

[AI 사피언스 시대]데이터 '이름표' 붙이기...일자리 창출 '보증수표'

발행일 : 2020.06.09

**AI 개발 핵심 '데이터라벨링' 수작업 검수
전문업체, AI 엔진으로 전처리 공정 병행
정부 'AI학습용 데이터 구축' 예산 2925억 확정
알디프로젝트.마인즈랩 등 외주화 추진도**

인공지능(AI)과 함께 부각되는 분야가 바로 '데이터라벨링'이다. 이름 그대로 데이터에 이름표를 붙이는 작업이다. 소위 AI판 '인형 눈알 붙이기' 작업으로 표현하기도 한다. 데이터라벨링은 AI 학습에 쓰이는 데이터에 '이름표'를 붙이는 작업이다. AI가 학습할 데이터에 정의를 내려준다. AI 개발을 위한 기초이자 필수 작업으로 꼽힌다. 실제로 AI 개발에서 데이터라벨링이 큰 비중을 차지한다. AI가 데이터를 학습하기 위해선 날 것 그대로의 데이터로는 부족하기 때문이다. 데이터라벨링 범주는 이미지, 음성, 텍스트를 포괄한다.

 **관련 통계자료 다운로드**

· 데이터라벨링 작업 과정



데이터라벨링은 노동집약적 산업 성격을 보인다. 예를 들면 이미지에서 나타난 사물 정체는 무엇인지를 작업자가 직접 확인, 입력하는 방식이다. AI는 데이터를 인식할 수 있게 된다. AI 학습용 데이터 하나하나에 이런 식의 수작업이 필요하다. 디지털 시대에 노동집약적 시장이 창출된 셈이다. 다만 노동력만으로 수행하기엔 데이터가 방대하다. 이 때문에 전문업체는 자체 AI 엔진을 통해 데이터 전처리 공정을 병행한다. AI 엔진이 선제적으로 데이터를 가공한다. 선 가공을 거친 데이터는 복수의 작업자가 이를 수차례 검수한다. 데이터에 대한 AI 학습효과까지 측정한다.

데이터라벨링의 일자리 창출 효과는 상당할 것으로 추정된다. AI 고도화에는 방대한 데이터가 요구된다. 데이터라벨링 수요는 큰 폭 성장할 것이란 게 업계 관측이다.

최근 정부가 발표한 '한국판 뉴딜'에서도 데이터라벨링을 통한 공공분야 데이터 일자리 창출 방안이 포함됐다. 과학기술정보통신부는 'AI 학습용 데이터 구축' 추경 예산 2925억원을 확정했다. 과기정통부에 따르면 데이터라벨링 클라우드 소싱 방식으로 10억원당 200명의 일자리를 창출했다.

국내 데이터라벨링 시장은 초기 단계다. AI 개발사가 자체적으로 데이터라벨링 작업을 소화하는 경우가 다수였다. 그러나 기업 AI 개발, 도입이 늘어나는 실정이다. AI의 데이터 처리량 또한 방대해지고 있다. 개발사가 자체적으로 데이터라벨링을 소화하기엔 한계가 있다. 실제 데이터라벨링 외주화를 전문으로 하는 전문 기업도 생겨났다. 국내에서는 알디프로젝트, 마인즈랩을 비롯한 AI 기술기업이 활동하고 있다. 스타트업인 알디프로젝트는 아프리카 가나에 2개의 센터를 운영한다. 영어를 공용어로 사용하면서도 인건비가 낮다는 점을 활용하기 위해서다.

유관 기술기업이 신규 사업으로 데이터라벨링 사업에 진출하는 분위기도 감지된다. 국내외에서 데이터라벨링의 고용 창출 효과를 주목해서다. 중국에서는 AI기업이 농민공과 같은 인력을 대거 흡수, 데이터라벨링에 투입하고 있다. 일자리 부족을 해소하는 한편, 새로운 부가가치를 창출하겠다는 의도다.

이영호기자 youngtiger@etnews.com

[관련 기획] AI 사피엔스 시대

이상직 'AI지식재산 특위 위원장 "AI IP 경쟁력 확보. 지금이 골든타임"
달궈지는 글로벌 AI 반도체 개발 경쟁
데이터 폭증 '커넥티드카 시대'...고성능·저전력 'AI 반도체' 부상
데이터 '이름표' 붙이기...일자리 창출 '보증수표'