

AI 정의 차량(ADV) 기반의 자율주행차용 AI반도체: 기술·시장 전망 및 전략적 시사점

AI Semiconductors for AI-Defined Vehicles: Technology Trends, Market Outlook, and Strategic Implications

송근혜 (G.H. Song, ghsong@etri.re.kr)

구본태 (B.T. Koo, koobt@etri.re.kr)

기술전략연구센터 선임연구원

지능형반도체연구본부 책임연구원/본부장

ABSTRACT

With advances in autonomous driving technology, artificial intelligence (AI) has emerged as a core component beyond mere auxiliary functions. However, the strategic and technical implications of AI semiconductors across the vehicle-edge-cloud continuum remain underexplored. Our findings reveal divergent design strategies across domains and a growing emphasis on edge-cloud synergy. This study explores the concept of an AI-defined vehicle (ADV), which enhances the role of AI in autonomous driving by integrating on-vehicle computation with edge cloud and data center AI processing. We reviewed the development of AI semiconductors tailored for autonomous vehicles, focusing on architecture, computing demands, and the evolving strategies of leading global and domestic companies. Based on these insights, we propose a structured outlook for the AI semiconductor market, segmented by vehicle, edge cloud, and data center applications. This study aims to provide strategic implications for strengthening competitiveness and establishing an integrated roadmap for AI-enabled autonomous vehicles.

KEYWORDS AI-Defined Vehicle, AI semiconductors, Autonomous vehicles, Physical AI

I. 서론

자율주행 기술이 고도화됨에 따라 AI(Artificial Intelligence)는 단순 보조 기능을 넘어 자율주행 시스템을 제어하는 핵심 기술로 부상하고 있다[1]. 자율

주행 시스템은 복잡한 주행 환경을 실시간으로 인식하고 최적의 주행 판단을 내릴 수 있어야 한다. 이 과정에서 차량 내·외부에서 대규모 데이터를 고속으로 처리할 수 있는 고성능 AI 연산 능력이 필수적으로 요구된다.

* DOI: <https://doi.org/10.22648/ETRI.2025.J.400402>

* 본 연구는 한국전자통신연구원 기본사업의 일환에서 수행되었음[25ZF1130, ICT 국가기술전략 정책연구].



차량의 컴퓨팅 구조는 대규모 AI 연산 능력을 개선하기 위해서 기능별로 전자제어장치(ECU: Electronic Control Unit)를 갖는 분산형 아키텍처에서 ECU를 통합하거나 고성능 중앙 컴퓨터(Zone Controller)가 차량 전체 기능을 통합적으로 제어하는 중앙 집중형 아키텍처로 전환되고 있다. 또한, 차량 시스템 전반을 소프트웨어로 제어하고 OTA(Over-the-Air) 업데이트를 통해 기능을 원격으로 개선하는 방식이 확산되고 있다[2].

현재 자율주행차는 SDV(Software-Defined Vehicle) 개념을 중심으로, 차량의 주요 기능을 소프트웨어로 구현하고 제어하는 분산형 구조를 채택하고 있다. SDV는 기능별 ECU를 가지고 있으며, OTA 방식을 통해 각 기능을 개별적으로 유지·보수한다는 점이 특징이다. 그러나 이러한 구조는 주로 기능 단위의 분산 제어에 초점을 두고 있어, 고성능 AI 연산이나 실시간 학습과 같은 고도화된 자율주행 요구사항을 통합적으로 처리하는 데 한계가 있다는 평가를 받아왔다[3].

이러한 한계를 극복하기 위해 AI를 중심에 둔 중앙집중형 연산 아키텍처를 적용한 ADV(AI-Defined Vehicle) 개념이 최근 등장하기 시작했다[4]. ADV는 AI를 자율주행 아키텍처의 핵심축으로 삼으며, 기존에 차량 내부에 국한됐던 연산 구조를 엣지 클라우드와 데이터센터까지 확장하는 E2E(End-to-End) AI 컴퓨팅 아키텍처를 지향한다. 이는 각 계층에 적합한 AI반도체를 배치해 연산 효율성과 확장성을 높이는 방식으로, 흔히 ‘3-Tier 구조’로도 불린다. 이러한 자율주행 아키텍처의 변화 속에서 AI반도체는 자율주행 기능 실행에 직접적으로 관여하는 핵심 구성요소로 인식되며, 시스템의 성능과 지능 수준을 좌우하는 핵심 기반 기술로 평가받고 있다.

본고에서는 ADV 개념의 도입과 함께 자율주행 아키텍처의 변화가 AI반도체 수요와 활용 방식

에 미치는 영향을 살펴보고, 글로벌 주요 기업들이 ADV 흐름을 선점하기 위해 추진 중인 기술 전략과 반도체 적용 사례를 검토하고자 한다. 이를 바탕으로 ADV 발전 방향과 자동차-클라우드 연계 관점에서 자율주행차용 AI반도체 시장의 성장 가능성을 전망하고 정책적 시사점을 제시하고자 한다.

II. 자율주행 아키텍처 전환과 AI반도체 수요의 증가

자율주행 시스템은 고도화될수록 차량 내부에 기능별로 분산되어 있던 연산 구조에서 벗어나 AI를 중심으로 한 통합형 연산 구조로 진화하고 있다. 최근에는 차량 내부뿐만 아니라 엣지 클라우드와 데이터센터까지 연산 자원을 연계하는 E2E 아키텍처가 주목받으며, ADV가 자율주행 시스템의 새로운 표준으로 부상하고 있다. 이는 단순히 AI 기술을 일부 기능에 적용하는 수준을 넘어, 자율주행 아키텍처 전반의 연산 흐름을 재편하려는 움직임으로 이어지고 있다. 이 과정에서 AI반도체는 자율주행 성능과 지능 수준을 좌우하는 핵심 인프라로 떠오르고 있다.

이러한 구조적 변화는 자율주행 기술 수준에 따라 ADV 1.0에서 ADV 2.0으로의 단계적 진화 흐름 속에서 확인되며, 그림 1은 ADV의 단계적 진화에 따른 연산 구조의 전환 과정을 시각적으로 보여준다.

ADV 1.0 단계(L1~L2 수준)는 SDV와 달리, AI를 초기부터 각 기능 수행의 핵심 연산자로 배치한다는 점에서 다르다. ADV 1.0 단계에서 AI는 주로 인식 및 예측 기능에 탑재되어 독립적인 연산을 수행한다. 기능 간 데이터 교환이 가능하지만, 기능별로 분산된 구조와 제한된 컴퓨팅 자원으로 인해 전체 시스템의 연산 최적화에는 한계가 있다. 이 단계에서는 APU(Accelerated Processing Unit), VPU(Vision Process-

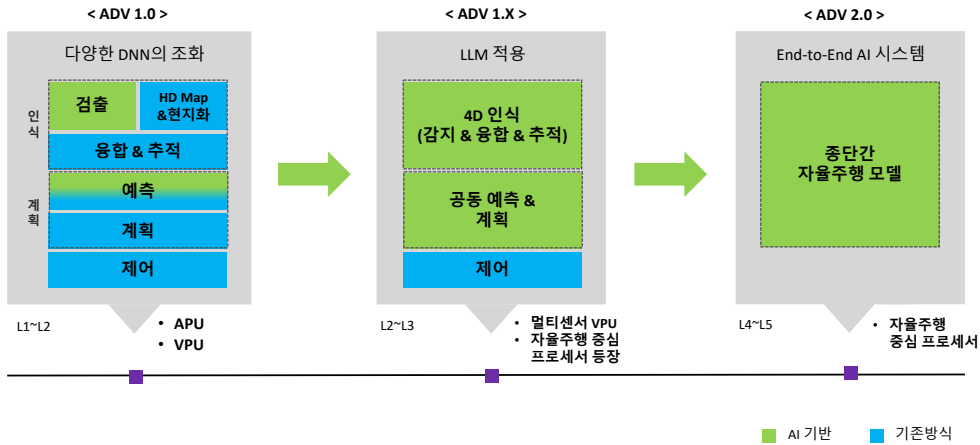


그림 1 ADV의 자율주행 아키텍처 진화

ing Unit) 등 기능별 전용 AI반도체가 주로 활용된다.

ADV 1.X 단계(L2~L3 수준)에서는 개별 모듈로 분산되었던 자율주행 기능이 ‘4D 인식-예측-제어’ 체계로 통합되면서, 연산 처리의 일관성과 기능 간 연계성이 한층 강화된다. 이 과정에서 일부 영역에는 트랜스포머 기반 AI모델이 도입되어, 3차원 공간과 시간 정보를 반영한 객체 움직임 예측이 가능해지고, 자율주행의 정밀도 또한 향상된다. 트랜스포머 모델은 대규모 시공간 데이터를 병렬적으로 처리하는 특성상 연산량이 급격히 증가하므로, 이를 효과적으로 처리하기 위한 고성능 AI반도체 기반의 중앙집중형 연산 구조로의 전환 요구도 함께 증가한다.

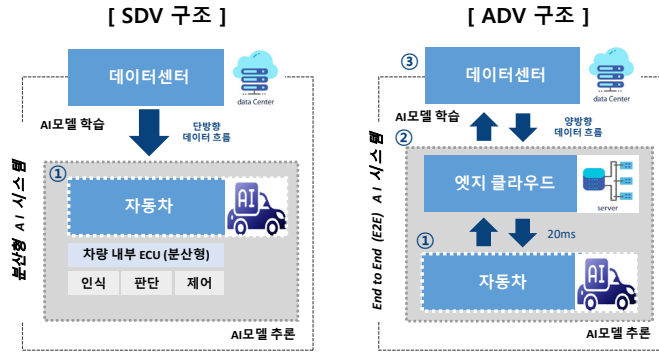
ADV 2.0 단계(L4~L5 수준)에서는 자율주행 시스템의 전 과정(센서 입력부터 차량 제어까지)을 단일 신경망으로 통합 처리하며, 생성형 AI 기반 시뮬레이션, 온디바이스 학습(On-Device Learning), 연속 학습(Continual Learning) 등이 핵심 기술로 도입된다. 고차원 AI 연산을 실시간으로 처리하기 위해서는 차량 내 연산 자원을 분산시키기보다는 고성능 AI반도체 기반의 중앙집중형 연산 구조가 보다 효과적이다. 실시간 AI모델 업데이트가 요구되는 이 단계에서, AI반도체는 학습과 추론을 동시에 수행하고 연속적

인 최적화 기능을 수행할 수 있는 성능을 지원해야 한다.

ADV 2.0 단계에서는 자율주행 AI모델의 학습과 배포가 차량과 데이터센터 간 유기적으로 연결된 구조에서 이루어진다. 차량이 주행 중 수집한 데이터는 데이터센터로 전송되어, AI모델의 학습-검증-배포 과정에 순환적으로 활용된다. 따라서, ADV 2.0 단계에서 자율주행 AI모델은 자율주행 차량 운용에 따라 지속적으로 고도화되는 특징이 있다.

구체적으로, 데이터센터에서는 대규모 데이터를 기반으로 자율주행용 AI모델을 학습하고, 다양한 주행 시나리오를 시뮬레이션하여 성능을 검증 및 최적화한 뒤, 이를 OTA를 통해 차량에 배포한다. 차량에서는 주변 환경과 운전자 특성 등에 대한 데이터를 실시간으로 수집하고, 온디바이스 학습과 연속 학습을 통해 AI모델을 해당 환경에 맞게 보정(Fine-Tuning)한다.

이렇게 보정된 모델은 돌발 변수와 새로운 주행 상황에 빠르게 적응하고, 성능 향상을 지속적으로 달성할 수 있다. 이후 차량에서 축적된 데이터는 다시 데이터센터로 전송되어 추가 학습에 활용되며, 고도화된 모델은 OTA를 통해 차량에 재배포된다.



출처: 게티이미지뱅크를 일부 활용하여 저자 그림. 무단 전재 및 재배포 금지

그림 2 SDV와 ADV의 자율주행 AI 컴퓨팅 아키텍처 비교

이러한 루프 기반의 순환형 AI 학습 구조는 ADV 2.0의 핵심 구성으로, 실시간성, 적응성, 확장성을 동시에 충족시키는 기술적 토대를 제공한다.

ADV 아키텍처는 기존 SDV 구조와 연산 방식, 학습 체계, OTA 운용 측면에서 뚜렷한 차이를 보인다 (그림 2). SDV는 차량 내 다수의 ECU를 중심으로 기능별 연산을 분산 처리하며, 데이터센터에서 미리 학습된 AI모델을 단방향으로 배포받아 제한적인 연산 기능을 수행한다. 반면, ADV는 차량-엣지 클라우드-데이터센터를 긴밀하게 연계한 E2E 구조의 통합형 학습·추론 체계를 구성한다. 이와 같은 연산 구조는 다양한 내·외부 데이터를 통합 분석하고, 실시간 상황 예측 및 고도화된 의사결정을 가능하게 하여 자율주행 성능의 지속적 향상을 뒷받침한다.

SDV와 ADV는 자율주행을 위한 AI모델의 개발 및 검증 방식에서도 차이를 보인다. SDV는 일반적으로 실제 도로 주행 데이터를 기반으로 AI모델을 개별 차량 내에서 최적화하는 방식을 갖는다. 반면, ADV는 실제 도로 주행 데이터뿐만 아니라 가상환경의 대규모 자율주행 시뮬레이션 데이터를 활용하여 데이터센터에서 학습·검증을 병행함으로써

보다 빠르고 정밀한 최적화가 가능하다. 이로 인해 ADV는 물리적 테스트의 한계를 벗어나, 가상환경에서의 다양한 코너케이스에 대한 대응력도 확보할 수 있다.

SDV와 ADV는 OTA 기능의 활용 목적과 방식에서도 차이가 있다. SDV에서의 OTA는 주로 사전 학습된 AI모델이나 소프트웨어 기능을 단방향으로 업데이트하는 유지보수 중심의 역할에 국한된다. 반면, ADV에서의 OTA는 차량이 실시간으로 수집한 데이터를 엣지 클라우드 및 데이터센터로 전송하여 AI모델을 순환적으로 재학습하고 재배포함으로써 자율주행 성능을 지속적으로 고도화하는 데 활용한다. 특히, ADV의 OTA 구조는 학습-배포-재학습의 프로세스가 반복되며 AI 모델의 성능이 지속적으로 개선되는 구조로, 다양한 주행 환경에 대한 적응성을 높일 수 있다.

자율주행 시스템이 ADV 중심으로 고도화됨에 따라, AI 컴퓨팅 수요도 기하급수적으로 증가하고 있다. 예를 들어, ADV 2.0 단계에서는 자율주행 시스템 전반의 누적 데이터 저장 수요가 약 10PB에서 3년 이내 100PB 규모로 확대될 것으로 전망되며, AI모델의 크기 또한 약 13배 이상 증가할 것으로 예상된다

[5]. 이에 따라 GPU(Graphic Processing Unit) 수요는 현재 L2 수준 대비 약 10배(8천 개 → 8만 개) 이상 확대될 것으로 예측되고 있다[4]. 이러한 변화는 자율주행 연산 구조 전반에 근본적인 전환을 요구하고 있다.

차량의 컴퓨팅 구조는 자율주행 연산 수요 급증에 대응하기 위해 기존의 분산형 ECU 체계에서 중앙집중형 구조로의 전환이 불가피해지고 있다. 분산형 구조에서는 각 기능이 독립적으로 작동하며 데이터를 개별 처리하지만, 중앙집중형 구조에서는 고성능 AI반도체가 주요 연산을 통합 수행함으로써 처리 지연을 최소화하고, 대규모 AI모델의 실시간 적용을 가능하게 하기 때문이다[2,6].

궁극적으로, 자율주행 시스템의 성능은 차량 내 연산 구조, 학습·추론의 통합 능력, 그리고 외부 클라우드 인프라와의 연계 가능성에 의해 결정되며, 이 세 가지 모두에서 AI반도체가 핵심 기반으로 작용한다. ADV 구조가 확산됨에 따라, AI반도체는 단순한 기능 보조를 넘어 자율주행 구현을 가능하게 하는 전략 기술로 자리매김하기 시작했다.

III. 자율주행차용 AI반도체 주요 기업

글로벌 주요 기업들은 자율주행차용 AI반도체 개발을 본격화하며, ADV 2.0 구현에 필요한 핵심 성능 요건 확보에 주력하고 있다. 테슬라, 엔비디아, 퀄컴은 AI 연산 성능, 에너지 효율, 학습·추론 통합 구조 등 고도화된 기술 요소를 중심으로 기술개발을 선도하고 있다. 또한, 국내 기업들도 자율주행차용 AI반도체 제품 개발을 추진해오고 있다.

1. 테슬라

테슬라는 자율주행 기능을 강화하기 위해 차량-데이터센터 간 직접 연계 구조를 구축하고 있다. 차

량에서 수집한 주행 데이터를 데이터센터(도조 슈퍼컴퓨팅)로 전송하고, 데이터센터에서 학습한 자율주행 AI모델을 OTA 방식으로 차량에 배포함으로써, AI 기반 자율주행 성능을 지속적으로 개선하는 순환 구조를 실현하고 있다[7]. 이 구조는 엣지 클라우드를 거치지 않으며, 자동차와 데이터센터 간 양방향 학습-배포 순환 구조를 직접 구현한다는 점에서, ADV 2.0의 핵심인 데이터 기반 모델 업데이트와 연속 학습 체계를 반영한다고 볼 수 있다.

차량-데이터센터 간 직접 연계하는 구조를 뒷받침하기 위해 테슬라는 차량용 AI반도체(FSD)와 데이터센터용 AI반도체(D1)를 자체 설계한 ASIC(Application-Specific Integrated Circuit) 형태로 제공하고 있다. FSD는 차량 내 실시간 추론을, D1은 대규모 AI모델 학습을 담당하는 이원 구조를 형성하며, 학습·추론 통합 및 순환적 업데이트 체계를 구현하는 기반으로 작용한다[8-10].

FSD 시리즈는 ADV 발전 단계에 따라 점진적으로 고도화 되어왔다. 초기형인 FSD1(2019년 출시)은 14nm 공정, 듀얼 NPU(Neural Processing Unit), LPD-DR4(Low-Power Double Data Rate) 메모리를 적용해 72TOPS(INT8)의 연산 성능을 제공한다. FSD1은 기능별 분산 연산에 최적화되어 있으며, ADV 1.0 수준과 유사한 SDV 구조에 적합한 것으로 평가된다. OTA 기능은 주로 소프트웨어 개선이나 일부 AI모델 파라미터 변경에 국한되었고, 전체 AI모델 업데이트를 처리하기에는 성능상 제약이 있었다.

FSD2(2023년 출시)는 5nm 공정, 트리플 NPU, GDDR6(Graphic Double Data Rate) 기반 고속 메모리를 적용해 118TOPS(INT8)의 연산 성능을 제공한다. FSD2는 트랜스포머 기반 모델, 고해상도 시각 정보, 예측-제어 통합 연산 등 고도화된 자율주행 기능을 차량 내에서 실행할 수 있는 성능 수준으로, ADV 2.0의 중앙집중형 연산 구조에 기술적으로 대

응할 수 있는 기반으로 평가된다. OTA 업데이트 범위가 확대되어, 복잡한 AI모델 전체를 보다 자주 차량에 안정적으로 배포할 수 있는 연산 여력을 확보하였다. 다만, FSD2 역시 학습 기능은 포함하지 않으며, AI모델의 학습 및 검증은 여전히 데이터센터에서 수행된 이후 OTA를 통해 차량에 배포되는 구조다.

데이터센터용 AI반도체인 D1은 362TFLOPS(BF16)의 연산 성능과 400W 전력 소모를 갖추고 있으며, 2023년부터 TSMC 7nm 공정을 통해 양산되고 있다. D1은 총 354개의 트레이닝 타일로 구성되며, 각 타일에는 1.25MB의 SRAM이 내장되어 있다. 도조 슈퍼컴퓨터에는 3,000개 이상의 D1이 장착되어 있으며, 전체 연산 성능은 1.1EFLOPS(Exa Floating Point Operations Per Second)에 달한다.

테슬라는 FSD2와 D1의 조합을 통해 차량-데이터센터 간 순환형 AI모델 업데이트 구조를 실현하고 있으며, 이를 바탕으로 ADV 2.0 체제로의 전환을 시도하고 있다. 특히 테슬라는 AI반도체 설계, 대규모 학습 인프라, 자율주행 소프트웨어, 완성차 제조 등 자율주행 전 주기를 내부 통제로 일원화하고 있으며, 경쟁사 대비 높은 수준의 수직 계열화를 갖추고 있다. 이러한 구조는 AI 기반 자율주행 하드웨어 분야에서 테슬라만의 차별화된 기술경쟁력 확보로 이어지고 있다.

2. 엔비디아

엔비디아는 자율주행 AI모델의 학습-배포-재학습 구조를 실현하기 위해 차량-엣지 클라우드-데이터센터를 연계하는 순환형 AI 컴퓨팅 구조를 구축하고 있다. 데이터센터(AI팩토리)에서 자율주행 AI모델을 학습하고, 엣지 클라우드(NVIDIA DRIVE)에서 최적화하여 차량에 배포하며, 차량에서 수집된

주행 데이터를 다시 데이터센터로 전송하여 AI모델을 지속적으로 고도화하는 방식이다. 이와 같은 순환 흐름은 ADV 2.0에서 요구하는 학습-추론 통합 구조와 동일한 기술적 방향을 지향하며, 엔비디아는 이를 통해 자율주행 AI플랫폼 전반에서의 영향력을 확대하고 있다[5].

차량-엣지 클라우드-데이터센터를 연계하는 구조를 실현하기 위해, 엔비디아는 차량용 AI반도체 Orin과 Thor, 그리고 데이터센터용 AI반도체 A100과 H100을 함께 제공하고 있다. Orin은 암페어(Ampere) 아키텍처 기반으로 254TOPS(INT8)의 연산 성능을 갖추고 있으며, 자율주행 L2~L3 또는 ADV 1.X 수준의 기능별 모듈 처리에 적합하도록 설계되었다. Thor는 1,000TFLOPS(FP8)의 연산 성능을 제공하는 블랙웰(Blackwell) 아키텍처 기반의 AI반도체로, 고해상도 인식과 통합형 AI모델 처리에 대응하기 위한 고성능 칩으로 개발되었으며, ADV 2.0 단계에서 실시간 연산 요구를 충족할 수 있을 것으로 평가되고 있다.

데이터센터용 AI반도체 A100과 H100은 각각 암페어 및 호퍼(Hopper) 아키텍처 기반 GPU로, 312TFLOPS(FP16) 및 1,513TFLOPS(FP16)의 연산 성능을 제공한다[11,12]. A100과 H100 모두 자율주행 AI모델의 학습과 시뮬레이션에 사용되며, 차량 내 추론용 AI반도체와 연계된 순환형 학습-추론 구조의 한 축을 담당한다.

엔비디아는 자율주행을 위한 전체 연산 구조를 GPU 기반 ASSP(Application-Specific Standard Product) 방식으로 구현하고 있으며, 다양한 자동차 제조사가 자사 소프트웨어 환경에 맞는 AI모델을 유연하게 통합할 수 있도록 'AI 하드웨어-플랫폼 통합 생태계'를 제공하고 있다. 특히 대규모 AI모델 업데이트와 실시간 연산이 요구되는 ADV 2.0 환경에서 엔비디아의 AI플랫폼은 고성능 연산과 신속한 AI모델

의 배포 간 균형을 제공하는 구조로 평가받고 있다. 현재 메르세데스-벤츠, 현대자동차 등 주요 완성차 업체들이 엔비디아의 AI팩토리와 NVIDIA DRIVE를 활용하고 있다.

엔비디아는 차량용 SoC(System on Chip)부터 데이터센터용 GPU까지 전 라인업을 아우르는 제품 체계를 기반으로, 자율주행 구현에 필요한 AI 연산 전반을 포괄하는 자율주행 AI플랫폼 전략을 구축하고자 한다. 학습-추론-재학습의 순환 구조를 AI플랫폼 수준에서 통합하면서, 단순 칩 경쟁을 넘어 자율주행 AI컴퓨팅 전반에서 독자적 생태계를 확장하고 있으며, ADV 2.0에 최적화된 수직 통합형 전략을 강화하고 있다.

3. 퀄컴

퀄컴은 자율주행차용 AI 시스템 구현을 위해 차량-엣지 클라우드-데이터센터를 연계하는 순환형 AI 컴퓨팅 아키텍처를 구축하고 있다. 데이터센터에서 학습된 AI모델을 차량 내 AI반도체에서 추론하며, 차량에서 수집한 주행 데이터를 다시 엣지 클라우드 및 데이터센터로 전송하여 AI모델을 지속적으로 고도화하는 구조를 채택하고 있다[13]. 이 과정에서 퀄컴은 자사 무선통신 기술과 엣지 클라우드를 결합해 차량과 외부 인프라 간 실시간 연결 기반의 데이터 교환을 지원하고 있으며, 이는 ADV 2.0에서 요구되는 통신 인프라의 실시간성 확보 측면에서 핵심 역할을 한다.

ADV 2.0 구현을 위해 퀄컴은 차량용 AI SoC인 스냅드래곤 라이드(Snapdragon Ride) 시리즈를 중심으로 연산칩, 통신기능, 소프트웨어 플랫폼을 통합한 AI 컴퓨팅 솔루션을 제공한다. 스냅드래곤 라이드 SoC는 4nm 공정 기반으로 최대 700TOPS의 연산 성능과 130W 소비 전력을 제공하며, 자동차 안전

무결성 기준(ASIL-D)을 만족하는 구조로 설계되었다[14,15]. 퀄컴은 GPU 기반 ASSP 방식과 달리, 독자적인 전용 연산 코어 기반 ASSP 방식을 채택하였으며, 높은 전력 효율과 실시간 추론 성능을 동시에 확보하고 있다. 이는 모바일 AP(Application Processor) 설계 경험에서 비롯된 저전력 최적화 기술을 반영한 결과다.

또한, 퀄컴은 5G 및 C-V2X(Cellular V2X) 기술을 기반으로 차량 간(V2V), 차량-인프라 간(V2I), 차량-클라우드 간(V2C) 연결을 실현하고 있으며, 이를 통해 엣지 클라우드와 유기적 연계를 강화하고 있다. 이 구조는 AI모델의 빠른 재학습과 유연한 개선을 가능하게 하며, 차량 내 연산 자원의 한계를 보완하는 동시에 ADV 2.0 환경에서 요구되는 외부 연산 연계 기반을 제공한다.

퀄컴은 이처럼 차량용 AI SoC 설계 기술뿐만 아니라 통신 인프라와 AI모델 운영 플랫폼을 포괄하는 통합형 자율주행 솔루션 전략을 바탕으로, ADV 2.0에 대응 가능한 저전력, 고효율, 고신뢰성 연산 생태계 구축하고자 한다. 현재 BMW, 혼다 등 주요 제조사들이 퀄컴 플랫폼을 바탕으로 자율주행 AI 모델의 개발, 학습, 배포를 통합적으로 수행 중이다. 퀄컴의 연산 구조는 차량-엣지 클라우드-데이터센터 간 AI 처리 유연성을 높임으로써, 연산 환경별 시스템을 개별 구축 및 운영해야 하는 부담을 줄이고 전체 시스템 관리의 복잡성도 완화한다. 이러한 전략을 통해 퀄컴은 고객사 입장에서 통합 구현 효율성과 확장성을 동시에 확보할 수 있는 플랫폼 경쟁력을 제공하며, 자율주행차용 AI반도체 시장에서 독자적인 시스템 통합 역량을 지속적으로 확대해 나가고 있다.

표 1은 테슬라, 엔비디아, 퀄컴이 추진 중인 자율주행차용 AI반도체 전략의 차이를 비교한 것으로, 각 기업의 ADV 2.0 대응 방식의 특징을 보여준다.

표 1 주요 기업의 자율주행차용 AI반도체 비교 분석

구분	테슬라	엔비디아	퀄컴
주요 사업	전기자동차, 자율주행 기술	AI반도체(GPU), 데이터센터, 자율주행 반도체	모바일 AP, 자동차 반도체 등
최근 매출 동향	2024년 상반기 전기차 수요 정체 및 경쟁 심화로 성장 둔화	데이터센터 매출 급증, 자동차용 반도체 매출 증가	자동차 반도체 매출 성장, 스마트폰 시장 침체 후 회복 중
ADV 2.0 전략	도조 슈퍼컴퓨터를 활용해 대규모 주행데이터를 학습 및 축적하여 자동차에 전송하는 AI 기반 자율주행 구현	데이터센터에서 학습한 자율주행 AI모델을 DRIVE 플랫폼을 통해 차량에 적용하고, 차량에서 수집한 데이터를 다시 AI팩토리(데이터센터)에서 학습	Snapdragon Ride 플랫폼을 통해 자율주행 기능 지원하며, 통신사 엣지 클라우드를 사용하여 자율주행에 필요한 다양한 통신 기능 지원
자동차 AI반도체 종류	ASIC	GPU 기반 ASSP	독점 코어 기반 ASSP
자율주행 프로세서 제품 현황	<p>FSD1: 72TOPS(INT8), 14nm, NPU 2개, LPDDR4</p> <p>FSD2: 118TOPS(INT8), 5nm, NPU 3개, GDDR6(16GB, 대역폭 224GB/s)</p> <p>D1: 362TFLOPS(BF16), 400W, 7nm</p>	<p>Orin: 최대 254TOPS(INT8), 45-70W, 5nm, 암페어 아키텍처 기반 GPU, 12 ARM Cortex-A78A CPU 코어 탑재</p> <p>Thor: 1,000 TOPS(INT8), 5nm, 블랙웰 아키텍처 GPU와 Arm Neoverse V3 CPU로 구성</p> <p>A100: 312TFLOPS(FP16), 300W, HBM2e(2TB/s)</p> <p>H100: 1,513TFLOPS(FP16), 350W, HBM2e(2TB/s)</p>	<p>Snapdragon Ride Flex SoC: 디지털 콕핏, ADAS 및 자율주행 기능 동시 탑재, 이기종 컴퓨팅 지원, 4nm</p> <p>Snapdragon Ride SoC: 30-700TOPS, 130W 성능으로 레벨 1-5 자율주행 지원</p>
SoC 면적당 AI연산기능 비중	전체 칩 면적의 50%	전체 칩 면적의 28%	비공개
AI 구현 특징	카메라 기반 시각정보 분석 자체 AI 학습 인프라 구축	센서 통합 AI 처리 GPU 기반 고성능 연산	저전력 고효율 설계 멀티모달 AI 구현
시장전략 (수직계열화 전략)	AI반도체+소프트웨어+차량 제조 완전 통합	AI반도체+소프트웨어+개발 환경 통합 제공	AI반도체+통신기능+소프트웨어 플랫폼 제공
파트너 생태계	독자적 생태계 구축	벤츠, 볼보, 현대차 등 주요 OEM과 협력	BMW, 폭스바겐, GM 등 주요 OEM과 협력
핵심 강점	완전한 기술 통제력 (테슬라 전용 반도체)	강력한 GPU 성능	저전력/고효율 설계

4. 국내 기업

국내 기업 중 자율주행차용 AI반도체를 개발하는 업체로 삼성전자, 텔레칩스, 넥스트칩, 딥엑스 등이 있다. 각 기업은 인포테인먼트, ADAS, 자율주행

SoC 및 엣지 AI 등 다양한 분야에서 AI반도체 기술을 고도화하며, 글로벌 시장을 겨냥한 전략적 행보를 보이고 있다.

삼성전자는 차량용 AI반도체 ‘엑시노스 오토’ 시리즈를 중심으로 자율주행 기술을 개발해 오고 있

다. 대표 제품으로는 엑시노스 오토 V9(2019년 1월 출시), 엑시노스 오토 V7(2021년 11월 출시), 엑시노스 오토 V920(2023년 6월 현대자동차와 협력 발표)이 있다. 가장 최신작인 엑시노스 오토 V920은 5nm 공정으로 설계되어 23.1TOPS의 AI연산 성능을 제공하며, 전작 대비 약 2.7배 성능 향상을 기록했다. 삼성전자는 또한, 자율주행차 내 대용량 AI 데이터를 안정적으로 저장·처리하기 위한 차량용 SSD(Solid State Device) ‘AM9C1’(8세대 V-NAND 기반)을 개발하고 있으며, 2TB 모델은 2025년 양산이 예정되어 있다. AM9C1은 PCIe 4.0 인터페이스 기반으로, 고속 데이터 입출력과 자동차 전용 인증(AEC-Q100, ASPICE CL3, ISO/SAE 21434)을 모두 충족하도록 설계되었으며, 고온 환경에서도 안정적인 동작이 가능한 고신뢰 저장장치로 평가받고 있다[16,17].

인포테인먼트용 AP 공급 중심에서 차량용 종합반도체 기업으로 변모 중인 텔레칩스는 AI반도체 유니콘 기업인 ‘리벨리온’(舊 사피온)과 협력하며 자율주행차용 AI반도체 개발을 본격화하고 있다[18]. 텔레칩스는 사피온(現 리벨리온)¹⁾으로부터 제공받은 NPU IP(Intellectual Property)를 기반으로 200TOPS급 AI 연산 성능을 갖춘 자율주행차용 AI반도체 ‘A2X’ 샘플을 2025년 4월 공개했으며, 향후 유연한 아키텍처 확장을 지원할 수 있도록 차량용 AI반도체 제품 라인업 확장을 계획 중이다[19].

넥스트칩은 ADAS 전용 SoC인 ‘아파치(Apache)’ 시리즈를 통해 차량용 AI반도체 시장에 진입하고 있다. 대표 제품인 아파치6은 삼성전자의 14nm 공정을 기반으로 최대 8TOPS 연산 성능을 제공하며, 현재 유럽 주요 완성차 기업과 PoC(Proof of Concept)를 진행하고 있다. 자율주행 L4 수준을 목표로 차세

대 제품인 아파치7을 개발 중이며, 아파치6 대비 약 10배 강화된 성능을 구현하고자 한다[20-22].

엣지용 AI반도체 전문 설계 기업인 딥엑스(DeepX)는 저전력 고성능 AI 연산에 최적화된 SoC를 개발하며 자율주행차용 AI반도체 시장 내 입지를 강화하고 있다. 대표 제품인 DX-V3는 비전 시스템에 특화된 SoC로, 자율주행 L2~L3 수준과 로봇 분야를 주요 타겟으로 하고 있다. DX-V3는 15TOPS급 AI 연산 성능을 제공하며, 평균 전력 소모 5W 미만으로 설계되어 차량용 엣지 디바이스에 적합한 고효율 솔루션으로 평가받고 있다. 2024년 말 글로벌 자동차 제조사를 대상으로 PoC를 진행했으며, 2026년 상반기 양산을 목표로 하고 있다. 딥엑스는 기존 NPU IP 기술력을 바탕으로, 후속 자율주행차용 AI반도체 제품인 DX-V5 등의 개발에 착수하였다[23,24].

전반적으로, 국내 기업들은 현재 ADV 1.0 또는 ADV 1.X 수준에 적합한 자율주행 기능 중심의 AI반도체 개발에 주력하고 있으며, 고성능 학습-추론 통합이 요구되는 ADV 2.0 수준의 제품은 아직 구현 사례가 부재한 것으로 보인다.

IV. 자율주행차용 AI반도체 시장전망

1. 시장전망 개요

자율주행 기술이 ADV 2.0으로 진화함에 따라 AI 연산 수요가 증가하고 있으며, 이에 대응하여 AI반도체 수요도 빠르게 성장할 것으로 기대되고 있다. 본고는 이러한 기술·산업적 흐름을 반영하여 자율주행차용 AI반도체 시장의 성장 가능성을 살펴보았다.

이를 위해 먼저, 자율주행차의 SW가 AI 기반으로 발전하는 흐름을 반영한 자율주행차용 AI반도체 시장을 목표시장으로 설정하였다(ADV 2.0 기준). 대상 반도체는 AI 연산 실행에 특화된 HW로 한정했

1) 사피온은 2024년 12월 리벨리온과 합병하여 통합 법인 ‘리벨리온’으로 공식 출범함

으며, 단순 연산 지원용 범용 반도체는 제외하였다. 즉, AI 연산 전용 독립형 칩 및 AI 연산 기능을 내장한 SoC, SiP, SoM으로 한정하였다. 그런 다음, 앞서 살펴본 ADV 발전 방향과 주요 기업의 비즈니스 전략을 바탕으로 자율주행 AI 연산이 발생하는 위치를 기준삼아 목표시장을 세분화하였다. 즉, 자율주행차가 자동차 내부, 엣지 클라우드, 데이터센터에서 발생하는 AI 연산 과정을 통해 구동된다는 특징을 반영하여 자율주행차용 AI반도체의 목표시장을 구성하였다.

자동차 내부용 AI반도체 시장은 실제 도로 주행 차량에 탑재된, AI 알고리즘 실행에 특화 설계된 HW 블록이 포함된 AI반도체의 매출액을 기준으로 하였다[25].

자동차 엣지 클라우드용 AI반도체 시장(X)은 전체 엣지 클라우드용 AI반도체 매출(Y)과 자동차 엣지 클라우드 시장의 비중(Z)을 곱한 값으로 추정하였다($X = Y \times Z$). 여기서, 자동차 엣지 클라우드 시장 비중(Z)은 전체 엣지 클라우드 시장에서 자동차 부분이 차지하는 구간별 추정 비중으로, 자동차 데이터센터용 AI반도체 시장 규모와 자동차 분야의 엣지 클라우드 도입 의향에 기반한 중요도의 가중 평균으로 산정하였다[26]. 전문가 자문을 통해 두 요소의 상대적 가중치는 7:3(시장 규모 : 중요도)으로 설정하였다. 즉, 자동차 엣지 클라우드용 AI반도체 시장은 자동차 산업의 도입 관심도보다 자동차 데이터센터용 AI반도체 시장 크기에 더 많은 영향을 받을 것으로 보았다.

* 본고는 OMDIA[25,27] 등의 2023년 말 기준 자료를 바탕으로 시장 수치를 구성하였으며, 이후 OMDIA[28]에 따르면 자동차 내부용 AI반도체의 연평균 성장률은 20.4%로 상향 조정되고, 엣지 클라우드용은 17.8%로 하향 조정되는 등 일부 수치 변화가 발생함. 이는 자율주행 기술의 고도화가 ADV 1.0 → 1.X → 2.0 단계를 거쳐 점진적으로 진행되는 흐름 속에서, 초기 수요가 차량 내부 연산 중심으로 선행 확산되고 있음을 보여주는 결과로 해석할 수 있음. 즉, 기술 구조는 E2E로 진화하고 있으나, 시장 확산은 내부 중심에서 외부 연산으로 점차 이행 중인 과도기적 양상을 반영함.

자동차 데이터센터용 AI반도체 시장은 전체 데이터센터용 AI반도체의 수직시장(Tier 1 공급업체 중심) 중 자동차 분야와 연결된 매출을 기준으로 추정하였다[27]. 데이터센터용 AI반도체는 클라우드 및 기업 데이터센터의 AI 워크로드 가속을 위한 전용 프로세서를 의미하며, 본고에서는 자율주행 AI모델 학습 등 자동차 특화 AI 연산 수요를 포함하는 경우로 한정하였다.

참고로, 본고에서 언급한 테슬라, 엔비디아, 퀄컴은 자율주행차용 AI반도체 기술 전략을 이해하기 위한 대표적 사례로 제시된 것으로, 이들의 매출 구조는 시장 수요 추정에 직접 반영하지 않았다. 시장 규모 산정은 산업 평균 구조를 기준으로, Tier 1 부품 공급업체 중심의 수요자를 기반으로 설정하였다.

2. 시장추정 결과

전 세계 자율주행차용 AI반도체 시장은 2023년 약 45.8억 달러 규모에서 연평균 16% 성장하여 2028년에는 약 96.7억 달러 규모를 형성하며, 전체 AI반도체 시장의 약 4.6~6.5% 비중을 차지할 것으로 전망되었다. 자율주행차용 AI반도체 시장은 초창기에는 자율주행차용 AI반도체 시장은 데이터센터용 제품이 가장 큰 비중을 차지하고 있으며, ADV 구조의 확산에 따라 차량 내부용과 엣지 클라우드용 AI반도체의 수요도 점진적으로 확대될 것으로 나타났다.

즉, 자율주행차용 AI반도체 시장은 데이터센터용 제품이 높은 비중을 지속하며 시장 성장의 중심축이 되고, 엣지 클라우드 및 내부용 AI반도체는 단계별 기술 진화에 따라 점차 수요를 확대해갈 것으로 예측되었다(표 2).

표 2 세계 자율주행차용 AI반도체 시장 전망

(단위: 백만 달러, %)

구분	2023	2024	2025	2026	2027	2028	CAGR (23-28)
자율주행차 내부용 AI반도체	1,745 (38.1%)	2,126 (32.9%)	2,403 (28.5%)	2,598 (28.8%)	2,815 (29.5%)	2,940 (30.4%)	11%
자율주행차용 엣지 클라우드 AI반도체	332 (7.2%)	499 (7.7%)	650 (7.7%)	763 (8.4%)	906 (9.5%)	1,058 (10.9%)	26%
자율주행차용 데이터센터 AI반도체	2,505 (54.7%)	3,843 (59.4%)	5,376 (63.8%)	5,673 (62.8%)	5,807 (60.9%)	5,673 (58.7%)	18%
합계	4,582	6,468	8,429	9,034	9,528	9,671	16%

출처 OMDIA[25,27]와 Gartner[26]의 자료를 동의하에 사용하였으며, 이를 바탕으로 저자 자료 생성, 2025. 7.

V. 결론 및 시사점

본고는 자율주행차에서 AI의 역할이 확대되는 기술 발전 추세와 글로벌 선도 기업들의 전략을 반영하여, 자율주행 구현에 직접적으로 관여하는 AI반도체 시장을 전망하였다. 특히, AI 연산 실행용 반도체만을 시장범위에 포함하여 보다 보수적인 관점에서 시장을 추정하였다. 또한, 분석 범위를 자동차 내부뿐만 아니라 외부에서 자율주행에 관여하는 AI반도체까지 확장함으로써 자율주행차의 AI 기능에 초점을 맞춘 시장분석을 시도하였다. 이를 통해 본고는 자율주행차용 AI반도체 시장을 새롭게 정의하고 성장 가능성을 확인했으며, 관련 정책 수립에 필요한 객관적 근거를 제시함으로써 자율주행차용 AI반도체를 미래 신성장 동력으로 삼기 위한 논리적 기반을 마련하고자 하였다.

이러한 분석을 바탕으로, 본고는 자율주행차용 AI반도체 기술개발 및 산업 활성화를 위한 정책적 시사점을 다음과 같이 제안하고자 한다.

첫째, 글로벌 선도기업들의 3-Tier 전략을 참고하여 ‘자동차-엣지 클라우드-데이터센터 연계형 전략’으로 자율주행차용 AI반도체 산업 생태계를 형성 및 고도화하는 방안을 고민할 필요가 있다. ‘자동

차-엣지 클라우드-데이터센터’ 영역별 전문기업 간 협의체를 구성하여 국내 산업구조에 적합한 3-Tier 연계 비즈니스 모델 발굴 및 공동 R&D 프로젝트를 추진하거나, 3-Tier 통합 솔루션의 실증 및 성능 검증을 위한 테스트베드 구축, 기업 공동 활용이 가능한 개발·검증 인프라 조성 등을 고민해볼 수 있다. 이를 통해 산업계 수요 기반의 자율주행 AI반도체 핵심 IP를 확보하고 설계 자산을 축적함으로써 국내 자율주행차용 AI반도체 기업의 기술경쟁력을 강화할 수 있을 것이다.

둘째, ADV 2.0 시대를 대비하기 위해서는 자율주행 시스템과 AI반도체를 유기적으로 연계한 로드맵을 수립하고 이행 방안을 논의할 필요가 있다. 예를 들어, 자동차-엣지 클라우드-데이터센터로 이어지는 데이터 처리 구조를 고려하여 자율주행차용 AI반도체 개발 전략을 수립하고, 이를 로드맵에 통합할 필요가 있다. 또한, ADV 2.0 달성을 위한 핵심 성능지표를 설정하고, ADV 1.0에서 2.0으로의 단계적 전환을 위한 요소 기술별 개발 전략을 구체화해야 한다.

자율주행차용 AI반도체 기술 개발 및 생태계 조성 전략 수립을 통해 산업 경쟁력 제고뿐만 아니라 글로벌 시장에서의 주도권 확보를 위한 기반을 마

련할 수 있을 것이다. 그러나 현재 국내 자율주행차용 AI반도체 기업들은 ADV 1.0에 해당하는 자율주행 특화 기능 중심의 AI반도체 개발에 주력하고 있으며, 글로벌 선도기업들 대비 ADV 2.0 대응을 위한 연산 구조 전환과 고성능 학습·추론 통합 기술력 확보는 다소 미흡한 실정이다. 향후 자율주행차 산업의 경쟁력이 AI반도체 역량에 의해 좌우될 가능성이 큰 만큼 경쟁력 있는 기술 확보를 위한 중장기적 전략 마련이 시급하다.

용어해설

ADV 인공지능이 차량의 인식, 판단, 제어를 주도하며, 스스로 학습하고 자율적으로 운행하는 차량으로, 원격 업데이트(OTA)를 통해 자율주행 성능을 지속적으로 향상

SDV 차량의 주요 기능을 소프트웨어로 구현하고, 원격 업데이트(OTA)를 통해 기능을 지속적으로 개선하거나 확장할 수 있는 차량

SRAM 전원이 공급되는 동안 데이터를 유지하는 메모리로, DRAM과 달리 주기적인 리프레시(Refresh)가 필요 없으며, 속도와 안정성이 요구되는 CPU 캐시, AI반도체 등 고성능 영역에 주로 적용

3-Tier 자율주행차의 데이터 처리 구조를 차량-엣지 클라우드-데이터센터로 구분해 각 계층에 적합한 AI반도체를 배치함으로써 연산 효율성과 처리 속도를 극대화하는 3단계 아키텍처

참고문헌

- [1] J.K. Jung et al., "Hardware accelerators for autonomous vehicles. In Artificial Intelligence and Hardware Accelerators," Artificial Intelligence and Hardware Accelerators, 2023, pp. 269-317.
- [2] V.M. Navale et al., "(R)evolution of E/E architectures," SAE Int. J. Passeng. Cars - Electron. Electr. Syst., vol. 8, no. 2, 2015, pp. 282-288.
- [3] D. Barnett and A. Mulay, "7 roadblocks for software-defined vehicles and how to overcome them," Luxoft, 2024. 7. 17. <https://www.luxoft.com/blog/7-roadblocks-sdv-solutions>
- [4] A. Narra, "Introduction to Autonomous Vehicles," NVIDIA GTC 24, 2024. 3. <https://www.nvidia.com/ko-kr/on-demand/session/gtc24-s62558/>
- [5] N. Marks, "Accelerating Autonomous Vehicle Development With High-Performance AI Computing," NVIDIA GTC 24, 2024. 3. <https://www.nvidia.com/ko-kr/on-demand/session/gtc24-se63003/>
- [6] P. Arejola et al., "The future of Automotive Computing," McKinsey&Company, 2022. 10. 6.
- [7] 임은영, 강희진, "테슬라 2022 AI Day Takeaways: 풍요로운 세상을 꿈꾸는 AI Giant," 삼성증권, 2022. 10. 4.
- [8] 이재일, "테슬라 AI Day, 자율주행 이상의 가능성," 유진투자증권, 2021. 8.
- [9] 테슬라 AI day 2021, <https://www.youtube.com/watch?v=jOz4FweCy4M>
- [10] 테슬라 AI day 2022, https://www.youtube.com/watch?v=ODSJsviD_SU
- [11] E. Juliussen, "What Is Nvidia Doing in Automotive?," EETimes, 2024. 6. 13. <https://www.eetimes.eu/what-is-nvidia-doing-in-automotive/>
- [12] NVIDIA Website, <https://www.nvidia.com/ko-kr/self-driving-cars/drive-platform/hardware/>
- [13] Qualcomm Website, <https://www.qualcomm.com/products/automotive/overview>
- [14] Qualcomm, "Unlocking on-device generative AI with an NPU and heterogeneous computing," 2024. 2.
- [15] F. Sideco, "Qualcomm Expecting To 'Flex' It's Automotive Muscles In 2024," Forbes, 2024. 1. 8.
- [16] 삼성전자 Website, <https://semiconductor.samsung.com/kr/processor/automotive-processor/exynos-auto-v920/>
- [17] 삼성전자 뉴스룸, "삼성전자, 현대자동차와 차량용 인포테인먼트 분야 첫 협력," 2023. 6. 7.
- [18] Telechips, "IR Presentation 2023," 2023. 12.
- [19] Telechips, "텔레칩스, 차량용 AI 가속기 공개...미래차 AI 기술혁신으로 SDV 기반기술 강화," 2025. 3. 6.
- [20] Nextchip, "IR Presentation 2023," 2023. 6.
- [21] 한주엽, "넥스트칩, 연내 ADAS칩 '아파치6' 싱글런 생산...유럽 완성차업체 공급 추진," Theelec, 2024. 5. 21.
- [22] Nextchip Website, <https://www.nextchip.com/kr/adas/adas.php?idx=4>
- [23] 딥엑스 Website, <https://deepx.ai/>
- [24] 정예린, "딥엑스, 'AI 초점' 차세대 칩 출시 초읽기...1세대 반도체 라인업 완성," The Guru, 2024. 9. 26.
- [25] A. Harrowell, "AI Processors for the Edge Forecast Report - 2023 Database," OMDIA, 2023. 12.
- [26] Gartner, "Emerging Technologies: Revenue Opportunity Projection for Edge Computing," 2024. 9. 23.
- [27] A. Harrowell and N. Dunay, "AI Processors for Cloud and Data Center Forecast Report," OMDIA, 2024. 7.
- [28] A. Harrowell, "AI Processors for the Edge Forecast Report - 2024 Database," OMDIA, 2024. 12.