
2025년 고성능컴퓨팅지원 사업

-NPU(국산 Si반도체 활용 사업)-

사용자 모집 안내서

2025. 2. 12.



과학기술정보통신부
Ministry of Science and ICT



정보통신산업진흥원
National IT Industry Promotion Agency

□ 사업목적

- AI 추론 서비스를 개발하려는 중소·벤처기업 등을 대상으로 국산 AI 반도체 기반의 개발용 서버 지원

□ 사업내용

◇ 본 사업을 통해서 **AI 학습은 불가**하며, 이미 AI 학습이 완료된 기업·기관에 한해서 신청할 수 있으며, AI 추론 서비스를 위한 NPU 개발 서버를 지원함

- 지원대상

구분	지원 대상	지원 상품	상품별 지원비율*	제외대상
민간부문	<ul style="list-style-type: none"> ■ (국내) 중소·벤처기업 ■ (국내) 창업기업 	60TOPS, 120TOPS,	50% 이상	■ 소속이 없는 개인, 대기업 제외
공공부문	<ul style="list-style-type: none"> ■ (국내) 대학교 ■ (국내) 대학병원 	240TOPS	50% 미만	■ 공공기관, 일반 협·단체 제외

※ 접수 현황 및 추진 상황에 따라 지원 비율은 변경될 수 있음

- (지원기간) 서버 할당 후 ~ 2025년 12월(또는 예산소진시 까지)
- (지원규모) 200개 내외 기업·기관 당 NPU 개발 서버 1식

NO	AI 가속화기 (INT8 기준)	CPU	메모리	스토리지	회선
1	60TOPS 이상	2.1GHz, 16Core 이상	128GB 이상	SSD 1TB 이상	1Gbps 이상
2	120TOPS 이상	2.1GHz, 16Core 이상	256GB 이상	SSD 1TB 이상	
3	240TOPS 이상	2.1GHz, 16Core 이상	256GB 이상	SSD 1TB 이상	

※ TOPS(Tera Operations Per Second) 초당 1조번의 정수 연산을 처리할 수 있는 성능

※ 상기 3종의 상품 중에서 1개 상품을 선택하여 신청하며, 상품에 따라 사용자 모집 예정 비율은 60TOPS(60%), 120TOPS(30%), 240TOPS(10%)임

* 모집 비율은 접수 현황 및 추진 상황에 따라 변동될 수 있음

※ 지원 대상별 NPU 서버 할당 정책

1. 민간부문 사용자 : 기업당 고성능 컴퓨팅 자원 1식 할당
2. 공공부문 사용자 : 교수당 고성능 컴퓨팅 자원 1식 할당

- (지원가능 AI 모델) 시각·언어 모델 중심으로 지원이 가능하며, 반드시 마지막 페이지 참조 “25년 국산 AI 반도체 활용 사업 지원 가능한 AI 모델” 확인한 후에 신청해야함

※ 지원 가능하지 않은 모델을 신청할 경우 선정 및 지원에 제한될 수 있음

- (기술지원) 중소·벤처기업 등이 원활한 AI 추론 서비스 제공할 수 있도록 커스터마이징 등 AI 반도체사의 직접 기술지원 예정
- (자부담금) 무상 지원
- (성과제출) 설문 및 성과조사 참여 미 준수 시 차년도 사업 참여 제한 및 기 사용한 자원의 사용비가 환수될 수 있음
 ※ 제출한 신청서 상의 사용계획에 대한 진행 경과를 월단위로 체크 할 예정

● 과제 진도 체크 계획

- 사용자 과제 진행 단계를 총 5단계로 구분하여 진도를 체크예정
 ※ (1단계) 모델선정 (2단계) 모델분석 (3단계) 모델변환(포팅), (4단계) 모델 최적화 (5단계) 추론서비스
- 진도체크 리스트 작성하여 단계별 진행현황을 운영기관에 보고하여야 하고, 2개월 이상 현단계 유지 시 유지사유서 추가로 제출(단, 4/5단계 제외)

※ 중간점검(8월) 및 결과점검(12월) 수행을 위한 자료를 제출해야 함

* 결과점검(신청서 상의 목표(성과지표)대비 실적점검) 결과 '미흡' 시 차년도 사업 참여제한의 조치를 받을 수 있음

2 신청방법

- (신청기간) 2025년 2월 12일(수) ~ 2025년 3월 17일(월) 15:00
- (신청방법) 첨부한 '25년 국산 AI 반도체 활용 사업 이용신청서 (붙임 사용계획서, 증빙서류 포함)'를 작성하여 이메일 주소(gpu@nipa.kr)로 제출

구분			작성양식	비고
1	공통	사용자 신청서	[붙임1-2], [붙임1-3]	민간/공공 상이
2		신청 현황표	[붙임2-1], [붙임2-2]	민간/공공 상이
3	민간부문	중소기업 확인서 또는 벤처기업 확인서 또는 창업기업 확인서(사업자등록증), 법인등기부등본(* 청년 기업일 경우)	발급처 양식에 따름	사용자 신청자격 확인이 가능한 자료 제출
4	공공부문	재학·재직 증명서	자체 양식	

◇ '25년 고성능 컴퓨팅 지원 사업의 학습용 GPU 컴퓨팅 자원을 신청 또는 선정된 기업·기관도 본 NPU(국산 AI반도체 활용) 사업에 신청 가능함

○ (신청자격) 중소·벤처 기업, 창업기업, 대학교(대학병원 포함)

- (중소·벤처) '중소기업 확인서' 또는 '벤처기업 확인서'를 발급받을 수 있는 법인 또는 개인사업자로서 중소기업·벤처기업 확인서 제출
- (창업기업) '창업기업 확인서'를 발급받을 수 있는 법인 또는 개인사업자로서 창업기업 확인서(또는 사업자등록증) 제출

* 중소기업을 창업하여 사업을 개시한 날부터 7년이 지나지 아니한 기업(법인과 개인사업자 포함)

※ (공통) 대표자(공동 대표인 경우 대표자 중 1인 이상)가 사용자 모집 공고일 기준 만 39세 이하 (1985.2.12. 이후 출생자)인 경우 **청년기업으로 인정하며, 별도 법인등기부등본을 제출** (법인이 아닌 개인 사업자의 경우 생년월일이 표기된 신분증 사본 제출)

- (대학교/대학병원) 대학교(원)에 재직자로서 소속을 증빙할 수 있는 재직증명서 제출
※ 신청자가 대학(원)생인 경우 학생 재학증명서와 (지도)교수 재직증명서를 함께 제출

1. 중소기업 (「중소기업기본법」 제2조 및 같은 법 시행령 제3조)

- 중소기업 확인서는 「중소기업기본법」 제2조에 따른 중소기업 여부를 확인하는 것으로서 「조세특례제한법」이나 「중소기업 판로지원법」 등 개별법에서 정한 중소기업 범위와 다름
- 중소기업 확인서의 유효기간은 기본적으로 직전 사업연도가 12개월 이상인 경우, 직전 사업연도 말일에서 3개월 경과한 날부터 1년간입니다. 다만, 직전(또는 당해) 사업연도에 창업, 합병, 분할한 기업 등은 유효기간이 달리 적용될 수 있음
- 중소기업현황정보시스템(sminfo.mss.go.kr)을 통해 중소기업 확인서 발급 가능

2. 벤처기업 (「벤처기업육성에 관한 특별법」 제2조)

- 다른 기업에 비해 기술성이나 성장성이 상대적으로 높아, 정부에서 지원할 필요가 있다고 인정하는 중소기업으로서 「벤처기업육성에 관한 특별법」 제2조의2제1항제2호에 따른 3가지 요건의 어느 하나에 해당하는 기업(① 벤처투자유형, ② 연구개발유형, ③ 혁신성장유형·예비벤처기업)을 말함
- 벤처기업임을 확인 받으려는 기업은 '벤처기업 확인 신청서'를 작성하여 벤처인(www.venturein.or.kr)을 통해 다음에 따른 확인기관의 장에게 벤처기업 확인을 신청해야 하며, 예비벤처기업으로 확인받으려는 경우에도 동일함

3. 창업기업 (「중소기업창업지원법」 제2조)

- "창업기업"이란 「중소기업창업지원법」 제2조에 따라 중소기업을 창업하여 사업을 개시한 날부터 7년이 지나지 아니한 기업(법인과 개인사업자를 포함한다)을 말한다. 이 경우 사업 개시에 관한 사항 등 창업기업의 범위에 관한 세부사항은 대통령령으로 정함
- 창업기업임을 확인 받으려는 기업은 '창업기업 확인 신청서'를 작성하여 창업기업확인시스템(https://cert.k-startup.go.kr)을 통해 다음에 따른 확인기관의 장에게 창업기업 확인을 신청해야함

4

선정 절차 및 기준

○ 추진 절차

선정 절차	주요 내용	일정
사용자 모집공고 신청서 접수	■ 사용자 모집 안내 공고	'25년 2월 12일 ~3월 13일
↓		
사업 설명회	<ul style="list-style-type: none"> • 2025년 고성능컴퓨팅 지원사업 설명회 - 일시 및 장소 : 2025.2.20.(목) 13:30 / 코엑스(아셈볼룸홀) - 참석대상 : 자원수요자, 공급기업, 유관기관 등 * 사전등록 불필요, 현장등록 가능, 주차지원 불가 * 사정에 의해 행사장 좌석 및 배포자료가 조기 소진 될 수 있음 	'25년 2월 20일(목)
↓		
서류 적격심사 / 사전 기술검토	<ul style="list-style-type: none"> ■ 지원 자격요건 및 증빙서류 등 검토 ※ 사용자의 AI 모델과 AI 반도체사 기술지원 가능 여부 사전 검토 	'25년 3월 중
↓		
선정평가	■ 신청서 서면 평가	'25년 3월 중
↓		
평가결과 통지	■ NIPA 홈페이지에 선정자 공지	'25년 3월 말
↓		
AI 반도체 및 클라우드사 배정	<ul style="list-style-type: none"> ■ AI 반도체사별로 지원 가능한 AI 모델 매칭 ■ 클라우드사 상품 배정 	'25년 4월 초
↓		
개발서버 할당·이용	■ 사용자별로 NPU 개발 서버 할당 및 이용	'25년 4월 ~ 12월

※ 세부 일정은 기관 사정에 따라 변경될 수 있음

- (선정방법) 제출한 이용신청서 기반으로 3종의 상품별로 서류심사를 통해 선정
- (선정기준) 지원 자격요건, AI 연구·개발대상 여부, 사용계획서(목적, 연구·개발분야, 사용내용, 필요성 등)를 중심으로 심사·평가

※ 공공 사용자(대학교, 대학병원)는 상품별 **50% 미만으로 선정 예정**

※ 청년기업의 경우 선정평가 시 **가점 1점을 부여** (반드시 증빙서류 제출)

※ 평가항목(신청서 상의 사용계획)

평가 부문	평가요소	배점
목적 (성과목표) (30)	<ul style="list-style-type: none"> 성과목표 설정, 컴퓨팅 자원 활용 목적의 적절성 과제수행을 위한 추진 내용의 구체성 및 현실성 	30
연구개발분야 (5)	<ul style="list-style-type: none"> 연구·개발 분야의 구체성 	5
사용내용 (50)	<ul style="list-style-type: none"> 컴퓨팅 자원 활용 계획의 구체성 및 적절성 컴퓨팅 자원 활용을 통한 목적(성과목표) 달성 내용 	50
필요성 (10)	<ul style="list-style-type: none"> 컴퓨팅 자원 필요 당위성 	10
기타 (5)	<ul style="list-style-type: none"> 컴퓨팅 자원 활용 일정 적절성 등 	5
합 계	-	100

○ (선정결과) '25년 3월이후 선정결과 통보 및 4월 자원할당* 예정

※ 사용자의 상품선택 의사를 반영하여 클라우드사와 AI 반도체사를 배정하지만, 클라우드사 및 AI 반도체사의 사정에 따라 타 클라우드사의 상품으로 변경될 수 있음

5 유의 사항

- 정부정책, 재정상황 등 환경 변화에 따라 사업 종료 시기가 변경될 수 있음
- AI 바우처 사업의 AI 반도체 트랙 등 유사 정부지원 사업에 중복 선정될 경우에는 선정대상에서 제외될 수 있음
- 선정 및 자원할당 이후 중도 포기한 사용자에게 대해서는 차년도 고성능 컴퓨팅 지원 사업에 참여를 제한할 수 있음
- 다음의 어느 하나에 해당하는 경우 지원 대상에서 즉시 제외될 수 있음
 - 접수마감일 기준 신청 기업 등이 과학기술정보통신부 등 정부지원사업 참여제한 등 제재를 받고 있거나 'AI 바우처' 사업 참여제한 대상기업
 - 접수마감일 기준 신청 기업 등이 ICT기금사업 참여에 제한을 받고 있는 경우
 - 개인정보보호법 등 법·규정 위반으로 제재가 확정된 경우, 신청사실에 허위가 있는 경우, 중복신청 경우
 - 대표자 또는 기업이 사회적 물의를 일으켜 지원이 합당하지 않다고 인정되는 경우 또는 사업계획서 이외 목적으로 사용(예: 가상화폐 채굴 등)
- 신청자격 요건(3. 신청자격)에 해당되는 증빙서류를 제출하지 않을 경우 선정 대상에서 제외함

□ 사업관련 문의처 : (대표전화) 02-580-0216, kaitcloud@kait.or.kr

구분		담당부서	담 당	연락처 / E-Mail
문의 처	국산 AI반도체 활용 (NPU) 자원 지원	한국정보통신진흥협회	안효주 대리	02-580-0213, 02-580-0212 / kaitcloud@kait.or.kr

참 고

'25년 NPU(국산 Si반도체 활용 사업) 지원 가능한 AI 모델

□ 기술지원 가능 AI 알고리즘

지원여부 분야	AI 알고리즘
시각 지능	ConvNeXt Series, ConvTasNet, DeepLabV3 Series, DeiT Series, DenseNet, Series, DPT-large, FCN_ResNet Series, GoogLeNet, Inception_V3, MC3_18, MNASNet Series, MobileNet_Series, R(2+1)D_18, R3D_18, RetinaFace, RegNet Series, ResNeXt Series, S3D, SAM2.1_hiera_large, UNet, VGG Series, ViT-large, YOLO Series
언어·음성 지능	Audio-Spectrogram-Transformer, BART Series, BGE-M3, BGE-Reranker Series, DistilBERT-base, E5-base-4K, EEVE-Korean-10.8b, EfficientNet Series, EXAONE-3.0 Series, Gemma Series, GPT2 Series, KoBART-base, Llama2 Series, Mi:dm Series, Mistral Series, Phi-2, Qwen Series, RoBERTa, Salamandra-7b, SecureBERT, SOLAR-10.7b, T5 Series, Wav2Vec2, Whisper Series, Wide_ResNet Series
생성 지능	BLIP2-2.7b, Llava-v1.6-mistral-7b, SDXL-turbo, Stable Diffusion Series,
분석 지능	EfficientNet Series, MotionBERT, Time-Series-Transformer,
기타	ShuffleNet_V2 Series, SqueezeNet1 Series

※ AI 알고리즘별 세부 지원규격은 선정된 후 “사용자와 AI 반도체사 협의”를 통해 기술지원 할 예정