Adapting ISO 21448

A Human-Centered Safety Lens for Cabin Climate Systems

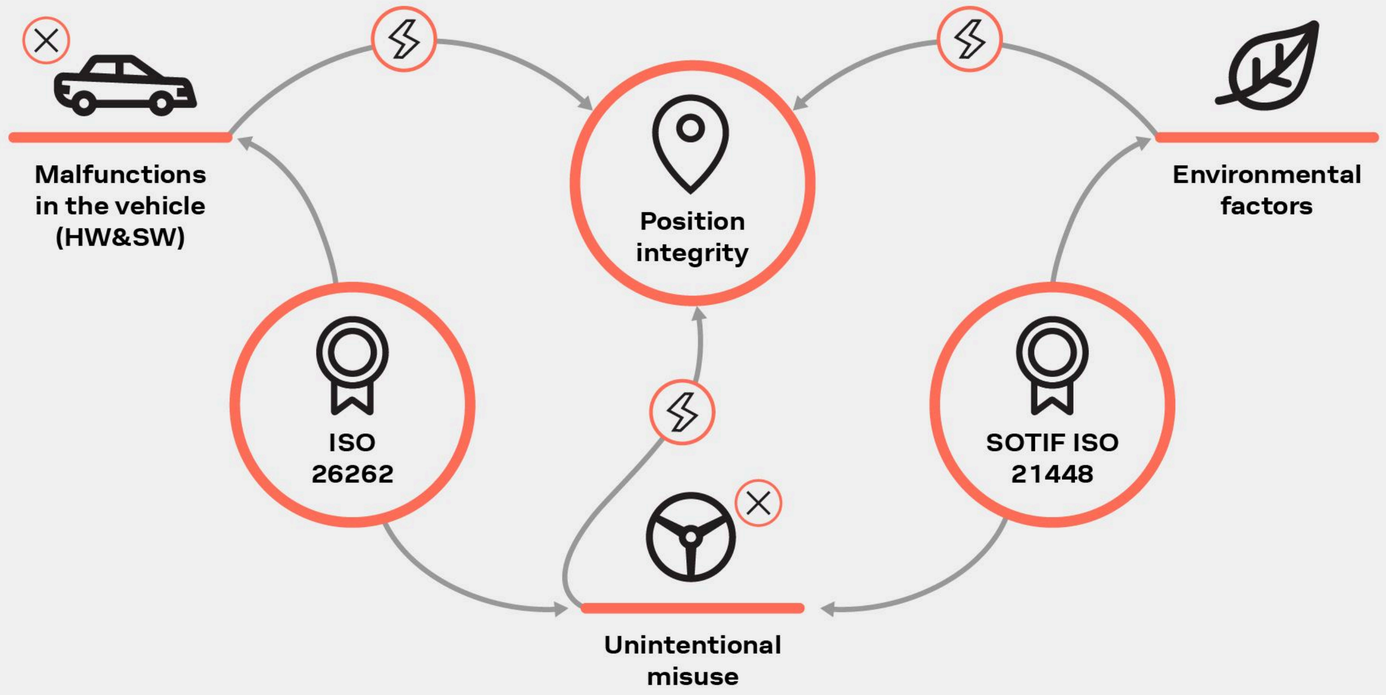
SAE International®
TMSS 2025

Contents

Context

What is ISO 21448?

The Safety of the Intended Functionality (SOTIF ISO 21448) concept was introduced when industry experts realized that ISO 26262 alone could not cover hazards that occur even when no system failure is present.



u-blox
Position integrity

ISO 21448: The two editions

Year	Document ID	Status	Key points from the scope
Jan 2019	**ISO/PAS 21448:2019** (Publicly Available Specification)	First public release	Defines SOTIF as “absence of unreasonable risk due to functional insufficiencies or reasonably foreseeable misuse.” Intended mainly for ADAS Levels 1–2 that depend on complex sensors/ algorithms providing situational awareness.
June 2022	**ISO 21448:2022** (First full International Standard)	Replaces and withdraws the 2019 PAS	Provides a complete argument framework and guidance for all levels of driving automation; becomes the authoritative SOTIF reference.

ISO/PAS 8800:2024 is a framework for assuring the safety of automotive systems that include artificial-intelligence or machine-learning elements. It tailors and extends the processes already defined in ISO 26262 (functional safety) and ISO 21448 (SOTIF) to cover risks that arise from AI-specific behavior such as performance limitations, data issues or model insufficiencies.

“ ISO 26262, ISO 21448, and ISO/PAS 8800 together cover a wide range of safety concerns – from hardware/software faults to functional insufficiencies and AI-related uncertainties, however we believe key gaps still remain.”

This raises the question -

What gaps still remain?

The meaning of “Controllability” in ISO 21448 SOTIF

ISO defines “Controllability” as the ability to avoid a specified harm or damage through the timely reactions of the persons involved (ISO 26262-1:2018, 3.25).

ISO 21448 (SOTIF) adopts all vocabulary from ISO 26262-1; Clause 3 opens with: “For the purposes of this document, the terms and definitions given in ISO 26262-1 and the following apply.”

Controllability classes used by ISO 21448 (SOTIF)

Class	ISO Definition
C0 — Controllable in general	The harm can be avoided by persons in general.
C1 — Simply controllable	The harm can be avoided with simple driver (or other person) reactions.
C2 — Normally controllable	Most drivers could act in time to prevent the harm.
C3 — Difficult to control / uncontrollable	It is difficult or impossible for people to avoid harm.

Traditional ASIL derivation under ISO 26262 assumes a predictable driver and defines Controllability (C0–C3) in terms of physical manoeuvring capability. However, in intelligent, AI-regulated thermal systems, affective and cognitive human states become central to safety outcomes.

Key Human Factors to consider

1. ISO Standard insufficiently account for human behavioral factors during Hazard Analysis and Risk Assessment (HARA).
2. Misalignment between perceived comfort and AI-driven system output can induce distraction or discomfort, posing latent safety risks.

Factors to consider

1. Cognitive overload
2. Irritability
3. Expectation management

Use Case

Consider a student driver with a rental car in a new environment.

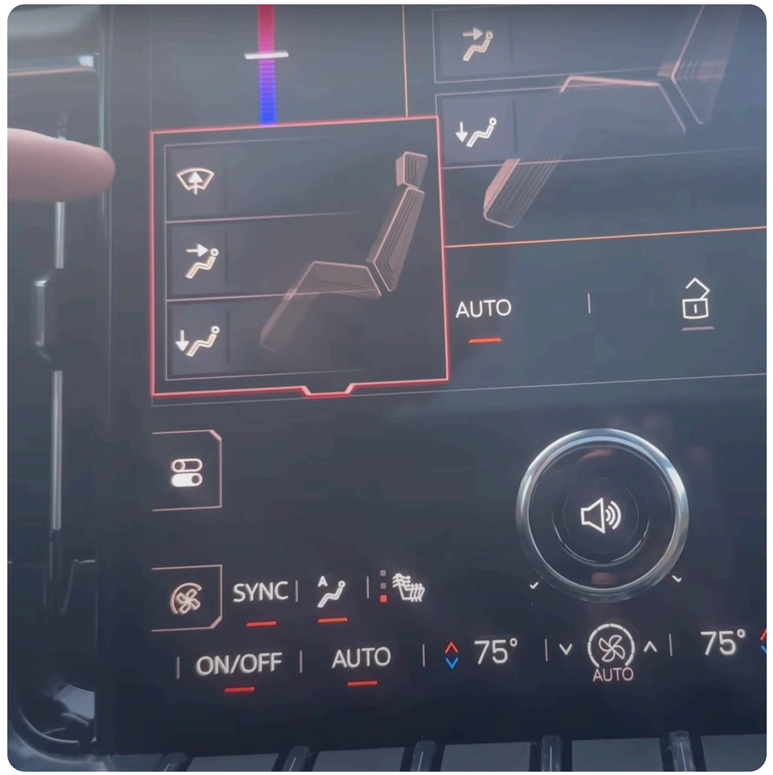
Human factors to consider

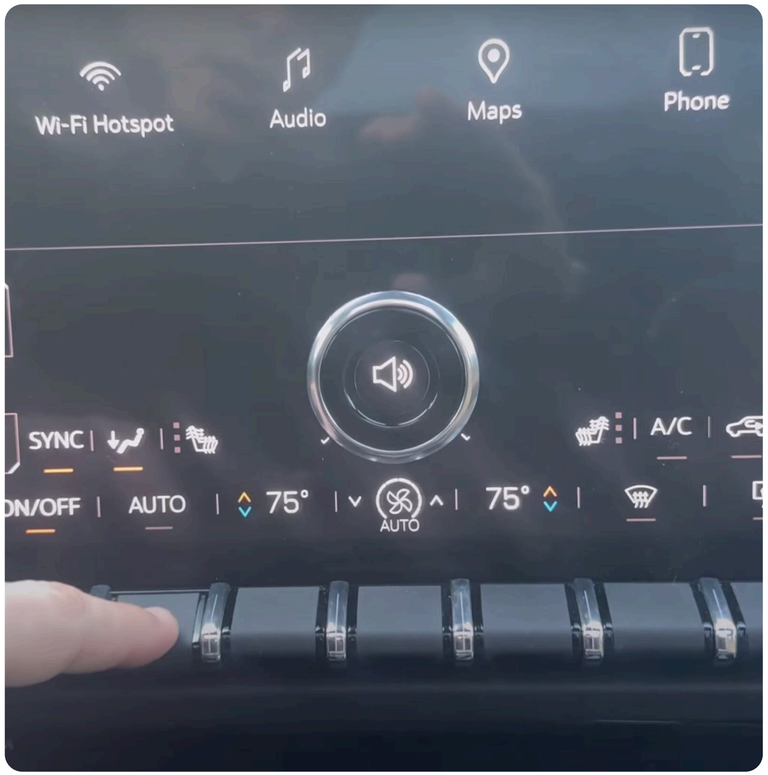
1. Age of the driver
2. Confidence in navigating new environments
3. Familiarity with the car
4. (Thermal) comfort in new surroundings
5. Alien setting (all cars are different, no standardization)

2025 Mercedes C300

2025 Genesis G80

Thermal controls in the GMY Yukon Denali. There are 3 options to control thermal systems - a) Full touchscreen; b) Focused touchscreen controls; and c) Physical buttons. This can get very confusing and increase cognition overload for users.





After accounting for the relevant human factors, does traditional controllability implicitly assume that all drivers will respond predictably and maintain control?

Introducing an additional layer of metric

“Degree of Perceptual Controllability” (DPC)

Degree of Perceptual Controllability (DPC)

1. We propose "DPC" as a qualitative behavioural augmentation metric of ISO’s Controllability scale (C0–C3).
2. This includes expectation, fidelity (trust) and the emotional response to system behaviour.

The framework

1. The “Degree of Perceived Controllability” (DPC) is a contextual, dynamic reflection of how a human operator perceives their control over an intelligent system response: based on expectation, affect, and cognitive load.
2. We define DPC = f(EO, B-Level, CL-Level).
3. Assign DPC to Controllability in ISO 21448.

Mapping to compute DPC

Expectation Outcome (EO)

E0 - Exceeded Expectation, Positively Astonishing

E1 - Met as Expected

E2 - Not Met, Frustrating but understandable

E3 - Not Met, Negative surprise

Behavioral Response Level (B-Level)

B0 - Comfortable and confident delegation to AI

B1 - Attempts workaround or override

B2 - Low trust, anxious override, confused

B3 - No perceived control; panic or helplessness.

Cognitive Load (CL-Level)

CL0 - No cognitive effort

CL1 - Minimal effort, low attention

CL2 - Moderate effort; requires short reasoning

CL3 - Overwhelming effort; cognitively paralyzing Or requires multitasking or risk calculation.

Table for Degree of Perceived Controllability (DPC)

		Cognitive Load (CL-Level)		
Behavioural Response (B-Level)	**Expectation Outcome (EO)**	**CL1**	**CL2**	**CL3**
B1	EO1	DPC0	DPC0	DPC0
	EO2	DPC0	DPC0	DPC0
	EO3	DPC0	DPC0	DPC1
B2	EO1	DPC0	DPC0	DPC1
	EO2	DPC0	DPC0	DPC1
	EO3	DPC1	DPC1	DPC2
B3	EO1	DPC1	DPC1	DPC2
	EO2	DPC1	DPC1	DPC3
	EO3	DPC1	DPC2	DPC3

Degree of Perceived Controllability (DPC)

DPC Level	Interpretation
DPC0	**Highly Controllable Perception:** AI enhances comfort, trust, and system acceptance.
DPC1	**Acceptable with Transparency Needs:** User remains onboard but desires more guidance or explainability.
DPC2	**Perceived Loss of Control:** AI decisions trigger user rejection, override attempts.
DPC3	**Breakdown of Human-System Interaction:** Leads to system shutdown, refusal, or avoidance.

=

Controllability Level	Controllability definition as per ISO
C0	Driver can always avoid the harm
C1	Driver can avoid the harm with routine actions
C2	Difficult for a typical driver to avoid harm
C3	Driver cannot reasonably avoid harm

Team members

Shradha Shejul

First Speaker

MOBIS Technical Center of North America

sshejul@umich.edu

+1 (734) 358-9619

Shrey Maheshwari

InterspectAI

+1 (215) 436-7603

shreym3110@gmail.com

Varad Vishwarupe

Ph.D. (pursuing), University of Oxford

+91 97645 47057

varad44@gmail.com

Thank You