

보도 일시	2023. 2. 22.(수) 09:00	배포 일시	2023. 2. 22.(수) 09:00
담당 부서	저작권국	책임자	과장 장경근 (044-203-2471)
	저작권정책과	담당자	사무관 임승현 (044-203-2476)
	문화정책관	책임자	과장 정원상 (044-203-2531)
	국어정책과	담당자	연구관 위진 (044-203-2538)
	콘텐츠정책국	책임자	과장 윤양수 (044-203-2411)
	문화산업정책과	담당자	사무관 강기호 (044-203-2415)

문체부, ‘한국어 잘하는 K-챗GPT’ 만들 문화 기반 마련한다

- ▲ 저작권 제도 개선, ▲ AI의 한국어 학습 지원, ▲ 콘텐츠 산업 활용 방안 다룬 3개의 워킹그룹 가동 -

- AI 시대 저작권 제도개선 및 ‘저작권 관점에서의 AI 산출물 활용 가이드(안)’ 마련 목표
- AI 학습에 쓰일 말뭉치 수요 파악하고, AI 모델의 한국어 능력 평가시스템 시범 운영
- 콘텐츠 산업에서의 AI 활용 방안 마련해, 디지털 기술이 뒷받침하는 K-컬처 도약 지원

문화체육관광부(이하 문체부)는 챗GPT로 대표되는 생성형 AI 기술이 가져올 거대한 변화에 필요한 문화적·제도적·산업적 기반을 선제적으로 마련하기 위해 3개의 워킹그룹을 발족한다. ▲ 저작권 제도의 개선, ▲ ‘한국어 잘하는 AI’를 위한 한국어 말뭉치 학습, ▲ 콘텐츠 창작과 산업에서의 AI 활용 등 세 개 분야의 워킹그룹을 통해 전문가와 문체부가 긴밀히 논의하며 현장의 요구를 듣고 정책적인 답을 찾는다.

박보균 장관은 “K-챗GPT의 성공을 앞당기기 위해서는 대기업은 물론 신생기업까지 저작권 문제와 데이터수집 비용을 걱정하지 않고 풍부한 자료를 학습하게 지원해야 한다.”라며, “고품질 한국어 학습데이터인 모두의 말뭉치 구축을 확대하고, AI 산업 발전과 창작자들의 공정한 권리 보장의 균형을 이룰 수 있는 저작권 제도개선 방안을 선제적으로 마련해 나갈 것”이라고 말했다.

AI 학습에 활용되는 저작물의 저작권 처리, AI 산출물의 저작권 법적 지위 문제 등 제도개선 방향 논의

먼저 AI 기술발전에 따른 저작권 제도개선 방향을 논의하기 위해 2월 24일(금) 오후 4시, 한국저작권위원회 서울사무소에서 AI-저작권법 제도개선 워킹그룹(이하 저작권 워킹그룹)을 발족한다. 저작권 워킹그룹에서는 ▲ AI 학습데이터에 사용되는 저작물의 원활한 이용 방안, ▲ AI 산출물의 법적 지위 문제 및 저작권 제도에서의 인정 여부, ▲ AI 기술 활용 시 발생하는 저작권 침해와 이에 대한 책임 규정 방안 등을 논의한다.

문체부는 '21년 7월부터 10월까지 AI 등 신기술 환경에서 저작권 제도의 개선 방향을 논의하는 협의체를 운영한 바 있으나, 최근 챗GPT 논의가 폭발적으로 이루어짐에 따라 관련한 저작권 문제가 부상하고 있어 이에 대해 다시금 선제적으로 논의하고자, 저작권 학계와 법조계, AI 산업계와 창작자 등 현장 전문가 및 이해관계자가 참여하는 워킹그룹을 구성했다.

워킹그룹에서는 산업계와 창작자 등의 현장감 있는 의견을 통해, 저작권 제도와 AI 기술이 융합할 수 있도록 변화된 시대에 맞는 제도의 방향에 대해 폭넓게 논의한다. 특히, 현행 「저작권법」 내에서 활용될 수 있는 (가칭)‘저작권 관점에서의 AI 산출물 활용 가이드(안)’을 마련하는 등 신산업으로서의 AI의 발전을 지원하면서도 인간 창작자들의 권리를 공정하게 보장할 수 있도록 다양한 합의점을 모색한다. 워킹그룹은 2월부터 9월까지 8개월간 운영된다.

말뭉치 지원으로 ‘한국어 잘하는 K-챗GPT’ 개발 뒷받침

챗GPT를 사용해본 많은 사람들은 한국어 대응이 아쉽다는 반응을 보인다. 영어권에서 개발되어 한국어 학습이 영어만큼 충분하지 않아서다. AI의 한국어 학습을 위해서는 한국어 특성을 반영한 고품질의 한국어 말뭉치가 필요하다. 문체부와 국립국어원은 ‘한국어를 잘하는 K-챗GPT’ 개발을 지원하기 위해 고품질 한국어 말뭉치 구축을 확대하고, 워킹그룹 운영을 통해 현장이 필요로 하는 말뭉치를 제공할 계획이다.

2018년부터 구축된 대규모 한국어 말뭉치 37종(약 22억 어절)은, 국립국어원 말뭉치 공개 사이트(모두의 말뭉치 corpus.korean.go.kr)를 통해 오픈소스로 제공되어 한국어 인공지능 개발에 활용되어왔다. 고품질의 말뭉치는 사람이 직접 말뭉치에 한국어 분석 정보(어휘 의미, 구문, 개체명, 감성 등)를 입력하고 검수하는 과정을 거친다. 이런 과정에 비용이 많이 들기 때문에 한국어 말뭉치 제공은 스타트업의 언어자료 구축 비용을 절감하고, 이미 개발된 인공지능 기술을 고도화하는 데 기여해왔다.

* 모두의 말뭉치 이용 건수('20년 8월~'22년 12월) : 누적 18,089건

** 주요 활용 사례 : 삼성전자(삼성전자 상담챗봇서비스), SK텔레콤(언어모델 개발(SKT-AI/KoBART), (주)코어소프트(단어 추천, 자동 문장 생성), 퀵스페이스(Quickspacer, 띄어쓰기 교정 모델), (주) 레이븐어스(아동용 한글 학습 도구 '우왕좌왕 세종대왕') 등

문체부는 앞으로도 인공지능 개발에 꼭 필요하지만 민간에서 구축이 어려운 분야를 빠르게 파악해 그 빈틈을 채울 계획이다. 이를 위해 인공지능 개발 현장 수요를 즉각적으로 파악할 수 있는 워킹그룹을 2월 23일(목), 준비 회의를 거쳐 3월부터 8월까지 운영한다. 워킹그룹은 인공지능과 말뭉치 구축 전문가, 학계 전문가, 문체부와 국립국어원 연구원으로 구성되며, ▲ 인공지능 기술 개발에 필요한 말뭉치 수요를 파악하고, ▲ 2027년까지 한국어 특성을 반영한 고품질 말뭉치 10억 어절 구축 계획을 세울 예정이다.

올해 문체부는 한국형 챗GPT가 빠르게 개발될 수 있도록 25종, 약 1억 2천만 어절의 고품질 한국어 말뭉치를 구축해 배포한다. 또한 인공지능 언어모델이 한국어를 잘 이해하고 생성하는지, 한국의 사회문화 지식을 갖추고 있는지를 검증하는 평가시스템을 시범 운영할 계획이다. 공신력 있는 평가 기준과 평가용 데이터 세트를 구축해 인공지능이 한국어 문장을 제대로 이해했는지 혹은 사진을 보고 생성한 문장이 얼마나 정확한지 등을 검증할 수 있도록 지원한다.

AI가 콘텐츠 창작과 산업 발전에 미치는 영향 분석하고 산업 지원 방안 마련

한편 문체부는 콘텐츠 산업 분야에서의 AI 활용에 능동적으로 대응하기

위해 우리나라 AI의 선구자인 김진형 한국과학기술원 명예교수를 비롯해 AI 관련 학계와 업계 관계자 9명*으로 구성된 ‘콘텐츠 분야 AI TF’를 구성했다.

* 김진형 한국과학기술원 전산학부 명예교수, 이만재 아주대 미디어학부 전 교수, 허문행 안양대 명예교수, 김윤명 경희대교수, 우창현 인천재능대학교 AI컴퓨터정보과 교수, 김태희 영산대학교 게임VR학부 교수, 서유미 딥브레인 AI 리더, 박지은 펄스나인 대표, 조영환 Two Bloc AI 대표

지난 2월 15일(수) 열린 첫 회의에서는 AI의 활용 범위가 문화, 콘텐츠로 빠르게 확대되는 추세에 발맞춰 콘텐츠 산업에 AI가 성공적으로 내재화할 수 있도록 상반기까지 정기적으로 회의를 개최해, ▲ AI를 활용한 신뢰 가능한 콘텐츠 제작환경 조성, ▲ 콘텐츠 기획과 제작자들의 AI 활용 마인드 확산, ▲ 민간주도 시장성장을 위한 AI 콘텐츠 스타트업 육성과 대기업 협업 등을 위한 방안을 본격적으로 논의하기로 했다.

특히 최근 이슈가 되고 있는 챗GPT 등 생성형 AI가 콘텐츠 산업 분야에 미칠 영향과 다양한 이슈를 파악하고, 콘텐츠 산업에 적용사례를 조사 분석해 이에 따른 정책적 지원 방안을 중점적으로 제안하기로 했다. 이 밖에 생성형 AI를 잘 이해하고, AI 콘텐츠 시장을 이끌 수 있는 인재 육성 방안도 모색한다. 문체부는 TF를 통해 청취한 의견과 제언을 종합해 AI를 활용한 콘텐츠산업 분야 지원 대책을 올해 5월까지 마련할 계획이다.

붙임 2023년 국립국어원 한국어 말뭉치 구축 계획



- 구축 종수 및 수량 : 총 25종(약 1.2억 어절 규모)
- 구축 말뭉치 목록

구분	구축 내용	목적
한국어와 사회문화 특성을 분석한 고품질 한국어 말뭉치	<ul style="list-style-type: none"> ○ 회의록 요약 ○ 주제 대화 요약 ○ 감정 분석 ○ 공격성 분석 ○ 함의 관계 분석 등 총 5종 	인공지능의 한국어 및 한국 사회문화 이해 기술 개발 지원
한국어 변화를 지역·연령(시기)·성별을 고려하여 반영한 한국어 균형 말뭉치	<ul style="list-style-type: none"> ○ 2022년 신문기사 1종(총 1억 어절 이상) * 매체별, 월별, 기사 주제별 균형을 고려하여 월별 1,000만 어절 이상 수집 구축 ○ 주제별 2인 이상 다자 대화 수집 1종 (500시간 이상) * 정제된 음성에 대한 발음전사 및 철자전사 병행 	
인공지능 모델의 언어능력 평가를 위한 평가용 말뭉치	<ul style="list-style-type: none"> ○ 언어 이해 과제(감정, 비유리, 함의) 3종 ○ 언어 생성 과제(문장 생성, 이야기 완성) 2종 	인공 지능의 한국어 처리 기술 성능 점검 지원
인공지능의 언어능력 평가 시스템 운영	<ul style="list-style-type: none"> ○ 인공지능의 언어능력 평가시스템 상시 운영 ○ 연간 제출된 인공지능 언어모델을 정성 평가하여 문체부 장관 포상 	
한국어-외국어, 수어, 점자 번역 말뭉치	<ul style="list-style-type: none"> ○ 한국어-외국어 병렬 말뭉치 8종 (총 11백만 어절) * 태국어, 베트남어, 인니어, 캄보디아어, 타갈로그어 (필리핀), 힌디어(인도), 우즈베크어, 러시아어 ○ 한국어-수어 병렬 말뭉치 1종(1백만 어절) ○ 한국어-점자 병렬 말뭉치 1종(1백만 어절) 	번역 기술 개발 지원 다른 언어, 문자 사용자 간 소통 지원 기술
글쓰기 채점 말뭉치	<ul style="list-style-type: none"> ○ 글쓰기 원자료(총 9,000건) ○ 채점 말뭉치 1종(총 5,000건) 	K-로봇 개발
한국어 학습자 말뭉치	<ul style="list-style-type: none"> ○ 한국어 학습자 말뭉치 1종 (총 120만 어절) * 영어권, 일본어권, 베트남어권, 타이어권, 스페인어권, 러시아어권 학습자 등 	한국어 학습 도구 개발