인사이트

웨이스는 정적 검사와 입력 시험을 통과한 AI 래더의 10% 이상이 가상 공장 시험에서 걸러진다고 예측했고, 관측 결과는 0개였습니다

방동걸 (주)웨이스 대표이사 2026년 10월 6일

제조 AI 운영체제 VEXPLOR를 만드는 웨이스가 10월 2일 기술보고서 TR-2026-04를 공개했습니다. AI 세 가지에게 사양서만 주고 PLC 래더(설비를 움직이는 제어 프로그램)를 짜게 한 뒤, 세 단계로 시험한 실험입니다. 판정한 래더 146개 가운데 109개가 정적 검사와 입력 시험을 통과했고, 그 109개는 가상 공장에 연결한 시험에서 하나도 실패하지 않았습니다. 웨이스는 실행 전에 「통과한 래더의 10% 이상이 가상 공장 시험에서 걸러진다」는 예측을 사전 등록하고, 그 문서의 해시값을 공개 저장소에 올려 두었습니다. 관측한 결과로는 그 예측이 틀렸습니다.

그래도 웨이스는 이 결과를 「현장 시운전이 필요 없다」로 읽지 않습니다. 보고서에 적은 대로 이 실험의 실물 대조는 0회이고, 가상 공장 시험이 보는 범위는 사람이 미리 정한 기준과 상황뿐이기 때문입니다. 이 글에서는 실험에서 AI가 한 일과 사람이 한 일을 나누고, 웨이스가 그래도 현장 시운전을 남겨 두는 이유를 적습니다.

01

웨이스는 사양서만 주고 AI 세 가지에게 래더를 짜게 했습니다

실험에 쓴 설비는 세 가지 가상 셀입니다. 왕복 컨베이어(셀 A, 27렁), 분류 라인(셀 B, 569렁), 분류 라인에 포장 공정을 더한 포장 셀(864렁)입니다. 렁(rung)은 래더 프로그램의 한 줄을 뜻합니다. 과제는 21개입니다. 셀마다 래더 전체를 짜는 과제가 1개, 래더 일부를 비워 두고 그 부분을 채우는 과제가 여러 개입니다.

생성기	종류	생성 수
G1	Claude Code	63
G2	Codex CLI	63
G3	로컬 공개 모델	21

로컬 공개 모델은 실험 컴퓨터에서 직접 실행한 Qwen3-Coder-Next입니다. 세 생성기 모두 운영체제 격리 환경 안에서 실행해 정답 래더, 가상 공장 모델, 다른 과제의 결과를 읽지 못하게 막았습니다. 실행 전에 Claude Code와 Codex에게 격리 환경 안에서 정답 래더를 읽어 보게 했고, 두 번 모두 거부되는 것을 확인했습니다.

02

웨이스는 판정 기준을 결과보다 먼저 정했습니다

웨이스는 판정 기준과 예측을 사전 등록 문서에 먼저 적고, 그 문서의 해시값(파일 내용을 대표하는 고유 문자열)을 공식 실행 전에 공개 저장소에 올렸습니다(2026-09-30 18:38, 커밋 f1c095a). 결과를 본 뒤에 기준을 바꾸지 않았다는 것을 누구나 확인할 수 있게 하려는 절차입니다. 사전 등록 문서의 초안은 AI 에이전트가 썼고, 설계와 판정 기준의 책임은 저자에게 있습니다.



래더 하나가 정적 검사, 입력 시험, 가상 공장 연결 시험을 차례로 거치는 3단계 판정 흐름

- **1단계 정적 검사** — 컴파일이 되는지, 한 출력에 코일이 두 번 쓰였는지, 입출력 이름과 주소가 사양과 맞는지, 금지된 작성 방식이 있는지를 봅니다
- **2단계 입력 시험** — 가상 공장 없이 PLC만 실행하고, 입력을 넣어 출력이 기대값과 같은지 62개 시험으로 비교합니다
- **3단계 가상 공장 연결 시험** — 가상 공장에 연결해 정상 생산, 안전 커튼 정지, 복귀 순서 같은 6가지 기준으로 판정합니다

실행 전에 세 단계가 실제로 틀린 래더를 잡는지도 확인했습니다. 1단계는 일부러 결함을 넣은 래더 7개를 모두 잡았습니다. 2단계는 13개 중 9개를 잡았고, 3단계는 앞 단계 실험의 실패 기록 7개를 모두 실패로 판정했습니다.

03

AI가 짠 래더 109개는 가상 공장 시험에서 하나도 실패하지 않았습니다

146개를 판정한 결과는 이렇습니다. 로컬 공개 모델은 21개 가운데 1개가 실험 도구 결함으로 완료 기록 없이 끝나 20개를 판정했습니다.

생성기	1·2단계 통과	3단계 실패
G1 Claude Code	51/63	0/51
G2 Codex CLI	57/63	0/57
G3 로컬 공개 모델	1/20	0/1

1·2단계를 통과한 109개가 모두 3단계도 통과했습니다. 웨이스의 예측은 「10% 이상이 3단계에서 실패하고, 그 실패의 절반 이상은 타이밍이나 물리 가정 때문」이었으므로 첫 부분부터 맞지 않았습니다.

이 숫자를 읽을 때 함께 볼 것이 셋 있습니다.

- **보고서의 본 지표는 109개가 아니라 54개입니다** — 빈칸을 채우는 과제에서 정답 래더의 주석 문구와 같은 문구가 나온 래더는 무효로 처리한다는 규칙을 사전 등록 문서에 적어 두었습니다. 다만 이 규칙은 실행 코드에 빠져 있어서, 결과를 본 뒤 그 문구대로 적용했고 대조 세부도 그때 정했습니다. 이 규칙으로 55개가 빠졌고, 남은 54개에서도 3단계 실패는 0개였습니다. 다만 과제에 함께 준 래더도 같은 형식의 주석을 쓰고 있어서, 이 규칙은 정답을 본 것과 형식을 따라 쓴 것을 구분하지 못합니다
- **로컬 공개 모델의 숫자는 모델의 실력이 아닙니다** — 시간 초과가 난 요청을 실험 도구가 끝까지 정리하지 못해, 뒤 과제들이 앞 요청을 기다리다 시간을 다 썼습니다. 보고서는 이 결함 때문에 판정되지 못한 답안을 최악으로 가정한 값도 함께 적었습니다. 저자가 결과 뒤에 정한 기준으로 59개 중 5개(8.5%), 한 답안을 더 실패로 가정하면 60개 중 6개(10.0%)로 예측선 10%에 닿고, 늦게 도착한 답안을 판정하기 전 기준으로는 61개 중 7개(11.5%)였습니다. 그래서 보고서는 이 실행만으로는 결함이 판정을 바꾸지 않았다고 말할 수 없다고 적었습니다
- **0개는 3단계가 보는 범위 안에서의 0개입니다** — 6가지 기준, 셀마다 1~6회의 시나리오 실행, 가정한 물리값이 그 범위입니다. 미리 적어 두지 않은 상황에서만 드러나는 결함은 나타날 기회가 없었습니다

04

이 실험에서 AI는 대부분의 작업을 맡았고, 사람은 판단을 맡았습니다

이 실험은 AI가 래더를 짜는 데서 끝나지 않았습니다. 웨이스는 보고서에 생성형 AI를 어디에 썼는지 그대로 적었습니다.



실험의 각 작업을 AI가 맡은 일과 사람이 결정한 일로 나눈 장면

작업	맡은 쪽	비고
래더 작성	AI 세 가지	연구 대상
셀 B·포장 셀 사양서	AI 에이전트	셀마다 다른 에이전트
입력 시험 62개	Claude Code	사양서만 보고 작성
사전 등록 문서·분석 스크립트	AI 에이전트	초안 작성
실험 코드·실행 감시	AI 에이전트	대표 지시로 작성
보고서 초안·재계산	AI 에이전트	서로의 결과를 읽지 않음
설계·해석·결과의 책임	저자(대표)	보고서에 명시

사람이 맡은 일은 적었지만 가볍지 않았습니다. 실행 도중에는 로컬 공개 모델의 반복 실행을 멈출지 정했습니다. 결과를 본 뒤에 내려야 했던 판단들(결함으로 수정 기회를 잃은 답안을 어떻게 집계할지, 무효 규칙을 본 지표로 둘지, 최악 상한에 어떤 답안을 넣을지 등)도 정했고, 이 판단은 「결과를 본 뒤에 정했다」고

보고서에 그대로 적었습니다. 검수도 같은 계열의 AI 에이전트가 했고, 사람 검수자의 동료 심사는 없다는 점도 밝혔습니다.

중국 정부가 「AI로 공장이 자동화할수록 사람의 기능은 더 복잡해질 것」이라고 한 것과 비슷한 모습입니다. AI가 작성과 반복을 맡으면 사람에게는 무엇을 맞다고 볼지 미리 정하는 일과 예상하지 못한 상황에서 결과를 어떻게 다룰지 정하는 일이 남습니다.

05

웨이스는 그래도 현장 시운전을 남겨 둡니다

3단계가 109개를 모두 통과시켰다고 해서 1·2단계만으로 충분하다거나 가상 공장 시험이면 된다고 말할 수는 없습니다. 보고서도 그렇게 결론 내리지 않았습니다.

- **실물 대조는 0회입니다** — PLC도 공장도 시뮬레이션이고, 벨트 속도 같은 물리값은 가정입니다
- **사양서에서 시작한 과제만 다뤘습니다** — 이미 쓰고 있는 래더를 고치는 일과 실제 현장 사양서의 품질은 다루지 않았습니다
- **양성 사례가 없습니다** — 1·2단계를 통과하고 3단계에서만 실패하는 래더가 실제로 있을 때 3단계가 잡는다는 것을 이번 실행 안에서 보이지는 못했습니다

이번 실험의 대상은 AI가 쓴 래더 프로그램이지만, 웨이스가 VEXPLOR에서 AI의 판단 명령을 다루는 방식도 같은 생각 위에 있습니다. VEXPLOR는 AI 판단이 디지털트윈의 가상 검증을 통과했을 때만 설비로 보내도록 설계했고, 이 부분은 지금 구현과 검증을 진행하고 있습니다. 웨이스는 그 가상 검증이 현장 시운전을 대신하지 않는다고 봅니다. 가상 공장이 걸러 내는 것은 사람이 미리 정한 기준 안의 결함이고, 그 바깥은 설비를 직접 다루는 사람이 확인합니다.



이 글을 한 장으로 요약하면 이렇습니다

과제 사양서 3편, 입력 시험 62개, 생성 래더, 재계산 스크립트는 보고서와 함께 공개했습니다.

- 기술보고서 TR-2026-04: https://doi.org/10.5281/zenodo.23096367
- 공개 저장소: https://github.com/waceinc/tech-reports

AI가 짠 래더가 가상 공장 시험을 모두 통과해도 실제 설비에서 맞는지는 현장 시운전에서 확인되고,
웨이스는 그 확인을 사람의 일로 남겨 둡니다.

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

06

함께 읽어보세요

- 중국은 AI 공장을 늘리면서 왜 숙련공을 국가대표처럼 키우는가
- 웨이스의 가상 공장 실험에서 PLC는 15 ms 정지 신호를 10번 중 7번 놓쳤습니다 — 입력을 계산 주기 시작 값으로만 받았을 때
- 웨이스가 VEXPLOR에서 AI의 판단을 PLC로 보내기 전에 같은 스캔의 가상 공장에서 먼저 실행하려는 이유
- 웨이스가 부정적 결과로 끝난 AI 명령 검증 실험을 기술보고서 TR-2026-01로 공개했습니다

원문 blog.vexplor.com/post/wace-ai-plc-ladders-prediction-wrong
발행 2026년 10월 6일 · 방동걸 (주)웨이스 대표이사

회사 주식회사 웨이스 (WACE Inc.) · 대표이사 방동걸 · 사업자등록번호 439-87-02852
주소 (18462) 경기 화성시 동탄구 동탄영천로 150, 에이동(A동) 1702-1708호 (영천동, 실리콘앨리동탄)
연락처 전화 070-4730-0629 · contact@wace.me
홈페이지 wace.me · linkedin.com/company/wace-inc

원문 홈페이지 링크드인