

참고1

스마트 디바이스 분야 정부 정책 · 전략 · 계획

□ 국산 AI반도체를 활용한 K-클라우드 추진방안('22. 12월)

- NPU, PIM 등 AI반도체의 3단계* 고도화와 함께 AI반도체 소프트웨어 개발, 실증체계 구축 등 종합적 기술개발 이행안 수립

* (1단계) 국산 NPU 고도화, (2단계) DRAM기반 PIM의 고성능 저전력 구현, (3단계) NVM기반 PIM의 극저전력화

- AI반도체, 인터넷기반 자원공유, AI서비스분야 연계 과제로 K-클라우드 생태계 조성



□ AI-반도체 이니셔티브('24. 4월)

- AI-반도체 전 영역에 걸친 기술혁신과 이를 뒷받침할 지속 가능한 AI 생태계 조성 및 추진체계 구축
 - AI모델, AI반도체 및 HW·SW-서비스 등 AI-반도체 가치사슬 분야별 강점을 분석하여 AI-반도체 9대 기술혁신 과제 도출



□ 반도체 현안점검회의('24. 4월)

- AI반도체를 중심으로 펼쳐지는 글로벌 경쟁에서 반도체 기술의 초격차를 확보하고 AI G3로 도약



- **(자율주행·행동기술) 자율주행 및 자율행동체 핵심기술 조기확보로 글로벌 기술경쟁력 확보**
 - **(자율주행)** 자율주행 AI·SW, 데이터 수집·가공 등 핵심기술의 리빙랩 추진을 통한 자율주행 ICT융합기술 통합실증 및 평가 추진
 - **(자율행동체)** 자율행동체 협력지능SW·엣지AI SW·로봇핸즈 프로토타입 개발 및 자율행동체 기반 환자 안전 이송·이송 실증기술 중점개발
- **(반도체) 국내 수출 1위의 반도체 분야 경쟁력 지속, AI·빅데이터 등 수요 대응***을 위해 민간 주도가 어려운 첨단기술, 인프라 투자 강화
 - * 대규모 데이터 저장 및 처리·연산, 데이터센터를 위한 전력반도체 등
 - **(차세대·AI) 초격차 혁신선도를 위한 차세대 반도체1」 투자 확대, 국산 AI반도체 기술의 분야별 맞춤형 개발2」 및 HW·SW 풀스택 기술개발3」 지원**
 - 1」 온디바이스 AI 반도체, 화합물 전력반도체, PIM 인공지능반도체 등
 - 2」 NPU(Neural Processing Unit), PIM(Processor In Memory), 뉴로모픽(Neuromorphic) 반도체 등 AI 반도체군에 대한 맞춤 지원
 - 3」 HW인프라, 시스템 SW, AI모델, 서비스 등 가치사슬 전반의 성장을 위한 기술개발
- **(지능형반도체) AI반도체를 비롯한 초격차 선도기술개발과 소·부·장 기반 구축 등 연구생태계 강화 지원 지속**
 - **(AI반도체) 국산 AI반도체의 분야별 맞춤형 개발***과 생태계 혁신을 위한 HW·SW 풀스택 기술개발** 등에 투자 지속
 - * NPU(Neural Processing Unit), PIM(Processor In Memory), 뉴로모픽(Neuromorphic) 반도체 등 AI 반도체군에 대한 맞춤 지원
 - ** HW인프라, 시스템 SW, AI모델, 서비스 등 가치사슬 전반의 성장을 위한 기술개발

참고3

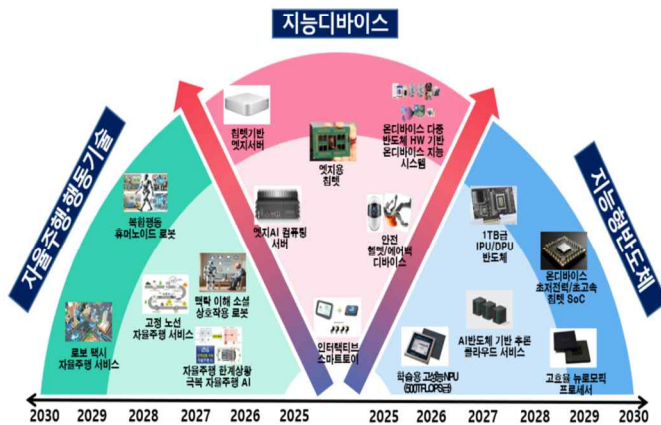
스마트 디바이스 분야 중장기 R&D 방향

미션	AI반도체 핵심기술 확보와 지능 디바이스, 자율주행·행동 기술의 개발·활용을 통한 스마트디바이스 생태계 강화		
비전	지능형 반도체와 온 디바이스 AI 기술 경쟁력 확보를 통한 AI 강국 도약		
미래상	As is (현재) <ul style="list-style-type: none"> ▶ 자율주행·행동기술의 제한적 기능 활용 ▶ 외부 서버 및 네트워크를 활용한 디바이스의 제한적 AI 처리 ▶ 데이터센터 중심의 고비용·고전력의 AI 반도체(GPU) 활용 	⇒	To Be (미래) <ul style="list-style-type: none"> ▶ 비정형·복잡한 환경에서 운영가능한 자율주행·행동 기술 ▶ AI의 디바이스 내재화를 통한 앰비언트 지능 디바이스 실현 ▶ 데이터센터, 온디바이스 등에 따라 맞춤형 고효율 AI반도체 활용
중장기 R&D 추진방향	구분	R&D 중장기 목표	R&D 추진방향
	자율주행·행동	▶ 자율주행·행동 지능 고도화·플랫폼 기술 확보	▶ 高신뢰·高안전의 자율주행·행동 지능 개발
	지능 디바이스	▶ 지능 디바이스 핵심기술 개발로 국내 시장 확보하고 글로벌 진출 마련	▶ 생성형 지능기반 디바이스 기술 ▶ 엣지 AI 컴퓨팅 시스템 기술
	지능형 반도체	▶ AI반도체 SW/HW 생태계 구축	▶ 고성능·저전력 AI반도체를 위한 HW 핵심기술 및 시스템SW 개발
기술확보 목표	자율주행·행동 기술	자율주행·행동 지능 기술 고도화	▶ 정형적 환경(고속도로 등) 내 제한적 수준의 자율주행·행동 지능을 복잡하고 비정형적인 환경(도심 등) 내 적용하도록 기술 고도화
		자율주행·행동 AI 플랫폼	▶ 특정 단일 지능 기반 자율주행·행동 지능의 한계 극복 및 멀티모달 AI 모델 기반의 범용 AI 플랫폼 기술 확보
	지능 디바이스	온디바이스 시스템 개발	▶ 국산 온디바이스 AI 반도체 및 개방형 반도체 지원과 판단형 지능에서 생성형 지능 지원으로 처리 지능규모 확장
		엣지 AI 컴퓨팅 기술 개발	▶ 온디바이스-엣지-클라우드 간 다계층·다형상의 엣지 컴퓨팅 시스템의 통합 및 연동 기술 확보
	지능형 반도체	초저전력·고성능 AI반도체 개발	▶ 지속가능한 데이터센터 기반 AI와 온디바이스 AI 구현이 가능한 전성비를 제공하는 AI 반도체 핵심기술* 확보 * 고성능 NPU, 저전력 PIM 및 초저전력 뉴로모픽 프로세서
		AI반도체 SW 생태계 구축	▶ 국산 AI반도체를 온디바이스 및 데이터센터에 효과적으로 적용하기 위한 SW 핵심기술 개발 * 반도체의 특징에 맞는 시스템 SW(컴파일러, Lib. SDK 등) 핵심기술 개발

□ 기술발전 전망과 R&D 핵심이슈

(발전 전망) AI혁명 시대의 도래에 따라 온디바이스부터 데이터센터까지 모든 스마트디바이스의 고성능·고신뢰의 지능화로 발전

<기술 발전 전망도>



<ICT R&D 핵심이슈>

✓ 자율주행·행동기술 ✓ 지능디바이스

- 자율주행·행동 지능의 자율성 강화 및 고신뢰성·고안전성 확보 등을 위해, 비지도 AI 학습 기술, 자가 학습 기술, 인간과의 상호작용 기술, 온디바이스 AI 프레임워크 및 실행기술 등 필요
- 온디바이스 AI 서비스 수요의 전 산업 확산과 다양한 온디바이스 AI 반도체 상용화, AI 워크로드에서 요구되는 고효율과 개인 정보보호 및 저지연성 해결을 통해 온디바이스 및 엣지 AI 기술 선점이 중요

✓ 지능형반도체

- 초거대 AI 성능의 핵심이 되는 컴퓨팅 자원 확보를 위해 주요국 및 빅테크를 중심으로 저전력·고성능 AI반도체 개발 경쟁 및 국가전략자산화 심화

□ R&D 추진방향

A S is



T o be

자율주행·행동기술

- 정형적 환경 내 제한적 수준의 자율주행·행동 지능
- 특정 단일 지능 기반 자율주행·행동 지능
- 자율주행·행동 기술의 시장 대비 낮은 신뢰성과 안정성 부족



- 복잡하고 비정형적 환경 내 고도화된 자율주행·행동 지능
- 멀티모달 AI 모델 기반 범용 자율주행·행동 AI 플랫폼 기술
- 자율주행·행동 지능의 기술적 신뢰성 확보 및 안정성 강화

지능디바이스

- 클라우드 기반 외부 서버 및 네트워크를 활용한 디바이스의 제한적 AI 처리



- AI의 디바이스 내재화를 통한 애플리케이션 지능 디바이스 실현
- 디바이스 근접 로컬화를 통한 최적화된 지능 제공

지능형반도체

- 연산성능 부족, 전력 과소비 문제로 데이터센터 일변도의 AI서비스 제공
- * (현재 수준) 데이터센터 국산화율 1%



- 데이터센터, 온디바이스 등 다양한 AI서비스 복잡도에 따른 맞춤형 AI반도체 지원
- * (6년후 수준) 데이터센터 국산화율 20%

□ 주요 마일스톤(스마트 디바이스)

		1단계			2단계	
구분		2025	2026	2027	'28년~'30년	
연도별 기술 로드맵	자율주행 기술	서비스	고정 노선 자율주행 서비스			로보 택시 자율주행 서비스
		제품	자율주행 한계상황 극복 자율주행 AI			자율주행 고신뢰 AI 핵심기술 및 자율주행 평가 검증 국가 표준
	자율행동 기술	서비스	자율행동지능 파운데이션 모델 기반 대화형 로봇 서비스			자율행동지능 파운데이션 모델 기반 복합행동 휴머노이드 서비스
		제품	맥락 이해 소셜 상호작용 로봇			복합행동 휴머노이드 로봇
		서비스	자율주행 한계상황 극복 자율주행 AI			고신뢰 자율주행 인증·평가 체계
		제품	자율주행 한계상황 극복 자율주행 AI			고신뢰 자율주행 인증·평가 체계
구분	2024	2025	2026	2027	'28년~'30년	
온디바이스 시스템	서비스	온디바이스 판단 지능 서비스	AI 반도체 가속 중심 온디바이스 판단/생성 지능 서비스			다중 반도체 융합 가속 중심 온디바이스 판단/생성 지능 서비스
	제품	인터넷 독립 및 연결 지능 디바이스	온디바이스 AI 반도체 HW기반 온디바이스 지능 시스템			온디바이스 다중 반도체 HW 기반 온디바이스 지능 시스템
	부품 소재 장비		온디바이스 AI 반도체 중심 지능 디바이스			

연도별 기술 로드맵	엣지 컴퓨 팅	서 비 스	산업용 엣지 컴퓨팅 서비스		지능형 모빌리티 디바이스를 위한 저전력·다계층 엣지 서비스		마이크로 엣지 데이터센터를 통한 초저지연 엣미엔트 AI 서비스		
		제 품	러기드 엣지서버	엣지서버 가속기		엣지AI 컴퓨팅 서버	엣지서버용 칩렛	칩렛기반 엣지서버	
		부 품 소 재 장 비 ★							
	신산 업서 비스	서 비 스	청각장애인대 화 지원 서비스	아동용 감정인식 교육 도구	위험상황 알림/예방 서비스	산업현장 안전사고 방지 서비스	3D프린팅 고정밀·초대 형출력 서비스	비침습 연속 건강관리 서비스	정신건강 관리 서비스
		제 품	청각장애인 의사소통 AR 안경	인터랙티브 스마트토이	위험상황 인지 디바이스	안전 헬멧/에어백 디바이스	3D프린팅 고정밀·초대 형출력 디바이스	비침습 연속 생체신호 측정 디바이스	정신건강 디지털 테라피 디바이스
	지능 반 도 체	서 비 스		AI반도체 기반 추론 클라우드 서비스			오픈소스 EDA SW 서비스	AI반도체 기반 학습 클라우드 서비스	
		제 품	엣지용 추론 NPU (80TOPS급)	학습용 고성능 NPU (500TFLOPS급)	온디바이스/데이 터센터 거대모델 개발 도구		고효율 뉴로모픽 프로세서	AI반도체 융합 연산 메모리 용 SoC	2TB/s급 직렬 인터페이 스
		부 품 소 재 장 비							
		서 비 스							
		제 품							
		부 품 소 재 장 비	100Gbps급 초고속 인터페이스	CXL(Compute Express Link) 3.1 올인원 프레임워크		AI반도체 칩파워공급	온디바이스 초저전력 칩렛 SoC	AI반도체 소자회 로 동시개 발	온디바이 스 초저전력 /초고속 칩렛 SoC
									

1. 자율주행·행동기술, 지능형 반도체

□ 글로벌 동향

- (규모) 스마트 디바이스 분야 시장은 '22년 2,679억 달러에서 연평균 약 20.5% 성장률로 '30년 약 1조 1,882억 달러 규모의 시장* 형성 전망
 - 국내 시장은 '22년 21.4조 원에서 연평균 16.7% 성장을 통해 '30년 73.5조 원 규모의 시장을 형성할 전망
 - * (자율주행/행동) 세계/국내: 1조4,411억 달러(CAGR 28.5%)/42조원(CAGR 14.9%)
 - * (지능디바이스) 세계/국내: 8,963억 달러(CAGR 18.9%)/20조원(CAGR 16.8%)
 - * (지능형반도체) 세계/국내: 2,774억 달러(CAGR 26.5%)/11.4조원(CAGR 26.5%)
- (시장) 스마트 디바이스분야 시장 선점과 AI 등 기술적 우위 원천 기술 확보 등을 위한 막대한 투자 전개
 - (자율주행·행동기술) 자율주행 레벨3 기술개발 본격화와 함께 로봇택시 등의 서비스가 확대되면서 AI기반 자율 모빌리티 경쟁 가속화 추세
 - * 자율주행 기술 종합순위로 웨이모, Ford Autonomous Vehicles, 지엠 크루즈, 바이두, 인텔 -모빌아이, 현대애틀브 순으로 제시(내비건트 리서치, 2020)
 - ⇒ 테슬라 등 자율주행 HW·SW 전반을 개발·생산하는 업체뿐만 아니라 엔비디아도 로봇용 AI플랫폼 '그루트'를 공개하며 시장 참여
 - (지능형반도체) GPU의 독점력에 대항하기 위한 反 NVIDIA 연합 전선이 형성, AI기술 확산을 위한 반도체 관련 기술에 관심 상승
 - ⇒ (오픈소스) NVIDIA CUDA 독점에 대항하기 위해 글로벌 클라우드社, 칩 벤더 중심으로 신규 오픈소스 구축 프로젝트 UXL* 추진
 - * 리눅스 재단 산하 컨소시엄으로 인텔, 삼성전자, ARM, 구글 등이 창립멤버로 활동 ('23.9 설립), 다양한 HW 플랫폼에서 실행가능한 오픈소스 SW 개발
 - ⇒ (칩렛) 초미세공정 도입으로 인한 비용 증가 및 수율 하락으로 인해, 무어의 법칙의 영속성을 위해 칩렛 기술의 성장 전망
 - * '23년 \$6.5B → '28년 약 \$148B로 성장전망(86.7% CAGR) (Marketandmarkets, '23)

⇒ (온디바이스) AI의 전 산업 확산으로 기기 자체에서 AI서비스를 제공하는 온디바이스 AI가 부각되며 AI반도체 신시장 창출

* '26년 모바일의 30%(3.9억대), PC의 60%(1.8억대) 등에 온디바이스 AI 탑재 전망(삼성증권)

○ (기술) 생성형 AI 기반 온디바이스 중심의 자율행동체 개발 및 경량화·저전력·고성능 AI 반도체 투자 확대

- (자율주행·행동기술) 생성형 AI 탑재를 통한 자가학습→자율행동으로 이어지는 복합 명령 수행 중심의 '온디바이스 AI 중심 자율행동체' 개발 확대

* 자율주행에서 데이터의 효율적인 학습, 멀티모달 데이터 처리, 상황 인지 및 예측 등 상호작용 기술로 발전 전망

⇒ (자율주행) 자율주행 성능을 개선하기 위해 데이터셋 확보 기술과 수집된 데이터를 효율적으로 처리할 인공지능 학습 인프라의 투자 활발

* (Tesla) 자율 시스템인 오토파일럿 기술 100만대 이상 차량 탑재('24), FSD(Full Self-Driving) 자율주행 기술 출시('23) 및 슈퍼컴퓨터 도조(Dojo) 생산 시작

* (Google) 캘리포니아, 애리조나, 네바다, 피닉스 등에서 자율주행차 테스트 및 총 2,000대 이상의 자율주행차를 운영('23.8)

⇒ (행동기술) 테슬라(美 옵티머스), 피겨AI(美 피겨02), 유니트리(中, G1·H1) 등을 중심으로 고도화된 AI기술 적용된 휴머노이드 로봇 개발이 활발히 전개 중


* (Tesla) 중량을 낮추면서 보행 속도를 높였고, 더욱 유연한 동작을 구사하는 'Optimus 2' 공개('23.12)

* (Amazon) 카디날이 갖고 있던 문제점을 개선한 인공지능 기반 자율행동체인 스페로우 발표('23.11)

- (지능형반도체) 천문학적 자금 투입을 통한 초거대 AI 개발 경쟁을 지속하는 가운데, 경량화 AI모델, 저전력·고성능 AI반도체 개발 지속

⇒ (자체 AI Chip) NVIDIA GPU의 고가, 품귀 문제 해결을 위해 주요 칩 벤더의 차세대 버전, 빅테크의 자사 맞춤형 AI반도체 개발 지속

< AI, 빅테크, 칩 벤더 AI반도체 개발 현황 >

 OpenAI	 Meta	 Microsoft
○ AI칩 개발을 위해, 7조 달러 규모 투자유치 계획('24.2) * 브로드컴·TSMC와 협력	○ 2세대 MTIA 발표('24.4) * Meta Training and Inference Accelerator	○ Arm 기반 CPU Cobalt100, Maia100 가속기 공개('23.11) * 내부 시스템용으로 사용 중
 amazon	 AMD	 intel
○ 추론용 인퍼런스아2('22.11), 학습용 트레이니움2 공개('23.11)	○ MI325X GPU 공개('24.10) * HBM3E 탑재 총 288GB 용량	○ Gaudi 3 출시('24.4) * H100 대비 학습/추론 1.5~2.3배 향상

⇒ (Full-Stack) 초거대 AI 기업들은 자사 중심의 AI생태계 지배를 위한 Full-Stack(AI모델+반도체·클라우드·서비스) 전략 구사 지속

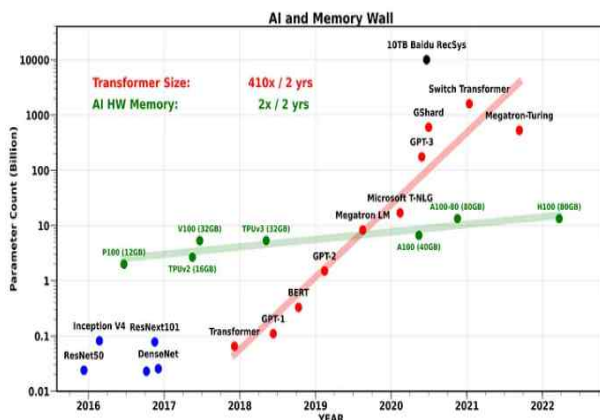
* (Google) TPUv5(AI반도체), 제미니(AI모델), 구글 클라우드(클라우드), 이마젠 3(AI서비스)
 (MS) Maia(AI반도체), MAI-1(AI모델), Azure(클라우드), Copilot(AI서비스)

< AI반도체 기반 AI생태계 >

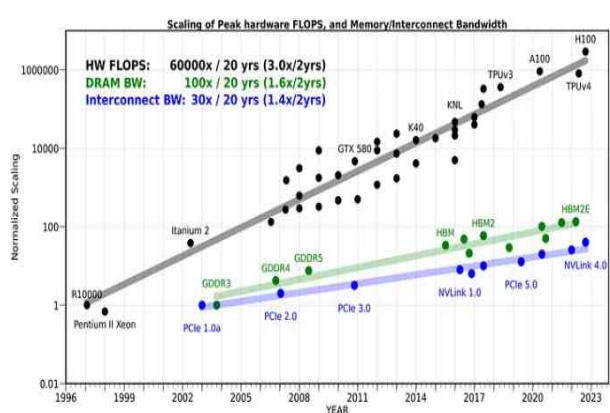


⇒ (메모리/인터페이스) AI모델 크기, GPU의 연산 성장속도에 비해, 메모리 용량 및 인터페이스의 대역폭 성능개선이 더뎠다, 이에 대한 중요성 부각

< AI모델과 Memory 발전속도 >



< GPU와 Memory 대역폭/인터페이스 발전속도 >



* AI모델 크기가 2년 동안 400배 이상 커지는 동안 메모리 발전은 2년에 2배로 발달

** 연산기 속도는 20년 동안 6만배 빨라졌으며, DRAM은 100배 커진데 반해, 대역폭은 30배 증가에 그침

⇒ (칩렛)* 기존 단일 칩 설계 기반의 반도체 설계 및 제조 비용 증가로 칩렛 기술이 미래 반도체 설계 패러다임으로 주목

* (칩렛) 시스템 반도체를 하나의 반도체로 생산하는 것이 아니라, 여러 모듈로 분할 생산한 다음 고성능 인터커넥트로 연결하여 하나의 큰 칩처럼 동작하게 하는 기술

** (NVIDIA) 단일 칩 중심으로 설계하였으나, B200 출시를 시점으로 칩렛 기술 도입 ('24), (Apple) M1 Max 칩 2개를 칩렛으로 붙여 M1 Ultra 제작('22)

⇒ (UCIe) 칩렛을 위한 칩 간 고속 통신 표준으로, 패키징 규격, 전송 속도, 다중 프로토콜, 오류 보정 등을 스펙에 포함, IP 개발 진행 중

- * (UCle 컨소시엄, Universal Chiplet Interconnect Express) 칩 간 연결 표준화를 위해 설립('22.3), 삼성전자, SK하이닉스, Arm, 퀄컴, 구글 등이 참여, 현재 UCle 1.1 스펙 마련
- * (해외) 시놉시스, 케이던스, 알파웨이브, (국내) 켈리타스, 오픈엣지 등이 일부 UCle PHY, 컨트롤러 IP 개발 및 제공

⇒ (뉴로모픽) 활용가능한 AGI 구현을 위해, 전력소모 등을 고려할 때, 초저전력 뉴로모픽 반도체 등 차세대 AI반도체 개발도 진행

- * (Intel) 뉴로모픽 기술 선점을 위해 Intel은 '로이히 2' 기반 '할라 포인트' 대규모 뉴로모픽 시스템 구축(초당 2경번 연산 지원, 와트 당 8비트 연산 15조번 가능)
- * (KAIST) SNN과 DNN을 선택적으로 사용해 트랜스포머 모델연산을 지원하는 상보형 트랜스포머(Complementary-Transformer) 개발('24.3) (A100 대비 약 600배 전력효율)

□ 국내 동향

- (자율주행·행동기술/R&D) 완전한 자율주행 수준의 서비스 제공을 위한 기반시설 마련 등을 포함한 '2030 미래자동차산업발전전략' 발표('19.11)

- * 자율주행기술개발혁신사업('21~'27, 1,650억), 복합지능자율행동체SW핵심기술개발사업('24~'27, 295억)

- (지능형반도체기술/R&D) 정부는 AI-반도체 육성 정책의 수립·이행을 통해 우리나라 AI-반도체 산업성장 및 경쟁력 강화를 위한 노력을 경주

< 정부 주요 정책 >



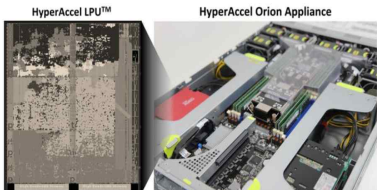


반도체 현안 점검회의('24.9) 주요내용
우리나라가 AI기술에서 G3로 도약하고 미래 AI반도체 시장을 석권하기 위해 'AI-반도체 이니셔티브'를 추진하고 '국가 AI위원회'를 대통령 직속으로 설치


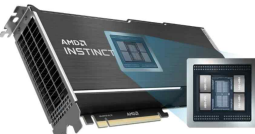
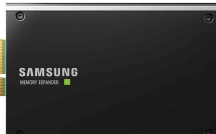
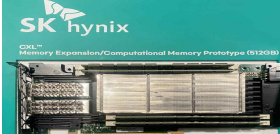
- 특히, 국산 AI반도체 기술을 단계별 고도화(NPU→PIM→뉴로모픽)하고 AI반도체 생태계 혁신을 위한 HW·SW 풀스택 기술개발 추진

- * 차세대지능형반도체('20~'29, 1조 96억원), PIM인공지능반도체('22~'28, 4,027억원)
- ** 데이터센터 인프라/HW, 컴퓨팅SW, 특화 클라우드 개발을 위한 K-클라우드 예타('25~'30, 4,031억원)

- (자율주행·행동기술시장 · 주요기업) 생성형 AI 확대로 자율주행/행동체의 학습 등에 인공지능을 적용한 기술 개발이 진행
 - * (자율주행) 현대차는 자율주행 기업 '모셔널'에 1조3000억원 자금 투자
 - * (자율행동체) AI를 탑재한 서비스 로봇(볼리(삼성전자), 클로이(LG전자)) 출시 및 로봇용 SW인 ARC(네이버랩스) 개발 등 진행
- (현대자동차) 미국, 영국 등 자율행동체 브랜드 X-ble 상표등록('23.1), 딥엑스와 MOU 체결('23.3) 및 사족보행 로봇 '스팟'이 재규어 전기차 공장 순찰, 장비검사 등에 배치('24.10)
- (통신사) 국내·외 주요 기업과의 협력을 통한 기술개발 및 서비스 출시
 - * (SKT) 퀄컴, 인티그리티와의 협력으로 개방형 로보틱스 데이터 플랫폼 개발 계획 발표('23.2), 뉴빌리티, SK실더스와 공동으로 '자율주행 AI순찰 로봇' 공동 개발 및 상용화 추진('23.3)
 - * (KT) 구글 안드로이드 기반의 'AAOS' 표준을 준수한 내비게이션을 포함 서비스를 고객에게 유연하고 확장성 있게 제공하는 'KT IVI 플랫폼' 개발
- (삼성전자) 지속적으로 사용자의 패턴을 학습해 진화하는 자율행동체 '볼리' 공개('24.1) 및 레인보우로보틱스에 590억원 투자 결정('23.1)
- (네이버) 자체 개발한 고정밀 공간 데이터 '오픈데이터셋'의 누적 다운로드 횟수가 2년간 8,500건 이상 달성('23.7)
- (지능형반도체시장 · 주요기업) 국내 AI반도체 기업은 모바일 등에서 일부 상용화, 데이터센터 부분은 자체 시범 적용 후 사업 본격화
 - (NPU社) 비전·음성 AI모델 이후 HBM을 탑재하여 LLM 모델을 타겟한 고성능 후속 버전 AI칩 개발, 서버 및 랙 솔루션 개발 중

퓨리오사AI	리벨리온	하이퍼엑셀
▶ HBM3 2개(48GB) 탑재한 2세대 'RNGD'를 Hot Chips에 발표(512TFLOPS, '24.8)	▶ HBM3E 4개(144GB)와 칩 4개를 집적한 '리벨-쿼드' '25년 발표 예정	▶ 16개 LPU(Latency Processing Unit)기반의 서버 솔루션 오리온 개발('23.11)
		

- (CSP社) 자사 파운데이션 모델 기반 생태계 확장을 위해 AI칩 개발사와 지속적인 협력 모색
 - * (KT클라우드) 자사 초거대 AI '믿음' 경량화 버전 서비스를 위해 '아톰' 버전부터 협력한 리벨리온에 KT그룹 차원에서 '리벨' 개발에 330억 투자('24.1)
 - * (네이버클라우드) AI 공동연구센터(NICL: NAVER Cloud Intel Co-Lab)를 설립하며 Intel 가우디 3기반으로 하이퍼클로바X 중심의 AI생태계 구축 노력
- (메모리社) DDR, HBM을 넘어 HBM-PIM과 CXL, MRDIMM 등 제2의 HBM, 차세대 메모리 제품 개발에 매진
 - * (SK하이닉스) ① PIM : GDDR6-AiM('22) 이후 가속기 카드 형태의 AiMX 시스템 개발('23.9) ⇒ GPU 대비 LLM 응답속도가 13배 향상, 전력 소모량 17% 낮은 성능 달성
 ② CXL : CXL 2.0 지원 96GB D램 샘플 개발('22.8), 머신러닝 및 데이터 필터링 연산기능을 포함한 메모리 솔루션 CXL CMS 2.0 발표('23.10)
 - * (삼성전자) ① PIM : AMD와 HBM-PIM 탑재한 GPU 'MI100-PIM' 성능을 ISSCC 2024에서 발표 ⇒ 기존 HBM 대비 GPU 성능 2.55배, 에너지 효율 2.67배 향상
 ② CXL : 업계 최초 CXL 2.0지원 128GB D램 개발('23.5), CMM-D, CMM-DC(CXL Memory Module DRAM Compute), CMM-H, CMM-HC(Hybrid Compute) 제품 출시 예고

PIM (Processing In Memory)		CXL (Compute eXpress Link)	
AiMX (SK하이닉스) 	MI100-PIM (삼성전자) 	CXL D램 (삼성전자) 	CMS (SK하이닉스) 

2. 지능 디바이스

□ 글로벌 동향

- (규모) 지능디바이스 세계시장 규모는 '22년 2,237억 달러에서 연평균 16.8%로 성장하여 '30년에는 8,964억 달러 규모를 달성할 것으로 전망
 - 지능디바이스 국내시장 규모는 '22년 5,805억 원에서 연평균 16.8%로 성장하여 '30년에는 2조 42억 원 규모로 성장할 것으로 전망
 - * (온디바이스시스템) 세계/국내: 990억 74백만억 달러(CAGR 29.8)/1.094조원(CAGR 14%)
- (시장) 온디바이스 AI 시장은 스마트폰, 자동차, 드론 등 다양한 기기에 탑재되어 신산업을 이끄는 차세대 AI 시장으로 성장 전망
 - (패러다임 전환) 서버 중심에서 온디바이스 중심으로, AI 데이터 학습에서 AI 추론(서비스)으로의 시장 패러다임 전환
- (기술) AI 워크로드에서 요구되는 고효율화, 개인 정보보호 및 저지연성 해결을 위한 온디바이스 및 엣지 AI 기술 선점의 중요성 확대
 - (AI 추론 최적화 기술) 범용적인 환경에서 동작이 필요한 온디바이스 AI 추론 최적화 기술 개발은 기술적 난이도가 높고 인프라적 성격이 강해 글로벌 대기업이 생태계 구축의 일환으로 기술 선도
 - (온디바이스 타겟 생성형 AI 기술) 메타, 애플, 테슬라 등 여러 기업들이 온디바이스에서 생성형 AI 지원을 위해 기술 개발 추진

□ 국내 동향

- (기술/R&D) - 「AI-반도체 이니셔티브」 이행을 위해 초기시장 공략 11대 분야 K-온디바이스 AI 플래그십 프로젝트 추진 중

< 정부 주요 정책 >

‘인공지능 G3’, K-반도체 새로운 신화창조를 위한 AI-반도체 이니셔티브(‘24.4)

유형	분야	주요 내용
민간 주도형	자동차	SDV, 자율주행에 특화되어 주행 안정성 향상(차량·사람과의 충돌 회피) 등
	자율행동로봇	간병·돌봄 등 맞춤형 온디바이스 AI서비스(상황 인지, 자율이동, 상호작용) 구현
	드론	데이터통신 표준 기반 드론 간 공중 충돌 회피 등 비행 고도화 시스템
	XR	산업 현장 작업 가이드·검사, 주변 환경·생활패턴 인식 등 맞춤형 서비스디바이스 개발
	지능형 홈·가전	개인정보보호가 강화된 지능형 홈 실증 및 온보드 AI모듈(칩셋 등) 개발
	웨어러블·헬스케어	착용자 안전(예: 산업안전, 일상생활 등) 보장 디바이스(예: 스마트렌즈·조끼·헬멧 등) 행동 패턴, 유전정보 분석을 통해 사용자의 질병 등 예측·진단·치료 디바이스
	디지털 제조	제조 현장 실시간 데이터 처리 기반 제품 품질관리, 부품 고장 진단 등
	유통·물류	24/7 고객 응대 및 수요예측을 통한 주문·발주 효율화, 배송 최적화 등
공공수요 연계형	에너지	전력소비 요인 확인, 전력수요 예측 등(한전·가스공사 등)
	CCTV	실증·성능평가 및 국산 AI반도체 적용 제품 공공구매 활성화(조달청 혁신제품 지정)
	국방	국방 반도체 목표 성능·수량 발표(방사청) 및 시범사업 추진

- (시장·주요기업) 온디바이스 AI 기술의 다양한 산업에 적용 확대로 지속적인 성장이 예상되고 있음
 - (삼성전자) 갤럭시 S24 시리즈 등 온디바이스AI 기반 실시간 통역 통화 기능을 제공하며 AI 모델 경량화를 위한 지식 증류 방식과 양자화 기술을 적용
 - (제텍에이아이) 기존 딥러닝 모델을 온디바이스 환경에서 동작할 수 있도록 경량화 도구 및 추론 가속 파이프라인을 포함한 제텍 멜란지 출시
 - (하드커널) AI 추론 성능 최적화를 위해서 자사 단일보드 컴퓨터에 복수 개의 Hailo 가속기를 장착하고 병렬화 수행

□ 2026년 중점 기획방향

- (자율주행·행동기술) 차량통신시스템 기술, 개방형 통신 인프라 기술을 개발하고 극혼잡·장거리 모빌리티 원격고속주행 자율주행 응용서비스로 자율협력 주행 안전 서비스 구현
 - (자율주행) 차세대 자율주행 AI SW 기술 및 데이터 공개를 통한 국내 자율주행 서비스 수준 향상 및 개방형 생태계 조성
 - * 자율주행 E2E 아키텍처, 데이터 자동화 파이프라인, Physics AI 기반 폐쇄루프 E2E 시뮬레이션, 멀티모달 파운데이션 등 중점 개발
 - (행동기술) 온디바이스 AI반도체 기반 자율행동체 온디바이스 AI 공통 적용 기술을 개발하고 응용 실증기술개발
 - * 자율행동체를 위한 온디바이스AI반도체 특화 라이브러리 기술개발 및 온디바이스 sLLM/SLM·VLA(Vision Language Action)·VLM(Vision Language Model) 모델개발 중점추진
- (지능형 반도체) 고성능·저전력 AI프로세서, 차세대 메모리 중심 지능형시스템, 통합형 클라우드 SW플랫폼을 연계, AI반도체 생태계 혁신을 위한 HW·SW 풀스택 기술개발 지속
 - (AI프로세서) 고대역폭·저지연 인터페이스, 고성능 메모리 설계 온디바이스 AI구현 기술 등 AI HW 핵심 요소 개발
 - * AI가속기 확장성을 위한 초고대역폭, 저지연 직렬 인터페이스 기술, 지능형 에이전트 구현을 위한 온디바이스 AI 행동·공간 학습용 AI반도체 기술 확보 등
 - (시스템SW) AI반도체 클라우드 연계 SW스택, 프로파일링, 집단 통신 최적화 등 운영효율화 컴퓨팅 SW 기술개발
 - * AI반도체 멀티 디바이스 최적 활용SW, 이종 AI반도체 간 SW성능 균질화 및 클라우드 네이티브 기반의 이종 AI반도체 통합 플랫폼 SW기술개발
 - (지능형시스템) PIM 아키텍처 기반 고성능 시스템 구현을 위한 메모리 설계, 고속 인터커넥트 등 HW 레벨 시스템 통합 기술개발
 - * PIM 아키텍처 기반의 고대역폭 송수신 인터페이스, 회로 설계 자동화, 초고속 인터페이스용 내장 메모리 회로 개발

○ (지능디바이스)

- AI반도체 및 AI모델 기술을 AI보드에 접목하여 제한된 성능에서도 AI를 구동할 수 있는 온디바이스 AI 핵심기술 확보
 - * 온디바이스AI 응용 서비스 개발 위해 필수적인 온디바이스AI 개발 보드(소형AI컴퓨터) 기술 개발
- 지능형 디바이스에 최적화된 고효율 무선 충전 적용을 위한 공용플랫폼* 개발 및 상용화 검증 지원
 - ※ 무선충전 기술(회로, 코일 펌웨어 등)의 쉬운 성능검증을 위한 모듈화된 솔루션
- * 온디바이스 자율무선충전 레퍼런스 솔루션 개발을 통한 소재부품 검증 지원 및 생태계 확산 지원 인프라 조성

□ 2026년 투자계획

(단위:백만원)

구분		'25년 예산	'26년 예산	비고
세부사업	내역사업			
자율주행을위한 차량통신협력기 술개발	자율주행을 위한 차량통신협력 기 술개발	-	5,625	-
딥러닝기반차세대자율주행 아키텍 처기술개발	딥러닝기반차세대자율주행 아키텍 처기술개발	-	9,500	-
자율주행온디바이스 응용지원핵 심기술개발	자율주행온디바이스 응용지원핵심 기술개발	-	6,000	-
PIM인공지능반도체 핵심기술개 발(설계)	PIM 설계기술	33,526	19,694	AI프로세서
	PIM 혁신기반기술	6,732	6,120	지능형시스템
	사업단운영비	1,200	1,200	-
차세대지능형반도체 기술개발(설계)	인공지능프로세서	12,354	10,353	AI프로세서
	초고속인터페이스	2,520	5,152	지능형시스템
	지능형반도체SW	4,704	2,807	시스템SW
	사업단운영비	690	630	-
AI반도체를 이용한 K-클라우드 기술개발	AI반도체 데이터센터 인프라 및 HW	10,800	17,201	AI프로세서
	AI반도체 데이터센터 컴퓨팅SW	17,000	27,800	시스템SW
	AI반도체 특화 클라우드	7,624	14,017	시스템SW
	사업단운영비	1,196	2,656	-
인공지능반도체SW통합플랫폼기 술개발	인공지능반도체SW통합플랫폼기 술개발	7,733	7,733	시스템SW
거대인공신경망인공지능반도체 SW기술개발	거대인공신경망인공지능반도체SW 기술개발	5,330	5,330	시스템SW
AI반도체기반데이터센터고도화 선도기술개발	AI반도체기반데이터센터고도화선 도기술개발	10,000	10,000	지능형시스템
AI반도체첨단이종집적 기술개발	AI반도체첨단이종집적 기술개발	8,250	7,425	지능형시스템
칩렛기반저전력온디바이스AI반 도체기술개발	칩렛기반저전력온디바이스AI반도 체기술개발	4,000	5,330	AI프로세서
통신융합합물반도체연구파운 드리기술개발	융합물반도체 인프라구축	13,835	6,900	지능형시스템
	파운드리 기술개발 및 지원	1,500	1,500	지능형시스템
(신규) 온디바이스 AIE이전트용 AI반도체 선도기술개발	온디바이스 AIE이전트용 AI반도 체 선도기술개발	-	6,000	AI프로세서
국산 온디바이스 AI 싱글보드 시스템 기술 개발	국산 온디바이스 AI 싱글보드 시 스템 기술 개발	-	4,500	'26년 신규
ICT융합 지능 디바이스 개발 활 성화 및 확산 기반구축	ICT융합 지능 디바이스 개발 활성 화 및 확산 기반구축	2,000	5,400	

* '26년 예산은 신청금액으로, 추후 변경 가능