

“계란값 묻자 엉뚱한 수치 내놓은 AI… 샘 올트먼도 황당할 듯”

안 국가데이터처장은…

△1968년 충북 제천 출생 △고려대 경제학과 △행시 40회 △통계청 경제동향심의관 △통계청 통계데이터허브국장 △통계청 통계정책국장 △통계청 경인지방통계청장 △통계청 교육원장 △국가데이터처장(현)



국가데이터처 제공

“생성형 인공지능(AI) 챗봇인 ‘챗GPT’ 등에 우리나라 경제성장률을 물어보면 엉뚱한 수치를 내놓고, 달걀 가격 추이 등을 물어도 정확한 답이 나오지 않습니다. 왜 틀린 답을 내놓느냐고 물으면, ‘임의로 만들어 봤다’고 해요. 국가데이터처는 적어도 통계 부문에선 AI의 이런 거짓 정보(할루시네이션)가 없도록 시스템을 구축할 겁니다.”

안형준(57) 국가데이터처장은 지난 12일 정부대전청사에서 이데일리와 만나 ‘AI 대전환을 위한 데이터처의 역할’에 대해 AI가 정확한 통계 정보를 제공할 수 있도록 하는 것이라며 이같이 말했다. 이를 바탕으로 전 국민의 통계 접근성과 활용성을 높일 수 있을 것으로 기대된다.

안 처장은 “세계은행(WB)이 올해 실제 국가별 성장률을 AI에 물어본 결과 챗GPT·제미니·클로드가 모두 다른 값을 제시한 사례도 있다. 일반인의 두뇌를 뛰어넘는다는 ‘똑똑한 AI’가 정작 가장 기본적인 통계를 읽지 못한다는 점은 아이러니하면서도 위험하다”고 했다.

안 처장은 AI 통계 오독(誤讀)의 원인을 “AI가 숫자의 구조를 이해하지 못하기 때문”이라고 진단했다. 생성형 AI는 기본적으로 텍스트의 다음 단어를 예측하는 방식으로 작동한다. 표의 행과 열이 어떤 의미인지, 분자·분모·비율의 구조가 무엇인지 이해하지 못하기 때문에 ‘확률적으로 그럴듯한 단어’를 조합할 뿐 정확한 수치를 찾아서 읽지 못한다는 것이다. 그는 “AI의 통계 할루시네이션은 국가 정책의 품질을 흔들 수 있는 위험 요소”라고 강조했다.

데이터처는 내부 논의 끝에 통계에 활용하는 AI를 ‘더 똑똑하게 만드는’ 접근이 아닌, ‘AI가 틀리지 않도록 데이터를 다시 설계하는’ 방법을 택했다. 수요자 중

안형준 국가데이터처장

소득·물가 등 국민 생활과 밀접 부문
똑똑한 AI보다 ‘안 틀리는’ AI 구축

내년 AI친화형 메타데이터 선보일 것
국민 누구나 쉽게 통계 활용케 하고
정부 정책 판단 속도 향상 뒷받침

심적인 처방이다. 통계 숫자의 구조를 AI가 정확히 이해할 수 있도록 표의 의미·관계·정의·단위 등을 촘촘히 설명한 ‘AI 친화형 메타데이터’를 구축겠다는 것이다. 메타데이터는 다른 데이터의 내용이나 구조, 속성, 이력 등을 표현해 데이터의 검색이나 활용을 돕는 역할을 한다. 이는 이미 국제적으로도 주목받고 있다.

안 처장은 “6월 경제협력개발기구(OECD) 통계청장 회의에서 인공지능통계 혁신과장이 통계 할루시네이션 해결법을 발표했는데, 스위스가 똑같은 방법론을 발표해 서로 놀랐다”며 “전 세계가 같은 문제의식을 갖고 있고, 해결책도 비슷하다는 점을 그때 알게 됐고 더 확신을 갖고 시스템을 구축할 수 있게 됐다”고 말했다.

안 처장은 “AI가 통계를 제대로 읽기 시작하는 순간, AI는 단순한 텍스트 생성기를 넘어 강력한 분석 도구로 변한다”고 강조했다. 지금까지는 숫자를 제대로 집어오지 못해 정책 분야에서 AI 활용이 제한적이었지만, 메타데이터 체계가 갖춰지면 상황이 완전히 달라진다

는 설명이다.

예를 들면 AI가 고용률·물가·임금·소득 자료를 정확히 읽고 상관관계를 분석하거나, 과거 10년 추세를 기반으로 향후 전망치를 산출하는 일이 가능해진다. 안 처장은 “누구나 AI에 ‘향후 5년 청년 고용 전망을 시각화해 달라’거나 ‘지역별 소비 패턴을 비교해 달라’고 요청하면 몇 초 만에 정확한 결과를 얻는 시대가 올 것”이라며 “통계는 더 이상 전문가만의 도구가 아닌 전 국민이 쉽게 활용할 수 있는 국가 경쟁력의 핵심 자산이 될 것”이라고 확신했다.

기존에는 전문 통계 프로그램을 이용하거나 연구기관에서 수많은 노력과 시간을 들여 수행하던 작업들이 AI를 통해 즉시·자동으로 이뤄지는 시대가 열리는 셈이다. 데이터처는 메타데이터 구축이 단순히 새로운 기술 개발을 넘어 통계의 접근성과 활용성을 국민 전체로 확장하는 작업이 될 것이라고 역설했다.

-통계 할루시네이션 해결 방법은 있다.

“AI 친화형 메타데이터’를 구축하는 것이다. 메타데이터는 통계 표를 설명하는 설명서로 이해하면 된다. 표의 행과 열이 뭘 의미하는지, 비율은 어떻게 계산되는지, 어떤 정의를 갖는지표인지 등을 AI가 이해할 수 있게 지도처럼 구조를 만들어 주는 역할을 한다.”

-통계 메타데이터 시스템은 언제쯤 선보일 계획인가.

“내년 중으로 고용동향 통계에 메타데이터를 적용해 AI가 실제로 숫자를 정확히 읽는지 시험적용할 계획이다. 예를 들어 실업률이나 고용률, 비경제활동인구 같은 지표를 AI가 서로의 관계에 맞게 읽어내는지, 표의 구조를 따라정답을 찾아내는지 확인하는 작업이다. 이게 성공

하면 후년에 전산 체계를 본격적으로 구축하고, 물가·임금·소득분배 등 다양한 통계에 확장할 수 있을 것으로 본다.”

-AI가 통계를 읽고 해석하는 시대가 오면 정책에도 큰 변화가 올 것 같다.

“그렇다. 지금은 통계 분석을 하려면 방대한 데이터를 정제하고 모델을 돌리고, 해석하는 절차가 필요하다. 그런데 메타데이터와 AI가 결합하면 이 같은 노력과 시간이 훨씬 줄어든다. 예를 들어 ‘지난 5년간 청년 고용의 추세를 보여달라’고 하면 AI가 정확한 숫자를 기반으로 결과를 뽑아줄 것이다. 나중에는 해석도 가능하리라고 본다. 이렇게 되면 정책 판단 속도도 훨씬 빨라지고 정책함도 높아질 것으로 기대하고 있다.”

-법적 정비가 필요한 부분은 없다.

“현행 통계법만으로는 부처별로 다 흩어져 있는 지금의 데이터 환경을 포괄할 수 없다. 국가데이터기본법이 생기면, 국가데이터의 종류와 품질, 민간 데이터는 어떻게 관리할 지 전체 그림이 그려진다. 그리고 신설되는 국가데이터위원회(국무총리 주재)를 통해 부처 간 정책 조율도 할 수 있게 된다. AI 시대에는 데이터가 정책의 기반이 되기 때문에 이러한 제도 정비가 필수이다.”

-마지막으로 데이터처 승격 이후 가장 중요하게 생각하는 목표가 있다면.

“통계청에서 데이터처로 승격했지만, 국민 인식은 아직 청에 머물러 있는 것만 같다. 그래서 ‘아, 데이터처가 뭔가 달라졌다’고 국민이 직접 느끼는 성과가 필요하다. AI 메타데이터는 그 시작이다. 통계에 AI를 접목해 국민이 체감하는 변화를 이끄는 게 가장 중요하다. 데이터처의 위상도 그런 성과를 통해 자리 잡을 것 같다.” 대전=강신우·하상렬 기자 yeswhy@