

ISSUE REPORT

2024-06호

온디바이스 AI 기술동향 및 발전방향



KEA | Korea
Electronics
Association

한국전자정보통신산업진흥회

ISSUE REPORT

2024-06호

DIGISIGHT

온디바이스 AI 기술동향 및 발전방향

I. 개요	06
II. 시장 및 기업 동향	09
III. 핵심기술 동향	12
IV. 미래 발전방향	20

KEA NOW

27

MEMBER NEWS

30

ESG TREND

33

STATS

35



온디바이스 AI 기술동향 및 발전방향

I. 개요

II. 시장 및 기업 동향

III. 핵심기술 동향

IV. 미래 발전방향

참고문헌

■ (정의) 온디바이스 AI(On-Device AI)란 데이터를 외부 서버나 클라우드에 전송하지 않고 디바이스 자체적으로 AI연산을 수행하는 기술

- ◎ 대용량 데이터 처리 제한, 개인정보 유출 위험, 실시간성 저하 등 기존 서버 기반 중앙집중형 구조의 여러 제약사항을 해결하기 위해 등장
- ◎ 높은 컴퓨팅 파워가 필요한 AI 연산을 다양한 하드웨어 및 소프트웨어 기술을 활용하여 최적화 된 스마트폰, 로봇, 드론 등 디바이스 내에서 AI 연산 수행
- ◎ 디바이스 내 자체 AI 연산을 통해 네트워크 환경에 독립적인 실시간 서비스, 개인정보 유출 최소화, 서버 운영비용 절감 등 강점
- ◎ NPU, AI 모델 최적화 등 고수준의 하드웨어 및 소프트웨어 기술 동시 요구로 인해 높은 진입장벽 존재

■ (시장동향) 글로벌 기업 및 각국 정부 투자 가속화로 고성장 예상

- ◎ 생성형 AI 및 온디바이스 AI 시장은 전 세계적으로 빅테크 기업 투자 열풍으로 급성장이 예상되며, 각국 정부들도 투자계획 수립을 서두르고 있어 향후 시장 확대가 가속화될 전망
- ◎ 미국을 포함한 유럽, 일본, 중국, 대만, 한국 등 주요 국가에서 디바이스 중심의 AI 관련 기술 개발 지원 정책을 지속적으로 발표
- ◎ AI칩 제조사들은 자사 칩 기반 생태계 구축을 위해 데이터·AI모델·추론·SDK 등 전방위적 기술 지원
 - 단순 하드웨어 드라이버 수준의 라이브러리를 넘어 AI 모델 경량화 및 최적화 도구 등 생태계 구축을 위해 다양한 소프트웨어 지원

**(기술동향) 온디바이스 AI 하드웨어부터 가속 추론 소프트웨어, 생성형AI 최적화까지
폭넓은 분야에 걸쳐 기술 개발 진행**

하드웨어

- 엔비디아는 GPU의 효율성을 지속적으로 향상시키고 있으며, 구글, 테슬라, 애플 등 디바이스 제조사에서는 NPU를 개발하여 자사 서비스에 적용
- 웰컴, AMD 등 글로벌 칩 제조사에서는 고효율의 범용 NPU를 지속적으로 개발하고 있으며, 최근에는 생성형 AI 지원 NPU를 경쟁적으로 출시
- 국내 대기업에서는 NPU를 상용화하여 자사 제품에 사용 중이며, NPU 관련 스타트업에서는 검증 단계를 넘어 양산 단계로 진입

소프트웨어

- 범용적인 환경에서 동작이 필요한 온디바이스 AI 연산 처리 소프트웨어 개발은 기술적 난이도가 높고 인프라적 성격이 강해 글로벌 대기업이 생태계 구축의 일환으로 기술 선도
- 국내에서는 대기업 중심으로 자사 제품에 활용 가능한 최적 AI 추론 기술 개발

생성형AI

- 미국 OpenAI가 전 세계 생성형 AI 기술을 주도하고 있는 한편 메타, 애플, 테슬라 등 여러 기업들이 온디바이스에서 생성형 AI 지원을 위해 기술 개발 추진
- 국내에서는 대기업에서 한국어 특화 생성형 AI 개발 및 경량화를 진행 중이며, 자사 제품이 적용을 목표로 기술개발 진행

(발전방향) 기술 선도 및 시장 선점을 위한 킬러 서비스 경쟁 심화 전망 및 국내 중소 기업 관련 지원 확대 정책 필요

- ◎ 온디바이스 AI가 향후 AI 시장의 게임 체인저 역할을 할 것으로 예상되며, 기술 확보를 위한 M&A, 인력 확보, 킬러서비스 확보 등 경쟁 심화 전망
 - 온디바이스 AI 기술을 활용해 사용자에게 혁신적인 경험을 제공하기 위한 킬러 서비스에 대한 경쟁 심화 예상되며 궁극적으로 개인화된 맞춤형 AI 서비스로의 발전 기대
- ◎ 기술 선도 및 시장 선점을 위해 생태계 확장이 필요하며 이를 위해 중소기업의 온디바이스 생성형 AI 시장 진입 장벽 해소를 위한 정책 필요
 - 온디바이스 AI 제품 및 서비스 개발을 위한 연구, 핵심 인력 양성·재직자 역량 강화 교육, 서비스 사례 확보를 위한 실증 지원 필요

I. 개요

1 개념정의

온디바이스 AI(On-Device AI)란 데이터를 외부 서버나 클라우드에 전송하지 않고 디바이스 자체적으로 AI 연산을 수행하는 기술을 의미

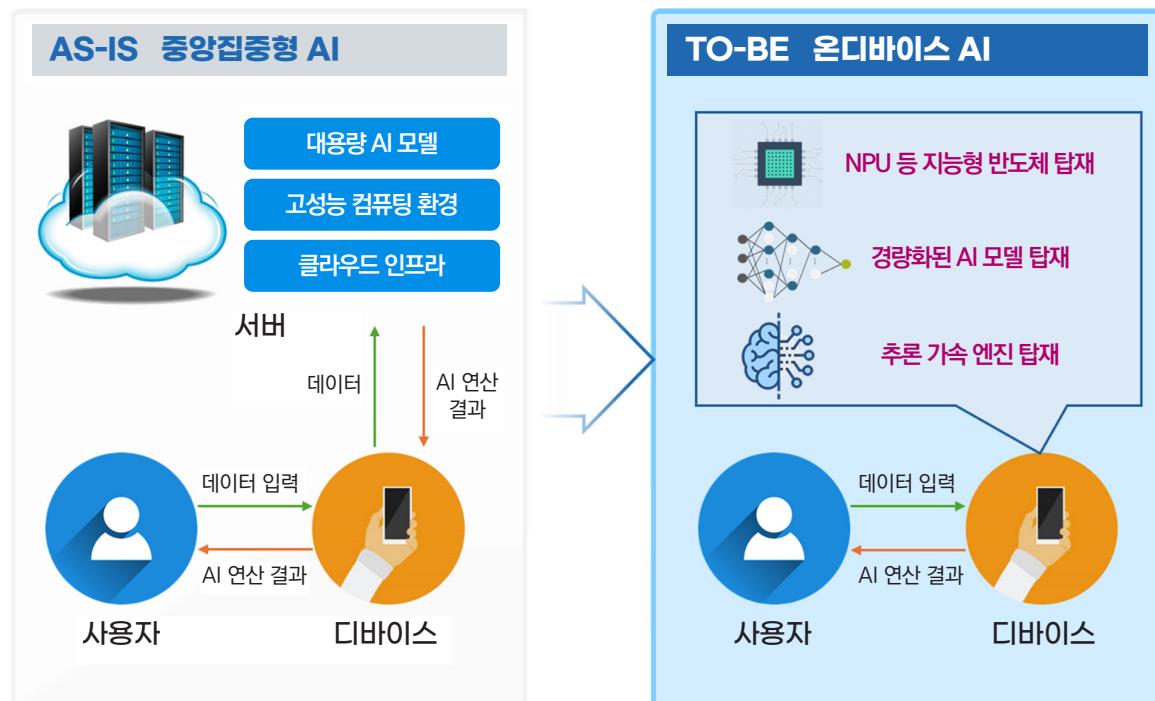
◎ 높은 컴퓨팅 파워가 필요한 AI 연산을 다양한 하드웨어 및 소프트웨어 기술을 활용하여 최적화된 스마트폰, 로봇, 드론 등 디바이스 내에서 AI 연산 수행

- 저전력으로 AI 연산 수행을 위해 최적화된 지능형 반도체 NPU(Neural Unit)*와 AI 모델 경량화를 위한 양자화, AI 모델 추론 가속을 위한 TensorFlow Lite 등 다양한 기술 활용

* 기존 AI 네트워크 연산에 사용되던 GPU와 유사한 구조나 AI 기술에 특화해 연산 혹은 에너지 효율을 높여 개발한 칩

- 일반적으로 온디바이스 AI는 추론에 집중되어 있으며 연산량이 훨씬 높은 모델 학습은 서버에서 진행

온디바이스 AI 기술 정의



2 등장배경

◎ 딥러닝 기반 AI 연산은 태생적으로 높은 연산량이 필요

- 딥러닝 모델은 매우 많은 양의 학습 데이터를 학습하여 높은 정확도의 예측을 수행하기 위해 수많은 노드와 깊은 레이어로 구성된 거대한 네트워크 구조
- 학습 뿐 아니라 추론 과정에도 높은 연산량이 요구됨에 따라 고성능 서버 중심의 중앙집중형 AI 서비스 구조가 보편화

◎ 디바이스 및 AI 기술 수준의 발달로 인해 기존 서버 중심의 중앙집중형 구조에서 다양한 제약사항 발생

- 디바이스의 역할이 단순한 센싱, 통신 등에 국한된 환경에서는 디바이스가 생산하는 데이터의 크기가 제한적이어서 일반적으로 서버로 모든 데이터를 송신한 다음, 서버에서 AI 연산을 수행하는 구조 가능
- 디바이스 및 AI 기술 발달로 인해 대용량 데이터 처리, 개인정보 유출 방지, 실시간 서비스 등의 요구사항이 발생하면서 기존 서버 기반 중앙집중형 구조에서 여러 제약사항 발생
 - **(고해상도 데이터 처리 제한)** 고성능 디바이스 수가 증가함에 따라 대용량 데이터 처리 시 네트워크와 클라우드 등에 부하가 증가하여 서비스 품질 저하 및 통신·서버 등 서비스 초기 구축 및 유지관리 비용 증가
 - **(개인정보 유출 우려)** 영상, 음성, 텍스트 등 민감 정보가 서버로 전송되므로 개인정보 남·오용 우려
 - **(실시간성 저하)** 데이터 전송 및 처리 결과 수신으로 인한 지연이 발생해 실시간 서비스가 힘들며 네트워크 연결 환경 강제

온디바이스 AI 기술의 필요성

디바이스 및 AI 기술 발달로 인한 환경 변화

고해상도 대용량 데이터 처리 필요	개인정보 유출 우려	실시간 서비스 요구
<ul style="list-style-type: none"> • 4K 카메라, LiDAR 등 고해상도 센서를 탑재한 고성능 디바이스 보급 • 고성능 디바이스가 대규모로 운용되어 통신망 및 클라우드 과부화 문제가 발생하는 인프라 환경 등장 (스마트시티 등) 	<ul style="list-style-type: none"> • CCTV, 전자도어락 등 민감정보가 많이 존재하는 스마트홈 등 개인 프라이버시 환경 • 개인정보 유출 방지를 위해 클라우드 개입을 최소화하는 디바이스 내 자체 서비스 수행 필요 	<ul style="list-style-type: none"> • AI 서비스가 보편화됨에 따라 사용자에게 보다 즉각적인 서비스 결과 제공 필요 • 스마트폰, 드론, 로봇, 자율차 등의 디바이스 환경에서 실시간 영상처리, 음성인식, 자연어처리 등의 서비스 요구사항 증대

“디바이스 내 자체적인 AI 연산 수행 필요”

◎ 이를 해결하기 위한 대안으로 온디바이스 AI 기술의 등장

- 이전부터 “엣지 AI”라는 용어로 디바이스 뿐만 아니라 분산 컴퓨팅 환경의 엣지 서버 엣지 게이트웨이 등 컴퓨팅 리소스가 제한된 환경에서 AI 연산 수행에 대한 연구 진행

- AI 연산에 최적화된 NPU의 등장으로 스마트폰, 로봇, 드론 등 배터리 기반의 이동형 기기에서도 자체적인 AI 연산 수행을 통한 서비스 제공이 실현되면서 디바이스 환경에서의 AI 연산에 보다 집중한 온디바이스 AI 기술 등장
- '23년 스마트폰에서 생성형 AI 서비스가 지원되며 온디바이스 AI 기술의 시장 확대 가속화

3 주요특징

■ 디바이스 내 자체 AI 연산 가능

◎ 네트워크 환경에 독립적인 고신뢰성 실시간 서비스 제공

- 외부와의 통신 없이 AI 연산 수행을 통해 통신 요청 및 응답으로 인한 지연 없이 실시간 서비스 제공
- 네트워크 환경에 따라 서비스의 품질이 좌우되지 않는 안정적인 서비스 제공

◎ 개인정보 유출 위험 최소화를 통한 사용자 만족도 및 서비스 비용 절감

- 사용자의 영상, 음성 등 민감한 개인정보가 서버로의 전송 없이 디바이스에서만 처리하여 개인정보 유출 위험이 낮아지며 이에 따라 사용자의 불안감 최소화
- 디바이스와 서버 간 개인정보 통신이 없어 데이터 전송에 대한 암호화, 보안 등이 불필요하여 AI 서비스 구축 및 유지관리를 위한 비용 절감

◎ 서버 의존성 탈피로 인한 경제적인 AI 서비스 제공

- 수많은 사용자의 AI 연산 수행을 서버에서 감당하기 위해 높은 비용의 클라우드 컴퓨팅 시스템을 운영하는 대신 디바이스에서 AI 연산을 각자 수행하여 비용 절감 가능

■ 고수준의 하드웨어 및 소프트웨어 기술

◎ NPU 기반 하드웨어 구성 및 활용을 위한 다양한 애로사항 존재

- 디바이스 AI를 위한 하드웨어 개발을 위해 저전력·저비용의 NPU 기반 디바이스 구성이 필요하나, NPU 시장의 성숙도가 높지 않아 중소기업이 NPU 기반 디바이스를 상용화 수준으로 개발하는데 많은 난관 존재
- 일반적으로 AI 서비스를 위해 사용하는 딥러닝 모델은 GPU 기반으로 개발되어 있어 이를 NPU에서 활용하기 위해서는 모델과 NPU의 특성에 대한 이해를 기반으로 하는 고난이도 이식 작업 필요

◎ AI 모델 경량화를 통해 정확도와 효율성 간 최적화 필요

- 높은 정확도의 딥러닝 모델을 디바이스 자원만으로 동작하기 위해서는 모델의 입력 데이터 크기를 줄이거나 모델 자체의 크기를 줄이는 등 경량화가 필요하나, 이는 정확도의 저하를 동반
- 최소한의 정확도 손실로 AI 모델을 최대한 경량화하기 위해 고수준의 최적화 필요

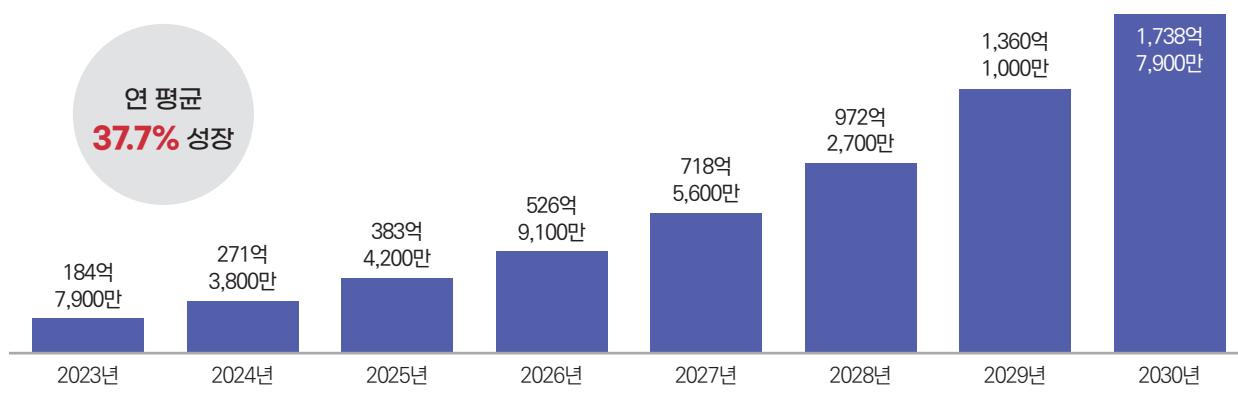
II. 시장 및 기업 동향

1 글로벌 시장

◎ 생성형 AI 및 온디바이스 AI 시장은 전 세계적으로 빅테크 기업 투자 열풍으로 급성장이 예상, 각국 정부들도 투자계획 수립을 서두르고 있어 향후 시장 확대가 가속화될 것으로 예상

- 디바이스 AI 시장은 스마트폰, 자동차, 드론 등 다양한 기기에 탑재되어 신산업을 이끄는 차세대 AI 시장으로 주목되어 '24년 271억 달러 규모에서 '30년 1,738억 달러로 연평균 37.7% 성장 전망

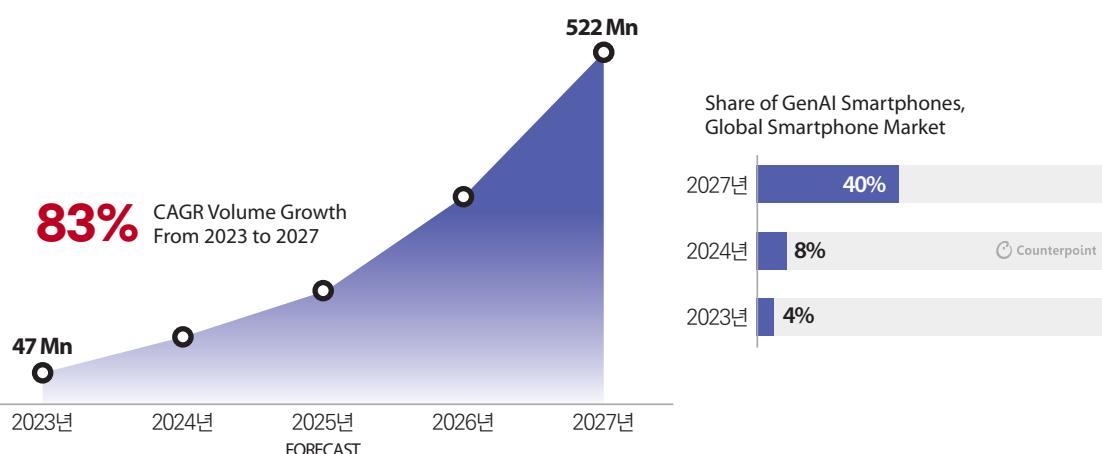
온디바이스 AI 시장 규모(달러)



* 출처 : 마켓츠앤마켓츠('23)

- 온디바이스 AI 시장을 열고 있는 생성형 AI 스마트폰 시장 규모는 '27년 5억 2,200만대로 추산되며, '24년부터 '27년 까지 연평균 성장률은 83%에 이르는 고성장 전망

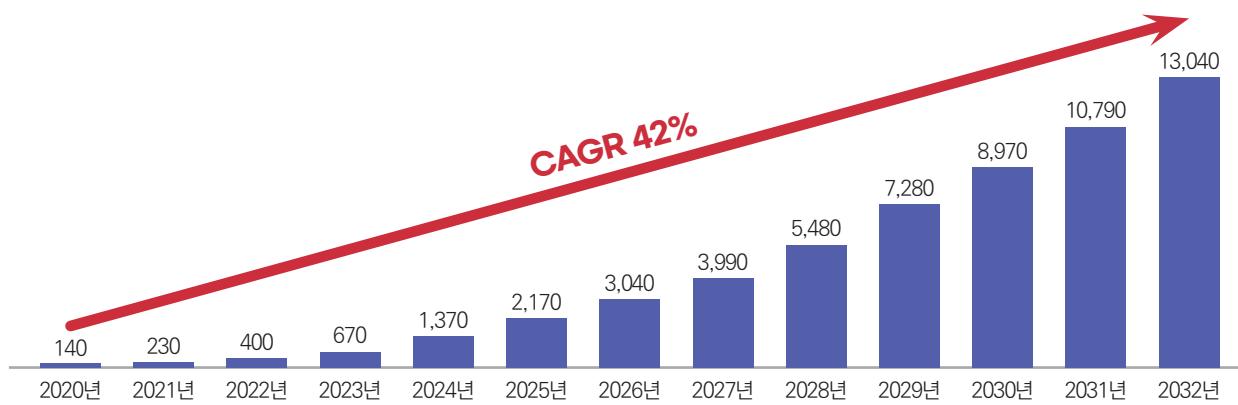
생성형 AI 스마트폰 시장 규모(백만대, %)



* 출처 : 카운터포인트리서치('23)

- 글로벌 생성형 AI 시장 규모는 '22~'32년에 연평균 42% 고성장, '32년의 시장 규모는 1조 3,040억 달러로 '22년 대비 30배 이상 성장 전망

생성형 AI 시장 규모(억 달러)



* 출처 : 블룸버그인텔리전스('23)

2 국가별 정책

◎ 미국을 포함한 주요 국가에서 디바이스 중심의 AI 관련 기술 개발 지원 정책을 지속적으로 발표

국가	정책 동향
미국 	<ul style="list-style-type: none"> 미국 과학기술정책국(OSTP)의 「국가 AI R&D 전략 계획」에 대한 '23년 업데이트 발표에서 장기적으로 투자할 차세대 AI 분야에 온디바이스 시스템과 같은 제한된 하드웨어 및 에너지 리소스를 고려하는 연구분야 포함('23) * 하드웨어 성능 개선 및 리소스 사용 최적화를 위한 AI 시스템 개발과 에너지 소비를 비롯한 지속 가능성을 고려하는 리소스 중심 AI 알고리즘 및 시스템 고려
유럽연합 	<ul style="list-style-type: none"> 독일은 「연방정부의 인공지능 전략 2020 업데이트」에서 연구지원 분야로 온디바이스 AI 기술의 활용도가 높은 헬스케어, 모빌리티, 에너지, 농업 분야에 대한 다양한 지원 정책 발표('20) 영국은 「인공지능 10개년 국가전략 계획」을 통해 영국AI 시장 트렌드를 선도하는 헬스케어 분야 온디바이스 AI 기술 개발 지원과 글로벌화 계획 발표('21)
일본 	<ul style="list-style-type: none"> 경제산업성은 「반도체·디지털 산업 전략」에서 개인별 최적화된 IT 디바이스 시스템 기반 디지털 기술 혁신을 통해 경제성장을 실현하기 위한 전략 발표('23.5) 「AI 전략 2021」을 발표하여 다양한 온디바이스 시스템이 적용될 건강 및 의료, 농업, 스마트시티, 제조업 등 다양한 분야에서 인공지능을 도입하고 상용화 가속 대응 방안 개선 추진 시작('21.6)
한국 	<ul style="list-style-type: none"> 국산 AI 반도체를 기반으로 초기시장 단계인 온디바이스 AI 시장 선점을 위한 「(가칭)온디바이스 AI 활성화 전략(안)」 계획 발표 (과기정통부, '24.2) 중소벤처기업부에서 LG전자 등과 손잡고 온디바이스 AI 스타트업 육성을 위해 「온디바이스(On-Device) AI 초격차 챌린지」 출범식 개최('24.3) NPU 고도화, PIM·뉴로모픽 혁신, 신소자 및 첨단패키징 등에 기반한 저전력 AI반도체로 클라우드 AI데이터센터 고도화, 온디바이스 AI 신격차 확보 전략을 포함하는 「AI·디지털 혁신성장 전략」 발표(과기정통부, '24.4)

국가	정책 동향
중국 	• 「중공중앙 국민경제사회발전 제14차 5개년 규획(2021~2025)」과 '35년 장기목표에 대한 초안'에서 차세대 IT 산업의 세부산업 육성 전략에 AI 분야에서 스마트 의료장비, 스마트 인식 시스템 등 온디바이스 AI 기술이 필요한 산업 분야를 선정 (21.3)
대만 	• 국가과학기술위원회(NSTC)는 자국에 특화된 AI 언어모델 TAIIDE 개발 프로젝트에 온디바이스 환경에서 동작 가능한 작은 모델 개발 계획이 포함됨을 발표 (24.1)

3 국가별 정책

◎ 자사 AI 칩 기반의 생태계 구축을 위해 데이터·AI 모델·추론기술·SDK*를 아우르는 전방위적 기술 지원

* 소프트웨어를 개발하는 도구로 소프트웨어 개발자가 특정 운영체제용 응용프로그램을 만들 수 있는 소스와 도구 패키지

◎ (애플) 자사의 하드웨어 환경에서 애플의 뉴럴 엔진을 활용하여 온디바이스 AI를 수행할 수 있도록 Core ML 라이브러리 제공

- 이 라이브러리는 이미지, 비디오, 사운드 등 미디어 분석을 위해 설계된 첨단 신경망과 같은 AI모델을 지원하며, TensorFlow나 PyTorch와 같은 라이브러리의 모델을 Core ML로 변환하는 기능도 제공

◎ (퀄컴) 자사 스냅드래곤 AP에서 효율적인 AI 연산 수행을 위한 라이브러리와 하드웨어·소프트웨어 환경에 맞게 사전 학습된 AI 모델을 제공하는 AI Hub 제공

- AI Hub는 개발자가 AI 연산을 수행할 환경을 선택하면 이에 따라 기 학습된 75여 가지의 인기 있는 AI 및 생성형 AI 모델을 제공하여 손쉬운 AI 서비스 구축 지원

◎ (AMD) 자사 칩셋에 최적화된 AI 연산 추론 라이브러리 ZenDNN과 AI 모델의 생성부터 최적화, 추론 가속까지 전과정을 지원하는 Vitis AI 플랫폼 제공

- Vitis AI 플랫폼은 AI 모델 경량화 뿐만 아니라 프로파일러를 통해 모델의 특성을 분석하고 최적화를 할 수 있는 도구를 제공하며, 기 학습된 AI 모델 또한 지원

◎ (인텔) 자사 칩셋에서 AI 연산을 최대한 효율적으로 수행하기 위해 OpenVINO 프로젝트를 출시했으며, 개발자를 위한 Edge AI Reference Kits를 오픈소스로 제공

- 기 학습된 모델부터 직접 모델을 개발할 수 있도록 최적화 도구 일체를 제공하며, 자사 칩셋 환경에서의 추론 가속 엔진 지원

◎ (엔비디아) AI 모델 경량화를 위한 Tensor RT 외 AI 연산을 위해 필요한 도구 일체 지원을 통해 성공적으로 생태계 구축

- '18년부터 엔비디아는 생태계 확장 및 기술개발 참여 유도를 위해 AI 추론 가속기 HW 및 SW 아키텍처를 NVDLA (NVIDIA Deep Learning Accelerator)라는 오픈소스로 공유

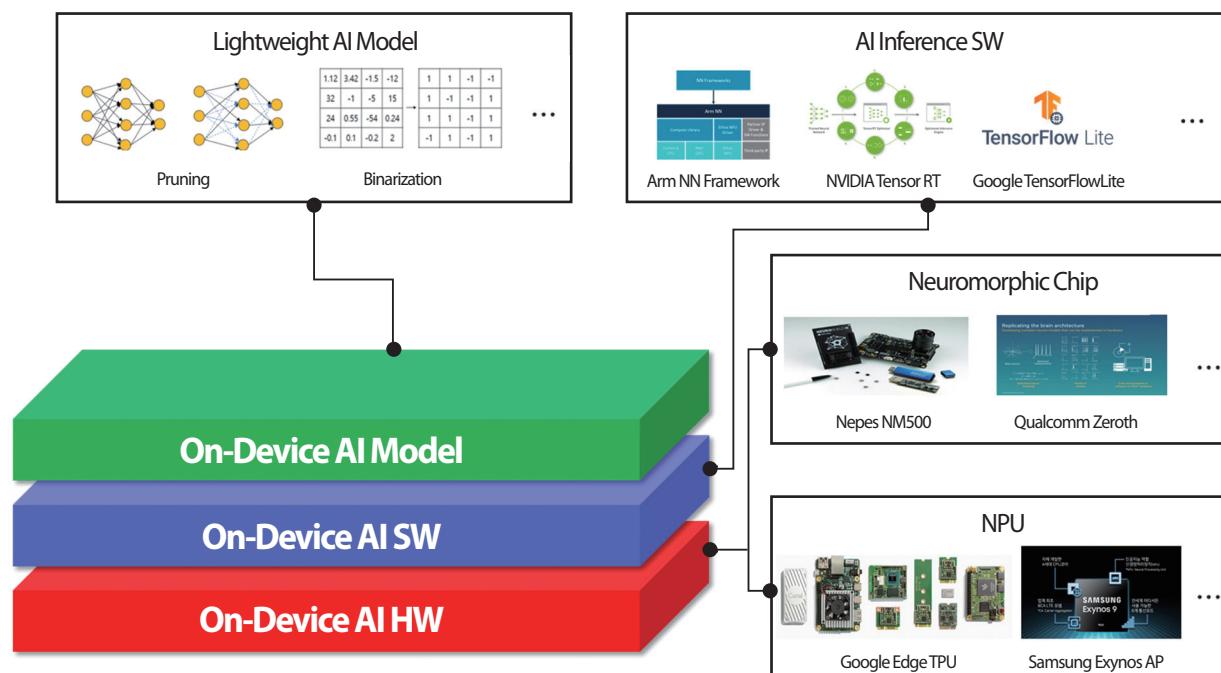
◎ 이 외 헤일로, 딥엑스, 모빌린트, 크네론 등 NPU 제조 스타트업도 단순히 하드웨어 드라이버 수준의 라이브러리를 넘어 AI 모델 경량화 및 최적화 도구 등 생태계 구축을 위해 다양한 소프트웨어 지원

III. 핵심 기술 동향

1 기술 개요

- ◎ 하드웨어부터 AI 모델까지 폭넓은 분야에 걸쳐 기술 개발 진행
- ◎ **(온디바이스 AI 하드웨어)** AI 연산에 최적화된 지능형 반도체인 NPU와 사람의 뇌 구조를 모사한 뉴로모픽 칩이 대표 기술
 - NPU는 GPU와 유사한 병렬연산에 최적화 된 반도체이나, 그래픽 등 일반적인 연산 타겟의 GPU와 달리 AI 연산만을 위해 설계되어 효율성은 높으나 GPU 대비 유연성은 부족
 - 뉴로모픽 칩은 행렬 연산 기반의 딥러닝 모델과는 달리 인간의 뇌와 유사한 스파이킹 뉴런 모델과 시냅스 모델로 동작하며, GPU 및 NPU보다 전력 소모량이 매우 적을 것으로 기대되나 아직 상용화 수준에는 미도달
- ◎ **(온디바이스 AI 소프트웨어)** 제한된 리소스 내 AI 모델을 최대한 효율적으로 동작시키기 위해 하드웨어·운영체제 등 시스템 레벨의 AI 연산 엔진
 - 대표적으로 구글의 TensorFlow Lite가 있으며, 기술 난이도가 높아 글로벌 대기업 위주로 기술 개발
- ◎ **(온디바이스 AI 모델)** AI 모델의 필요 없는 파라미터를 제거하는 프루닝, 데이터 해상도를 낮추는 양자화 등 다양한 기술을 통해 온디바이스 환경에서 동작 가능한 수준의 경량화된 딥러닝 모델

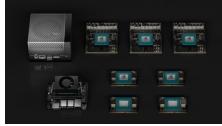
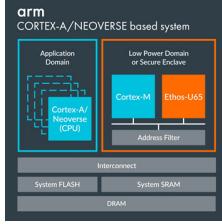
온디바이스 AI 기술 분야

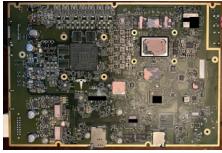


2 하드웨어 기술

해외

- 엔비디아는 GPU의 효율성을 지속적으로 향상시키고 있으며, 구글, 테슬라, 애플 등 디바이스 제조사에서는 NPU를 개발하여 자사 서비스에 적용
- 퀄컴, AMD 등 칩 제조사에서는 고효율의 범용 NPU를 지속적으로 개발하고 있으며, 최근에는 생성형 AI 지원 NPU를 경쟁적으로 출시

기업명	기술 동향	
엔비디아 	<ul style="list-style-type: none"> 임베디드 환경에서의 온디바이스 AI를 위한 Jetson Orin 시리즈는 다성능의 디바이스를 지원하기 위해 20 TOPS 성능의 Nano부터 200 TOPS 성능의 AGX까지 3단계 성능의 솔루션 출시 	
애플 	<ul style="list-style-type: none"> 얼굴인식, 음성인식으로 대표되는 자사의 AI 서비스 품질 향상을 위해 뉴럴엔진이라는 NPU를 독자적으로 개발하여 AP(Application Processor)와 통합된 형태로 제품을 출시해 왔으며, '23년 A17 Pro에 탑재된 7세대 뉴럴엔진에서는 전작 대비 약 2배 성능을 향상시켜 35 TOPS 달성 	
인텔 	<ul style="list-style-type: none"> 서버 환경에서의 AI 가속을 위한 체온 프로세스 등 고성능 CPU 기술 개발 경험을 토대로 신경망 프로세서 및 비전 처리 장치를 개발함으로써 점차적으로 온디바이스 AI 수행을 위한 하드웨어 기술로 확장 디바이스 환경에서 영상처리 관련 AI 수행에 최적화된 Movidius Myriad X VPU(Vision Processing Unit)를 USB 기반 모듈 형태로 만든 뉴럴컴퓨트스틱2 출시, '23년에는 3세대 3700VC VPU를 출시했으며, VPU가 내장된 14세대 CPU 메테오레이크 공개 	
구글 	<ul style="list-style-type: none"> 저전력 프로세서 기반 AI 연산을 위한 Edge TPU (Tensor Processing Unit)를 출시했으며, USB 및 PCIe/M.2 형태의 HW를 제공하여 라즈베리파이, 아두이노와 같은 IoT 기기에서의 AI 추론을 간편한 형태로 지원 	
퀄컴 	<ul style="list-style-type: none"> '18년 스냅드래곤855부터 전용 텐서가속기가 hexagon DSP에 탑재, 최근 hexagon DSP 대신 hexagon NPU라고 명칭을 변경하여 AI 가속에 집중 	
ARM 	<ul style="list-style-type: none"> 온디바이스 AI를 위한 microNPU제품군 개발, 클라우드부터 엣지에서 사용 가능한 고 사양의 Ethos-N 시리즈와 디바이스에서 사용 가능한 저전력의 Ethos-U 시리즈를 공개했으며 '24년 Ethos-U85 출시 	

기업명	기술 동향	
AMD 	<ul style="list-style-type: none"> 로봇, 드론 등 임베디드 환경에서의 AI 연산 수행을 위해 NPU와 DSP를 포함하는 Versal AI Edge 시리즈 출시, 5 TOPS부터 202 TOPS까지 7단계의 선택지 제공 	
헤일로 	<ul style="list-style-type: none"> '20년 26TOPS 성능의 Hailo-8을, '23년에는 외부 메모리가 필요 없어 구성이 손쉬우며 저전력으로 동작하는 13TOPS 성능의 Hailo-8L을 발표, M.2, mPCIe 등 다양한 인터페이스의 엣지형 NPU 모듈 판매 '24년 생성형 AI를 지원하는 Hailo-10 출시 	
테슬라 	<ul style="list-style-type: none"> 자사의 핵심 기능인 자율 주행을 위해 NPU를 탑재한 자체 칩인 FSD 칩(Full Self-Driving) 개발 중, '22년 향상된 성능의 HW4.0 출시 	
미디어텍 	<ul style="list-style-type: none"> 모바일 AP 타겟의 NPU인 APU (AI Processing Unit)을 지속적으로 개발: '24년 생성형 AI 타겟의 APU790을 탑재한 모바일 AP 디멘시티 9300+ 공개 	
키나라 	<ul style="list-style-type: none"> 저전력의 Ara-1과 생성형 AI를 지원하는 고성능의 Ara-2 NPU 출시, USB, M.2, PCIE 3 가지 폼팩터의 모듈 판매 레노버와 협력해 소형 컨텐츠 제작용 데스크톱 씽크센터에 NPU 탑재 	
NXP 	<ul style="list-style-type: none"> 자사의 다양한 제품에 탑재하여 임베디드 환경 AI 연산 수행을 위한 eIQ Neutron NPU 출시, 용도에 따라 다양한 성능으로 구성 가능 MCX N MCU, i.MX 95 AP 등이 NPU를 탑재한 대표적인 제품 	

국내

◎ 대기업에서는 NPU를 상용화하여 자사 제품에 사용 중이며, NPU 관련 스타트업에서는 검증 단계를 넘어 양산 단계로 진입

기업명	기술 동향	
삼성전자	<ul style="list-style-type: none"> NPU 시장선점을 위해 지속적인 투자를 통해 리딩, '18년 11월 공개한 엑시노스9820에 처음으로 NPU 탑재를 시작으로 최신 20시리즈 (엑시노스2100~2400)까지 모두 NPU 탑재 최근에는 차량용 프로세서인 엑시노스오토 V920을 발표하여 현대자동차와 협력해 '25년 실적용을 목표로 개발 	
LG전자	<ul style="list-style-type: none"> 가전 등에서 온디바이스AI를 수행하도록 비전 관련 작업 가속기 및 음성 관련 작업 가속기를 탑재해 가전에서 필요한 영상 및 음성 AI 분석을 지원하는 LG8111 AI SoC 솔루션 개발, '23년에는 신규 NPU 개발 칙수를 공개해 개발 완료 후 스마트 TV의 알파10 칩에 적용 예정 	

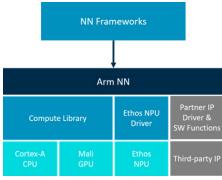
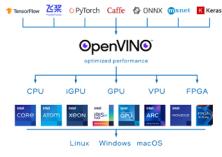
기업명	기술 동향	
사피온	<ul style="list-style-type: none"> SKT에서 분사하여 데이터센터 및 디바이스 대상 NPU를 개발 중, '20년에 데이터센터용 NPU X220, '23년 X330 출시 '24년 디바이스 및 자율주행차 대상 X300 Edge, X300 Auto 등 출시 예정 	
딥엑스	<ul style="list-style-type: none"> 엣지향 및 서버향 NPU를 개발하는 팹리스 기업으로 세계 최고 수준의 다양한 신규 알고리즘들을 높은 전성비/정확도를 기반으로 동작하는 NPU 개발을 목표로 함. '23년 DX-L1/L2/M1/H1의 12 TOPS부터 1600 TOPS급의 4가지 NPU를 발표했으며, '24년 양산 예정 	
넥스트칩	<ul style="list-style-type: none"> 차량, 보행자, 차선 등을 인식 및 검출하는 프로세서로, 최첨단 운전자 지원 시스템(ADAS)향 스마트 센싱 카메라에 적합한 영상기반 이미지 처리 지원 비전프로세서 개발 	
모빌린트	<ul style="list-style-type: none"> 가격 측면에서 최소 4배 이상 엔비디아 제품보다 뛰어난 성능으로 구현한다는 목표를 세우고 ARIES NPU를 개발, 온디바이스 AI 중에서도 비교적 큰 기기, 산업용 로봇이나 서빙 로봇, 스마트 팩토리 등의 시장을 타깃으로 MLA100, MLX-A1 2가지 품팩터의 제품 출시 '24년 3월 디자인하우스 세미파이브와 협력하여 칩 양산에 돌입했다고 발표 	

3 소프트웨어 기술

해외

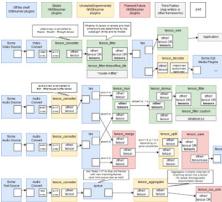
◎ 범용적인 환경에서 동작이 필요한 온디바이스 AI 연산 처리 소프트웨어 개발은 기술적 난이도가 높고 인프라적 성격이 강해 글로벌 대기업이 생태계 구축의 일환으로 기술 선도

기업명	기술 동향	
구글	<ul style="list-style-type: none"> 경량 딥러닝 네트워크 추론 엔진인 TensorFlow Lite를 개발해 스마트폰과 같은 엣지디바이스에서 리소스 사용을 최소화하면서 AI 추론 지원 최근 생성형 AI의 손쉬운 지원을 위해 MediaPipe, TensorFlow Lite, Gemini 나노를 포함하는 엔드 투 엔드 스택인 Google AI Edge 출시 수 KB 수준의 메모리를 가진 MCU 타겟의 TensorFlow for Microcontrollers를 실험 기능으로 제공 	
엔비디아	<ul style="list-style-type: none"> 엔비디아 TensorRT는 고성능 딥러닝 추론을 위한 SDK로 임베디드 환경에서 학습된 모델을 기반으로 실시간 추론 및 실시간 모델 업데이트 기능 제공 	

기업명	기술 동향	
ARM 	<ul style="list-style-type: none"> 자사의 하드웨어에 최적화하여 온디바이스환경 비전 및 AI 분석을 지원하는 라이브러리 ACL(Arm Compute Library) 제공 자사의 Cortex-A CPU, Mali GPU, Ethos NPU에 최적화된 머신러닝구동을 위해 NN Frameworks를 오픈소스로 제공 	
퀄컴 	<ul style="list-style-type: none"> 모바일에서 자주 사용되는 비전 분석 기능을 최적화한 FastCV라이브러리를 개발, 자사 스냅드래곤 AP에서 고효율 AI 모델 처리를 지원하는 SNPE (Snapdragon Neural Processing Engine)과 AI 모델 압축 및 양자화를 수행하여 온디바이스 AI를 지원하는 AIMET(AI Model Efficiency Toolkit) 개발 	
인텔 	<ul style="list-style-type: none"> 자사 칩셋 타겟의 OpenVINO 프로젝트를 통해 Xeon CPU부터 Arm 코어까지 다양한 디바이스 환경에서 AI 연산을 효과적으로 수행하기 위한 OpenVINO Runtime API 제공 최근 생성형 AI 지원을 위한 기능 추가 	

국내

◎ 대기업 중심으로 자사 제품에 활용 가능한 최적 AI 추론 기술 개발

기업명	기술 동향	
삼성전자	<ul style="list-style-type: none"> 뉴럴 네트워크 연산을 효율적으로 처리하기 위한 AI 추론 엔진인 ONE(On-device Neural Engine) 개발, 효율적인 추론을 위해 모델 컴파일을 위한 컴파일러와 모델 수행을 위한 런타임 파트로 나누어 개발 	
삼성리서치	<ul style="list-style-type: none"> 리눅스의 영상 파이프라인 프레임워크인 GStreamer를 기반으로 뉴럴네트워크 파이프라인 프레임워크인 NNStreamer를 오픈소스로 개발, 온디바이스학습을 위한 NNTrainer도 오픈소스로 공개 	
노타	<ul style="list-style-type: none"> AI 하드웨어의 특성에 따라 AI 모델을 경량화하는 넷츠프레소 솔루션 출시, 접근성 향상을 위해 웹 기반 넷츠프레소 사용 지원을 위한 LaunchX 웹 어플리케이션 서비스 출시 	
제텍에이아이	<ul style="list-style-type: none"> 기존 딥러닝 모델을 온디바이스 환경에서 동작할 수 있도록 경량화 도구 및 추론 가속 파이프라인을 포함한 제텍 멜란지 출시 	

4 생성형 AI 기술

해 외

- 미국 OpenAI가 전 세계 생성형 AI 기술을 주도하고 있는 한편 메타, 애플, 테슬라 등 여러 기업들이 온디바이스에서 생성형 AI 지원을 위해 기술 개발 추진

기업명	기술 동향
OpenAI 	<ul style="list-style-type: none"> 이미지를 인식할 수 있는 멀티모달 지원 'ChatGPT-4' 출시. 이를 기반으로 온라인 서비스인 플러그인 기능을 추가해 현재 오픈 AI 플러그인 스토어에 1,000여 개의 플러그인이 탑재되는 등 AI 생태계의 주력 트렌드로 자리 잡음. 최근 응답 속도를 비약적으로 향상시킨 'ChatGPT-4o' 출시
퀄컴 	<ul style="list-style-type: none"> 현 시점에서는 GPT2 기반 모델 크기 500M~1G급으로 약 20토큰 처리하는 데 10s 정도 수준의 기술 보유, 조만간 GPT 모델 경량화와 domain specific model finetuning 지원 온디바이스 GPT 지원 계획 발표. '23년 메타와 협력하여 메타의 생성형 AI인 거대언어모델(LLM) 라마2를 온디바이스에서 지원하기 위해 최적화를 수행하기로 발표한 후 당해 10월 생성형 AI를 반도체 칩에 내장한 '스냅드래곤 8' 3세대 발표 '24년 메타가 공개한 최신 LLM 라마3를 스냅드래곤 플래그십 플랫폼의 온디바이스 환경에서 동작시키기 위해 최적화 수행 계획 발표
메타 	<ul style="list-style-type: none"> 대규모 AI 언어 모델인 '라마(LLaMa)'를 공개하면서 최적화를 통해 특정 분야에 한정하여 작은 용량으로 고수준의 언어 처리를 지원하는 sLLM도 함께 공개, 이를 기반으로 노트북이나 휴대폰에서 동작하는 sLLM 응용도 공개
테슬라 	<ul style="list-style-type: none"> 자사에서 설립한 AI 스타트업 xAI가 ChatGPT와 같은 생성형 AI '그록(Grok)' 공개
구글 	<ul style="list-style-type: none"> 딥마인드와 협력하여 텍스트, 오디오, 이미지 등을 지원하는 멀티모달 생성형 AI 모델 '제미나이'를 개발했으며, 온디바이스 AI 지원을 위해 경량화된 '제미나이 나노'도 함께 개발 '24년 삼성전자와 협력하여 '제미나이 나노 2.0'을 갤럭시 S24에 탑재
애플 	<ul style="list-style-type: none"> '24년 iOS에서 개인 맞춤형 생성형 AI를 지원하는 애플 인텔리전스를 발표하면서 자체 개발 생성형 AI 모델인 '에이잭스(Ajax)'와 함께 오픈 AI와 협력하여 자사 음성 AI 비서 시리에 ChatGPT-4o를 접목한다고 발표 단, 모든 처리를 디바이스에서 처리하는 대신 상황에 따라 강화된 보안이 적용된 프라이빗 클라우드에서 AI 연산 수행 가능
MS 	<ul style="list-style-type: none"> 대부분의 MS 제품에 생성형 AI를 결합하여 'MS365 Copilot', 'Security Copilot' 등 신제품 출시 중, '24년 온디바이스 AI를 지원하는 인공지능 피씨인 Copilot+ PC 공개
지브라 	<ul style="list-style-type: none"> '23년 디지털 솔루션 기업으로 퀄컴의 플랫폼을 기반으로 자사 핸드헬드 모바일 컴퓨터 및 태블릿 등 온디바이스 환경에서 클라우드와의 연결 없이 대형언어모델 및 생성형 AI를 구현하는데 성공했다고 공개
화웨이 	<ul style="list-style-type: none"> 자체클라우드 LLM인 Pangu와 디바이스 내 AI비서인 Cella가 연결되며, '23년 8월 HarmonyOS 4에 LLM을 탑재할 것을 발표
오포 	<ul style="list-style-type: none"> 자체 LLM인 AndesGPT를 기반으로 AI비서인 Xiaobu를 외부테스트를 수행 중이며, 향후 스마트폰 등에 탑재 예정

국내

◎ 대기업에서 한국어 특화 생성형 AI 개발 및 경량화를 진행 중이며, 자사 제품이 적용을 목표로 기술개발 진행

기업명	기술 동향
삼성전자	• '23년 온디바이스 환경에서 생성형 AI 서비스 제공을 위한 '삼성 가우스'를 공개, '24년부터 갤럭시S24 등의 제품에 탑재
카카오	• 한국어 특화 AI 언어모델인 'KoGPT'를 '21년 공개했으며 구글의 텐서 처리 장치를 활용, 연산속도 고도화. '24년 1월 새 언어모델 '하니비'를 공개하였으며 문자와 이미지를 동시에 학습하고 연산할 수 있는 멀티모달(다중모델) 기반의 대규모 언어모델(LLM)을 오픈소스 형태로 구현
네이버	• 블로그, 카페, 지식 IN 등 자사 서비스의 대규모 데이터를 활용해 한국어 대용량 데이터를 구축한 '하이퍼 클로버'를 '21년 공개
LG전자	• 멀티모달 연산이 가능한 초거대 AI 'EXAONE 2.0' 공개
LGCNS	• 대규모 언어모델(LLM)을 탑재한 생성형 AI 서비스 플랫폼 'DAP GenAI' 구축
LGU+	• '24년 LG AI 연구원과 협업해 통신, 플랫폼, 금융 등 다양한 분야에 적용 가능한 생성형 AI 모델인 '익시젠(ixi-GEN)'개발 계획 공개. 딥엑스와 협력하여 추후 온디바이스 환경에서도 익시젠 기반 서비스를 제공하기로 발표
SKC&C	• 네이버와 협력하여 '한국형 AI 서비스'를 공동개발하여 기업 맞춤형 보고서 제작 솔루션 개발
삼성SDS	• 기업 내 업무 생산성을 높이기 위한 생산형 AI 플랫폼 '브리티 코파일럿(Brity Copilot)'과 '팹브릭스(Fabrix)'공개

5 활용 서비스 사례

◎ 사용자와 상시 연결되어 즉각적인 AI 서비스 제공이 가능한 모바일·웨어러블 기기를 중심으로 온디바이스 생성형 AI를 탑재한 제품 개발 및 출시

적용분야	사례 및 주요내용
웨어러블 온디바이스	 휴메인(Humane) / AI 핀 • 월컴 모바일 AI반도체 기반 생성형 지능 탑재한 퍼스널 디지털 디바이스('23.12, 발표, '24년 국내 출시 예정) • 생성AI 탑재를 통해 통번역, 전화 송수신, 건강·영양, 사진촬영, 검색, 이메일 요약, 음악재생 등 서비스 제공
	 틴에이저엔지니어링(TeenageEngineering) / Rabbit R1 • 대형 액션모델 기반으로 휴대폰 앱구동 AI에이전트 장치('24.1, 출시) • 사용자 음성을 통해 쇼핑, 호텔 예약, 메시지 송신 등 서비스 제공

적용분야	사례 및 주요내용	
웨어러블 온디바이스	<p> 메타(Meta) / 스마트 (AI) 안경</p> <ul style="list-style-type: none"> • 스마트안경 레이밴 스토리 출시('23.10 발표, 미출시) • 메타 생성형 AI 탑재를 통해 사용자 시선의 사물에 대한 실시간 정보 제공 	
	<p> 삼성전자 / 갤럭시링</p> <ul style="list-style-type: none"> • 반지 형태의 스마트링('24.1 발표, 8월 출시 예정) • 심박수, 혈압, 산소포화도, 수면 품질 등 실시간 측정, 지능형 헬스 기능을 통한 맞춤형 건강 가이드 제공 	
	<p> 탭AI(TabAI) / AI 펜던트</p> <ul style="list-style-type: none"> • 작은 원형 패션 펜던트 ('24.1, 국내 미출시) • 30시간 지속 배터리, 마이크와 블루투스 사용해 오디오를 수집 및 스마트 폰과 연동하여 클라우드 전송, ChatGPT를 활용하여 다양한 통찰력 제공 	
	<p> 오우라헬스(Oura Health) / 오우라링(3세대)</p> <ul style="list-style-type: none"> • 반지 형태의 스마트링('21년말, 국내 미출시) • 수면 추적(체온, 심박, 심박변이 등), 활동 추적(3D가속도를 통한 운동·활동 모니터링 등) 등 	
모빌리티 온디바이스	<p> 유토피아(UTOPIA) / 전기자전거 퓨전(Fusion)</p> <ul style="list-style-type: none"> • 생성형AI 연동 앱 디자이너와 협업된 제품 전기자전거('23.7, 발표) • 사용자 음성을 통한 정보(수리, 영양플랜, 부상회복), GPS(여행경로, 도난방지), 건강(심박) 등 서비스 제공 	
스마트 오피스·홈 온디바이스	<p> 플라우드 노트(PLAUD NOTE) / 플라우드 노트</p> <ul style="list-style-type: none"> • ChatGPT 기반의 스마트 레코더('23.11, 출시) • 녹음(통화, 일반) 및 녹음내용의 실시간 텍스트 번역, 요점정리, 마인드맵 정리, 메일 발송 등 서비스 제공 	
	<p> LG전자 / 가전</p> <ul style="list-style-type: none"> • 멀티모달 센서와 온디바이스 AI 적용한 초개인화 가전 스마트 헤이전트 공개('24.1, CES) • 가전용 AI 칩 개발 	
스마트폰 온디바이스	<p> 삼성전자 / 갤럭시S24</p> <ul style="list-style-type: none"> • 삼성전자의 생성형 AI 모델(삼성 가우스)을 탑재한 스마트폰('24.1, 출시) • 통화중 실시간 통역(13개국), 실시간 문자 번역(13개국), 사용자 언어를 통한 검색, 문서 요약 등 서비스 제공 	
	<p> 구글 / 픽셀8 Pro</p> <ul style="list-style-type: none"> • LLM 온디바이스 모델(나노)를 탑재한 스마트폰('23.12, 출시) • 생성형 AI 기반 망연동 없는 녹음내용 요약, 대화 맥락 분석을 통한 답변 제안, 영상 교정 등 서비스 지원 	

IV. 미래 발전 방향

1 향후 전망

(시장) 온디바이스 AI 기술은 AI 시장의 중심이 될 것으로 기대

- ◎ 개인정보 보호, 안정적인 실시간 서비스, 서버 운영비용 절감 등 온디바이스 AI의 다양한 강점으로 인해 향후 AI 시장의 게임 체인저 역할 예상
 - 온디바이스 AI 기술은 대규모 클라우드 서버 운영이 필요 없어 비용 절감 측면에서 혁신적인 기술이며, 나아가 데이터 센터의 에너지 소모 절감 가능
 - 오픈AI는 '23년 매출이 2조 원이 넘었으나 하루 약 9억 원으로 추산되는 막대한 운영비용으로 수익 실패
 - 글로벌 빅테크들은 경쟁적으로 온디바이스 AI 기술 실현에 앞장서고 있으며, 지속적인 투자 계획 발표
- ◎ 현재 온디바이스 AI 분야로 대표되는 스마트폰 외 웨어러블 기기, 로봇, 디지털 헬스케어, 스마트 홈 등 다양한 산업에서 폭넓게 활용될 것으로 전망
 - AI 서비스가 일상에 깊숙이 침투함에 따라 개인정보 보호에 대한 중요성은 나날이 증대
 - 지능형 블랙박스·인공지능(AI) 스피커·지능형 CCTV 등 일상 속 AI 서비스에는 사용자의 음성, 영상 등이 입력 데이터로 활용되며, 이러한 개인정보에 대한 보호 조치 필요
 - 온디바이스 AI를 활용해 개인정보 보호 문제없이 사용자 맞춤형 AI 서비스를 제공할 수 있어 활용성 클 것으로 기대

(업계) 기술확보를 위한 M&A 및 AI 관련 인력 확보 전쟁 심화

- ◎ 미래 산업으로의 전환을 위해 전 산업에서의 AI분야 M&A 활발히 진행
 - AI가 타산업의 발전을 가속화하는 기반 기술로 부상하며 로봇, 헬스케어, 모빌리티, 게임·엔터 등 전 산업에 적용되고 영향력 확대됨에 따라 AI를 중심으로 기술 융합과 혁신 진행
- * 최초 보급 후 10년차 침투율(PWC·Bessemer VC, %) : (생성형AI)66, (스마트폰)55, (SW 내 클라우드)31

산업별 AI 영향도

산업	영향도
로봇	<ul style="list-style-type: none"> AI 접목된 로봇은 최종적으로는 인간의 개입을 필요로 하지 않는 '무인화' 로봇, 즉 스스로 활동할 수 있는 '휴머노이드 로봇'으로 발전
헬스케어	<ul style="list-style-type: none"> (신약개발) 딥러닝을 통한 단백질 구조 해독 등으로 신약개발에 소요되는 시간 획기적 단축 (진단/치료) 딥러닝으로 환자 유전자 데이터를 분석, 대규모 의료영상 빠르게 처리 → 질병 판독 및 의료진 치료결정에서 불확실성 감소 (예방) 헬스케어 기관들이 보유한 방대한 의료 데이터를 활용하여, 개인별 질병 발생 확률 예측 & 사전 예방형 치료 제공 (서비스) 헬스 디바이스의 발전과 함께 주요 건강지표들의 지속적 모니터링 & 챗봇을 통한 의료상담 등이 가능해지면서, 원격의료 본격화
모빌리티	<ul style="list-style-type: none"> 인포테인먼트·커넥티드를 넘어 자율주행의 시대 도래 <ul style="list-style-type: none"> - 도로 주행데이터 축적 및 AI 분석능력 향상으로 완성도 높은 자율주행시스템구현
게임·엔터	<ul style="list-style-type: none"> (개발) AI를 통한 코딩 자동화를 통해 개발자들의 중복되는 코드 작업 최소화. 인건비 감소는 기업들의 제작비용을 감소시켜 더 많은 콘텐츠 개발 및 진입 장벽을 낮춰 신생업체들의 대규모 유입 유도 (콘텐츠) AI는 메타버스와 결합해 게임/엔터/교육의 궁극적인 변화 이끌 전망 <ul style="list-style-type: none"> - AI를 활용한 디자인 프로세스 간소화로 3D 부문의 급속한 발전 가능. 홀로그램 활성화로 '가상-현실의 경계'가 모호해질 것으로 전망 - 메타버스 활성화로 게임/엔터/교육/광고 분야에 큰 수혜 예상
테크	<ul style="list-style-type: none"> (하드웨어) 개인 IT 기기(핸드폰, 태블릿 등)에 온디바이스 AI 탑재되며 신규 수요 창출 및 XR(확장현실)기기 등 신제품 출시 (소프트웨어) AI 모델에 대한 시장 선점을 위한 글로벌 기업들의 개발 경쟁 지속 및 AI를 활용한 개인 맞춤형 서비스를 제공할 수 있는 어플리케이션 발전 (반도체) AI 반도체에 대한 중요성 부각되며, 시스템 반도체 부문 핵심 경쟁력을 확보하기 위한 글로벌 거대 기업들의 투자 지속
금융	<ul style="list-style-type: none"> 챗봇(상담 업무)를 넘어서 금융 비서로 AI 업무 영역 확장 <ul style="list-style-type: none"> - 자연어 처리 역량이 음성인식과 맞물릴 경우 모바일 환경에 익숙하지 않은 연령층까지 고객군 확대 가능 - 단순 상담 업무 뿐 아니라 자산관리 및 투자 부분에서도 AI의 역할 확대 예상

* 출처 : PWC('24)

- 경기 불확실성에도 첨단기술인 AI(온디바이스AI, 메타버스 기술 및 하드웨어)에 대한 투자에 집중해 미래 산업 대응
- (테크) 온디바이스 AI, 메타버스, 하드웨어 등 AI 관련 기술 확보를 위한 M&A 진행
- (통신) 데이터센터, AI, 로봇, 클라우드 등 미래 기술에 대한 투자 지속 및 디지털화·커넥티드 역량 확보를 위해 소프트웨어, 모빌리티, 기계 업종 등과의 제휴 증가
- (미디어) 빅테크의 파운데이션 개발 투자, 스타트업 지분 투자·M&A 통해 생성형 AI 기반 다양한 수익모델 시도

◎ AI관련 핵심 인재 확보를 위한 빅테크 기업 간 경쟁 심화

- OpenAI, 구글, MS, 애플, 메타 등 빅테크 기업들은 AI 인력에게 연봉, 스톡옵션 등 막대한 조건을 제시하여 공격적인 영입 제시하며 AI 인재 확보에 총력
- * 글로벌 빅테크 AI 연구원('23년 신규 박사급 AI 연구원 기준) 초봉(한국경제·로라, 만 달러)
 - : (오픈AI)86.5, (엔스로픽)85.5, (인플렉션AI)82.5, (테슬라)78, (아마존)71.9, (구글브레인)69.5

- 국내 AI 인력은 경쟁국에 비해 매우 부족한 상황으로 민간의 인재 유치 강화 전략 및 정부 맞춤 인재육성 정책 절실

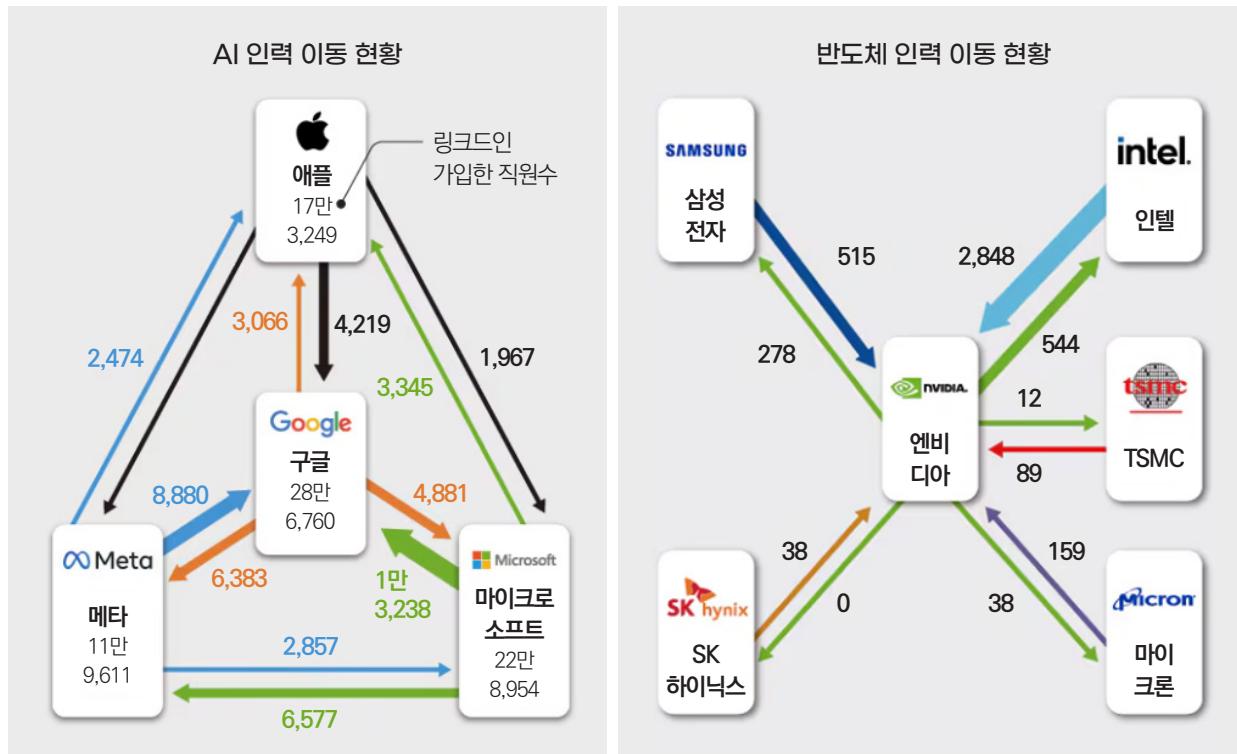
* '20년 국가별 글로벌 AI 전문인재 확보 현황(한국일보·엘리멘트AI, %)

: (1위)미국 39.4, (2위)인도 15.9, (3위)영국 7.4, (4위)중국 4.6....(22위)한국 0.5

- AI 반도체 시장 주도를 위해 반도체사들의 인력전쟁도 엔디비아를 중심으로 진행되어 국내업체에서의 인력 유출 심화

빅테크간 AI 및 반도체 인력 전쟁

(단위 : 명, '24. 6. 17 링크드인 가입자 기준)



* 출처 : 조선일보('24)

■ (기술·서비스) 킬러 서비스 확보를 위한 경쟁 심화 및 개인화된 맞춤형 AI로 점진적 발전 예상

◎ 경량 생성형 AI 모델, NPU와 같은 온디바이스 AI 실현을 위한 기반 기술의 성숙도가 높아짐에 따라 이를 활용해 사용자에게 혁신적인 경험을 제공하기 위한 킬러 서비스 확보 경쟁 심화 예상

- 온디바이스 AI 기술 자체는 이전부터 활용되고 있어 사용자에게 새로운 경험 제공이 어려우며, 온디바이스 생성형 AI 기반의 차별화된 서비스 개발 필요
 - 스마트폰의 파노라마 사진, 사진에서 객체 제거, 실시간 영상 합성 기능을 비롯해 블랙박스의 차량 출발 감지 기능, 지능형 CCTV의 객체 인식 기능 등 다양한 온디바이스 AI 서비스 기 보편화
- 사용자가 체감 가능한 온디바이스 생성형 AI 기술은 스마트폰 중심의 실시간 통역과 음성 대화의 텍스트 요약 수준
- 글로벌 빅테크들은 스마트폰·AI PC·웨어러블 기기 내 온디바이스 생성형 AI 기술을 활용하여 대표 서비스 개발에 박차를 가하고 있으며 '24년 말부터는 갤럭시 링을 포함하여 단계적 출시 예상

◎ 맞춤형 AI 서비스를 위해서는 개인의 데이터를 기반으로 AI 모델의 점진적 업데이트 필요

- 현재는 서버에서 학습된 AI 모델을 모든 사용자가 동일하게 사용하고 있어 사용자 특성이 반영된 서비스 제공 불가
- 디바이스에서 수집한 사용자의 데이터 통해 지속적으로 AI 모델을 업데이트하여 사용자별 특화된 서비스 제공 가능
- AI모델학습은 추론 대비 훨씬 많은 컴퓨팅 자원이 필요해 온디바이스 환경에서의 AI모델학습을 위해 많은 난관 극복 필요
- 최근 서버향 타겟의 NPU에서 학습을 지원하는 등 기술 발전이 이루어지고 있어 향후 맞춤형 AI 서비스의 실현 기대

2 정책 제언

I 온디바이스 AI 시장 선점을 위해 중소기업 시장 참여를 통한 생태계 확장 필요

◎ 국내 온디바이스 생성형 AI 기술은 저전력 고효율 지능형 반도체 개발이라는 후방 산업에 집중되어 있고 서비스는 일부 대기업 위주로만 개발 중으로, 혁신적인 서비스 개발이 이루어지지 않아 전방 산업 부진

- 삼성전자의 스마트폰에 중점적인 온디바이스 시스템을 위한 기반 역량은 보유하고 있으나 폐쇄된 대기업 기술로 중소기업과 역량 차이 매우 큼

◎ 국내 중소기업 기술역량 한계로 높은 난이도의 온디바이스 AI 기술 도입 난항

- '21년 12월 말 기준 전체 기업체(20만 7,000여개)의 AI 기술 및 서비스 이용률은 2.7%로 매우 낮음
- 국내 기업 100곳 중 3곳만이 AI 기술을 도입했으며 AI를 도입한 기업 대부분은 대기업(91.7%). 중소기업의 AI 도입 시 가장 큰 걸림돌은 수요에 맞는 AI 기술 및 인력 부족(72.1%)

◎ 사용자 수요 증대를 통해 전방기술 시장을 확대하고 지속적인 성장을 견인하기 위해서 다수의 중소기업이 생태계에 참여해 도전적이고 참신한 서비스 발굴 필요

II 중소기업의 온디바이스 AI 시장 진입 장벽 해소를 위한 기술·서비스 개발, 인력양성 등 맞춤형 정책지원 필요

◎ (기술 개발 및 연구 기반 지원) 온디바이스 생성형 AI 제품 및 서비스 개발을 위한 연구 기반 지원 필요

- 온디바이스 생성형 AI의 특성상 서비스 개발을 위해서는 온디바이스 AI 하드웨어부터 소프트웨어 및 경량화된 생성형 모델까지 필요하나 이는 중소기업이 시장에 진입하는 것을 가로막는 주요 걸림돌
- 저전력 레퍼런스 보드, 개방형 생성형 AI 모델, 한국어 타겟 학습 데이터 공개 등 중소기업의 기술적 제약을 해소해 혁신적인 제품 및 서비스를 개발할 수 있도록 기반 지원 필요

◎ (인력 양성 및 교육 지원) 온디바이스 AI 인재 부족 현상을 해소하기 위해 국가 차원의 인력 양성 및 재직자 역량 강화를 위한 교육 지원 필요

- 서버 시스템 지원 AI 소프트웨어 및 하드웨어 인력은 국가의 전략적인 연구개발 투자 및 인력 양성으로 지속적으로 증가
- 온디바이스 AI 소프트웨어 및 하드웨어 시스템에 대한 역량을 가진 인재는 매우 부족하여 AI 및 임베디드 시스템 기술 전문성이 융합된 온디바이스 AI 인력 양성이 핵심
- 온디바이스 AI 인재 풀 확대와 기술 수준 향상을 위해 대학 등을 대상으로 한 인력 양성 뿐만 아니라 중소기업 재직자 대상 교육 지원 등 다각도에서 정책적 지원 필요

◎ (서비스 개발 및 실증 사업 지원) 온디바이스 생성형 AI 적용 서비스 사례 확보를 위한 실증지원 절실

- 국내 중소기업은 AI 기술 및 NPU 확보, AI 인력, 외부 전문가 협업 등 자원이 부족하여 서비스 출시 및 사례 확보가 어려워 신산업 진출, 글로벌 수출 차질
- 특히, 학습 데이터 확보, 온디바이스 AI 하드웨어 및 최적화된 소프트웨어 개발 등 서비스 개발을 위해 시간과 비용이 많이 필요한 온디바이스 생성형 AI 특성 상 종래 단기간 내 실증 사업은 실효성 부족
- 온디바이스 생성형 AI 서비스 개발과 실환경 장기간 검증을 수행할 수 있도록 2~3년차의 중기간 사업을 통한 실증지원 필요

참고문헌

- AI반도체 시장 현황 및 전망, 2024, 한국수출입은행
- 생성형 AI 열풍으로 성큼 다가온 온디바이스 AI, 2024, LG경영연구원
- 저전력 온디바이스 비전 SW 프레임워크 기술 동향, 2021, 한국전자통신기술연구원
- AI·디지털 혁신성장 전략(요약), 2024, 과학기술정보통신부(관계부처 합동)
- 온디바이스 AI 시장 규모, 2023, 마켓츠앤마켓츠·파이낸셜뉴스
- 생성형 AI 스마트폰 시장 규모, 2023, 카운터포인트리서치
- 생성형 AI 시장 규모, 2023, 블룸버그인텔리전스
- Large Language Model Cost Analysis, 2023, 세미에널리시스
- Global M&A Industry Trends : 2024 Outlook, 2024, 삼일PwC경영연구원
- 빅테크 기업 ‘AI·반도체 인재 쟁탈전’, 2024, 조선일보
- 글로벌 빅테크 AI 연구원 조봉, 2024, 한국경제·로라
- 글로벌 AI 인재보고, 2024, 한국일보·엘리먼트AI
- 정보화통계조사, 2022, 한국지능정보사회진흥원
- AI에 대한 기업체 인식 및 실태조사, 2021, 한국개발연구원

KEA NOW

KEA 주요 소식을 한눈에

1

KEA, KTC와 업무협약 체결



한국전자정보통신산업진흥회(KEA)는 6월 5일 한국기계전기전자시험연구원(KTC)과 가전·전자기업을 포함한 ICT유망기업의 공동 발굴 및 지원프로그램 협력강화를 위한 업무협약을 체결하였다.

양 기관은 이번 협약을 통해 △ICT유망기업 육성을 위한 협력 네트워크 구축 △스마트가전, IoT사이버보안, 인공지능, 소프트웨어, 5G/6G 등 분야의 시험·인증 컨설팅 지원 △KEA 전자혁신제조 지원센터, XR실증센터, 빅데이터센터의 기업지원 프로그램을 공동 활용하여 미래 유망기업의 혁신성장을 지원한다.

특히, IoT보안인증에 대응하여 수출기업이 심각한 애로에 직면하지 않도록 사이버보안 관련 해외인증 정보와 컨설팅을 제공한다.

이번 협약은 양 기관의 협업 네트워크를 기반으로 ICT유망기업의 제품·서비스 개발부터 시험·인증 및 제품 출시까지 수요자 중심의 끊김없는(Seamless) 지원체계를 구축하는데 의미가 크다.

양 기관은 업무협약의 첫 행사로 6월 27일, 상암동 전자회관에서 가전·전자기업을 위한 ‘해외 IoT사이버 보안 컨퍼런스’를 개최할 예정이다.

박철원 KEA 부회장은 “이번 협약이 ICT유망기업의 혁신성장을 위한 지원기관 간 협력을 강화하는데 의미가 크다며, KEA는 전자·IT산업 진흥기관으로써 수요자 중심의 기업지원플랫폼이 되도록 유관기관과 협력을 지속해 나가겠다.”고 밝혔다.

2

KEA, 전자제조 스타트업 육성



한국전자정보통신산업진흥회(KEA)는 6월 9일 아프리카 케냐 콘자개발청(KoTDA) 관계자들이 서울 용산에 위치한 전자제조센터를 방문, 전자제조 스타트업 육성 모델을 벤치마킹하고 교류 방안을 논의했다고 밝혔다.

KEA 전자제조센터는 도심형 제조사설로, 연간 200여종 전자제품 생산을 지원한다. 40여종 전문장비 시설을 갖추고 있어 스타트업이 어려움을 느끼는 시제품 설계와 생산 지원에 특화돼 있다. KEA 기술진이 전자분야 제품 개발부터 생산, 홍보, 전시 과정을 지원한다. 케냐 정부는 국민 삶의 질 향상과 산업화 지원으로 중산층 국가로 거듭나기 위한 비전 2030을 추진하고 있다. 이의 일환으로 KoTDA는 수도 나이로비에서 60km 떨어진 지역에 스마트시티를 꾸리고 첨단기술 스타트업을 육성하기 위한 생태계를 조성하는 ‘콘자 테크노폴리스’ 프로젝트를 추진하고 있다. 첨단기술과 제조·연구시설을 갖춰 아프리카의 기술 허브로 발전시키는 것이 목표다. 전자제조센터에는 KoTDA 위원회 이사장과 부청장이 방문하여 케냐 콘자의 도시 모델에 반영하기 위해 한국의 선진 전자산업 시설과 산업 생태계를 둘러봤다. KEA는 전자제조센터의 국내 혁신제품 전초기지 역할, 지원 프로그램, 한국 스타트업 생태계 구조를 설명했다. 내부 시설을 소개하고, 지원한 주요 스타트업 제품을 소개했다. KoTDA는 콘자 테크노폴리스 사업 내 과학기술원과 ICT 스타트업 육성에 KEA 전자제조센터 모델을 반영할 방침이다.

박철원 KEA 부회장은 “케냐와 유기적 파트너십 구축을 기대한다”며 “KEA 지원 모델이 개발도상국을 포함해 세계적으로 확산되도록 노력하고 스타트업의 전자제품 전시와 홍보에도 협력하겠다”고 밝혔다.

3

KEA, 전자정보분야 해외기술규제 설명회 개최



한국전자정보통신산업진흥회(KEA)는 6월 11일 양재 엘타워에서 ‘2024년도 제1차 전자정보분야 해외기술규제 설명회’를 개최했다.

최근 산업계 전반에서는 생성형AI·머신러닝 등의 기술개발 과정에서 투명성을 강화해 자국 산업을 보호하고, 제품의 생애주기 정보 등을 공개해 공급망을 관리하고 있다.

‘전자정보분야 해외기술규제 설명회’는 이 같은 추세를 반영해 AI, 사이버보안, 화학물질 등 전자정보 분야의 주요 해외 기술규제에 대한 산·학·연의 대응력 강화를 위해 개최했다.

이번 설명회는 ▲AI 사이버 시큐리티 기술동향 및 대응방안 ▲EU 사이버복원력법(CRA) 규제동향 및 대응방안 ▲EU DPP 규제동향 및 대응방안 ▲분석장비를 활용한 과불화화합물(PFAS) 규제 대응방안 순으로 발표를 진행했다.

KEA 박청원 부회장은 “오늘 설명회는 전자·IT업계 학우가 되는 해외기술규제 대응력 강화를 위한 설명회로서 우리나라 전자·IT기업들이 중대한 해외기술규제에 선제적으로 대응하는데 도움이 될 중요한 자리라고 생각한다”며 “전자·IT업계에서 무역기술장벽을 넘기 위해 애쓰는 기업들이 지속적으로 성장해 나갈 수 있도록 적극 노력하겠다”고 말했다.

4

KEA, 영덕·의성에서 가전제품 무상점검 행사



한국전자정보통신산업진흥회(KEA)는 6월 12일과 13일 양일간 영덕군과 의성군에서 가전제품 무상점검과 건강진단 기반 의료기기 체험 행사를 실시했다. 행사에는 삼성전자, 세라젬, LG전자, 오텍캐리어, 쿠쿠전자, 쿠첸, 휴롬엘에스, 경동나비엔, 귀뚜라미가 참여했다.

전자기업은 휴대폰 서비스 차량과 실내 행사장에 부스를 설치하고 농어민, 고령자, 다문화가정 등 취약계층 소비자의 노트북, 밥솥, 청소기, 공기청정기 등 소형가전 제품 무상수리와 의료기기 체험을 제공했다. 노인복지관, 보육원, 장애인복지관 등 사회배려시설을 사전 방문해 보일러와 대형가전 제품을 점검·수리했다.

행사와 연계해 어르신 장수사진 촬영, 취약계층 소비자의 피해 예방 교육과 자산관리 상담, 고령자 대상 건강상태 진단과 의료기기 체험 행사도 실시했다. 마을 단위로 수집한 소형폐가전도 회수했다.

본 행사는 한국소비자원과 경상북도 영덕군, 의성군 등 지자체가 공동으로 주관하였으며, 한국전자정보통신산업진흥회, 기업소비자전문가협회, 한국자동차모빌리티산업협회, 한국석유관리원, 한국주택금융공사, LG생활건강 등이 참여하여 행사를 지원하였다.

MEMBER NEWS



KEA 회원사의 성과와 동향

MEMBER NEWS

* 기사를 클릭하면 보다 자세한 내용을 확인하실 수 있습니다.

세라젬,

헬스케어 도전 및 매출 3조 목표

1

국내 안마의자 1위 기업인 세라젬의 이경수 대표는 한국경제신문과의 인터뷰에서 “7대 사업부문으로 영역을 확장해 글로벌 종합 헬스케어 회사로 발돋움하겠다”며 이같이 말했다. 2022년 7501억원이던 회사 매출을 2028년까지 3조원으로 끌어올리겠다는 포부도 밝혔다.

이어 “아직 공개하지 않은 제품이 계속 연달아 출시될 것”이라며 “세븐케어라고 이름 붙인 우리 7개 사업군을 국내뿐 아니라 해외에서도 체험할 수 있도록 5년 안에 전 세계 1만 개 오프라인 점포를 여는 게 목표”라고 말했다. 세라젬은 아시아, 미주뿐 아니라 유럽, 아프리카, 중동 등 70여 개국에 진출해 있다.

올해 새로 나올 제품은 혈액 순환을 돋는 의료기기, 공기청정기, 맞춤형 영양제, 우울증 치료기, 불면증 치료기 등이 있다. 최근 출시한 파우제 M6는 척추 스캔 기술, 온열 기능을 담은 실리콘 볼, 팔다리 안마 기능 등을 다 담았다.



디지월,

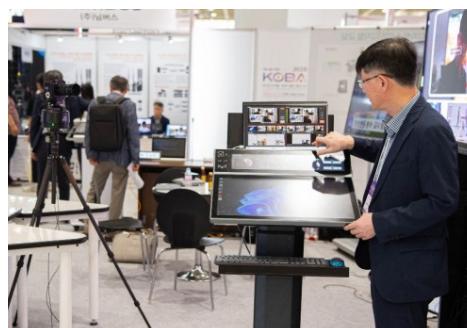
양방향 AI 최첨단 교육 시스템 '클래스 아이' 출시

2

최근 (주)디지월(대표 홍석환)이 AI 기반 실시간 양방향 공유 에듀테크 시스템 ‘클래스 아이(Class interaction)’를 출시해 주목을 받고 있다. 이번에 출시된 ‘클래스 아이’는 수업 모듬형 조별 학습장치, 콘텐츠·학습관리 시스템(LMS), AI학습분석 솔루션, 멀티미디어 학습장치(미디어 허브) 등 하나로 통합 관리가 가능한 시스템이다.

현재 전남교육청과 협력해 여수와 목포에 사전 시범 운영하고 있으며, 오는 29일 여수세계박람회장에서 ‘2024 대한민국 글로컬 미래교육박람회’에서 미래교실을 시연할 예정이다.

홍석환 디지월 대표이사는 “지속 가능한 성장과 사회적 책임 달성을 위해 고객 요구에 적극 대응하며 기술 혁신에 중점을 두고 있다”며 “회사가 추구하는 미래 가치를 바탕으로 지속적인 연구개발을 통해 급변하는 디지털 시대에 발맞춰 나가겠다”고 밝혔다.



MEMBER NEWS

* 기사를 클릭하면 보다 자세한 내용을 확인하실 수 있습니다.

삼영전자공업, 친환경 모빌리티 시장 최적화 하이브리드 콘덴서 개발

3

알루미늄 전해콘덴서 전문생산업체 삼영전자는 일본케미콘과 기술협력을 통해 장수명 하이브리드 콘덴서를 개발했다고 밝혔다.

이번에 개발된 FPV/FPW/FRV시리즈는 고온도(125°C 이상) 환경에서 전해콘덴서의 보증 수명을 해결하여 ADAS첨단운전자 지원시스템, HUD헤드업디스플레이, Telematics인터넷차량정보통신장치, IBU통합바디제어기 등 친환경 자동차 분야와 외부 노출로 큰 기온 변화를 견뎌야 하는 5G 통신용기지국 같은 고신뢰성이 필요한 특수 산업 시장에서 활용할 수 있다.

업체 관계자는 “이번 장수명 하이브리드 콘덴서 개발로 인해 전량 수입에만 의존했던 하이브리드 콘덴서의 국산화가 가능해졌다며 친환경 자동차시장의 성장과 더불어 하이브리드 콘덴서의 글로벌 시장 규모도 매년 30% 이상 성장하고 있어 향후에도 지속적인 투자와 연구개발로 신제품 개발을 이어나갈 것이다”고 말했다.



코맥스, 대한적십자사 의료원에 디지털 도어락 기부

4

스마트홈 전문기업 코맥스가 대한적십자사 의료원에 5000만원 상당의 스마트 디지털 도어락을 기부했다고 밝혔다. 이번 기부는 코맥스가 사회공헌활동의 일환으로 국민의 건강을 위해 노력하는 대한적십자사 의료원과 해당 시설을 이용하는 환우들에게 도움을 주기 위한 목적으로 진행됐다.

코맥스의 스마트 디지털 도어락은 카드, 비밀번호 뿐 아니라 지문 인식, 얼굴 인식이 가능한 스마트 바이오 등 강력한 보안 기능들이 탑재돼 병원 내 안전과 보안을 크게 향상시킬 것으로 기대된다. 또 이를 통해 의료진과 환우의 편리한 생활이 보장되고 병원 운영의 효율성도 높아질 것으로 전망된다.

이기상 코맥스 전략기획부문장은 “불철주야 노고가 많은 의료진과 응원이 필요한 환우 및 가족들에게 작은 격려라도 해드리고 싶다”며 “이번 기부를 통해 새로운 인연을 맺고, 지속적인 사회공헌활동으로 사회적 기업의 책임을 다하겠다”고 강조했다.



ESG TREND



ESG Regulations 주요현황

1



환경부

**'24년 내 환경성보장제
전품목 확대 계획**

출처 : 환경정책협의회



환경부(한국환경공단)는 전기·전자제품 환경성보장제를 전품목으로 확대할 방침이며, 지난 5월 30일 전기전자 업계 대상으로 확대 계획 설명회를 개최하였다.

환경성보장제는 전기·전자제품의 재활용을 촉진하기 위해 제조·수입업자의 유해물질 사용 억제, 재활용 정보의 제공, 폐제품 회수·재활용 및 재질구조개선사항 평가 등 6대 의무를 부여하는 제도로서, 2008년 최초 도입 이래 현재 50개 품목에 적용 중이다.

환경부는 시장 발전에 따라 다양해진 품목을 관리하고 국가 회수·재활용 목표의 원활한 달성을 위해 환경성 보장제를 전품목으로 확대할 예정이며, 의류건조기, 의류관리기, 전기레인지 및 디지털카메라 등 신규 품목에 대해 2026년부터 회수·재활용 의무를 적용하고 2028년부터 유해물질 사용제한도 시행할 전망이다.

더 나아가 현재 49개 품목 대상 시행 중인 재질·구조 개선사항 평가제도의 대상 품목도 장기적으로 확대될 것으로 예상된다.

환경부는 올 하반기 중 근거 법률인 전자제품등 자원순환법 개정안을 공개하고 12월까지 법률 개정을 완료할 계획이며, KEA는 향후 입법 동향을 모니터링하고 전자업계의 의견이 반영될 수 있도록 노력할 방침이다.

2



우즈베키스탄

**수입 가전제품 에너지효율등급
인증서 발급 지연**

출처 : KEA TBT 사무국



우즈베키스탄으로 수출되는 가전제품에 대한 에너지 효율등급 시험기관의 부족 및 운영 차질로 수출기업의 인증서 발급 지연과 통관 애로가 우려된다.

우즈베키스탄 정부는 올해 1월부터 에너지대책의 일환으로 에너지효율 B등급 이하 가전제품의 수입 금지 조치를 시행 중이며, 현재는 냉장고, 에어컨 2개 품목에 적용 중이나 향후 대상 품목을 세탁기, 청소기, TV, 모니터, 전자레인지 등 여러 품목으로 확대할 계획이다.

우리나라 수출기업은 통관 시 현지 시험기관이 발급한 에너지효율등급 적합성 인증서(CoC)*를 제출해야 하나, 이용 가능한 시험기관이 에어컨 2개소, 냉장고 1개소에 불과할 뿐 아니라 일부 기관은 내부 사정으로 CoC 발급 업무가 중단됨으로 인해 CoC 발급 지연이 우려되는 실정이다. 나아가 대상 품목이 확대될 경우 인증 수요를 더 해소하기 어려워질 전망이다.

* CoC(Certificate of Conformity)

KEA는 국가기술표준원(이하 국표원)의 전자산업 TBT 기술규제대응 지원기관으로서 업계와 대응방안을 여러 차례 논의하였으며, 국표원을 통해 대상 품목 추가 시 인증서 발급기간을 확보할 수 있도록 우즈베키스탄 정부에 충분한 유예기간 부여를 요청하였다.

STATS



데이터로 읽는 전자·IT 업황

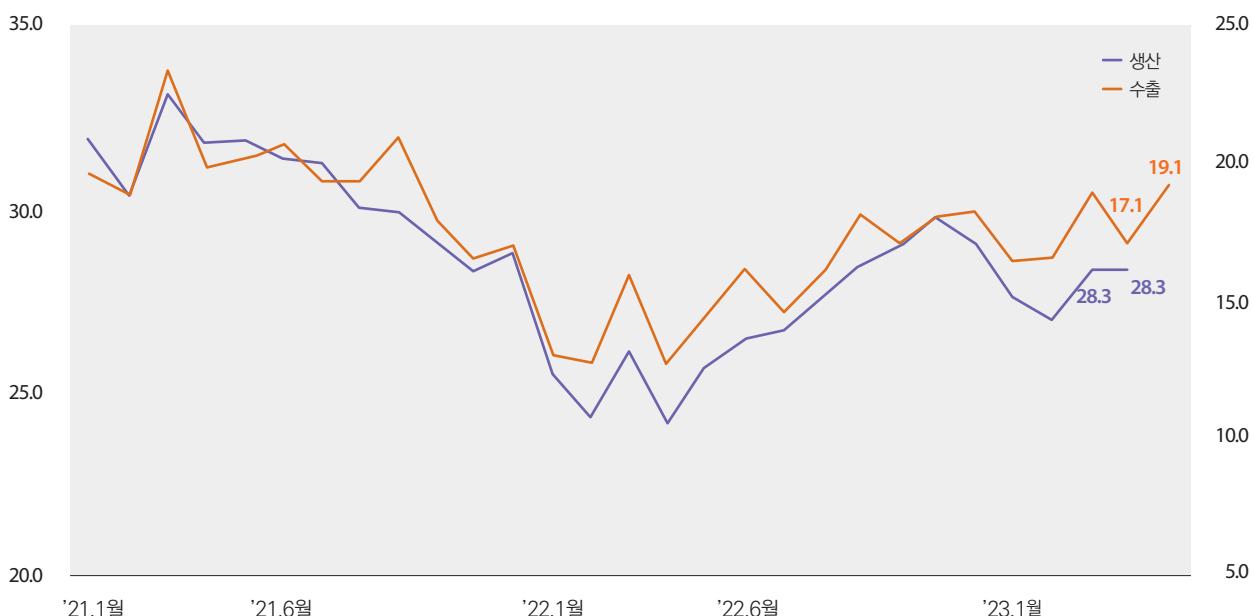
1. 전자·IT 제조업 생산, 수출

• **(생산)** '22년 6월부터 17개월간 전년동월대비 감소세를 지속해오다 '23년 11월부터 증가세로 전환됐으며
'24년 4월 생산은 전년동월대비 16.8% 증가한 28.3조원

• **(수출)** '22년 7월부터 16개월간 전년동월대비 감소세를 지속해오다 '23년 11월부터 증가세로 전환됐으며
'24년 5월 수출은 전년동월대비 31.8% 증가한 19.1억불

구분		1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월	계
생산	2022	31.8	30.4	33.1	31.8	31.8	31.4	31.2	30.1	29.9	29.1	28.3	28.8	367.8
	2023	25.6	24.4	26.0	24.2	25.7	26.4	26.7	27.6	28.4	29.0	29.7	29.1	322.9
	2024	27.6	27.0	28.3	28.3									
수출	2022	19.6	18.9	23.3	19.9	20.2	20.6	19.3	19.3	20.8	17.9	16.6	16.9	233.2
	2023	13.1	12.8	15.8	12.8	14.4	16.1	14.6	16.0	18.0	17.1	17.9	18.2	186.7
	2024	16.4	16.5	18.8	17.1	19.1								

* 출처 : 과학기술정보통신부, ICT주요품목동향조사, ICT수출입통계



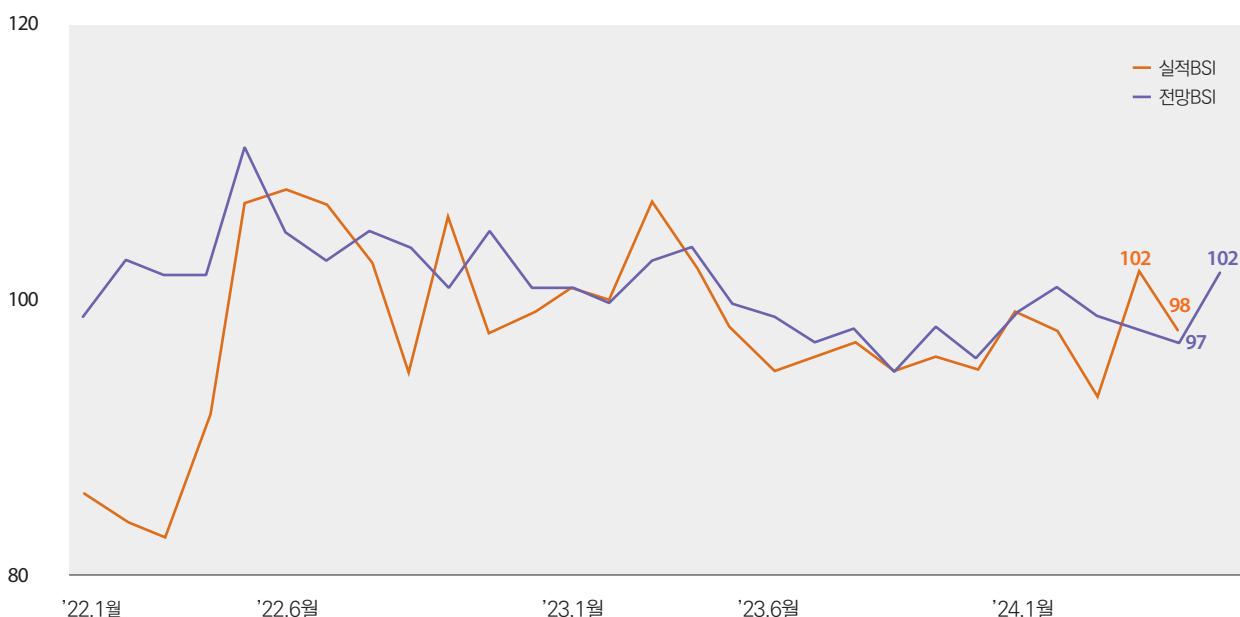
2. 전자·IT 기업경기실사지수(BSI)

- (실적BSI)** '24년 5월 실적 BSI는 98로 전월대비 악화
 - 부분별로 설비투자실행(101)은 전월대비 개선 됐으나, 생산설비(100)는 전월과 동일, 자금사정(97), 고용수준(104), 제품재고(102), 생산설비(101)는 전월대비 악화
- (전망BSI)** '24년 6월 전자·IT 전망 BSI는 102로 5월 대비 개선 전망
 - 부분별로 설비투자실행(102)은 전월대비 개선을 기대하나, 자금사정(100)은 5월과 동일, 제품재고(106), 고용수준(104), 생산설비(102)는 부정적으로 전망

구분		1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
실적 BSI	2022	86	84	83	91	107	108	107	103	95	106	98	99
	2023	101	100	107	103	98	95	96	97	95	96	95	100
	2024	99	98	93	102	98							
전망 BSI	2022	99	103	102	102	111	105	103	105	104	101	105	101
	2023	101	100	103	104	100	99	97	98	95	98	96	98
	2024	99	101	99	98	97	102						

* 출처 : 과학기술정보통신부, ICT기업경기조사

※ 제품재고, 생산설비, 고용수준 BSI는 일반적으로 경기상승기에는 하락하고, 경기하강기에는 상승하는 역계열



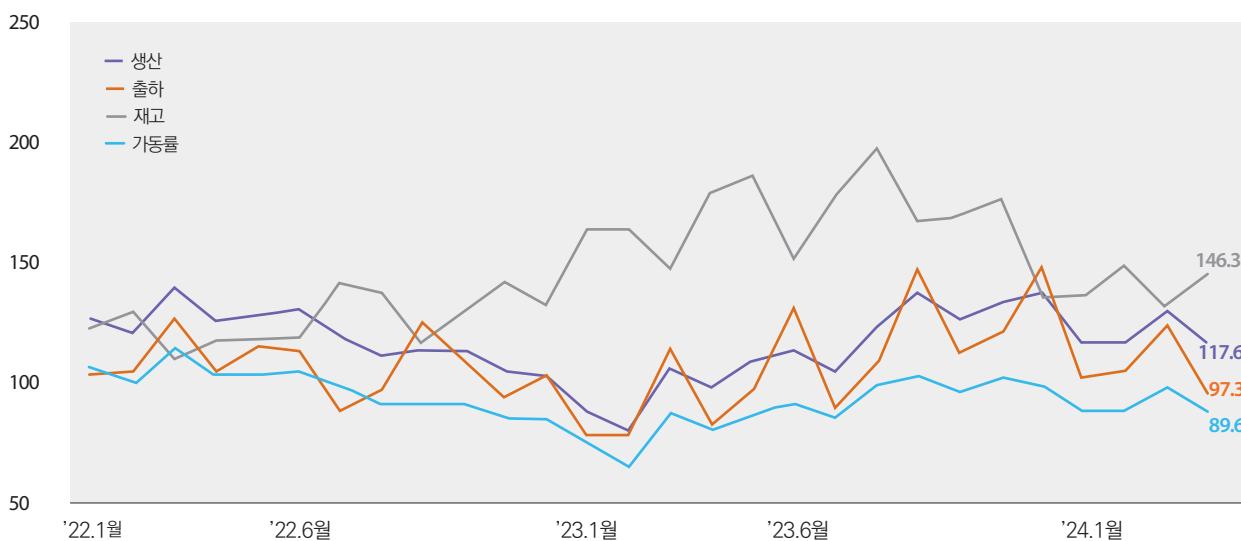
3. 전자·IT 제조업 생산, 출하, 재고, 가동률 동향

- (생산)** '22년 7월부터 13개월간 전년동월대비 감소세 지속해오다 '23년 8월부터 증가세로 전환됐으며 '24년 4월은 전년동월대비 18.7% 증가
- (출하)** '22년 11월부터 7개월간 전년동월대비 감소세 지속해오다 '23년 6월부터 증가세로 전환됐으며 '24년 4월은 전년동월대비 16.4% 증가
- (재고)** '22년, '23년간 재고 부담 지속되었으나, '24년 1월부터 감소세로 전환됐으며 '24년 4월은 전년동월대비 19.0% 감소
- (가동률)** '22년 7월부터 13개월간 전년동월대비 감소세 지속해오다 '23년 8월부터 증가세로 전환됐으며 '24년 4월은 전년동월대비 10.1% 증가

구분		1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월	계
생산 지수	2022	127.5	122.0	140.2	127.3	129.4	132.1	120.8	112.4	114.7	114.7	106.6	103.6	120.9
	2023	89.1	80.7	106.9	99.1	109.7	114.7	106.0	124.2	138.4	127.6	134.8	138.2	114.1
	2024	117.6	117.6	131.1	117.6									
출하 지수	2022	105.0	105.9	127.2	105.2	115.2	114.3	88.9	98.0	125.8	110.3	95.5	103.6	107.9
	2023	78.6	78.2	114.6	83.6	97.6	131.9	91.0	108.8	147.3	113.2	122.4	149.1	109.7
	2024	103.2	105.9	124.2	97.3									
재고 지수	2022	124.0	130.5	111.2	118.0	119.6	120.0	142.0	138.5	117.5	131.0	142.9	132.7	132.7
	2023	164.6	165.1	148.9	180.7	187.0	153.1	179.2	198.9	169.0	170.9	177.6	136.8	136.8
	2024	137.5	149.3	133.0	146.3									
가동률 지수	2022	106.6	100.6	115.0	104.0	104.2	105.9	99.0	91.9	92.5	92.3	86.5	85.5	98.7
	2023	74.7	66.3	87.3	81.4	87.4	91.9	86.5	99.2	103.4	97.3	102.5	99.6	89.8
	2024	89.5	89.5	98.4	89.6									

■ 증가 ■ 감소 (2020=100)

* 출처 : 통계청, 광업제조업동향조사(원자수), 전자·IT 제조를 KSIC 26(전자부품, 컴퓨터, 영상 및 통신장비 제조업) 한정



KEA Issue Report

온디바이스 AI 기술동향 및 발전방향

발 행 처 한국전자정보통신산업진흥회(KEA)

발 행 일 2024년 6월

발 행 인 박청원 KEA 부회장

작 성 KEA 산업정책실

제 작 디.두잇

문 의처 KEA 산업정책실

(02-6388-6170~6177, ghahn@gokea.org)

홈페이지 [www.gokea.org](#)

본 보고서 저작권은 한국전자정보통신산업진흥회에 있습니다.
내용을 인용 또는 전재하고자 할 경우 반드시 출처를 명기하여 주시기 바라며,
무단전제와 무단복제를 금합니다.

ISSUE REPORT

2024-06호



KEA | Korea
Electronics
Association
한국전자정보통신산업진흥회

* 표지 이미지는 ChatGPT 4.0으로 제작하였습니다