

[보도자료]**바이브컴퍼니, 'AI 학습용 데이터 구축' 사업자로 선정**

바이브컴퍼니(대표 송성환, (구)다음소프트, 이하 바이브)가 자사의 독보적인 기술 플랫폼 소피아(Sofia)를 기반으로 '인공지능(AI) 학습용 데이터 구축 사업(2차)'에 나선다고 21일 밝혔다.

총 2,925억 원 규모의 '인공지능 학습용 데이터 구축 사업'은 올해 정부가 적극 추진하는 디지털 뉴딜의 대표사업으로, 과학기술정보통신부가 주최하고 한국정보화진흥원(NIA)이 주관한다.

이는 포스트 코로나와 인공지능 시대에 발맞춰, AI 핵심 경쟁력인 대규모 데이터를 단기간에 확보해 AI 선도국가로 도약하는 한편, 일자리 창출과 경제성장 동력을 확보하는 등 관련 산업 생태계를 조성하기 위해 실시하는 사업이다.

바이브는 지정 공모 48개 과제 중 57억 규모의 '요약 데이터 AI 데이터'와 76억 규모의 '한국어 텍스트 데이터 AI 데이터' 과제에 각각 수행기관과 참여기관의 자격으로 오는 12월까지 4개월 간 사업을 수행한다.

이는 ▲논문자료 요약 데이터 ▲도서자료 요약 데이터 ▲한국어 대화 요약 데이터 등을 포함하는 분야로 바이브는 이중 한국어 대화 요약 데이터 구축을 담당한다. 뉴스, 기사 등의 문어체에 비해 생략이나 변형이 많고 문맥을 고려해야 하는 일상 대화, 토론, 회의록, 상담 내역 등 다양한 유형의 한국어 대화 원문 데이터를 기반으로 한국어 대화 요약 AI 기술 개발을 위한 데이터셋을 구축한다.

또한, '한국어 텍스트 AI 데이터' 과제에서도 한국어 구어체 텍스트 기반의 정보검색, 대화분석, 질의응답, 명령어 이해, 언어모델 학습 등 자연어처리 AI 기술 개발을 위한 한국인의 일상대화 메신저 채팅 데이터를 구축하는 데 참여한다.

바이브는 자체 개발 기술 플랫폼인 소피아 내 대용량 데이터를 추출하는 '트렌드맵'과 자연어처리 기술을 기반으로 한 문서 요약 기술을 보유하고 있다. 크게 키워드 입력을 통해 검색된 문서를 요약하는 데 특화된 TextRank 기술, 한국어의 특성을 고려한 추상적 요약 기술, 리뷰 등 데이터 해석 기반의 요약 기술 등을 '썸트렌드(Sometrend)' 등 자사 서비스에 적용해 사업을 진행 중이다. 또, 지난 4월에는 기술특례상장을 위한 기술평가에서 전문 평가기관인 이크레더블로부터 최고 등급인 'AA'등급을 받으며 기술의 우수성을 입증 받기도 했다.

바이브 이기황 이사는 "바이브가 오랜 시간 동안 데이터를 수집, 분석해 서비스를 구축해 온 경험을 토대로 이번 구축 사업 참여를 통해 디지털 뉴딜에 기여할 뿐만 아니라 자사에서 서비스 예정인 데이터 API 관련 비즈니스도 더욱 확장해 나갈 수 있을 것."이라고 말했다.