

「독자 AI 파운데이션 모델」 프로젝트 2차 단계평가 결과를 상세히 공개합니다.

- 독자 AI 파운데이션 모델 프로젝트(이하 독파모)는 단순한 순위 경쟁이나 탈락 팀 선정을 목적으로 하는 것이 아니라, 우리 AI기업들이 경쟁을 넘어 성장하고, 국내 AI 생태계의 질적 성장과 확장을 견인하는 데 그 취지가 있습니다.
 - 이에, 평가 결과 공개에 따른 일부 기업의 예기치 않은 직·간접적 피해를 최소화하는 한편, 평가 과정·결과에 대한 공정성과 투명성이라는 가치 역시 균형 있게 고려하여, 결과 공개 범위에 신중을 기해 왔습니다.
 - 다만, 8월 27일(목) 모티프테크놀로지스(이하 모티프)가 입장문을 통해 세부 평가 결과 공개를 요청하였고, 다른 정예팀들도 이에 동의함에 따라, 독파모 2차 단계평가 세부 점수를 공개합니다.
- 독파모는 빠른 AI기술 진화 상황 등을 고려하여 6개월 마다 AI모델 개발 목표를 변경 중이며(무빙타겟), 이에 따라 단계평가 기준과 방식도 매차시마다 정예팀 협의를 통해 마련 중입니다.
 - 이번 2차 단계평가 기준과 방식 역시, 정예팀들과 수차례 협의 거쳐 마련하였으며, 평가 전 최종안을 공식 안내한 결과, 정예팀 이견이 없었고, 이를 토대로 평가를 진행하였습니다.
- 지난 보도자료(8.18일), 보도설명자료(8.20일) 등을 통해 상세 평가 기준, 주요 점수 등을 공개한 바 있으며, 상세 점수를 추가 공개합니다.
 - 2차 단계평가는 ①벤치마크 평가(①-1AAII 벤치마크 평가, ①-2NIA 벤치마크 평가), ②전문가 평가, ③사용자 평가(③-1AI 전문 사용자진 평가, ③-2일반국민 평가)를 병행하였습니다.

□ ①⁻¹ AAI(Artificial Analysis Intelligence Index, 이하 AAI) 벤치마크 평가는

- AA(Artificial Analysis)에서 6월말 공표한 기준을 채택해 진행하였으며, AA가 직접 평가 결과를 도출해 제공하였고, AAI 벤치마크 점수는 사전에 공지한 대로 25점 만점 스케일링해 최종 점수에 반영하였습니다.

※ 예 : AAI 점수 60점일 시, 최종 점수는 15점($=60\text{점} \times (25/100)$) 반영

- 이를 종합한 AAI 벤치마크 평가 점수는 모티프 정예팀이 11.9점으로 1위, 업스테이지 정예팀이 9.4점으로 2위, SKT 정예팀이 8.8점으로 3위, LG AI연구원이 7.8점으로 4위를 기록하였습니다.

구분	업스테이지	SK텔레콤	LG AI연구원	모티프테크놀로지스
AAI 벤치마크 평가	9.4점	8.8점	7.8점	11.9점

□ ①⁻² NIA(한국지능정보사회진흥원, 이하 NIA) 벤치마크는

- 7개 분야(△수학, △지식, △장문이해, △안전성(사회적 안전성 포함), △신뢰성, △한국어, △지시이행) 벤치마크 데이터셋으로 NIA에서 평가 결과를 도출하고 15점 만점 스케일링해 최종 점수에 반영하였습니다.

※ 예 : NIA 벤치마크 평가 점수 60점일 시, 최종 점수는 9점($=60\text{점} \times (15/100)$) 반영

- 이를 종합한 NIA 벤치마크 평가 점수는 SKT 정예팀이 13.4점으로 1위, 업스테이지 정예팀이 13.3점으로 3위, LG AI연구원 정예팀이 12.8점으로 3위, 모티프 정예팀이 12.7점으로 4위를 기록하였습니다.

구분	업스테이지	SK텔레콤	LG AI연구원	모티프테크놀로지스
NIA 벤치마크 평가	13.3점	13.4점	12.8점	12.7점

□ ② 전문가 평가는

- 외부 전문가(산업계 3명, 학계 5명, 연구계 2명)로 평가위원회를 구성해 추진하고, 정예팀이 제출한 평가서류 등을 바탕으로 약 5일 간의 서면 평가와 QnA를 병행해 진행하였습니다.

- 전문가 평가 기준은 ¹⁾개발 전략 및 기술 10점, ²⁾개발 성과 및 계획 10점, ³⁾파급효과 및 기여계획 15점이며(독자성 최소기준 점점 병행), 각 평가위원 별 점수는 이를 합산하여 산정하였습니다.

※ 정예팀별 평가점수 산출은 위원별 평가점수 중 최고·최저 점수를 제외한 나머지 평가점수를 산술 평균

- 이를 종합한 전문가 평가 점수는 LG AI연구원 정예팀이 29.5점으로 1위, SKT 정예팀이 29.3점으로 2위, 업스테이지 정예팀이 29.1점으로 3위, 모티프 정예팀이 27.1점으로 4위를 기록하였습니다.

구분	업스테이지	SK텔레콤	LG AI연구원	모티프테크놀로지스
전문가 평가	29.1점	29.3점	29.5점	27.1점

□ ③-1 AI 전문 사용자진 평가는

- AI에 대한 전문적인 식견·경험 등을 갖추고, 일반 국민 대비 실제 다양한 AI 개발 과정에서 AI모델을 폭넓게 사용해 본 AI 스타트업 대표들로서 이해충돌 확인 등을 거쳐 50명이 선정되어, 실제 49명이 평가에 참여하였습니다.

- 각 팀의 AI 사용 웹사이트를 기반으로 AI 사용성 등에 대한 평가를 진행해 최저 1점부터 최고 5점까지 점수를 부여하도록 하였고, 이를 합산해 15점 만점으로 스케일링하여 최종 점수를 산출하였습니다.

※ 1점(매우미흡) - 2점(미흡) - 3점(보통) - 4점(우수) - 5점(매우우수)

※ 예 : 전문 사용자진(49명)의 점수 합산 후 평균 점수가 3.7점일 시, 최종 점수는 11.1점(=3.7점×(15/5)) 반영

- 이를 종합한 AI 전문 사용자진 평가는 SKT 정예팀이 11.6점으로 1위, LG AI연구원 정예팀이 11.3점으로 2위, 업스테이지 정예팀이 10.8점으로 3위, 모티프 정예팀이 8.4점으로 4위를 기록하였습니다.

구분	업스테이지	SK텔레콤	LG AI연구원	모티프테크놀로지스
AI 전문 사용자진 평가	10.8점	11.6점	11.3점	8.4점

□ ③-2 일반 국민 평가는

- 일반 국민에 대한 대표성 제고를 위해, 충분한 기간 동안 공모한 후, 국내 주민등록인구 통계 기준으로 성별·연령별 쿼터를 설정하고, 시스템을 통해 무작위로 200명이 선정되었고, 실제 185명이 평가에 참여하였습니다.

※ 상기 인원 모두는 4개 정예팀과 이해충돌 확인 등을 거쳐 확정된 인원임

- 국민 평가단의 평가 역시 정예팀별 절대평가로 진행되어, 최저 1점부터 최고 5점까지 점수를 부여하도록 하였고, 이를 합산한 후 10점 만점으로 스케일링하여 최종 점수를 산출하였습니다.

※ 1점(매우미흡) - 2점(미흡) - 3점(보통) - 4점(우수) - 5점(매우우수)

※ 예 : 일반 국민(185명)의 점수 합산 후 평균 점수가 3.7점일 시, 최종 점수는 7.4점 (=3.7점×(10/5)) 반영

- 이를 종합한 일반 국민 평가는 LG AI연구원이 7.6점으로 1위, SKT 정예팀이 7.5점으로 2위, 업스테이지 정예팀이 7.3점으로 3위, 모티프 정예팀이 5.7점으로 4위를 기록하였습니다.

구분	업스테이지	SK텔레콤	LG AI연구원	모티프테크놀로지스
일반국민 평가	7.3점	7.5점	7.6점	5.7점

- 앞선 평가 점수를 종합한 결과, SKT 정예팀이 70.6점으로 1위, 업스테이지 정예팀이 69.9점으로 2위, LG AI연구원 정예팀이 69.0점으로 3위, 모티프 정예팀이 65.8점으로 4위를 기록하였으며, 세부 점수는 붙임으로 공개해 드립니다.

구분	업스테이지	SK텔레콤	LG AI연구원	모티프테크놀로지스
벤치마크 평가	22.7점	22.2점	20.6점	24.6점
AAI 벤치마크 평가	9.4점	8.8점	7.8점	11.9점
NIA 벤치마크 평가	13.3점	13.4점	12.8점	12.7점
전문가 평가	29.1점	29.3점	29.5점	27.1점
사용자 평가	18.1점	19.1점	18.9점	14.1점
AI 전문 사용자진 평가	10.8점	11.6점	11.3점	8.4점
일반국민 평가	7.3점	7.5점	7.6점	5.7점
계	69.9점	70.6점	69.0점	65.8점

- 이번 결과 공개 이후에도, 우리 기업들이 단순한 점수나 순위로 평가받기 보다, 보유한 저력과 역량을 바탕으로 국내외 AI생태계에 폭넓게 기여하고, 우리 AI생태계가 새롭고 경쟁력 있는 AI기업들이 끊임없이 도전하는 역동적인 생태계로 발전해 나가기를 기대합니다.

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



내일을 만드는 과학기술
내일을 채우는 디지털·AI

대한민국
지정브리핑



붙임1

AAII 벤치마크 평가 결과

벤치마크		업스테이지	SK텔레콤	LG AI연구원	모티프테크놀로지스
GDPval-AA v2		31.0	30.5	24.0	38.7
τ^3 -Banking		21.6	16.1	11.5	35.3
Terminal-Bench v2.1		44.2	39.0	40.4	74.9
SciCode		45.6	38.5	41.0	40.6
AA-LCR		68.3	66.0	57.7	72.3
AA-Omniscience	Accuracy	18.8	18.6	13.1	30.1
	Hallucination Rate	25.4	33.0	22.6	28.4
HLE (Humanity's Las Exam)		28.5	29.6	18.6	37.0
GPQA Diamond		85.7	85.7	82.9	83.4
CritPt		5.7	8.9	0.9	6.6
계		37.4	35.0	31.0	47.4
환산(25점 만점)		9.4	8.8	7.8	11.9

붙임2

NIA 벤치마크 평가 결과

벤치마크(배점)	업스테이지		SK텔레콤		LG AI연구원		모티프테크놀로지스	
	원점수	환산점수	원점수	환산점수	원점수	환산점수	원점수	환산점수
수학(2)	90.9	1.819	91.7	1.834	82.1	1.641	87.6	1.753
지식(3)	78.4	2.352	83.1	2.493	71.4	2.143	74.4	2.231
장문이해(3)	84.4	2.533	83.0	2.491	85.0	2.549	80.3	2.408
지시이행(1)	100.0	1.000	99.8	0.998	99.8	0.998	100.0	1.000
한국어(1)	97.3	0.973	97.5	0.975	92.7	0.927	93.2	0.932
신뢰성(2.5)	95.2	2.380	95.3	2.383	86.6	2.165	85.9	2.146
안전성(1.5)	86.2	1.293	79.2	1.188	89.5	1.342	82.3	1.235
사회적안전성(1)	100.0	1.000	99.0	0.990	98.8	0.988	99.2	0.992
계(15)	732.4	13.349	728.7	13.352	705.8	12.753	702.8	12.697
환산(15점 만점)	13.3		13.4		12.8		12.7	

붙임3

전문가 평가 결과

구분	업스테이지	SK텔레콤	LG AI연구원	모티프테크놀로지스
위원1	27	28	29	24
개발 전략 및 기술	8	8	9	7
개발 성과 및 계획	8	8	8	7
파급효과 및 기여계획	11	12	12	10
위원2	26	25	25	29
개발 전략 및 기술	8	7	7	9
개발 성과 및 계획	8	7	8	10
파급효과 및 기여계획	10	11	10	10
위원3	32	30	31	26
개발 전략 및 기술	10	9	10	9
개발 성과 및 계획	9	8	9	7
파급효과 및 기여계획	13	13	12	10
위원4	32	30	29	32
개발 전략 및 기술	9	8	8	9
개발 성과 및 계획	9	8	8	9
파급효과 및 기여계획	14	14	13	14
위원5	34	31	33	29
개발 전략 및 기술	10	9	9	8
개발 성과 및 계획	9	8	9	7
파급효과 및 기여계획	15	14	15	14
위원6	31	32	31	31
개발 전략 및 기술	9	9	9	9
개발 성과 및 계획	9	10	9	9
파급효과 및 기여계획	13	13	13	13
위원7	30	20	24	17
개발 전략 및 기술	9	6	7	6
개발 성과 및 계획	9	6	7	5
파급효과 및 기여계획	12	8	10	6
위원8	26	30	28	22
개발 전략 및 기술	7	9	8	5
개발 성과 및 계획	7	8	8	6
파급효과 및 기여계획	12	13	12	11
위원9	29	30	33	28
개발 전략 및 기술	8	8	10	8
개발 성과 및 계획	7	9	10	8
파급효과 및 기여계획	14	13	13	12
위원10	26	30	30	28
개발 전략 및 기술	7	8	9	9
개발 성과 및 계획	7	8	8	6
파급효과 및 기여계획	12	14	13	13
계	29.1	29.3	29.5	27.1

※ 위원별 평가점수 중 최고·최저 점수를 제외한 나머지 평가점수를 산술 평균함

붙임4

AI 전문 사용자진 평가 결과

구분	업스테이지	SK텔레콤	LG AI연구원	모티프테크놀로지스
전문 사용자진 1	3	4	2	1
전문 사용자진 2	2	4	3	5
전문 사용자진 3	5	5	4	4
전문 사용자진 4	3	4	2	3
전문 사용자진 5	2	5	4	1
전문 사용자진 6	2	5	1	3
전문 사용자진 7	4	5	5	3
전문 사용자진 8	3	4	4	3
전문 사용자진 9	4	3	2	1
전문 사용자진 10	4	3	4	2
전문 사용자진 11	5	4	5	4
전문 사용자진 12	4	4	4	4
전문 사용자진 13	4	5	5	3
전문 사용자진 14	4	3	2	4
전문 사용자진 15	4	5	4	1
전문 사용자진 16	2	3	2	4
전문 사용자진 17	4	4	5	2
전문 사용자진 18	3	3	4	2
전문 사용자진 19	1	5	3	4
전문 사용자진 20	4	2	3	3
전문 사용자진 21	3	3	2	3
전문 사용자진 22	4	5	5	4
전문 사용자진 23	5	5	5	3
전문 사용자진 24	4	4	3	4
전문 사용자진 25	5	5	5	4
전문 사용자진 26	5	3	5	4
전문 사용자진 27	3	4	5	4
전문 사용자진 28	4	4	5	2
전문 사용자진 29	5	4	3	2
전문 사용자진 30	3	4	5	3
전문 사용자진 31	5	3	4	1
전문 사용자진 32	4	4	4	3
전문 사용자진 33	4	3	4	4
전문 사용자진 34	2	1	2	1
전문 사용자진 35	5	5	4	5
전문 사용자진 36	5	4	4	2
전문 사용자진 37	4	2	3	1
전문 사용자진 38	5	5	5	1
전문 사용자진 39	4	3	4	3
전문 사용자진 40	3	4	4	3
전문 사용자진 41	3	4	5	3
전문 사용자진 42	4	5	4	3
전문 사용자진 43	3	4	4	2
전문 사용자진 44	3	4	5	3
전문 사용자진 45	1	3	4	4
전문 사용자진 46	4	3	3	4
전문 사용자진 47	3	4	3	2
전문 사용자진 48	2	4	3	1
전문 사용자진 49	4	3	4	1
계	176	189	184	137
환산(15점 만점)	10.8	11.6	11.3	8.4

붙임5

일반 국민 평가 결과

구분	업스테이지	SK텔레콤	LG AI연구원	모티프테크놀로지스
일반국민 1	4	2	5	3
일반국민 2	3	4	3	3
일반국민 3	4	5	4	4
일반국민 4	4	5	3	4
일반국민 5	5	3	4	1
일반국민 6	5	5	5	4
일반국민 7	2	5	5	3
일반국민 8	5	4	5	3
일반국민 9	4	4	3	3
일반국민 10	4	4	3	2
일반국민 11	3	4	4	3
일반국민 12	3	4	5	3
일반국민 13	4	4	5	2
일반국민 14	4	4	5	3
일반국민 15	2	4	2	4
일반국민 16	4	4	5	3
일반국민 17	4	3	4	1
일반국민 18	5	5	5	4
일반국민 19	4	3	4	2
일반국민 20	1	4	4	2
일반국민 21	4	3	5	4
일반국민 22	4	4	3	3
일반국민 23	2	3	5	2
일반국민 24	4	4	4	3
일반국민 25	3	4	5	2
일반국민 26	1	5	3	2
일반국민 27	5	4	4	2
일반국민 28	3	3	4	3
일반국민 29	4	3	5	3
일반국민 30	4	5	5	4
일반국민 31	4	3	1	3
일반국민 32	4	4	5	3
일반국민 33	5	3	4	3
일반국민 34	3	3	4	4
일반국민 35	2	4	5	2
일반국민 36	5	5	5	3
일반국민 37	4	5	5	3
일반국민 38	4	3	5	4
일반국민 39	4	3	5	1
일반국민 40	4	3	4	2
일반국민 41	3	3	2	3
일반국민 42	3	4	2	2
일반국민 43	4	4	3	3
일반국민 44	4	4	5	3
일반국민 45	3	3	3	2
일반국민 46	4	3	2	4

구분	업스테이지	SK텔레콤	LG AI연구원	모티프테크놀로지스
일반국민 47	3	2	3	2
일반국민 48	3	4	3	3
일반국민 49	4	4	5	4
일반국민 50	4	4	3	4
일반국민 51	2	4	4	2
일반국민 52	3	3	3	2
일반국민 53	5	4	4	3
일반국민 54	3	3	4	2
일반국민 55	2	5	3	2
일반국민 56	4	3	5	1
일반국민 57	4	4	4	2
일반국민 58	4	4	4	3
일반국민 59	4	5	5	4
일반국민 60	2	1	3	2
일반국민 61	3	5	4	2
일반국민 62	2	3	3	4
일반국민 63	4	5	2	5
일반국민 64	2	3	2	3
일반국민 65	4	5	4	3
일반국민 66	5	5	5	5
일반국민 67	5	5	5	4
일반국민 68	3	4	3	3
일반국민 69	2	5	3	2
일반국민 70	2	4	2	3
일반국민 71	3	3	4	2
일반국민 72	3	4	5	3
일반국민 73	5	4	4	3
일반국민 74	5	5	3	4
일반국민 75	4	4	3	5
일반국민 76	4	4	5	3
일반국민 77	4	3	4	3
일반국민 78	4	3	5	3
일반국민 79	4	5	5	4
일반국민 80	3	5	4	2
일반국민 81	2	5	2	3
일반국민 82	3	3	4	2
일반국민 83	3	5	5	3
일반국민 84	3	4	5	2
일반국민 85	4	3	3	2
일반국민 86	4	3	2	1
일반국민 87	4	2	1	3
일반국민 88	5	4	3	1
일반국민 89	4	5	3	4
일반국민 90	3	4	4	2
일반국민 91	3	5	1	2
일반국민 92	4	4	3	4
일반국민 93	3	2	3	1

구분	업스테이지	SK텔레콤	LG AI연구원	모티프테크놀로지스
일반국민 94	5	4	4	3
일반국민 95	3	4	4	3
일반국민 96	4	4	5	3
일반국민 97	5	4	5	3
일반국민 98	5	4	4	3
일반국민 99	4	4	3	2
일반국민 100	2	3	4	5
일반국민 101	4	5	3	4
일반국민 102	3	5	4	2
일반국민 103	5	4	4	5
일반국민 104	4	3	4	2
일반국민 105	4	3	4	2
일반국민 106	4	3	2	2
일반국민 107	3	4	4	3
일반국민 108	3	4	2	3
일반국민 109	2	3	2	3
일반국민 110	3	4	5	2
일반국민 111	4	5	5	4
일반국민 112	4	3	4	2
일반국민 113	3	4	5	4
일반국민 114	3	4	5	3
일반국민 115	4	4	5	4
일반국민 116	3	4	3	3
일반국민 117	4	3	4	3
일반국민 118	2	4	4	2
일반국민 119	3	3	5	2
일반국민 120	4	3	2	3
일반국민 121	4	3	1	2
일반국민 122	5	4	5	3
일반국민 123	5	4	4	3
일반국민 124	5	5	4	4
일반국민 125	3	4	2	3
일반국민 126	2	3	4	3
일반국민 127	3	5	4	2
일반국민 128	4	5	4	4
일반국민 129	4	5	5	4
일반국민 130	3	3	5	2
일반국민 131	4	2	4	2
일반국민 132	3	2	3	1
일반국민 133	4	4	2	3
일반국민 134	4	4	4	3
일반국민 135	4	3	4	4
일반국민 136	3	2	2	3
일반국민 137	5	4	4	2
일반국민 138	4	3	4	2
일반국민 139	3	5	5	4
일반국민 140	3	3	4	2

구분	업스테이지	SK텔레콤	LG AI연구원	모티프테크놀로지스
일반국민 141	5	4	2	4
일반국민 142	4	4	4	2
일반국민 143	3	4	5	3
일반국민 144	2	4	3	2
일반국민 145	5	2	1	3
일반국민 146	4	3	4	2
일반국민 147	4	3	4	3
일반국민 148	5	4	4	3
일반국민 149	3	3	4	2
일반국민 150	5	5	4	4
일반국민 151	3	4	5	3
일반국민 152	3	2	4	2
일반국민 153	4	3	5	3
일반국민 154	5	3	4	4
일반국민 155	5	5	4	3
일반국민 156	4	4	4	3
일반국민 157	3	3	4	2
일반국민 158	3	3	4	3
일반국민 159	4	4	1	3
일반국민 160	5	5	5	3
일반국민 161	4	3	3	3
일반국민 162	3	2	3	1
일반국민 163	5	5	4	4
일반국민 164	3	2	3	1
일반국민 165	2	4	5	3
일반국민 166	4	5	5	3
일반국민 167	4	2	3	3
일반국민 168	3	4	1	2
일반국민 169	3	3	2	3
일반국민 170	4	3	3	2
일반국민 171	4	1	3	1
일반국민 172	3	5	3	3
일반국민 173	4	4	4	3
일반국민 174	5	5	4	4
일반국민 175	4	4	4	2
일반국민 176	5	5	4	4
일반국민 177	3	3	5	2
일반국민 178	4	3	4	2
일반국민 179	3	4	4	4
일반국민 180	3	4	4	3
일반국민 181	3	5	4	3
일반국민 182	4	4	5	4
일반국민 183	3	4	5	3
일반국민 184	5	5	5	5
일반국민 185	5	4	4	5
계	675	698	701	528
환산(10점 만점)	7.3	7.5	7.6	5.7