2025 2학기 컴퓨터소프트웨어학부

CSE융합세미나

장소 ITBT관 911호

날짜 / 시간 2025.12.03 16:00~18:00

Upcoming Talk

이가영

연구원/ NAVER AI Lab

이미지 생성 모델의 안전 및 윤리 문제

최근 이미지 생성 모델의 성능이 급격히 향상되면서, 이러한 모델이 현실과 구분하기 어려운 고품질 이미지를 생성할 수 있게 되었습니다. 그러나 이와 동시에 모델이 부적절하거나 유해한 콘텐츠를 생성하는 사례도 늘어나고 있어, 생성형 AI의 안전성에 대한 사회적 우려가 커지고 있습니다.
본 세미나에서는 이미지 생성 모델이 야기할 수 있는 잠재적 위험 요소를 다각도로 분석하고, 편향된 데이터, 저작권 침해, 프라이버시 침해, 허위 정보 생성과 같은 주요 윤리적 문제들을 살펴봅니다. 또한 데이터 필터링, 안전 스캐닝, 모델 언러닝(Unlearning), 프롬프트 제어 기법 등 최근 연구에서 제안된 다양한 해결 방법들을 소개합니다.

한양대학교 HANYANG UNIVERSITY