

「방송영상 AI 학습용 데이터 구축」 컨소시엄 선정

- 방송사업자와 AI·데이터 기업·기관이 협력하는 총 4개 컨소시엄 선정
- 방송영상으로 약 21,000시간의 고품질 AI 학습용 데이터 구축 본격 착수

과학기술정보통신부(장관 배경훈, 이하 ‘과기정통부’)와 한국전파진흥협회(회장 홍범식)는 「방송영상 인공지능(AI) 학습용 데이터 구축」 사업 지원대상으로, **문화방송(MBC)** 컨소시엄, **엠비씨충북** 컨소시엄(지역MBC 총 15개사), **케이티이엔에이(KT ENA)** 컨소시엄, **한국방송공사(KBS)** 컨소시엄 등 총 4개 컨소시엄을 선정했다고 밝혔다.

< 2025년 「방송영상 AI 학습용 데이터 구축」 사업 지원대상 컨소시엄 선정 결과 >

연번	주관기관	참여기관
1	문화방송	아이엠비씨, 클라우드웍스, 데이터메이커, 엘지경영개발원 AI연구원(4개)
2	엠비씨충북	엠비씨강원영동 등 지역MBC 14개사, 도스트11, 제머나이소프트, 한국전자통신연구원, 유핀(18개)
3	케이티이엔에이	디오비스튜디오, 매트릭스, 아카이브웍스, 포스트에이아이, 어니컴(5개)
4	한국방송공사	KBS미디어텍, 스펙스페이스, 네이버클라우드, 국가유산진흥원(4개)

※ 컨소시엄 순서는 주관기관 가나다 순

동 사업은 영상 등을 생성하는 AI 모델 개발 지원을 위해 방송영상을 활용한 고품질 AI 학습용 데이터를 구축하는 사업으로, 2025년 1차 추경으로 반영되어 신규 추진된다. 지원대상은 방송법에 따른 방송사업자(주관기관)와 AI·데이터 기업·기관 등(참여기관)으로 구성된 컨소시엄으로, 4개 컨소시엄 선정에 12개 컨소시엄이 지원하여 3:1 경쟁률을 보였다. AI·데이터 등 전문가로 구성된 심사위원회 평가를 통해, 우리나라 고유의 가치, 특성이 반영된 방송콘텐츠를 기반으로 한 AI 학습용 데이터 구축 및 활용 계획, AI 기술역량 등이 우수한 컨소시엄 4개를 최종 선정했다.

과기정통부는 선정된 4개 컨소시엄에 각각 48.3억원을 지원하며, 4개 컨소시엄은 저작권 이슈가 해소된 국내 방송영상 원본 총 42,000시간을 활용하여 총 21,000시간의 고품질 방송영상 AI 학습용 데이터를 구축한다. 과기정통부는 각 컨소시엄이 고품질 데이터를 구축하도록, 데이터 품질검증 전문기관을 통해 품질기준 설정, 단계별 품질 관리·컨설팅, 품질기준 달성 검증 등 데이터 품질을 전주기 관리·검증한다.

컨소시엄별 사업 내용은 다음과 같다.

(문화방송 컨소시엄) 버추얼 스튜디오 등 다양한 방송콘텐츠 제작 환경에서 사용될 고정밀·고품질의 배경영상을 생성하는 AI의 학습용 데이터를 구축한다. 보도, 시사·교양, 예능, 드라마 등 10,000시간의 원본 영상을 활용하여, 3,633시간의 영상데이터, 49.2만건의 이미지데이터 등 총 5,000시간 분량의 AI 학습용 데이터를 구축한다.

(엠비씨충북 컨소시엄[지역MBC 총 15개사*]) 우리나라 각 지역의 고유한 문화와 생활양식 등을 반영한 이미지 및 영상을 생성하는 AI의 학습용 데이터를 구축한다. 보도, 시사·교양, 다큐 등에서 각 지역의 다양한 자연·풍경, 생활·문화, 역사·사회 등과 관련된 12,000시간의 원본 영상을 활용하여, 3,600시간의 영상데이터, 84.24만건의 이미지데이터 등 총 5,940시간 분량의 AI 학습용 데이터를 구축한다.

* MBC충북, MBC강원영동, 광주MBC, MBC경남, 대구MBC, 대전MBC, 목포MBC, 부산MBC, 안동MBC, 울산MBC, 원주MBC, 여수MBC, 전주MBC, 제주MBC, 포항MBC

(케이티이엔에이 컨소시엄) 감정이 표현되는 우리나라 인물, 우리나라 배경, 예능 자막 등을 생성하는 AI의 학습용 데이터를 구축한다. 예능, 시사·교양, 다큐 등 10,000시간의 원본 영상을 활용하여, 3,600시간의 영상데이터, 50.4만건의 이미지데이터 등 총 5,000시간 분량의 AI 학습용 데이터를 구축한다.

(한국방송공사 컨소시엄) 촬영구도 및 편집 자동화, 사극 및 시대극에서 활용되는 소품(의복, 무기, 유물 등)을 3D로 생성하는 등 방송콘텐츠 제작에 필요한 AI의 학습용 데이터를 구축한다. 보도, 예능, 다큐, 스포츠 등 10,000시간의 원본 영상을 활용하여, 4,500시간의 영상데이터, 1,500건의 3D데이터*, 20만건의 이미지데이터 등 총 5,067시간 분량의 AI 학습용 데이터를 구축한다.

* 길이, 너비, 높이의 3차원 공간 정보를 담아 사극 소품의 입체 형상과 구조를 디지털로 표현

동 사업을 통해 구축된 데이터는 AI 모델 개발 등에 활용된다. 먼저, 각 컨소시엄은 데이터를 활용하여 방송제작 현장에 필요한 특화 AI 개발 및 적용을 지속 추진한다. 아울러, 구축된 데이터는 「독자 AI 파운데이션 모델」 프로젝트에 선정된 정예팀 요청 시 제공되고, AI 허브 내 안심존 등에 연구·교육용 AI 개발을 위해 활용될 수 있도록 약 10~35% 이상 개방될 예정이다.

방송영상 AI 학습용 데이터가 거래 등을 통해 활용도를 높일 수 있도록 협의제도 운영된다. 선정된 컨소시엄, AI·데이터 전문가, 데이터 수요기업 등이 참여하여, 데이터 구축 현황을 공유하고 거래기준 및 체계 등을 논의할 예정이다.

과기정통부 배경훈 장관은 “데이터는 GPU와 함께 AI 모델 개발의 핵심 연료로 그 중요성이 커지고 있으며, 우리나라 인물, 사회, 역사, 자연 등이 풍부하게 담겨 있는 고품질 방송영상은 독자 AI 모델을 위한 최적의 데이터”라고 강조하면서,

“이번 과제를 통해 방송영상이 AI 데이터로 적극 활용되어 방송사가 AI 기술을 다양하게 접목해 제작 효율성을 높임과 동시에 고품질 서비스를 제공하는 기반이 마련되기를 기대한다”고 밝혔다.

담당 부서	과학기술정보통신부 디지털방송정책과	책임자	과 장	어정옥 (044-202-6550)
		담당자	사무관	임창민 (044-202-6556)
관련 기관	한국전파진흥협회 AI미디어빅데이터사업팀	책임자	팀 장	김지수 (02-317-6197)
		담당자	대 리	안철웅 (02-317-6042)

