

# 보도자료

보도 희망일	제한없음(즉시)
문의	AI연구원
	원유리 전화: 880-4259, 이메일: yrwon@snu.ac.kr

배포일: 2024.1.23.(화)

## 제3기 율촌AI장학생 김현우 박사, 국제 최우수 AI학회 논문상 수상

- 고품질 대규모 대화 데이터셋 간편하게 구축하는 방법론 개발
- 세계 최대 규모 데이터셋 'SODA' 공개

□ 서울대 AI연구원(원장 장병탁)은 율촌 AI 장학생 김현우 박사가 국제 최우수 AI 학회인 EMNLP 2023에서 한국인 제1저자로 논문상을 최초 수상했다.

□ 김현우 박사는 농심그룹 율촌재단에서 지원하고 서울대 AI연구원이 운영하는 '율촌 AI 장학생' 3기이다. AI 원천기술 분야에서 잠재력을 보이는 학생에게 지원하는 '율촌 AI STAR' 장학금을 지원받은 바 있다.

□ 서울대 컴퓨터공학부 김현우 박사와 김건희 교수가 미국 시애틀 앨런 인공지능 연구소 및 워싱턴 대학교의 최예진 교수 등과 함께 진행한 연구가 국제 최우수 자연어 처리 학술대회 EMNLP 2023에서 Outstanding Paper Award를 수상하였다. 국내 대학 연구진이 EMNLP에서 수상을 한 것은 이번이 최초다.

□ 기존 인공지능 대화 분야는 고질적으로 데이터 부족 문제에 시달려왔다. 일상에서 이루어지는 대화들은 데이터로 기록되지 않는 경우가 대부분이며, 여러 법적 제약들이 존재하기 때문에 연구에 활용하기에는 어려웠다. 이에 대한 대안으

로 클라우드 소싱을 통해 대화 데이터셋을 제작했지만 이는 비용과 크기에 대한 문제점이 있다. 그 결과, 기존 대화 관련 연구들은 모두 소수의 작은 데이터셋 위에서 이루어져 범용성에 큰 제약이 있었다.

□ 수상 논문인 ‘사회 상식 맥락화 기법을 통한 백만 단위의 대화 증류(SODA: Million-scale Dialogue Distillation with Social Commonsense Contextualization)’는 이를 해결하기 위해 거대 언어모델(large language model; LLM)과 기호 기반 상식 그래프(symbolic commonsense knowledge graph)를 함께 사용하여 데이터셋을 효율적으로 자동 구축하는 방법론인 CO\_3 증류 프레임워크(distillation framework)을 제시하였다.

□ 기존에 클라우드 소싱을 활용하면 대화 하나당 길게는 몇 분이 걸리던 작업을 CO\_3 프레임워크는 6초 내외로 단축한다. 연구팀은 이를 통해 세계 최초로 150만 규모의 세계 최대 고품질 일상 대화 데이터셋인 SODA를 무료로 공개하였는데, 이는 기존 데이터셋 대비 100배 큰 규모다.

□ 크기뿐만 아니라, 대화의 자연스러움, 일관성, 구체성 등 품질 또한 기존 클라우드 소싱 기반 대화 데이터셋에 비해 2배 이상 뛰어난 것으로 연구 결과가 나타났다. 그 결과, SODA에 학습시킨 챗봇 COSMO는 파라미터 수가 30억 개로 작은 크기임에도 불구하고 기존 큰 챗봇들(파라미터 수 70억)보다 대화 맥락에 더 적합하고 자연스러운 답변이 가능하다.

□ 해당 논문은 미국 시애틀의 앨런 인공지능 연구소(Allen Institute for AI; AI2), 워싱턴 대학교 (University of Washington), 카네기멜론 대학교 (Carnegie Mellon University), 서던 캘리포니아 대학교 (University of Southern California), 그리고 피츠버그 대학교 (University of Pittsburgh)와 함께 국제 공동협력 연구로 이루어 낸 결과다.

□ 김현우 박사는 서울대 박사과정 중 미국 시애틀의 AI2에 인턴을 수행하며 국제 협력 연구를 이끌었으며 졸업 후 AI2에서 박사후연구원으로 취업하여 후속 연구 이어가고 있다.

□ 한편, 서울대 AI연구원은 2020년 농심그룹 율촌재단과 업무협약을 체결하고 협약에 따라 율촌재단은 AI 인재 양성을 위해 매년 2억 원 규모의 장학금을 지원하고 있다. 김현우 박사와 같이 세계적인 수준의 AI 연구성과의 잠재력을 가진 대학원생을 대상으로 매년 엄격한 절차를 거쳐 선발하고 있다. 올해 율촌 AI 장학생은 6월 모집 예정이다.