Check for updates

Received: 12 March 2026 | Revised: 10 July 2026 | Accepted: 11 August 2026

DOI: 10.4218/etrij.2026-0204

ORIGINAL ARTICLE

ETRI Journal WILEY

STPA-based analysis of driver misuse risk scenarios in level 3 automated driving

Yangkoo Lee[1] | Eun Hye Jang[1] | Mi Chang[1] | Kezia Amanda Kurniadi[1] | Jiwoo Han[1] | Woojin Kim[1] | Daesub Yoon[2]

[1]AI Robot UX Research Section, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea

[2]AI Robot & Mobility Research Division, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea

Correspondence
Yangkoo Lee, AI Robot UX Research Section, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea.
Email: yk_lee@etri.re.kr

Funding information
Korea Institute for Advancement of Technology, Grant/Award Number: 20018248; Ministry of Trade, Industry and Energy; Korea Evaluation Institute of Industrial Technology

Abstract

This paper proposes an analytical framework that integrates the SOTIF perspective with System-Theoretic Process Analysis (STPA) to systematically analyze driver misuse in Level 3 automated driving environments. In Lv.3 systems, drivers must intervene when Takeover Requests (TOR) occur, making misuse a critical safety risk even when the system operates as intended. Driver misuse is modeled as inadequate control actions within the control structure, and the four psychological states that induce misuse—overtrust, distrust, distraction, and interference—are treated as the causal factors of those unsafe control actions. Following the STPA procedure, 41 causal scenarios were derived and evaluated using severity, likelihood, and controllability. Clustering analysis was then applied to identify recurring risk patterns and representative Risk Archetypes. The results show that many high-risk scenarios are strongly related to drivers' failures to perceive or properly respond to TOR. The proposed framework provides a structured basis for understanding driver misuse risks and supports the design of driver monitoring, improved HMI strategies, and multilayer safety assistance mechanisms.

KEYWORDS

driver misuse, Level 3 automated driving, risk pattern analysis, safety of the intended functionality (SOTIF), system-theoretic process analysis (STPA)

1 | INTRODUCTION

Automated driving technology has been rapidly advancing toward minimizing direct driver intervention while improving driving efficiency and user convenience, and ultimately enhancing the overall safety of road transportation systems [1]. Such technological progress is clearly reflected in the varying levels of technical maturity across automated driving stages and the corresponding changes in the driver's role. In particular, within the internationally recognized classification framework for automated driving levels, as the level increases, the scope of dynamic driving tasks performed by the vehicle system expands, while the role that must be directly executed by the driver is progressively reduced [2, 3]. In this transition, automated driving systems move beyond simple driver assistance functions and begin to operate as the primary agent of driving under specific conditions.

This is an Open Access article distributed under the term of Korea Open Government License (KOGL) Type 4: Source Indication + Commercial Use Prohibition + Change Prohibition (http://www.kogl.or.kr/info/licenseTypeEn.do).
1225-6463/$ © 2026 ETRI

Level 3 (Lv.3) conditional automated driving is defined as a stage in which the system performs all aspects of the Dynamic Driving Task (DDT) within a specified Operational Design Domain (ODD). In Lv.3 environments, the driver is relieved from continuous vehicle control activities such as steering, acceleration, and braking and is permitted to engage in Nondriving-Related Tasks (NDRT) while the system manages the driving operation. However, this degree of freedom is not unlimited. When the system reaches a situation in which it can no longer sustain automated driving, a Takeover Request (TOR) is issued, and the driver must immediately respond by assuming the role of a fallback-ready user. This characteristic fundamentally distinguishes Lv.3 automation from Level 2 systems, in which steering operation and continuous monitoring of the driving environment are required. In Lv.3, the driver's gaze diversion and attentional disengagement are structurally permitted, thereby introducing new forms of human-system interaction challenges.

In Lv.3 environments, even when the system operates as intended in accordance with its design specifications, there remains a significant possibility that severe accidents may occur if the driver fails to perceive a Takeover Request (TOR) or does not assume control at an appropriate moment [4–7]. As the level of automation increases, drivers tend to disengage from the driving supervision loop, and such an out-of-the-loop condition has been reported to result in degraded situation awareness and delayed responses in emergency situations [8, 9]. These risks are not limited to simple operational errors or system malfunctions, but rather manifest as structural problems arising from interaction failures between the driver and the automated system.

Such risks fall outside the scope of the functional safety standard ISO 26262, which defines safety primarily in relation to hardware or software failures [10]. Accordingly, ISO/PAS 21448, namely, the Safety of the Intended Functionality (SOTIF) standard, was established to systematically address nonfault-based risks that may arise even when intended functions operate correctly [11]. SOTIF defines system performance limitations, uncertainties in the operational environment, and foreseeable driver misuse as principal risk factors, and explicitly identifies the driver's cognitive and behavioral characteristics as variables that must be considered during the design phase.

From this perspective, the Level 3 automated driving environment embodies a dual-demand structure that permits the driver to disengage from the driving task while simultaneously requiring immediate intervention at specific moments. This structural characteristic inherently contains vulnerabilities through which driver misuse may propagate into accidents. Therefore, in Lv.3 environments, driver misuse should not be treated merely as an isolated behavioral error, but rather defined as a risk factor emerging within the system control structure, requiring a systematic and structured analytical approach.

Based on this problem recognition, this study applies the System-Theoretic Process Analysis (STPA) methodology to systematically analyze potential risk scenarios arising from driver misuse in Level 3 automated driving environments [12, 13]. STPA is a methodology that analyzes complex systems not in terms of component failures, but from the perspective of control structures and interactions, thereby enabling the structural identification of diverse risk factors that may emerge in environments where humans and systems are integrated.

In applying STPA, this study places the driver as the primary controller of the control structure and analyzes driver misuse at the levels of Unsafe Control Actions (UCA) and their causal factors, while system-level hazards are defined as hazardous vehicle states in accordance with the standard STPA convention. To interpret the mechanisms of misuse from a SOTIF perspective, the psychological states that induce driver misuse are classified into four categories—overtrust, distrust, distraction, and interference—and are treated as the causal factors of UCAs rather than as elements of the UCA definitions themselves. The main contribution of this study lies in integrating these driver psychological states into the STPA-based derivation of misuse-related UCAs and causal scenarios, whereas SOTIF, STPA, risk scoring, and clustering have each been established individually in prior work. Based on this integration, 41 causal scenarios are derived, quantitatively assessed, and statistically clustered to identify representative and high-impact risk patterns, thereby proposing an analytical framework for effectively addressing driver misuse in Lv.3 automated driving environments.

2 | RELATED WORK

The assurance of safety in automated driving systems has traditionally been discussed based on the functional safety standard ISO 26262, which focuses primarily on failures of electrical and electronic systems. However, as automated driving technologies have advanced and artificial intelligence-based perception and decision-making functions have been extensively introduced, limitations in its scope of applicability have gradually become apparent. Specifically, even in the absence of explicit faults in system components, accidents may increasingly occur due to uncertainties in the external environment, limitations in perception performance, and interaction failures between the driver and the system.

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

To systematically address such nonfault-based risks, ISO/PAS 21448, namely, the Safety of the Intended Functionality (SOTIF) standard, was established [11], defining system performance limitations, operational environmental complexity, and foreseeable misuse as primary considerations [11, 14]. Prior studies have applied the SOTIF concept to automated driving systems through sensor perception performance validation, analysis of AI-based object recognition limitations, and risk assessments using simulation environments. While these approaches have expanded the technological risk domains faced by automated driving systems, the primary analytical focus has largely remained on system performance or environmental conditions, with driver misuse addressed only in a limited manner as one of several contributing risk factors [9, 15].

System-Theoretic Process Analysis (STPA) is a system-theoretic hazard analysis method proposed by Nancy Leveson to address such limitations [12]. Unlike traditional approaches that interpret accidents as the result of individual component failures or isolated human errors, STPA explains accidents as outcomes of inadequate control across the system and failures in interactions among system components. In the field of automated driving, STPA has been utilized to model vehicle control structures, analyze potential errors in sensor and control signals, and derive safety requirements [16, 17]. Some studies have combined STPA with SOTIF to analyze risks arising from interaction failures, incorporating the driver as a component within the control structure [18]. However, even in these prior studies, the primary focus has largely been placed on technical control structures and the derivation of system requirements, with insufficient structural explanation of how driver behavioral characteristics or psychological states transition into Unsafe Control Actions and subsequently propagate into accident risks.

Driver misuse in automated driving environments has been extensively addressed in Human Factors research [19, 20], investigating HMI design, warning system effectiveness, driver distraction management, and response time in transition situations. Driver trust has been conceptualized as a calibration issue between actual system performance and the user's perception, with excessive trust leading to overreliance and distrust resulting in unnecessary intervention [21–23].

In summary, research on automated driving safety has evolved along three primary directions: technical risk analysis from the SOTIF perspective, structural accident analysis based on STPA, and studies of driver misuse from a Human Factors perspective. However, these lines of research have largely been conducted independently, with insufficient attention given to integrating the SOTIF perspective with STPA-based structural analysis while positioning driver misuse as a central risk factor in Level 3 conditional automated driving environments. Building upon this research gap, the present study establishes driver misuse as the primary subject of analysis and systematically derives structural scenarios through which misuse propagates into inadequate control actions and accident risks by applying STPA, proposing a more concrete and practically applicable risk analysis framework for SOTIF-based safety design.

3 | STPA-BASED ANALYSIS OF DRIVER MISUSE RISK SCENARIOS IN LEVEL 3 AUTOMATED DRIVING

This section analyzes driver misuse in Level 3 (Lv.3) automated driving environments using the System-Theoretic Process Analysis (STPA) procedure. Rather than treating misuse as an isolated behavioral error, the analysis interprets it as inadequate control within the driver–system interaction structure that can drive the vehicle into hazardous system states, and systematically derives the related risk scenarios.

3.1 | Definition of analysis target and scope

The subject of analysis is driver misuse in Level 3 (Lv.3) conditional automated driving environments. The scope focuses on control transition situations within the Operational Design Domain (ODD), including the configuration of control transition settings, the occurrence of and response to Takeover Requests (TOR), the interpretation of warning information, and driver responses during manual takeover.

Foreseeable driver behaviors such as delayed TOR response, misinterpretation or disregard of warnings, improper control takeover, and temporally inappropriate intervention are defined as misuse. These behaviors are considered expected user responses during system use rather than intentional violations. Risks arising from hardware or software faults are excluded from the scope of this analysis.

In the STPA application, the driver is deliberately positioned as the primary controller of the control structure, and driver misuse is analyzed at the levels of Unsafe Control Actions and their causal factors. Accordingly, the four psychological conditions that induce misuse—overtrust, distrust, distraction, and interference—are used as causal factors explaining why unsafe control actions occur, not as components of the hazard or UCA

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

definitions. System-level hazards are correspondingly defined as hazardous states of the vehicle, consistent with the standard STPA convention.

3.2 | Definition of losses and system-level hazards

In the first step of the STPA analysis, the potential losses associated with driver misuse and the corresponding system-level hazards and safety constraints are defined. Losses were defined as human fatality or personal injury, loss of or damage to the vehicle, injury to persons or damage to objects outside the vehicle, and loss or degradation of driver trust in the automated driving system.

Consistent with the STPA definition, a system-level hazard is specified in this study as a state of the vehicle that, under worst case environmental conditions, can lead to a loss, rather than as a driver behavior or a perception failure. On this basis, four system-level hazards were identified: loss of a longitudinal safe distance from a preceding or following object, departure from the lane or loss of a safe lateral path, entry into a dynamically unstable state beyond the stable operating envelope, and continuation of automated driving mode without any agent exercising appropriate control in a situation requiring manual control, that is, a state in which control authority is not established. These hazards may arise even in the absence of any hardware or software fault. Driver perception failures and psychological states such as overtrust, distrust, distraction, and interference, which the previous formulation embedded at the hazard level, are not treated as hazards in this study; instead, they are analyzed downstream as unsafe control actions and their causal factors in Sections 3.4 and 3.5, preserving the traceability from hazardous system states to operational unsafe control actions and their psychological causes.

Each hazard was paired with a top-level safety constraint expressed as a system obligation to prevent the vehicle from entering the corresponding hazardous state. Table 1 summarizes the relationships among losses, hazardous system states, and safety constraints and serves as the reference structure for the subsequent analysis.

TABLE 1 Hierarchical relationship among defined losses, system-level hazards, and corresponding safety constraints in the Lv.3 automated driving environment.

Category	ID	Description
Loss	L-1	Human fatality or personal injury
	L-2	Loss of or damage to the vehicle
	L-3	Injury to persons or damage to objects outside the vehicle
	L-4	Loss or degradation of driver trust in the automated driving system
System-level hazards	H.1	The vehicle fails to maintain a longitudinal safe distance from a preceding or following object. [L-1, L-2, L-3]
	H.2	The vehicle departs from its lane or fails to maintain a safe lateral path. [L-1, L-2, L-3]
	H.3	The vehicle exceeds its stable operating envelope and enters a dynamically unstable state. [L-1, L-2, L-3]
	H.4	In a situation requiring manual control, the vehicle remains in automated driving mode without any agent exercising appropriate control, i.e., control authority is not established. [L-1, L-2, L-3]
System-level constraints	SC.1	The system shall ensure that the vehicle always maintains a longitudinal safe distance from a preceding or following object.
	SC.2	The system shall ensure that the vehicle maintains a safe lateral path and does not depart from its lane.
	SC.3	The system shall ensure that the vehicle does not exceed its stable operating envelope or enter a dynamically unstable state.
	SC.4	The system shall ensure that the vehicle does not remain in automated driving mode without established control authority when manual control is required.

3.3 | Control structure modeling

In the second step of the STPA analysis, the control structure of the system was modeled to analyze how driver misuse may propagate into control failures in Lv.3 conditional automated driving environments. The driver was defined as the primary controller of the system, and the interactions among the driver, the automated driving system (ADS), the vehicle, and the driving environment were configured as a hierarchical control structure, reflecting the premise that the driver ultimately bears responsibility for safety in situations where a control transition is required.

In the control structure model, the driver receives information regarding the ADS, vehicle status, and driving environment as feedback delivered through the HMI, and based on this information, perceives the situation, makes judgments, and generates control commands. The driver's control actions are directed toward the ADS or,

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for direct maneuvers such as steering, acceleration, and braking, toward the vehicle, and the outcomes are fed back to the driver through the HMI and the observable vehicle behavior. In particular, warning and TOR information are delivered unidirectionally from the ADS to the driver through the HMI and are represented as feedback rather than as driver control actions. This separation makes the driver's automation-state awareness dependent on the HMI feedback channel, so that omissions, delays, or distortions in this channel may induce mode confusion or an incomplete transition of control authority, even when the driver's own control actions are executed appropriately.

A total of nine primary control actions performed by the driver were defined and grouped into three categories: transition-related configuration (control transition setting and the multichannel warning and warning-repetition guidance settings), continuous driving operations (steering, acceleration, braking, and manual system control), and responses to system notifications (TOR response and emergency mode response). Here, control transition setting denotes the driver's preconfiguration of the conditions and mode governing the handover of control authority between the ADS and the driver; it is distinct from TOR response, which is the driver's act of accepting a takeover request and initiating the transition of control authority, and from manual system control, which is the driver's sustained operation of the vehicle after the transition has been completed. Figure 1 presents the identified primary control actions and the hierarchical control structure among the driver, the ADS, the vehicle, and the environment. Based on this integrated control structure model, the subsequent steps identify potential Unsafe Control Actions for each control action and analyze the logical pathways through which such control failures propagate into system-level hazards and accidents.

3.4 | Identification of Unsafe Control Actions (UCA)

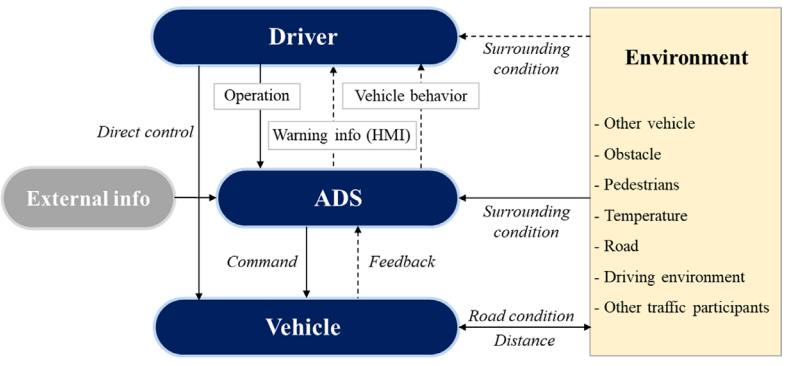
In the third step of the STPA analysis, Unsafe Control Actions (UCA) that may propagate driver misuse into system-level hazards were identified based on the control structure and primary control actions defined in the previous stage. The four standard UCA types proposed in STPA were applied to the nine primary control actions [13]: cases in which a control action is not provided and thereby causes a hazard (Not Providing Causes Hazard), cases in which a control action is provided but leads to a hazard (Providing Causes Hazard), cases in which the timing or order of the control action is inappropriate (Too Early, Too Late, and Out of Order), and cases in which the duration of the control action is inappropriate (Stopped too Soon, Applied too Long).

The four UCA types were applied nonuniformly according to the operational nature of each control action. For the continuously operated control actions (CA-1 through CA-4 and CA-9), the duration-related type was subdivided into "applied too long" and "stopped too soon," yielding five UCA entries per control action. For the discrete, event- or setting-type control actions (CA-5 through CA-8), the "stopped too soon" case does not constitute a physically meaningful misuse, because these actions are discrete handover or on/off acts; the corresponding entries were therefore explicitly recorded as not applicable (N.A) with a documented rationale, yielding four valid UCA entries per control action. This procedure produced a total of 45 UCA entries, of which 41 are valid UCAs and 4 are N.A entries.

Each UCA was defined in a structured form linking the control action, the UCA type, the operational context in which the action becomes unsafe, and the associated system-level hazard. In contrast to the original

Control Action	
Operation	CA-1 Control transition setting
	CA-2 Steering control
	CA-3 Acceleration control
	CA-4 Braking control
	CA-5 TOR response
	CA-6 Multi-channel warning setting
	CA-7 Emergency mode response
	CA-8 Warning repetition guidance setting
	CA-9 Manual system control

(A)



(B)

FIGURE 1 (A) Nine primary driver control actions defined as analytical units for Unsafe Control Action (UCA) identification, (B) Hierarchical control structure model representing interactions among the driver, automated driving system, vehicle, and environment.

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

formulation, the driver's psychological states—overtrust, distrust, distraction, and interference—were deliberately excluded from the UCA definitions and are treated as causal factors in the causal-scenario analysis of Section 3.5, consistent with the STPA distinction between an unsafe control action and the reasons why it occurs. For the warning-related setting actions (CA-6 and CA-8), the link to H.4 operates through a mediated pathway: An inappropriate warning-channel or warning-repetition configuration deprives the driver of the feedback required to respond, so that no agent establishes effective control when manual intervention is required. Table 2 presents the complete set of UCA definitions for the nine control actions together with their associated hazard links. The 41 valid UCAs serve as the input for the causal scenario analysis conducted in the subsequent step.

3.5 | Derivation of causal scenarios

Based on the 41 valid Unsafe Control Actions defined in the previous stage, this step derives the causal scenarios through which driver misuse propagates into actual accident risks. In STPA, a causal scenario explains why an unsafe control action may occur; in this study, the four driver psychological states—overtrust, distrust, distraction, and interference—are treated as the causal factors of interest. To examine the candidate space exhaustively rather than selectively, each of the 45 UCA entries (41 valid UCAs and 4 N.A entries) was combined with the four psychological states, producing a grid of 180 candidate combinations that was then evaluated case by case.

Each combination was evaluated against two refinement criteria, applied in addition to the structural N.A category. First, combinations in which the driver misuse could not be linked to a violation of the defined safety constraints were removed (18 combinations). Second, combinations converging to the same risk mechanism were integrated into a representative scenario (105 combinations). Convergence was determined by an element-wise comparison of the three outcome elements of each combination—the hazardous behavior, the linked hazard (H.1–H.4), and the resulting harm. Two combinations differing only in the psychological state were judged convergent when all three elements are substantively identical, and nonconvergent when any one of the elements differs. The disposition of every combination under this rule is documented in the 180-combination grid provided as Supporting Information. The 16 combinations inherited from the N.A entries in Table 2 were excluded as structurally not applicable, independently of the two refinement criteria. Through this process, 41 causal scenarios directly associated with violations of the safety constraints were finally derived from the 180 candidate combinations ($180 - 16$ N.A $- 18$ unlinked $- 105$ converged $= 41$). Table 3 illustrates how each disposition category was applied, using actual combinations from the 180-combination grid, including a convergence pair in which only the psychological state differs.

Analysis of the derived causal scenarios revealed that situations repeatedly emerged in which the driver failed to appropriately respond to a Takeover Request (TOR), with the system entering a hazardous state directly without supplementary safety intervention. This suggests that the overall control structure relies excessively on a single driver response—the perception of and reaction to the TOR—implying a structural limitation through which risk may propagate to the entire system. Table 4 presents representative examples of the causal scenarios, illustrating how each scenario links a control action and UCA type with the driver psychological state acting as its causal factor and with the associated system-level hazard.

4 | RISK ASSESSMENT AND SCENARIO ANALYSIS

4.1 | Definition of risk assessment criteria and calculation method

This section defines the criteria and calculation method for quantitatively assessing the risk of the 41 driver misuse causal scenarios derived in Section 3. The risk assessment was conducted based on three standard risk variables: Severity (S), Likelihood (L), and Controllability (C). Severity refers to the extent of potential human injury and material damage that may result if a given scenario occurs. Likelihood represents the probability that the situation may arise in real driving environments. Controllability is defined in this study as the degree of difficulty of avoiding an accident or mitigating its consequences through intervention by the driver or surrounding traffic participants after the hazardous behavior has begun; a higher Controllability score therefore indicates that the hazardous event is more difficult to control and contributes to a higher overall risk. Each variable was independently rated on the 1-to-10 integer scale by four evaluators according to the scoring rubric summarized in Table 5, with reference to relevant prior cases, and the final score of each scenario was obtained by averaging the individual ratings.

This three-factor evaluation approach is based on the risk assessment framework widely used in traditional

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Unsafe Control Actions (UCAs) identified for the nine driver control actions, classified by the four standard UCA types with associated system-level hazards. N.A denotes UCA types that are not applicable to discrete control actions.

Control action	Not providing causes hazard	Providing causes hazard	Too early, Too late, Out of drder	Stopped too Soon, Applied too Long
CA-1 Control transition setting	The driver does not provide the control transition setting in a situation where a control transition is required. [H.4]	The driver provides the control transition setting in a situation where a control transition is not required. [H.3]	The driver provides the control transition setting too late. [H.4]	The driver provides the control transition setting for too short a duration. [H.4] The driver continuously provides the control transition setting. [H.4]
CA-2 Steering control	The driver does not provide steering control in a situation where lane keeping or evasive steering is required. [H.2]	The driver provides steering control in a situation where steering is not required. [H.2]	The driver provides steering control too late. [H.2]	The driver provides steering control for too short a duration. [H.2] The driver continuously provides steering control. [H.3]
CA-3 Acceleration control	The driver does not provide acceleration control in a situation where acceleration is required. [H.1]	The driver provides acceleration control in a situation where acceleration is not required. [H.1]	The driver provides acceleration control too late. [H.1]	The driver provides acceleration control for too short a duration. [H.1] The driver continuously provides acceleration control. [H.1]
CA-4 Braking control	The driver does not provide braking control in a situation where braking is required. [H.1]	The driver provides braking control in a situation where braking is not required. [H.1]	The driver provides braking control too late. [H.1]	The driver provides braking control for too short a duration. [H.1] The driver continuously provides braking control. [H.1]
CA-5 TOR response	The driver does not perform the takeover response when a TOR is issued. [H.4]	The driver performs a takeover response when no TOR has been issued. [H.3]	The driver performs the takeover response too late. [H.4]	N.A The driver remains in a state of not responding to the takeover request. [H.4]
CA-6 Multichannel warning setting	The driver does not set the multichannel warning in a situation where warning reception is required. [H.4]	The driver sets the multichannel warning unnecessarily. [H.4]	The driver sets the multichannel warning too late. [H.4]	N.A The driver keeps the multichannel warning setting continuously enabled. [H.4]
CA-7 Emergency mode response	The driver does not perform the emergency mode response when an emergency mode transition occurs. [H.4]	The driver performs an emergency mode response in a nonemergency situation. [H.3]	The driver performs the emergency mode response too late. [H.4]	N.A

(Continues)

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 (Continued)

Control action	Not providing causes hazard	Providing causes hazard	Too early, Too late, Out of drder	Stopped too Soon, Applied too Long
				The driver remains in a state of not responding to the emergency mode. [H.4]
CA-8 Warning-repetition guidance setting	The driver does not set the warning-repetition guidance function in a situation where repeated warnings are required. [H.4]	The driver sets the warning-repetition guidance function unnecessarily. [H.4]	The driver sets the warning-repetition guidance function too late. [H.4]	N.A The driver keeps the warning-repetition guidance setting continuously enabled. [H.4]
CA-9 Manual system control	The driver does not provide manual system control in a situation where manual control is required. [H.4]	The driver provides manual system control in a situation where manual control is not required. [H.3]	The driver provides manual system control too late. [H.4]	The driver provides manual system control for too short a duration. [H.4] The driver continuously provides manual system control. [H.3]

TABLE 3 Representative dispositions of UCA–psychological state combinations in the 180-combination grid. The first two rows form a convergence pair in which only the psychological state differs.

UCA (CA, type)	Causal factor	Disposition	Rationale (C)	Result
UCA-01 (CA-1, Not providing)	Overtrust	Valid	Constitutes an independent risk mechanism linked to a safety-constraint violation	Adopted as SCN 1
UCA-01 (CA-1, Not providing)	Distrust	Excluded (convergence)	Changing the psychological state does not substantially alter the hazardous behavior, the hazard, or the harm	Converged into SCN 1
UCA-02 (CA-1, Providing)	Overtrust	Excluded (no constraint linkage)	The misuse cannot be linked to a violation of the defined safety constraints under this psychological state	Removed
UCA-25 (CA-5, Stopped too soon)	Overtrust	N.A	The UCA type itself is not applicable to a discrete response action, independently of the psychological state	Not applicable

Failure Mode and Effects Analysis (FMEA) [24]. However, the conventional multiplicative Risk Priority Number (RPN) used in FMEA has limitations in adequately reflecting the relative importance of each risk factor. To address this limitation, this study applies a Multicriteria Decision-Making (MCDM) approach, adopting a Weighted Sum Model for risk aggregation [25].

$$Weighted\ Risk\ Score = (S \times W_S) + (L \times W_L) + (C \times W_C)$$

The highest weight was assigned to Severity in order to reflect the difficulty of recovery once an accident occurs. Referring to the severity-dominant weighting configurations applied in prior FMEA-based risk aggregation studies [26], the weights were assigned as $W_S = 0.5, W_L = 0.3$, and $W_C = 0.2$ based on the judgment of the participating researchers, and these values are treated as working assumptions of the present analysis. A sensitivity check in which the weights were perturbed by up to ± 0.1, including the equal-weight case, confirmed that the resulting rankings remain highly correlated with the baseline (Spearman's $\rho = 0.94$–0.99) and that the majority of the top 10 scenarios (typically eight or nine) are preserved, indicating that the prioritization is robust to moderate changes in the weight assumptions.

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Representative examples of derived causal scenarios, linking each UCA to its psychological causal factor, the associated hazardous system state, and the accident outcome.

SCN IDX	UCA ID	UCA	Causal Factor	Triggering Event	Hazardous Behavior	Haz.	Scenario	Control	Harm
21	CA-5	The driver does not perform the takeover response when a TOR is issued.	Overtrust	The vehicle enters a situation in which automated driving is no longer feasible (e.g., a construction zone or complex road environment), and a TOR is correctly issued by the system.	The driver does not perform the takeover response.	H.4	While in automated driving, the vehicle enters a complex intersection or construction zone, and the system correctly issues a TOR after judging that automated driving can no longer continue. The driver, however, assumes that the system can handle the entire situation and does not perform the takeover response, so the vehicle proceeds into the hazardous section without terminating automated driving and is exposed to a collision risk.	The driver fails to take over control and cannot control the vehicle.	Frontal collision with a forward obstacle.
22	CA-5	The driver performs a takeover response when no TOR has been issued.	Distrust	The vehicle passes, in automated mode, through a section where lane markings temporarily disappear.	The driver performs an unnecessary takeover response (abrupt braking or steering) in a situation where takeover is not required.	H.3	While the vehicle passes in automated mode through a section with temporarily faint lane markings, the driver judges that the system does not properly recognize the situation. Although no TOR has been issued, the driver treats the situation as critical and performs a takeover response with abrupt steering or braking, causing the vehicle to become unstable, depart from its lane, and enter the adjacent lane.	Distrusting the automated driving system, the driver intervenes unnecessarily and fails to maintain stable control of the vehicle.	Side collision due to lane departure.
23	CA-5	The driver performs the takeover response too late.	Distraction	The vehicle reaches a situation where automated driving is no longer feasible, and a TOR is correctly issued.	The driver performs the takeover response later than required.	H.4	In automated driving, the vehicle enters a sharp curve and the system issues a TOR at an appropriate time, but the driver is looking at a smartphone and cannot perform the takeover response immediately. Because of the delayed response, the vehicle loses stability in the	The driver fails to respond promptly to the TOR and cannot control the vehicle's progression.	Road departure and collision with roadside structures.

(Continues)

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 (Continued)

SCN IDX	UCA ID	UCA	Causal Factor	Triggering Event	Hazardous Behavior	Haz.	Scenario	Control	Harm
							curved section and departs from the lane.		
24	CA-5	The driver remains in a state of not responding to the takeover request.	Distraction	The vehicle enters a situation where automated driving conditions are no longer satisfied, and a TOR is correctly issued.	The driver remains in a state of not performing the takeover response.	H.4	The vehicle loses its automated driving conditions while entering a complex urban road, and the system correctly issues a TOR to request the transfer of control. However, due to smartphone use, headset use, or drowsiness, the driver remains in a state of not performing the takeover response; the vehicle keeps driving in automated mode even in the hazardous section, then sways and departs from the lane when entering a curved segment.	The driver does not perceive the TOR, fails to take over control, and cannot stably control the vehicle.	Lane departure or side collision.

4.2 | Risk assessment results and implications

Across the 41 causal scenarios, the Severity scores ranged from 5.75 to 9.0 (mean 7.27), the Likelihood scores from 3.75 to 7.25 (mean 5.95), and the Controllability scores from 5.25 to 8.25 (mean 6.37). The results of the risk assessment indicate that a substantial proportion of the high-risk scenarios are closely associated with situations in which the driver fails to appropriately respond to the system's Takeover Request (TOR), quantitatively demonstrating that overall safety in Level 3 automated driving environments depends critically on the quality of interaction with the driver. In particular, even when a TOR is issued correctly, the risk level increases significantly if the driver does not respond to it appropriately or in time, suggesting that forms of misuse rooted in driver distraction, disregard of warnings, and improper configuration may function as single points of failure.

Table 6 presents the distribution of Severity, Likelihood, and Controllability scores for the top 10 highest risk causal scenarios. These leading scenarios share consistently high Severity levels, ranging from 8.25 to 9.0, while their Likelihood scores are more widely distributed. Controllability scores were evaluated at relatively high levels, between 6.5 and 8.25, which—under the definition adopted in this study—indicates that once the hazardous condition occurs, recovery is structurally difficult. These findings suggest that effective risk mitigation requires not only ensuring the proper issuance of TORs but also supporting the driver's immediate perception and intervention through improved HMI design, enhanced driver state monitoring, and the introduction of multilayer safety mechanisms.

5 | RISK ANALYSIS BASED ON SCENARIO CLUSTERING

5.1 | Clustering method and configuration

In this section, clustering techniques were applied to the 41 driver misuse scenarios in order to identify recurring Risk Archetypes by grouping scenarios that exhibit similar risk characteristics—structural similarities not readily identifiable through individual risk scores alone. Each scenario was represented by a feature vector composed of the three quantitative risk variables: Severity, Likelihood, and Controllability. The qualitative attributes of each scenario—the UCA type, the cause of driver misuse, and the type of accident—were used after clustering to characterize and interpret the resulting clusters. All three

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 5 Scoring rubric for Severity, Likelihood, and Controllability on the 1–10 scale.

Score	Severity (S)	Likelihood (L)	Controllability (C)
1–2	No injury; negligible vehicle or property damage	Occurrence is extremely rare in real driving environments	Almost all drivers can easily avoid the accident or mitigate its consequences
3–4	No or negligible injury; minor vehicle or property damage	Occurs rarely, only under uncommon combinations of conditions	Most drivers can avoid the accident through ordinary corrective action
5–6	Light to moderate injury; moderate vehicle or property damage	Occurs occasionally under specific driving conditions	The accident can be avoided only through skilled and timely intervention or under favorable conditions
7–8	Severe injury possible; major vehicle or property damage	Occurs frequently in comparable driving situations	Avoidance is difficult; mitigation depends largely on intervention by surrounding traffic participants
9–10	Fatal or life-threatening injury	Occurs almost always once the triggering conditions arise	The accident is practically unavoidable once the hazardous behavior has begun

TABLE 6 Risk evaluation results for the highest ranked causal scenarios based on Severity, Likelihood, and Controllability scores.

Rank	SCN IDX	UCA summary	Scenario summary	S	L	C	Weighted risk score
1	31	Late response to emergency mode transition	The vehicle switches to emergency mode due to a sudden stop of the preceding vehicle. The driver recognizes the transition too late, resulting in a collision.	8.75	7	6.75	7.825
2	39	Delayed provision of manual system control	The driver ignores a system malfunction warning. If speed intervention is delayed, the vehicle collides with a stationary obstacle.	8.75	5.75	8	7.7
3	23	Late takeover response to the TOR	While using a smartphone, the driver responds to the TOR too late, leading to road departure in a curved section.	8.25	7.25	7	7.7
4	1	Failure to set control transition conditions	The TOR condition is not configured, resulting in the vehicle approaching an obstacle without warning and causing a collision.	9	5.25	7.75	7.625
5	3	Control transition set too late	The TOR condition is configured too late, leaving insufficient response time and resulting in a collision with an obstacle.	8.5	6.25	7.25	7.575
6	21	Failure to provide the takeover response to the TOR	The driver does not respond to the TOR and enters a construction zone, resulting in a collision.	8.75	5.5	7	7.425
7	25	Failure to configure multichannel warning	With only visual warnings enabled, the driver misses the forward warning and collides with an obstacle.	8.25	5.5	6.5	7.075
8	35	Late configuration of warning-repetition function	The warning-repetition function is set too late, causing the driver to miss a lane departure warning and resulting in a side collision.	8	5.75	6.5	7.025
9	37	Failure to execute manual control	Due to drowsiness or headset use, the driver fails to recognize the TOR, leading to road departure or collision.	8.5	3.75	8.25	7.025
10	5	Control transition set too short	The TOR warning is issued for too short a duration for the driver to recognize it, resulting in a collision.	8.25	4.5	7.5	6.975

features were standardized to zero mean and unit variance prior to distance computation, so that variables with larger numerical ranges would not dominate the distance metric.

The K-means algorithm was applied for clustering. Since K-means requires the number of clusters K to be specified in advance, this study utilized the Silhouette Score to compare the degree of separation and cohesion

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

according to different values of K. Figure 2 presents the Silhouette Scores corresponding to varying values of K, showing the highest score of 0.48 at $K = 2$, which indicates that classification into two clusters provides the clearest separation. In addition, $K = 4$ yielded a relatively high score of 0.44. Therefore, both $K = 2$ and $K = 4$ were adopted as criteria for the clustering analysis. Given the limited sample size of 41 scenarios, the clustering results are interpreted throughout this study as an exploratory analysis of recurring risk patterns rather than as a definitive classification.

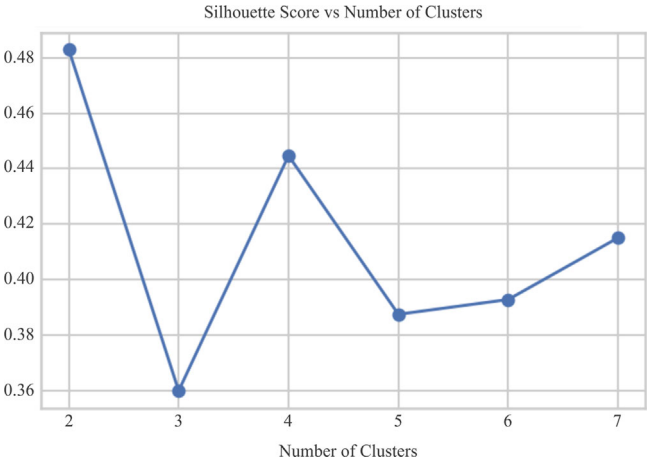
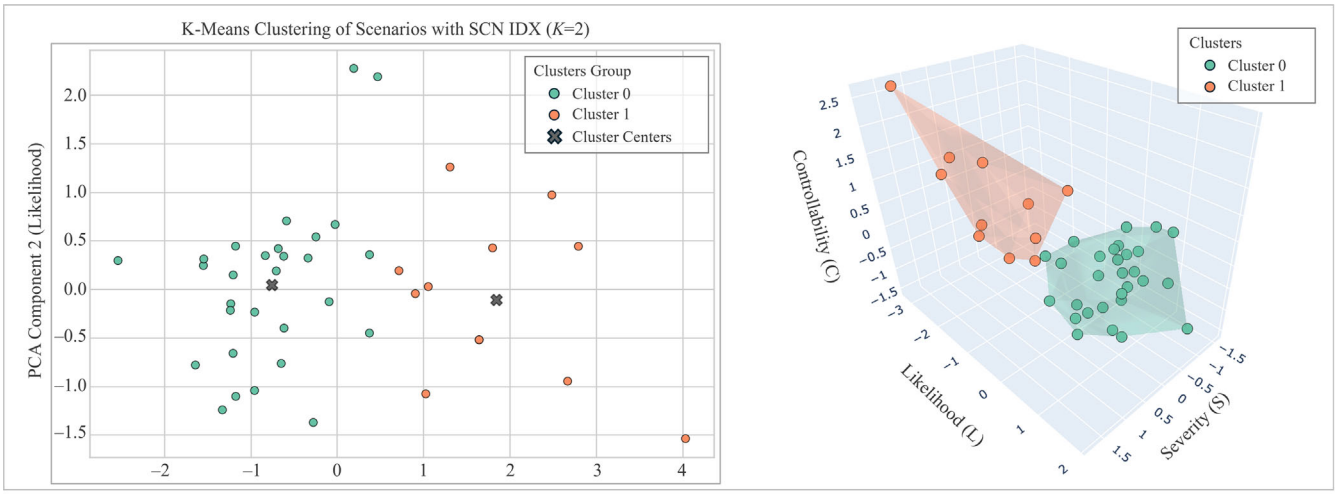


FIGURE 2 Silhouette score analysis for determining the optimal number of clusters in scenario classification.

5.2 | Clustering results and interpretation of risk patterns

The clustering results based on $K = 2$ divided all scenarios into two groups with distinct risk profiles. One group consisted of relatively lower risk scenarios in which the hazardous situations remained comparatively easy to control, reflected in lower Controllability scores, including many cases in which driver intervention remained feasible. The other group comprised high-risk scenarios involving failure to respond to Takeover Requests (TOR) or rapid risk escalation, exhibiting a high likelihood of progressing to accidents within a short time once a hazardous condition arises, indicating that the success or failure of driver intervention constitutes a critical factor in determining overall risk levels. Figure 3 presents the clustering results visualized in two-dimensional and three-dimensional spaces, using Principal Component Analysis (PCA) for the two-dimensional projection of the standardized feature space.

For a more detailed analysis, clustering with $K = 4$ was performed, and the results indicate that each cluster exhibits a distinct Risk Archetype. The first cluster consists of scenarios in which the driver fails to perceive the TOR itself, making immediate response impossible. These scenarios arise less frequently but, once triggered, are severe and structurally difficult to recover from. The second cluster is characterized primarily by cases in which warning signals are not delivered due to improper



Risk Attribute	Cluster 0 (Green) – 29	Cluster 1 (Orange) – 12
Severity	6.81±0.756	8.25±0.544
Likelihood	6.21±0.425	5.44±0.873
Controllability	6.00±0.451	7.21±0.572
Weighted Risk Score	6.47±0.445	7.20±0.397
Cluster Characteristics	A group of scenarios in which risk exists; however, severity and controllability are relatively balanced, making them comparatively manageable.	A high-risk scenario group characterized by relatively lower likelihood but including severe and difficult-to-control situations.

FIGURE 3 Visualization of clustering results for $K = 2$ using dimensionality reduction techniques.

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

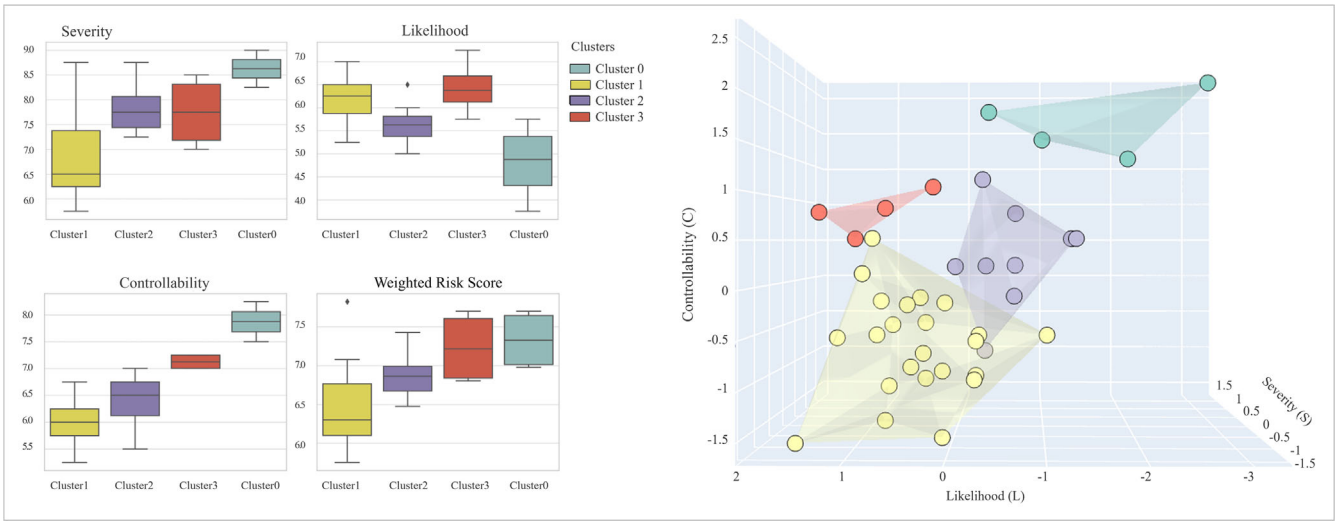


FIGURE 4 Distribution of risk variables and three-dimensional visualization of scenario clusters for $K = 4$.

TABLE 7 Comparative summary of mean risk variables and defining characteristics for each identified cluster.

Risk attribute	Cluster 1 (Yellow) – 23	Cluster 2 (Purple) – 10	Cluster 3 (Red) – 4	Cluster 0 (Green) – 4
Severity	6.74 ± 0.824	7.81 ± 0.513	7.75 ± 0.736	8.63 ± 0.323
Likelihood	6.20 ± 0.446	5.63 ± 0.5	6.44 ± 0.625	4.81 ± 0.875
Controllability	5.97 ± 0.428	6.38 ± 0.518	7.13 ± 0.144	7.88 ± 0.323
Weighted Risk Score	6.42 ± 0.481	6.87 ± 0.295	7.23 ± 0.472	7.33 ± 0.384
Risk Summary	Scenarios that may occur relatively frequently but are characterized by lower severity and lower difficulty of control, representing general risk.	Scenarios that are difficult to control but whose frequency and severity are adjustable, representing moderate-risk scenarios.	High-risk scenarios requiring immediate response and associated with a high probability of accidents.	Scenarios that occur less frequently but involve high severity and are difficult to control once the hazardous condition arises.
Scenario Characteristics	This cluster consists of scenarios in which the ADS properly issues TORs and driver responses are relatively appropriate. It includes typical road situations, with severity and controllability remaining balanced, making the risks comparatively manageable.	This cluster includes scenarios where delayed warning channel configuration or failure to enable repeated warnings weakens driver response. Although these cases do not immediately result in accidents, risk may escalate in complex environments if driver intervention fails.	This cluster comprises scenarios involving delayed TOR recognition or deactivation of warning-repetition functions, leading to delayed control intervention. Controllability is reduced, and the likelihood of accident propagation is high, representing a high-risk group.	This cluster includes scenarios in which the driver fails to perceive warnings or is in a sensory-blocked condition. The control loop collapses and immediate intervention becomes difficult; although these situations arise less frequently, their consequences are severe and structurally difficult to recover from.
Risk Type	General TOR Response	Warning Configuration Error	Warning Disregard / Delayed Response	Recognition Failure

configuration of warning functions. The third cluster comprises distraction-oriented scenarios in which responses are delayed or warnings are ignored even after recognition. The final cluster includes relatively lower risk scenarios involving typical TOR response situations. Figure 4 presents a visualization of the distribution of risk variables and the spatial separation among clusters. Table 7 provides a comprehensive summary of the mean

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 8 Selected representative scenarios corresponding to distinct Risk Archetypes.

SCN IDX	Risk type	Misuse cause	Key issue	Cluster characteristics summary	Selection rationale
10	Steering operation failure	Interference	Steering intervention is not maintained, causing the vehicle to transition into an uncontrolled state	A low-risk scenario that can repeatedly occur in general road environments	A representative case where failure to maintain steering leads to loss of vehicle control and potential real-world accidents
25	Omission of warning channel configuration	Overtrust	Only visual warnings are configured, resulting in failure to perceive alerts	A moderate-risk scenario in which configuration errors and perception failures combine to limit warning effectiveness	Demonstrates how mistakes in the warning design stage can lead to accidents, highlighting the importance of UI/UX design
23	Delayed TOR recognition	Distraction	Failure to recognize TOR issuance and immediately assume control	A high-risk scenario where TOR output is normal but response delay occurs due to smartphone use or similar distractions	A representative HMI interaction failure case where warning recognition is delayed due to distraction
37	Manual control failure	Interference	Failure to perform manual intervention due to TOR recognition failure	A high-risk scenario where system abnormalities or warning recognition failure prevent control intervention	A case where external interference factors lead to TOR recognition failure and inadequate manual intervention
26	Overreaction due to excessive warnings	Distrust	The driver attempts unnecessary maneuvers due to surprise caused by excessive warnings	A moderate-risk scenario where excessive warnings or high sensitivity induce startle responses and overcontrol	A case where excessive system reactions unintentionally induce unnecessary driver intervention, leading to accidents

risk-variable values and scenario characteristics for each cluster, demonstrating that each cluster can be interpreted as a Risk Archetype sharing distinct risk generation mechanisms and misuse characteristics.

Based on the clustering results, representative scenarios were derived to reflect each Risk Archetype by comprehensively considering the misuse factors and accident types that repeatedly appeared within each cluster. In total, five representative scenarios were identified, providing a balanced representation of various misuse causes and accident types while effectively encompassing the key risk patterns derived from the analytical process. Table 8 summarizes the results of the representative scenario selection.

6 | CONCLUSION

This study proposed an analytical framework that integrates the SOTIF perspective with the System-Theoretic Process Analysis (STPA) methodology to systematically analyze driver misuse in Level 3 (Lv.3) conditional automated driving environments. Driver misuse was interpreted not as a simple behavioral error but as inadequate control within the driver–system interaction structure. Based on this perspective, 41 causal scenarios were derived and structured with explicit traceability linking hazardous system states, unsafe control actions, psychological causal factors, and safety-constraint violations.

The Weighted Risk Score analysis showed that many high-risk scenarios are closely associated with failures to respond appropriately and in time to Takeover Requests (TOR), indicating that safety in Lv.3 automated driving depends not only on technological capability but also on the quality of interaction with the driver. Clustering analysis further showed that driver misuse can be grouped into distinct Risk Archetypes, including perception failure, warning configuration error, and delayed response patterns. These findings suggest that driver misuse risks should be understood as recurring patterns rather than isolated cases and provide useful implications for driver monitoring, HMI design, and multilayer safety assistance systems.

This study is limited by its reliance on expert judgment and scenario-based analysis rather than real-world driving data or experimental validation. In addition, the risk- variable weights are analysis assumptions assigned by the participating researchers, and the clustering results are exploratory given the limited number of scenarios. Future research should validate the proposed procedure using driving logs, driver experiments, and simulation data, and derive design requirements and mitigation strategies for each identified Risk Archetype.

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ACKNOWLEDGMENTS

This work was supported by the Korea Evaluation Institute of Industrial Technology (KEIT) grant funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea) (No. 20018248, Development of safety of the intended functionality from insufficiency of perception and decision-making).

ORCID

Yangkoo Lee https://orcid.org/0000-0001-7295-5103
Eun Hye Jang https://orcid.org/0000-0001-8348-2663
Mi Chang https://orcid.org/0000-0002-6699-076X
Kezia Amanda Kurniadi https://orcid.org/0000-0002-3167-074X
Jiwoo Han https://orcid.org/0009-0004-7842-1611
Woojin Kim https://orcid.org/0000-0001-9589-784X
Daesub Yoon https://orcid.org/0000-0002-2442-0080

REFERENCES

1. H. Guo, H. Qiu, Y. Zhou, and Y. Deng, *Analysis of influencing factors of level 3 automated vehicle takeover: a literature review*, Sustainability **16** (2024), no. 19, 8345.
2. SAE On-Road Automated Vehicle Standards Committee, *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems*, SAE Standard J3016, 2014.
3. SAE International, *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles*, SAE J3016, 2021.
4. A. Eriksson and N. A. Stanton, *Takeover time in highly automated vehicles: Noncritical transitions to and from manual control*, Hum. Factors **59** (2017), no. 4, 689–705.
5. C. Gold, D. Damböck, L. Lorenz, and K. Bengler, *Take over! how long does it take to get the driver back into the loop?* Proc. Hum. Factors Ergon. Soc. Annu. Meet. **57** (2013), no. 1, 1938–1942.
6. Y. Liu, X. Wang, Z. Li, and Y. Zhang, *Safety evaluation and prediction of takeover performance in automated driving considering drivers' cognitive load*, Transport. Res. F: Traffic Psychol. Behav. **99** (2024), 102–115.
7. H. Kim, W. Kim, J. Kim, S.-J. Lee, D. Yoon, O.-C. Kwon, and C. H. Park, *Study on the takeover performance of level 3 autonomous vehicles based on subjective driving tendency questionnaires and machine learning methods*, ETRI J. **45** (2023), no. 1, 75–92.
8. M. R. Endsley and E. O. Kiris, *The out-of-the-loop performance problem and level of control in automation*, Hum. Factors **37** (1995), no. 2, 381–394.
9. N. Merat, A. H. Jamson, F. C. H. Lai, M. Daly, and O. Carsten, *The out-of-the-loop concept in automated driving: Proposed definition, measures and implications*, Cogn., Technol. Work **21** (2019), no. 1, 87–98.
10. ISO, Road vehicles—Functional safety, ISO 26262, 2018.
11. ISO, Road vehicles—Safety of the intended functionality (SOTIF), ISO/PAS 21448, 2022.
12. N. G. Leveson, *Engineering a safer world: Systems thinking applied to safety*, MIT Press, 2011.
13. N. G. Leveson and J. P. Thomas, *STPA Handbook*, MIT-STAMP-001, 2018.
14. D. W. Kang, J. Kim, S. Lee, and J. Park, *Modeling and validating foreseeable misuse in safety of the intended functionality (ISO 21448): a case study of automated emergency braking*, SAE JCAV **9** (2026), no. 1, 1–10.
15. L. Birkemeyer, C. King, and I. Schaeger, *Is scenario generation ready for SOTIF? A systematic literature review*, arXiv preprint (2023), DOI 10.48550/arXiv:2308.02273.
16. A. Abdulkhaleq, M. Baumeister, H. Böhmert, and S. Wagner, *Missing no interaction–using STPA for identifying hazardous interactions of automated driving systems*, Int. J. Safety Sci. **2** (2018), no. 1, 115–124.
17. S. Chen, S. Khastgir, I. Babaev, and P. Jennings, *Identifying accident causes of driver-vehicle interactions using system theoretic process analysis (STPA)* (Proceedings of the IEEE International Conference on Systems Man, and Cybernetics, Toronto, Canada), 2020, pp. 3247–3253.
18. S. Zhang, T. Tang, and J. Liu, *A hazard analysis approach for the SOTIF in intelligent railway driving assistance systems using STPA and complex network*, Appl. Sci. **11** (2021), no. 16, 7714.
19. M. Blanco, J. Atwood, H. M. Vasquez, T. E. Trimble, V. L. Fitchett, J. Radlbeck, G. M. Fitch, S. M. Russell, C. A. Green, B. Cullinane, and J. F. Morgan, *Human factors evaluation of level 2 and level 3 automated driving concepts*. DOT HS 812 182, 2014.
20. R. Häuslschmid, M. von Bülow, B. Pfleging, and A. Butz, *Supporting trust in autonomous driving* (Proceedings of the 22nd International Conference on Intelligent User Interfaces, Limassol, Cyprus), 2017, pp. 319–329.
21. J. D. Lee and K. A. See, *Trust in automation: Designing for appropriate reliance*, Hum. Factors **46** (2004), no. 1, 50–80.
22. R. Parasuraman and V. Riley, *Humans and automation: Use, misuse, disuse, abuse*, Hum. Factors **39** (1997), no. 2, 230–253.
23. Z. Lu, *Human factors of transitions in automated driving*, Delft University of Technology, 2020.
24. J. Huang, J. You, and Y. Li, *Failure mode and effect analysis improvement: A systematic literature review and future research agenda*, Reliab. Eng. Syst. Saf. **199** (2020), 106885.
25. E. Triantaphyllou, *Multi-criteria decision making methods: A comparative study*, Springer, 2000.
26. L. Lu, Z. Rong, and R. de Souza, *Enhanced FMEA for supply chain risk identification*, Proc. Hambg. Int. Conf. Logist. (HICL) **25** (2018), 311–330.

AUTHOR BIOGRAPHIES



Yangkoo Lee received his MS and PhD degrees in computer science from Chungbuk National University, Cheongju, Republic of Korea, in 2004 and 2010, respectively. Since 2011, he has been with the Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea, where he is currently a principal researcher. His main research interests include interoperability among heterogeneous systems, human–robot interaction, human factors, and digital twin-based future mobility applications.

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



Eun Hye Jang is a principal researcher at ETRI AI Robot UX Lab. She received her Ph.D. in psychology from Chungnam National University in 2009 and has worked at ETRI since then. Her research focuses on users' intention and behavior detection using human factors and psychophysiological signals.



Mi Chang is currently a senior researcher at the Electronics and Telecommunications Research Institute, Republic of Korea. She received her PhD degree from the Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology, Republic of Korea, in 2023. She received her MDes degree in product design and her BS degree in electronic communication engineering, in 2018 and 2011, respectively.



Kezia Amanda Kurniadi received her B.S. degree in Industrial Engineering from Petra Christian University, Surabaya, Indonesia, in 2011. She received her MS and PhD degree in Industrial Engineering from Pusan National University, Busan, Republic of Korea, in 2013 and 2022, respectively. Since 2024, she has been with ETRI. Her current research interests include driver monitoring systems, SOTIF, virtual human or Metahuman, and agentic AI for smart factories.



Jiwoo Han received his BS degree in computer science from the Vietnam National University, Ho Chi Minh City, Vietnam and his MS degree in computer science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea. He is currently a senior researcher in the AI Robot UX Research Section at the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea. His research interests include data management, system architecture, simulation-based evaluation, and the development of user-centric mobility technologies.



Woojin Kim received the B.S. degree from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea, in 2008, and the Ph.D. degree from Seoul National University, Seoul, Republic of Korea, in 2014. Since 2014, he has been with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea, where he is currently the Director of the AI Robot UX Research Section. His research interests include AI-based robot control, human–robot interaction, and intelligent robotic systems.



Daesub Yoon received his M.S. and Ph.D. degrees in Computer Science and Software Engineering from Auburn University, USA. From 2001 to 2005, he was a Research Assistant at the Intelligent & Interactive Systems Laboratory at Auburn University. He joined the Electronics and Telecommunications Research Institute (ETRI), Republic of Korea, in 2005. His research interests include AI, robotics, future mobility, eye tracking, attentive user interfaces, mental workload, and human factors. He is currently the Assistant Vice President of the AI Robot & Mobility Research Division at ETRI.

How to cite this article: Y. Lee, E. H. Jang, M. Chang, K. A. Kurniadi, J. Han, W. Kim, and D. Yoon, *STPA-based analysis of driver misuse risk scenarios in level 3 automated driving*, ETRI Journal (2026), 1–16, DOI 10.4218/etrij.2026-0204

22337326, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.4218/etrij.2026-0204 by Electronics and Telecommunications, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License