운영 준비

◇ 민·관 협력 등을 통해 최대한 GPU 연내 확보·서비스 추진 계획

< 목표 일정 >

구분	주요내용
`25. 5월	▶ 첨단·대규모 GPU 구매·구축 대행 클라우드 기업(CSP) 공모
`25. 6월	▶ CSP 평가·심사(현장실사 등) 및 선정·협약(직권조정 등)
`25. 7월	
`25. 8월	▶ GPU 구매 발주 (CSP, GPU 전량 확보(~10월))
`25. 9월	▶ 데이터센터 제반 인프라 (전력, 설계 등) 사전 준비(CSP)
`25. 10월	
`25. 11월	▶ GPU 구축, 서비스 안정화 이후 GPU 서비스 순차 개시
`25. 12월	

※ GPU 기종에 따라 순차 확보될 경우 일부 변동 가능

3 사용

기본 방향

국가 AI컴퓨팅 센터(SPC) 중심으로 GPU(H200·B200)를 사용하며,
국내 산학연 + 국가적 프로젝트 지원 등에 전략적 GPU 배분

① **(주안)** 국내 산·학·연의 GPU 접근성을 높이면서도, 국가 전략적 필요에 따른 최적화된 GPU 지원체계 수립·시행

① 국내 산·학·연의 상향식 GPU 지원 요구에 기반한 GPU 지원
(상시 지원체계 마련 추진 + 도전적·혁신적 AI 연구·서비스 중점지원)

※ 대학·연구기관의 장기·지속 수요 + 산업계 단기 수요

② 국가적 프로젝트(^{가칭} WBL 프로젝트 등) 등 정책 우선순위에 따른 집중형 GPU 지원 병행

③ GPU 가동율 모니터링 등 통해 GPU 유휴자원 최소화 추진

② **(조정)** 시장 수요(총조사·분석), 정책 우선순위(WBL 총수요 파악 등), 실적·성과(GPU 이용에 따른 AI 연구·서비스 실적·성과) 등 종합 감안,

○ 국가 AI컴퓨팅 센터(SPC) 중심으로 GPU 지원 비중 탄력 조정

③ **(산학연 매칭)** 산학연 등의 GPU 사용 부담은 적정 수준 설정

※ 기관규모 + 기존 他사업(GPU 임차·지원사업 자부담 : 학·연 무상 & 중소기업 10% 미만) 사례 등 종합 고려하여 설정, 현금 자부담은 SPC 설립 취지 맞게 사용 추진

- 국가 AI컴퓨팅 센터^{SPC}에서의 GPU 사용 전까지 구체적 부담 수준 마련(~'25.9월)

⇒ GPU 구매·구축 과정 中, 국내 GPU 수요·용처('25.9월까지 종합)에 근거한 맞춤형 GPU 지원 방안 수립 및 이행(GPU 구축 후) 추진

⇒ 국가 AI컴퓨팅 센터^{SPC}로 GPU 이관 추진(비용* 등 고려, 시기·방안 별도 검토)

* 자본 이동에 따른 과세, 물리적 이동에 따른 소요비용 등을 폭넓게 감안

[참고] 국내 GPU 수요 기초조사(1차)

- ◇ 405개 산학연 대상 GPU 수요 조사시 1만장 수준 수요 파악
→ 점진적 조사 확대 통해 수요 구체화 및 지원 방안 마련

1] 조사개요

- (조사대상) GPU 수요가 있는 산·학·연(대기업 제외)
- (조사기간/방법) `25. 4. 28. ~ 5. 1. / 온라인 설문조사
- (응답기관) 총 405개 기관(산업계 115, 학계 273, 연구계 10, 기타 7)
※ 학계 응답이 다수이나, 이를 학계 수요 우위로 직결시키는 해석에는 유의 필요

2] 조사결과

- (시급성) GPU 수요가 **“매우 시급”(6개월 이내)**하다는 답변이 가장 많으며(45.9%), 5기관 중 4기관이 1년 이내 필요(79.2%)

구분	매우 시급 (6개월 이내 필요)	시급 (1년 이내 필요)	보통 (필요하나 유동적)	장기 고려 (미래 수요 존재)
응답 수 (비중)	186명 (45.9%)	135명 (33.3%)	75명 (18.5%)	9명 (2.2%)

- (수요목적) **생성형AI**(64.3%) 분야가 가장 높은 비중 차지 복수선택

구분	생성형 AI	컴퓨터 비전	자연어 처리	로봇/ 자율주행	기타	음성인식	추천 시스템
응답 수 (비중)	261명 (64.3%)	246명 (60.6%)	216명 (53.2%)	75명 (18.5%)	51명 (12.3%)	47명 (11.6%)	40명 (9.9%)

- (수요기간) **1년 7개월 ↑ GPU 장기 수요**가 과반 차지(55.8%)

구분	장기 (1년 7개월 이상)	중기 (7개월 ~ 1년 6개월)	단기 (6개월 이하)	기타 (추후확인 등)
응답 수 (비중)	226명 (55.8%)	145명 (35.8%)	22명 (5.4%)	12명 (3.0%)

- (수요규모) 소수 응답기관(405개)만 해도 **GPU 수요 1만장 상회**

※ 최신 기종인 블랙웰은 사용례가 없어 수요가 과소 집계되었을 가능성

기종 구분	산업계	학계	연구계	기타	계
암페어(A100)	180장	673장	-	-	853장
호퍼(H100-H200)	1,933장	4,673장	124장	163장	6,893장
블랙웰(B100-B200)	609장	2,228장	206장	-	3,043장
총계					10,789장

※ 수요가 범위(n~m장)로 응답된 경우, 실제 수요의 충분성 담보를 위해 최대치(m장)로 집계

[참고] 가칭 WBL 프로젝트 추진방향

※ 5월 중 기본계획 수립 예정

1 개요

- 실력있는 국내 AI 기업에 ① GPU + ② 데이터 + ③ 인재 집중 지원 → 글로벌 수준의 AI 파운데이션 모델 개발 지원

2 주요내용(안)

- (선발) 복수의 정예팀(권소시엄, 최대 5개팀) 공모·선발

< 정예팀 선발 방향(안) >

구분	파운데이션 모델 개발 방향·방식	예시
유형1	새로운 파운데이션 모델 개발	A社 A모델 이상의 신규 모델 개발
유형2	기존 파운데이션 모델 고도화·확장	B社 B모델 고도화 + 특화모델로 확장

- 기업규모, 오픈소스 수준 등에 따라 자부담 최소화

- (지원) ^{최대}3년 간 정예팀 희망 ①GPU, ②데이터, ③인재 지원, 연차평가로 지속 지원할 정예팀 압축(ex. '25. 5 → '26. 3 → '27. 2팀)

※ 정예팀이 ①·②·③ 중 필요사항, 규모 제시 → 적정규모 검토 후 지원(매년 검토)

구분	개요	방식
GPU	▶ 정예팀이 AI 파운데이션 모델 개발에 필요로 하는 GPU 지원	▶ '25년에는 민간 GPU 임차·지원, '26~'27년은 국가 AI컴퓨팅 센터의 GPU(정부구매분) 지원
데이터	▶ AI 모델 개발에 필요로 하는 데이터 구축·가공·활용 지원	▶ 저작물 데이터 공동활용 지원(공통수요) + 개별팀 데이터 구축·가공 지원(개별수요)
인재	▶ 해외 우수 AI 연구자 등 유치 관련 인건비·체제비 등 지원	▶ 정예팀이 해외 우수 연구자(재외한인 포함)를 주도적 유치할 경우, 매칭 지원

- (후속) 개발된 AI 모델 생태계 지원(정부사업 등 다각연계)

3 목표일정(안)

- 정예팀 공모 및 평가·심사 거쳐 정예팀 선정: ~7월
- AI 모델 ^{글로벌 수준} 개발 지원(GPU+데이터+인재) : 8월~
- 1차 연차평가(정예팀들과 협의 통해 평가체계 확정) : 12월

- ◇ **첨단 GPU 확보**와 병행하여 국가적 AI컴퓨팅 인프라 확충·지원 체계의 구심점 역할을 할 **국가 AI컴퓨팅 센터(SPC) 구축** 추진

1 국가 AI컴퓨팅 센터 개요

- (주요내용) 민·관 공동 출자*를 통해 SPC^{특수목적법인}를 설립하고, 정책금융 대출** 등을 활용하여 국가 AI컴퓨팅 센터 구축

* 정부·정책금융기관 2천억원 + 민간 2천억원 총 4천억원 출자 목표('25~'30)

** 총 2.5조원('25년 6,250억원) 규모의 AI컴퓨팅 인프라 지원상품 신설('25~'27)

- (지분구조) 공공(정부, 정책금융기관) 51%, 민간 49%
- (입지/규모) 비수도권('27년 개소) / 총 1EF 이상 규모(~'30년)

2 추진경과 및 현황

- 국가AI위원회 AI컴퓨팅 인프라 특별위원회 개최(1.21)
 - ※ 「국가 AI컴퓨팅 센터 구축(SPC 설립) 실행계획(안)」서면 심의
- 「국가 AI컴퓨팅 센터 구축(SPC 설립) 실행계획」 발표(1.22, 경장)
- 국가 AI컴퓨팅 센터 SPC 민간 참여사 공모(1.23~5.30)

[SPC 민간 참여사 공모 주요 추진현황]

- ◇ 제1차 사업설명회 개최(2.7)
 - ※ 기업·기관 등 460명 참석, 사업 개요 및 공모 절차·일정 등 안내
- ◇ 사업참여계획서 접수(~2.28) : 108개 기업·기관 등
- ◇ 제2차 사업설명회 개최(3.7, 비공개)
 - ※ 사업참여의향서 제출 기업 등 149명 참석, 공모지침서 주요 내용 등 안내

◇ 서면질의 답변(~5.16) : 총 208건 질의 중 208건 답변 완료(5.8 기준)

① (공공기관 지정) SPC의 공공지분이 51%인데 공공기관으로 지정되는 것인지?

- 공공기관 지정 시 경영 공시 및 실적 평가, 국회 예결산, 감사원 감사 등의 의무가 부과되어 민간 자율적 운영이 어려워질 우려

⇒ SPC는 공공지분이 51%로 공공기관운영법에 따른 공공기관 지정 가능 요건에 해당

- 다만, 지정 여부는 아직 결정된 바 없으며, 향후 공공기관운영위 심의(기재부 주관, '26.1월 예상)를 통해 결정될 사항
- 민간의 자율성을 최대한 보장하는 방향으로 관계부처 협의 추진

② (매수청구권*) 불가피한 상황 발생 시 민간참여자에게 공공지분 매수 의무를 부과하는 매수청구권으로 인해 민간의 사업 참여 부담 우려

* 공공참여자가 공공지분(보장수익률 포함)을 민간참여자에게 팔 수 있는 권리(뜻옴션)

⇒ 매수청구권은 공공 출자금 회수와 민간의 적극적인 참여를 위해 필요

- 다만, 불가피한 상황이 발생하지 않고 SPC가 안정적으로 운영·유지될 수 있도록 SPC에 대한 인센티브 등 종합적 지원정책* 병행 추진

* 정부 사업의 AI컴퓨팅 인프라 수요 연계, 정부구매 GPU 수익금 인센티브 등

③ (입지) 국가 AI컴퓨팅 센터는 입지를 비수도권으로 제한하고 있는데, 정부구매 GPU를 SPC로 이관할 경우 입지 요건은?

⇒ 국가 AI컴퓨팅 센터는 기존 계획대로 비수도권에 단일 센터(동일 부지 내 1개 또는 복수의 건축물)로 집적하여 구축 필요

- 다만, 정부구매 GPU 등 추가 구축 인프라는 타 부지에 구축·운영 가능

④ (사용권 협약) 정부구매 GPU를 SPC가 사용할 계획이나 SPC 공모(1.23) 시 관련 내용이 미반영되어 SPC와 협의 필요

⇒ SPC가 정부구매 GPU를 사용·관리할 계획임을 사전에 안내하고, 사업자 선정 과정(실시협약 등)에서 수익금 활용방안 협의 추진 예정

③ 향후 계획

○ 사업참여계획서 접수(5.19~5.30)

○ 1단계 기술·정책평가(6월) 및 2단계 금융심사(~8월)

○ 민간 참여사 선정(실시협약 체결*) 및 SPC 설립 추진(~10월)

* SPC의 1EF 이상 GPU·NPU 구매 일정, 정부구매 GPU 사용·관리 방안 등 협의

○ 사업 착수 및 국가 AI컴퓨팅 센터 구축 추진('27년 개소 목표)

IV. 추진 일정

주요내용	목표일정
<p>▶ CSP 공모</p> <p>※ 첨단·대규모 GPU 구매·구축 대행할 클라우드 기업 공모</p>	5~6월
<p>▶ CSP 등 대상 설명회</p> <p>※ 참석 : 과기정통부, NIPA, 국내 주요 클라우드 기업 등</p>	5월 中
<p>▶ 평가 및 심사</p> <p>※ 서류 + 대면 평가 및 현장 실사 병행</p>	6~7월
<p>▶ 선정 및 협약</p> <p>※ 선정된 CSP와 추후 권리 등에 관한 협약(5년 검토) 체결</p>	7월~
<p>▶ GPU 구매 발주</p> <p>※ 협약 체결 후, 선정된 CSP의 GPU 구매발주 신속 추진</p>	7월~
<p>▶ GPU 확보·구축 및 순차 서비스</p> <p>※ GPU 기종에 따라 확보·구축 일정 등 변동 가능성 상존</p>	10월~