

# 지각 특화 AI 반도체 투자 타당성 검토

- 기술 니즈와 시장 수요 중심으로 -

이성준·박정렬·이승환

본 보고서는 ETRI ICT전략연구소 기본사업인  
“국가 지능화 기술정책 및 표준화 연구”를 통해 작성된  
결과물입니다.



- \* 본 보고서의 내용은 연구자의 견해이며 ETRI의 공식 의견이 아님을 알려드립니다.
- \* 보고서의 시장 자료는 출판사의 사용 허가를 받아 사용되었습니다. 재사용을 원하시면 출판사에 문의하시기 바랍니다.
- \* 본 보고서는 기 발간된 '피지컬 AI 시대의 지각 특화 AI 반도체 개념 및 전략적 투자 방향'(전자통신동향분석 Vol.40, 2025.12.)의 개념 및 내용을 확장하여 작성한 결과물입니다.

# 목 차 C O N T E N T S

<b>Executive Summary</b> .....	<b>i</b>
<b>I. 피지컬 AI 시대의 도래와 기술 패러다임 전환</b> .....	<b>1</b>
1. 피지컬(Physical) AI 확산 배경과 기술적 의미 .....	1
2. 물리법칙(Physical World) 기반 지각 연산 환경의 변화 .....	8
3. 중앙집중형 구조의 한계와 새로운 연산 요구 .....	13
<b>II. 지각 특화 AI 반도체 개념과 특성</b> .....	<b>19</b>
1. 지각(Perception) 특화 AI 반도체의 정의와 역할 .....	19
2. 필요 기술 및 기능적 차별성 .....	23
<b>III. 목표시장과 확산 가능성 분석</b> .....	<b>33</b>
1. 목표시장 정의 및 추정 프레임워크 .....	33
2. 활용 분야별 확산 전망과 시장 진입 시점 분석 .....	48
3. 목표시장 규모 추정과 산업 생태계 현황 분석 .....	55
<b>IV. 신시장 창출과 성장 견인을 위한 정부 역할</b> .....	<b>69</b>
1. 국가전략기술로서 정부의 투자 필요성 .....	69
2. 신시장 창출과 성장 견인을 위한 정부의 투자전략 방향성 .....	73
<b>참고문헌</b> .....	<b>79</b>
<b>부 록</b> .....	<b>87</b>





## Executive Summary

첫째. 물리 세계와 직접 상호작용하는 피지컬 AI 시대의 도래에 따라, 이에 대응할 새로운 반도체 전략 수립이 시급한 시점임

둘째. 지각(Perception) 특화 반도체는 센서와 메인 SoC 사이의 기능적 공백을 메우고, 데이터 병목과 전력 문제를 해결하는 핵심 중간 계층임

셋째. 목표시장(휴머노이드 로봇, 자율주행차, XR기기)은 기존 주력 반도체에 버금가는 420억 달러(`35년)로 성장, 기술개발(`26년)부터 상용화(`30년~)까지 5년 소요 예상

넷째. 시장 대중화 속도를 고려할 때 `31년 적기 진입이 승부처이며, 이를 위해 `26년 선제적 R&D 착수로 골든타임을 확보해야 함

다섯째. 글로벌 기술 패권 경쟁 대응 및 시스템반도체 신성장동력 확보를 위하여, 국가 차원의 전략적 투자와 생태계 조성이 필요하며, 범용성 및 호환성 확보를 위한 설계 및 공정 기술의 개발과 파급력이 큰 단기 전략과 확장성 기반의 중장기 시장 진입전략이 실행되어야 함

## 1. 피지컬 AI 시대의 도래와 기술 패러다임 전환

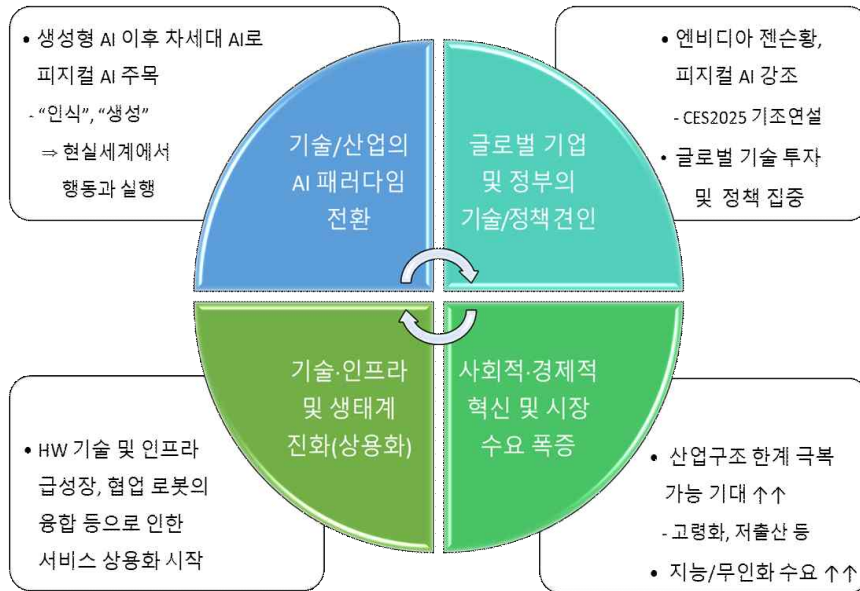
### (1) 피지컬(Physical) AI 확산 배경과 기술적 의미

#### ● 피지컬 AI 정의 및 도래

- \* 현실 세계에서 직접 상호작용(물리적 세계를 이해하여 추론하고 행동)하는 AI
- \* 산업·사회 전반에서 AI가 현실 현장에 직접 개입·동작하는 새로운 패러다임
- \* 이는 생성형 AI 중심의 디지털 혁신을 넘어, 국가·기업 차원의 차세대 기술패권 경쟁의 핵심 축으로 부상할 가능성이 높음
- \* 로봇, 자율주행, XR, 물류, 제조 등에서 이미 초기 적용사례가 확산 중

● 피지컬 AI 시장 성장의 가시화

- ※ 피지컬 AI 시장은 도약기(`20~`25년) → 본격 도입·확산기(`25년~`30년) → 대중화기(`30년 이후)로 구분 가능
- ※ `25~`30년 : 시범서비스·상용화 활성화 시점, 지금이 기술/제도 대응 마련 적기
- ※ `27년~`32년 전후 : 새로운 생태계 형성 및 기술 성숙 단계 돌입 전망



[ 피지컬 AI가 최근 주목받는 이유 ]

● 피지컬 AI 기술 특성 : 현재 언어중심 생성형 AI와의 차별성

< 현재(언어중심) AI와 미래(피지컬) AI의 차별화 포인트 >

차별화 포인트	현재(언어중심) AI	미래(피지컬) AI
물리 세계와의 상호작용	비물리적·가상 데이터 기반(텍스트·이미지·음성) 분석	센서·로봇·엣지컴퓨팅 결합 → 현실 환경에서 직접 행동·개입
지능 루프1) 구조2)	데이터 입력 → SW 판단 또는 예측(의사결정 지원 중심)	센싱 → 상황 인지·판단·계획 → 행동 실행 → 피드백의 전 주기 통합시스템
물리 법칙 내재화3)	텍스트·이미지·음성 등 비물리적 데이터 위주로 학습	마찰·힘·충돌·관성 등 물리 법칙을 모델링·예측하여 안전한 행동 수행
실시간성 및 적응력	서버·오프라인 중심 데이터 분석	실제 변화에 대응하는 실시간 센싱-반응-행동 수정
자기 진화 능력	강화학습 적용되나, 주로 시뮬레이션 기반	현실 상호작용 기반 자기생성 데이터 및 실시간 강화학습으로 빠른 성능 향상
중심산업 및 사회적 효과4)	디지털 산업 및 업무 효율화 중심 효과	현실 산업 현장 지능화 및 일상 생활 변화 중심 효과

## (2) 물리법칙(Physical World) 기반 지각 연산 환경의 변화

### 가. 피지컬 AI 시대의 물리법칙 환경의 센싱 데이터 특성 변화

#### ● 물리법칙 환경(Physical World)이 반영된 센싱 데이터 특성 변화

- ※ 데이터 규모 폭증, 데이터 유형의 멀티모달화, 데이터 생성주기의 단축 및 불규칙성, 멀티모달 데이터의 비정합

#### ● 고품질 지각을 위한 센서의 다중화 및 고집적화

- ※ 센서 단가 하락 + 지각 정밀도 요구 상승 → 센서 수의 증가 및 시스템 내 고집적 트렌드 가속

#### ● 물리 환경의 복잡성 증가에 따른 다중 센서 융합 필요성 확대

### 나. 일반 엣지(Edge) 디바이스를 초과하는 피지컬 AI 시스템의 요구성능

#### ● 독립적이고 신속한 의사결정을 위한 고성능 지각 처리 필요

#### ● 엣지 디바이스 내 강력한 AI 연산 능력 요구

- ※ 클라우드/데이터센터 의존 시 지연 발생 → 안전·제어·물리 상호작용에 부적합
- ※ 현장의 대량/다양 센서정보를 로컬에서 즉시 해석·추론 가능한 AI 처리 능력 요구

#### ● 높은 신뢰성 및 안정성과 에너지 효율성 확보 필요

- ※ 피지컬 AI는 사람·사물과 직접 상호작용하므로 오작동 리스크 최소화 필요
- ※ 로봇, 자율주행, XR, 드론 등 다수가 배터리 기반 → 고연산·저전력 동시 달성 필수

#### ● 물리적 제약 극복을 위한 내장형 전용 하드웨어 필요

- ※ 물리 환경에서의 관성·마찰·중력·동역학 모델을 실시간 계산해야 하므로 일반 엣지 수준을 초과하는 연산 요구

1) 엔비디아 GROOT. Cosmos 플랫폼 등이 센싱-판단-계획-동작-피드백의 실시간 루프를 갖춘 대표적인 사례

2) Brian Caulfield. (2025.1.). CES 2025: AI Advancing at 'Incredible Pace,' NVIDIA CEO Says

3) 엔비디아 보도자료. (2025.1.). NVIDIA Expands Omniverse With Generative Physical AI

4) 전 세계 물리 AI 시장은 726억 달러 규모이며, 2030년까지 두 배 이상 성장하여 1,660억 달러에 달할 것으로 예상되며, 연평균 성장률은 16.4%에 달할 것으로 예상 (출처: Eximius Ventures. (2025.5.). Physical AI and Its Role in Shaping Industry 5.0)

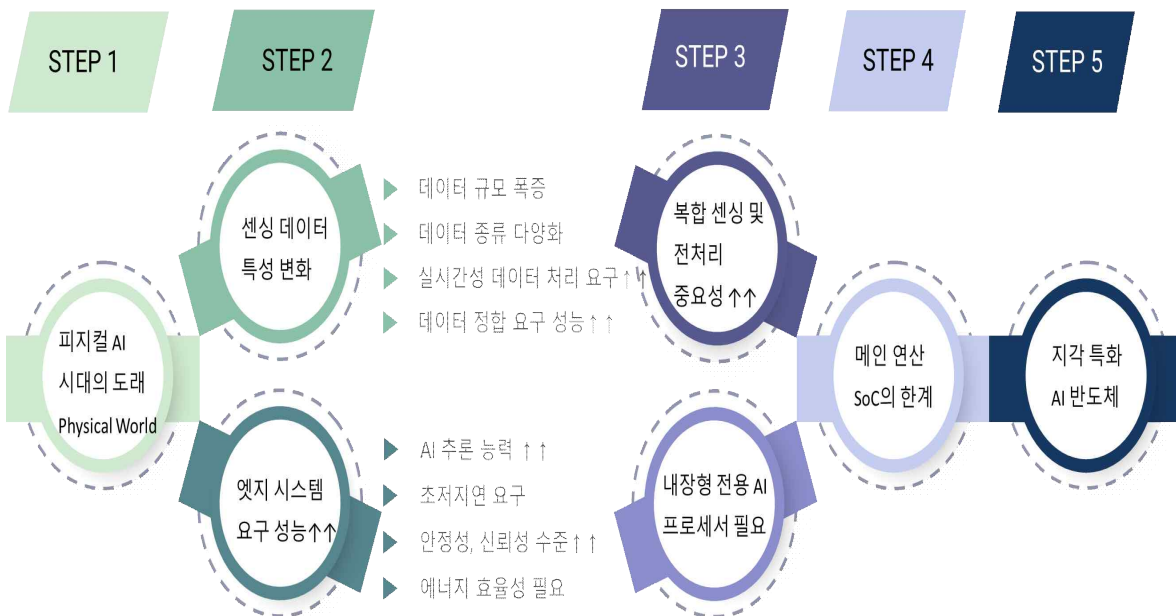
### (3) 중앙집중형 구조의 한계와 새로운 연산 요구

#### 가. 중앙집중형 구조(메인 연산 SoC)의 처리 한계

- 통신 및 전송 과정에서의 지연 누적
- 메모리 대역폭 포화 및 데이터 이동 병목
- 전력·발열 증가로 인한 에너지 효율성 저하
- 모델 업데이트의 경직성 및 경량화의 부작용

#### 나. 새로운 연산 구조 : 지각 특화 AI 반도체의 필요성

- 센서 인근에서 데이터 해석을 수행하는 전용 반도체 필요
- 실시간성, 안정성, 에너지 효율성 요구 충족 필요
  - ※ 센서 근처에서 지각 연산을 수행함으로써 데이터 이동을 최소화하고, 전력 및 열 효율 개선
- 상호작용(Interaction) 증가에 따른 지각 특화 하드웨어의 중요성 확대



[ 피지컬 AI 시대 도래에 따른 지각 특화 AI 반도체 필요성 ]

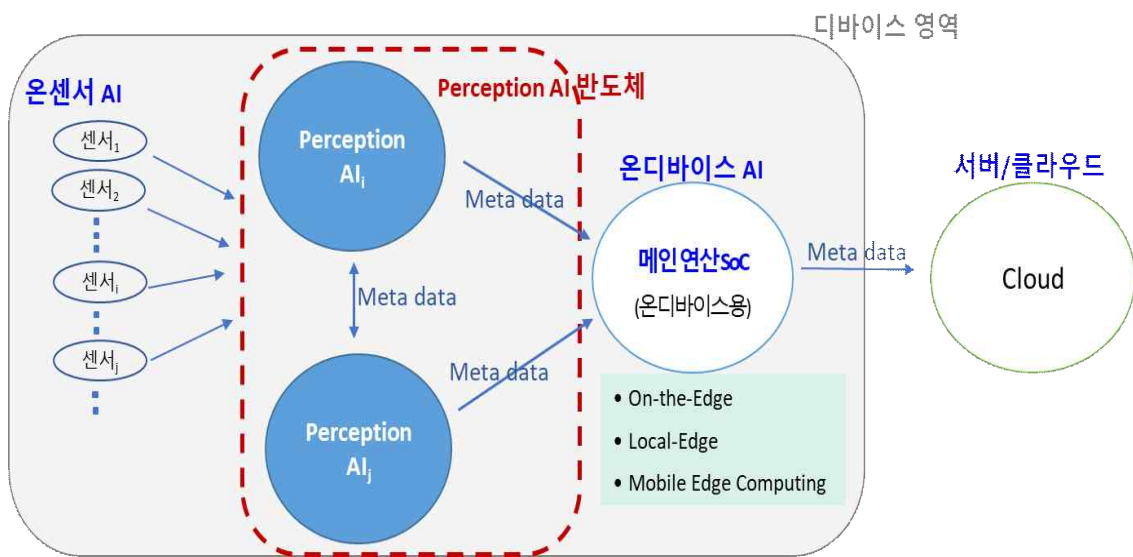
## 2. 지각(Perception) 특화 AI 반도체 개념과 특성

### (1) 지각 특화 AI 반도체의 정의와 역할

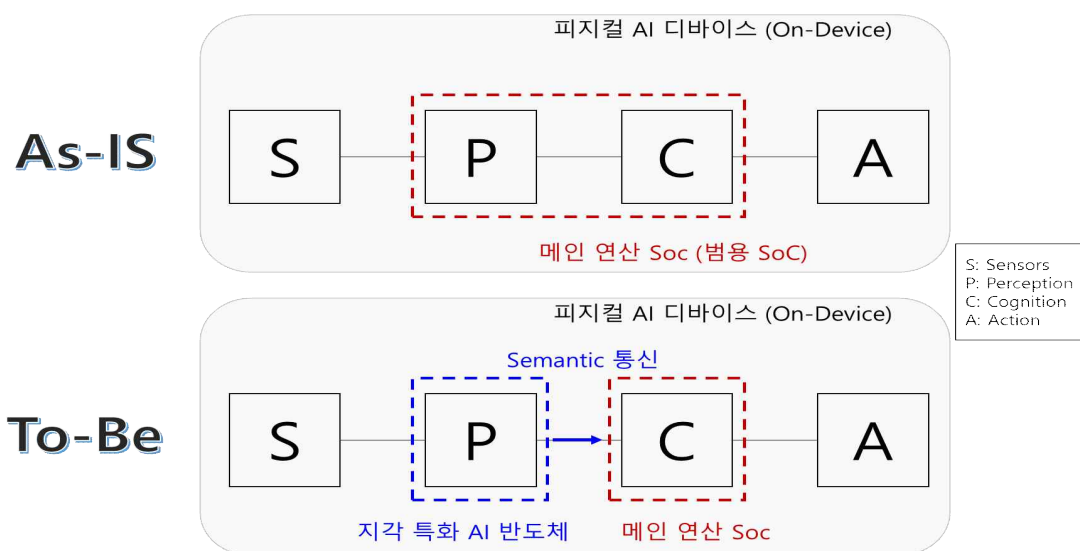
#### 가. 개념 정의

##### ● 지각(Perception) 특화 AI 반도체 정의

※ 디바이스 내에서 멀티 센서의 복합 데이터를 해석하여, 메타데이터를 생성하고, 엣지 내 메인 연산 SoC 또는 상위 클라우드로 전달하는 AI 반도체



[ 지각(Perception) 특화 AI 반도체 개념 ]



[ 지각 특화 AI 반도체 도입에 따른 피지컬 AI 디바이스 구조 변화 ]

## 나. 처리 프로세스 및 주요 기능

### ● 처리 프로세스

- ※ 디바이스 내에서 다양한 센서들로부터 수집된 복합 데이터를 해석 → 해석된 정보들은 메타데이터로 의미있는 단위로 축약, 추출되어 생성 → 생성된 메타데이터는 디바이스 내 메인 연산 SoC는 물론이고 필요에 따라 다른 디바이스 또는 상위 엣지/클라우드로 전송

### ● 센서 퓨전 (Sensor Fusion)

- ※ 다양한 종류의 센서 데이터 또는 여러 위치의 동종 센서 데이터를 융합하여 하나의 종합적인 환경 인식을 형성

### ● 온-디바이스 지각 처리 (On-Device Perception)

- ※ 센서 신호 처리와 AI 추론을 클라우드가 아닌 디바이스 자체에서 수행

### ● 메타데이터 생성 및 요약 (Semantic 통신)

- ※ 원시 센서 데이터에서 의미 있는 메타데이터를 추출하고, 상위 시스템이 활용하기 쉽도록 데이터 구조화/요약

### ● 엣지-클라우드 연계 (Interaction; AI-to-Human, AI-to-AI)

- ※ 엣지 디바이스 내에서 신호 전처리와 1차 지각(Perception)/인지를 수행
- ※ 필요 정보와 요약 결과만 엣지 서버나 클라우드로 전송, 상위계층 AI 등과 협업

## (2) 필요 기술 및 기능적 차별성

### 가. 주요 기능별 필요 기술

#### ● 센서 인터페이싱 및 퓨전 알고리즘

- ※ 다양한 종류의 센서 데이터 또는 여러 위치의 동종 센서 데이터를 융합하여 복합적으로 해석하기 위해서는 센서 퓨전 알고리즘과 인터페이싱 기술이 필요

#### ● 경량화된 지각(Perception) AI 추론 모델

- ※ 지각 특화 처리를 위한 AI 알고리즘은 엣지 디바이스에서 실시간 동작하는 전처리 모델과 추론 모델 기술이 필요

#### ● 메타데이터 전송 및 칩 보안 기능

- ※ 필요 이상으로 원본 데이터를 저장하지 않고 즉시 필요한 정보만 추출하도록 시스템 설계

● **엣지 환경에 최적화된 전력 및 성능 관리**

- ※ 전력 효율 극대화를 위한 기술 고도화 및 최적화된 성능 관리

● **칩렛(Chiplet) 기반 또는 설계자산 IP(Intellectual Property) 기반 모듈화**

- ※ 여러 기능 칩(칩렛<sup>5)</sup>)을 한 패키지에 집적하거나, IP<sup>6)</sup>(설계자산) 블록 형태로 모듈화하여 SoC<sup>7)</sup> 구성 가능

**나. 기능적 차별성 및 주요 장점**

● **온센서·온디바이스 AI의 한계**

- ※ 온센서 AI의 한계 : 센서 수 증가 및 고해상도화로 인하여 DSP 수준의 내장 연산 성능과 제한된 전력으로는 실시간 처리 및 센서 융합 불가
- ※ 온디바이스 AI의 한계 : 대규모 센싱 데이터 처리 시 모달 확장성이 제한되고, 메모리 및 대역폭의 제약, 전력 부담으로 실효 추론 성능이 급격히 저하
- ※ 피지컬 AI에서는 기능별 역할 분업이 필수적이며, 지각 전용 반도체가 없으면 지연, 전력 소모, 안정성 측면에서 시스템 전반의 성능 곡선이 악화될 위험 초래

● **기능적 장점**

- ※ 실시간 지각 한계 극복 : 엣지 환경의 불규칙·고속 센싱 데이터를 안정적으로 처리
- ※ 멀티센서 지각 최적화 : 다양한 센서 신호를 전용 하드웨어 가속으로 고속·저전력 처리
- ※ 물리환경 지각 성능 극대화 : LLM/VLM 대비 압도적인 지각 전용 연산 효율 제공

● **시장성 측면의 장점**

- ※ 센서 내부 AI 기능 확장 시 전력·비용·발열 증가 문제를 외부 전용 칩으로 해소
- ※ 센서와 독립된 고성능·저가의 지각 전용 칩 공급이 가능해 다수 산업에서 범용화 기대
- ※ 안정적 공정·IP 모듈 기반으로 로봇, 자율차, XR 등 다양한 피지컬 AI 시장에 확대 적용 가능

● **하드웨어·시스템적 차별성**

- ※ 범용 GPU/CPU 대비 전력 대비 연산 효율 극대화
- ※ (확장성 및 경량성 확보) 칩렛 기반 구성으로 모듈 추가/삭제 및 제품군 파생이 용이, IP 모듈화를 통한 고객 맞춤형 구성 지원, 경량 버전으로 저전력·저비용 시장(가전·소형 로봇 등)까지 확장 가능

5) Chiplet : 하나의 큰 칩(SoC)을 여러 개의 작은 칩(칩렛)으로 쪼개어, 각각의 칩렛이 특정 기능만 담당하도록 만든 뒤, 이 칩렛들을 하나의 패키지 안에 조립하는 방식. IP로 설계된 여러 기능 블록들을 실제로 만든 각각의 칩 또는 그 칩들을 패키지에서 조립하는 방식

6) Intellectual Property: 칩(반도체) 안에 들어가는 '기능 블록'을 미리 설계해 놓은 설계 자산

7) System on Chip으로 여러 부품들이 하나의 반도체에 통합되어 단일 시스템을 구성하는 형태

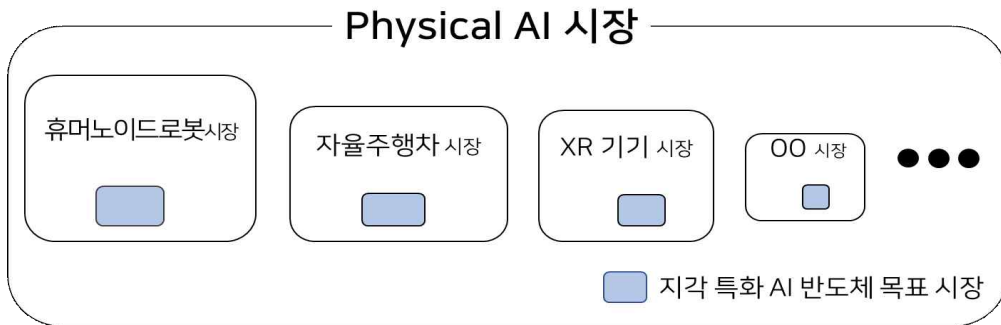
### 3. 목표시장과 확산 가능성 분석

#### (1) 목표시장 정의 및 추정 프레임워크

##### 가. 목표시장 정의

###### ● 목표시장

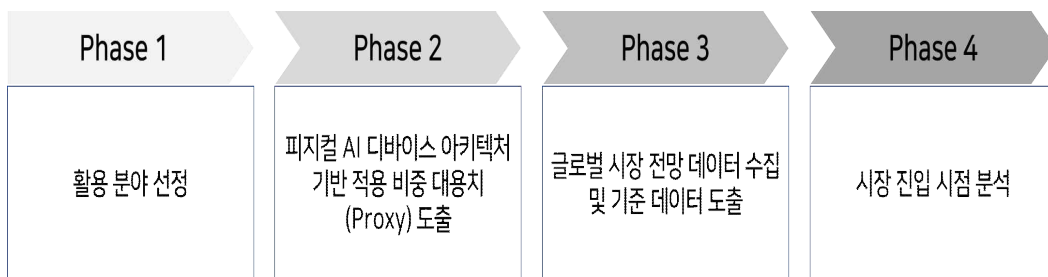
- ※ 데이터가 발생하는 지점(센서 단)에서 이중의 멀티모달 데이터를 실시간으로 융합·추론하고 지능적으로 전처리함으로써, 메인 SoC의 부하를 경감하는 분산적·계층적 처리 구조의 반도체 시장으로 정의



[ 지각 특화 AI 반도체의 목표시장 ]

##### 나. 단계별 시장 규모 추정 방법 및 결과

###### ● 4단계 분석을 수행하여 목표시장 규모 추정



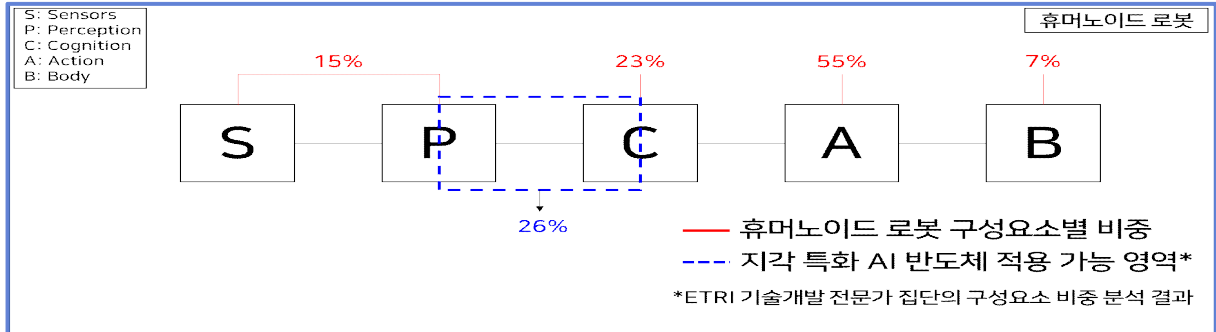
[ 시장규모 추정 순서도 ]

###### ● (Phase 1) 핵심 활용 분야(Key Application) 선정

- ※ ① 실시간 환경 인식, ② 멀티모달 대용량 데이터 처리, ③ 초저지연·고효율 엣지 임베디드 연산이라는 3대 기술적 임계(Criticality)를 동시에 요구하는 휴머노이드 로봇, 자율주행차, XR 기기를 선정함

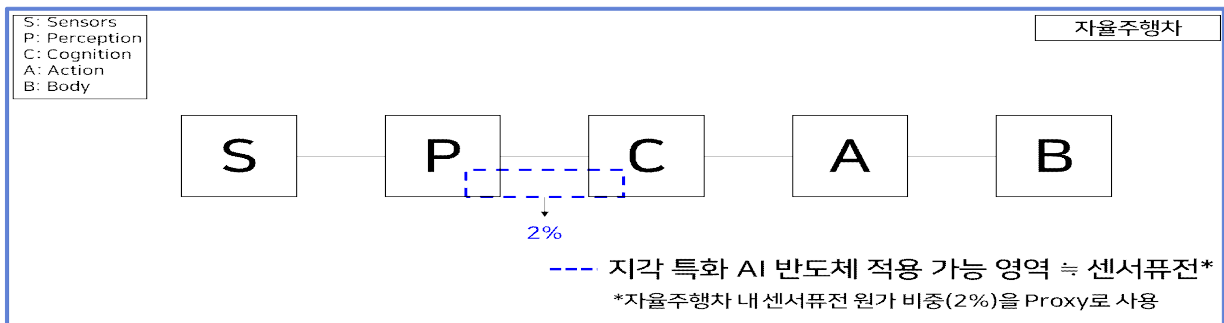
● (Phase 2) 피지컬 AI 디바이스 아키텍처 기반 적용 비중 대용치(Proxy) 도출

※ (휴머노이드 로봇) 피지컬 AI 디바이스의 S-P-C-A 구조 중 지각(S+P: 15%), 인식(C: 23%) 영역의 비중을 기반으로 도출



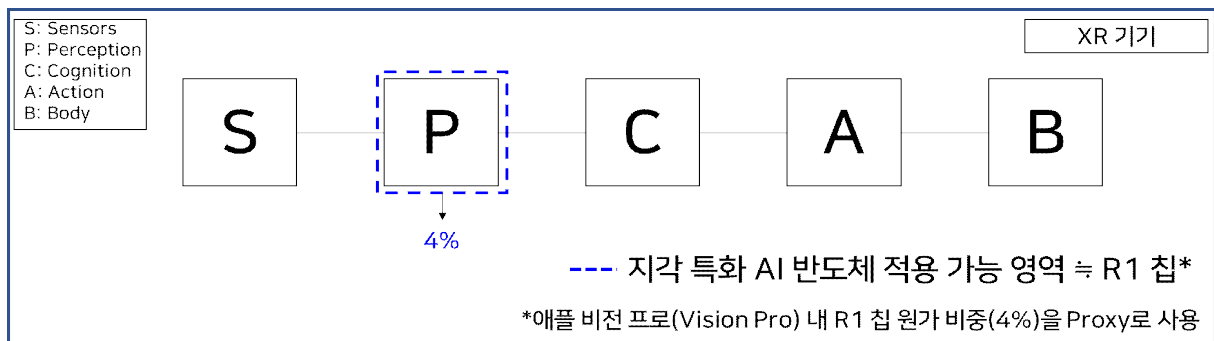
[ 휴머노이드 로봇 내 지각 특화 AI 반도체 적용 비중 도출 방법 및 결과 ]

※ (자율주행차) L2+ 자율주행차 평균 판매가격에서 센서 퓨전 모듈의 비중 (1.1-1.5%)과 `35년 글로벌 자율주행차 시장 규모 추정치에서 자동차용 센서 퓨전 시장이 차지하는 규모 2.4%를 고려하여 최종적으로 2%를 목표시장 규모 추정을 위한 대용치로 사용



[ 자율주행차 내 지각 특화 AI 반도체 적용 비중 도출 방법 ]

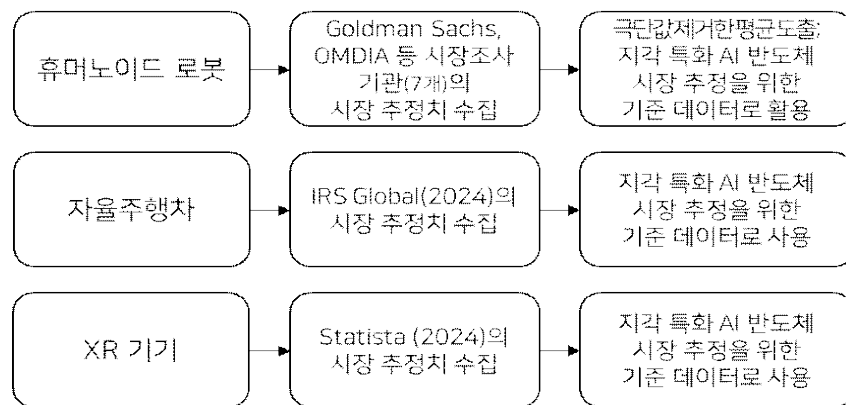
※ (XR 기기) 애플 비전 프로(Vision Pro)의 전체 원가 중 센서 데이터 처리를 전담 하는 연산 장치(R1칩)가 차지하는 원가 비중인 4%를 시장 규모 추정을 위한 대용치로 사용



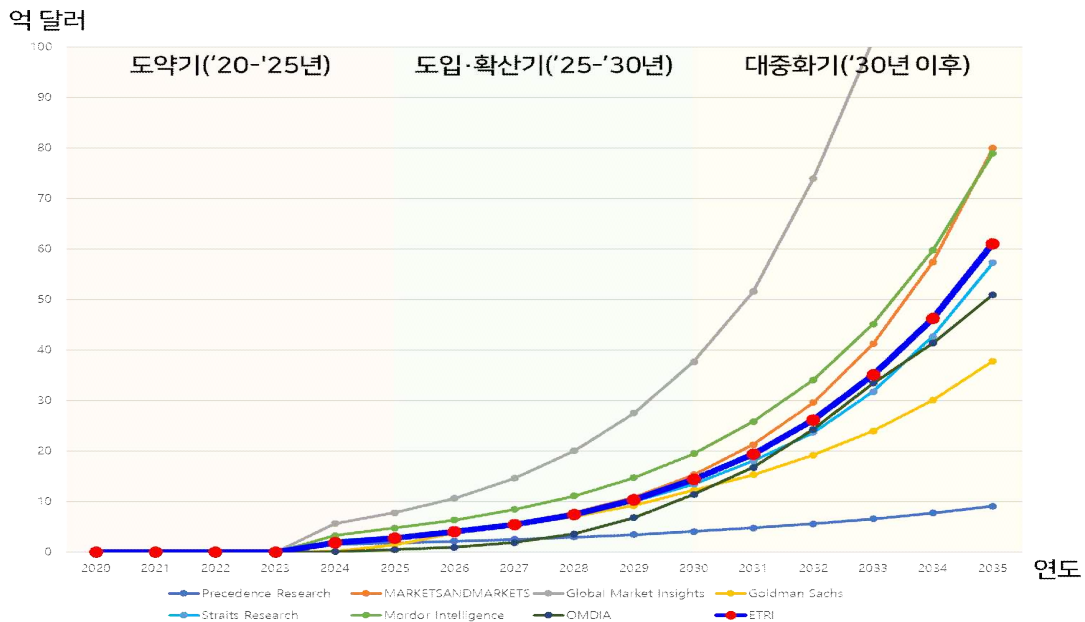
[ XR 기기 내 지각 특화 AI 반도체 적용 비중 도출 방법 ]

● (Phase 3) 글로벌 시장 전망 데이터 수집 및 기준 데이터 도출

- ※ (휴머노이드 로봇) 시장조사기관별 예측 편차를 보정하기 위해 상·하위 극단값을 제거한 절사평균(Trimmed Mean)을 적용하여 `25년 280억 달러(39.2조 원)<sup>8)</sup>에서 `35년 610억 달러(85.4조 원)로 연평균 37.6% 성장 전망
- ※ (자율주행차) 글로벌 자율주행차 시장은 `25년 1,500억 달러(210조 원)에서 `35년 1조 1,200억 달러(1,568조 원)로 연평균 22% 성장 전망
- ※ (XR 기기) 글로벌 XR 기기 시장은 `25년 466억 달러(65.2조 원)에서 `35년 951억 달러(133.1조 원)로 연평균 8.3% 성장 전망



[ 활용 분야별 시장 규모 추정 데이터 수집 방법 ]



[ 글로벌 휴머노이드 로봇 시장 규모 추정 결과 (주요 기관별 예측치 포함) ]

8) 환율 1,400원 기준, 이하 동일

● (Phase 4) 시장 진입 시점 분석

- ※ 기술 공급(Supply): 개발 리드타임 고려
  - NTIS 과제(3.5년), TRL 3-7단계 R&D 기간(3.8년), 상용 제품 사례(5년) 분석 결과, 5년 이상의 개발 기간 필요
- ※ 시장 수요(Demand): 확산 주기 고려
  - `25-`27년(산업용 도입) ⇒ `28-`31년(서비스용 대중화) 본격 시장 확대
- ※ 최적 시점(Targeting): 선제적 대응 전략
  - (위험) 시장 개화 후 착수 시, 개발 기간 소요로 인해 핵심 성장기(`28-`31년) 진입 불가
  - (전략) `26년 R&D 착수 → `30년 기술성숙도·아키텍처 완성 → `31년 시장 진입의 선도적 일정 수립을 통해 골든타임 확보

(2) 목표시장 규모 추정과 산업 생태계 현황 분석

가. 지각 특화 AI 반도체 목표시장 규모 추정

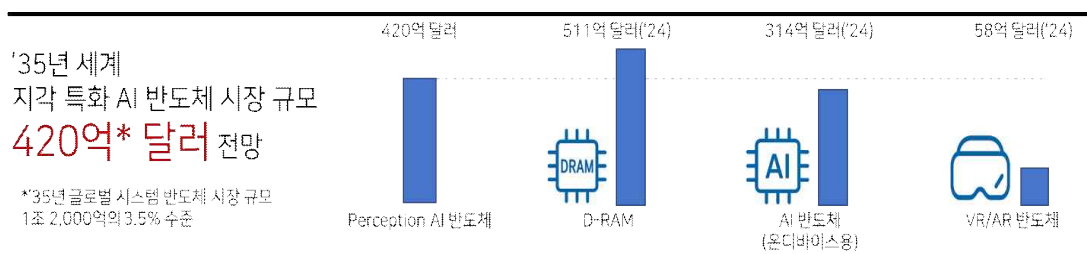
- 휴머노이드 로봇과 자율주행차의 수요 확대에 힘입어 `31년 180억 달러(25.2조 원) → `35년 420억 달러(58.8조 원)(CAGR 23.7%) 성장할 전망

< 지각 특화 AI 반도체 목표시장 규모 (단위: 십억 달러/%) >

구분	`30	`31	`32	`33	`34	`35	CAGR
휴머노이드 로봇	-	5 (27.8%)	7 (31.8%)	9 (33.3%)	12 (35.3%)	16 (38.1%)	33.1%
자율주행차	-	10 (55.5%)	12 (54.6%)	15 (55.5%)	18 (52.9%)	22 (52.4%)	22.3%
XR 기기	-	3 (16.7%)	3 (13.6%)	3 (11.2%)	4 (11.8%)	4 (9.5%)	7.4%
합계	-	18 (100.0%)	22 (100.0%)	27 (100.0%)	34 (100.0%)	42 (100.0%)	23.7%

● 유사 시장 비교를 통한 잠재력 평가

- ※ `35년 지각 특화 AI 반도체 시장은 420억 달러(58.8조 원)로 전망



[ 지각 특화 AI 반도체 목표시장 규모 ]

## 나. 산업 생태계 현황 분석을 통한 신시장 창출 기회 탐색

### ● 전략적 공백(White Space) 공략 : 과점화된 양 끝단 사이의 기회

- ※ (지배구조) 온센서 AI와 온디바이스 AI는 이미 고도로 과점화된 시장으로, 기존 강자들이 독점적 하드웨어 표준과 특허 장벽을 통해 폐쇄적 생태계를 공고화하고 있어 신규 진입자의 기술 주도권 확보가 제한적임
- ※ (공백 영역) 두 거대 생태계 사이에서 대용량 멀티모달 데이터의 융합·전처리를 담당하는 중간 계층은 지배적 강자가 부재한 전략적 공백 영역임
- ※ (기회) 기존 강자들 사이의 아키텍처 공백을 선점함으로써, 특정 하드웨어에 종속되지 않는 개방형 생태계 구축이 가능

### ● 경쟁우위 확보 가능성 : 설계 효율 중심의 경쟁

- ※ (핵심 경쟁력) 옛지 지각 시장의 성패는 미세 공정 경쟁보다 이종 멀티모달 데이터의 실시간 정합 및 특징 추출을 위한 전용 아키텍처 설계 역량에 의해 좌우될 것으로 판단
- ※ (국내 기업의 기회) 데이터센터용 AI 대비 자본집약도가 낮고 설계 효율성이 핵심이어서, 국내 팹리스 기업이 기술 차별화를 통해 글로벌 점유율을 확보하기에 유망한 분야임
- ※ (포지셔닝) 기존 강자의 점유율을 직접 대체하기보다, 시스템 전체 효율을 높이는 기능적 보완재(Functional Complement)이자 파트너로 포지셔닝할 필요가 있음
- ※ (진입 시점) 상용화 소요기간을 고려할 때, 시장 개화기인 `31년을 목표로 현시점부터 조속한 R&D 착수가 필요

## 4. 신시장 창출과 성장 견인을 위한 정부 역할

### (1) 국가전략기술로서 정부의 투자 필요성

#### ● 시장 성장성과 산업 파급력

- ※ 지각(Perception) 특화 AI 반도체 시장 추정 규모는 글로벌 시스템반도체 시장(약 1조 2,000억 달러)의 3.5% 수준으로, 국가 성장동력으로서 높은 매력도를 보유

#### ● 피지컬 AI 시대의 기술 패권 경쟁 대응

- ※ 센서 수 증가와 데이터 폭증으로 GPU 등 고성능 칩 수요가 급증하며, 이는 공급망 종속 및 기술 디커플링 위험으로 직결
- ※ 센서 및 AI 모델 변화에 범용적으로 대응 가능한 ‘AI 센서 칩/플랫폼’ 확보는 기술 주권 관점에서 국가적 필수 과제임

#### ● 시스템반도체 분야 신성장동력 발굴 측면

- ※ 한국은 시스템반도체 글로벌 점유율이 약 3% 수준으로 제한적임
- ※ 로직·센서·광학 계열과 연계된 지각 특화 AI 반도체는 국내 시스템반도체 산업의 전략적 타깃 시장을 명확히 설정하고, 글로벌 진입 기반을 마련할 핵심 분야임

### (2) 신시장 창출과 성장 견인을 위한 정부의 투자전략 방향성

#### ● 기술 개발 전략 방향성

- ※ 대량 양산을 위한 설계 최적화 및 공정 기술 확보, 칩렛·IP 기반의 모듈형 패키징 기술이 필수적임
- ※ HW-SW 통합 플랫폼화 및 표준화 전략을 통해 시장 확산의 구조적 기반을 마련해야 함

#### ● 시장 진입 전략 방향성

- ※ 우선 활용시장(로봇, 자율차 등)을 중심으로 호환성과 확장성을 강화하여, 다양한 피지컬 AI 제품군으로 확산해야 함
- ※ 공정 최적화 → 수율 제고 → 대량생산 체계 구축으로 비용 우위를 확보하고 글로벌 경쟁력을 강화해야 함



# I 피지컬 AI 시대의 도래와 기술 패러다임 전환

## 1. 피지컬(Physical) AI 확산 배경과 기술적 의미

- CES 2025를 기점으로 피지컬(Physical) AI가 차세대 AI 패러다임으로 부상하며, 인공지능이 물리적 세계와 직접 상호작용하는 시대가 도래
- 휴머노이드 로봇, 자율주행차, XR 기기 등은 이제 단순한 데이터 처리 수준을 넘어, 센서를 통해 현실을 지각하고 즉시 판단·행동하는 실시간 지능 시스템으로 진화 중
- 이러한 피지컬 AI 확산 배경과 기술적 특성을 고찰하여, 지각(Perception)에 특화된 AI 반도체의 필요성을 도출하고자 함

### ■ 피지컬(Physical) AI 정의

- 현실 세계에서 직접 상호작용(물리적 세계를 이해하여 추론하고 행동)하는 AI
- 물리적 장치(로봇, 센서, 액추에이터 등)와 결합해 현실 세계에서 직접 상호작용하는 AI 기술  
- 지각/인지(센서 등으로 환경 파악) → 생각(상황 판단 및 계획) → 행동(실제 물리장치 제어)의 전 과정을 하나로 포괄하는 개념

<표 1> 피지컬 AI 정의

	구분	정의
학계	Miriyev & Kovac(2020)	지능형 유기체와 관련된 작업을 수행할 수 있는 물리적 시스템을 만드는 이론과 실제
	Yingbo Li(2021)	센서·액추에이터·AI 알고리즘이 통합된 시스템으로, 물리적 세계에서 자율적으로 작동하는 지능형 시스템
	Yang Liu(2024); SJ Oks(2024); Zikai Zhao(2024)	기존의 Embodied AI(체화된 AI), 소프트 로보틱스(Soft Robotics), 사이버물리시스템(Cyber-Physical System, CPS), 적응형 AI(Adaptive AI) 등의 개념들과 밀접하게 연결
연구계	AI4EU(2020)	물리적 환경과의 직접 상호작용이 수반되는 문제를 해결하는 AI 기술을 활용
	WEF(2025)	기계공학, AI, 센서, 연결성의 융합을 통해 현실 세계에서 자율 작동하는 시스템
산업계	NVIDIA(2025e)	현실 세계의 복잡한 행동을 인식, 이해 및 수행하는 자율시스템
	Google Deepmind(2025)	AI가 인간처럼 물리 환경에서 안전하고 유연하게 상호작용할 수 있도록 진화해야 한다는 관점
	USAI(2025)	AI가 데이터 처리를 넘어 실제 물리적 상호작용을 지능적으로 수행하는 기술

출처: 이해수 외(2025)<sup>9)</sup>의 내용

■ 피지컬(Physical) AI 시대의 도래

- 피지컬 AI는 AI가 디지털 영역을 넘어, 실물 세계의 다양한 산업 현장에서 직접적인 의사결정과 행동을 수행하는 새로운 패러다임의 전환점을 의미함
  - AI가 물리적 환경과 결합하여 진화한 복합적 기술 진화의 결과물
- 센서, 로봇, 시뮬레이션, 학습 기술의 융합을 통해 산업 구조 전반의 근본적 변화를 촉발할 기술적 전환점으로 평가
  - 휴머노이드 로봇, 자율주행차, 무인 드론 등이 주요 구현 사례로 주목받고 있음
  - 핵심적인 기술 요소 : 다양한 센서, 실시간 제어 및 환경 적응 학습, 강화학습 및 모방학습 기반의 AI 모델, HW(로봇, 모터, 액추에이터 등)와의 실시간 융합 등
- 이러한 특성으로 인해 피지컬 AI는 전혀 새로운 차원의 기술 패권 경쟁 국면을 형성할 가능성이 높음<sup>10)</sup>



[그림 1] 피지컬 AI 시대의 기술 패권 경쟁

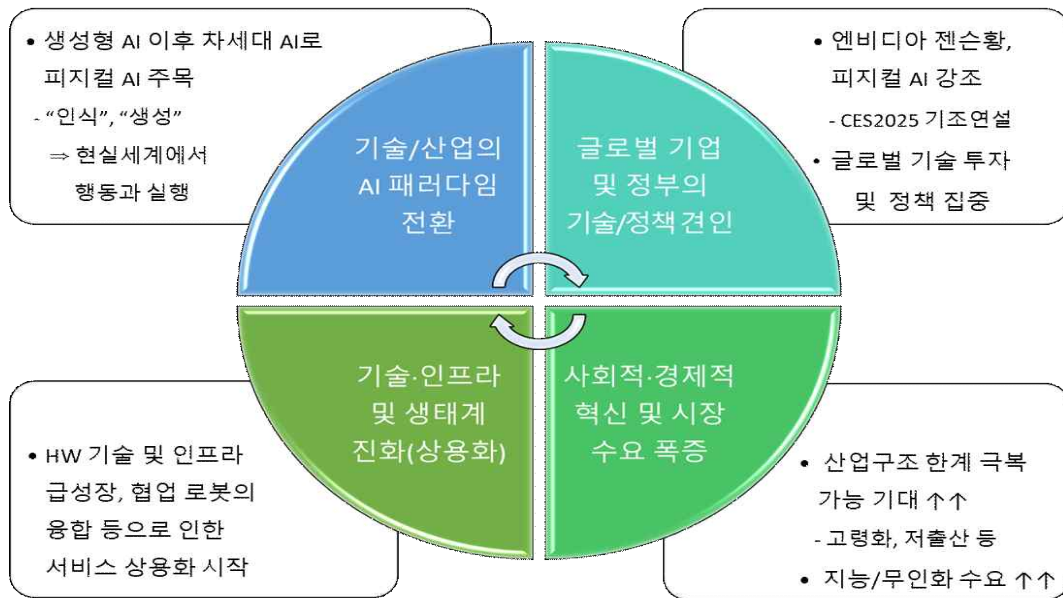
■ 피지컬(Physical) AI 최근 동향

- 피지컬 AI가 최근 주목받는 이유
  - (기술/산업의 AI 패러다임 전환) 인식 및 생성 중심의 생성형 AI 이후 현실세계에서 행동과 실행 중심의 차세대 AI로 피지컬 AI 주목
  - (글로벌 기업 및 정부 기술/정책 견인) [그림 1]에서 확인할 수 있듯이, 글로벌 기술 기업 및 정부 차원의 집중적 투자 및 정책 견인

9) 이해수 외. (2025.4.). 피지컬 AI의 현황과 시사점. 소프트웨어정책연구소

10) 양성욱. (2025.8.). 피지컬 AI 개발 전략 中

- (사회적·경제적 혁신 및 시장 수요 폭증) 고령화·저출산 등 사회 구조 변화에 따른 지능화·무인화 수요 급증
- (기술·인프라 및 생태계 진화) HW·인프라 발전과 로봇 융합을 통한 서비스 상용화 본격화
- 이러한 요소들로 인하여 차세대 산업 및 사회 인프라 혁신의 핵심으로 피지컬 AI가 부상



[그림 2] 피지컬 AI가 최근 주목받는 이유

○ 피지컬 AI 관련 최근 국내외 움직임

- 피지컬 AI 글로벌 얼라이언스 콜로키움(`25.7) 및 얼라이언스 공식 출범(`25.9)
  - 자동화 기계와 로봇, 소프트웨어의 융합 기술인 피지컬 AI(Physical AI)를 중심으로 한 글로벌 협력체
  - 대한민국 피지컬 AI 생태계의 전방위 혁신을 이끌고 산·학·연·관 협력 플랫폼 역할을 목표로 5개 생태계 분과와 5개 도메인 분과로 구성<sup>11)12)</sup>
  - `25년도 2차 추가경정 예산안에 피지컬 AI 핵심기술 PoC 사업 예산안 포함 등 ; 426억 원<sup>13)</sup>
- 피지컬 AI 협회 출범 (`25.6)
  - 산업 자동화와 국방 혁신 핵심 기술로 평가되는 피지컬 AI 산업 활성화를 위한 민간 주도 목적<sup>14)</sup>

11) 로봇신문 (2025.9.)  
 12) 공동의장 7인: 과학기술정보통신부 장관, 산업통상자원부 장관, 중소벤처기업부 장관, 국회 과학기술정보방송통신위원회 정동영 의원, 최형두 의원, 한국인공지능소프트웨어산업협회장, 한국자동차모빌리티산업협회장  
 13) 조선비즈 (2025.7.)  
 14) 전자신문 (2025.6.)

- 피지컬 AI SW, AI 반도체, 로봇/완제품, 소재·부품·장비, 그리고 제조·농업·건설·국방·의료 등 주요 산업 전반을 회원사로 포괄하는 구조
- CES 2025 핵심 주제 : “Physical AI is Here.”
  - 휴머노이드 로봇, 산업로봇, 자율주행 등 전 산업에서 실제 데이터를 활용한 AI와 스마트 로봇이 화두<sup>15)</sup>
  - 엔비디아 : `25년 CES에서 ‘코스모스(Cosmos)’, ‘블랙웰(Blackwell)’ 등 피지컬 AI를 위한 플랫폼 및 차세대 하드웨어를 공개하며, 로봇, 자율주행 등 다양한 영역에서 폭 넓은 기술 진보를 주도
  - 테슬라, BMW 등: 글로벌 제조 및 자동차 기업은 휴머노이드 로봇과 자율주행 솔루션 시범 도입 중
- 엔비디아 GTC 2025 : 최대 AI·로봇 컨퍼런스 GTC(워싱턴 D.C.)에서 피지컬 AI 주제 세션 예정 (`25.10)<sup>16)</sup>

#### ■ 피지컬 AI 시장 성장 추세 및 도래 시기

##### ○ 폭발적 도약기 (`20년~`25년)

- 글로벌 피지컬 AI 및 로봇틱스 시장은 `20년 약 50억 달러에서 `25년 225억 달러로 급성장(AI Matters, 2025.5.)
- `30년에는 약 643억 달러(약 85조 원, 연평균성장률 : 23.3%)로 전망

##### ○ 실질적 도입·확산기 (`25년~`30년)

- 제조, 물류, 헬스케어 등에서 휴머노이드 로봇, 자율주행 시스템, 스마트 제조 등이 본격적으로 상용화되는 시기로 구체적 변화가 나타남
  - 테슬라, 엔비디아, 아마존 등이 AI와 로봇, 자율제어 SW의 결합 솔루션 출시
- 로봇-물류-공장-헬스케어 현장에서 실제 배치 사례가 급증

##### ○ 대중화기 (`30년 이후)

- 피지컬 AI 기반 휴머노이드, 지능형 로봇, 자율 운송 등이 병원, 건설, 농업, 서비스 등 다양한 산업에 확산되면서 사회 구조적 혁신을 이끄는 주축으로 성장
- `35년에 전 세계적으로 13억 대 이상의 AI 로봇이 `50년까지는 40억 대 이상이 실제 사회에 확산될 것으로 전망(<표 2> 참고)
  - 휴머노이드 로봇 시장의 경우에는 기술적 가능성, 사용 사례, 그리고 경제성을 고려할 때 `50년까지 6억 4,800만 대로 성장할 것으로 예상<sup>17)</sup>

15) Global X (2025.1.)

16) 엔비디아 GTC 웹페이지. Physical AI and Robotics Conference Sessions

17) Citi Group. (2024.12.)의 예측치이며, 본 연구에서는 해당 기관에서 로봇 유형을 정의하고 구분하는 기준 자체보다

- 로봇의 단위 가격의 하락을 고려한 총 시장 규모는 `35년까지 2,090억 달러, `50년까지 7조 달러에 이를 것으로 예상

〈표 2〉 유형별 로봇 수 예측 전망치

로봇 유형	`24	`25	`30	`35	`50	CAGR
Autonomous Vehicle	27	34	126	380	1,858	17.4%
Domestic Cleaning Robot	286	326	541	793	1,188	5.3%
Humanoids	0	0	1	13	648	60.7%
AGV & AMR	2	3	9	28	181	17.7%
Drone	37	40	54	73	149	5.4%
Caring Robot	0	1	6	18	71	20.0%
Commercial Cleaning Robot	1	2	6	14	25	10.6%
Food & Grocery Delivery Robot	0	0	4	11	19	19.2%
Food Service Robot	0	0	3	10	15	17.2%
Caring Overlap	0	0	(1)	(5)	(18)	
합계	354	405	749	1,337	4,136	9.7%

출처: Citi Group, THE RISE OF AI ROBOTS: Physical AI is Comming for you, 2024.12.

- 이외에도 새로운 피지컬 AI 생태계는 `27년~`32년에 시장과 기술이 본격적인 성숙 단계로 돌입할 것으로 전망<sup>18)</sup>되고 있음
  - SXS<sup>2025</sup><sup>19)</sup> 행사에서 현장 발표 및 다수의 벤처 캐피털리스트들의 분석에 의하면, AI 현재 인프라 구축 단계는 데이터센터에 집중되어 있고, 2~7년 안에 성숙될 것으로 예상되는 피지컬 AI 단계로 연결
  - 5~10년 후 예상되는 피지컬 AI가 과학의 한 분야로 정착될 것으로 예상

**■ 피지컬 AI의 기술 특성 : 현재 언어중심 생성형 AI와의 차별성**

- 피지컬 AI 기술은 시스템(디바이스)의 직접적인 행동, 실행, 적응 등이 결합된 실제 물리 법칙이 적용되는 현실 세계 지능을 구현하기 위해서 물리적 현실과 디지털 AI의 본격적인 융합 AI의 핵심 요소들로 구성
  - 이러한 특성 때문에 언어중심의 현재 AI와 <표 3>과 같은 차별적 특성을 지님

전체 합계의 예측치에 대해서 고찰함

18) Forbes. (2025.4.). The Dawn Of Physical AI : Industry Leaders At SXS<sup>2025</sup> Predict Its Impact. Melody Brue

19) South By Southwest가 매년 개최하는 세계적인 유명한 융복합 컨퍼런스

<표 3> 현재(언어중심) AI와 미래(피지컬) AI의 차별화 포인트

차별화 포인트	현재(언어중심) AI	미래(피지컬) AI
물리 세계와의 직접적 상호작용	디지털 데이터 분석, 언어 이해, 이미지 인식 등 가상적·비물리적 환경에서만 작동	센서·로봇·엣지컴퓨팅 등 물리적 장치와 결합해 실제 환경에서 직접적으로 행동하고 현실 세계에 개입가능
환경 지각 → 인지·판단·계획 → 실시간 행동의 완전한 루프 <sup>20)</sup> 구현 <sup>21)</sup>	데이터 입력 → SW적 판단이나 예측만 제공 (의사결정 지원, 예측, 추천 등)	센서로 실시간 환경 지각→ 상황 인지 및 판단(행동 계획)→ 실제로 행동(물체 조작 또는 이동 등)까지 전 주기의 통합 시스템
물리 법칙 내재화 <sup>22)</sup>	텍스트·이미지·음성 등 비물리적 데이터 위주로 학습	마찰, 힘, 충돌, 관성 등 실제 물리 법칙을 내재적으로 이해·예측하여, 현실에서 다양한 변수에 안전하게 대응
실시간성 및 적응력	대부분 오프라인 또는 서버에서 데이터 분석	실시간 현실 변화를 센서·로봇이 즉각 반영, 예측 결과에 따라 행동을 실시간 수정하며 환경에 유연하게 적응
자율적 자기 진화 (강화학습 등)	자기생성 데이터와 자체 피드백 기반의 강화학습을 통해 실시간 업데이트 및 진화	현실 세계와 상호작용하며 환경 변화에 실시간 대응하는 자기진화형 물리학습(강화학습)
중심산업 및 사회적 효과 <sup>23)</sup>	디지털 산업 및 업무 효율화 중심 효과	현실 산업 현장 지능화 및 일상 생활 변화 중심 효과

출처: ETRI 기술전략연구본부 작성, 2025.11.

20) 엔비디아 GROOT, Cosmos 플랫폼 등이 센싱-판단-계획-동작-피드백의 실시간 루프를 갖춘 대표적인 사례

21) Brian Caulfield. (2025.1.). CES 2025: AI Advancing at 'Incredible Pace,' NVIDIA CEO Says

22) 엔비디아 보도자료. (2025.1.). NVIDIA Expands Omniverse With Generative Physical AI

23) 전 세계 물리 AI 시장은 726억 달러 규모이며, 2030년까지 두 배 이상 성장하여 1,660억 달러에 달할 것으로 예상되며, 연평균 성장률은 16.4%에 달할 것으로 예상 (출처: Eximius Ventures. (2025.5.). Physical AI and Its Role in Shaping Industry 5.0)

[참고 1] 엔비디아 젠슨 황이 말하는 피지컬 AI의 핵심 기술<sup>24)25)</sup>

○ 물리적 세계 이해(Physical Reasoning)

- 물리 법칙 통합 : 마찰, 관성, 인과관계, 물체의 지속성(Object Permanence) 등 물리학의 기본 원리를 AI가 내재화해야 한다고 강조. 이로써 AI는 실제 환경에서 사물의 움직임, 충돌, 힘 조절 등의 물리적 상호작용을 예측, 처리 가능

○ 시뮬레이션 및 예측 기반 데이터 처리

- 3D 시뮬레이션 플랫폼 : 엔비디아가 발표한 ‘코스모스(Cosmos)’와 ‘옵니버스(Omniverse)’ 등은 실제 세계의 동적 특성을 이해하도록 AI를 훈련시키는 핵심 인프라이며, 코스모스는 2,000만 시간 이상의 물리 기반 비디오 데이터로 학습하며, 미래 상태를 영상으로 예측해 실제 로봇/AV(자율주행) 시스템의 의사결정에 활용 가능
- 물리 기반 강화학습: AI가 가상 환경에서 반복적으로 행동과 피드백을 겪으며 실제 환경의 물리법칙을 스스로 익히도록 하는 것이 중요

○ 로봇 및 엣지컴퓨팅 하드웨어

- 실세계 적용 하드웨어 : 강력한 GPU(Blackwell, Rubin 등), AI 칩, 엣지 컴퓨팅 시스템이 피지컬 AI 모델이 실시간 의사결정과 제어를 하도록 지원 ; 엔비디아의 DGX 컴퓨터, RTX 50 시리즈, 그리고 신형 로봇 하드웨어가 대표적임
- 로봇-센서 융합 : 다양한 센서(LiDAR, 카메라, 힘 센서 등)와 머신비전 기술로 로봇이 주변환경을 정밀하게 인지하고, 실시간 데이터와 물리적 판단을 결합

○ 지능적 에이전트와 자율성

- Agentic AI(주도적 AI) : 복합 환경에서 목표를 설정 및 분해하고, 외부 데이터와 도구를 능동적으로 활용하며, 연속적인 계획-행동-최적화 과정을 반복할 수 있는 능력이 필요

○ 표준화와 생태계

- 오픈 생태계 및 데이터셋 : 오픈소스 모델, 물리 시뮬레이션 데이터셋, 세계적 기업 간 협력 등 생태계 조성도 피지컬 AI의 성숙에 필수적인 기반

24) CNBC. (2025.7.). Nvidia CEO: If I were a 20-year-old again today, this is the field I would focus on in college  
 25) The Economic Times. (2025.7.). Nvidia CEO Jensen Huang would not have studied computer science today if he were a student today. He urges mastering the real world for the next AI wave

## 2. 물리법칙(Physical World) 기반 지각 연산 환경의 변화

### 가. 피지컬 AI 시대 물리법칙 환경의 센싱 데이터 특성 변화

- 피지컬 AI는 앞서 설명한 바와 같이 휴머노이드 로봇, 자율주행차, XR(확장 현실) 기기 등 물리적 환경과 상호작용하며 실시간으로 지각, 판단, 반응하는 AI 시스템
- 기존의 중앙 집중형 SoC(System-on-Chip) 구조는 향후 폭증하는 센싱 데이터를 처리하기에는 데이터 병목, 지연, 전력 소모량 증가 등의 구조적 한계를 노출할 가능성이 존재함
- 물리적 세계와 직접 상호작용하는 피지컬 AI는 실시간성, 신뢰성, 복잡한 환경 대응을 동시에 요구하므로 그 설계 및 구현 난이도가 현저히 높아짐

### ■ 물리법칙 환경(Physical World)이 반영된 센싱 데이터 특성 변화

- 데이터 규모 폭증
  - 밀리초 단위로 변화하는 현실 환경을 지속적으로 센싱함에 따라 데이터 규모가 급증함
  - 고해상도 비전 센서와 센서 수 증가로 데이터 생성량은 기하급수적으로 확대
  - 자율주행차는 차량 1대당 초당 수 GB~수십 GB의 데이터를 생성
    - 예를 들어, 구글의 Waymo처럼 비전센서의 퓨전 형태로 운행하는 차량은 초당 약 10GB 이상의 데이터를 생성
- 데이터 유형 다양화 : 단일(이미지) → 멀티모달
  - 이미지 중심 단일 데이터에서 벗어나 영상, 음향, 모션, 촉각 등 멀티모달 데이터가 동시에 생성
    - 비전(카메라, 레이더, 라이다 등), 모션, 음향 등 다양한 센서 출력이 서로 다른 형식·해상도·스케일로 존재
    - 예를 들어, 카메라의 경우에도 용도에 따라 RAW<sup>26)</sup>, YUV<sup>27)</sup>, MPEG<sup>28)</sup>, AER<sup>29)</sup> 등 매우 다양함
  - 센서별 출력 형식, 해상도, 시간 해상도가 상이해 이질적 데이터 처리와 통합이 필수적임

26) Raw Image Data : 이미지 센서가 취득한 가공 전 원시 데이터

27) Luminance-Chrominance Color Space : 밝기(Y)와 색차(U, V)를 분리한 영상 색공간 포맷

28) Moving Picture Experts Group : 영상 압축 표준을 정의한 국제 표준 규격

29) Address-Event Representation : 픽셀 변화 이벤트의 주소와 시간 정보만 비동기적으로 전송하는 표현 방식

○ 데이터 생성주기의 단축 및 불규칙성

- 인간 및 환경과의 상호작용으로 데이터가 실시간·비주기적으로 생성
- 다수 센서의 동시 동작으로 데이터 생성 주기가 불규칙해짐

○ 멀티모달 데이터의 비정합과 동기화 문제

- 센서 간 시간·공간 차이로 데이터 비정합이 발생하며, 정밀한 동기화가 필수 요건으로 부상
- 가장 느린 센서가 전체 처리 지연을 유발하는 병목 요소로 작용할 수 있음
  - 카메라 센서의 경우도 서로 다른 카메라의 프레임 속도와 타임스탬프 불일치로 인해 퓨전 데이터의 정합성이 저하되는 경우가 많음

■ 센서 다중화와 실시간 처리 요구의 증가

○ 고품질 지각을 위해 카메라, LiDAR, Radar, 마이크, IMU 등 다양한 센서의 다중화 및 고집적이 가속화

- 완전자율주행차, 휴머노이드 로봇, XR 기기 등과 같이 현실(3D)과 실시간 상호작용하는 기기들이 증가 : 이러한 디바이스들은 카메라, LiDAR, Radar, 마이크, IMU 등 10종 이상 센서를 동시에 사용할 것으로 예상

○ 센서 단가 하락과 기술 발전으로 다수 센서를 탑재한 피지컬 AI 디바이스가 보편화되고 있으나, 이는 실시간 처리 병목을 동시에 유발

○ 자율주행, 로봇, 무인체계 등에서는 ms 단위의 초저지연성과 실시간 처리가 필수적임

- 특히 생명과 안전에 직결되는 중대(Mission-Critical) 서비스, 인간 또는 기계와의 실시간 상호작용, 자율 협력 또는 지능화 무인 협업 등을 위해서는 센싱 정보의 실시간 처리 필요성 증가
  - 자율주행차의 경우 0.1초 이내 판단·제어 실패가 안전 문제로 직결
  - 스마트팩토리의 경우 시간 지연 자체가 불량이거나 수율 저하 등 제품 품질에 치명적인 문제를 발생
- 다중 센서 퓨전 환경에서는 모든 센서 처리가 동시에 고속으로 이루어져야 함

■ 복합 물리 환경에서의 다중 센서 융합 필요

○ 피지컬 AI는 카메라, LiDAR, IMU 등 다양한 센서 정보를 통합 분석해 환경을 인식해야 함

- 자율주행차의 경우, 도로의 차선, 보행자, 신호등, 그리고 다른 차량의 움직임을

동시에 인식하고 이를 바탕으로 안전한 주행 경로를 계산해야 함

- 이는 단순 생성형 AI 대비 훨씬 높은 연산 복잡도와 신뢰성을 요구함
  - UWS(무인전투체계)의 경우, 피아식별은 매우 정교한 판단이 필요한데, 적과 아군을 잘못 구분하면 인명 피해로 이어질 수 있으므로, 시스템은 열악한 환경(예: 안개, 연기, 전자전 환경)에서도 정확한 판단을 내려야 함
- 원시 데이터 증가로 인해 전처리(노이즈 제거, 정규화, 특징 추출)의 중요성이 급격히 확대됨

## 나. 일반 엣지(Edge) 디바이스보다 더 높은 퍼지컬 AI 시스템의 요구성능

### ■ 독립적이고 신속한 의사결정 요구

- 퍼지컬 AI는 네트워크 지연이나 불안정과 무관하게 디바이스 내부에서 독립적인 지각·판단 능력을 확보해야 함
  - 데이터센터 또는 클라우드 환경에서의 의사결정이 아닌 대부분 엣지 디바이스 형태의 퍼지컬 AI 시스템에서는 디바이스 내에서 독립적인 지각 능력 필요
  - 네트워크의 지연 및 불안정에도 실시간성 판단 및 대응이 필요하며, 이를 통한 퍼지컬 AI 시스템의 안정성과 연속성 확보가 가능
- 예측 불가능한 물리 환경에서 다중 센서 데이터를 실시간 융합·처리해 즉각적인 의사결정이 필요함
  - 물리적 환경은 예측 불가능한 변수(날씨, 조명, 장애물의 갑작스러운 등장 등)로 다양하게 구성
  - 물리적 환경은 고차원적이고 다중 모달(Multi-Modal) 데이터를 다루므로, AI는 다양한 데이터 스트림을 융합하고, 이를 기반으로 즉각적인 의사결정이 필요
  - 이는 연산 부담을 크게 증가시키며, 지연이 1ms만 증가해도 치명적 결과를 초래
  - 생성형 AI가 단일 모달(예: 텍스트나 이미지)에 초점을 맞춘 것과 달리, 훨씬 더 복잡한 데이터 처리 프로세스를 필요로 함

### ■ 엣지 디바이스 내 강력한 AI 처리 능력 요구

- 클라우드(데이터센터) 의존도를 최소화하고 지연을 줄이기 위해 엣지 디바이스 단에서의 고성능 AI 연산이 필수적

- 퍼지컬 AI는 동적·불확실한 환경에 지속적으로 적응해야 하며, 온라인 학습 및 적응 제어 등 고급 알고리즘이 요구됨
  - 물리적 환경은 날씨, 조명, 장애물, 인간의 행동 등 예측 불가능한 요소로 구성
    - 자율주행차의 경우, 강우 시 도로 반사광을 오인식하지 않아야 하며, UWS는 전자전 환경에서의 신호 간섭을 극복해야 함
  - 이는 학습 데이터셋과 같이 상대적으로 정적인 환경에서 동작하는 기존 생성형 AI와 달리, 퍼지컬 AI가 동적·불확실한 환경에 지속적으로 적응해야 함을 의미
  - 퍼지컬 AI는 환경 변화에 따라 실시간 학습 또는 적응이 요구
    - 자율주행차는 새로운 도시 환경에 배치될 때마다 현지 데이터 기반의 모델 미세 조정(Fine-tuning)이 필요
  - 이러한 불확실성은 AI가 새로운 데이터를 지속적으로 수용·처리해야 함을 의미
  - 이는 높은 연산 성능과 온라인 학습·적응 제어 등 복잡한 알고리즘을 요구하여 시스템 설계의 복잡성을 크게 증가시킴

■ 신뢰성, 안정성, 에너지 효율의 동시 확보 필요

- 잘못된 판단은 인명 피해로 이어질 수 있어 높은 신뢰성과 안전성이 필수 요건임
  - 자율주행차나 로봇, UAV, UWS는 잘못된 판단이 사람의 생명을 위협할 수 있음
    - 예를 들어, 테슬라의 오토파일럿 시스템이 장애물을 잘못 인식해 사고를 일으킨 사례는 퍼지컬 AI의 높은 신뢰성 요구를 보여줌
  - 특히, 사람과 직접 상호작용하거나 AI 간의 자율협력 상황에서는 훨씬 더 높은 신뢰성과 안정성이 필요
  - 반면, 생성형 AI가 잘못된 텍스트나 이미지를 생성하더라도 인명과 직결되는 문제까지 확장되진 않으며, 대체로 심각한 결과를 초래하지는 않음
- 배터리 기반 시스템이 대부분이므로 에너지 효율과 성능의 균형이 중요함
  - 퍼지컬 AI는 다수의 시스템이 배터리로 작동하므로 높은 연산량을 처리하면서도 에너지 효율이 매우 중요
    - 드론의 경우, 제한된 배터리로 장거리 비행과 복잡한 연산을 동시에 수행해야 하므로, 하드웨어와 소프트웨어 모두 최적화가 필요함
    - 엔비디아의 DRIVE 플랫폼은 초당 테라플롭스(TFLOPS) 단위의 연산을 수행하면서도 전력 소모를 최소화해야 함
  - 이는 클라우드 기반의 생성형 AI가 에너지 효율에 덜 민감한 것과 대조적임

- 장애 발생 시 안전 상태로 전이하는 Fail-Safe/Fault-Tolerant 설계가 요구되며, 이를 위해 중복 설계와 실시간 모니터링이 필요함
  - 피지컬 AI는 장애발생 시 안전대응(Safe-Fail)<sup>30)</sup> 메커니즘이 필수적이며, 시스템이 99% 이상의 신뢰성을 유지해야 함
  - 이를 위해 백업시스템(Redundancy), 장애허용(Fault Tolerance), 실시간 모니터링이 필요하며, 이는 시스템 설계를 더욱 복잡하게 만듦

■ 피지컬 AI 시스템의 한계 극복을 위한 내장형 전용 AI 프로세서 필요

- 피지컬 AI는 이동, 조향, 제동 등 물리적 제어를 직접 수행하므로, 연속적·고차원 상태 공간을 실시간으로 처리해야 함
  - 물리법칙 환경의 현실 세계에서 대표적인 피지컬 AI 시스템인 휴머노이드의 현재 기술 수준의 한계를 극복하기 위한 요구사항들이 출현<sup>31)</sup>
    - 인간 수준 물리적 운동 능력 확보를 위해서 액추에이터, 센서, 제어 알고리즘의 정확한 실시간 동기화 필요
    - 인간이 이해하는 지적 능력 확보를 위해서는 실시간 감각 데이터 인지·처리·반응, 다중감각 입력 통합, 해석 필요
  - 물리적 제어는 연속적이고 고차원적인 상태 공간에서 작동하며, 이를 위해 강화학습(Reinforcement Learning)<sup>32)</sup>이나 제어이론(Control Theory)<sup>33)</sup>을 활용한 복잡한 알고리즘이 필요
- 강화학습, 제어이론 기반 알고리즘은 막대한 연산 부담과 에너지 효율 문제를 동반
- 따라서 엣지 환경(디바이스 내)에서 요구 성능을 충족하기 위해서는 피지컬 AI에 최적화된 전용 AI 프로세서가 필수적임

30) 시스템 고장, 오류, 이상 상황 발생 시 기능 유지를 목표로 하지 않고, 사람·장비·환경의 안전을 최우선으로 한 상태로 전이하는 설계 개념. 평균 성능보다 최악의 상황 대응을 더 중시하는 개념임

31) 오현정 외 (2025.6.)

32) 에이전트가 환경과의 상호작용을 통해 보상을 최대화하는 행동 정책을 학습하는 기계학습 방법

33) 시스템의 상태를 수학적으로 모델링하여 안정성·응답성·성능을 만족하도록 입력을 설계하는 이론적 방법

### 3. 중앙집중형 구조의 한계와 새로운 연산 요구

#### 가. 중앙집중형 구조의 고성능 가속기(메인 연산 SoC)의 처리 한계

- 피지컬 AI 환경에서 요구되는 센싱 데이터의 폭증과 실시간 처리 요구로 인해, 중앙집중형 구조의 메인 연산 SoC는 성능 및 구조적 한계를 노출
- 현재 다수의 엣지 디바이스는 메인 SoC가 데이터 전처리부터 지각·판단·제어까지를 일괄 수행하는 구조를 채택하고 있으나, 피지컬 AI 환경에서는 다음과 같은 한계가 발생

##### ■ 통신과 전송 경로에서의 지연 누적

- 메인 연산 SoC는 평균적인 처리 성능은 우수하나, 최악 지연(Latency)<sup>34)</sup>과 지터(Jitter)<sup>35)</sup>가 존재하는 환경에서는 치명적인 한계를 가질 수 있음
  - 피지컬 AI에서는 평균 지연보다 최악 상황의 지연과 변동성이 핵심 성능 지표로 작용
    - 자율주행차의 긴급 제동, 휴머노이드 로봇의 균형 제어와 같이 순간적 오차가 사고로 직결되는 상황에서는 수 ms 단위 지연도 위험 요인이 될 수 있음
- 중앙집중형 구조에서는 센서 데이터가 메인 SoC까지 전달되는 과정에서 다중 버스 전송, 메모리 적재, 네트워크 라우팅을 거치며 지연이 누적
  - 네트워크 불안정성은 지터를 가중시켜 예측 불가능한 응답 지연을 유발
- 즉각적 대응이 필수적인 피지컬 AI 환경에서는 평균 처리 속도 향상만으로는 한계
- 최악 지연을 최소화하고 안정적인 실시간성을 보장할 수 있는 구조적 보완이 필요

##### ■ 메모리 대역폭 포화 및 데이터 이동의 병목

- 중앙 메모리에 모든 데이터를 적재하는 구조는 메모리 월(Memory Wall) 현상을 야기하여 시스템 성능을 제한
  - 연산 성능은 급격히 향상되었으나, 메모리 대역폭과 데이터 이동 효율이 이를 따라가지 못해 실제 성능의 상당 부분이 낭비<sup>36)</sup>
- 멀티모달 센서 데이터는 DRAM 대역폭을 초과하여 읽기·쓰기 병목, 연산 지연, 전력 소모 증가를 유발
  - 고성능 HBM과 같은 차세대 메모리도 수급·제조 공정상 병목 가능성이 존재<sup>37)</sup><sup>38)</sup>

34) 데이터 패킷이 한 지점에서 다른 지점으로 이동하는 데 걸리는 시간

35) 데이터 패킷 간 전송 지연 시간이 일정하지 않고 변동되는 현상

36) Adam Kovac. (2024.9.). Higher Density, More Data Create New Bottlenecks In AI Chips. Semiconductor Engineering

- 다수의 고해상도 카메라와 LiDAR 데이터가 동시에 유입될 경우, 공유 버스 및 인터페이스의 한계로 대역폭 포화가 빠르게 발생
  - 이로 인해 연산 유닛 활용도가 저하되고, 대기열 증가와 연쇄적 병목이 시스템 전반으로 확산
    - 8개의 4K 카메라(각 1Gbps 이상)와 LiDAR(2Gbps 이상) 데이터가 동시에 전송된다고 가정하면, PCIe 4.0(16GB/s)의 대역폭이라 해도 빠르게 포화
- 결과적으로 메모리 병목은 단순 성능 저하를 넘어, 상황 인식 지연, 제어 명령 지체, 안전성 저하로 직결되는 핵심 한계 요인임

**■ 에너지 효율성과 발열 문제**

- 대량의 데이터 이동은 연산 자체보다 훨씬 높은 에너지 비용을 수반
  - 메모리-프로세서 간, 칩 내부·외부 인터커넥트에서의 데이터 전송은 전력 소모와 발열을 급격히 증가시킴
- 범용 연산에 최적화된 메인 SoC는 멀티모달 실시간 처리에 비효율적이며, 지속적인 고부하 연산으로 인해 전력 소모·열 밀도 증가 문제가 발생
  - 추가 냉각 장치, 배터리 용량 증가 요구는 소형·이동형 디바이스에 치명적 제약임
- 배터리 기반 엣지 디바이스에서는 전력 소모 증가가 가동 시간 단축, 발열 증가는 열 관리 부담 및 성능 저하(Thermal throttling)<sup>39)</sup>로 이어짐
  - 저전력 모드 최적화는 정확도 저하<sup>40)</sup>를 동반해 실시간 응답성과 안전성 확보에 한계가 있음

**■ 모델 업데이트 한계 및 경량화의 부작용**

- 중앙집중형 구조에서는 AI 모델이 메인 SoC에 통합되어 있어, 환경 변화나 응용 요구에 따른 모델 업데이트 및 교체가 복잡함
  - 다수의 현장 디바이스에 대한 동시 업데이트는 운영 중단, 버전 불일치 등의 위험을 수반<sup>41)</sup>
- 성능·전력 절감을 위한 모델 경량화(양자화, 프루닝)는 연산 부담을 줄일 수 있으나, 무분별한 적용 시 추론 정확도 저하와 상황 인식 오류를 유발할 가능성이 있음<sup>42)</sup>

37) HBM은 기존 DRAM과 달리 3D 적층(3D-Stack) 구조와 TSV(Through Silicon Via) 기술을 활용하여 폭발적인 데이터 전송속도가 가능

38) The Economist. (2024.10.). Memory chips could be the next bottleneck for AI

39) 반도체 온도가 허용 한계 초과 시, 발열과 손상을 방지하기 위해 클럭 속도 및 성능을 자동으로 낮추는 제어 기법

40) Wevolver. 2023 Edge AI Technology Report

41) milvus.io. (2025). What are the challenges of implementing edge AI?

- 피지컬 AI는 ms 단위 판단 오류가 물리적 사고와 안전 문제로 직결되므로, 경량화 과정에서 허용 가능한 오차 범위가 극히 제한적임
- 따라서 피지컬 AI 시스템에서는 유연한 모델 업데이트 체계와 정밀 검증을 동반한 경량화 전략이 병행되지 않는 한, 실시간성과 안전성을 동시에 만족하기 어려움

<표 4> 메인 연산 SoC 기술적 한계

한계 요소	기술적 한계 내용 <sup>43)</sup>
실시간 처리 보장의 구조적인 어려움	- 멀티모달 센서들이 생성하는 비동기적 인터럽트와 데이터 스트림을 중앙 SoC에서 처리할 때, 각 센서의 엄격한 타이밍 요구사항을 동시에 만족시키기 어려움
메모리 대역폭의 물리적 한계로 시스템 성능 제약	- 고성능 MCU 기준으로도 메모리 대역폭은 1-2GB/s, VGA 해상도 카메라 하나만으로도 27MB/s, AI 추론 과정에서 50-100MB/s의 대역폭 소비 - 원시 데이터의 버퍼링, DMA 전송, 캐시 일관성 유지를 위한 오버헤드가 추가되면 실질적으로 처리가능한 센서 수는 2-3개로 제한
데이터 동기화의 복잡도가 기하급수적 상승	- 이종 센서들의 서로 다른 샘플링 주기, 클럭 드리프트 <sup>44)</sup> , 데이터 형식을 중앙에서 정합하려면 복잡한 타임스탬프 관리와 보간 알고리즘이 필요 - 센서별 지연 편차가 누적되면 센서 퓨전의 정확도가 급격히 저하되며, 이를 보정하기 위한 추가 연산이 또 다른 병목을 유발하는 악순환이 발생
전력 효율의 근본적 한계에 직면	- 엣지 디바이스에서 데이터 이동에 소요되는 에너지는 연산 자체보다 10-100배 높으며, 모든 센서 데이터를 중앙으로 전송하는 구조는 전체 전력 예산의 60-80%를 데이터 전송에 소비 - 특히 배터리 구동 디바이스에서는 이러한 비효율이 동작 시간을 수 시간에서 수 분으로 단축
확장성과 유지보수성이 현저히 저하	- 센서 수가 N개로 증가할 때 중앙 처리 복잡도는 $O(N^2)$ <sup>45)</sup> 로 증가 - 센서 추가 시 전체 시스템의 타이밍과 리소스 할당 재설계 필요 - 펌웨어 업데이트 시 모든 센서 드라이버와 처리 로직이 단일 바이너리에 포함되어 부분 업데이트가 불가능 - 하나의 센서 오류가 전체 시스템을 정지시킬 위험이 상존

출처: Wevolver(2023) 등을 참고하여 ETRI 기술전략연구본부 작성, 2025.11.

42) Wevolver. 2023 Edge AI Technology Report

43) ETRI 내 관련 기술부서 전문가들에게 기술 자문 의뢰한 내용 중 일부 내용 발췌

44) Clock Drift : 서로 다른 센서·장치의 내부 클럭이 미세한 주파수 오차로 인해 시간에 따라 점차 어긋나는 현상. 장 시간 운용 시 타임스탬프 불일치와 데이터 정합 오류를 유발

45) 입력 크기가 N으로 증가할 때 연산량이나 복잡도가  $N^2$ 으로 증가

[참고 2] ETRI가 개발중인 자율주행 버스의 例<sup>46)</sup>

- 라이다 5대, 카메라 7대가 탑재되며, 모든 센서 데이터를 10Hz로 동기화해 수집
- 매 초 약 2.3GB에 달하는 데이터가 한꺼번에 중앙 SoC로 집중
- 내부 네트워크 대역폭은 급격히 포화, 순간적인 데이터 폭주로 인해 지연 발생
- 이러한 대용량 raw 데이터를 SoC에서 직접 처리하는 과정에서 연산 자원과 메모리 접근량이 집중되어 전력 소모 크게 증가
- 이후 중앙 SoC는 각 센서의 캘리브레이션 정보(내·외부 파라미터와 시간 동기화 값)를 활용해 이 메타데이터를 정렬·융합하고, 더 높은 수준의 상황 판단과 제어에 집중

[ ETRI가 개발 중인 자율주행 버스의 데이터 수집량 ]

센서 종류	대수	포인트 수/프레임	데이터량/프레임(1대)	데이터율 (1대, 10Hz)	총 데이터율 (N대 합)
HESAI OT128 라이다	2대	3,456,000	약 55.3 MB	553 MB/s	1,106 MB/s
HESAI QT128 라이다	3대	1,152,000	약 18.4 MB	184 MB/s	552 MB/s
3MP 카메라	7대	3MP x 3 Byte	약 9 MB	90 MB/s	630 MB/s
합계	12대	-	-	-	약 2.3 GB/s

46) ETRI 내 관련 기술부서 전문가들에게 기술 자문 의뢰한 내용 중 일부 내용 발췌

## 나. 메인 연산 SoC와 상호보완적으로 결합하는 지각 특화 AI 반도체의 필요성

### ■ 메인 연산 SoC의 의존성 탈피를 위한 전용 지각 특화 연산 반도체 필요

- 피지컬 AI 시스템은 메인 연산 SoC의 성능에 전적으로 의존하기보다, 물리 환경에서 발생하는 방대한 멀티모달 센서 데이터를 실시간으로 지각·해석할 수 있는 전용 디바이스 내장형 AI 프로세서가 필요함
- 이를 효율적으로 수행하기 위해서는 다양한 센서 데이터를 신속히 융합, 추론, 요약하는 지각 전용 반도체가 요구되며, 해당 반도체는 메인 연산 SoC와 상호보완적으로 결합함으로써 피지컬 AI 시스템의 전체 성능을 극대화할 수 있음
- 현재 상용화된 온센서 AI와 온디바이스 AI는 다음과 같은 구조적 한계를 지님
  - 온센서 AI는 단일 센서 기반의 간단한 이벤트 감지에는 유효하나, 경량화된 자원 제약으로 인해 다중 센서 융합, 복합 추론, 의미 있는 메타데이터 생성에는 한계가 있음
  - 온디바이스 AI는 범용 연산과 응용 처리에는 적합하나, 멀티센싱 데이터 처리를 위한 자원 할당이 제한적이며, 대규모 데이터 이동으로 인한 전력 소모·발열 문제로 실효 성능이 제약
- 결과적으로 센서 내부의 최소 연산(온센서 AI)과 중앙 SoC의 대규모 범용 연산(온디바이스 AI) 사이에는 다중 센서·저전력·초저지연 지각 연산을 담당할 전용 하드웨어 계층이 부재함
- 이 공백을 메우기 위해 메인 연산 SoC와 결합되는 지각 특화 AI 반도체가 필수적으로 필요

### ■ 피지컬 AI 요구 성능 충족을 위한 지각 특화 AI 반도체의 역할

- 피지컬 AI 시스템의 실시간성, 안정성, 에너지 효율성을 보장하기 위해서는, 센서 인근에서 멀티모달 데이터를 즉시 융합, 추론, 압축, 의미화할 수 있는 지각 특화 AI 반도체가 필요
- 지각 특화 AI 반도체는 단순히 메인 SoC의 부하를 경감하는 보조 장치가 아니라, 피지컬 AI의 지능적 반응성과 안전성을 실질적으로 뒷받침하는 핵심 구성요소임
- 프로세스·소프트웨어 측면의 최적화만으로는 한계가 있으며, 하드웨어 측면에서는 지각 전용 반도체를 도입하는 것이 가장 효과적인 구조적 해법으로 판단됨
- 해당 반도체는 센서와 메인 SoC 사이에서 전처리-지각-요약 기능을 담당함으로써, 앞서 제시된 지연·병목·전력·안전성 문제를 구조적으로 완화할 수 있음

■ AI와 인간, AI-to-AI 상호작용(Interaction) 증가에 따른 지각 특화 AI 반도체의 필연성

○ AI와 인간(AI-to-인간)의 상호작용

- 휴머노이드가 음성, 제스처, 표정을 자연스럽게 이해하기 위해서는, 음성·영상·촉각 신호를 10ms 이내의 초저지연으로 통합 처리해야 함

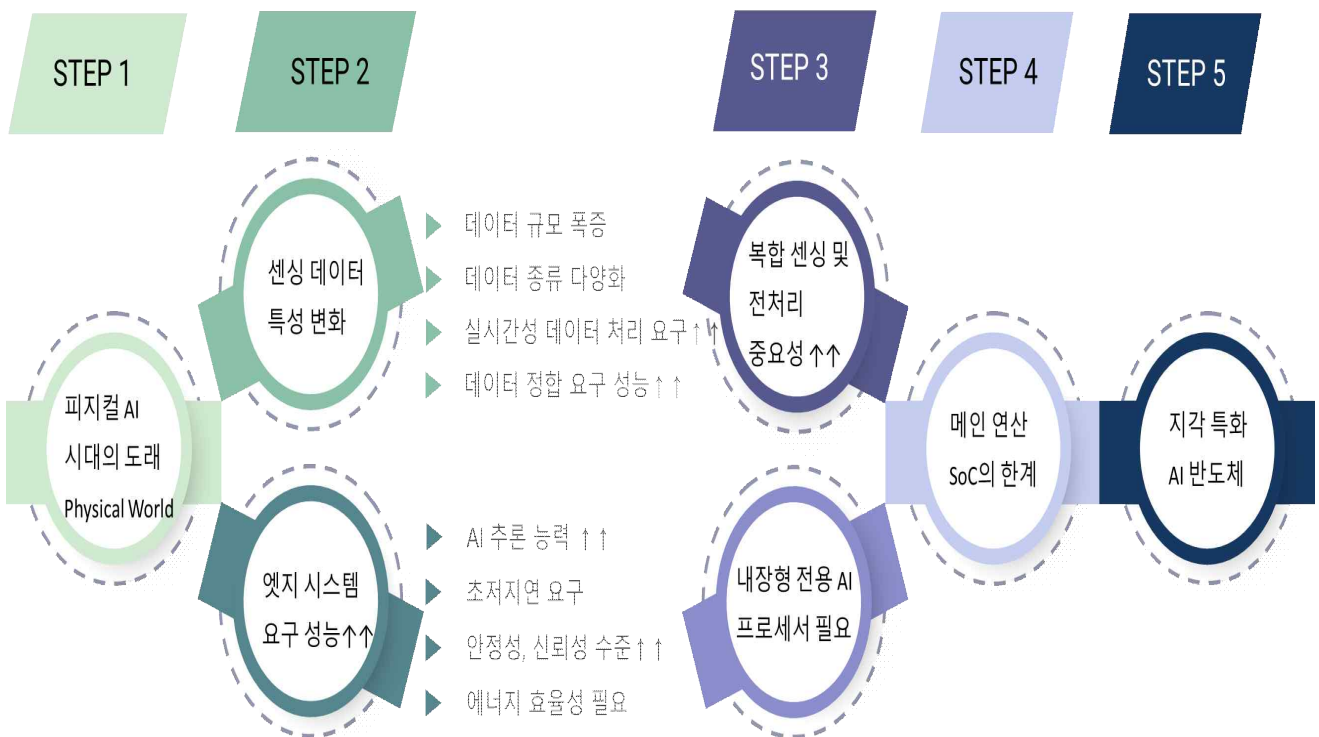
○ 이를 위해서는 센서 근처에서 신호 융합과 위험 판단을 수행하는 지각 특화 하드웨어가 필수적임

○ AI와 AI(AI-to-AI)의 자율 협력

- 완전자율주행차나 로봇 간 협력(V2X 등)을 위해서는, 각 시스템이 센서로 인지한 상황을 의미 단위 메타데이터(예: 객체 위치·속도)로 요약·공유해야 함

- 지각 특화 AI 반도체가 생성하는 표준화된 메타데이터 없이는 기기간 협력이 구조적으로 제한됨

○ 상호작용의 빈도와 복잡성이 증가할수록 센서 근처에서 데이터를 의미 단위로 신속히 압축 및 전달하는 지각 특화 AI 반도체의 가치와 수요는 더욱 확대될 것으로 판단됨



[그림 3] 피지컬 AI 시대 도래에 따른 지각 특화 AI 반도체 필요성

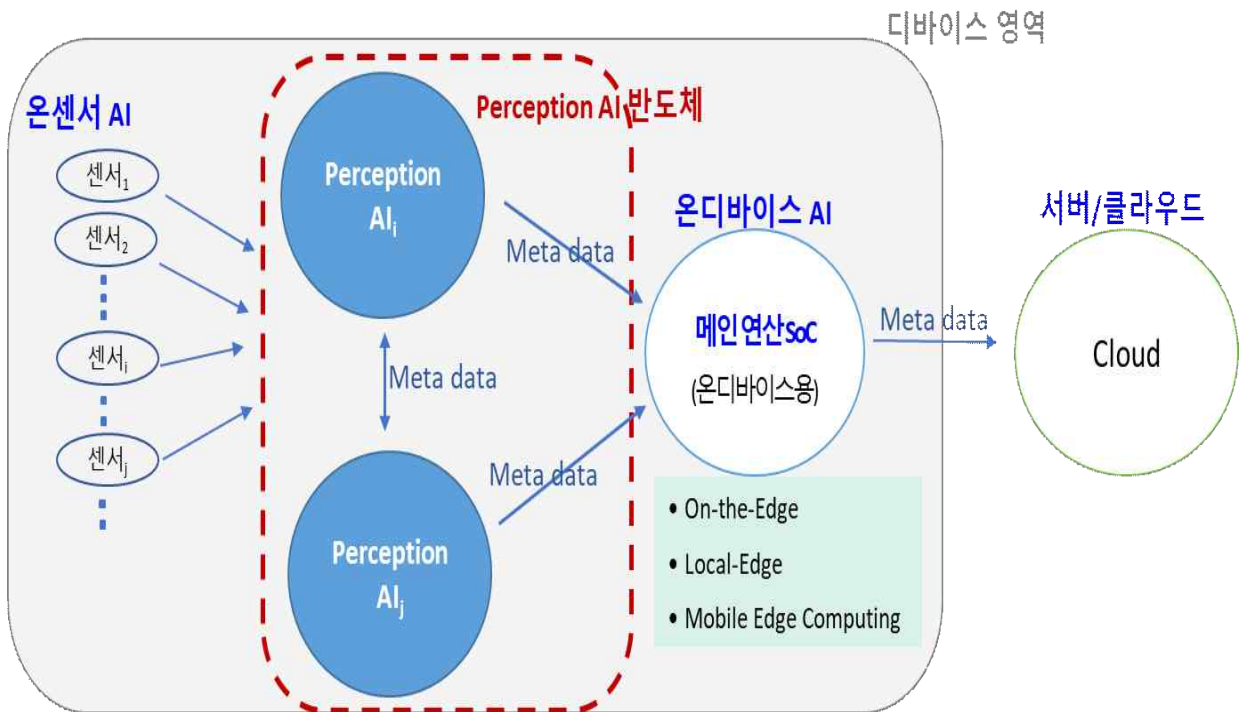
## II 지각 특화 AI 반도체 개념과 특성

### 1. 지각(Perception) 특화 AI 반도체 정의와 역할

#### 가. 개념 정의

##### ■ 지각(Perception) 특화 AI 반도체 정의

- 지각(Perception) 특화 AI 반도체는 디바이스 내에서 멀티 센서의 복합 데이터를 해석하여, 메타데이터를 생성하고, 엣지 내 메인 연산 SoC 또는 상위 클라우드로 전달하는 AI 반도체
  - 지각(Perception) 개념은 센서 정보를 처리하고 AI로 추론하는 기능을 포함
  - 이러한 AI 반도체는 하나의 디바이스 내에 요구되는 성능 수준, 센서의 모달 유형, 그리고 탑재되는 폼팩터의 니즈에 따라 다수가 탑재 가능

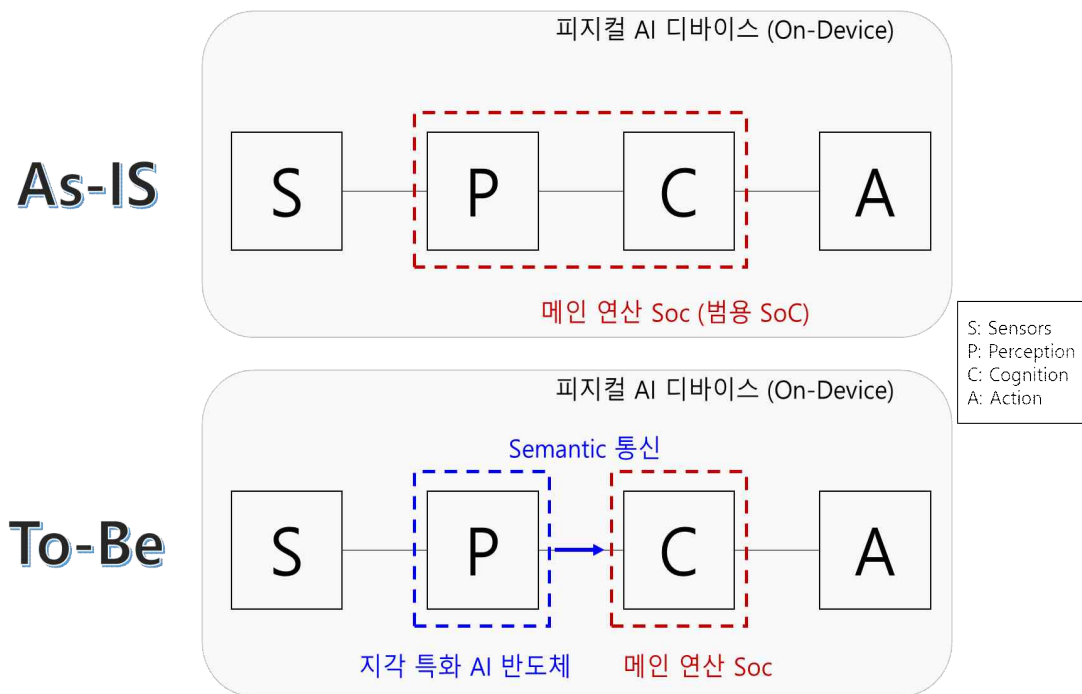


[그림 4] 지각(Perception) 특화 AI 반도체 개념

<표 5> 개념 정의에 사용된 관련 용어

용어	개념
온디바이스 AI	① 통상적으로 클라우드 서버나 외부 네트워크에 의존하지 않고 디바이스 내부에서 AI 연산을 직접 수행하는 기술을 의미 ② 본 고에서는 디바이스 내부에서 고성능 가속기로 구현된 메인 (범용)연산을 위한 중앙 칩인 SoC의 개념으로 사용
온센서 AI	센서 내부에 직접 탑재되어 단일 센서 수준에서 AI가 수행되는 기술 또는 칩
지각 (Perception)	① 휴머노이드 로봇에서 환경 사물의 인식(P, Perception), 판단(C, Cognition), 행동(A, Action)에서 인식(P)에 해당하는 용어로 사용 ② 본 고에서는 인식은 물론 즉각적인 행동을 위한 판단(추론)의 영역도 일부 포함하는 개념으로 사용

출처: ETRI 기술전략연구본부 작성, 2025.11.



[그림 5] 지각 특화 AI 반도체 도입에 따른 피지컬 AI 디바이스 구조 변화

## 나. 주요 기능

### ■ 처리 프로세스

- 디바이스 내에서 다양한 센서들로부터 수집된 복합 데이터를 해석
  - 이러한 해석은 데이터를 실시간 센싱하고 AI 알고리즘으로 추론하는 기능
- 해석된 정보들은 메타데이터로 의미있는 단위로 축약, 추출되어 생성
- 생성된 메타데이터는 자율 협력 기능의 협업 AI간 또는 AI와 인간의 상호작용을 위해서 디바이스 내 메인 연산 SoC는 물론이고 필요에 따라 다른 디바이스 또는 상위 엣지/클라우드도 전송

### ■ 센서 퓨전 (Sensor Fusion)

- 다양한 종류의 센서 데이터 또는 여러 위치의 동종 센서 데이터를 융합하여 하나의 종합적인 환경 인식을 형성
- 개별 센서로 얻은 정보의 불확실성을 감소시키고 보다 신뢰도 높은 판단
- 카메라, LiDAR, ToF(Time-of-Flight)<sup>47)</sup>, 레이더, 초음파, GPS 등 다양한 센서 데이터를 결합<sup>48)</sup>
- 멀티센서 데이터의 동기화와 실시간 병합을 위한 하드웨어 가속 및 알고리즘을 내장

### ■ 온-디바이스 지각 처리 (On-Device Perception)

- 센서 신호 처리와 AI 추론을 클라우드가 아닌 디바이스 자체에서 수행
- 통신 지연 없이 실시간 응답 가능하고, 외부와 연결이 끊겨도 독자적인 지각 기능수행
- 데이터 생성 즉시 실시간 처리가 가능하여, 소비전력과 지연시간을 크게 단축
- 단순 압축을 넘어, 딥러닝 네트워크를 통해 센서 원시 데이터를 저차원의 특징 벡터로 변환하거나, 객체 검출, 이벤트 감지 등의 결과를 요약하여 메타데이터 형태로 전달
- 로컬에서 필요한 연산을 모두 수행함으로써 엣지 단에서의 자율성을 확보

47) 신호(빛 또는 전자기파)를 방출한 뒤 반사되어 돌아오는 시간을 측정하여 거리·깊이 정보를 계산하는 센서 방식

48) Simon C. Morris. (2024). Entering the Era of Multimodal Perception, 2024 embedded vision summit

■ 메타데이터 생성 및 요약 (Semantic 통신)

- 원시 센서 데이터에서 의미있는 메타데이터를 추출하고, 상위 시스템이 활용하기 쉽도록 데이터 구조화/요약
- 저용량의 고수준 정보만 전송하게 해주어 통신 부하를 감소시키는 효과
- 뉴로모픽 방식의 칩들은 아예 HW 단계에서 이벤트 기반 데이터로 변환하여, 불필요한 데이터를 필터링
  - Ex. BrainChip社의 Akida : 연속 입력신호를 이벤트로 변환 처리함에 있어서, 최대 10배 데이터 양 감소 후 연산<sup>49)</sup>
- 실시간성은 유지하면서도 시스템 전반의 효율을 높이는 핵심 기능
- 센서 원본에 대한 부가 정보를 추가하여 이후 단계에서 다시 맥락에 맞게 활용도 가능

■ 엣지-클라우드 연계 (Interaction; AI-to-Human, AI-to-AI)

- 엣지 디바이스 내에서 신호 전처리와 1차 지각(Perception)/인지를 수행
- 필요한 정보와 요약 결과만 엣지 서버나 클라우드로 전송하여 상위 계층 AI 등과 협업
- 이 분산 컴퓨팅 구조를 통해 전체 시스템 성능을 최적화
  - Ex. 자율주행 네트워크는 차량, 도로 인프라, 클라우드가 연결된 복합 하이브리드 구조로 발전
- 통신 모듈, 프로토콜 스택, 그리고 지능형 데이터 업로드/다운로드 제어 등이 포함<sup>50)</sup>
- 정보를 로컬에서 처리하고 어떤 것을 상위로 보낼지를 동적으로 결정
- 대역폭 효율을 높이고 필요시 클라우드의 고도화된 AI와 연동하면서도 현장 자율성을 잃지 않도록 설계<sup>51)</sup>

49) Brainchip.com. (2025). Technology Performance where it matters most

50) Alex Vakulov. (2024.12.). Addressing Data Processing Challenges in Autonomous Vehicles. IoT For All

51) Alex Vakulov. (2024.12.). Addressing Data Processing Challenges in Autonomous Vehicles. IoT For All

## 2. 필요 기술 및 기능적 차별성

### 가. 주요 기능별 필요 기술<sup>52)</sup>

#### ■ 센서 인터페이싱 및 퓨전 알고리즘

- 다양한 종류의 센서 데이터 또는 여러 위치의 동종 센서 데이터를 융합하여 복합적으로 해석하기 위해서는 센서 퓨전 알고리즘과 인터페이싱 기술이 필요
- 센서 퓨전
  - 서로 다른 센서 시간차를 보정하고 잡음을 제거하며, 여러 센서의 출력을 동시에 처리할 수 있는 하드웨어 지원 병렬처리 구조가 요구
  - 센서 허브/데이터 프로세서가 탑재 : 다채널 센서 스트림을 수집/전처리
  - 수집된 데이터는 퓨전 알고리즘에 의해 통합 : 퓨전 알고리즘에는 저수준 데이터 융합(센서 신호 결합)부터 고수준 인식 결과 융합 (Ex: 개별 AI 모듈 결과의 종합)까지 여러 단계가 있음<sup>53)</sup>
  - 이러한 알고리즘들을 실시간으로 동작시키려면 FPGA/ASIC 기반 가속회로<sup>54)</sup> 사용
  - 동시에 SW로써 유연한 조정이 가능하도록 DSP<sup>55)</sup>나 AI 프로세서 코어도 포함
- 인터페이싱 및 동기화
  - 센서 주행/동작 중 온도, 진동으로 변화하는 센서 정렬을 자동으로 보정하는 온라인 캘리브레이션 기술
  - 동기화/타임스태핑(PTP/TSN 급 정밀도)과 드리프트 보정이 내장 필요
  - 다중 센서 시간 정렬, 타임스태프 HW 지원, 저지연버스 기술 등

#### ■ 경량화된 지각(Perception) AI 추론 모델

- 지각 특화 처리를 위한 AI 알고리즘은 엣지 디바이스에서 실시간 동작하는 전처리 모델과 추론 모델 기술이 필요
  - 디바이스 내 제한된 자원으로 동작하기 위해서는 경량화된 딥러닝 모델 및 알고리즘이 필수
  - 적은 연산으로도 높은 인지 정확도를 얻는 모델 설계와 최적화가 지각 특화 AI 반도체 성능을 좌우

52) 해당 상세기술 내용은 ETRI 내 관련 기술부서(온디바이스연구본부, AI로봇연구본부, 지능형반도체연구본부) 전문가들에게 기술 자문 의뢰한 내용 포함

53) Simon C. Morris. (2024). Entering the Era of Multimodal Perception, 2024 embedded vision summit

54) 특정 연산을 범용 프로세서보다 빠르고 효율적으로 수행하도록 설계된 하드웨어 가속 회로. FPGA (Field-Programmable Gate Array) 기반은 제조 후에도 하드웨어 구조를 재구성할 수 있는 반도체이며, ASIC (Application-Specific Integrated Circuit) 기반은 특정 용도에 맞춰 고정 설계된 전용 반도체로 성능·전력·지연 측면에서 최적화에 유리함

55) Digital Signal Processor : 신호 필터링, 변환, 특징 추출 등 실시간 신호 처리를 효율적으로 수행하도록 설계된 전용 프로세서

○ 지각 특화 처리

- 핵심 공통 전처리 알고리즘 : 경량화된 CNN 네트워크 설계, 경량의 트랜스포머 가속기, 멀티모달정합(칼만/팩토그래프·이벤트 기반), VoxelNet<sup>56</sup>/BEV<sup>57</sup>기반 포인트 클라우드 요약 등
- 노이즈 필터링, FFT<sup>58</sup>/DCT<sup>59</sup> 변환, 엣지 검출, 히스토그램 균등화 등 전통적 신호 처리와 SIFT<sup>60</sup>, SURF<sup>61</sup>, ORB<sup>62</sup> 등의 고전적 특징 추출을 기본 수행

○ 경량화된 AI 추론

- 가속 처리, 실시간 추론, 파인튜닝, 최적화를 위한 양자화 및 푸루닝 등
- MobileNet, SqueezeNet 수준의 분류 모델<sup>63</sup>, YOLO Nano나 SSD MobileNet 등의 경량 객체 검출<sup>64</sup>, 키워드 스포팅<sup>65</sup>, 제스처 인식 등 특화 모델
- 메모리 최적화를 통해 온칩 메모리 활용을 극대화하는 노력이 필요
- 뇌 신경을 모방한 스파이크 신경망(SNN) 등의 뉴로모픽 알고리즘도 경량/저전력 지각에 활용
- PIM<sup>66</sup> 기술로 메모리 병목을 줄인 연산을 구현하는 기술도 검토 필요

■ 메타데이터 전송 및 칩 보안 기능

○ 메타데이터 전송

- 처리 정보를 다른 시스템과 교환하기 위해 유·무선 통신 모듈을 통합하거나 외부 인터페이스와 연동<sup>67</sup>할 수 있는 표준 스키마 및 표준 전송 기술 등이 필요
- 이를 위하여 필요 이상으로 원본 데이터를 저장하지 않고 즉시 필요한 정보만 추출하도록 시스템 설계도 고려해야 함<sup>68</sup>)

○ 칩 보안 기능

- 민감한 정보가 많아 데이터 암호화와 접근 제어를 칩 단계에서 구현 필요
- AI 추론 보안을 위해 모델 및 데이터에 대한 동형암호, 안전 멀티파티 연산 등의 기술이 필요

56) 3차원 공간을 복셀(Voxel, 3D 격자)로 분할하고, 각 복셀 내 포인트들을 학습 기반으로 특징 추출·요약하는 방식  
 57) Bird’s-Eye View : 3차원 포인트 클라우드를 상공에서 내려다본 2D 평면(BEV)으로 투영하여 요약  
 58) Fast Fourier Transform : 신호를 시간/공간 영역에서 주파수 영역으로 빠르게 변환하는 알고리즘  
 59) Discrete Cosine Transform : 신호를 코사인 성분 기반의 주파수 표현으로 변환하는 기법  
 60) Scale-Invariant Feature Transform : 크기·회전에 안정적인 특징점(keypoint)을 검출·기술하는 알고리즘  
 61) Speeded-Up Robust Features : SIFT를 연산 속도 중심으로 경량화된 특징 추출 기법  
 62) Oriented FAST and Rotated BRIEF : FAST 코너 검출과 BRIEF 기술자를 결합한 고속·저연산 특징 추출 기법  
 63) 모바일 및 온디바이스 환경을 위해 연산량과 파라미터 수를 최소화한 경량 CNN 기반 분류 모델  
 64) 실시간 처리를 목표로 한 경량 객체 검출(Object Detection) 모델  
 65) Keyword Spotting : 연속 음성 중 특정 짧은 명령어(웨이크 워드)를 인식하는 음성 인식 기술  
 66) Processing In Memory, 메모리 내 프로세싱  
 67) Alex Vakulov. (2024.12.). Addressing Data Processing Challenges in Autonomous Vehicles. IoT For All  
 68) Anand Ramachandran. (2025.4.). Advancing Robotic Perception through Multimodal Sensor Fusion and Advanced AI: Breakthroughs, Challenges, and Future Directions. Linked in

- 프라이버시 및 무결성 보장을 위한 부팅 체인, 키 관리 등의 온칩 보안 기술도 함께 구현되어야 함

**■ 옛지 환경에 최적화된 전력 및 성능 관리**

○ 전력 효율 극대화를 위한 기술 고도화

- 전력 효율 극대화 : Fine-grained Power Gating<sup>69)</sup>, Adaptive Voltage Scaling<sup>70)</sup>, Clock Domain Crossing<sup>71)</sup> 최적화 등
- 효율적인 메모리 관리 기술 또는 pre-Loading 기술 등을 고려한 저전력 관점의 설계 필요

○ 최적화된 성능 관리

- 런타임 시스템 고도화 : 동적 메모리 관리자, 멀티코어 작업 스케줄러, 인터럽트 기반 이벤트 처리기 등을 통합
- 컴파일러 및 최적화 도구 : 연산 그래프 최적화, 메모리 접근 최소화, HW 특성을 고려한 Auto-tuning 기술, 오픈소스 컴파일러 경량화 등
- Energy-Aware Scheduling<sup>72)</sup> 등을 통한 성능과 전력의 최적 균형점 도출

**■ 칩렛(Chiplet) 기반 또는 설계자산 IP(Intellectual Property) 기반 모듈화**

○ 여러 기능 칩(칩렛<sup>73)</sup>)을 한 패키지에 집적하거나, IP<sup>74)</sup>(설계자산) 블록 형태로 모듈화 하여 SoC<sup>75)</sup> 구성 가능

- CPU, GPU, NPU, 센서 인터페이스, 메모리 등을 각각 최적 공정으로 만든 후, 고속 인터커넥트(UCIe<sup>76)</sup> 등)로 연결하여 하나의 시스템처럼 동작하게 가능
- 칩 집적도의 한계 시 면적 문제 해소, 불량 시 부분 교체나 여러 버전 조합 등 확장성 제공
- 모달별 칩렛을 필요에 따라 조합 구성 고려
- 칩렛과 뉴로모픽 칩을 결합하여 엔비디아의 CUDA와 같은 가상의 통합 플랫폼 형성 효과 획득 가능

---

69) 회로 블록 단위로 사용하지 않는 부분의 전원을 세밀하게 차단하여 누설 전력을 최소화하는 기술  
 70) 부하·온도·성능 요구에 따라 동작 전압을 동적으로 조절하여 전력 소모를 최적화하는 기술  
 71) 서로 다른 클록 주파수·위상을 사용하는 회로 간에 데이터를 안전하게 전달하기 위한 설계 기법  
 72) 작업의 중요도·마감시간·연산 부하·전력 상태를 고려하여 성능과 에너지 효율을 동시에 최적화하도록 실행 순서와 자원 배분을 결정하는 스케줄링 기법  
 73) Chiplet : 하나의 큰 칩(SoC)을 여러 개의 작은 칩(칩렛)으로 쪼개어, 각각의 칩렛이 특정 기능만 담당하도록 만든 뒤, 이 칩렛들을 하나의 패키지 안에 조립하는 방식. IP로 설계된 여러 기능 블록들을 실제로 만든 각각의 칩 또는 그 칩들을 패키지에서 조립하는 방식  
 74) Intellectual Property: 칩(반도체) 안에 들어가는 '기능 블록'을 미리 설계해 놓은 설계 자산  
 75) System on Chip으로 여러 부품들이 하나의 반도체에 통합되어 단일 시스템을 구성하는 형태  
 76) 각 칩렛을 고속으로 꽂아주는 표준 핀 (데이터/전원 연결 규격)

- IP 기반 모듈화 : 센서 처리, AI 가속, 통신보안 등 각 기능을 모듈 IP로 제공하고, 필요 부분만 조합하여, 맞춤형 SoC를 제작하는 방식
  - 적용 분야와 요구 성능이 다양하므로, 칩렛/IP 기반의 유연한 설계로 제품군을 효율적 확장
  - AI 모델의 변화에도 교체, 장착이 용이한 방식 고려
  - SW 개발 환경 및 SW 플랫폼 제공 방안 고려
- 센서 전처리부터 AI 연산까지의 시스템 아키텍처 구성 방안
  - 개방형 아키텍처를 전제, 현재 가용한 패키징 및 소자 기술로 시작하여, 미래 지향적인 최신 아키텍처 지향
  - Near-Sensor Processing 레이어, Sensor Fusion 레이어(Perception AI), AI Inference 레이어 계층 구성

## 나. 기능적 차별성 및 주요 장점

### ■ 관련 유사 제품 현황 및 차별점 (<표 6> 참조)

- Sony IMX500(온센서 AI 제품)
  - 세계 최초의 AI 엣지프로세싱 기능 내장 비전 센서
  - 이미지 센서와 DSP, 온칩 SRAM이 적층된 원칩구조
  - 외부 서버 없이 이미지 인식 및 AI 추론 가능
  - 센서 정보 처리에 특화
  - 다중 센서 융합 처리 불가
- Apple R1(초저지연 XR 지각전용 제품)
  - Apple Vision Pro(혼합현실 헤드셋) 용으로 설계된 전용 프로세서
  - M2 칩과 함께 동작, 12개 카메라, 5개 센서, 6개 마이크 센싱 데이터를 초저지연 (12ms) 처리
  - 메인 AI 연산 이외에 저지연 처리
  - XR 기기의 광학 처리에 한정
- NVIDIA Thor(지각 기능이 포함된 메인 SoC, 온디바이스 AI 제품)
  - 차량 내 모든 컴퓨팅을 통합하는 중앙집중식 차세대 슈퍼칩
  - Blackwell 아키텍처 기반

- Grace CPU<sup>77)</sup>, Ada Lovelace GPU<sup>78)</sup>, Hopper Multi-Instance GPU(MIG)<sup>79)</sup> 등 최신 기술 집약
- 온디바이스 AI 제품의 최신, 최고 추론 성능 보유
- 고전력 소요 및 발열 등으로 실효 성능 한계

○ BrainChip Akida(이벤트 기반 전처리)

- 세계 최초 상용화된 뉴로모픽(뇌 모방형) AI 프로세서
- SNN 아키텍처로 초저전력, 초고효율 실시간 AI 추론 및 학습 지원
- LLM/VLM의 실시간 물리 환경 인식의 한계를 극복 가능
- 에너지 효율에 특화
- 매우 제한적인 추론 성능 한계

<표 6> 지각 특화 AI 반도체와 유사한 제품들의 특징

제품	디바이스 내 위치	지연 (Latency)	성능 (TOPS)	소요 전력	특징 및 한계	역할 의미
 <p>Sony IMX500</p>	센서 내	1-3ms	mW	mW	카메라 특화 (단일 이미지)	센싱데이터 크기 1/10 <sup>3</sup> 로 축소 <sup>80)</sup>
 <p>Apple R1</p>	XR HMD	12ms <sup>81)</sup>	-	5W 미만	XR (광학처리) 한정	지각 → M2(메인연산) 저지연 처리 <sup>82)</sup>
 <p>Jetson Thor</p>	보드	10-30ms	2,000	100W	고전력(100W)	다중 Fusion, ASIL-D <sup>83)</sup> <sup>84)</sup>
 <p>BrainChip Akida</p>	이벤트 <sup>85)</sup>	μs-ms	DSP	mW	대형 CNN 불가	Always-on, 배터리 IoT <sup>86)</sup>

출처: ETRI 기술전략연구본부 작성, 2025.11.

77) NVIDIA가 설계한 고성능·고전력효율 Arm 기반 데이터센터용 CPU

78) NVIDIA의 고성능 GPU 아키텍처

79) Hopper 아키텍처 GPU를 여러 개의 독립된 가상 GPU로 분할하는 기술

80) <https://developer.sony.com/imx500>

81) 인간이 멀미를 느끼지 않는 한계(약 20 ms 이내)를 만족하기 위한 설계 목표값

82) <https://support.apple.com/en-us/117810>

83) Automotive Safety Integrity Level - D 등급

84) <https://nvidia.com/newsroom/nvidia-unveils-the-first-centralized-car-computer-using-dual-infotainment-automated-driving-and-parking-in-a-single-cost-saving-system>

85) SNN 코어로서 이벤트 입력이 없으면 연산이 0, 배터리 또는 IoT에 최적

86) <https://brainchip.com/technology/>

■ 온센서 AI, 온디바이스 AI 제품 이외에도 지각 특화 AI 반도체가 추가로 필요한 이유

○ 물리법칙(Physical World) 환경에서 온센서 AI(지능형센서 포함), 온디바이스 AI 제품의 한계

- (온센서 AI) 센서가 다수로 늘고, 해상도와 프레임이 폭증하면 센서 내부의 DSP 등으로는 연산, 전력, 센서 융합이 불가능
- (온디바이스 AI) 폭증한 데이터를 메인 연산 SoC인 온디바이스 AI가 단독으로 처리하기에는 모달, 메모리, 지연성 부족, 소요 전력 과다 등으로 인한 실효적인 추론 성능의 한계

○ 지각 특화 AI 반도체가 추가로 필요한 이유

- 피지컬 AI 시스템 내 각 역할이 분업화되어 처리 데이터, 소요 전력, 허용 안전 곡선을 오히려 전체적으로 낮추는 결과 가능
  - 사람, 로봇, 차량 등이 상시적으로 상호작용하기 위해서는 10ms급 멀티센서 의미화 레이어 필요
- 기능의 안정성 관점에서 이점 : 센서 오작동 시 격리 가능, 메인 연산 SoC-Fail Operational<sup>87)</sup>에도 대비
- 비용 증가에 대한 우려 해소 가능 : 비용 절감 우위 달성 가능
  - 칩렛, IP 시대의 효율화 요인 : 지각 기능을 하나의 칩렛으로 제작하여 여러 제품군에서 공유 가능
  - 센서-스태킹 원가의 인하(대량 생산·공정 미세화와 IP 소자 표준화)로 저가 생산 가능
  - 통신·스토리지 업무 부하 감소 및 비용 절감으로 인한 상쇄 효과 : 지각 특화 AI 반도체 채택 시 저가의 칩과 전력비용 절감으로 충분히 상쇄
- 유연성 및 업데이트 용이 : 센서와 분리하여 유지보수 비용 절감, SW 업데이트 또는 교체 용이

■ 지각 특화 AI 반도체가 추가로 필요한 시나리오 예

○ 완전자율주행차의 횡단보도 긴급정지 상황 (<표 7> 참조)

- 온센서 AI : 단일 센서에서 노이즈와 불필요한 픽셀 제거 (센싱한 원본 데이터를 1/10<sup>3</sup>로 축소)
- 지각(Perception) 특화 AI : 다양한 다수 센서를 융합해 의미있는 데이터만 추출 (데이터 용량을 1/10<sup>2</sup> 크기로 추가 축소)
- 온디바이스 AI(메인 SoC) : 축소 및 의미화된 데이터를 바탕으로 의사결정

87) 메인 연산 SoC에 고장이나 일부 오류가 발생하더라도 핵심 기능을 중단하지 않고 제한된 성능으로 계속 동작하도록 설계하는 개념으로 단순 정지(Fail-safe/Safe-fail)와 달리 기능 지속성 확보가 목적

- 총 지연시간이 42ms 내외로 규제 기준치(150ms, UN-R 152<sup>88</sup>)의 지각-제동 한계시간)보다 훨씬 빠름
- 지각(Perception) 특화 AI 반도체가 비록 5W의 전력이 추가로 소모되지만, 온디바이스 AI인 메인 SoC의 전력 소모는 크게 감소(85W → 60W)

〈표 7〉 완전자율주행차 긴급정지 상황에서 단계별 데이터 처리 시나리오 예

단계	단계별 기능 시나리오	지각 특화 AI 반도체 無			지각 특화 AI 반도체 有		
		데이터 용량	소요 시간	소모 전력	데이터 용량	소요 시간	소모 전력
① 센서	전방 8K 카메라 6대(1대당 12-bit RAW, 프레임당 50MB, 120fps), LiDAR가 RAW 데이터 생성	36GB/s	0ms	32W	36GB/s	0ms	32W
② 온센서 AI	카메라 픽셀 노이즈 제거, Tiny-CNN으로 사람 후보 500개 추출 (10 <sup>3</sup> 배 축소)	36GB/s	+2ms	0.6 W	36GB/s	+2ms	0.6W
③ 지각 특화 AI 반도체	카메라-LiDAR-Radar 동기화 → 3D Fusion "보행자, 거리 15m, 속도 1.2m/s" 의미 패킷 1개만 생성 (10 <sup>2</sup> 배 추가 축소)	-	-	-	3.6 MB/s	+12 ms	5W
④ 메인 SoC (온디바이스 AI)	보행자-차량 위치, 신호 등 상태 종합 → 경로 재계산 및 제동 명령	36 MB/s	+53 ms	85 W	3.6 MB/s	+30 ms	60W
⑤ 제어기	브레이크 작동 <sup>89)</sup>	-		*	-	*	
합 계			+55 ms	~120 W		+44 ms	~100 W

출처: ETRI 기술전략연구본부 작성, 2025.11.

■ 기능 측면에서의 지각 특화 AI 반도체의 장점 ; 옛지 디바이스 내 실시간 물리환경 지각의 한계 극복

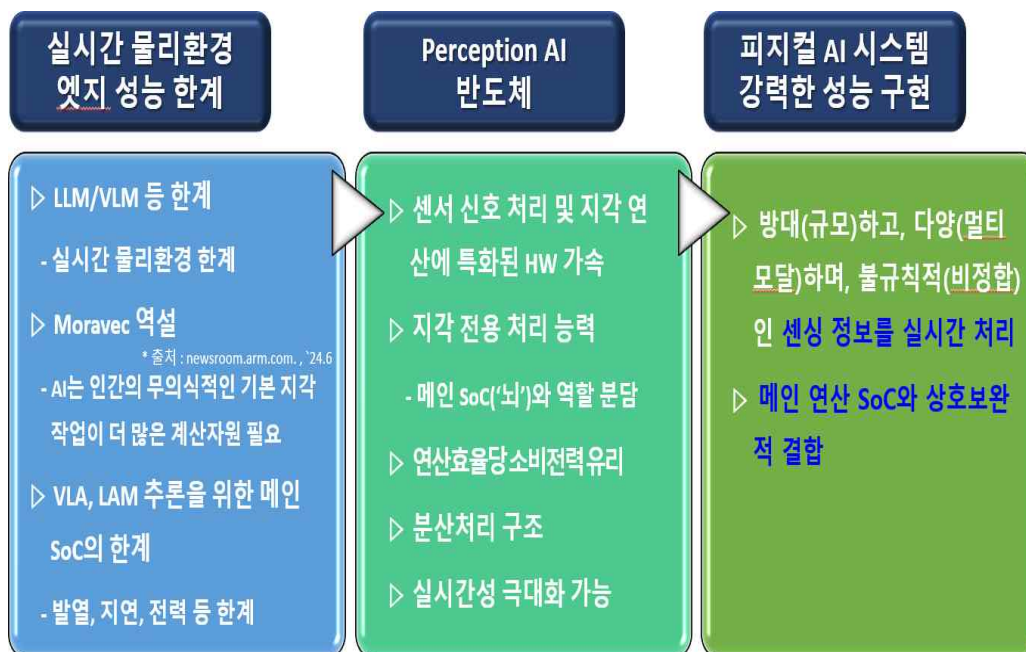
○ 다양한 센싱 정보 처리 및 지각 전용 처리 성능 극대화 ; 센서 신호 처리 및 인지 연산에 특화된 하드웨어 가속을 제공

- LLM/VLM의 실시간 물리 환경 인식의 한계를 극복 가능

- 거대언어모델(LLM)이나 비전-언어모델(VLM)은 방대한 매개변수를 바탕으로 고차원적인 추론과 생성 능력을 보여주지만, 실시간 물리 환경 인식에는 부적합하다는 한계가 존재

88) UN Regulation No.152 : 유엔 유럽경제위원회(UNECE)가 제정한 자동 비상 제동 시스템(AEBS) 관련 국제 자동차 안전 규정  
89) 상시 작동되는 제어 소모 전력은 별도

- Ex. 수십억 파라미터의 LLM이 문자로 환경을 설명하는 작업을 할 수는 있어도 초당 수십 프레임의 고해상도 비디오 스트림을 실시간 처리하며 수 ms 단위로 제어 신호를 생성하기는 어려움
- 모라벡(Moravec) 역설 : AI는 인간의 무의식적인 기본 지각 작업이 더 많은 계산 자원 필요<sup>90)</sup>
- 센서 신호 처리 및 인지 연산에 특화된 하드웨어 가속을 제공
  - 동등한 연산을 범용 프로세서나 거대 모델로 수행하는 것보다 훨씬 효율적이고 신속하게 처리
  - Ex. 지각 전용 칩은 설계 단계부터 병렬 이미지/신호 처리, 저지연 추론에 맞춰져 있는 작업 수행
- 기존 LLM 및 VLM 등과 대비한 지각 전용 처리 능력
  - LLM/VLM이 뇌의 '생각' 역할을 한다면, 지각 전용 AI 칩은 '감각기관 + 각피질'과 유사한 역할
  - 두 역할을 분담함으로써 전체 시스템 효율과 성능을 제고
- 연산 효율당 소비전력에서 일반적인 데이터센터 GPU나 CPU 대비 훨씬 유리
  - 엣지 환경의 제약을 견딜 수 있는 큰 차별성 가능



[그림 6] 지각 특화 AI 반도체의 기능 측면에서의 장점

90) Chloe Ma. (2024.6.). Transforming the Future of AI and Robotics with Multimodal LLMs. Arm Newsroom

- 엣지 환경에서의 퍼지컬 AI 시스템의 강력한 성능 구현 ; 로봇/휴머노이드 통합을 위한 분산 처리 구조 및 실시간성 극대화 가능
  - 하나의 시스템에 다수의 칩을 분산 배치함으로써 실시간성과 신뢰성을 극대화 가능
    - 휴머노이드 로봇 한 대에 카메라용 지각 칩, 음향용 지각 칩, 촉각용 지각 칩 등을 다수 탑재한 환경 하에서, 이들 칩이 서로 병렬로 동작하면서 다양한 감각 정보를 동시에 처리 가능
    - 하나의 거대 중앙 연산장치가 모든 센서 데이터를 순차 처리하는 구조보다 지연을 줄이고, 각 모듈 고유의 최적화를 가능
    - 센서 가까이에서 데이터를 처리함으로써 네트워크 지연의 획기적 감소 및 시스템 응답성을 제고<sup>91)</sup>
  - 분산 구조는 모듈별 독립성이 있어 일부 장애가 전 시스템 마비로 이어질 위험 감소
    - 필요한 부분만 업그레이드하거나 교체할 수 있어 유지보수에도 유리
  - 분산 실시간 지각 구조 : 각 센서부에 분산시켜 인간의 말초신경계처럼 설계 가능
    - 로봇이 주변 환경에 반사적이고 즉각적으로 대응 가능
    - 중앙의 메인 연산 AI가 고수준의 의사결정을 하는 동안 주변부 지각 모듈들이 저수준 자율성을 갖고 실시간 반응하는 체계를 구축 가능

■ 시장성 측면에서의 지각 특화 AI 반도체의 장점

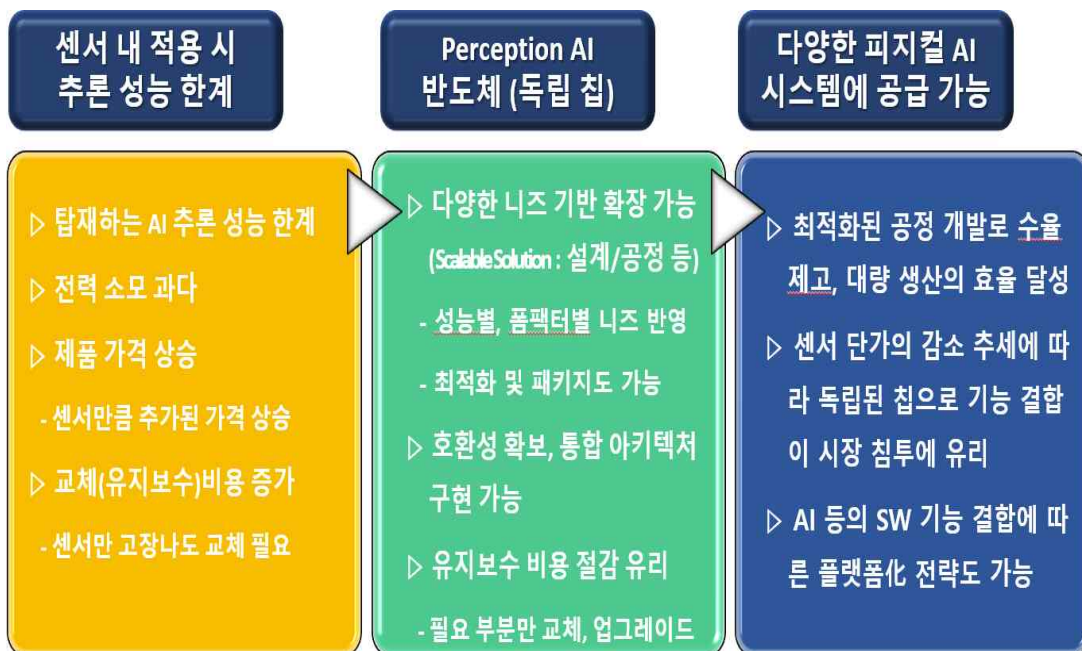
- 다양한 요구사항을 반영할 수 있는 확장 및 경량(Lightweight) 버전 구현 가능
  - 용도, 예산, 전력 제약에 따라 성능, 규모, 구성 규모를 단계적으로 확장 또는 축소 할 수 있는 스케일러블(Scalable) 솔루션 설계 가능
    - Ex. 고성능 로봇용으로는 다센서 채널을 처리하는 복수의 칩을 연결하거나 고사양 칩(칩렛)을 조합하여 최고의 성능을 내고, 소형 IoT 기기용으로는 일부 기능만 넣은 경량 버전의 칩을 제작하여 저전력 최소 구성으로 제공 가능
  - 칩렛 기반 디자인은 필요한 모듈의 탈부착을 통해 제품군 파생을 용이하게 함
    - 기존 범용 AI 프로세서는 미세 단위로 성능, 자원을 조절하기 어려움
  - IP 모듈화를 통해 하나의 기본 설계를 보유하면 고객 요구에 따라 맞춤형 구성 가능
  - 하드웨어적인 유연성으로 향후 다양한 분야(자율주행, 산업용 로봇, 스마트 가전, AR/VR 등)에 최적화된 버전을 전개 가능
    - Ex. 이미지센서 업체인 픽셀플러스는 온센서 AI 칩 기술을 확장하여 자동차용, 소비자용 등 다양한 라인업으로 발전시킬 계획 발표<sup>92)</sup>

91) Anand Ramachandran. (2025.4.). Advancing Robotic Perception through Multimodal Sensor Fusion and Advanced AI: Breakthroughs, Challenges, and Future Directions. Linked in

- 경량(Lightweight<sup>93</sup>) 버전 칩들은 비용과 전력 소모를 낮추면서 핵심 시각 기능만 제공함으로써 엣지 디바이스 AI 대중화에 크게 기여할 수 있음

○ 시장성 측면에서의 장점([그림 7] 참조)

- 센서 내 적용 시 탑재하는 AI 추론 성능, 전력 소모, 비용 상승의 한계 극복
- 센서와 분리된 더 좋은 AI 추론 성능의 저가 칩 공급 ⇒ 시장성 및 범용화 기대
- 최적화 공정 개발 등으로 다양한 피지컬 AI 시스템에 공급 가능



[그림 7] 시각 특화 AI 반도체의 시장성 측면의 장점

92) 디지털데일리. (2024.5.). 픽셀플러스, 온센서 AI 구현한 포토닉 칩렛 기술 개발

93) 성능·기능을 필요 최소 수준으로 최적화하여 연산량, 전력 소모, 면적, 비용을 줄인 설계 특성

### III 목표시장과 확산 가능성 분석

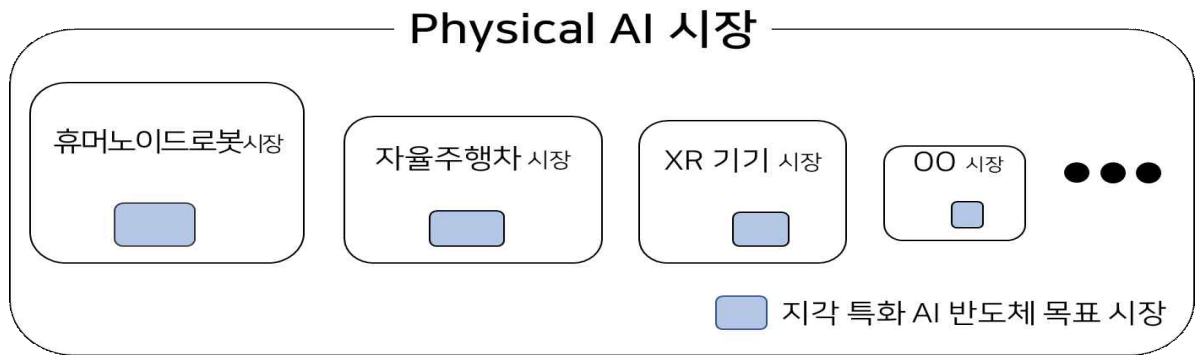
#### 1. 목표시장 정의 및 추정 프레임워크

본 절에서는 지각 특화 AI 반도체를 기존 온센서 AI 및 온디바이스 AI 시장과 구분되는 독자적 시장 영역으로 정의하고, 해당 시장을 정량적으로 추정하기 위한 개념적·방법론적 프레임워크를 제시한다. 아울러 피지컬 AI 시대의 기술적 요구와 기존 반도체 아키텍처의 한계를 기반으로 목표시장 확정의 논리적 근거를 정립한다.

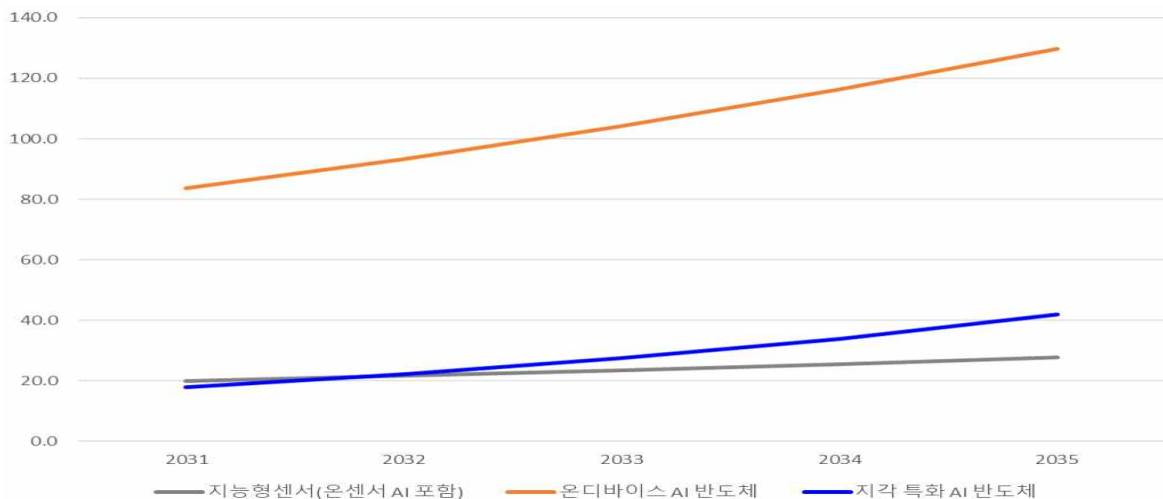
- 지각 특화 AI 반도체 시장은 기존 시장의 연장선이 아니라, 피지컬 AI 시대의 도래라는 거시적 환경 변화에 의해 새롭게 부상하는 독자적 시장 영역임
  - 기존의 온센서 AI(지능형센서 포함) 시장과 온디바이스 AI(메인 SoC) 시장의 교차점에 위치하지만, 두 시장이 해결하지 못하는 기능적 공백(Functional gap)을 메우는 핵심 역할로서 정의할 수 있음
- 온센서 AI는 개별 센서 단위에서 제한적인 연산만을 수행하도록 설계되어, 여러 센서로부터 들어오는 방대한 데이터를 통합적으로 처리하는 데에는 역부족
  - Sony의 IMX 500 이미지 센서나 Knowles의 스마트 마이크처럼 센서 자체에 AI 프로세서를 내장하는 기술은 키워드 인식이나 단순 객체 감지와 같은 단일 모달리티 작업에는 매우 효율적임
  - 그러나 휴머노이드 로봇이나 자율주행차가 요구하는 다수의 카메라, 라이다, 레이더 등의 센서 데이터를 실시간으로 융합하고 복잡한 3D 환경을 인지하는 데 필요한 연산 능력과 다중 입력 처리 능력은 부족함
- 온디바이스 AI는 기기 내 중앙에 위치한 연산장치(SoC)가 모든 원시 데이터를 처리하는 중앙 집중형 구조로, 방대한 양의 데이터를 생성하는 환경에서는 데이터 처리 지연, 병목 및 발열 등의 문제가 발생 가능
  - 여러 센서의 원시 데이터를 융합하는 데이터 레벨 융합은 방대한 정보 공간을 생성하여 의사결정 절차를 지연시킬 수 있음
  - 특정 센서가 고장날 경우 전체 시스템이 모호한 결과를 도출할 수 있는 취약점을 가짐
- 피지컬 AI 시대는 멀티모달 센서(카메라, LiDAR, Radar, 마이크, 촉각 등)를 통해 생성되는 이종(Heterogeneous) 데이터를 즉각적으로 분석하고 판단하는 새로운 아키텍처 요구
  - 지각 특화 AI 반도체는 이 요구를 충족시키기 위해 센서와 중앙의 메인 연산 SoC 사이의 중간 계층을 담당하는 핵심 노드가 될 수 있음

### 가. 목표시장 정의

- (정의) 데이터가 발생하는 지점(센서 단)에서 이중의 멀티모달 데이터를 실시간으로 융합·추론하고 지능적으로 전처리함으로써, 메인 SoC의 부하를 경감하는 분산적·계층적 처리 구조의 반도체 시장
  - 기존 센서 시장이나 AI 반도체 시장의 일부가 아니라, 지각 능력이 핵심 경쟁력으로 작용하는 휴머노이드 로봇, 자율주행차, XR 기기 등 미래 산업 전반에서 독자적인 가치를 창출하는 신규 시장으로 정의할 수 있음
  - 센서와 메인 연산 SoC 사이의 중간 계층으로서 시스템 아키텍처의 근본적인 전환을 촉진하며, 지연 시간 단축, 데이터 병목 해소, 에너지 효율 향상이라는 차별화된 가치를 창출 가능



[그림 8] 지각 특화 AI 반도체의 목표시장



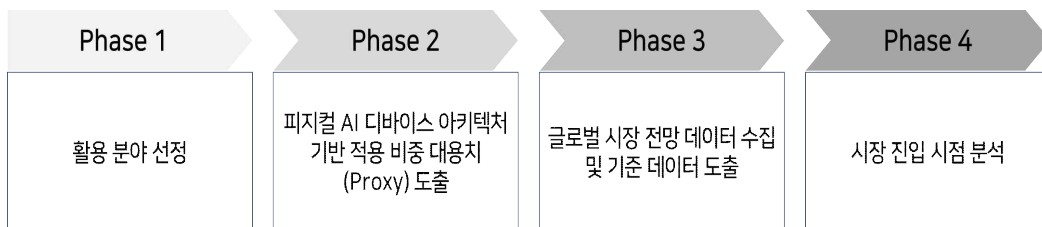
[그림 9] 유사 시장과의 규모 비교

- 지각 특화 AI 반도체는 시장 초기('31년 기준) 온센서 AI, 온디바이스 AI 반도체보다 상대

- 적으로 규모가 작으나, '35년에는 온센서 AI 시장 규모를 넘어설 것으로 전망([그림 8] 참조)
- 지각 특화 AI 반도체의 목표시장은 초기인 '31년에는 77억 달러로 시작하여 온센서 AI 시장(199억 달러), 온디바이스 AI 반도체 시장(837억 달러)보다 상대적으로 규모가 작음
    - (온센서 AI) 199억 달러('31년) → 277억 달러('35년, CAGR 8.5%)<sup>94)</sup>
    - (온디바이스 AI) 837억 달러('31년) → 1,298억 달러('35년, CAGR 11.6%)<sup>95)</sup>
    - (지각 특화 AI 반도체<sup>96)</sup>) 180억 달러('31년) → 420억 달러('35년, CAGR 23.7%)

## 나. 시장 규모 추정 방법

- 본 시장은 새롭게 부상하는 초기 단계로 아직 기존 시장조사기관의 정량적 자료가 부재하며, 지능형센서나 온디바이스 AI 등 유사 시장 자료를 인용하는 것은 기능적 공백과 독자적 아키텍처 가치를 제대로 반영하기 어렵다는 한계가 존재함
- 이러한 한계를 극복하기 위해, 피지컬 AI 시대의 니즈가 반영된 활용 분야를 선정하여 그 시장 규모를 도출하고, 활용 분야별 지각 특화 AI 반도체의 비중에 대한 대응치(proxy)를 도출하여, 목표시장 규모를 추정하였으며, 다음의 4단계로 구분
  - 이러한 접근은 기존 시장조사기관의 자료를 단순히 인용하는 수준을 넘어, 기능별 원가 기여도와 적용 가능성, 나아가 기술의 상용화 시점까지 반영한 정량적 추정이라는 점에서 차별성을 가짐



[그림 10] 시장규모 추정 순서도

- (Phase 1) 핵심 활용 분야(Key Application) 선정
  - ① 실시간 환경 인식, ② 멀티모달 대용량 데이터 처리, ③ 초저지연·고효율 엣지 임베디드 연산이라는 3대 기술적 임계(Criticality)를 동시에 요구하는 휴머노이드 로봇, 자율주행차, XR 기기를 선정함
  - (분석 대상 제외 기준) 스마트폰·웨어러블은 단일 센서 기반의 단순 감지에 그치며, 산업용 로봇은 통제된 환경 내의 정밀 제어가 목적이므로, 현존하는 온센서 AI나

94) SKYQUEST. (2024.12.). Microcontroller Units (MCUs) Market Size, Share, and Growth Analysis. : 본 고에서는 가장 유사한 HW 특성을 가진 32-bit MCU 시장 데이터를 시장 규모 대응치(proxy)로 채택

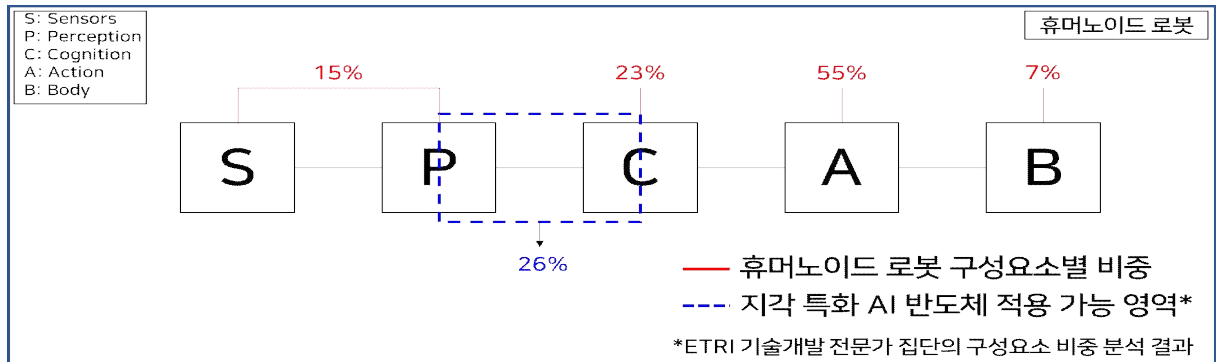
95) 한국수출입은행. (2024). AI반도체 시장 현황 및 전망

96) 본 연구 목표시장의 추정 규모로서 뒷 부분에서 상세하게 설명

MCU로 대응 가능한 영역은 분석에서 제외함

○ (Phase 2) 퍼지컬 AI 디바이스 아키텍처 기반 적용 비중 대용치(Proxy) 도출

- (휴머노이드 로봇) 퍼지컬 AI 디바이스의 S-P-C-A 구조 중 지각(S+P: 15%), 인식(C: 23%) 영역의 비중을 기반으로 도출 ([그림 11] 참조)
- ETRI 관련 기술 전문가들의 심층 분석을 통해 지각 특화 AI 반도체의 실질적 기여분을 총 26%로 도출 : 휴머노이드 로봇 시장에서의 비중 대용치로 사용



### Humanoid Robots – Key Components and Potential Suppliers Worldwide

**Mental Plane - Decision Making -** Accounts for ~23% of total cost

- AI training & AIGC solutions
- OS
- Autopilot
- SoC
- Machine vision solutions
- Remote control solutions, etc.

**Sensing Plane - Data Collection -** Accounts for ~15% of total cost

- 6D torque sensors
- Lenses and camera modules
- CMOS image sensors
- Tactile sensors
- Environment sensors
- MEMS microphones, etc.

**Movement Plane - Joint Activities -** Accounts for ~55% of total cost

- Coreless motors
- Brushless DC motors
- Planetary roller screws
- Roller screws
- Harmonic drives
- Metal structural parts, etc.

**Power Plane - System Operation -** Accounts for ~7% of total cost

- Batteries and control modules
- Charging modules
- Cooling solutions
- Power transmission
- Drivers
- Encoders, etc.

Source: TrendForce, Feb. 2025

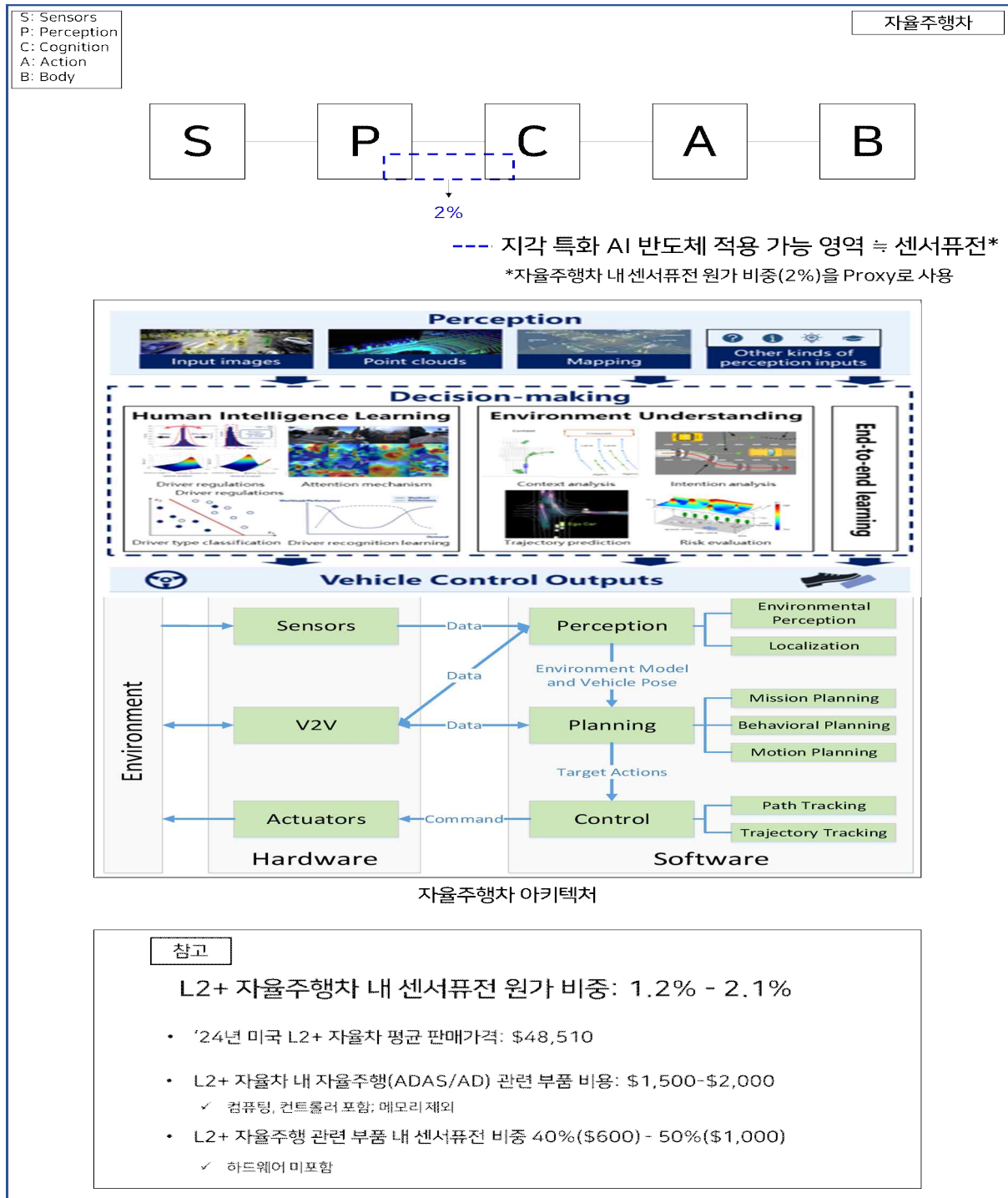
휴머노이드 로봇 구성요소

	Mental Layer - Decision Processing	관련 적합도					Sensing Layer - Data Collection	관련 적합도				
		없음	하	중	상	일치		없음	하	중	상	일치
SW	Autonomous navigation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	MCUs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	AI training and AIGC Solutions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Lens and camera modules	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Machine vision solution	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	VCSEL	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Control and remote control solutions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	EEL	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	HMI solutions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CMOS Imagine sensors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HW	OS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Tactile sensors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	SoC	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Torque sensors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	RF component and communication modules	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6D torque sensors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	DRAM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	IMUs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	NAND Flash	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Environmental sensors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	ISP	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ePapaer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	DSP	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	MEMS microphone	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Security Chip	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

휴머노이드 로봇 구성요소별 세부 기술과 지각 특화 AI 반도체 간의 관계 분석표

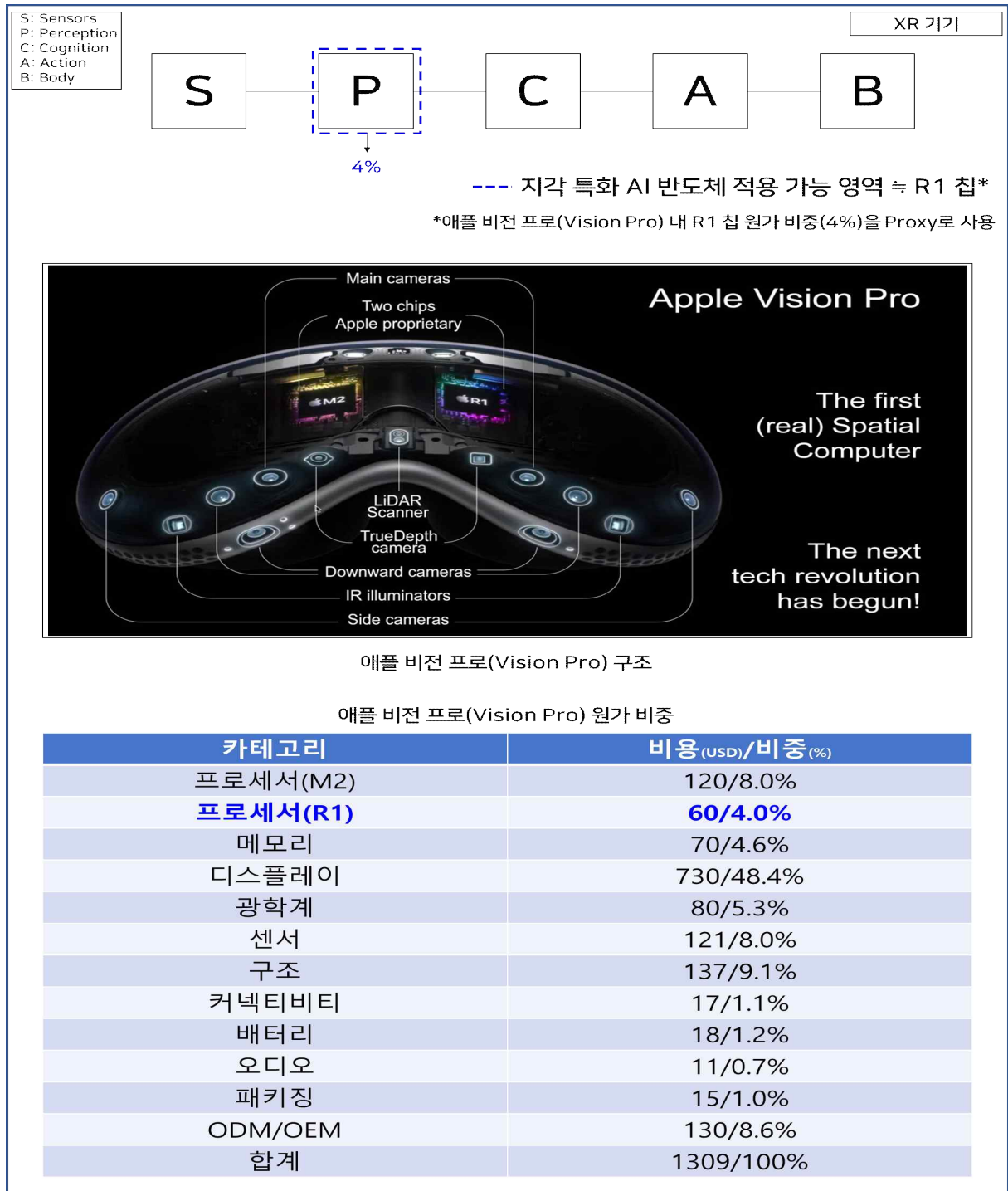
[그림 11] 휴머노이드 로봇 내 지각 특화 AI 반도체 적용 비중 도출 방법

- (자율주행차) L2+ 자율주행차 평균 판매가격에서 센서 퓨전 모듈의 비중 (1.1-1.5%)과 `35년 글로벌 자율주행차 시장 규모 추정치에서 자동차용 센서 퓨전 시장이 차지하는 규모 2.4%를 고려하여, 최종적으로 2%를 자율주행차 시장에서의 비중 대응치로 사용([그림 12] 참조)



[그림 12] 자율주행차 내 지각 특화 AI 반도체 적용 비중 도출 방법

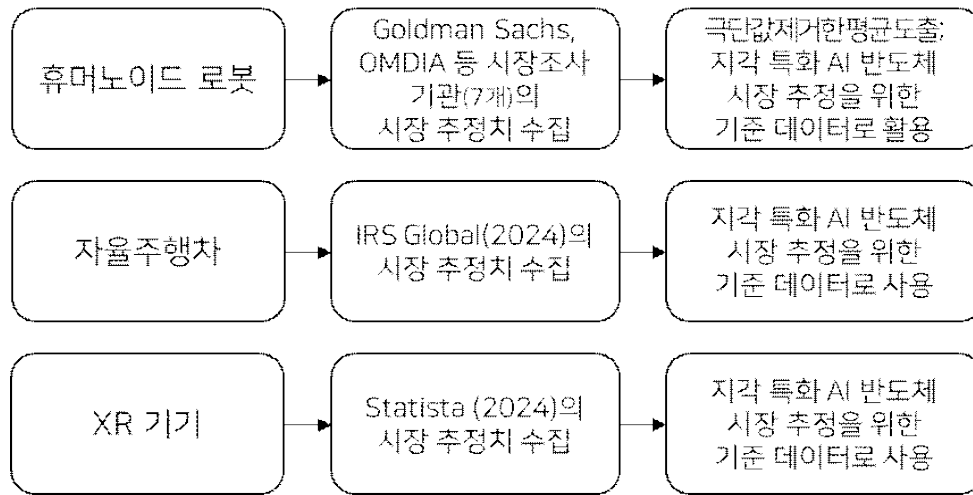
- (XR 기기) 애플 비전 프로(Vision Pro)의 전체 원가 중 센서 데이터 처리를 전담하는 연산 장치(R1칩)가 차지하는 비중인 4%를 XR 기기 시장에서의 비중 대용치로 사용 ([그림 13] 참조)



[그림 13] XR 기기 내 지각 특화 AI 반도체 적용 비중 도출 방법

○ (Phase 3) 글로벌 시장 전망 데이터 수집 및 기준 데이터 도출

- 각 활용 분야의 전체 시장 규모에 대한 글로벌 시장조사기관들의 전망치를 수집하고, 비교하여 기준 데이터를 도출



[그림 14] 활용 분야별 시장 규모 추정 데이터 수집 방법

○ (Phase 4) 시장 진입 시점 분석

- 기술 공급과 시장 수요 측면을 종합적으로 고려하기 위해, 유사 기술의 연구개발 및 양산 소요 기간, 기술 성숙도(TRL), 목표시장의 대중화 시기 등을 고려하여 기술이 상용화될 최적의 시점을 분석

다. 활용 분야 선정 및 분야별 지각 특화 AI 반도체 적용 비중 도출

■ 지각 특화 AI 반도체가 활용되는 분야(시장) 선정

- 지각 특화 AI 반도체는 ① 실시간 환경 인식, ② 멀티모달 대용량 데이터 처리, ③ 초저지연·고효율 엣지 임베디드 연산이라는 세 가지 특성이 동시에 요구되는 분야에서 그 가치가 극대화될 수 있음
  - (실시간 환경 인식) 물리적 환경 속에서 주변 상황을 지속적으로 감지하고, 수집된 정보를 바탕으로 즉각적인 판단이 가능해야 함
  - (멀티모달 대용량 데이터 처리) 카메라, 레이더, 마이크 등 이종의 다수 센서로부터 방대한 양의 데이터를 동시에 수집하고 통합적으로 처리할 수 있어야 함
  - (초저지연·고효율 엣지 임베디드) 에너지와 연산 자원이 제한된 엣지 디바이스 환경에서 동작하며, 매우 짧은 반응 지연 시간과 높은 에너지 효율성을 보장해야 함
- 다양한 잠재적 활용 분야 중 이상의 세 가지 특성이 동시에 필요한 활용 분야로 휴머노이드 로봇, 자율주행차, XR 기기를 우선 선정
  - (휴머노이드 로봇) 지각-인식-행동(P-C-A) 구조에서 센서 단에서 영상·음성 데이터를

전처리하고 핵심 정보만 메인 SoC로 전달함으로써, 반응 속도를 향상시키고 연산 부하와 전력 소모를 줄여 배터리 지속시간을 확대할 수 있음

- (자율주행차) 물체 인식과 거리 산출 등 지각 연산을 센서 단에서 실시간으로 수행하고, 메인 SoC는 메타데이터 기반 판단에 집중함으로써 시스템 전체의 응답성과 안정성을 향상시킬 수 있음
- (XR 기기) 멀티모달 센서 데이터를 센서 인접 엣지에서 실시간 처리함으로써 지연을 최소화하고 몰입감을 높이는 동시에 전력 효율·발열 관리 개선 및 폼팩터 소형화를 가능하게 할 수 있음
- 해당 활용 분야들은 실시간 센싱·판단·행동이 필수적이며, 연산 반도체 성능이 안전성과 몰입감에 직결되어 지각 특화 AI 반도체의 가치가 극대화되는 핵심 목표시장으로 판단
- 스마트폰, 산업용 로봇, 드론·소형 가전 시장은 기술 요구 수준의 차이로 인해 본 연구의 주요 활용 분야에서 제외함
  - (스마트폰/웨어러블) 단일 센서(이미지, IMU) 기반의 단순 이벤트 감지가 주를 이루며, 기존 온센서 AI(Sony IMX500 등) 또는 모바일 AP 수준으로 충분히 대응 가능함
  - (산업용 로봇) 통제된 환경에서의 반복 작업 수행이 목적이며, 고차원 환경 인식보다는 정밀 제어가 핵심으로 기존 고성능 MCU 시장 영역에 해당함
  - (드론/소형 가전) 배터리·중량 제약으로 기능이 제한되고, 복합 센서 융합보다는 특정 기능 중심으로 구현되어 단기 진입 시장으로는 유효하나, 대규모 시장 형성을 견인할 핵심 동력으로는 한계가 있음

#### ■ 휴머노이드 로봇 시장에서의 지각 특화 AI 반도체 적용 비중 도출

- 휴머노이드 로봇은 동적·비정형 환경에서 지각-인식-행동(P-C-A) 루프를 지속적으로 수행해야 하는 피지컬 AI 기술의 집약체로, 지각 특화 AI 반도체의 필요성을 가장 명확히 보여주는 활용 분야임<sup>97)</sup>
  - (개발·배포 흐름) 물리적으로 정확한 시뮬레이션을 통해 합성 데이터를 생성하고, 이를 실제 데이터와 결합해 범용 로봇 파운데이션 모델을 학습한 뒤, 로봇 내장형 컴퓨팅 유닛에 배포하여 실제 로봇을 구동
  - (핵심 과제) 시뮬레이션과 실제 환경 간의 간극이 주요 난관이며, 지각 특화 AI 반도체는 다수 센서의 원시 데이터를 실시간으로 융합, 정제함으로써 일관된 3D 환경 표현을 생성함으로써 이를 완화할 수 있음

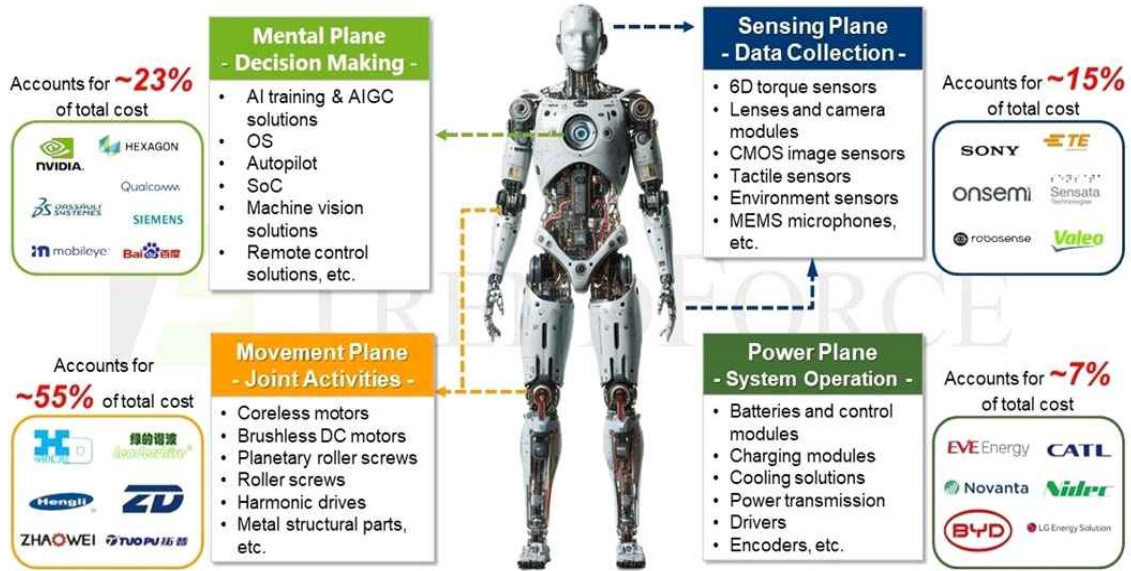
97) Bhide, A., Dzhumamuratova, A., and Welsh, J. (2025). Build Synthetic Data Pipelines to Train Smarter Robots with NVIDIA Isaac Sim

- (역할 분담 효과) 센서 데이터 전처리 및 초기 판단을 지각 특화 AI 반도체가 담당함으로써, 메인 컴퓨팅 유닛은 고수준 추론과 행동 생성에 집중할 수 있음
- 휴머노이드 로봇의 구성요소<sup>98)</sup>는 지각(Perception) 15%, 인식(Cognition) 23%, 행동(Actuation) 55%, 운영(Operation) 7%로 구분되며, 지각 특화 AI 반도체는 지각 영역 전반과 인식 영역의 초기 의사결정·특징 추출을 포괄함
- 적용 비중 산정을 위해 보수적·낙관적 시나리오를 검토하였으나, 각 시나리오는 로봇 고유의 인식 연산 또는 전처리·판단 기능을 충분히 반영하지 못하는 한계가 있음
  - 보수적 시나리오 : 자율주행차 센서 퓨전 원가 비중을 준용해 3% 적용
    - 자율주행차의 센서 총비용 중 센서 퓨전 모듈이 차지하는 원가 비중 20%<sup>99)</sup>를 벤치마킹
  - 낙관적 시나리오 : 지각 영역 하드웨어 대체율 가정으로 15% 적용
    - 휴머노이드 로봇의 지각 영역에 들어가는 센서 및 부품들의 기능을 대부분 수행 가능하다고 판단
- 본 연구에서는 휴머노이드 로봇 핵심 기술 요소와의 기능적 연관성을 정밀 분석한 ETRI 관련 기술(인공지능, 반도체, 온디바이스 AI, 로봇 등) 전문가 의견들을 반영하여, 지각 10% 비중과 인식 16% 비중을 합산한 총 26%(상세 내용은 부록 1 참고)를 지각 특화 AI 반도체의 최종 적용 비중으로 채택함
  - 보수적/낙관적 시나리오의 한계
    - 공통적으로 단순 참조에 따른 기술적 특성 반영 미흡
    - 보수적 시나리오(3%, 자율주행차의 센서 퓨전 원가 비중을 적용)의 한계 : 로봇 고유의 인식(Cognition) 관련 연산 부하 미반영
    - 낙관적 시나리오(15%, 기존 센서 하드웨어(Sensing Layer) 대체율에 기반)의 한계 : 지각 특화 AI 반도체의 전처리 및 판단 기능 미포괄
  - 전문가 의견 기반 산정 근거
    - 단순 센서 기능을 넘어 지능형 연산(전처리 및 판단)까지 포괄
    - 적용 범위의 확장성과 기술 현실성을 동시에 반영
    - 영역별 적합도 분석 (관련 기술개발 전문가 의견 수렴)

98) TrendForce. (2025.7.). Humanoid Robot Quarterly Report

99) PatentPC. (2025.10.). The Cost of Self-Driving Technology: How Much Do AV Components Really Cost? (Market Breakdown)

Humanoid Robots – Key Components and Potential Suppliers Worldwide



Source: TrendForce, Feb. 2025

TRENDFORCE

[그림 15] 휴머노이드 로봇의 구성요소 및 비중

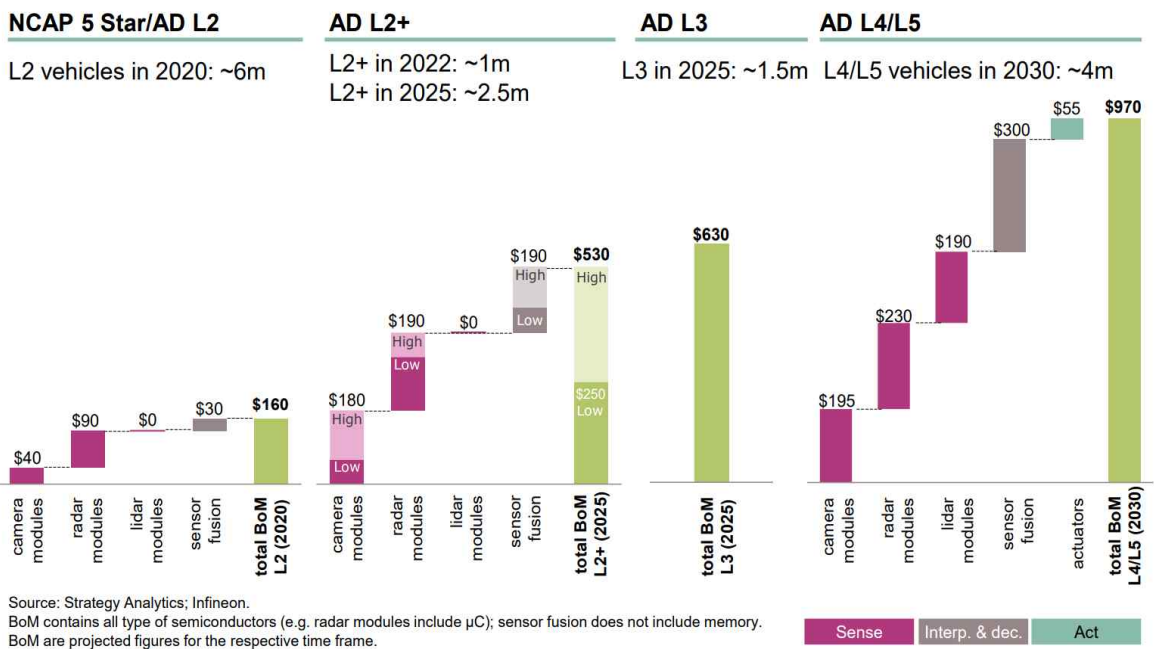
■ 자율주행차 시장에서의 지각 특화 AI 반도체 적용 비중 도출

- 자동차 산업은 다수의 분산형 ECU에서 소수의 고성능 중앙 컴퓨터로 통합되는 아키텍처로 전환 중이며, 안전과 직결된 자율주행 특성상 대규모 센서 데이터를 처리하는 지능형 엣지 프로세싱의 중요성이 증대
  - NVIDIA의 차량용 SoC 'DRIVE Thor'는 단일 칩으로 자율주행, 인포테인먼트, 클러스터를 통합 처리하는 중앙집중형 구조를 지향하며, 개발자 키트 기준 다수의 카메라와 라이다/레이더 입력을 위한 대규모 데이터 대역폭이 요구됨<sup>100)</sup>
- 자율주행차 시스템은 지각-판단-제어의 계층적 구조를 가지며, 지각 계층은 다수 센서 데이터를 실시간으로 융합해 환경을 구조화하는 핵심 영역임
  - 센서 퓨전은 단일 센서의 한계를 상호 보완하여 인식 정확도와 신뢰성을 향상시킴
- 지각 특화 AI 반도체는 센서로부터 수집된 원시 데이터를 센서 단에서 실시간 융합·추론하여, 메인 SoC의 연산 부담을 경감할 수 있음
  - (데이터 통합·동기화) 분산 센서 데이터를 동기화하고 초기/중간 레벨 융합 후 메인 SoC로 전달
  - (안전성·이중화) 메인 SoC 장애 시 최소 안전 기능을 독립 수행하는 예비 경로(fallback) 제공

100) <https://developer.nvidia.com/drive/agx>

[참고 3] 센서 퓨전이 자율주행차 평균 판매가격에서 차지하는 비중

- `24년 미국의 L2+ 자율주행차(LiDAR 포함) 평균 판매가격은 48,510달러<sup>101)</sup>
- L2+(LiDAR 포함) 자율주행차 내 ADAS/AD 반도체 비용은 약 1,500~2,000달러<sup>102)</sup>
- L2+ 자율주행차 내 ADAS/AD 반도체 BoM에서 센서 퓨전\*이 차지하는 비중은 35.8%<sup>103)</sup>
  - L2(`20년 기준): 18.8%, L2+(`25년 기준) : 35.8%, L4/L5(`30년 기준): 30.9%
  - \* 프로세서와 컨트롤러 포함, 센서 및 메모리 제외
- 따라서 센서 퓨전이 L2+ 자율주행차 평균 판매가격에서 차지하는 비중은 1.1% - 1.5% 수준임



[ 자율주행 단계별 반도체 원가 구조 및 센서 퓨전 비중 변화 ]

- 센서 퓨전에 포함되는 요소
  - 프로세서/컨트롤러(MCU, SoC) : 센서로부터 받은 데이터를 통합하고 판단/제어를 수행하는 중앙 컴퓨터 유닛(ex. 인피니언 AURIX™ 계열 등)
  - 센서 데이터 처리 인터페이스 : 센서 모듈로부터 들어오는 데이터 스트림을 받아 처리하기 위한 입출력(패키지, 버스, 연결망) 비용 및 연관 반도체
  - 판단/융합 알고리즘 운영 하드웨어 : 센서 데이터 융합, 객체 인식, 경로 판단 등을 위해 실시간 운영되는 하드웨어(프로세서/가속기) 비용 일부
- 센서 퓨전에 포함되지 않는 요소
  - 센서 모듈 : 카메라, RADAR, LiDAR 등은 별도 비용 항목
  - 메모리 : Flash, DRAM, SRAM 등 제외
  - 센서 하우징/마운트, 센서 통합 구조, 시스템 소프트웨어(운영체제, 지도데이터, 업데이트 기능) 등 미포함
  - 액추에이터(steering, brake, etc) 및 센서 외 주변 전장품 미포함

- 자율주행차 내 지각 특화 AI 반도체의 비중 산정을 위해, 기능적으로 유사한 센서 퓨전 모듈의 원가 비중을 대용치로 적용
  - 평균 판매가격 대비 센서 퓨전 비중은 1.1~1.5%
    - `24년 미국 L2+ 승용차(LiDAR 포함)의 평균 거래가격(48,510달러) 대비, 전체 ADAS/AD 부품 비용(1,500~2,000달러) 내에서 센서 퓨전이 차지하는 비중 35.8%를 역산한 결과
  - `35년 기준 글로벌 센서 퓨전 시장(180억 달러)은 자율주행차 시장(1.1조 달러)의 약 2.4%로 확인됨<sup>104)105)</sup>
- 상기 근거를 종합하여, 자율주행차 시장 내 지각 특화 AI 반도체의 비중을 2%로 채택
  - 자율주행 레벨 고도화(L4-L5)에 따라 센서 단 지능 처리 수요가 증가할 가능성이 높아, 향후 비중 상향 조정 여지가 있음

■ XR 기기 시장에서의 지각 특화 AI 반도체 적용 비중 도출

- XR 기기의 사용자 경험을 좌우하는 핵심 기술 지표는 모션-투-포톤(Motion to Photon, MTP) 지연시간으로, 사용자의 움직임이 디스플레이에 반영되기까지의 시간을 의미함<sup>106)</sup>
  - MTP 지연이 증가할 경우 시각 정보와 전정기관 감각 간 불일치로 인해 멀미 등 사용자 불편이 발생
  - 이를 최소화하기 위해 시선 추적, 손동작 인식, 공간 매핑 등 다양한 센서 입력에 대한 실시간 반응이 필수적임
- 이러한 초저지연(Ultra-Low Latency) 요구를 충족하기 위해, XR 기기는 메인 연산 SoC와 분리된 센서 데이터 전달 보조 프로세서를 필요로 함
- Apple Vision Pro는 메인 프로세서(M2)와 별도로, 다수의 센서 데이터를 실시간 처리하는 R1 칩을 탑재하여 분산 아키텍처를 구현함
  - R1 칩은 센서 데이터를 처리해 약 10~12ms<sup>107)</sup> 이내로 디스플레이에 반영함으로써, 사용자에게 지연 없는 몰입형 경험을 제공
- 지각 특화 AI 반도체는 R1 칩과 유사하게 중간 계층에서 센서 융합 및 전처리를 담당함

101) Center for Automotive Research. (2024.8). Affordability: The Twenty-Five Thousand Dollar Electric Vehicle, Center for Automotive Research  
 102) <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/autonomous-drivings-future-convenient-and-connected>  
 103) Infineon. (2019.12). Semiconductors enabling Automated Driving, Infineon, Liberum Future of E-Mobility Conference  
 104) Statistics Market Research Consulting. (2024.10.). Automotive Sensor Fusion Market Forecast to 2030  
 105) IRS Global. (2024.12.). 차세대 자동차 Global Market Data: Current Status & Outlook(2025)  
 106) OPTOFIDELITY. (2024.2.). Apple Vision Pro Benchmark Test 2: Angular Motion-to-Photon Latency in VR  
 107) 10ms : 0.01초

으로써, 메인 SoC의 부하를 경감하고 초저지연 처리, 전력 효율 및 발열 관리 개선, 폼팩터 소형화를 가능하게 함

- XR 기기 내 시각 특화 AI 반도체의 비중 산정을 위해, Apple Vision Pro의 원가명세서 (BoM, Bill of Materials) 분석에서 R1 칩이 차지하는 원가 비중(약 4%<sup>108</sup>)을 대응치로 적용
  - 향후 XR 기기의 입력 채널이 고도화되고 현실-가상 상호작용이 복잡해질수록, 시각 특화 AI 반도체의 비중은 추가 확대될 가능성이 있음

<표 8> 애플 비전 프로(Vision Pro)의 원가 비중

Category	Component	Specification /Model	Supplier	Unit Price (USD)	Qty	Total (USD)
Processor	Main Processor	M2 Series	Apple	120	1	120
<b>Processor</b>	<b>Coprocessor</b>	<b>Vision Image Processing Chip</b>	<b>Apple</b>	<b>60</b>	<b>1</b>	<b>60</b>
Memory & Storage	ROM	UFS4.0 512G	Samsung/Kioxia	20	1	20
Memory & Storage	RAM	LPDDR5 12G	Hynix/Samsung/Micron	30	1	30
Memory & Storage	WiFi SIP	WiFi 6	Broadcom/Skyworks	6	1	6
Memory & Storage	BLE	Bluetooth 5.3	Broadcom/Skyworks	2	1	2
Memory & Storage	PMIC	-	Apple/ST/TI etc.	-	-	4
Memory & Storage	Others	Codec, Audio Amp, LED Driver, Motor Driver, Capacitor etc.	-	-	-	8
Display	External Display	Custom AMOLED	LG	30	1	30
Display	Internal Display	1.3" Micro-OLED	Sony	350	2	700
Optics	Optical Lens	Pancake 3P	Yujing/Young Optics	30	2	60
Optics	IPD Adjustment Module	-	GSEO	10	2	20
Sensors	6DOF Tracking	Sony IMX418	Lens: Largan/GSEO; Prism: Crystal-Optech	5	4	20
Sensors	VST Camera	RGB	Lens: Largan/GSEO; Prism: Crystal-Optech	8	2	16
Sensors	Eye Tracking	WLO Package	Sony	12	2	24
Sensors	Facial Tracking	WLO Package	Sony	12	2	24
Sensors	Body Tracking	Sony IMX418	Lens: Largan/GSEO; Prism: Crystal-Optech	5	2	10
Sensors	Hand Tracking	Structured Light XR+TX	Lens: Largan/GSEO;	10	1	10

108) Wellsenn XR. (2023). Apple MR disassembly and BoM list

			Prism: Crystal-Optech			
Sensors	TOF	dTOF Sony IMX611	Lens: Largan/GSEO; LG	10	1	10
Sensors	IMU	-	TDK	3	1	3
Sensors	Haptics	-	-	2	2	4
Structure	Structural Parts	Middle frame, Outer shell, parts with carbon fiber/magnesium alloy	Lens Technology/Goertek	120	1	120
Structure	Thermal Module	Heatsink & Fan	-	9	1	9
Structure	Others	Adhesives, Foam	-	8	1	8
Connectivity	PCB	-	Pengding	8	1	8
Connectivity	FPC	-	Pengding	6	1	6
Connectivity	External Power Cable	-	-	3	1	3
Battery	Headset Battery	Approx. 5000mAh	-	3	1	3
Battery	External Battery	Approx. 17Wh	Desay Battery	15	1	15
Audio	Microphone	Omnidirectional	Goertek	1	3	3
Audio	Speaker	-	Goertek	2	4	8
Packaging	Packaging Parts	Outer box, charger accessories	Luxshare Precision	15	1	15
ODM/OEM	Assembly	-	Luxshare Precision	130	1	130
Total						1,509

출처: Apple MR disassembly and BoM list, Wellsenn XR, 2023.

## 2. 활용 분야별 확산 전망과 시장 진입 시점 분석

본 절에서는 휴머노이드 로봇, 자율주행차, XR 기기를 중심으로 지각 특화 AI 반도체의 활용 분야별 시장 확산 경로와 성장 특성을 분석한다. 또한 기술 공급 측면과 시장 수요 측면을 종합적으로 고려하여, 본 기술의 현실적인 시장 진입 시점과 최적의 진입 타이밍을 도출한다.

### 가. 휴머노이드 로봇 시장

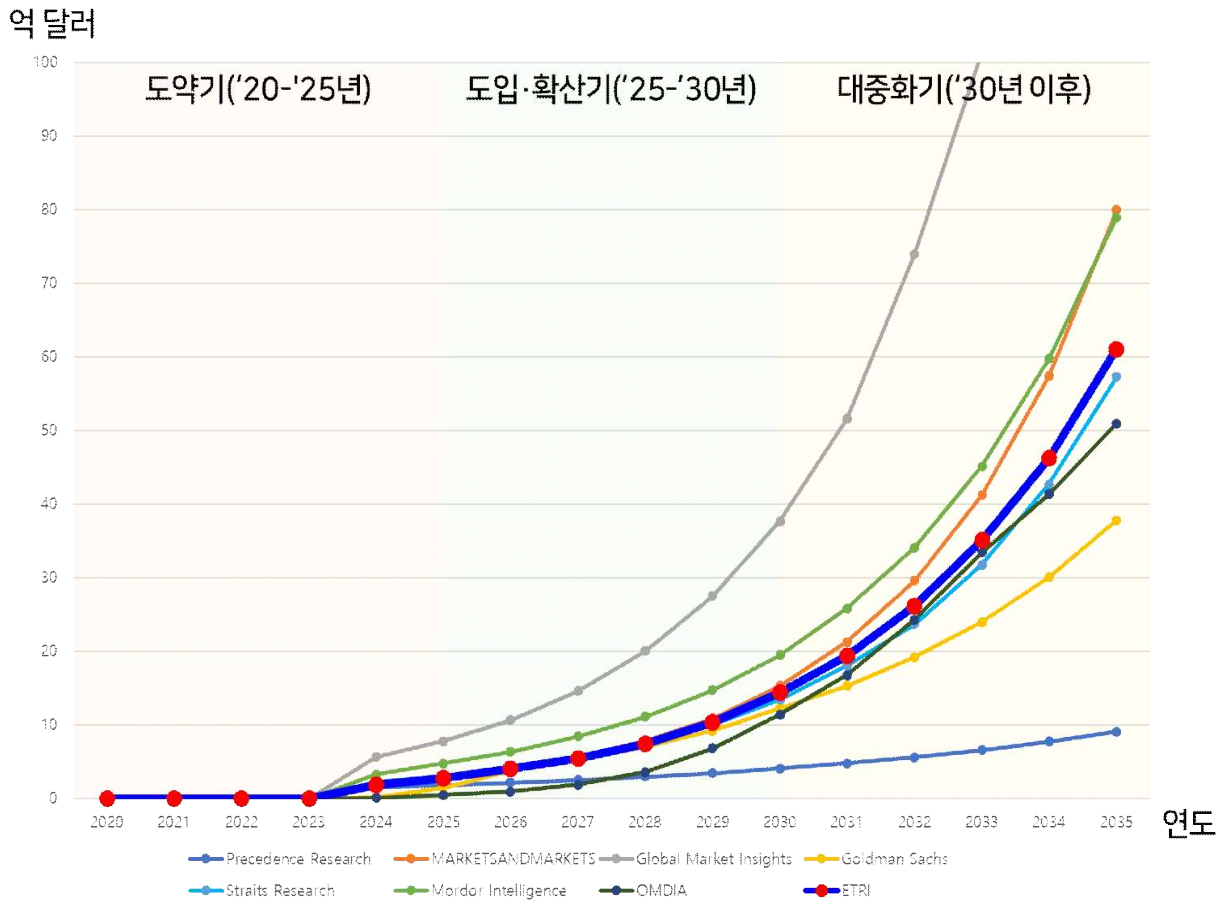
- 글로벌 휴머노이드 로봇 시장은 아직 본격적인 대중화 이전 단계로, 예측 방법론, 시장 획정, 상용화 시점 등의 차이로 인해 기관별 전망치 간 편차가 큼
  - 이는 시장의 불확실성과 기술 도입 속도에 대한 기관별 시각 차이에 기인함

〈표 9〉 글로벌 시장조사기관별 시장 규모 추정 방법

기관	지표	시장획정	주요 성장요인
Precedence Research	·공급자 매출 기반 외삽 ·초기 매출 실적 기반 성장 추세 예측	·휴머노이드 로봇 : 이족보행+휠 -HW(센서·모터·배터리) -SW(AI제어·OS)	·인간-로봇 상호작용 ·클라우드 연동
MARKETSAND MARKETS	·기기 출하량·단가 ·산업규모 비중	·휴머노이드 로봇 : 이족보행+휠 -HW(센서·액츄에이터·제어) -SW(AI·제어)	·AI 제어 ·객체 인식 ·음성 및 상황 인지
Global Market Insights	·부품+SW 매출 합산 ·수요·기술 보급률	·휴머노이드 로봇 : 이족보행+휠 -HW(센서·액츄에이터·전원) -SW(제어·OS)	·AI/ML 제어 ·객체 인식 ·IoT/디지털트윈
Goldman Sachs	·시물레이션 기반 분석 (기술확산·비용·보급률 등)	·휴머노이드 로봇 : 인간 유사(머리·팔·다리) -HW(바디·센서·모터·배터리) -SW(AI제어·LLM지각엔진·클라우드OS)	·부품 단가 하락 ·AI, 모션, 인지 향상 ·고위험 노동 대체
Straits Research	·정성분석(설문·사례) ·정량분석(통계)	·휴머노이드 로봇 : 이족보행+휠 -HW(액츄에이터·센서·전력) -SW(AI제어·자연어·모션제어)	·고령화 대응 ·아태지역 제조/정책 지원
Mordor Intelligence	·출하량·단가 ·전문가 인터뷰 ·기술트렌드 분석	·휴머노이드 로봇 : 이족보행+휠 -HW(센서·액츄에이터·배터리) -SW(AI제어·자연어처리·모션인식)	·AI HW 비용 하락 ·고령화/노동력 격차 ·아태지역 투자 확대
OMDIA	·1차(산업 관계자 인터뷰)+ 2차(보도자료/기업자료 등) 조사 및 전시회/세션 참여를 결합해 데이터 구성	·휴머노이드 로봇 : 이족보행 -HW(센서·액츄에이터·전력·반도체) -SW(AI/ML 제어·파운데이션 모델 기반 지각/조작)	·고령화(노동력 부족)+제조업 리쇼어링 ·물류 중심 AGV·AMR 대규모 도입 ·휴머노이드·4족 고성장+AI (경량 ML~파운데이션 모델) 결합

출처: 각 시장조사기관들의 자료를 종합하여 ETRI 기술전략연구본부 작성, 2025.11.

- 본 연구에서는 객관적이고 안정적인 기준값 도출을 위해 절사평균(Trimmed Mean)을 적용하여 휴머노이드 로봇 시장 규모를 재추정함
  - 절사평균은 극단치의 영향을 배제함으로써 일반평균 대비 통계적으로 강건한 추정치를 제공<sup>109)</sup>
- 추정 결과, 글로벌 휴머노이드 로봇 시장 규모는 `25년 32억 달러(4.5조 원<sup>110)</sup>)에서 `35년 635억 달러(88.9조 원)로 성장할 것으로 전망([그림 16] 및 <표 10> 참조)



[그림 16] 글로벌 휴머노이드 로봇 시장 규모 추정 결과(주요 기관별 예측치 포함)

109) Lugosi, G., and Mendelson, S. (2021). "Robust multivariate mean estimation: The optimality of trimmed mean," Ann. Statist., vol. 49, no. 1, 2021, pp. 393-410

110) 환율 1,400원 기준, 이하 동일

〈표 10〉 글로벌 휴머노이드 로봇 시장 전망치 (단위: 십억 달러)

구분	`24	`25	`26	`27	`28	`29	`30	`31	`32	`33	`34	`35	CAGR
Precedence Research	1.6	1.8	2.2	2.5	3.0	3.5	4.1	4.8	5.6	6.6	7.8	9.1	('24-'34) 17.3%
MARKETS AND MARKETS	1.8	2.9	4.0	5.6	7.8	10.9	15.3	21.3	29.6	41.3	57.4	80.0	('25-'30) 39.5%
Global Market Insights	5.7	7.8	10.7	14.7	20.1	27.5	37.7	51.6	74.0	101.4	138.9	190.3	('24-'34) 37.6%
Goldman Sachs	0.2	1.5	3.8	5.4	7.0	9.2	12.3	15.3	19.2	24.0	30.1	37.8	('24-'35) 61.0%
Straits Research	2.3	3.1	4.1	5.6	7.5	10.0	13.4	18.0	23.7	31.8	42.7	57.3	('24-'32) 33.9%
Mordor Intelligence	3.3	4.8	6.4	8.4	11.2	14.8	19.5	25.8	34.1	45.1	59.7	79.0	('25-'30) 32.3%
OMDIA	0.1	0.5	1.0	1.9	3.6	6.9	11.4	16.8	24.3	33.5	41.3	51.0	('24-'35) 75.5%
ETRI	1.8	2.8	4.1	5.5	7.4	10.3	14.4	19.4	26.2	35.1	46.3	61.0	('24-'35) 37.6%

- 주1) 회색 음영은 CAGR 기반 예측치
- 주2) Precedence Research, Humanoid Robot Market Size, Share, and Trends 2024, to 2024(2024.11.) 시장 전망치(`24년 16억 달러 → `34년 78억 달러)를 기반으로 `35년 ETRI 기술전략연구본부 추정, 2025.11.
- 주3) MARKETANDMARKETS, Humanoid Robot Market Size, Share & Trends(2025.4.) 시장 전망치(`25년 29억 달러 → `30년 153억 달러)를 기반으로 `24년, `26-`29년, `31-`35년 ETRI 기술전략연구본부 추정, 2025.11.
- 주4) Global Market Insights, Humanoid Robot Market(2024.8.) 시장 전망치(`24년 57억 달러 → `32년 740억 달러)를 기반으로 `25-`31년, `33-`35년 ETRI 기술전략연구본부 추정, 2025.11.
- 주5) Goldman Sachs, Global Automation Humanoid Robot: The AI Accelerant, 2024.1.
- 주6) Straits Research, humanoid robot market size(2024) 시장 전망치(`24년 23억 달러 → `32년 237억 달러)를 기반으로 `25-`31년, `33-`35년 ETRI 기술전략연구본부 추정, 2025.11.
- 주7) OMDIA, Robotics Hardware Market Forecast, 2025.
- 주8) 7개 시장조사기관 전망치의 편차를 고려하여 상·하위 극단치 1개씩(총 2개)을 제외한 절사평균값임

## 나. 자율주행차 및 XR 기기 시장

- 글로벌 자율주행차 시장 규모는 `25년 1,500억 달러(210조 원)에서 `35년 1조 1,200억 달러(1.6경 원)로 연평균 22.3% 성장 전망<sup>111)</sup>
  - 글로벌 자동차용 센서 퓨전 시장 규모는 `25년 18억 달러(2.5조 원)에서 `35년 265억 달러(37.1조 원)로 연평균 30.8% 성장 전망<sup>112)</sup>

<표 11> 글로벌 자율주행차 및 자동차용 센서 퓨전 시장 규모 전망치 (단위: 십억 달러)

구분	`25	`26	`27	`28	`29	`30	`31	`32	`33	`34	`35	CAGR
자율주행차	150.0	183.4	224.2	274.2	335.2	409.9	501.1	612.7	749.2	916.0	1,120.0	22.3%
자동차용 센서 퓨전	1.8	2.2	2.7	3.3	4.0	6.9	9.0	11.8	15.5	20.2	26.5	30.8%

주1) 회색 음영은 CAGR 기반 예측치

주2) ISR Global, 차세대 자동차 Global Market Data: Current Status & Outlook(2024.12.) 시장 전망치(`25년 1,500억 달러 → `35년 1조 1,200억 달러)를 기반으로 `26-`34년 ETRI 기술전략연구본부 추정, 2025.11.

주3) Statistics Market Research Consulting, Automotive Sensor Fusion Market Forecast to 2030(2024.10.) 시장 전망치(`25년 18억 달러 → `30년 69억 달러)를 기반으로 `31-`35년 ETRI 기술전략연구본부 추정, 2025.11.

- 글로벌 XR 기기 시장 규모는 `25년 466억 달러(65.2억 원)에서 `35년 951억 달러(133.1조 원)로 연평균 7.4% 성장 전망<sup>113)</sup>

<표 12> 글로벌 XR 기기 시장 규모 전망치 (단위: 십억 달러)

구분	`25	`26	`27	`28	`29	`30	`31	`32	`33	`34	`35	CAGR
XR 기기	46.6	51.0	54.7	58.3	62.0	66.6	71.5	76.8	82.5	88.6	95.1	7.4%

주1) 회색 음영은 CAGR 기반 예측치

주2) Statista, AR & VR - Worldwide(2024.4.) 시장 전망치(`25년 466억 달러 → `29년 620억 달러)를 기반으로 `30-`35년 ETRI 기술전략연구본부 추정, 2025.11.

111) IRS Global. (2024.12.). 차세대 자동차 Global Market Data: Current Status & Outlook(2025)

112) Statistics Market Research Consulting. (2024.10.). Automotive Sensor Fusion Market Forecast to 2030

113) Statista. (2024.4.). AR & VR - worldwide

## 다. 시장 진입 시점 분석

### ● 기술 공급 측면에서의 진입 시점

- 지각 특화 AI 반도체의 현실적인 시장 진입 시점을 예측하기 위해, 기술 개발부터 양산까지 소요되는 기간을 다각도로 분석함

#### ■ 관련 기술개발 과제 기간 분석

- NTIS(국가과학기술지식정보서비스) 연구과제 데이터베이스를 활용해 온디바이스 AI, AI 반도체, 휴머노이드 로봇 등 관련 기술 분야의 총 1,485개 연구과제를 분석한 결과, 평균 연구개발 기간은 약 3년 5개월로 산출
  - (휴머노이드 로봇) `03-`25년 총 과제 수 395건, 평균 3년 10개월
  - (온디바이스 AI) `20-`25년 총 과제 수 284건, 평균 3년 2개월
  - (AI 반도체) `18-`25년 총 과제 수 364건, 평균 3년 6개월
  - (멀티모달 센서) `02-`25년 총 과제 수 275건, 평균 3년 3개월
  - (센서 퓨전) `03-`25년 총 과제 수 167건, 평균 3년 3개월

#### ■ 유사 기술성숙도 단계 연구개발 기간 분석

- 지각 특화 AI 반도체는 기술성숙도 기준으로 개념정립 단계(TRL 2)에 해당하며, 실험실 검증(TRL 3)부터 양산 전 실증(TRL 7)까지의 단계적 연구개발이 필요
  - 산업통상자원부 로봇산업기술개발사업 사례 분석 결과, TRL 3~7 단계의 평균 연구개발 기간은 약 3년 8개월로 확인
- 국가 R&D 지원사업의 기술 상용화 기준에 따르면, 기술성숙도가 상용화 직전 단계(TRL 7~8)에 도달하고 양산 준비, 인증, 마케팅, 유통망 구축이 병행될 경우, 연구개발 완료 후 평균 3년 이내 시장 진입이 가능함

#### ■ 유사 상용화 제품의 소요 기간 분석

- 지각 특화 AI 반도체와 유사한 상용화 제품들의 기술개발 발표부터 상용화까지는 평균적으로 5년의 기간이 소요
  - Sony의 IMX 500은 이미지캡처와 AI 추론을 센서 내에서 수행하는 비전센서로 개발 착수부터 출시까지 5년(`16년 기술 개발 착수, `20년 AI 내장 비전 센서 출시)의 기간이 소요<sup>114)</sup>

- BrainChip의 Akida는 이벤트 기반 처리, 실시간 반응 및 초저전력 추론이 가능한 AI 프로세서로 개발 착수(`17년)부터 양산 시작(`21년)까지 5년의 기간이 소요(Sunato, 2021.4.)

## ● 시장 수요 측면에서의 진입 시점

- 시장 진입 시점은 기술 공급 요인뿐 아니라, 실제 수요가 본격적으로 형성되는 시점 (시장 견인, Market Pull)을 함께 고려하여 판단할 필요가 있음
- (초기시장 : `25년-`27년) 휴머노이드 로봇이 산업 현장을 중심으로 실증 및 초기 도입 단계에 진입하는 시기로 전망<sup>115)</sup>
  - 기술 성숙도와 비용 대비 효익(ROI)이 명확한 조립·운반·검사 등 제한된 업무 영역부터 도입이 확산될 것으로 예상
  - 주요 기업들은 제조·물류 현장을 중심으로 PoC 및 실증을 진행하며 초기 수요를 형성 중임
    - 테슬라는 옵티머스 로봇을 기가팩토리(Gigafactory) 생산 라인에 투입해 조립·운반·검사 업무를 자동화하고 있으며, Figure AI사의 Figure 01 모델도 글로벌 제조사 및 물류 기업과의 PoC 단계를 거쳐 실증 중
- (본격 확산 : `28년-`31년) 가정·서비스 영역으로 적용 범위가 확대되며 휴머노이드 로봇의 대중화가 시작되는 시기로 전망(Goldman Sachs, 2024)
  - 돌봄, 가사 보조, 교육, 엔터테인먼트 등 사람과 직접 상호작용하는 분야가 주요 수요처로 부상할 것으로 예상
  - 이 단계에서는 작업 효율성뿐 아니라 안전성, 친밀감, 사회적 수용성이 핵심 평가 기준으로 작용할 것임

## ● 분석 결과

- (타이밍 불일치) 시장은 `28년부터 본격적으로 형성될 것으로 예상되나, 기술 개발에는 평균 3년 이상의 기간이 소요되어 현 시점에서 선제적 대응이 없을 경우 초기 시장 선점 기회를 상실할 우려가 있음
  - 일반적인 선형적 개발 프로세스를 적용할 경우, `28년 이후 가시화되는 초기 성장 국면에 적기 진입이 어려울 가능성이 큼
  - 시장 형성 초기에 주요 고객사와의 협력 관계를 구축하고 기술 표준을 선도하기 위해서는 조속한 R&D 착수가 필요

114) Ryoji, E. (2022.10). Realizing the world's first image sensor with built-in AI engine! Discover the stories behind the drive to create the world's first of its kind

115) AI Matters. (2025). What's Next After ChatGPT? Tesla and Google's Physical AI War

- (최적 시장 진입 시점) 본 고에서는 기술 성숙도와 시장 대중화 시점이 교차하는 `31년을 최적의 시장 진입 시점으로 설정
  - (1단계 : `26-`29년) 핵심 기술 확보를 위한 R&D 착수 및 TRL 3~7 수준의 기술 검증 완료
  - (2단계 : `30년) 양산 공정 확립, IP 포트폴리오 구축, 아키텍처 최적화 등 상용화 준비
- 이러한 단계적 준비를 통해 `31년에 본격적인 시장 진입이 가능하며, 이는 시장 확산 주기와 기술 상용화 속도를 정합적으로 연계함으로써, 초기 수요 선점과 글로벌 경쟁우위 확보를 가능하게 하는 전략적인 진입 시점으로 판단됨
  - `31년 시장 진입은 단순한 제품 출시를 넘어, 글로벌 표준화 주도, 핵심 활용 분야와의 동반 성장, 국가 AI 반도체 산업의 신규 성장축 창출이라는 전략적 의미를 내포함

### 3. 목표시장 규모 추정과 산업 생태계 현황 분석

본 절에서는 앞서 제시한 추정 프레임워크 기반으로 지각 특화 AI 반도체의 글로벌 목표시장 규모와 중장기 성장 전망을 제시하고, 온센서 AI 및 온디바이스 AI 중심의 기존 산업 생태계 분석을 통해 본 시장이 위치하는 경쟁구도와 전략적 공백 영역을 함께 도출한다. 이를 통해 지각 특화 AI 반도체가 기존 시장과 차별화되는 신규 가치 창출 영역으로서 갖는 산업적 의미와 진입 가능성을 종합적으로 검토한다.

#### 가. 목표시장 규모 추정

##### ■ 활용 분야별 목표시장 규모 추정

- 앞서 도출한 휴머노이드 로봇 내 지각 특화 AI 반도체의 반영 가능한 비중(보수적 3%, 낙관적 15%, ETRI 기술 전문가 집단 26%)을 활용하여 글로벌 휴머노이드 로봇용 지각 특화 AI 반도체의 목표시장 규모를 추정 → 최종적으로 ETRI 전문가 집단 의견을 반영한 26%를 휴머노이드 로봇 시장에서의 비중으로 적용하여 최종 목표시장 규모를 추정(<표 13> 참조)
  - (보수적 3%) `31년 6억 달러 → `35년 18억 달러로 성장 전망(CAGR 33.1%)
  - (낙관적 15%) `31년 29억 달러 → `35년 92억 달러로 성장 전망(CAGR 33.1%)
  - (기술 전문가 26%) `31년 51억 달러 → `35년 159억 달러로 성장 전망(CAGR 33.1%)

<표 13> 휴머노이드 로봇용 지각 특화 AI 반도체 목표시장 규모 추정 (단위: 십억 달러)

구분		`30	`31	`32	`33	`34	`35	CAGR
휴머노이드 로봇용 지각 특화 AI 반도체	보수적(3%)	-	0.6	0.8	1.1	1.4	1.8	33.1%
	낙관적(15%)	-	2.9	3.9	5.3	6.9	9.2	33.1%
	기술전문가(26%)	-	5.1	6.8	9.1	12.0	15.9	33.1%

출처: ETRI 기술전략연구본부 추정, 2025.11.

- 자율주행차용 지각 특화 AI 반도체의 목표시장 규모는 `31년 10억 달러에서 연평균 22.3% 성장하여 `35년 224억 달러로 성장할 것으로 예측(<표 14> 참조)
  - 앞서 도출한 자율주행차 내 지각 특화 AI 반도체의 비중 2%를 기반으로 목표 시장 규모를 추정

<표 14> 자율주행차용 지각 특화 AI 반도체 목표시장 규모 추정 (단위: 십억 달러)

구분	`30	`31	`32	`33	`34	`35	CAGR
자율주행차용 지각 특화 AI 반도체	-	2.0	4.8	8.8	14.4	22.4	22.3%

출처: ETRI 기술전략연구본부 추정, 2025.11.

- XR 기기용 지각 특화 AI 반도체의 목표시장 규모는 `31년 29억 달러에서 `35년 38억 달러로 연평균 7.4% 성장할 것으로 예측(<표 15> 참조)
  - 앞서 도출한 XR 기기 내 지각 특화 AI 반도체의 비중 4%(Apple Vision Pro 내 센서 데이터 전처리 칩의 비중)를 활용하여 목표시장 규모를 추정

<표 15> XR 기기용 지각 특화 AI 반도체 목표시장 규모 추정 (단위: 십억 달러)

구분	`30	`31	`32	`33	`34	`35	CAGR
XR 기기용 지각 특화 AI 반도체	-	2.9	3.1	3.3	3.5	3.8	7.4%

출처: ETRI 기술전략연구본부 추정, 2025.11.

■ 지각 특화 AI 반도체 목표시장 규모

- 지각 특화 AI 반도체의 활용 분야를 합산한 목표시장 규모는 31년 180억 달러에서 35년 420억 달러로 연평균 23.7% 성장할 것으로 전망(<표 16> 참조)
  - 초기 시장 규모는 제한적이나, 중장기적으로 고성장이 예상되는 유망 기술 분야임을 확인
  - `35년 기준 시장 구성은 휴머노이드 로봇 : 자율주행차 : XR 기기 = 4 : 5 : 1의 비중으로 확산될 것으로 예상.
  - 본 시장은 특정 산업에 한정되지 않고, 상호작용 품질이 핵심인 다양한 산업 전반으로 확산될 잠재력을 보유

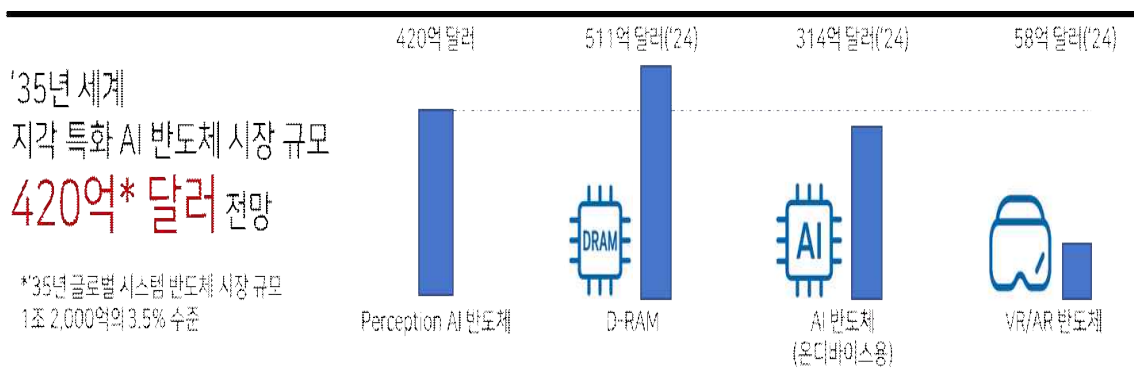
<표 16> 지각 특화 AI 반도체 목표시장 규모 (단위: 십억 달러, %)

구분	`30	`31	`32	`33	`34	`35	CAGR
휴머노이드 로봇	-	5 (27.8%)	7 (31.8%)	9 (33.3%)	12 (35.3%)	16 (38.1%)	33.1%
자율주행차	-	10 (55.5%)	12 (54.6%)	15 (55.5%)	18 (52.9%)	22 (52.4%)	22.3%
XR 기기	-	3 (16.7%)	3 (13.6%)	3 (11.2%)	4 (11.8%)	4 (9.5%)	7.4%
합계	-	18 (100.0%)	22 (100.0%)	27 (100.0%)	34 (100.0%)	42 (100.0%)	23.7%

출처: ETRI 기술전략연구본부 추정, 2025.11.

○ 지각 특화 AI 반도체 시장은 고성장 신흥 반도체 분야로서 새로운 투자처로 주목할 만한 잠재력을 보유([그림 17] 참조)

- 시장 규모는 온디바이스 AI 반도체 대비 1.3배, VR/AR 반도체 대비 7.2배, D-RAM 대비 약 82% 수준으로, 중장기 성장성이 높은 영역으로 확인
- 초기 시장 규모는 제한적이나, 높은 성장률(53.2%)을 바탕으로 거대 시장으로 확대될 가능성이 높아 선제 투자 가치가 큼
- 온디바이스 AI 및 지능형 센서 시장이 빅테크 플랫폼 중심으로 고착화된 것과 달리, 본 시장은 산업 다변성과 신규 진입 여지 측면에서 차별성을 가짐
- 휴머노이드 로봇, 자율주행차, XR 기기 등 주요 활용 분야는 정책 지원·산업 확산·생태계 확대가 동반되는 구조로, 고부가가치 수요 창출이 기대됨



[그림 17] 지각 특화 AI 반도체 시장 추정 규모

## 나. 산업 생태계 현황 분석

### ● 온센서 AI (지능형센서 포함) 생태계 현황 및 분석

#### ■ 온센서 AI 정의

- 센서는 물리적·화학적·생물학적 정보를 감지하여 이를 전기적 신호로 변환하는 장치이며<sup>116)</sup>, 최근에는 센서 내부에 AI 연산 기능을 결합해 데이터 획득과 동시에 추론까지 수행하는 온센서 AI(On-Sensor AI)가 지능형센서의 핵심 트렌드로 부상
- 센서 기술은 MEMS 미세화와 디지털 집적, MCU/SoC 내장 등을 통해 고도화되며, 다음과 같이 4세대로 구분<sup>117)</sup>
  - 1세대(Discrete Sensor) : 센싱 소자와 신호처리 회로가 분리된 구조
  - 2세대(Integrated Sensor) : MEMS 도입으로 소자와 신호처리 회로가 결합
  - 3세대(Digital Sensor) : 디지털 회로 집적으로 비선형 오차 보정 및 네트워킹 가능
  - 4세대(Smart Sensor) : MCU/SoC 내장으로 자동 보정, 자가진단, 제어·판단 기능 수행
- 온센서 AI는 4세대 Smart Sensor를 넘어, 센서 칩 내부에 경량 AI 모델을 탑재해 센싱-전처리-추론을 센서 단에서 완결하는 구조
  - 32-bit MCU 또는 전용 NPU를 내장해 대용량 원시 데이터 전송 없이 처리함으로써 저전력·초저지연 실시간 처리가 가능함
- (구조적 한계) 온센서 AI는 단일 센서(단일 모달) 중심 구조와 MCU 기반 구현의 연산·메모리 제약이 존재함
  - 이종 센서 간 실시간 융합이나 고차원 상황 인식에는 한계가 존재
  - 피지컬 AI가 요구하는 비정형·고병렬 지각 처리(TOPS, 메모리 대역폭)를 MCU 단독으로는 충족하기 어려움

#### ■ 시장 전망 및 경쟁 현황

- 온센서 AI는 센서와 프로세서가 결합된 구조로, 고성능 32-bit MCU 시장과 AI 기능이 탑재되는 주요 센서 시장을 중심으로 형성
  - MCU는 AI/ML(TinyML) 기능을 내장한 고성능 32-bit 중심으로 재편 중
  - 통신, 보안 기능을 통합한 지능형 SoC로 진화해 온센서 AI의 핵심 하드웨어로 자리매김하고 있음

116) Urbina, M., Acosta, T., Lazaro, J., Astarloa, A., & Bidarte, U. (2019). Smart sensor: SoC architecture for the Industrial Internet of Things. IEEE Internet of Things Journal, 6(4), pp. 6567-6577.

117) 권정은, 조유리. (2022). 지능형 센서 플랫폼, KISTEP 브리프 36.









<표 22> 글로벌 차량용 MEMS 센서 시장 점유율 (단위: 백만 달러)

공급업체	`23	`24	`24 점유율
██████████	████	████	████
██████████	████	████	████
██████████	████	████	████
██████████	████	████	████
██	████	████	████
████████████████████	████	████	████
████████████████████	████	████	████
██████████	████	████	████
██████████	████	████	████
████████████████████	████	████	████
██████████	████	████	████

출처: OMDIA, MEMS & Sensor Market Tracker, 2025.

■ 주요 기업 동향 : 원칩(One-Chip) 솔루션 및 생태계 확장

- 주요 기업들은 센서 다이(Die)에 전용 AI 프로세서를 적층 또는 내장하는 원칩(One-Chip) 솔루션을 통해 데이터 전송량을 최소화하고 초저전력, 실시간 처리를 구현하는 데 주력하고 있음
  - 범용 AI가 아닌 비전·음성·3D 심도 등 특정 작업에 특화된 전용 프로세서를 적용해, 옛지 디바이스에 요구되는 mW급 전력 소모와 즉시 반응성을 핵심 경쟁력으로 강조함
    - Ex. BrainChip(Akida), Knowles(음성 전용 DSP), Intel(3D 심도 전용 ASIC)
- 이러한 추세는 전통적인 센서 기업뿐 아니라 MCU 기업에서도 센서-프로세서 융합 형태로 확산되고 있음을 보여줌
  - (센서 주도형 통합) Sony는 이미지 센서 내부에 DSP를 집적한 IMX500으로 AI 센서를 상용화했으며, BrainChip은 뉴로모픽 아키텍처를 적용해 센서 단에서 학습·추론을 직접 수행
  - (MCU 주도형 통합) Bosch는 IMU와 32-bit MCU를 통합한 BHI260AP로 모션 인식·학습을 센서에서 수행하고, STMicroelectronics는 MEMS 다이에 DSP를 집적한 ISPU 기술로 신호처리와 머신러닝 연산을 내재화
- 아울러 기업들은 칩 판매를 넘어 SDK와 클라우드 플랫폼(예: Sony AITRIOS)을 제공해 개방형 생태계를 구축하고, 스마트폰을 넘어 로보틱스, XR, 자동차 등 옛지 AI 응용 분야로 확장을 가속화하고 있음

<표 23> 온센서 AI(지능형센서 포함) 주요 기업별 기술개요, 개발동향 및 시장적용 분야

기업(제품)	기술개요(핵심특성)	개발동향	시장적용(타겟분야)
Sony (IMX500)	이미지센서 내부에 DSP를 직접 집적하여 영상 데이터를 센서 내부에서 AI로 처리	세계 최초 AI센서(`20) 출시, AITRIOS 플랫폼 공개	리테일 분석, 스마트시티 교통 등 엣지 비전 분야
BrainChip (Akida)	뉴로모픽 아키텍처로 센서 단에서 데이터 처리, 학습, 추론을 수행	1세대(`20)→2세대(`23) 성능 향상, Microchip과 데모	항시동작 센서(음성키워드, 이상탐지)에 활용 - mW급 소모
Luxonis (OAK camera)	카메라 모듈 내부에 VPU(Myriad X)를 내장하여 영상 분석을 현장에서 처리; GPU 없이 센서 단에서 AI 수행	OpenCV와 공동개발, 오픈소스 생태계 형성 - 클라우드펀딩 성공	개발자용 AI 비전 키트로 로봇·드론 등 프로토타이핑 활용
Intel (RealSense D4)	스테레오 깊이 계산 전용 ASIC으로 3D 데이터를 카메라 내부에서 직접 생성·처리	`17년 D435 출시, `21년 사업축소로 신규 개발 중단	로봇·AR의 3D센싱 용도로 사용 (자율주행, 제스처인식 등)
Knowles (SmartMic)	MEMS 마이크 내부 DSP에서 음성인식·화자식별을 직접 수행하여 메인 프로세서 개입 없이 동작	`18년 첫 제품 → 스마트폰·이어버드 채용 확대	보이스 UI 디바이스 전반 (스마트폰, 웨어러블, 스마트TV 등)
삼성전자 (Vizion 63D)	iToF 센서 내부에 ISP를 통합하여 심도 연산을 수행; 센서 단에서 3D 맵을 직접 계산	`23년 로보틱스/XR용 3D센서 2종 동시 발표 (ToF+글로벌셔터)	XR기기 공간인식, 로봇 3D비전, 모바일 얼굴인식
Pixelplus (PX2100)	이미지센서+ISP+AI 코어를 적층 통합하여 센서 자체에서 객체 인식 수행	`23년 車용 ASIL-B 인증 달성, 2MP 센서 양산 및 5MP 준비	차량 주변·실내카메라, 보안/스마트홈 카메라
Bosch (BHI260AP)	6축 IMU와 32비트 MCU를 하나의 패키지로 통합, MCU 내부에서 모션 인식·학습 수행; 센서+제어형 구조.	`20년 출시, Self-learning AI 펌웨어로 진화	웨어러블·폰 모션인식, 산업 진동모니터링 센서
ST (ISPU)	MEMS 다이에 DSP 코어를 직접 통합, 기본 신호처리와 간단한 ML 연산을 수행하는 센서 내 MCU형 아키텍처	`22년 세계첫 발표, 가속도 센서에 적용 및 NanoEdge AI 툴 제공	IoT 기기의 임베디드 AI센서 (자가진단, 낙상감지 등)

출처: ETRI 기술전략연구본부 작성, 2025.11.

## ● 온디바이스 AI 생태계 현황 및 분석

### ■ 온디바이스 AI 정의

- 온디바이스 AI는 클라우드나 외부 네트워크에 의존하지 않고 디바이스 내부에서 AI 연산을 직접 수행하는 기술을 의미하며, 본 고에서는 이를 고성능 가속기(NPU)를 내장한 중앙 SoC 기반의 메인 연산 구조로 정의
  - 네트워크를 거치지 않아 충분한 연산 성능이 확보될 경우 빠른 응답성을 제공<sup>121)</sup>
- 구현의 핵심은 SoC 내 AI 가속기(NPU)와 이를 구동하는 SW 스택(런타임, 프레임워크, 모델 최적화)이며, 스마트폰 AP를 중심으로 NPU 탑재가 확산되며 기기 연산 능력이 상향 표준화<sup>122)</sup>
- 온디바이스 AI의 확산은 프라이버시 강화 요구, 실시간 반응성, 비용·네트워크 의존성 축소가 맞물린 결과임
  - 민감 데이터의 외부 전송을 회피할 수 있어 규제 대응에 유리<sup>123)</sup>
  - 고성능·저전력 NPU 발전으로 실시간 추론이 가능<sup>124)</sup>
  - 오프라인 환경에서도 동작해 통신·클라우드 비용 절감 효과<sup>125)</sup>
- (구조적 한계) 중앙 SoC에 연산이 집중되는 구조로 인해 피지컬 AI 환경에서는 센서 증가에 따른 대역폭 병목, 원시 데이터 집중으로 전력 및 발열 증가, 센서→SoC 데이터 이동에 따른 지연 누적, 배터리 기반 로봇·모빌리티에서의 에너지 비효율이 발생할 수 있음

### ■ 시장 전망 및 경쟁 현황

- 글로벌 온디바이스 AI 시장 규모는 `24년 86억 달러에서 `30년에는 366억 달러로 연평균 27.8% 성장이 전망<sup>126)</sup>(<표 24> 참조)
  - 스마트폰·태블릿 등 모바일 기기 중심의 하드웨어 수요가 성장을 견인하며, `24년 기준 하드웨어 비중은 약 60%임(<표 25> 참조)

121) 한국저작권위원회. (2024.4.). 온디바이스 AI(On-Device AI) 산업현황 보고서

122) Edge AI and Vision Alliance. (2024.3.)

123) 삼성 SDS. (2025.4.)

124) Edge AI and Vision Alliance. (2024.3.)

125) Poddar. (2025.6.)

126) GRAND VIEW RESEARCH, On-device AI Market (2025-2030)

<표 24> 글로벌 온디바이스 AI 시장 규모 (단위: 십억 달러)

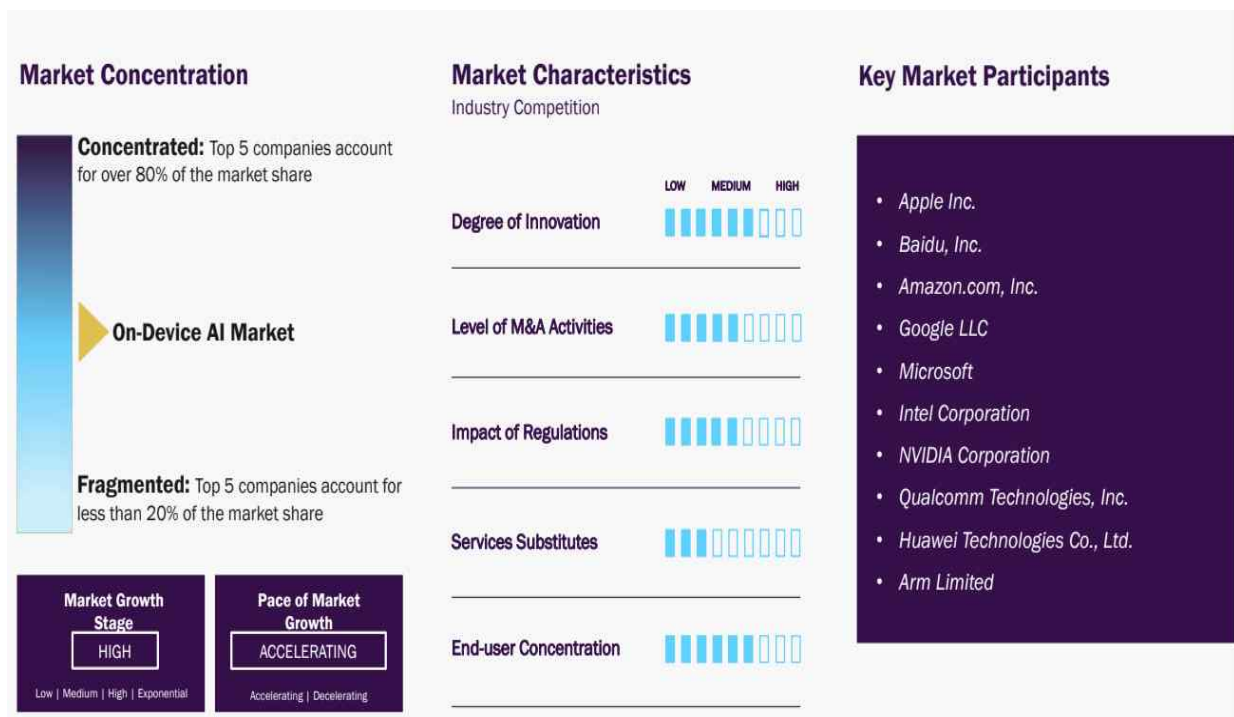
구분	`25	`26	`27	`28	`29	`30	`31	`32	`33	`34	`35	CAGR
하드웨어	6.1	7.6	9.5	12.1	15.3	19.7	25.0	31.6	40.0	50.6	64.0	26.5%
소프트웨어	4.6	5.9	7.6	9.8	12.8	16.8	21.8	28.2	36.5	47.1	61.0	29.3%
합계	10.7	13.5	17.1	21.9	28.1	36.5	46.8	59.8	76.5	97.7	125	27.8%

출처: GRAND VIEW RESEARCH, On-Device AI MARKET ESTIMATES AND TREND ANALYSIS, 2025의 `25-`30년 추정 데이터를 기반으로 `31-`35년 ETRI 기술전략연구본부 추정, 2025.11.

<표 25> 글로벌 온디바이스 AI 배포 방식별 시장 규모 (단위: 십억 달러)

구분	`25	`26	`27	`28	`29	`30	`31	`32	`33	`34	`35	CAGR
클라우드	5.7	7.1	8.9	11.3	14.4	18.5	23.4	29.6	37.4	47.3	59.8	26.4%
자체 서버 (On-premises)	5.0	6.4	8.2	10.6	13.8	18.1	23.3	30.2	39.0	50.4	65.1	29.2%
합계	10.7	13.5	17.1	21.9	28.2	36.6	46.7	59.8	76.4	97.7	124.9	27.8%

출처: GRAND VIEW RESEARCH, On-Device AI MARKET ESTIMATES AND TREND ANALYSIS, 2025의 `25-`30년 추정 데이터를 기반으로 `31-`35년 ETRI 기술전략연구본부 추정, 2025.11.



[그림 18] 글로벌 온디바이스 AI 시장 주요 플레이어 (GRAND VIEW RESEARCH, 2025)

- 시장은 Apple, Google, NVIDIA 등 빅테크 중심의 고집중 구조(상위 5개사 80% 이상)로 형성<sup>127)</sup>([그림 18] 참조)
  - 시장 지배력은 칩셋-OS-AI 프레임워크-디바이스 생태계의 수직통합 수준에 의해 결정되어 진입 장벽이 높음
  - 신흥 기업은 특정 버티컬, 개방형 플랫폼을 중심으로 성장하나 영향력은 제한적임

<표 26> 온디바이스 AI 반도체 주요 기업별 기술개요, 개발동향 및 시장적용 분야

기업(제품)	기술개요(핵심특성)	개발동향	시장적용(타겟분야)
Apple (R1)	12대 카메라와 센서 입력을 통합·전처리하는 전용 프로세서로, 센서 내부가 아닌 기기 레벨에서 AI 연산 수행	MR 생태계 위한 visionOS·RealityKit 지원	Vision Pro 헤드셋의 실시간 공간컴퓨팅
NVIDIA (DRIVE Thor)	차량 내 여러 센서 데이터를 중앙 SoC에서 병합·추론하는 고성능 AI 프로세서로 센서 외부에서 수행하는 엣지 AI	Orin후속 Thor로 전환, 성능 2배↑·FP8도입	자율주행차 ECU 통합 - 2025년 ZEEKR 적용 예정
DEEPIX (DX Chip)	엣지 NPU SoC로 센서 외부에서 다중 AI 모델을 실시간 병렬 추론; 배터리 구동 기기용 고효율 AI 가속기	`21 칩 발표, 현대차 등 투자 유치 → 차량용 확장	IoT카메라, 센서모듈 등 배터리 구동기기 (PoC 단계)
한화비전 (Wisenet 9)	듀얼 NPU 아키텍처로 영상 분석을 카메라 SoC 수준에서 수행; 센서 내부가 아닌 디바이스 단위의 연산	`21 출시, CCTV 9세대 탑재 - 국산 AI카메라 확립	지능형 CCTV (산업안전, 교통, 보안 모니터링)
NXP (Radar SoC)	77GHz 트랜시버와 프로세서를 통합한 자동차용 레이더 SoC로 레이더 신호를 센서 외부에서 추론	`24 CES Safeguard 86xx 발표 - 성능↑/OTA지원	ADAS 레이더(전방·측후방), 자율주행 센서네트워크
넥스트칩 (Apache5)	CPU+NPU 통합 ADAS SoC로, 카메라 입력을 처리하나 센서 내부가 아닌 차량용 시스템 수준에서 수행	`21 시제품 완료→`24 양산, 차기작 Apache6 L3 자율주행용 개발	자동차 카메라 (전방, 서라운드, DMS 등)

출처: ETRI 기술전략연구본부 작성, 2025.11.

■ 주요 기업 동향 : 전용 AI 하드웨어 기반 원칩(SoC) 고도화 및 생태계 확장

- 메인 SoC에 전용 AI 하드웨어(NPU/커스텀 SoC/보조 프로세서)를 내장·통합해, 클라우드 없이 초저지연·초고성능·초저전력 추론을 구현하는 원칩 중심 전략을 강화 중(<표 26> 참조)

127) GRAND VIEW RESEARCH, On-device AI Market (2025-2030)

- 범용 성능 경쟁보다 응용처별 KPI 최적화가 핵심 전략으로 전환됨

- Ex. Apple(보조 프로세서 기반 초저지연), NVIDIA(자율주행용 초고성능), DEEPX(엣지 초저전력), NXP/넥스트칩(전장), 한화비전(보안)

○ 그리고 하드웨어 고도화에 더해 전용 OS·SDK·개발도구·플랫폼을 결합한 통합 솔루션을 제공하며, 개발자·고객 락인(Lock-in)을 통한 생태계 주도권 확보 전략을 강화 중임

- 스마트폰을 넘어 자율주행, XR/공간컴퓨팅, 지능형 보안, IoT 엣지 등 고신뢰, 고부가 엣지 시장으로 적용 범위를 확대하고 있음

### ● 산업 생태계 분석(기술·시장·경쟁 상황 고찰)을 통한 신시장 창출 기회 탐색

#### ■ 전략적 공백(White Space) 공략 : 과점화된 양 끝단 사이의 기회

○ (지배구조) 온센서 AI와 온디바이스 AI는 이미 고도로 과점화된 시장으로, 기존 강자들이 독점적 하드웨어 표준과 특허 장벽을 통해 폐쇄적 생태계를 공고화하고 있어 신규 진입자의 기술 주도권 확보가 제한적임

- 온센서 AI 시장 : Sony, Infineon, NXP 등 상위 5개사 77% 점유

- 온디바이스 AI 시장 : NVIDIA, Apple 등 상위 5개사 80% 이상 점유

○ (공백 영역) 두 거대 생태계 사이에서 대용량 멀티모달 데이터의 융합·전처리를 담당하는 중간 계층은 지배적 강자가 부재한 전략적 공백 영역임

○ (기회) 기존 강자들 사이의 아키텍처 공백을 선점함으로써, 특정 하드웨어에 종속되지 않는 개방형 생태계 구축이 가능

#### ■ 경쟁우위 확보 가능성 : 설계 효율 중심의 경쟁

○ (핵심 경쟁력) 엣지 지각 시장의 성패는 미세 공정 경쟁보다 이종 멀티모달 데이터의 실시간 정합 및 특징 추출을 위한 전용 아키텍처 설계 역량에 의해 좌우될 것으로 판단

○ (국내 기업의 기회) 데이터센터용 AI 대비 자본집약도가 낮고 설계 효율성이 핵심이어서, 국내 팹리스 기업이 기술 차별화를 통해 글로벌 점유율을 확보하기에 유망한 분야임

○ (포지셔닝) 기존 강자의 점유율을 직접 대체하기보다, 시스템 전체 효율을 높이는 기능적 보완재(Functional Complement)이자 파트너로 포지셔닝할 필요가 있음

○ (진입 시점) 상용화 소요기간을 고려할 때, 시장 개화기인 `31년을 목표로 현시점부터 조속한 R&D 착수가 필요

## N 신시장 창출과 성장 견인을 위한 정부 역할

### 1. 국가전략기술로서 정부의 투자 필요성

- 지각 특화 AI 반도체는 피지컬 AI 시대의 핵심 기반기술로, 국가 경쟁력 확보를 위해 정부의 선제적·전략적 투자가 요구됨
- 특히 반도체-로봇-자동차로 이어지는 국가 주력 산업 생태계의 지속성·경쟁력·기술 자립성을 확보하기 위해, 국가 차원의 마중물 투자가 필요한 시점임

#### ■ 시장 매력도 측면

- 피지컬 AI 시장의 급속한 확산
  - 로봇, 자율주행, XR 분야를 중심으로 실시간 센싱 및 AI 처리 수요가 급증
  - 피지컬 AI는 보고, 판단, 행동하는 모든 기계에 적용 가능한 범용 기술로, 산업 전반에 파급력이 매우 큰 초대형 시장으로 전망<sup>128)129)</sup>
- 지각 특화 AI 반도체의 높은 성장 잠재력
  - 휴머노이드 로봇, 자율주행차, XR 등 주요 분야에서 약 20% 수준의 반도체 수요 비중을 차지할 것으로 예상
  - `35년 목표시장 규모는 약 420억 달러로, 시스템반도체 시장의 약 3.5%에 해당하며 타 반도체 분야 대비 성장 잠재력이 높은 것으로 분석됨

#### ■ 기술 패권의 국가 경쟁력 측면

- 피지컬 AI 시대의 경쟁구도 변화
  - 데이터 : 텍스트, 이미지 중심 → 동작, 물리 변수, 시간 변동성이 큰 현실 세계 데이터로 확장
  - 처리 방식 : 오프라인 분석 → 실시간 판단, 추론, 적응
  - 기술 요구 : 추론 중심 → 지각-행동-학습 통합 처리
  - 이는 전용 반도체, 로봇 제어 SW, 시뮬레이터, 센서 등 복수 기술의 동시 경쟁을 의미

128) Peter Repetto. (2025.8.). "iCapital Market Pulse: AI Gets a Body - The Coming Rise of Humanoids", iCapital  
 129) 최홍섭. (2025.9.). "Physical AI 시대, 온디바이스 AI 기술과 전략", 2025 e4ds Tech Day

○ 미국과 중국의 기술 패권 경쟁 심화(<표 27><sup>130</sup>) 참고)

- 미국 : AI 모델, SW, HW, 로봇 기업 중심으로 피지컬 AI 풀스택 투자 강화
- 중국 : 정부 주도의 대규모 투자로 로봇, AI 칩, VLA 연구 집중 육성
- 양국은 피지컬 AI에서 ‘AI 모델-칩-로봇 플랫폼’을 통합적으로 육성하며 미래 주도권 경쟁 중

<표 27> 미국과 중국의 피지컬 AI 글로벌 패권 경쟁 현황

구분	미국	중국
정부 정책	·AI Action Plan(`25.7)	·피지컬 AI 정부 주도 : 대규모 투자 및 조달
	·AI 풀스택 수출 : 모델 SW, HW, 응용, 표준 ·피지컬 AI 전략적 투자 : 제조, 물류, 국방 등	·차세대 파괴적 산업으로 육성(`27년 세계최고 목표) ·북경, 상해, 항주, 선전 로봇 스타트업 투자
피지컬 AI·SW	·글로벌 파운데이션 모델	·자체 파운데이션 모델 확산
	·구글 RT-2, 메타(축가), 아마존(DeepFleet) ·스타트업/오픈소스 : 피지컬 인텔리전스 파이제로/OpenVLA	·자체 모델 : Unitree(Unifo-LM), AgiBot(Genie Operator-1) ·DeepSeek(R1), Alibaba(Qwen) ; AI 모델 제공 협력 Tencent(Robotics X’s VLA), ByteDance(GR-3 VLA)
컴퓨팅 플랫폼	·저전력, 고성능 개발 중	자체 개발 중
	·풀스택 피지컬 AI : AI 가속기 기반 모델, 시뮬레이터 제공 ·NVIDIA Jetson Orin/Thor (~1,000 TOPS / (40~130W)) ·Qualcomm DragonWing (~40 TOPS / BW)	·Huawei : 피지컬 AI 풀스택 추진(AI, SW, HW) ·UBTech(`25.5), 16개 로봇사(`24.11)와 AI, 칩셋 협력 ·Horizon Robotics (~560TOPS / (40~60)BW)
휴머노이드 로봇	·제조공장 투입	·대량 생산 예정(`25년말, 10,000 대 이상)
	·Tesla Optimus : AI, SW, HW, 로봇 풀스택 ·스타트업 Figure AI, Apptронik Agility Robotics	·상하이 WAIC(`25.7), 휴머노이드 로봇 60대 이상출품 ·스타트업 : UBTech, Unitree, AgiBot, Fourier 등 ·휴머노이드 로봇 가격 인하 주도 : Unitree R1 @\$5,900

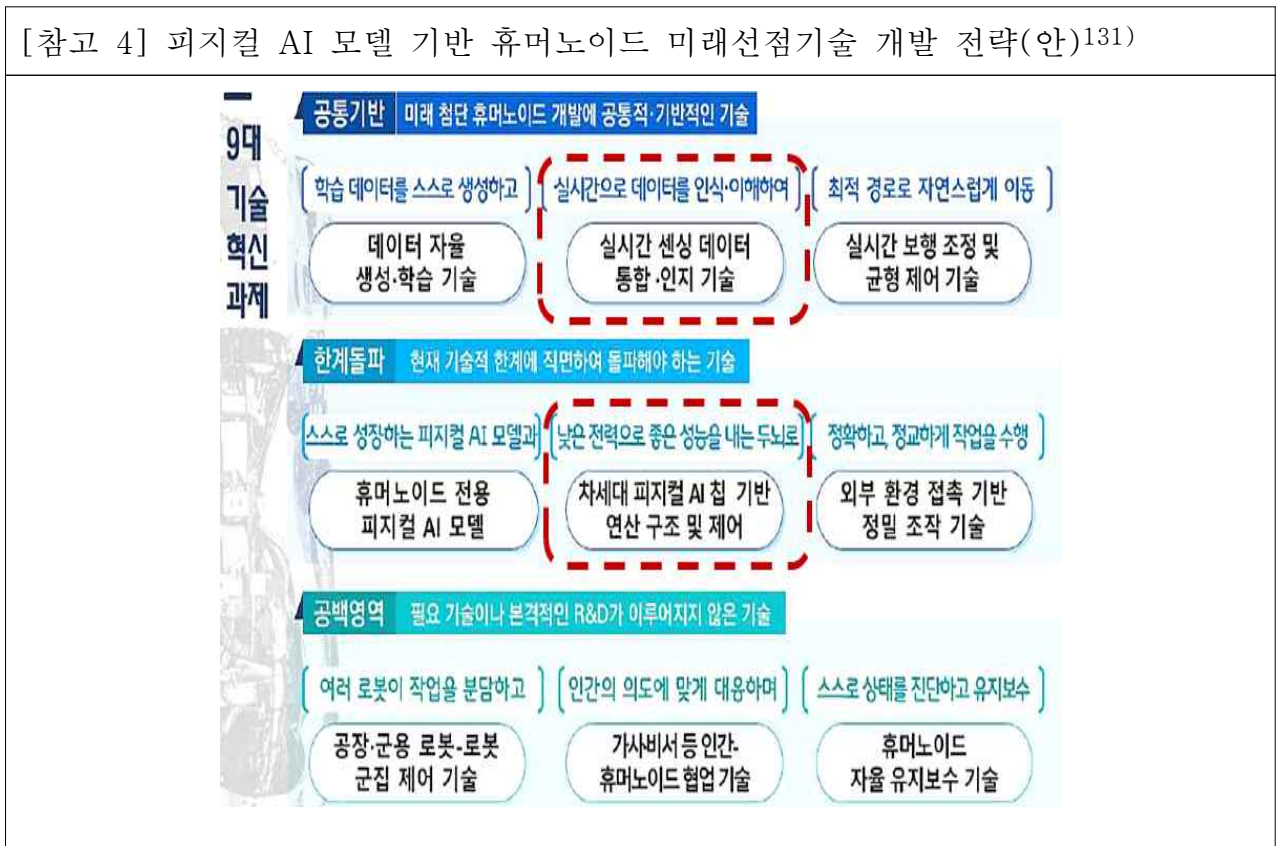
출처: 김욱, 2025.11.

130) 김욱, (2025.10.). 피지컬 AI 혁신, 고려대 미래성장연구원 AX 전략 포럼, IITP Tech&Future Insight Forum, 고려대 MOT 공동 주최 세미나

○ 주력산업 경쟁력 및 공급망 리스크 대응

- 로봇, 자동차 등 고센싱 산업에서 데이터 폭증으로 인해 AI 가속기(GPU) 수요 급증
- 기존 GPU 공급망에 대한 높은 의존도는 기술 디커플링 및 공급망 단절 위험으로 연결될 가능성이 존재
- 이에 대응하기 위해 범용성과 센서 및 AI 모델 적응력을 갖춘 지각 특화 AI 반도체 개발이 핵심 전략으로 부상

[참고 4] 퍼지컬 AI 모델 기반 휴머노이드 미래선점기술 개발 전략(안)<sup>131)</sup>



■ 시스템반도체 분야 신성장동력 발굴 측면

○ 국내 시스템반도체 경쟁력 제고의 필요성

- 글로벌 시스템반도체 시장은 메모리 대비 2배 이상 규모이나, 국내 점유율은 약 3% 수준으로 낮음<sup>132)</sup>
- 최근 20여 년간(`04~`25) 시장자료 분석 결과(<표 28>참고), 로직 IC, 센서, 광학 반도체가 성장률, 변동성, 시장성 측면에서 가장 매력적 영역임

131) 과학기술정보통신부. (2025.5.)

132) 경희권, 김상훈. (2025). “세계 비메모리반도체 시장 지형과 정책 시사점,” KIET 산업경제, pp. 22-23.

○ 지각 특화 AI 반도체의 전략적 의의

- 지각 특화 AI 반도체는 로직 IC와 센서, 광학 반도체의 특성을 결합한 분야로 시스템반도체 산업의 타겟팅 고도화와 글로벌 확장 기회를 제공

○ 정부의 마중물 역할 강화 필요

- 지각 특화 AI 반도체를 중심으로 시스템반도체 시장 진입을 가속화하면 국가 차원의 글로벌 점유율 확대 및 반도체 산업 전반의 경쟁력 강화가 가능

<표 28> 온디바이스 AI 반도체 주요 기업별 기술개요, 개발동향 및 시장적용 분야

차별요소	로직 IC	마이크로 IC	아날로그 반도체	광학 반도체	센서 반도체
시장규모	가장 큼 	큼 	큼 	작음 	작음 
성장률	고성장 	저성장 	저성장 	고성장 	고성장 
변동성	낮음 	낮음 	낮음 	낮음 	낮음 
음(-) 성장확률	낮음 	높음 	높음 	-	-
종합검토					

출처: 이주완, 반도체 패권전쟁, 에텐하우스, 2025, pp.141-151.

## 2. 신시장 창출과 성장 견인을 위한 정부의 투자전략 방향성

### 가. 기술 개발 전략 방향성

- 지각 특화 AI 반도체 구현에는 **복합 기술의 집약**이 필수적이며, 생성 데이터 처리, 실시간 AI 추론, 초저전력 구동을 위한 소자·회로·설계 기술과 시스템 소프트웨어까지 **전주기 기술의 통합적 확보**가 요구됨
  - 단기 실증을 넘어, 다양한 피지컬 AI 응용으로 확장 가능한 **중장기 기술 축적 전략**을 병행할 필요가 있음
  - 이를 위해 핵심 기술을 선별하고, 중점 개발 항목을 중심으로 **단계적·체계적 R&D 전략**을 수립해야 함

### ■ 설계·공정 기술의 전략적 중요성

- 지각 특화 AI 반도체는 범용성, 호환성, 대량 생산성 확보가 필수적이므로, **설계(Design Technology)와 공정(Process Technology)이 가장 중요한 경쟁 요소**
- **설계 기술 개발 요건**
  - 생성 데이터 특성과 목표 응용의 요구 성능을 반영한 **모듈화·계층화 설계**가 가능해야 함
  - 에너지 효율과 발열 억제는 공통 핵심 제약으로, 설계 단계에서 선제적으로 반영되어야 함
    - 자율주행차 : 장시간 안정 동작
    - 휴머노이드 로봇 : 배터리 소모 최소화
    - XR 기기 : 발열 억제 및 착용 안전성
  - 데이터 이동 최소화, 전력 효율 극대화, **메모리-연산 구조 최적화**는 설계 단계의 필수 목표임
- **공정 기술 개발 요건**
  - 공정 최적화를 통한 **수율 제고와 양산성 확보**로 비용 경쟁력을 확보해야 함
  - 센서 수, 종류, 계층 구조의 다양성을 유연하게 수용할 수 있도록, **고성능, 저전력, 저지연 대응 공정 플랫폼** 마련이 필요함
- **범용성 및 호환성 확보**
  - 지각 특화 AI 반도체는 제품 성능 수준에 따라 **계층별로 유연하게 탑재** 가능해야 함
  - 이종시스템 간 상호운용성을 보장하는 **플랫폼 중심의 범용·호환 구조**를 지향해야 함

[참고 5] 관련 유사 기술간 주요 성능 비교<sup>133)</sup>

구분	온-센서 AI (지능형센서)	온-디바이스 AI (메인 SoC)	로봇 제어 MCU	Perception AI 반도체
주 역할	센서 1개 기준 전처리·간단 인식	고수준 앱/맵/경로계획/대형 모델	하드 실시간 제어 (모터·안전)	다센서 동기·퓨전·지각·시맨틱 요약
데이터 범위	로컬 원시/피처 (단일 센서)	시스템 전체 상태·지도·학습 로그	센서/엔코더·상태 값	멀티모달 (카메라·라이다·레이다·IMU 등) 통합
지연/결정성	매우 낮음 (센서 주기)	변동 있음 (멀티태스킹/대용량)	매우 낮음 (마이크로초~ms, 하드 RT)	낮음 (지각 파이프라인 10~수십 ms급)
대역 정책	로컬 축약 (ROI/이벤트)	대용량 메모리/버스 중심	제어 버스 저대역	시맨틱 통신 (객체/이벤트/맵+신뢰도/출처)
장점	전력·원가 유리, 센서 바로 처리	대규모 연산·메모리, 유연한 SW	결정성·안전성 최고	데이터 이동·대역·지연 동시 절감, 글로벌 컨텍스트로 정확도↑
한계	센서 간 글로벌 맥락 부족, 확장성 제한	데이터 이동 병목·전력/열 부담	ML 연산 한계, 지각엔 부적합	메인 SoC·MCU와 연계 설계 필요
대표 KPI	mW급/센서, 단일탐지 지연	앱 성능, 대모델 처리량	제어 주기/안전 무결성	센싱→지각 <~50ms, ≥10x 대역 절감
주 사용자	스마트 카메라/레이더 모듈	IVI/도메인 컨트롤러/폰·HMD	모터·브레이크·팔 제어	ADAS/AMR/드론/XR 등 지각 허브

## 나. 시장 진입 전략 방향성

### ■ 유사 시장의 기술 및 수요 추세

- 온센서 AI, 온디바이스 AI, 제어 MCU 등 기존 시장에서도 센싱 및 추론 기능 강화 수요가 증가하고 있으나, 고성능 지각 처리를 충족하지 못하는 구조적 한계가 존재
- 온센서(On-Sensor) AI 시장
  - 센서 내 AI 기능이 확대되고 있으나, 단일 센서·용도 특화 구조에 머물러 이중 센서 융합과 복합 상황 인식에 한계가 있음
  - 글로벌 소수 기업 중심의 과점 구조로, 국내 생태계는 초기 단계임
- 온디바이스 AI 시장(메인 SoC)
  - 엣지 중심 확산과 경량 추론 최적화가 진행 중이나, 발열, 전력 소모, 고가 SoC 의존성이라는 구조적 제약이 지속됨
  - 중앙 집중 구조로 인해 센서 폭증 시 대역폭 병목과 처리 지연 문제가 발생

133) ETRI 내 관련 기술부서 전문가들에게 기술 자문 의뢰한 내용 중 일부 내용 발췌

○ **로봇 제어 MCU 시장**

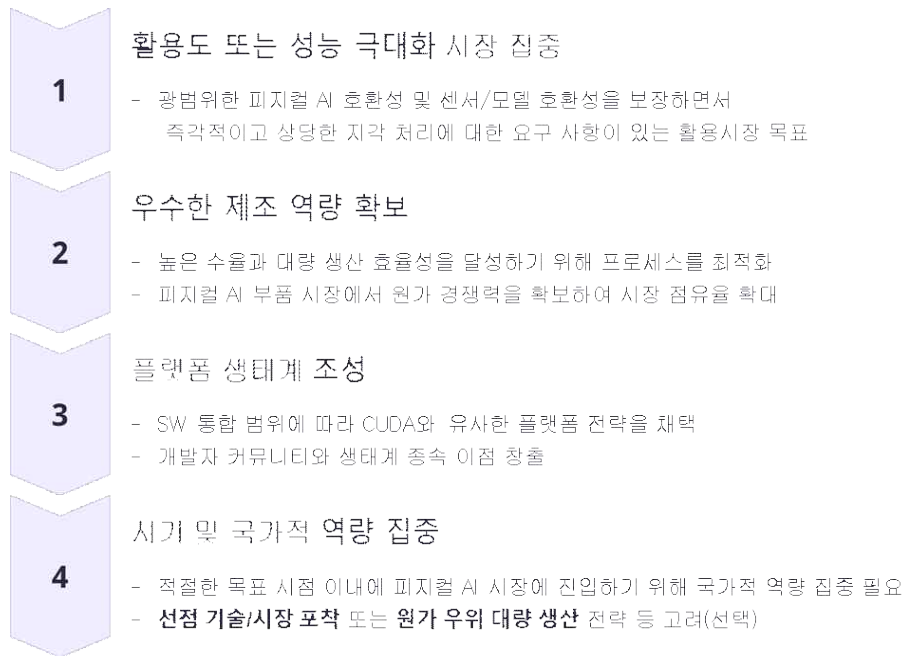
- 기존에는 액추에이터 제어 중심의 정형화된 MCU 시장이 형성
- 결정론적 제어에 최적화되어 있어, **고난도 딥러닝 추론을 위한 병렬 연산 및 메모리 대역폭이 부족함**
- AI MCU에 대한 초기 시도는 있으나, 멀티센서 및 비정형 환경을 지원하는 생태계는 미성숙함

■ **시장 포지셔닝 및 차별화 기회**

- 세 시장 모두 지각 특화 AI 반도체의 필요성은 인식하고 있으나, 핵심 영역 밖에서 제한적으로 접근하고 있어 기능적·성능적 공백(Market Gap)이 존재
- 이 공백은 지각 특화 AI 반도체가 **신규 시장을 창출하고 차별화할 수 있는 전략적 기회**로 작용함
  - (온센서 AI 및 MCU와의 차별점) 다중 센서 데이터를 통합해 **고성능 융합 추론 후 정제된 결과만 전달**
  - (온디바이스 AI와의 차별점) SoC 전단에서 원시 데이터를 해석해 **메인 시스템 부하와 지연을 획기적으로 감소**

■ **단계별 시장 진입 전략**

- **시장 파급력이 높은 대규모 응용부터 우선 진입**
  - 휴머노이드 로봇, 완전자율주행차 등 **시장 규모가 크고 파급효과가 큰 분야**를 우선 공략하여 다양한 기술 성능을 실증하여 기술 신뢰성 확보
- **기술 적용 효과가 큰 고민감도 시장부터 단기 진입**
  - 드론, XR 기기, 국방 로봇 등 **지연, 전력, 중량 제약이 극심한 시장**은 지각 특화 AI 반도체의 성능상 이점이 극대화될 수 있으므로 **단기 진입 시장**으로 유리
- **범용성·확장성 기반의 중장기 시장 확대 전략**
  - 다양한 피지컬 AI 제품군에서의 성공적 실증 사례를 축적하고, **범용성·확장성 중심의 생태계 확산 전략**이 필요함
- **정책적 전략 선택**
  - 피지컬 AI 시장 초기 진입 시 국가 차원에서 다음 전략의 조합 검토가 필요함
    - **선도 전략(First Mover)** : 고난도 기술 조기 확보를 통한 글로벌 시장 선점
    - **효율성 전략(Fast Follower)** : 대량 양산 기반 비용·신뢰성 우위 확보
  - 국가 투자 여건과 산업 생태계 역량을 고려한 **최적의 전략 조합**을 선택해야 함



[그림 19] 성공적인 시장 진입 전략에 필요한 요소

## 다. 정부 투자를 위한 추후 과제

### ■ 피지컬 AI 시대 대비 국가 전략기술 제안의 의의

- 본 연구는 피지컬 AI가 요구하는 물리환경 기반 센싱 데이터의 특성과 처리 요구를 고려하여, 센서-알고리즘-엑추에이터 간 기술 조합 방향성을 제시했다는 점에 의의가 있음
- 또한 차세대 AI로 부상하는 피지컬 AI 분야에서 기술 주도권 확보 가능성과 시스템반도체 산업의 신성장 동력으로서 지각 특화 AI 반도체의 시장 매력도와 전략적 필요성을 검증

### ■ 국가 차원의 전략적 투자 방향 및 역량 결집 필요성

- 시장 과급력 확대, 기술 패권 경쟁 심화, 시스템반도체 신성장동력 확보를 고려할 때, 지각 특화 AI 반도체에 대한 국가 차원의 전략적 투자가 필요
- 기술 개발 측면에서는 피지컬 AI 시스템의 요구 성능과 다양한 폼팩터를 지원하는 설계 기술, 확장·업데이트가 용이한 아키텍처, 최적화된 공정 기술이 핵심 요소임
- 시장 진입 전략 측면에서는 선도형 시장 창출 전략과 대량 양산 기반 비용 경쟁력 전략 중 국가적 방향성을 명확히 설정하고, 도메인 특화 시장에서의 적용 사례 축적이 중요함
- 기술 패권 대응 측면에서는 주력 산업의 초격차 유지, 부품 공급망·밸류체인에서의 기술 종속성 완화, 제조 역량 강화가 필수 과제로 제시

### ■ 정부 기술 개발 계획과 연계한 후속 기획 필요

- 지각 특화 AI 반도체의 구현 범위, 성능 지표, 추진 체계는 향후 정부 기술 개발 계획과 연계하여 상세한 후속 기획이 필요함
- 휴머노이드 로봇, 온디바이스 AI, 피지컬 AI 등 정부의 추진 중 또는 추진 예정인 기술 개발 정책과 연계하여, 지각 특화 AI 반도체 기술 개발과 생태계 조성 정책이 구체화되고 활성화될 필요가 있음



 **참고문헌**

**◆ 국내 문헌**

과학기술정보통신부. (2023). 제4차 융합연구개발 활성화 기본계획(‘23~’27)  
 관계부처합동. (2024). K-로봇경제 실현을 위한 제4차 지능형로봇 기본계획(2024~2028)  
 김선형. (2018). 지능형 로봇의 현재와 미래. 한국정보기술학회지, 16(1), pp. 7-12  
 김은영. (2024). 빠르게 진화하는 중국 휴머노이드 로봇산업. 중국전문가포럼, 대외경제정책연구원  
 신동윤. (2024). 세계 최초 인간을 닮은 범용로봇 개발한 생추어리 AI. MIT Technology Review  
 엄위섭 외. (2013). 지능형 로봇의 발전 동향. 항공우주산업기술동향, 11(1), pp. 150-160  
 로봇신문. (2025). ‘피지컬 AI 글로벌 얼라이언스’ 공식 출범  
 양성욱. (2025). 피지컬 AI 개발 전략. KIST. 국가 공공분야 소버린AI 토론회  
 이해수 외. (2025). 피지컬 AI의 현황과 시사점. 소프트웨어정책연구소  
 전자신문. (2025). 한국 피지컬 AI 협회 출범...마음 AI, 초대 협회장사 맡아  
 조선비즈. (2025). ‘피지컬 AI’ 주도권 잡는다... 산·학·연 뭉친 글로벌 얼라이언스 출범  
 KBS 뉴스. (2025). 돌아온 ‘AI칩 황제’...“로봇의 챗GPT 순간’이 오고 있다”  
 권정은, 조유리. (2022), 지능형 센서 플랫폼, KISTEP 브리프 36  
 연구개발특구진흥재단. (2021), 유망시장 Issue Reports, 스마트센서  
 한국수출입은행. (2024). AI반도체 시장 현황 및 전망  
 한국저작권위원회. (2024.4.). 온디바이스 AI(On-Device AI) 산업현황 보고서  
 e4ds. (2020.10.). 센서의 미래, 지능형 센서 “스마트 IT 융합 서비스의 핵심 부품”  
 IRS Global. (2024.12.). 차세대 자동차 Global Market Data: Current Status & Outlook(2025)

**◆ 국외 문헌**

A. Kalouguine, V. De-León-Gómez, C. Chevallereau, Sébastien Dalibard, & Y. Aoustin. (2020). Definition of a Walking with Starting and Stopping Motions for the Humanoid Romeo. In International Conference on Informatics in Control, Automation and Robotics  
 AI Matters. (2025). What’s Next After ChatGPT? Tesla and Google’s Physical AI War  
 Citi Group. (2024). THE RISE OF AI ROBOTS : Physical AI is Coming for You  
 Global X. (2025). CES 2025: Physical AI Is Here

Albers, Albert & Brudniok, Sven & Otnad, Jens & Sauter, Christian & Sedchaicharn, Korkiat. (2007). Design of Modules and Components for Humanoid Robots, 10, 5772/4857

Andtfolk, M., Nyholm, L., Eide, H., & Fagerström, L. (2021). Humanoid robots in the care of older persons: A scoping review. *Assistive Technology*, 34, pp. 518–526

Lugosi, G., and Mendelson, S. (2021). “Robust multivariate mean estimation: The optimality of trimmed mean,” *Ann. Statist.*, vol. 49, no. 1, 2021, pp. 393–410.

Goldman Sachs.(2024.1.). Global Automation Humanoid Robot: The AI accelerant

GRAND VIEW RESEARCH, On-device AI Market (2025–2030)

IEEE Spectrum. (2024.10.). Brain-like Computers Tackle the Extreme Edge Start-up BrainChip announces a new chip design for a milliwatt-level AI inferenceDina Genkina01 Oct 2024 3 min read

OMDIA. (2025). MEMS & Sensors Market Tracker

OMDIA. (2025). Microcontroller (MCU) Market Tracker

Precedence Research. (2025.5.). Smart Sensors Market Size, Share, and Trends 2025 to 2034

SKYQUEST. (2024.12.). Microcontroller Units (MCUs) Market Size, Share, and Growth Analysis

SKYQUEST. (2025.9.). GLOBAL MICROCONTROLLER UNITS(MCUS) MARKET

TrendForce. (2025.7.). Humanoid Robot Quarterly Report

Urbina, M., Acosta, T., Lazaro, J., Astarloa, A., & Bidarte, U. (2019). Smart sensor: SoC architecture for the Industrial Internet of Things. *IEEE Internet of Things Journal*, 6(4), pp. 6567–6577

Wellsenn XR. (2023). Apple MR disassembly and BoM list

Sony. (2021). Sustainability Report 2021

Statista. (2024.4.). AR & VR – worldwide

Statistics Market Research Consulting. (2024.10.). Automotive Sensor Fusion Market Forecast to 2030

## ◆ 웹사이트

공공 데이터 자원 개발 및 활용 가속화에 관한 의견(关于加快公共数据资源开发利用的意见) (2024.10.9., [https://www.gov.cn/zhengce/202410/content\\_6978910.htm](https://www.gov.cn/zhengce/202410/content_6978910.htm), 최종접속일: 2025.2.17)

글로벌이코노믹. (2025.2). 중국, AI 발전 위해 '공공 데이터' 개방...'AI 플러스' 전략 시동 ([https://www.g-enews.com/article/Global-Biz/2025/01/2025011620281793890c8c1c064d\\_1](https://www.g-enews.com/article/Global-Biz/2025/01/2025011620281793890c8c1c064d_1), 최종접속일: 2025.2.17.)

엔비디아 GTC 웹페이지. Physical AI and Robotics Conference Sessions

넥스트칩 웹사이트. (<https://www.nextchip.com/en/adas/adas.php?idx=5#:~:text=APACHE5%20%28NVS2900%29%20is%20the%20vision,autonomous%20automotive%20applications%20such%20as>)

삼성전자. (2023). 삼성전자, '미래 첨단산업의 눈' 아이소셀 비전 2종 공개. (<https://semiconductor.samsung.com/kr/news-events/news/samsung-unveils-two-new-isocell-vision-sensors-tailored-for-robotics-and-xr-applications/#:~:text=%EB%B6%84%EC%95%BC%EC%97%90%20%EC%B5%9C%EC%A0%81%ED%99%94%EB%90%90%EB%8B%A4,Image%20Signal>)

삼성 SDS(2025.4.). 클라우드 AI vs 온디바이스 AI: 공존과 진화의 방향. (<https://www.samsungsds.com/kr/insights/ondevice-ai-and-cloud-ai.html>)

서울경제. (2023.11.). "누구나, 어디서든 쓸 수 있는 'AI 민주화' 이룰 것". (<https://www.seaily.com/NewsView/29X2GGUCZH>)

시큐리티월드. (2025.3.). 한화비전, 산업별 맞춤 기능 제공하는 툴킷 'AI PACK' 등 주목. (<https://www.boannews.com/media/view.asp?idx=136571>)

시큐리티월드. (2025.9.). 한화비전, AI 앞세운 전시·체험관으로 남미 보안시장 점점 확대. ([https://m.boannews.com/html/detail.html?tab\\_type=1&idx=138999#:~:text=Image%20%E2%96%B2%ED%95%9C%ED%99%94%EB%B9%84%EC%A0%84%20CCTV%20%2C000%EC%97%AC%EB%8C%80%EB%A5%BC%20%EB%8F%84%EC%9E%85%ED%95%9C,%EC%9E%90%EB%A3%8C%3A%20%ED%95%9C%ED%99%94%EB%B9%84%EC%A0%84](https://m.boannews.com/html/detail.html?tab_type=1&idx=138999#:~:text=Image%20%E2%96%B2%ED%95%9C%ED%99%94%EB%B9%84%EC%A0%84%20CCTV%20%2C000%EC%97%AC%EB%8C%80%EB%A5%BC%20%EB%8F%84%EC%9E%85%ED%95%9C,%EC%9E%90%EB%A3%8C%3A%20%ED%95%9C%ED%99%94%EB%B9%84%EC%A0%84))

한화비전 웹사이트. (<https://www.hanwhavision.com/ko/technology/wisenet9/>)

ALL ABOUT CIRCUITS.(2024.1.). NXP Serves up a 77 GHz Radar SoC for Software Defined Vehicles. (<https://www.allaboutcircuits.com/news/nxp-serves-up-a-77-ghz-radar-soc-for-software-defined-vehicles/#:~:text=The%20goal%20of%20automotive%20vision,capability%20via%20NXP%E2%80%99s%20S32%20architecture>)

androidcentral. (2025.5.). Sony's new IMX500 is the world's first camera sensor with AI smarts. (<https://www.androidcentral.com/sony-imx500-worlds-first-camera-sensor-ai-smarts>)

Apple-A. (2023.6.). Introducing Apple Vision Pro: Apple's first spatial computer. (<https://www.apple.com/newsroom/2023/06/introducing-apple-vision-pro/>)

Apple-B. (2023.6.). Developer tools to create spatial experiences for Apple Vision Pro now available. (<https://www.apple.com/newsroom/2023/06/developer-tools-to-create-spatial-experiences-for-apple-vision-pro-now-available/>)

ArduCam. (2025.). OAK-D: OV9282×2+IMX378+Intel Movidius Myriad X for Depth AI OAK. (<https://www.arducam.com/opencv-ai-kit-oak-d-ov9282x2imx378intel-oak-d.html>)

- Bhide, A., Dzhumamuratova, A., and Welsh, J. (2025). Build Synthetic Data Pipelines to Train Smarter Robots with NVIDIA Isaac Sim. (<https://developer.nvidia.com/blog/build-synthetic-data-pipelines-to-train-smarter-robots-with-nvidia-isaac-sim/>)
- BrainChip. (2024.1.). BrainChip to Unveil Akida Neuromorphic Processor Enabled by Microchip's 32-bit MPU at CES 2024. (<https://brainchip.com/brainchip-to-unveil-akida-neuromorphic-processor-enabled-by-microchips-32-bit-mpu-at-ces-2024/>)
- BrainChip. (2025.). Akida in Space. (<https://brainchip.com/akida-in-space/>)
- BrainChip. (2025.8.). BrainChip launches Akida Cloud for instant access to latest Akida neuromorphic technology. (<https://brainchip.com/brainchip-launches-akida-cloud-for-instant-access-to-latest-akida-neuromorphic-technology/>)
- CEVA. (2025.4.). Ceva Neural Processing Unit IP for Edge AI Selected by Nextchip for Next-Generation ADAS Solutions. (<https://www.ceva-ip.com/press/ceva-neural-processing-unit-ip-for-edge-ai-selected-by-nextchip-for-next-generation-adas-solutions/>)
- CODICO. (2018.8.). KNOWLES New Family of SMART Microphones. (<https://www.codico.com/en/en/current/news/knowles-new-family-of-smart-microphones#:~:text=%E2%80%9CAlways>)
- COOLENJOY. (2024.1.). DEEPX의 DX-M1 칩, CES 2024에서 선도적인 사물 AI 솔루션으로 인정. (<https://coolenjoy.net/bbs/38/5216646#:~:text=DEEPX%EC%9D%98%20DX,%EB%AA%A8%EB%8D%B8%EA%B9%8C%EC%A7%80%20EA%B4%91%EB%B2%94%EC%9C%84%ED%95%9C%20AI%20%EB%AA%A8%EB%8D%B8%EC%9D%84%20%EC%A7%80%EC%9B%90%ED%95%A9%EB%8B%88%EB%8B%A4>)
- DEEPX 웹사이트. DEEPX AI Chip Line-up Powering the Future of Edge AI. (<https://deepx.ai/products/ai-chips/#:~:text=DX>)
- Edge AI and Vision Alliance. (2024.3.). What is an NPU, and Why is It Key to Unlocking On-device Generative AI?. (<https://www.edge-ai-vision.com/2024/03/what-is-an-npu-and-why-is-it-key-to-unlocking-on-device-generative-ai/>)
- Edge AI + vision alliance. (2025.7.). Sony Semiconductor Demonstration of AI Vision Devices and Tools for Industrial Use Cases. (<https://www.edge-ai-vision.com/2025/07/sony-semiconductor-demonstration-of-ai-vision-devices-and-tools-for-industrial-use-cases/>)
- EDGE IMPULSE. (2024.4.). Picking the Brain of AI: BrainChip on the State of Neuromorphic Processing. (<https://www.edgeimpulse.com/blog/brainchip-state-of-neuromorphic/>)
- eeNEWS. (2025.4.). DeepX plans 2nm edge AI chip. (<https://www.eenewseurope.com/en/deepx-plans-2nm-edge-ai-chip/>)
- EETimes. (2023.1.). NXP introduces new Radar chipsets for improved ADAS applications. (<https://www.eetimes.com/nxp-introduces-new-radar-chipsets-for-improved-adas-applicati>)

- ons/#:~:text=,industry%20for%20safety%20and%20security)
- electropages. (2022.5.). Self-learning AI smart sensor for fitness tracking. (<https://www.electropages.com/2022/05/self-learning-ai-smart-sensor-fitness-tracking#:~:text=Mouser%20now%20stocks%20the%20BHI260AP,functionalities%20in%20a%20SiP%20solution>)
- embedded. (2025.2.). Sensors and AI for Smart Sensing. (<https://www.embedded.com/sensors-and-ai-for-smart-sensing/>)
- emotion3D. (2022.1.). Nextchip and emotion3D deliver ultra-efficient driver and occupant monitoring package based on APACHE5 platform. (<https://emotion3d.ai/nextchip-and-emotion3d-deliver-ultra-efficient-driver-and-occupant-monitoring-package-based-on-apache5-platform/>)
- emotion3D. (2022.7.). aiMotive and emotion3D deliver optimized driver monitoring solutions for aiWare NPU on Nextchip APACHE5. (<https://emotion3d.ai/aimotive-and-emotion3d-deliver-optimized-driver-monitoring-solutions-for-aiware-npu-on-nextchip-apache5/>)
- Hyungsoo, P. (2024.5.). Featured Stock] PixelPlus Overcomes Technical Challenges of OnSensor AI Technology... AI Chip Vertical Packaging New Technology Up. (<https://cm.asiae.co.kr/en/article/2024050809162698309#:~:text=PixelPlus%20has%20launched%20the%20image,reduction%20and%20automatic%20exposure%20correction>)
- Knowles. (2018.6.). Vivo Selects Knowles SmartMic in New Flagship “vivo NEX” AI Smartphone. (<https://investor.knowles.com/news/news-details/2018/Vivo-Selects-Knowles-SmartMic-in-New-Flagship-vivo-NEX-AI-Smartphone-06-12-2018/default.aspx#:~:text=Featuring%20Knowles%20SmartMic%20,wake>)
- Luxonis. (2025.). Robotics Vision Core 4(RVC4). (<https://docs.luxonis.com/hardware/platform/rvc/rvc4/>)
- MARTIN. D. (2025.7.). RealSense Almost Died At Intel. Now It’s Spinning Out With \$50M In Funding. (<https://www.crn.com/news/internet-of-things/2025/realsense-almost-died-at-intel-now-it-s-spinning-out-with-50m-in-funding#:~:text=For%20the%20next%20few%20years%2C,other%20businesses%20during%20Gelsinger%E2%80%99s%20tenure>)
- Medium. (2024.3.). BrainChip’s Akida: Neuromorphic Processor Bringing AI to the Edge. (<https://medium.com/@neurocortexai/brainchips-akida-neuromorphic-processor-bringing-ai-to-the-edge-0aed37968a02>)
- NVIDIA. (2022.9.). NVIDIA Unveils DRIVE Thor — Centralized Car Computer Unifying Cluster, Infotainment, Automated Driving, and Parking in a Single, Cost-Saving System. (<https://nvidianews.nvidia.com/news/nvidia-unveils-drive-thor-centralized-car-computer-unifying-cluster-infotainment-automated-driving-and-parking-in-a-single-cost-saving-system>)

- NVIDIA. (2025.8.). NVIDIA Blackwell-Powered Jetson Thor Now Available, Accelerating the Age of General Robotics. (<https://nvidianews.nvidia.com/news/nvidia-blackwell-powered-jetson-thor-now-available-accelerating-the-age-of-general-robotics>)
- NXP. (2024.1.). NXP Extends Industry-First 28 nm RFCMOS Radar One-Chip Family to Enable ADAS Architectures for Software-Defined Vehicles. (<https://www.nxp.com/company/about-nxp/newsroom/NW-NXP-EXTENDS-INDUSTRY-FIRST-28#:~:text=,meets%20the%20latest%20security%20requirements>)
- OPTOFIDELITY. (2024.2.). Apple Vision Pro Benchmark Test 2: Angular Motion-to-Photon Latency in VR. (<https://www.optofidelity.com/insights/blogs/apple-vision-pro-benchmark-test-2.-angular-motion-to-photon-latency-in-vr>)
- PatentPC. (2025.10.). “The Cost of Self-Driving Technology: How Much Do AV Components Really Cost? (Market Breakdown)”. (<https://patentpc.com/blog/the-cost-of-self-driving-technology-how-much-do-av-components-really-cost-market-breakdown>)
- Poddar, N. (2025.6.). The State of On-Device AI: What’s Missing in Today’s Landscape. (<https://www.nimbleedge.com/blog/state-of-on-device-ai>)
- RealSense 웹사이트. Introducing the RealSense™ D400 Product Family. (<https://realsenseai.com/stereo-depth/introducing-intel-realsense-d400-product-family/#:~:text=We%20have%20greatly%20accelerated%20computing,devices%20such%20as%20mobile%20phones>)
- RFID JOURNAL. (2024.1.). New NXP Radar Features Connect Software-Defined Vehicles. (<https://www.rfidjournal.com/news/new-nxp-radar-features-connect-software-defined-vehicles/202077/#:~:text=The%20SAF86xx%20is%20a%20radar,of%20ICs%20for%20radar%20sensors>)
- ROOS, G. (2023.12.). Samsung launches two sensors for robotics and XR, EDN. (<https://www.edn.com/samsung-launches-two-sensors-for-robotics-and-xr/#:~:text=The%20Vizion%2063D%20sensor%20claims,noise%20for%20minimal%20motion%20blur>)
- Ryoji, E. (2022.10.). “Realizing the world's first image sensor with built-in AI engine! Discover the stories behind the drive to create the world’s first of its kind”. (<https://www.sony-semicon.com/en/feature/2022101101.html>)
- Sony. (2021.10.). Sony to Launch Edge AI Platform Service AITRIOS™ To Streamline the Development and Implementation of AI Camera-Driven Sensing Solutions. (<https://www.sony-semicon.com/en/news/2021/2021100601.html>)
- ST. (2022.2.). STMicroelectronics’ “Intelligent Sensor Processing Unit” Integrates Brains into Sensors to Launch Onlife Era. (<https://newsroom.st.com/media-center/press-item.html/p4427.html#:~:text=Intelligent%20Sensor%20Processing%20Unit%20,sensor%20on%20the%20same%20silicon>)

- ST. (2022.6.). STMicroelectronics' new inertial modules enable AI training inside the sensor. (<https://newsroom.st.com/media-center/press-item.html/p4460.html#:~:text=STMicroelectronics%20,edge%20to%20%E2%80%9Cin%E2%80%9D%20the%20edge>)
- Sunarto, E. (2021.4.). "Brainchip begins mass production of its neuromorphic AI chips". (<https://stockhead.com.au/tech/brainchip-begins-mass-production-of-its-neuromorphic-ai-chips/>)
- TechCrunch. (2023.6.). First impressions: Yes, Apple Vision Pro works and yes, it's good. (<https://techcrunch.com/2023/06/05/first-impressions-yes-apple-vision-pro-works-and-yes-its-good/>)
- THEELEC. (2021.5.). Nextchip to supply ADAS SoC to Japanese customer. (<https://www.theelec.net/news/articleView.html?idxno=2828#:~:text=Nextchip%20will%20supplying%20its%20advanced,1%20supplier>)
- Volvo Cars. (2024,9.). From car to cloud: Volvo Cars expands collaboration with NVIDIA. (<https://www.media.volvocars.com/global/en-gb/media/pressreleases/331848/from-car-to-cloud-volvo-cars-expands-collaboration-with-nvidia>)
- Yole. (2023.11.). 2023 Microcontroller Industry Outlook. (<https://www.yolegroup.com/yole-group-actuality/2023-microcontroller-industry-outlook/>)



## 부록 1

## 지각 특화 AI 반도체의 기술 자문 결과 : ETRI 전문가 (온디바이스AI연구본부, AI로봇연구본부, 지능형반도체연구본부) 답변 내용 요약

### ■ 전문가 자문 내용 및 구성

○ ① 지각 특화 AI 반도체의 필요성 및 개념 관련 기술 자문, ② 휴머노이드 로봇 구성요소 내 항목별 지각 특화 AI 반도체 적용 가능 비중 의견

#### ○ 전문가 구성

- ETRI 인공지능, 로봇, 온디바이스 AI 등 관련 기술개발 전문가 13명 서면 자문
- ETRI 온디바이스AI연구본부, AI로봇연구본부, 지능형반도체연구본부

### ■ 멀티모달 센싱이 고도화될수록, 기존의 중앙집중형 SoC 구조의 기술적 한계

#### ○ 중앙집중형 SoC 구조의 한계

##### - 성능 병목 및 데이터 병목

- 고해상도 카메라·LiDAR·마이크 등 센서 데이터가 폭증하면서 SoC로 집중 전송 시 대역폭(Bandwidth) 한계, 전력 소모, 지연(latency) 문제가 심화됨
- 예를 들어, VGA 해상도 카메라 1대만으로도 초당 27MB, AI 추론 시 50~100MB/s의 대역폭이 필요하며, 실제 시스템은 2~3개 센서만 안정적으로 처리 가능

##### - 지연과 프레임 손실

- 다수 센서의 동기화·타임스탬프 보정 과정에서 프레임 드롭과 수십 ms의 응답 지연 발생
- 자율주행·휴머노이드 등 10ms 이하 실시간 응답이 필요한 응용에는 치명적 제약

##### - 전력 소모 및 발열

- 데이터 이동에 필요한 에너지가 연산보다 10~100배 높음. 전체 전력의 60~80%가 데이터 전송에 소비되며, 배터리 기반 로봇·차량에서는 작동시간 급감
- SoC의 클럭을 높여 처리 시 에너지 소비와 발열, 시스템 불안정성이 증가

##### - 확장성 및 유지보수 한계

- 센서 수 증가 시 중앙처리 복잡도는  $O(N^2)$ 로 증가
- 새로운 센서 추가나 펌웨어 업데이트 시 전체 시스템 재설계 필요
- 단일 SoC 장애로 전체 시스템이 중단되는 Single Point of Failure 문제 상존

○ 기술적 근본 원인

- 대역폭 및 버스 경합(Contention)

- 다수 센서의 데이터 스트림이 동시에 중앙 버스를 점유해 충돌과 지연이 발생

- 데이터 동기화 복잡도

- 센서별 샘플링 주기, 클럭 드리프트 차이로 정합(Alignment)·보간 과정이 복잡하며, 오차 누적으로 정확도 저하

- 시스템 전성비(TOPS/W) 한계

- NVIDIA Jetson 등 SoC는 세대별 성능(TOPS)이 비선형적으로 향상되지만, 연산 요구가 더 빠르게 증가해 GPGPU와의 격차 유지
- CNN→Transformer 기반 GenAI로의 전환은 연산량과 메모리 요구량을 폭증시켜 온디바이스 처리를 어렵게 함

○ 해결 방향 및 기술 전환

- 엣지 컴퓨팅(Edge Computing)

- 데이터 발생지(센서 근처)에서 전처리·요약·필터링을 수행해 중앙 SoC의 부담을 완화
- 데이터 이동 거리 단축으로 전력 효율 및 실시간성 개선

- 센서별 SoC 또는 System-on-Sensor 구조

- 센서마다 독립적 연산 모듈을 내장해, 연산을 분산처리함으로써 대역폭·지연 문제를 해소
- 모듈형 설계로 확장성과 유지보수성 향상

- 하드웨어·소프트웨어 공동 최적화

- 데이터 이동 최소화를 위한 메모리 계층 설계, 저전력 인터랙트(UCIe 등), 병렬처리 아키텍처 필요
- AI 패러다임 전환에 대응한 SoC-Edge-Heterogeneous Fusion 구조로 진화 중

■ 메인 SoC의 기술적 한계 극복을 위해서 지각 특화 AI 반도체와 같은 구조가 실질적인 기술개발 측면의 대안이 될 수 있을까에 대한 가능성

○ 지각 특화 AI 반도체의 대안 가능성

- 센서 데이터 폭증과 중앙집중형 SoC 의 전력·대역폭 한계로 기존 시스템은 자율주행·로봇 등 Embodied AI 응용에서 실시간 처리 및 지연 요구를 만족하기 어려움

- 이를 완화하기 위해 지각 특화 AI 반도체가 센서 근처에서 대용량 데이터를 전처리·요약 후 고차원 임베딩 형태로 중앙 SoC에 전달하는 센서 근처의 분산 지각 아키텍처로 제시 가능

○ 기술적 구조 및 필요 기능

- 공통 전처리 : MobileNet·경량 Transformer 기반 멀티모달 어텐션 임베딩 추출
- OTA(Over-the-Air) 업데이트 : 학습 파라미터를 주기적으로 갱신 가능해야 다양한 태스크에 적응
- 하드웨어 기반 동기화 및 QoS 보장: PTP/TSN 시간 정렬, 데드라인 인식 스케줄링, semantic communication 내장으로 멀티모달 지터 및 우선순위 역전 억제
- 모듈형 확장성: 센서-MCU 쌍을 독립 모듈로 플러그-앤-플레이 연결, 개별 펌웨어 업데이트 및 고장 격리 가능

○ 분산 지각 특화 AI 아키텍처의 효과

구분	개선 내용
지연 (Latency)	데이터 생성지 근처 처리로 RT 응답성 향상
전력 소모	이벤트 기반 적응형 샘플링 → $\mu$ A 수준 대기전력
대역폭	메타데이터만 전송 → 네트워크 부하 100-400배 감소
실시간성	RTOS 우선순위 스케줄링, DMA 병렬처리
신뢰성	센서 모듈 단위 독립 동작 → 부분 고장 시 시스템 유지

○ 과제 및 한계

- 중앙 SoC 역시 연산 성능이 빠르게 향상되고 있어 차별화 필요
- 지각 특화 AI와 메인 SoC의 이중 학습·연동 복잡성 증가
- 고정밀 경량 모델 및 데이터셋 정합 문제 해결 필요
- Gen AI 수준의 복합 센서 추론까지 확장하려면 추가적 모델 경량화 및 연산 최적화가 필요

■ 지각 특화 AI 반도체 개념 정의에 대한 기술적 타당성 및 실현 가능 방향성 의견

○ 기술적 타당성

- (HW 측면) 국내에서도 NPU 상용화가 시작되면서 SoC 기반 지각 특화 AI 칩 설계가 가능한 수준에 도달. 다만, 패키징 공정 및 제조 비용, 발열·전력 관리 등의 세부 기술적 검증이 필요
- ASIC 기반 엣지 AI 칩은 기존 CPU·GPU 대비 수십~수천 배의 와트당 연산 성능 (W/W)을 달성하며, 밀리초(ms) 단위의 지연 시간을 구현 가능
- (SW 측면) 현재 수준에서도 개념 검증용 가상화 및 시뮬레이션 기반 선행연구가

가능하며, 향후 HW 완성 단계에서 최적화된 AI 컴퓨팅 SW로 확장 가능

- CNN, Transformer 등 '인지 중심' AI 모델의 연산 특성(행렬 연산 중심)에 맞춘 NPU·PIM·fusion accelerator 설계는 성능 및 전력 효율 면에서 합리적임

○ 기술 동향 및 실현 가능성

- (멀티모달 융합(fusion) 고도화) 단일 이미지 인식에서 벗어나, 다양한 센서 데이터의 Point·Voxel·Region·Time-series 수준 통합이 활발히 연구 중
- 이는 task-agnostic 구조로 일반화되고 있으며, rare·long-tail 시나리오 증강에 Generative AI/MM-LLM이 보조적으로 활용
- (산업 적용 사례) Tesla FSD는 자율주행 스택을 휴머노이드 로봇에 확장 적용하며, E2E OTA 시스템 반도체 개념을 구현 중
- 분산 시스템-중앙집중형 시스템 발전의 역사적 흐름상, 고성능 분산형 인지 시스템으로의 수렴은 기술적으로 타당한 방향으로 평가
- (관련 기술군) Neuromorphic, in/near-sensor computing, compute-in-memory 등은 최신 논문과 리뷰에서 공통적으로 지각 특화 AI 반도체 구현에 필요한 핵심 기반 기술요소로 제시

○ 성능 보완 및 과제

- 일부 언어-지시형 테스트(IFEval, TruthfulQA 등)는 양자화(quantization) 민감도가 높아, 지각 특화 AI SoC 내에서도 제어 경로 및 후단 모듈 정밀도 향상이 필요
- 특정 모델(FSD, VLM 등)에 종속될 경우 빠른 세대 교체로 경쟁력 상실 우려가 있으므로, 모듈화·범용화된 HW/SW 인터페이스 설계가 바람직함

■ 기술적 실현을 위한 필수 선결과제

○ 기술적 필수 요소

- (연산 성능, 전력 효율, 발열 관리) 엣지 환경에서 실시간 처리를 유지하면서 저전력화를 달성해야 함
  - PIM/NVM 구조로 데이터 이동 최소화 및 와트당 성능(TOPS/W) 극대화
  - 저정밀 연산(2-8bit), 희소성(Sparsity)활용으로 효율 향상
  - Thermal Throttling 방지를 위한 고급 냉각 및 패키징 기술 필요
- (대역폭 및 메모리 아키텍처) DRAM 왕복 최소화, SRAM-CIM, chiplet+HBM/PoP 구조를 통한 처리 효율 개선
- 센서-메모리-연산 간 병목을 줄이는 System-Level co-Design이 중요

- 첨단 패키징 및 모듈화

- Heterogeneous Integration을 통한 센서·AI 코어·메모리의 집적화
- 자율주행·로봇용 소형 고신뢰 패키지 구현 (충격, 진동, 온도 내성)

○ AI 구조와 학습 전략

- 경량화 모델 설계(On-device AI)

- MobileNet, SqueezeNet, Depthwise Separable Conv, NAS 적용을 통한 초경량 구조 설계
- MCU 수준에서도 실용적 성능을 확보해야 하며 단순 모델 압축이 아닌 엣지 최적 설계 필요

- 실시간 운영체제 및 스케줄링

- Rate Monotonic, EDF 기반 RTOS(FreeRTOS, Zephyr 등)에 AI 워크로드 대응형 스케줄러 적용
- Priority Inheritance Protocol과 Sporadic Server로 지연과 우선순위 문제 해결

- 데이터 품질 및 학습 체인

- Simulation·Augmentation으로 Bias·Long-tail 보완, Foundation Model 재활용
- 경량 SNN 및 온디바이스 학습 툴 체인 구축 필수

○ 시스템 통합 및 소프트웨어

- (Fusion Accelerator) BEV, cross-attention, filter 등 핵심 커널 최적화로 멀티모달 정렬 품질 유지

- (Real-time Sync & QoS) PTP/TSN 기반 타임스탬프와 NoC 스케줄링으로 지연 보장

- (Safety & Diagnostics) Watchdog/self-test/reflex path로 Fail-Operational 시스템 지향

- (SW 생태계) NPU 최적화 컴파일러·디바이스 드라이버·네트워크 스택 통합 필요.

○ 지각 특화 AI 응용과 표준화

- 지각-추론-제어 통합(Physical Intelligence)

- Pi Zero 프레임워크처럼 VLA (vision-language-action) 모델을 로봇 인지·제어에 적용
- 대규모 VLM을 활용해 언어-시각 결합 기반 로봇 학습 및 동작 생성

- Neuro-Symbolic AI 접목

- 심볼 표현 기반의 표준화 및 메타데이터 공유가 필요<sup>134)</sup>

- 현장 적용성(Edge Deployment)

- 스마트팩토리, 병원, 가정 환경에서 전력 제약, 네트워크 변동, 안전 요구 대응을

134) Nawaz et al.(2025)의 "Neuro-symbolic AI for advanced cognitive systems" 모델이 참고 사례임

위한 OTA(Over-the-Air) 업데이트 및 자율 재조정 기능이 필수

■ 시각 특화 AI 반도체의 주요 기능에 대한 기술적 타당성과 추가로 필요한 기능

○ 기술적 타당성

- 센서 퓨전(Sensor Fusion)

- 이미 검증된 기술 영역으로, 확장 칼만 필터(EKF), 파티클 필터, 그래프 기반 SLAM 등이 MCU 수준에서도 구현 가능
- Ex. ARM Cortex-M4 기반에서 IMU-카메라 퓨전 5ms, 라이다-카메라 정합 20ms 이내 처리 가능
- 하드웨어 타임스탬프, 인터럽트 기반 데이터 수집 등은 이미 상용 MCU 표준 기능

- 온디바이스 시각처리(On-Device Perception)

- TinyML 기반으로 객체 감지, 키워드 인식, 이상 탐지 등 실시간 가능
- INT8 양자화 및 프루닝으로 100KB 이하 모델에서도 90% 이상의 정확도 확보
- Google Edge TPU, Coral Dev Board 등에서 5ms 이내 추론 성능 달성 사례 존재

- 메타데이터 생성 및 요약 기능

- 원시 데이터 대비 전송량 100배 이상 절감 가능
- JSON, Protocol Buffers 등 경량 직렬화 포맷을 통해 고수준 의미정보를 효율적으로 전송

○ 추가 필요 기능

- 적응형 모델 선택 및 동적 재구성 (Adaptive Model Switching)

- 환경·전력·지연 조건에 따라 모델 복잡도 조정
- Ex. 고전력 시 고정밀 모델, 저전력 시 경량 모델 자동 선택

- 온디바이스 학습 및 연합학습 (On-Device / Federated Learning)

- 로컬 데이터로 모델을 지속 개선하면서 프라이버시 보호
- Transfer Learning, Few-shot Learning을 통한 빠른 환경 적응

- 자가진단 및 품질 모니터링 (Self-Diagnosis)

- 센서 드리프트, 모델 성능 저하, 하드웨어 이상 실시간 감지
- 불확실성·신뢰도 정보를 메타데이터에 포함시켜 상위 의사결정 지원

- 보안 및 무결성 (Security & Integrity)

- AI 모델 지식재산 보호 및 Adversarial 공격 방어
- 데이터·추론 무결성 확보 위한 하드웨어 보안 모듈 필요

- 동적 워크로드 및 전력 제어 (DVFS 기반 관리)

- 동적 주파수/전압 스케일링, 모델 업데이트 가속 등으로 전력 효율 최적화

■ 시각 특화 AI 반도체를 구현하기 위한 필요 요소기술

○ 주요 기술 구성 요소

- 하드웨어 아키텍처

- (Neuromorphic Computing) Event-driven, 비동기(asynchronous) 방식의 SNN 구조를 통해 Von Neumann 병목을 완화하고, 시공간(spatiotemporal) 패턴 처리를 고효율로 수행
- (Heterogeneous Acceleration) GPU/NPU/FPGA/ASIC과 Neuromorphic 코어의 공동 배치(Co-deploy)로 복합 연산 수요 대응
- (Near-sensor Processing) 데이터 전송량을 줄이기 위해 센서 근처에서 특징(feature)을 추출하는 SPU(Pixel Processor Array)기반 저전력 전처리 구현

- 센서 인터페이스 및 동기화

- (표준화 인터페이스) MIPI CSI-2, GMSL/FPD-Link, LVDS/SLVS-EC 등 다중 센서 간 정합성과 신뢰성 확보
- (시간 정렬 및 TSN 지원) 정밀한 타임스탬프 동기화를 위한 하드웨어 타이머 내장형 이더넷 사용(10Gbps 이상)
- (ISP 우회(raw data 입력)) 이미지 센서의 raw 데이터(Bayer)를 NPU로 직접 입력하여 ISP 지연 제거 및 고품질 결과 확보

- 소프트웨어 및 런타임 스택

- (AI 반도체용 컴파일러 및 런타임) 모델 변환, 최적화, 하드웨어 매핑을 지원하는 통합 SW 스택 구축 필요
- (PyTorch 등 프레임워크 연계 어댑터) 오픈소스 기반 생태계와의 호환성 확보
- (모델 최적화 기술) 양자화(quantization), 프루닝(pruning), 파인튜닝(fine-tuning) 기반의 실시간 경량화 및 추론 가속
- (HW-SW Co-design 커뮤니티) 다양한 분야 연구자가 참여하는 개방형 개발 구조 제안

○ 핵심 알고리즘 및 기능

- Sensor Fusion

- Transformer 기반 융합(TransFusion 등)으로 멀티센서 간 정보의 상호보완적 통합

- Online Calibration

- 온도·진동 등으로 인한 센서 정렬 변화의 실시간 보정

- On-device AI

- 로컬 추론 가속 및 지속적 학습(on-device learning)

- Metadata 표준화

- 센서별 메타데이터 구조 및 전송 표준 정립

- Edge-Cloud 연계

- 분산 지능형 실행을 위한 클라우드 오프로드·플랫폼 연계 기술

○ 구현을 위한 필수 요소기술

- 하드웨어 설계 및 아키텍처 기술 : Neuromorphic, Heterogeneous Core 등
- 센서 인터페이스 및 동기화 기술 : MIPI, TSN Ethernet, Time Alignment 등
- AI 컴퓨팅 SW 및 알고리즘 기술 : Compiler, Runtime, Fusion, Calibration 등

■ 지각 특화 AI 반도체가 수행해야 할 추론 연산 수준

○ 핵심 구성 및 연산 단계

- 1단계) Near-Sensor Encoding & Preprocessing

- ISP, 노이즈 필터링, FFT/DCT, 엣지 검출, 히스토그램 균등화, 특징 추출 (SIFT/SURF/ORB 등)

- 2단계) On-Device Fusion & Inference Heads

- BEV projection, cross-attention, correlation, 검출·추적 등

- 3단계) Host/Cloud Global Reasoning

- SLAM, HD map, occlusion recovery, retraining/distillation

- XR·자율주행 등 초저지연 응용(≈10ms)의 경우 Level 1~2는 on-device로 수행하고, Level 3은 클라우드로 위임하는 계층적 분산 구조

○ 연산 성능 요구사항

- (NPU 성능) 수~수십 TFLOPS급, CNN/RNN/Transformer 가속 지원
- (DSP/MCU) point cloud 및 센서 데이터 실시간 처리용 내장 고속 연산 장치
- (추론 수준) MobileNet, SqueezeNet 등 경량 AI 모델 기반의 중간 수준 inference 수행
- (데이터 특성) 디지털 외에도 센서 기반 아날로그 데이터의 추론 지원 필요
- (추론 결과 활용) 모델 개선·재학습(retraining)으로의 확장 가능성 고려

○ 최적화 및 설계 원칙

- (Data-in-Place) DRAM 왕복 최소화 및 on-chip 연산 비중 극대화
- (Bandwidth 절감) 카메라 원본(예: RAW) 대비 수배 축소된 feature uplink 구조
- (하드웨어-소프트웨어 통합 최적화) Pruning, Quantization 등 경량화 기법과 NPU 아키텍처의 연동

- (초저지연(Ultra-Low Latency)) ms 단위 실시간 반응성 확보

○ 기술적 의의

- 지각 특화 AI 반도체는 센서 인지·추론·결정이 통합된 지능형 엣지 컴퓨팅 칩

- ① 데이터 이동 최소화, ② 연산 효율 극대화, ③ 실시간성 및 신뢰성 향상을 달성

- 이는 향후 자율주행, XR, 로봇틱스, 스마트시티 등 Physical AI 시대의 핵심 기반 기술로 자리매김 가능

■ 센서 전처리부터 AI 연산까지 통합 설계를 위한 시스템 아키텍처 구성방안

○ 센서-연산-통신-안전성이 긴밀히 통합된 시스템 아키텍처가 필요

○ 계층별 시스템 구조

- Near-Sensor Processing Layer

· (역할) 센서 인접 MCU/FPGA가 아날로그-디지털 변환, 신호 컨디셔닝, 타임스탬핑, 이벤트 감지 수행

· (특징) 마이크로초 수준의 저지연, 마이크로와트급 초저전력으로 의미 있는 데이터만 상위로 전달

- Sensor Fusion Layer

· (역할) 고성능 MCU/DSP가 다중 센서의 시공간 정합 및 상태 추정 수행 (칼만 필터 등)

· (특징) 하드웨어 가속기를 이용한 매트릭스 연산 최적화로 밀리초 단위 처리 지연 보장

- AI Inference Layer

· (역할) NPU/GPU 기반 딥러닝 추론(객체 인식, 장면 이해, 행동 예측 등)

· (특징) 적응형 모델 선택 및 동적 자원 할당으로 성능-전력 균형 유지

○ 핵심 기술 요소

- (고대역-저지연-저전력 균형) TSV 적층 구조로 대역폭·지연·에너지 효율을 동시에 확보

- (정밀 동기화) PTP/TSN 기반 하드웨어 타임스탬프를 이용해 sub- $\mu$ s 수준의 동기화 유지

- (Semantic Codec) 객체, 키포인트, BEV, 포인트 시맨틱을 태스크 기반 비트스트림으로 전송해 통신, 연산을 공동 최적화

- (전력/열 제어) Multi-domain DVFS, Clock/Power Gating, On-die Thermal Sensing(ms 단위 피드백)을 활용하여 Tail-Latency 억제

- (안전성 확보) Reflex Path, Watchdog, Self-Test 등으로 Perception 결과의 무결성 상시 점검

○ 하드웨어 설계 방향

- 칩렛 구조는 AI 반도체의 유연한 업그레이드·교체를 가능하게 하며, AI 모델 진화 (CNN→Generative AI 등)에 대응하기 유리함
- 특정 AI 모델에 종속된 반도체 설계는 위험하므로, 플랫폼 수준의 지각 특화 AI HW 기반 위에 다양한 AI 칩 교체형 구조를 채택할 필요가 있음

항목	단일 SoC	칩렛 기반	HBM/TSV 구조
센서 수·대역 확장성	중간	높음(모듈화 용이)	매우 높음
지연/결정성	양호	양호~우수	우수
성능 확장성	제한적	매우 유연	고성능 한정

■ 초저전력/초지연 조건을 만족하기 위해 고려해야 할 반도체 설계 방식

○ 초저전력과 초저지연을 동시에 달성하기 위한 구조적 혁신이 요구

- 이벤트 기반 센서(Event-based Sensor), 근사연산(Approximate Computing), PIM(Processing-in-Memory)을 중심으로 한 새로운 아키텍처 제시

○ 주요 기술 구성

- 이벤트 기반 처리(Event-driven Processing)
  - Dynamic Vision Sensor(DVS)를 적용하여 변화가 발생한 픽셀만 처리함으로써 데이터량을 최대 1,000배 감소, 지연을 마이크로초( $\mu$ s) 수준으로 단축
  - Asynchronous Event Processing을 통해 센서 이벤트 발생 시 즉시 처리 시작 → 버퍼링 지연 제거
  - Priority-based Event Queue로 중요 이벤트를 우선 처리, Event-driven State Machine으로 복잡한 시퀀스도 효율적으로 제어
- 신경망 경량화 및 근사 연산 기법
  - Early Termination : 신뢰도 임계치 도달 시 연산 조기 종료
  - Cascade 구조 : 입력 복잡도에 따라 점진적으로 연산 경로를 확장
  - Tiny-Attention 및 Sparse CNN/SNN(Spiking Neural Network)을 L1·L2 수준에 분리 배치하여 reflex 경로를 단축
  - 단순 입력의 경우 얇은 네트워크만 활성화하여 전력 소비 최소화

- 고급 전력 관리 및 회로 최적화

- Heterogeneous Multi-core구조 : 고성능 코어는 burst 처리, 저전력 코어는 상시 모니터링 전용
- Fine-grained Power Gating: 미사용 블록을 클럭 사이클 단위로 차단
- Power Island 분할 설계로 독립적 전력 제어
- Retention Register로 상태를 유지하며 절전 모드 전환

- 메모리 중심 설계 (PIM: Processing-in-Memory)

- 지연과 전력의 주범은 연산 자체가 아니라 데이터 이동(Data Movement)임
- 연산을 메모리 근처(혹은 내부)로 이동시켜 전력과 지연을 동시에 절감
- 아날로그 PIM 기반 AI 칩을 통해 연산·메모리 융합, 데이터 이동 최소화, 전력 효율 극대화

○ 기술 간 상호보완 관계

- 초저전력과 초저지연은 상충이 아닌 보완적 관계
- 저전력화를 위해서는 불필요한 메모리 접근 및 컴포넌트 구동을 줄여야 하나, 재기동 시 복구 지연이 발생할 수 있음
- 따라서 효율적 메모리 관리, 모델 사전 적재(pre-loading), 전력 상태 전환 최적화 등이 필수적임

○ 응용 및 기대효과

- XR motion-to-photon, 근접 충돌 회피(reflex motion)sub-10ms급 반응이 요구되는 응용에 적합
- 정적 배경 중심 감시 카메라, 체스처 인식, 자율주행 지각 모듈등에서 데이터 폭증 억제 및 실시간성 확보
- 결과적으로, 에너지 효율성(Energy Efficiency) 및 처리 속도(Latency Reduction)를 동시에 달성하는 AI-SoC 진화 방향성을 제시

■ 지각 특화 AI 반도체의 기술 범위 및 목표시장 영역을 고려할 때, 온센서 AI(지능형센서 포함), 온디바이스AI(디바이스 내 메인 연산 SoC), 로봇틱스의 제어 유닛 MCU 등(택일 가능)과의 개념적, 기술적 차별성

○ 개념 및 역할

- 지각 특화 AI 반도체는 센서로부터 수집된 다중 모달 데이터를 실시간으로 전처리·융합(fusion)·요약하여, 상위 SoC(System-on-Chip)의 연산 부담을 줄이는 중간 지각 허브로, 센서 근처에서 저전력·저지연으로 데이터의 의미(semantic

information)를 생성함으로써, 메인 연산 칩과 제어 유닛 간의 효율적 협력 구조를 형성함

- 온센서 AI : 센서 내부의 단순 신호 전처리·탐지 기능 중심 (예: DVS, event-based sensor)
- 온디바이스 AI (Main SoC) : 범용 고성능 연산 중심으로 대규모 모델과 통신·그래픽 등 전체 시스템을 담당
- 로보틱스 제어 MCU : 실시간 제어와 안전 로직 중심, 지각 연산 능력은 낮음
- 지각 특화 AI : 상기 세 계층을 연결하는 센서 융합·지각 중심 반도체, 피지컬 AI의 감각 뇌(sensory brain) 역할 수행

○ 주요 기술적 차별성

구분	온센서 AI	지각 특화 AI	메인 SoC (On-device AI)	MCU
주요 기능	센서별 단일 신호처리	멀티센서 데이터 융합·지각	범용 연산 및 통합 판단	실시간 제어·안전
처리 위치	센서 내부	센서 근처(엣지 전단)	중앙 연산부	액추에이터 근처
장점	전력·원가 유리	대역폭·지연 절감, 글로벌 컨텍스트 확보	대규모 연산 가능	높은 결정성·안정성
한계	확장성 제한	상위·하위 칩 연계 필요	전력/열 부담	지각 처리 부적합
대표 KPI	mW급/센서	센싱→지각 <50ms, ≥10× 대역 절감	대모델 처리량	제어 주기·안전 무결성
주요 적용	스마트센서	ADAS, 드론, XR, 로봇 등	IVI·HMD·폰	모터/팔 제어

○ 설계 원리 및 구조적 특징

- Multimodal Fusion 최적화

- 이벤트 기반(event-driven)·프레임 기반 신호의 타임스탬프 정렬 및 지터 억제를 통해 bounded latency 확보

- Fusion 전용 가속기 내장

- BEV projection, cross-attention, Kalman filtering(ESKF/UKF) 등의 핵심 커널을 하드웨어화

- On-chip memory hierarchy 설계
  - Compute-in-Memory(CIM), scratchpad SRAM 활용으로 DRAM 접근 최소화
- Semantic communication 지원
  - feature/keypoint 기반의 QoS-aware 인터페이스로 edge-cloud/로봇 간 협업 인지 강화
- 안전 지향(fail-operational) 설계
  - reflex-like fast path, watchdog, self-test 기능을 포함해 안정적 동작 보장

○ 시스템 내 상보적 역할

- (지각 특화 AI ↔ SoC) SoC의 연산 부하·대역폭 소모를 줄이며, 고수준 판단 연산을 보조
- (지각 특화 AI ↔ MCU) 지각 기반 제어 신호를 전달하여 상황인식형 제어 (closed-loop, context-aware control) 구현
- (온센서 AI ↔ 지각 특화 AI) 전처리된 데이터를 융합하여 더 정교한 인지 결과를 생성

■ 유사 기술(온센서 AI, 메인 연산 SoC, 제어 유닛 MCU 등)의 산업에서 지각 특화 AI 반도체의 기능과 관련된 최근 기술/산업 동향

○ 온센서(On-Sensor AI) 및 지능형 센서 동향

- (엣지화·저전력화) 센서 다이 내 AI 연산을 통합하거나 3D 적층을 통해 전송 전력을 최소화하며, 이벤트 기반·저비트 양자화 기술을 활용해 전력 소모를 기존 대비 1/10 수준으로 절감
- 대표 사례
  - Sony IMX500- 이미지센서 내 객체 인식 및 이벤트 감지 수행, 데이터 전송량 감소
  - Bosch BME688- 환경 센서에 AI 프로세서 내장, 온도·습도·가스 등 실시간 분석
  - 뉴로모픽·이벤트 카메라- 마이크로초 단위 저지연, 초저전력 동작으로 로봇틱스·자율주행에 적용 확대
- (의의) 지각 특화 AI 칩과 결합해 근-센서 데이터 요약 및 전처리 계층을 형성, 시스템 전체의 효율 향상

○ 메인 연산 SoC(On-Device AI) 동향

- (고성능·멀티모달 처리 강화) 이미지·텍스트·음성·열·IMU 등 다양한 센서 신호를 통합하는 Transformer 기반 모델이 차량·로봇 내에서 직접 구동
- 주요 제품 :

- NVIDIA Jetson-Thor- 2000 TOPS(4fp), 130 W로 6개 카메라 이상 동시 처리. 자율주행·비전 LLM 지원
- Qualcomm Ride Flex- 단일 SoC로 ADAS~L4 자율주행, mixed-criticality 워크로드 통합
- Mobileye EyeQ Ultra- 176 TOPS, 센서 독립형 구조, REM 맵과 연동 + ASIL-D 안전성
- Google Coral, TI TDA4VM, Xilinx Kria등 엣지 AI SoC 다수 출시
- (추세) NPU 성능 및 에너지 효율 상승, LLM·비전 트랜스포머 온디바이스 처리 보편화

○ 제어 유닛 MCU 및 로보틱스 통합

- Deterministic Control 중심으로 지각 특화 AI 칩이 생성한 semantic 정보를 소비하는 구조로 진화
- (핵심 요구) functional safety, data integrity 확보를 위한 경량 모니터링 경로 및 실시간 제어 인터페이스
- (산업 추세) ST STM32N6, NXP Neutron NPU, ARM Ethos-U85 등 저전력 AI MCU 출시로 센서-제어 간 지연 감소 및 엣지 AI 확산 가속

○ 패키징·통합 및 HW-SW 협업 동향

- (기술 핵심) HBM·TSV·칩렛 기반 고대역폭 통합, Memory-Compute co-location, Event-Driven Front-End + Tensor-Centric Back-End 구조
- (SW 생태계) MCU 제조사별 AI SDK ( ST, NXP 등 ) 및 MIT MCUNet v1-v3 연구로 최적화 진행
- (통합 효과) 지연·전력 절감 및 멀티모달 데이터플로우 최적화로 실시간성 강화

○ 시사점 및 전망

- 지각 특화 AI 반도체는 On-Sensor AI의 전송 효율, Main SoC의 집중 추론, MCU의 결정론적 제어를 연결하는 지각 허브로 진화
- 근접 특징추출 + 멀티모달 융합 + 시간인식 스케줄링 + 안전 피드백을 통합한 새로운 분산 지각 층을 형성함
- 이를 통해 대역폭·지연·에너지를 근본적으로 절감하고, Embodied AI 시대의 신뢰도 높은 지능형 센서 플랫폼으로 확장될 것으로 예상

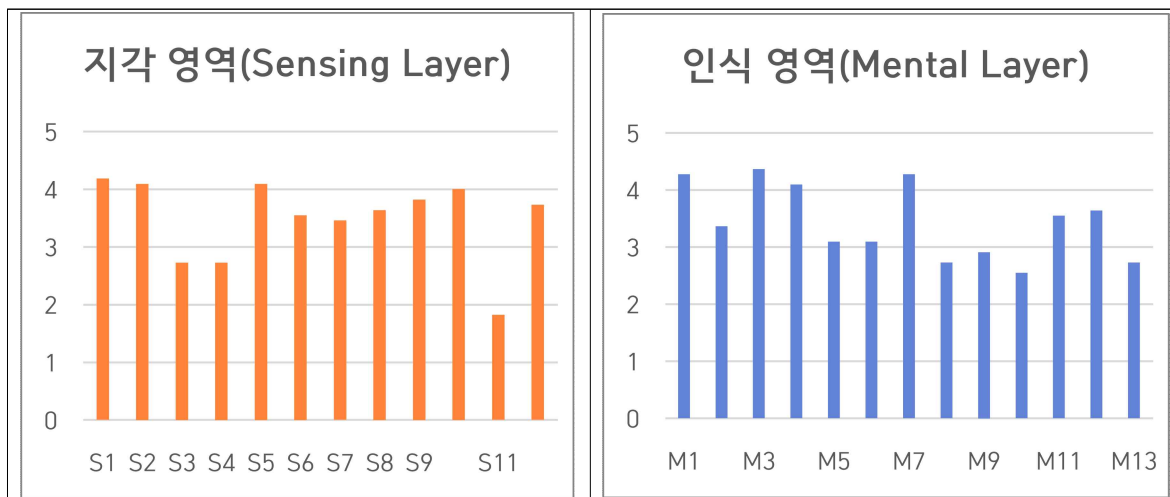
■ 휴머노이드 로봇 구성요소 내 항목별 지각 특화 AI 반도체 적용 가능 비중

○ 휴머노이드 로봇 구성요소 항목별 지각 특화 AI 반도체의 관련 적합도

Mental Layer - Decision Processing		관련 적합도					Sensing Layer - Data Collection	관련 적합도				
		없음	하	중	상	일치		없음	하	중	상	일치
SW	Autonomous navigation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	MCUs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	AI training and AICG Solutions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Lens and camera modules	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Machine vision solution	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	VCSEL	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Control and remote control solutions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	EEL	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	HMI solutions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	CMOS Imagine sensors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	OS	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Tactile sensors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HW	SoC	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Torque sensors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	RF component and communication modules	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6D torque sensors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	DRAM	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	IMUs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	NAND Flash	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Environmental sensors	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	ISP	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ePaper	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	DSP	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	MEMS microphone	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Security Chip	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						

[ 휴머노이드 로봇 구성요소별 세부 기술들과 지각 특화 AI 반도체 간의 관계 분석표 ]

- 전문가 분석 결과 지각 영역(Sensing Layer)의 평균값은 3.48, 인식 영역(Mental Layer)의 평균값은 3.48로 나타남
- 휴머노이드 로봇 구성요소 내 지각 영역 비중 15%, 인식 영역 23%에 평균값 반영하여 지각 특화 AI 반도체의 적용 가능 비중을 26%(지각: 10%, 인식: 16%)로 도출



[ 지각 및 인식 영역 전문가 자문 결과 ]



## 부록 2

## 시장조사기관별 글로벌 휴머노이드 로봇 시장 추정 방법

## ◆ Precedence Research

## ○ 시장 정의

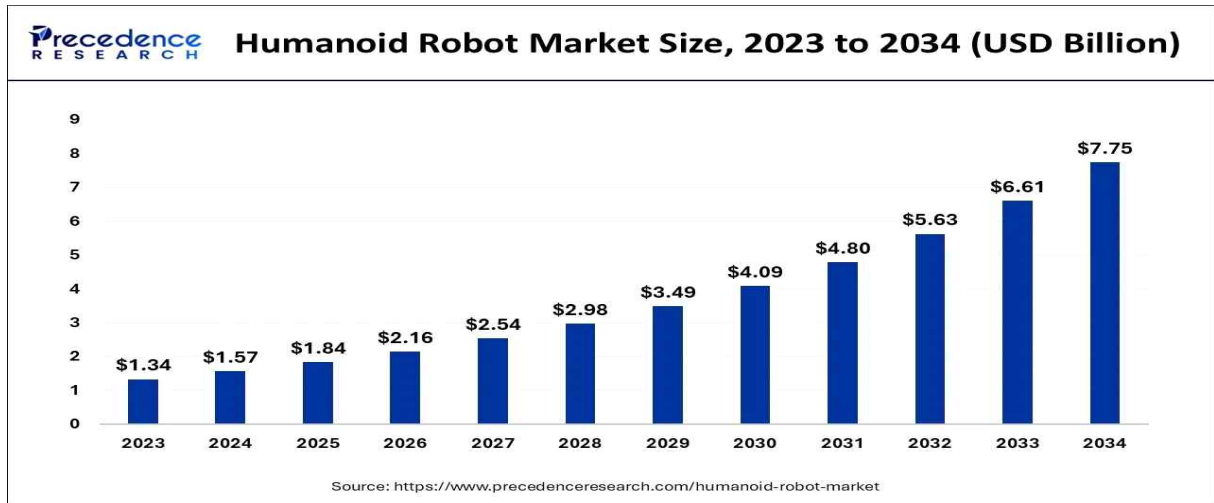
- (유형) 이족보행형과 휠 기반형으로 구분; 이족보행형은 사람과 유사한 보행 동작을 구현할 수 있어 서비스·케어 분야에서 유리하며, 휠 기반형은 구조적으로 안정성과 에너지 효율이 높아 물류·산업 현장에 적합한 형태로 정의함
- (구성요소) 하드웨어와 소프트웨어로 나누었으며, 하드웨어에는 센서(비전, 자이로, 토크 등), 모터, 배터리와 같은 핵심 구동·인지 장치가 포함; 소프트웨어에는 AI 제어 알고리즘과 운영체제가 포함됨
- (응용분야) 초기에는 개인도움·케어와 같은 생활 밀착형 서비스가 시장 형성을 주도할 것이며, 시간이 지남에 따라 물류, 보안, 고객센터 등 산업 및 상업 분야로 확장될 가능성을 제시함

## ○ 시장 규모 추정

- (규모) 글로벌 휴머노이드 시장 규모는 2024년 15.7억 달러에서 2034년 77.5억 달러에 이를 것으로 전망함(CAGR 17.3%)
- (성장특징) 구성요소별로는 소프트웨어 부문이 2023년 기준 가장 큰 시장 비중을 차지하고 있으며, 향후에도 하드웨어 대비 빠른 성장세를 유지할 것으로 예측함; 지역별로는 북미가 2034년 기준 최대 매출을 기록하였으나, 아시아태평양 지역은 중국·일본·한국의 제조 역량과 기술 보급 속도를 기반으로 가장 높은 성장률을 기록할 것으로 분석됨
- (성장동인) 기술 발전, 인공지능 통합 확대, 그리고 인간-로봇 상호작용(HRI) 기술의 고도화가 장기적 성장을 견인할 핵심 요인으로 제시됨; 특히 고령화에 따른 케어 수요 증가와 위험 환경에서의 로봇 활용 확대가 시장 수요를 지속적으로 창출할 것으로 판단함

## ○ 방법론

- (방식) 공급자 매출 실적 기반의 외삽 모델을 사용하여 시장규모를 예측함; 과거 실적 데이터를 기반으로 연도별 성장률을 산출하고, 보수적 성장 시나리오를 적용하여 단기적 변동성보다 장기적 평균 성장률에 중점을 둔 접근 방식을 채택함
- (데이터) 주요 로봇 제조사의 제품 카탈로그, 투자자 보고서, 산업 동향 보고서, 과거 보급 추세 데이터를 다각적으로 활용함
- (가정) 시장형성 초기 단계의 데이터를 바탕으로 예측 모델을 구축하였으며, 기술 발전 속도와 가격 하락률이 점진적으로 반영되는 비교적 리니어한 성장 곡선을 가정함



[ 글로벌 휴머노이드 로봇 시장 규모 전망치 (Precedence Research) ]

출처: Precedence Research, Humanoid Robot Market Size, Share, and Trends 2024, 2024.11.

## MARKETSANDMARKETS

### ○ 시장 정의

- (유형) 이족보행형과 휠 구동형으로 구분, 이족보행형은 인간과 유사한 움직임을 통해 다양한 환경에서 높은 적응성을 보이는 형태로, 서비스·교육·케어 등 사람 중심의 응용에 적합하다고 정의함; 휠 구동형은 안정성과 기동성을 중시하며, 제조·물류 환경에서의 연속적 작업에 강점을 가진 형태로 정의함
- (구성요소) 하드웨어 부문에는 센서(자이로, 가속도, 위치, 비전, 토크 등), 액추에이터(전기·공압·유압·압전 방식), 제어시스템이 포함되며, 소프트웨어 부문에는 AI 기반 제어, 객체 인식, 음성 인지 및 상황 인지 기능을 담당하는 알고리즘과 플랫폼이 포함됨
- (응용 분야) 교육·연구, 헬스케어, 제조·물류 산업 등에서 수요 확대가 예상되며, 특히 산업 현장에서는 자동화 효율성 제고와 인력 대체 수요, 서비스 영역에서는 맞춤형 상호작용과 보조 기능 수요 증가 등을 주요 성장 요인으로 제시함

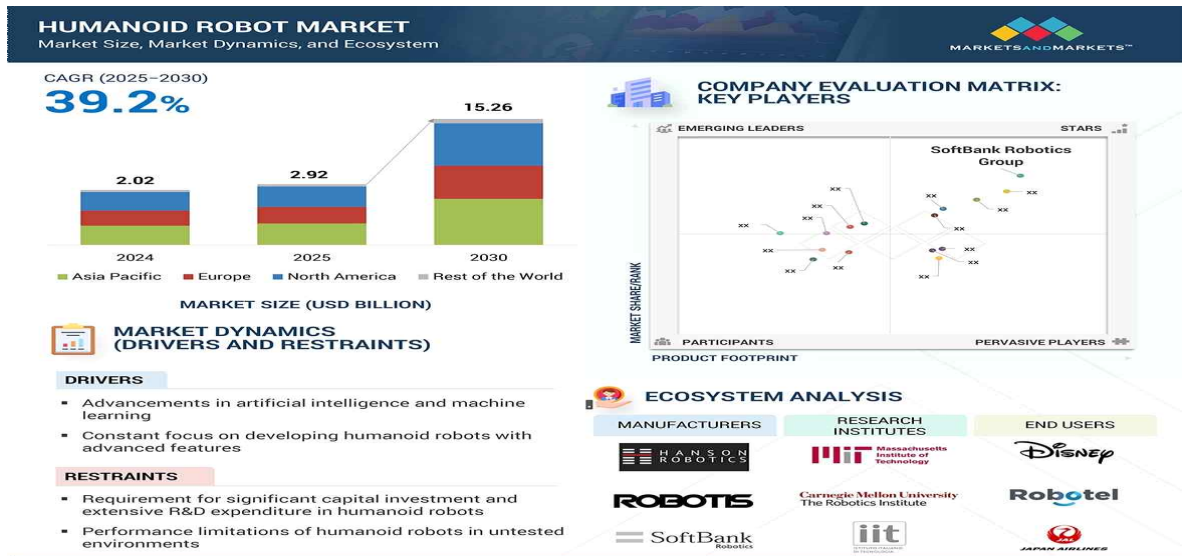
### ○ 시장 규모 추정

- (규모) 글로벌 휴머노이드 로봇 시장 규모는 2024년 약 29.2억 달러에서 2030년 약 153억 달러로 성장할 것으로 전망됨(CAGR 39.2%)
- (성장특징) 지역별로 북미는 2024년 기준 최대 시장 점유율을 차지하였으며, 아시아태평양 지역이 향후 가장 높은 성장률을 기록할 것으로 전망함; 제품 유형 중에서는 이족보행형 휴머노이드가 높은 시장 비중을 유지하며, 소프트웨어 부문이 하드웨어보다 빠른 성장세를 보일 것으로 분석함
- (성장동인) AI 제어·객체 인식·음성 인지 기능의 발전, 제조 및 물류 산업에서의

자동화 수요 증가, 하드웨어 단가 하락 등이 시장 확산을 가속화할 주요 요인으로 제시함

○ 방법론

- (방식) 상향식(기기 출하량·평균단가 합산)과 하향식(산업 전체 규모 대비 휴머노이드 비중) 접근 방식을 병행하여 시장규모를 산정함
- (데이터) 주요 기업 보고서, 신제품 출시 자료, 산업 연간 보고서, 재무 보고서, 기술 컨퍼런스 발표 자료, 유료 통계 데이터 등을 활용함
- (가정) 산업별 적용 비율과 기술 확산 속도가 연도별로 점진적으로 증가한다는 전제를 두었으며, 단기 성장세를 반영하기 위해 초기 기간(2024~2030년)의 성장률을 상대적으로 높게 설정함



[글로벌 휴머노이드 로봇 시장 규모 전망치 (MARKETSANDMARKETS) ]

출처: MARKETSANDMARKETS, Humanoid Robot Market Size, Share & Trends, 2025.4.

◆ Global Market Insights

○ 시장 정의

- (유형) 이족보행형과 휠 기반형으로 구분, 이족보행형은 인간의 보행 및 동작을 모사하여 사람과 동일한 작업 환경에서 높은 적응성을 발휘하는 형태로, 휠 기반형은 이동 효율성과 안정성이 뛰어나 산업 현장에서 장시간 운용에 유리한 형태를 말함
- (구성요소) 하드웨어에는 센서(비전, 위치, 힘·토크 등), 액추에이터, 전원 장치 등이 포함되며, 소프트웨어에는 AI 제어 알고리즘과 머신러닝 기반 운영 시스템이 포함됨; 이러한 구성요소는 IoT, 디지털 트윈, AI·ML 기반 제어 기능과의 결합을 통해 기능 확장과 성능 고도화를 지원하는 것으로 정의함

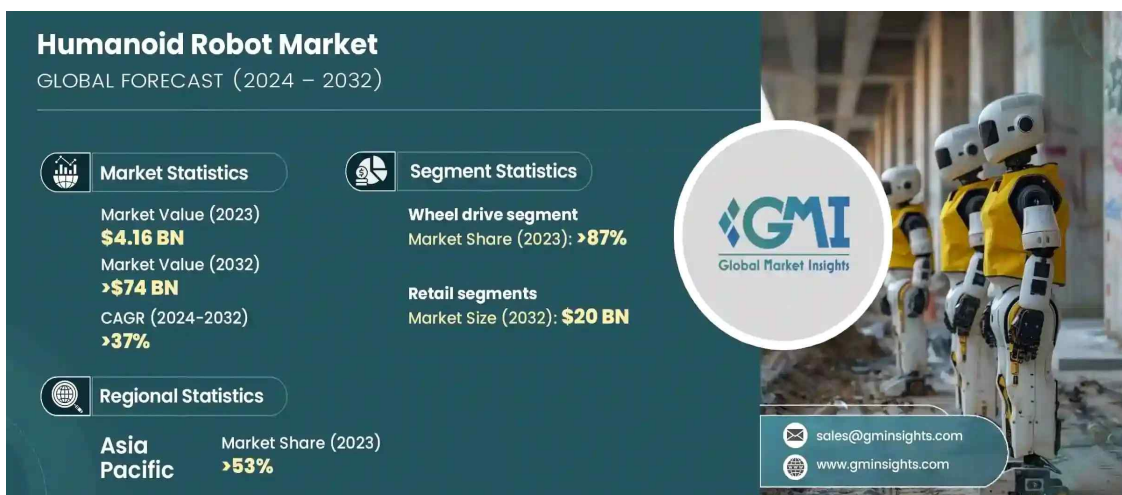
- (응용분야) 국방, 건설, 헬스케어, 교육, 연구, 리테일 등 다양한 산업 분야를 대상으로 하며, 각 분야별로 상이한 요구 사항과 환경 조건에 맞추어 휴머노이드 로봇의 설계와 기능이 최적화되는 것으로 설정함

○ 시장 규모 추정

- (규모) 글로벌 휴머노이드 로봇 시장은 2023년 약 42억 달러에서 2032년 약 740억 달러로 성장할 것으로 전망됨(CAGR 37.0%)
- (성장특징) 산업 자동화 수요의 증대가 전체 시장 확대의 핵심 원인으로 분석하였으며, 지역별로는 북미와 유럽이 안정적이고 성숙한 시장 구조를 유지하는 반면, 아시아태평양은 제조·서비스 산업에서의 빠른 기술 도입을 기반으로 가장 높은 성장률을 기록할 것으로 전망함
- (성장동인) AI·ML 기반 제어 기능, 고도화된 객체 인식 기술, 그리고 IoT 및 디지털 트윈 기술과의 융합이 주요 성장 촉진 요인으로 제시; 이러한 기술 통합은 휴머노이드 로봇의 자율성, 정밀성, 실시간 데이터 처리 능력을 강화하여 산업 전반에서의 활용 가치를 높일 것으로 언급함

○ 방법론

- (방식) 산업 매출 통계와 주요 업체별 실적 데이터를 기반으로 하드웨어와 소프트웨어 매출을 합산하고, 수요 확산 추세를 반영한 CAGR을 적용
- (데이터) 글로벌 주요 로봇 제조사 실적 자료, 시장 매출 통계, 기술 동향 보고서, 산업별 수요 분석 자료를 활용함
- (가정) 시장 확산 속도가 기술 발전 및 가격 하락과 맞물려 가속화된다는 가정을 두었으며, AI·ML·IoT·디지털 트윈 등 첨단기술의 통합이 중장기적 수요를 지속적으로 견인할 것으로 전제함



[ 글로벌 휴머노이드 로봇 시장 규모 전망치 (Global Market Insights) ]

출처: Global Market Insights, Humanoid Robot Market, 2024.8.

◆ Goldman Sachs

○ 시장 정의

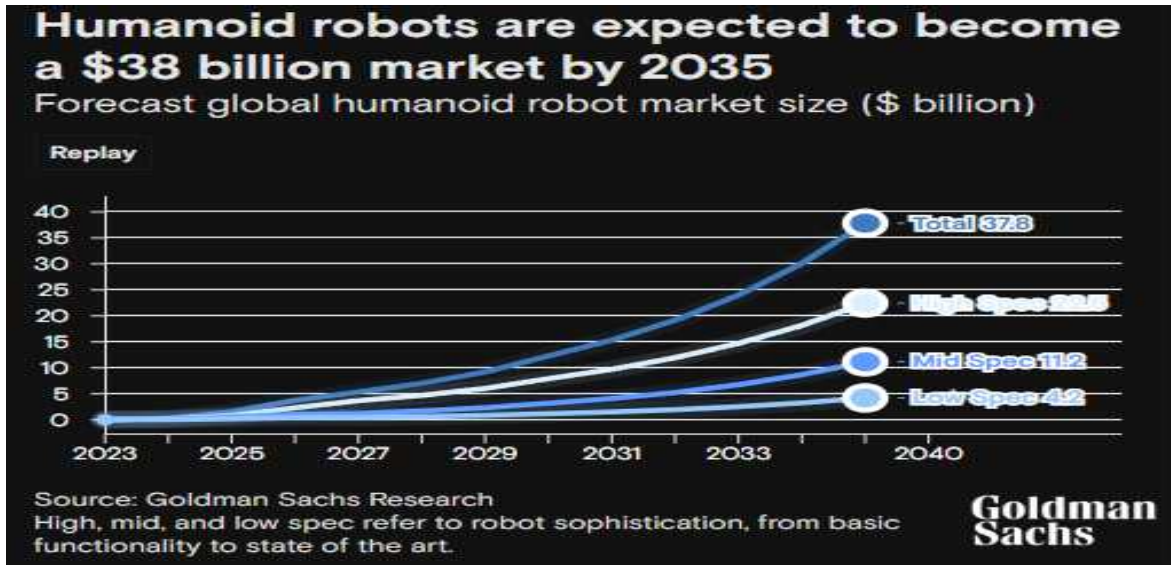
- (유형) 머리·팔·다리 등 인간과 유사한 신체 구조를 갖추고, 물리적 상호작용과 노동 대체가 가능한 로봇으로 정의함; 이러한 정의는 단순 서비스 로봇과 구분되며, 실제 산업 및 생활 환경에서 사람과 유사한 형태로 다양한 작업을 수행할 수 있는 능력을 전제로 함
- (구성요소) 하드웨어에는 메카니컬 바디, 센서, 카메라, 고정밀 모터, 배터리와 같은 핵심 부품이 포함되며, 소프트웨어에는 AI 제어 시스템, 대규모 언어모델(LLM) 기반 지각 엔진, 클라우드 연동 운영체제가 포함됨; 이러한 구성은 로봇의 자율성, 상황 인지 능력, 그리고 실시간 의사결정 역량을 강화하는 데 초점을 맞춤
- (응용분야) 초기에는 제조·물류 등 구조화된 환경에서 수요가 집중될 것으로 예상하였으며, 기술 성숙 이후에는 케어·서비스 분야로 확산될 것으로 분석함; 특히 대규모 제조 기반을 갖춘 중국과 미국이 주요 초기 수요처이자 성장 동력이 될 것으로 판단함

○ 시장 규모 추정

- (규모) 세계 휴머노이드 로봇 시장은 2024년 약 2억 달러에서 2035년 약 378억 달러로 성장할 것으로 전망됨(CAGR 61.0%)
- (성장특징) 초기 수요는 제조 및 물류 산업에서 집중될 것이며, 이는 기존 자동화 설비의 한계를 보완하고 작업 효율성을 극대화하기 위함임; 이후 로봇 기술의 신뢰성 및 안전성이 확보되면, 고령화 사회 대응을 위한 케어 서비스 분야와 상호작용 기반의 서비스 산업으로 확산될 것으로 예측함
- (성장동인) 부품 단가의 급격한 하락, AI·모션 제어·인지 기능의 혁신적 향상, 그리고 고위험·고반복 작업의 대체 수요를 핵심 성장 요인으로 분석함; 특히 LLM 기반 지각 엔진과 클라우드 연동 시스템의 발전이 로봇의 적응성과 활용성을 크게 확장할 것으로 판단함

○ 방법론

- (방식) 시나리오 기반의 시뮬레이션 모델을 사용하여 기술 확산 속도, 부품 단가 변화, 보급률 등을 복합적으로 고려한 예측을 수행함
- (데이터) 주요 로봇 제조사의 제품 사양, 출시 시기, 가격 정보와 함께 부품 단가 추세, AI 및 LLM 기술 발전 경로 등을 통합 분석함
- (가정) 기술 발전과 부품 단가 하락이 예측보다 빠르게 진행될 경우, 휴머노이드 로봇의 보급이 가속화될 것이며, 초기 시장은 특정 산업에 집중되나, 장기적으로는 다수의 산업군에서 병렬적으로 수요가 증가할 것으로 가정함



[ 글로벌 휴머노이드 로봇 시장 규모 전망치 (Goldman Sachs) ]

출처: Goldman Sachs, Global Automation Humanoid Robot: The AI Accelerant, 2024.1.

## ◆ Straits Research

### ○ 시장 정의

- (유형) 이족보행형과 휠 기반형으로 구분, 이족보행형은 인간의 보행 패턴과 유사한 동작 구현이 가능하여 사람과의 상호작용이 필요한 서비스·교육·헬스케어 분야에 적합하다고 정의함; 휠 기반형은 안정성과 에너지 효율성이 높아 물류, 산업·제조 환경 등 연속 작업에 강점을 가진 형태로 설명함
- (구성요소) 하드웨어에는 액추에이터, 센서, 전력 시스템이 포함되며, 소프트웨어에는 AI 제어, 자연어 처리(NLP), 모션 제어 시스템이 포함됨; 이러한 구성요소는 로봇의 자율성, 상황 이해, 정밀 동작 구현을 가능하게 하는 핵심 기반으로 제시됨
- (응용분야) 교육·연구, 헬스케어, 엔터테인먼트·게임, 개인도움·컴패니언, 산업·제조 등에 활용될 수 있을 것으로 봄; 이를 통해 휴머노이드 로봇이 단일 산업에 국한되지 않고, 서비스와 제조 영역을 아우르는 광범위한 활용 가능성을 갖고 있음을 강조함

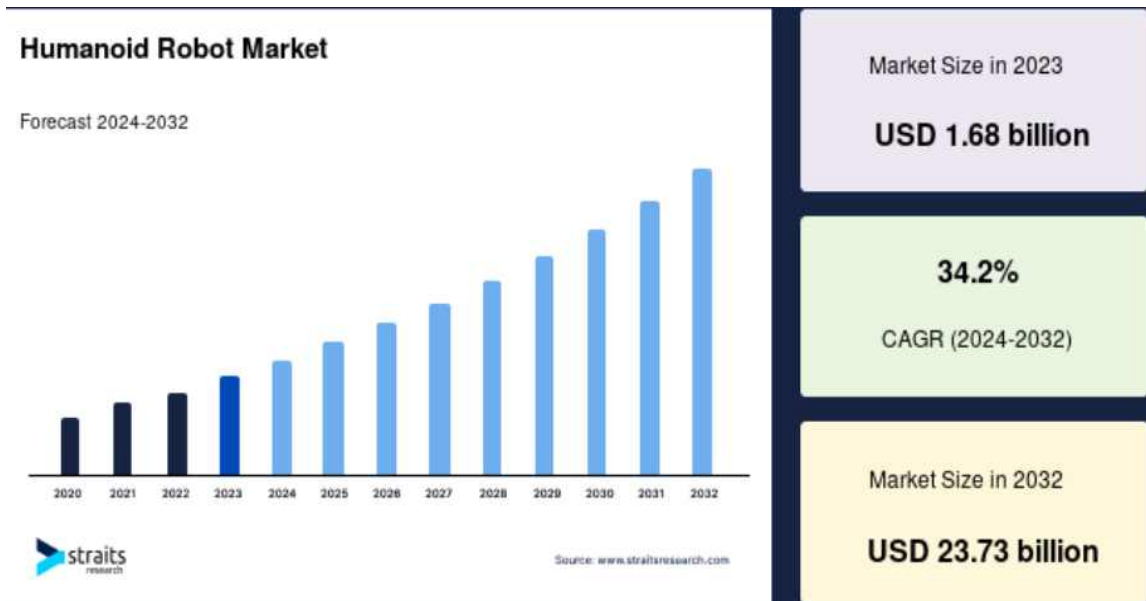
### ○ 시장 규모 추정

- (규모) 세계 휴머노이드 로봇 시장은 2024년 약 23억 달러에서 2032년 약 237억 달러로 성장할 것으로 전망됨(CAGR 34.2%)
- (성장특징) 지역별로 아시아태평양이 최대 시장 규모를 형성할 것으로 예상되며, 유럽이 가장 높은 성장률을 기록할 것으로 분석함; 이러한 전망은 아시아태평양의 제조 역량과 빠른 기술 도입 속도, 유럽의 로봇 규제 완화와 산업 혁신 정책에 기인한다고 평가함

- (성장동인) 고령화 및 인력 부족 문제 대응, 서비스·헬스케어·교육·물류·고객 서비스 등 다양한 분야로의 응용 확장이 시장 성장을 촉진할 것으로 제시함; 특히 엔터테인먼트·게임과 같은 신규 수요처 발굴 가능성이 장기 성장을 뒷받침하는 요인으로 언급됨

○ 방법론

- (방식) 전문기관 설문조사, 주요 기업 사례 분석, 공개 통계자료를 바탕으로 시장 수요 추세를 정성적으로 분석한 후, 이를 기반으로 시장 점유율을 연도별로 추정함
- (데이터) 주요 수치, 세그먼트별 하이라이트, 지역별 시장 동향을 포함함
- (가정) 각 산업 분야별 기술 확산 속도와 채택률이 점진적으로 증가한다는 전제를 두었으며, 서비스 및 제조 분야 모두에서 병행 성장이 이루어진다고 가정함



[ 글로벌 휴머노이드 로봇 시장 규모 전망치 (Straits Research) ]

출처: Straits Research, Humanoid Robot Market Size, Share & Trends Analysis Report, 2024.

**Mordor Intelligence**

○ 시장 정의

- (유형) 이족보행형, 휠 기반형 로봇으로 구분, 이족보행형은 사람과 동일한 환경에서 다양한 작업을 수행할 수 있는 범용성을 강조하며, 휠 기반형은 높은 이동 효율성과 구조적 안정성을 바탕으로 산업 및 물류 현장에서 장시간 활용에 적합하다고 설명함
- (구성요소) 하드웨어에는 센서, 액추에이터, 배터리가 포함되며, 소프트웨어에는 AI 제어 시스템, 자연어 처리(NLP), 모션 인식 알고리즘이 포함됨; 이러한 구성

요소는 로봇이 물리적 환경을 인지하고 실시간 의사결정을 수행하며, 사람과의 상호작용을 원활하게 하는 핵심 기반으로 설명함

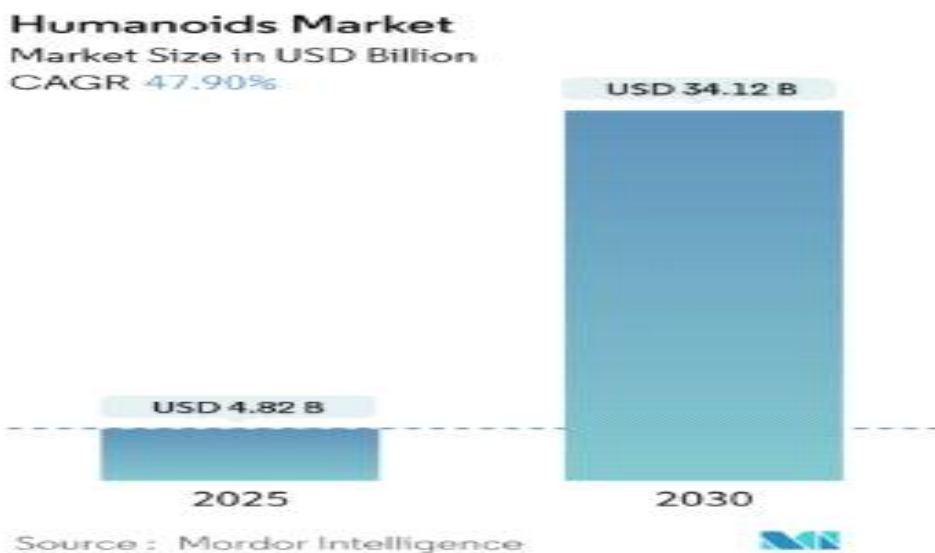
- (응용분야) 헬스케어, 서비스, 제조, 물류 등 다수의 산업 분야에 적용될 수 있으며, 특히 고령화 사회에서의 케어·서비스 로봇 수요 확대를 주요 활용 영역으로 제시함

○ 시장 규모 추정

- (규모) 세계 휴머노이드 로봇 시장은 2025년 약 48억 달러에서 2030년 약 341억 달러로 성장할 것으로 전망됨(CAGR 47.9%)
- (성장특징) 아시아태평양 지역이 정책 지원과 제조 역량 강화에 힘입어 가장 빠른 성장세를 보일 것으로 예측함; 특히 중국, 일본, 한국과 같은 제조 강국이 초기 확산의 중심지가 될 것으로 분석함
- (성장동인) AI 하드웨어 비용의 급격한 하락, 선진국과 중국을 비롯한 주요 국가들의 고령화로 인한 노동력 부족 문제, 그리고 아시아태평양 지역의 정부 주도 정책 및 투자 확대를 핵심 성장 요인으로 제시함

○ 방법론

- (방식) 상향식 접근법(기기 출하량 및 평균 단가 기반 시장규모 추정)에 정성적 조사(전문가 인터뷰, 기술 트렌드 분석 등)를 사용함
- (데이터) 산업 보고서, 주요 기업 공개자료, 기술 로드맵, 국제 기구 통계, 전문가 설문과 인터뷰를 포함하여 데이터를 수집함
- (가정) 기술 발전 속도와 부품 단가 하락률이 예측 기간 동안 지속적으로 유지된다는 가정 하에, 헬스케어 및 서비스 분야를 중심으로 한 수요 확대가 전체 시장 성장을 견인할 것으로 전제함



[ 글로벌 휴머노이드 로봇 시장 규모 전망치 (Mordor Intelligence) ]

출처: Mordor Intelligence, Humanoids Market Size & Share Analysis, 2025.7.

부록 3

온센서 AI(지능형센서 및 MCU) 기업 동향

◆ Sony (IMX500 중심으로)

- (기술개요) 온센서 AI 이미지 센서
  - 세계 최초로 AI 프로세싱 기능을 이미지 센서에 내장한 제품으로, 촬영 데이터의 전처리-추론-결과 출력을 센서 내부에서 수행(Sony, 2021.10.)
  - 12.3MP CMOS와 전용 DSP를 단일 칩 스택 구조로 적층해, 센싱과 연산을 물리적으로 결합한 통합 구조
- (핵심기술 구성/성능) 센서 내 실시간 추론 + 고성능 영상 처리 동시 제공
  - 4K 60fps, FHD 240fps 수준의 센싱 성능을 유지하면서 객체 분류·검출 등 딥러닝 추론을 센서에서 직접 수행
  - 외부 프로세서·메모리 의존을 줄여 데이터 이동량과 지연·전력 소모를 낮추고, 원본 대신 '의미정보' 중심 처리 흐름을 구현
- (개발동향) IMX500 확산 및 AITRIOS 플랫폼 기반 개발환경 구축
  - '20년 출시 이후 IMX500 기반 엣지 AI 카메라 생태계가 형성되었고, '21년 클라우드 연동 플랫폼 AITRIOS를 공개
  - '23년 라즈베리파이 협업 모듈 출시로 접근성을 높였으며, YOLO 등 범용 모델이 센서에서 구동되도록 SW 스택을 지속 개선(Edge AI+Vision alliance, 2025.7.)
- (시장적용) 매장·교통·관제 등 '대규모 영상+현장 판단' 수요처 공략
  - 리테일 매장 분석, 스마트시티 교통/혼잡 모니터링 등 대규모 비전 데이터가 발생하는 현장 중심 응용을 타깃으로 초기 시장을 개척
  - 로마시 교통 실증 적용 사례를 제시했고, 센서 샘플 가격을 약 1만 엔 수준으로 제시해 가격·도입 장벽을 낮춤(Androidcentral, 2025.5.; Sony, 2021)
- (기업전략) '센서+AI' 개방형 생태계 확장(파트너·오픈소스·플랫폼)
  - Microsoft, 라즈베리파이 등과 협업하여 센서-플랫폼-서비스까지 연결되는 생태계 확장을 추진
  - SW 스택 오픈소스 공개 및 개발 키트 제공으로 개발자 기반을 넓히고, 클라우드 연계로 운영·배포 편의성을 강화(Sony, 2021.10.)
- (정책/산업 시사점) 원본 최소화형 비전 아키텍처·표준 메타데이터 체계가 확산의 관건
  - 공공·산업 현장 도입 시 원본 영상 전송/저장 중심에서 현장 메타데이터 생성 중심으로 조달·성능지표를 전환할 필요가 있음(지연·비용·프라이버시 동시 대응)

- 센서 내 추론이 늘어날수록 메타데이터 스키마·전송 규격·보안 부팅/키관리 등 표준 체계가 중요해지므로, 인터페이스·데이터 표준화를 병행해야 함

### ◆ BrainChip (Akida 중심으로)

- (기술개요) 스파이킹 뉴로모픽 기반 초저전력 엷지 AI
  - 인간 두뇌 신경망을 모방한 스파이킹 뉴로모픽 프로세서로, 초저전력 always-on 엷지 AI 연산을 지향(BrainChip, 2024.1.)
  - 이벤트 기반 입력을 중심으로 ‘필요할 때만 연산’하는 구조를 통해 센서 상시 감지·각성 시나리오에 최적화
- (핵심기술 구성/성능) 이벤트 기반 병렬 처리 + 온칩 학습(edge learning)
  - 센서 이벤트를 스파이크 형태로 변환해 뉴런 코어에서 병렬 처리하며, 필요 시에만 연산을 수행해 에너지 효율을 극대화(Medium, 2024.3.)
  - 칩 자체에서 학습과 추론을 모두 수행하는 온칩 학습 기능을 제공하고, 1mW 수준 동작을 통해 상시 키워드 인식·이벤트 탐지를 강조(IEEE Spectrum, 2024.10.)
- (개발동향) 2세대 Akida IP 고도화 및 클라우드 평가환경 제공
  - `20년 1세대 이후 적용사례·피드백을 축적했고, `23년 말 2세대 Akida IP 공개로 성능·유연성을 강화(Edge Impulse, 2024.4.)
  - `25년 Akida Cloud 출범으로 하드웨어 없이도 테스트·개발이 가능하도록 하여 진입장벽을 낮춤(BrainChip, 2025.8.)
- (시장적용) 웨어러블·산업 모니터링·차량 상시 감지 영역 공략
  - 배터리 수명이 핵심인 보청기·웨어러블 헬스·이상 이벤트 탐지 등 항상 대기형 센싱 시장에서 적용 가능성이 부각
  - 산업 설비 모니터링 및 차량 실내 상시 모니터링 등 ‘이벤트 기반 지각’ 응용으로 확장을 모색
- (기업전략) IP 라이선싱+SDK/보드/클라우드로 생태계 확장
  - Akida IP를 다양한 SoC에 라이선스로 탑재하는 전략을 병행하며, 상용 MCU/MPU 결합을 통해 도입 장벽을 낮춤
  - SDK·교육자료·평가보드 유통과 Akida Cloud를 통해 개발·평가 비용을 낮추고 커뮤니티 기반 확산을 강화
- (정책/산업 시사점) 초저전력 always-on 지각은 센서-주연산 공백을 메우는 보완 축
  - 미션 크리티컬·상시 감지 서비스에서는 평균 성능보다 최악 지연/지터·전력 예산이

핵심이므로, 초저전력 지각 코프로세서군을 별도 전략으로 육성할 필요

- 확산 병목은 성능보다 튜체인·벤치마크·검증환경이므로, 모델 변환·성능/전력 측정·신뢰성 평가의 공용 인프라 구축이 중요

◆ **Luxonis (OAK - OpenCV AI Kit 중심으로)**

- (기술개요) VPU 내장 스마트 깊이 카메라(키트형 플랫폼)
  - 카메라 내 전용 VPU를 내장한 스마트 깊이 카메라 모듈로, 호스트 GPU 없이도 비전 AI를 오프로딩하는 구조(Luxonis, 2025)
  - 컬러+IR 기반 스테레오 심도 인식과 AI 추론을 결합해 3D 환경 인지 중심의 개발 키트 생태계를 제공
- (핵심기술 구성/성능) Depth 계산+AI 추론을 카메라에서 통합 수행
  - 초기 모델은 Myriad X 기반으로 깊이 계산과 신경망 추론을 동시에 수행하고, 컬러/IR 일체화로 3D 인식을 지원(Arducam, 2025)
  - 오픈소스 API(DepthAI)로 Python 등에서 파이프라인·신경망 제어가 가능해 빠른 PoC/프로토타이핑에 유리(Luxonis, 2025)
- (개발동향) 자체 칩(RVC4) 전환 및 로봇 통합(ROS) 강화
  - `20년 OAK-DC 공개 이후 커뮤니티 피드백을 반영하며 하드웨어·소프트웨어를 지속 혁신
  - `22년 자체 설계 RVC4 개발, `23년 OAK-4 출시 후 ROS 패키지·YOLO 최적화 및 멀티센서 융합(mmWave/열화상 등)을 추진
- (시장적용) 교육-PoC-산업 적용까지 확산되는 개발자 기반 플랫폼
  - 로봇틱스 스타트업의 SLAM·장애물 인식, 농업 드론 영상 분석 등에서 손쉬운 엣지 비전 구현 수단으로 활용
  - 보안 카메라·IoT 업체들이 오픈소스 스택과 호환성을 활용해 자체 솔루션에 통합하며 응용 범위를 확대
- (기업전략) OpenCV 커뮤니티 기반 개방·확장 전략
  - OpenCV 생태계와의 긴밀한 연계를 통해 하드웨어 개조·활용이 가능한 개방형 플랫폼 전략을 유지
  - 상대적으로 낮은 가격대와 빠른 개선 주기로 사용자 기반을 확대하고, 그 피드백으로 제품 고도화를 가속
- (정책/산업 시사점) 키트형 확산 모델은 실증 속도를 높이고 표준 SW 스택의 중요성을 부각

- 현장 적용의 첫 관문은 PoC 기간·비용이므로, 레퍼런스 키트/모듈 기반 실증은 도입 리스크를 크게 낮출 수 있음
- 확산은 하드웨어 성능보다 API·ROS·모델 최적화·배포체계에 의해 좌우되므로, SW 스택/SDK까지 포함한 산업 생태계 설계가 필요

**◆ Intel (RealSense Vision Processor D4 중심으로)**

- (기술개요) 스테레오 기반 3D 심도 계산 특화 ASIC
  - 스테레오 카메라 기반 3D 심도 센싱을 위해 설계된 맞춤형 ASIC으로, 실시간 깊이 계산에 특화된 프로세서
  - 모바일·드론·로봇 등 소형 기기에서 3D 인식을 임베디드하기 위한 전력·발열·패키지 최적화를 지향
- (핵심기술 구성/성능) SGM 하드웨어 구현 기반 고속 심도 처리
  - SGM 알고리즘을 하드웨어로 구현해 초당 3,600만 깊이 포인트 계산 및 90fps 이상 달성을 제시(RealSense, 2025.11. 접속 기준)
  - 6.3×6.3mm 패키지에서 22nW/포인트 이하 동작을 내세우며 동급 CPU 대비 전력 효율을 강조
- (개발동향) 신규 개발 중단, 오픈소스 SDK 중심 유지
  - `17년경 D4 공개 이후 `18~`19년 펌웨어 개선으로 저조도 품질·오차 성능을 개선
  - `21년 사업 축소로 추가 개발은 중단됐으나, librealsense 등 오픈소스 SDK 지원은 유지
- (시장적용) 로봇·드론·AR/VR에서 레거시 채택 지속
  - 자율주행/장애물 회피, 3D 스캔, 제스처 인식 등 다양한 응용에서 계속 활용 (Martin, 2025.7.)
  - 물류 자동화, 무인점포, 출입통제 등에서 3D 인식 수요가 증가하며 기존 장비 활용이 지속
- (기업전략) HW 축소, SW 최적화(OpenVINO) 중심으로 재정렬
  - RealSense 신제품은 제한적이나, Edge AI/비전 최적화를 위해 OpenVINO 등 소프트웨어 지원을 강화
  - 오픈소스·스타트업 생태계가 심도카메라 개발을 이어가는 구조에서 간접 기여 형태로 유연성을 확보
- (정책/산업 시사점) 공급·지속성 리스크를 반영한 대체/다중센서 전략 필요
  - 사업 축소 사례는 공공·산업 도입에서 단종·장기 유지보수 리스크로 직결되므로, 대체 기술/공급망 옵션을 사전에 설계해야 함

- 3D 지각은 로봇·XR의 기반 기능이므로, ToF/스테레오/레이더 등 멀티모달 조합과 표준 인터페이스·검증 시나리오 확보가 중요

#### ◆ Knowles (IA611 SmartMic 중심으로)

- (기술개요) DSP 내장 MEMS 마이크 기반 always-on 오디오 센서
  - MEMS 마이크에 DSP 코어를 내장해 음성 인식 작업을 마이크 단계에서 처리하는 초저전력 오디오 프로세서(Knowles, 2018.6.)
  - 상시 대기 상태에서 키워드 검출·화자 식별을 수행하는 ‘스마트 마이크’ 개념을 상용화한 사례
- (핵심기술 구성/성능) 마이크 단계 전처리+KWS로 AP 부하·전력 절감
  - 오디오 전처리(노이즈 억제·빔포밍)와 키워드 스팟팅을 통합해 ‘깨우기 단어’를 마이크에서 인식
  - 수  $\mu$ A급 전류로 always-on 동작을 지향하며, 필요 시에만 데이터 전송해 지연·프라이버시 부담을 완화
- (개발동향) 스마트폰 채택 이후 이어버드·IoT로 확산
  - `18년 Vivo 채택 이후 다수 플래그십 스마트폰에 적용되며 화면 꺼짐 상태 음성 호출 UX를 확산
  - 최근 AI 알고리즘을 고도화한 차세대 제품을 출시하고, SDK/튜닝 도구 제공으로 맞춤형 적용을 지원
- (시장적용) 보이스 UI 확산에 따른 전 산업 적용 확대
  - 무선 이어버드 상시 음성 명령, 스마트TV 리모컨 음성 검색, 차량 음성비서 모듈 등으로 적용이 확대
  - 현장 처리형 음성 AI 수요가 증가하며 응답지연·개인정보 부담을 줄이는 구조로 채택이 확대
- (기업전략) AP 비중속성·OEM 맞춤형 협업으로 채택 확대
  - 특정 AP 종속을 낮춰 OEM이 기종/플랫폼 변화에도 일관된 음성 UX를 구현하도록 지원(Knowles, 2018.6.)
  - 주요 OEM과 맞춤형 공동개발을 통해 채택을 가속하고 모바일-IoT-차량으로 외연을 확장
- (정책/산업 시사점) 음성 센서단 처리 확대에 따라 온디바이스 우선 보안·거버넌스 중요
  - always-on 오디오는 민감 정보와 직결되므로, 온디바이스 처리·최소 전송 원칙과

함께 안전 부팅/키관리 등 온칩 보안 체계가 필수화됨

- 표준 KWS/Voice ID 품질평가, 소음 환경 검증, 개인정보 처리·보관 가이드라인을 패키지로 마련해야 산업 확산이 가능

### ◆ 삼성전자 (ISOCELL Vizion 63D ToF 센서 중심으로)

- (기술개요) 로봇·XR 지향 iToF 기반 3D 심도 센서
  - 빛의 위상차를 이용하는 간접 ToF(iToF) 방식으로 3차원 입체 정보를 측정하며 모바일 외 로봇·XR 응용으로 확장(삼성전자, 2023)
  - 정밀 3D 인식이 필요한 미래 기기의 핵심 입력(‘눈’)을 목표로 센서 포트폴리오를 산업용으로 확장
- (핵심기술 구성/성능) ISP 내장 원칩 iToF로 심도맵 생성 효율화
  - 원칩 iToF 센서에 심도 연산 최적화 ISP를 내장해 AP 지원 없이 심도맵 촬영이 가능하다고 제시
  - 근거리 정밀 인식(손동작/공간 인식, 로봇 장애물 감지, 안면인증 등)에 필요한 지연·전력·집적도를 개선
- (개발동향) 모바일 ToF 경험 기반으로 산업용 라인업 동시 전개
  - `20년대 초 모바일 ToF 센서 공급을 기반으로, 산업용 확장을 위해 63D를 개발
  - `23년 말 63D(iToF)와 931(글로벌 셔터) 동시 공개로 스마트폰 중심 구조에서 미래 산업으로 다각화(ROOS, 2023.12.)
- (시장적용) XR·로봇·인증 중심 3D 인식 수요 대응
  - XR 헤드셋에서 손동작·공간 인식, 서비스 로봇에서 근거리 장애물 감지·맵핑 등 핵심 기능 구현에 활용 가능
  - 스마트폰 3D 얼굴인식/AR, 차량 실내 모니터링(체스처) 등으로 응용 확장 가능성이 제시됨
- (기업전략) 미래 산업용 센서로 성장축 전환 및 협력 확대
  - 스마트폰 시장 성숙에 대비해 로봇·XR용 고성능 센서를 성장축으로 육성(삼성전자, 2023)
  - 조직·R&D 강화와 광학 혁신 기술(메타표면 등) 협력을 통해 차별화 요소를 확보하려는 전략
- (정책/산업 시사점) 3D 지각은 안전·신뢰 성능이 핵심—통합 성능지표와 검증체계 필요
  - 심도 센서는 미션 크리티컬 응용(로봇 안전, XR 멀미/지연)에 직결되므로, 평균 성능보다 최악 지연/지터·신뢰성 중심 평가체계가 필요

- 센서-ISP-융합 SW를 포함한 통합 솔루션화가 진행되므로, 인터페이스/메타데이터 표준과 시험·검증 시나리오를 함께 구축해야 함

**Pixelplus (PX2100/Photonic Chiplet 중심으로)**

- (기술개요) 3D 적층 패키징 기반 온센서 AI 통합 센서
  - 이미지 센서와 ISP, AI 가속기를 패키지 차원에서 적층 통합하는 Photonic Chiplet 기술로 온센서 AI를 구현
  - `24년 PX9210K를 출시하며 CMOS+ISP+AI 코어를 수직 3D 패키징한 통합 센서 솔루션을 제시
- (핵심기술 구성/성능) 데이터 이동 최소화로 지연·전력 동시 절감
  - 센서 데이터를 로컬에서 AI 연산해 전력 소모와 지연을 줄이고, 원본 전송 부담을 완화하는 구조
  - 노이즈 제거·자동노출·HDR 등 ISP 기능과 기본 객체 인식을 센서단에서 처리하도록 구성해 모듈화를 강화
- (개발동향) 열·구조 난제 대응 및 차량용 확장 추진
  - 유리 커버 구조 제약 및 ISP 발열에 따른 화질 열화 문제를 열 전도 설계 등으로 극복했다고 제시
  - 자동차용 센서 샘플(PK5130KA 등) 공개와 양산 예고를 통해 Tier-1 직납 및 차량 시장 확장을 추진
- (시장적용) 보안/차량 카메라에서 스마트홈·테스트 공급으로 확장
  - 보안 카메라·차량 카메라에서 가성비 기반 채택 경험을 보유하며, 온센서 AI로 고부가 전환을 시도
  - AI 가전·스마트 도어벨 등에서 실시간 객체 인식·사용자 식별 기반 기능을 제공하는 방향으로 응용을 확대
- (기업전략) 자동차·컨슈머 양측 고부가 라인업 확대
  - Photonic Chiplet 기반 온센서 AI 제품군을 자동차와 컨슈머 가전으로 확대해 제품 믹스를 고부가로 전환
  - 디자인하우스·파운드리 협력 및 고객 공동개발로 개발 속도·맞춤 대응력을 강화하는 전략
- (정책/산업 시사점) 첨단 패키징(3D 적층)·열관리·신뢰성 인프라가 상용화의 병목
  - 온센서 AI는 설계만으로 완성되지 않으며, 3D 적층·열해석·신뢰성/수율 확보가

경쟁력을 좌우하므로 패키징 생태계 연계가 중요

- 차량/보안 등 고신뢰 시장 진입을 위해 인증·품질 데이터·양산 검증 체계를 공용화 하면 제품화 속도를 높일 수 있음

#### ◆ Bosch (BHI260AP 중심으로)

- (기술개요) IMU+MCU 통합 초저전력 지능형 센서
  - 6축 IMU와 32비트 MCU를 하나의 패키지에 통합한 초저전력 지능형 센서로, 센서 자체의 로컬 분석을 지원(electropages, 2022.5.)
  - 걸음 인식·동작 분류 등 알고리즘을 센서 내부에서 수행해 호스트 부하·전력 소모를 줄이는 구조
- (핵심기술 구성/성능) 센서 내 ML 실행 + 개발툴(AI-Studio) 기반 배포
  - 센서가 운동 데이터를 실시간 처리·분석해 ‘의미 이벤트’만 호스트로 전달하는 방식으로 데이터 이동을 최소화
  - 사용자 정의 ML 모델 탑재를 지원하고, 개발툴을 통해 임베디드 모델 학습·배포를 가능하게 하는 플랫폼을 제공
- (개발동향) 펌웨어 업데이트·자기학습 기능으로 패턴 확장
  - `20년경 출시 후 펌웨어 업데이트로 자전거/팔 동작 등 신규 패턴 감지를 추가하며 기능을 확장
  - `21년 Self Learning AI 개념을 공개해 사용자 고유 패턴 적응을 강화하고, 환경 센서 등으로 개념 확장을 추진
- (시장적용) 웨어러블·모바일·산업 센싱에서 로컬 지각 확산
  - 스마트워치/피트니스 밴드에서 트래킹 정확도를 높이고, 스마트폰 공동프로세서 형태로 상시 모션 감지를 수행
  - 산업 분야에서는 진동 패턴 기반 예지보전 센서로 시범 적용되며, 배터리 기반 장기 운영 시나리오에 적합
- (기업전략) ‘센서+프로세서+SW’ 통합으로 솔루션 프로바이더 전환
  - 센서 칩 공급을 넘어 로컬 AI 기능과 개발 툴을 결합해 부가가치가 높은 스마트 센서 시장을 창출(embedded, 2025.2.)
  - 파트너 개발용 SDK/툴 제공으로 생태계를 묶고, 다양한 응용에 재사용 가능한 플랫폼화를 추진
- (정책/산업 시사점) 저전력 무선 IoT 확산의 핵심은 센서단 이벤트화+장기운영 검증

- 배터리 기반 장기 운영은 데이터 전송보다 현장 처리·이벤트화가 유리하므로, 로컬 ML 센서의 도입 경제성이 커짐
- 확산을 위해서는 모델 배포 표준·장기 신뢰성 평가·현장 데이터 기반 알고리즘 라이브러리 등 운영 인프라가 함께 필요

#### ◆ STMicroelectronics (ISPU 중심으로)

- (기술개요) MEMS 다이 내 DSP 집적 지능형 센서(ISPU)
  - MEMS 센서 다이 위에 DSP 코어를 집적해 센서 신호를 로컬에서 즉시 처리·분석하는 지능형 센서(ST, 2022.2.)
  - 별도 MCU 없이 필터링·FFT·ML 추론까지 센서 내부에서 수행해 지연과 전력, 데이터 이동을 줄이는 구조
- (핵심기술 구성/성능) 센서 내 신호처리+ML로 ‘의미정보’ 출력
  - 32bit RISC DSP 기반으로 원시 센서 데이터를 실시간 처리하고, 결과(제스처/이상징후 등) 중심으로 호스트에 전달
  - 단일 칩 내 센서+프로세서 결합으로 데이터 전송 지연·전력 소모를 줄이고 초소형 엣지 디바이스에 적합
- (개발동향) ISM330IS 등 제품 적용 및 개발자 툴 연계 확대
  - `22년 ISPU 발표 후 가속도 센서 등에 적용 제품을 출시(ST, 2022.6.)
  - `23년 NanoEdge AI Studio 플러그인 공개 등으로 코딩 없이 알고리즘 생성·배포를 지원하고 생태계 확산을 추진
- (시장적용) 스마트팩토리·웨어러블 이상징후 감지 등 초기 확산
  - 모터 진동 분석 기반 예지보전 센서 등에서 배터리 기반 무선 센서를 구현하는 용도로 적용이 검증 중
  - 웨어러블에서는 낙상 감지 등 이벤트 기반 기능에 적용 가능하며, 개발 키트 중심으로 도입 검토가 확대
- (기업전략) 센서 지능화를 차별화 축으로 삼아 생태계 락인 강화
  - 센서 자체가 컴퓨터라는 방향으로 일반 센서 시장의 가격 경쟁을 회피하고 고부가 스마트 센서 시장을 확대
  - MCU 생태계(STM32)와 툴체인을 연계해 모델 이식·개발 편의성을 높이고 자사 플랫폼 결속을 강화
- (정책/산업 시사점) 센서 내 연산 확산은 메타데이터 표준·경량 ML 워크플로가 성패

- 센서가 의미정보를 생성할수록 기기간 협력을 위한 표준 메타데이터 스키마·전송 규격이 중요해짐(융합·협력의 전제)
- TinyML 개발-검증-배포 체계를 표준화하고, 공정/소자/IP 표준화와 함께 대량생산 비용 구조를 낮추는 전략이 필요

## 부록 4 온디바이스 AI 반도체 기업 동향

### Apple (Vision Pro의 R1 중심으로)

- (기술개요) XR용 센서 융합 전담 저지연 프로세서
  - Vision Pro에 탑재된 전용 프로세서로 다중 센서 입력을 실시간 처리하며, M2와 듀얼 칩으로 동작(Apple-A, 2023.6.)
  - 센서 데이터 전처리·융합을 전담해 XR 핵심 품질요소(지연/멀미)를 구조적으로 해결하는 방향
- (핵심기술 구성/성능) 12ms급 저지연 목표 + 고대역 센서 파이프라인
  - 12대 카메라·5개 센서·6개 마이크 입력을 처리해 시야 지연을 12ms 이내로 제한하는 목표를 제시(Apple-A, 2023.6.)
  - 256GB/s 메모리 대역폭 등 초저지연 구조로 반응성을 강화해 몰입감과 사용성을 높임(TechCrunch, 2023.6.)
- (개발동향) visionOS/ARKit 결합으로 개발 생태계 구축
  - visionOS, RealityKit, ARKit 등 개발도구 제공으로 서드파티가 공간 데이터와 기능을 활용하도록 지원(Apple-B, 2023.6.)
  - R1 세부 아키텍처는 비공개이나, 센서 처리·ISP·전용 가속기 혼합의 특화 설계로 확장 가능성이 제시됨
- (시장적용) Vision Pro 핵심 기능 구현 및 라인업 확장 기반
  - 눈동자 추적·손짓 인식·공간 지도화 등 XR 핵심 기능을 실시간 구현하는 데 활용
  - 향후 AR/VR 기기 라인업 확장 시 유사 센서 처리 전담칩이 반복 탑재될 가능성이 큼
- (기업전략) 수직통합(칩+OS+SDK)로 경험 품질과 락인 강화
  - 칩 설계 역량과 OS/SDK를 결합해 경쟁사 대비 차별화된 MR 경험을 제공하는 전략
  - 개발도구·스토어·콘텐츠까지 연결해 생태계 락인과 사용자 경험 표준화를 동시에 추진
- (정책/산업 시사점) 센서융합 전담칩+SW 스택 결합이 XR 경쟁력의 핵심 구조로 고착
  - XR/로봇 등에서는 평균 성능이 아니라 최악 지연/지터가 품질을 좌우하므로, 센서 전담 처리 구조와 검증 지표 체계가 중요해짐
  - 확산은 하드웨어 단품이 아니라 데이터 파이프라인·SDK·개발도구에 의해 좌우되므로, 풀스택 패키지 경쟁을 전제로 산업 전략을 설계해야 함

◆ NVIDIA (DRIVE Thor 중심으로)

- (기술개요) 차량용 통합 AI 프로세서(중앙 컴퓨팅 지향)
  - 자율주행차 시대를 겨냥한 통합 AI 프로세서로 최대 2,000 TOPS 성능을 제공 (NVIDIA, 2022.9.)
  - 클러스터·IVI·ADAS 등 차량 기능을 하나의 칩에서 처리하는 중앙집중형 아키텍처를 지향
- (핵심기술 구성/성능) GPU+Arm CPU 통합 및 격리/확장 아키텍처
  - 최신 GPU 아키텍처와 Arm 기반 CPU(Grace 코어)를 통합하고 멀티 OS·MIG로 워크로드 격리/병렬을 지원
  - NVLink-C2C로 다중 칩을 확장 가능하게 설계해 차량 중앙 컴퓨터의 성능/확장성 요구에 대응
- (개발동향) Atlan→Thor 전환 및 소프트웨어 스택/시뮬레이션 강화
  - Orin 후속 Atlan을 대체해 Thor로 전환되었고, Hopper/Transformer 엔진·FP8 도입으로 추론 성능 향상을 강조(NVIDIA, 2022.8.)
  - DriveWorks SDK 및 Omniverse 기반 시뮬레이션을 제공해 OEM 개발·검증 체계를 함께 강화
- (시장적용) '25년형 양산차 목표로 OEM 채택 확대
  - 다수 OEM에 제안되었고, ZEEKR 등 고객 발표를 통해 '25년 신차 적용 계획이 공개됨(NVIDIA, 2022.9.)
  - 볼보 등 주요 완성차 업체가 차세대 모델에서 Thor 기반 중앙 컴퓨팅을 활용하는 계획을 제시(Volvo Cars, 2024.9.)
- (기업전략) HW+SW+검증 인프라로 차량 AI 플랫폼 표준화
  - SoC 공급을 넘어 레퍼런스 설계·개발 툴·시뮬레이션까지 결합해 플랫폼 지위를 강화
  - 파트너 생태계와 개발자 기반을 확대해 '플랫폼 전환 비용'을 경쟁 우위로 활용
- (정책/산업 시사점) 차량은 안전·보안·검증이 결합된 시장—스택/인증 역량이 경쟁력
  - 차량 도메인은 기능안전·사이버보안·OTA 검증이 필수이므로, 칩 성능만으로는 경쟁이 어렵고 스택/검증체계가 핵심
  - 센서 데이터 폭증 환경에서 지각-판단-제어 루프의 최악 지연/지터를 관리하는 표준 성능지표·공용 검증 인프라가 필요

◆ DEEPX (DX series AI 중심으로)

- (기술개요) 엣지용 고효율 NPU 칩셋(전성비 중심)
  - 엣지 디바이스용 고효율 NPU로 2~5W 전력에서 GPU급 AI 성능을 지향 (DEEPX, 2025.11. 접속 기준)
  - 보안 감시·로봇·스마트팩토리 등 현장 실시간 영상 분석 수요를 목표로 제품 라인업을 전개
- (핵심기술 구성/성능) 다중 모델 병렬 + 다양한 비전 모델 지원
  - 전용 DRAM 대역폭/메모리 관리 최적화로 다중 모델 동시 처리를 지원하고, YOLO~ViT 등 비전 모델 실행을 강조
  - INT8 양자화 등으로 지연을 낮추면서 정확도를 유지하고, 호스트 CPU 아키텍처 비의존 통합을 지향
- (개발동향) 라인업 확장 및 차세대 공정/생성형 추론 대비
  - DX-V1/L2/M1/H1 등 제품군을 공개하며 엣지부터 서버까지 확장 구성을 제시
  - 차세대 DX-M2를 2nm 공정으로 추진하며 경량 생성형 모델 실시간 추론까지 범위를 확장한다는 방향을 제시(eeNEWS, 2025.4.)
- (시장적용) 다채널 영상 분석·로봇 지각 등 현장형 비전 적용
  - 다수 CCTV를 동시 처리하는 관제·감시 시나리오에서 병목 완화 수단으로 제시
  - 물류/서비스 로봇에서 다중 카메라 입력을 통합 분석해 주변환경 인지 및 경로 계획을 지원하는 적용을 강조
- (기업전략) 전성비/비용 효율과 SW 스택 고도화로 상용 확산
  - 절대 성능보다 와트당 성능·비용당 성능을 강조하며 ‘모든 디바이스에 AI 탑재’ 기조를 추진
  - 컴파일러/툴체인 투자로 모델 지원 속도를 높이고, 해외 법인·전시회 등으로 영업 채널을 다각화
- (정책/산업 시사점) 전성비 NPU 확산의 병목은 ‘모델 이식·검증·배포’ 체계
  - 현장형 서비스는 에너지·비용 제약이 강하므로 전성비는 핵심이지만, 실사용 확산은 SDK/컴파일러/벤치마크·성능 검증체계가 좌우
  - 공공·산업 도입을 촉진하려면 레퍼런스 설계(보안/관제/로봇)와 함께 데이터·성능 지표·보안요건을 표준화해 도입 비용을 낮춰야 함

◆ 한화비전 (Wisenet 9 중심으로)

- (기술개요) 보안 카메라용 엣지 AI SoC(듀얼 NPU)
  - 영상 보안 카메라용 AI SoC로 듀얼 NPU 아키텍처를 채택해 화질 개선과 객체 분석을 분업 처리(한화비전, 2025.11. 접속 기준)
  - 카메라 내부에서 실시간 분석을 수행해 원본 전송을 줄이고, 현장 판단 중심의 운영 모델을 강화
- (핵심기술 구성/성능) 4K 입력+ISP/암호화+분업 NPU로 실시간 분석
  - 한 NPU는 노이즈 제거 등 화질 개선을, 다른 NPU는 객체 검출·추적 등 분석을 전담해 동시 성능을 확보
  - 저조도 ISP·영상 암호화 엔진 등을 원칩에 통합해 보안 카메라 요구 기능과 AI 처리를 결합
- (개발동향) 세대별 칩 고도화 및 'AI 팩' 모듈 제공 확대
  - Wisenet 5/7을 거쳐 Wisenet 9에서 AI 기능을 전면 배치하며 자체 영상처리칩 개발을 지속
  - 교통·리테일·산업안전 등 업종별 AI 소프트웨어 모듈(AI 팩)을 제공해 설치환경 맞춤형 분석을 지원(시큐리티월드, 2025.3.)
- (시장적용) 도시 방범·교통·산업안전 등 공공/산업 현장 확산
  - 중남미·중동 등 도시 방범 사업에 채택되어 관제 인력 부담 없이 위험상황 탐지를 지원(시큐리티월드, 2025.9.)
  - 공장·건설현장에서 위험행동 감지, 안전모 착용 여부 판별 등 산업안전 적용을 확대
- (기업전략) HW-SW 통합 및 프리미엄 전략으로 차별화
  - 하드웨어부터 소프트웨어까지 통합 개발해 신뢰성·운영 편의성을 강화하고, 업종별 솔루션 패키지로 확장
  - 가격 경쟁보다 기술 고도화·브랜드 신뢰를 기반으로 프리미엄 전략을 지향하며 통합 보안 플랫폼화를 추진
- (정책/산업 시사점) 지능형 CCTV 확산은 '원본 최소화+프라이버시/보안 기준'이 전제
  - 공공 분야 확산을 위해서는 사건/행동 등 메타데이터 중심 운영과 함께 접근통제·암호화·감사체계 등 온칩/운영 보안 기준이 필요
  - 업종별 AI 모듈(팩) 확산은 표준 성능지표·데이터 거버넌스(보관/활용)·책임소재 정립과 결합될 때 도입 저항을 낮출 수 있음

◆ NXP (자동차용 Radar SoC 중심으로)

- (기술개요) 77GHz 레이더 트랜시버+프로세서 원칩 통합 Radar SoC
  - 최신 SAF85xx/86xx는 77GHz 레이더 트랜시버와 고성능 프로세서를 단일 칩에 통합한 자동차용 반도체(ALL ABOUT CIRCUITS, 2024.1.)
  - 소형 패키지에서 고해상도 레이더 센싱과 신호처리를 동시에 수행하는 SDV 지향 센서 노드 구조를 제공
- (핵심기술 구성/성능) 모놀리식 집적+다채널 MIMO+차량 네트워킹/보안 통합
  - 28nm RF-CMOS 기반으로 RF부·Arm Cortex-M7·DSP 등을 단일 다이에 집적해 신호처리 효율을 확보(RFID JOURNAL, 2024.1.)
  - 4Tx/4Rx, ADC, 2MB SRAM 및 Gigabit Ethernet MACsec 등으로 센서 네트워크 구성과 보안 통신을 지원(EETimes, 2023.1.)
- (개발동향) 3세대 플랫폼 확장 및 OTA/센서 네트워크 대응
  - `18년 28nm RF-CMOS 레이더 트랜시버 양산 이후 세대별 플랫폼을 확장해왔고, SAF86xx로 통합도·통신 대역폭을 강화(NXP, 2024.1.)
  - 스트리밍(raw)과 스마트센서(온칩 처리) 아키텍처를 모두 지원해 Tier-1의 용도별 맞춤화를 가능케 함
- (시장적용) AEB/BSD부터 L2+/L3 ADAS까지 대량 적용
  - NCAP 기반 안전 기능부터 고급 ADAS까지 폭넓게 적용되며, 다중 레이더 구성을 통한 360도 인지에 활용
  - 글로벌 부품사가 NXP 레이더 SoC를 채택해 4D 이미징 레이더 등 차세대 레이더 솔루션을 개발 중
- (기업전략) 레거시 양산 실적 기반 ‘표준 플랫폼’ 지향
  - 레이더 양산 실적과 완성차 공동개발 경험을 바탕으로 시장 신뢰를 강화하고, 센서 네트워크 기반 SDV 흐름을 선도
  - 레이더 외 LiDAR/V2X/MCU 등과 결합한 센서 퓨전 포트폴리오로 차량 인지 전 영역 관여를 확대
- (정책/산업 시사점) 안전 규제 시장은 ‘표준 인터페이스+보안 통신+검증 시나리오’가 핵심
  - 레이더는 규제·안전과 직결되어 신뢰성/보안 요건이 높으므로, 센서 데이터·메타 데이터·전송 규격 표준화가 경쟁력으로 작동
  - SDV로 갈수록 센서 네트워크와 OTA가 필수이므로, MACsec 등 보안 통신과 기능안전 검증 체계의 공용화가 산업 확산을 촉진

◆ 넥스트칩 (Apache5 ADAS SoC 중심으로)

- (기술개요) 차량용 비전 인식 SoC(NPU 내장) 기반 ADAS 프로세서
  - 쿼드코어 ARM Cortex-A53와 NPU를 통합해 ADAS/자율주행 보조 애플리케이션을 지원(넥스트칩, 2025.11. 접속 기준)
  - 카메라 기반 인지(차선/표지판/보행자 등) 기능을 실시간 처리하는 분산형 ADAS 시장을 타깃
- (핵심기술 구성/성능) 고해상도 입력+NPU 효율 중심 실시간 추론
  - 5.7M 해상도 카메라 입력까지 처리하며 1.6 TOPS NPU를 기반으로 높은 이용 효율(95~98%)을 강조
  - AEB, DMS/OMS, FSD 등 비전 기반 기능을 실시간 구현하고 표준 딥러닝 모델 제공으로 적용을 단순화
- (개발동향) Apache4 공급 경험 기반으로 Apache5/6 로드맵 전개
  - Apache4를 일본 Tier-1에 공급한 경험을 바탕으로 Apache5에서 딥러닝 NPU를 본격 탑재(THEELEC, 2021.5.)
  - Apache6은 L3 대응을 목표로 CPU/GPU 고도화와 ViT 등 신모델 지원을 추진하는 방향이 제시됨
- (시장적용) ADAS 카메라 모듈 채택 및 유럽 규제 강화 수혜 가능
  - 전방 ADAS, SVM, DMS/OMS 등 카메라 모듈에 탑재되어 실차 적용이 진행 (emotion3D, 2022.7.)
  - 유럽 ADAS 의무화 단계 시행에 따라 수요 확대 가능성이 제시되며, 표준 모델 선택형 제공으로 중소 OEM에도 접근
- (기업전략) 유연한 SW 정책+글로벌 IP 협력으로 경쟁력 확보
  - 고객이 원하는 AI 모델을 구동할 수 있도록 SDK/툴체인을 개방하고 필요 시 알고리즘 공동개발을 지원
  - 자체 NPU뿐 아니라 글로벌 IP 기업과 협력(예: NeuPro-M)해 최신 AI 기술을 빠르게 흡수하고 시장을 확장(CEVA, 2025.4.)
- (정책/산업 시사점) 차량 도메인 확산의 관문은 ‘인증·평가데이터·레퍼런스’
  - ADAS는 기능안전(ASIL)·사이버보안·실차 검증이 필수이므로, 공용 테스트베드·평가 데이터·인증 지원이 산업 확산을 좌우
  - 모델 포팅/최적화의 표준 워크플로와 차량용 품질·신뢰성 데이터 축적이 선행될 때, 글로벌 채택 비용을 구조적으로 낮출 수 있음

---

## 저자 소개

**이성준** ETRI ICT전략연구소 기술전략연구본부 미래전략연구실 책임연구원  
e-mail: sungjun2@etri.re.kr Tel. 042-860-6536

**박정렬** ETRI ICT전략연구소 기술전략연구본부 미래전략연구실 연구원  
e-mail: jrpark16@etri.re.kr Tel. 042-860-1349

**이승환** ETRI ICT전략연구소 기술전략연구본부장  
e-mail: lsh@etri.re.kr Tel. 042-860-3876

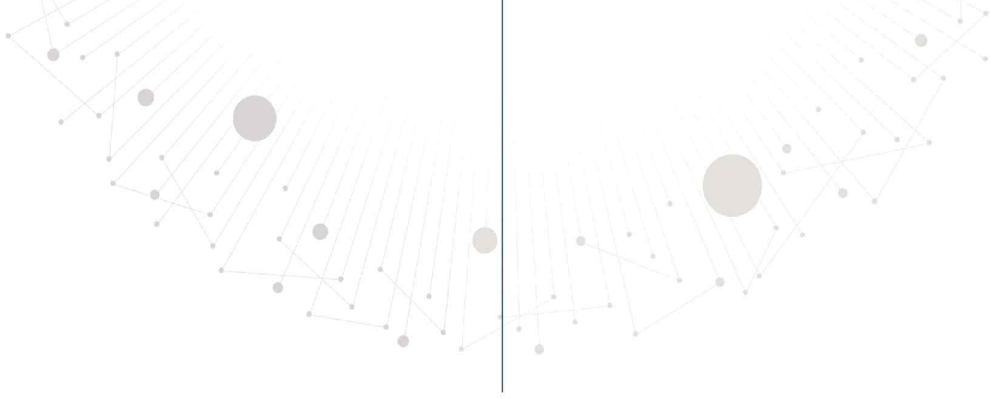
---

## 지각 특화 AI 반도체 투자 타당성 검토 : 기술 니즈와 시장 수요 중심으로

발행인 한 성 수

발행처 한국전자통신연구원 ICT전략연구소

발행일 2025년 12월 31일



[www.etri.re.kr](http://www.etri.re.kr)

본 저작물은 공공누리 제4유형:  
출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수  
있습니다.



**ETRI** Electronics and Telecommunications  
Research Institute

34129 대전광역시 유성구 가정로 218  
TEL.(042) 860-6114 FAX.(042) 860-6504

