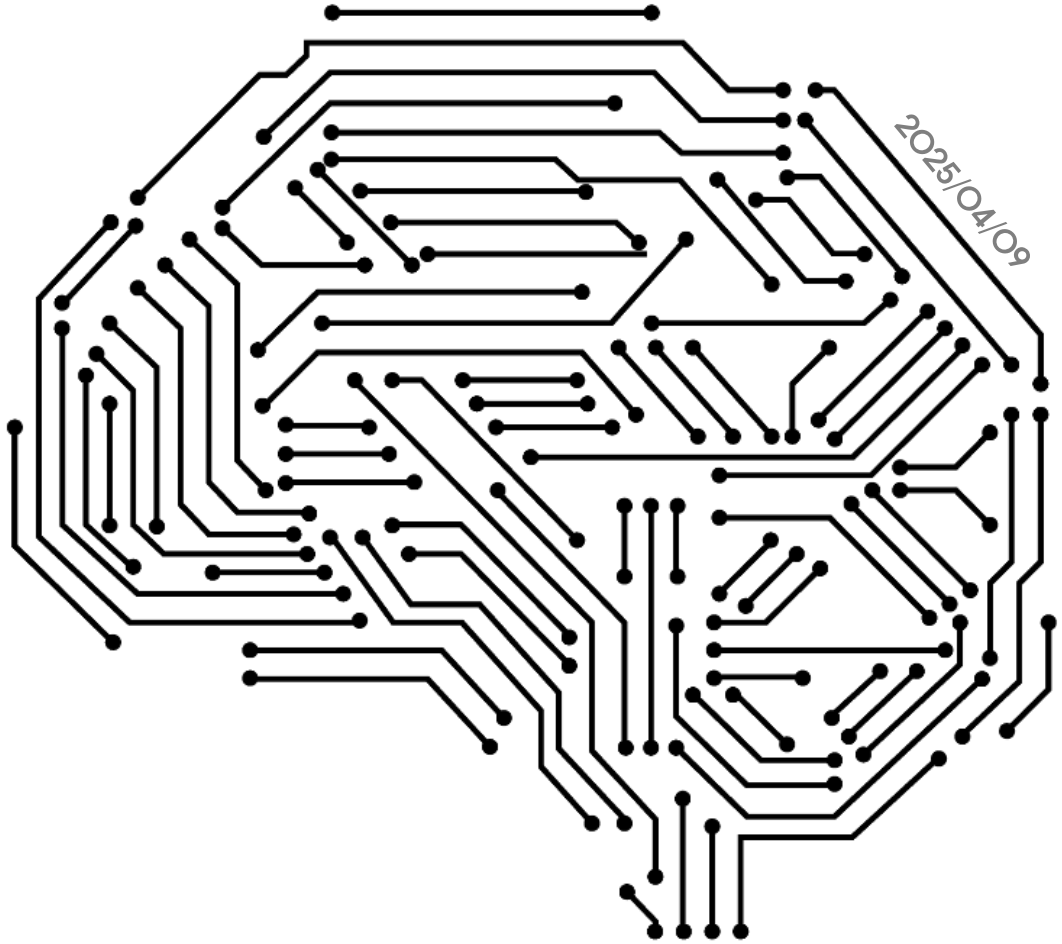
성장산업 조태나 02)368-6174 tena@eugenefn.com



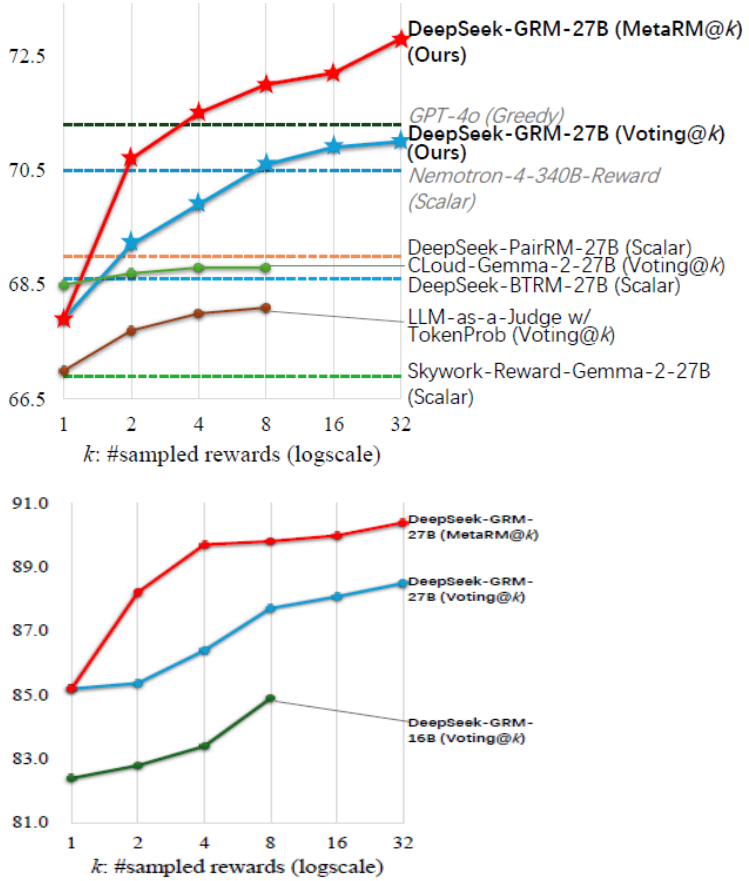
2025/04/09

AI WEEKLY 15

Key AI Insights of the Week

이제 스스로 기준을 만들고 판단하는 AI가 옵니다
feat. DeepSeek R2

DeepSeek는 2025년 4월 3일, 칭화대학교와 함께 새로운 보상 모델인 DeepSeek-GRM(Generalist Reward Modeling)을 공개. 해당 모델은 GRM과 SPCT(Self-Principled Critique Tuning)을 결합한 모델로, 답변에 대해 유연하게 점수를 매긴다는 부분에서 기존 모델과 차별화됨. 모델이 스스로 원칙을 만들고 비판적 사고를 하게 만듦. 주어진 질문에 대해 여러 번 샘플링을 해서 다양한 보상 점수를 생성하고, 점수를 합산, 선별(Voting)하여 최종 점수를 결정하는 과정을 거침. 이번 모델 역시 모델 크기 보다는 추론 시간을 확장하는 것이 더 효과적임. 중요한 점은 훈련 비용은 감소하지만, 기존 단일 추론 대비 추론 비용은 기하급수적으로 증가할 수 있다는 것.- 샘플링을 반복할수록 추론 비용 증가. 1-2주 내 공개 될 R2 모델에 해당 기술이 적용되었을 가능성 높음. 스스로 생각하는 AI가 한 단계 더 진화.



출처: Inference-Time Sacaling for Generalist Reward Modeling, 5 Apr 2025

NEWS

- ✓ 딥시크 R2에 적용될 고효율 보상 모델(RM) 공개
- ✓ 메타, 멀티모달 모델 Llama4 출시
- ✓ MS, '기억력'가진 AI 어시스턴트 공개
- ✓ A&O Shearman, 로펌 업무 속도 향상 위한 AI 개발
- ✓ 앱러빈 틱톡 글로벌 인수 참전
- ✓ 서비스형 GPU 고객군 중소기업, 스타트업으로 다양화
- ✓ 솔트룩스, AWS 마켓플레이스에 LLM 출시
- ✓ IBM, AI 시대를 위한 새로운 메인프레임 출시
- ✓ IBM, AI 컨설팅 투자 확대를 위해 '하코다' 인수
- ✓ 인도에서 chatGPT 사용자 급증, 수익화는 미진
- ✓ 트럼프 행정부, 데이터센터 전력 공급에 석탄 활용 계획

GLOSSARY

"GPUaaS"

GPU as a Service로, GPU 컴퓨팅 파워를 클라우드에서 제공하는 IaaS의 한 형태. 사용자는 하드웨어를 소유하지 않고도 GPU를 활용해 작업이 가능.
필요한 만큼 GPU를 사용하고, 사용한 만큼만 비용을 지불하며, 하드웨어의 유지보수, 냉각, 전력 관리 등을 서비스 제공자가 해결해주기 때문에 사용자는 작업에만 집중할 수 있음.

IaaS가 서버, 스토리지, 네트워크 등 인프라 전체를 임대하는 서비스라면, GPUaaS는 GPU에 특화된 인프라만 제공하는 특수한 형태. 두 가지 서비스 모두 클라우드 기반이고 온디맨드라는 점은 같음.

PEER TABLE

		1W	1M	3M	1Y	YTD	Mcap($B)	price($)	12FWD PER
Foundation Model									
Alphabet	GOOGL	-6.6	-15.6	-24.3	-5.2	-22.5	1,803.6	146.75	15.5
Meta	META	-11.9	-17.5	-15.5	-0.6	-11.8	1,308.0	516.3	19.0
Alibaba	BABA	-20.1	-24.6	26.6	49.1	25.0	253.3	106.0	10.5
Infra / Cloud & Datacenter									
Microsoft	MSFT	-6.4	-9.0	-15.7	-15.7	-15.1	2,660.3	357.9	24.5
Amazon	AMZN	-8.8	-12.0	-21.1	-5.4	-20.1	1,857.4	175.3	23.6
Oracle	ORCL	-10.4	-18.0	-22.1	2.3	-23.7	356.6	127.2	19.4
IBM	IBM	-9.8	-13.7	1.2	18.9	2.7	209.4	225.8	20.8
Coherent	COHR	-17.1	-20.6	-43.6	-7.9	-44.2	8.2	52.8	12.6
Datadog	DDOG	-12.7	-19.5	-37.8	-29.9	-38.3	30.4	88.2	47.5
Equinix	EQIX	-8.5	-13.0	-21.4	-5.1	-20.7	72.8	747.7	54.9
Digital Realty Trust	DLR	-6.4	-8.4	-24.4	-4.7	-22.9	46.9	136.7	94.0
Hardware									
Nvidia	NVDA	-11.4	-13.4	-30.3	12.1	-27.3	2,382.4	97.6	20.4
TSMC	TSM	-13.3	-17.5	-29.4	2.4	-26.0	758.1	146.2	15.4
broadcom	AVGO	-8.5	-20.9	-32.8	15.4	-33.5	724.8	154.1	21.6
AMD	AMD	-18.6	-16.6	-31.4	-50.8	-30.8	135.9	83.6	16.6
ARM	ARM	-17.2	-29.4	-39.0	-31.4	-28.2	93.2	88.6	43.5
Micron	MU	-22.9	-26.5	-31.2	-44.4	-18.8	76.4	68.4	7.3
Marvel Tech.	MRVL	-18.6	-28.0	-56.9	-29.1	-53.8	44.2	51.0	17.0
Amphenol	APH	-7.8	-3.2	-15.2	4.4	-12.3	73.8	60.9	25.6
Intel	INTC	-11.2	-5.2	-1.6	-48.5	-2.4	85.3	19.6	29.6

출처: Bloomberg, 유진투자증권

PEER TABLE

		1W	1M	3M	1Y	YTD	Mcap($B)	price($)	12FWD PER
Salesforce	CRM	-9.6	-13.7	-25.3	-19.1	-27.0	234.7	244.2	21.4
Palantir	PLTR	-8.1	-8.3	14.1	238.4	2.9	182.6	77.8	132.6
Cisco	CSCO	-12.0	-14.9	-8.1	12.8	-8.1	216.5	54.4	13.9
Servicenow	NOW	-9.4	-13.5	-29.9	-6.4	-30.6	152.3	735.5	42.6
SAP	SAP	0.0	0.0	0.0	0.0	0.0	198.8	161.8	21.9
Applovin	APP	-17.9	-14.1	-29.5	197.5	-28.3	78.9	232.2	30.9
Datadog	DDOG	-12.7	-19.5	-37.8	-29.9	-38.3	30.4	88.2	47.5
Snowflake	SNOW	-12.4	-17.5	-18.6	-15.4	-15.1	43.8	131.0	106.0
Reddit	RDDT	-15.5	-30.6	-45.9	101.5	-43.1	16.8	93.0	37.0
Dexcom	DXCM	-8.5	-20.8	-23.0	-56.0	-20.8	24.2	61.6	28.8
Twilio	TWLO	-14.7	-21.1	-24.6	36.6	-22.5	12.8	83.7	18.4
Illumina	ILMN	-6.3	-15.2	-47.4	-41.9	-45.0	11.6	73.5	15.8
Duolingo	DUOL	-9.4	0.3	-8.2	35.4	-8.5	13.4	296.5	101.4
C3.ai	AI	-10.7	-15.0	-42.5	-23.3	-44.4	2.5	19.1	-
Digital Turbin	APPS	-19.1	-22.4	23.2	6.0	34.9	0.2	2.3	-

출처: Bloomberg, 유진투자증권

당사는 자료 작성일 기준으로 지난 3개월 간 해당종목에 대해서 유가증권 발행에 참여한 적이 없습니다. 당사는 본 자료 발간일을 기준으로 해당종목의 주식을 1% 이상 보유하고 있지 않습니다. 당사는 동 자료를 기관투자가 또는 제3자에게 사전 제공한 사실이 없습니다. 조사분석담당자는 자료작성일 현재 동 종목과 관련하여 재산적 이해관계가 없습니다. 동 자료에 게재된 내용들은 조사분석담당자 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭 없이 작성되었음을 확인합니다. 동 자료는 당사의 제작물로서 모든 저작권은 당사에게 있습니다. 동 자료는 당사의 동의 없이 어떠한 경우에도 어떠한 형태로든 복제, 배포, 전송, 변형, 대여할 수 없습니다. 동 자료에 수록된 내용은 당사 리서치센터가 신뢰할 만한 자료 및 정보로부터 얻어진 것이나, 당사는 그 정확성이나 완전성을 보장할 수 없습니다. 따라서 어떠한 경우에 동 자료는 고객의 주식투자의 결과에 대한 법적 책임소재에 대한 증빙자료로 사용될 수 없습니다.