기술

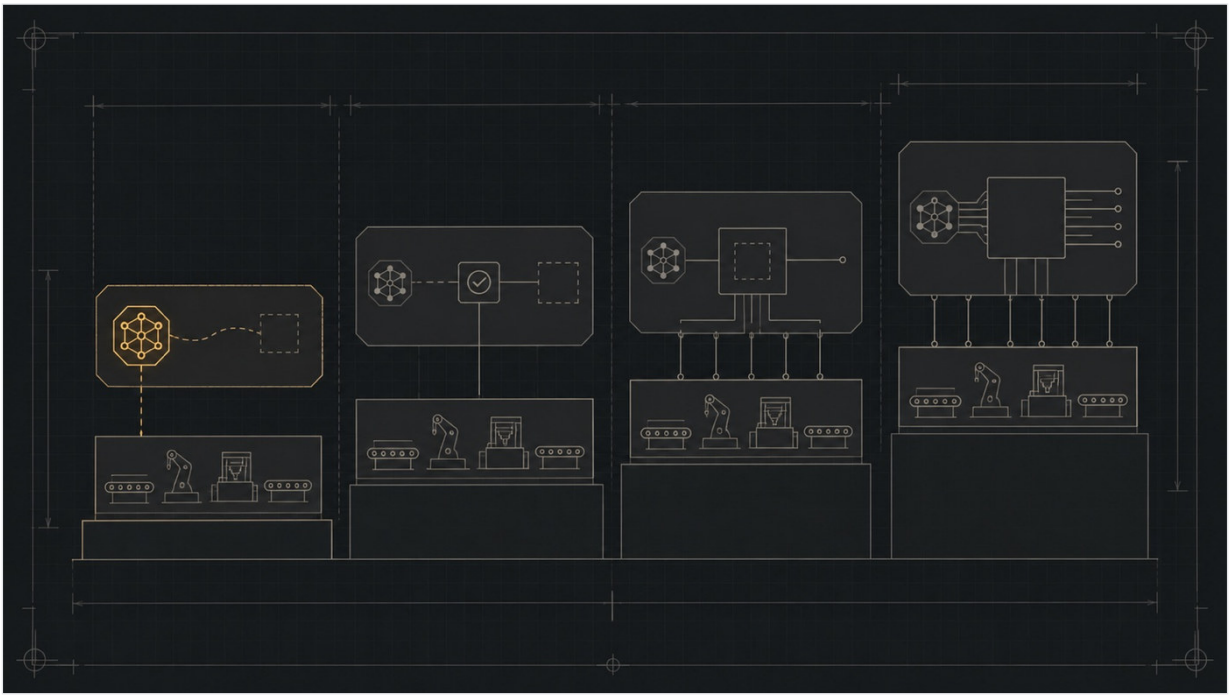
"AI가 알아서 라인 세웠습니다": 이 한마디가 왜 공포인가

방동걸 (주)웨이스 대표이사 2026년 6월 21일

새벽 두 시, 생산팀 담당자 휴대폰에 알림이 뜹니다. "이상 징후 감지. 3호기 자동 정지 완료." 좋은 소식처럼 들리시나요? 현장에 한번 계셔 보면 알게 됩니다. 이 문장을 본 담당자의 첫 반응은 안도가 아니라 등골이 서늘해지는 쪽이에요. 누가, 무슨 근거로, 무슨 권한으로 라인을 세웠지? 정지 자체가 옳았더라도, 사람이 모르는 사이에 AI 혼자 결정하고 실행했다는 사실 그 자체가 무섭습니다. 웨이스가 VEXPLOR에서 AI에 실행 권한을 주는 설계를 하며 가장 무겁게 들은 말입니다.

저희가 현장에서 "AI가 판단해서 실행한다"는 그림을 그려 보이면, 다들 처음엔 눈이 반짝합니다. 그러다 잠깐 뒤에 표정이 굳어요. "그래서, 그게 틀리면요?" 솔직히 이게 자율 AI의 진짜 장벽입니다. 모델이 모자라서가 아니에요. 요즘 모델은 충분히 똑똑합니다. 문제는 잘못된 자동 실행 하나가 곧장 불량으로, 라인 정지로, 심하면 안전 사고로 번진다는 거죠. 자동차나 의료처럼 규제가 빡빡한 곳일수록 "AI가 알아서 한다"는 말은 매력이 아니라 그대로 리스크입니다.

여기서 흔히 나오는 반론이 있어요. "그럼 다 사람이 승인하면 되잖아." 맞는 말 같지만, 그렇게 하면 확장이 막힙니다. 모든 판단에 사람 결재가 붙으면 AI를 붙인 의미가 사라져요. 새벽 두 시에 사람을 깨워 승인받느라 대응이 30분 늦으면, 그게 더 큰 사고가 됩니다. 그러니 답은 둘 중 하나가 아니라 그 사이에 있습니다. 전부 자동도, 전부 수동도 아닌, "무엇까지 AI에게 맡길지"를 단계로 그어 두는 거버넌스요. 오늘은 저희가 그 선을 어떻게 긋는지 얘기해 보려고 합니다.



권한은 한 번에 주는 것이 아니라 한 칸씩 넓혀 주는 것입니다.

01

권한을 한 칸씩 나눠서 줍니다

VEXPLOR는 AAF(Agent Authority Framework)로 AI의 실행 권한을 다섯 단계로 정의합니다. 자율성을 단계로 쪼갠다는 발상 자체는 새로운 게 아니에요. human-in-the-loop이라는 이름으로 오래 논의돼 온 일반 원리입니다. 저희가 한 건, 그 원리를 제조 운영이라는 까다로운 현장에 맞게 권한 사다리로 구체화한 거예요.

레벨	권한	의미
L0	조회	데이터 읽기만
L1	제안	분석 후 추천
L2	승인 후 실행	사람 승인 → 실행
L3	자동 실행	정책 범위 내 자동
L4	긴급 대응	안전 위협 시 즉시 조치

핵심은 하나입니다. 업무와 위험도에 따라 레벨을 다르게 준다는 것. 단순 조회는 L0면 충분하고, 개선 아이디어 정도 내는 건 L1, 영향이 큰 변경은 L2로 묶어 사람 승인을 받게 하고, 정책으로 안전이 보장되는 반복 작업만 L3 자동에 올립니다. 안전 위협 상황은 L4가 즉시 대응하고요. 그러니까 하나의 AI가 "전부

자동"으로 풀려 있는 게 아닙니다. 작업마다 허락된 권한 안에서만 손을 댈 수 있어요. 아까 새벽 두 시 알림이 무서웠던 건, 정지라는 행동이 사실 어느 레벨 권한이었는지 담당자가 몰랐기 때문입니다. AAF가 깔려 있으면 그 답이 처음부터 명확합니다.

02

같은 "설비 정지"인데, 자동차와 의료가 다릅니다

말로만 단계라고 하면 추상적이니, 시나리오 하나를 따라가 볼게요. "이상 감지 시 설비를 정지시킨다"는 똑같은 액션을 두 공장에 가져갑니다. 한쪽은 자동차 부품 라인, 다른 쪽은 의료기기 생산 라인입니다.

자동차 쪽은 IATF 16949 체계 안에 있습니다. 여기서도 추적성과 검증은 당연히 중요하지만, 특정 안전 한계를 넘어선 이상 신호에 대해서는 "사람을 기다리지 말고 일단 멈추는" 쪽을 더 안전하다고 볼 여지가 있어요. 그래서 잘 정의된 정책 범위 안에서라면 정지 액션을 L3 자동, 혹은 안전 위협 상황의 L4 긴급 대응으로 둘 수 있습니다. 사람은 사후에 감사 추적을 보고 왜 멈췄는지 확인하는 식이죠.

같은 액션을 의료기기 라인으로 옮기면 그림이 달라집니다. 이쪽은 FDA, ISO 13485의 검증·승인·문서화 요건이 훨씬 무겁습니다. 공정 변경 하나하나가 추적되고 승인 경로를 거쳐야 하는 세계예요. 그러니 똑같은 "설비 정지"라도 AI가 단독으로 실행하는 L3로 두기보다, 사람 승인을 거치는 L2 쪽으로 권한을 낮춰 거는 게 규제 맥락에 맞습니다. 같은 AI, 같은 기능인데 허용 레벨이 산업에 따라 통째로 바뀌는 거죠.

이게 권한 단계의 진가가 드러나는 지점입니다. VEXPLOR는 AAF를 규제 산업별 거버넌스에 매핑합니다. IATF 16949(자동차), FDA(의약·의료), ISO 13485(의료기기) 같은 체계요. 한 가지만 분명히 해 둘게요. 이건 저희가 어떤 인증을 취득했다는 얘기가 아닙니다. 권한 레벨을 규제 요건에 맞춰 다르게 매핑한다는 뜻이에요. 추적성·검증·승인 요건이 엄격한 영역은 자동화 레벨을 낮추고, 위험이 낮고 반복적인 영역은 자동으로 올리는 식으로요. 그래야 "규제를 지키면서도 자동화한다"가 말장난이 아니라 실제로 돌아갑니다.

03

L1을 담당하는 얼굴, AI 어드바이저

권한 사다리에서 L1(제안)을 맡는 게 AI 어드바이저입니다. 하는 일은 공정을 분석해 병목을 짚어내고, 스키마와 워크플로우 개선안을 제안하는 거예요.

여기서 재밌는 부분이 있습니다. 제안의 대상이 데이터에 그치지 않는다는 거예요. 어드바이저는 시스템 구조 자체를 들여다봅니다. 어디가 병목인지 찾고, 그걸 풀려면 스키마나 워크플로우를 어떻게 바꿔야 하는지까지 제안해요. 그렇다고 어드바이저가 멋대로 구조를 갈아엎느냐 하면, 절대 아닙니다. 어디까지나 L1, 제안까지입니다. 그 제안을 받아 검토하고 승인할지(L2), 아니면 정책으로 못 박아 자동화할지(L3) 결정하는 건 사람과 정책의 몫이에요. AI는 운영을 더 낫게 만들자고 손을 들 뿐, 실제로 적용할지 말지의 통제권은 사람 쪽에 그대로 남겨 둡니다.



같은 '설비 정지'라도 산업마다 사람이 쥐는 결재선이 다릅니다.

04

성숙도를 올린다는 건, 권한을 넓힌다는 것

VEXPLOR의 AI 성숙도는 네 칸으로 올라갑니다. L1 수동 질의응답에서 시작해 L2 능동 선제개입, L3 자율 판단·실행, 그리고 L4 AI OS 공장운영까지요. 한 줄로 줄이면 "AI가 얼마나 스스로 공장을 운영하느냐"의 단계입니다.

그런데 성숙도를 올린다는 말을 다르게 풀면, AI에게 주는 권한을 넓힌다는 뜻이에요. 그리고 권한을 넓힐 때마다 무서워지는 게 자연스럽죠. 앞에서 본 새벽 두 시의 그 알림이 점점 잦아진다는 얘기니까요. AAF가 바로 이 상승을 안전하게 받쳐 주는 장치입니다. 권한이 한 칸 올라갈 때마다 그에 맞는 정책·승인·긴급대응 규칙이 따라붙어요. 그래서 자율성이 올라가도 통제가 같이 따라 올라갑니다.

게다가 VEXPLOR는 도입과 동시에 L1부터 L3까지의 토대가 자동으로 깔립니다. 처음 켜는 순간부터 어드바이저가 제안하고, 사람이 승인하고, 정책 범위 안에서 자동이 도는 골격이 서 있다는 거예요. 그 위에서 "어디까지 자동에 맡길지"는 AAF로 그때그때 조절합니다. 자율성과 통제, 이 둘이 따로 노는 게 아니라 한 몸으로 같이 가는 구조입니다.

05

그런데 왜 표준화가 먼저냐

"정책 범위 내 자동 실행"(L3)을 믿을 만하게 정의하려면, 그 정책이 가리키는 데이터와 관계가 표준화돼 있어야 합니다. 생각해 보면 당연해요. "이 조건이면 이 작업을 자동으로 허용한다"는 규칙을 쓰려면, 그 "조건"과 "작업"이 시스템 안에서 흔들림 없이 같은 의미여야 하잖아요. 데이터 구조와 온톨로지가 표준으로 깔려 있어야 이 규칙을 일관되게 표현하고 검증할 수 있습니다.

결국 거버넌스도 저희가 입버릇처럼 말하는 "표준화 없이 AI 없다"의 연장선입니다. 표준이라는 단단한 바닥 위에서만 권한과 정책이 제자리에서 작동해요. 바닥이 흔들리면 아무리 정교한 권한 사다리를 그려도 그 위에서 같이 흔들립니다.

06

같이 풀 문제, 그리고 같이 만들 사람

AI 거버넌스에는 솔직히 머리 아프지만 재밌는 난제가 잔뜩 남아 있습니다. 권한 정책을 정형으로 표현하고 검증하는 법, 규제 요건을 권한 레벨로 자동 매핑하는 법, 자동 실행의 감사 추적(audit trail)과 설명가능성을 어떻게 보장할지, 긴급 대응(L4) 트리거를 어떻게 안전하게 설계할지, 사람 승인 부하와 자동화 사이의 최적점을 어디에 둘지 같은 것들이요. 하나하나가 "AI가 행동하되 안전하게"라는 한 문장을 진짜로 만드는 일입니다.

이걸로 6편짜리 시리즈를 마무리합니다. 돌아보면 결국 같은 이야기를 여섯 갈래로 한 셈이에요. 검색을 네 갈래로 쪼개 숫자를 안 틀리게 만들고, 관계를 온톨로지로 표준화하고, 그 위에서 노코드로 운영을 짜고, 예지보전으로 미리 손쓰고, 그리고 마지막으로 AAF로 "그 모든 자동화를 어디까지 풀어 줄지"를 안전하게 통제하는 일. 처음부터 끝까지 관통하는 한 줄은 변하지 않습니다. 표준 위에서, 사람이 통제권을 쥔 채로, AI가 현장에서 실제로 일하게 만든다는 것.

VEXPLOR는 제조 AI OS를 만드는 (주)웨이스의 제품입니다. 위 문제들이 "어 이거 재밌겠는데" 싶으셨다면, 아마 잘 맞으실 거예요.

- 관심 분야: AI 에이전트·오케스트레이션 / 정책·거버넌스 엔진 / 규제 대응(컴플라이언스) / 산업 안전·신뢰성
- 더 알아보기: 채용·회사 소개 https://wace.me · 기술 문의 contact@wace.me

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

원문 blog.vexplor.com/post/ai-advisor-aaf-governance
발행 2026년 6월 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) 1702-1708호 (영천동, 실리콘앨리동탄)
연락처 전화 070-4730-0629 · contact@wace.me
홈페이지 wace.me · linkedin.com/company/wace-inc

원문 홈페이지 링크드인