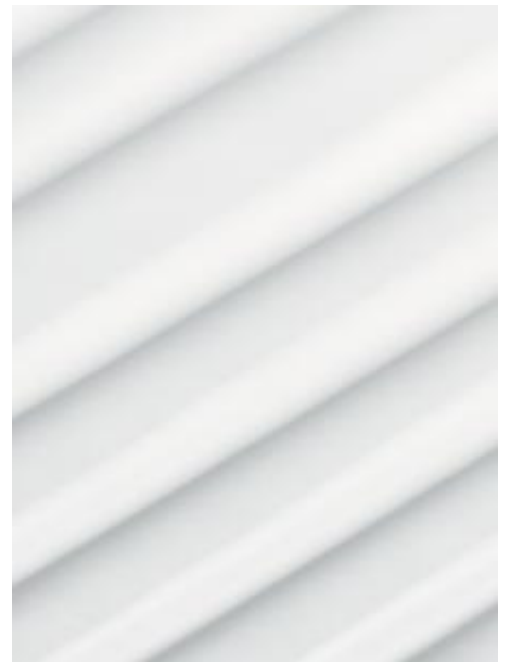


Chip News

September 2026

반도체 산업 최신 이슈를

카드뉴스로 한눈에 확인하세요.



HBM을 넘어선 새로운 시도, HBF와 HBC

The Next Step Beyond HBM : HBF and HBC

엔비디아, 소버린 AI 분야에서도 점유율 92%

NVIDIA Takes 92% of Sovereign AI Infrastructure

엔트로픽, 자체 칩 설계 본격화

Anthropic Begins Its Own Custom Silicon Effort

HBM을 넘어선 새로운 시도, HBF와 HBC

HBM의 대역폭, 용량, 전력 한계가 동시에 노출되며 새로운 메모리 계층 등장
SK하이닉스, 샌디스크의 HBF와 퀄컴의 HBC가 각기 다른 해법 제시

	HBM	HBF	HBC
Announced	2013 (JEDEC JESD235)	Aug 4, 2026 (FMS 2026)	Jun 24, 2026 (Qualcomm Investor Day)
Led by	  	 	  
Position	The workbench next to the GPU Incumbent standard for training	A capacity tier between HBM and SSD	An alternative to the HBM + CoWoS package
Ships	Now	TBD	AI250, 2027
Solves	Bandwidth	Capacity	Power

8월 4일 SK하이닉스와 샌디스크가 FMS 2026에서 HBF(High Bandwidth Flash) 첫 표준 규격을 OCP를 통해 공개했습니다. HBM과 SSD 사이의 포지션으로, 낸드를 TSV로 적층해 최대 512GB 용량과 약 0.4~3.0TB/s 대역폭을 제공하고 UCIe로 CPU·GPU에 붙습니다.

지난 6월에는 퀄컴이 HBC(High Bandwidth Compute)를 공개했습니다. 인터포저 없이 LPDDR 스택 아래에 컴퓨팅 다이를 두는 구조로, HBM 대비 와트당 대역폭 6배를 주장합니다. 삼성전자와 SK하이닉스 모두 기존 HBM 로드맵과 병행해 개발에 참여합니다.

The Next Step Beyond HBM: HBF and HBC

New memory emerge as HBM hits bandwidth, capacity and power limits
HBF from SK hynix and Sandisk, HBC from Qualcomm

	HBM	HBF	HBC
Announced	2013 (JEDEC JESD235)	Aug 4, 2026 (FMS 2026)	Jun 24, 2026 (Qualcomm Investor Day)
Led by	  	 	  
Position	The workbench next to the GPU Incumbent standard for training	A capacity tier between HBM and SSD	An alternative to the HBM + CoWoS package
Ships	Now	TBD	AI250, 2027
Solves	Bandwidth	Capacity	Power

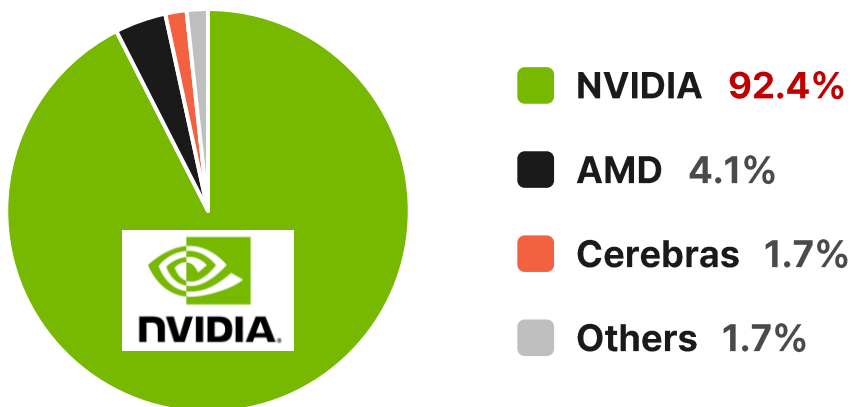
Kimi K3 built, optimized, and verified a 4 mm² chip in a single 48-hour autonomous run, using open-source EDA tools on the Nangate 45nm library — closing timing at 100 MHz with 1.46M standard cells and an INT4 MAC array.

Paired with the ACE-RTL agent, NVIDIA's Nemotron 3 Ultra averages a 97.1% pass rate across nine CVDP categories (100% on debugging) while using 71% fewer tokens per iteration than Kimi K2.6.

엔비디아, 소버린 AI 분야에서도 점유율 92%

미국, 중국 제외 55개국의 자체 LLM 92%가 엔비디아 칩 기반으로 구축
호스팅은 내재화 했으나 AI 칩은 여전히 CUDA 생태계에 종속

Counterpoint Research
AI 360 Practice: Sovereign AI Research, July 2026



Source : Counterpoint Research

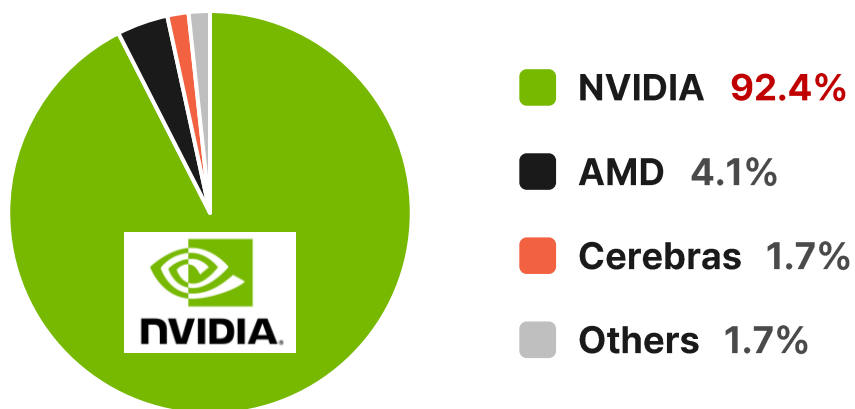
카운터포인트리서치가 8월 공개한 소버린 AI LLM 보고서에 따르면, 미국·중국을 제외한 80여 개국 가운데 자체 LLM을 개발 중인 약 55개국의 170개 이상 모델을 분석한 결과 92%가 엔비디아 칩 위에서 학습 및 추론되고 있습니다.

보고서의 요지는 "호스팅은 자국화됐지만 AI 칩은 그렇지 않다"는 것입니다. 다만 추론 시장에서 균열이 발생하고 있습니다. 퀄컴은 사우디 휴메인과, 리벨리온은 SK텔레콤과 손잡았고, 국내에서는 모레와 업스테이지가 AMD 플랫폼에서 운영되고 있습니다.

NVIDIA Takes 92% of Sovereign AI Infrastructure

92% of sovereign LLMs outside the US and China built on NVIDIA silicon
Hosting has localized, yet AI chips remain bound to the CUDA ecosystem

Counterpoint Research AI 360 Practice: Sovereign AI Research, July 2026



Source : Counterpoint Research

Counterpoint Research's August sovereign AI LLM report found that 92% of the models surveyed are trained and served on NVIDIA chips. It covered 55+ countries outside the US and China.

Hosting has localized while AI silicon has not. The cracks are opening in inference: Qualcomm works with Saudi Arabia's HUMAIN, Rebellions with SK Telecom, and in Korea, Moreh and Upstage run on AMD.

엔트로픽, 자체 칩 설계 본격화

클로드 전용 커스텀 칩 설계를 위한 사내 실리콘 팀 구성 공식 확인
엔비디아 이탈이 아닌 원가 절감을 겨냥한 멀티칩 전략의 연장선

ANTHROPIC

8월 5일 엔트로픽이 클로드 전용 커스텀 칩을 설계할 사내 실리콘 팀 구성을 공식 확인했습니다. 하드웨어와 모델을 함께 설계해 추론 속도와 비용을 개선하는 것이 목표이며, 앞서 6월에는 오픈AI 자체칩 프로그램 출신 인력을 영입한 것으로 알려졌습니다.

회사는 이를 "멀티칩 전략의 다음 단계"로 규정하며 AWS, 구글, 엔비디아, AMD 하드웨어를 계속 쓴다고 밝혔습니다. 7월에는 삼성전자 2nm 파운드리와 첨단 패키징 활용을 논의 중이라는 보도가 나왔으나, 칩의 용도와 성능은 아직 확정되지 않았습니다.

Anthropic Begins Its Own Custom Silicon Effort

In-house silicon team confirmed to design custom chips for Claude
An extension of its multi-chip strategy targeting inference cost

ANTHROPIC

On August 5, Anthropic confirmed it is assembling an in-house silicon team to design custom chips for Claude. The aim is to co-design hardware and models so inference runs faster and cheaper.

Anthropic frames this as the latest step in a multi-chip approach and says it will keep deploying hardware from AWS, Google, NVIDIA and AMD. In July, reports said it was discussing Samsung's 2nm foundry and advanced packaging, though the chip's purpose and performance targets are reportedly still undecided.