

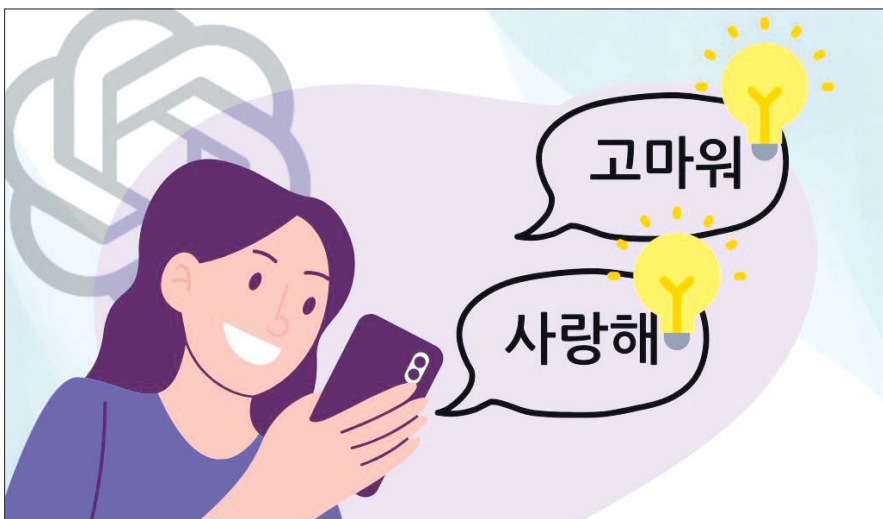
AI에 무심코 건넨 인사… 수천만 달러 전기요금 발생시킨다

입력 텍스트, 문자열 단위로 토큰화
대화가 길어질수록 연산량 급증
짧은 인사에도 수천번의 연산 실행
평균 이용량, 전구 11시간 켜는 수준

인공지능(AI) 모델에 무심코 건넨 ‘고마워’ 한마디가 전구를 켤 정도의 전력을 소모하는 것으로 나타났다.

AI 모델에게 사람들이 무심코 건넨 공손한 인사말이 막대한 비용을 발생시킨다는 증언이 나왔다. 이는 AI 모델이 작동할 때, 사람이 입력한 문자와 모델이 생성한 응답을 모두 ‘토큰(token)’ 단위로 처리하며, 이 과정에서 상당한 연산 자원이 요구되기 때문이다. 토큰은 AI 모델이 이해하거나 생성할 수 있는 최소한의 문자다.

22일 <메트로경제신문> 취재를 종합해보면 AI 모델이 이용자의 요구에 따라 답변을 출력하기 위한 과정에서 고난이도·대용량 연산이 비용부담을 더 가중시키는 것으로 확인됐다. ‘고마워’



챗GPT 이용 중 ‘고마워’와 같은 공손한 인사말이 막대한 전력량을 소모한다는 증언이 나왔다. /김서현 기자

‘미안해’와 같은 공손한 표현을 덧붙이는 것도 AI 모델 특유의 알고리즘으로 인해 비용을 추가로 발생시켰다.

최근 샘 올트먼 오픈AI CEO는 SNS 플랫폼 X(구 트위터)에서 한 이용자가 “사람들이 챗GPT에 ‘제발’ ‘고맙습니다’를 반복할 때마다 전기 비용이 얼마나 들었는지 궁금하다”고 묻자 “수천만

달러의 전기요금을 발생 시켰다”고 답했다.

사람들의 공손한 표현만으로도 수천만 달러의 비용이 소모되는 이유는 생성형 AI가 텍스트를 분해해서 이해하고 연산, 출력하는 과정의 특수성 때문이다.

GPT 모델은 입력 텍스트를 자주 등

장하는 문자열 단위로 분해해 토큰화하며, 한국어의 경우 보통 한 음절 또는 짧은 단어 하나가 1토큰 정도로 처리한다. 영어는 4~5자당 1토큰 정도가 소모된다. 이에 따라 ‘사과’는 2토큰, ‘사랑해’는 3토큰이 된다. 여기에 출력시에도 토큰을 하나씩 생성하는 방식이기 때문에, 대화가 길어질수록 연산량이 기하급수적으로 늘어난다. 샘 올트먼의 말대로 ‘고마워’ 한 마디도 수천 번의 연산이 이뤄질 수밖에 없는 셈이다.

이러한 연산 과정에서 탄소 발생 논란도 일어난다. GPT-4 등의 대형 모델은 연산을 위해 수천 개의 고성능 GPU가 연결된 대규모 데이터센터를 이용하는데, 토큰 수가 많아질수록 해당 GPU 자원의 점유율과 동작 시간도 함께 증가하고, 필요로 하는 전력량 또한 비례해 늘어난다.

대표적인 AI 모델인 챗GPT를 기준으로 할 때 1토큰은 단순계산 기준 입출력 난이도에 따라 최소 0.000001kWh(1μWh)에서 최대 0.000005kWh(5μWh)

를 소모한다. 이는 전구를 약 0.4초에서 2초 가량 켤 수 있는 수준이다. 즉, ‘사과’는 9W 전구를 0.4초, ‘사랑해’는 4초 가량 켤 수 있게 된다.

챗GPT를 다양한 고난이도 연산에 사용하는 이용자들(개발자 등)의 평균 이용량인 1일 1만~2만 토큰은 0.01~0.1kWh로, 9W 전구를 약 1시간 6분에서 11시간 6분 켤 수 있는 수준에 달한다. 이는 생수 40병 분량의 물을 1℃ 데우는 데 드는 전력량이다.

토큰에 따른 비용이 막대하게 늘어난다. GPT-4 등의 대형 모델은 상업적으로 운영하는 기업은 토큰 사용량에 따라 정해진 AI 모델 이용 비용을 줄이기 위해 토큰나이저 최적화나 출력 제한 기술도 고도화하는 추세다.

IT 업계 관계자는 “최근 AI 기술은 전력 소모량을 줄이고 데이터 입출력을 효율화 하는 데에 중점을 두고 있다”며 “사람들이 AI에도 ‘고마워’ ‘사랑해’ 등 인사를 하는 현상은 무척 흥미롭다”고 말했다. /김서현 기자 seoh@metroseoul.co.kr

“AI, 이용자가 제공하지 않은 정보도 학습” 카카오엔터, AI로 숏폼 제작 시간·비용 절감

SPRI “LLM 기반 AI가 발전할수록
개인 식별 가능성 ↑… 개인정보 침해”
관련 법적근거·정보 은폐기술 등 필요

생성형 AI(인공지능)의 학습에 사용자가 제공하지 않은 데이터가 이용되면서 정보주체의 자기정보통제권이 약화되고 있다는 진단이 나왔다. 정보주체의 권리 보장을 위한 새로운 법적·기술적 대응이 필요하다는 주장이 제기된다.

22일 소프트웨어정책연구소(SPRI)가 이달 펴낸 ‘소프트웨어 중시사회’ 보고서에 따르면, 거대언어모델(LLM)을 기반으로 한 생성형 AI가 발전할수록 개인 식별 가능성이 증가해 개인정보 침해 위험이 커진다.

보고서는 “현재 기업들은 개인이 제공하지 않은 정보까지 수집해 개인 식별 가능성을 고도화하는 중”이라며 “LLM 개발회사는 데이터 기업이나 데이터 결합 기관 등에서 획득한 정보뿐만 아니라 온라인에서 확보한 다양한 데이터까지 무차별적으로 수집해 새로운 개인정보를 생성해내고 있다”고 분석했다.

AI 기업이 정보주체보다 개인을 더 잘 아는 시대가 도래했다는 것이다. 개인을 이해하려는 욕구는 조직의 생존과

직결된 문제다. 기업은 개인에게 최적화된 맞춤형 서비스 제공으로 이익을 극대화하기 위해 사용자 정보를 무차별적으로 수집하고 있다.

장재영 한국인터넷진흥원 개인정보 안전활용본부 연구위원은 보고서에서 “생성형 AI는 이용자가 제공하지 않은 정보도 학습하고, 이용자가 모르는 이용자에 관한 정보까지 생성하고 있다”며 “이러한 정보의 존재 자체를 이용자가 직관적으로 인식하지 못할 가능성이 크기 때문에 개인정보 자기통제권의 실질적 행사가 본질적으로 제한될 위험이 높다”고 분석했다.

예컨대 AI가 학습한 데이터에서 특정인의 직업·소득 수준·정치적 성향 등을 추론해 제공할 경우 사용자는 이런 정보가 생성됐다는 사실조차 인지하지 못하게 된다. 또 당사자가 자신의 정보를 제공한 적이 없다면 사용자에 관한 데이터임에도 AI가 생성한 정보에 대해 소유권을 주장하기 어려워진다.

보고서는 개인정보 자기통제권이 약화되면 사용자의 주체성이 침해되고 소비자와 기업 간 데이터 처리 관련 정보 격차가 심화되며 자동화된 의사결정으로 인한 이용자 차별 등의 문제가 발생

할 수 있다고 지적했다.

정보주체의 개인정보 자기통제권을 보호하기 위한 방안으로 보고서는 ▲개인정보 생성 관련 법적 근거 손질 ▲학습 데이터에 개인정보 최소화 ▲정보주체 권리 보장 기술 적용을 제안했다.

장재영 연구위원은 “공개된 정보의 활용과 사용자가 제공하지 않은 정보를 이용한 개인정보 생성에 대한 규율 근거를 정비해야 한다”면서 “또 합성 데이터, 차등 프라이버시 등 개인정보를 은폐할 수 있는 기술을 사용하면 개인정보 침해 예방에 효과가 있을 것”이라고 제안했다.

합성 데이터는 컴퓨터 시뮬레이션이나 알고리즘을 이용해 원본 데이터의 구조적·통계적 속성을 재현한 데이터이고, 차등 프라이버시는 데이터에 노이즈를 추가해 개별 데이터의 기여도를 숨기면서 전체 데이터 패턴을 유지하는 기술이다.

장 연구위원은 “LLM은 한번 학습된 정보를 쉽게 삭제할 수 없어 개인정보 삭제 요구권을 효과적으로 행사하기 어렵다”면서 “언러닝 기술을 적용해 특정 개인정보를 포함하는 학습 데이터를 제거하는 방법도 개발할 필요가 있다”고 조언했다. /김현정 기자 hjk1@

‘헬릭스 숏츠’ 공식 론칭

카카오엔터테인먼트가 AI(인공지능) 기술로 숏폼 제작에 들어가는 비용과 시간을 획기적으로 줄였다.

카카오엔터는 22일 ‘헬릭스 숏츠’를 공식 론칭했다. 헬릭스 숏츠는 웹툰을 짧은 영상으로 자동 제작하는 AI 기술 기반 서비스다. 기존에는 숏츠 한 편 제작에 약 3주의 기간과 200만원 가량의 비용이 투입됐는데, 헬릭스 숏츠로 3시간여만에 6만원 수준으로 제작 가능해졌다.

사용자들은 작품 줄거리, 관전 포인트 등이 담긴 영상을 카카오페이지 앱 내 화면에서 만나볼 수 있다.

헬릭스 숏츠에는 텍스트 중심의 기존 AI와 달리 이미지와 감정까지 정교하게 분석하는 고도화된 인공지능 기술이 적용됐다. 웹툰의 ▲대사 ▲표정 ▲컷

구성 ▲말풍선 등을 심층 분석해 줄거리를 요약하고 내레이션을 만들어 낸다. 웹툰 해설 문장은 문자 음성 변환 기술로 구현되며, AI가 스토리 감전에 따라 컷과 이미지 효과(줌 인, 좌우 삽입 등), 배경 음악을 조합해 몰입감 높은 숏츠를 완성한다.

카카오엔터는 지난 2023년 헬릭스 숏츠 기술 개발을 시작해 이듬해 8월 첫 테스트를 실시하고, 같은해 9월 관련 기술에 대한 특허 출원을 냈다. 이어 작년 11월부터 헬릭스 숏츠를 본격적으로 도입했고, 올 2월을 기점으로 최근 30일 내 웹툰 열람 이력이 있는 전체 사용자에게 헬릭스 숏츠 적용을 완료했다. 현재 카카오페이지에 노출되는 전체 숏츠 가운데 약 40%가 헬릭스 숏츠로 만들어진 영상이다.

카카오엔터는 헬릭스 숏츠비율을 점차 늘려나갈 예정이다. /김현정 기자

‘CJ SW 창의캠프’, 국무총리표창 수상

CJ올리브네트웍스 사회공헌 프로그램

CJ올리브네트웍스가 사회공헌 프로그램 ‘CJ SW 창의캠프’로 국무총리표창을 수상했다.

CJ올리브네트웍스는 과학기술정보통신부와 방송통신위원회가 공동으로 주최한 ‘2025 과학·정보통신의 날’ 기념식에서 정보통신 유공 정부 포상으로 국무총리표창을 받았다고 22일 밝혔다. 이는 정보방송통신 분야의 발전에 기여한 개인 또는 단체의 공로를 인정해 수여하는 상이다.

CJ SW 창의캠프는 회사가 소외계층 청소년의 디지털 교육 격차 해소를 위해 2015년부터 운영해온 프로그램이다. IT(정보기술) 전문성을 바탕으로 웹



유인상 CJ올리브네트웍스 대표가 ‘2025 과학·정보통신의 날’ 기념식에서 정보통신 유공 정부 포상으로 국무총리표창을 받고 있다. /CJ올리브네트웍스

고 재미있게 소프트웨어를 경험하며 논리적 사고력을 키울 수 있도록 지원한다. 그간 CJ SW 창의캠프에서 교육받은 아동·청소년은 1만4700여명에 달한다. /김현정 기자

LG U+, AI 기술로 환자 문의 실시간 대응

바른세상병원에 AICC 도입

LG유플러스가 경기도 성남시 분당구 소재 관절·척추 특화 병원인 바른세상병원에 인공지능 기반 상황인지 응대형 의료상담 컨시어지(AI 컨택센터, AICC) 서비스를 본격 운영한다고 22일 밝혔다.

바른세상병원에 도입된 AICC는 AI

콜봇을 기반으로 환자의 문의에 실시간 대응하는 서비스다. 병원 대표번호로 걸려오는 전화 중 예약 확인·변경·취소, 병원 위치 안내, 제증명 발급 등 단순 반복적인 상담은 AI가 자동 처리하고, 상담 직원은 보다 복잡하고 전문성이 필요한 상담에 집중할 수 있도록 설계됐다.

LG유플러스는 전국 200개 이상의

병원 컨택센터에 통신과 상담 시스템을 공급한 경험을 바탕으로 이번 AICC를 제작했다. 전문 협력사인 나스카랩과 함께 OCS(처방전달시스템), EMR(전자 의무기록) 등 병원 고유 시스템과 연동 가능하도록 했다.

특히 LG AI연구원이 자체 개발한 초거대 AI 모델 ‘엑사원(EXAONE)’을 기반으로 한 STT/TA(음성인식·텍스트분석) 기술을 활용해 자연어 이해도와 상담 응답 정확도를 높였다.

/김서현 기자