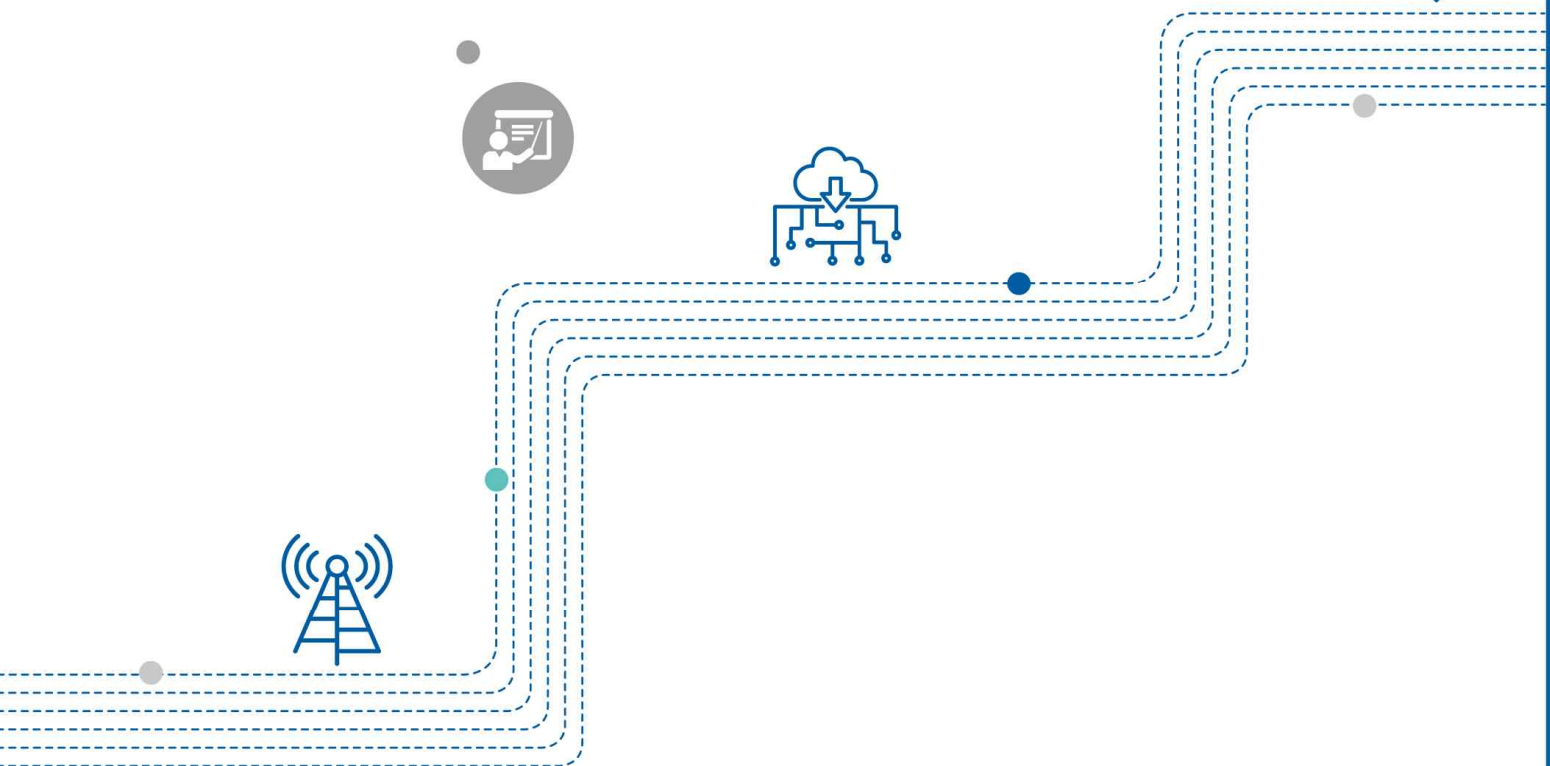


지능형 반도체의 주요 응용분야 시나리오와 핵심가치

하영욱·김태완



본 보고서는 ETRI 기술경제연구본부 주요사업인 “ICT R&D 경쟁력 제고를 위한 기술경제 연구”의 결과물입니다.



본 보고서의 내용은 연구자의 견해이며 ETRI의 공식 의견이 아님을 알려드립니다.

본 보고서의 일부 내용이나 표현은 ETRI(2017)의 “인공지능 반도체 산업동향 및 이슈 분석” 보고서를 인용하거나, 수정·보완하였습니다.



Contents

요약	1
I. 개요	7
1. 개념, 범위 및 대상 시장	7
2. 기술 분류	8
3. 중요성 및 특징	9
II. 주요 동향 및 전망	13
1. 산업 구조	13
2. 사업자 동향	14
3. 기술 동향	21
4. 국내 R&D 투자 동향	24
III. 주요 응용분야별 핵심가치	29
1. 도출 프레임워크	29
2. 지능형 반도체의 니즈 도출	30
3. 핵심가치 제안	41
참고문헌	43

표목차



[표 1] 지능형 반도체 기술분류	8
[표 2] 지능형 반도체 세계시장전망 (억 달러)	9
[표 3] 주요 벤더 - 반도체 업체	17
[표 4] 주요 벤더 - SW/서비스 업체	19
[표 5] 주요 벤더 - 디바이스 업체	20
[표 6] 국내 지능형 반도체 관련 주요 공공 R&D 현황 및 계획	25



[그림 1] 지능형 반도체의 개념 (협정의 범위)	7
[그림 2] 지능형 반도체 Ecosystem	13
[그림 3] 인공지능 발전과 컴퓨팅 패러다임의 변화	21
[그림 4] 인공지능 연산의 흐름	22
[그림 5] 핵심가치 도출 프레임워크	29
[그림 6] 분석 대상 applications 선정	30
[그림 7] 마케팅 - search advertising의 수요 변화	31
[그림 8] 마케팅 - social media advertising의 수요 변화	33
[그림 9] 의료 - patient data & risk analysis의 수요 변화	34
[그림 10] 의료 - medical imaging & diagnostics의 수요 변화	35
[그림 11] 보안 - identity and access management의 수요 변화	37
[그림 12] 보안 - antivirus/antimalware의 수요 변화	38
[그림 13] 제조 - predictive maintenance & machinery inspection의 수요 변화	39
[그림 14] 제조 - quality control의 수요 변화	40

요약

개요

[개념, 범위 및 대상 시장]

- (개념) 시스템에 내재화되어 인식·추론·학습·판단 등 지능형서비스를 구현하기 위한 SW와 SoC가 융합된 반도체
- (범위) 기계학습, 자연어, 상황인지컴퓨팅, 비전 등의 인공지능 SW와 프로세서, 메모리, 네트워크 등의 HW SoC가 융합된 인공지능 반도체라는 협의의 범위를 대상
 - ※ 광의로는 간단한 기능 구현을 위한 일반 SW와 SoC가 융합된 반도체까지 포함
- (대상 시장) 지능형 반도체용 프로세서(HW), 지능형 반도체(HW+SW)

[중요성 및 특징]

- (고성장 전망) Marketsandmarkets(2018)은 지능형(인공지능) 반도체 시장을 70.6억 달러('18)에서 592.6억 달러('25)로 연평균 35.5%의 고성장을 전망
- (인공지능의 높은 파급 효과) 최근 인공지능은 거의 모든 산업분야에서 활용가치와 파급력을 재평가 받으며, 모든 기술이슈를 잡아먹는 소용돌이
- (컴퓨팅 파워 진보) 인공지능의 3대 핵심요소인 데이터, 알고리즘, 컴퓨팅 파워 중, 가장 큰 제약요소였던 컴퓨팅 파워가 기술적으로 빠르게 진보

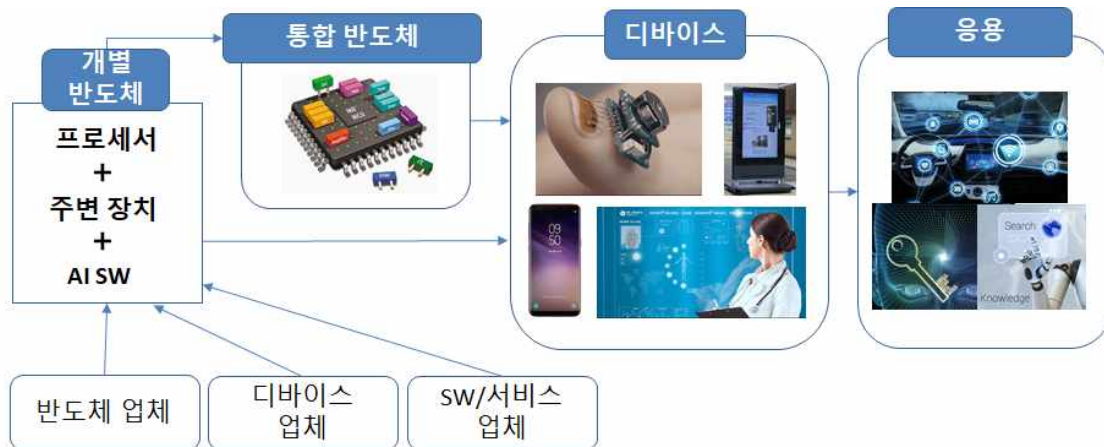


주요 동향 및 전망

[산업 구조 및 경쟁 동향]

- (산업 구조) 지능형 반도체 산업은 음성인식, 비전인식 등 특정용도를 위한 개별 반도체, 모바일 AP 등 통합 반도체, 디바이스, 응용 등으로 구성
- (경쟁 동향) 개별 반도체 영역에 기존 반도체 벤더뿐만 아니라, SW/서비스 벤더, 그리고 디바이스 벤더 등이 시장 선점을 위해 적극 경쟁
 - (반도체 업체) CPU, GPU, FPGA 등 기반 기술을 지능형 반도체 분야로 확장하고 있으며, 현재의 반도체 기술 한계를 넘어서기 위해 뉴로모픽 분야 등에 적극 투자
 - (디바이스 업체) HW 측면의 고객 접점을 최대한 확보하면서 시장 진입
 - (SW/서비스 사업자) 서버 비용 감소 및 효율적 고객 서비스 제공 등을 위해 지능형 반도체를 도입하고 있으며, SW와 빅데이터 보유라는 강점으로 시장 확장

지능형 반도체 Ecosystem



[기술 동향]

- 인공지능을 위해 대규모 데이터의 고속처리가 필요한데, 이를 효율적으로 감당하기 위한 방향으로 컴퓨터 하드웨어가 진화 전망
 - (異기종 통합) 기존 CPU에 특정 데이터 연산에 특화된 가속프로세서가 추가되며, 프로세서와 메모리의 병렬 연결을 확대하기 위해 2D에서 3D 구조로 진화
 - (同기종 통합) 병렬성 극대화를 위해 뉴로모픽칩과 같이 프로세서와 메모리를 한 곳에 복합하는 구조로 진화
- 또한 서버의 대규모 데이터를 edge로 분산하거나, edge에서 독립적으로 인공지능 작업을 수행하기 위한 기반 기술로서 edge용 지능형 반도체가 발전
 - 데이터센터의 서버는 주로 인공지능 학습/추론용으로 사용하고 edge는 추론용에 국한되어 사용되고 있으나, 향후 edge에서 학습까지 수행하는 방향으로 진화
 - Edge에서는 배터리 및 크기의 제약 때문에 단독형으로 탑재보다는 모바일 AP에 인공지능의 기능 일부가 통합되는 형태가 주류가 될 것으로 전망

[국내 R&D 투자 동향]

- (프로세서 코어) 현재 프로세서 코어는 거의 대부분 수입에 의존하고 있으나, 이를 국산화하기 위한 기술 개발 추진
- (차세대 프로세서) 차세대 CPU 코어 원천 기술인 PIM (Processor-in-memory) 및 연산기능과 메모리기능이 합쳐진 지능형 반도체인 뉴로모픽 칩 개발 추진
- (공정/설계) 미세 공정, 초절전 및 경량 설계를 위한 과제 진행 중
- (보안) 코드의 안전한 실행을 하드웨어적으로 보장하는 보안 아키텍처인 TEE (Trusted Execution Environment)를 기반으로 보안 솔루션 개발
- (응용) 자율주행자동차, 안전/환경, 개인행동 모니터링 등에서의 응용을 위한 기술 개발 계획 중



핵심가치 제안

[General suggestions]

- Processor 코어의 진화 및 혁신
 - (초고속/저전력 코어) 초고속 processor, 저전력 → GPU와 FPGA 기반은 선도적인 업체들이 포진해 있는 상황이며, 저비용/특정 목적용 ASIC 개발은 차별화 가능
 - (차세대 코어) 초초고속 processor, 초저전력, 발열 방지 (뉴로모픽, 3차원 구조)
 - 다수의 선도 기업/기관들이 기술 개발 중으로 기술 개발자의 차별화 가능성 사전 점검 필요하며, 프로그램의 유연한 수정이 가능한 뉴로모픽 칩 등 개발 고려
 - (IoT용 코어) 고속 processor, 초저전력, (에너지 하베스팅), 초저비용 → 다량의 칩이 소요되는 Niche 시장으로서 추진 가능
- Application specific chip 개발
 - 높은 성장이 예상되는 각 응용분야별 특화된 지능형 반도체의 개발을 통한 시장의 선점이 필요
 - 현재 선도 기업/기관들은 자율주행자동차 분야에 집중적인 투자를 하고 있는 상황이나, 마케팅, 의료, 보안, 제조 등의 분야에도 관심을 가질 필요
- 추론 모델의 통합 (at Edge)
 - Case1: Edge와 cloud 모두 training, cloud에서 학습된 추론 모델을 edge에서 추가
 - Case2: Cloud에서만 training, cloud에서 학습된 추론 모델을 edge에서 추가

[Niche suggestions]

- Training at Edge & 'Training + Inference' in one chip
 - Training in edge를 통한 개인화: 특정한 음성인식률/언어이해율을 향상하거나, 동네

병원 의사들의 개별적인 노하우 축적으로 차별화 가능

- Training + inferencing in one chip으로 하나의 칩에서 학습 및 추론을 수행함으로써 소형화와 실시간 학습 및 추론 프로그램 업그레이드 제공

■ Edge에서 시스템으로 작동하기 위한 통합 칩 설계 (유망 응용분야)

- (대화형 인터페이스) 디지털 사이니지에 내장되는 대화형 인터페이스 칩으로, 다국어 지원, 대화형을 위해 자연어처리/언어이해, 소란스러운 장소에서의 음성인식을 향상, 화자의 상황이해, cloud와 edge간 실시간 통신(query + response) 가능
- (인공지능 주치의) 환자 유전정보 기반 개인별 위험 관리 모델, 장시간 착용 및 인체 무해, 24시간 통신, 외부 충격에 강인하며 센서 성능 저하 시 보정 가능
- (고장관리) 방대한 센서 데이터를 유의미한 데이터만 필터링 + 센서 정보 보정, 수집된 자료(유의미한 데이터)의 실시간 cloud 전송, 추론 엔진의 주기적 갱신

■ 차별화 모델

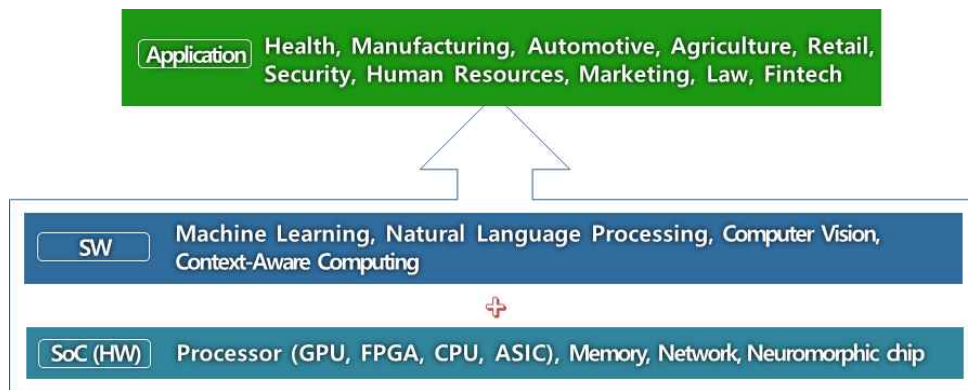
- 음성인식의 차별화: 비전인식 + 지향성 음성인식 in public space
- 개인 맞춤형 AI 주치의: 개인 유전정보 등을 기반으로 환자 개인별 추론 모델을 edge에 포함

I 개요

1 개념, 범위 및 대상 시장

- (개념) 시스템에 내재화되어 인식·추론·학습·판단 등 지능형서비스를 구현하기 위한 SW와 SoC가 융합된 반도체
- (범위) 기계학습, 자연어, 상황인지컴퓨팅, 비전 등의 인공지능 SW와 프로세서, 메모리, 네트워크 등의 HW SoC가 융합된 인공지능 반도체라는 협의의 범위를 대상
 - ※ 광의로는 간단한 기능 구현을 위한 일반 SW와 SoC가 융합된 반도체까지 포함
- (대상 시장) 지능형 반도체용 프로세서(HW), 지능형 반도체(HW+SW)

< 그림 1 > 지능형 반도체의 개념 (협의의 범위)



* 출처 : Marketsandmarkets(2018) 참고하여 작성

(참고) 기존 연구에서의 지능형 반도체의 개념

- (ETRI, 2018.) 시스템에 내재화되어, 관련 제품·서비스(스마트 자동차, 착용형 스마트 기기, IoT 등)의 지능형 서비스를 구현함으로써 새로운 부가가치를 창출하는 SW와 SoC이 융합된 기술
- (IITP, 2018a.) ICT 디바이스에서 인식·추론·학습·판단 등 지능형 서비스에 최적화(지능화, 저전력화, 안정화)된 SW와 SoC가 융합된 반도체



2 기술 분류

- 기관별 목적에 따라 다양하게 기술을 분류하고 있는데, 최근에 작성된 IITP(2018a)에서는 지능형 반도체를 지능화, 저전력, 고신뢰 기술로 크게 구분

< 표 1 > 지능형 반도체 기술분류

중분류	소분류	개념
지능화 기술	초병렬 매니코어 프로세서	병렬로 수천 개 이상의 쓰레드를 동시 실행할 수 있는 매니코어 구조 프로세서 및 뉴럴넷 코어를 융합한 인공지능 전용 어플리케이션 프로세서 설계기술
	딥러닝 뉴럴넷 코어	딥러닝을 위한 뉴럴넷(Neural Network) 연산에 최적화한 프로세서 코어 및 컴퓨팅 기술
	초고속 온칩 프로세서-메모리	인공지능 컴퓨팅이 요하는 대용량의 데이터를 저장 및 높은 대역폭의 프로세서-메모리간 전송 기술
	재구성형 인공지능 회로설계	인공지능 프로세서의 상용화를 위한 데이터 타입 최적화 컴퓨팅 설계
	지능화 소자 및 공정 기술	뉴런 및 시냅스의 가소성(Plasticity)을 보이는 디바이스 제조를 위한 반도체 공정 및 장비 기술
저전력 기술	저전력 로직 아키텍처 및 IP	약인공지능 프로세싱 기능을 갖춘 프로세서, 마이크로 컨트롤러 등에 널리 사용되는 핵심IP로서 전력 소모량과 면적 최소화 회로 설계 기술
	저전력 차세대 메모리 아키텍처	온칩 및 오프칩 메모리의 전력소모량을 줄이기 위한 동일전력 대용량 메모리 집적, 동적전압 조절, 메모리 아키텍처 변경에 의한 저전력 메모리 기술
	저전력 무선 통신 설계	무선 통신을 통하여 칩간의 데이터 전송시 소모되는 전력소모량을 최소화 하는 IoT, IoE용 통신 반도체
	저전력 소자 및 공정 기술	반도체의 전력소모량을 최소화 하기 위한 반도체 소자(Device) 기술로서 트랜지스터 레벨의 저전력 기술
고신뢰 기술	기능안전 기반 반도체 설계	안전 동작이 필수적인 자율이동체(자율주행차, 가정용/산업용로봇, 드론 등)에서 설계 의도대로 기능(Function)이 상시 안전하게 동작하며, 고장시에는 사용자 또는 외부에 즉각 보고하고 안전한 상태로 반도체를 유지하는 기술
	정보보안 반도체 설계 기술	반도체의 신뢰성 있는 상시동작을 위하여 온도, 열, 전압, 전류 등의 변화에 둔감하여 신뢰성 있는 동작을 보장하는 소자의 설계 및 신뢰성 검증 기술
	고신뢰 반도체 소자 설계	반도체가 온도, 열, 전압, 전류 등의 변화에 둔감 하여 상시 신뢰성 있고 안정적인 동작을 보장하는 소자의 설계 및 신뢰성 검증 기술

* 출처 : IITP(2018a)

(참고) 기타 지능형 반도체의 기술 분류

- (IITP, 2018a.) 인체 부착형 지능형 반도체, 지능형 차량용 SWSoc, 지능형 IoT SW-SoC, SW 통합 매니코어 반도체, 인지 컴퓨터 뉴럴-딥러닝 반도체, 바이오생체분석 반도체 등
- (ETRI, 2018.) 인공지능 프로세서, 매니코어 프로세서, 지능형 엣지, 고속 신호처리 SoC
- (중소벤처기업부, 2018.) IITP(2018a)와 유사하게 초지능화, 초경량화, 안정화의 3가지 기술로 구분

3 중요성 및 특징

가. 시장규모 및 성장률

- Marketsandmarkets(2018)은 지능형(인공지능) 반도체 시장을 70.6억 달러('18)에서 592.6억 달러('25)로 연평균 35.5%의 고성장을 전망

< 표 2 > 지능형 반도체 세계시장전망 (억 달러)

		2015	2016	2017	2018	2019	2021	2023	2025	CAGR ('18-'25)
Total		28.1	38.2	51.9	70.6	95.9	177.0	327.4	592.6	35.5%
by HW	Processor	23.2	31.3	42.2	56.9	76.7	139.1	252.7	448.8	34.3%
	Memory	3.1	4.4	6.3	8.9	12.7	25.5	51.1	100.0	41.2%
	Network	1.8	2.5	3.5	4.8	6.5	12.4	23.5	43.8	37.3%
by SW	Machine Learning	13.2	18.3	25.2	34.8	47.9	90.8	171.9	318.8	37.2%
	Natural Language	8.7	11.6	15.4	20.4	26.9	46.3	78.9	132.5	30.6%
	Context Aware	5.1	6.9	9.3	12.5	16.7	30.2	54.6	96.6	34.0%
	Computer Vision	1.0	1.4	2.1	3.0	4.4	9.7	22.0	44.7	47.2%
by App.	Marketing	10.2	13.6	17.8	23.3	30.1	49.4	78.4	120.6	26.5%
	Security	7.8	10.5	14.0	18.5	24.3	41.1	67.2	106.3	28.4%
	Fintech	2.7	3.9	5.6	8.0	11.3	22.2	41.9	76.5	38.1%
	Healthcare	1.8	2.6	3.8	5.5	8.1	17.7	39.2	88.1	48.6%
	Retail	1.9	2.6	3.7	5.3	7.5	15.4	32.2	68.3	44.2%
	Automotive	1.2	1.7	2.6	4.1	6.2	13.6	27.9	51.1	43.6%
	Manufacturing	0.7	1.0	1.5	2.3	3.7	9.9	27.6	59.2	58.8%
	Law	1.0	1.2	1.5	1.9	2.4	4.0	6.9	12.3	30.6%
	Agriculture	0.7	0.9	1.1	1.3	1.6	2.6	4.1	6.4	25.5%
Human Resource	0.2	0.3	0.4	0.5	0.6	1.1	2.0	3.8	34.0%	

* 출처 : MarketsandMarkets(2018) 재구성

도표의 저작권은 MarektsandMarkets에 있으며, MarektsandMarkets의 동의하에 사용되었습니다. 추후 이용 시 MarketsandMarkets에 문의하시기 바랍니다.

- (HW 기준) Processor, Memory, Network로 구분 가능하며, 시장 비중에서는 Processor가 가장 높으며 (76%, '25), 성장률은 Memory가 가장 높을 것으로 전망(41.2%, '18-'25)
- (SW 기준) Machine learning, Natural language processing, Context-aware computing, Computer vision으로 구분 가능하며, 시장 비중에서는 Machine learning이 가장 높으며 (54%, '25), 성장률은 Computer vision이 가장 높을 것으로 전망(47.2%, '18-'25)
- (Application 기준) Marketing, Security 등으로 구분 가능하며, 시장 비중에서는 Marketing이 가장 높으며 (20%, '25), 성장률은 Manufacturing이 가장 높을 것으로 전망(58.8%, '18-'25)



나. 성장 동인

- 최근 인공지능은 거의 모든 산업분야에서 활용가치와 파급력을 재평가 받으며, 모든 기술이슈를 잡아먹는 소용돌이
 - 글로벌 ICT 기업들은 발 빠르게 변화에 대응하며 자사의 전략 방향 확립 중
 - ※ MS는 누구나 AI 툴 사용해 개발하는 'AI 민주화' 목표로, 혁신, 속도, 신뢰라는 3가지 방향성 제시 (에너지경제, '18.11.7)
 - ※ 구글은 '모든 제품에 적용한다, 누구나 쓰게 한다'는 AI 대원칙을 발표하는 등 AI를 구글의 최종 도착지가 될 것이라는 목표를 밝힘 (조선비즈, '17.11.28.)
 - 인공지능은 개인, 산업, 사회에 광범위한 파급 효과를 유발하며, 인간 삶의 방식을 근본적으로 변화시킬 것으로 전망
 - ※ Bank of America는 로봇과 인공지능 확산으로 인해 많은 산업에서 30% 가량 생산성이 향상되고, 제조 노무비가 18~33% 절감될 것이라고 전망
 - ※ 세계경제포럼(WEF, 2016)은 세계 주요 15개국에서 '20년까지 510만 개의 일자리 감소 전망

- 또한 사물인터넷(IoT)이 현실화됨에 따라 인공지능을 모든 사물에 적용하기 위한 니즈가 빠르게 확산되면서 인공지능 반도체의 수요 증대 전망
 - 사물인터넷의 센서, 통신, 프로세싱 기능을 담당하는 반도체에 인공지능을 탑재한 저전력 고성능 인공지능 반도체 수요가 증대될 것으로 예상됨

- 인공지능의 3대 핵심요소인 데이터, 알고리즘, 컴퓨팅 파워 중, 가장 큰 제약요소였던 컴퓨팅 파워가 기술적으로 빠르게 진보
 - 인공지능은 그 개념이 발현된 지 반세기만에 최근에 빅데이터, 딥러닝 알고리즘, GPU등 연산가속 HW 발달 등으로 이전과 다른 새로운 발전 국면을 맞이
 - ※ 인공지능(Artificial Intelligence)이란 용어는 1956년 존 매카시(John McCarthy)가 '기계가 인간 행동의 지식과 같이 행동하게 만드는 것'으로 최초 정의
 - (알고리즘) 최근 각광받고 있는 심층신경망(DNN: Deep Neural Network)은 50년이 넘는 오랜 기간 동안 진전되어온 인공신경망 알고리즘에 기초

- ※ 딥러닝의 시발점을 Hinton et al. (2006)으로 삼기는 하나, 신경망의 기본개념인 퍼셉트론(Perceptron)은 '50년대 Frank Rosenblatt에 의해 제안되었고, 딥러닝 연산의 핵심인 다층신경망에서의 역전달(back propagation)알고리즘은 80년대 Hinton과 그 동료들에 의해 제안
- (데이터) '90년대 인터넷 시대가 열리면서 방대한 디지털 데이터의 생산 및 축적이 가능해졌고, 이는 '00년대 이후, ICT산업이 부흥으로 급격히 가속화
 - ※ 세계 데이터규모는 '16년 16ZB에서 '25년 163ZB로 연평균 29%의 고성장 전망 (IDC,2017)
- (컴퓨팅 파워) '06년에 대규모 병렬 연산처리에 효과적인 방안이 제시되면서, 인공지능 발전의 가장 큰 병목이었던 컴퓨팅 파워 문제의 새로운 돌파구가 열림
 - ※ '06년 그래픽연산 프로세서(GPU) 생산업체인 엔비디아는 GPU를 일반적인 병렬 연산 처리에 활용하는 GP-GPU 개념과 개발 툴(CUDA)를 발표하면서, 기존 CPU 중심의 연산처리의 한계를 극복하는 새로운 방향성을 제시
 - ※ '12년 ILSVRC(Imagenet Large Scale Visual Recognition Challenge, 약칭 이미지넷)¹⁾에서 GPU를 활용한 딥러닝 시스템인 Alexnet이 현격한 성능우위로 우승을 하면서 가능성 확인

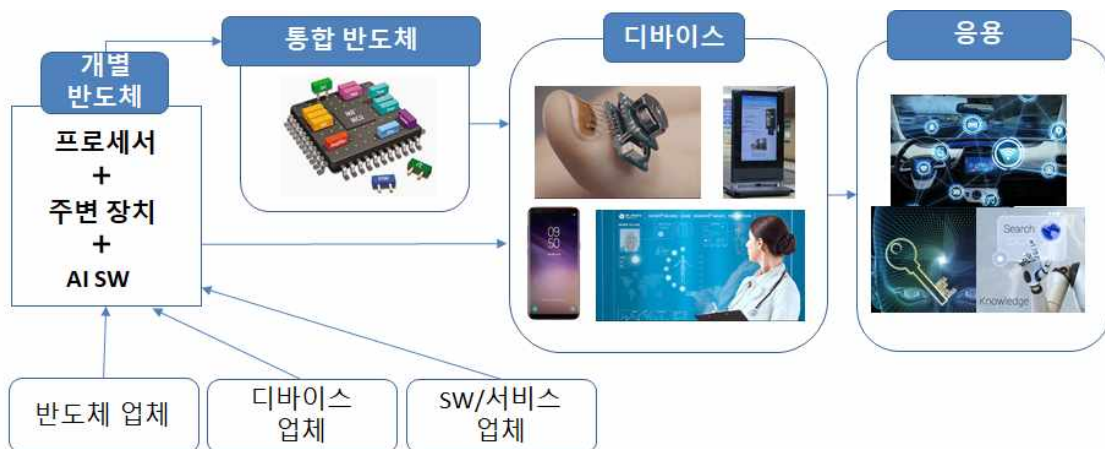
1) 이미지넷은 1000개가 넘는 카테고리로 분류된 100만개의 이미지를 인식해 그 정확도를 겨루는 대표적인 시각지능 대회

II 주요 동향 및 전망

1 산업 구조

- 지능형 반도체 산업은 음성인식, 비전인식 등 특정용도를 위한 개별 반도체, 모바일 AP 등 통합 반도체, 디바이스, 응용 등으로 구성
 - 특정용도의 개별 반도체는 통합 반도체에 포함되어 디바이스에 탑재되거나, 개별적으로 디바이스에 탑재
- 개별 반도체 영역에 기존 반도체 벤더뿐만 아니라, SW/서비스 벤더, 그리고 디바이스 벤더 등이 시장 선점을 위해 적극 경쟁
 - (반도체 업체) CPU, GPU, FPGA 등 기반 기술을 지능형 반도체 분야로 확장하고 있으며, 현재의 반도체 기술 한계를 넘어서기 위해 뉴로모픽 분야 등에 적극 투자
 - (디바이스 업체) HW 측면의 고객 접점을 최대한 확보하면서 시장 진입
 - (SW/서비스 사업자) 서버 비용 감소 및 효율적 고객 서비스 제공 등을 위해 지능형 반도체를 도입하고 있으며, SW와 빅데이터 보유라는 강점으로 시장 확장

< 그림 2 > 지능형 반도체 Ecosystem



* 출처 : 저자 작성 (세부 그림은 google 이미지 검색)



2 사업자 동향

가. 반도체 업체(Chip maker)

- (Intel) 서버용 CPU 시장점유율 세계 1위, PLD 및 FPGA 분야 세계 2위 업체인 Intel은 차세대 서버와 IoT 시장 대응을 위한 기술기업 인수 및 기술 확보 중
 - (CPU) 절대적인 경쟁력을 보유한 CPU 아키텍처를 cloud의 인공지능 워크로드 가속화와 edge용으로 활용하고 있으며, 하드웨어 기반 보안 솔루션인 SGX를 도입
 - ※ Intel Xeon Phi (cloud), Intel Core i9-9900K processor (edge)
 - ※ SGX (보안): CPU 지시 코드들의 내장으로, 블록체인과 인공지능 워크로드에서 처리되는 민감 정보를 보호
 - (FPGA) FPGA 제조업체인 Altera를 인수('15)하고 전력 효율을 높일 수 있는 FPGA 기술에 투자를 확대하고, FPGA를 자사의 강점인 CPU와 통합한 프로세서를 제공
 - ※ Stratix 10SX FPGA (cloud), Intel Arria 10 FPGAs (edge), Xeon Gold 6138P (CPU + FPGA)
 - (ASIC) Movidius (edge)는 Intel의 주목할 만한 제품으로, 드론, 스마트카메라, AR/VR 등 시각 기반 기기의 딥러닝과 AI 가속화를 지원하고 있으며, Mobile eye (edge)는 자율주행자동차 전용으로 Mobile eye와 공동 개발
 - ※ Movidius Myriad X: compute capacity: > 4 TOPS, On-chip accelerators: 20+ image/vision processing accelerators, Neural network capability: Neural compute engine (up to 1 TOPS), 공정: 16nm FFC (TSMC)
 - (뉴로모픽 등) 기업 인수를 통해 뉴럴 네트워크 프로세서 기술을 확보하였고, 프로세서 자체에 인공 뉴런을 탑재하는 뉴로모픽 칩 개발에 적극 투자
 - ※ Nervana NNP(Neural network processor, cloud): 인공지능 워크로드 가속화를 위한 텐서코어를 핵심으로 하는 칩 개발 스타트업인 너바나 시스템즈를 인수(2016), 기업 대상 추론용 전용칩(NNP-I, 2017) 및 훈련용 전용칩(NNP-T, 2019 예정) 개발
 - ※ Loihi (cloud): 뉴로모픽칩 (프로세서 자체에 인공 뉴런 탑재)
- (Nvidia) PC용 GPU에서 시작, GPU 분야 세계 시장 점유율 1위 기업으로, 서버용 그래픽 가속기 개발 및 CPU와 GPU 통합 프로세서로 모바일 분야 진입

- 그래픽카드의 가격 하락에도 불구하고 GPU 기반 인공지능 반도체가 새로운 성장 동력이 되고 있으며, GPU 기반의 cloud 및 edge 인공지능 칩 기술을 선도
- (GPU) GeForce(게임), Quadro(디자인), Tesla and DGX (AI data 연구), GRID (클라우드 기반 비주얼 컴퓨팅) 등의 전문 시장에 제품 제공
 - ※ Volta: Tensor core GPU architecture
 - ※ Tesla (cloud): Volta 기반, 빠른 속도로 큰 규모의 시뮬레이션 실행에 필요한 성능 제공
 - ※ Nvidia TensorRT (cloud): 이기종 GPU로 구성된 인프라 환경에서 최적의 성능과 자원 활용 유도
 - ※ Xavier (edge): 자율 주행자동차용 컴퓨터, 90억 개 이상의 트랜지스터 이용 30 TOPS 이상 제공, 6종의 고성능 프로세서 보유(512 Volta tensor core GPU, 8코어 ARM64 CPU, 듀얼 NV임 심층 학습 가속기, 이미지 프로세서, 비전 프로세서, 비디오 프로세서)
 - ※ DGX-1 (cloud): 기존 NVIDIA GPU의 클라우드 딥러닝 스택에 비해 4배 이상의 training 속도 제공
- (Tegra processors) 전체 컴퓨터를 단일 칩에 통합하고, GPU 및 멀티 코어 CPU를 통합하여 자율 로봇, 무인 항공기 및 자동차와, 모바일 게임 및 엔터테인먼트 디바이스를 구동하는 테그라 (Tegra) 브랜드 제품을 제공
 - ※ Nvidia Tegra (edge, CPU + GPU): ARM 프로세서와 GPU를 통합한 단일 칩 시스템으로 2008년 최초 출시
 - ※ Tegra x1(2015년 출시): ARM Cortex-A57 코어 4개, ARM Cortex-A53 코어 4개, 20 nm 공정, ARM v8 프로세서, 256코어 맥스웰 GPU (16bit 1024GFLOPS)
- (Xilinx) FPGA를 포함한 PLD 분야에서 세계 시장 점유율 1위 기업으로, '데이터센터 퍼스트'를 핵심전략으로 FPGA 기반의 cloud 시장 공략에 주력하고 있으며, 응용분야로는 ADAS, Drones, Medical 등으로 확장
 - (3D IC) 단일 패키지에 FPGA, 트랜시버, 와이드 메모리 인터페이스를 결합하여 단일 소자의 용량과 대역폭을 초과하는 3D IC를 설계 및 개발
 - (ACAP) 적응형 컴퓨팅 가속화 플랫폼으로, FPGA, AP, HBM, HW, SW 등을 하나로 통합하여 다양한 응용분야와 작업 요구에 대응하기 위해 설계된 제품으로, 기존 FPGA 및 CPU+GPU의 기능을 훨씬 능가 전망
 - ※ 현재의 데이터센터 시장은 기존 CPU + GPU를 기반으로 한 Intel 및 Nvidia가 시장 주도 중



- FPGA의 장점(데이터 센터에서의 워크로드의 추가/제거 용이)이 요구되는 cloud 및 응용분야(자동차 등)에서의 경쟁력 강화 전망
 - ※ (Alveo) 데이터센터용 적응형 컴퓨터 가속 플랫폼
 - ※ (Versal) 7nm FinFET 공정 기술(TSMC)을 기반으로, 스칼라 프로세싱 엔진, 적응형 하드웨어 엔진, 지능형 엔진을 최첨단 메모리 및 인터페이스 기술과 결합하여, 응용 서비스에 강력한 이종 가속 기능 제공 (2019년 출시 예정)
 - ※ (Zinq) 듀얼-코어 ARM Cortex-A9 MPCore 프로세서, FPGA, 핵심 주변기기를 하나의 칩에 통합한 것으로 ADAS 등의 분야에 응용
- (AMD) 세계 CPU 시장점유율 2위 업체로, 자사의 CPU 및 GPU를 연계한 고성능 이종 시스템 아키텍처를 통해 HPC 시장 확장 중
 - (RYZEN) 젠 아키텍처를 이용한 AMD의 CPU
 - (Radeon graphic card) 데이터 센터 시각화 워크로드를 가속화하기 위한 듀얼 GPU 인프라
 - ※ Radion Pro WX 8200 graphics card(2018): 실시간 시각화, 가상 현실 (VR) 및 사실적 렌더링을 위한 워크 스테이션 그래픽 성능 제공
 - (7nm Vega GPU) 7nm 공정 및 GPU 기반 Vega 그래픽 카드
 - (APU) 범용 연산의 CPU와 3차원 기하학 연산의 GPU를 통합 (2011년 공개)
- (ARM) IP에 집중하는 기업으로, 모바일 프로세서 코어, 그래픽 프로세서, 그리고 이 둘을 하나의 칩에 통합하여 시장 공략
 - ※ 칩 회사인 ARM은 인터넷 인프라를 보유한 소프트뱅크에게 인수('16)되어 IoT 시장의 시너지 창출 기대
 - (Cortex CPU) 모바일 프로세서 코어로, Qualcomm, Xilinx, Huiwei, Samsung 등 다수의 대기업에서 기반 코어로 활용 중
 - (ARM Mali GPU) ARM 아키텍처 기반의 AP에서 사용하기 위한 GPU
 - (보안) 별도의 보안 하드웨어 칩을 쓰지 않고 하나의 CPU에서 Secure World와 Normal World를 분리하여 동작
- (Qualcomm) 스마트폰 시장의 과도한 집중을 탈피하고, 자동차, 헬스케어, 의료, 로봇 응용 등 기계 학습 솔루션에 투자 집중

- (Snapdragon) ARM 기반 CPU, Modem, Adreno GPU 등을 통합한 One-chip Solution으로, GPU 코어 부분은 AMD의 소유였던 Adreno GPU를 인수 및 자체 설계
- (Spectra) Snapdragon에 탑재되는 이미지 센서 프로세서
 - ※ Spectra 270: 1,600만 화소 센서, 4K, 60fps 영상 듀얼 스트리밍 지원
- (QCS) 10nm 공정으로, ARM 기반 CPU, Adreno GPU, Spectra 등을 탑재하여, 360도 카메라, 로봇 청소기, 스마트 디스플레이 등의 IoT에 활용
 - ※ QCS400은 원거리 음성 인식 기능이 포함 (2019 출시)
- (DDP: Drive data platform) Snapdragon neural processing engine 기반의 주행 정보 플랫폼
 - ※ 차량 반도체 분야 시장 점유율 1위인 영상 처리 엔진 업체 NXP 인수를 통해 자동차 (커넥티드카) 분야 강화 예상

< 표 3 > 주요 벤더 - 반도체 업체

	Processor/Chips	주요 응용분야
Intel	- NNP (cloud) - CPU (cloud, edge) - FPGA (cloud, edge) - ASIC (cloud, edge) - 뉴로모픽 (cloud)	Datacenter, Self-driving cars, AR/VR, Drones, Surveillance
NVIDIA	- GPU (cloud, edge) - ASIC	Datacenter, Self-driving cars, Retail Analytics, Smart Cities, Surveillance
Qualcomm	- AP (CPU+GPU, edge) - ISP (Image sensor processor)	Computer vision, Self-driving cars
XILINX	- FPGA (cloud, edge)	Datacenter, ADAS, Drones, Medical
AMD	- CPU (cloud) - GPU (cloud, edge)	Datacenter, Game
ARM (IP Licensor)	- CPU (edge) - ASIC (edge)	Voice assistants, Consumer robots, Smartphones
Cambricon (IP Licensor)	- ASIC (edge)	Self-driving cars

* 출처 : PWC Strategy&(2018), IITP(2018b), KISTEP(2019) 등 참고



나. SW/서비스 업체

- (Google) 자사의 데이터센터(전력 감소) 등 클라우드용의 인공지능 반도체를 개발하고 있으며, 고객 데이터를 Google cloud에 저장하고, TPU를 통해 머신러닝 학습을 가속화하고, edge TPU에서 추론을 하는 등 cloud와 edge의 연계 전략 추진
 - (TPU) ASIC 기반의 인공지능 가속기로, 클라우드 서비스의 신경망/기계 학습용으로 개발 (2016 공개)
 - ※ 현재 최대 180 테라플롭(Teraflops) 연산 제공
 - ※ 구글의 데이터센터는 TPU 칩을 탑재하여 에너지 비용을 15% 절감
 - (Edge TPU) ASIC 기반의 소형 인공지능 가속기로, IoT 기기에서 Google의 머신러닝 오픈소스 텐서플로를 구동할 수 있도록 고안 (2017 공개)
- (Microsoft) 효율적인 서버 운용과 응용분야의 경쟁력 확보를 위해 인공지능 반도체 투자
 - (Catapult) FPGA 기반의 칩을 탑재한 서버메인보드
 - ※ 검색엔진 Bing의 성능은 CPU 기반의 경우에 비해, 30%의 비용절감과 10% 절전 가능
 - ASIC 기반의 HPU (Holographic Processing Unit)을 개발하여 자사 홀로렌즈에 탑재 (2017)
 - ※ HPU는 현실세계와 증강현실에 대한 데이터를 통합 처리하는 프로세서로, 음성 및 제스처도 처리
- (Facebook) ASIC 기반 인공지능 반도체 칩 개발 착수
 - 음성 제어, 자연어 처리, 신경 네트워크 방식을 통한 전력 절감
 - ※ 스마트 디지털 도우미: 어떤 주제에 대해서도 상식적인 토론이 가능
- (Amazon) 인공지능 칩의 외부 조달에서 자체 칩 제작으로 전략 수정
 - (Inferentia chip) 음성의 텍스트 변환 등이 가능한 인공지능 칩으로 자체 클라우드 고객에게 서비스(2020년 예정)
 - ※ 기존에는 Intel과 Nvidia 등의 주요 공급업체에서 인공지능 칩을 조달
 - (Graviton) ARM 기반 기술을 데이터센터에 적용하기 위한 기술 개발 중

- (Alibaba) 중국 반도체 설계 기업 C-스카이 마이크로시스템을 인수(2018)하고 인공지능 반도체 개발에 본격 진입
 - (Ali-NPU) 영상비디오 분석, 머신러닝 등 인공지능 추론 계산 등에 적용할 수 있는 신경망 칩으로, 이미지 분석, 동영상 식별, 클라우드 컴퓨팅 등에 활용 (2018년 개발 계획 발표)
 - ※ 기존 CPU + GPU 결합 AI 칩대비 10배 성능, 제조비용 절반, 전력 소비 절반으로 설명
- (Baidu) 구글을 모델로 사업을 확장하고 있으며, 중국 내 확고한 입지 구축 및 세계 시장 진입 가능성이 있는 Baidu는 Nvidia의 GPU 기반 데이터센터용 AI 가속기를 도입하였으며, 자체 칩 개발에도 적극적
 - (Kunlun) FPGA 기반의 AI 칩으로 오픈소스 딥러닝 알고리즘을 지원하며, 데이터센터, 음성 인식, 자연어 처리, 대규모 추천, 자율 주행 등에 활용
 - ※ 14nm 공정(삼성전자), 260TOPS, 100W 전력 소비 (구글의 TPU는 45TOPS)
 - (Apolong) 중국 기반 자동차 회사인 Xiamen King Long United Automotive Industry와 협력으로 자율주행자동차용으로 개발하였으며, 중국과 일본의 일부 대도시에서 미니 버스(4레벨 자율주행)에 배치 계획 (2019 예정)

< 표 4 > 주요 벤더 - SW/서비스 업체

	Processor/Chips	주요 응용분야
Google	- ASIC (TPU, cloud, edge)	Datacenter, Image search, Voice search, Translate, Smart home, Smart factory
Microsoft	- ASIC (Hololens, edge) - GPU (NVIDIA 기반) - FPGA (Intel 기반)	Digital transformation, Intelligent assistant
Amazon	- ASIC (Alexa, edge) - GPU (NVIDA 기반) - FPGA (Xilinx 기반)	Facial recognition, Text-to-speech, Smart assistant
Alibaba	- Neural Engine (알리-NPU)	
Baidu	- FPGA (Xilinx 기반) (cloud)	Search, Voice assistant, Computer vision

* 출처 : PWC Strategy&(2018), IITP(2018b), KISTEP(2019) 등 참고



다. 디바이스 업체 (Device maker)

- (Apple) 개인정보 유출 방지와 처리속도 개선을 위해 인공지능 칩 개발
 - (NPE: Neural processing engine) 빠른 이미지 인식/처리가 가능한 엔진으로 A11 칩셋에 포함 (2017)
- (Samsung) Apple과 Huawei 대비 인공지능 AP의 상용화가 늦었지만, 세계 최초 7nm EUV 공정 적용한 획기적 성능이 기대되며, 서버용 인공지능 반도체 개발도 진행 중
 - (엑시노스) 초고속 통신모뎀, 이미지/얼굴인식 및 실시간 번역 등이 가능한 신경망 기능의 인공지능 설계기술 적용한 AP로 향후 자동차 전장 분야로 확장 적용 예정
 - ※ 갤럭시S10(2019)에 적용 예정이며, 7nm EUV(차세대 노광 장비) 공정 적용, ARM 기반 독자기술의 CPU(몽구스), 초고속 모뎀, 지능형 이미지 처리 등이 포함
- (Huawei) AI, 칩 설계 등에서 세계 최고 수준의 기술을 확보하기 위한 노력 중
 - (Kunpeng920) 7nm 공정(TSMC) 사용, ARM 기반 CPU를 자체 설계
 - (Kirin 980) 스마트폰에서 프로세서용으로 사용되는 7nm 인공지능 칩으로, CPU, GPU 외 인공지능 연산만 따로 처리하는 NPU를 탑재
 - ※ NPU(신경망 처리장치)는 Cambricon이 공급
 - ※ 동영상을 배경과 오브젝트로 분리: 동영상의 경우 초당 30프레임 이상의 사진에 대한 연산 필요
 - (Ascend) 서버 및 edge용 AI 가속기로 자사 제품에만 활용 계획
 - ※ Ascend910(cloud): 교육/통신 등, Ascend310(edge): 스마트폰, 스마트워치 등

< 표 5 > 주요 벤더 - 디바이스 업체

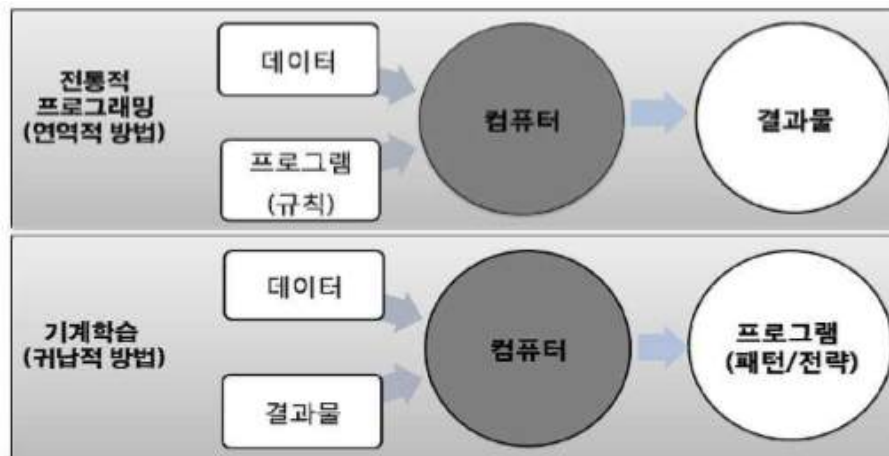
	Processor/Chips	주요 응용분야
Apple	AP(edge), Neural Engine	Facial recognition, Animated emoji
Samsung	AP(edge), Neural Engine	Facial recognition, Animated emoji
Huawei	CPU (ARM 기반), NPU (edge), AP (edge)	Datacenter, 5G
Tesla	GPU (NVIDIA 기반), ASIC (AMD 협력, edge)	Self-driving cars

* 출처 : PWC Strategy&(2018), IITP(2018b), KISTEP(2019) 등 참고

3 기술동향

- 인공지능의 발전은 기존 컴퓨팅 패러다임의 변화를 야기
 - 기존 컴퓨터가 사람이 정해놓은 규칙을 따라 데이터를 처리, 결과물을 도출했다면, 인공지능은 입력 데이터와 결과 값이 주어지면, 스스로 규칙을 찾아 학습하는 방식
 - ※ 일반적으로 학습을 통해 사람이 만든 모델보다 좋은 결과를 보여주며, 컴퓨터에게 일일이 규칙(프로그램)을 작성해 주지 않아도 되기 때문 효율적

< 그림 3 > 인공지능 발전과 컴퓨팅 패러다임의 변화



* 출처 : STEPI(2016)

- 인공지능 연산은 신경망의 최적 가중치를 찾는 학습(training)과 신규 데이터를 적용하는 추론(inference)의 2단계 과정으로 구분
 - 학습(training) : 딥러닝의 학습은 정의된 신경망에 대규모 데이터를 돌려가며, 신경망의 최적 가중치를 찾아가는 과정으로, 주로 데이터센터 서버에서 진행
 - ※ 신경망 복잡도가 높을수록, 학습되는 데이터가 많을수록 분석 정확도 높아지나, 요구 컴퓨팅파워는 높아짐. '12년 이미지넷에서 우승한 AlexNet은 8개 층을 가진 신경망이었으나, '15년 우승한 ResNet은 152개 층을 갖는 심층망으로 구성
 - 추론(inference) : 학습이 끝난 신경망에 신규 데이터를 적용하는 과정으로, 실행



(Run, execution)으로도 표현. 신경망 최적화 수준에 따라 품질이 좌우

※ 데이터 센터에서 진행이 되어 사용자 단말로 전달되는 것이 일반적 서비스 제공 방식이었으나, 향후에는 단말에서의 추론처리가 일반화 될 것으로 전망

< 그림 4 > 인공지능 연산의 흐름



- 인공지능은 단순한 연산의 수없는 반복으로 구성되어 계산량이 매우 큰 특징이 있으나, 現컴퓨터 아키텍처는 이를 효과적으로 지원하는 대규모 병렬연산에 한계
 - 두 가지 측면에서 한계가 존재: ① 현재 컴퓨터 구조의 데이터 병목 문제, ② 무어의 법칙으로 대변되는 반도체 집적도의 물리적 한계
 - 폰 노이만(von-Neumann) 구조의 데이터 병목 문제: 주기억 장치, 중앙 처리 장치, 입·출력 장치로 이어지는 직렬처리(Serial Processing)는 최근 요구되는 고속 병렬 연산 에서 심각한 데이터 병목 현상이 발생
 - ※ 하나의 CPU가 중앙에서 모든 데이터를 처리·제어하므로, 연산량이 많아질수록 메모리와 CPU사이의 병목현상 심각
 - 무어의 법칙(Moore's Law)²⁾의 물리적 한계: 반도체 집적도를 높임으로서, 데이터 처리 속도를 높일 수 있었으나, 점차 미세공정 고도화의 물리적 한계 봉착 中
 - ※ 집적도를 높일수록 심화되는 발열과 간섭 등의 문제점도 해결과제임

2) 트랜지스터의 집적도가 18개월 내지 24개월마다 2배씩 늘어나는 것을 의미

- 인공지능을 위해 대규모 데이터의 고속처리가 필요한데, 이를 효율적으로 감당하기 위한 방향으로 컴퓨터 하드웨어가 진화 전망³⁾
 - (헛기종 통합) 기존 CPU에 특정 데이터 연산에 특화된 가속프로세서를 추가하는 헛기종 시스템 구조가 대규모 데이터 고속처리의 해결책으로 대두
 - ※ 헛기종 시스템 구조: CPU는 시스템 부팅 및 컨트롤 등을 담당하는 호스트 프로세서를 담당, 특정 연산에 특화된 GPU 등을 CPU와 메모리를 공유하여 하나의 연산 장치처럼 활용
 - 헛기종 시스템 구조는 프로세서와 메모리의 병렬 연결 확대를 위해 3D HBM + PIM의 구조로 진화 전망
 - ※ (HBM: high bandwidth memory) GPU와 메모리 사이 수 백만 개의 병렬 연결선 확보를 위해 2D구조에서 3D구조로 진화, 연결선은 전자빔 도파관 구조 등으로 1개 연결선당 100Gbps 이상 확보
 - ※ (PIM: processor in memory) 메모리 위에 프로세서 설치
 - (동기종 통합) 프로세서와 메모리를 한곳에 복합(예: 뉴로모픽칩)한 구조로 진화
 - 병렬 데이터의 외부 유출입 필요가 없어 병렬성 극대화와 동시에 전력 소모 및 시간 지연 문제가 해결 가능하나, 인공지능 알고리즘을 마음대로 바꾸지 못하는 등 시스템 유연성(flexibility)이 떨어져 응용 분야가 제한될 수 있음

- 또한 서버의 대규모 데이터를 edge로 분산하거나, edge에서 독립적으로 인공지능 작업을 수행하기 위한 기반 기술로서 edge용 지능형 반도체가 발전⁴⁾
 - 데이터센터의 서버는 주로 인공지능 학습/추론용으로 사용하고 edge는 추론용에 국한되어 사용되고 있으나, 향후 edge에서 학습까지 수행하는 방향으로 진화
 - ※ Edge에서 인공지능을 실행 시 실시간 성능과 처리 속도가 보장되고, 네트워크에 연결되어 있지 않아도 서비스가 가능하며, 프라이버시 문제 해결과 클라우드 서버의 부하를 줄일 수 있다는 장점이 있음
 - Edge에서는 배터리 및 크기의 제약 때문에 단독형으로 탑재보다는 모바일 AP에 인공지능의 기능 일부가 통합되는 형태가 주류가 될 것으로 전망

3) 뉴스핌(2019.2), 인공지능 반도체의 미래 (<http://www.newspim.com/news/view/20190217000152>) 참고

4) KISDI(2018), KISTEP(2019) 참고



4 국내 R&D 투자 동향

가. 공공 R&D

- 프로세서 코어
 - 현재 프로세서 코어는 거의 대부분 수입에 의존하고 있으나, 이를 국산화하기 위한 기술개발 추진 중
 - ※ ETRI, KETI, KAIST 등은 다양한 구조 및 성능의 프로세서 코어 기술을 개발하고 이를 상용화로 진입시키기 위한 노력 진행 중
 - (한국형 코어) 중저급의 국산 프로세서 코어의 보급 확산을 위하여, 정부 R&D를 통해 “한국형 CPU 코어 상용화” 사업 진행 중
 - (GPU) 국내 범용 GPU 설계 기술은 극소수의 중소기업에 의해서 시도되고 상용화 되었으나 매출 규모는 미미
- 차세대 프로세서
 - 초저전력, 고성능의 차세대 CPU 코어 원천 기술인 PIM (Processor-in-memory) 기술 개발 지원중
 - 뉴로모픽 기술 관련하여 정부는 최근, 뇌신경계 정보처리 원리 모사 컴퓨터 개발을 중점연구 후보분야로 선정, 시냅스 모방소자 개발에 투자 진행
 - ※ 대학에서는 산발적인 뇌신경계 시냅스 모사 연구, 뉴로모픽 컴퓨팅 소자/칩, 인지시스템 기술개발을 진행
- (공정/설계) 미세 공정, 초절전 및 경량 설계를 위한 과제 진행 중
- (응용) 자율주행자동차, 안전/환경, 개인행동 모니터링 등에서의 응용을 위한 기술 개발 계획 중

< 표 6 > 국내 지능형 반도체 관련 주요 공공 R&D 현황 및 계획

구분	세부기술	Source	계획							
			17	18	19	20	21	22	23	
코어	CPU/GPU 혼합 실시간 처리 (14~15)	미래성장동력								
	지능형 프로세서 아키텍처	혁신성장동력		○						
	초절전 매니코어 프로세서	혁신성장동력		○	○	○				
	다중 태스크 강화학습 프로세서	중장기로드맵(IITP)			○	○	○	○	○	○
차세대 코어 - PIM	(초고속) 메모리/프로세서 통합설계	시스템반도체 산업 경쟁력 강화방안('17)	○							
	메모리/프로세서 통합설계	기술로드맵(중기청)		○	○	○				
차세대 코어 - 신경모사	신경모사 인지형 모바일 컴퓨팅 지능형 반도체	혁신성장동력		○	○					
	시냅스 소자기반 패턴인식 HW	혁신성장동력		○	○	○				
	- 영상 및 음성 데이터 인식, 이해, 판단용 인공신경망 - 학습 DB, 시맨틱 이해가 가능한 프로세서 아키텍처	중장기로드맵(IITP)				○	○	○	○	
	- 뉴런/시냅스 소자 회로 모델 lib. 화 및 설계 - 뉴런/시냅스 소자 기반 공정/설계 tool 및 신뢰성	중장기로드맵(IITP)			○	○	○	○	○	○
	- 뉴런/시냅스 소자-상위 회로 모델 통합 인터페이스 - 인공신경망 반도체	중장기로드맵(IITP)			○	○	○	○	○	○
	3D 인터커넥션 공정 기반 대규모 인공신경망 반도체	중장기로드맵(IITP)			○	○	○	○	○	○
설계	초저전력/경량SW-SoC 플랫폼아키텍처 (16~)	미래성장동력								
	(초경량) 공용 설계(IP) 미세화 및 경량 SoC	시스템반도체 산업 경쟁력 강화방안('17)	○	○	○	○	○			
	유무선 데이터통신, 데이터 서비스 지원 위한 SoC 설계, 부품	기술로드맵(중기청)		○	○	○				
	메모리를 SoC에 내장하는 기술	기술로드맵(중기청)		○	○	○				
	초전력, 플렉서블 SoC 설계	기술로드맵(중기청)		○	○	○				
	IoT의 강력한 온보드 프로세서를 통한 전력 절감	기술로드맵(중기청)		○	○	○				
	광대역 데이터 전송 및 저전력 소모 위한 SoC 설계	기술로드맵(중기청)		○	○	○				
	전력소모감소, 이기종 컴퓨팅을 적절한 크기의 칩에 설계, 집적 회로 설계	기술로드맵(중기청)		○	○	○				



공정	차세대 반도체 소재·공정 원천기술	시스템반도체 산업 경쟁력 강화방안(17)	○						
	10nm 이하 차세대 반도체 소자	혁신성장동력		○	○	○	○	○	○
응용 - 비전	Image sensor processor (차량용) (14~15)	미래성장동력							
	비전인식 신호처리 (차량용)	혁신성장동력		○	○				
	영상화질 및 인식 최적화	혁신성장동력		○	○	○	○		
응용 - 자율차	자율주행차용 ADAS	혁신성장동력		○	○	○	○	○	
	초고신뢰성 상황 인식 및 주행 판단 반도체	중장기로드맵(IITP)			○	○	○	○	
응용 - 객체인지	- 딥러닝 기반 다중센서 퓨전 및 객체인지 반도체 - 딥러닝 기반 대용량 센서 데이터 분석/압축반도체	중장기로드맵(IITP)			○	○	○	○	○
응용 - 상황인지	인체 움직임 감지 반도체	중장기로드맵(IITP)			○	○	○		
	유연 센서기반 촉각 인식용 반도체	중장기로드맵(IITP)			○	○	○	○	
	운전자의 안전편의 상황 인지 반도체	중장기로드맵(IITP)			○	○	○	○	○
	생체 신호 기반 행동상황 인지 휴먼케어 반도체	중장기로드맵(IITP)					○	○	○
응용 - 네트워크	차세대 지능형 버스형 엣지 네트워크 및 엣지/코어 네트워크 최적 설계	중장기로드맵(IITP)			○	○			
	저전력 지능형 고속 네트워크	중장기로드맵(IITP)					○	○	
	고장/공격 감지, 손상 회피/감소	중장기로드맵(IITP)					○	○	
	차세대 지능형 네트워크의 어플리케이션별 최적화	중장기로드맵(IITP)							○
	능동형 IoT 디바이스 및 실시간 IoT 제어 무선 플랫폼	중장기로드맵(IITP)				○	○		
응용 - 통신	저전력 저비용 근거리 무선통신 Connectivity	기술로드맵(중기청)		○	○	○			
응용 - 보안	정보 보호, 빅데이터 보안, 고신뢰성 지원 SoC 기술	기술로드맵(중기청)		○	○	○			
응용 - 안전	구조물 위험 감지 및 진단용 반도체	중장기로드맵(IITP)			○	○	○	○	
응용 - 환경	유해환경 감지 및 진단용 반도체	중장기로드맵(IITP)			○	○	○	○	○
	초고신뢰성 오토파일럿 BMS(배터리 관리) 반도체	중장기로드맵(IITP)				○	○	○	○

* 출처 : IITP(2018a), 관계부처 합동(2016, 2018), 중소벤처기업부(2017) 자료를 통합 재구성

나. 민간 R&D⁵⁾

■ 프로세서 코어

- 시스템반도체 2010 사업에서 개발된 프로세서 기술이 상용화되어 있으나 동작 주파수는 수백 MHz 급이며 소프트웨어 및 HW 성능 개선을 위한 추가 투자 필요
- 삼성전자는 ARM 아키텍처 라이선스 확보를 통해 독자 개발한 CPU 코어인 몽구스(Mongoose)를 갤럭시S10의 엑시노스(AP)에 탑재 (2019 출시)할 예정이며, ARM의 코어를 대체할 수 있는 인공지능 IoT용 프로세서도 개발 중
- 중소기업으로는 넥셀이 자체 GPU 개발 및 상용화를 지속하고 있으며, 에이디칩스는 자체 기술을 기반으로 프로세서 코어의 IP사업 추진 중

■ 차세대 프로세서

- 삼성전자, SK하이닉스는 차세대 반도체산업으로 연산기능의 프로세서와 기억에 해당하는 메모리기능 합쳐진 인공지능형 반도체인 뉴로모픽 칩 개발 추진
- SK하이닉스는 HP사와 공동으로 Memristor⁶⁾ 개발 진행, 삼성전자는 뉴로모픽 시스템에 대한 연구를 진행

■ 보안

- 코드의 안전한 실행을 하드웨어적으로 보장하는 보안 아키텍처인 TEE (Trusted Execution Environment)를 기반으로 보안 솔루션 개발
- ARM TrustZone이나 Intel SGX 등에서는 CPU 기반의 TEE가 구현되어 있으나, 국산의 경우, TEE를 위한 하드웨어 지원이 전무
- IoT 시장을 겨냥하여 프로세서 외부에 TPM(Trusted Platform Module) 같은 제한적 보안모듈을 부착하여 SoC 형태로 구현된 보안 플랫폼은 존재
- 네오와인, 라닉스, 이더블유비엠 등 국내 팹리스 중심으로 보안 IP 및 SoC 기술 개발 진행 중

■ 응용

- 프로세서 코어를 적용한 제품을 개발하는 중견, 중소기업체를 중심으로 한국형 프로

5) 중소벤처기업부(2017, 2018) 등 참고

6) Memristor: 메모리(memory)와 레지스터(resistor)의 합성어로 이전의 상태를 모두 기억하는 메모리 소자



세서 응용사업 추진

- 최근 빅데이터 처리, 가상현실, 딥러닝, 자율 주행 등 높은 데이터 연산 처리 능력에 대한 수요가 증가하면서 GPU를 High-Performance Computing (HPC) 분야에서 널리 활용
- 바이오, 의료분야의 반도체는 일부 상용화에 이른 것도 있으나, 대부분은 연구개발 단계
 - ※ 삼성, LG, 인포피아, 아이소텍, 바이오메드랩 등에서 DNA칩, 레이저 채혈 측정기, 혈색소 측정 장비 등 의료기기용 반도체 개발 진행

Ⅲ 주요 응용분야별 핵심가치

1 도출 프레임워크

- 본 연구는 application 분야별 칩셋에 대한 상이한 요구사항을 고려한 차별화된 칩셋의 value를 제시
 - 인공지능 반도체 관련 기술의 발전은, 칩셋(프로세서 및 주변 기기들)의 발전에 따른 응용분야의 발전, 혹은 응용분야의 요구사항 변화에 따른 칩셋의 발전의 두 가지 방법으로 접근 가능
 - 기존의 연구들은 주로 칩셋 기반 접근 방법 중심으로, 일반적인 접근법의 한계에 따라 범용의 기술 요구 제시에 그치는 경향
- 응용분야별 핵심가치 도출은 분석 대상 applications 선정, 서비스 시나리오 작성, 기술 needs 도출, value 제안의 4개 단계로 구성

< 그림 5 > 핵심가치 도출 프레임워크



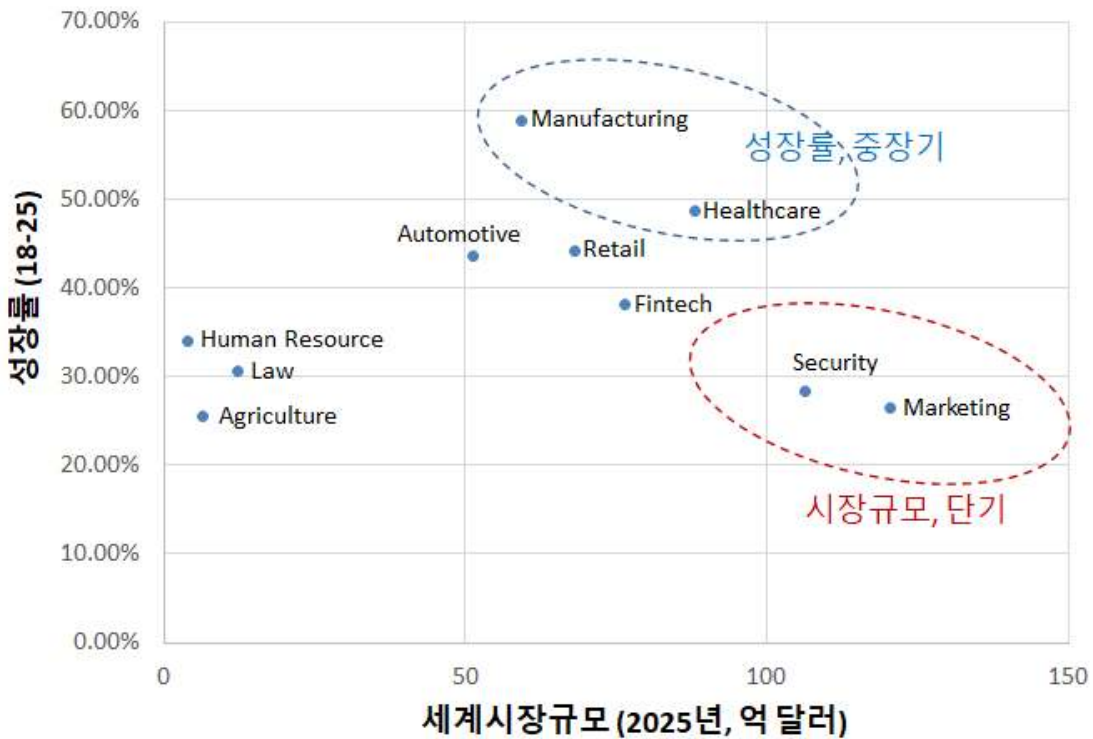


2 지능형 반도체의 니즈 도출

가. 분석 대상 applications 선정

- 시장규모 및 성장률을 기준으로 Marketing, Security, Manufacturing, Healthcare의 4개 응용분야를 선정하고, 응용분야별 2개의 세부 응용분야를 분석대상으로 선정
 - (시장규모, 단기) 응용분야 중 시장규모(2025년)가 가장 클 것으로 예상되는 분야로, Marketing, Security의 2개 분야 선정
 - (성장률, 중장기) 응용분야 중 성장률(18-25)이 가장 높은 것으로 예상되는 분야로, Manufacturing, Healthcare의 2개 분야 선정

< 그림 6 > 분석 대상 applications 선정



* 출처 : Marketsandmarkets(2018)의 시장자료를 참고하여 작성

시장자료의 저작권은 MarektsandMarkets에 있으며, MarektsandMarkets의 동의하에 사용되었습니다. 추후 이용 시 MarketsandMarkets에 문의하시기 바랍니다.

나. 응용분야별 니즈 도출⁷⁾

■ 마케팅 - search advertising

- 세계 시장은 \$875.9 mil ('18)에서 \$3,530.0 mil ('25)로 연평균 22.0% 성장 전망
- AI 기반의 음성 검색이 가장 빠르게 성장하고 있으며, 자연어 처리 기술의 진보가 이러한 음성 검색의 빠른 성장을 견인
- (수요 변화) 소비자 측면에서는 음성/자연어인식을 통한 손쉬운 검색이 요구될 이며, 광고 업체 측면에서는 광고 운영의 자동화 및 효율화 달성 요구 전망

< 그림 7 > 마케팅 - search advertising의 수요 변화



- (Scenario 1: Personal device) 음성인식 칩(Edge)을 통해 아날로그 음성을 디지털로 전환, 자연어 처리/언어 이해(Edge) → 데이터 전송(Edge → Cloud) → 광고 업체는 빅데이터 기반 광고 검색(Cloud) → 서비스 제공(Cloud → Edge)
- (음성인식, Edge) 학습(특정인의 음성 인식 모델), 추론(음성 인식), 수시 갱신
- (자연어인식, Edge) 학습(특정인의 언어 패턴 모델), 추론(자연어 이해), 수시 갱신
 - Training in edge를 통한 개인화 (특정인 음성인식률/언어이해율 향상)
 - Training + inferencing in one chip으로 하나의 칩에서 학습 및 추론
 - 경량, 초저전력, 고속 processor & chip
- (DB 검색, Cloud) 대용량 DB 검색, 대량 데이터 이동
- (검색 모델, Cloud) 학습(검색 모델), 추론(실시간 검색률 변경)
 - 초고속, 저전력, 발열방지 processor & chip

7) 세부분야별 세계 시장 전망은 MarketsandMarkets(2018) 인용



- (Scenario 2: Digital signage) 음성 인터페이스 (대화형, any language, Edge), 자연어 처리/언어 이해(Edge) → 데이터 전송(Edge → Cloud) → 상황 기반 광고 결과 (Cloud)
- (대화형 인터페이스, Edge) 다국어 지원, 대화형을 위해 자연어처리/언어이해, 소란스러운 장소에서의 음성인식률 향상, 화자의 상황이해, cloud와 edge간 실시간 통신 (quiry + response)
 - 음성인식률 향상: 비전인식 + 지향성 음성 인식
 - 상황이해: 비전 + 음성인식
 - 실시간 통신: 초고속 통신 (cloud <--> edge 간) 모듈
 - 통합 설계: 저전력, 저지연, 고인식률(음성), 고이해율(자연어), 맞춤형 정보 제공
- (상황기반 검색, Cloud) 학습(Edge별 상황인지 기반 검색 모델), 추론(Edge별 실시간 검색률 변경), 유용 정보 구분 및 DB 관리 자동화(추론), Edge별 이용자의 상황기반 DB 검색
 - 초고속, 저전력, 발열방지 processor & chip

■ 마케팅 - social media advertising

- 세계 시장은 \$476.6 mil ('18)에서 \$2,533.6 mil ('25)로 연평균 27.0% 성장 전망
- SNS 이용자에게 개인화된 콘텐츠 제공으로 인게이지먼트 향상을 위한 AI의 이용 증가 예상
- (수요 변화) 개인 맞춤형 및 실시간 유해 콘텐츠 필터링 요구

< 그림 8 > 마케팅 - social media advertising의 수요 변화



- (Scenario) 개인의 SNS 이용상황과 자연어 분석 기반 개인 맞춤형 광고 서비스 제공, 실시간 유해정보/영상의 판단 및 필터링 (필터링 수준은 개인이 설정 가능) (cloud)
- (필터링, Cloud) 유해 콘텐츠 필터링을 위한 실시간 사진/영상정보분석
- (상황기반 광고 생성, Cloud) 상황과 광고의 매핑 모델 (학습 + 추론)
- (보안, Cloud) 개인 데이터를 확보하고 cloud에서 안전하게 보호
 - 초초고속, 저전력, 실시간 processor & chip
 - 정보보호 processor or chip



■ 의료 - patient data & risk analysis

- 세계 시장은 \$145.3 mil ('18)에서 \$2,076.5 mil ('25)로 연평균 46.2% 성장 전망
- 게놈 정보 등을 활용한 예측 어널리틱스는 의사들이 환자들의 위험을 파악할 수 있도록 도와주며, 위험을 회피할 수 있는 방법들에 대해서도 조언을 해주게 됨
- 예를 들어 AI 칩을 웨어러블로 착용하여, (병원 혹은 병원 밖에서) 고혈압, 당뇨 등 만성 질환자를 상시 관리함으로써 중증 질병 발생 위험을 사전 포착하여 치료
- (수요 변화) 365일 24시간 환자를 관리하는 인공지능 주치의 요구

< 그림 9 > 의료 - patient data & risk analysis의 수요 변화

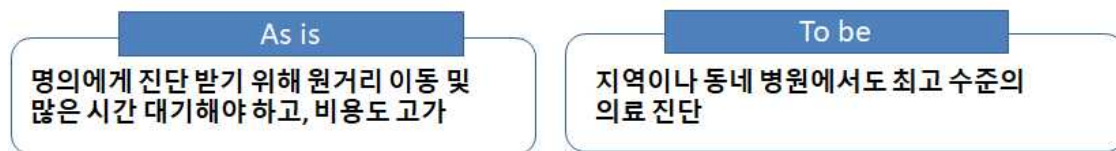


- (Scenario) 센서퓨전칩(웨어러블 센서 + AI 칩)에서 환자 data 수집(edge) → edge에서 cloud 전송(스마트폰 경유 or 직접) → 유전/환자정보 + 수집 데이터 기반 위험 분석 (cloud) → cloud에서 edge로 위험 정보 전송
- (센서퓨전, Edge) 환자 유전정보 기반 개인별 위험 관리 모델, 장시간 착용 및 인체 무해, 24시간 통신, 외부 충격에 강인하며 센서 성능 저하 시 보정
 - 개인 맞춤형(인공지능 주치의), 경량
 - 초저전력, 에너지 하베스팅
 - 인체 무해 통신 (환자 위험 관리를 cloud에서 수행 시 실시간 데이터 송수신)
 - 외부 충격 및 자극에 강한 구조 혹은 유연 구조, 센서 정보 보정
 - 통합설계: 개인 맞춤형, 경량, 초저전력, 에너지 하베스팅, 인체 무해, 센서정보 보정, Flexible
- (환자위험모델, Cloud) 환자 의료정보 + 유전정보 등 대용량 데이터 기반 환자 위험 모델(학습), 환자 개인별 위험 관리(추론, 환자 위험 관리를 cloud에서 수행 시)
 - 초고속, 저전력

■ 의료 - medical imaging & diagnostics

- 세계 시장은 \$71.2 mil ('18)에서 \$1,536.2 mil ('25)로 연평균 55.1% 성장 전망
- 다수의 healthcare AI 벤처들이 2015년 1월 이후 기계학습을 이용 이미징 및 진단 분야에 투자
- (수요 변화) 편한 시간과 생활 공간과 가까운 곳에서 최고 수준의 의료 진단

< 그림 10 > 의료 - medical imaging & diagnostics의 수요 변화



- (Scenario 1: 대형병원의 지역 진단센터) 대형병원 지역 진단센터에서 환자 데이터 수집 및 cloud 전송 → 진단 결과 생성 (cloud) → 명의의 최종 진단
- (비전, Edge) 초고해상도, 고화질 영상 및 동영상 처리
 - 초고속 비전 프로세서
- (통신, Edge) 각종 측정 자료 등 대용량 데이터 송신
 - 대용량 데이터 통신
- (의료진단, Cloud) 영상 정보, 개인 유전정보, 개인 상황 등 대용량 데이터 기반 진단 학습 및 추론, 인도 등 저가의 원격 진단 서비스(경쟁 서비스) 대비 경쟁력 있는 진단 비용
 - 초고속, 저전력, 저비용
- (Scenario 2: 동네병원 의사 보조) 측정 기구 및 의사가 직접 측정한 의료 데이터 입력 → 추론 및 진료 보조 결과 생성 (edge), training은 동네병원 별 차별화를 위해 edge에서 직접 수행
- (의료진단, Edge) 동네병원 별 차별화를 위한 의사 개인 노하우 축적 (Training in edge), 중앙(cloud)에서 신규 학습한 노하우의 edge 이전
 - Training in edge를 통한 개인화 (특정인 음성인식률/언어이해율 향상)



- Training + inferencing in one chip으로 하나의 칩에서 학습 및 추론
- 편리한 training interface를 위한 simple한 구조
- 추론 모델의 손쉬운 추가 이식: edge training과 cloud training의 결합, 혹은 dual로 처리
- (의료진단모델, Cloud) 개인 유전정보 + 의료 데이터 기반 학습
 - 초고속, 저전력

■ 보안 - identity and access management

- 세계 시장은 \$414.9 mil ('18)에서 \$2,761.6 mil ('25)로 연평균 31.1% 성장 전망
- 최근 규정 준수 요구 사항(regulatory compliance requirements)이 점차 더 엄격하고 복잡해짐에 따라 ID 및 액세스 관리 솔루션이 널리 보급되고 중요
- (수요 변화) 자동화된 자원관리로 효율성 증가 요구

< 그림 11 > 보안 - identity and access management의 수요 변화



- (Scenario) PC, 모바일의 웹 or 전용통신망 기반 접근, 개별 기업/조직 단위에서 시스템 구축 (control center, 각 resource server에 인공지능 칩) 혹은 구글 등 전문기업에서 웹기반 서비스 제공 (Cloud)
- (보안, Cloud) History 기반 실시간 학습 + 상황 정보 기반 추론, 접근에 대한 정확한 식별 및 제어, 정밀/실시간 자원 관리 (자동화)
 - HW 차원의 보안 (TEE)
 - 초고속, 저전력, 발열 방지



■ 보안 - antivirus/antimalware

- 세계 시장은 \$551.2 mil ('18)에서 \$2,701.4 mil ('25)로 연평균 25.5% 성장 전망
- 바이러스 및 악성 프로그램 공격의 유형과 유형이 급속하게 증가함에 따라, 인공지능이 그러한 공격을 막는 솔루션으로 급성장
- (수요 변화) 악성 위협을 사전 예측하고 자동적으로 대응을 요구

< 그림 12 > 보안 - antivirus/antimalware의 수요 변화

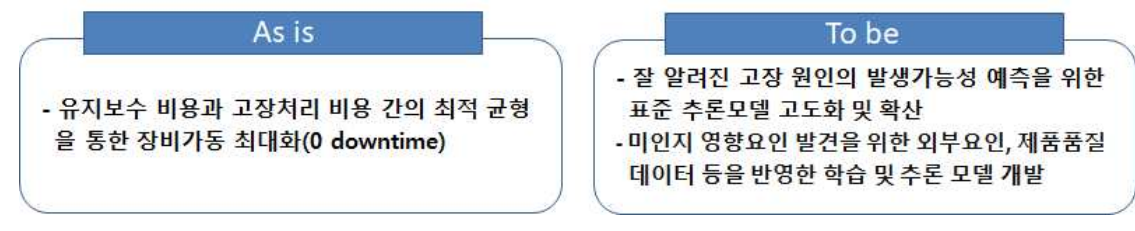


- (Scenario) 개인이 소지한 단말 (PC, 노트북, 스마트폰 등)의 칩에 인공지능 기능이 내장되어 있으며(integrated edge), 서비스 제공업체는 새로운 유형의 바이러스를 학습(cloud)하고, 개인 단말은 추론 프로그램을 정기적으로 update (edge)
- (보안, Edge) 실시간 예측 및 대응
 - 경량, 저전력
- (통신, Edge) 추론 프로그램의 정기적 업데이트
 - 고속 통신
- (보안 모델, Cloud) 실시간/대용량 데이터 학습
 - 초고속, 저전력, 발열 방지

■ 제조 - predictive maintenance & machinery inspection

- 세계시장은 \$113.3mil('18)에서 \$3,071.4mil('25)로 연평균 60.2% 성장 전망
- 제조 분야에서 최고 시장점유율(48.8%, '18)을 차지하는 응용분야로 장비 조기고장 진단으로 운용비용을 절감하기 위한 요구가 증대되면서 빠른 시장 성장 전망
- (수요 변화) 정밀하고 자동적인 관리 기법을 통한 장비가용성과 유지보수 비용 절감

< 그림 13 > 제조 - predictive maintenance & machinery inspection의 수요 변화



- (Scenario) 기계 장비 상태 + 다양한 환경 데이터 수집 (Edge) → 방대한 데이터 중 유의미한 데이터만 필터링 (Edge) → 데이터 전송 (Edge → Cloud) → 고장 예측 (추론) 및 고장 예측 모델 수정 (학습) → 추론 모델 갱신 (Cloud → Edge)
 - ※ 소리, 진동, 온도, 압력, 가스 등 기계적 특성 외에 음성, 영상, 동작 등 센서 대상 확대
- (센서퓨전, Edge) 방대한 센서 데이터를 유의미한 데이터만 필터링 + 센서 정보 보정, 다량의 칩 소요
- (통신, Edge) 수집된 자료(유의미한 데이터)를 실시간 cloud로 전송, cloud에서 만들어진 추론 엔진을 주기적 갱신
 - 경제적 가격의 센서퓨전 칩
 - 고속 통신 모듈
 - 통합 설계: 센서, 센서정보 보정, 진단, 저전력, 고속 통신
- (고장 예측, Cloud) 방대한 데이터 기반 학습/추론
 - 초고속, 저전력, 발열 방지



■ 제조 - quality control

- 세계시장은 \$38.4mil('18)에서 \$1,348.3mil('25)로 연평균 66.2% 성장 전망
- 인공지능 기반의 품질관리 시스템은 의약품, 식품&음료, 반도체 등 다양한 산업에서 폭넓게 활용될 수 있어 급속한 시장 성장 전망
- (수요 변화) 인공지능의 활용으로 제품 불량률의 획기적 감소 요구

< 그림 14 > 제조 - quality control의 수요 변화



- (Scenario) 제품 이미지 센싱 및 불량품 진단 (Edge) → 유의미한 데이터 필터링 및 송부 (Edge → Cloud) → 통합 품질 관리 (학습 및 추론, Cloud)
- (품질관리, Edge) 제품의 다각도 디지털 이미지 촬영, 3차원 모형 구성을 통한 이미지를 센싱, 품질관리자의 이동 및 원격관리를 고려한 모바일형 품질관리 단말, 불량품 진단 추론 기능 탑재
 - 초고속 비전 프로세서
- (품질관리, Cloud) 통합 품질 관리 + 통합 품질 관리 모델 개발을 위한 대용량 데이터 처리
 - 초고속, 저전력, 발열 방지

3 핵심가치 제안

가. General suggestions

- Processor 코어의 진화 및 혁신
 - (초고속/저전력 코어) 초고속 processor, 저전력 → GPU와 FPGA 기반은 선도적인 업체들이 포진해 있는 상황이며, 저비용/특정 목적용 ASIC 개발은 차별화 가능
 - (차세대 코어) 초초고속 processor, 초저전력, 발열 방지 (뉴로모픽, 3차원 구조)
 - 다수의 선도 기업/기관들이 기술 개발 중으로 기술 개발자의 차별화 가능성 사전 점검 필요하며, 프로그램의 유연한 수정이 가능한 뉴로모픽 칩 등 개발 고려
 - (IoT용 코어) 고속 processor, 초저전력, (에너지 하베스팅), 초저비용 → 다량의 칩이 소요되는 Niche 시장으로서 추진 가능
- Application specific chip 개발
 - 높은 성장이 예상되는 각 응용분야별 특화된 지능형 반도체의 개발을 통한 시장의 선점이 필요
 - 현재 선도 기업/기관들은 자율주행자동차 분야에 집중적인 투자를 하고 있는 상황이나, 마케팅, 의료, 보안, 제조 등의 분야에도 관심을 가질 필요
- 추론 모델의 통합 (at Edge)
 - Case1: Edge와 cloud 모두 training, cloud에서 학습된 추론 모델을 edge에서 추가
 - Case2: Cloud에서만 training, cloud에서 학습된 추론 모델을 edge에서 추가

나. Niche suggestions

- Training at Edge & 'Training + Inference' in one chip
 - Training in edge를 통한 개인화: 특정한 음성인식률/언어이해율을 향상하거나, 동네 병원 의사들의 개별적인 노하우 축적으로 차별화 가능
 - Training + inferencing in one chip으로 하나의 칩에서 학습 및 추론을 수행함



으로써 소형화와 실시간 학습 및 추론 프로그램 업그레이드 제공

- Edge에서 시스템으로 작동하기 위한 통합 칩 설계 (유망 응용분야)
 - (대화형 인터페이스) 디지털 사인지에 내장되는 대화형 인터페이스 칩으로, 다국어 지원, 대화형을 위해 자연어처리/언어이해, 소란스러운 장소에서의 음성인식률 향상, 화자의 상황이해, cloud와 edge간 실시간 통신(query + response) 가능
 - (인공지능 주치의) 환자 유전정보 기반 개인별 위험 관리 모델, 장시간 착용 및 인체 무해, 24시간 통신, 외부 충격에 강인하며 센서 성능 저하 시 보정 가능
 - (고장관리) 방대한 센서 데이터를 유의미한 데이터만 필터링 + 센서 정보 보정, 수집된 자료(유의미한 데이터)의 실시간 cloud 전송, 추론 엔진의 주기적 갱신
- 차별화 모델
 - 음성인식의 차별화: 비전인식 + 지향성 음성인식 in public space
 - 개인 맞춤형 AI 주치의: 개인 유전정보 등을 기반으로 환자 개인별 추론 모델을 edge에 포함



참고문헌

◆ 국외자료

Hilton et al. (2006), A fast learning algorithm for deep belief nets.

IDC (2017), GPUs, FPGAs, ASICs, or Many-Core processors. Which acceleration technology do data centers need.

Nakhare (2017), Hardware options for machine/deep learning, MS&E 238 Blog, <https://mse238blog.stanford.edu/2017/07/gnakhare/hardware-options-formachine-deep-learning/>.

PWC Strategy& (2018), Artificial intelligence - The next big growth driver for the semiconductor industry.

WEF (2016), Future of jobs survey.

◆ 국내자료

IITP (2018a), ICT R&D 기술로드맵 2023 - 지능형반도체.

IITP (2018b), 반도체 산업의 차세대 성장엔진, AI 반도체 동향과 시사점.

KISDI (2018), 기술혁신의 관점에서 바라보는 인공지능 생태계 - 지능형 반도체의 혁신 방향을 중심으로.

KISTEP (2019), 인공지능(반도체).

STEPI (2016), 신기술 발전에 따른 산업 지형의 변화 전망과 대응 전략.

관계부처 합동 (2016), 2016년도 미래성장동력 종합실천계획(안).



관계부처 합동 (2018), 혁신성장동력 시행계획.

김정호(2018), Introduction to terabyte interconnection and package laboratory & 3D IC.

융합연구정책센터 (2018), 뇌연구 국내외 정책동향.

중소벤처기업부 (2017), 중소기업 기술로드맵 2018-2020 - 지능형 센서.

중소벤처기업부 (2018), 중소기업 기술로드맵 2019-2021 - 지능형 반도체.

◆ ETRI 내부자료

ETRI (2018), ETRI 중장기 기술개발계획 2025.

ETRI 기술경제연구본부 (2017), 인공지능 반도체 산업동향 및 이슈 분석.



저자소개

하영욱 ETRI 미래전략연구소 기술경제연구본부 기술경제연구그룹 책임연구원
e-mail: hahaa@etri.re.kr Tel. 042-860-6173

김태완 ETRI 미래전략연구소 기술경제연구본부 기술경제연구그룹 책임연구원
e-mail: twkim@etri.re.kr Tel. 042-860-6368

지능형 반도체의 주요 응용분야 시나리오와 핵심가치

발행인 한 성 수

발행처 한국전자통신연구원 미래전략연구소 기술경제연구본부

발행일 2019년 6월 15일



www.etri.re.kr

ETRI 한국전자통신연구원 미래전략연구소

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

