

회사소식

웨이스가 부정적 결과로 끝난 AI 명령 검증 실험을 기술보고서 TR-2026-01로 공개했습니다

방동걸 (주)웨이스 대표이사

2026년 9월 21일

제조 AI 운영체제 VEXPLOR를 만드는 웨이스가 2026년 9월 21일 첫 기술보고서 TR-2026-01을 공개했습니다. 제목은 「Behavioral Equivalence Cannot See Dormant Logic」, 우리말로 옮기면 「동작 일치 검사는 잠자는 로직을 보지 못한다」입니다. 영문 보고서이고, 회사 GitHub 저장소 waceinc/tech-reports에 PDF와 마크다운으로 올려 두었으며, 라이선스는 CC BY 4.0이라 출처를 밝히면 누구나 옮기고 인용할 수 있습니다. 저자는 웨이스 대표이사입니다.

보고서가 다루는 것은 AI나 MES가 PLC에 쓰는 명령을 설비 앞에서 한 번 더 확인하는 장치입니다. 지난 글에서 「실행 전 가상 검증」이라고 소개한 그 장치를 5가지 구성으로 나누어 같은 명령 869건을 넣었고, 결론은 짧습니다. 검증용 프로그램이 현장과 같은지를 동작으로 확인하는 검사를 붙여도, 놓치는 위험 명령은 한 건도 줄지 않았습니다. 불인 쪽과 뺀 쪽이 똑같이 280건 가운데 23건을 통과시켰습니다. 웨이스는 이 실험의 내용에 특허를 출원하지 않기로 하고 방어적 공개(남이 같은 내용으로 특허를 받지 못하도록 먼저 공개하는 것)로 냈습니다. 이 글은 무엇을 실험했는지, 23건이 어디에서 왔는지, 실험 규칙이 없었다면 어떤 숫자가 나갔을지, 그리고 다음에 무엇을 하는지를 적습니다.

정정 안내 (2026-09-22). 웨이스가 이 보고서를 v1.1 로 바로잡았습니다. 아래 「사람 노출 8건 가운데 인터록 제거 뒤 가동 명령 4건」은 실제로는 **PLC 가 정지해 있는 동안 들어온 가동 명령**이었습니다(4건 가운데 현장 프로그램에서 인터록이 빠진 것은 1건). 가동 중인 PLC 에서 인터록이 빠진 상태의 위험 명령은 시험 세트에 12건 있었고 모두 막혔습니다. PLC 정지 중에 들어온 명령은 이후 넣은 스캔 카운터 확인이 24건 모두 막힙니다. 「조건식 형태 4건은 7회차에서 0이 될 것」이라는 예상은 틀렸습니다 — 현장 프로그램에서만 타이머 설정값이 바뀔 경우라 프로그램 서명 대조로만 잡힙니다. 다시 낸 결과는 두 구성 모두 278건 가운데 19건 통과(사람 노출 4건)이고, 결론인 「동작 일치 검사를 붙여도 놓치는 위험 명령은 줄지 않는다」는 그대로입니다. 「최소 1초 전에 동기화된 검증용 프로그램의 값」이라는 설명도 틀렸습니다 — 검증용 프로그램은 판정할 때마다 동기화되고, 1초는 동작 일치 검사가 값을 확인하는 간격입니다. 정정 전문: [TR-2026-01 v1.1](#) · [정정 목록](#)

01

5가지 구성을 하나씩 더해 가며 같은 명령 869건을 넣었습니다

실험 대상은 명령을 내는 쪽(AI 계획기·MES·HMI)과 PLC 사이에 두는 장치입니다. PLC의 상태를 그대로 옮겨 둔 검증용 프로그램에 맞춰 두고, 들어온 명령을 검증용 프로그램에서 몇 주기 앞서 실행해 본 뒤 조건이 깨지면 보내지 않습니다. 이 방식 자체는 공개된 선행 기술이 있고, 웨이스가 물은 것은 그 위에 검사 하나를 더 붙이면 무엇이 달라지는가였습니다.

- 프로그램 — 자체 작성 래더 프로그램 3개(컨베이어 인터록·비상정지·배치, 3~19줄).
- 명령 — 869건. 각 명령에 「검증용 프로그램이 현장과 같음/어긋남」 조건과 위험 등급(사람 노출·제품·기록·지표)을 실행 전에 붙였고, 명령 목록·규칙·등급 파일의 해시를 고정했습니다.
- 구성 — 5가지. 게이트 없음 → 주소·값 규칙만 → 미리 실행(출력만 동기화) → 미리 실행(내부 상태 동기화) → 여기에 동작 일치 검사 추가. 넷째와 다섯째의 차이가 검사 하나의 효과입니다.
- 실행 — 공식 6회. 유효 5회, 무효 1회. 채택한 것은 6회차이고 모든 커밋과 기록을 남겼습니다.

구성	놓친 위험 명령	잘못 막은 명령	정지 명령 차단(94건 중)
게이트 없음	246/284 (86.6%)	0	—
주소·값 규칙만	95/284 (33.5%)	128/309 (41.4%)	47/94
미리 실행, 출력만 동기화	11.9%	13/309 (4.2%)	2/94
미리 실행, 내부 상태 동기화	23/280 (8.2%)	6/309 (1.9%)	2/94
위 구성 + 동작 일치 검사	23/280 (8.2%)	6/309 (1.9%)	2/94

기준이 되는 280건은 위험 명령 284건에서, 불러온 상태가 이미 규칙을 위반해 판정 대상이 아니었던 4건을 뺀 값이고, 그 4건의 번호는 결과 파일에 있습니다. 셋째 구성은 결과 파일에 전체 건수가 없어 비율만 적었습니다. 판정 지연은 16코어 노트북 한 대 기준으로 95번째 백분위(p95)가 36밀리초 안팎입니다.

둘째 구성이 막은 정지 명령 47건은 실험을 무효로 만드는 차단이었습니다. 셋째 구성부터 막힌 2건은 불러온 상태가 이미 규칙을 위반한 명령이고, 실험을 무효로 만드는 사유로 막힌 정지 명령은 0건입니다. 다만 이 숫자가 「게이트는 정지 명령을 절대 막지 않는다」는 뜻은 아닙니다. 판정은 그 명령이 위반을 줄이는지를 보지 않기 때문에, 비상정지 계열은 게이트를 거치지 않게 배선하는 것을 설치 조건으로 둡니다.

5가지 구성을 하나씩 더해 가며



5가지 구성 — 왼쪽부터 부품을 하나씩 더한 순서이고, 넷째와 다섯째의 차이가 동작 일치 검사 하나의 효과입니다

02

놓친 23건 가운데 사람 노출 8건, 그중 4건은 규칙을 고쳐도 잡히지 않습니다

23건을 위험 등급으로 나누면 사람 노출(사람이 다칠 수 있는 명령) 8건, 제품 6건, 기록 5건, 지표 4건입니다. 사람 노출 8건은 원인이 2가지로 나뉩니다.

원인	건수	해결 방법
조건식이 순간의 이벤트 형태	4	상태값으로 고치면 됨(7회차 예상)
인터록 제거 뒤 가동 명령	4	규칙으로는 안 됨

앞의 4건은 경보음 지속 시간이나 로봇 팔 대기 시간이 바뀐 상태에서 통과한 명령입니다. 조건을 「이벤트가 있었는가」로 적어 두어 게이트가 동기화 이전의 이벤트를 알 수 없었던 것이고, 같은 사실을 PLC 타이머의 값으로 고쳐 적으면 판정할 수 있습니다. 7회차에서 0이 될 것으로 예상하지만 아직 결과가 아니라 예상입니다.

뒤의 4건이 보고서 제목이 된 이유입니다. 인터록 한 줄이 빠진 프로그램은 가동 명령이 올 때까지 아무 태고도 움직이지 않습니다. 움직임이 없으니 동작을 관찰하는 검사는 경계 상태로 들어가지 않고, 가동 명령이 왔을 때 게이트가 읽는 허용 조건은 최소 1초 전에 동기화된 검증용 프로그램의 값입니다. 검증용 프로그램 안의 인터록은 정상이므로 명령은 통과합니다. 이것은 구현의 결함이 아니라 동작을 관찰해

최신인지 확인하는 검사 전체의 경계이고, 그래서 보고서는 프로그램 서명값(체크섬)을 확인하는 정적 검사를 동작 검사보다 앞에 두라고 적었습니다. 서명값이 없는 프로토콜(Modbus)에서는 그 검사가 성립하지 않으므로 PLC 쓰기 보호와 변경 관리가 그 역할을 맡아야 합니다.

막을 수 없다고 확인한 위험도 따로 적었습니다. 원격 리셋과 원격 알람 확인은 래더 프로그램이 정당하게 제공하는 조작이라 게이트가 운전원이 누른 1과 상위 시스템이 쓴 1을 구분하지 못하고, 두 워드로 쓰는 32비트 설정값은 한 워드만 반영되면 아무도 명령하지 않은 값이 됩니다(정상 변경 30건 가운데 17건 미반영, 4건 반만 반영). 3종류, 시험 명령 34건입니다.

03

실험 규칙이 없었다면 나갔을 숫자와, 실제로 낸 숫자는 이렇게 다릅니다

이 보고서에서 웨이스가 보탠 것은 검사의 경계를 확인한 것과 함께, 그 숫자를 만든 실험 규칙입니다. 규칙은 실행 전에 적어 두었고, 실행하면서 무엇을 잡았는지도 표로 남겼습니다.

규칙	나갈 뻔한 값	실제로 낸 값
판정 3회 불일치면 무효	4회차 「유효」	4회차 무효
정지 차단 0건 기준	「227건 차단」	1회차 미채택
등급 집계는 도구로만	사람 4	사람 8
건수·전체·제외 공개	「8.2%」만	23/280 + 제외 4
사후 재분류 금지	4건 뺀 건수	4건 포함 건수
제외 전후 함께 공개	제외 효과만	포함·제외 나란히
출처 검사	출처 틀린 결과	고쳐서 재실행

몇 줄은 설명이 필요합니다. 「사람 4」는 실험 책임자가 손으로 세어 본 값이었고, 도구가 낸 값은 8이었습니다. 규칙이 손 집계를 금지하지 않았다면 4가 나갔을 것입니다. 인터록 제거 4건을 「막을 수 없는 위험」으로 옮기면 놓친 명령 수가 4건 줄어드는데, 사후 재분류를 금지했기 때문에 그대로 두고 별도 항목으로 적었습니다. 4건을 측정 제외로 옮긴 것은 4회차가 무효가 된 뒤에 사전 등록에서 벗어나 도입한 변경이고, 그 사유와 시각을 기록했으며, 포함한 판(23회차)과 제외한 판(56회차)을 나란히 냈습니다. 제외로 움직인 숫자는 기본 조건의 정지 명령 차단 4→2 하나입니다. 잘못 막은 비율과 일치 상태의 판정 불가 비율도 5회차부터 달라졌는데(3.2%→1.9%, 38.7%→38.2%), 원인은 건별 비교가 남아 있어 보고서는 값만 적었습니다.

6회를 실행한 기록도 그대로 있습니다. 농친 명령 23건은 2회차부터 6회차까지 변하지 않았고, 1회차에서 2회차로 갈 때는 오히려 불리한 쪽(0.078→0.082)으로 움직였습니다. 게이트 소스는 1회차와 6회차가 같고 명령 목록의 해시는 6회 전부 같습니다.

이 원칙들은 새것이 아닙니다. 연구자의 자유도와 사전 등록, 결과 변수의 사전 코딩, 보안 평가의 함정을 다룬 선행 연구가 이미 같은 말을 합니다. 웨이스가 더한 것은 그 원칙을 기계가 강제하도록 한 것입니다. 완화된 항목이 무효 판정의 근거가 된 항목과 겹치면 실행 자체가 거부되고, 보고서 첫머리의 숫자가 공개된 구성 요소에서 다시 계산되지 않으면 기록 파일이 만들어지지 않습니다.



규칙이 잡은 지점 — 나갈 뻔한 값과 실제로 낸 값이 달라진 7곳 가운데 3곳(등급 집계·건수 공개·4회차 판정)

04

특허 대신 방어적 공개를 고른 이유는 3가지입니다

- **방식 자체가 선행 기술입니다.** 검증용 프로그램에서 미리 실행해 보는 게이트는 공개된 특허와 연구가 있고, 웨이스가 새로 낸 것은 「검사를 하나 더 붙이면 무엇이 달라지는가」의 측정입니다. 측정은 반복될수록 쓸모가 있고, 반복되려면 공개돼야 합니다.
- **부정적 결과는 묻히기 쉽습니다.** 「효과가 없었다」는 결과는 논문으로도 제품 설명으로도 잘 실리지 않습니다. 그래서 같은 검사를 붙이려는 다음 팀이 같은 실험을 다시 하게 됩니다. 실험 설계와 전체 건수, 제외한 항목 번호까지 공개하면 그 팀은 그 지점에서 시작할 수 있습니다.
- **주장할 수 없는 것을 주장하지 않기 위해서입니다.** 시뮬레이션 PLC 한 대, 자체 작성 래더 프로그램 3개, 외부 작성 프로그램 0개, 실기 PLC 대조 0회입니다. 이 조건에서 남들이 못 하는 기술이라거나 현장에서

검증된 안전이라고 말할 수 없고, 보고서도 그렇게 적었습니다. 학회 발표는 보류했고, 게이트 소스와 원시 결과 파일은 이 저장소에 넣지 않았습니다. 코드 공개는 별도로 결정합니다.

보고서에는 생성형 AI 관여도 적었습니다. 소스 코드·실험 도구·설계 문서의 상당 부분을 LLM 코딩 에이전트가 저자의 지시로 작성했고, 실험 규칙의 독립 판정과 검토 일부도 같은 모델의 에이전트가 맡았습니다. 설계 승인, 사전 등록 확정, 실행 유효 판정, 해석, 보고서의 모든 주장은 사람 저자가 확인했고 책임집니다. 「독립 판정」은 서로의 출력을 읽지 않은 에이전트 실행을 뜻하며 사람 검토자를 뜻하지 않습니다.

05

다음 실험 5가지와 현장에 요구하는 조건 4가지를 먼저 정해 두었습니다

보고서는 여기서 멈추지 않고 다음 순서를 적어 두었습니다. 실험은 5가지입니다.

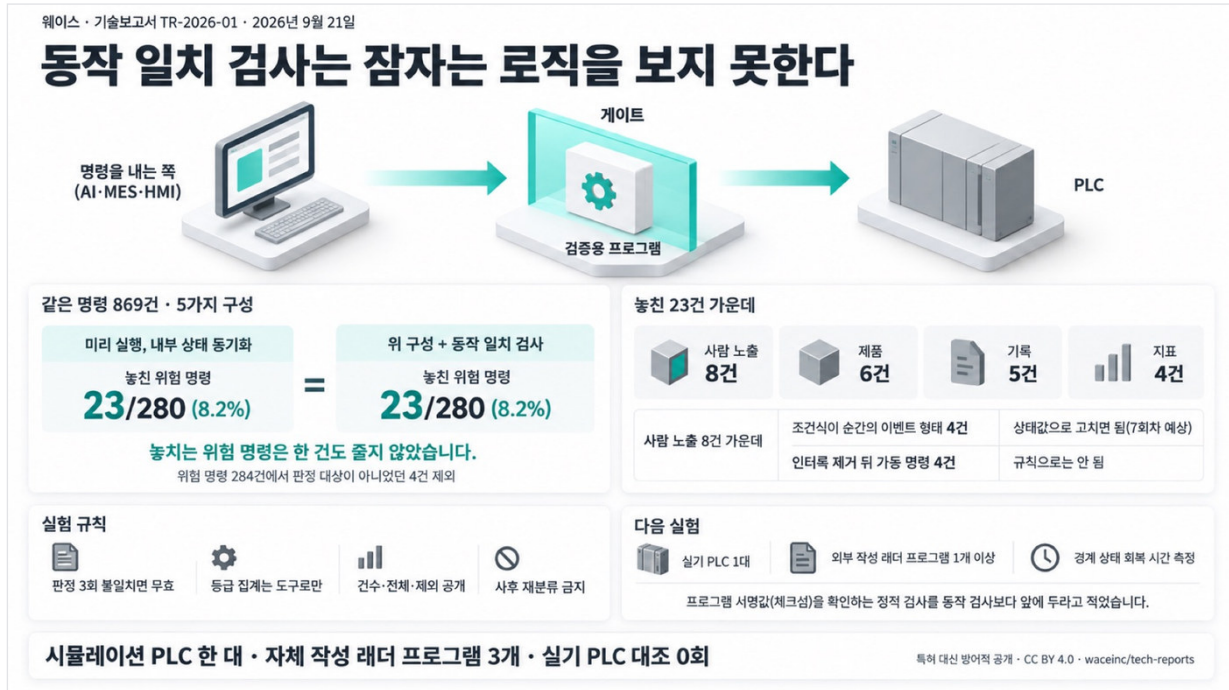
- **실기 PLC 1대** — 가동 중 올린 25건, 인터록 제거 4건, 정지 명령 94건, 강제 설정 비트.
- **강제 설정 비트·값 덮어쓰기** — 프로그램은 같은데 값만 바뀔 경우. 동작 일치 검사가 효과를 내야 하는 유일한 종류인데 이번 명령 집합에는 없었습니다.
- **7회차** — 순간의 이벤트로 적은 조건식을 상태값으로 고치고, 서명값 검사를 앞에 두고, 허용 조건을 PLC에서 직접 읽습니다. 예상 방향(조건식 4→0, 인터록 제거 4→변함없음)을 실행 전에 등록했습니다.
- **외부 작성 래더 프로그램 1개 이상.**
- **경계 상태 회복 시간 측정** — 하한 8초, 상한 미확인.

현장에 요구하는 조건은 4가지입니다.

- **정지·차단 동작은 게이트를 거치지 않게 배선합니다.** 게이트는 불러온 상태가 이미 규칙을 위반한 명령을, 그것이 설비를 멈추는 명령이어도 막기 때문입니다.
- **원격 리셋·원격 알람 확인 비트는 상위 시스템 권한에서 뺏니다.**
- **32비트 설정값은 다중 레지스터 쓰기(FC16)로 보냅니다.**
- **PLC 스캔이 멈추면(정지 모드·감시 타이머 정지·운전 중 편집 일시정지) 상위 명령을 설계상 막습니다.** PLC가 스캔 중인 채로 라인만 멈춘 경우는 막지 않습니다.

타이머와 카운터 내부값은 통신으로 노출돼야 하고, 무엇이 위반인지는 그 공장에 맞게 적어야 합니다.

알려진 한계도 그대로 적혀 있습니다. 사람 노출 농침은 6회차 기준 8건이고 그중 4건은 규칙 변경으로 해결되지 않으며, Modbus에서는 잠자는 로직 변경을 볼 수 없고, 1초 안에 바뀌는 입력은 시간 기반 해제 없이 게이트를 경계 상태에 붙들어 둘 수 있습니다.



이 글을 한 장으로 요약하면 이렇습니다 — 게이트의 위치, 5가지 구성, 농진 23건, 실험 규칙이 잡은 값, 다음 숫자

공장이 스스로 판단하게 하려면 판단이 틀렸을 때 설비 앞에서 멈추는 장치가 있어야 하고, 그 장치가 무엇을 못 보지는 만든 쪽이 먼저 측정해 적어야 합니다. 웨이스는 그 첫 숫자를 기술보고서로 냈고, 다음 숫자는 실기 PLC에서 냅니다.

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

06

함께 읽어보세요

- 기술보고서 원문(영문, CC BY 4.0): [WACE Technical Report TR-2026-01 · PDF](#)
- [AI에게 설비를 맡기려면, 틀린 명령이 나가기 전에 멈출 수 있어야 합니다](#)
- [설비 제어 로직을 소프트웨어로 실행해 보면, 처음 틀리는 곳은 3군데입니다](#)



원문 blog.vexplor.com/post/tr-2026-01-negative-result
발행 2026년 9월 21일 · 방동걸 (주)웨이스 대표이사

회사 주식회사 웨이스 (WACE Inc.) · 대표이사 방동걸 ·
사업자등록번호 439-87-02852
주소 (18462) 경기 화성시 동탄구 동탄영천로 150, 에이동(A동)
1702·1706호 (영천동, 실리콘엘리동탄)
연락처 전화 070-4730-0629 · contact@wace.me
홈페이지 wace.me · linkedin.com/company/wace-inc



원문



홈페이지



링크드인