전자공학종합설계프로젝트 제안서

AI 반도체 NPU (Neural Processing Unit)

NPU의 설계와 검증

지도 교수 : 강준우 (인)

제출 일자 : 2021. 02. 28

학번 : 201804531

소속 : 전자공학과

성명 : 김형준

1. 연구의 필요성

가. 국내외 기술개발 현황

SoC는 System on Chip으로 반도체 기술을 하나로 집적화 시키는 역할을 한다. CPU, GPU 등이 하나의 칩으로 구현되어 반도체로의 역할을 하는 것이다. SoC의 장점으로는 시스템 블록을 하나의 칩으로 구현할 수 있게 한다는 것이다. 이는 SoC가 들어가는 기기의 크기를 줄이며, 제조 비용 자체를 절감하는 효과를 가져온다.

SoC는 모든 기능을 제한된 상황에서 효율적인 일 처리를 가능하게 하는데, 작은 스마트폰에 적용되는 고성능 SoC와 냉장고에 들어가서 간단한 연산을 진행하는 MCU(Micro Controller Unit)라는 칩도 있다.

시스템반도체 기업 순위('17)

단위: 억 달러

	기업	국가	매출	점유율	주요 제품
1	인텔	미국	656	26.3%	CPU
2	퀄컴	미국	164	6.6%	AP
3	브로드컴	미국	164	6.6%	통신칩
4	텍사스인스트루먼트	미국	149	6.0%	
5	엔비디아	미국	104	4.2%	GPU
6	NXP	네덜란드	80	3.2%	
7	미디어텍	대만	79	3.2%	AP
8	하이실리콘	중국	74	3.0%	AP
9	ST마이크로일렉트로닉스	스위스	66	2.6%	
10	AMD	미국	62	2.5%	CPU
11	삼성전자	한국	60	2.4%	DDI, AP
12	아날로그디바이스	미국	59	2.4%	
13	르네사스 일렉트로닉스	일본	57	2.3%	
14	애플	미국	54	2.2%	AP
15	마이크로칩	미국	50	2.0%	

자료: IC Insights, Trendforce 및 수출입은행.

다음과 같이 시스템 반도체 기업 순위를 볼 수 있다. AI 반도체 기술은 CPU, GPU 그리고 NPU 반도체로 변화를 예상하므로 이 기업들의 순위를 지켜볼 필요가 있다. 이 표를 통해 지켜보면 시스템반도체를 주력사업으로 하는 기업들 중 우리나라 기업은 삼성전자를 제외하고는 찾아볼 수 없다. 주로 미국 기업 그 다음은 유럽 기업으로, 점유율을 통한 비교를 보면 양극화는 더욱 더 심한편이다.

주요국 시스템반도체 기술수준

	한국	미국	일본	유럽	중국
시스템반도체	80.8	100.0	89.6	91.0	74.2
AI 반도체	84.0	100.0	88.0	89.8	89.3

주: 시스템반도체는 2017년 기준, AI반도체는 2018년 기준
자료: 한국산업기술평가관리원 및 정보통신기획평가원.

기술수준의 강세 역시나 미국이 차지하고 있고 그 다음은 유럽이 차지하고 있다.

주요국 시스템 반도체 기술수준에 관한 표를 보면 현재 우리나라의 AI 반도체는 경쟁국가들에 비해서 뛰어나지 못한 편이다. 우리나라는 현재 메모리 반도체에 편향이 되어있다. 메모리 반도체와 시스템 반도체의 균형적인 발전을 위해 시스템 반도체의 육성이 필요하나 우리나라 반도체 산업은 종합반도체 기업 중심의 사업구조로 시스템 반도체를 위한 맹목적인 투자를 기대하기는 어려운 편이다.

근소한 차이로 중국이 발전하고 있다. 특히 중국의 발전율은 상당히 높다. 중국 AI 반도체의 시장 규모는 약 50%에 가깝게 기록하고 있다. 이러한 추세로 보면 중국이 유럽 다음의 AI 반도체 강국이 되는 것은 시간문제이다.

중국은 정부의 AI 반도체에 대한 적극적인 정책을 내세우며 종합반도체 기업 중심의 시장인 우리나라와는 달리 중국은 AI 반도체 스타트업을 기반으로 시장이 형성되고 있다.

AI반도체는 GPU를 시작으로 대규모 병렬 처리가 가능한 장점에서 시작되었다. 그러나, 과도하게 높게 측정되는 소비전력과 발열량으로 CPU를 완전히 대체할 수 없었다. 이를 보완하기 위해 ASIC가 출시되었다. ASIC는 인공지능 처리 성능, 에너지 효율이 우수하고 GPU의 단점을 해결하였지만 주문형 반도체로 고객의 수요를 완전히 충족시키기 힘들었고 초기 투입 비용이 높았다. 그러나 이에 따른 상황에서는 FPGA가 있다. FPGA는 회로를 변경하거나 재설계가 가능하고 ASIC에 비해 초기 투입 비용이 적어 시장이 쉽게 정착했다. 그러나, FPGA에도 단점은 있는데 GPU 대비 낮은 연산 성능과 주요 AI 시스템에서 활용 빈도가 낮은 편이다. 이러한 칩들은 CPU를 포함하여 장점, 단점과 각자의 역할이 뚜렷하여 공존하고 있다.

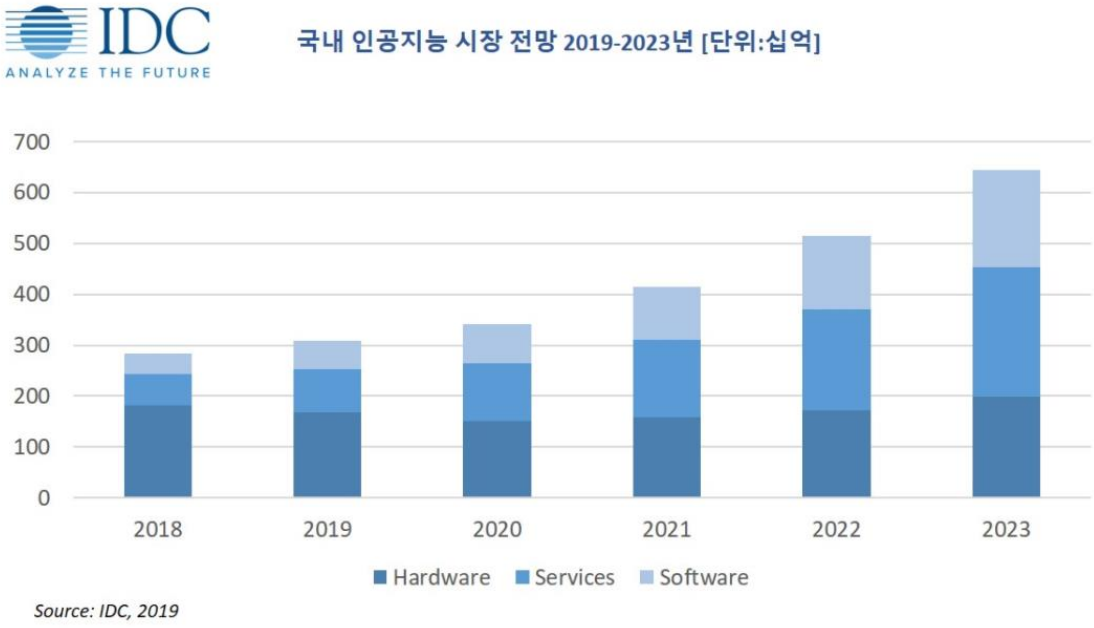
우리나라는 NPU를 구현하는데 다른 나라에 비하면 상당히 상황이 좋지 못한 편이다. 그래도 삼성의 경우 뉴로모픽 반도체를 중심으로 AI 연구에 투자하는 편이고, 해외 기업과 국내 대학의 협업을 진행하며 연구에 몰두하는 중이다.

최근에는 인공지능을 이용하여 이미지를 추론하고 검색하며 판단하는 기능이 우리의 삶과 가깝다고 볼 수 있다. 쉽게 스마트폰의 카메라에서 상황에 맞게 카메라의 모드와 구도를 변경하며 스스로 상황을 판단하는 것이다. 하지만, 이러한 인터넷에 연결되어 기기 자체의 인공지능 능력이 아닌 기기에 탑재된 인공지능 온 디바이스 AI를 위해서는 더 많은 NPU의 발전이 필요할 것이다.

외국에서 NPU는 비교적 시장에서 많이 활용되고 있는데, 차량용 반도체의 전면부 감지 운전자 모니터링, ADAS 등에 NPU가 쓰여 자율주행 자동차의 발전에 기여하고 있다. NXP(네덜란드), 인피니온(독일), 르네사스 일렉트로닉스(일본)등의 기업들이 자율주행 자동차에 수많은 도로상황을 학습시켜 스스로 문제를 판단하는 자동차를 만들기 위해 노력중이다.

나. 문제점 및 앞으로 전망

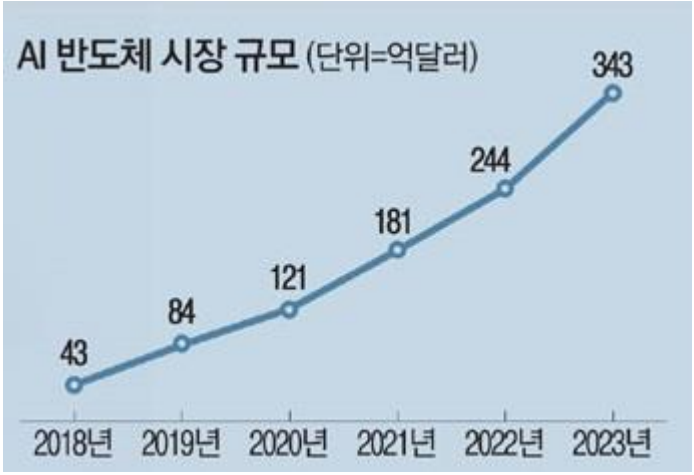
- 국내 인공지능 시장 전망은 다음과 같다



위와 같이 인공지능은 미래에 우리의 삶에 가장 큰 영향을 끼칠 과학분야로 뽑히고 있다. 모든 반복 수작업, 기술, 인력의 부분에서 빠른 적응력을 보이고 있기에 많은 분야에서의 투자가 유입되고 있다. 그에 따라, 인공지능의 시장 자체도 점차 커질 것으로 예상된다.

하지만, 세계적으로 비교해보았을 때, 미국이 30.7%인 반면, 우리나라는 17.8%로 아직 우리의 경쟁 상대와 비교해 보았을 때는 큰 성장률은 아니다.

NPU의 현재 산업 활용은 상대적으로 CPU나 GPU에 비해서는 짧은 사용기간을 거쳤기에 주로 사용되기엔 기본 틀이 갖춰지지 않은 경우가 많다. NPU는 인공지능에 매우 적합하고 이를 특화하여 딥러닝에 잘 적용하여 사용이 되면 시장에 매우 큰 반응을 얻을 것이다.



지금의 NPU는 기존의 큰 반도체 회사들보다는 스마트폰으로 유명한 회사들이 주로 공략하고 있는 분야이기도 하다. NPU를 스마트폰에 적용시켜 수많은 피드백을 바로 얻고, 기기 자체에 인공지능을 심는 역할을 해낼 수 있는 최고의 NPU를 얻기 위한 노력 중이다. 이러한 노력이 결과를 얻게 된다면 기존의 인공지능 서비스에서 생겼던 정보처리에 소요되는 시간과 정보의 보안성에 대한 문제를 해결할 수 있을 것이다. 클라우드를 통한 인공지능이 아니라 기기 자체의 인공지능이 탑재가 되어 더 빠르고 안전한 인공지능을 볼 수 있을 것이다. 그리고 자체적 딥러닝 수행이 가능해질 것이다.

다. 연구 개발의 중요성

전자공학 설계 프로젝트에서 수행하려는 내용은 AI 반도체에 대하여 칩을 만드는 과정이다. 우리가 이러한 칩을 이론적으로 배우는 것 외에 실습을 통해 칩의 구동이나 배치를 배울 수 있는 기회는 없었기에, 회로를 설계를 해볼 수 있는 좋은 기회가 된다.

SoC 설계 방안에 대해서는 첫번째로, IP를 기반으로 시스템 자체를 구성하는 블록 기반 설계 방안이 있고, 플랫폼 자체를 재구성하여 사용하는 플랫폼 기반 설계 방법이 있다. 이를 각각 BBD와 PBD라고 한다. 여기서 SoC 플랫폼을 기반으로 설계하는 것은 IP를 매번 검증하고 배치하는데 드는 시간과 자본을 줄이기 위해 방안해낸 방법이다. 이는 제품을 제한된 시간내에 제작하는데 어려움을 초래했다. 이 때, SoC 플랫폼을 통해 원래 HDL의 방식을 컴퓨터 언어를 통해 쉽게 표현을

가능하게 하였고, 이는 많은 상황에 대한 시간과 비용을 단축하게 해주는 방법이 되었다. 현재 NPU는 단일 뉴럴넷을 NPU에 적용시켜 작동하는 아주 간단한 일 밖에 할 수 없고, 우리가 원하는 다양한 인공지능 학습에 적용은 쉽지 않다. 그리고 NPU를 만드는 개발자에 따라서도 많은 차이점이 생겨 아직 NPU는 개발이 필요한 분야라고 생각한다. 그래서 NPU 장치에 대해 여러 SoC 플랫폼을 기반으로 하여 다양화된 실습을 진행해봐야 한다고 생각한다.

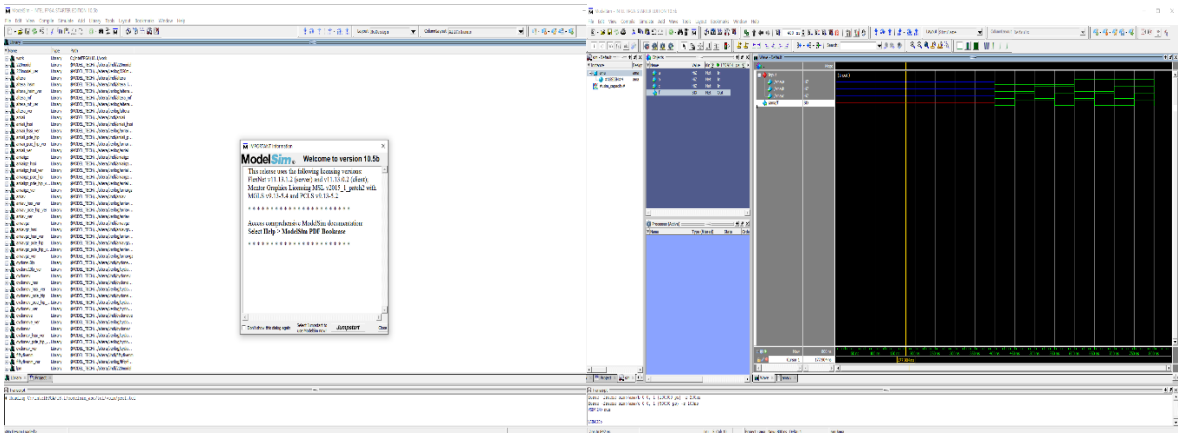
2. 연구 개발 목표

가. Neural Processing Unit의 설계와 검증

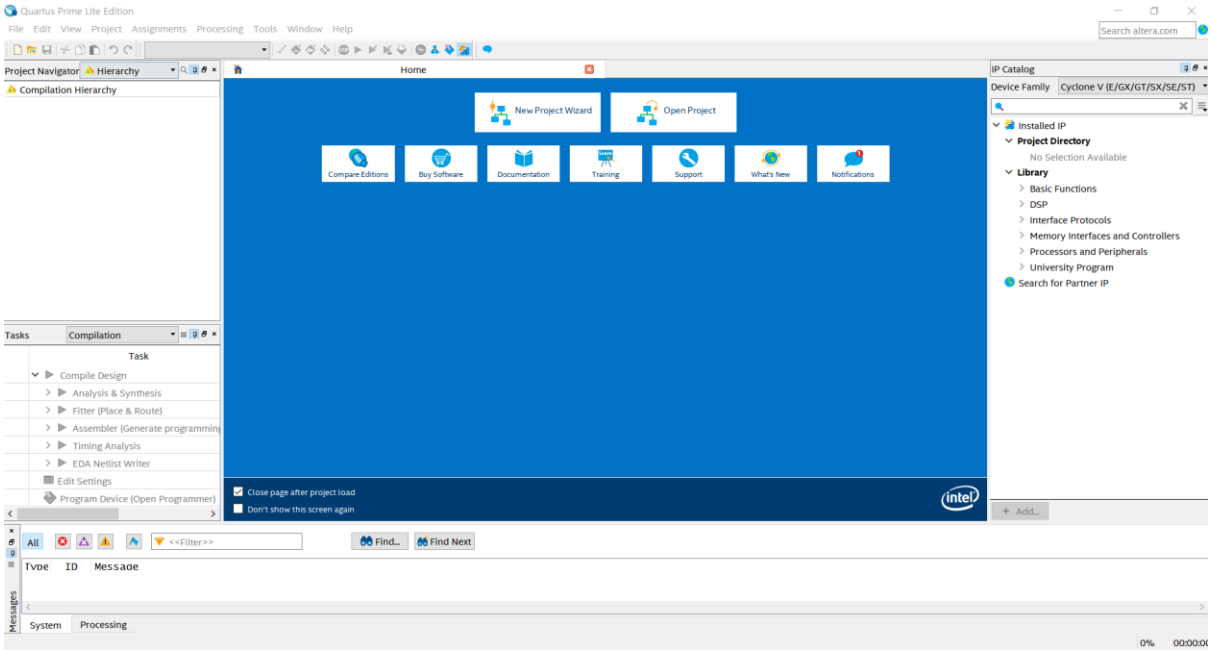
NPU 반도체를 안정적이며 효율적인 칩으로 구현하고, 이를 통해 인공지능 기술을 이용하여 다양한 이미지나 정보에 대해 받아들인 뒤 사고하여 빠른 해답을 내릴 수 있는 칩을 구현해보고 싶다.

나. 연구 내용 및 범위

설계 프로젝트를 통해서 기본적인 NPU 프로세서를 다양한 상황에 안정적으로 적용이 되고, 생체 인식을 통해 사람을 판단할 줄 아는 기능을 구현해볼 것이다. 예를 들어, 스마트폰의 카메라에서 핸드폰 주인의 얼굴을 인지하고 만약 카메라를 켰을 때, 주인이 움직이게 되면 그 움직임에 맞춰서 스마트폰에 연결된 삼각대를 이용하여 같이 움직이게 되는 방식이다. 즉, 스마트폰의 주인 얼굴을 반복적으로 인식하고 스스로 인지하여 카메라 화면에서 벗어나거나 구석에 있을 때, 스스로 판단을 내리고 이동을 할 줄 아는 기능을 만들 것이다.



위 사진은 ModelSim 프로그램으로 우리가 구현하고자 하는 회로칩을 test bench를 통해 미리 시뮬레이션을 할 수 있는 프로그램이다. 간단한 논리회로를 코딩을 통해 생성한 뒤, 시뮬레이션을 작성하여 프로젝트를 만들어 여러가지 경우의 수를 확인할 수 있다.



위 사진은 Quartus2 프로그램으로 시뮬레이션을 돌려볼 수 있는 기능이 있다. 작성한 프로젝트를 컴파일하여 불러내면 다양한 설정을 통해 그에 맞는 waveform과 Block Diagram을 얻을 수 있다.

다. 연구 계획표

프로젝트 구상	3 월				4 월					5 월				6 월		
	1 주	2 주	3 주	4 주	1 주	2 주	3 주	4 주	5 주	1 주	2 주	3 주	4 주	1 주	2 주	3 주
SoC 구조 이해																
NPU 소자의 동작 이해																
제안서 준비 및 발표																
NPU 의 HDL 프로그램 작성																
HDL 모델 작성과 시뮬레이션																
NPU 를 검증 키트로 구현																
최종보고서 작성 및 발표																

3. 연구결과

연구 및 실험에 대한 예상과 결과는 매주 보고서를 통해 정리된다. 실험 전에는 실험에 대한 목표와 주제 그리고 예상되는 결과 또는 문제점을 정리하여 진행될 실험에 대해 다시 한번 고려해보게 된다. 실험이 진행되는 동안 실험의 다양한 결과는 기록이 되고, 실험을 통해 얻은 결과의 의미와 변수요인을 확인한다.

디지털집적회로설계 수업을 통해 실험을 진행하면서 직접 간단한 칩을 만들어보고 결과물을 받아볼 수 있을 것이다. 그러나, 모든 실험에 대한 칩을 만들고 확인할 수는 없기에 위에서 언급했던 다양한 프로그램을 활용할 예정이다.

SoC를 간단하게 설계할 수 있도록 하는 Quartus2 프로그램을 이용하여 여러 상황에 대해 시뮬레이션을 가동이 가능하게 하는 Modelsim 프로그램과 이용하여 칩의 결과를 얻을 것이다.

4. 추진전략 및 방법

위에서 언급했듯이 NPU는 현재 시장에 많은 분야로 활용이 될 수는 있지만, 아직 널리 사용되기엔 기반이 구축되어 있지는 않은 편이다. 그래서 기술에 대한 정보수집은 인터넷과 전문가 확보에 중점을 두어야한다.

SoC의 설계는 반도체의 소자에 대한 이해와 이론적으로 SoC의 구조와 동작 원리에 대해 빠르게 파악하고, 위에 언급했던 프로그램들을 이용하여 다양하고 최적화된 칩을 구성하는데 노력할 것이다.

5. 기대성과 및 활용방안

가. 기대성과

→ 아직 완전히 정착하지 못한 NPU가 시장에 안정적으로 정착되어 CPU, GPU만큼의 틀을 갖춰 많은 방면으로 활용되기를 기대한다.

→ 경쟁 국가에 비해 아쉬운 국내 AI 관련 SoC 산업에 인공지능 반도체 시장이 활발해질 것을 기대한다.

→ 메모리 반도체에 비해 미약한 국내 시스템 반도체에 대해 연구하고 배우는 기회가 될 것이다.

→ 2022년에 553억 달러까지 커지는 세계 자율주행 자동차 전용 반도체에 수집분석, 통신 등에 쓰일 NPU를 가질 것이다.

나. 활용방안

AI 알고리즘을 통해 이루어지는 대량 연산에 현재 가장 넓게 사용되는 GPU의 단점으로 시스템 특성에 맞게 활용하지 못한다면 비효율적인 점과 낮은 확장성이 있다. 그렇다고 이러한 GPU를 아예 대체하는 NPU가 되는 것은 아니지만, 모바일 AP는 같이 포함되어 사용된다. NPU에는 NPU 제어기를 포함하고 이 NPU 제어기에는 CPU, SRAM 등이 포함된다. On-Chip 학습을 위한 저전력 인공지능프로세서에 대한 수요가 증가하여 이는 앞으로도 많은 방향으로 활용이 될 예정이다.

6. 참고문헌

[1] 윤민호, 박정근, 강태삼, "FPGA를 이용한 압전소자 작동기용 단일칩 제어기 설계", 제어로봇시스템학회, 513-518쪽, 2016년 7월

[2] 김상철, 김용연, 김태호, "다중 이중 NPU 장치에서 다중 뉴럴넷을 추런하기 위한 NPU 운영체제 플랫폼 개발", 한국정보과학회, 735-737, 2020년 7월

[3] 이상엽, 이동규, "인공지능(AI)의 경제적 영향과 향후 정책방향에 대한 시사점 : 조세 및 사회보장 제도를 중심으로", 한국조세연구포럼, 61-88쪽, 2020년 9월

[4] 윤성재, 박기혁, 김원종, 조한진, "SoC 가상화 플랫폼 기반 효율적인 주변장치의 소비 전력 추정 기술", 대한전자공학회, 2017년 6월

[5] 최규명, 정의영, 엄준형, 어수관, "[특집:SoC Platform 설계기술] 플랫폼을 기반으로 하는 SoC 설계방법", 대한전자공학회, 28-35쪽, 2003년 9월

[6] 이미영, 정재훈, "인공지능프로세서 기술동향과 ETRI 결과물", TECHWORLD ONLINE NEWS, http://www.epnc.co.kr/news/articleView.html?idxno=98308, 2020년 6월 18일

[7] 장준영, 한진호, 배영환, 조한진, "SoC Platform 기반 Design Methodology", KoreaScience, http://www.koreascience.kr/article/JAKO200444948052408.kr, 2004년 8월 20일

[8] 양대규, "정부, AI용 반도체,센서 2235억원 투자 - AI반도체 시장 매년 26.5%↑", AI타임스, http://www.aitimes.com/news/articleView.html?idxno=136132, 2021년 2월 1일

[9] 정윤아, 최승근, "한국 IDC, 국내 인공지능 시장 2023년까지 연평균 17.8% 성장 전망",

https://www.idc.com/getdoc.jsp?containerId=prAP46186820&utm_medium=rss_feed&utm_source=Alert&utm_campaign=rss_syndication, 2020년 4월 2일

[10] 김연균, "'온 디바이스 설명 가능한 AI' 기술 개발 가속도", https://www.koit.co.kr/news/articleView.html?idxno=77759, 2020년 1월 8일