nipa 글로벌ICT포털

ICT Global Market Analysis

품목별 ICT 시장동향

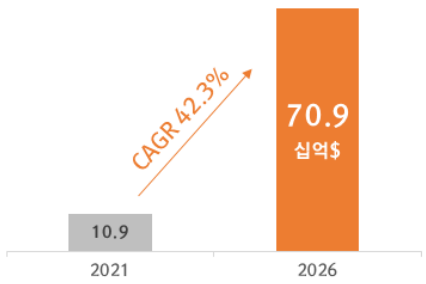
AI반도체

nipa 정보통신산업진흥원

CONTENTS

(2021.7 ~ 2022.6) AI반도체 품목 동향

▸ (2019~2022) AI반도체 발전 현황



Point ① AI반도체 산업 2030년까지 연평균 42.3% 성장

- 2026년 기준 시장규모 709억 달러 전망

Point ② AI반도체 산업 발전현황

- (2019~2020) 글로벌 빅테크 기업 중심으로 AI 반도체 경쟁 치열
- (2021~2022) AI 규제 제안 및 미디어 분야 인공지능 도입 중요성 대두

▸ (2022) AI반도체 선도 기업

Point ① AI반도체 시장 진출에 적극

- NVIDIA : AI반도체 칩 개발 관련 대규모 확장 추진
- Cerebras Systems : AI반도체 칩 'Cerebras WSE-2' 출시

Point ② 고급 AI반도체 솔루션

- SambaNova Systems: 고성능, 고정밀 AI 하드웨어-소프트웨어 프로그램 시스템 개발

▸ (2021.7 ~ 2022.6) 주요 급성장 AI반도체 기술 키워드

1위	고대역폭 메모리
2위	심층신경망
3위	명령어 집합 구조
4위	신경망처리장치
5위	양자컴퓨팅

- AI반도체 수요, HBM 성장 주도
- 심층신경망 기술 AI반도체에 응용
- AI반도체 업계, 오픈소스 명령어 세트 '리스크파이브'에 주목
- NPU, AI 연산에 유용
- AI 기술, 양자 컴퓨팅으로 비약적 도약 가능 전망

▸ (2021.7 ~ 2022.6) 주요 급성장 AI반도체 수요처 키워드

1위	미디어
2위	물류
3위	에너지
4위	리테일
5위	공공

- 엔비디아, AI 3D 모델링 'NeRF' 기술 공개
- 물류 분야 AI반도체 수요 증가 추세
- 인도공과대학, 초저전력 AI반도체 개발
- 각국 공공분야 AI반도체 활용 프로젝트 실시
- 리테일 관련 AI반도체 기업에 투자 활발

품목 개요

1. AI반도체 발전 현황
2. AI반도체 시장 규모
3. AI반도체 선진국가
4. AI반도체 신흥국가

Ⅰ. 품목 개요

1. AI반도체 발전 현황

(2019~2020) 글로벌 빅테크 기업 중심으로 AI 반도체 경쟁 치열

- 2019년 화웨이(Huawei), 알리바바(Alibaba), 알파벳 GV(Alpahbet GV), 바이두(Baidu), 삼성전자(Samsung Electronics) 등 글로벌 반도체 대기업을 중심으로 AI반도체 생산에 돌입함
- 2020년 AI 반도체 시장의 경쟁이 치열해지는 양상을 보임. 퀄컴(Qualcomm)이 AI반도체를 출시했으며, 구글(Google)은 AI반도체을 설계함. 엔비디아(NVIDIA)가 AI반도체 시장 선두를 달리고 있으며 그 뒤를 인텔(Intel)과 그래프코어(GraphCore)가 쫓고 있음

(2021~2022) AI 규제 제안 및 미디어 분야 인공지능 도입 중요성 대두

- 2021년에는 AI반도체 시장이 더욱 활성화되는 양상을 보임. 빅테크 기업들은 자체 칩 개발을 서두르고 있으며 바이두는 2세대 쿤룬 AI(Kunlun AI) 칩 양산을 돌입함. 또 이스라엘 AI반도체 제조업체 Hailo는 유니콘에 등극함
- 2022년은 AI반도체 관련 새로운 기술 및 사용성에 주목됨. 엔비디아 신형 칩은 AI, 자율주행차, 메타버스를 지원함. MIT가 개발한 모듈식 AI반도체는 전자폐기물 감소가 기대됨

[표 1] 2019~2022년 AI 산업 주요 핵심 이슈

구분	주요 이슈
2019	▶ 중국 자체 반도체 기술 추진하는 가운데 화웨이·알리바바 AI반도체 발표
	▶ Alphabet GV, 광학 AI반도체 스타트업 Lightmatter에 투자
	▶ 바이두-삼성전자, AI반도체 양산 협력
2020	▶ Qualcomm, 클라우드 AI반도체 출시
	▶ 엔비디아 시장 선두, Intel과 GraphCore 등 도전
	▶ 구글, AI 활용하여 AI반도체 설계
2021	▶ 빅테크 기업, 자체 칩 개발 서둘러
	▶ Baidu, 2세대 Kunlun AI반도체 양산 돌입
	▶ 이스라엘 AI반도체 제조업체 Hailo, 유니콘 등극
2022	▶ 엔비디아 신형 칩, AI, 자율주행차, 메타버스 지원
	▶ Intel, NVIDIA에 도전하는 새로운 AI반도체 출시
	▶ MIT, 레고 블록 닮은 모듈식 AI반도체 개발…전자폐기물 감소 기대

출처 : 주요 글로벌 ICT 매체 발표 기사 취합

Ⅰ. 품목 개요

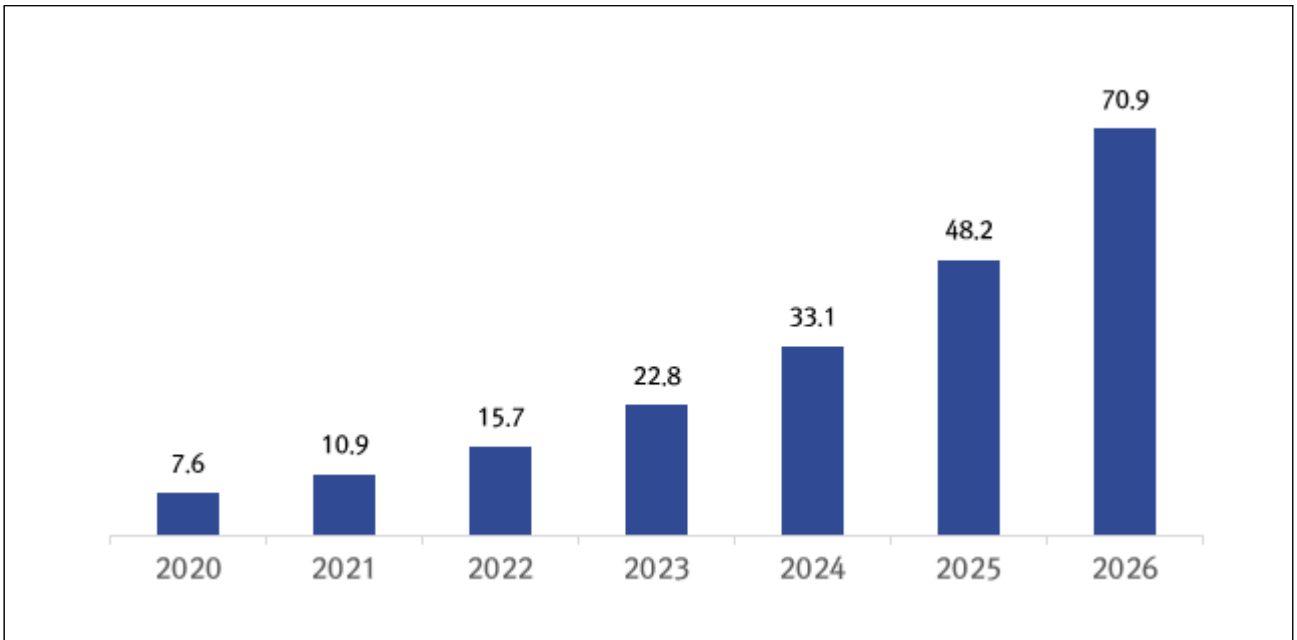
2. AI반도체 시장 규모

AI반도체 시장 규모, 2026년 $709억 전망

- 2021년 AI반도체 시장 규모는 109억 달러(약 14조 2,681억 원)로 추정됨. 2020~2026년 AI반도체 시장은 연평균 복합 성장률(CAGR) 42.3%로 성장해 2026년 709억 달러(약 92조 8,081억 원)에 이를 것으로 예상됨
- AI 기술을 지원하기 위해서는 머신러닝 알고리즘에 최적화된 특수 AI반도체가 필요함. 향후 AI반도체는 스마트폰, 태블릿, 스마트 스피커, 웨어러블 등 다양한 소비자 기기에 사용될 전망임. 이 외에도 로봇, 센서, 기타 IoT 장치 등 여러 시장에서 사용될 것으로 예상되며, 스마트시티 이니셔티브가 활발히 진행됨에 따라 추가적인 시장 성장이 기대됨
- 인텔(Intel), 삼성(Samsung), 브로드컴(Broadcom), 퀄컴(Qualcomm) 등 반도체 선도 기업들은 AI반도체 개발에 막대한 투자를 진행하고 있음. 이 외에도 어드밴스드 마이크로 디바이시스(AMD)와 엔비디아(NVIDIA)의 행보가 주목됨. 애플(Apple), 구글(Google)과 같은 빅테크 기업 또한 AI 개발을 위해 AI반도체 분야에서의 혁신을 모색하고 있음

[그래프 1] 2020~2026년 AI 반도체 시장 규모

단위 : 십억 달러



분류	2020	2021	2022	2026	CAGR (20~26)
AI반도체	87.04	119.78	164.99	1591.03	42.3%

출처 : Statista

Ⅰ. 품목 개요

3. AI반도체 선진국가

■ 미국과 중국, 전 세계 AI 반도체 특허 신청 양분

- 2006~2020년 AI 반도체 특허 신청량은 미국이 37%, 중국이 36%로 두 나라를 중심으로 양분되고 있음. 그 뒤를 이어 한국 8%, 일본 6%, 대만 2%, 기타 10%로 나타남

■ 중국, AI 반도체 특허 관련 양적 우위 차지

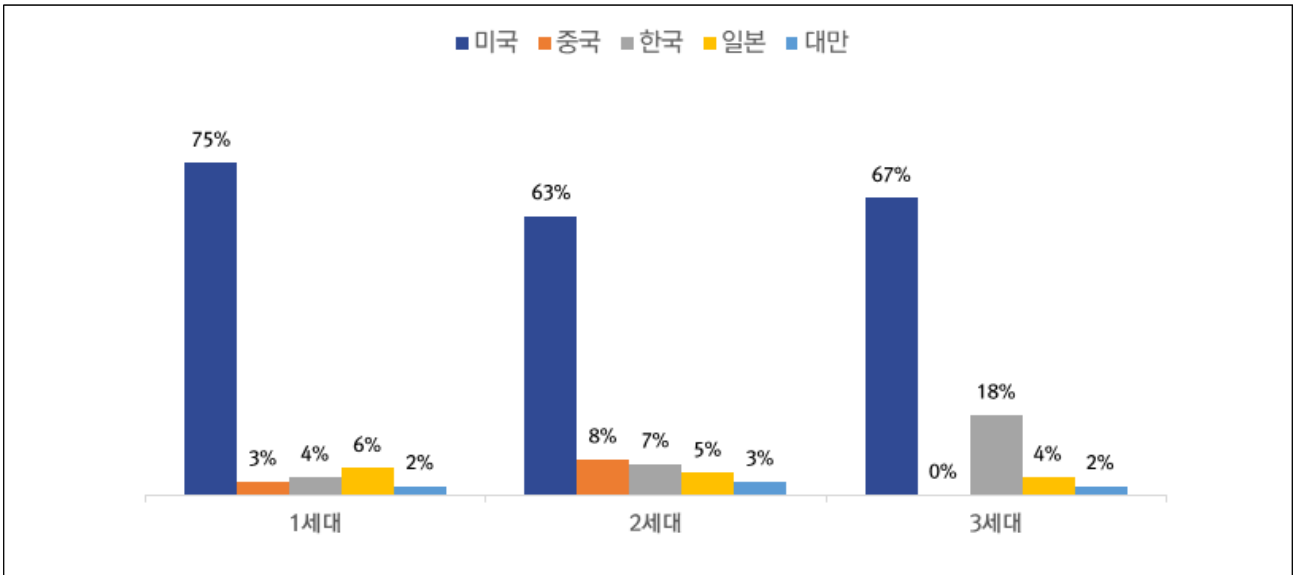
- 2019년 중국이 특허 신청량에서 미국을 추월했으며, 그 이후 중국이 특허신청량에서 세계 1위를 유지하고 있음

■ 미국, 중국 대비 질적 우위 차지

- 미국은 해외 특허신청비율, 평균 패밀리 국가 수 등 질적 지표에서 중국 대비 우위를 차지함. 중국은 특허신청량은 많으나 자국 중심으로 이루어져 한계가 나타남
- AI 반도체 세계 최대 규모인 미국 시장을 겨냥해 각국의 특허신청 경쟁이 치열함

[그래프 2] 2020~2026년 AI 반도체 시장 규모

단위 : 십억 달러



분야	2026	CAGR (20~26)
해외 특허신청량	5,124	1,049
해외 특허신청비율	50%	11%
평균 패밀리 국가수	3.2	1.3

출처 : 특허청

Ⅰ. 품목 개요

4. AI반도체 신흥국가

일본 메가칩스, 글로벌 AI반도체 시장 진입

- 2022년 4월 일본 커스텀 주문형 직접회로(ASIC) 기업 메가칩스(MegaChips)가 AI 전문가 없이도 AI 기능을 통합할 수 있는 AI 파트너 프로그램을 출시했다고 발표함. 이는 메가칩스가 글로벌 에지 AI(Edge AI) 칩 시장에 진입한 것을 의미함
- 메가칩스는 일본 내에서 성공을 거둔 후 미국 시장으로 확장해 풀 서비스 ASIC 솔루션을 제공하고 있음

이스라엘 AI반도체 칩 기업에 주목

- 스마트 카메라, 스마트 자동차 등 에지 디바이스(Edge Device)에 AI 기술을 제공하는 이스라엘 칩 제조업체 헤일로(Hailo)가 2021년 10월 시리즈 C 펀딩을 통해 1억 3,600만 달러(약 1,780억 2,400만 원) 유치에 성공함. 투자를 통해 헤일로의 가치는 10억 달러(약 1조 3,090억 원) 이상으로 평가되며 유니콘 반열에 등극함
- 2019년 12월 인텔(Intel)은 이스라엘 AI 반도체 스타트업 하바나 랩스(Habana Labs)를 20억 달러(약 2조 6,180억 원)에 인수함. 인텔은 AI 비즈니스 강화를 위한 전략의 일부로 인수를 택했으며, 이를 통해 컴퓨팅 성능을 높이고 데이터 센터 효율성을 개선하고자 함
- 이스라엘 반도체 신생기업 Neuronix AI Labs는 비용 및 전력 소비의 95% 이상을 축소하는 것을 목표로 하는 AI반도체 제조업체임. 해당 기업은 시드 라운드에서 150만 달러(약 19억 6,350만 원) 유치에 성공함. 또 다른 AI반도체 스타트업 FortifyIQ는 부채널 공격(side-channel attack) 관련 최대한의 보안을 제공하는 것에 중점을 두며 칩셋을 생산하는 것이 특징임

선도 기업

1. 인공지능 선도 기업
2. 선도 기업 분석
   ① NVIDIA
   ② SambaNova Systems
   ③ Cerebras Systems

nipa 정보통신산업진흥원

II. 선도 기업

1. AI반도체 선도 기업

AI반도체 선도 기업

- NVIDIA : 컴퓨터 GPU 설계
  - AI반도체 칩 개발 관련 대규모 확장 추진
- SambaNova Systems : AI 하드웨어-소프트웨어 시스템 개발
  - 고성능, 고정밀 AI 하드웨어-소프트웨어 프로그램 시스템 개발
- Cerebras Systems : 의료 분야 AI 딥러닝 애플리케이션 위한 컴퓨팅 시스템 구축
  - 새로운 AI 반도체 칩 'Cerebras WSE-2' 출시로 대규모 AI 모델 가능하도록 지원

[표 2] 2022년 AI반도체 글로벌 선도기업

구분	기업명	설립연도	시가총액/기업가치($)
주요 AI반도체 제조사	Google Alphabet	1998	1조 4,800억
	NVIDIA	1993	3,940억
	Intel	1968	1,579억
	Advanced Micro Devices(AMD)	1969	1,314억
	IBM	1911	1,258억
주요 스타트업	SambaNova Systems	2017	51억
	Cerebras Systems	2015	40억
	Graphcore	2016	28억
	Groq	2016	10억
	Mythic	2012	N/A

출처 : 주요 글로벌 ICT 매체 취합

II. 선도 기업

2. AI반도체 선도 기업 분석

① NVIDIA

NVIDIA : 컴퓨터 GPU 설계

- 데이터 사이언스 및 고성능 컴퓨팅을 위한 GPU, API를 설계
- 모바일 컴퓨팅 및 자동차 시장을 겨냥한 단일 칩 시스템(SoC) 생산
- AI반도체 칩 개발 관련 대규모 확장 추진

[표 3] NVIDIA 기업 분석

구분		내용		
기업 정보	기업명(국적)	NVIDIA (미국) / www.NVIDIA.com		
	시가총액/기업가치	$3,940억 5,000만	설립년도	1993
	기업 유형	AI반도체		
발전 단계		▸ 2004년 고효율 병렬 프로세서 '쿠다(CUDA)' 투자 이후 AI 사업 빠르게 성장 ▸ 2016~2021년 매출 신장률 233% 기록. 2019년 총 매출의 25%를 차지한 데이터 센터 매출이 2020년 36%로 증가 ▸ 2021년 4월 독자적으로 개발한 데이터 센터용 ARM 기반 CPU 발표 ▸ 2022년 3월 AI 반도체 칩 및 소프트웨어 앱 개발 규모 대규모 확장 계획 발표		
개발 기술		▸ 대표 솔루션 ① DGX™ A100 - 데이터 센터용으로 설계된 주력 AI반도체 - 8개의 GPU와 최대 640GB GPU 메모리를 통합 ▸ 대표 솔루션 ① H100 GPU - 새로운 GPU 아키텍처 '호퍼(Hopper)'가 적용된 제품 - 디지털 트윈, 딥러닝 추론, 인공지능(AI) 언어, 딥러닝 추론 등 전방위적 산업에 사용 가능 ▸ 대표 솔루션 ② 그레이스(Grace) GPU - 컴퓨텍스 2022(Computex 2022) 기조연설을 통해 자체 개발한 ARM 기반 CPU '그레이스(Grace)' 탑재 서버 라인업 공개 - 그레이스 탑재 서버는 클라우드 게임과 그래픽에 최적화된 CGX, 디지털 트윈과 옴니버스 구축을 위한 OVX, 고성능 컴퓨팅을 위한 HGX 등 총 4개로 구성		

출처 : NVIDIA

II. 선도 기업

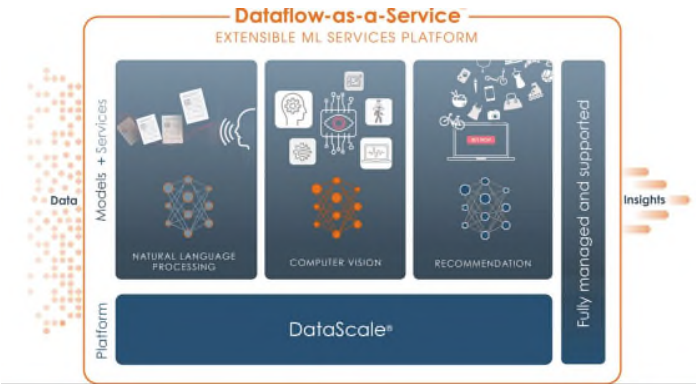
2. AI반도체 선도 기업 분석

② SambaNova Systems

■ SambaNova Systems : AI 하드웨어-소프트웨어 시스템 개발

- 고성능, 고정밀 AI 하드웨어-소프트웨어 프로그램 시스템 개발
- 데이터센터에서 에지에 이르기까지 AI 및 데이터 집약적인 앱을 실행하는 시스템을 제공
- AI 프로세서를 판매하는 대신 데이터 센터를 구축해 임대하는 사업 진행

[표 4] SambaNova Systems 기업 분석

구분		내용		
기업 정보	기업명(국적)	SambaNova Systems (미국) / sambanova.ai		
	시가총액/기업가치	$51억	설립년도	2017
	기업 유형	AI 하드웨어 및 소프트웨어, AI 반도체, 데이터센터		
발전 단계		▸ ELEVAITE 멤버십 프로그램 'GPT-as-a-service' 발표 ▸ 2020년 2월 유니콘 등극 ▸ 2021년 4월 시리즈 D 펀딩으로 6억 7,600만 달러(약 8,871억원) 투자 유치 ▸ 2022년 기준 11억 달러(약 1조 4,435억원) 이상 자금 조달 성공		
개발 기술		▸ 대표 솔루션 : ① SN10 RDU – 맞춤형 데이터 흐름 파이프라인을 구축할 수 있는 유연성과 대용량 모델을 효율적으로 실행할 수 있는 대용량 메모리 데이터 프로세서 칩 ▸ 대표 솔루션 : ② GPT-as-a-service – 모델 커스터마이징 및 트레이닝, 배포, 운영 및 유지 관리 등 고객이 필요로하는 모든 작업을 대신해주는 프로세스 구축 		

출처 : SambaNova Systems

II. 선도 기업

2. AI반도체 선도 기업 분석

③ Cerebras Systems

■ Cerebras Systems : 의료 분야 AI 딥러닝 애플리케이션 위한 컴퓨팅 시스템 구축

- 신경의학 분야 뉴런, 시냅스 연결 관련 방대한 데이터를 AI 딥러닝 기술을 활용해 처리할 수 있도록 지원
- 새로운 AI 반도체 칩 'Cerebras WSE-2' 출시로 대규모 AI 모델 가능하도록 지원
- 아스트라제네카(AstraZeneca), 글락소스미스클라인(GlaxoSmithKline) 등 제약 회사와 협력

[표 5] Cerebras Systems 기업 분석

구분		내용		
기업 정보	기업명(국적)	Cerebras Systems (미국) / www.cerebras.net		
	시가총액/기업가치	$40억	설립년도	2015
	기업 유형	AI반도체, 컴퓨팅 시스템		
발전 단계		▸ 2018년 시리즈 D 라운드를 통한 8,800만 달러(약 1,156억 560만 원) 자금 조달로 유니콘 등극 ▸ 2021년 11월 시리즈 F 펀딩으로 2억 5천만 달러(약 3,284억 2,500만 원) 투자 유치 성공. 기업가치 40억 달러(약 5조 2,548억 원) 이상 평가 ▸ 2022년 3월 방대한 양의 신경의학 데이터를 처리 가능한 'Cerebras WSE-2' 칩 출시		
개발 기술		▸ 대표 솔루션 ① Cerebras WSE-2 - 전 세계에서 가장 큰 규모의 컴퓨터 칩. 해당 칩 클러스터를 통해 현존 AI 모델 대비 100배 더 거대한 AI 모델 실행 가능 - 850,000개의 코어와 2조 6,000억 개의 트랜지스터를 갖춘 것이 특징 - 생물학적 뉴런과 시냅스 간의 120조 개 연결 상호작용 관련 수학적 시뮬레이션 실행 가능 		

출처 : Cerebras Systems

유망 기술

1. 유망 기술 선정
2. 급성장 기술 키워드
   ① 고대역폭 메모리
   ② 심층신경망
   ③ 명령어 집합 구조
   ④ 신경망처리장치
   ⑤ 양자컴퓨팅

nipa 정보통신산업진흥원

Ⅲ. 유망 기술

1. 유망 기술 선정

2021년 7월~2022년 6월, 주요 급성장 AI반도체 기술 키워드

- 고대역폭 메모리(HBM) : AI반도체 수요, HBM 성장 주도
- 심층신경망(DNN) : 심층신경망 기술 AI반도체에 응용
- 명령어 집합 구조(ISA) : AI반도체 업계, 오픈소스 명령어 세트 '리스크파이브'에 주목
- 신경망처리장치(NPU) : NPU, AI 연산에 유용
- 양자컴퓨팅(quantum computing) : AI 기술, 양자 컴퓨팅으로 비약적 도약 가능 전망

[표 6] 2021년 7월~2022년 6월 급성장 AI반도체 기술 키워드

순위	키워드		발생률1)	성장률2)
	국문	영문		
①	고대역폭 메모리	HBM	0.38%	695%
②	심층신경망	DNN	0.38%	692%
③	명령어 집합 구조	ISA	0.18%	563%
④	신경망처리장치	NPU	0.49%	479%
⑤	양자컴퓨팅	quantum computing	12.32%	435%
⑥	불휘발성 메모리	nonvolatile memory	0.15%	424%
⑦	주문형 반도체	ASIC	2.06%	285%
⑧	DRAM	DRAM	1.28%	256%
⑨	반도체 설계 자동화	EDA	1.17%	198%
⑩	강자성체 메모리	MRAM	0.09%	189%

출처 : 2021년 7월~2022년 6월, IT 뉴스매체 분석 결과

1) 발생률 : 2021년 7월~2022년 6월 AI반도체 기술 키워드 전체 발생량 124,446건 중 해당 키워드의 발생 비율을 뜻함
2) 성장률 : (후반 6개월 키워드 발생량) - (전반 6개월 키워드 발생량) / (전반 6개월 키워드 발생량)

Ⅲ. 유망 기술

2. 급성장 기술 키워드

① 고대역폭 메모리(HBM)

(*) 고대역폭 메모리란?
고대역폭 메모리(HBM)는 2013년 발표된 적층형 메모리 규격으로, 고성능 그래픽스 가속기 및 네트워크 장치와 결합하기 위해 사용되는 고성능 램(RAM) 인터페이스를 의미

AI반도체 수요, HBM 성장 주도

- AI반도체 시장 성장으로 HBM 발전 견인
  - AI반도체 시장이 급성장하면서 고대역폭 메모리(HBM) 기술이 주류로 자리잡고 있는 추세
  - AI반도체의 경우 HBM 등 메모리 반도체에 기반한 성능 향상 지속
  - HBM은 현재 최고급 게임용 그래픽 카드 대부분에 사용되는 GDDR 메모리 대비 훨씬 더 높은 대역폭과 낮은 전력 소비를 제공해 GDDR을 대체하는 용도로 활용
- HBM 규격 개발 진행
  - HBM 규격은 HBM, HBM2, HBM2E, HBM3로 구분
  - HBM3은 대역폭의 급격한 향상으로 개발이 늦어지고 있으며, HBM2E가 현재 대체재로 투입

주요 AI반도체 기업 HBM 개발 주도

- 삼성전자, 하이닉스 등 반도체 기업 HBM 개발 주도
  - 삼성전자는 2020년 2월 스택당 최대 8-Hi, 최대 3.2GT/s, 410GB/s, 총 16GB를 지원하는 플래시볼트 HBM2E를 양산
  - SK하이닉스는 2020년 7월 스택당 최대 8-Hi, 최대 3.6GT/s, 460GB/s, 총 16GB를 지원하는 HBM2E를 개발하여 대량 생산에 돌입
  - 엔비디아(NVIDIA) 데이터센터 GPU ‘A100’은 2TB/s의 메모리 대역폭으로 80GB의 HBM2E 성능을 제공
  - 인텔(Intel)은 차세대 데이터 센터에 적용되는 제온(Xeon) 프로세서의 차세대 서버용 칩 ‘사파이어래피즈(Sapphire Rapids)’ 제품군에 HBM을 소개

[표 7] HBM 사양

구분	HBM2 / HBM2E(현재)	HBM	HBM3
핀 최대 전송속도	3.2Gbps	1Gbps	-
최대 용량	24GB	4GB	64GB
최대 대역폭	410 GBps	128 GBps	512 GBps

출처 : Toms hardware

Ⅲ. 유망 기술

2. 급성장 기술 키워드

② 심층신경망(DNN)

(*) 심층신경망(DNN)이란?
심층신경망이란 입력층과 출력층 사이에 여러개의 은닉층들로 이뤄진 인공신경망으로, 복잡한 비선형 관계들을 모델링 가능하며, 각 물체가 분석 대상의 기본적 요소들에 대한 계층적 구성으로 표현 가능

■ 심층신경망 기술 AI반도체에 응용

- 심층신경망 기술에 AI 부문 주목
  - 기계학습의 하위 유형인 심층신경망(DNN)은 AI 부문에서 가장 주목하는 기술로 인기
  - 대규모 작업을 수행하는 AI 프로세스는 병렬 작업이 적합하며, 심층신경망은 병렬 계산이 가능
- 심층신경망 기술 발전으로 AI 문제 해결
  - 심층신경망 기술은 AI 학습의 정확성을 높일 수 있으나 기존 컴퓨터로 감당이 어려워 실용화에 어려움 존재
  - 엔비디아(NVIDIA)는 대량 병렬 연산 기능을 가진 GPU 활용으로 심층신경망 실용 가능성 향상
  - AI 분야 심층신경망 기술 사용의 실용성 증대로 AI가 가진 문제 해결이 수월
  - 심층신경망 응용 기술 발전으로 향후 AI 반도체가 방대한 계산을 감당할 수 있을 것으로 예측

■ 신티언트, AI반도체 'NDP200' 출시

- 신티언트(Syntiant), 심층신경망 프로세서 가능한 AI반도체 출시
  - 미국 AI 프로세서 반도체 스타트업 신티언트가 딥러닝용 AI반도체 'NDP200'을 출시
  - 딥러닝 및 반도체 설계가 NDP200 칩 솔루션에 결합되면 초저전력, 고성능 심층신경망 프로세서를 실행할 수 있는 것이 특징
- 1mW 미만에서 정확한 추론으로 시각적 처리를 수행하며, 전 모델 'NDP100' 대비 25배 이상의 처리량이 특징
- 이전 모델 대비 더 많은 신경 컴퓨팅을 도입하여 디바이스 인텔리젠스(device intelligence)를 가능하게 하는 것이 목표

Ⅲ. 유망 기술

2. 급성장 기술 키워드

③ 명령어 집합 구조(ISA)

(*) 명령어 집합 구조(ISA)란?
명령어 집합 구조란 소프트웨어 하드웨어 사이의 약속으로, 여러 명령어를 정의하는 것을 의미. 현재 시스템의 구성 상태를 알 수 있으며, 명령어 실행 시 상태가 어떻게 바뀌는지 확인 가능

■ AI반도체 업계, 오픈소스 명령어 세트 '리스크파이브'에 주목

- 리스크파이브(RISC-V) 시장 규모 2024년 $10억 예상
  - UC 버클리에서 2010년부터 개발하고 있는 오픈소스 명령어 세트 아키텍처 '리스크파이브'에 반도체 업계 관심 집중
  - 리스크파이브 시장 매출 규모는 2021년 4억 달러(약 5,246억 원) 미만에서 2024년 10억 달러(약 1조 3,115억 원) 규모로 성장할 것으로 예상
  - 딜로이트 글로벌(Deloitte Global)은 리스크파이브 프로세싱 코어 시장 규모는 2022년 전년 대비 두 배로 성장할 것이라 예상. 또 2023년에는 해당 시장이 다시 두 배로 급성장할 것이라 예측
- RISC-V, ARM 대체재로 반도체 업계 관심 상승
  - 리스크파이브는 상업적 장점이 높아 스마트폰, 임베디 장치 CPU로 독점적 성격을 가지고 있는 ARM과 경쟁할 수 있을 것으로 기대
  - 리스크파이브는 ARM 칩과 비교해 성능이 유사하며, 칩 면적은 30%~50% 수준으로 작고, 소비전력은 60%를 감소할 수 있다는 장점이 존재

■ 삼성SDS, 리스크파이브 AI반도체 테스트 진행

- 삼성SDS ET-SoC-1 진행
  - 삼성SDS가 리스크파이브 기반 고성능·저전력 컴퓨팅 솔루션 개발 기업 에스페란토 테크놀로지스(Esperanto Technologies)와 AI 추론 가속기 성능 테스트 'ET-SoC-1'를 진행
  - ET-SoC-1는 64비트의 리스크파이브 프로세서 코어 1,088개를 탑재한 것이 특징

■ 중국 반도체 기업, 리스크파이브 채택 증가

- 중국 기업, 라이선스 비용 절감 위해 리스크파이브 선호
  - 라이선스 비용을 절감하기 위해 리스크파이브를 채택하는 중국 기업이 많아지고 있는 추세
  - 중국 알리바바(Alibaba)는 자체 엔지니어가 설계한 리스크파이브 CPU 코어를 사용

Ⅲ. 유망 기술

2. 급성장 기술 키워드

④ 신경망처리장치(NPU)

(*) 신경망처리장치(NPU)란?
신경망처리장치란 자극을 종합판·판단해 명령을 내리는 인간의 뇌를 모방해 만든 데이터 처리 장치로, 심층신경망을 사용하는 딥러닝에서 복잡한 행렬 곱셈 연산을 수행

NPU, AI 연산에 유용

- NPU, AI컴퓨팅 및 AI 애플리케이션 구현에 탁월
  - NPU는 CPU, GPU 대비 AI 컴퓨팅 및 AI 애플리케이션 구현에 장점이 존재
  - 데이터 기반 병렬 컴퓨팅 아키텍처로 동영상 및 이미지와 등 대용량 멀티미디어 데이터 처리에 탁월
  - 화웨이(Huawei)는 세계 최초로 휴대폰에 NPU를 적용
  - 삼성 갤럭시(Galaxy)의 NPU는 모바일 프로세서에 내장돼 고급 신경망을 활용, 높은 수준의 시각 지능 제공

각국 기술 기업에서 신경망처리장치 개발 활발

- ST마이크로일렉트로닉스(STMicroelectronics), NPU 적용된 '마이크로 컨트롤러' 출시 예정
  - ST마이크로일렉트로닉스가 NPU를 갖춘 최초의 마이크로컨트롤러를 2022년 말 출시할 계획임
  - 출시 예정인 마이크로컨트롤러는 가속기가 탑재된 쿼드 코어 프로세서와 동일한 AI 성능을 제공하지만 비용은 10분의 1, 전력 소비는 12분의 1 수준일 것이라 밝힘
  - ARM 코어는 M55 또는 M85가 사용될 것으로 예측됨
- 한국전자통신연구원(ETRI), '딥러닝 컴파일러' 개발
  - ETRI는 AI 핵심 시스템 소프트웨어 딥러닝 컴파일러 '네스트(NEST-C)'를 개발함
  - 네스트의 개발로 AI 응용프로그램과 AI 반도체 간 이질성을 해소하여 AI 반도체 개발이 용이해짐
  - 네스트는 CPU, GPU, NPU 프로세서 모두 호환돼 범용성이 높은 것이 특징임

[표 8] 처리장치 특성

구분	특성
NPU	▶ 회로 레이어의 뉴런을 시뮬레이션하고 시냅스 가중치에 의해 저장 및 연산 통합 실현 ▶ 통신분야, 빅데이터, 이미지 처리에 주로 사용
CPU	▶ 트렌지스터의 70%가 캐시 구축 및 제어 장치의 일부로 사용 ▶ 논리 제어 작업에 적합한 연산장치
GPU	▶ 대규모 병렬 컴퓨팅에 적합하며, 계산 복잡성이 적은 연산장치에 트랜지스터를 사용. ▶ 빅데이터, 백엔드 서버, 이미지 처리에 주로 사용
FPGA	▶ 논리 프로그래밍이 가능하고 높은 연산 효율성이 특징으로 ▶ 스마트폰, 휴대용 모바일 기기, 자동차에 주로 사용

출처 : utmel

Ⅲ. 유망 기술

2. 급성장 기술 키워드

⑤ 양자컴퓨팅(quantum_computing)

(*) 양자 컴퓨팅(quantum_computing)이란?
양자 컴퓨팅이란 아원자 입자의 물리학을 활용해 정교한 병렬 계산을 수행하는 방법으로, 오늘날 컴퓨터 시스템에서 사용되는 단순한 형태의 트랜지스터를 대신하는 것이 특징

AI 기술, 양자 컴퓨팅으로 비약적 도약 가능 전망

- 양자 컴퓨팅, 복잡한 AI 연산 해결
  - 양자 컴퓨팅 기술로 복잡한 AI 연산을 빠르게 해결 가능해 AI 분야의 비약적 도약 가능
  - 양자 컴퓨팅은 고전 이진 컴퓨터가 해결할 수 없는 계산 문제 잠재적으로 해결 가능
  - 2020년 3월 구글(Google)은 양자 머신러닝 라이브러리 'TensorFlow Quantum' 공개

IBM, 양자 컴퓨팅 개발에 집중

- IBM, 세계 최대 초전도 양자컴퓨터 '이글(Eagle)' 공개
  - 2021년 11월 IBM은 127개 큐비트를 탑재한 세계 최대 초전도 양자컴퓨터 이글을 공개함
  - 이글의 막대한 연산 능력은 새로운 분자 및 물질의 모델링 작업, 금융 사기 탐지 등에 활용 가능
  - IBM은 IBM 퀀텀 네트워크(IBM Quantum Network)에 가입한 파트너를 대상으로 이글 프로세서를 설치해 전세계적인 양자 생태계 구축 및 상용화 계획
- IBM, 싱크2022(Think2022)에서 양자 로드맵 설명
  - IBM은 싱크2022 컨퍼런스에서 양자 로드맵을 발표
  - 2022sus akf 433 큐비트 프로세서 'IBM 오스프리(IBM Osprey)' 출시 예정
  - 2023년 1,000큐비트 이상의 세계 최초 범용 양자프로세서 'IBM 콘도르(IBM Condor)'를 공개 예정
  - 2025년까지 4,000큐비트의 양자컴퓨터 시스템을 구축할 계획

텐센트, 초전도 양자 칩 특허 취득

- 텐센트(Tencent), 초전도 양자 칩 관련 특허 출원
  - 중국 테크 기업 텐센트가 '큐비트의 주파수 제어 신호 처리 방법, 초전도 양자 칩' 특허를 출원
  - 특허를 통해 초전도 양자 비트의 주파수 제어 신호 왜곡을 측정 가능

유망 수요처

1. 유망 수요처 선정
2. 급성장 수요처 키워드
   ① 미디어
   ② 물류
   ③ 에너지
   ④ 리테일
   ⑤ 공공

nipa 정보통신산업진흥원

Ⅳ. 유망 수요처

1. 유망 수요처 선정

2021년 7월~2022년 6월 주요 급성장 AI반도체 수요처 키워드

- 미디어(Media) : 엔비디아, AI 3D 모델링 'NeRF' 기술 공개
- 물류(Logistics) : 물류 분야 AI반도체 수요 증가 추세
- 에너지(Energy) : 인도공과대학, 초저전력 AI반도체 개발
- 리테일(Retail) : 각국 공공분야 AI반도체 활용 프로젝트 실시
- 공공(Public) : 리테일 관련 AI반도체 기업에 투자 활발

[표 9] 2021년 7월~2022년 6월 급성장 AI반도체 수요처

순위	키워드		발생률[3]	성장률[4]
	국문	영문		
①	미디어	media	12.38%	157%
②	물류	logistics	1.23%	131%
③	에너지	energy	3.97%	121%
④	리테일	retail	4.30%	110%
⑤	공공	public	13.42%	105%
⑥	교육	education	2.87%	105%
⑦	소비자	consumer	8.93%	104%
⑧	운송	transport	2.49%	86%
⑨	국방	defense	2.27%	85%
⑩	보건	health	8.85%	74%

출처 : 2021년 7월~2022년 6월 IT 뉴스매체 분석 결과

3) 발생률 : 2021년 7월 ~2022년 6월 AI 반도체 수요처 키워드 전체 발생량 81,158건 중 해당 키워드의 발생 비율을 뜻함
4) 성장률 : (후반 6개월 키워드 발생량) - (전반 6개월 키워드 발생량) / (전반 6개월 키워드 발생량)

Ⅳ. 유망 수요처

2. 급성장 수요처 키워드

① 미디어(Media)

■ 엔비디아, AI 3D 모델링 'NeRF' 기술 공개

- 엔비디아 GTC(NVIDIA GTC)에서 AI 3D 모델링 기술 공개
  - 2022년 3월 엔비디아는 엔비디아 GTC를 통해 사진 몇 장으로 짧은 시간 안에 폴리곤 이미지를 만드는 '엔비디아 인스턴트 NeRF'를 공개
  - 같은 공간에서 동시에 촬영된 다각도의 사진을 AI가 분석한 후 3D 모델링을 완성하며, NeRF 기술은 초고속 신경망 훈련 기술이 적용돼 AI의 3D 모델링 속도를 1,000배 이상 향상

■ 그래프코어, 3D WoW 적용한 AI 반도체 출시

- 그래프코어(Graphcore), 보우 IPU(Bow IPU) 출시
  - 2022년 4월 그래프코어는 TSMC와 협력을 통해 세계 최초 3D 웨이퍼-온-웨이퍼(Wafer-on-Wafer) 반도체 'Bow IPU(보우 IPU)'를 출시
  - 보우 IPU는 차세대 Bow POD AI 컴퓨터 시스템의 핵심으로, 주요 AI 애플리케이션에 걸쳐 기존 프로세서 대비 40% 향상된 성능과 16% 뛰어난 전력 효율을 제공
  - 미국 에너지부 산하 퍼시픽 노스웨스트 국립 연구소(PNNL)에서 사이버 보안 및 컴퓨터 화학 관련 애플리케이션에 Bow를 적용

■ 세레모픽, 스텔스 모드 종료

- 세레모픽(Ceremorphic), 실리콘 시스템 제공 계획 발표
  - 2022년 1월 AI반도체 디자인 스타트업 세레모픽이 스텔스 모드를 종료하고 실리콘 시스템을 제공하겠다 발표
  - 반도체 칩 제조는 물론 워크로드를 줄이기 위한 알고리즘을 개발

Ⅳ. 유망 수요처

2. 급성장 수요처 키워드

② 물류(Logistics)

■ 물류 분야 AI반도체 수요 증가 추세

- 전자상거래 급증으로 자동화 요구 증가
  - 코로나19 발발로 전자상거래 규모 성장 가속화
  - 2019년 전체 리테일 매출 중 전자상거래는 14% 미만이었으나 2021년 약 20% 수준으로 상승
  - 물류 창고의 80%가 자동화를 도입하고 있지 않으며, 전자상거래 급증으로 자동화 요구가 증가함. 이에 따라 AI반도체 수요 함께 상승

■ 신티어트, $5,500만 신규 자금 조달 완료

- 에지 AI반도체 생산 기업 신티어트(Syntiant), 자금 유치 성공
  - 2022년 3월 에지 AI반도체 생산 기업 신시어트가 5,500만 달러 신규 자금 유치에 성공
  - 르네사스일렉트로닉스, 인텔캐피탈, 어플라이벤처스 등 투자 참여
  - 신시어트는 딥러닝과 반도체 설계를 결합한 고성능 에지 AI반도체을 생산하며, 물류, 보안시스템, 산업, 제조, 운송 등 다양한 분야에 적용

[표 10] 물류 산업 AI반도체 활용 사례

기업명	제품명	내용
퀄컴 (Qualcomm)	QCM6490/ QCS6490 SOC	▶ Kryo 670 CPU, Adreno 642L GPU, Adreno 633 VPU를 기반으로 하며, Hexagon DSP AI 엔진을 포함 ▶ 커넥티드 헬스케어, 물류 관리, 창고 등에 활용
보스턴 다이나믹스 (Boston Dynamics)	Stretch	▶ 그리퍼가 장착된 거대한 창고 로봇 ▶ 컴퓨터 비전 시스템을 사용하여 자율적으로 작동하며 높은 수준의 상자 식별 능력 보유 ▶ 물류 회사 DHL Supply Chain과 1,500만 달러 계약 체결
OTTO Motors	Lifter	▶ 자체적으로 최상의 경로를 선택할 수 있는 자율지게차 ▶ AI와 딥러닝으로 구동되며 복잡한 환경에서 실시간 결정 가능

출처 : 주요 글로벌 ICT 매체 발표 기사 취합

Ⅳ. 유망 수요처

2. 급성장 수요처 키워드

③ 에너지(Energy)

■ 인도공과대학, 초저전력 AI반도체 개발

• 인도공과대학(Indian Institute of Technology), 스파이킹 신경망 저전력 AI반도체 개발

- 2022년 6월 인도공과대학 연구진은 스파이킹신경망(SNN) 기술을 활용한 초저전력 AI반도체 개발
- 대규모 트랜지스터 전류가 필요해 전력 소비가 높은 기존 SNN의 단점을 해결
- BTBT(band-to-band-tunneling) 전류 소스를 사용해 초저 전류로 SNN 커패시터 충전이 가능

■ 세레브라스, 에너지 부문 첫 고객 유치

• 세레브라스(Cerebras), 토탈에너지 SE(TotalEnergies SE)에 제품 판매

- 세레브라스가 토탈에너지 SE에 자사 'CS-2' 컴퓨터를 판매하며 에너지 부문 첫 고객 유치에 성공
- CS-2로 수행한 첫 번째 프로젝트는 석유와 가스를 시추하는데 사용되는 '지진 모델링' 작업으로, 기름, 가스 등을 구분

■ 구글 딥마인드, 핵융합 토카막 플라즈마 제어 기술 개발

• 구글 딥마인드(Google Deepmind), 강화학습으로 신기술 연구

- 구글 딥마인드가 심층 강화학습을 사용해 핵융합 토카막 플라즈마 제어 기술을 개발
- 핵융합을 통해 깨끗한 에너지를 공급 가능하나 엄청난 에너지를 방출해 위험성 존재, 토카막을 이용해 고온의 플라즈마를 제어하는 기술 연구
- 강화학습을 통해 복잡한 상태로 연속적으로 이루어지는 핵융합 과정을 학습하고 전자 코일을 제어하는 방법을 자율적으로 발견하는 시스템 개발

■ WBG 반도체, 전기차 성장 주도

• 광대역갭(WBG) 반도체, 전기차 성능 향상 지원

- WBG 반도체는 하이브리드 및 완전 전기차의 성능을 향상시켜 전기차 성장을 주도
- 여러 자동차 업체에서 전력 반도체 SiC, GaN에 투자하고 있으며, 반도체 제조 업체는 전기 자동차 시장의 급성장에 힘입어 사업을 전환하는 추세

Ⅳ. 유망 수요처

2. 급성장 수요처 키워드

④ 리테일(Retail)

■ 리테일 관련 AI반도체 기업에 투자 활발

- 블레이즈(Blaize), $7,100만 자금 조달 성공
  - AI반도체 기업 블레이즈가 시리즈D 펀딩을 통해 7,100만 원(약 931억 5,200만 원) 자금 조달에 성공
  - 블레이즈는 고성능 AI 워크로드를 위한 특수 칩을 제조하며, 고유한 칩 아키텍처 '그래프 스트리밍 프로세서(Graph Streaming Processor)'를 사용
  - 블레이즈의 모듈 중 일부는 소매점 등 원격에 위치한 데스크탑 컴퓨터 및 서버용으로 설계
  - 리테일 등 산업 부문에서 고성능, 저전력, 저비용 AI 하드웨어 및 소프트웨어에 대한 수요 증가를 충족하기 위한 제품 로드맵에 신규 자금 사용 계획
- 신티어트(Syntiant), 4,200만 파운드 투자 유치 마감
  - 신티어트가 글로벌 에지 AI반도체 공급 확대를 위해 시리즈 D 라운드에서 4,200만 파운드(약 660억 7,524만 원)를 유치함
  - 신티어트의 'Neural Decision Processor'는 웨어러블 기기용으로 설계됐으며 듣고, 말하고, 보고, 느끼는 다양한 방식 활용 가능

■ 엔비디아, 지멘스와 메타버스 협력 구축

- 산업용 메타버스 시장 공략
  - 2022년 6월 엔비디아(NVIDIA)는 메타버스 개발을 위해 지멘스(Siemens)와 파트너십을 체결
  - 지멘스는 개방형 디지털 비즈니스 플랫폼 '엑셀러레이터(Xcelerator')에 엔비디아의 고성능 GPU 등 활용해 기업용 솔루션 제공
  - AI를 활용해 소매 업체에서 매장 내 프로모션 가격 결정, 고객 개인화 및 추천 등 고객에게 더 나은 쇼핑 경험 제공할 수 있도록 지원

■ Flex Logic, AL·ML 소프트웨어 출시

- 플렉스 로직(Flex Logic), '이지 비전(EasyVision)' 플랫폼 출시
  - 2022년 6월 플렉스 로직이 AI 인터페이스 액셀러레이터 칩을 위한 ML 플랫폼 이지 비전을 출시
  - 소매 분석, 보안, 산업 등 광범위한 시장을 위한 에지 컴퓨터 비전 제품을 신속하게 출시할 수 있도록 설계

Ⅳ. 유망 수요처

2. 급성장 수요처 키워드

⑤ 공공(Public)

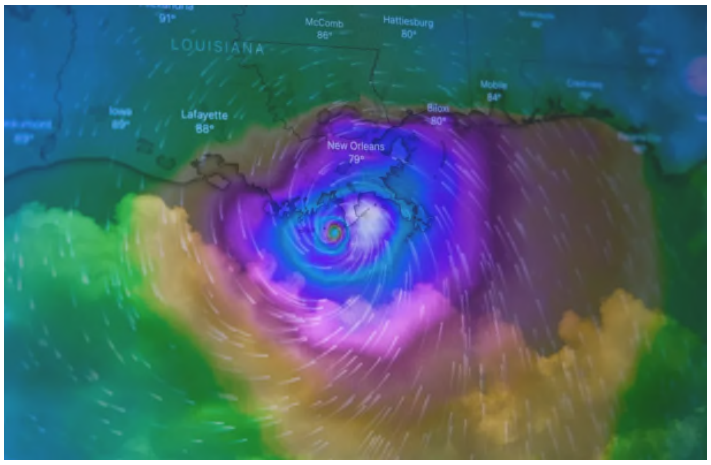
■ 각국 공공분야 AI반도체 활용 프로젝트 실시

- 싱가포르, AI 연구에 1억 8,000만 싱가포르 달러 추가 투입
  - 2021년 11월 싱가포르 정부는 AI 연구에 1억 8,000만 싱가포르 달러(약 1,697억 940만 원)를 추가 배정
  - AI 연구를 통해 새로운 응용 프로그램을 발전시키는 것이 목적
- 캐나다 최초의 공공 연구용 양자 컴퓨터 탄생
  - 캐나다 양자 컴퓨팅 회사 Anyon Systems는 캐나다 최초의 공공 연구용 양자 컴퓨터 Monarch를 고성능 컴퓨팅 센터 칼큘 퀘벡(Calcul Québec)에 인도
  - Monarch는 칼큘 퀘벡의 기존 슈퍼컴퓨팅 인프라와 통합

■ 슈퍼컴퓨터로 정확한 기상 예측

- 국립해양대기국(NOAA), 슈퍼컴퓨터 교체로 일기예보 개선
  - 미국 상무성(DOC) 하부기관 국립해양대기국이 신식 슈퍼컴퓨터 두 대를 도입하며 일기예보 개선
  - 새로 도입한 HPE 크레이(HPE Cray) 슈퍼컴퓨터는 최대 12.1페타플롭을 작동할 수 있는 327,680개의 코어를 제공하는 2,560개의 AMD Epyc Rome 64코어 7742 서버 CPU를 제공
  - 신식 슈퍼컴퓨터를 통해 높은 해상도의 날씨 모델을 생성하여 기존에 발견되지 않았던 날씨 패턴, 이상 현상, 뇌우 등 예측 가능

[그림 1] 슈퍼컴퓨터 'HPE 크레이' 활용 날씨 모델 생성 이미지



출처 : tweaktown

[참고문헌]

■ 참고 자료

1. 특허청, 「우리나라 인공지능 반도체 기술, 특허로 저력 확인」, 2022.03

■ 참고 사이트

1. amd(www.amd.com)
2. eetasia(www.eetasia.com)
3. tomshardware(www.tomshardware.com)
4. embedded(www.embedded.com)
5 electronicdesign(www.electronicdesign.com)
6. sciencedirect(www.sciencedirect.com)
7. skhynix(news.skhynix.com)
8. linkedin(www.linkedin.com)
9. medicaldevice-network(www.medicaldevice-network.com)
10. tti(www.tti.com)
11. spectrum(spectrum.ieee.org)
12. cdotrends(www.cdotrends.com)
13. theregister(www.theregister.com)
14. deloitte(www2.deloitte.com)
15. eetimes(www.eetimes.com)
16. eenewseurope(www.eenewseurope.com)
17. utmel(www.utmel.com)
18. techxplore(techxplore.com)
19. datanami(www.datanami.com)
20. reuters(www.reuters.com)
21. analyticsindiamag(analyticsindiamag.com)
22. pandaily(pandaily.com)
23. cnet(www.cnet.com)
24. nextplatform(www.nextplatform.com)
25. zdnet(www.zdnet.com)
26. iotworldtoday(www.iotworldtoday.com)
27. freethink(www.freethink.com)
28. siliconangle(siliconangle.com)
29. nvidia(www.nvidia.com)
30. edgecomputing-news(www.edgecomputing-news.com)
31. thehindubusinessline(www.thehindubusinessline.com)
32. yahoo(finance.yahoo.com)
33. tweaktown(www.tweaktown.com)
34. statista(www.statista.com)
35. nedo(www.nedo.go.jp)
36. timesofisrael(www.timesofisrael.com)
37. cardumencapital(cardumencapital.medium.com)

· 발행·편집 : 정보통신산업진흥원
· 발행일자 : 2022.7.22