STPA-WA: A Safe System-Oriented Extension of STPA for Dependability in Automated Driving

Manabu Okada[1], Barbara Gallina[2]

[1] *TIER IV, Inc., Jacom Building, 1-12-10 Kitashinagawa, Tokyo, 140-0001, JAPAN*
manabu.okada@tier4.jp

[2] *Mälardalen University, P.O. Box 883, Västerås, SE-72123, Sweden*
barbara.gallina@mdu.se

ABSTRACT

To nearly achieve vision zero, the United Nations have introduced the safe system approach. This approach aims at accommodating human factors by incorporating redundancy within the different identified safety pillars (safe speed linked to regulations, safe vehicles, safe roads, safe emergency care). To contribute to vision zero, ADSs (Automated Driving Systems) need to contribute to road safety, in addition to ensuring safety of the intended functionality. Achieving safe and harmonious ADSs necessitates a comprehensive safety approach beyond the vehicle-centric / ADS-centric perspective. It also calls for roadmanship which is a concept that centers on whether the vehicle plays well with others.

System Theoretic Process Analysis (STPA) is a process for conducting hazards analysis. STPA has been recently included within the automotive safety standards and regulations as suggested process to be adopted. However, its application has been considered challenging and proposals for adaptations have been provided. In this paper, we provide yet another novel adaptation, called STPA-WA[1], which embraces roadmanship and the different perspectives represented by the different pillars. STPA-WA deepens safety/dependability understanding.

1. INTRODUCTION

The road traffic system is a complex socio-technical system. Automated Driving Systems (ADSs) are part of the road traffic system. Its safety (road traffic safety, ISO 39001:2012 [12]) comprises vehicles, roads, the emergency medical system and road users. To nearly achieve casualty-free road traffic (vision zero), the safe system approach was introduced by the United Nations. This approach aims at accommodating human factors by incorporating redundancy within the different identified safety pillars (safe speed linked to regulations, safe vehicles, safe roads, safe emergency care). Hence, achieving safe and harmonious ADSs necessitates a comprehensive safety approach beyond the vehicle-centric / ADS-centric perspective, provided within ISO 21448:2022 [8], standard on Safety of the Intended Functionality (SOTIF). It also calls for roadmanship which is a concept that centers on whether the vehicle plays well with others. System Theoretic Process Analysis (STPA) is a process for conducting hazards analysis. STPA has been recently included within the automotive safety standards (ISO 21448:2022) and regulations (UN Regulation No. 157, [22]) as suggested process to be adopted. However, its application has been considered challenging and proposals for adaptations have been provided. In this paper, we provide yet another but novel adaptation, which we call STPA-WA. STPA-WA embraces roadmanship and the different perspectives represented by the different pillars.

STPA-WA deepens safety, or better, dependability understanding by focusing on: (1) achieving traffic harmony within the existing traffic environment (WA1), by analyzing ADSs' interaction with the other pillars, while considering roadmanship-specific metrics; and (2) addressing stakeholder feedback loops (WA2), to ensure continuous learning and adaptation in design, operation, and policy development. By integrating this socio-technical perspective, STPA-WA offers valuable inputs for evolving SOTIF, aiming to build dependable ADSs and foster positive emergent properties within the overall road transportation system.

The rest of the paper is organized as follows. In Section 2, we provide essential background information. In Section 3, we apply STPA based on the SOTIF-related automotive practices. We also draw our lessons learned. In Section 4, we present our adaptation of STAMP/STPA which extends through two complementary dimensions: WA1 and WA2. In Section 5, we discuss related work. Finally, in Section 6, we provide concluding remarks and sketch directions for future work.

2. BACKGROUND

In this section, we recall essential background information. **The Safe System Approach (SSA)** comprises five interdependent Pillars (P), which are: **P1**) road safety management (including safe speed), which consists of organizational structures, policies, and processes for coordinating safety efforts; **P2**) safe roads and infrastructure, which consists of the infrastructure designed to reduce crash risks and protect vulnerable users; **P3**) safe vehicles, which consists of vehicle standards, technologies, and inspection systems; **P4**) safe road users, which consists of education, training, awareness, and enforcement; **P5**) post-crash response, which consists of emergency services, medical care, and rehabilitation. SSA recognizes that responsibility is shared

[1] Japanese wa is a homophone: 1)和 "harmony" and 輪 "loop/ring". We chose "WA" to evoke harmony and feedback loops.

among the different pillars, with the goal of creating a road traffic system that accommodates human factors without resulting in serious harm.

Functional Safety (FuSa) is the absence of unreasonable risk due to hazards caused by malfunctioning behavior of E/E systems [7].

Safety Of The Intended Functionality (SOTIF) is absence of unreasonable risk due to hazards resulting from functional insufficiencies (i.e., insufficiency of specification or performance insufficiency) of the intended functionality or its implementation [8]. SOTIF builds on top of FuSa. SOTIF focuses on hazards that can occur even when a system is functioning as designed. SOTIF applies to intended functionalities that require proper situational awareness, covering SAE driving automation levels 1-to-5. It provides a framework for identifying and addressing: functional insufficiencies in the specification of the intended functionality, performance limitations of the implemented system, and reasonably foreseeable misuse by the user. However, as highlighted in [18], there are two significant gaps in the current SOTIF framework.

Road safety [12] comprises not only road vehicles but also other systems (roads, the emergency medical system and road users) and their interactions. Road safety deals with the conditions and factors related to road traffic crashes and other road traffic incidents that have an impact on, or have the potential to have an impact on death or serious injury of road users.

Roadmanship [6] captures the ability to drive on the road safely without creating hazards and responding well (regardless of legality) to the hazards created by others. The roadmanship concept centers on whether the vehicle "plays well with others," even if others are not around. Playing well may embrace: avoiding near misses (i.e., incidents, events with potential safety-related effects that, at the end, are prevented from evolving into actual crash accidents), blocking traffic incidents, which instead seem to be frequent in the available data from the Austin database [3]. Specifically, the data reveal that safety-related incidents (classified as "Near Miss" and "Safety Concern") constitute more than half of all reported autonomous vehicle incidents. Significantly, "Blocking Traffic" ranks as the third most frequent incident type, while "Ignore APD (Austin Police Department) Direction" appears fifth on the list.

Prospective safety performance can be measured. In ISO TS 21934-2 [11], two metrics have been defined: Time To Collision (TTC) and Post Encroachment Time (PET).
Time To Collision (TTC) is the remaining time until a collision is going to occur in case the movement (direction, velocity) of the involved road users does not change. To calculate t_{TTC}, we use formula 1, which takes only the longitudinal distance (d_{VUT}) and velocity (v_{VUT}).

$$t_{\mathrm{TTC}} = \frac{d_{\mathrm{VUT}}}{v_{\mathrm{VUT}}} \tag{1}$$

Post Encroachment Time (PET): is the time period between the first road user leaving the conflict zone t_e^{out} and a second vehicle entering the conflict zone t_o^{in}. The conflict zone is defined by the vehicles' dimensions and their trajectories. We use PET as defined in formula 2.

$$\mathrm{PET} = t_o^{\mathrm{in}} - t_e^{\mathrm{out}} \tag{2}$$

Space-sharing conflicts occur when multiple road users attempt to occupy the same space at the same time [15]. This is in-line with the historical definition of **traffic conflict**, i.e., observable situations in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged [2]. These space-sharing conflicts are categorized into five prototypical scenarios.

- **Obstructed path (*OP*)**: A road user's path is physically or operationally obstructed, leading to space sharing with another road user's travel space.
- **Merging paths (*MP*)**: Multiple road users on different paths are forced to merge or travel alongside each other.
- **Crossing paths (*CP*)**: Road users on perpendicular paths creating potential conflict at intersection points.
- **Unconstrained head-on paths (*UHP*)**: Road users approaching from opposite directions with freedom to adjust their positions.
- **Constrained head-on paths (*CHP*)**: Road users approaching from opposite directions with limited position adjustment options.

These categorized conflicts provide a structured way to analyze how ADSs shall interact with other road users, especially in complex traffic environments. Proper analysis of these interactions is essential for designing ADSs that can integrate safely into existing traffic systems, i.e., that can exhibit roadmanship.

Essential terms are recalled in what follows. *Environment*: dynamic (i.e., moving actors (other than the ego vehicle) that are relevant to the ego vehicle) and static (i.e., geo-spatially stationary elements, also the presence of buildings near the road side that act as a view-blocking obstruction) parts of the scenario [4]. *Scene* [8]: a snapshot of the environment including the scenery, dynamic elements, and all actors' and observers' self-representations, and the relationships among those entities. *Scenario*: describes the temporal development between several scenes, usually including the ADS(s)/subject vehicle(s) and their interactions in the process of performing a Dynamic Driving Task (DDT). In [8], a scenario is the description of the temporal relationship between several scenes in a sequence of scenes, with goals and values within a specified situation, influenced by actions and events. *Current Operational Domain (COD)* [10] is a specific set of operating conditions, which exist presently in the immediate vicinity of an ADS, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics. Hence, COD defines real-time real-world conditions that the ADS experience.

6-layer model To structure an environment description for scenes and scenarios, valid for all cases (including urban traffic), a 6-Layer Model (6LM) was introduced [21]. The 6 layers are: Street Layer (PL1), Traffic Infrastructure (PL2), Temporal Modifications of PL1 and PL2 (PL3), Movable Objects (PL4), Environment Conditions (PL5), Digital Information (PL6). For details regarding the layers, the reader may refer to [21].

System-Theoretic Process Analysis (STPA) is a four-step process to perform hazard analysis. STPA is based on system theory and builds on top of STAMP, which stands for Systems-Theoretic Accident Model and Process. STAMP provides the foundational theory. While STPA provides a process for applying the theory. For details on STAMP/STPA, the reader may refer to [14] as well as the recent scoping review [19] aimed at allowing readers to gain a holistic understanding of how and where STAMP and STPA have been applied and tested.

The first step of STPA establishes the foundation for the hazard analysis. This first step is a composite, which consists in identifying unacceptable losses (i.e., unacceptable outcomes that must be avoided), translating the losses into system-level hazards, establishing the corresponding safety constraints preventing the hazards.The second step consists in modeling the control structures. The third step consists in identifying the unsafe control actions. The fourth step consists in identifying the loss scenarios.

Three-Layer Scenario Framework [1] extends STPA by constructing control structure into a three-layer as follows.
- Physical layer : road geometry, signal layout, weather, ego/other-actor positions and other objective environmental factors.
- Behavior & Recognition layer : agents' control actions (e.g., acceleration, braking, steering) and what the ADS/sensors or humans are attending to or perceiving.
- Prediction & Intention layer : each agent's forward predictions and intentions (e.g., yield/advance, honk, gesture), making implicit/explicit communication analyzable.

3. Application STPA and Lessons Learned

In this section, we introduced the application of SOTIF-focused STPA(3.1) , based on the automotive practices. Then, based on our findings, we give our lessons learned(3.2).

3.1. Application of SOTIF-focused STPA

Figure 1. Scene of application

In what follows, we report on the application of SOTIF-focused STPA on the scene, given in Figure 1, which shows that the ADSs-equipped vehicle (ADSEV) is approaching a crossroad, where a group of pedestrians is crossing. The scenario, which captures the crossing paths at the intersection, where no traffic signal is present, is categorized *CP* in the space-sharing conflict patterns, recalled in Section 2.

STPA STEP1

Losses (L): L-1: Loss of life or serious injury to people. L-2: Loss of vehicle property due to damage or surrounding property.

Hazards (H): H-1: ADSEV violates the required safety distance with other road users. (leads to L-1) H-2: ADSEV does not maintain safety distance from objectives, road structures, infrastructure, or terrain. (leads to L-2).

System-Level Safety Constraints (SC): SC-1: ADSEV must always maintain a safe distance from all detected road users (related to H-1). SC-2: ADSEV must maintain safety distance from objectives, road structures, infrastructure, or terrain. (related to H-2).

STPA STEP2 . We created a high-level control structure, shown in Figure 2, which consists of: ADS as controller, vehicle's platform as controlled process. The ADS and the vehicle's platform interact with the environment (e.g., a crossroad, a pedestrian group, weather, etc.), external to the vehicle. The ADS's control actions (CA) are: ADS-CA1: "maneuver to drive" (including from creeping to acceleration, steering, and restarting from stopped state) and ADS-CA2: "maneuver to stop" (including from light braking for speed maintenance to complete stopping and maintaining a stopped state), which addresses the scene in Figure 1. The control structure represents a problem space of ADSEV. This is aligned with the conventional STPA application within the SOTIF framework.

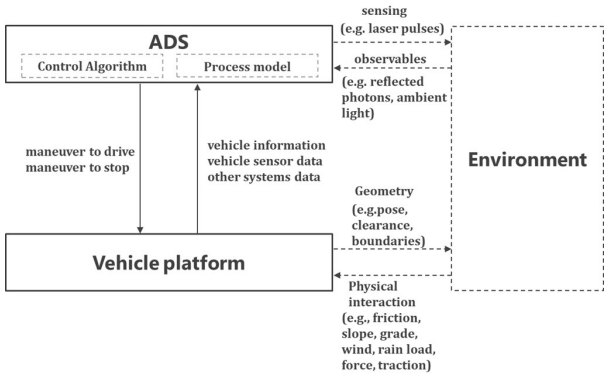

Figure 2. High-level control structure of ADS

STPA STEP3 Based on the control structure in Figure 2, we analyzed CAs to identify UCAs shown in table 1 Table includes acronyms for the following categories: not providing causes hazard (NP), providing causes hazard (P), providing too early (T-E) or too late (T-L), stop too soon(D-S) or applied too long(D-A).

STPA STEP4 Here, we analyze the hazard causal scenarios (HCS) of UCAs. Taking UCA ADS-T2-L1 as an example, the hazard causal scenarios analysis reveals: HCS ADS-T2-L1-1: ADS believes that ADSEV can pass without "maneuver to stop" (e.g., braking) based on its prediction of the pedestrian speed, location and expected trajectory (cause of UCA ADS-T2-L1).

3.2. Achievements, Limitations, and Lessons Learned

Achievements (A): A1: STPA effectively addressed SOTIF-related scenarios. A2: STPA successfully analyzed the causes of functional insufficiencies by focusing on interactions among system components.A3: STPA enabled the derivation of safety requirements based on identified hazard causal scenarios.

Limitations (L): L1: Since the SOTIF focuses on collisions, when we apply STPA to address SOTIF and identify only collision-related losses in STPA STEP1, we may miss non-collision-related losses related to roadmanship. L2: Since the SOTIF has vehicle-centric view, when we model the control structure in STPA STEP2 with only ADS as the controller, we may miss indirect interactions among road users that potentially

Table 1. Unsafe Control Actions for ADS maneuvers

Control Action	Not providing causes hazard	Providing causes hazard	Providing too early, too late	Stop too soon, applied too long
ADS-CA1 maneuver to drive (including from creeping to acceleration, steering, and restarting from stopped state)	no UCA	**ADS-P1-1** ADS provides maneuver to drive when the expected trajectories conflict within safe stopping distance.(H-1)	**ADS-T1-E1** ADS provides maneuver to drive (acceleration or restart or creeping) too early when the expected trajectories conflict within safe stopping distance.(H-1)	**ADS-D1-A1** ADS continues to maneuver to drive too long when the expected trajectories conflict within safe stopping distance.(H-1)
ADS-CA2 maneuver to stop (including from light braking for speed maintenance to complete stopping and maintaining a stopped state)	**ADS-NP2-1** ADS does not provide maneuver to stop when the expected trajectories conflict within safe stopping distance.(H-1)	no UCA	**ADS-T2-L1** ADS provides maneuver to stop too late when the expected trajectories conflict within safe stopping distance.(H-1)	no UCA

affect subject vehicle. L3: STPA may not support the prioritization of safety requirements and measures, especially when existing environmental safety features (e.g., speed bumps) are already in place.

Lessons Learned (LL): LL1: SOTIF is crucial for ADSs, and STPA provides a solid foundation. However, achieving harmony with the existing traffic environment is also essential. In some cases, there is a trade-off between safety margins and traffic smoothness. Therefore, the concept of roadmanship shall be incorporated into STPA. LL2: To analyze roadmanship, it is important to identify road users within a defined problem space-referred to as the Area of Interest (AOI)—where road users are expected to share space. Roadmanship tends to emerge in such shared-space scenarios. LL3: The STPA handbook also recommends stakeholder analysis. By incorporating SSA into context definition, we can gather relevant safety-related information specific to AOI. This enables prioritization of safety requirements and countermeasures.

4. SSA-ORIENTED ANALYSIS:STPA-WA

To address the limitations *L* and the lessons learned *LL*, identified in the previous section, we propose STPA-WA, which highlights a methodological adaptation aimed at deepening dependability analysis for ADSs by integrating different viewpoints: the Safe system-centric viewpoints within a socio-technical perspective. By integrating these viewpoints within a socio-technical perspective, STPA-WA contributes to valuable insights to the evolution of vehicle-centric SOTIF, aiming to promote dependable ADS behavior and the emergence of positive properties in real-world traffic environments. STPA-WA focuses on two key harmony (WA)-related dimensions:

WA1: WA-Traffic to address trade-offs
-Extends the analysis from vehicle-centric to Safe System-centered viewpoints based on conventional STPA results.
-Derives and allocates safety requirements to Safe System stakeholders.
-Evaluates safety measures via roadmanship-specific metrics.
WA2: Stakeholder Feedback Loops & WA Evolution
-Explicitly models safety control structures including feedback mechanisms among stakeholders.
-Enables continuous learning and adaptation.
-Informs system design, operation, and policy development.

As illustrated in Figure 3, a key and scope-focused distinction exists between conventional STPA application within the SOTIF framework and STPA-WA.

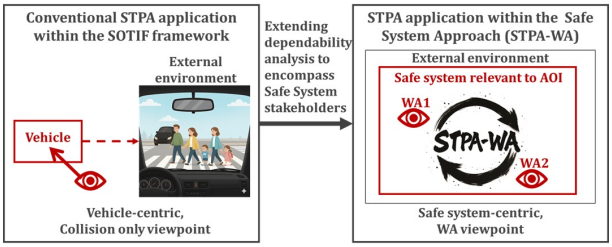

Figure 3. SOTIF-focused STPA & STPA-WA relationship

The conventional STPA application within the SOTIF framework focuses primarily on collisions to derive vehicle-centric safety requirements. STPA-WA extends the scope to identify and address trade-off (e.g., traffic harmony and safety distances with road users) relationships for ADS as follows; among multiple requirements and disciplines aspects. STPA-WA consists of four steps, which are detailed in what follows.

4.1. WA1-Step1: Analysis context adaptation

WA-Step1 extends STPA by broadening the analysis context beyond the SOTIF-focused STPA. Specifically, it shifts the analytical viewpoint from a vehicle-centric to a Safe-System centric viewpoint.

4.1.1. AOI definition

In the initiation of STPA-WA1, we need to define AOI as the specific region within the road environment, where the analysis is conducted. AOI encompasses both P1 and P2 within the SSA. P1 includes objectives. P2 is the result of the implementation of P1. AOI is defined as the combination of the items selected in P1 and P2, as shown in Figure 4. Concretely, P1 consists in the assigned laws, regulations, intended design of traffic (e.g.,context-appropriate roadway design, appropriate speed-limit setting, targeted education, outreach campaigns,

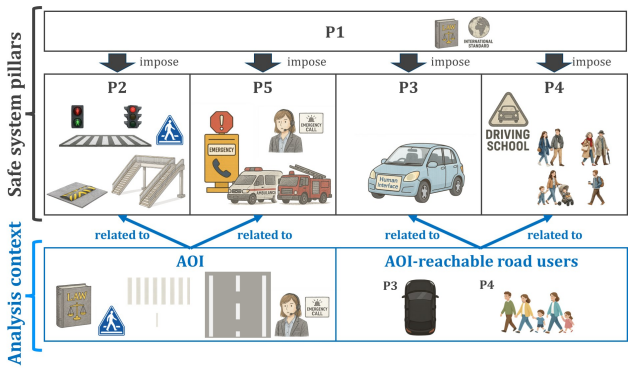

Figure 4. Relationships between safe system pillars and AOI

and enforcement), and appropriate safety capabilities for a safe ADS feature [9]. P2 coincides with the measures to comply with those laws (e.g., traffic signals, signs on the road surface), including design of roadway environment(e.g.,space sharing pattern *OP*, *CP*, etc.) to mitigate human mistakes and account for injury tolerance etc.

Result of AOI definition:
To exemplify the result of AOI definition, P1 represent traffic laws and, more specifically, P1 represent these two rules 1) "when approaching a pedestrian, who is crossing the lane, or a bicycle, which is crossing the lane, a vehicle (or streetcar) must proceed at a speed that will enable it to stop immediately in front of the crossing pedestrian (or the crossing bicycle), unless there are clearly no pedestrians (nor bicycles) ahead of the vehicle (or streetcar)". 2) "pedestrians must cross at the pedestrian crossing, when there is one in the vicinity".
To contribute to the implementation of P1, P2 implements the painting of the crosswalk (i.e., the part of a road that has been marked with road signs or road markings indicating it to be a place for pedestrians to cross). P2 marks intentionally designed and explicitly designates this area as an intended Crossing Paths (CP) space-sharing zone. This design highlights signals to all road users that perpendicular trajectories will intersect here, requiring coordination and yielding behaviors.

4.1.2. Identification of AOI-reachable Road Users

4.1.2 identifies the AOI-reachable road users, which potentially share the space with other road users within AOI. The AOI-reachable road users include P3 and P4 within the SSA. AOI-reachable road users can be determined by focusing on expected trajectories of each road user, including vulnerable road users, movable objects and subject vehicles. Potentially, AOI-reachable road users vary based on their intentions, expectations, and interpretation of anticipatability. Also, the AOI-reachable road users vary depending on the scenes within the scenarios in timely manner.

4.1.3. Scene Analysis (Scenario Decomposition)

4.1.3 decomposes scenarios into multiple constituent scenes along a timeline. 4.1.2 and 4.1.3 are also associated with each scene. This scene-based analysis allows for more extensive analysis, such as potential hazards during decision making and the context-dependent factors that may lead to space-sharing conflicts within the AOI. Figure 5 shows an example of a 4-scene decomposition of the scenario, which was studied in 3.1. However, conventional SOTIF-focused STPA typically captures only the final scene where hazardous behavior occurs—scene t in Figure 3. The temporal progression through backward scenes (t → t-1→ t-2→t-3) and their contribution to establishing hazardous conditions may not be explicitly modeled, limiting the ability to identify early intervention points.

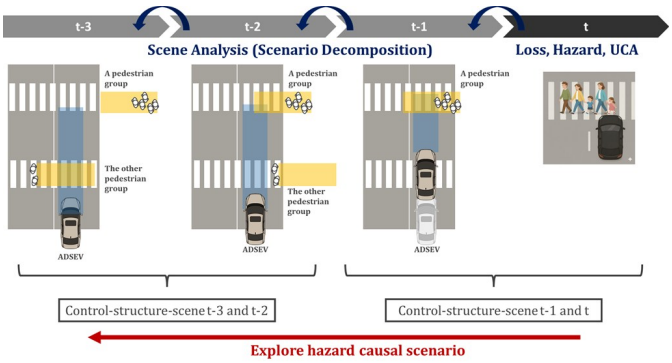

Figure 5. Scene analysis

- **Scene t:** Before a pedestrian group (PedG) finishes crossing, ADSEV performs an avoidance maneuver, departing from its original path to circumvent the intersection area, while the pedestrians are still on the crosswalk.
- **Scene t-1:** A PedG is mid-crossing when ADSEV reaches the intersection. ADSEV recognizes the ongoing pedestrian crossing and keeps on approaching the crosswalk very slowly.
- **Scene t-2:** A PedG begins crossing the road while ADSEV continues approaching to the crosswalk area. The other PedG finishes crossing the road.
- **Scene t-3:** ADSEV approaches the crossroad while a PedG is crossing the crosswalk, and another PedG is positioned at opposite side of the crosswalk, preparing to cross.

4.1.4. Requirements analysis

4.1.4 derives safety requirements, which are allocated to multiple stakeholders across the SSA pillars in roadmanship and traffic perspectives. This structure extends beyond the scope of vehicle-centric SOTIF. Within this step, the trade-offs between safety margin and traffic smoothness are considered. A derived requirement, within the scenario under consideration for illustrative purposes, is: Req tra-1: ADSEV shall proceed through the intersection without conflicting with other road users (or within T1) after entering the intersection and recognizing that its trajectory is clear. T1 is initially defined as an acceptable time to be spent in the intersection in a functional viewpoint. We will specify this criteria in 4.3.1.

4.1.5. Adaptation of the analysis purpose

4.1.5 builds on the analysis results (i.e., losses (L-1, L-2), hazards (H-1, H-2), safety constraints (SC-1, SC-2)), which were identified through conventional STPA application within the SOTIF framework (see Section 3). 4.1.5 extends the Losses as follows: **Additional Losses:** L-3: Violation of traffic laws leading to legal consequences.
L-4: Significant disruption to traffic flow, undermining social acceptance.
Correspondingly, 4.1.5 extends the hazards to address the addi-

tional losses: **Additional Hazards:** H-3: The ADS operates in a manner that violates critical traffic regulations (link to L-3). H-4: The ADS's behavior is overly cautious or unpredictable, causing significant traffic gridlock (link to L-4). **Additional safety constraints:** SC-3: The AV must comply with all applicable traffic laws within its ODD. (related to H-3) SC-4: The AV must balance safety with traffic efficiency, avoiding unnecessary stops or overly defensive maneuvers. (related to H-4)

This extension reflects the multiple requirements including the ones related to roadmanship, acknowledging that safe ADSs should not only avoid collisions but also play harmoniously within the AOI's traffic environment.

4.2. WA1-Step2: Extension of control structure modeling

In WA1-Step2, we adapt STPA by extending the existing control structures. Specifically, within the existing control structures, we incorporate sequential and context aspects. STPA-WA Step 2 extends conventional control structure modeling in two ways: (1) introducing an orchestrator for space-sharing coordination among multiple controllers (ADSEV and pedestrians), and (2) explicitly identifying safety-efficiency trade-offs using roadmanship-specific metrics. However, we can leverage the existing results in conventional STPA application within the SOTIF framework focuses on control structures for the system under control and identifies unsafe control actions.

4.2.1. Context analysis within the AOI

4.2.1 analyzes the context within the AOI using a three-layered framework [1], which serves as the foundation for control structures modeling. Since WA1 features a context-aware safety-traffic analysis aspect, the control structures are constructed for each scene identified in 4.1.3. This three-layered framework enables systematic analysis of how AOI-reachable road users interact and share space across different scenes. In what follows, we explain the layers.

Physical Layer captures the ground truth of the traffic environment within the AOI, including; PL1, PL2, PL3 and PL5, and the positions of all AOI-reachable road users. This layer provides the objective physical state that forms the basis for modeling the control structure. STPA-WA guides analysts to define the assumptions, hypothetical conditions, states, and equipment of analysis context in existing control structure's conditions before analysis. At the top of the control structure, which will be discussed in 4.2.2, an ideal orchestrator monitors and analyzes interactions among Safe System stakeholders (P1-P5), enabling context-aware safety-traffic analysis, based on the physical reality of the AOI.

Recognition Layer captures how each road user communicates, perceives, interprets, and forms beliefs through the interaction with other road users' behavior. This layer highlights the process/mental model in each controller of the control structure. These behaviors can be seen as causes of effects of UCAs which we will analyze in 4.4.1. More specifically, recognition embraces: intentions and expectations for space-sharing within the AOI, recognition of other road users' presence and movements, interpretation of traffic rules, signals, and explicit communications (including PL6), implicit communications, knowledge, experience, customs, and strategies applied to the situation, communication and negotiation behaviors. Since intentions and expectations of AOI-reachable road users vary depending on the temporally changing context, an ideal orchestrator needs to enable context-aware safety-traffic analysis.

Action Layer captures the actual control actions executed by each road user based on their recognition and decision-making, including: movement execution (acceleration, braking, steering), communication actions (signals, gestures, horn usage), speed control and trajectory adjustments, space-sharing coordination behaviors. This layer mainly focuses on the control algorithm in each controller of the control structure.

These three layers and scene analysis in 4.3 are considerations for creating the control structure in 4.2.2.

4.2.2. Extended control structure

Since the orchestrator must be context-aware for multi-agent coordination, STPA analyzing this orchestrator must inherently be context-aware as well. The control structure need to be created by considering 4.2.2 constructs control structures based on the sequence aspect:scenario decomposition from 4.1.3 and physical-recognition-action aspects: three-layered context analysis from 4.2.1. As shown in Figure 5, we created two control structures covering each scene to capture the context-dependent interactions within the AOI. As the context evolves from scene t-3 to t, this sequential progression enables more precise identification of causal scenarios and more effective evaluation of safety measures.

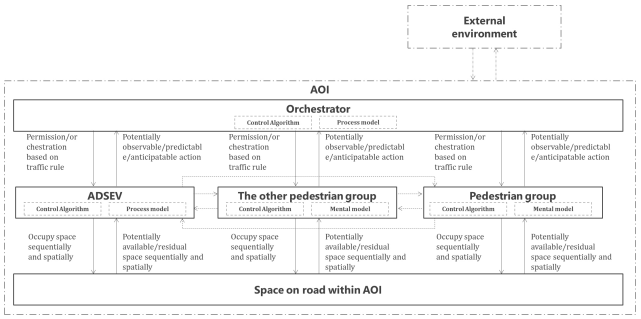

Figure 6. Control structure within the scene t-3 and t-2

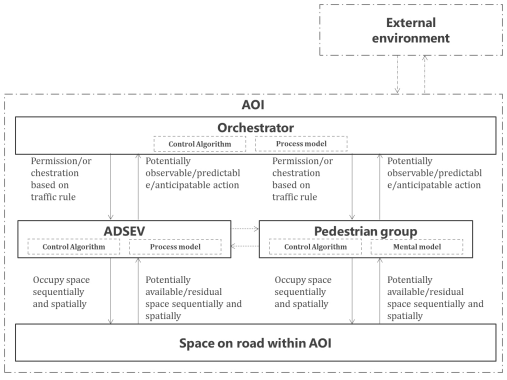

Figure 7. Control structure within the scene t-1 and t

Figure 6 and Figure 7 illustrate the control structures of each scene. As the top of the controller, an ideal orchestrator component coordinates space sharing, based on traffic rules (P2).

Controllers in control structures As mentioned, the orchestrator monitors and coordinates space-sharing among road users based on traffic rules and regulations. It provides permission

or orchestration for road users to share space sequentially and spatially within the AOI.

Underneath of the orchestrator, there is a consumer layer of space within the AOI, which includes components for the AOI-reachable road and some movable objects. For example, **ADSEV**: The subject vehicle equipped with an automated driving system, **Pedestrian Groups**, comprising a Control Algorithm and Process Model.

The components within the consumer layer try to perceive the potentially available/residual space within the AOI and execute control actions to occupy space sequentially and spatially. Both ADSEV and Pedestrian group execute control actions to *occupy space sequentially and spatially* within the AOI.

External Environment: The broader environment outside the AOI that influences conditions within the AOI. For example, climates, noises, and digital signals. The External Environment influences the AOI through disturbances and contextual changes, affecting the overall system behavior.

Feedback: The Space on Road provides feedback about *potentially available/residual space sequentially and spatially* to both ADSEV and Pedestrian group. Road users observe and perceive potentially observable/predictable/anticipatable actions of other road users through their process models. These control loops are relevant for the context-aware safety-traffic analysis.

Lateral Interactions (Horizontal arrows): Bidirectional communication exists between ADSEV and the pedestrian group, enabling mutual observation and prediction of each other's intentions and actions. This interaction supports implicit and explicit communication to negotiate space-sharing conflicts.

This control structure allows for the modeling of distributed safety responsibilities across multiple road users within the AOI, beyond the scope of vehicle-centric SOTIF.

4.3. WA1-Step3: Extension of UCAs identification

WA1-Step3 focuses on context-aware analysis especially for sharing spaces with multiple agents, we need to set the Metrics for orchestrator.

4.3.1. Trade-offs analysis

Based on the UCAs identified in STPA STEP3 3.1, 4.3.1 analyzes these UCAs from a roadmanship perspective to explicitly identify trade-offs between safety and traffic efficiency. For example, since Hazards and safety constraints are updated in 4.1.5, the UCAs as shown in Table 1 are expected to be updated.

- **ADS-NP1-1:** ADS does not privide manever to drive even when the trajectory is clear(H-4).
- **ADS-D2-A1:** ADS continues to maneuver to stop too long even when the trajectory is clear(H-4).

We observed that the boundaries between safety and traffic efficiency becomes increasingly blurred, as our UCA-based analysis of extended losses, hazards, and safety constraints revealed both positive and negative emergent properties, requiring simultaneous consideration of both aspects.

Metrics serve as roadmanship-specific criteria. Since assumption of Safety Measure which are already implemented in AOI, recognizes that the crosswalk design itself contributes to safety, belonging to the "roads" pillar (P2) within the SSA, expected trajectories of the ADSEV and the pedestrian group are in a*CP*. The appropriate safety metrics are adopted as follows.

Metric 1: *T1*, as previously defined 4.1.4 in requirement analysis. *T1* is formulated as *Tstay* in Formula 3 based on PET.

$$T_{(\mathrm{i})}^{stay} = t_i^{\mathrm{out}} - t_i^{\mathrm{in}} \tag{3}$$

where $T_{(\mathrm{i})}^{stay}$ denotes the acceptable time to be spent in the intersection; t_i^{in} and t_i^{out} denote the time when the road user i enters, respectively exits the AOI.

Metric 2: PET is set as metric to achieve roadmanship (related to 4.1.4 Req tra-1).

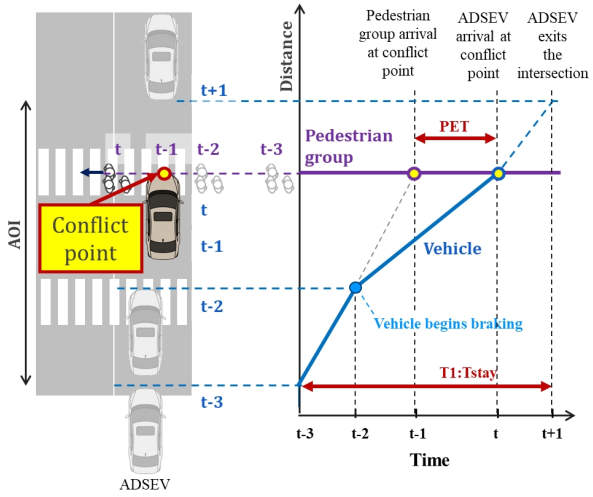

Figure 8. Road Safety Socio-technical System&Stakeholders

As shown in Figure 8, each behavior of the ADSEV and the pedestrian may roughly be distinguished.

As an example, conservative driving behavior: the vehicle stops even when the pedestrian is 2m away from the crosswalk. This results in larger *Tstay* and *PET* values. As another example, aggressive driving behavior: the vehicle continues driving even when the pedestrian is 1m away from the crosswalk. This results in smaller *Tstay* and *PET* values.

As the same concept, SC-1 3.1 (larger *Tstay* and *PET*) and Req tra-1 4.1.4(larger smaller*Tstay* and *PET*) are in a tradeoff relationship.

Through the identification of these trade-offs, 4.3.1 allows for comprehensive dependability analysis. This balances safety, efficiency, and social acceptance.

These trade-offs may be triggerd by 1) other road users' actions rather than by ADSEV itself, and 2) interactions with other road users rather than ADSEV's behavior alone. We identify UCAs for each road user from physical, recognition, and behavioral aspects, as established in the three-layered framework. To find trade offs between requirements and safety constraints (including UCAs), we conduct trade-off analysis. This context-aware analysis enables us to hypothesize the causal process of failures in decision-making that may lead to hazards across different scenes.

Table 2. Unsafe Control Actions for Pedestrian behaviors

Control Action	Not providing causes hazard	Providing causes hazard	Providing too early, too late	Stop too soon, applied too long
PED-CA1 crossing (including entering and traversing the crosswalk)	no UCA	**PED-P1-1** PedG crosses when ADSEV is approaching within unsafe distance (H-1, H-2)	**PED-T1-E1** PedG starts crossing too early when ADSEV is approaching within unsafe distance (H-1, H-2)	**PED-D1-A1** PedG continues crossing too long when ADSEV is approaching within unsafe distance (H-1, H-2)
PED-CA2 waiting/stopping (including preparing to cross and maintaining stopped state)	**PED-NP2-1** PedG does not wait or stop when ADSEV is approaching within unsafe distance (H-1, H-2)	no UCA	**PED-T2-L1** PedG stops too late when ADSEV is approaching within unsafe distance (H-1, H-2)	no UCA

Based on the control structures, as shown in Figure 6 and 7, we added the UCA related to pedestrian group (abbreviated PedG in what follows) as shown in Table 2.

These pedestrian's UCAs and ADSEV's in STPA STEP3 3.1 only show directly lead to hazard equally. However, there might be limitations in prioritizing UCAs in the scene (e.g., from t to t-3) or understanding their contextual explanation within the scenario timeline. To find contextual relationships, we refined UCAs without specific conditions in Table 3. The behaviors, listed in Table 3, are useful to identify HCSs in 4.4.

Table 3. Behaviors of Pedestrian Group (PedG) and ADS

Behavior ID	Description
Beh-ADS-NP2-1	ADS does not maneuver to stop.
Beh-ADS-P1-1	ADS maneuvers to drive.
Beh-ADS-T1-E1	ADS maneuvers to drive too early.
Beh-ADS-T2-L1	ADS maneuvers to stop too late.
Beh-ADS-D1-A1	ADS continues to maneuver too long.
Beh-PED-NP2-1	PedG does not wait or stop.
Beh-PED-P1-1	PedG crosses.
Beh-PED-T1-E1	PedG starts crossing too early.
Beh-PED-T2-L1	PedG stops too late.
Beh-PED-D1-A1	PedG continues crossing too long.

4.4. WA1-Step4: Extention of loss scenarios identification

WA1-Step4 identified causal factors of UCAs, where safety measures are planned and their effectiveness is predicted.

4.4.1. Hazard Causal Scenario Identification

Since STPA-WA aims at extending the analysis to the stakeholders of each pillar of the safe system, within this subsection, we extend the analysis to identify hazard causal factors caused by 1) other road users' actions rather than by ADSEV itself, and 2) interactions with other road users rather than ADSEV's behavior alone.

This section presents the HCS analysis using ADS-T2-L1 and HCS-ADS-T2-L1-1 as an example. The combination of behavior shown in Table 3 supports the identification and explanation of pedestrian's behavior, which potentially triggers ADSEV's hazardous behaviors as HCSs.

WA1-HCS-ADS-T2-L1-P1)-1 ADS believes that the path is clear and ADSEV can proceed without braking. Motivation: the crosswalk had faded road markings due to aging, and remained listed as 'planned' status in the map database without updates and ADS could not reliably detect the faded markings under low-light conditions.

WA1-HCS-ADS-T2-L1-P2)-1 ADS believes that ADSEV can proceed without braking, based on the absence of detected obstacles and map data (e.g., temporary construction), failing to account for the reduced sensor range (blind spot) caused by the road geometry (e.g., upward gradient).

WA1-HCS-ADS-T2-L1-P3)-1 ADSEV has already spent time in the intersection beyond T1, ADS tries to pass when the pedestrian group exits from the expected trajectory of ADSEV.

WA1-HCS-ADS-T2-L1-P3)-2 ADS believes that ADSEV can pass without maneuver to stop (e.g., braking) based on its prediction of the pedestrian speed, location and expected trajectory. But, the predictions are insufficient to provide decisions for ADS stop, due to the following pedestrian factors from WA1-HCS-ADS-T2-L1-P4)-1 to WA1-HCS-ADS-T2-L1-P4)-4.

WA1-HCS-ADS-T2-L1-P4)-1 (related to Beh-PED-T1-E1) - PedG runs toward the crosswalk and attempts to cross swiftly, because PedG is in a hurry and misinterpreted the ADS's slow speed as an indication of yielding.

WA1-HCS-ADS-T2-L1-P4)-2 (related to Beh-PED-P1-1) PedG runs toward the crosswalk and attempts to cross swiftly, because it knows ADSEV behavior (e.g., when ADSEV would restart after some symptoms are observed) and it has recognized the same symptoms, meaning there is some time until ADSEV starts moving.

WA1-HCS-ADS-T2-L1-P4)-3 (related to Beh-PED-P2-1) PedG stops in the middle of the crosswalk to pick up a child's shoe that fell off.

WA1-HCS-ADS-T2-L1-P4)-4 (related to Beh-PED-NP2-1) - PedG runs toward the crosswalk and attempts to cross swiftly, because it observed that ADSEV had stopped waiting for another PedG to finish crossing, which made time for it to cross the crosswalk.

In some cases, UCAs occurring at earlier time points may become causes of UCAs at later time points. By organizing these inter-UCA causal relationships, it becomes possible to verify the effectiveness of safety measures by identifying optimal intervention points in the causal chain.

4.5. WA2: Stakeholder Feedback Loops & WA Evolution

WA2 evaluates the effectiveness of the safety measures allocated to the pillars of the safe system. Safety measures within P1 include maintenance and continuous monitoring. This encom-

Table 4. Effectiveness–Capability–Categories Matrix

Effectiveness	Capability	Categories
L	Detectable	Sensing assistance (Visual: road surface)
L	Detectable	Sensing assistance (Visual and auditory)
L	Detectable	Sensing assistance (Electronic system-based)
L	Detectable	Sensing assistance necessary for driving
M	Controllable / Cooperative	Sensing assistance and speed control necessary for driving
M	Controllable	Sensing assistance (road elevation) and speed suppression
H	Coordinated	Traffic control (management of time and spatial occupancy)
H	Avoidable	Spatial separation (temporary)
H	Avoidable	Spatial separation (permanent)
N/A	Mitigatable	Damage mitigation upon collision
N/A	Communicable	Communication support

passes continuous evaluation of the effectiveness of measures allocated to other pillars of the safe system and their updates as needed. After implementing safety measures in the AOI, the upper layer (P1) in Figure 8 evaluates whether updates are necessary based on inputs such as public opinion, long-term monitoring of near-misses and accidents, and other feedback.

4.5.1. Safety measures effectiveness evaluation

Since safety measures allocated to people always face human errors (seen as symptoms of safety design), WA2 focuses on safety mechanisms that are allocated to P2 and P3. We propose to classify the effectiveness of the safety mechanisms which means to increase the calculated results of TTC and PET based on Formula 1 and 2: High (H): Avoidable, Coordinated, Middle (M): Controllable, Cooperative, Low (L): Detectable, Not applicable (N/A): Mitigatable.

Instinctive safety measure: H
P2) Pedestrian bridges, underpasses (complete grade separation), P2) Movable bollards, automatic barriers, P2) Traffic signals, push-button type. The The safety related uncertainties are expected to be addressed by P1 (e.g., maintenance, periodical checks and monitoring).

Safety measure: M
P2) Speed bumps and steps, P2) V2I, P3) Sign recognition and map integration, P3) Pedestrian detection and alert by cameras and sensors. The residual risks are expected to be addressed by P3 (e.g., SOTIF, functional safety).

Safety measure: L
P2) Road paint, 3D optical illusion, P2) Light (beacon, embedded type in Osaka, nighttime lighting) + sound, P3) Perception by cameras and sensors.

The road safety management shall be distributed over the SSA pillars. The evaluation of effectiveness aims to: 1) identify opportunities for further enhancements of safety measures, 2) plan strategies and actions for safety related uncertainties and 3) address uncertainties and exacerbating factors that may affect the performance of safety measures. This predictive approach could be aligned with ISO 21448 [8], clause 13 operational safety.

5. Related Work

In the literature, several works report about the application of STPA. Some of these works aim at describing how STPA can be applied to specific automotive systems; some also provide suggestions for adaptation in terms of e.g., further guidance including prescriptive guidance in terms of models, methods to be used at each STPA's step. Khastgir et al. [13] propose a systems approach to creating test scenarios for automated driving systems by linking STPA analysis directly to test scenario generation. Their approach demonstrates how safety analysis results can be translated into testable scenarios. This aligns with our goal of creating structured scenarios from STPA results. Oginni et al. [17] enrich the STPA methodological approach with a set of system engineering models such as rich pictures, pig diagrams, context diagrams, etc. Their work focused on rail systems. However, their prescriptive guidance may find application also in the automotive domain. Zhang et al. [24] develop a methodology that extends the STPA framework by incorporating a causal scenario classification system and a complex network-based evaluation to analyze SOTIF-related hazardous factors. Their work focuses on intelligent railway driving assistance systems but provides valuable insights for our extension to ADSs. Agatsuma et al. [1] propose a scenario model for the integrated evaluation of safety and smooth road traffic interactions in ADSs. Their work addressed similar concerns regarding the trade-offs between safety and traffic flow, using a three-layer scenario model comparable to our approach. Nouri et al. [16] propose to extend STAMP/STPA to target distributed development and multi-abstraction levels. Yang et al. [23] apply STPA focusing on steps 2 and 3. They advocate that while conducting the STPA the scope shall be limited to ensure a reduced number of scenarios to be considered. Their contribution consists in the identification of control structures and corresponding unsafe control actions. The identified hierarchical control structures are based on the strategic, tactical, and operational functions defined in SAE J3016 [20]. Their focus is on the ego-vehicle while our focus goes beyond the ego-vechicle. Elizebeth et al. [5] apply STPA on an Automated Lane Keeping Systems (ALKS, [22]). Their contribution lies in demonstrating the applicability of STPA on complex systems and its effectiveness as a means for analysing unsafe control actions which then could be used by policy makers for enriching regulations. However, their focus is also on ego-vehicle.

6. Conclusion and Future Work

To conduct risk analysis in the automotive domain, one suggested process is STPA. However, applying STPA in the automotive domain has been reported to be challenging. Various works have provided adaptations and extensions. In this paper, we explained our perspective and why a novel adaptation is needed. Then, we provided yet another novel adaptation

of STPA, which we called STPA-WA. STPA-WA embraces roadmanship and the different perspectives represented by the different pillars belonging to the safe system approach. By so doing a more effective analysis can be conducted. As future work, we aim at validating STPA-WA by conducting case-based research considering other ADSs-related cases. We also intend to augment the analysis with scenario-based simulations and measure roadmanship's performance and show the delta between conventional application of STPA and STPA-WA. Finally, we also aim at providing evidence regarding its generalisability within, as well as beyond, the automotive domain.

7. Acknowledgment

We used: Claude to translate some Japanese paragraphs in English; ChatGPT, Gemini to create images; Writefull to check spelling and grammar. We performed final reviews and edits.

References

[1] R. Agatsuma, T. Takai, and M. Okada. Construction of a scenario model for the integrated evaluation of safety and smooth road traffic interactions in automated driving systems. IEICE Technical Report, 2023.

[2] F. H. Amundsen and C. Hyden. . In First workshop on traffic conflicts. Oslo, Norway, 1977.

[3] City of Austin. Autonomous vehicles, n.d. URL `https://www.austintexas.gov/page/autonomous-vehicles`.

[4] de Gelder, E. and Op den Camp, O. and de Boer, N. Scenario categories for the assessment of automated vehicles. Technical report, CETRAN, VMAD-SG1-11-03, 2020.

[5] M. J. Elizebeth, S. Khastgir, and P. Jennings. Hazard analysis of an automated lane keeping system using systems-theoretic process analysis. Accident Analysis & Prevention, 221:108171, 2025. ISSN 0001-4575.

[6] L. Fraade-Blanar, M. S. Blumenthal, J. M. Anderson, and N. Kalra. Measuring Automated Vehicle Safety: Forging a Framework. RAND Corporation, Santa Monica, CA, 2018. doi: 10.7249/RR2662.

[7] ISO/TC 22/SC 32. ISO 26262:2018 Road vehicles – Functional safety, 2018.

[8] ISO/TC 22/SC 32. ISO 21448:2022 Road Vehicles – Safety of the intended functionality, 2022.

[9] ISO/TC 22/SC 32. ISO/TS 5083:2025 Road vehicles – Safety for automated driving systems — Design, verification and validation, 2025.

[10] ISO/TC 22/SC 33. ISO 34503:2023 Road Vehicles – test scenarios for automated driving systems – specification for operational design domain, 2023.

[11] ISO/TC 22/SC 36. ISO TS 21934-2:2024 Road Vehicles – Prospective safety performance assessment of pre-crash technology by virtual simulation, 2024.

[12] ISO/TC 241. ISO 39001:2012 Road traffic safety (RTS) management systems - Requirements with guidance for use, 2012.

[13] S. Khastgir, S. Brewerton, J. Thomas, and P. Jennings. Systems approach to creating test scenarios for automated driving systems. Reliability Engineering & System Safety, 215:107610, 2021.

[14] Leveson, N. G. and T., John P. STPA handbook. Cambridge, Massachusetts, U.S., 2018.

[15] G. Markkula et al. Defining interactions: a conceptual framework for understanding interactive behaviour in human and automated road traffic. Theoretical Issues in Ergonomics Science, 21(6):728–752, 2020.

[16] A. Nouri, C. Berger, and F. Törner. On STPA for Distributed Development of Safe Autonomous Driving: An Interview Study. In 2023 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pages 5–12, 2023.

[17] D. Oginni, F. Camelia, M. Chatzimichailidou, and T. L.J. Ferris. Applying system-theoretic process analysis (stpa)-based methodology supported by systems engineering models to a uk rail project. Safety Science, 167:106275, 2023. ISSN 0925-7535.

[18] M. Okada and B. Gallina. Safety of the intended functionality of external human interfaces: Gaps and research agenda. In 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), pages 578–583. IEEE, 2024.

[19] R. Patriarca, M. Chatzimichailidou, N. Karanikas, and G. Di Gravio. The past and present of system-theoretic accident model and processes (stamp) and its associated techniques: A scoping review. Safety Science, 146: 105566, 2022. ISSN 0925-7535.

[20] SAE J3016. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems, 2021.

[21] M. Scholtes et al. 6-layer model for a structured description and categorization of urban traffic and environment. IEEE Access, 9:59131–59147, 2021.

[22] UNECE. *UN Regulation No 157 – Uniform provisions concerning the approval of vehicles with regards to Automated Lane Keeping Systems [2021/389]*, January 2021.

[23] Y. Xuezhu, D. Zhongju, J. Pimentel, R. Johansson, G. Gruska, X. Ruoyu, and X. Fu. Identifying Effective STPA Control Structures to Characterize SOTIF Areas 1, 2, 3, and 4 in Automated Vehicles. In 27th International Technical Conference on the Enhanced Safety of Vehicles (ESV), 2023.

[24] 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. Applied Sciences, 11(16):7714, 2021.