

# 6G를 위한 Edge AI 기술 동향과 향후 연구방향

연세대학교 ■ 이웅섭·윤태용·강예성·김항민

## 1. 서 론

무선통신에 대한 활용도가 높아지면서 5G 기술이 더욱 광범위하게 사용되고 있다. 5G 무선통신 시스템에서는 eMBB, uRLLC 및 mMTC를 기반으로 다양한 시스템에서 새로운 차원의 무선 서비스를 제공한다. 최근 홀로그램/AR/VR/자율이동체와 같은 새로운 무선 서비스를 제공해주기 위해서 5G를 넘어선 6G 기술에 대한 연구가 진행되고 있다. 이러한 새로운 서비스를 효율적으로 제공해주기 위해서 6G 이동 통신에서는 최대 전송속도 1Tbps, 사용자 체감속도 1Gbps 및 0.1ms 이내의 종단간 레이턴시 등 5G 무선 통신 기술의 10~100배 수준의 서비스를 제공해야 한다 [1].

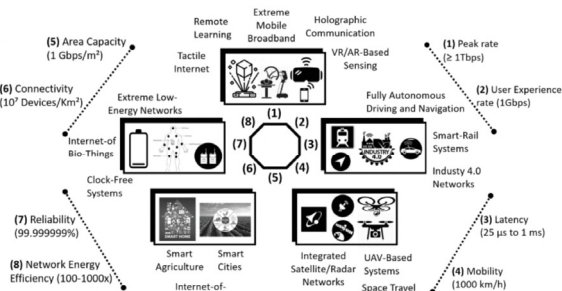


그림 1 6G 무선통신의 요구조건 [1]

이러한 도전적 성능 향상을 위해서 6G에서는 다양한 기계학습 (Machine Learning, ML) 및 인공지능 (Artificial Intelligence, AI) 기반의 기술사용이 고려되고 있다. 특히 ML/AI기술은 복잡한 네트워크를 지능적으로 관리하고, 시스템 최적화 문제를 해결하며, 사용자 경험을 개선하는 데 효율적으로 활용될 수 있다. 현재 ML/AI 기반 무선통신 기술들이 다양한 표준에

서 본격적으로 논의되고 있으며 5G와 6G의 중간단계인 5G Advanced 3GPP 18 Rel. 에서는 본격적인 ML/AI 적용방안에 대한 논의가 진행되고 있다 [2].

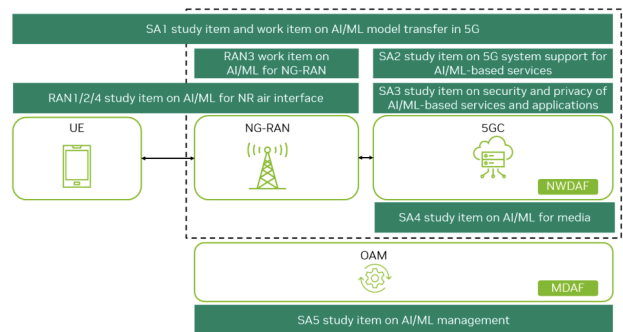


그림 2 5G Advanced에서 AI/ML 연구 [2]

ML/AI 적용을 통해서 무선통신 시스템의 성능을 크게 향상할 수 있지만, 실제 시스템에 적용을 위해서는 여러 고려할 사항들이 있다. 특히 ML/AI의 경우 높은 계산 복잡도 문제가 있고, 클라우드에서 ML/AI 처리를 할 경우 지연시간 문제가 발생할 수 있다. 더불어 네트워크 말단에서 수집된 다양한 사용자 데이터를 통한 학습이 필요할 수 있으므로 데이터에 대한 프라이버시 문제가 발생할 수 있다.

이를 해결하기 위해서 무선통신 시스템에 Edge AI 기술을 접목하는 다양한 연구들이 진행되고 있다. Edge AI에서는 클라우드가 아닌 종단 장비에서 저복잡도로 ML/AI 기능을 수행하여 기존의 문제를 해결할 수 있다. 특히 Edge AI 기술은 계산 능력이 한정적이며 지연시간 감소가 중요한 지능형 사물인터넷(AIoT)에서 유용하게 활용될 수 있다.

본 원고에서는 Edge AI 기법에 대한 기본적인 내용 및 연구 동향에 대해서 알아보고 무선통신에 적용 시 한계점에 대해서 논의한다. 2장에서는 Edge AI 기술의 주요 특징에 대해서 살펴보고, 3장에서는 Edge AI의 기술 동향에 대해서 알아본다. 4장에서는 Edge AI

† This work was carried out with the support of “Cooperative Research Program for Agriculture Science & Technology Development (Project title: Advancement of Hanwoo beef tracing and management technique based on data, Project No. PJ0170202022)” Rural Development Administration, Republic of Korea.

의 구현을 위한 도전과제 및 향후 연구 방향에 대해서 살펴보고, 마지막으로 5장에서 본 원고의 결론을 맺는다.

## 2. Edge AI 정의 및 특징

Edge AI는 인공지능 알고리즘과 데이터 처리를 중앙 데이터 센터나 클라우드가 아닌 데이터 생성 원천, 즉 네트워크 말단에 있는 Edge 장비에서 수행하는 기술을 의미한다. Edge AI는 클라우드 서비스의 데이터 센터와 같은 중앙 집중식 위치가 아닌 사용자 근처에서 AI 계산을 수행한다는 점에서 클라우드 기반 AI와 뚜렷이 차별화될 수 있다 [3]. 새로운 AI 기술의 발전과 사물인터넷 사용의 확산 및 Edge 컴퓨팅 기술을 통해 혁신적인 Edge AI 기술들이 개발되고 있다. 특히 많은 산업분야에서 Edge AI를 활용함으로써 더 나은 서비스 제공이 가능함에 따라, Edge AI에 대한 관심이 더욱 높아지고 있다.

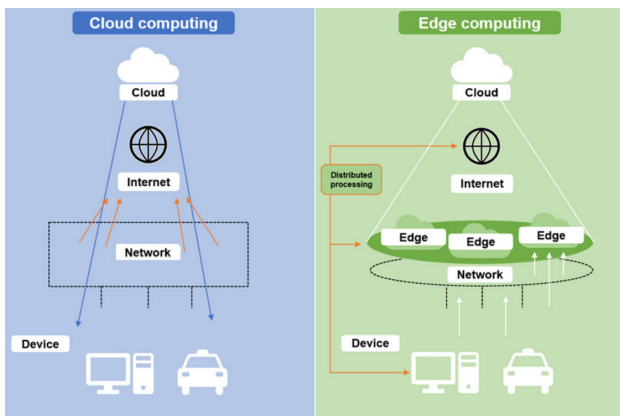


그림 3 클라우드 컴퓨팅 vs Edge 컴퓨팅 [3]

Edge AI 기술은 스마트폰, IoT 센서, Edge 서버와 같이 실제 데이터를 수집하고 생성하는 장치와 물리적으로 가까운 곳에서 실행될 수 있다. 따라서 Edge AI 기술에서는 데이터를 로컬에서 처리함으로써 클라우드 기반 AI 대비 다음과 같은 특징과 이점을 갖는다.

### 2.1 Edge AI의 특징

**[낮은 지연시간]** Edge AI는 데이터가 로컬 디바이스에서 처리 가능하기 때문에 네트워크 전송에 따른 지연시간을 크게 줄일 수 있다 [4]. 특히 이미지 데이터와 같이 큰 데이터를 처리할 경우 Edge AI를 활용하여 지연시간을 대폭 감소시킬 수 있다.

**[네트워크 대역폭 절감]** Edge AI에서는 데이터를

네트워크로 보내기 전에 로컬에서 필요한 분석을 수행하여, 전송할 데이터 양을 최소화할 수 있다. 로컬에서 분석된 결과 중 필요한 정보들만 네트워크를 통해 전송하게 함으로써 사용되는 네트워크 대역폭을 크게 줄일 수 있다.

**[분산된 아키텍처 및 로컬 데이터 처리]** 중앙 서버에 모든 데이터를 집중시키지 않으므로, 시스템 전체가 단일 장애점에 의한 영향을 덜 받게 된다 [6]. 또한, 민감한 데이터를 로컬에서 처리함으로써 데이터 전송 중 발생할 수 있는 보안 위험을 줄이고, 데이터 유출의 위험을 최소화할 수 있다.

**[독립성과 자율성]** Edge AI 시스템은 네트워크 연결이 끊기더라도 독립적으로 작동할 수 있다 [7]. 또한, Edge 장비들이 데이터를 스스로 분석하고 필요한 동작을 기기 자체에서 결정함으로써 사람의 개입 최소화하여 고도화된 자율성을 구현할 수 있다.

**[분산 처리 동작]** Edge AI는 로컬 컴퓨팅 리소스에 의존하기 때문에 처리 능력에 한계가 있을 수 있다. 이는 복잡한 AI 모델이나 대규모 데이터 처리가 요구되는 경우에 걸림돌이 될 수 있다.

### 2.2 Edge AI의 이점

**[실시간 데이터 처리]** 데이터가 로컬에서 처리되기 때문에 네트워크 전송 시간 지연이 거의 없으며, 실시간에 가까운 데이터 처리가 가능하다. 이를 통해 즉각적인 인사이트 도출 및 긴급한 상황에서의 대처가 가능하다.

**[네트워크 비용 및 운영 비용 절감]** 컴퓨팅 자원을 네트워크 말단으로 이동하여 처리함으로써 네트워크 비용을 크게 절감할 수 있고 운영 비용을 줄이는 데 기여할 수 있다.

**[데이터 기밀성 강화]** Edge AI에서는 프라이버시에 민감한 데이터를 로컬에 보관하고 분석 및 관찰 결과만 클라우드로 전송함으로써 데이터의 기밀성을 강화할 수 있다 [8].

**[가용성]** Edge AI는 탈중앙화되어 있어 인터넷 연결 없이도 작동할 수 있으므로 인터넷이 정상적으로 동작하지 않는 환경에서도 활용 가능하다.

## 3. Edge AI 기술 동향

Edge AI 기술은 데이터가 생성되는 네트워크 가장 자리에서 수행되는 특징으로 인해 여러 제약이 있다. 특히 사물인터넷 기기들과 같은 경량 Edge 장비들은 제원에 따른 성능과 연산 속도, 크기, 소비전력 등에 큰 제약이 있다. 더불어 사용자의 데이터를 다루므로

민감한 개인정보 탈취를 위한 공격이 있을 수 있으므로 다양한 보안요건을 고려해야 한다. Edge AI와 관련된 다양한 문제를 해결하기 위해 하드웨어와 소프트웨어 측면에서 여러 기술에 대한 개발 및 연구가 진행되고 있으며, 무선통신과 융합하여 활용하는 다양한 시도가 이루어지고 있다.

### 3.1 Edge AI의 하드웨어 기술 동향

Edge AI의 하드웨어 기술은 효율성 및 지연 시간 측면에서 빠르게 발전하고 있다. 특히 성능적 한계를 극복하기 위해 저전력 AI 프로세서가 개발되고 있고, 하드웨어 가속 기술 또한 동시에 연구되고 있다. 특히 최근에는 Edge 장비에 NPU (Neural Processing Unit) 나 GPU와 같은 연산 가속 프로세서를 통합하여 효율적인 AI 기능 구현이 가능케 한 On-device intelligence 기술에 관한 연구 및 개발이 활발하게 진행되고 있다.

표 1 GPU와 NPU의 비교

	GPU	NPU
정의	그래픽 처리 장치	신경망 처리 장치
특화 연산	그래픽 연산 특화	인공지능 연산 특화
범용성	범용성 중간	범용성 낮음
전력 소모	전력 소모 높음	전력 소모 낮음
그래픽	그래픽 기능 있음	그래픽 기능 없음
특징	CUDA, OpenCL과 같은 GPGPU를 통해 인공지능 연산에 사용 가능	특정 인공지능 프레임워크(Tensorflow 등)에 최적화. GPU에 비해 전력대비 연산속도가 높음.

최근 On-device intelligence를 구현하기 위해 디바이스에서 CNN (Convolutional Neural Network) 및 RNN (Recurrent Neural Network)과 같은 다양한 종류의 딥러닝 모델을 사용할 수 있도록 NPU와 같은 AI 프로세서가 개발되고 있다. NPU는 모바일 기기와 임베디드 시스템에서 실시간 AI 처리를 위해 개발된 프로세서로, 삼성, 퀄컴, 구글, 테슬라 등 여러 칩셋 제조사들이 개발 중에 있다. NPU와 GPU는 CPU와 다르게 병렬 연산이 가능하고, 특히 NPU는 그래픽 연산을 위해 개발된 GPU와 다르게 AI 모델 가속을 전담하도록 개발되어 GPU보다 에너지 효율성이 좋다. Edge AI에 사용할 수 있는 NPU의 예로써 Google Coral 프로젝트에서 개발된 프로세서인 Edge TPU가 있다. Edge TPU는 Tensorflow Lite 기반의 하드웨어 가속만 지원하지만, 2와트의 전력으로 8비트 정수(Int8)로 양자화된 모델을 초당 4조 번의 연산이 가능

하다 (2TOPS/W) [9]. NPU외에도 ASIC (Application Specific Integrated Circuit)과 FPGA (Field Programmable Gate Array) 같은 다양한 반도체 솔루션을 통한 AI 프로세서 개발이 활발하게 진행되고 있다.

### 3.2 Edge AI의 소프트웨어 기술 동향

Edge AI의 소프트웨어 기술도 하드웨어 기술과 마찬가지로 효율성, 지연 시간, 개인정보 보호 및 보안을 고려하여 알고리즘, 프레임워크, 모델 압축 등 관련 연구가 활발히 진행되고 있다. 특히 그래픽 연산을 위해 설계된 GPU를 범용 연산에 사용할 수 있도록 하는 GPGPU (General-Purpose computing on Graphic Processing Unit) 프레임워크, 연산 작업을 분산시켜 처리 성능을 향상시키는 분산학습, 서버와 데이터를 공유하지 않고 가중치만을 공유하여 개인정보를 보호하며 학습할 수 있는 연합학습 및 사물인터넷에 적용할 수 있는 보안 알고리즘이 개발되고 있다.

우선 Edge의 GPU를 AI 계산에 활용하는 GPGPU 프레임 워크가 활발하게 연구되고 있다. 대표적인 예로 NVIDIA Cuda, OpenCL, Apple Metal 등이 있다. 또한 분산학습을 통해 중앙 서버에 의존하지 않고 여러 Edge 장치에서 공통의 AI 모델을 훈련하는 연구도 진행되고 있다. 특히 기울기 값을 적은 수의 비트로 양자화하여 데이터 크기를 줄이는 Gradient quantization, 천천히 변화하는 기울기를 식별하고 재사용하는 방법인 LAG (Lazily Aggregated Gradient) 등을 통해서 분산학습에서 통신 오버헤드를 줄이는 다양한 연구가 진행되고 있다 [10][11]. 더불어 연합학습을 사용하여 원본 데이터 대신 모델의 가중치만을 공유하여 데이터 개인정보 보호를 유지하면서 동시에 Edge 장비에서 AI 모델을 훈련하는 연구가 진행되고 있다. 가중치를 평균화하여 훈련하는 FedAvg (Federated Averaging) 알고리즘, 통계적 이질성을 해결하여 견고성을 향상시키는 FedProx (Federated Proximal) 알고리즘 등 학습 과정에서 보안을 고려한 다양한 연합학습 알고리즘이 연구되고 있다 [12][13]. 마지막으로 Edge AI에서는 네트워크 종단에서 다양한 데이터를 수집하기 때문에 이러한 데이터의 기밀성을 유지하는 것이 중요하다. 이를 위해서 데이터에 노이즈를 추가하여 개인정보 문제를 해결하는 Differential privacy 메커니즘, 여러 사람이 가진 입력값을 공개하지 않고 입력값을 연산하여 공동으로 계산한 결과를 가질 수 있도록 하는 Secure multiparty computation 프로토콜 등 Edge AI에 적용할 수 있는 다양한 보안 알고리즘이 연구되고 있다 [14][15].

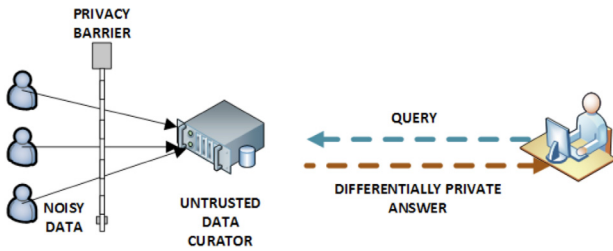


그림 4 로컬 Differential privacy 메커니즘 [14]

### 3.3 Edge AI를 활용한 다양한 사물인터넷 최적화 방안

최근 인공지능을 활용한 무선통신 기술개발이 진행됨에 따라서 사물인터넷 통신에서 Edge AI를 활용하여 무선통신의 성능을 향상시키는 (Learning for communication) 다양한 연구가 진행되고 있다. 특히 많은 수의 사물인터넷 기기들이 동시에 존재하는 환경에서 최적의 자원할당을 수행하거나 주변 스펙트럼 환경을 빠르게 인지하는데 Edge AI 기술이 유용하게 활용될 수 있다. [16]의 연구에서는 Dynamic Nested Neural Network와 Markov Decision Process를 활용한 Edge AI 기반의 동적 협력 자원할당 알고리즘이 제안되었다. 특히 실시간 사물인터넷 환경을 고려하여 알고리즘 계산 시간 단축을 목표로 한 알고리즘을 제안하였다. 또한 [17]의 연구에서는 NVIDIA Jetson Nano와 SDR을 이용한 실시간 저전력 광역 통신망 (Low-Power Wide-Area Network, LPWAN)을 위한 다중 대역 기술 인지방안을 개발하였다. 제안 방안에서는 CNN을 활용하며 무선 대역을 인지하며 SDR 및 NVIDIA Jetson Nano 보드를 활용하여 실제 환경에서 성능검증을 수행하였다. 이를 통해서 제안 방안이 Low SNR 환경에서도 99%의 높은 정확도로 무선 대역을 인지하는 것을 확인하였다. [18]의 연구에서는 Edge AI를 활용한 ISAC (Intertrated of Sensing and Communication) 시스템에서의 핸드오버 기술이 제안되었고, 이를 위한 네트워크 구조도 제안되었다.

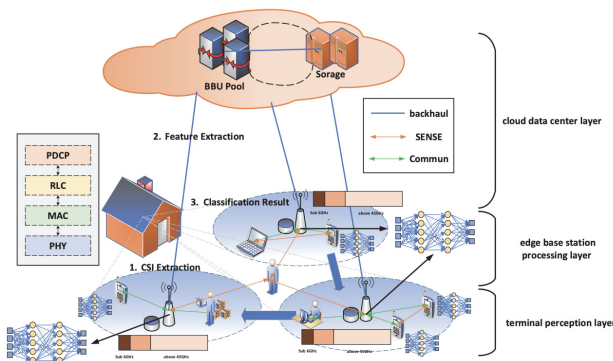


그림 5 Edge AI기반 ISAC 시스템 구조 [18]

## 4. Edge AI 도전과제 및 향후 연구 방향

Edge AI 구현과 관련하여 1) 하드웨어 측면, 2) 네트워크 측면, 그리고 3) 데이터 기밀성 및 보안 측면에서 다양한 도전 과제들이 있다.

### 4.1 하드웨어 측면

Edge AI가 동작하는 Edge 기기와 Edge 게이트웨이는 중앙 클라우드와 비교하여 낮은 처리 능력과 제한된 메모리 용량을 가지고 있다. 또한 저전력 및 제한된 전력으로 동작해야 하는 제약조건이 있다. 따라서 높은 성능을 보이는 기존의 대형 AI 모델들 (예: LLM)의 활용이 제한될 수 있으며 이를 고려하여 효율적으로 AI모델을 경량화 하고, Edge 기기들을 위한 하드웨어 가속기 기술이 필요하다. 우선 모델 경량화 및 최적화 연구에서는 AI 모델을 경량화하거나 훈련된 모델을 더 작게 만들기 위한 방안을 개발한다. 이를 통해서 모델의 크기와 연산 요구사항을 줄이면서도 성능을 유지하는 것을 목표로 한다. 더불어 하드웨어 가속기 연구는 Edge 디바이스에 AI 연산을 가속할 수 있는 전용 하드웨어의 개발을 통해서 Edge 기기들의 연산 능력을 증대시켜 복잡한 AI 모델의 사용이 가능케 한다. 또한 많은 Edge AI 기기들의 경우 한정된 전력환경에서 사용될 수 있으므로 전력소모를 고려한 하드웨어 디자인도 중요하다.

### 4.2 네트워크 측면

Edge AI는 네트워크의 엣지에서 동작하지만 구현 방식에 따라서 여전히 클라우드와의 데이터 송수신이 필요할 수 있다. 또한 낮은 성능을 지닌 Edge 기기 간의 연합학습 및 추론에서 Edge AI를 활용하는 경우에서도 네트워크 활용이 필요하다. 이때 네트워크 접근이 제한되거나 일관되지 않을 경우 Edge AI의 원활한 동작이 어려울 수 있고 이미지 데이터와 같은 큰 용량의 데이터 전송 시 대역폭 문제가 발생할 수 있다. 더불어 연합학습에서는 Edge AI 기기 간 학습된 모델 파라미터 공유에 많은 통신 리소스가 사용될 수 있다.

이러한 네트워크 측면의 문제를 해결하기 위해서 로컬 데이터 처리 및 Edge AI 간 데이터 전송 최적화 연구가 필요하다. 특히 로컬 데이터 처리를 통해서 최대한 많은 데이터를 엣지에서 처리하여 기기 간의 데이터 송수신 오버헤드를 줄임으로써 네트워크 의존성을 줄일 수 있다. 또한 공유 데이터를 효율적으로 선택함으로써 네트워크의 트래픽을 효율적으로 줄일 수 있다. 더불어 데이터 전송 최적화를 통해서 Edge AI 간 데이터 전송시 네트워크 대역폭을 효율적으로 활

용할 수 있다. 또한 전송속도를 높임으로써 Edge AI 기술의 지연 시간을 줄일 수 있다.

#### 4.3 데이터 기밀성 및 보안 측면

Edge AI는 사용자의 건강 기록등과 같이 개인 정보 데이터를 다루는 경우가 많으므로 활용 데이터의 기밀성과 보안을 보장하는 것이 중요하다. 이러한 문제를 해결하기 위해 효율적 데이터 선택 및 데이터 암호화 등에 대한 연구가 필요하다. 효율적 데이터 선택에서는 수집된 데이터 중 개인정보에 민감한 데이터를 처리하여 활용함으로써 데이터의 기밀성 및 보안을 유지하는 방식이다. 특히 연합학습등을 통해서 다수의 Edge AI 노드들이 학습 데이터 및 업데이트된 그라디언트를 효율적으로 전처리 함으로써 데이터의 보안성을 지킬 수 있다 [19]. 또한 데이터 암호화는 데이터가 전송되거나 저장되는 동안 암호화하여 기밀성을 유지하는 방식이다. Edge 기기들을 활용한 Edge AI에서는 노드들의 낮은 계산 성능 및 전력 소모 제약조건들에 의해서 복잡한 암호화 기술을 활용할 수 없으므로 이에 적합한 경량 암호화 기술의 개발이 필요하다.

### 5. 결 론

본 원고에서는 Edge AI 기술의 개념과 관련 연구 동향을 살펴보았다. 현재 다양한 ML/AI 기술들이 무선 통신에 접목되면서 성능을 크게 변화시키고 있으므로 Edge AI의 효율적인 접목을 통해서 무선통신, 특히 6G 및 지능형 사물인터넷 기술의 성능을 비약적으로 향상시킬 수 있을 것이다. 이를 위해서는 다양한 현실적 제약조건을 고려한 Edge AI기술 연구가 필요하다.

### 참고문헌

[1] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proceedings of the IEEE*, vol. 109, no. 7, pp. 1166-1199, Jul. 2021.

[2] X. Lin, "Artificial intelligence in 3GPP 5G-Advanced: A survey," *Global Communications*, 2024.

[3] J.-H. Huh and Y.-S. Seo, "Understanding edge computing: Engineering evolution with artificial intelligence," *IEEE Access*, vol. 7, pp. 164229-164245, 2019.

[4] E. Li, L. Zeng, Z. Zhou, and X. Chen, "Edge AI:

On-demand accelerating deep neural network inference via edge computing," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 447-457, 2019.

[5] R. Marculescu, D. Marculescu, and U. Ogras, "Edge AI: Systems design and ML for IoT data analytics," in *Proc. of 26th ACM SIGKDD*, Virtual Event, CA, USA, Aug. 2020.

[6] N. Moustafa, "A new distributed architecture for evaluating AI-based security systems at the edge: Network TON\_IoT datasets," *Sustainable Cities and Society*, vol. 72, p. 102994, May 2021.

[7] W. Su, L. Li, F. Liu, M. He, and X. Liang, "AI on the edge: a comprehensive review," *Artificial Intelligence Review*, vol. 55, no. 8, pp. 6125-6183, Mar. 2022.

[8] M. Shafique, A. Marchisio, R. V. W. Putra, and M. A. Hanif, "Towards energy-efficient and secure edge AI: A cross-layer framework ICCAD special session paper," in *Proc. of 2021 IEEE/ACM ICCAD*, Munich, Germany, Nov. 2021.

[9] Google Coral EdgeTPU, [Online]. <https://coral.ai/technology/>

[10] D. Alistarh, D. Grubic, J. Z. Li, R. Tomioka, and M. Vojnovic, "QSGD: Communication-efficient SGD via gradient quantization and encoding," in *Proc. of NeurIPS*, Long Beach, CA, USA, Dec. 2017.

[11] J. Sun, T. Chen, G. B. Giannakis, and Z. Yang, "Communication-efficient distributed learning via lazily aggregated quantized gradients," in *Proc. of NeurIPS*, Vancouver, Canada, Dec. 2019.

[12] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. of PMLR*, Apr. 2017.

[13] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. of MLSys*, Austin, TX, USA, Apr. 2020.

[14] M. A. Husnoo, A. Anwar, R. K. Chakraborty, R. Doss and M. J. Ryan, "Differential privacy for IoT-enabled critical infrastructure: A comprehensive survey," *IEEE Access*, vol. 9, pp. 153276-153304, 2021.

[15] O. O. Olakanmi and K. O. Odeyemi, "Trust-aware and incentive-based offloading scheme for secure multi-party computation in Internet of Things," *Internet of Things*, vol. 19, Apr. 2022.

[16] K. Lin, Y. Li, Q. Zhang, and G. Fortino, "AI-driven



collaborative resource allocation for task execution in 6G-enabled massive IoT,” *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5264-5273, Apr. 2021.

- [17] J. Fontaine, A. Shahid, R. Elsas, A. Seferagic, I. Moerman, and E. De Poorter, “Multi-band sub-GHz technology recognition on NVIDIA’s Jetson Nano,” *Proc. of IEEE VTC2020-Fall*, Victoria, BC, Canada, Nov. 2020.
- [18] X. Liu, H. Zhang, K. Sun, K. Long, and G. Karagiannidis, “AI-driven integration of sensing and communication in the 6G era,” *IEEE Network*, Early Access, 2024
- [19] J. Huh and W. Lee, “Privacy-preserving consumer churn prediction in telecommunication through federated machine learning,” in *Proc. of 2024 IEEE BigComp*, Bangkok, Thailand, Feb. 2024.

## 약 력



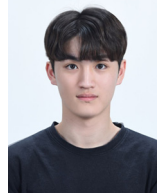
### 이 웅 섭

2006 KAIST 전기및전자공학과 졸업(학사)  
 2012 KAIST 전기및전자공학과 졸업(박사)  
 2012 KAIST BK21 (박사후연구원)  
 2013 독일 FAU (박사후연구원)  
 2014~2022 경상국립대학교 정보통신공학과 부교수  
 2023~현재 연세대학교 정보대학원 부교수  
 관심분야 : AIoT, LLM, 인공지능 기반 무선통신  
 Email : woongsup.lee@yonsei.ac.kr



### 윤 태 웅

2020 충북대학교 축산학과 졸업(학사)  
 2024 전남대학교 데이터사이언스학과 졸업(석사)  
 2024~현재 연세대학교 정보대학원 박사과정  
 관심분야 : AIoT, Data Analysis, Smart Livestock Management System  
 Email : tyun24@yonsei.ac.kr



### 강 예 성

2024 한세대학교 컴퓨터공학과 졸업(학사)  
 2024~현재 연세대학교 정보대학원 석사과정  
 관심분야 : AIoT, TinyML, Edge AI  
 Email : kkds002@yonsei.ac.kr



### 김 황 민

2020 중앙대학교 지식경영학부 졸업(학사)  
 2023~현재 연세대학교 정보대학원 석사과정  
 관심분야 : BigData Analysis, Prompt Engineering, LLMs  
 Email : khm0106@yonsei.ac.kr