Institut pour la Maîtrise des Risques
Sûreté de Fonctionnement - Management - Cindyniques

Quel rôle pour la SOTIF (ISO 21448:2022) dans la démonstration de sécurité des trains autonomes ?

What role for SOTIF (ISO 21448:2022) in the safety assurance of autonomous trains?

Ouail HIMRANE
IRT Railenium
Valenciennes
ouail.himrane@railenium.eu

Abhimanyu TONK
IRT Railenium
Valenciennes
abhimanyu.tonk@railenium.eu

Abderraouf BOUSSIF
Université Gustave Eiffel
Villeneuve d'Ascq
abderraouf.boussif@univ-eiffel.fr

Résumé — Dans le domaine automobile, la sécurité fonctionnelle est régie par la norme ISO 26262 :2018. Cette norme établit un processus et définit des exigences concernant la conception et le déploiement des systèmes électriques/électroniques (E/E) embarqués sur les véhicules afin de garantir leur sécurité. Le principal focus des différentes parties de cette norme concerne les défaillances aléatoires (du matériel) et systématiques (matérielles et logicielles) des systèmes E/E embarqués. Avec l'arrivée des véhicules autonomes, et leurs problématiques sécuritaires, une extension normative (ISO 21448 :2022) nommée SOTIF (*Safety Of The Intended Functionality*) a été élaborée pour adresser principalement les risques liés à la mauvaise interaction du système avec les acteurs externes, les défaillances liées aux insuffisances en termes de spécifications, ainsi que les limitations en termes de performances des E/E systèmes. Dans le domaine ferroviaire, la spécification et la démonstration de la sûreté de fonctionnement des systèmes ferroviaires sont régies par la série de normes EN50126/8/9, ainsi que la méthode de sécurité commune relative à l'évaluation des risques (CSM-RA) au niveau européen. Aujourd'hui, l'automatisation et l'autonomisation des systèmes ferroviaires (i.e., trains autonomes) incitent à réévaluer l'adéquation de s'appuyer (uniquement) sur les normes actuelles pour assurer un niveau de sécurité acceptable des systèmes ferroviaires autonomes. Ainsi, le besoin de l'étude de la faisabilité (ainsi que la méthodologie) de consolider les normes EN5012x avec les aspects de la SOTIF est bien justifié. L'objet de cette communication est de contribuer à établir un cadre générique permettant d'intégrer la SOTIF dans le processus de spécification et de démonstration de sécurité des systèmes ferroviaires autonomes. En outre, nous présentons une analyse de cas portant sur le système de monitoring de l'environnement afin de décrire l'application pratique de la SOTIF dans le contexte spécifique des trains autonomes.

Mots-clefs — Assurance de sécurité, sécurité des systèmes autonomes, voitures autonomes, trains autonomes, sécurité de la fonctionnalité attendue (SOTIF)

Abstract — In the automotive sector, functional safety is governed by the ISO 26262:2018 standard. This standard establishes a V-cycle process and defines requirements to ensure the safe design and deployment of on-board electrical/electronic (E/E). The main focus of the various parts of this standard is on random (hardware) failures and systematic (hardware and software) faults in E/E systems. With the advent of autonomous vehicles, and their safety issues, a normative extension (ISO 21448 :2022) named SOTIF (*Safety Of The Intended Functionality*) was developed to address mainly risks related to system misinteraction with the external actors, failures related to specification insufficiencies, and limitations in terms of E/E system performance. In the railway sector, the specification and demonstration of the safety of railway systems are framed by EN50126/8/9 standards, as well as the Common Safety Method for Risk Assessment (CSM-RA) at European level. Today, the automation and autonomization of rail systems (i.e., autonomous trains) are prompting a reassessment of the suitability and the sufficiency of relying (solely) on current standards to ensure an acceptable level of safety for autonomous rail systems. Thus, question of whether (and how) to reinforce EN5012x standards with the SOTIF framework is well justified. The purpose of this paper is to contribute to establishing a generic framework for integrating SOTIF into the process of specifying and demonstrating the safety of autonomous railway systems. In addition, we include a case analysis of the environmental monitoring system to describe the practical application of SOTIF in the specific context of autonomous trains.

Keywords — Safety assurance, safety of autonomous systems, autonomous vehicles, autonomous trains, safety of the intended functionality (SOTIF)

I. Introduction

Les systèmes de transport sont des systèmes critiques dont les défaillances peuvent entraîner des pertes considérables (i.e. dommages aux équipements et à l'environnement, blessures graves aux personnes ou même la perte de vies humaines) (Himrane, 2021,2023). Toutefois, les systèmes de transports ferroviaires sont reconnus comme des moyens de transport très sûrs. Ceci est notamment attribuable au fait que la sécurité ferroviaire repose sur des principes conservateurs établis et renforcés par un cadre normatif et règlementaire strict. En particulier, ce cadre exige la mise en place d'une démarche de gestion de risque et de mise en sécurité préalablement à toute modification/évolution dans le système ferroviaire existant ou toute conception d'un nouveau système. Concrètement, cette démarche vise à assurer la non-régression du niveau de sécurité global du système ferroviaire, notamment au travers d'un processus rigoureux ayant pour objectif de cadrer les activités fondamentales d'identification des dangers, d'évaluation des risques, ainsi que la mise en place des mesures de sécurité adéquates.

Dans le contexte actuel, marqué par une poussée vers l'autonomisation motivée par l'intégration accrue des systèmes basés sur l'intelligence artificielle (IA) substituant des fonctions traditionnellement assurées par les humains, une transformation majeure se dessine dans les systèmes de transport. En particulier, les systèmes de conduite autonome (*ADS – Automated Driving Systems*) se profilent comme une perspective prometteuse. Dans le domaine ferroviaire, l'ambition est de tirer profit de l'utilisation de l'intelligence artificielle pour améliorer la ponctualité des trains, augmenter la capacité des voies ferrées (sans nécessité de poser de nouvelles voies), accroître l'efficacité opérationnelle et réduire la consommation d'énergie. Toutefois, la sécurité demeure la première préoccupation, conformément à la règle immuable selon laquelle ``*il est strictement interdit de dégrader le niveau de sécurité du système*''.

D'autre part, ce contexte d'évolution rapide et de transformation technologique disruptive a fait apparaitre certaines limites des normes de sécurité actuelles face aux défis posés par l'autonomisation et l'utilisation des systèmes d'intelligence artificielle (SIA). En effet, les activités d'assurance et de démonstration de sécurité des systèmes complexes autonomes se heurtent à plusieurs obstacles à surmonter (e.g., la spécification du comportement du système autonome ; l'impact des interactions des systèmes autonomes avec leur environnement opérationnel, qui peuvent présenter des situations inconnues et potentiellement dangereuses).

Dans cette communication, notre contribution vise à étudier l'apport de la considération de la norme SOTIF (*Safety Of The Intended Functionality* - ISO 21448:2022), initialement développée pour le domaine automobile, pour faire face aux défis (en matière de sécurité) inhérents aux SIA dans le domaine ferroviaire. L'objet de cette communication est de contribuer à établir un cadre générique permettant l'adoption et l'intégration de la SOTIF dans le cadre normatif actuel de la sécurité ferroviaire, particulièrement, dans le processus de spécification et de démonstration de sécurité des trains autonomes. Dans cette optique, nous commençons par synthétiser le processus de démonstration de sécurité ferroviaire tel que défini par cadre normatif et réglementaire actuel avant d'identifier les difficultés et les limites de s'appuyer uniquement sur ce processus quand il s'agit des systèmes ferroviaires autonomes. De plus, nous analysons la norme SOTIF et nous montrons comment elle pourrait compléter le cadre normatif actuel pour répondre aux défis sécuritaires des systèmes de transport autonomes. Au travers d'une analyse de cas portant sur le système de monitoring de l'environnement dans le contexte spécifique des trains autonomes, nous illustrons l'adoption concrète de la SOTIF dans le secteur ferroviaire. Enfin, nous incluons une discussion portant sur la proposition d'un cadre général permettant d'intégrer les activités de la SOTIF dans le cycle global de la démonstration de sécurité ferroviaire.

II. Contexte normative de la securite des systemes de transports

A. La norme mère IEC 61508

La IEC 61508 est la norme de référence en matière de sécurité fonctionnelle applicable aux systèmes relatifs à la sécurité dans tous les domaines qui intègrent des dispositifs électriques/électroniques/électroniques programmables (E/E/PE). C'est également la norme mère qui a été utilisée pour rédiger les normes de sécurité spécifiques aux divers domaines industriels, telles que l'EN 50126 pour le domaine ferroviaire et l'ISO 26262 pour l'automotive.

Selon la norme IEC 61508, le concept fondamental de sécurité est que tout système relatif à la sécurité doit soit fonctionner correctement. Dans le cas d'une défaillance, cette dernière doit faire passer le système dans un état sûr (*fail-safe*). Cette norme de sécurité couvre spécifiquement les dangers qui surviennent lorsque les fonctions de sécurité échouent. L'objectif principal de la IEC 61508 est donc de réduire le risque associé à une défaillance dangereuse à un niveau acceptable.

La IEC 61508 repose sur deux piliers fondamentaux : (i) le cycle de vie de la sécurité, qui vise à réduire ou à éliminer les défaillances dues à des causes systématiques au cours du développement et de l'exploitation du système, et (ii) l'approche probabiliste pour traiter les défaillances aléatoires dangereuses des composants matériels par le biais des niveaux d'intégrité de la sécurité (SILs).

Ce concept est renforcé par le fait que le système doit être développé, validé et évalué en fonction d'exigences spécifiques résultant de l'identification des dangers et de l'analyse des risques.

B. Cadre normatif de la sécurité ferroviaire

La mise en service d'un système dans le transport ferroviaire est liée à la mise en œuvre du référentiel CENELEC (EN50126/28/29) couvrant principalement les aspects systèmes, matériel et logiciel. Bien qu'applicable initialement aux systèmes de signalisation, ce référentiel reste généralisable aux autres (sous-)systèmes ferroviaires.

Le cadre général pour assurer la sécurité des systèmes ferroviaires est défini par la norme EN 50126 portants sur la spécification et la démonstration de la fiabilité, disponibilité, maintenabilité et sécurité (FMDS). La norme EN 50126 considère le système

ferroviaire (y compris les opérateurs humains) au sein d'un environnement physique et opérationnel donné. La norme prend en considération (dans sa partie 1) les aspects génériques du cycle de vie FMDS du système (y compris le rôle du facteur humain dans ces phases), et prescrit (dans sa partie 2) une démarche de gestion des risques dans le cadre du cycle de vie du système.

Dans ce cadre de référence, la démonstration de sécurité doit être argumentée et présentée dans un dossier de sécurité, pour une évaluation indépendante par une tierce partie. Pour ce faire, le système est développé conformément aux normes EN 50126 (pour le système global) et EN 50128 (pour les logiciels sécuritaires). En outre, un système relatif à la sécurité dans le domaine ferroviaire doit être conçu pour un environnement opérationnel spécifique, ce qui nécessite une définition claire des règles d'exploitation et de maintenance, ainsi que la prise en compte des facteurs externes (telles que le climat). Par conséquent, le document « conditions d'utilisation relatives à la sécurité (SRAC) » doit englober : les conditions, règles et contraintes relatives à la conception, à l'exploitation et à la maintenance du système, ainsi que la manière de les vérifier, conformément à la norme EN 50129. D'autre part, la sécurité du système en cas d'influences externes doit également être démontrées dans le cadre de la démonstration de sécurité. En conséquence, le dossier de sécurité n'est valable que dans le cadre spécifié des influences externes, tel que défini dans la spécification des exigences du système.

C. Cadre normatif pour les véhicules routiers

Le cadre normatif de la sécurité des véhicules routiers concerne principalement la sécurité fonctionnelle des systèmes et équipements E/E. Cet aspect est traité par la norme multipartite ISO 26262. Cette norme est une adaptation de la norme de sécurité fonctionnelle IEC 61508 pour assurer la sécurité fonctionnelle des systèmes E/E à bord des véhicules. Suivant un modèle en V pour le développement au niveau du système, du matériel (HW) et du logiciel (SW), la norme ISO 26262 porte sur la gestion de la sécurité fonctionnelle des systèmes et équipements E/E automobiles vis-à-vis des défaillances aléatoires et systématiques (du matériel et logiciel) tout au long du cycle de vie.

Concrètement, la norme ISO 26262 vise à éliminer les risques causés par le dysfonctionnement des systèmes électriques et électroniques des véhicules. Un tel dysfonctionnement du système peut être dû à une défaillance ou à un comportement non intentionnel du système par rapport à la conception prévue. Dans ce contexte, le risque associé aux situations opérationnelles dangereuses est évalué qualitativement au moyen des niveaux d'intégrité de la sécurité automobile (ASIL). Par ailleurs, des mesures de sécurité doivent être définies afin d'éviter ou de contrôler les défaillances systématiques et de détecter ou de contrôler les défaillances matérielles aléatoires ou bien d'en atténuer les effets. À ce titre, la norme ISO 26262 décrit la manière de concevoir, de vérifier et de valider un système de sécurité automobile.

D. Cadre normatif pour les véhicules autonomes

Avec l'émergence des systèmes de conduite automatisée (*ADS*) comme la composante technique principale permettant l'automatisation de la conduite des voitures et la substitution aux tâches réalisées auparavant pour un conducteur humain, il s'est avéré que se conformer uniquement à la norme ISO 26262 ne suffit pas à garantir la sécurité des voitures autonomes. En effet, la complexité inhérente de l'ADS, tant au niveau des fonctionnalités (perception de l'environnement, compréhension de la situation, évaluation de risques, planification, etc.) que technologies utilisés (intelligence artificielle et techniques d'apprentissage) engendre de nouveaux dangers qui dépassent le cadre de la sécurité fonctionnelle. Ainsi, plusieurs initiatives de standardisation ont été proposées pour supporter la norme ISO 26262 dans la démonstration de sécurité des voitures autonomes.

La norme ANSI/UL4600[1] a comme objectif de garantir qu'une prise en compte suffisamment approfondie de la sécurité d'un véhicule autonome a été effectuée pendant le processus de développement et continuera effectivement tout au long du cycle de vie du système. Ainsi, elle vise à couvrir les principes de la sécurité, les méthodes et le processus de cycle de vie nécessaires pour la préparation et l'évaluation de la démonstration de sécurité (à travers le dossier de sécurité) des voitures autonomes. Concrètement, la norme ANSI/UL4600 prescrit les éléments sur lesquels le dossier de sécurité des véhicules autonomes doit se concentrer et la manière dont le dossier de sécurité devra être évalué. Il est à noter que la norme ANSI/UL4600 est alignée avec la norme ISO 26262 et est spécifiquement destinée à la conduite autonome. De manière générale, la norme ANSI/UL4600 ne prescrit pas les technologies ou les architectures à employer, mais elle exige que le dossier de sécurité présente des arguments convaincants en faveur de l'assurance sécurité de l'ADS, en particulier sur la base de vérification formelle, de simulations, d'essais en laboratoire et d'essais sur la voie publique.

Par ailleurs, il est noté que la norme ISO 26262 porte sur la sécurité fonctionnelle des équipements E/E automobiles en cas de défaillance du matériel et du logiciel. Toutefois, cette norme ne couvre pas la sécurité du véhicule en l'absence de défaillance de l'équipement E/E, par exemple en cas de dysfonctionnement de l'ADS dû à une erreur du conducteur ou à des changements imprévus dans un environnement opérationnel complexe. C'est précisément dans ce but qu'a été élaborée la norme ISO 21448, désignée par l'acronyme SOTIF (*Safety Of The Intended Functionality* - Sécurité de la fonctionnalité attendue). La norme ISO 21448 a été proposée dans sa première version (2019) comme une norme complémentaire à l'ISO 26262 avec un focus particulier sur les véhicules équipés de système d'aide à la conduite (niveaux 1 et 2). La nouvelle version de la norme (2022) s'étend pour couvrir la conduite autonome aussi (niveaux 3, 4, et 5). Cette norme, qui représente aujourd'hui une pierre angulaire dans la démonstration de sécurité des voitures autonomes, s'intéresse aux dangers et risques engendrés par l'insuffisance des fonctionnalités de l'ADS, lors des opérations d'une voiture autonome dans son domaine opérationnel.

[1] *ANSI/UL 4600 (2023) Standard for Safety for the Evaluation of Autonomous Products.*

Concrètement, la SOTIF propose un cadre de gestion des risques engendrés par l'ADS tout au long de cycle de vie du véhicule autonome.

Une autre composante de la sécurité système des véhicules automobiles est la cybersécurité, qui est globalement couverte par la norme ISO/SAE 21434[2]. Pour répondre aux problématiques des voitures autonomes, cette norme est ainsi complétée par le rapport technique ISO/TR 4804 visant à décrire (et prescrire) les étapes de développement et de validation de la sécurité et la cybersécurité des ADS (niveaux 3 et 4). De plus, il s'intéresse à l'intégration et au croisement des objectifs de la cybersécurité avec ceux de la sécurité.

III. Securite de la fonction attendue (SOTIF) – ISO 21448

A. Le besoin de l'étude de la sécurité de la fonction attendue

Dans les véhicules conventionnels, et en présence d'un conducteur humain, il incombe à ce dernier d'assurer la conduite en conditions nominales et de réagir à des situations de conduite imprévues (non couvertes par les procédures établies). Si le conducteur humain devait être remplacé par un système de conduite automatisée (i.e., ADS), ce dernier aurait alors à faire face à des scénarios opérationnels divers issus d'un environnement opérationnel complexe. Il est évident que la conduite d'un véhicule (e.g. train) dans un si vaste éventail de scénarios opérationnels ne peut pas être défini par des règles, étant donné qu'il est pratiquement impossible d'obtenir une spécification complète des tâches exécutées par l'homme. C'est pourquoi, l'utilisation de divers systèmes et software d'IA, dotés des capacités de perception de l'environnement, de traitement et de compréhension de la situation de conduite, ainsi des capacités de décision et d'exécution des tâches de conduite tout en assurant la sécurité des usagers, est devenue une évidence. En effet, L'apport des systèmes basés sur l'IA pour la conduite autonome réside dans leur remarquable capacité d'adaptation, car ils peuvent apprendre à exécuter des tâches impliquant un large éventail de scénarios opérationnels et de grandes quantités de données (représentant des millions de situations réelles et des milliards de scénarios simulés) sans nécessiter de règles explicites pour le comportement du système.

De plus, étant donné que la mise en œuvre de l'ADS dépend de composants physiques (par exemple, des caméras ou un système de télédétection par laser LIDAR- *LIght Detection And Ranging*) pour exécuter ses fonctions, la bonne réalisation des fonctionnalités attendues peut être altérée en raison des limites de performances du matériel, en particulier lorsqu'il est confronté à des scénarios opérationnels imprévus. En outre, des limitations comparables liées aux performances des fonctionnalités peuvent également apparaître pendant la phase de développement de l'algorithme d'apprentissage automatique (ML).

Cette situation a incité le secteur automobile à se pencher sur les comportements dangereux des systèmes résultant de insuffisances de spécifications et de performances dans la conception et la mise en opération du système. C'est dans ce contexte qu'a été élaborée la norme ISO 21448, appelée SOTIF (*Safety Of The Intended Functionality* - Sécurité de la fonctionnalité attendue).

B. Contexte et périmètre de la SOTIF

Un niveau de sécurité acceptable d'un véhicule se justifie par l'absence de tout risque non raisonnable engendré par des dangers associés à des défaillances des fonctionnalités du système et/ou de leur implémentation. Si pour les voitures conventionnelles (avec conducteurs), la principale préoccupation concerne la sécurité fonctionnelle des systèmes techniques (en termes de défaillances matérielle et logicielle) ; alors pour les voitures autonomes, cette préoccupation s'étend aussi pour inclure la sécurité de la fonctionnalité attendue (SOTIF) (en termes d'insuffisance de spécification et de limitations de performance des systèmes techniques).

Les dangers associés à la SOTIF sont causés par des défaillances systématiques liées aux insuffisances de la fonctionnalité attendue. Cette insuffisance fonctionnelle se manifeste, au niveau du véhicule, par une potentielle insuffisance de spécification, et au niveau des éléments (système E/E) soit par une insuffisance de spécification ou bien par une limitation en termes de performances des systèmes E/E (voir Figure 1).

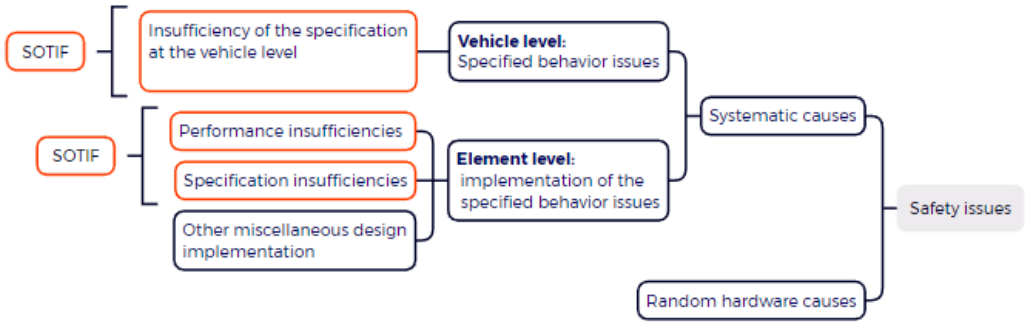

Fig. 1. Contexte et périmètre de la SOTIF

[2] *ISO/SAE 21434 :2021 Road Vehicles – Cybersecurity engineering.*

L'insuffisance de la fonctionnalité attendue est principalement révélée et déclenchée par des conditions opérationnelles et environnementales (*triggering conditions)* spécifiques à des scénarios opérationnels de conduite. Ces conditions de déclenchement peuvent concerner aussi des mauvais usages ou interactions (directes ou indirectes) des opérateurs et acteurs humains avec les systèmes techniques implémentant la fonctionnalité attendue. A noter que les mauvais usages se différencient par leurs liens causals avec le danger. En effet, un mauvais usage direct de la fonctionnalité attendue peut entraîner une condition déclenchante, tandis qu'un mauvais usage indirect de la fonctionnalité attendue peut entraîner une réduction de la contrôlabilité ou une augmentation de la gravité d'un événement dangereux.

La norme ISO 21448 est proposée comme un cadre générique avec un ensemble de processus, activités et méthodes de bonnes pratiques permettant d'assurer et de maintenir les objectifs de la SOTIF au niveau de l'ADS ainsi qu'au niveau du véhicule.

L'objectif de la norme SOTIF consiste à suivre une approche systémique pour l'identification et l'implémentation de mesures de sécurité permettant l'élimination des dangers ou la réduction des risques engendrés par l'insuffisance de la fonctionnalité attendue, durant les trois macro-phases de cycle de vie d'un système : (1) la spécification et le développement, (2) la vérification et la validation, et (3) la mise en opération.

La norme SOTIF présente un ensemble de modèles de cause-effet, au niveau véhicule ainsi qu'au niveau système E/E, permettant d'illustrer l'enchainement des conditions et évènements menant potentiellement à l'occurrence d'une insuffisance fonctionnelle ou d'un danger.

Deux approches sont envisageables pour la réalisation d'une analyse SOTIF :

- Une approche orientée système (*System-based analysis*), i.e., à partir de l'identification d'une potentielle insuffisance fonctionnelle vers la détermination des conditions de déclenchement, ou bien
- Une approche orientée scénarios (*Scenario-based analysis*), i.e., à partir de l'identification des conditions de déclanchement vers la détermination des éventuelles insuffisances fonctionnelles ;

La SOTIF ne définit pas de principes d'acceptation de risques, toutefois, elle fait référence à un ensemble de principes/critères d'acceptation de risques qui peuvent être utilisés ; parmi eux, les trois critères d'acceptation de risques ferroviaires ALARP (As Low As Reasonably Practicable), GAME (Globalement Au Moins Equivalent), et MEM (Minimum Endogenous Mortality). De plus, contrairement aux normes de la sécurité fonctionnelle, la SOTIF ne définit pas d'exigences relatives aux niveaux d'intégrité de sécurité. Ainsi, aucun ASIL (*Automative Safety Integrity Level*) n'est alloué aux fonctions de sécurité à la suite du processus d'appréciation de risques.

C. Objectifs de la SOTIF par rapport aux insuffisances de performances selon les scénarios opérationnels

En évoquant les défis liés aux scénarios opérationnels non anticipés, il est souligné que le monde réel ne peut pas être perçu à 100 % avec exactitude. Cette lacune peut être considéré comme un défi pour la conscience situationnelle (*situational awareness*) du système de conduite autonome (Becker, 2020).

Par ailleurs, une connaissance situationnelle adéquate repose sur une perception suffisamment complète et précise des conditions environnementales et opérationnelles pertinentes. En outre, une compréhension correcte de la scène et un modèle de prévision concernant l'état de chaque acteur impliqué sont nécessaires.

La connaissance situationnelle peut être renforcée par la prise en compte d'informations supplémentaires telles que la localisation et la communication avec d'autres éléments de l'environnement. Ces considérations doivent être intégrées lors de la spécification du domaine de conception opérationnelle (ODD) et pendant le développement du système afin de garantir une fonctionnalité sûre pendant l'exploitation

Selon la SOTIF, les scénarios opérationnels sont classés en quatre groupes : (i) connus et non dangereux, (ii) connus et dangereux, (iii) inconnus et dangereux, et (iv) inconnus et non dangereux. Ces catégories sont représentées sous forme de zones dans la Figure 2.

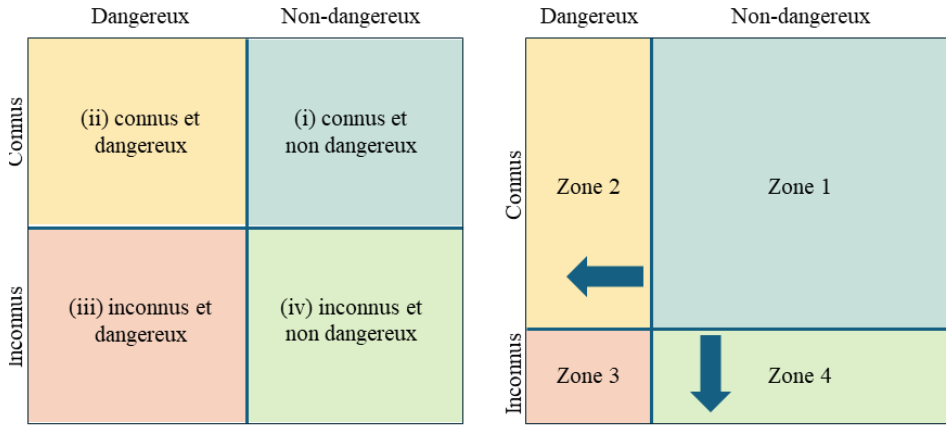

Fig. 2. Catégorisation des scénarios opérationnels.

Au regard de cette catégorisation des scénarios opérationnels, l'objectif principal inhérent à la SOTIF est de minimiser la zone 3 et de maximiser la zone 1. De cette manière, nous visons à maximiser la part des scénarios opérationnels connus et non dangereux en minimisant la part des scénarios inconnus et dangereux.

Pour atteindre cet objectif, la SOTIF prévoit l'analyse des fonctionnalités attendues dans le cadre des scénarios connus (domaines 1 et 2) afin d'évaluer l'acceptation des risques et de proposer des modifications de conception si nécessaire. De cette manière, les scénarios dangereux connus peuvent être contrôlés pour devenir non dangereux (passage de la zone 2 à la zone 1).

Dans le cas de scénarios relevant des catégories inconnues (domaines 3 et 4), une stratégie de vérification et de validation adéquate doit être élaborée afin de les réduire. Cela peut notamment être réalisé en explorant un large éventail de scénarios à l'aide de simulations et de tests.

D. Synthèse du process de la norme SOTIF

La norme ISO 21448 aborde le concept de la SOTIF en mettant en place un un processus systématique destiné à réaliser l'analyse et l'évaluation des risques et d'assurer la sécurité de la fonctionnalité attendue pour les systèmes de conduite automatisés. Ce processus repose sur des recommandations, organisées sous la forme de multiples clauses, traitant chacune d'un aspect spécifique des activités liées à la SOTIF.

Concrètement, les premières clauses (à savoir **1, 2, 3 et 4**) représentent la partie introductive de la norme ISO 21448 et couvrent respectivement le champ d'application, les références normatives, les termes et définitions, ainsi qu'une vue d'ensemble de l'organisation des activités SOTIF.

Ensuite, la **clause~5** marque le démarrage effectif du processus SOTIF en traitant la spécification des fonctionnalités et de la conception du système.

Les **clauses~6** et **7** représentent la partie des activités d'évaluation des risques qui peuvent être réalisées à travers une analyse itérative. Le premier objectif de la clause 6 est l'identification et l'évaluation des dangers. En fonction des résultats de cette évaluation, et dans le cas où un danger est jugé non maîtrisable de manière générale ou susceptible d'entraîner des dommages, le processus SOTIF préconise la spécification de critères d'acceptation pour le risque résiduel, dans le cadre de la clause~6. Dans le même ordre d'idées, la clause~7 traite de l'identification et de l'évaluation des insuffisances fonctionnelles potentielles et des conditions déclencheuses potentielles de ces insuffisances.

D'autre part, la partie du processus comprenant les **clauses~9, 10 et 11** représente les activités de la SOTIF liées à l'évaluation au moyen de la vérification et de la validation (V&V). Plus précisément, ces activités traitent de la définition de la stratégie de vérification et de validation ainsi que de l'évaluation des scénarios connus et inconnus.

Par la suite, la **clause~12** marque la fin de la phase de développement en examinant et en évaluant la bonne exécution du processus SOTIF, et l'absence de risque non raisonnable ouvre la voie aux activités de la phase opérationnelle couvertes par la **clause~13**.

Enfin, il est à souligner que le processus SOTIF proposé dans la norme ISO~21448 est un processus itératif. De ce fait, la **clause~8**, qui est menée parallèlement aux activités mentionnées précédemment, aborde les modifications fonctionnelles qui se rapportent aux risques liés à la SOTIF. Ces modifications peuvent se révéler nécessaires selon les résultats des clauses 7 à 12 et dans le cas où une des affirmations suivantes est avérée. À savoir : (i) La réponse attendue du système aux conditions de déclenchement n'est pas acceptée. (ii) Le risque résultant des scénarios connus n'est pas suffisamment faible. (iii) La probabilité de rencontrer un scénario inconnu entraînant un comportement dangereux n'est pas suffisamment faible. (iv) Présence d'un risque déraisonnable.

IV. Integration de la SOTIF dans le cadre de la securite ferroviaire

A. Étude comparative de la terminologie.

Dans la perspective d'intégrer les activités de la SOTIF dans le contexte de la sécurité ferroviaire, nous choisissons de conduire une étude comparative entre les concepts fondamentaux de la sécurité selon les références automobiles et ferroviaires.

D'une manière générale, la gestion de la sécurité des systèmes techniques est principalement étudiée sous l'angle de la sécurité fonctionnelle. La norme IEC 61508 définit la *sécurité fonctionnelle* comme « *une partie de la sécurité globale du système qui dépend du fonctionnement correct des systèmes EE relatifs à la sécurité et d'autres mesures de réduction des risques* ». Selon cette définition, la sécurité fonctionnelle ne prend en compte que les défaillances causées par des défauts dans les composants E/E.

Dans le même esprit, et conformément à la norme ISO 26262, toute défaillance de la fonctionnalité attendue d'un système automobile due à des environnements opérationnels complexes ou à des lacunes dans les spécifications des exigences ne relève pas du champ d'application de la norme ISO 26262 (sécurité fonctionnelle) et doit être couverte par la norme ISO 21448.

La norme ISO 21448 définit la sécurité de la fonctionnalité attendue comme l'absence de risque déraisonnable dû à des dangers résultant d'insuffisances fonctionnelles de la fonctionnalité attendue et de sa mise en œuvre. Il est important de noter que le terme « fonctionnalité prévue » fait référence à la fonctionnalité attendue. Par ailleurs, une insuffisance fonctionnelle peut découler de problèmes de spécification ou de performance. Alors qu'une insuffisance de spécification est due à un manque d'exhaustivité, les insuffisances de performance résultent de limitations techniques vis-à-vis des conditions opérationnelles et environnementales.

En revanche, dans le domaine ferroviaire, la sécurité est définie comme l'absence de risque inacceptable. Cette définition suggère que l'évaluation de la sécurité s'articule autour du concept de risque. La méthode de sécurité commune relative à l'évaluation des risques (CSM-RA) définit le risque comme « *la fréquence d'occurrence d'accidents et d'incidents causant un dommage (dû à un danger) et le degré de gravité de ce dommage* ». De même, la norme EN 50126 définit le risque comme « *la combinaison de la fréquence attendue d'une perte et du degré de gravité attendu de cette perte* ». Bien que les deux définitions tiennent compte d'une combinaison de fréquence et de gravité, les termes « *perte* » et « *dommage* » ne correspondent pas directement. Dans le présent article, nous ne considérerons ces termes que sous l'angle de la sécurité.

Par ailleurs, nous notons que selon les normes EN 5012x et IEC 61508, la cause d'une défaillance due à l'environnement opérationnel relève d'un défaut systématique dans la conception du système.

En complément, les liens entre divers autres termes liés à l'évaluation des risques établis par la norme EN 50126 sont illustrés dans la Figure 3.

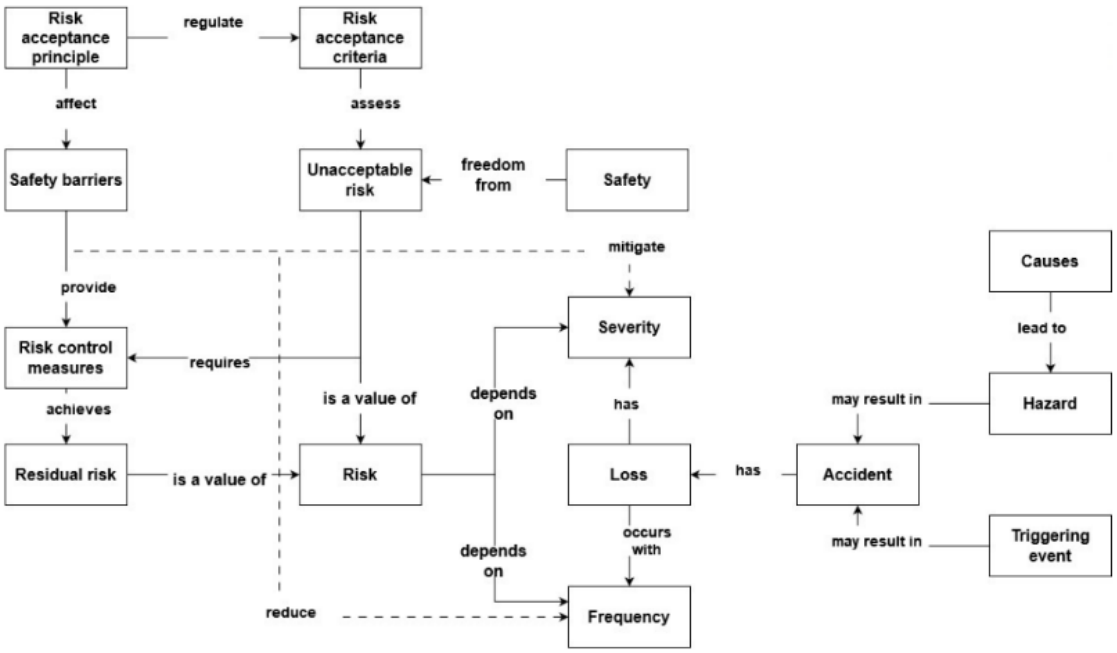

Fig. 3. Liens entre les terminologies relatives aux risques (selon EN 50126)

B. Liens entre la SOTIF et les de la norme EN 50126-1

Les activités du processus SOTIF (ISO 21448) sont intentionnellement destinées à compléter les activités conventionnelles de la sécurité fonctionnelle (déjà prévues dans le cycle de vie de la sécurité des systèmes). Dans le cas des systèmes automobiles, les activités de sécurité fonctionnelle sont définies dans la norme ISO 26262 et comprennent l'analyse des dangers et l'appréciation des risques (HARA - *Hazard Analysis and Risk Assessment*), les concepts de sécurité fonctionnels et techniques, ainsi que les tests de vérification et de validation. Dans ce contexte, la norme ISO 21448 intègre un modèle en forme de V pour souligner les liens entre les activités de la norme ISO 26262 et le processus SOTIF.

De manière comparable, il est tout à fait envisageable que des liens entre la norme ISO 21448 et les normes EN 5012x soient établis. C'est dans cette optique que nous proposons dans la Figure 4 une illustration de notre contribution qui vise la manière à travers de laquelle les activités SOTIF peuvent être intégrées aux processus de sécurité ferroviaires.

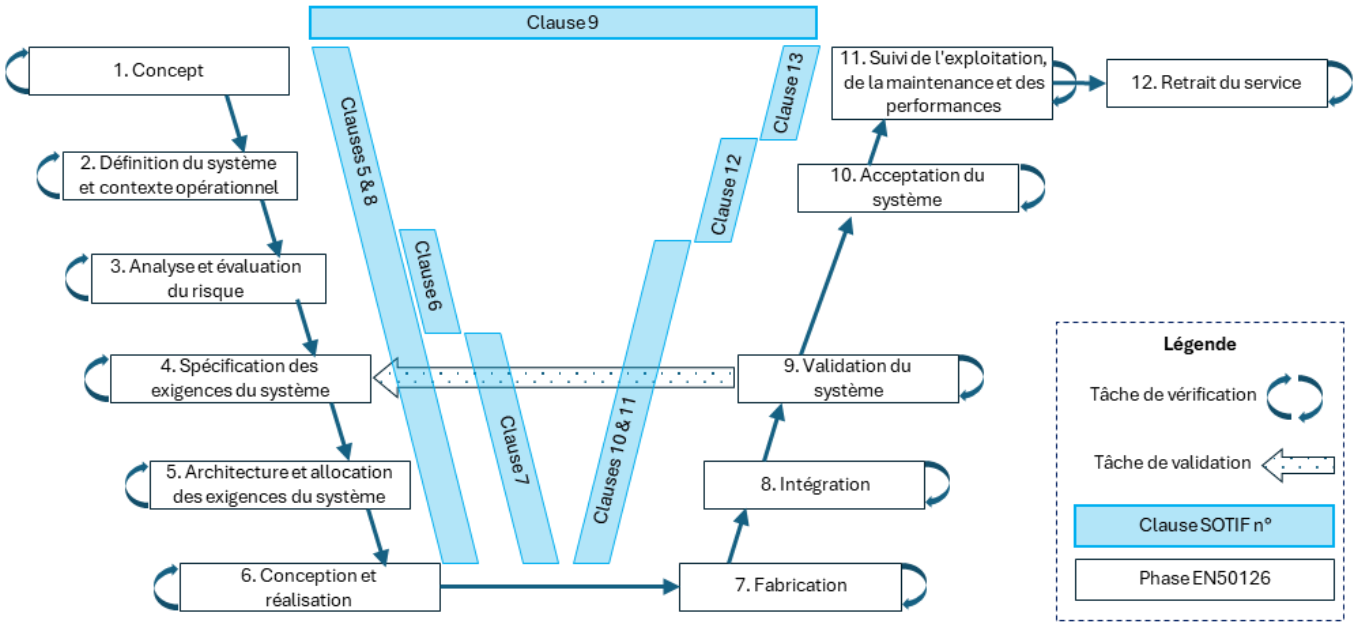

Fig. 4. Liens entre les activités SOTIF et les activités du cycle de vie du système ferroviaire selon la norme EN 50126

Une description plus détaillée de cette proposition d'intégration des aspects SOTIF dans le processus d'analyse et de démonstration de sécurité ferroviaire est présentée au travers du cas d'étude applicatif abordé dans la suite de cet article.

C. Présentation du contexte du cas d'étude

Le cas applicatif étudié dans le cadre de ce papier s'insère dans le contexte des sous-systèmes de perception à base d'IA pour l'autonomie dans le domaine ferroviaire "train autonome".

En effet, dans la perspective d'opérations ferroviaires relevants du niveau d'automatisation GoA-4 (Grade Of Automation), le train doit accomplir les tâches de conduite de manière complètement autonome, sans nécessiter d'intervention humaine. Ainsi, le système de train autonome devra être responsable à lui seul de l'exécution de fonctions telles que le *respect de la signalisation latérale, la détection d'obstacles, et le monitoring de l'environnement*.

En particulier, le périmètre des fonctions de détection des obstacles et de monitoring de l'environnement est définie en fonction de la position géographique de la zone à surveiller par rapport à la ligne ferroviaire (voir la Figure 5). Ainsi, les objets ou êtres vivants situés dans le gabarit du train sont considérés comme étant des obstacles relevant de la fonction de détection d'obstacle. D'autre part, les éléments situés en dehors du gabarit du train sont traités par la fonction de monitoring de l'environnement.

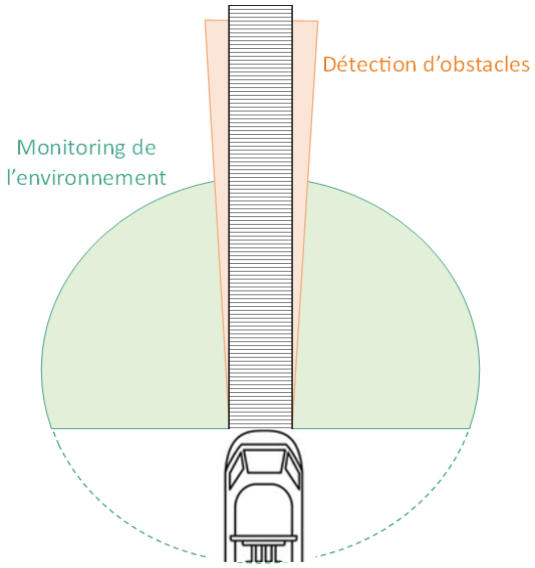

Fig. 5. Périmètre de la fonction de monitoring de l'environnement

D. Démarche de travail intégrant les aspects SOTIF

Dans le cadre de ce cas applicatif, la fonction de perception pour le monitoring de l'environnement est considérée comme une fonction sécuritaire, dont la défaillance peut entraîner l'occurrence d'un évènement indésirable ou l'atteinte d'une situation dangereuse pour le train autonome. Ainsi, afin d'assurer la sécurité de la fonction monitoring de l'environnement, un ensemble d'activités de sécurité doivent être conduites tout au long du cycle de développement du système implémentant la fonction. La structure de notre démarche de travail ainsi qu'un aperçu des interactions entre les diverses activités sont illustrés dans la Figure 6.

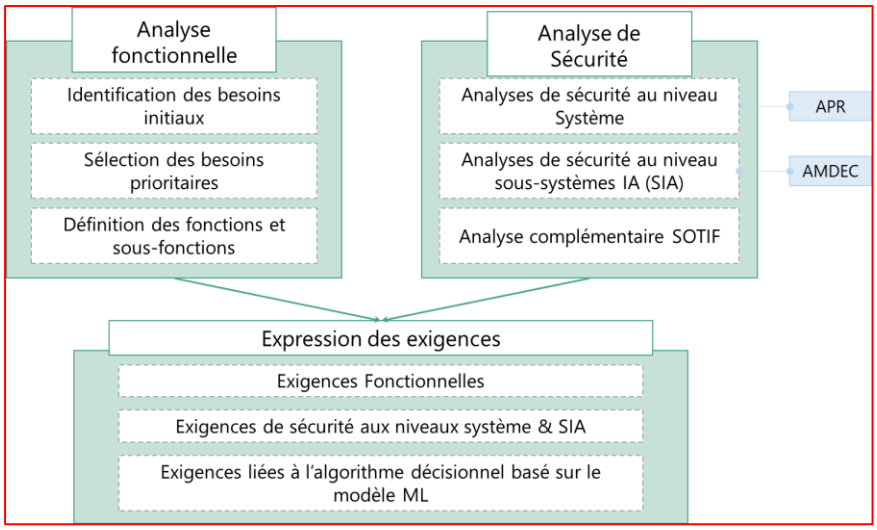

Fig. 6. Démarche de travail intégrant les aspects SOTIF

Concrètement, la démarche adoptée s'articule autour de trois parties principales.

L'analyse fonctionnelle représente le premier volet de la démarche. L'objectif est d'analyser le besoin initial (exprimé par les partenaires industriels du projet) afin de définir clairement les frontières du sous-système de monitoring de l'environnement et d'identifier l'objectif final escompté par la fonction de monitoring de l'environnement. Le besoin initial est ensuite décomposé en un ensemble de besoins principaux qui permettent de mieux préciser le périmètre du besoin initial en énumérant les différents éléments dont la surveillance est pertinente pour la fonction de monitoring de l'environnement. Dans la continuité, l'ensemble des sous-fonctions qui sont nécessaires à la réalisation des fonctions principales sont identifiées. Finalement, l'architecture fonctionnelle du système de monitoring de l'environnement est établie en définissant les interactions, liens et dépendances entre les sous-fonctions précédentes.

Le deuxième volet de la démarche consiste en une **analyse de sécurité** qui est conduite parallèlement aux activités de l'analyse fonctionnelle. Dans un premier temps, cette étude suit le cycle d'activité de sécurité classique, tel que défini par la norme EN 50126. Plus précisément, une analyse préliminaire des risques (APR) est réalisée au niveau du système pour identifier un ensemble de risques qui doivent être couvert tout au long du cycle de vie du système. En outre, des analyses de sécurité sont effectuées au niveau du sous-système à l'aide de méthodes et techniques, telles que l'AMDEC et/ou l'ADD (Arbre de Défaillance). Parallèllement, une analyse de sécurité complémentaire est conduite dans l'optique de traiter de manière particulière la problématique de la sécurité de la fonctionnalité attendue (SOTIF).

À la lumière des résultats obtenus suite aux analyses précédentes, la troisième étape de la démarche s'intéresse à la spécification et à l'allocation des **exigences** relatives aux différents éléments et niveaux du système étudié. Plus précisément, les résultats de l'analyse fonctionnelle forment une liste d'exigences fonctionnelles. Les résultats issus de l'analyse préliminaire des risques permettent d'identifier un certain nombre d'exigences de sécurité au niveau système. Ensuite, des exigences de sécurité plus détaillées peuvent être spécifiées et associées aux variantes fonctions et aux éléments intervenant dans le monitoring de l'environnement à la suite des analyses au niveau sous-système. Finalement, la considération des aspects relatifs à la SOTIF dans le cadre des analyses permet de compléter cette liste des exigences.

V. Discussion sur les resultats du cas d'etude integrant la SOTIF.

Dans cette section, nous présentons notre retour d'expérience à la suite de la considération des aspects couverts par la SOTIF dans le cadre de ce cas applicatif. Ainsi, cette discussion portera sur les avantages et liens identifiés au niveau de diverses étapes de la démarche.

Tout d'abord, en se rapportant à la représentation en cycle en V des activités de cycle de vie du système ferroviaire selon la norme EN 50126, nous notons que les activités décrites dans notre démarche de travail pour ce cas d'étude se rapportent aux activités de la partie descendante du cycle en V (phases 1 à 6). Dans ce contexte, les orientations et les recommandations de la (clause 5) de la SOTIF s'intègrent dans le cadre de l'analyse fonctionnelle de notre démarche et permettent de traiter la spécification des fonctionnalités et de la conception du système.

D’autre part, les orientations des clauses 6 et 7 de la SOTIF concernent les activités du deuxième volet de notre démarche (i.e., analyse de sécurité). En particulier, les indications de la SOTIF relatives à la catégorisation des scénarios opérationnels permettent de mieux distinguer et de classifier l’ensemble des scénarios identifiés au niveau système (notamment au travers de l’APR). En outre, les instructions de la SOTIF nous ont permis d’identifier de manière plus complète les dangers liés à l’interaction du sous-système de perception pour le monitoring de l’environnement avec son environnement opérationnel. En particulier, la clause 7 de la SOTIF (traitant de l'identification et de l'évaluation des insuffisances fonctionnelles potentielles et des conditions de déclenchement potentielles de ces insuffisances) nous a permis d’identifier de manière explicite certaines causes de défaillances (i.e., liées aux insuffisances de spécifications et de performances) non couvertes lors de l’analyse AMDEC classique.

Concrètement, une des sous-fonctions identifiées pour le monitoring de l’environnement stipule que le système doit « détecter la présence d’une personne au sein de l’emprise ferroviaire ». Pour l’étude de cette fonction, l’analyse AMDEC conventionnelle permet de couvrir des cas tels que la “Perte de la fonction à cause d’une défaillance HW de la caméra utilisée pour la perception”. De manière complémentaire, l’analyse intégrant la SOTIF pour cette sous-fonction permet de mettre en avant d’autres causes de défaillances potentielles. Notamment :

L’analyse des causes de défaillance liées aux “Insuffisance de spécifications” se traduit, de manière non-exhaustive, par :

- Définition manquante ou incomplète de ce qu’est ‘un piéton’
- Définition manquante ou incomplète de ce qu’est ‘une emprise ferroviaire’

De plus, l’identification des causes relatives aux “Insuffisance fonctionnelle liée aux performances” permet de couvrir les aspects incluant :

- Limites de performances intrinsèques des équipements utilisés (e.g., portée du capteur)
- Altération des performances nominales des équipements utilisés par des perturbations externes liées à l’environnement opérationnel (e.g., conditions météorologiques ; niveau de luminosité)
- Insuffisance fonctionnelle liée à la performance de l’algorithme décisionnel (e.g., sensibilité vis-à-vis des scénarios inconnus)

Concernant le volet de spécification d’exigences, cet aspect regroupe les résultats des activités conduites précédemment (incluant les clauses 5,6,7, et 8 de la SOTIF). À ce stade, nous rappelons que selon la norme EN 5012x, la cause d'une défaillance due à l'environnement opérationnel est considérée comme un défaut systématique dans la conception du système. À ce titre, les causes SOTIF peuvent être considérées comme étant des défaillances systématiques. Ainsi, les mesures de couvertures liées à ces causes portent principalement sur les exigences relatives à la qualité du process de développement qui doit être respecté de manière rigoureuse.

En particulier, les performances des modèles ML dépendent largement de la qualité des données utilisées pour leur apprentissage, mais également de leur quantité, de la couverture des cas limites, etc. afin d’éviter le sur-apprentissage et le sous-apprentissage. En effet, bien que les modèles ML (e.g., réseaux de neurones NN- Neural Networks) soient des solutions puissantes pour effectuer des tâches complexes par rapport aux humains, ils présentent néanmoins une forte sensibilité au bruit et aux petites erreurs dans les données d'apprentissage et de test. Par conséquent, une attention particulière doit être portée sur les jeux de données d'apprentissage et de validation lors de la spécification d'exigences.

VI. Conclusion

Dans le contexte actuel, l’émergence de technologies de rupture à base d’IA impactera vraisemblablement la manière de penser, concevoir, et implémenter un système complexe. Néanmoins, dans le cas des systèmes critiques (e.g., systèmes de transport) le paradigme d’analyse et de démonstration de sécurité restera considérablement inchangé même pour ces nouveaux systèmes. Concrètement, ce paradigme repose sur trois piliers fondamentaux : (i) l’analyse et l’évaluation des risques. (ii) La spécification et mise en œuvre de mesure de contrôle de ces risques. (iii) La démonstration que les risques résiduels sont maîtrisés. Ainsi, les activités relatives à chaque aspect devront être assurées préalablement à toute autorisation et mise en service d’un système de transport autonome intégrant l’IA.

Cependant, le recours aux modèles IA pour assurer certaines fonctions relatives à la sécurité engendre l’apparition de nouveaux dangers. La nature incertaine de ces dangers impose la révision et l’adaptation du cadre réglementaire et normatif actuel portant sur la sécurité des systèmes. C’est dans ce contexte que les normes UL4600 et ISO 21448 ont vu le jour dans le domaine automobile. Par ailleurs, le cadre réglementaire européen évolue également, notamment avec l’AI Act, qui définit les « High Risk AI » et les exigences associées. Ainsi, le cadre normatif au niveau de l'Europe devra répondre à ces nouvelles exigences.

Dans ce papier, nous examinons le rôle potentiel de la SOTIF (ISO 21448) dans l’analyse et la démonstration de sécurité des systèmes ferroviaire autonomes. Cette intégration de la SOTIF dans le domaine ferroviaire est explorée au travers d’un cas applicatif relatif à la “ fonction de perception pour le monitoring de l’environnement à base d’IA dans le cadre d’un train autonome”. En particulier, ce cas d’étude nous a permis d’évaluer l’apport de la considération de la SOTIF pour l’étude et la couverture des risques liés à l’interaction du système avec l’environnement externe, les défaillances liées aux insuffisances en termes de spécifications, ainsi que les limitations en termes de performances des systèmes E/E (i.e., performances des capteurs et des algorithmes d’IA).

D'autre part, nous notons que les travaux abordés dans ce papier concernent la partie descendante du cycle en V (selon la norme EN 50126-1). Dans ce cadre, l’intégration de la SOTIF a été considérée pour compléter l’identification des dangers ainsi que la spécification des exigences (mesures de sécurité à mettre en œuvre).

En revanche, les activités de la partie montante et stratégie de V&V ne sont pas abordées dans cette contribution. Or, ces aspects revêtent une importance primordiale dans la démonstration de sécurité des systèmes critiques. Ainsi, une suite logique de cette contribution devra porter sur l'étude de l'intégration de la SOTIF dans la stratégie de V&V.

En particulier, nous soulignons que la norme ISO 21448 aborde la notion d'ODD (Domaine de conception opérationnelle - *Operational Design Domain*). Un domaine de conception opérationnelle (ODD) est un instrument utilisé pour définir les conditions spécifiques, y compris l'environnement opérationnel dans lequel un système est conçu pour fonctionner. L'ODD d'un système permet à ses développeurs d'identifier, de spécifier et de réaliser une évaluation des risques liés aux scénarios opérationnels auxquels le système pourrait être soumis pendant ses opérations. De cette manière, l'ODD aide les développeurs du système à identifier les fonctions que le système basé sur l'apprentissage devra exécuter dans différentes situations opérationnelles.

Dans le secteur ferroviaire, la présence du concept comparable de « conditions opérationnelles et environnementales » est observée. Néanmoins, ce dernier n'a pas le niveau de structuration que l'on retrouve dans le domaine automobile (ISO 34501, ISO 34502, ISO 34503). Il est donc utile d'explorer la définition et la structuration d'un ODD spécifique au secteur ferroviaire (Tonk, 2021). Plus précisément, l'ODD pourrait jouer un rôle central dans le cadre d'une stratégie de validation basée sur des scénarios. Concrètement, un aspect clé de la justification de la sécurité se concentrera sur la démonstration d'un degré suffisant d'inclusion des éléments décrits dans l'ODD tout au long des activités de validation par le biais de tests et/ou de simulations. Par conséquent, la justification d'une couverture étendue des scénarios opérationnels définis dans l'ODD constituera un argument probant en faveur de la démonstration de sécurité des systems à base d'IA.

REMERCIEMENTS

Ce travail a été réalisé dans le cadre du projet ASTA (Assurance Sécurité du Train Autonome), co-financé par la SNCF. Nous tenons à remercier les partenaires du projet pour leur soutien et leur collaboration fructueuse, qui ont permis de mener à bien ce travail.

RÉFÉRENCES

Becker, C., Brewer, J. C., & Yount, L. (2020). Safety of the intended functionality of lane-centering and lane-changing maneuvers of a generic level 3 highway chauffeur system (No. DOT HS 812 879). United States. National Highway Traffic Safety Administration. Electronic System Safety Research Division.

IEC 61508 :2010 - Sécurité fonctionnelle des systèmes électriques/électroniques/électroniques programmables relatifs à la sécurité (2010)

CSM-RA - Règlement d'exécution (UE) n°402/2013 concernant la méthode de sécurité commune relative à l'évaluation et à l'appréciation des risques. *Agence de l'Union Européenne pour le ferroviaire* (2013)

EN 50126-1:2017 - Applications ferroviaires - Spécification et démonstration de la fiabilité, de la disponibilité, de la maintenabilité et de la sécurité (FDMS) - Partie 1 : processus FMDS générique.

EN 50126-2:2017 - Applications ferroviaires - Spécification et démonstration de la fiabilité, de la disponibilité, de la maintenabilité et de la sécurité (FDMS) - Partie 2 : approche systématique pour la sécurité.

EN 50128:2011 - Applications ferroviaires - Systèmes de signalisation, de télécommunication et de traitement - Logiciels pour systèmes de commande et de protection ferroviaire.

EN 50129 :2018 - Applications ferroviaires - Systèmes de signalisation, de télécommunications et de traitement - Systèmes électroniques de sécurité pour la signalisation.

Filip, A. (2022). Synergies between road and rail transport in the development of safe self-driving vehicles. International journal of transport development and integration, 6(3), 313-325.

Himrane, O., Beugin, J., & Ghazel, M. (2021). Toward formal safety and performance evaluation of GNSS-based railway localisation function. IFAC-PapersOnLine, 54(2), 159-166.

Himrane, O., Beugin, J., & Ghazel, M. (2023). Implementation of a model-oriented approach for supporting safe integration of GNSS-based virtual balises in ERTMS/ETCS level 3. IEEE Open Journal of Intelligent Transportation Systems.

ISO 21448:2022 - Véhicules routiers – Sécurité de la fonction attendue (2022)

ISO 26262:2018 - Véhicules routiers – Sécurité fonctionnelle (2018)

ISO 34501:2022 - Véhicules routiers - Scénarios d'essai pour les systèmes de conduite automatisée - Vocabulaire (2022)

ISO 34502:2022- Véhicules routiers - Scénarios d'essai pour les systèmes de conduite automatisée - Cadre d'évaluation de la sécurité basé sur des scenarios (2022)

ISO 34503:2023 - Véhicules routiers - Scénarios d'essai pour les systèmes de conduite automatisée - Spécification du domaine de conception opérationnelle (2023)

Madala, K., Avalos-Gonzalez, C., & Krithivasan, G. (2021). Workflow between ISO 26262 and ISO 21448 standards for autonomous vehicles. Journal of System Safety, 57(1), 34-42.

Stellet, J. E., Brade, T., Poddey, A., Jesenski, S., & Branz, W. (2019, June). Formalisation and algorithmic approach to the automated driving validation problem. In 2019 IEEE Intelligent Vehicles Symposium (IV) (pp. 45-51). IEEE.

Tonk, A., Boussif, A., Beugin, J., & Collart-Dutilleul, S. (2021, September). Towards a specified operational design domain for a safe remote driving of trains. In Proceedings of the 31st European Safety and Reliability Conference, Angers, France (pp. 19-23).

UL4600 - Standard for Safety for the Evaluation of Autonomous Products (2020)