

인사이트

홍콩시립대 연구진의 AI 에이전트는 공정 최적화 모델을 세울 때 제약 조건에서 가장 약했습니다

「탱크는 넘치면 안 된다」 같은 조건은 문서에 잘 적히지 않습니다

방동걸 (주)웨이스 대표이사

2026년 10월 18일

공장의 에너지를 아끼려면 공정을 최적으로 제어해야 하고, 그러려면 먼저 공정을 수식으로 옮긴 최적화 모델이 있어야 합니다. 이 일은 지금까지 공정 엔지니어가 해 왔습니다. 공정 설명서와 배관 도면을 읽고, 무엇을 조절할 변수로 둘지, 무엇을 목표로 삼을지, 어떤 조건을 넘으면 안 되는지를 정합니다. 홍콩시립대 연구진은 2026년 9월 공개한 논문에서 이 일을 돕는 AI 에이전트 'little m'을 만들고 시험했습니다.

제조 AI 운영체제 VEXPLOR를 만드는 웨이스가 이 논문에서 주목한 것은 결과표의 가장 낮은 칸입니다. little m은 비교한 다른 AI 모델보다 나은 평가를 받았지만, 세 모델 모두 넘지 말아야 할 조건, 곧 제약 조건을 세우는 데서 가장 약했습니다. 이 글에서는 실험이 어떻게 짜였는지, 왜 제약 조건이 어려운지, 그리고 공장이 AI에게 공정 모델을 맡기기 전에 사람이 무엇을 적어 줘야 하는지를 정리합니다.

01

연구진은 공정 엔지니어가 일하는 순서대로 AI 에이전트를 3단계로 짰습니다

연구진이 먼저 짰은 문제는 이렇습니다. 범용 AI 모델에게 공정 설명을 주고 바로 최적화 모델을 쓰라고 하면, 수학적으로는 그럴듯하지만 물리적으로 맞지 않는 제약 조건을 넣기 쉽습니다. 공정 설명서에는 적혀 있지 않은 상식이 많기 때문입니다. 연구진이 든 예가 「탱크는 넘치면 안 된다」입니다. 엔지니어에게는 당연해서 아무도 적지 않지만, 이 조건이 모델에서 빠지면 최적화가 탱크를 넘치게 하는 답을 낼 수도 있습니다.

그래서 little m은 한 번에 수식을 쓰지 않고, 공정 엔지니어가 프로젝트를 진행하는 순서를 따릅니다.

- **1단계 정보 정리** — 공정 설명과 도면에서 설비·입출력·운전 상태를 뽑아 정리하고, 빠진 정보를 찾습니다. 사람이 확인해야 다음 단계로 넘어갑니다.
- **2단계 제어 전략** — 무엇을 목표로 삼을지, 어떤 제어 방식을 쓸지, 어떤 제약이 필요한지 정합니다. 여기서도 사람이 확인합니다.
- **3단계 수식 작성** — 목표 함수·결정 변수·고정 값·제약 조건을 나눠 쓰고, 모든 기호에 그 값이 어디서 오는지 적게 합니다.

2·3단계에서는 공정 최적화 지식 저장소에서 비슷한 문제의 풀이 방향과 제약 유형을 찾아 붙입니다. 저장소에는 공정 산업 전반에 되풀이되는 일반적인 풀이 구조와, 연구진이 실제 산업 프로젝트에서 정리한 사례가 함께 들어 있고 항목은 55개입니다. 추론에는 구글의 Gemini 2.5 Pro를 썼습니다.

논문에 실린 예를 보면 흐름이 잘 보입니다. 사용자가 「증발기에서 나오는 제품 농도를 안정시키는 것이 목표」라고 적고, 공정 도면과 함께 「스팀 유량, 태그 FIC-01, 단위 kg/hr, 1초 간격 수집」 같은 데이터 항목을 넘깁니다. little m은 이를 바탕으로 조절할 변수, 측정값, 외란을 나누고, 예측 제어 같은 제어 방식을 고른 뒤, 목표 함수와 상하한 제약을 수식으로 적습니다. 사람은 단계마다 그 결과를 보고 고칩니다.

02

연구진은 little m을 교과서 문제 50개로 다른 AI 모델 2개와 비교했습니다

시험 문제는 공정 제어와 화학공학 최적화 교과서에서 뽑은 50개입니다. 반응·분리·열교환 같은 공정이 들어 있고, 대부분 글과 공정 도면을 함께 읽어야 풀 수 있습니다. 비교 대상은 공개 모델 Qwen3와 DeepSeek-V3.2입니다.

평가는 두 가지로 했습니다. 하나는 사람 평가입니다. 컴퓨터과학·데이터과학·제어공학 전공의 박사과정생과 박사후연구원 8명이 20문제에서, 어느 모델이 만들었는지 가린 채 섞어 놓은 세 모델의 답 가운데 가장 나은 것을 골랐습니다. 다른 하나는 기계 평가로, 50문제 전부에서 각 모델의 답이 전문가가 검증한 정답 모델과 얼마나 맞는지를 0에서 1 사이 점수로 매겼습니다.

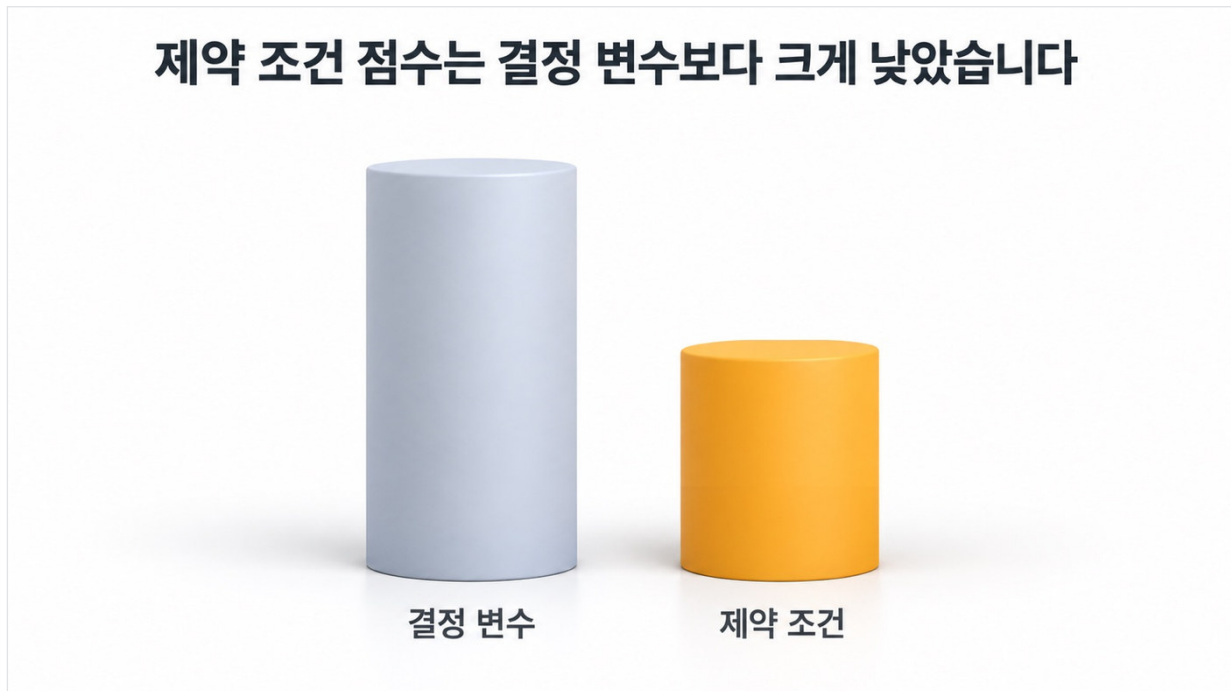
평가	Qwen3	DeepSeek	little m
사람 평가 전체 선택률	16%	18%	66%
결정 변수 점수	0.673	0.699	0.733
목표 함수 점수	0.526	0.553	0.518
제약 조건 점수	0.389	0.395	0.418

사람 평가에서는 little m이 모든 항목에서 가장 많이 뽑혔습니다. 그런데 기계 평가를 보면 차이가 크지 않고, 목표 함수 점수는 오히려 DeepSeek이 높았습니다. 연구진도 이 결과가 「모든 항목에서 낫다」는 주장을 뒷받침하지 않는다고 적었습니다.

03

세 모델 모두 제약 조건에서 가장 낮은 점수를 받았습니다

표에서 가장 눈에 띄는 것은 세로줄이 아니라 가로줄입니다. 세 모델 모두 제약 조건 점수가 0.390.42로 가장 낮았습니다. 무엇을 조절할지(결정 변수)는 0.670.73으로 비교적 잘 찾았지만, 무엇을 넘으면 안 되는지(제약 조건)는 점수가 0.5에도 못 미쳤습니다. 연구진은 문제 규모와 관계없이 제약 조건이 계속 병목이었다고 적었습니다.



세 모델 모두 결정 변수보다 제약 조건 점수가 크게 낮았습니다

지식 저장소를 뺐을 때도 가장 크게 떨어진 것이 제약 조건이었습니다. 0.418에서 0.381로 내려갔습니다. 연구진은 저장소가 운전 제약을 세우는 데 필요한 분야 지식의 틀을 제공한다고 해석했습니다.

공정 도면을 뺐을 때는 목표 함수 점수가 0.518에서 0.468로 가장 크게 떨어졌습니다. 어떤 설비가 어디로 연결되는지를 알아야 무엇을 목표로 삼을지도 정해진다는 뜻입니다. 공정 종류별로 보면 열교환 같은 열 공정이 세 모델 모두에게 가장 쉬웠고, little m은 반응과 열 공정에서, DeepSeek은 분리 공정에서 앞섰습니다. 사람 평가를 맡은 전문가들은 제약을 어떻게 묶어 쓰느냐 같은 형식에서는 의견이 엇갈렸지만, 빠진 조건이나 근거 없는 관계는 대체로 똑같이 지적했습니다.

04

빠진 정보를 AI가 물어보게 하자 제약 조건 점수가 0.314에서 0.423으로 올랐습니다

연구진은 일부러 정보를 빼고도 시험했습니다. 공정 구조, 운전 한계, 제어 정보 가운데 일부를 가린 요청을 주고 세 가지 조건으로 비교했습니다.

조건	결정 변수	제약 조건
정보를 다 줌	0.733	0.418
정보를 빼고 바로 작성	0.721	0.314
빼고, AI가 되물어 보충	0.802	0.423

정보를 빼고 바로 쓰게 하자 제약 조건 점수가 0.314로 떨어졌습니다. 하지만 AI가 빠진 것을 사람에게 되물어 보충하게 하자 0.423까지 올라, 정보를 다 줬을 때와 비슷해졌습니다. 다만 이 실험의 「사람」은 관련 있는 질문에만 가린 정보를 알려 주는 시뮬레이터였습니다. 실제 현장 엔지니어가 그렇게 답해 줄지는 따로 확인해야 합니다.

연구진은 한계도 분명히 적었습니다. 이 평가는 수식이 정답과 얼마나 비슷한지를 본 것이고, 그 모델을 실제 최적화 프로그램으로 풀어 보지도, 물리적으로 맞는지 따져 보지도, 공장에서 제어에 썼을 때 어떻게 되는지 확인하지도 않았습니다. 문제가 교과서에서 왔기 때문에 AI가 학습 중에 이미 봤을 가능성도 배제할 수 없다고 했습니다.

05

웨이스도 AI가 근거 없는 값을 채우지 못하게 막는 규칙부터 정했습니다

이 논문의 3단계에는 「모든 기호에 그 값이 어디서 오는지 적게 한다」는 규칙이 있습니다. 웨이스가 업무 문서 자동화를 만들 때 먼저 정한 것도 같은 규칙입니다. AI가 요구사항 문서를 채우게 하되, 근거가 확인되지 않은 값은 채우지 못하게 막는 규칙입니다. AI가 빈칸을 그럴듯한 값으로 메우면 사람이 그 값을 다시 검증해야 하고, 그 과정에서 틀린 값이 그대로 남기 쉽기 때문입니다.

웨이스가 10월 8일 공개한 가상 시험 설비 실험(기술 보고서 TR-2026-08)에서도 비슷한 결과를 확인했습니다. 고장 코드만으로 원인을 하나로 정할 수 없는 165문항에서, 고장 코드를 본 단순한 규칙은 90문항, 고장 코드를 가린 기록만 받은 추가 학습 없는 4B 언어 모델은 40문항을 맞혔습니다. 두 쪽이 본 정보가 달랐던 비교이지만, 고장 코드를 읽는 단순한 규칙이 AI가 넘어야 할 기준선이었습니다. 공정 모델의 제약 조건도 마찬가지입니다. 제약 조건은 현장에서 오래 일한 사람의 머릿속에 있고, 그것을 글로 옮겨 두지 않으면 AI가 알 방법이 없습니다.

현장의 제약은 글로 옮겨 두어야 AI가 쓸 수 있습니다



현장의 제약은 글로 옮겨 두어야 AI가 쓸 수 있습니다

06

공장이 AI에게 공정 모델을 맡기기 전에 적어 둘 제약 5가지

논문의 실험 설계와 한계를 공장 쪽 질문으로 바꾸면 이렇습니다. 적어 둔 것이 많을수록 AI가 세우는 모델의 제약 조건에 빈 곳이 줄어듭니다.

1. 넘치거나 비면 안 되는 것 — 탱크·호퍼·버퍼의 상한과 하한. 너무 당연해서 빠지기 쉬운 제약입니다.
2. 설비가 버티는 한계 — 온도·압력·회전수·유량의 운전 범위와, 그 범위를 정한 이유.
3. 바꾸는 속도의 한계 — 설정값을 한 번에 얼마까지 바꿀 수 있는지. 최적화 결과로는 급격한 변화가 종종 나옵니다.
4. 시작과 정지 절차 — 기동·정지 때만 지켜야 하는 순서. 연구진도 기동 절차를 사람이 바로잡는 예로 들었습니다.
5. 값의 출처 — 각 한계값이 설비 사양서에서 왔는지, 시험에서 왔는지, 숙련자의 경험에서 왔는지.



이 글을 한 장으로 요약하면 이렇습니다

little m은 공정 엔지니어가 일하는 순서를 AI에게 주고 단계마다 사람이 확인하게 하면 범용 AI보다 나은 초안이 나온다는 것을 보여줬습니다. 동시에 가장 어려운 부분은 여전히 넘지 말아야 할 조건이었고, 그 지식은 문서가 아니라 사람에게 있었습니다. AI에게 공정 최적화를 맡기려는 공장이 먼저 할 일은 현장의 제약을 글로 옮겨 두는 것입니다.

WACE(웨이스)는 공장이 스스로 판단하고 멈추지 않게 하는 제조 AI 운영체제 VEXPLOR를 만드는 회사다.

07

함께 읽어보세요

- [웨이스 시험 설비에서 AI 모델은 설비가 멈춘 원인을 단순한 규칙보다 적게 맞혔습니다](#)
- [AI가 채운 값은 왜 늘 실제보다 좋게 나올까요](#)
- [10년 뒤 로봇을 누구나 살 수 있게 되면, 중소 제조기업은 무엇으로 앞서는가 — 숙련공의 판단을 지금 기록하는 법](#)
- [지멘스 연구진의 AI 에이전트는 가상 공장의 생산 과제를 93% 풀었습니다 — 고장 난 설비가 「성공」이라고 답할 때는 구조마다 결과가 달랐습니다](#)

참고한 자료

- Yongchao Ye·Xinyu He·Dutliff Boshoff·Way Kuo·Lishuai Li, "little m: An AI Agent for Industrial Process Optimization", arXiv:2609.16680v2, 2026년 9월 16일 (홍콩시립대)



원문

blog.vexplor.com/post/little-m-process-constraints

발행

2026년 10월 18일 · 방동걸 (주)웨이스 대표이사

회사

주식회사 웨이스 (WACE Inc.) · 대표이사 방동걸 ·
사업자등록번호 439-87-02852

주소

(18462) 경기 화성시 동탄구 동탄영천로 150, 에이동(A동)
1332·1333·1334·1335호 (영천동, 실리콘엘리동탄)

연락처

전화 070-4730-0629 · contact@wace.me

홈페이지

wace.me · linkedin.com/company/wace-inc



원문



홈페이지



링크드인