

# "앱 없는 스마트폰 시대 온다"...韓 'AI 최적화' 기업 주목

팡동현 입력 2026. 5. 25. 15:28 수정 2026. 5. 25. 18:23

AI에이전트 중심 인터페이스 변혁 임박  
온디바이스AI 경량화·최적화 필요성 ↑

앱을 일일이 실행하지 않아도 인공지능(AI)이 사용자의 의도를 파악해 스스로 작업을 수행하는 이른바 '앱 없는 AI스마트폰' 시대가 다가오고 있다. AI에이전트가 모바일 경험의 중심으로도 파고들면서, 휴대용 기기에서 AI모델을 효율적으로 구동하기 위한 경량화·최적화 기술 또한 관심을 받는다.

25일 업계와 외신에 따르면 오픈AI는 퀄컴 및 미디어텍과 AI 중심 스마트폰용 프로세서 개발을 추진하고 있으며, 중국 렉스웨어가 시스템 공동 설계·제조 파트너로 거론된다. 앱 실행 중심의 기존 사용방식에서 벗어나 AI기반의 맞춤형·자동화가 전제되는 이 'AI에이전트 스마트폰'은 내년 상반기 출시를 목표로하는 것으로 전해졌다.

이런 '앱리스 경험'을 위해선 운영체제(OS)·하드웨어·AI가 통합된 구조가 필요하다는 점에서 산업 전반의 구조적 변화를 예고한다. 앞서 삼성전자도 갤럭시 S26 시리즈에 빅스비, 구글 제미나이, 퍼플렉시티를 함께 탑재하고 각각 역할을 수행할 수 있도록 했다. 구글 역시 '제미나이 인텔리전스'를 앞세워 '멀티 앱 자동화' 기능을 강화하는 등 스마트폰 인터페이스가 AI 중심으로 재편되는 추세다.

스마트폰이 단순한 앱 실행기기를 넘어, 풍부한 실시간 데이터를 바탕으로 AI에이전트의 상시 작동 플랫폼으로 거듭나는 셈이다.

이를 위해선 하드웨어와 소프트웨어 영역을 아우르는 온디바이스AI 구현 역량이 요구된다. 클라우드 등 인터넷 연결 없이도 기기 내부에서 AI 연산을 처리함으로써 효율과 보안을 동시에 확보할 수 있다.

하지만 AI 연산량이 늘어날수록 스마트폰 발열, 전력 소모, 지연 시간, 메모리 한계 등 과제도 심화된다. 이에 단순히 고성능 칩을 탑재하는 것을 넘어, AI모델을 더 작고 빠르게, 낮은 전력으로 구동하도록 만드는 AI 경량화·최적화 기술이 핵심적인 요소로 떠오르고 있다. 국내 AI스타트업들도 이런 수요를 발 빠르게 공략하는 모습이다.

먼저 노타는 자사 AI 최적화 플랫폼 '넷츠프레소'를 앞세운다. 회사에 따르면 AI모델 크기를 90% 이상 줄이면서도 정확도를 유지할 수 있다. 삼성전자와 계약을 맺고 모바일 애플리케이션 프로세서(AP) 엑시노스 2400부터 2600까지 최적화 기술을 지속적으로 공급하고 있으며, AI 개

발 툴체인인 '엑시노스 AI 스튜디오' 차세대 버전 개발에도 참여해 양자화, 컴파일 최적화 등을 자동화한 파이프라인을 구축하고 있다.

AI 경량화 분야에서는 스쿼즈비츠 역시 주요 플레이어로 꼽힌다. 스쿼즈비츠는 AI 모델의 압축·양자화·추론 최적화 기술을 바탕으로 GPU와 NPU 환경 모두에서 거대언어모델(LLM) 서빙 효율을 높이는 솔루션을 개발하고 있다. 창업 4년 만에 누적 매출 50억원을 돌파했고, 2년 연속 영업이익 흑자를 달성했다.

아울러 클리카는 AI모델을 제한된 하드웨어 환경에 맞게 압축·최적화하는 기술을 개발, 중소벤처기업부와 LG전자가 진행한 '온디바이스AI 챌린지'에서 우수기업으로 선정된 바 있다. 또 제텍 AI는 AI모델 최적화, 변환, 실제 기기 벤치마킹, 배포를 자동화해 CPU·GPU·NPU 등 다양한 하드웨어에서 온디바이스AI를 구현할 수 있게 해주는 소프트웨어 인프라를 제공한다. 이밖에 모바일 AP의 미세공정 전환과 NPU 고도화도 온디바이스AI 경쟁을 가속한다.

향후 스마트폰 내에서 AI의 비중이 커질수록 지속적·안정적 에이전트 구동 성능도 관건이 될 전망이다. 업계 관계자는 "이제 스마트폰은 사용자가 조작하는 기기가 아니라, 사용자의 의도를 이해하고 실행하는 AI 플랫폼으로 진화하고 있다"며 "온디바이스AI 최적화 기술을 확보한 기업이 이 경쟁의 핵심 플레이어가 될 것"이라고 말했다.

팡동현 기자 dhp@dt.co.kr

