10th Anniversary

GISC 2026

Global ICT Standards Conference 2026

(세션3) 인공지능의 미래: Core AI에서 피지컬 AI까지의 사실표준화 동향

안전한 피지컬 AI 시대를 위한 AI 신뢰성 평가 및 안전 기준 표준화

정재룡 실장
슈어소프트테크

ICT Standards and Intellectual Property:
Shaping the Physical AI

주최 과학기술정보통신부 지식재산처 주관 국립전파연구원 TTA IITP KEA kista ETRI

Index

01 피지컬 AI 안전·신뢰 개요

02 안전 입증 방법론

03 요약 및 정리

Abstract

발표 요약

ICT Standards

피지컬 AI의 오류는 '틀린 답'이 아니라 물리적 사고로 이어진다.
본 발표는 피지컬 AI 관련 공식표준과 사실표준의 현황을 정리하고,
공식표준이 아직 비어 있는 영역을 사실표준과 현장 방법론이 어떻게 메우고 있는지 살펴본다.
확률적으로 동작하고 학습으로 변하는 시스템의 안전과 신뢰를
Safety Case 기반으로 '증명'하는 방법(SOTIF·ODD, 안전 외피, 시나리오 기반 평가)을 사례와 함께 설명한다.
끝으로, 방법론만으로는 해결되지 않는 과제에서 다음 표준화 의제를 도출하고 방향을 제언한다.

01. 개요

현실화된 피지컬 AI 사고

국내

2023.11 · 경남 고성

선별장 로봇이 작업자를 상자로 오인해 압착, 사망

시운전 전 센서를 점검하던 제조사 직원이 피해

실패 유형 인식 오류

해외

2023.10 · 美 샌프란시스코

로보택시가 보행자를 약 6m 끌고 간 뒤 정차

캘리포니아 DMV 운행 허가 정지 · NHTSA 150만 달러 제재

실패 유형 판단·대응 오류

해외

2025 소송 · 美 테슬라 공장

정비 중 로봇 암이 갑자기 작동해 작업자 중상

작업자가 테슬라·FANUC 상대 약 5,100만 달러 손해배상 소송

실패 유형 안전 기능 실패

해외

2025.11 · 美 Figure AI

휴머노이드 충격력 '두개골 골절 수준' 경고

전 제품안전 책임자의 내부고발 소송 (사측은 부인)

실패 유형 안전 검증 누락

생성형 AI의 오류는 '틀린 답', 피지컬 AI의 오류는 '물리적 사고'

규제 시행 — 한국 AI 기본법(2026.1 시행) · EU AI Act · EU 기계규정(2027.1 적용)

→ 기능 및 성능을 넘어 안전과 신뢰를 '증명'해야 하는 상황

02. '안전'의 정의

안전의 정의와 분류

ISO/IEC Guide 51

안전 = 허용할 수 없는 위험이 없는 상태

위험이 '0'인 상태가 아니라 위험이 **'수용 가능한 수준'**인 상태 → 무엇이 수용 가능한지 논증 필요

위험의 원인으로 나눈 안전의 네 층

층	원인	대표 표준	관통 예시 (사람 감지·정지)
물리 안전	기계적 위험: 충돌, 끼임, 낙상	ISO 10218 · ISO 3691-4 · ISO 25785-1	감속이 늦어 사람과 충돌
기능안전	부품·소프트웨어의 고장	IEC 61508 · ISO 13849 · ISO 26262	카메라 고장으로 감지 실패
의도된 기능의 안전	고장이 없는데도 기능이 불충분	ISO 21448	역광에서 사람을 못 봄
AI 안전	학습된 행동의 불확실성	TR 5469 · TS 22440 · ISO/PAS 8800	처음 보는 작업복을 사람으로 인식 못함

AI는 "아무것도 고장 나지 않았는데 사고가 나는" 문제를 만든다 → 기존 기능안전만으로는 부족한 이유

03. '신뢰성'의 정의

AI에서 얘기하는 신뢰성은 "Trustworthiness"

'신뢰성'의 정의: 우리말의 함정

Reliability · 신뢰도

주어진 조건과 시간 동안 요구된 기능을 고장 없이 수행하는 능력

IEC 60050-192 · 전통적 신뢰성 공학(고장률, MTBF)

Trustworthiness · 신뢰가능성

이해관계자의 기대를 검증 가능한 방식으로 충족하는 능력

ISO/IEC 22989 · 하위 특성은 ISO/IEC TR 24028

Trustworthiness (신뢰가능성)

신뢰도 | 가용성 | 회복탄력성 | 강건성 | 안전 | 보안

프라이버시 | 투명성 | 설명가능성 | 제어가능성 | 책무성

신뢰도(Reliability)와 안전(Safety)은 모두 신뢰가능성의 하위 특성.
안전은 동시에 기능안전 표준군이라는 독자적 규범 체계를 가진다.

04. 표준 현황

피지컬 AI의 범위와 표준 성숙도

■ 공식표준 ISO·IEC 등 공적 국제표준화기구 ■ 사실표준 UL·IEEE·업계단체·연구기관의 표준·방법론

형태	물리·기능 안전 표준 공식	AI 관련 표준 공식	사실표준·방법론 사실	성숙도
자율주행	ISO 26262 · ISO 21448(SOTIF) · ISO 34502(시나리오 기반 평가)	ISO/PAS 8800	UL 4600 · IEEE 2846 · PEGASUS · ASAM OpenSCENARIO	가장 앞섬
협동로봇	ISO 10218-1/-2:2025 (TS 15066 협동 요구사항 통합)	공통 AI 표준 준용	전용 없음 → 공통 방법론 차용	물리 안전 성숙 · AI 초기
이동로봇	ISO 3691-4:2023 (3판 DIS 투표 2026.9 종료) · ISO 13482	공통 AI 표준 준용	ANSI/A3 R15.08-1/-2	중간
휴머노이드	ISO 25785-1 (CD, 2026말~2027 발행 전망) · ISO 13482	공통 AI 표준 준용	전용 없음 → 공통 방법론 차용	가장 초기
공통 AI	—	TR 5469 → TS 22440 · 24029 · 25223 · 8200	AMLAS · GSN(SCSC) · CAE	AI 기능안전 규범 확정 전

자율주행이 10년 넘게 쌓은 방법론(SOTIF · 시나리오 기반 평가 · Safety Case)이 사실표준을 통해 로봇으로 이식 중

04. 표준 현황

피지컬 AI 관련 표준 현황

■ 공식표준 ISO·IEC 등 공적 국제표준화기구 ■ 사실표준 UL·IEEE·업계단체·연구기관의 표준·방법론

트랙	~2024	2025	2026	2027 (전망)
로봇·이동 공식	ISO 3691-4:2023 ISO 13482:2014	ISO 10218-1/-2:2025 발행 (TS 15066 통합) ISO 25785-1 초안(5월)	ISO 25785-1 CD → DIS ISO/WD 26264-1(6월) ISO/DIS 3691-4 투표 종료(9월)	ISO 25785-1 발행 ISO 3691-4 3판
자율주행 공식	ISO 26262:2018 ISO 21448:2022 ISO/PAS 8800(2024)	축적된 방법론이 로봇 분야로 이식	—	—
공통 AI 공식	TR 5469(2024.1) TR 24029-1(2021) · 24029-2(2023) TS 8200(2024)	—	TS 22440-1~3 CD 등록 및 의견수렴 AWI TS 25223 PWI 26751(AI Safety)	TS 22440 발행 (상반기~중반)
사실표준·방법론 사실	UL 4600 3판(2023) · IEEE 2846(2022) R15.08-2(2023) · AMLAS · GSN v3(2021)	공식표준 확정 전까지 현장 입증의 기준 Safety Case 작성 · SPI 기반 운영 감시 · 시나리오 기반 평가		

2026~2027 = 로봇 물리 안전과 AI 기능안전 공식표준이 동시에 확정되는 구간

04. 표준 현황

피지컬 AI 관련 표준 현황

공식표준이 비어 있는 곳을, 사실표준이 먼저 채우고 있다

공백 영역	공식표준 현황 공식	메우고 있는 사실표준 사실	역할
자율 제품의 Safety Case 평가	AI 포함 시스템 Safety Case 요구 규범 미확정 (TS 22440 개발 중)	UL 4600	Safety Case 필수 요소 · SPI 기준
ML 컴포넌트 보증 절차	ISO/PAS 8800은 자동차 한정	AMLAS (영국 요크대)	단계별 논증 패턴 · 산출물 정의
논증 표기법	ISO/IEC/IEEE 15026-2는 구조만 규정	GSN(SCSC) · CAE	실무 표기 · 도구 간 호환
시나리오 기술	ISO 34502 · 34503은 자율주행 중심, 로봇 전용 없음	PEGASUS · ASAM OpenSCENARIO	시나리오 3단계 추상화 · 교환 형식
런타임 안전 규칙	행동 규칙을 수식으로 정의하는 기준 없음	Mobileye RSS → IEEE 2846	안전 거리 규칙의 형식화

연구·업계 방법론
RSS (2017)

→

사실표준화
IEEE 2846 (2022)

→

공식표준·규제로 흡수
향후 로봇 표준에서 반복될 경로

05. 표준의 공백

표준의 공백을 방법론으로 대응

VLA·파운데이션 모델 로봇의 행동 수준 평가

LIBERO, SimplerEnv 같은 벤치마크는 성능 비교용이지 안전 검증 기준이 아니다

시뮬레이션 결과의 신뢰성

sim-to-real 타당성을 어떻게 인정할지는 자동차(NATM)에서만 일부 논의 중

학습·적응하는 시스템의 재검증

운영 중 모델이 업데이트될 때 무엇을 다시 검증해야 하는지 기준이 없다

표준이 비어 있는 곳에서, 현장은 '방법론 패턴'으로 대응하고 있다

06. 현장의 방법론

Safety Case를 기준으로 연결



07. Safety Case

안전을 입증하는 방식

Safety Case란 무엇이고, 왜 AI에 필요한가

정의 · 영국 국방표준 Def Stan 00-56 (요지)

특정 운영 환경에서 특정 용도로 시스템이 안전하다는 것을, **증거로 뒷받침하여** 설득력 있고 이해 가능하며 타당하게 보여주는 **구조화된 논증**

규칙 준수 방식 (Prescriptive)

전통 소프트웨어 · V-모델

"이 시험을 통과하면 안전"이라는 전제

행동을 미리 명세하고 전수 검증할 수 있을 때 유효

요구사항을 명세하고, 커버리지 100%로 검증 가능

목표 기반 방식 (Goal-based)

피지컬 AI

"왜 안전하다고 믿을 수 있는가"를 논증

AI는 가능한 모든 행동 예측 불가 → 전수 검증 불가 →
"남은 위험이 수용 가능하다"는 논증으로 전환

07. Safety Case

구조와 표기법



두 가지 대표 표기법

요소	CAE	GSN
주장	Claim	Goal
논증	Argument (+ Warrant)	Strategy (+ Justification)
증거	Evidence	Solution
맥락	Context	Context
가정	Assumption	Assumption

CAE: Bloomfield 등(Adelard) · GSN: Goal Structuring Notation (SCSC 표준)

핵심 맥락: “어디서나 안전”이 아니라 “이 조건(ODD)에서 안전”을 주장한다

07. Safety Case

CAE 트리 예시

C0 · 최상위 주장
정의된 ODD 안에서 사람 감지 기반 감속·정지 기능은 허용 가능한 위험 수준 내에서 동작한다

맥락
ODD(조도 범위, 작업자 복장, 최대 속도 등) · 위험 수용 기준

논증 전략 · 생명주기 단계별 분해 → "각 단계에서 위험이 충분히 통제되면 전체 위험이 통제된다"

C1 · 설계	C2 · 데이터	C3 · 모델	C4 · 런타임	C5 · 운영
AI 단일 실패에 의존하지 않는 구조다	학습·검증 데이터가 ODD를 충분히 덮는다	인지 모델이 ODD 안의 변동에 강건하다	OOD 상황에서 안전 상태로 전환된다	현장 성능을 지속 감시하고 갱신한다
증거 아키텍처 분석 · FMEA/FTA · 사용 수준 분류 기록	**증거** 데이터 커버리지 분석 · 데이터 품질 지표	**증거** 형식 검증 결과 · 섭동 시험 보고서	**증거** 불확실성 모니터 시험 · 시나리오 시뮬레이션	**증거** 안전 성능 지표(SPI) · 아차사고 로그
TR 5469 / TS 22440	ISO/IEC 5259	ISO/IEC 24029	TS 25223 · 시나리오 기반 평가	UL 4600 SPI · 운영 모니터링

07. Safety Case

TS 22440(AI 시스템 기능안전)

AI 사용 수준 × 기술 등급 분류

사용 수준 · AI가 시스템에 미치는 영향

수준	의미
A1	시스템 안에서 AI가 직접 결정 (예: 로봇이 움직일지 스스로 판단)
A2	AI는 인지·분석만, 최종 결정은 기존 알고리즘
B1	개발 단계에서 AI가 코드·테스트 등 제품에 들어가는 산출물에 영향
B2	개발 단계에서 AI가 엔지니어를 보조만
C	시험·위험 저감 도구처럼 간접적 영향
D	격리되어 안전에 영향 없음

기술 등급 · 기존 표준으로 대응 여부

Class I
기존 기능안전 표준(IEC 61508 등)으로 개발·검증 가능

Class II
그대로는 안 되지만 추가 조치(데이터 품질 관리, 강건성 검증, 설명가능성 등)로 보완 가능

Class III
표준에 맞춰 개발·검토할 수 없고, 추가 조치로도 부족

두 축의 조합별 상대적 위험도를 정성 표로 제시

07. Safety Case

설계에 따른 상대적 위험도

설계 ①



사용 수준: AI가 감지 결과로 직접 정지를 결정. → A1
기술 등급: 심층신경망은 보통 Class II~III
조합: A1 / Class II or III → 기능 안전 요구 수준 높음

설계 ②



사용 수준: AI는 감지만, 정지 결정은 비AI 안전 기능. → C
기술 등급: 사용 수준 C이므로 기술 등급 무관
조합: C / Class - → 기능 안전 요구 수준 낮음(간접 영향 분석 및 관리)

같은 AI라도 아키텍처를 어떻게 짜느냐에 따라 검증 부담이 달라진다

07. Safety Case

강건성과 불확실성 평가 (ISO/IEC 24029 · TS 25223)

강건성 · ISO/IEC 24029 시리즈

방법	내용
통계적	성능 지표를 대표 시험 데이터로 측정
형식적	입력 변동 범위에서 출력이 유지됨을 수학적으로 증명 (24029-2)
경험적	현장·시뮬레이션 시험, 전문가 판단

불확실성 · AWI TS 25223 (개발 중)

종류	원인
우연적 (aleatoric)	센서 잡음·한계, 환경 변동, 사람 행동의 예측 불가능성 — 피할 수 없음
인식적 (epistemic)	학습 데이터 부족 → 학습 분포 밖(OOD) 상황, 즉 엣지 케이스

예시 A · 형식 검증

카메라 밝기·노이즈가 정해진 범위 안에서 변해도 사람 감지 결과가 **바뀌지 않음**을 증명

예시 B · OOD 대응 시험

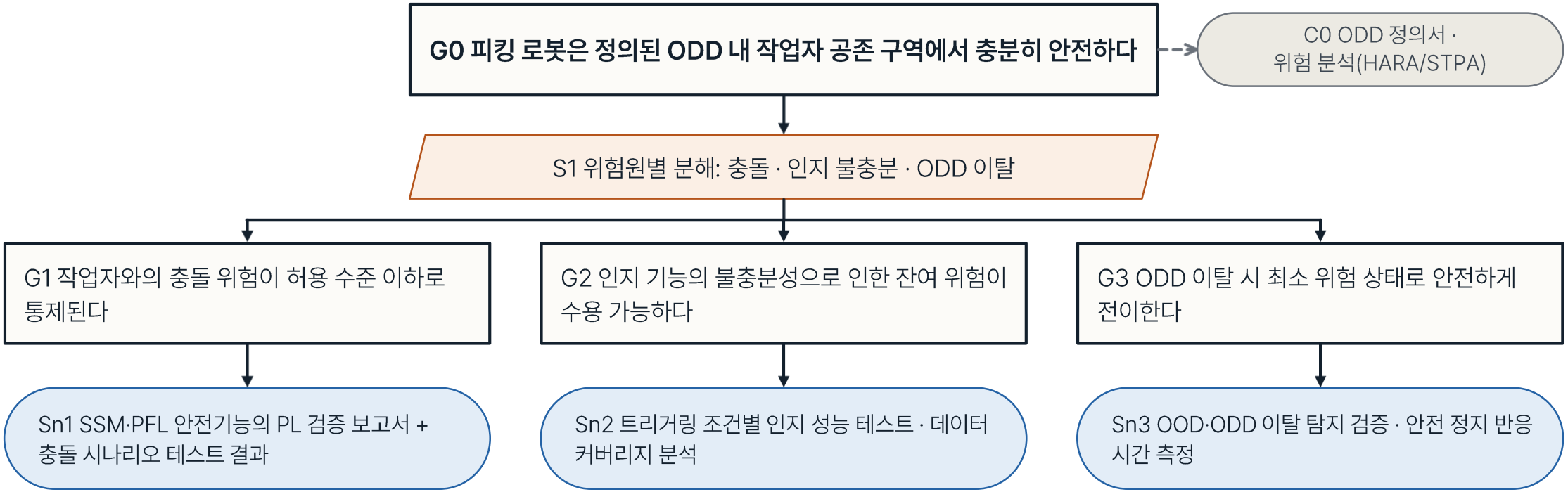
처음 보는 반사 조끼 → 불확실성 점수 상승 → **최소위험동작**으로 전환되는지 확인

"정확도 몇 %"가 아니라 "어떤 조건 범위에서 안전한가"를 보여주는 것이 신뢰성 평가의 본질

07. Safety Case

GSN 트리

주장을 하위 목표로 쪼개고, 말단에 증거를 붙인다



G 목표(주장) · S 전략(논거) · C 맥락 · Sn 증거 — GSN: Goal Structuring Notation

07. Safety Case

AMLAS(Assurance of Machine Learning for use in Autonomous Systems)

GSN을 적용한 인공지능 및 자율시스템 머신러닝 보증 프레임워크

단계 1	단계 2	단계 3	단계 4	단계 5	단계 6
ML 안전 보증 범위 설정	ML 안전 요구사항 보증	데이터 관리 보증	모델 훈련	모델 검증	운영 배포 안전 보증

무엇을 제공하나

단계마다 논증 패턴과 만들어야 할 산출물을 정의 →
ML 부분의 Safety Case를 재사용 가능한 틀로 작성

CAE와의 매핑

단계 3~5(데이터 관리 · 모델 학습 · 모델 검증)가
CAE 트리의 C2~C4 근거를 짜는 틀이 된다

각 단계별 GSN을 적용하여 명확하고 방어 가능한 안전성 주장을 생성

07. Safety Case

좋은 Safety Case의 세 가지 조건

① 증거 더미는 논증이 아니다

시험 보고서 100장을 쌓아도, 그것이 왜 주장을 뒷받침하는지 설명하지 못하면 Safety Case가 아니다.

→ 가장 흔한 실패 패턴

② 반박(Defeater)을 찾아라

확증 편향을 막으려면 "이 주장이 틀릴 수 있는 이유"를 적극적으로 찾는다

예: "야간에 반사 조끼를 입은 작업자는 감지율이 떨어진다" → 해당 조건 추가 시험을 증거로 붙이거나, ODD에서 명시적으로 제외

③ 살아있는 문서여야 한다

배치 후에도 환경이 바뀌고 모델이 갱신
→ 운영 데이터(C5)로 계속 갱신.
UL 4600의 SPI가 이 역할

신뢰성 = 이해관계자의 기대를 검증 가능한 방식으로 충족하는 능력

신뢰성(Trustworthiness)이란 결국 잘 짜인 Safety Case로 표현되는 것

08. SOTIF(Safety Of The Intended Functionality)

트리거링 조건 · ODD 분석

고장이 없어도 생기는 위험을 다룬다

구분	기능안전 · ISO 26262 / IEC 61508	SOTIF · ISO 21448
다루는 위험	고장으로 생기는 위험	고장 없이 정상 동작하는데도 생기는 위험
원인	하드웨어 랜덤 고장 · 체계적 결함	기능의 불충분성 — 설계 한계, ML 인지 오류
예시	관절 엔코더 고장으로 속도 감시가 틀어짐	센서·모델 모두 정상인데, 쪼그려 앉은 작업자를 사람으로 인식하지 못함
대응	진단 · 이중화 · SIL/PL 달성	트리거링 조건 식별 · ODD 제한 · 데이터와 모델 개선

트리거링 조건 = 기능 불충분성을 촉발하는 외부 조건 (역광, 반사 표면, 예상 밖 자세 등)

ODD(Operational Design Domain) = 자율주행 시스템이나 특정 안전 기능이 안전하게 작동하도록 설계된 운행 조건 및 환경의 범위

08. SOTIF(Safety Of The Intended Functionality)

SOTIF 4영역 모델



예시: 쪼그려 앉은 작업자

- **영역 3** 개발 단계에서 이 자세를 고려하지 못함
- **탐색** 자세·거리·조도를 바꿔 가는 탐색 테스트에서 사람 미검출 발견
- **영역 2** 트리거링 조건으로 등록, 세이프티 케이스에 반영
- **대응** 해당 자세 데이터 보강 + 인지가 불확실하면 감속
- **영역 1** 재검증 통과 → 알려진 안전

목표: 영역 2·3을 줄여 잔여 위험을 수용 기준 이하로

핵심 목적 : 알려지지 않은 위험을 찾아내 줄여 간다

08. SOTIF(Safety Of The Intended Functionality)

ODD와 트리거링 조건

ODD가 곧 책임 범위이자 테스트 공간의 경계

ODD 정의 예시 · 피킹 로봇

항목	범위 (예시 값)
바닥	평탄한 에폭시 바닥, 경사 2° 이하
조도	150 – 1,000 lux
사람	통로당 최대 3명, 안전조끼 착용 작업자 (방문객 제외)
물체	박스 · 비닐 포장 SKU, 최대 10 kg
속도	공존 구역에서 최대 1.0 m/s
운영	관제 연결 유지, 비상정지 접근 가능

트리거링 조건 식별 · 3분류

센서 한계

도크 셔터 개방 시 역광, 반사 랩 포장, 투명 PET 물체

환경

먼지 · 분진, 바닥 진동, 젖은 바닥

대상 특성

쪼그려 앉거나 사다리 위의 작업자, 휠체어 사용자, 어린이 방문객(ODD 밖)

ODD를 모호하게 정의하면 이후 모든 증거가 무엇을 입증하는지 불분명해진다 — 실무에서 가장 자주 부실해지는 단계

08. SOTIF(Safety Of The Intended Functionality)

실무 적용 절차

01

ODD 정의

시스템이 동작하도록 설계된 조건의 범위
(ISO 34503, BSI PAS 1883)

예: 바닥 · 조도 · 사람 밀도 · 물체 종류 · 작업 공간 배치

→

02

트리거링 조건 식별

센서 한계 · 환경 · 대상 특성으로 체계적으로 도출

예: 역광, 반사 랩 포장, 투명 물체, 쪼그린 작업자

→

03

수용 기준 · 검증 목표

정량 목표를 정하고, 입증에 필요한 테스트 양을 역산

예: 공존 구역 위험 상황 [목표값]/시간 이하 → 필요한 테스트 시간·시나리오 수 산출

→

04

데이터 요구사항 연결

학습·검증 데이터가 ODD와 트리거링 조건을 얼마나 덮는지 명세로 관리 (ISO/PAS 8800)

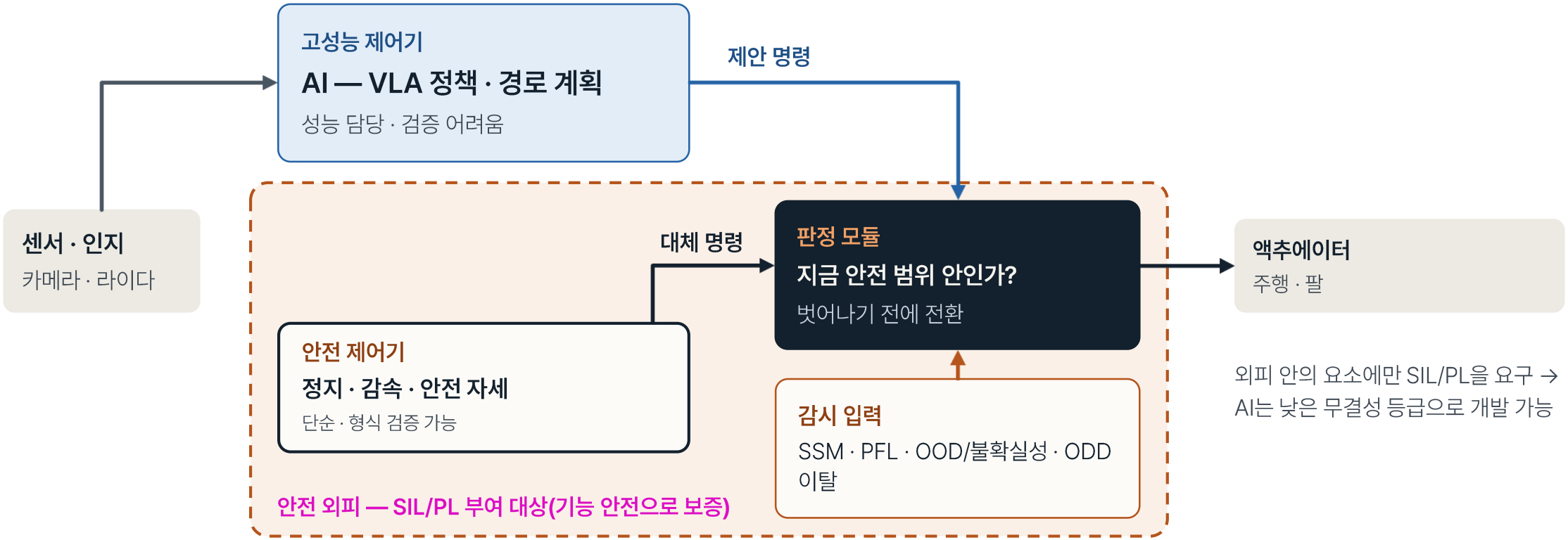
예: 저조도 × 쪼그린 자세 샘플 수를 커버리지 표로 추적

범위를 정하고, 조건을 찾고, 목표를 수치로 박는다

09. Safety Envelope

안전 외피 아키텍처(Simplex·런타임 보증)

AI를 완벽하게 검증하는 대신, 검증 가능한 단순한 안전 장치가 AI의 출력을 감시·제한하도록 설계



외피 안의 요소에만 SIL/PL을 요구 →
AI는 낮은 무결성 등급으로 개발 가능

09. Safety Envelope

구현 예시

산업 로봇
속도·이격 감시(SSM), 힘·출력 제한(PFL), 안전 레이저 스캐너 + 안전 PLC. AI가 어떤 경로를 계획하든 물리적 한계를 강제

자율주행
Mobileye의 RSS(Responsibility-Sensitive Safety) — 안전 거리 규칙을 수식으로 정의, IEEE 2846의 기반

로봇 제어 연구
제어 장벽 함수(CBF)로 안전 집합을 수학적으로 정의하고, 제어 입력이 그 안에 머물도록 실시간 보정

AI 상태 감시
OOD(분포 밖 입력) 탐지 · 불확실성 추정 · ODD 이탈 감지 → 최소 위험 상태(MRC)로 전이

피킹 로봇 예 — AI가 작업자 쪽으로 향하는 경로를 계획했을 때

VLA가 작업자 방향 경로 계획 → SSM: 경고 구역 진입 → 감속 → 보호 구역 진입 → 보호 정지 → 작업자 이탈 확인 → 작업 자동 재개

09. Safety Envelope

장점과 한계

입증 부담을 덜어 주지만, 모든 위험을 막지는 못한다

왜 실무에서 선호되나

안전 무결성 요구(SIL/PL)를 단순한 감시 계층에만 부여
→ AI 컴포넌트는 낮은 무결성 등급으로 개발할 수 있다

ISO/IEC TS 22440이 "기존 기술 기반 안전기능으로 AI 시스템의 안전을 보장"하는 경우를 별도 범주로 다룬다 = 이 패턴의 공식화

한계 1 · 보수성

안전 범위를 좁게 잡으면 AI의 성능 이점이 사라진다

예: 사람 근처에서 계속 멈추는 로봇 → 생산성 저하

한계 2 · 의미적 위험은 못 막는다

속도·거리 같은 물리량 위반은 잡지만, 과업 수준의 오류는 단순 감시기로 걸러내기 어렵다

예: 칼을 날 쪽으로 건넴, 잘못된 약을 집음 — VLA 로봇에서 커지는 문제

한계 3 · 휴머노이드의 딜레마

동적 안정 로봇은 멈추면 쓰러진다 → 정지 ≠ 안전 상태, 최소 위험 상태 설계가 어렵다

→ ISO 25785-1이 별도로 필요한 이유

10. 시나리오 기반 XIL 테스트

시나리오 기반 X-in-the-Loop 테스트

위험한 상황을 골라서 집중 테스트한다

출발점

수억~수십억 Mile

실주행만으로 안전성을 통계적으로
입증하려면 필요한 거리 (RAND 등의 분석)

→ 무작위로 많이 달리는 대신, 위험한 상황을 골라 집중 테스트하는 방식
→ 로봇에도 동일하게 적용

시나리오 3단계 추상화 — 독일 PEGASUS 프로젝트에서 정립

기능 시나리오 · 자연어 서술
작업자가 랙 사이에서 로봇 경로로 갑자기 진입한다

↓

논리 시나리오 · 파라미터와 범위
진입 속도 0.5–2 m/s · 진입 각도 0–90° · 로봇 속도 0.3–1.5 m/s · 조도 50–500 lux · 자세 {서기, 쪼그림}

↓

구체 시나리오 · 값 하나로 확정
1.4 m/s · 60° · 1.2 m/s · 80 lux · 쪼그림 → 실행 가능한 테스트 케이스 1건

10. 시나리오 기반 XIL 테스트

시나리오를 얻는 경로와 테스트 환경

세 경로로 시나리오를 모으고, 계층별 환경에서 돌린다

지식 기반

위험 분석(HARA, STPA)과 SOTIF 트리거링 조건에서 도출

예: '반사 랩 포장 + 역광' 조합 시나리오

데이터 기반

현장 로그에서 아차 사고 · 이상 상황 추출

예: 보호정지 직전 구간의 로그를 시나리오로 재구성

탐색 기반 (falsification)

최적화·강화학습으로 파라미터 공간을 탐색해 실패 지점을 능동적으로 찾음
— SOTIF 영역 3 발굴의 핵심 수단

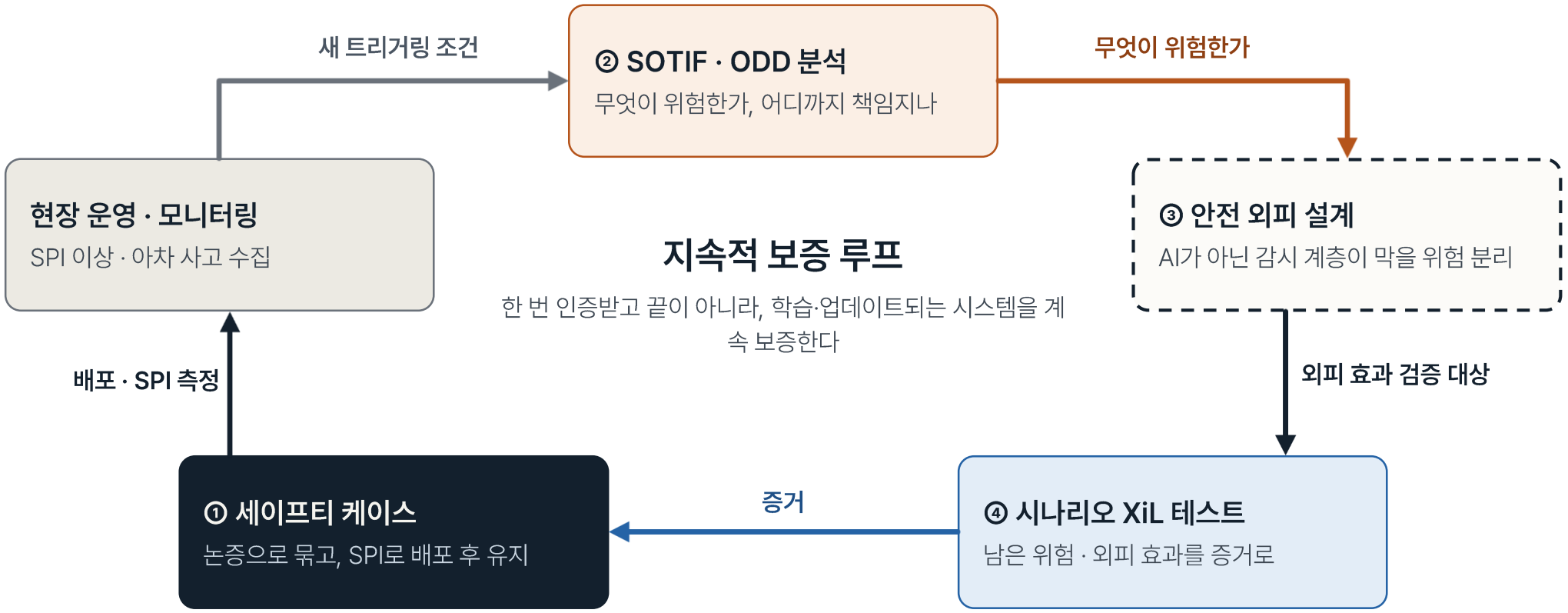
예: 최소 이격 거리가 가장 작아지는 진입 각도·속도 조합 탐색



실환경 결과로 시뮬레이션을 보정하는 순환이 중요

11. 방법론 정리

지속적 보증 루프



11. 방법론 정리

루프 사례

아차 사고 하나가 루프를 한 바퀴 도는 과정

1 · 현장 운영

특정 도크 구역에서 오후 시간대 보호정지가 늘고 (SPI), 아차 사고 1건 보고

→

2 · ② SOTIF · ODD

원인 분석 → 새 트리거링 조건 등록: 역광 + 반사 랩 포장 시 사람 오인식

→

3 · ③ 안전 외피

해당 구역·시간대 감속 규칙 추가, 인지 불확실성이 높으면 감속

→

4 · ④ 시나리오

역광 각도·포장 반사율을 반영한 시나리오 추가, 탐색 테스트 → 데이터 보강 후 회귀 테스트

→

5 · ① 세이프티 케이스

세부 목표 아래에 반론과 반박 증거 추가, SPI 기준 갱신 → 재배포

각 단계의 산출물이 다음 단계의 입력이 된다 — 이 연결이 끊기지 않아야 '지속적 보증'이다

11. 방법론 정리

현장 대응 및 남은 과제

표준의 공백		현장의 대응		남은 과제 = 다음 표준화 의제
VLA·파운데이션 모델 로봇의 행동 수준 평가	→	안전 외피로 물리량 위반 차단 (SSM · PFL · OOD 감시)	→	과업 수준 '의미적 위험' 평가 기준 (칼 방향, 약 오인 등)
시뮬레이션 결과의 신뢰성	→	XiL 계층 테스트 + 실환경 결과로 시뮬레이션 보정	→	로봇 분야 sim-to-real 타당성 인정 기준 (자동차 NATM 수준의 논의도 부재)
학습·적응 시스템의 재검증	→	지속적 보증 루프 — SPI로 Safety Case 갱신	→	모델 변경 시 재검증 범위를 정하는 변경 영향 분석 기준

방법론은 공백을 '관리'할 뿐, 메우지는 못한다 → 남은 과제가 곧 다음 표준화 의제

12. 결론

표준화 제언

01

사실표준 → 공식표준 연결

현장에서 검증된 방법론이 로봇·AI 표준에 반영되도록 의견수렴 단계부터 참여
Safety Case · SOTIF · 시나리오 기반 평가 → TS 22440 · ISO 25785-1

02

남은 과제를 표준화 의제로

행동 수준 안전 평가 · 시뮬레이션 신뢰성 · 재검증 기준
→ 국내 단체표준으로 선행 정립
→ 국제 표준으로 제안

03

표준을 시험 절차로

표준 요구사항 → 실제 시험·평가 절차로 구현 + 시험기관 역량 확보

→ 표준이 현장에서 작동

피지컬 AI의 안전과 신뢰는 성능이 아니라 증명에서 나온다.

표준은 그 증명의 문법이고, Safety Case는 그 문장이다.

10th Anniversary

GISC 2026

Global ICT Standards Conference 2026

- 감사합니다 -

정재룡 실장 / 슈어소프트테크

jyjung@suresofttech.com

ICT Standards and Intellectual Property:
Shaping the Physical AI

주최 과학기술정보통신부 지식재산처 주관 국립전파연구원 TTA IITP KEA kista ETRI