과학기술정보통신부 **보도자료** 대체불가 대한민국

보도시점 배포시점 배포 2026. 8. 18.(화)

「독자 AI 파운데이션 모델」 프로젝트 2차 단계평가 결과 발표

- 4개 정예팀 모두 세계적 수준의 독자 AI모델 개발 성과 입증
- 3개 정예팀(업스테이지, SK텔레콤, LG AI연구원 정예팀, 가나다順) 다음 단계 진출

【관련 국정과제】 21. 세계에서 AI를 가장 잘 쓰는 나라 구현

과학기술정보통신부(장관 겸 부총리 배경훈, 이하 과기정통부)와 정보통신산업진흥원(원장 박윤규, 이하 NIPA)은 「독자 AI 파운데이션 모델」 프로젝트 2차 단계평가 결과를 공개하였다.

[독자 AI 파운데이션 모델 프로젝트 2차단계(`26년 상반기) 성과]

「독자 AI 파운데이션 모델」 프로젝트 2단계 참여 정예팀들은 '26년 상반기 동안 글로벌 빅테크 기업과의 자원 격차 등을 딛고, 글로벌 경쟁력 있는 독자 AI 파운데이션 모델을 개발해 오픈소스로 공개하였으며, 이를 통해 국내 AI 기술의 글로벌 경쟁력과 개방형 생태계 조성 역량을 입증했다.

정예팀별로 공개한 금번 독자 AI모델 개요(출처·설명 : 해당 정예팀)

구분	모델명(파라미터 수)	주요특징
모티프 테크놀로지스	Motif 3 (314B-A13B)	美·中 외 국가의 AI모델 중 세계 최고 성능을 기록한 에이전트 특화 AI모델로(AAII 기준), 코딩·에이전트 등 분야에 특화된 전문 AI 능력을 모델 하나로 통합
업스테이지	Solar Open2 (250B-A15B)	에이전트 최적화 AI모델로, 한 번의 프롬프트 안에 수백 페이지 이상의 문서 작업 처리와 엔비디아 H200 칩 2장으로 구동 가능해 비용 효율성 측면 강점 보유
SK텔레콤	A.X K2 (688B-A33B)	한국어·수학 벤치마크에서 글로벌 오픈소스 모델 중 최고 수준 성능 기록, 제조·국방·바이오 등 산업 분야 적용 확대 및 전 국민의 AI 전환 본격 지원
LG AI연구원	K-EXAONE 2.0 (750B-A37B)	국내 최대인 7,500억 개 파라미터 규모 모델로, 아파치 2.0 상용 라이선스 및 10개 국어 지원, 전작(K-EXAONE 1.0) 대비 10%↑ 개선 성능 달성(벤치마크 종합)

작년 말 1차 공개된 5개 AI모델에 이어, 2단계 참여 4개 정예팀이 최근 공개한 AI모델 4개 모두 '에포크 AI(Epoch AI, 미국 비영리 AI 연구기관)'의 '주목할만한 AI 모델(Notable AI Models)' 등재되었다. 에포크 AI가 '26년에 선정한 주목할만한 모델을 보유한 국가는 현재 미국·중국·한국 등 3개국에 불과하다.

Epoch AI의 주목할만한 모델(Notable Model) 등재 현황

Notable AI models | Frontier AI models | Large-scale AI models | All AI models

	Model	Organization	Publication date	Domain	Task
1	Motif-3	Motif Technologies	2026-08-07	Language	Language modeling/genera
2	Muse Spark 1.2	Meta AI	2026-08-05	Language	Language modeling/genera
3	K-EXAONE 2.0	LG AI Research	2026-07-31	Language	Language modeling/genera
4	A.X K2	SK Telecom	2026-07-29	Language	Code generation Language
5	Claude Opus 5	Anthropic	2026-07-24	Language Vision Multim	Language modeling/genera
6	Qwen 3.8 Max	Alibaba	2026-07-19	Language	Language modeling/genera
7	Kimi K3	Moonshot	2026-07-16	Language Multimodal Vi	Language modeling/genera
8	Inkling	Thinking Machines	2026-07-15	Language Multimodal Au	Language modeling/genera
9	Solar Open2 250B	Upstage	2026-06-28	Language	Language modeling Reaso
10	Claude Mythos 5	Anthropic	2026-06-09	Language Multimodal	Language modeling/genera

또한, AI 모델 평가 전문 플랫폼인 'Artificial Analysis'가 측정·발표하는 글로벌 AI 모델의 지능과 성능에 대한 종합 평가지수인 'AAII(Artificial Analysis Intelligence Index)'도 4개 정예팀 모두 31점 이상을 기록하며, 유럽, 캐나다 등의 주요 AI모델(예, Mistral Medium 3.5 등)을 모두 추월한 것으로 나타났으며, 40점 초과 점수에서는 미국·중국에 이어 우리 한국 모델이 포함되었다.

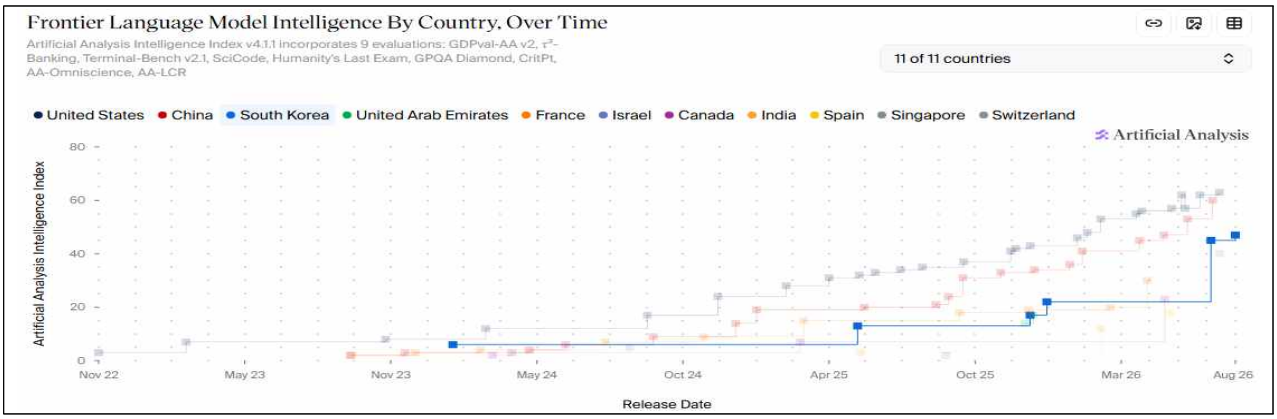
최상위 AI모델과 AAII 점수 격차도 축소*되고, 정예팀 AI모델이 오픈 웨이트(가중치 공개 모델) 모델 중 AAII 전세계 6위에 오르는** 등 미국·중국에 이은 AI 3강 지위를 보다 공고히 한 것으로 평가된다.

* 1차 평가 대비 최상위 모델과 격차 축소 : '26.1월 19점차(GPT-5.2 51점 vs 국내 정예팀 최고점 32점) → '26.8월 16점차(클로드 오퍼스5 63점 vs 국내 정예팀 최고점 47점) ※ AAII 기준 변동된 점도 감안 필요

** 1위Kimi K3(60점) → 2위Qwen3.8 2.4T A95B(58점) → 3위GLM 5.2(53점) → ... → 6위국내 정예팀 최고점(47점)

아울러, 현재 국가별 최고 수준(Frontier) 언어모델(Language Model) 성능을 비교하였을 때(AAII 기준), 우리나라는 미국·중국에 이은 3위에 위치하고 있다.

국가별 프론티어 AI모델 성능 추이(출처 : AA)



※ 주요 국가별로 AAII 최고 점수를 획득한 AI모델의 AAII 추이 분석(AA)

한편, AAII에서, 전 세계 AI기업별 최고 득점 AI모델을 기준으로 할 경우, 모티프테크놀로지스 정예팀이 개발한 독자 AI모델인 'Motif 3'가 글로벌 Top10에 오르는 성과도 기록하였다.

전세계 AI기업별 AAII 최고 득점 AI모델 순서(참고 : AAII, 현재 기준)

순서	기업	모델	AAII 점수
1	Anthropic	Claude Opus 5 (max)	63점
2	OpenAI	GPT-5.6 Sol (max)	61점
3	SpaceXAI	Grok 4.6 (high)	61점
4	Kimi	Kimi K3 (max)	60점
5	Alibaba	Qwen3.8 Max	58점
6	Meta	Muse Spark 1.2 (xhigh)	57점
7	Google	Gemini 3.7 Flash (high)	56점
8	DeepSeek	DeepSeek V4 Pro 0813 (max)	53점
9	Z AI	GLM-5.2 (max)	53점
10	**Motif Technologies**	**Motif 3**	**47점**

업스테이지 정예팀이 개발한 독자 AI모델 'Solar Open2'는 한 번에 기억하고 처리할 수 있는 텍스트 양(Context window)이 문서 수백 페이지 이상에 해당하는 최대 100만(1M) 토큰에 달해, 전 세계 최상위급 AI모델과 동일한 수준의 정보처리 능력을 확보하였다.

※ 현재 전 세계 최상위급 모델인 Claude Opus 5, GPT-5.6 Sol 등은 Context window가 모두 1M 토큰임

SK텔레콤 정예팀이 개발한 독자 AI모델 'A.X K2'는 국제수학올림피아드(IMO) 2026년 문제 평가에서 금메달 기준에 해당하는 점수를 기록, MathArena AIME 2026* 리더보드에서도 97.1%의 정확도로 공동 최고점을 달성하며, 세계 최정상급 수리 추론 역량을 입증하였다.

* 美 고교 수학 경시대회인 AIME(American Invitational Mathematics Examination) 문제 기반 평가

LG AI연구원 정예팀이 개발한 독자 AI모델 'K-EXAONE 2.0'은, AAII 벤치마크 지표 중 'AA-Omniscience Non-Hallucination Rate'(AI모델 환각(Hallucination) 최소화 지표*)에서 전 세계 AI모델 중 9위를 기록하며, 신뢰성 높은 AI모델 성능을 선보였다.

* AI모델이 모르는 정보를 마주했을 때 그럴듯한 거짓말(환각)을 하지 않고 안전하게 모른다고 대답하는 능력을 측정하는 지표(모르는 질문에 아는 척 추측하여 틀리는 성향을 강하게 감점)

이에 더해, 글로벌 최대 오픈소스 플랫폼인 '허깅페이스(Hugging Face)'는 「독자 AI 파운데이션 모델」 프로젝트를 정부 주도 오픈소스 정책의 우수 사례로 평가하며, 미국·중국 외 '대안국'으로서 AI역량을 보유한 우리나라와 공동 홍보 등 폭넓은 협력을 제안해 와 구체적 추진 방안을 논의 중에 있다.

이와 같이, 이번 상반기 AI모델 개발 경쟁 및 평가 등을 거쳐 대한민국이 자체 개발한 독자 AI모델들이 글로벌 최상위권 AI모델들과 견줄 수 있는 경쟁력을 확보했음을 국내외에 입증했다는 점에서 그 의미가 있다.

4개 정예팀 외에도 KT, 카카오, 네이버, NC AI 등 다양한 국내 AI기업들이 각자의 영역에서 의미있는 성과를 지속 창출하고 있으며*, 이 같은 K-AI 기업들은 ▵반도체 공정, ▵판례 분석, ▵신물질 개발 등 다양한 산업 분야는 물론, ▵정부 행정망, ▵R&D 예산 심의, ▵국민 AI경진대회 등 공공 분야에서도 폭넓게 접목되며**, 우리나라 AI 기업·생태계의 역량을 입증하고 있다.

* (KT) KT의 첫 번째 Omni AI모델, 믿:음(Midm) K 2.0 Omni Basic 개발, `26.7월
(카카오) 온디바이스용 경량 파운데이션 모델 4종 공개(Kanana-2-1.3B-base 등), `26.7월
(네이버) 국방 최적화 옴니모달 '하이퍼클로바X 시드 4B' 공개, `26.6월
(NC AI) 3D 생성 모델 '바르코 3D 2.0' 공개, 피지컬 AI 기반 강화 시동, `26.6월

** K-AI 현장 접목사례 10회 릴레이 홍보 보도참고자료(`26.5월` ~ `26.7월) 참조

이는 「독자 AI 파운데이션 모델」 프로젝트를 통해 우리나라의 AI모델 개발 역량과 경쟁력이 제고되고, 이러한 성과가 4개 정예팀을 넘어 국내 AI 생태계 전반의 AI모델 개발·확산으로 파급되면서, 결국 대한민국 AI 생태계 전체의 역량 강화와 성장·확장으로 이어지고 있음을 보여준다.

[2차 단계평가 개요]

이번 2차 단계평가는 ①벤치마크 평가(배점 40점), ②전문가 평가(배점 35점), ③사용자 평가(배점 25점)를 병행하는 입체적 평가로 진행되었다.

이 같은 2차 단계평가는 1차 단계평가(`26.1월) 대비 ▵대외 공신력 높은 고난도 글로벌 벤치마크 평가 도입(벤치마크 평가), ▵AI모델 개발에서 나아가 현장 AX 접목·확산 평가 확대(전문가 평가), ▵국민 참여형 사용성 평가 신규 도입(사용자 평가) 등을 통해 평가 체계를 한 층 정교화·고도화 하였다.

1·2차 단계평가 기준 변화

구분	1차 단계평가		2차 단계평가
벤치마크 평가	글로벌 공통·개별 벤치마크 NIA 벤치마크	⇨	AAII 벤치마크 NIA 벤치마크(분야 확장)
전문가 평가	개발전략 및 기술(10점) 개발 성과 및 계획(15점) 파급효과 및 기여계획(10점)	⇨	개발전략 및 기술(10점) 개발 성과 및 계획(10점) 파급효과 및 기여계획(15점)
사용자 평가	AI 전문 사용자진 평가	⇨	AI 전문 사용자진 평가 일반 국민 평가(추가)

이는 단계평가가 단순 정예팀 간 경쟁을 통해 압축·선별하는 데에 의의가 있는 것이 아니라, 정부가 평가 기준을 적극적으로 설계·진화시켜 나감으로써 정예팀들의 ▵글로벌 수준 AI모델 개발 역량 강화, ▵독자 AI모델 국내 AI생태계 현장 접목·확산, ▵국민의 눈높이에 부합하는 AI모델 개발 등을 유도하였다는 점에서 의미가 있다.

특히 이 평가 기준은 국내 최고 수준의 AI전문가들로 구성된 4개 정예팀과 충분한 협의를 거쳐 마련되었다는 점에서 더욱 중요한 의미를 갖는다.

[①벤치마크 평가 개요 및 결과]

①벤치마크 평가는 Artificial Analysis Intelligence Index(이하 AAII) 벤치마크 평가(배점 25점)와 NIA 벤치마크 평가(배점 15점)로 구성되었다.

AAII 벤치마크 평가는, AAII를 구성하는 ▵에이전트(Agents), ▵코딩(Coding), ▵일반(General), ▵과학적 추론(Scientific Reasoning) 분야의 고난도 벤치마크 9종으로 평가를 진행하였으며, Artificial Analysis(AA)社와의 협력 기반으로 AA社가 직접 벤치마크 평가를 진행해, 평가의 신뢰성·공정성을 제고하였다.

* AAII 영역별 벤치마크
  · 에이전트(Agents) : GDPval-AA v2, τ^3-Banking
  · 코딩(Coding) : Terminal-Bench v2.1, Scicode
  · 일반(General) : AA-LCR, AA-Omniscience
  · 과학적 추론(Scientific Reasoning) : Humanity's Last Exam, GPQA Diamond, CritPt

AAII는 ▵비정기적으로 벤치마크를 업데이트 해, 평가의 변별력을 제고하며, ▵일부 벤치마크는 문제·정답 등을 비공개하여, 벤치마크 오염(Contamination)을 최소화하는 한편, ▵다수의 최상위급 AI모델들도 고득점을 받기 어려운 벤치마크들로 구성하는 등 전 세계 유수의 글로벌 빅테크들도 고득점을 받기 어렵고 공신력 높은 벤치마크로 알려져 있다.

과기정통부는 정예팀들의 AI모델 개발 역량이 글로벌 프론티어 수준에 도달할 수 있도록 견인하고, 벤치마크 평가의 대내외 공신력을 확보하기 위해 이 같은 AAII 벤치마크 평가를 도입하였다.

NIA 벤치마크 평가는 ▵수학, ▵지식, ▵장문이해, ▵지시이행, ▵한국어 뿐 아니라, ▵안전성, ▵신뢰성 분야까지 포괄하여 벤치마크를 구성하였으며, 이를 통해 종합적 성능을 평가하는 데 주안을 두었다.

※ 1차 단계평가 당시 NIA 벤치마크 대비 ▵지시이행, ▵한국어 평가 분야 확장

NIA 벤치마크는 ▵전체 문제와 정답이 공개되지 않아 벤치마크 오염 우려가 매우 낮고, ▵AAII 벤치마크가 글로벌향 AI모델 범용 성능 평가에 중점을 뒀다면, NIA 벤치마크는 한국어·한국적 맥락에서의 성능과 안전성·신뢰성까지 종합 평가함으로써, 일부 벤치마크만으로는 검증하기 어려운 국내 AI 모델의 실질적 경쟁력을 가늠할 수 있는 지표로 의의가 있다.

벤치마크 평가(1차 단계평가)		벤치마크 평가(2차 단계평가)
글로벌 공통·개별 벤치마크	⇨	AAII 벤치마크
NIA 벤치마크		NIA 벤치마크(분야 확장)

①벤치마크 평가(배점 40점) 4개 정예팀 전체 평균은 22.5점이며, 1위와 4위 간 점수 격차는 4.0점이다.

[②전문가 평가 개요 및 결과]

②전문가 평가는 외부 AI전문가로 구성된 전문가위원회를 구성하여, 정예팀이 제출한 평가서류 등을 바탕으로 약 일주일 내외 간의 평가를 진행하였다.

※ 원활한 평가를 위해, 전문가위원회(평가자) - 정예팀(피평가자) 간 QnA 병행

정예팀들의 ▵개발 전략 및 기술(배점 10점), ▵개발 성과 및 계획(배점 10점), ▵파급효과 및 기여계획(배점 15점)을 중심으로 심층 평가하였다.

전문가위원회는 AI 알고리즘, 서비스, 데이터 분야 등의 전문성을 보유한 전문가들로, 이해충돌 문제가 없는 10명이 평가에 참여하였다.

단순한 독자성 확보 여부에서 나아가, '독자성 및 이를 기반으로 한 성능 향상 수준'을 전문가 평가 항목에 반영함으로써, 독자적 기술이 실제 AI모델 성능 향상으로 이어지는 경우, 이를 높게 평가할 수 있도록 하였다.

또한 국내외 AI생태계 파급효과·기여계획에 대한 평가 비중을 확대함으로써, 단순 고성능 AI모델 개발 평가에서 나아가, 실질적인 현장 확산 등 국내 AX(AI Transformation, AI전환) 등의 확대를 유도하는 데에 주안을 두었다.

전문가 평가(1차 단계평가)		전문가 평가(2차 단계평가)
개발전략 및 기술(10점)		개발전략 및 기술(10점)
개발 성과 및 계획(15점)	⇨	개발 성과 및 계획(10점)
파급효과 및 기여계획(10점)		파급효과 및 기여계획(15점)

②전문가 평가(배점 35점) 4개 정예팀 전체 평균은 28.8점이며, 1위와 4위 간 점수 격차는 2.4점이다.

[③사용자 평가 개요 및 결과]

③사용자 평가는 AI 전문 사용자진 평가(배점 15점)와 일반 국민 평가(배점 10점)으로 구성되었으며, 이를 기반으로 AI 사용성(Usability) 등에 대한 평가를 진행하였다.

AI 전문 사용자진 평가는 49명의 AI 전문 사용자(AI스타트업 대표 등)가 참여하여 평가를 진행하였으며, 일반 국민 평가는 185명(최종 200명 선정 후, 185명이 평가에 참여)의 국민이 직접 참여하여 평가를 진행하였다.

사용자 평가는 ▵단순 AI모델이 보여주는 성능에서 나아가, 실제 다양한 사용자들이 AI모델을 직접 사용하며 체감하는 사용성과 효능감까지 종합적으로 평가하기 위해 마련되었으며, ▵이를 통해 정예팀들이 AI모델 개발을 넘어 실사용자 관점의 AI서비스 제공 역량까지 함께 갖춰 나가도록 유도할 수 있다는 점에서 중요한 의의를 갖는다.

이에, 지난 1차 단계평가와 비교하여, AI전문가들의 눈높이 뿐 아니라, 일반 국민의 눈높이에서도 AI 사용성 평가가 이뤄질 수 있도록 다각적인 평가 체계를 구축하는 데 중점을 두었다.

사용자 평가(1차 단계평가)		**사용자 평가(2차 단계평가)**
AI 전문 사용자진 평가	⇨	AI 전문 사용자진 평가
		일반 국민 평가(추가)

각 평가단들에게는 프롬프트(Prompt) 가이드라인을 제공하여, 깊이있는 AI 사용성 평가를 지원하였으며, 평가단들의 평가는 정예팀별 비교평가가 아닌 절대평가로 이뤄졌다.

③사용자 평가(배점 25점) 4개 정예팀 전체 평균은 17.6점이며, 1위와 4위 간 점수 격차는 5.0점이다.

[2차 단계평가 종합 결과]

앞서와 같은, ①벤치마크 평가(배점 40점), ②전문가 평가(배점 35점), ③사용자 평가(배점 25점)를 병행하는 2차 단계평가를 종합한 결과, 근소한 차이로, 기존 4개 정예팀 중 업스테이지, SK텔레콤, LG AI연구원 정예팀(가나다順)이 다음 단계로 진출하게 되었다.

전문가위원회 평가 의견에 따르면,

업스테이지 정예팀은 “개발된 성과를 국민이 체감할 수 있도록 포털 ‘Daum’ 및 ‘Timely’ 플랫폼과 연계 추진”, “퓨리오사AI NPU 협업을 통해 외산

하드웨어 종속성(Lock-in)을 완화하고 실제 'Daum' 서비스 환경에서 PoC 형태로 실증한 것은 국산 AI SW와 HW 생태계 결합을 보여준 선도적 사례" 등의 호평을 받았다.

SK텔레콤 정예팀은 "수리 추론과 한국어 영역에서 비교군 최상위 성능을 확보하였고, 대규모 상용 서비스에 실제 탑재되어 사용성·실용성 측면의 우위 뚜렷", "국방 분야용 A.X K1 모델 제공, 제조 산업 특화 에이전트 실증, 법무 및 세무 분야 시연 등을 통해 개발 모델의 적용 가능성과 실용성을 다양한 산업 분야에서 입증" 등의 높은 평가를 받았다.

LG AI연구원 정예팀은, "글로벌 파급력을 높일 수 있도록 글로벌 국제기구 등과의 협업 전략을 수립하고 있다는 점과 Agentic AI에 대한 차별성이 있음". "안전하고 신뢰할 수 있는 AI 개발을 위한 고민과 관련 활동이 충실하게 나타나며, 모델의 신뢰성 확보와 위험요인 관리에 대한 인식을 통해 지속가능성까지 고려되고 있음" 등의 의미있는 평가를 기록하였다.

모티프테크놀로지스 정예팀은 "아키텍처부터 토크나이저, 옵티마이저, 커널까지 자체 개발하여 외부 종속을 제거", "아키텍처 개선 및 학습 효율화 기법은 Motif 3 모델에 성공적으로 적용되어 글로벌 평가에서 성과를 거둠", "독자 기술 스택을 적용하는 과정에서 도출된 기술적 한계점과 시행착오 데이터 역시 국가적 자산으로서의 가치가 높음" 등의 평가가 이어지는 등 깊이 있는 역량을 보유한 것으로 평가되었다.

특히 모티프테크놀로지스 정예팀은 AAII에서 미국·중국 외 국가에서 개발한 AI 모델 중 세계 최고 성능을 기록하여, 에이전트·코딩 등의 고난도 벤치마크 분야에서 글로벌 최선두 빅테크들에 버금가는 높은 수준을 선보이는 등 독자 AI모델 개발 분야에서 의미있는 성과를 보여준 만큼, 앞으로도 국내 AI 생태계에 다각도로 기여할 것으로 기대된다.

[시사점 및 향후계획]

「독자 AI 파운데이션 모델」 프로젝트는 유례없는 대규모 국내 산학연 협력을 이끌어냈음은 물론, 이들의 역량을 결집하여 우리나라의 AI 모델 개발 저력을

높이고 국내외에 알렸다는 점에 있어 의미있는 성과를 기록하며, 주요국에서도 이 같은 우리나라 정책에 깊이있는 관심을 보이고 있는 것으로 평가된다.

다음 단계에 진출한 정예팀에는 고성능 GPU(B200) 지원을 확대할 계획이며(`26.上 B200 768장 수준 → `26.下 B200 1,000장 수준), 이들 정예팀은 이를 통해 더 높은 성능의 독자 AI모델 개발에 주력할 것으로 기대된다.

※ GPU 지원 추이(B200 기준) : (`25.下) 500장 수준 → (`26.上) 768장 수준 → (`26.下) 1,000장 수준

한편, 최근 AI모델을 둘러싼 글로벌 경쟁이 보다 치열해지고, AI모델의 범분야 성능이 눈에 띄게 도약하는 가운데, 주요 선도국에서 AI모델의 국가안보 전략자산화 움직임이 현실화·가시화되고 있다.

이에, 과기정통부도 최상위급 독자 AI모델 개발 필요성을 깊이 인식하고, 기존의 방식을 뛰어넘는 새로운 지원체계를 관계부처와 논의 중에 있으며, 추후 구체적 내용을 준비해 발표할 계획이다.

담당 부서	인공지능정책실 인공지능기술기반정책과	책임자	과 장	양기성 (044-202-6560)
		담당자	사무관	이현우 (044-202-6566)

참고1 2차 단계평가 개요

구분		주요 내용
벤치마크 평가 (40점)	개요	"**글로벌 + NIA 벤치마크** 기반으로 **AI모델 성능 평가**"
	방안	① (**글로벌 벤치마크 평가**) 글로벌 주요 리더보드(AAII*)에서 활용하는 벤치마크를 활용하여 AI모델 성능 평가 * Artificial Analysis Intelligence Index ② (**NIA 벤치마크 평가**) ▵**수학**, ▵**지식**, ▵**장문이해**, ▵**안전성**, ▵**신뢰성**, ▵**한국어**, ▵**지시이행** 분야를 아우르는 NIA 벤치마크 기반 평가
	배점	▸ ① 글로벌 벤치마크 평가 **25점** + ② NIA 벤치마크 평가 **15점**
전문가 평가 (35점)	개요	"**전문가 평가위원회**를 구성하여, 제출 **서류** + **AI 사용 심층 평가**"
	방안	① (**평가위원회 구성**) **외부 전문가**로 평가위원회 구성 ② (**평가**) 정예팀의 **평가 서류** 등을 기반으로 AI 개발 성과, 향후 계획 무빙타겟 기반, 사용성, 생태계 파급계획·효과 등 **심층 평가**(QnA 병행) ※ 팀별 점수는 위원별 평가 점수 중 최고·최저를 제외한 나머지 평가점수 산술평균
	배점	▸ ① 개발 전략 및 기술 **10점** + ② 개발 성과 및 계획 **10점** + ③ 파급효과 및 기여계획 **15점** ※ 독자성 최소기준 점검 등 병행
사용자 평가 (25점)	개요	"**AI User** 집단 구성, **AI 사용성 평가**"
	방안	① (**집단 구성**) AI **전문 사용자진**(50명 이하) + **일반 국민**(200명 이상) 참여진 구성 ② (**평가**) 정예팀 AI모델 적용한 **AI서비스 웹사이트** 기반 AI 사용 평가 ※ 웹사이트별 UI/UX가 아닌 AI모델의 성능을 중심으로 평가되도록 Guide
	배점	▸ 각 사용자가 **절대평가***하여 점수를 합산하고, **전문 사용자진 평가 15점**, **일반 국민 평가 10점** 만점으로 스케일링하여 최종점수 반영 * 1점매우미흡 → 2점미흡 → 3점보통 → 4점우수 → 5점매우우수

참고2 전문가 평가 기준

분류	평가항목	세부항목	배점
개발 전략 및 기술	전략	▸ 개발 전략 수립 및 실행 - 당초 계획 대비 수행 충실도 및 성실성 - 개발 과정 중 리스크 관리 능력 및 대응의 적절성 - 컨소시엄 참여기관 간 협력 네트워크 실적	10
	기술력	▸ 개발 기술의 우수성·혁신성 - 기술적 독자성 및 이를 기반으로 한 성능 향상·수준 - 자원 활용·처리의 혁신성 및 효율성 - 안전하고 신뢰성있는 AI모델 개발 기술력 ※ 독자성 평가요소 예시 : ▲하이퍼파라미터 변경 등 양적 변형, ▲모듈·블록 레벨 등 질적 변형, ▲데이터 파이프라인의 효율적 개선 및 혁신 ▲아키텍처 수준 혁신 등	
개발 성과 및 계획	성과	▸ 개발 성과의 달성도 및 확장 가능성 - 개발 성과의 우수성 및 성능 달성도 - 개발 모델의 사용성 및 실용성 - 개발 모델의 다양한 도메인/시나리오 적용 가능성	10
	계획	▸ 향후 개발 계획 - 무빙타겟 기반 대응 전략의 혁신성 및 구체성 - 차기 목표·계획의 도전성·우수성 및 실현 전략 - 차세대 기술 도입 및 혁신 계획 - 지속적 개선을 위한 피드백 체계 및 대응 전략	
파급효과 및 기여계획	생태계 파급	▸ 국내 AI생태계 등 파급효과 및 기여 계획 - 국민 AI 접근성 증진 실적 및 지원 계획 - 전분야 AX 지원 등 AI생태계 파급실적 및 기여계획 - AI 모델 공개를 통한 AI생태계 확장 실적 및 계획 - 대학·스타트업, 국산 AI반도체 생태계 지원 실적·계획	15
	글로벌 파급	▸ 글로벌 파급력 및 지속 가능성 - 개발 모델의 글로벌 파급력 및 향후 파급계획 - 국가 AI 경쟁력 강화에 대한 기여 수준 및 전략 - 글로벌 시장 진출 실적 및 계획	
계			35

독자성 최소기준 점검	▸ 본 프로젝트로 금번 개발한 AI 파운데이션 모델이 ① 기술적, ② 정책적, ③ 윤리적 측면의 독자성 기준을 미충족함(체크 시 미충족 판단)	□