

2024년도 대한전자공학회
하계종합학술대회 특별세션 강연 발표정보

■ 발표자 정보

성명	전석훈	사진	
소속(학교)	한국전자기술연구원(KETI)		
부서(학과)	SoC플랫폼연구센터		
직위	선임 연구원		
발표분야	인공지능 소프트웨어 기술		
약력	<p>학력 및 경력</p> <p>2010~2012: 송실대학교 정보통신공학과 석사 2013~2017: Seagate Korea Design Center 2017 ~: 한국전자기술연구원SoC플랫폼연구센터 선임연구원</p> <p>연구실적</p> <ul style="list-style-type: none">- 시뮬레이션 기반 고속/고정확도 데이터센터 워크로드/시스템 분석 플랫폼 개발 (IITP, 2024.04 ~ 2027.12, 주관 연구책임자)- 거대인공신경망 처리 PIM-NPU 지원 시스템 SW 기술개발 (IITP, 2023.04 ~ 2027.12, 공동 연구책임자)- 재구성형 PIM 디바이스 기반의 Memory-Centric 아키텍처 개발 (IITP, 2022.04 ~ 2028.12, 공동 연구책임자) <p>주요연구 및 관심분야</p> <ul style="list-style-type: none">- 통합 인공지능 소프트웨어 플랫폼 개발- LLM 성능 분석 및 프로파일러 개발- 양자컴퓨팅 소프트웨어 프레임워크		

■ 강연 정보

제 목	인공지능 소프트웨어 기술 개발 동향
Abstract	<p>본 강연에서는 인공지능 소프트웨어의 기술개발 동향에 관해 설명한다. 한국전자기술연구원 SoC 플랫폼 센터에서 진행한 연구 및 기술개발 히스토리를 통해 최근 중요하게 떠오르는 HW-SW co-design과 optimization에 관한 내용을 소개한다. 먼저 엣지 디바이스를 고려한 인공지능 실행 및 양자화 방법 그리고 개발을 위한 MLOPS SW플랫폼 기술에 대해 소개한다. 다음으로 대규모 언어 모델 (LLM)의 특성과 성능 분석 그리고 프로파일링을 통한 분석 내용을 통해 HW를 고려한 특성 분석 및 접근방법에 관해 다룬다. 본 센터의 연구 사례를 바탕으로 하드웨어에 최적화된 인공지능 실행을 위한 소프트웨어 기술개발의 중요성에 대해 정리한다.</p>