



배성호
경희대학교

'VLM/LLM 기반 미디어 처리' 특집호를 내며

이번 특집호에서는 최근 빠르게 발전하고 있는 대규모 언어 모델(LLM)과 비전-언어 모델(VLM) 기술이 미디어 처리 전반에 가져온 변화와 연구 동향을 집중 조명하고자 합니다.

텍스트, 영상, 이미지, 코드 등 다양한 데이터가 융합되는 환경에서 LLM과 VLM은 단순한 정보 이해를 넘어 생성, 편집, 평가까지 아우르는 범용 인공지능 기술로 자리매김하고 있습니다. 본 특집호는 이러한 흐름 속에서 기술적 진보와 산업 응용, 그리고 도메인 특화 과제를 다루는 여섯 편의 기고문을 담았습니다.

첫 번째 기고문("VLM 등장 및 기술 동향: LLM에서 산업 응용까지")은 대규모 언어 모델의 발전사를 돌아보고, VLM의 출현과 학습 전략, 인스트럭션 투닝, RLHF 등 주요 기술을 정리하며 산업 현장에서의 활용 가능성을 소개합니다. LLM이 언어 처리에서 출발해 멀티모달 인공지능으로 확장된 큰 흐름을 조망할 수 있습니다.

두 번째 기고문("코드 거대 언어 모델 연구 동향")은 코드 이해와 생성이라는 특수 영역에서의 LLM 적용을 다룹니다. 코드 검색, 클론 탐지, 버그 검출, 취약점 분석 등 다양한 과업에 어떻게 대규모 언어 모델이 사용되고 있는지, 그리고 데이터 품질, 의미적 격차, 평가 지표의 한계와 같은 도전 과제를 함께 분석합니다.

세 번째 기고문("Semantic Fidelity Meets Efficiency: Reviewing Vision-Language Models (VLMs) and Compression Approaches")은 VLM과 데이터 압축 기술을 교차하는 새로운 연구 영역을 체계적으로 정리합니다. 토큰 프루닝·머징, 압축 도메인 추론, 의미 기반 압축, 엣지-클라우드 분산 처리 등 여섯 가지 연구축을 제시하며, 효율성과 의미 보존의 균형을 강조합니다.

네 번째 기고문("The Role of Diffusion Models in Modern Visual Content Creation")은 확산 모델(Diffusion Model)이 이미지 및 영상 생성·편집에 미친 영향을 조명합니다. 잠재공간 확산(LDM), 조건 제어(ControlNet), 비디오 확장 기법 등 최신 연구를 소개하며, 미디어 제작 환경에 확산 모델이 필수 도구로 자리잡고 있음을 보여줍니다.

다섯 번째 기고문("Leveraging Vision-Language Models for Remote Sensing VQA: Analysis, Gaps, and Directions")은 원격탐사(RS) 영상이라는 특수 도메인에서 VLM의 성

능을 분석합니다. RS-LLaVA와 일반 LLaVA를 비교 실험하여 현황과 격차를 제시하며, 다원천 데이터와 공간·스펙트럼 정보에 최적화된 학습 전략의 필요성을 강조합니다.

여섯 번째 기고문(“Evaluating Vision-Language Models for Scanned Image Quality Assessment”)은 문서 스캔 이미지 품질 평가라는 현실적 과제에 VLM을 적용한 연구입니다. 다양한 열화 요인에 대해 VLM 기반 평가자의 가능성과 한계를 검토하며, 정량적 지표와 인공지능 기반 품질 판정의 접목 방안을 탐색합니다.

이번 “VLM/LLM 기반 미디어 처리” 특집호는 기초 기술의 전반적 동향부터 효율화, 새로운 생성 패러다임, 특수 도메인 응용, 그리고 품질 평가에 이르기까지 다층적인 연구 지형을 담고 있습니다. 바쁘신 가운데도 원고를 준비해 주신 저자분들께 깊은 감사를 드리며, 본 특집호가 독자 여러분께 향후 연구와 산업 응용의 나침반이 되기를 기대합니다.