

쿠팡, NVIDIA와 혁신을 위한 기술 협력 강화

2026. 10. 8.



지난 9월, 쿠팡은 AI 스타트업, 독립 소프트웨어 벤더, 대기업 등 20개 기업의 리더 약 40명을 초청해 라운드테이블 행사를 개최했습니다. 이날 행사에 참석한 참석자들은 GPU 추론 인프라와 에이전틱 AI 생태계를 실제 비즈니스 의사결정과 연결하는 방안에 대해 논의했습니다.

쿠팡은 엔비디아 클라우드 파트너(NVIDIA Cloud Partner, NCP)로서 쿠팡 인텔리전트 클라우드(Coupang Intelligent Cloud, 이하 CIC)를 통해 대규모 GPU 인프라를 운영하고 있으며, 국내 AI 스타트업 및 기업들과의 협력을 꾸준히 확대하고 있습니다. 이번 행사는 참석 기업들과 쿠팡 및 CIC의 인프라 역량, 그리고 엔비디아의 폭넓은 기술 생태계를 연결하고, 실질적인 협력과 새로운 비즈니스 기회를 모색할 수 있도록 마련됐습니다.

쿠팡은 AI, 로보틱스, 클라우드 서비스 등 첨단 기술 전반에 걸쳐 지속적으로 인프라에 투자하며 기술 혁신을 이어가고 있습니다. 이러한 노력의 일환으로 쿠팡은 엔비디아(NVIDIA)와의 기술 협력을 강화하고 있습니다.

이날 행사에서 엔비디아는 'Inference Tech Trend & NVIDIA Solutions(추론 기술 동향 및 NVIDIA 솔루션)'를 주제로 기술 발표를 진행하며, 추론 기술의 최신 동향과 관련 솔루션을 소개했습니다. 엔비디아는 AI가 실제 서비스와 고객 접점의 영역으로 확장되면서 산업 전반의 중심이 모델 학습(Training)에서 추론(Inference)으로 이동하고 있다고 설명했습니다.

NVIDIA Cloud Partner 모델에 기반해 운영되는 CIC는 갈수록 고도화되는 추론 워크로드를 지원하기 위해 인프라 역량을 확대하고 있습니다. CIC는 NVIDIA GB300 NVL72 시스템 도입을 통해 대규모 AI 모델과 고도화된 추론 워크로드를 보다 효율적으로 지원하는 한편, 엔터프라이즈 규모의 AI 인프라를 구축하고 운영하는 과정에서 발생하는 복잡성을 줄이는 것을 목표로 하고 있습니다.

또한, 이번 행사는 쿠팡과 엔비디아가 이어온 기술 협력을 확대하는 자리이기도 했습니다. 쿠팡은 지난 3월 NVIDIA AI Conference & Expo(GTC)에서 양사의 협력 성과를 소개했습니다. 당시 쿠팡은 CIC와 엔비디아 DGX SuperPOD를 결합해 'AI 팩토리(AI factory)'를 구축했다고 설명했습니다. AI 팩토리는 엔지니어와 데이터 사이언티스트가 AI 모델을 보다 빠르게 테스트하고 구축할 수 있도록 지원하는 셀프서비스 AI 생태계입니다.

이러한 인프라 개선을 통해 물류센터 스케줄링과 빈패킹(bin-packing) 최적화에 활용되는 AI 모델의 GPU 활용률은 65%에서 95%로 향상됐습니다. 쿠팡은 또한 에이전틱 추론을 위해 설계된 오픈소스 소프트웨어인 엔비디아 다이내모(NVIDIA Dynamo)의 런칭 파트너로 참여하며, 엔비디아와의 기술 협력을 지속적으로 확대하고 있습니다.

CIC는 첨단 AI 인프라를 엔지니어와 데이터 사이언티스트가 대규모로 보다 쉽게 활용할 수 있도록 설계됐습니다. GPU 컴퓨팅 자원과 UI, API, 커맨드라인 도구를 통해 이용할 수 있는 셀프서비스 기능을 결합해 AI 워크로드 운영에 수반되는 복잡성을 줄이고, 각 팀이 AI 모델의 개발, 테스트 및 배포에 보다 집중할 수 있도록 지원합니다. 현재 CIC는 미국 시애틀과 캘리포니아 마운틴뷰를 비롯한 쿠팡의 글로벌 엔지니어링 조직이 회사 전반의 AI 애플리케이션을 개발하고 운영하는 데 활용되고 있습니다.

쿠팡의 기술 혁신은 대외적으로도 꾸준히 인정받고 있습니다. 쿠팡은 올해 3월 글로벌 정보서비스 기업 렉시스넥시스(LexisNexis)가 발표한 ‘2026 글로벌 혁신기업 톱 100’에 2년 연속 이름을 올렸습니다. 2025년에는 패스트컴퍼니(Fast Company)가 선정한 ‘세계에서 가장 혁신적인 기업’ 리테일 부문에서 2위를 차지했습니다.

쿠팡은 이번 라운드테이블을 계기로 엔비디아와의 기술 협력을 더욱 강화하는 한편, CIC의 인프라 및 운영 역량을 지속적으로 고도화해 나갈 계획입니다. 이를 통해 국내 AI 스타트업과 기업들이 실험 단계에 머무르지 않고 AI 서비스를 실제 프로덕션 환경으로 확장할 수 있도록 지원할 방침입니다.

취재 문의 media@coupang.com