

생성형 AI 시대, 새로운 퍼스널 디바이스 개발 경쟁 시작

유지호 · 김영혁

jihoryu@lgbr.co.kr · kimyoungghuk@lgbr.co.kr

생성형 AI를 온전히 활용하기에 가장 적절한 디바이스 후보 중 하나는 역시 스마트폰이다. 그러나 스마트폰의 여러 장점에도 불구하고, 현재의 스마트폰은 생성형 AI의 생산성 향상 가치를 온전히 누리기 어렵다는 것과 사용자의 상황, 맥락을 온전히 센싱하는 'Always-on' 센서로 활용하기에는 폼팩터적 측면에서 약점이 있다고 지적된다.

이를 보완하기 위해서는 스마트폰의 '귀'가 되어주는 액세서리 형태의 디바이스와, 특정 업무 수행에 있어 '눈'과 '손'을 갖춰 생산성을 극대화하는 독립 디바이스의 형태가 제시되고 있다. 이러한 시도를 통해 다양한 형태나 UI의 디바이스들이 출몰하고, 기존 밸류체인이 변화가 나타날 수 있다. 관련 고객, 기술, 산업의 변화 움직임을 면밀히 살펴봐야 할 때다.

2023년 9월, ChatGPT 열풍의 주역인 OpenAI의 샘 알트만(Sam Altman)이 아이폰 디자이너로 유명한 조나단 아이브(Jonathan Ive), 그리고 ARM의 소유주인 소프트뱅크 손정의 회장과 만남을 가졌다. 생성형 AI 기반의 새로운 퍼스널 디바이스에 대한 논의를 위한 자리였다. 이 만남이 언론에서 화제가 되자, 샘 알트만은 이는 초기 아이디어 검증 차원의 논의에 불과하다며 억측 자체를 요청했지만, OpenAI가 주도하는 AI 기반, 새로운 퍼스널 디바이스에 대한 기대는 해가 바뀌었음에도 여전히 크다. 기존의 스마트폰, PC, 가전, 자동차 및 각종 스마트기기 사업자들의, 생성형 AI 기술 접목 시도가 활발한 이 때, 미래의 퍼스널 디바이스(하드웨어)는 어떤 모습일지 많은 이들이 주목하는 것이다.

스마트폰은 생성형 AI의 잠재력을 모두 담아낼 수 있을까?

생성형 AI가 적용된 미래 표준 퍼스널 디바이스의 가장 유력한 후보는 역시 스마트폰이다. 하루 종일 사용할 수 있는 배터리, 다양한 기능을 제공할 수 있는 고성능 칩셋, 정보를 입출력할 수 있는 다양한 센서와 터치 스크린 기반의 직관적 UI를 가진 스마트폰은 이미 완성형 디바이스이고, 이는 생성형 AI의 잠재력을 구현하는데 부족함이 없어 보인다. 그러나, 스마트폰은 여러 장점에도 불구하고, 생성형 AI를 온전히 활용하지 못할 것이라는 우려도 나오고 있다.

먼저, 생산성과 관련된 개인 업무 처리를 위한 기기로서의 한계다. 생성형 AI는 콘텐츠 생성, 아이디어 브레인스토밍, 문서의 교정·요약, 문서 내 키워드 추출 및 분류 등의 다양한 작업에 있어 생산성을 획기적으로 높이는데 특히 유용할 것으로 기대되고 있다. 그러나 현재 스마트폰은 사용 시간의 60%~85%가 소셜 미디어 접속과 게임, 동영상 관람 등 미디어 소비에 쓰이고 있을 정도로 생산성 향상을 위한 기기와는 거리가 멀어 보인다. 이는, 현재 스마트폰이 가지고 있는 디스플레이, 터치식 입력 인터페이스, 앱을 찾고 실행하는 UI, 폼팩터 등의 요소들이 생산성 향상 업무보다는 콘텐츠 소비 및 커뮤니케이션 용도에 최적화되어 있음을 말해준다. 실제 문서 작성, 콘텐츠 생성 등 생산성과 관련된 많은 업무들은 일반적으로 스마트폰보다 PC나 노트북, 태블릿 또는 오프라인 작업을 통해 이뤄지는 것이 선호되는 상황이다.

또한, 스마트폰은 개인 사용자의 상황과 맥락을 상시 파악하기 위한 센싱 기기로는 부족한 측면을 가지고 있다. 지금까지는 사용자가 디바이스를 학습하는 것이 일반적이었지만, 우리가 마주하게 될 생성형 AI 시대에는 디바이스가 사용자와 사용자를 둘러싼 맥락을 학

습하여 사용자에게 개인 맞춤형 서비스를 제공하게 될 가능성이 크다. 이를 위해서는 생성형 AI 디바이스가 사용자의 시선을 항상 따라 다니며 주변의 모든 상황과 소리를 담아내는 'Always-on' 센서의 기능을 갖추는 것이 필수적인데, 현재의 스마트폰은 이와 같은 역할을 수행하기에 크기나 무게, 형태 등 폼팩터적인 측면에서 약점을 가지고 있다.

즉, 생성형 AI를 활용한 업무 처리 프로세스가 ①센싱·인풋(눈과 귀) → ②사고·연산(머리) → ③액션·아웃풋(손)의 3단계라고 했을 때, 스마트폰은 가지고 있는 고성능 칩셋의 AI 연산 기능 발전에 따라 머리의 역할은 충분히 수행할 수 있겠으나, 눈과 귀, 손의 기능을 온전히 담당하기에는 폼팩터와 인터페이스, 구동방식 등에 있어 한계가 있을 것으로 예상된다.

퍼스널 디바이스에 귀와 눈, 그리고 손을 이식하기 위한 시도 활발

위와 같은 스마트폰의 한계를 돌파하기 위해 다양한 기업들은 새로운 시도를 진행 중이다. 지금까지 나타나는 시도들은 크게 두 가지의 방향성을 보인다. 기존 스마트폰을 메인 디바이스로 두고 부족한 센싱 기능은 액세서리 디바이스를 통해 보완하는 방향과, 특정 업무 수행에 있어 생산성과 효율성을 극대화할 수 있는 새로운 유형의 독립 디바이스를 선보이는 방향이다.

▲ 눈과 귀가 되어주는 보완형 액세서리 디바이스

유명 벤처 투자자 안드리센 호로위츠(Andreessen Horowitz)가 투자한 스타트업 Limitless AI(舊, Rewind AI)는 스마트폰과 PC에 연동하여 사용할 수 있는 펜던트 형태의 기기를 출시했다. 사용자 및 사용자 주변의 모든 소리를 녹음 후 해당 내용을 컴퓨터에 연동시키는 기능밖에 없는 간단한 기기이지만, 이와 같은 Always-on 센싱 기기가 스마트폰·PC 등과의 연동을 통해 구현할 수 있는 기능은 꽤 위력적이다.

먼저 이 기기는 사용자가 참여한 과거 대화, 회의 내용 등을 스마트폰 등을 통해 언제든지 초 단위로 검색할 수 있게 한다. 또한, 사용자의 대화 내용을 학습·분석하여 사용자의 관심사나 취향 등을 파악한 상태에서 초개인화 서비스를 제공한



Limitless AI의 AI 펜던트, 사용자 주변 소리를 녹음 후 스마트폰에 연동하여 생성형 AI 서비스에 활용
출처: Limitless AI

다. 작동 원리는 다음과 같다. 펜던트를 통해 녹음된 모든 내용은 먼저 암호화한 후 텍스트로, 데이터화되어 스마트폰의 생성형 AI에 보내진다. 이 때 스마트폰에 설치된 Limitless AI 앱은 사용자의 스마트폰 사용 현황을 모두 스크린 샷으로 찍어 저장하는 작업도 수행한다. 생성형 AI는 이렇게 펜던트와 스마트폰에서 획득한 모든 정보를 학습하여 사용자의 맥락을 이해하고, 이를 바탕으로 사용자의 요청에 응답한다. 예컨대 사용자가 자신의 하루와 업무 정리를 요청하면 Limitless AI는 확보한 개인 데이터에 기반하여, 본인이 오늘 어떤 일을 했고, 미팅의 내용은 무엇이었고, 향후 어떻게 계획을 세워 나갈지 등에 대해 알려줄 수 있다.

웨어러블 AI 스타트업 업체 Tab 또한 이와 유사한 AI 펜던트 기기 출시를 앞두고 있다. Tab의 펜던트 역시 사용자의 모든 일상을 녹음하여 스마트폰으로 전송하는데, Tab은 낮은 품질로 녹음하고, 최소의 전력으로 이 기능을 구현하는데 집중하는 것으로 보인다. Tab의 창업자 아비 쉬프만(Avi Schiffmann)은 “AI에는 새로운 인터페이스가 필요한 것이 아니라 Always-on 디바이스가 필요한 것”이라 말한다. 스마트폰에 귀의 역할을 보완해줄 수 있는 음성 센싱 액세서리만 더해지면 스마트폰이 최적의 생성형 AI 디바이스로 자리매김할 수 있다는 그의 믿음을 엿볼 수 있다.



Tab의 AI 펜던트, Always-on 보완형 액세서리를 추구
출처: Tab

▲ 눈과 귀, 손이 있는 독립형 디바이스

애플 출신 직원들이 설립한 Humane은 스마트폰 없이도 구동될 수 있는 퍼스널 디바이스, ‘AI Pin’을 선보였다. 이름에서 유추가 가능하듯이 AI Pin은 브로치와 같이 옷에 부착



Humane의 AI Pin, 브로치 형태로 사용자 주변 음성 및 시각 정보 센싱
출처: Humane

할 수 있는 형태로 디자인된 소형 디바이스로, 사용자 및 주변의 음성 정보뿐 아니라 시각 정보까지 센싱하는 데 특화되어 있는, ‘눈’이 달린 디바이스다. 비록 불법 촬영, 프라이버시 등의 이슈로 평상시에는 Always-on 기능이 비활성화되어 있지만, 원할 때면 언제나 바로 사용자와 시선을 함께하며 ChatGPT 기반의 개인 비서 역할을 수행한

다. 자연스러운 자연어 대화 기반의 Q&A는 물론 외국어 통번역이나 카메라로 인식한 음식의 영양소 정보 전달 등의 작업을 수행할 수 있다. AI Pin은 작업 중 중요 정보를 출력할 수 있도록 간단한 LED 프로젝터를 내장하고 있지만, 그 외 모든 인터페이스는 음성을 기반으로 이뤄진다.

2024년 CES에서 큰 주목을 받은 Rabbit이라는 업체는 여기서 한발 더 나아가 생성형 AI에 ‘눈과 귀’뿐 아니라 ‘손’의 기능까지 부여한 디바이스를 출시했다. R1이라는 이름의 이 독립형 디바이스는 일반 스마트폰 디스플레이의 절반 크기밖에 되지 않는 작은 E-ink 디스플레이와 카메라, 스크롤키와 PTT(Push to Talk) 버튼 등의 최소화한 물리적 장치만을 가지고 있다. 이것이 가능한 이유는, 이 기기가 스마트폰의 GUI(Graphical User Interface) 기반의 클릭으로 구동되는 방식을 과감히 버리고, 음성 소통 기반의 대형 액션 모델(LAM, Large Action Model)을 통해 원하는 작업들을 자동 수행하게 하는 작동방식과 인터페이스를 채택하였기 때문이다.



Rabbit의 R1, 생성형 AI 기반 새로운 인터페이스 적용
출처: Rabbit

이 기기를 활용한다면 사용자는 택시 호출, 음식 주문, 항공편 및 숙박 예약 등의 작업을 수행하기 위해 스마트폰 화면에서 앱을 찾고, 실행하고, 로그인하고, 검색하고, 여러 단계에 걸쳐 실행 버튼을 누르는 일련의 중간 과정들을 일일이 수행하지 않아도 된다. 간단한 음성 지시 몇 마디만으로 원하는 작업의 결과물을 얻을 수 있다. R1의 LAM(앱을 동작시키는 액션을 출력하는 알고리즘)이 내 의도를 실행 시켜주는 손이 되어주는 것이다. 예컨대 회사에서 야근 후 R1에게 “피곤하니 택시타고 집에 가자”라고 말 한마디만 하면, R1은 그동안 학습된 개인 맥락 데이터를 기반으로 나의 취향 및 성향에 맞게 내가 평소에 자주 쓰는 택시 플랫폼에서, 가장 저렴한 요금의 택시를, 나의 집 주소로, 내가 평상시 탑승을 선호하는 회사 앞의 구체적인 장소로 실제 불러준다. 뿐만 아니라, 결제 시에는 회사에서 야근을 했으니 자율적인 판단에 따라 회사 법인 카드로 결제해줄 수 있다. 디바이스가 나의 손이 되어 귀찮고 번거로운 중간 과정의 일들을 대신 처리해 주고 실제 액션으로까지 연결시켜 주는 것이다.

R1의 LAM 기술은 아직 초기 단계에 불과하다. 따라서, 음성 지시를 통해 일련의 맞춤형 행동들이 자동 실행되게 하기 위해서는 아직까지 사용자가 ‘The Rabbit Hole’이라고 불리는 웹 포털을 통해 사전에 필요한 계정 정보 등을 미리 입력하고, ‘Teach Mode’를 통해 사

용자의 선호와 행동 패턴을 학습시키는 등의 사전 준비 과정이 필요하다. 그렇지만, R1의 LAM과 같이 사용자의 손이 되어 주는 새로운 구동 방식은 향후 지속적인 발전을 이룰 것으로 전망된다.

또한 Rabbit의 R1은 무조건 음성 기반의 인터페이스만을 고집하지 않는다는 특징이 있다. 필요시에는 적절하게 E-ink 디스플레이를 통해서, 또 AI의 눈이 되어주는 카메라를 통해 사용자와 소통한다. 앞서 언급한 택시 호출과 같은 상황에서도 R1은 디스플레이를 통해 호출 상세 내용을 요약하여 보여줌으로써 AI가 나의 의도에 맞춰 올바르게 실행했는지 여부를 재확인할 수 있게 해준다. 자연어 대화 기반으로 문서 처리 및 이메일/SNS 소통을 대신해 줄 때도 작업의 효율을 높이기 위해 텍스트를 화면에 보여준다. 또, 카메라를 통해 사용자와 함께 방 안에서 잃어버린 물건을 찾는거나, 냉장고 안의 재료를 통해 만들 수 있는 요리를 추천해주는 등 오프라인의 물리적 세계와 연계된 다양한 작업을 가능하게 한다.

두 제품 모두가 현재 스마트폰의 대화면 디스플레이의 비중을 줄이거나 없애는 등의 방식을 채택한 것은 의미심장하다. 생성형 AI를 통해 가능해진 자연어 대화 기반의 상호작용 방식이 적어도 특정 기능을 수행하는데 있어서는 현재 스마트폰의 디스플레이 및 앱 UI 기반 방식보다 생산성에 유리할 수 있다는 점을 보여주기 때문이다. 또한 두 기기들은 필요한 기능들이 모두 클라우드 기반으로 구동되어 디바이스를 최대한 가볍게 만드는데 집중하여 스마트폰 대비 가격 경쟁력을 갖추고 있다. 이에 따라 Rabbit의 R1은 199 달러에, AI Pin의 경우에는 보급형 스마트폰 가격 수준인 699 달러에 판매가를 책정할 수 있었다.

생성형 AI 품은 퍼스널 디바이스, 새로운 게임 룰 제시 가능

사람이 디바이스를 학습하지 않고 디바이스가 사람을 학습하는 시대, 디바이스가 단순히 머리만 좋은 것이 아니라 눈과 귀, 그리고 손이 있어 스스로 맥락을 파악하고 행동이 필요한 부분은 직접 행동하는 시대가 온다는 것은 많은 변화 가능성을 내포한다.

첫째, 다양한 형태의 디바이스 출현이 예상된다. 자연스럽고 직관적인 자연어 기반 커뮤니케이션에서는 특정 작업을 수행하기 위한 중간 과정이 생략되어 고객이 앱을 알 필요도 없기 때문에, 지금의 대화면 터치 디스플레이 중심의 스마트폰 형태를 고집할 필요가 없게 된다. 만약 이런 미래가 펼쳐진다면, 미래의 퍼스널 디바이스는 스마트폰과 같은 단 하나의 표준 폼팩터나 디자인을 채택하는 것이 아니라, 사람들이 자주 활용하는 작업에 따라

다양한 형태로 구성될 수 있다. 클라우드 기반으로 하나의 머리를 쓰면서 여러 개의 손을 가지고 있는 것처럼 말이다. 미래 퍼스널 디바이스의 경쟁에 있어 성능적인 측면보다는 얼마나 사용하기에 편리한 인체공학적(ergonomic) 디자인을 가졌는가 등의 하드웨어적인 측면이 더욱 부각될 수 있다.

둘째, 퍼스널 디바이스 시장의 밸류체인에서 지각변동이 일어날 수 있다. 새로운 퍼스널 디바이스의 한 갈래에서는 비용의 많은 부분을 차지하는 디스플레이의 역할이 축소되어 하드웨어 가격이 매우 저렴해질 수 있다. 다른 한 갈래에서는 마이크나 카메라와 같은 센서 데이터의 중요성이 크게 부각되어 새로운 방식의 저전력 고감도 always-on 센서 기술이 중요해질 수 있다. 또한, LAM과 같이 ‘손’에 해당하는 액션으로 연결시킬 수 있는 기술, 예컨대 다양한 앱 및 응용 프로그램과 상호작용하는 기술이나 개인 정보가 노출될 가능성이 없으면서도 개인 정보의 활용도는 극대화할 수 있는 보안 기술 등이 중요해질 수 있다. 더 나아가, 물리적인 ‘손과 발’이 실제 AI에게 부여되는 형태인 휴머노이드, 스마트홈 AI 에이전트 로봇 등이 활성화한다면 이와 관련한 Manipulator(인간의 팔과 유사한 동작을 제공하는 기계 장치) 설계, 제어, 핸들링 기술을 가지고 있는 HW 기반 전문 솔루션 업체들이 디바이스 시장에서 부상할 수 있다.

이처럼 생성형 AI를 품은 퍼스널 디바이스는 다양한 갈래로 나비효과를 불러 일으킬 수 있는 파괴력을 지니고 있기 때문에 이를 둘러싼 변화의 움직임에 면밀히 살펴보아야 한다. 남보다 한 걸음 앞서 미래 변화 방향을 제대로 포착하고, 반 걸음 앞서 대응하기 위해서는 무엇보다 기술적 인사이트와 시장 변화에 대한 깊은 이해를 바탕으로 새로운 사용자 인터페이스와 상호작용 방식을 지속적으로 탐구해야 할 것이다. LG경영연구원