



김 정 창
국립한국해양대학교

‘온디바이스 AI 기반 미디어 기술’ 특집호를 내며

이번 특집호에서는 “온디바이스 AI 기반 미디어 기술”을 주제로 미디어 기술의 혁신을 가져오고 있는 온디바이스 AI 기술과 그 응용에 대해서 살펴보고자 합니다.

최근 급격히 발전하고 있는 인공지능 기술은 클라우드에서 디바이스로 이동하고 있으며 이에 따라 미디어 산업과 데이터 처리 방식에 큰 변화가 일어나고 있습니다. 이번 특집호의 기고문에서 알 수 있듯이, 온디바이스 AI는 미디어 기술의 혁신을 가져오는데 핵심적인 역할을 하고 있습니다. 초저전력 임베디드 AI, 경량화된 생성형 AI, 데이터셋 응축 등 각 기고문은 온디바이스 AI의 핵심적인 기술과 응용, 그리고 확장성에 대해서 소개하고 있습니다. 클라우드에 대한 의존성에서 벗어나 개인정보 보호, 실시간, 에너지 효율, 개인화 등 새로운 가치를 창출하는 온디바이스 AI의 발전은 미디어 산업뿐만 아니라 다양한 산업 분야의 지능화와 디지털 전환을 앞당겨 줄 것입니다. 다양한 하드웨어 플랫폼에서 AI 가속화는 초소형 디바이스에서도 고성능의 AI 추론을 가능하게 하며, 다양한 경량화 기법을 통해 AI 모델의 크기와 연산량을 줄이면서도 성능 저하를 최소화함으로써 자원 제약 환경에서도 실제 사용 가능한 AI의 구현이 가능해지고 있습니다. 또한, 생성형 AI의 온디바이스 경량화를 통해 실시간 영상/음성 합성과 AR/VR 콘텐츠 제작 등 미디어 산업의 자동화와 개인화, 몰입형 경험이 향상되고 있습니다. 데이터를 사용자 디바이스에서 처리함으로써 개인정보 보호와 보안성을 크게 향상시킬 수 있으며, 이는 사용자 신뢰와 서비스 확산에 크게 기여할 수 있습니다. 이로 인해 스마트홈, 스마트시티, 제조, 의료, 교육 등 다양한 산업 분야에서 온디바이스 AI 기반 미디어 기술이 빠르게 적용되고 있습니다. 본 특집호에서는 초경량 모델에서 LLM 적용까지 온디바이스 AI 기술 동향, 온디바이스 미디어 제작을 위한 생성형 인공지능 경량화 기술 현황 및 전망, TinyML 기술 동향과 산업 응용, 서브셋 매칭 알고리즘을 이용한 효율적인 데이터 응축 기술을 소개한 온디바이스 AI 분야 전문가들의 원고 4편을 준비했습니다.

첫 번째 기고문(“온디바이스 AI 기술 동향: 초경량 모델에서 LLM 적용까지”)에서는 온디바이스 AI의 하드웨어와 소프트웨어 플랫폼, 모델 경량화 기법, 주요 응용 사례, 그리고 경량화된 대규모 언어모델 적용 가능성을 다룹니다. 마이크로컨트롤러(MCU), 모바일 AP, NPU 등 다양한 하드웨어와 TensorFlow Lite, CMSIS-NN 등 경량 추론 프레임워크의 발전으로 온디바이스 AI의 적용 가능성이 높아지고 있습니다. Pruning, Quantization, Knowledge Distillation 등 모델 경량화 기술을 통해 음성 인터페이스, 영상 인식, 시계열 예측과 이상치 진단 등 다양한 분야에서 온디바이스 AI가 적용되고 있습니다. 본 기고문에서는 온디바이스 AI 구현을 위한 핵심 기술과 플랫폼, 경량화 기법과 주요 활용 사례를 소개합니다.

두 번째 기고문(“온디바이스 미디어 제작을 위한 생성형 인공지능 경량화 기술 현황 및 전망”)

에서는 온디바이스에서 생성형 AI 기반 미디어 제작 기술을 구현하기 위한 경량화 기술을 집중적으로 분석합니다. CNN 기반 경량화 기법의 원리, 트랜스포머와 디퓨전 모델 등 생성형 AI에 특화된 구조적 최적화 방법을 상세히 소개하고 있습니다. Meta의 Spark AR, 애플 Clips, 구글 Pixel, 네이버 제페토 등 국내외 온디바이스 생성형 AI 서비스 사례를 제시하며, 개인정보 보호, 실시간, 개인화 등 온디바이스 미디어 제작의 강점과 미래 전망을 제시합니다. 하드웨어 친화적 모델 설계, 혼합 정밀도 연산, AR/VR 융합 등 차세대 미디어 기술의 발전 방향도 함께 논의하고 있습니다.

세 번째 기고문(“TinyML 기술 동향과 산업 응용”)에서는 초소형·초저전력 마이크로컨트롤러에서 머신러닝을 구현하는 TinyML의 최신 동향과 산업 응용을 다루고 있습니다. 최근 퀄컴의 Edge Impulse 인수, ST마이크로일렉트로닉스·NXP·TI 등 글로벌 MCU 제조사의 AI 가속기 내장 제품 출시 등 TinyML 생태계의 급격한 변화를 소개하고 있습니다. TinyML은 스마트홈, 스마트시티, 산업 현장 등에서 실시간과 초저전력 지능형 시스템 구현을 가능하게 하며, 특히 음향 센서 기반 시계열 데이터 분석에서 활용되고 있습니다. 국내에서는 LG전자, 어보브반도체, 감바렘스 등도 초경량 온디바이스 AI 기술 개발에 적극적으로 나서고 있습니다. 본 기고문에서는 TinyML의 최신 하드웨어 및 소프트웨어 동향, 주요 응용 사례, 국내 기업들의 대응 현황을 살펴보고, 미래에 TinyML이 탄소배출 감축, 프라이버시 보호, 디지털 격차 해소에 기여할 수 있는 가능성을 다루고 있습니다.

네 번째 기고문(“서브셋 매칭 알고리즘을 이용한 효율적인 데이터 응축 방법”)에서는 대용량 이미지 데이터셋을 소규모로 압축해 효율적으로 학습할 수 있도록 하는 데이터셋 응축 분야의 최신 연구를 소개합니다. 기존 그래디언트 매칭 기반 응축법의 한계(지역적 텍스처 정보 손실)를 극복하기 위해 원본·합성 데이터셋을 여러 부분집합으로 분할해 각 부분집합 내에서 독립적으로 그래디언트 매칭을 수행하는 ‘서브셋 매칭’ 기법을 제안합니다. 이 방법은 미세한 텍스처와 형태 정보를 더 잘 보존하며, 다양한 실험에서 기존 방법 대비 우수한 성능을 입증하였고, 온디바이스 AI 환경에서 데이터 저장·학습 효율을 극대화할 수 있는 실질적 해법을 제시합니다.

이번 “온디바이스 AI 기반 미디어 기술” 특집호 출간을 위해서, 바쁘신 와중에서도 원고를 준비해 주신 모든 저자분들께 깊이 감사드립니다. 네 편의 기고문은 온디바이스 AI의 기술적 향상, 산업적 파급력, 그리고 확장성을 구체적으로 보여주며, 독자 여러분께 최신 동향과 실질적 인사이트를 제공하리라 기대합니다. 아무쪼록 본 특집호가 온디바이스 AI 기반 미디어 기술에 대한 동향 파악과 관련 연구에 조금이나마 도움이 되기를 기대합니다.