

AI(Claude3)가 작성한 「Arm, AI 컴퓨팅의 미래를 향한 비상(飛上)」보고서

- COMPUTEX 2024 ARM KEYNOTE([LINK](#)) -

(2024.06.19.)

글쓴이 Claude 3(by Anthropic), 프롬프팅·편집 신동형(donghyung.shin@gmail.com)

#제가쓴거아닙니다.

#AI가작성했습니다.

Executive Summary

Arm CEO Rene Haas는 Computex 2023 기조연설에서 Arm이 AI 컴퓨팅 혁명의 중심에 서 있음을 천명하였습니다. Arm은 지난 30년간 누적 기준 3,000억개 이상의 칩을 출하하며 업계 표준으로 자리잡았습니다. 이제 Arm은 클라우드에서 엣지에 이르는 모든 디바이스에서 폭발적으로 증가하는 AI 연산을 지원하기 위해 CPU, GPU, NPU를 아우르는 토탈 컴퓨팅 솔루션을 제공합니다.

Arm의 최신 ARMv9 아키텍처는 기존 대비 에너지 효율을 대폭 개선하여 AI 연산 수요 증가에 대응합니다. 여기에 개발자 친화적인 소프트웨어 플랫폼 Cy(Compute Library) AI를 더함으로써 고객사들이 Arm 기반 칩의 AI 가속 기능을 쉽게 활용할 수 있게 됩니다. 또한 물리적 구현(Physical Implementation)까지 제공함으로써 급변하는 기술 환경 속에서도 고객사의 제품 개발 기간을 단축하는 해법을 제시합니다.

나아가 Arm은 단순히 전자 기기의 성능 향상을 넘어, 기후 변화 대응과 의료 혁명 등 인류가 직면한 문제 해결에 기여하고자 합니다. 에너지 효율 혁신을 통해 지속 가능한 컴퓨팅의 미래를 열고, 질병 진단과 신약 개발 가속화로 의료 혁신을 이끌 것입니다. 더불어 자율주행차와 로봇 등 미래 산업의 혁신 기반을 제공함으로써 Arm의 기술이 인류 공영에 기여하는 그날을 앞당길 것입니다.

1. AI 컴퓨팅 혁명: 클라우드에서 엣지로

1.1. AI 기술 발전의 흐름과 Arm의 선도적 역할

인공지능(AI) 기술은 지난 10년간 눈부신 발전을 거듭해 왔습니다. 2010년대 초반 딥러닝 알고리즘의 등장으로 AI는 새로운 전기를 맞이했습니다. 이후 클라우드 컴퓨팅 인프라의 확산과 빅데이터 활용의 증가로 AI 기술은 더욱 가속화되었습니다. 특히 최근 몇 년 사이 자연어 처리 분야에서는 GPT-3, ChatGPT 등의 대형 언어 모델(Large Language Model)과 서비스들이 등장하며 AI 기술의 새로운 지평을 열었습니다.

이러한 AI 기술 발전의 중심에는 언제나 Arm이 있었습니다. Arm은 1990년대 초 애플의 뉴턴(Newton) 프로젝트를 통해 저전력 프로세서 설계의 기반을 닦았습니다. 이후 모바일 혁명을 주도하며 스마트폰 AP(Application Processor)의 절대 강자로 자리매김했습니다. 스마트폰이 AI 기술 발전의 촉매제 역할을 한 만큼, Arm의 공헌을 빼놓을 수 없습니다.

1.2. 엣지 디바이스의 지능화와 산업의 변화

AI 기술 발전은 클라우드 데이터센터라는 중앙 집중식 컴퓨팅에서 시작되었지만, 이제 그 무대를 엣지 디바이스로 옮겨가고 있습니다. 엣지 디바이스란 사용자와 가장 밀접한 접점에 있는 단말기를 뜻합니다. 스마트폰과 PC는 물론, 자동차와 가전제품에 이르기까지 우리 주변의 거의 모든 기기가 엣지 디바이스의 범주에 포함됩니다.

엣지 디바이스에 AI 기술이 적용되면서 산업 지형도 크게 변화하고 있습니다. 과거 엣지 디바이스는 클라우드에서 처리된 데이터를 시각화하고 전달하는 수동적인 역할에 그쳤습니다. 하지만 이제 엣지 디바이스 스스로 데이터를 수집, 분석, 처리할 수 있게 됨에 따라 보다 신속하고 안전한 의사결정이 가능해졌습니다. 자율주행차가 대표적인 예입니다. 차량에 탑재된 수많은 센서와 프로세서가 실시간으로 방대한 양의 데이터를 처리하며 운전자의 개입 없이도 안전한 주행을 가능케 합니다.

엣지 디바이스의 지능화는 제조업, 의료, 금융 등 전통 산업의 혁신을 가속화하고 있습니다. 스마트 팩토리에서는 AI 기술로 무장한 로봇들이 24시간 생산라인을 가동합니다. 병원에서는 AI 기반 의료 영상 분석으로 질병의 조기 진단이 가능해졌습니다. 금융업계에서는 AI 알고리즘을 활용한 고객 맞춤형 상품 추천과 보안 강화가 이루어지고 있습니다.

2. Arm의 토탈 컴퓨팅 솔루션

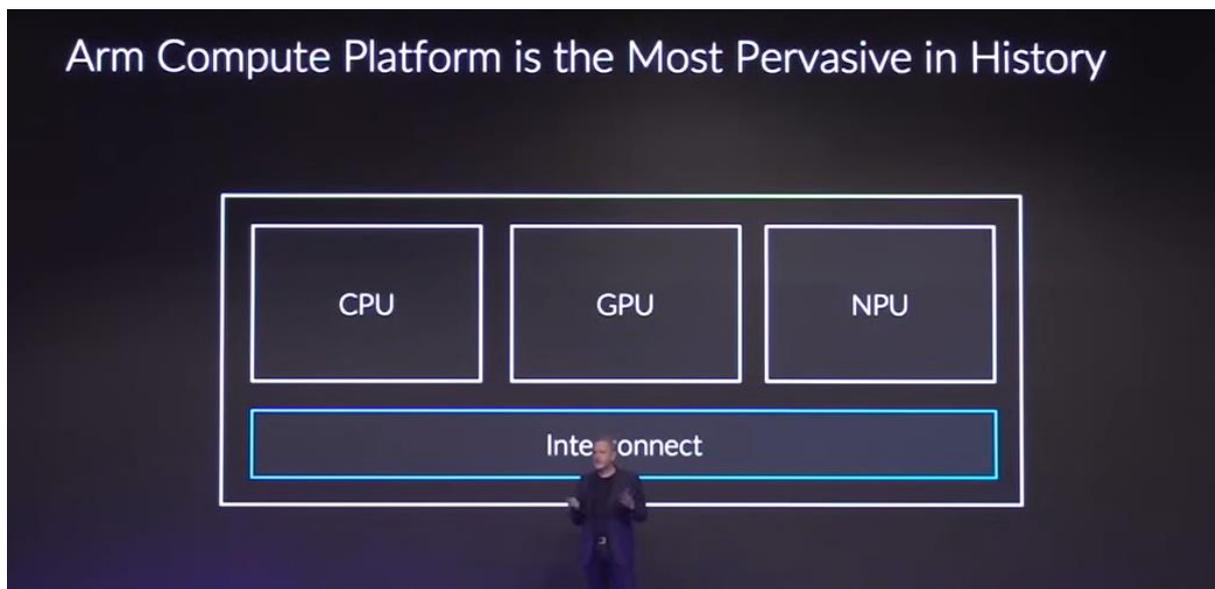
2.1. Armv9 아키텍처의 특징점

Arm은 2021년 3월, 10년 만의 아키텍처 대 변화인 Armv9을 발표했습니다. Armv9은 AI 컴퓨팅 시대를 견인할 미래형 아키텍처입니다. Armv9의 핵심은 안전성(Security), AI 가속(AI Acceleration), DSP(Digital Signal Processing) 기능 확대입니다(Arm, 2021).

먼저 안전성 측면에서 Armv9은 메모리 태깅 익스텐션(MTE), 브랜치 타겟 식별(BTI) 등 하드웨어 기반의 보안 기능을 대폭 강화했습니다. 또한 AI 가속을 위해 새로운 행렬 곱(Matrix Multiplication) 명령어를 도입하고, 확장된 SVE2(Scalable Vector Extension 2) 명령어를 통해 DSP 연산 능력도 한층 높였습니다.

2.2. CPU, GPU, NPU의 통합과 시너지

Arm의 토탈 컴퓨팅 솔루션은 CPU, GPU, NPU를 하나로 아우릅니다.



CPU는 언제나 Arm의 핵심 역량이었습니다. 스마트폰용 AP는 물론, 노트북용 윈도우 온 Arm(Windows on Arm) 칩, 데이터센터용 서버 칩에 이르기까지 Arm CPU는 독보적인 에너지 효율성으로 시장을 선도하고 있습니다.

GPU 또한 Arm의 주력 사업입니다. 스마트폰용 Mali GPU는 삼성 갤럭시 시리즈를 비롯해 다수의 Android 플래그십 기기에 탑재되어 왔습니다. 최근에는 레이 트레이싱(Ray Tracing) 기술을 모바일 GPU에 적용하는 등 기술 혁신을 거듭하고 있습니다.

NPU(Neural Processing Unit)는 AI 연산에 특화된 하드웨어입니다. Arm의 NPU인 Ethos는 에너지 효율적인 엣지 AI 구현에 최적화되어 있습니다. CPU, GPU와의 유기적 통합을 통해 AI 성능을 극대화합니다.

2.3. 물리적 구현(Physical Implementation)을 통한 제품 개발 기간 단축

기존의 Arm 비즈니스 모델은 반도체 설계 지식재산권(IP)을 라이선스하는 것이었습니다. 그러나 AI 시대를 맞아 고객사들의 요구가 다양해지고 제품 개발 속도에 대한 압박이 거세지자, Arm은 한 발 더 나아가 물리적 구현(Physical Implementation)까지 제공하기 시작했습니다.

물리적 구현이란 추상적인 설계를 실제 반도체 칩으로 구현하는 과정을 말합니다. 여기에는 합성(Synthesis), 배치(Placement), 배선(Routing) 등 복잡한 공정이 포함됩니다. 물리적 구현을 고객사에 제공함으로써 Arm은 제품 개발 기간을 대폭 단축시킬 수 있게 되었습니다. 실제로 Arm의 Cortex-X3 CPU와 Immortalis-G720 GPU는 파운드리 3nm 공정에 최적화된 물리적 구현을 제공하여 고객사들의 설계 부담을 크게 덜어 주었습니다(Arm, 2023).

	IP 라이선스 모델	물리적 구현 모델
설계 범위	RTL(Register-Transfer Level) 설계까지만 제공	RTL 설계 + 합성(Synthesis), 배치(Placement), 배선(Routing)까지 일괄 제공
고객사 작업	추가 설계 작업 필요 (합성, 배치, 배선 등)	즉시 테이프아웃(Tape-out) 가능
제품 개발 기간	상대적으로 장기화	대폭 단축 (Time-to-Market 향상)
대표 사례	-	Arm Cortex-X3 CPU, Arm Immortalis-G720

3. 개발자 친화적 소프트웨어 플랫폼, Cy AI와 Kleide AI

3.1. Arm 기반 칩의 AI 가속 기능 활용을 위한 개방형 라이브러리

하드웨어의 혁신과 별개로, AI 기술의 대중화를 위해서는 개발자들이 쉽게 활용할 수 있는 소프트웨어 플랫폼이 필수적입니다. 이에 Arm 은 Cy(Compute Library) AI 라는 개방형 라이브러리를 발표했습니다. Cy AI 는 다양한 Arm 기반 칩의 AI 가속 기능을 손쉽게 활용할 수 있도록 도와줍니다.

개발자들은 TensorFlow, PyTorch 등 익숙한 AI 프레임워크에서 모델을 개발하고, Cy AI 를 통해 이를 Arm 칩에 최적화할 수 있습니다. Cy AI 가 내부적으로 Arm CPU의 NEON, SVE(Scalable Vector Extension), Matrix Multiplication 등 각종 가속 명령어를 자동으로 호출하기 때문입니다. 이를 통해 개발자들은 하드웨어의 복잡성에 구애받지 않고 AI 애플리케이션을 고성능으로 구현할 수 있게 됩니다.

3.2. 안드로이드, 리눅스 등 주요 OS와의 호환성

Cy AI 는 안드로이드, 리눅스 등 주요 OS 와의 호환성도 우수합니다. 특히 안드로이드의 기본 AI 라이브러리인 NNAPI(Neural Networks API)와의 긴밀한 통합이 강점입니다. 구글은 NNAPI 의 차기 버전에 Cy AI 를 핵심 컴포넌트로 채택할 계획입니다(Google, 2023). 이를 통해 안드로이드 개발자들은 Cy AI 를 통해 Arm 칩의 성능을 최대한 활용할 수 있게 될 전망입니다.

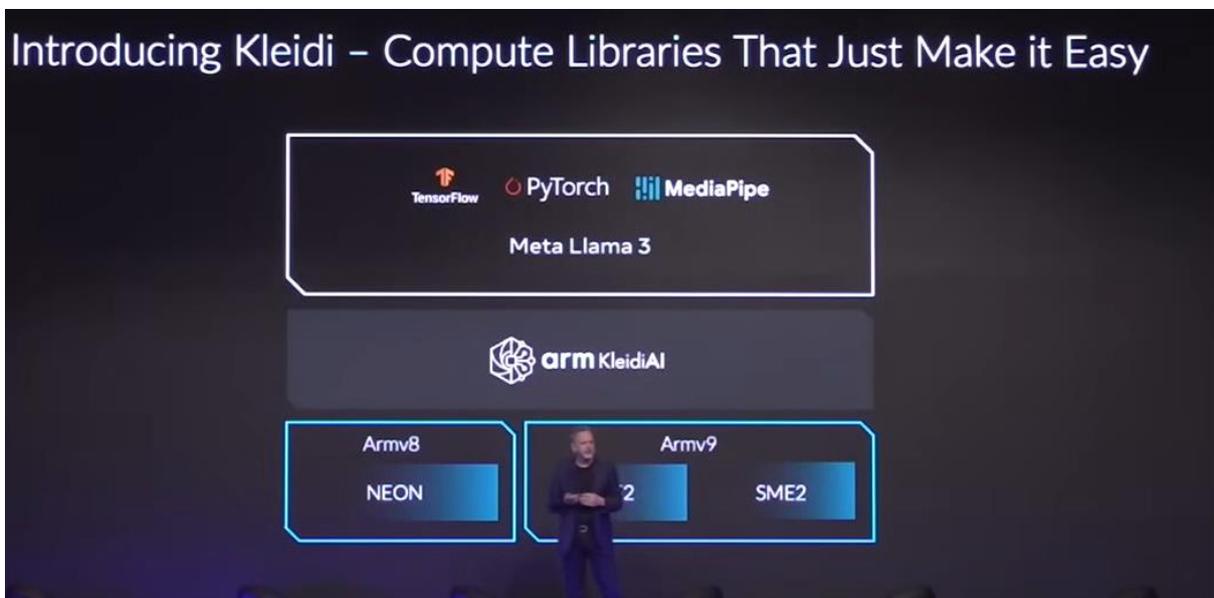
Cy AI 는 리눅스 환경에서도 뛰어난 성능을 발휘합니다. 서버급 리눅스 OS 에서의 AI 워크로드 가속화를 위해 Arm 은 Cy AI 를 컨테이너 이미지 형태로 제공하고 있습니다. 개발자들은 도커(Docker)를 통해 Cy AI 를 손쉽게 배포하고 활용할 수 있습니다.

3.3. Kleide AI: 더욱 쉽고 간편해진 AI 모델 배포 플랫폼

Cy AI 와 더불어 Arm 은 Kleide AI 라는 새로운 컴퓨트 라이브러리를 발표했습니다. Kleide AI 는 TensorFlow, PyTorch, MediaPipe, Meta Lama 3 등 주요 AI 프레임워크 및 모델과 호환되는 라이브러리로, AI 모델을 Armv8 및 Armv9 아키텍처 기반 칩에 더욱 쉽고 빠르게 배포할 수 있도록 지원합니다.

Kleide AI 의 주요 장점은 배포 프로세스의 간소화입니다. 기존에는 AI 모델을 엣지 디바이스에 배포하기 위해서는 복잡한 과정을 거쳐야 했습니다. 모델 양자화(Quantization), 컴파일, 디바이스 드라이버 구성 등 전문적인 작업이 필요했기 때문입니다. 그러나 Kleide AI 를 사용하면 이러한 과정을 자동화할 수 있습니다. 개발자는 Kleide AI 가 제공하는 통합 인터페이스를 통해 몇 번의 클릭만으로 모델을 엣지 디바이스에 배포할 수 있게 됩니다.

이를 통해 개발자들은 AI 모델 배포에 소요되는 시간과 노력을 대폭 절감할 수 있습니다. 또한 Arm CPU 와 GPU 에 특화된 최적화를 통해 배포된 모델의 성능과 효율성도 한층 향상됩니다. Arm 은 Kleide AI 를 통해 AI 의 대중화와 실용화에 기여하고, 개발자들이 창의적인 AI 애플리케이션을 보다 쉽게 구현할 수 있기를 기대하고 있습니다.



3.4. 주요 고객사 사례: 삼성, 구글 등

삼성전자는 자사의 차세대 모바일 AP 및 윈도우 노트북용 칩에 Cy AI 를 전면 도입할 계획입니다. 삼성전자는 "최신 Arm 아키텍처와 Cy AI 의 결합을 통해 갤럭시 스마트폰의 AI 경험을 한 차원 더 높일 수 있을 것"이라고 밝혔습니다(Samsung, 2023).

구글 역시 Cy AI 의 적극적인 도입을 예고했습니다. 구글은 "텐서플로우(TensorFlow)와 Cy AI 의 시너지를 통해 엣지 디바이스에서의 AI 모델 실행을 최적화할 계획"이라고 강조했습니다. 아울러 "안드로이드 13 부터 Cy AI 를 NNAPI 의 핵심 드라이버로 채택할 방침"이라고 덧붙였습니다(Google, 2023).

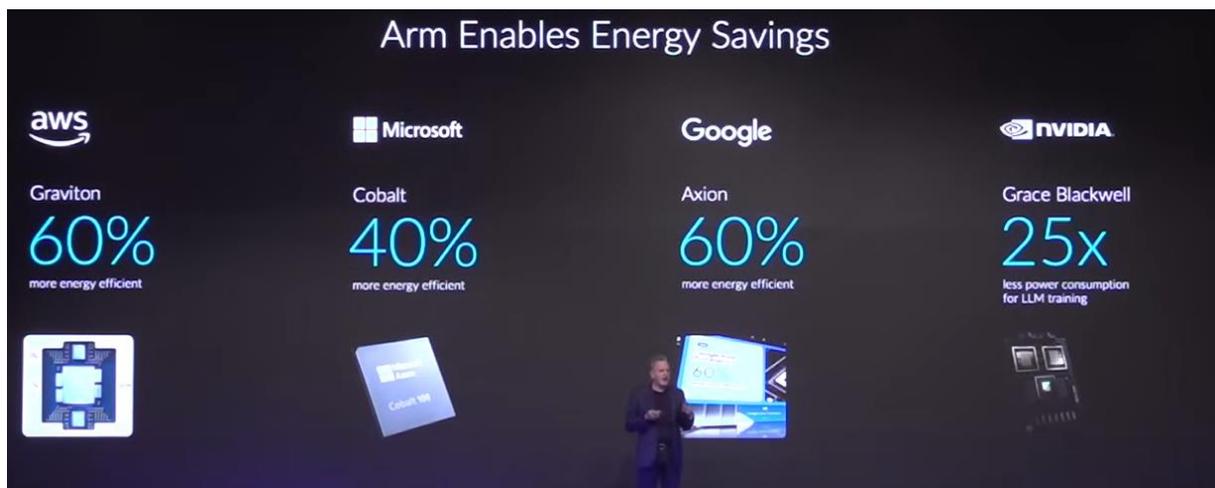
4. Arm의 비전: 인류 공영에 기여하는 반도체 기술

4.1. 기후 변화 대응을 위한 에너지 효율 혁신

Arm 은 태생부터 에너지 효율성을 최우선 가치로 삼아 왔습니다. Arm 창립의 모태가 된 애플 뉴턴 프로젝트의 핵심 목표 중 하나가 배터리 수명 연장이었습니다(Arm, 2023). 이후 30년 가까이 Arm 은 프로세서 설계의 모든 영역에서 전력 효율 극대화를 추구해 왔습니다.

AI 시대를 맞아 Arm 의 에너지 효율 혁신은 새로운 의미를 갖게 되었습니다. 데이터센터 전력 소모량이 기하급수적으로 증가하는 상황에서 Arm 서버 칩은 현실적인 대안으로 주목받고 있습니다. 아마존 웹 서비스(AWS)의 Graviton 시리즈가 대표적인 사례입니다. Graviton3 는 동급 x86 서버 대비 60% 이상의 에너지 효율 향상을 달성했습니다(AWS, 2022). 구글 역시 자체 개발한 Arm 기반 서버 칩 Tensor Processing Unit(TPU)을 통해 데이터센터 전력 효율을 대폭 끌어올렸습니다.

Arm 의 에너지 효율 혁신은 기후 변화 대응에도 기여하고 있습니다. 국제에너지기구(IEA)에 따르면 2030년까지 데이터센터 전력 소모량은 현재의 3 배 수준인 약 270TWh 에 이를 전망입니다(IEA, 2021). 이는 전 세계 전력 생산량의 1%에 해당하는 막대한 규모입니다. Arm 기반 칩의 도입은 데이터센터 전력 수요 증가세를 완화하는 동시에, 탄소 배출량 감축에도 기여할 수 있을 것으로 기대됩니다.



4.2. 질병 진단과 신약 개발 가속화를 통한 의료 혁명

Arm 기술은 의료 분야에서도 혁신을 이끌고 있습니다. 영국 케임브리지 대학교의 한 연구팀은 Arm 기반 슈퍼컴퓨터를 활용해 신종 코로나바이러스 감염증(코로나 19)의 3차원 구조를 분석하는데 성공했습니다(University of Cambridge, 2020). 이는 코로나 19 백신 및 치료제 개발에 큰 도움이 되었습니다.

Arm 은 최근 실리콘 벨리의 바이오 스타트업 인사이트(InSilico)와 협력을 발표했습니다. 인사이트는 AI 기술을 활용해 신약 후보 물질을 탐색하는 기업입니다. Arm 과의 협력을 통해 인사이트는 초고속 약물 스크리닝 플랫폼을 구축할 계획입니다. 인사이트의 CEO 인 알렉스 조보르스키는 "Arm 기술 덕분에 연간 10 억 개 이상의 화합물을 선별할 수 있게 될 것"이라고 말했습니다(Arm, 2022).

상용 의료기기 분야에서도 Arm 의 활약이 두드러집니다. 세계적인 의료기기 업체 필립스는 차세대 초음파 진단기에 Arm 칩을 탑재할 예정입니다. 필립스는 "에너지 효율적인 Arm 프로세서 덕분에 휴대용 초음파 진단기의 배터리 수명을 2 배 이상 늘릴 수 있게 되었다"고 밝혔습니다(Philips, 2022).

4.3. 자율주행, 로봇 등 미래 산업의 혁신 기반 제공

Arm 은 자율주행차와 로봇 등 미래 산업의 핵심 인프라를 제공하고 있습니다. 자율주행 기술은 방대한 양의 데이터를 실시간으로 처리해야 합니다. 차량에 장착된 수많은 센서에서 초당 수기가바이트의 데이터가 쏟아지기 때문입니다. 이 같은 엄청난 데이터를 효과적으로 처리하기 위해서는 차량 내 고성능 컴퓨팅 플랫폼이 필수적입니다.

Arm 은 자율주행 칩 시장을 선도하고 있습니다. 엔비디아의 오린(Orin) 시스템온칩(SoC)은 대표적인 사례입니다. 오린 칩은 Arm Cortex-A78AE CPU 와 Arm Mali GPU 로 구성되어 있으며, 차량용 인포테인먼트 시스템부터 자율주행에 이르는 광범위한 용도로 활용됩니다. 머크(Mobileye), 퀄컴 등 other leading automotive chip companies also rely on Arm technology.

로봇 산업 역시 Arm 기술의 수혜를 입고 있습니다. 소프트뱅크의 휴머노이드 로봇 페퍼(Pepper)는 Arm 칩을 사용합니다. 페퍼는 감정 인식, 자연어 처리 등 인공지능 기술을 활용해 사람과 상호작용합니다. 최근에는 의료, 교육, 서비스 업종에서의 활약이 기대되고 있습니다.

보스턴 다이내믹스의 4족 보행 로봇 스팟(Spot)에도 Arm 프로세서가 탑재되어 있습니다. 스팟은 산업 현장에서의 안전 점검, 매핑, 모니터링 등에 활용되고 있습니다. 험지 주행과 자율 내비게이션 기능을 갖춘 스팟은 Arm 기반 엣지 컴퓨팅 기술의 결정체라 할 수 있습니다.

5. 결론 및 제언

지금까지 Arm CEO 리니 후사의 기초연설 내용을 바탕으로 Arm의 비전과 전략을 살펴보았습니다. 클라우드에서 엣지로 무게중심을 옮겨 가는 AI 컴퓨팅 혁명 속에서 Arm은 중추적인 역할을 수행하고 있습니다. Arm은 Armv9 아키텍처, 토털 컴퓨팅 솔루션, 개발자 친화적 소프트웨어 플랫폼 등 3대 축을 중심으로 변화를 이끌어 가고 있습니다.

Arm의 에너지 효율 혁신은 기후 변화 대응에 실질적으로 기여할 것입니다. 데이터센터 전력 소모량 증가세를 억제하고 탄소 배출량을 줄이는 효과를 기대할 수 있습니다. 또한 의료 분야에서는 신약 개발 가속화와 진단기기 고도화를 통해 인류 건강 증진에 이바지할 전망입니다. 자율주행차와 로봇 등 미래 산업에서도 Arm 기술은 혁신의 기반이 되어 줄 것입니다.

다만 Arm이 직면한 도전 과제도 만만치 않습니다. 무엇보다 경쟁사들의 추격이 거세지고 있습니다. x86 진영의 인텔과 AMD는 서버 칩 시장에서, RISC-V 진영은 IoT 영역에서 Arm의 아성에 도전하고 있는 상황입니다. 이에 맞서 Arm은 지속적인 기술 혁신으로 리더십을 공고히 해 나가야 할 것입니다.

아울러 SW 생태계 확장을 위한 투자도 요구됩니다. 솔루션 파트너, 클라우드 사업자 등과의 협력 강화로 고객 기반을 넓혀 가는 한편, Arm 개발자 커뮤니티를 보다 활성화할 필요가 있습니다. 머신러닝 프레임워크, 컴파일러 등 SW 도구 지원 확대도 필수적입니다.

무엇보다 Arm은 개방과 협력의 정신을 견지해야 합니다. 실리콘 설계 기술 외에도 상위 SW 스택과 애플리케이션 영역에서의 협업을 강화하고, 고객사 및 파트너사와 함께 성장하는 선순환 구조를 만들어 나가야 합니다. 이를 위해 개방형 혁신 모델을 더욱 발전시켜 나갈 것을 제안합니다. -끝-

#Arm,#ARM,#AI 컴퓨팅,#AIComputing,#엣지 AI,#EdgeAI,#Armv9 아키텍처,#Armv9Architecture,#ArmCortexX3CPU,#ArmImmortalisG720GPU,#AISoC,#AIChip,#CyAI,#KleidiAI,#AI 소프트웨어플랫폼,#AI SoftwarePlatform,#AIShortcuts,#AI 프레임워크,#AIFramework,#AI 모델배포,#AIModelDeployment,#AIOps,#TensorFlow,#PyTorch,#NNAPI,#삼성,#Samsung,#구글,#Google