AI 뉴스와 용어
월·수·금 오전 7시

METAL
First, AI Magazine

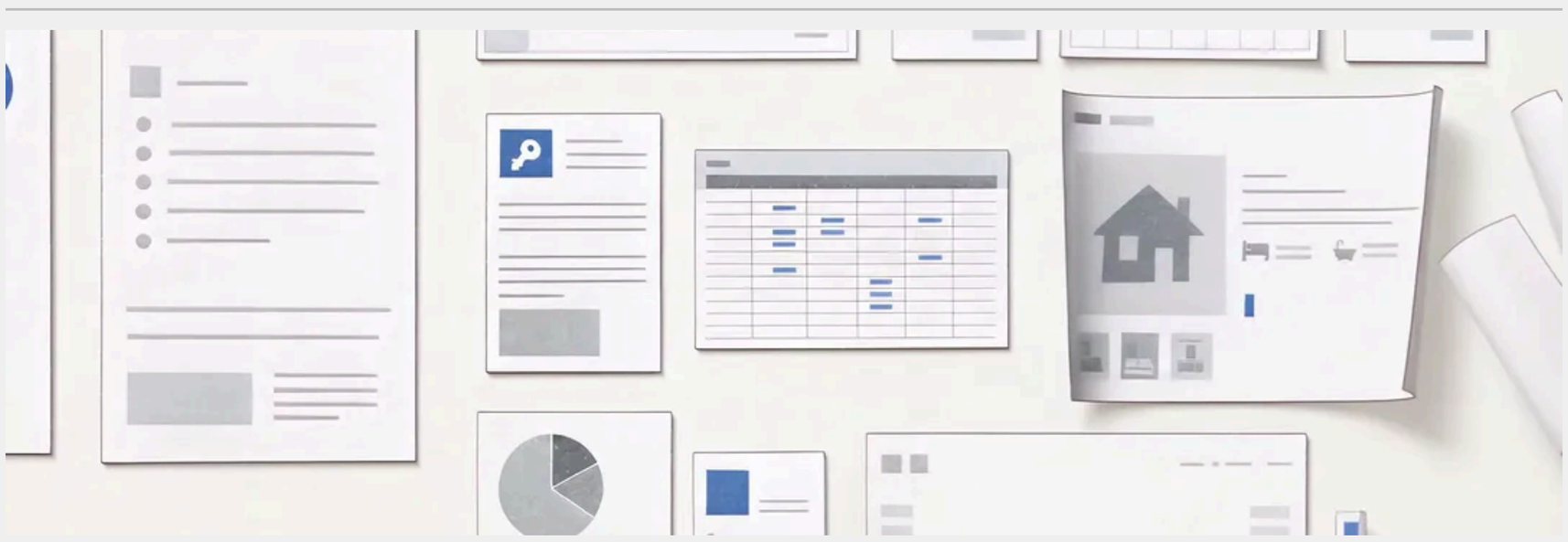
무료 · 언제든 수신거부
metallab.ai

제 43 호 | 2026년 10월 13일 화요일 | 오늘의 AI

AI

메타, 프론티어 모델 Muse Wonder 공개

메타가 첫 프론티어 모델 Muse Wonder와 Muse Wonder Full을 공개했다. 자체 측정한 개인 초지능 지수와 법률·금융 에이전트 과제에서 Claude Opus 5.5와 GPT-6 Astra를 앞섰다고 밝혔다.



AI · 메타, 프론티어 모델 Muse Wonder 공개

메타가 10월 12일 자사 첫 프론티어 모델인 Muse Wonder와 Muse Wonder Full을 공개했다. 메타는 발표문에서 두 모델을 지금까지 내놓은 모델 가운데 가장 뛰어난 모델이라고 소개하고, 개인 업무와 전문 업무를 대신 처리하는 에이전트 과제에서 최고 수준 성능을 낸다고 밝혔다. Muse Wonder는 공개 당일부터 Meta Model API와 코딩 도구 Muse Code에서 쓸 수 있고, Muse Wonder Full은 Muse Code 구독자에게만 열렸다. 두 모델 모두 컨텍스트 창이 100만 토큰이다.

이번 공개는 메타 AI 조직 개편의 첫 성적표다. 메타는 14개월 전 AI 사업을 갈아엎기 시작해 기반을 다시 세우는 단계에서 프론티어를 미는 단계로 왔다고 설명했다. 보도에 따르면 이번 모델은 스케일AI 창업자 알렉산더 왕이 지난해 8월 메타의 AI 조직을 재편한 뒤 처음 나온 프론티어급 모델이고, 메타는 그사이 경량 모델 Muse Spark와 개인 에이전트 Muse만 내놓았다. 메타는 연구 블로그에서 "두 모델은 우리 스케일링 사다리의 다음 단계이자 개인 초지능을 향한 큰 도약"이라고 썼다. 보도에 따르면 학습은 올해 가동한 오하이오주 중부의 프로메테우스 데이터센터에서 기가와트 규모로 이뤄졌다.

메타가 겨냥한 일은 코딩만이 아니다. 발표문은 보험 청구 처리, 임대 계약서 이해, 개인 재무 정리 같은 일을 예로 들며 두 모델이 법률·금융·의료 같은 전문 영역과 프론티어급 코딩을 함께 다룬다고 밝혔다. 메타 개발자 페이지는 Muse Wonder가 긴 작업을 계획하고 일부를 하위 에이전트에게 나눠 맡기며, 도구를 병렬로 부르고 오류에서 스스로 회복해 몇 시간씩 이어지는 다단계 작업을 끊김 없이 수행한다고 설명한다. 브라우저와 데스크톱, 모바일 앱 화면을 보고 직접 조작하는 컴퓨터 사용 기능도 들어갔다. 보도에 따르면 두 모델은 글과 이미지, 영상, 음성을 모두 입력으로 받는다.



중국, 남는 재생에너지로 AI 데이터센터 확충

메탈이 확인한 메타 개발자 페이지의 벤치마크 표는 두 모델을 Claude Fable 5.1, GPT-6 Astra, Claude Opus 5.5와 14개 항목에서 나란히 비교했다. 표의 점수는 메타가 공개한 값이다. 개인·지식 업무를 묶은 메타 내부 지표인 개인 초지능 지수에서 Muse Wonder Full은 59.1점으로 Claude Opus 5.5(55.2점)와 GPT-6 Astra(54.5점), Claude Fable 5.1(49.9점)을 앞섰다. 법률 에이전트 과제인 Harvey LAB-AA에서는 13.1점으로 비교 모델 가운데 가장 높았고 Claude Opus 5.5는 4.2점이었다. 금융 분석 과제 Vals Finance Agent V2에서 59.1점, 업무 자동화 과제 AutomationBench-AA에서 45.4점으로 역시 1위였다.

모든 항목을 이긴 것은 아니다. 같은 표에서 Claude Opus 5.5는 의료 질의 HealthBench Pro(65.6점), 전문 도구 사용 JobBench(75.4점), 터미널 코딩 Terminal-Bench 4.0-AA(59.6점), 처음부터 프로그램을 짜는 ProgramBench(70.5점)에서 가장 높았고, 외부 기관 지표인 Artificial Analysis 지수에서도 58점으로 Muse Wonder Full(56점)보다 위였다. 코딩에서 메타가 내세운 것은 긴 호흡의 작업이다. Muse Wonder Full은 장기 코딩 과제 DeepSWE 1.1에서 79.0점, 초장기 엔지니어링 과제 SWE-Marathon에서 57.5점으로 비교 모델을 모두 앞섰고, 메타는 이 결과가 독립 제3자 측정에서도 확인됐다고 밝혔다.

을 재는 MemoryBench에서 Muse Wonder는 85.6점으로 표 안의 다섯 모델 가운데 가장 높았다. 카비야 코파라푸 메타 연구원은 발표 영상에서 "사용자에게 중요한 것을 며칠, 몇 주, 몇 달, 그 너머까지 모델이 기억하도록 엄청난 공을 들였다"고 말했다.

외부 측정도 비슷한 그림이다. 보도에 따르면 Artificial Analysis 지능 지수에서 Muse Wonder Full은 56점, Muse Wonder는 52점으로 전작 Muse Spark 1.3의 48점보다 크게 올랐고, Terminal-Bench 4.0 점수는 33%에서 57%로 뛰었다. 수학 증명 비공개 벤치마크 ProofBench에서는 100문제 중 94문제를 풀어 전작의 58문제를 크게 넘었다. 사실 지식은 약점으로 꼽혔다. 지식 시험 AA-Omniscience에서 Muse Wonder Full은 30점으로 Claude Opus 5.5의 46점, GPT-6 Astra의 43점, Gemini 4 Argon의 42점에 못 미쳤다. 대신 환각 비율은 전작의 33%에서 23%로 내려갔다. 메탈은 메타가 Muse Spark 1.3을 내놓으며 코딩 벤치마크를 앞세운 소식을 보도한 바 있으며, 이번에는 그 경량 모델 위에 대형 모델 층을 얹은 셈이다.

안전 장치도 함께 내놓았다. 메타는 개인 에이전트가 메시지와 일정, 문서를 넘나들며 일하는 만큼 두 모델이 허락 없이 개인정보를 공유하지 않고 실제로 해로운 행위를 돕는 요청을 거절하도록 훈련했다고 밝혔다. 보안 팀이 자기 코드의 결함을 찾는 작업은 그대로 할 수 있게 남겨 뒀다. 메타는 가장 강력한 모델을 메타 초지능 스케일링 프레임워크의 기준에 맞춰 훈련·평가·출시하며, 추가 안전성 평가를 담은 전체 보고서는 곧 공개한다고 덧붙였다.

다음 무대는 개인 에이전트 Muse다. 메타는 Muse Wonder가 곧 Muse의 더 많은 기능을 구동하며 구독자부터 적용된다고 밝혔다. 메탈은 메타가 보안 가상머신 위에서 도는 개인 에이전트 Muse를 내놓은 소식을 보도한 바

오늘 지면

1 AI
메타, 프론티어 모델 Muse Wonder 공개

2 비즈니스
마이크로소프트, AI 추론 칩 Maia 300 공개

3 비즈니스
밍 루, AI 시대 혁신의 역설 진단

4 기후
미국 에너지부, FERC의 PJM 개혁 요구 지지

5 기후
중국, 남는 재생에너지로 AI 데이터센터 확충

이 지면은 2026년 10월 13일 화요일 아침 구독자 메일함으로 나간 레터를 신문 판형으로 다시 조판한 것입니다. 기사 전문은 metallab.ai 에서 읽을 수 있습니다.

기후



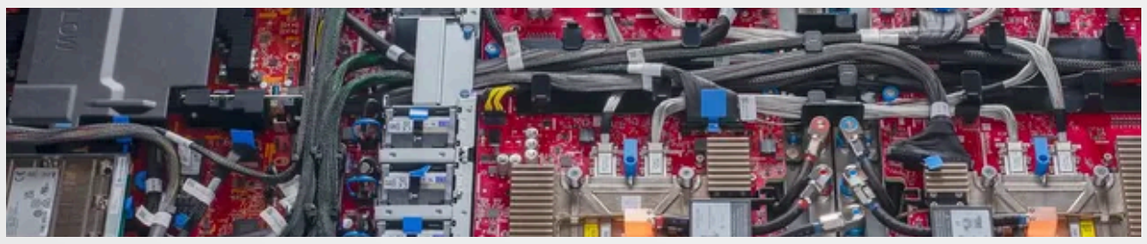
미국 에너지부, FERC의 PJM 개혁 요구 지지

미국 에너지부가 연방에너지규제위원회(FERC)의 PJM 개혁 요구를 지지하고 나섰다. 에너지부는 10월 9일 발표문에서 FERC에 개입 통지서와 입장서를 냈다고 밝히고, 미국 최대 전력망 운영기관 PJM이 데이터센터 같은 대형 전력 수요처에 새 발전소와 기반시설 비용을 물리도록 규칙을 고쳐야 한다고 요구했다. 보도에 따르면 에너지부가 FERC에 입장서를 낸 것은 적어도 5년 만에 처음이다. 쟁점은 PJM이 추진하는 일회성 신규 용량 조달, 이른

▶ 2면에 계속

비즈니스

마이크로소프트, AI 추론 칩 Maia 300 공개



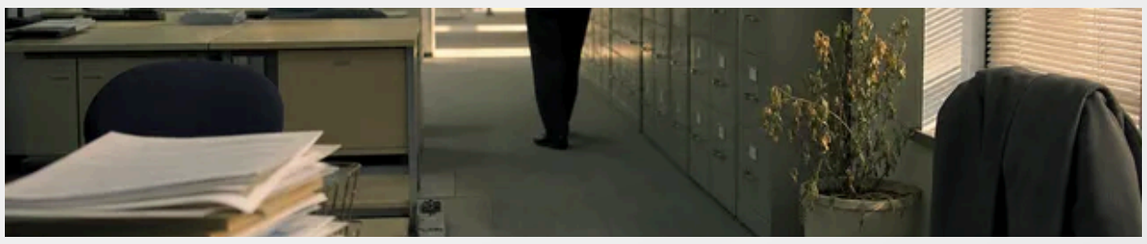
마이크로소프트가 10월 12일 3세대 자체 AI 추론 칩 Maia 300을 공개하고, 2027년부터 Maia 제품군을 수 기가와트 규모로 배치해 자사 제품과 서비스의 주력 추론 시스템으로 삼겠다고 밝혔다. 회사는 Maia 300이 선도적인 GPU 기반 시스템보다 메가와트당 40% 많은 토큰을 낼 것으로 내다봤다. 가속기 96개를 다 채운 랙 구성에서는 랙 단위 토큰 처리량이 전 세대 Maia 200의 최대 7배다.

발표의 전제는 이미 돌고 있는 2세대다. 마이크로소프트에 따르면 Maia 200은 현재 Microsoft Copilot 운영 워크로드를 처리하며 사내 MAI 팀의 차세대 모델 개발을 지원하고 있고, 연말까지 GitHub Copilot과 Microsoft Foundry가

▶ 2면에 계속

비즈니스

밍 루, AI 시대 혁신의 역설 진단



기업 임원과 사모펀드 파트너로 40년 넘게 일한 밍 루가 10월 11일 미국 경제지에 실은 기고문에서 AI 시대 혁신의 역설을 진단했다. 그는 지난 50년 동안 디지털 혁신이 소비자 보급에서는 눈부신 속도를 냈지만 생산성 증가율은 계속 떨어졌다며 이 현상을 혁신의 역설이라고 불렀다. 그의 정의는 한 문장이다. "우리가 빠르게 혁신할수록 성장은 느려진다."

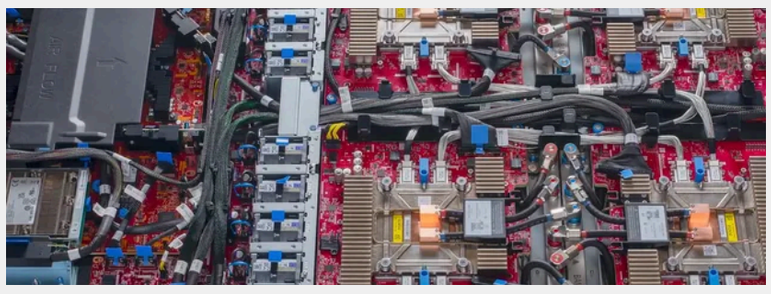
출발점은 경영진의 체감이다. 루는 맥킨지의 2026년 AI 현황 조사를 인용해 세계 대기업 최고재무책임자(CFO)의 88%가 AI가 실적을 바꾸고 있느냐는 물음에 별로 그렇지 않다고 답했다고 적었다. 같은 조사에서 기업의 88%가 최소 한 개 부서에서 AI를 쓰지만 이자·세전이익(EBIT)에 5% 넘는

▶ 2면에 계속

◀ 1면에서 계속

비즈니스

마이크로소프트, AI 추론 칩 Maia 300 공개



1 마이크로소프트는 10월 12일 Maia 300을 공개하고 2027년부터 Maia 제품군을 수 기가와트 규모로 배치해 자사 제품·서비스의 주력 추론 시스템으로 삼겠다고 밝혔다.

2 Maia 300은 HBM4 스택 8개로 18.0TB/s 대역폭을 내는 950와트 칩이며, 가속기 96개를 담은 167킬로와트 랙의 처리량은 Maia 200의 최대 7배다.

3 이미 운영 중인 Maia 200은 Copilot 워크로드를 처리하며 Teams·Copilot·Cowork 기능에서 선도 GPU 인프라보다 와트당 최대 1.7배 많은 토큰을 낸다.

Maia 300의 설계 목표는 추론, 그중에서도 긴 문맥과 메모리를 많이 쓰는 디코드 단계다. 칩 하나에 HBM4 메모리 스택 8개를 얹어 최대 메모리 대역폭 18.0TB/s를 내고, 전력 설계 기준(TDP)은 950와트다. 연산과 메모리 컨트롤러, 입출력 기능을 따로 떼어 낸 멀티 칩렛 구조를 처음 썼고 연산 칩렛은 TSMC의 N2P 공정으로 만든다. 칩 안의 SRAM을 소프트웨어가 관리하는 임시 저장 공간으로 써서 HBM의 부담을 덜고, 개선된 MXFP4 같은 저정밀 형식으로 연산과 KV 캐시 저장에 드는 메모리를 줄였다. 사우라브 디게 애저 하드웨어 시스템 기획·아키텍처 부사장은 이 변화를 멀티 칩렛과 HBM4, 새 정밀도 형식, 차세대 스케일업 네트워크, 더 크고 전력 효율이 높은 랙까지 다섯 갈래로 정리했다.

랙 단위 숫자가 처리량 7배의 근거다. Maia 300 랙 하나는 가속기 96개를 167킬로와트 전력 안에 묶고 HBM 27.6TB와 합산 HBM 대역폭 1.7PB/s를 하나의 고속 스케일업 영역으로 쓴다. 가속기 사이 연결은 텐서 병렬 처리에 맞췄고, 이더넷 기반의 자체 AI 전송 방식으로 랙 하나에서 데이터센터 규모까지 같은 토폴로지로 넓힌다. 메탈이 확인한 마이크로소프트 공개 사양표에 따르면 칩 한 개의 HBM 용량은 288GB, 양방향 스케일업 대역폭은 3.2TB/s이며 스케일업 영역 하나에 Maia 가속기 약 1만 2000개가 묶인다.

회사가 숫자를 와트와 메가와트로 내세우는 데는 이유가 있다. 라니 보카 애저 하드웨어 시스템·인프라 사장은 "AI의 경제성은 최대 연산 성능이 아니라 시스템 전체가 얼마나 효율적으로 대규모 토큰을 만들어 내느냐로 점점 더 결정된다"고 말했다. 마이크로소프트는 대형 추론 시스템의 병목이 연산 능력이 아니라 데이터 이동이라고 보고, 소프트웨어 정의 데이터플로 구조와 AI 보조 커널 최적화로 데이터를 덜 옮기는 쪽을 택했다. 칩 분석가 오스틴 라이언스는 회사가 배포한 분석가 논평 자료에서 에이전트가 계획하고 도구를 부르고 다시 추론하는 동안 문맥이 유지돼야 하기 때문에 추론이 점점 데이터 이동 문제가 된다고 짚었다.

◀ 1면에서 계속

비즈니스

밍 루, AI 시대 혁신의 역설 진단

1 밍 루는 10월 11일 기고문에서 지난 50년 디지털 혁신이 소비자 보급은 빨랐지만 생산성 증가율은 떨어뜨렸다며 이를 혁신의 역설이라고 불렀다.

2 그가 인용한 맥킨지 2026년 조사에서 기업의 88%가 AI를 쓰지만 EBIT에 5% 넘는 효과를 낸 고성과 기업은 6%로 2025년과 같았다.

3 라이터의 2026년 조사에서도 경영진 79%가 도입 문제를, 54%가 AI 도입이 회사를 갈라놓는다고 답했고, 루는 깊이와 넓이 있는 조직 재설계를 해법으로 들었다.

루는 이 격차를 이중 K자 확산으로 설명한다. 첫 번째 K자는 소비자와 기업 사이에 있다. 스마트폰은 7년 만에 미국인의 80%에게 퍼졌지만 전사적자원관리(ERP) 시스템이 미국 기업의 57%에 닿기까지는 33년이 걸렸다. 소비자 AI 사용은 챗GPT 출시 3년 만에 미국 성인의 53%로 늘었지만 실제 업무에 배치된 기업 AI는 10%다. 두 번째 K자는 기업 안에서 벌어진다. 루는 사용 기업 88%와 고성과 기업 6% 사이의 82%포인트 격차를 기술 부족이 아니라 조직 전환의 부족으로 봤다. 고성과 기업은 업무 흐름을 처음부터 끝까지 다시 설계했을 가능성이 세 배 높고, 그렇게 한 기업은 전체 도입 기업의 21%뿐이다.

주식시장도 같은 모양이다. 루에 따르면 S&P500 안의 AI 기업 36곳이 지수 시가총액의 45%를 차지한다. 최근 3년 S&P500 수익률은 76%였지만 AI 종목을 빼면 32%로 내려간다. 메탈은 EY 수석이코노미스트가 AI 생산성 붐이 승자독식 경제를 만든다고 경고했다고 보도한 바 있으며, 루의 숫자는 그 쏠림이 이미 시장 가격에 들어와 있음을 보여 준다.

비교 대상은 20세기 초의 전기화다. 전기와 통신은 소비자와 기업에 비슷한 속도와 비슷한 깊이로 퍼졌다. 전기 모터는 공장을 경직된 배치에서 연속 흐름 생산으로 바꿨고 전화는 조정 비용을 낮췄다. 제너럴모터스의 사업부제와 테일러의 과학적 관리 같은 조직 재설계가 뒤따랐다. 디지털 기술은 서비스가 GDP를 지배하는 경제로 들어갔다. 경제학자 윌리엄 보몰이 사람 사이의 상호작용에 기대는 탓에 생산성이 잘 오르지 않는다고 지목한 의료·교육·공공행정·건설 부문은 선진국 GDP의 약 50%를 차지한다.

◀ 1면에서 계속

기후

미국 에너지부, FERC의 PJM 개혁 요구 지지



1 미국 에너지부는 10월 9일 FERC에 입장서를 냈다고 밝히고 PJM이 대형 전력 수요처에 신규 발전·기반시설 비용을 물리는 개혁을 서둘러야 한다고 요구했다.

2 FERC는 9월 29일 PJM의 신뢰도 백스톱 조달안을 받아들이되 5개월 정지하고 비용 배분·송전사업자 탈퇴 규정·담보 요건을 심리에 부쳤다.

3 PJM 용량 경매는 2027/2028년 6,623메가와트가 모자랐고 부하 증가 대부분이 데이터센터 몫이어서, 비용을 누가 내느냐가 동부 전기요금을 가른다.

배경에는 PJM의 용량 부족이 있다. 메탈이 확인한 FERC의 9월 29일 명령서(97쪽)에 따르면 PJM의 2027/2028년 기본 용량 경매는 2025년 12월 경매 역사상 처음으로 신뢰도 기준을 크게 밑돌아 6,623메가와트가 모자랐다. 그해 예상 부하 증가 5,250메가와트 가운데 약 5,100메가와트가 데이터센터 몫이었다. 낙찰가는 하루 메가와트당 333.44달러의 임시 상한에 걸렸고, 2026년 7월 2028/2029년 경매도 상한 325달러에서 신뢰도 기준에 미달했다. 명령서는 20여 년 전 도입된 기존 신뢰도 백스톱 경매가 한 번도 쓰이지 않았다며 현행 약관이 부당할 수 있다고 봤다.

PJM의 RBP는 2028/2029년 경매 부족분 6,831메가와트를 새 발전원으로 채우는 계획이다. 낙찰 자원은 최장 15년, 2042/2043년 공급연도까지 계약을 받고 2032년 6월 1일까지 상업 운전을 시작해야 한다. PJM 이사회는 낙찰가 가중평균 상한을 하루 메가와트당 555달러로 정했다. 수요 쪽 관심도 컸다. 명령서에 인용된 사전 조사에서 21개 기업이 대형 부하 사업 47건을 냈고 평균 규모는 850메가와트였다. 공급 쪽은 87개 기업이 397건, 약 136기가와트를 제시했다.

FERC는 계획을 받아들이되 5개월 정지해 발효일을 2027년 2월 28일로 미루고서면 심리를 열었다. 비용 배분, 송전사업자 탈퇴 규정, 부하공급사업자(LSE) 담보 요건 세 가지가 부당할 수 있다고 봤다. FERC는 PJM이 30일 안에 새 신청을 내면 심리를 보류할 수 있다며 신속한 재신청을 권했다. PJM은 9월 30일로 잡았던 조달 개시를 미뤘다. PJM은 "FERC가 세 가지 개별 쟁점을 짚었지만 이를 빠르게 풀고 조달을 진행할 분명한 길도 제시했다"는 입장을 냈다.

기후

중국, 남는 재생에너지로 AI 데이터센터 확충

중국은 올해 상반기 풍력·태양광 발전량의 약 9%를 버렸고, BNEF는 이를 살릴 전력망에 2049년까지 3조9000억 달러가 필요하다고 추산했다. 서부의 남는 전력을 AI 데이터센터로 돌리는 계획의 관건은 송전선과 배터리, 그리고 가격으로 움직이는 시장이다.



1 중국의 공식 출력 제한 통계로 올해 상반기 풍력·태양광 발전량의 약 9%가 버려졌고, CREA는 상반기에만 401테라와트시를 더 생산할 수 있었다고 분석했다.

2 BNEF는 중국이 2049년까지 전력망에 3조9000억 달러를 투자해야 한다고 추산했고, 간쑤·내몽골이 2028년까지 최대 컴퓨팅 거점이 되며 데이터센터 용량이 5년 안에 두 배로 는다고 전망했다.

3 배터리 저장 용량은 세계 최대지만 독립형 배터리 충방전은 연 299회로 국제 기준 350회에 못 미치고, 단기 전력거래 비중은 12%에 머물러 시장 개혁이 관건으로 남았다.

중국이 바람과 햇빛으로 만들고도 쓰지 못해 버리는 전력을 AI 경쟁력으로 돌리려 하고 있다. 공식 출력 제한 통계로 보면 올해 상반기 풍력·태양광 발전량의 약 9%가 전력망 용량이나 수요가 모자라 버려졌다. 에너지 시장 분석기관 BNEF는 남는 전력을 실어 나르고 담아 둘 송전선과 대형 배터리, 장주기 저장장치를 갖추려면 중국이 2049년까지 전력망에 3조9000억 달러를 쏟아부어야 한다고 추산했다. 유럽이나 미국에 예상되는 투자보다 4분의 1 넘게 많다. 보도에 따르면 서부의 남는 재생에너지로 AI 데이터센터를 확충하는 일이 미국과의 AI 경쟁을 가를 변수로 떠올랐다.

규모는 간쑤성에서 드러난다. 성 북부 주취안에는 지난 20년 동안 수십 개 회사가 고비사막 끝자락에 세운 20기가와트 넘는 세계 최대 풍력단지가 있다. 중국 가정 최대 2000만 가구에 전기를 댈 수 있는 용량이지만 가구도 공장도 가까이 있지 않아 전력 상당량을 수백 킬로미터 떨어진 도시로 보낸다. 해마다 생산은 늘고 버려지는 양도 함께 늘어난다.

공식 9%는 실제보다 작을 가능성이 크다. 에너지·청정대기연구센터(CREA)는 날씨를 보정한 기대 발전량과 실제 발전량을 비교해 중국이 상반기에만 401테라와트시를 더 생산할 수 있었다고 분석했다. 스페인과 필리핀의 연간 발전량을 합친 것과 맞먹는다.