

HARA & SOTIF Guide for Autonomous Systems

A practitioner's walkthrough of how to actually run a HARA on an autonomous system — the automotive method (ISO 26262), its SOTIF counterpart for performance-limitation risk, and the robotics method (ISO 12100 + ISO 13849) for machinery and mobile robots, plus where cybersecurity threats become safety hazards. Companion to the Functional Safety & Autonomy Playbook, with its own worked examples, decision matrices, and ready-to-use worksheets.

ISO 26262-3

ISO 21448 SOTIF

ISO 12100

ISO 13849-1

ISO/SAE 21434

ANSI/A3 R15.08



Patrick Salem

Technical Program & Engineering Leader · Autonomous Systems
& Robotics

[linkedin.com/in/patricksalem](https://www.linkedin.com/in/patricksalem)

October 2026

EVERY SAFETY GOAL STARTS HERE. START IT BEFORE THE DESIGN DOES.

A 9-step map for running a HARA on an autonomous system.



AUTOMOTIVE TRACK

ISO 26262 · ISO 21448 (SOTIF)

Severity × Exposure × Controllability → ASIL



ROBOTICS TRACK

ISO 12100 · ISO 13849-1

Severity × Frequency × Avoidance → PLr

THE 9-STEP HARA MAP

1 PREPARE THE ANALYSIS

1

Define the item

Functions, boundaries, ODD

2

Fill the room

Systems, test, field data

3

Lock the method

S×E×C or S×F×P, set early

2 ANALYZE & RATE

4

Find the hazards

Guidewords × situations

5

Rate the risk

ASIL or PLr

6

Add SOTIF

Triggering conditions

3 CONNECT & DEFEND

7

Add cyber threats

TARA → HARA entries

8

Write safety goals

Free of implementation

9

Pass the review

Complete and traceable

ONE DEFENSIBLE THREAD

HAZARD → HAZARDOUS EVENT → RATING → SAFETY GOAL

EACH STEP IS DETAILED IN THE PAGES THAT FOLLOW

Starting map—ratings must come from your own ODD, data, and safety plan.

Table of Contents

Seven sections from "what is a HARA" through an assessor's checklist, four decision matrices, four ready-to-use worksheets, and a glossary — page references below.

01	What a HARA is (and isn't)	3
02	Before you start — inputs & participants	5
03	Automotive HARA walkthrough	6
04	SOTIF's version of HARA	8
05	Robotics HARA walkthrough	10
06	Cybersecurity-triggered hazards	13
07	Pitfalls & an assessor's checklist	15
	Decision matrices	16
	Worksheets (automotive, SOTIF, robotics, TARA overlap)	18
	Glossary	23

01 What a HARA is (and isn't)

A Hazard Analysis and Risk Assessment (HARA) is the systematic process of identifying the hazardous events an item can cause through its own malfunctioning behavior, and rating the risk of each one to produce a safety integrity target — ASIL in ISO 26262, Performance Level (PL) in ISO 13849, SIL in IEC 61508/62061 — and a Safety Goal that everything downstream has to satisfy. It is the single document every other safety work product traces back to: architecture decisions, verification rigor, and the safety case itself all exist to show that a Safety Goal from the HARA has been met.



The HARA process from item definition to Safety Goal, and the feedback loop that re-opens it. The rating step branches by domain — Section 03 walks the automotive ($S \times E \times C$) path, Section 05 the robotics ($S \times F \times P$) path — but the surrounding process is the same shape either way.

HARA VS. FMEA VS. FTA — THREE DIFFERENT QUESTIONS

Technique	Question it asks	Direction / level	When it runs	Produces an ASIL?
HARA	If this function misbehaves in this situation, how bad could it get, and how much rigor does that demand?	Top-down, item/function level	Before a design exists — implementation-independent	Yes — this is where it's set
FMEA	If this specific component fails this specific way, what happens?	Bottom-up, component level	After there's an architecture to analyze	No — verifies an architecture meets the ASIL HARA already set
FTA	For this known hazardous event, what fault combinations could cause it?	Top-down, quantitative	After the hazardous event is known, to verify its probability is low enough	No — verifies against a target, doesn't set one

In short: HARA decides *what counts as hazardous and how seriously to take it*; FMEA and FTA are used afterward to show the architecture you built actually earns that rating. Running FMEA before HARA — "let's see what could break, then decide if it matters" — inverts the standard's intent and usually produces a weaker safety argument that an assessor will push back on.

A NOTE ON TERMINOLOGY — HAZARD VS. HAZARDOUS EVENT

ISO 26262 draws a precise line between the two that's worth stating explicitly, because Section 03 below rates one, not the other: a **hazard** is "a potential source of harm caused by the malfunctioning behavior of the item" — on its own, with no operational context (e.g. "unintended full braking"). A **hazardous event** is that hazard combined with a specific operational situation (e.g. "unintended full braking, at high speed, in dense following traffic, on a wet motorway"). Risk is always rated at the hazardous-event level, never the bare hazard — the same hazard can carry very different Severity, Exposure, and Controllability depending on the situation it's paired with, which is exactly why Section 03's Step B (pairing with operational situations) can't be skipped or rushed.

A COMPLEMENTARY TECHNIQUE WORTH KNOWING — STPA

Classical HARA's guideword technique identifies hazards from a function's malfunctioning behavior, one function at a time. For highly automated and autonomous systems, a growing body of practice (and a number of published studies on applying it alongside ISO 26262) supplements this with **STPA (System-Theoretic Process Analysis)** — a systems-theory-based technique that models the full control structure (sensors, controllers, actuators, and the human or fleet operator) and looks for *unsafe control actions* arising from interactions between components, not just single-component malfunction. It's better suited than guideword HARA at surfacing emergent, system-level hazards — the kind that arise only from how perception, planning, and control interact, not from any one part failing on its own. STPA doesn't replace the ISO 26262-mandated HARA and doesn't produce an ASIL by itself, but running it alongside classical HARA on a complex autonomous stack is a reasonable way to catch what a function-by-function guideword pass is structurally prone to miss.

WHEN TO RUN IT

As early as the item definition is stable enough to describe its intended functions and interfaces — ideally before architecture is locked in, because the ASIL/PL a HARA produces is supposed to drive architecture decisions (how much redundancy, what diagnostic coverage, whether decomposition is worth it), not the other way around. Running HARA after the architecture exists means you're rationalizing a design rather than deriving requirements from risk.

WHAT TRIGGERS A RE-OPENING

- A change to the item definition — new function, new interface, new dependency on another system.
- An ODD change — new roads/speeds/weather, a new workspace layout, a new duty cycle or environment the original analysis didn't cover.
- An architecture change that moves a hazard's root cause across a boundary (e.g. a function moves from a dedicated ECU to a shared compute platform).
- Field or incident data showing a different Exposure or Controllability rating than assumed, or revealing a hazardous event nobody identified.
- A new variant or configuration the original HARA's assumptions don't cover.
- A cybersecurity finding (TARA) that opens a new path to an existing or new hazardous event — see Section 06.

Treat the HARA as a living document re-baselined at defined gates, not a workshop output you file away after the first pass. An assessor will ask when it was last reviewed against the current item definition — "at kickoff, eighteen months ago" is a finding, not an answer.

COMMON PITFALL

Treating the HARA as a one-time workshop deliverable instead of a living document. The most common root cause of a HARA that fails a Functional Safety Assessment isn't a wrong rating — it's a HARA that was never reopened after the item definition or ODD changed out from under it.

02 Before you start — inputs & participants

WHO NEEDS TO BE IN THE ROOM

- **Safety engineering** — facilitates, owns the rating methodology, and keeps the room implementation-independent.
- **Systems engineering** — owns the item definition and can speak to actual functions, interfaces, and boundary conditions.
- **A domain subject-matter expert** per function area in scope — perception, controls, mechanical/actuation — to keep malfunctioning-behavior brainstorming grounded in what can plausibly go wrong, not just what's easy to imagine.
- **Test or field-data representative** — brings real exposure and controllability evidence instead of pure engineering judgment; this is usually where Exposure ratings get sharpened or corrected.
- **An independent reviewer** for any hazardous event that's likely to land at ASIL C/D or PLr d/e — someone who wasn't party to the design decisions. ISO 26262 names two independence levels for this: **I2** (a person from a different team) and **I3** (a person from a different department or organization) — the confirmation review of an ASIL D HARA requires I3, and many programs use an external assessor. A peer who sits two desks away, on the same team, doesn't satisfy either level.

INPUTS YOU NEED BEFORE THE WORKSHOP, NOT DURING IT

- A baselined **Item Definition** — functional boundaries, interfaces, and dependencies on other vehicle, robot-cell, or platform systems.
- The **Operational Design Domain (ODD)** or application envelope — speed range, road type or workspace layout, weather, lighting, human proximity, duty cycle. This single input drives Exposure and Severity reasoning more than anything else in the room.
- A functional list stated at the item boundary — intended behavior, not implementation. If the list already names a sensor model or a control algorithm, push it back up a level before the workshop starts.
- Known field data or incident history from similar items — informs Exposure ratings with evidence instead of pure gut feel, and is the fastest way to catch an optimistic Controllability assumption before it becomes a finding later.

- An **assumptions log**, started now and kept open for the life of the HARA. Every assumption behind a rating — "driver is attentive," "workspace is access-controlled," "pedestrian density is low" — gets written down as it's made. An assessor's first question is usually "show me the assumptions log," not "show me the ratings."

DECIDE THE RATING METHOD BEFORE THE ROOM IS FULL

Automotive teams use the Severity × Exposure × Controllability matrix (Matrix 1). Robotics teams use the ISO 13849 risk graph (Severity × Frequency/exposure × Possibility of avoidance). Get agreement on which method applies, and on any program-specific tailoring of the rating definitions, *before* the workshop — litigating the method itself while a cross-functional room is waiting burns the most expensive resource in the room: everyone's attention.

COMMON PITFALL

Running the workshop without a field-data or test representative in the room. Exposure and Controllability ratings decided purely by engineering judgment, with no one present who has actually seen how the system behaves with real operators or in real traffic/workspace conditions, skew optimistic far more often than they skew conservative.

03 Automotive HARA walkthrough

ISO 26262-3 Cl.6

STEP A — IDENTIFY MALFUNCTIONING BEHAVIORS

For each function named in the item definition, work through a guideword-style technique adapted from HAZOP: *no function* (function doesn't activate when needed), *unintended function* (activates when not needed), *incorrect timing or duration* (too early, too late, stuck on, stuck off), and *incorrect value or direction* (wrong magnitude, wrong direction, degraded output presented as valid). Applying the same four guidewords to every function is what keeps the brainstorm systematic instead of anecdotal — a room that only discusses the failure modes someone happened to think of will miss the ones nobody thought to worry about.

STEP B — PAIR WITH OPERATIONAL SITUATIONS

Combine each malfunctioning behavior with a representative operational situation drawn from the ODD: road type, traffic density, speed band, weather, lighting. Don't try to enumerate every literal combination — that's combinatorial and unnecessary. Instead pick a small set of situations that bound the worst-case Severity, Exposure, and Controllability for that behavior (e.g. "highway, dense following traffic, high speed" for an unintended-braking behavior, rather than also separately rating "highway, light traffic" if it can't produce a worse outcome).

STEP C — WRITE THE HAZARDOUS EVENT

State it as a plain-language sentence with a consistent structure: *[malfunctioning behavior] + [operational situation] → [hazardous event]*. Keep it implementation-free — "unintended full braking with no obstacle present, on a highway in dense following traffic, causes a following vehicle to rear-end the host" is a hazardous event; "the radar false-positives on a stationary object" is a root cause, and belongs in the eventual FMEA/FTA, not here.

STEP D – RATE SEVERITY, EXPOSURE, CONTROLLABILITY

Severity (S)	Exposure (E)	Controllability (C)
S0 — no injuries	E0 — incredibly unlikely	C0 — controllable in general
S1 — light to moderate injuries	E1 — very low probability	C1 — simple to control
S2 — severe to life-threatening, survival probable	E2 — low probability	C2 — normally controllable
S3 — life-threatening to fatal	E3 — medium probability	C3 — difficult to control or uncontrollable
	E4 — high probability, most operating conditions	

These summarize ISO 26262-3's class definitions — quote your licensed copy's exact wording in your rating rationale. There is no official correspondence table between S0–S3 and clinical injury scales such as AIS; informal research mappings exist, but citing one as an ISO 26262 requirement is a factual error an assessor will catch.

STEP E – LOOK UP THE ASIL

Read the S/E/C combination off Matrix 1 in the Decision Matrices section. If different plausible operational situations for the same malfunctioning behavior would produce different ratings, rate the one that produces the *highest* ASIL — the hazardous event inherits the worst case, not the average case.

STEP F – WRITE THE SAFETY GOAL

One sentence per hazardous event (or per group of hazardous events with a shared root cause), stated at the top level, free of implementation detail, inheriting the highest ASIL among the hazardous events it addresses. "The lane-centering function shall not command a steering torque the driver cannot override within [X] seconds" is a Safety Goal; "the EPS controller shall saturate torque commands below [X] Nm" is a Technical Safety Requirement derived from it two steps later.

Worked example — multi-row HARA extract: L2 lane-centering assist

#	Malfunctioning behavior	Operational situation	Hazardous event	S	E	C	ASIL
1	Unintended steering torque toward lane boundary	Highway, adjacent vehicle in blind spot	Host clips adjacent vehicle before driver can countersteer	S3	E3	C3	C
2	No steering torque when lane departure is imminent	Highway, driver briefly inattentive	Host drifts out of lane into adjacent traffic	S3	E3	C2	B
3	Steering torque persists after driver applies override	Any speed, driver actively countersteering	Driver fights the system while attempting an evasive maneuver	S3	E2	C3	B
4	Delayed torque command (timing fault)	Highway curve, high speed	Correction arrives too late to track the curve, host departs lane	S2	E3	C2	A

Row 1 sets the governing Safety Goal's ASIL (C) — row 2's lower rating doesn't lower it; a Safety Goal inherits the *highest* ASIL among the hazardous events it covers. Ratings here are illustrative, not validated field data — your program's actual E/C ratings must come from your own ODD and controllability studies.

DOCUMENTATION STRUCTURE AN ASSESSOR EXPECTS

- **Scope & item reference** — which baselined Item Definition this HARA covers, and its version.
- **Method & rating criteria** — which matrix, which rating definitions, any program-specific tailoring, agreed before the workshop.
- **Assumptions log** — every assumption behind every rating, not just the controversial ones.
- **Hazard list** — every malfunctioning behavior × operational situation × hazardous event × S/E/C × ASIL, including the ones that rated QM — an assessor checks for completeness, not just correctness of the ones that rated high.
- **Safety goals list** — one per hazardous event or hazard group, each traced back to the hazard list entries it covers.
- **Traceability** — to the Item Definition's function list, so coverage can be checked function by function.
- **Review & sign-off record** — who rated, who reviewed independently (required for ASIL C/D), and when.
- **Change history** — every re-opening, what triggered it, and what changed as a result.

COMMON PITFALL

Rating Controllability by assuming an attentive, trained driver, when the ODD includes the general public. "The driver can simply let go of the wheel" is not a valid C1 rationale for a mass-market L2 system — it's valid, if at all, only for a system whose ODD and user population genuinely support that assumption.

Classical HARA asks what happens when something *malfunctions* — a fault. SOTIF asks a different question: what happens when nothing is broken, but the system's own performance limits or a foreseeable misuse produce a hazardous outcome anyway? A camera-based perception stack that correctly processes every frame it receives, but whose training data never included low sun-angle glare on wet pavement, isn't malfunctioning in the ISO 26262 sense — it's exhibiting a performance limitation that classical HARA's guideword technique isn't built to find.

THE SOTIF CYCLE, AND HOW IT MAPS ONTO HARA'S STEPS

Area 1 – Known / Safe

Triggering condition identified and shown acceptable. The SOTIF equivalent of a hazard analysis entry that rated QM — document it and move on.

Area 2 – Known / Unsafe

Triggering condition identified and hazardous. The SOTIF equivalent of a rated hazardous event — needs a mitigation: requirement, design change, or ODD restriction.

Area 3 – Unknown / Unsafe

Not yet identified — found later via testing, simulation, or field data. Has no classical-HARA equivalent: there's no guideword technique that finds an unknown unknown.

Area 4 – Unknown / Safe

Also not yet identified, but benign. Can't be targeted directly — managed by shrinking Area 3 and watching what falls out as actually safe.

In practice: identify triggering conditions the same systematic way you identified malfunctioning behaviors in Section 03 (sensor degradation modes, foreseeable misuse, edge-of-ODD scenarios), classify each into an Area, and for anything landing in Area 2, reduce it exactly the way you'd reduce an ASIL-rated hazard — through requirements, design, or by narrowing the ODD. The output format can mirror a HARA hazard list almost line for line; the difference is entirely in what populates the "cause" column.

WHERE SOTIF DIVERGES OPERATIONALLY

- **No ASIL.** SOTIF doesn't produce an Automotive Safety Integrity Level — a triggering condition's acceptability is argued against a defined acceptance criterion, not looked up on Matrix 1. Programs commonly reach for one of a handful of named risk-acceptance philosophies to frame that argument: **GAMAB** ("globalement au moins aussi bon" — the system as a whole performs at least as well as what it replaces, typically a human driver), **ALARP/ALARA** (risk reduced as low as reasonably practicable, balancing further reduction against cost and benefit), **MEM** (minimum endogenous mortality — bounding added risk against the baseline mortality rate from everyday hazards), and **Positive Risk Balance** (the system must demonstrably reduce overall risk compared to the baseline it replaces, not just avoid introducing new risk). None of these is mandated by ISO 21448 itself — the standard leaves the acceptance criterion to the program — but an assessor will expect you to have picked one deliberately and named it, rather than arguing acceptability ad hoc per triggering condition.

- **Continuous, not a single workshop.** A classical HARA is baselined and then re-opened on defined triggers. SOTIF analysis for an ML-based perception stack keeps finding new triggering conditions throughout the vehicle's operational life — every disengagement, near-miss, and simulation campaign is a potential source of a new Area 3 finding that needs classifying. Budget for a standing SOTIF review cadence, not a one-time analysis.
- **Ties directly to AI/ML assurance.** Where perception or decision components are ML-based, SOTIF's triggering-condition analysis and ISO/PAS 8800's data/model/verification/operation assurance activities are two views of the same underlying risk — a model's known failure modes on out-of-distribution inputs are exactly the triggering conditions SOTIF is trying to find and classify.

Worked example — SOTIF triggering-condition extract: camera-based lane detection

#	Triggering condition	Scenario	Hazardous behavior	Area	Mitigation
1	Low sun angle, wet pavement glare	Dawn/dusk, westbound highway, rain within last hour	Lane markings misclassified, false lane-departure correction	2	Confidence-based fallback + sun-angle/wetness ODD restriction
2	Lane markings absent (worn/repaving)	Construction zone, temporary markings	No reliable lane reference, function should disengage gracefully	2	Marking-confidence threshold triggers handover, not silent failure
3	Novel marking pattern (e.g. new paint scheme)	Rare, not in training distribution	Unknown — flagged only after field disengagement cluster	3	Add to training/test set; reclassify once characterized

COMMON PITFALL

Running SOTIF analysis once at a program milestone and treating it as closed. Area 3 only shrinks with a continuous field-data and simulation feedback loop — a SOTIF analysis that stops updating is a SOTIF analysis that's quietly going stale while the fleet keeps encountering triggering conditions nobody classified.

05 Robotics HARA walkthrough

ISO 12100 + ISO 13849-1

Robotics and machinery risk assessment starts from a different base document. ISO 26262 and ISO 21448 are electrical/electronic-systems standards with their own built-in risk-assessment method (HARA). ISO 13849 has no risk-assessment method of its own — it only tells you how to *design a safety-related control system* once a hazard has already been identified and a safety function has already been decided on. The identification and risk-estimation process itself comes from **ISO 12100**, the base machinery safety standard, and it's the step most commonly skipped by teams that jump straight to "what Performance Level do we need."

ISO 12100'S FOUR-STEP RISK ASSESSMENT

- **1. Determine the limits of the machine** — use limits (intended use and reasonably foreseeable misuse), space limits (reach, workspace, exclusion zones), and time limits (service life, maintenance intervals).
- **2. Identify hazards** — systematically, across the full taxonomy below, not just the obvious pinch points.
- **3. Estimate the risk** — for each hazard, combine severity of potential harm with probability of occurrence (itself a function of frequency/duration of exposure, probability of the hazardous event occurring, and possibility of avoiding or limiting harm).
- **4. Evaluate the risk** — is it acceptable? If not, proceed to risk reduction below. This is iterative: reduce, re-estimate, re-evaluate, until the residual risk is acceptable.

THE HAZARD TAXONOMY (STEP 2, IN FULL)

Mechanical

crushing, shearing, cutting,
entanglement, impact, ejection

Electrical

contact, arcing, static

Thermal

burns, scalds from hot/cold surfaces
or materials

Noise & vibration

hearing damage, whole-body or
hand-arm vibration

Radiation

laser, UV, RF emissions from
sensors or welding

Material / substance

fluids, dust, fumes, battery
chemistry

Ergonomic

awkward posture induced by the
machine's layout or pace

Environment of use

slips, trips, lighting, confined space

Combination

two hazards together producing a
worse outcome than either alone

RISK REDUCTION HIERARCHY (APPLY IN THIS ORDER, NOT IN PARALLEL)

- **1. Inherently safe design** — eliminate the hazard or reduce the risk by design: lower force or speed, remove pinch points, round edges. Always preferred; nothing downstream of this is as reliable.
- **2. Safeguarding / protective measures** — guards, light curtains, presence sensing, and safety-rated control systems. This is the step where a Performance Level gets assigned — the safety function itself (e.g. "stop the arm if a person enters the zone") needs a PLr.
- **3. Information for use** — warnings, training, PPE. The least preferred measure, valid only for residual risk after 1 and 2 have been exhausted — not a substitute for either.

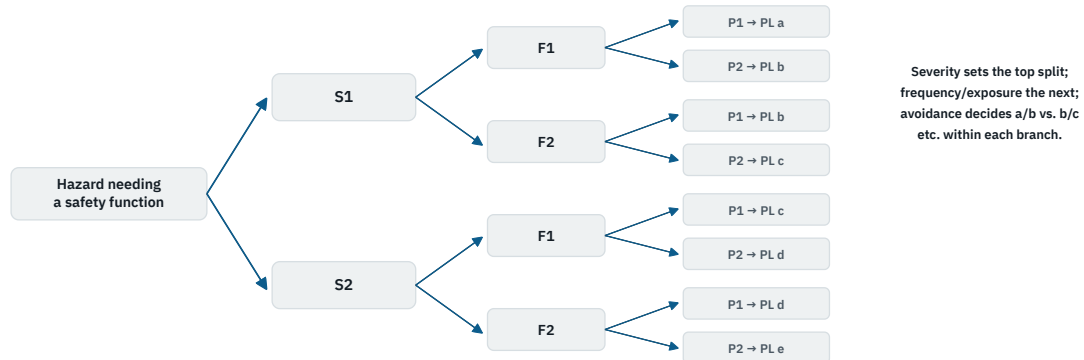
DETERMINING THE REQUIRED PERFORMANCE LEVEL (PLR) – THE RISK GRAPH

Once a hazard requires a safety function (risk reduction step 2), ISO 13849-1 Annex A's risk graph converts three binary parameters into the Performance Level that safety function must achieve:

- **Severity (S):** S1 — slight, normally reversible injury. S2 — serious, normally irreversible injury or death.
- **Frequency & duration of exposure (F):** F1 — rare to less often, and/or short exposure time. F2 — frequent to continuous, and/or long exposure time. Secondary sources quote different rule-of-thumb thresholds for defaulting to F2 when there's no other justification — some say more often than roughly once every 15 minutes, an older IFA/BGIA guideline to the standard says more often than once per hour — which tells you

the number itself isn't standardized across editions and shouldn't be treated as a precise cutoff. Check your edition's own informative guidance, and when it's close, default to F2 rather than argue for F1 on a borderline case.

- **Possibility of avoidance (P):** P1 — possible under specific conditions (hazard is slow, visible, or an escape route exists). P2 — scarcely possible (hazard is sudden, fast, or there's no practical escape).



The eight S/F/P combinations and their resulting PLr. This redraws the logical structure of ISO 13849-1 Annex A's risk graph rather than reproducing the standard's own copyrighted figure — the S1/S2 → F1/F2 → P1/P2 → five-outcome (a–e) structure shown here has been independently cross-checked against IFA/BGIA's own published guideline to the standard (the German institute that co-developed this risk-graph methodology), which confirms the same branching and the same five terminal Performance Levels. For the standard's own figure artwork and page-specific informative notes, consult your organization's licensed copy of ISO 13849-1. PLr also only tells you the safety function's *required* PL; whether a given architecture *achieves* it still depends on Category, MTTFd, diagnostic coverage (DCavg), and common-cause failure (CCF) — verify the final design against the standard's full method (the SISTEMA tool is the common way to do this), not against this summary alone.

PLr only tells you what the safety function has to achieve. Getting from PLr to a verified, achieved PL is a separate design-verification exercise — out of scope for the risk-assessment stage this guide covers, but don't let a HARA-equivalent workshop's output be mistaken for a finished safety-function design.

Worked example — multi-row risk assessment extract: warehouse AMR fleet

#	Hazard	Situation	S	F	P	PLr	Safeguard
1	Collision with pedestrian at blind intersection	Shared aisle, no right-of-way markings	S2	F2	P2	PL _e	Safety-rated LiDAR zone + hard stop, Category 3/4 architecture
2	Pinch at lift/conveyor interface	Maintenance access, infrequent	S2	F1	P1	PL _c	Interlocked guard + lockout-tagout procedure
3	Loss of localization → wrong-aisle entry	SLAM degraded by lighting change	S1	F2	P1	PL _b	Confidence-based speed reduction + geofence enforcement
4	E-stop unreachable during entrapment	Operator between robot and fixed rack	S2	F1	P2	PL _d	Redundant wireless e-stop + presence sensing, no single reachable-button dependency

AMR-SPECIFIC LAYERS ON TOP OF THE GENERIC PROCESS

ISO 3691-4 (driverless industrial trucks/AMRs) and, in the U.S., **ANSI/A3 R15.08** add hazard categories the generic ISO 12100/13849 process doesn't name on its own: travel in reverse, multi-robot interaction and traffic management, charging and battery hazards, and degraded-mode operation when a robot loses connectivity to its fleet manager. Run the generic four-step process first, then check it against these AMR-specific standards' hazard lists for gaps — don't substitute one for the other.

COMMON PITFALL

Jumping straight to "what PL do we need" without first running the ISO 12100 identification and risk-reduction-hierarchy steps. A safety-rated control system assigned PL_d for a hazard that inherently safe design could have eliminated (e.g. simply slowing the mechanism below an injury threshold) is over-engineering a problem that shouldn't have reached the control-system layer at all.

06 Cybersecurity-triggered hazards

ISO/SAE 21434 TARA

ISO 21434's Threat Analysis and Risk Assessment (TARA) identifies threat scenarios and rates each by impact (safety, financial, operational, privacy) combined with attack feasibility, producing a Cybersecurity Assurance Level (CAL, 1–4). Critically, **TARA does not assign an ASIL** — a threat scenario whose impact category is "safety" still has to flow into the HARA as a new candidate malfunctioning-behavior trigger, rated by the same S/E/C method as any other hazard. Spoofed GPS causing an AMR to leave its geofence, or a replay attack inducing unintended braking, are malicious triggers for hazardous events that the HARA — not the TARA — ultimately rates and owns.

	HARA (ISO 26262)	TARA (ISO 21434)
Question asked	How bad could this malfunction be, in this situation?	How likely is this attack, and how bad if it succeeds?
Trigger considered	Random fault or performance limitation	Deliberate, malicious action
Rating inputs	Severity × Exposure × Controllability	Impact severity × attack feasibility
Output	ASIL (QM, A–D)	CAL (1–4)
Owns the safety-impact rating?	Yes — any safety-impact threat still gets its ASIL from here	No — flags it, hands it off (Section 06)

RUNNING TARA AND HARA WITH A SHARED INTAKE POINT

- Any TARA finding tagged with safety impact gets a corresponding HARA entry — same hazardous-event format, with the trigger described as "malicious: [threat scenario]" instead of "random fault" or "performance limitation."
- Any HARA hazardous event with a plausible malicious trigger path gets a corresponding TARA entry, so a security control credited with mitigating a threat isn't double-counted against a hazard the HARA's Exposure rating already assumed was random.
- Keep CAL and ASIL as separate numbers for the same hazard where both apply — they're built on different risk models (CAL from impact × attack vector, ASIL from severity × exposure × controllability) and don't convert into each other.

Worked example — TARA/HARA overlap

THREAT SCENARIO	Spoofed wheel-speed sensor signal, injected via an exposed diagnostic port, causes the ABS/ESC controller to compute an incorrect vehicle speed.
TARA RATING	Impact: Safety. Attack feasibility: Medium (requires physical port access). CAL 3.
HARA ENTRY	Malfunctioning behavior: "incorrect vehicle speed input to ESC." Operational situation: "wet road, moderate speed." Hazardous event: "ESC applies incorrect correction, host loses lateral control." $S3 \times E2 \times C3 \rightarrow$ ASIL B.
SHARED MITIGATION NOTE	Diagnostic-port authentication (TARA control) reduces attack feasibility, not the HARA's Exposure rating — the HARA's ASIL B stands independent of whether the trigger is credibly "random" or "malicious."

COMMON PITFALL

Treating a cybersecurity mitigation as if it lowers the hazard's ASIL. A port-authentication control makes the *attack* harder, not the *underlying hazardous event* less severe or more controllable — the HARA rating and Safety Goal stand regardless of whether the trigger turns out to be a random fault or a malicious one.

07 Pitfalls & an assessor's checklist

PITFALLS WORTH REPEATING IN ONE PLACE

- Rating Controllability (or Avoidance) optimistically — assuming an attentive, trained, or physically capable operator when the actual ODD or workspace includes the general public.
- Combinatorial under-coverage — skipping operational situations that seem low-probability but still produce the worst-case Severity if they occur.
- Treating PLr (or ASIL) as a finished answer rather than a target — PLr still has to be verified as *achieved* through Category, MTTFd, DCavg, and CCF analysis; an ASIL still has to be verified as achieved through the hardware metrics in the companion playbook's Matrix 2.
- No independent review for ASIL C/D or PLr d/e ratings — the single most common Functional Safety Assessment finding.
- Freezing the HARA after the first workshop and never re-opening it against ODD, item-definition, or architecture changes (Section 01).
- Running SOTIF as a one-time analysis instead of a continuous field-data feedback loop (Section 04).
- TARA and HARA run in separate silos with no shared intake point, leaving a safety-relevant threat covered by neither document (Section 06).
- Citing an informal AIS-to-S0–S3 crosswalk as if it were an official part of ISO 26262 — no such official correspondence exists; use the standard's own S0–S3 definitions and your own clinical/engineering judgment instead.

AN ASSESSOR'S CHECKLIST

ISO 26262 verification of a HARA comes down to four named criteria — an assessor is, in effect, checking the hazard list against each of these in turn:

- **Completeness** — every function in the Item Definition appears somewhere in the hazard list, including the ones that rated QM, and every plausible operational situation per function has been considered, not just the first one someone thought of.
- **Consistency** — the same rating definitions are applied the same way across every hazardous event; an S2 in row 3 means the same thing as an S2 in row 40.
- **Correctness** — each S/E/C or S/F/P rating is technically justified against a stated source (data, standard, or named engineering judgment), not asserted.
- **Traceability** — every hazardous event traces forward to a Safety Goal, and every Safety Goal traces back to the hazardous events it covers (and inherits the highest ASIL among them, even when only one of several addresses it).

Beyond those four, an assessor will also check:

- Is the HARA explicitly versioned against a specific, baselined Item Definition?

- Is the rating method (matrix or risk graph) stated once and actually followed throughout — not switched mid-document?
- Is there an assumptions log, and does it explain the non-obvious ratings, not just restate the obvious ones?
- Is there I2 or I3 independent-reviewer sign-off (Section 02) on every ASIL C/D or PLr d/e entry?
- Is there a change history showing re-ratings tied to specific triggers (ODD change, field data, architecture change) rather than silent edits?
- For ML-based functions, is there a corresponding SOTIF triggering-condition analysis, not just a classical HARA entry?
- For functions reachable over a network or diagnostic interface, is there a corresponding TARA entry with a documented hand-off to the HARA for any safety-impact finding?

An assessor is not grading whether your ratings are generous or conservative in isolation — they're checking whether the ratings are *defensible*: traceable to a stated method, backed by a documented rationale, and consistent with the evidence (field data, test data, or an explicit, named engineering judgment) behind them. A HARA with a few aggressive ASIL A ratings and airtight rationale will pass review faster than one with uniformly high ASILs and no documented reasoning at all.

Decision matrices

Reference matrices for this guide: the ASIL determination table (shared with the companion Functional Safety & Autonomy Playbook for consistency), the SOTIF risk-acceptance framework, and how ASIL, PL, and SIL relate across domains.

☐ QM — no safety requirement ☐ ASIL A ☐ ASIL B ☐ ASIL C ☐ ASIL D — highest rigor

MATRIX 1 — ASIL DETERMINATION (ISO 26262-3, SEVERITY × EXPOSURE × CONTROLLABILITY)

Severity	Exposure	C1 (simple)	C2 (normal)	C3 (difficult)
S1 — light/moderate injury	E1	QM	QM	QM
	E2	QM	QM	QM
	E3	QM	QM	A
	E4	QM	A	B

S2 — severe/life-threatening (survivable)	E1	QM	QM	QM
	E2	QM	QM	A
	E3	QM	A	B
	E4	A	B	C
S3 — life-threatening/fatal	E1	QM	QM	A
	E2	QM	A	B
	E3	A	B	C
	E4	B	C	D

E1 = very low probability of exposure · E4 = high probability (occurs in most drives). C1 = drivers can typically control the situation · C3 = difficult or impossible to control.

MATRIX 2 — SOTIF RISK ACCEPTANCE (ISO 21448 AREAS)

Area 1 — Known / Safe

Behavior is understood and acceptable. No design action; keep in regression test suite.

Area 2 — Known / Unsafe

Hazardous behavior already identified. Must be reduced through requirements, design change, or ODD restriction before release.

Area 3 — Unknown / Unsafe

Discovered via testing, simulation, or field data. Priority target: drive toward Area 1 through added test coverage and design mitigation.

Area 4 — Unknown / Safe

Cannot be fully evaluated in advance. Managed through continuous field monitoring and scenario capture, not upfront elimination.

The SOTIF release argument is essentially: shrink Area 3 through verification until residual risk is acceptable, and instrument the fleet to keep catching what Area 4 hides.

MATRIX 3 — RISK CLASSIFICATION ACROSS DOMAINS

Domain	Governing standard	Risk class	Scale	What drives the class
Automotive / ADAS	ISO 26262	ASIL	QM, A, B, C, D	Severity × Exposure × Controllability (Matrix 1)
Machinery & robot arms	ISO 13849	Performance Level (PL)	a, b, c, d, e	Severity, frequency/exposure, and possibility of avoidance

Process & general industry	IEC 61508 / 62061	Safety Integrity Level (SIL)	SIL 1–4	Probability of dangerous failure on demand (PFD/PFH)
Cybersecurity	ISO/SAE 21434	Cybersecurity Assurance Level (CAL)	CAL 1–4	Impact severity × attack vector feasibility

There's no exact numeric conversion between ASIL, PL, SIL, and CAL — they're built on different risk models and answer different questions (safety vs. safety vs. safety vs. security). Requalify against the governing standard for each hazard rather than translating a number from one scale directly into another.

MATRIX 4 – S/E/C AND S/F/P RATING CRITERIA SIDE BY SIDE

Parameter	Automotive (ISO 26262)	Robotics (ISO 13849)
Severity	4 levels: S0 (none) – S3 (life-threatening/fatal)	2 levels: S1 (slight, reversible) – S2 (serious, irreversible or fatal)
Frequency / exposure	5 levels: E0 (incredibly unlikely) – E4 (high, most operating conditions)	2 levels: F1 (rare/short) – F2 (frequent/continuous, rule of thumb >1 per 15 min)
Third parameter	Controllability: C0 (controllable) – C3 (difficult/uncontrollable)	Possibility of avoidance: P1 (possible) – P2 (scarcely possible)
Output	ASIL via lookup table (Matrix 1)	PLr via risk graph (Section 05 figure)

Automotive uses more granular Severity/Exposure scales because the ASIL matrix spans a wider population and operating-time range; robotics collapses both to binary because the risk graph is meant to be walked quickly on the shop floor during a design review, not computed from field telemetry.

Worksheets

Four ready-to-use worksheets, one per method covered in this guide. Each starts on its own page — print it, photocopy it, or paste the structure into your own tool.

WORKSHEET 1 OF 4

Automotive Multi-Hazard HARA Worksheet

When to use it: in the cross-functional HARA workshop (Section 03) — systems, test, and field-data representatives in the room, not filled out solo at a desk. Use one sheet per function, covering every malfunctioning behavior × operational situation pair you brainstorm for it.

How to fill it in: apply the four guidewords (no function, unintended function, incorrect timing, incorrect value/direction) to the function named at the top, pair each with a representative operational situation, and rate S/E/C using the official definitions in Section 03. Look up the ASIL on Matrix 1 and carry it into the Safety Goal column.

FUNCTION UNDER ANALYSIS

ITEM DEFINITION REF. & VERSION

DATE / FACILITATOR

#	Guideword	Malfunctioning behavior	Operational situation	Hazardous event	S	E	C	ASIL	Safety goal
Ex	Unintended function	Full braking, no obstacle present	Highway, following traffic, high speed	Following vehicle can't stop in time, rear-ends host	S3	E4	C3	D	AEB shall not induce braking that following traffic can't reasonably react to
1									
2									
3									
4									
5									
6									

Run one of these per function in the item definition before calling the HARA complete — partial function coverage is the single most common gap an assessor finds.

WORKSHEET 2 OF 4

SOTIF Triggering-Condition Worksheet

When to use it: for every ML-based or performance-limited function — perception, prediction, planning — alongside (not instead of) the classical HARA worksheet above. Re-use this sheet at every standing SOTIF review, not just at program kickoff.

How to fill it in: brainstorm triggering conditions (sensor/model degradation modes, edge-of-ODD conditions, foreseeable misuse), classify each into a SOTIF Area (Matrix 2), and write the mitigation or verification action that moves an Area 2 or Area 3 entry toward Area 1.

FUNCTION / MODEL UNDER ANALYSIS

REVIEW CADENCE / DATE

SOURCE (DESIGN REVIEW / FIELD DATA
/ SIM)

#	Triggering condition	Scenario	Hazardous behavior	Area	Mitigation / verification action	Owner
1						
2						
3						
4						
5						
6						

An Area 3 entry is not a failure to document — it's the expected output of a healthy field-data loop. Track how long an entry stays in Area 3 before it's characterized; a growing backlog is the metric to escalate.

WORKSHEET 3 OF 4

Robotics Risk Assessment Worksheet (ISO 12100 + ISO 13849)

When to use it: during the machinery/robot risk assessment (Section 05) — after the machine's limits (use, space, time) are defined, and before any safety-rated control system is designed.

How to fill it in: identify each hazard against the nine-category taxonomy, note which risk-reduction step actually resolves it (inherent design, safeguarding, or information-for-use), and only fill in S/F/P/PLr for hazards that reach step 2 (safeguarding) — a hazard eliminated by inherent design has no PLr to assign.

MACHINE / CELL / ROBOT UNDER
ANALYSIS

USE LIMITS (INTENDED USE /
FORESEEABLE MISUSE)

DATE / FACILITATOR

#	Hazard category	Hazard & situation	Risk reduction step applied	S	F	P	PLr	Safeguard / rationale
Ex	Mechanical	Operator reach into pinch zone during jam clearance	2 — Safeguarding	S2	F1	P2	PLd	Interlocked guard, dual-channel monitoring
1								
2								
3								
4								
5								
6								

For AMRs specifically, cross-check this list against ISO 3691-4 / ANSI A3 R15.08's hazard categories (travel in reverse, multi-robot interaction, charging) before calling the assessment complete.

WORKSHEET 4 OF 4

TARA–HARA Overlap Checklist

When to use it: alongside both the TARA (ISO 21434) and the HARA, at the point where a threat scenario's impact rating or a hazardous event's trigger analysis is being finalized — the shared intake point described in Section 06.

How to fill it in: for each threat scenario with a safety-impact rating, confirm there's a matching HARA entry (and vice versa for any hazard with a plausible malicious trigger), and record whether the mitigation closes the TARA finding, the HARA finding, or — rarely — both.

#	Threat scenario (TARA)	CAL	Matching hazardous event (HARA)	ASIL	Mitigation closes	Owner
1						
2						
3						
4						
5						

Role	Name	Signature	Date
Safety engineering (HARA)			
Cybersecurity engineering (TARA)			

An empty row in either direction — a TARA finding with no HARA match, or a HARA entry with a plausible attack path and no TARA match — is the gap this checklist exists to surface. Don't close it out until both columns are accounted for.

Glossary

Terms used throughout this guide, spelled out once.

Term	Meaning
HARA	Hazard Analysis and Risk Assessment — the process that finds hazardous events and rates them to produce a safety integrity target and Safety Goal.
ASIL	Automotive Safety Integrity Level (QM, A, B, C, D) — ISO 26262's risk class, from Severity × Exposure × Controllability.
PL / PLr	Performance Level (a–e) — ISO 13849's risk class for machinery safety functions; PLr is the <i>required</i> level from the risk graph, PL is what an architecture is verified to actually <i>achieve</i> .
SIL	Safety Integrity Level (1–4) — IEC 61508/62061's risk class, based on probability of dangerous failure (PFD/PFH).
ODD	Operational Design Domain — the conditions (roads/workspace, speeds, weather, lighting, human proximity) a function is designed to handle.
SOTIF	Safety Of The Intended Function (ISO 21448) — risk from a system's performance limits or foreseeable misuse, not from a fault.
Triggering condition	SOTIF's unit of analysis — a specific condition (e.g. sensor degradation, edge-of-ODD scenario) that can expose a performance limitation.
Hazard	A potential source of harm caused by the item's malfunctioning behavior, considered without operational context — distinct from a hazardous event (Section 01).
Hazardous event	A hazard combined with a specific operational situation — the actual unit that gets rated $S \times E \times C$ (or $S \times F \times P$) and produces an ASIL or PLr.
STPA	System-Theoretic Process Analysis — a systems-theory hazard analysis technique that models the full control structure and looks for unsafe control actions arising from component interactions; a complement to, not a replacement for, classical HARA on complex autonomous systems.
GAMAB / ALARP / MEM / PRB	Named risk-acceptance philosophies commonly used to argue SOTIF residual-risk acceptability: "at least as good as" a baseline, reduced as low as reasonably practicable, bounded against baseline mortality, or a demonstrated net risk reduction, respectively. Not mandated by ISO 21448 — the program selects and states one.
FMEA	Failure Mode and Effects Analysis — bottom-up: for each component, what happens if it fails this way? Verifies an architecture against an ASIL/PL the HARA already set.
FTA	Fault Tree Analysis — top-down: for this known hazardous event, what fault combinations could cause it?
Severity (S)	How bad the harm could be. 4 levels in ISO 26262 (S0–S3); 2 levels in ISO 13849 (S1–S2).
Exposure (E)	ISO 26262's probability-of-operating-situation parameter (E0–E4).

Controllability (C)	ISO 26262's parameter for how likely the driver/operator is to avoid harm once the hazardous event starts (C0–C3).
Frequency / exposure (F)	ISO 13849's two-level exposure parameter (F1–F2); the frequency threshold for defaulting to F2 varies by source and edition (quoted as high as once per hour in some IFA/BGIA guidance, as low as once per 15 minutes elsewhere) — check your edition's own text rather than relying on a fixed number.
Possibility of avoidance (P)	ISO 13849's two-level parameter (P1–P2) for whether the hazard can realistically be escaped once underway.
Safety Goal	A top-level, implementation-free safety requirement derived from one or more hazardous events, inheriting the highest ASIL among them.
MTTFd	Mean Time To dangerous Failure — one of the inputs (with Category and DCavg) used to verify an achieved PL against a required PLr.
DCavg	Average Diagnostic Coverage — how much of a safety function's dangerous failures a diagnostic can detect; another PL-verification input.
CCF	Common Cause Failure — a single root cause defeating multiple redundant channels at once; must be ruled out or bounded when verifying PL.
SISTEMA	A free IFA tool commonly used to compute an achieved PL from a safety function's Category, MTTFd, DCavg, and CCF measures.
Category (B,1,2,3,4)	ISO 13849's structural requirement for a safety-related control system's architecture (e.g. single-channel vs. redundant with diagnostics) — one of the PL-verification inputs.
TARA	Threat Analysis and Risk Assessment (ISO 21434) — identifies cybersecurity threat scenarios and rates them by impact × attack feasibility.
CAL	Cybersecurity Assurance Level (1–4) — ISO 21434's risk class, from impact severity × attack vector feasibility; doesn't convert to ASIL/PL.
AIS	Abbreviated Injury Scale — a clinical injury-severity scale sometimes informally compared to ISO 26262's S0–S3, but with no official standard crosswalk.
ISO 12100	The base machinery safety standard — defines the 4-step risk assessment process and 3-step risk reduction hierarchy that ISO 13849 builds on.
AMR	Autonomous Mobile Robot — governed by ISO 3691-4 internationally and ANSI/A3 R15.08 in the U.S., layered on top of the generic ISO 12100/13849 process.
Item Definition	The baselined document describing a function's boundaries, interfaces, and dependencies — the HARA's primary input and traceability anchor.
Assumptions log	A running record of every assumption behind a HARA rating — typically the first artifact an assessor asks to see.
I2 / I3	ISO 26262's independence levels for review: I2 is a reviewer from a different team, I3 is a reviewer from a different department or organization. ASIL D HARA reviews require I3.

Companion to the Functional Safety & Autonomy Playbook — validate ASIL/PL targets and rating criteria against your program's own safety plan, tailored definitions, and OEM- or customer-specific requirements before use in a formal safety case.