Linköping University | Department of Electrical Engineering
Master's thesis, 30 ECTS | Elektroteknik
2024 | LIU-ISY/LITH-EX-A--24/5632--SE

Early-Stage Validation of Autonomous Vehicles in Ambiguous Environments

– A Systems-Theoretic Process Analysis (STPA) of an Autonomous Military Defense Industry Vehicle

---

Validering av autonoma fordon i oklara miljöer under tidiga utvecklingsstadier
– En säkerhetsanalys med analysmetoden STPA genomförd på ett autonomt militärt fordon inom försvarsindustrin

Maria Axelsson

Supervisor : Daniel Jung
Examiner : Erik Frisk

LINKÖPING UNIVERSITY

Linköpings universitet
SE–581 83 Linköping
+46 13 28 10 00 , www.liu.se

Abstract

This report delves into the early developmental phase of an autonomous vehicle designed for defense applications. Navigating diverse terrains, this unmanned ground vehicle (UGV) poses unique challenges, particularly in the absence of clearly defined directives found in typical traffic scenarios. The analysis employs the Systems-Theoretic Process Analysis (STPA) to identify and anticipate risks inherent in the conceptual stage of product development.
Beyond the specific UGV case, the report explores the broader landscape of validating autonomous systems. It discusses prevalent methods, emphasizing adaptability to different contexts and stages of development. By shedding light on the risks and challenges of autonomy in vehicles and examining effective validation strategies, this report aims to contribute valuable insights to the ongoing discourse surrounding autonomous vehicle development.

Acknowledgments

I would like to express my gratitude to AFRY and Milrem Robotics for making this project possible. A special thanks to Johan Persson and Jaanus Urb from Milrem Robotics for always being available to answer questions and for enabling a trip to Tallinn to witness the examined vehicle. I also want to thank to my supervisor at AFRY, Jens Gunnarsson, who has been a constant presence throughout the project, providing valuable feedback and assisting me in navigating every step of the analysis. I would also like to acknowledge Peter Sidenbladh from AFRY, whose support started the entire project and made it possible for me to be at AFRY and do this project.

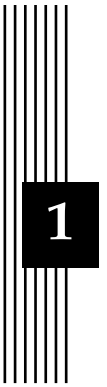
Thanks also go to my supervisor at Linköping University, Daniel Jung, who, through regular meetings, has facilitated the progress of this project, offering continuous support, feedback, and insightful tips. Additionally, my gratitude extends to my examiner at Linköping University, Erik Frisk, for good feedback and useful ideas, along with being supportive and available throughout the project.

Acronyms

1. **AFRY:** AFRY is a global engineering and design company that provides expert services in infrastructure, industry, and energy sectors.
2. **Milrem Robotics:** Milrem Robotics is an Estonian robotic vehicle manufacturer specializing in developing autonomous vehicles within the defense industry, including products such as THeMIS and the upcoming Next Generation Vehicle (NGV).
3. **THeMIS:** THeMIS represents a partially autonomous military vehicle developed by Milrem Robotics. It stands as one of their significant products within the defense industry, showcasing features of autonomous functionalities combined with military applications.
4. **STPA:** STPA stands for Systems-Theoretic Process Analysis and is a method utilized for hazard analysis and safety assurance in complex systems. It emphasizes a systematic approach to identify potential hazards, examining the interactions between system components and their control structures. STPA is recognized for its efficacy in early-stage development, extracting substantial insights from limited data to ensure the safety and reliability of complex systems, particularly in fields like autonomous vehicle technology.
5. **NGV:** NGV refers to the Next Generation Vehicle developed by Milrem Robotics, signifying an upcoming prototype in the realm of autonomous vehicles. It represents an advanced stage in the evolution of their autonomous vehicle technology and is anticipated to build upon the functionalities of their previous product, THeMIS. It is important to note that NGV is a conceptual vehicle and does not currently exist in physical form, marking an anticipated advancement in their autonomous vehicle lineup.
6. **Six-Step Model:** A structured approach integrating safety and security measures across functions, structure, failures, attacks, safety countermeasures, and security countermeasures in the life cycle of autonomous vehicles.
7. **SAE:** Society of Automotive Engineers.
8. **ISO 26262:** ISO 26262 is an international standard for functional safety in road vehicles. It outlines requirements for the entire safety lifecycle of automotive systems, including development, production, operation, service, and decommissioning.
9. **SAE J3016:** SAE J3016 is a standard published by the Society of Automotive Engineers (SAE) that defines levels of driving automation ranging from Level 0 (no automation) to Level 5 (full automation). It provides a common language for discussing and categorizing the capabilities of automated vehicles.
10. **SAE J3061:** SAE J3061 is an SAE standard focused on the cybersecurity aspects of vehicle systems. It provides guidance for the entire vehicle cybersecurity lifecycle, including design, implementation, operation, and decommissioning.
11. **UGV:** Unmanned Ground Vehicle.
12. **MIFIK:** Milrem's Intelligent Functions Kit.
13. **SC:** System Constraint.
14. **UCA:** Unsafe Control Action.
15. **UI:** User Interface.
16. **HMI:** Human-Machine Interaction.
17. **ECU:** Engine Control Unit.
18. **AI:** Artificial Intelligence.
19. **AV:** Autonomous Vehicle.
20. **CPS:** Cyber-Physical Systems.

21. **PHA:** Process Hazard Analysis.
22. **TARA:** Threat Analysis and Risk Assessment.
23. **FMEA:** Failure Modes and Effects Analysis.
24. **FTA:** Fault Tree Analysis.

Contents



1 Introduction

1.1 Background

This Master's Thesis will be within the field of electrical engineering, marking the conclusion of a 5-year engineering program in applied physics and electrical engineering at Linköping University.

The research will be conducted in collaboration with AFRY, a company that has previously engaged in employing this analysis methodology across various clients and different vehicle platforms. With guidance from a supervisor at AFRY, the analysis will constitute a central component of this Master's Thesis.

Milrem Robotics, an Estonian robotic vehicle manufacturer specializing in developing autonomous vehicles within the defense industry, and a client of AFRY, will undergo a safety analysis within this specific domain for one of their products. Notably, their products are designed for the defense industry, and despite the existence of military standards, it exists an interest in conducting this type of analysis for their vehicles. Access to technical drawings and expert assistance will be provided by Milrem Robotics. Moreover, the university will also appoint a supervisor who will offer guidance, ensuring the alignment and approval of the thesis project. An examiner from the university will be present to pass all mandatory components of the project.

1.2 Motivation

This Master's Thesis aims to explore critical insights at the outset of self-driving vehicle development. The primary focus lies in uncovering valuable information available in the initial phases of creating these vehicles. Additionally, it seeks to highlight potential risks associated with self-driving vehicles, especially in areas lacking clear regulatory frameworks.

Moreover, this research investigates means of ensuring the reliability and safety of self-driving systems. It explores robust validation techniques to discern the most informative and accurate methods. These inquiries form the core motivation behind this research endeavor.

1.3 Functionality of THeMIS and NGV

THeMIS, a partially autonomous military vehicle developed by Milrem Robotics, (an Estonian robotic vehicle manufacturer specializing in developing autonomous vehicles within the defense industry,) is an Unmanned Ground Vehicle (UGV) designed to support various missions with the aim of reducing the number of troops on the battlefield Figure 1.1. It is equipped with Milrem's Intelligent Functions Kit (MIFIK), which includes wired and wireless follow-me functionality, waypoint navigation, and obstacle detection.

1.3.1 THeMIS

THeMIS is a robust and versatile UGV that can be adapted for various missions. THeMIS can be remotely controlled by an operator or programmed to follow predefined waypoints.

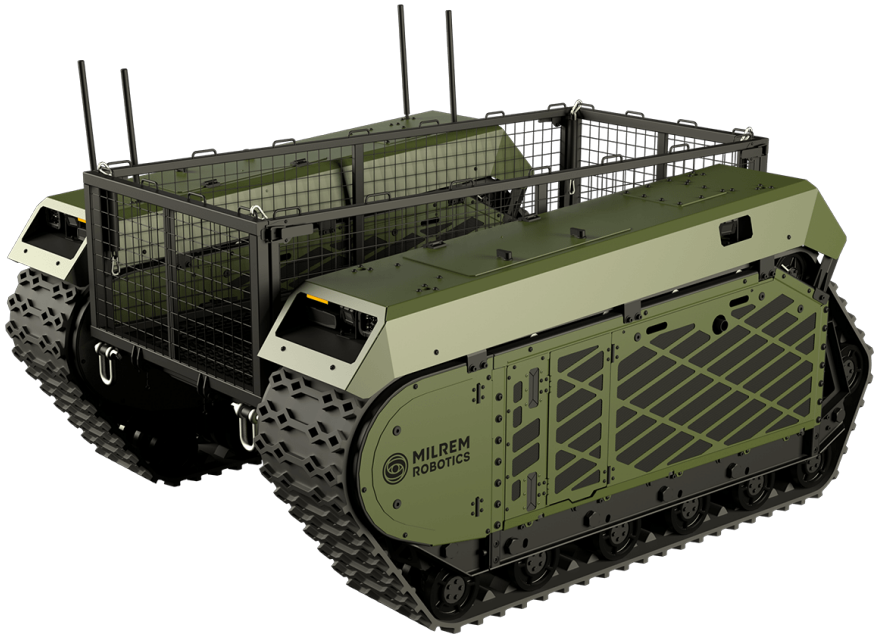


Figure 1.1: THeMIS [1]. Used with permission from Milrem Robotics.

1.3.2 NGV

Next Generation Vehicle (NGV) represents the future iteration of THeMIS, currently in the conceptual phase with no physical implementation yet. It is envisioned to refine and enhance the autonomous functionalities already present in THeMIS. The three primary autonomous domains available in THeMIS are set to undergo substantial improvements and will be fully realized in NGV:

- **Obstacle Avoidance**: NGV integrates sensors, such as cameras and Lidar, to detect and evade obstacles in its operational environment.
- **Waypoint Navigation**: NGV's programming allows it to autonomously traverse the terrain by following a sequence of predefined waypoints. This capability is crucial for scenarios involving the transportation of goods between specific locations or continuous movement between designated areas.

- **Follow-me Mode**: In this operational mode, NGV can autonomously trail an operator or another vehicle. This feature is valuable in instances requiring human supervision or when the UGV needs to track a manned vehicle.

It is important to note that while these autonomous functionalities are present in THeMIS, they are anticipated to undergo significant enhancements and be fully operational and optimized in NGV. The continuous development aims to achieve a more efficient and advanced level of autonomy in NGV compared to its predecessor, THeMIS.

1.4 The system's work area and work environment

The system is designed to operate in complex and unregulated terrains, particularly in support of defense and military missions. It is designed to function in various types of environments, such as rugged terrains, dirt roads, and forests.

1.5 Aim

This Master's Thesis aims to extract valuable insights available during the early phases of self-driving vehicle development. Its objective is to uncover and analyze information obtainable in the initial stages of crafting autonomous products. Furthermore, it aims to identify and evaluate the general risks inherent in self-driving vehicles, especially in contexts lacking explicit guidelines.

Additionally, a significant goal of this research is to investigate effective methods for validating self-driving systems. It intends to examine various validation techniques, aiming to identify the most practical approaches for ensuring the reliability and safety of autonomous vehicles.

By focusing on these key areas, this research aims to emphasize the importance of early-stage information in autonomous vehicle development. Moreover, it aims to outline effective strategies for mitigating risks and implementing robust validation methods to ensure the dependability and safety of self-driving systems, specifically leveraging the potential of an Systems-Theoretic Process Analysis (STPA). The STPA method stands as an effective tool in early-stage development, allowing the extraction of substantial insights from limited data. This study will apply the STPA method to investigate and derive critical insights from the upcoming NGV developed by Milrem Robotics. The NGV represents a crucial stage in the production of a self-driving vehicle, making it conducive to exploring the research questions outlined in this study.

1.6 Research questions

This study aims to address two pivotal research questions:

1. What risks can be concretely identified in the early stages of developing a self-driving vehicle?

This question delves into the identification and concretization of potential risks inherent in the initial phases of self-driving vehicle development. It seeks to understand the specific challenges and vulnerabilities that emerge early in the development process.

2. How can we ensure the validation of an autonomous control system? This includes considering the use of the STPA method and other applicable approaches.

This question focuses on establishing a robust validation process for an autonomous control system. The NGV marks an early stage in the production of a self-driving vehicle, making it conducive to investigating these research questions. The study will leverage the STPA method, renowned for its effectiveness in early-stage development. This method allows extracting substantial information from limited data, making it particularly suitable for this investigation.

The STPA method is chosen for its power in extracting valuable insights during the initial stages of development. Milrem Robotics continuously develops and prototypes new models. One of their significant products is THeMIS, a partially autonomous military vehicle. Their next anticipated prototype, NGV, will be investigated in this study as it represents an early stage in the production of a self-driving vehicle. The study aims to utilize the STPA method due to its capacity to derive significant insights from minimal information, which is vital in the early stages of development.

1.7 Delimitations

This study delineates specific boundaries and limitations to ensure a focused analysis within a defined scope. It's crucial to note the following constraints guiding the scope of this research:

- **Ethical Priorities and Decision-making:** The exploration of ethical dilemmas concerning prioritization of lives and related ethical decisions in autonomous driving scenarios is immensely intricate and multifaceted. Therefore, this study refrains from delving into these ethical dimensions, focusing primarily on the risks associated with driving itself. While ethics remains a vital aspect in autonomous driving, this analysis centers solely on risks inherent to driving maneuvers.
- **Defense-related Strategies:** Given that the autonomous vehicle under scrutiny pertains to the defense industry, this study excludes discussions involving warfare-related tactics or defense strategies. The emphasis remains specifically on off-road driving scenarios, disregarding any content associated with military tactics or strategy.
- **Focus on Terrain Driving:** The primary focus of this analysis revolves around off-road driving scenarios and associated risks. While certain aspects of cybersecurity might be pertinent concerning potential future applications, the study predominantly concentrates on risks inherent to off-road driving.
- **Hardware Specifications:** Although the product assumes certain technical components, such as sensors and LiDAR, as integral parts, this research refrains from detailed discussions or preferences regarding specific hardware. It acknowledges their presence within the product but doesn't emphasize or specify particular hardware choices. The primary focus remains on the foundational STPA analysis, which serves as a basis for decision-making during the product development phase.

These delineations establish a clear perimeter for this study, focusing solely on the risks associated with off-road driving scenarios of autonomous vehicles in the defense industry, excluding ethical considerations, defense-related strategies, and specific hardware preferences.

1.8 STPA Method

Conducting a Systems-Theoretic Process Analysis (STPA) [2] is a powerful approach to investigate and identify safety aspects associated with the introduction of autonomy in a vehicle type like the NGV. Here are the fundamental steps to include in the STPA analysis:

1. Define Purpose of the Analysis
    - Identify Losses
    - Identify System-Level Hazards
    - Identify System-Level Constraints
    - Refine Hazards (Optional)
2. Model the Control Structure
    - Definition: A hierarchical control structure is a system model composed of feedback control loops. An effective control structure enforces constraints on the overall system's behavior.
    - Use of Microsoft Visio for Control Structure Modeling. Alternatively, another tool for visualizing/creating clear block diagrams or control structures.
3. Identify Unsafe Control Actions
    - Once the Control Structure has been modeled, the next step is to identify Unsafe Control Actions.
4. Identify Loss Scenarios
    - Once Unsafe Control Actions have been identified, the next step is to identify Loss Scenarios.

The choice of the STPA method for analyzing the safety of the NGV system is particularly advantageous when compared to traditional safety analysis approaches for several reasons:

1. **Comprehensive Understanding:** STPA offers a comprehensive approach to safety analysis by considering not only technical aspects but also human factors, external influences, and system interactions. This holistic view is essential for identifying potential risks comprehensively.
2. **Early-Stage Analysis:** STPA is especially well-suited for systems like NGV that are still in the development phase and may lack extensive operational data. It allows for safety analysis even when limited information about the system is available.
3. **Identification of Complex Scenarios:** NGV operates in diverse and unregulated environments, making safety analysis challenging. STPA's ability to uncover safety-critical scenarios and their associated risks is vital in such complex contexts.
4. **Actionable Recommendations:** STPA doesn't just identify hazards; it can also suggests control measures and recommendations for improving safety. This makes it a practical method for decision-makers and engineers.

And selecting the level of ambition is straightforward. The STPA analysis is designed for multiple iterations of each step, offering flexibility to determine the preferred level of detail or what aligns with the specific context of the project. By employing the STPA method, even in scenarios where the system is relatively unknown, one can proactively address safety concerns and enhance the overall safety of the NGV system [2].

1.9 Method

1.10 Data Collection

Throughout the course of the research, guidance and assistance will be sought from experts at Milrem Robotics to ensure the relevance of the analysis and the alignment of its outcomes with their product. Additionally, support and insights will be sought from systems safety experts at AFRY, enriching the analysis and ensuring its robustness.

The foundational phase involves gathering technical drawings, consulting experts, and reviewing documented specifications from Milrem Robotics. These invaluable resources offer intricate insights into THeMIS NGV's design, functionalities, and safety measures.

1.11 STPA

The core of this research lies in executing a STPA specifically targeting the NGV concept to address research question 1. (What risks can be concretely identified in the early stages of developing a self-driving vehicle?) The outcome of this STPA analysis primarily revolves around delineating numerous specific scenarios that must be avoided. Additionally, it involves creating tables resembling requirements, forming a robust foundation for subsequent detailed requirement analyses.

1.12 Comparative Analysis

Analyzing the results derived from each step of the STPA process will be thoroughly discussed, both within the analysis section and subsequently in the overall discussion of the report. Comprehensive comments at each stage aim to facilitate easy comprehension and alignment with the analysis. The findings stemming from the STPA analysis will be compared with results from other analyses to address research question 2. (How can we ensure the validation of an autonomous control system? This includes considering the use of the STPA method and other applicable approaches.) This comparative study will delve into the strengths and weaknesses of each analysis method applied, shedding light on their efficacy within NGV's unique terrain challenges and autonomous control system validation.

2 STPA analysis

This chapter presents the System-Theoretic Process Analysis (STPA) for the safety assessment of the implementation of autonomy in the NGV. The STPA methodology is a systematic approach to identifying and analyzing safety-related aspects in complex systems. The analysis consist of four general steps.

1. Define purpose of the analysis
2. Model the control structure
3. Identify unsafe control actions
4. Identify loss scenarios

The foundation of the analysis itself is primarily derived from the STPA Handbook [2].

2.1 Define purpose of the analysis

This chapter marks the first step in the systematic analysis of the safety aspects associated with the implementation of autonomy in the NGV. In this phase of the analysis, the following key tasks will be accomplished:

1. **Identify Losses:** Define critical and known undesirable outcomes to establish a clear safety objective. These losses represent the major events that must be avoided during the NGV's autonomous operation.
2. **Identify System-level Hazards:** Explore potential system-level failures that could lead to the previously identified losses. This step focuses on understanding what could go wrong at the broader system level.
3. **Identify System-level Constraints:** Based on the insights gained from identifying hazards, establish limitations and restrictions on the system to prevent the occurrence of undesirable outcomes. This step essentially defines what should not happen to ensure safety

The analysis will begin by defining losses related to the autonomous operation of the NGV, followed by the identification of system-level hazards and constraints. The optional step of refining hazards will be undertaken as necessary to ensure a comprehensive evaluation of the NGV's safety. This initial step will also provide a sufficient basis for the first iterations of the control structure that will be developed in the next step of the STPA analysis.

2.1.1 Defining the purpose of the analysis

The primary purpose of this analysis is to systematically evaluate the system safety aspects associated with the implementation of autonomy in the NGV. This analysis focuses on identifying potential losses, hazards at the system level, and system limitations. This analysis will have a specific focus on autonomy and its integration into the system. Autonomy is integrated in the following three areas, which will be of greater significance in later stages of the analysis:

- **Obstacle Avoidance**: Autonomy is integrated into the NGV's ability to detect and avoid obstacles using sensors like cameras and Lidar.
- **Waypoint Navigation**: The NGV can be autonomously programmed to follow a predefined series of waypoints, facilitating autonomous terrain navigation for tasks such as transporting goods between specific locations or continuous driving along a predetermined route.
- **Follow-me Mode**: In this mode, the NGV can be autonomously programmed to track and follow an operator or another vehicle, making it valuable for scenarios requiring human supervision or autonomous following of a manned vehicle.

2.1.2 Identifying stakeholder losses

In the initial phase of the analysis, the primary focus is on identifying stakeholder losses. This entails outlining the significant and overarching undesirable outcomes that needs to be avoided within our system. These losses represent the pivotal events that must be prevented at all costs.The subsequent stages of the analysis are built upon these stakeholder losses, as they serve as a concrete foundation for the entire process. Losses in the context of NGV's autonomous operation can be categorized as shown in Table 2.1. In the table, the first column specifies the type of loss. The next column indicates the system under examination, and the third column outlines the corresponding stakeholder loss along with a brief description.

Table 2.1: Losses in Autonomous NGV Operation.

Loss	System	Description
L-1	NGV	**Loss of Life:** The NGV may pose a risk to human life if its autonomy system malfunctions or makes incorrect decisions during missions.
L-2	NGV	**Personal Injury:** Autonomous NGV operations may result in injuries to personnel in the vicinity if safety protocols are not rigorously enforced.
L-3	NGV	**Vehicle and System Damage:** Unintended incidents involving the NGV could lead to damage to infrastructure, valuable equipment, or the vehicle itself, incurring significant financial costs and potential malfunctions.
L-4	NGV	**Environmental Damage:** The NGV's actions may have adverse effects on the environment, leading to significant ecological and financial consequences.
L-5	NGV	**Data Security Vulnerability:** The NGV's data systems may have weaknesses that could be exploited, potentially exposing sensitive information and raising concerns about privacy and security.

2.1.3 Further explanation of definitive stakeholders

In the context of stakeholder losses associated with NGV's autonomous operation:

- **L-1 (Loss of Life):** The primary concern revolves around potential risks to human life due to malfunctions or incorrect decisions made by the NGV's autonomy system during its missions. Definitive stakeholders in this category include individuals directly exposed to the NGV's operational sphere. The NGV's sensor suite, incorporating cameras and Lidar, is tasked with detecting obstacles and potential risks in real-time. Ensuring precision in sensor data interpretation and rapid decision-making algorithms within the NGV's autonomy system is crucial to mitigate such risks.
- **L-2 (Personal Injury):** This category encompasses potential harm to individuals in the NGV's operational area. The definitive stakeholders involve anyone who might face injuries due to the NGV's actions. The NGV's ability to follow predefined waypoints or an operator in 'Follow-me Mode' necessitates robust safety protocols to prevent collisions or incidents, demanding continuous monitoring and integration of safety features within the NGV's autonomous functions.
- **L-3 (Vehicle and System Damage):** This category encompasses a broad range of stakeholders. Definitive stakeholders include entities associated with critical infrastructure such as electrical grids, telecommunication networks, and valuable equipment. Additionally, the NGV itself is a key stakeholder within this category; any unintended incidents involving the NGV that cause damage could lead to significant financial costs, disrupt essential services, and potentially result in malfunctions within the NGV's operational system. The NGV's autonomy, incorporating obstacle avoidance sensors and waypoint navigation, needs precise calibration and rigorous testing to prevent collisions or damage to both itself and surrounding entities.
- **L-4 (Environmental Damage):** Definitive stakeholders here involve ecological systems and entities impacted by the NGV's actions. These stakeholders include local ecosystems and communities affected by adverse environmental consequences due to the NGV's operation. Furthermore, it is essential to consider the preservation of fragile ecosystems and specific wildlife, avoiding disruptions that could harm or disturb endangered species and nature's balance. Robust obstacle avoidance and navigation systems within the NGV's autonomy must be developed to ensure minimal impact on fragile ecosystems, necessitating specific protocols to avoid disturbances in environmentally sensitive areas.
- **L-5 (Data Security Vulnerability):** This category encompasses individuals or entities concerned with data security and privacy. Definitive stakeholders involve organizations managing the NGV's data systems, potentially facing catastrophic consequences if subjected to cyberattacks. The possibility of malicious entities gaining control over the NGV's systems can have devastating outcomes, compromising sensitive information or even allowing adversaries to take control. Strengthening cybersecurity measures within the NGV's systems, incorporating encryption protocols and continuous vulnerability assessments, is imperative to prevent potential breaches that could compromise critical data.

These five categories encompass critical aspects that the analysis will prioritize. While acknowledging the possibility of other potential risks, these key areas cover the major domains where potential damages are likely to occur. They provide a comprehensive framework for preventing and mitigating potential hazards, ensuring a generally secure and robust operational system.

2.1.4 Identifying system-level hazards

System-level hazards for autonomous NGV operation can be identified as in Table 2.2. In the table, the first column designates the type of hazard. The next column indicates the system under scrutiny, the third column specifies the particular system hazard defined, and the last column outlines the corresponding stakeholder loss that could occur in the event of this system-level hazard.

Table 2.2: System-level Hazards and Affected Stakeholder Losses.

Hazard	System	Unsafe Condition	Affected Stakeholder Losses
H-1	NGV	Fails to maintain controlled movement	[L-1, L-2, L-3, L-4]
H-2	NGV	Exceeds safe speed	[L-1, L-2, L-3, L-4]
H-3	NGV	Fails to maintain safe distance to obstacles	[L-1, L-2, L-3, L-4]
H-4	NGV	Misinterprets the authorized operator's commands	[L-1, L-2]
H-5	NGV	Fails to detect hazardous internal conditions	[L-3, L-4]
H-6	NGV	Fails to maintain structural integrity when driving in rough terrain	[L-3, L-4]
H-7	NGV	Fails to authenticate operator, safety officer or both	[L-5]
H-8	NGV	Fails to maintain safe distance to hazardous terrain	[L-1, L-2]
H-9	NGV	Fails to accurately identify soft obstacles in terrain, such as tall grass or dense undergrowths	[L-1, L-2, L-3, L-4]
H-10	NGV	Misinterprets information displayed on the user interface, leading to incorrect decisions by the operator	[L-1, L-2, L-3, L-4]

2.1.5 Further explanation of each definitive hazard

H-1 The NGV Fails to Maintain Controlled Movement

Hazard H-1 involves the NGV's potential failure to maintain controlled movement, leading to various stakeholder losses:

- **Loss of Life (L-1):** Controlled movement is critical for avoiding accidents and ensuring the safety of human life. Unpredictable behavior in the autonomous NGV could pose a threat to individuals nearby.
- **Personal Injury (L-2):** This constraint is rooted in the same principle as L-1, emphasizing the vital importance of maintaining controlled movement to prevent injuries to humans in the vicinity. Unexpected actions or erratic behavior of the autonomous NGV can jeopardize the safety of individuals in its proximity.
- **Property Damage (L-3):** Uncontrolled movement or unexpected behavior may result in stress or damage to the vehicle, potentially causing malfunctions or mechanical failures.

H-2 The NGV Exceeds Safe Speed

Hazard H-2.1 involves the NGV exceeding safe speed limits, resulting in stakeholder losses:

- **Loss of Life (L-1):** Exceeding safe speeds could pose a threat to individuals nearby as the system isn't designed to handle such speeds, leading to unpredictable consequences if control is lost.

- **Personal Injury (L-2):** This constraint is rooted in the same principle as L-1, operating at unsafe speeds could result in harm to individuals nearby.
- **Property Damage (L-3):** Excessive speeds can cause damage to the NGV system or infrastructure due to the system's limitations in handling high velocities.
- **Environmental Damage (L-4):** High speeds can have adverse effects on the environment, impacting ecosystems and habitats due to the increased force and impact on surroundings.

H-3 The NGV Fails to Maintain Safe Distance to Obstacles

Hazard H-3 involves the NGV's failure to maintain a safe distance to obstacles, resulting in stakeholder losses:

- **Loss of Life (L-1):** Inadequate distance maintenance from obstacles poses a risk to human life, especially in hazardous conditions where unexpected events may occur.
- **Personal Injury (L-2):** Failing to keep a safe distance may lead to hazards or unexpected encounters, potentially causing injuries to bystanders.
- **Property Damage (L-3):** Insufficient distance maintenance can result in potential damage to the NGV's mechanical components or infrastructure due to unanticipated encounters.
- **Environmental Damage (L-4):** Inadequate distance maintenance may lead to environmental impact, particularly in different terrains or ecosystems where maintaining safe distances is crucial for minimizing harm.

H-4 Misinterprets the Authorized Operator's Commands

Hazard H-4 involves the risk of the NGV misinterpreting the authorized operator's commands, leading to stakeholder losses:

- **Loss of Life (L-1):** Misinterpretation of commands could lead to critical situations, potentially endangering human life due to unexpected NGV behavior.
- **Personal Injury (L-2):** Incorrect interpretations may lead to accidents, causing injuries to bystanders due to unpredictable NGV actions.

H-5 Fails to Detect Hazardous Internal Conditions

Hazard H-5 involves the NGV's failure to detect hazardous internal conditions, resulting in stakeholder losses:

- **Property Damage (L-3):** Failure to detect hazardous conditions may lead to potential damage to the NGV's components or infrastructure.
- **Environmental Damage (L-4):** Inadequate detection of hazardous conditions may result in the NGV operating in environmentally sensitive areas. This could cause disturbances or damage to ecosystems and habitats due to the vehicle's unintended actions in fragile environments.

H-6 Fails to Maintain Structural Integrity When Driving in Rough Terrain

Hazard H-6 involves the NGV's potential failure to maintain structural integrity when driving in rough terrain, leading to stakeholder losses:

- **Property Damage (L-3):** The rugged terrain might cause wear and tear or severe damage to the vehicle's mechanical parts, impacting its structural integrity.
- **Environmental Damage (L-4):** Operations in hazardous terrain can result in environmental damage, disrupting delicate ecosystems or habitats. The NGV's presence in such areas might lead to disturbances for nature and animals or cause environmental damage.

H-7 Fails to Authenticate Operator, Safety Officer, or Both

Hazard H-7 involves the risk of the NGV failing to authenticate the operator, safety officer, or both, leading to stakeholder losses:

- **Data Security Vulnerability (L-5):** Unauthorized access to internal functions and sensitive information could compromise critical data, potentially allowing adversaries to manipulate the NGV's operations. This breach may result in significant data security risks, compromising the integrity and confidentiality of vital information. Moreover, hostile control over the NGV could lead to adverse actions, posing serious threats to security and causing severe harm to individuals and infrastructure.

H-8 Fails to Maintain Safe Distance to Hazardous Terrain

Hazard H-8 involves the NGV's failure to maintain a safe distance to hazardous terrain, resulting in stakeholder losses:

- **Loss of Life (L-1):** Incorrect interpretation of terrain may result in accidents or incidents in 'Follow Me' mode. If, for instance, a person falls and lies on the ground, it's important that the vehicle does not proceed further.
- **Personal Injury (L-2):** Misinterpretation of terrain features may cause accidents, resulting in injuries to personnel.

H-9 Fails to Accurately Identify Soft Obstacles in Terrain

Hazard H-9 involves the NGV's failure to accurately identify soft obstacles in terrain, such as tall grass or dense undergrowths, leading to stakeholder losses:

- **Loss of Life (L-1):** Incorrect identification of obstacles, such as a person standing in tall grass, may pose a risk to human life.
- **Personal Injury (L-2):** Failure to accurately identify obstacles, like individuals hidden in complex environments, may result in injuries to personnel.
- **Property Damage (L-3):** Inaccurate identification of obstacles can lead to damage to infrastructure.
- **Environmental Damage (L-4):** Failure to identify obstacles correctly may have adverse effects

H-10 Misinterprets Information Displayed on the User Interface

Hazard H-10 involves the risk of the NGV misinterpreting information displayed on the user interface, resulting in stakeholder losses:

- **Loss of Life (L-1):** Misinterpretation of crucial information on the UI may lead to erroneous decisions by the operator, potentially endangering human life due to unexpected NGV behavior.
- **Personal Injury (L-2):** Incorrect interpretations of displayed information could lead to accidents, causing injuries to bystanders due to unpredictable actions by the NGV.
- **Property Damage (L-3):** Misinterpreting UI information might result in erratic vehicle behavior, causing damage to the NGV itself or infrastructure due to unexpected maneuvers or collisions.
- **Environmental Damage (L-4):** Erroneous interpretations leading to NGV actions in sensitive environmental areas or ecosystems could cause disturbances or harm to habitats, impacting the environment due to the vehicle's unintended actions.

From a technical standpoint, the system-level hazards identified within the autonomous NGV stem from the complexities inherent in its operations. These complexities highlight the challenges in ensuring its safe and effective functionality. The NGV's intricate design, specifically tailored for use in complex and unregulated terrains, demands a meticulous examination of potential risks. This has led to the compilation of specific hazards and their corresponding implications for stakeholders.

While hazards such as H-1 and H-2 were delineated to encompass various facets of risks, ensuring a comprehensive understanding without overwhelming the analysis, other hazards underwent refinement and development in their singular forms. Hazards H-8 and H-9, while specific situations, were deemed critical scenarios essential to encompass within this framework.

Hazard H-7, addressing potential cyber threats, holds significant potential for further exploration. However, in this analysis, the focus remains on prioritizing autonomy over delving deeply into cybersecurity specifics. Yet, acknowledging the relevance of cybersecurity within autonomous NGV operations, this hazard remains a crucial consideration despite limited elaboration.

While the potential for additional hazards or aspects to consider exists, these selected ten hazards offer a multifaceted approach, covering various stakeholder losses. They establish a robust foundation for further analysis and decision-making processes. While certain hazards could benefit from further elaboration, maintaining conciseness within this scope aims to ensure a comprehensive yet manageable analysis.

These hazards now serves as the foundation representing the risks in the NGV's operations in the analysis.

2.1.6 Defining system-level constraints

System-level limitations for the autonomous NGV are as shown in Table 2.3. In the table, the first column identifies the type of constraint. The next column specifies the system under consideration, the third column outlines the specific constraints, and the fourth column indicates the associated hazard.

Table 2.3: System-level Constraints and Associated Hazards.

Constraint	System	Condition to Enforce	Associated Hazard
SC-1	NGV	Must not fail to maintain controlled movement	[H-1]
SC-2	NGV	Must not exceed safe speed	[H-2]
SC-3	NGV	Must not fail to maintain safe distance to obstacles	[H-3]
SC-4	NGV	Must not misinterprets the authorized operator's commands	[H-4]
SC-5	NGV	Must not fail to detect hazardous internal conditions	[H-5]
SC-6	NGV	Must not fail to maintain structural integrity when driving in rough terrain	[H-6]
SC-7	NGV	Must not fail to authenticate operator, safety officer or both	[H-7]
SC-8	NGV	Must not fail to maintain safe distance to hazardous terrain	[H-8]
SC-9	NGV	Must not fail to accurately identify soft obstacles in terrain, such as tall grass or dense undergrowth	[H-9]
SC-10	NGV	Must not misinterpret information displayed on the user interface	[H-10]

2.1.7 Further explanation of the definitive System Constraints

In accordance with the handbook, identifying system constraints involves a thorough review of all identified hazards, ensuring that each system constraint (SC) effectively mitigates or counters these hazards. This method involves a detailed examination of the potential risks associated with each hazard, such as risks to life, injury, property, and the environment. From this analysis, system constraints are formulated to directly address these risks.

Through a comprehensive analysis of hazards and their potential impacts on stakeholders, system constraints are derived as preventive measures or protective measures against these potential adverse outcomes. The subsequent section lists the specific system constraints devised to mitigate or eliminate identified hazards, ultimately enhancing the overall safety and functionality of the NGV system.

2.2 Model the control structure

The second step in the STPA focuses on developing the control structure as a foundation for safety analysis and assessment. This step builds upon the identification of losses, system-level hazards, and constraints from the first step.

The control structure captures interactions within the NGV's autonomy system and human operators, emphasizing critical control actions, monitoring, and mitigation strategies. It forms the basis for further safety analysis, including identifying critical scenarios, ensuring fault tolerance, and assessing mitigation strategies.

2.2.1 Enhanced Control Structure Understanding

A hierarchical control structure is a key component. It consists of feedback control loops, as illustrated in Figure 2.1 and 2.2.

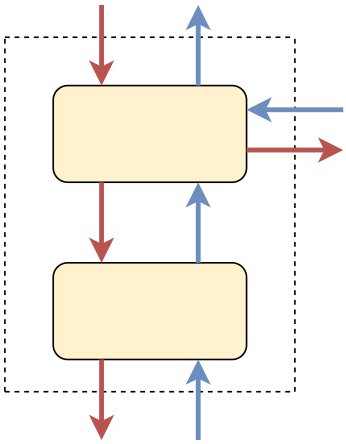


Figure 2.1: Modeling the control structure.

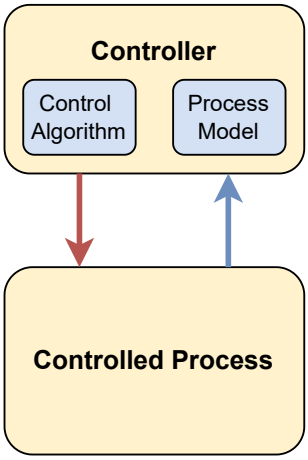


Figure 2.2: Control loop example.

These control loops involve controllers, control algorithms, process models, and feedback mechanisms. They are critical for enforcing constraints on the NGV's autonomy system behavior and ensuring safe operations.

Identifying problems within these control loops, such as inconsistencies in process models or sensor failures, is crucial for maintaining system safety and mitigating potential risks.

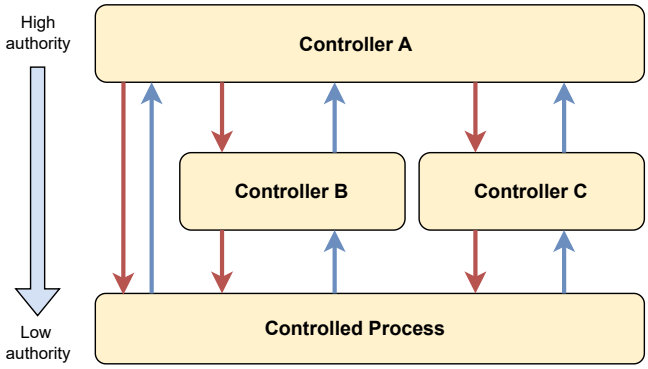


Figure 2.3: Example of a generic hierarchical control structure.

In a typical hierarchical control structure, at least five key elements are present:

1. Controllers
2. Control Actions
3. Feedback
4. Inputs and Outputs to and from components (neither control nor feedback)
5. Controlled processes

The vertical axis in this hierarchical control structure, see Figure 2.3, represents the system's control and authority distribution. It establishes a control hierarchy, with higher-level controllers positioned at the top and lower-level entities situated at the bottom. Each entity holds control and authority over those immediately subordinate to it, while concurrently being subject to the control and authority of higher-level entities.

For example, within the context of aircraft automation, an entity serves a dual role as both a controller, responsible for issuing control actions to aircraft systems, and a controlled process, which executes control actions received from the flight crew and provides feedback.

This vertical arrangement simplifies the management of system complexity and enhances the clarity of control relationships and feedback loops. It also facilitates the identification of issues such as entities providing control actions without receiving essential feedback, feedback being directed to entities incapable of acting upon it, or multiple controllers issuing conflicting commands to the same component without mechanisms for conflict resolution. Detecting and addressing these issues within the control structure diagram is an essential aspect of the systematic STPA methodology.

2.2.2 Modeling the control structure

The methodology that has been adopted here involves going through each System Constraint individually and then creating associated loops with specific parameters to support their respective functionalities. The control structure comprises a primary loop with a controller and a controlled process, following the same structure mentioned in the previous section. The controlled process is here represented as the NGV. The structure also includes an internal loop dedicated to autonomy, emphasizing its significance within the analysis.

This chapter will begin with an of the overview structure of the model, followed by detailed

explanation of each box and every defined System Constraints connection and functionality within the model.

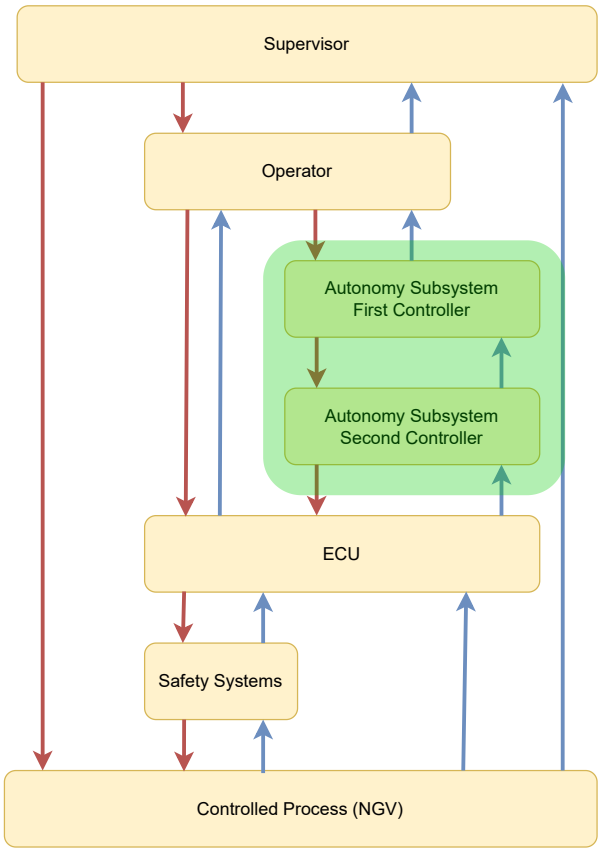


Figure 2.4: Overview structure of the control model.

In Figure 2.4, the developed general control structure of the system is illustrated. To summarize, the control model's structure is as follows: situated at the hierarchy's apex is a Supervisor capable of halting the entire system instantaneously. Following this, an Operator takes charge, directly steering the vehicle or selecting from the three available autonomy modes. Within the green marked area lies the autonomy functionality, covering all three autonomy modes, forming an internal loop where the First Controller internally manages the Second Controller. Internally within the green area, the Second Controller represents a controlled process, while outside the green area, it serves as a controller directing commands towards the ECU.

The Operator and the Autonomous System jointly control the ECU, serving as the core program. Before issuing direct commands to the vehicle, the ECU converts instructions from the controls into executable commands for the NGV. Prior to reaching the NGV, instructions traverse through a block called Safety Systems, acting as a filter between the vehicle and the ECU. This ensures adherence to limitations and filters out unsuitable instructions. The control action arrows are depicted in red, while the feedback arrows are represented in blue, as discussed in the previous section. The NGV acts as the final controlled process. Subsequently, a detailed exploration of each subsystem will provided, elucidating their contents,

explaining the information conveyed by each arrow, and outlining the reasons behind their transmission.

2.2.3 Each block in detail

In Figures 2.5 - 2.10 a more detailed description of each block and its associated arrows is presented.

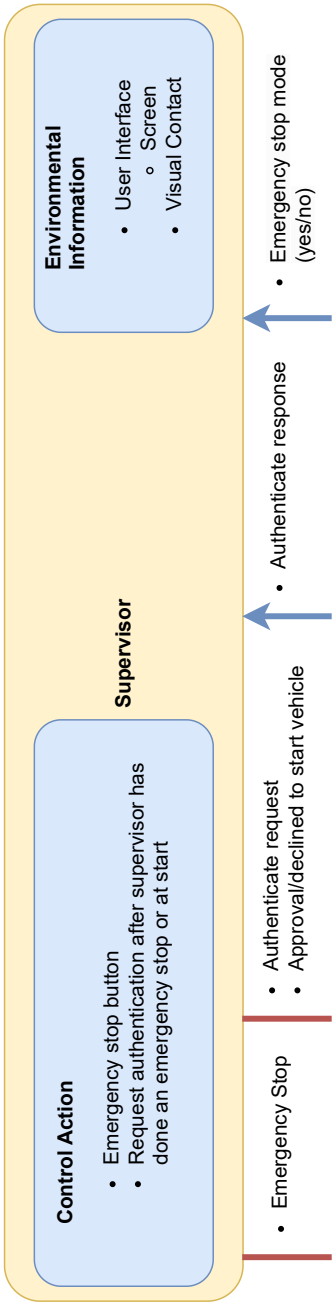


Figure 2.5: Supervisor block.

Located at the top of the hierarchy is the Controller Supervisor, the highest-ranking overseer within the system. This is presented in Figure 2.5. Its instructions take precedence above all else. This role encompasses crucial functionalities, including an emergency stop feature capable of halting the entire system. Additionally, an authenticator ensures that the Operator, the next in line within the hierarchy, verifies their identity before gaining control of the system. The Supervisor is an abstract concept; it remains unclear whether it represents an individual, whether it overlaps with the Operator role, or if it denotes an officer or another entity entirely. Moreover, the Supervisor holds environmental information about the system, potentially accessible through a UI or visual interface. Details such as authentication status and instances of an emergency stop by the NGV constitute the type of information it processes.

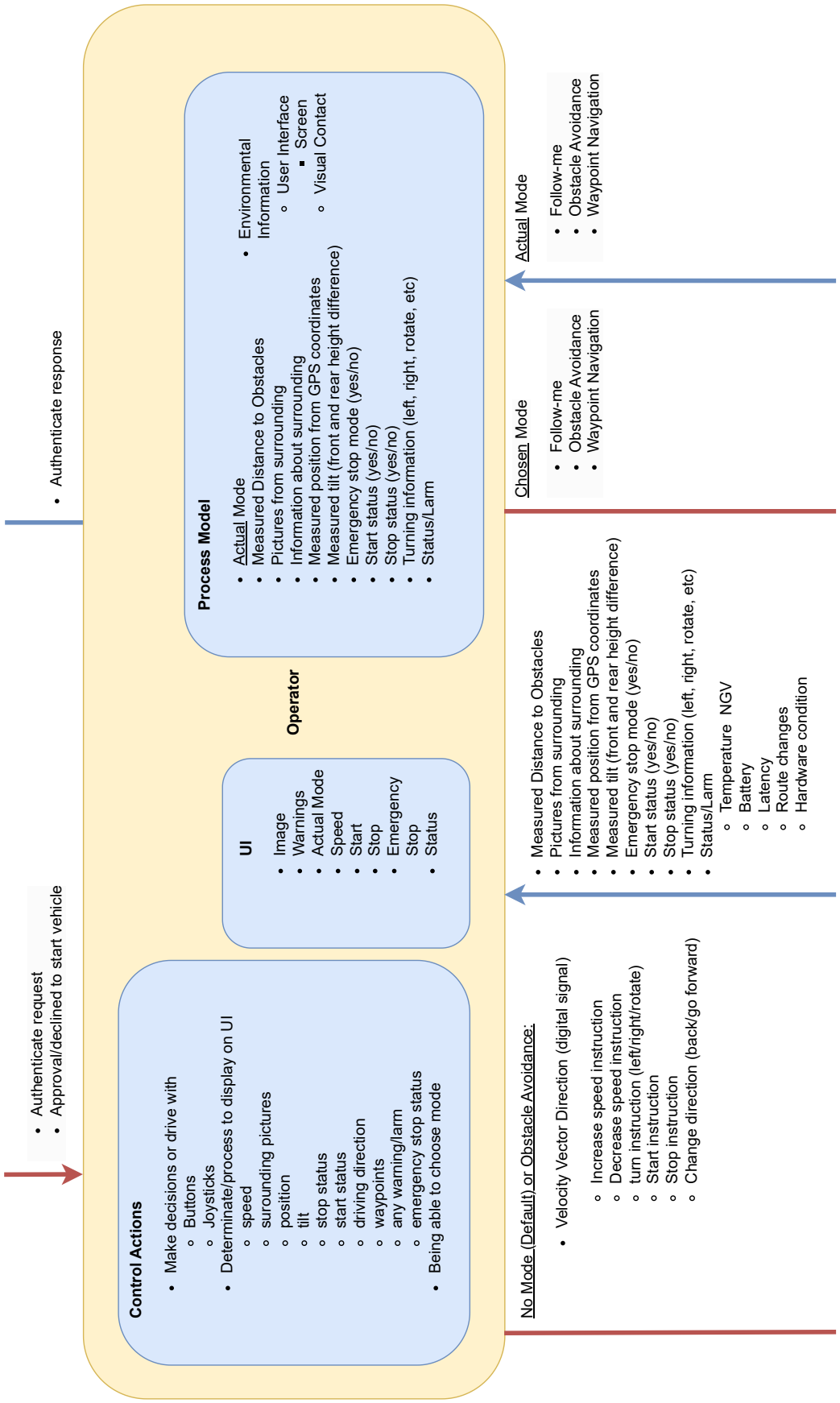


Figure 2.6: Operator block.

Next in line within the hierarchy is the Operator. This is presented in Figure 2.6. This role represents the individual controlling the NGV through a console equipped with a screen displaying various information: speed, potential alarms, position, tilt, emergency stop status, and more. All this data is relayed to the Operator from the ECU, as illustrated in Figure 2.4. As mentioned earlier regarding the Supervisor, the Operator must undergo some form of verification by the Supervisor to assume control of the NGV. Control of the NGV is managed through buttons and joysticks in this block. Additionally, this block encompasses the screen encoder and all UI-related functions. All necessary data displayed here is sent from the ECU, and when the Operator maneuvers the NGV, the signals are sent directly back to the ECU. Figures 2.4 and 2.6 indicate arrows extending toward the autonomous system. This allows the Operator to choose between directly controlling the NGV from their console or activating one of the three autonomous modes available. When an autonomous mode is engaged, the autonomous subsystem takes over control of the NGV.

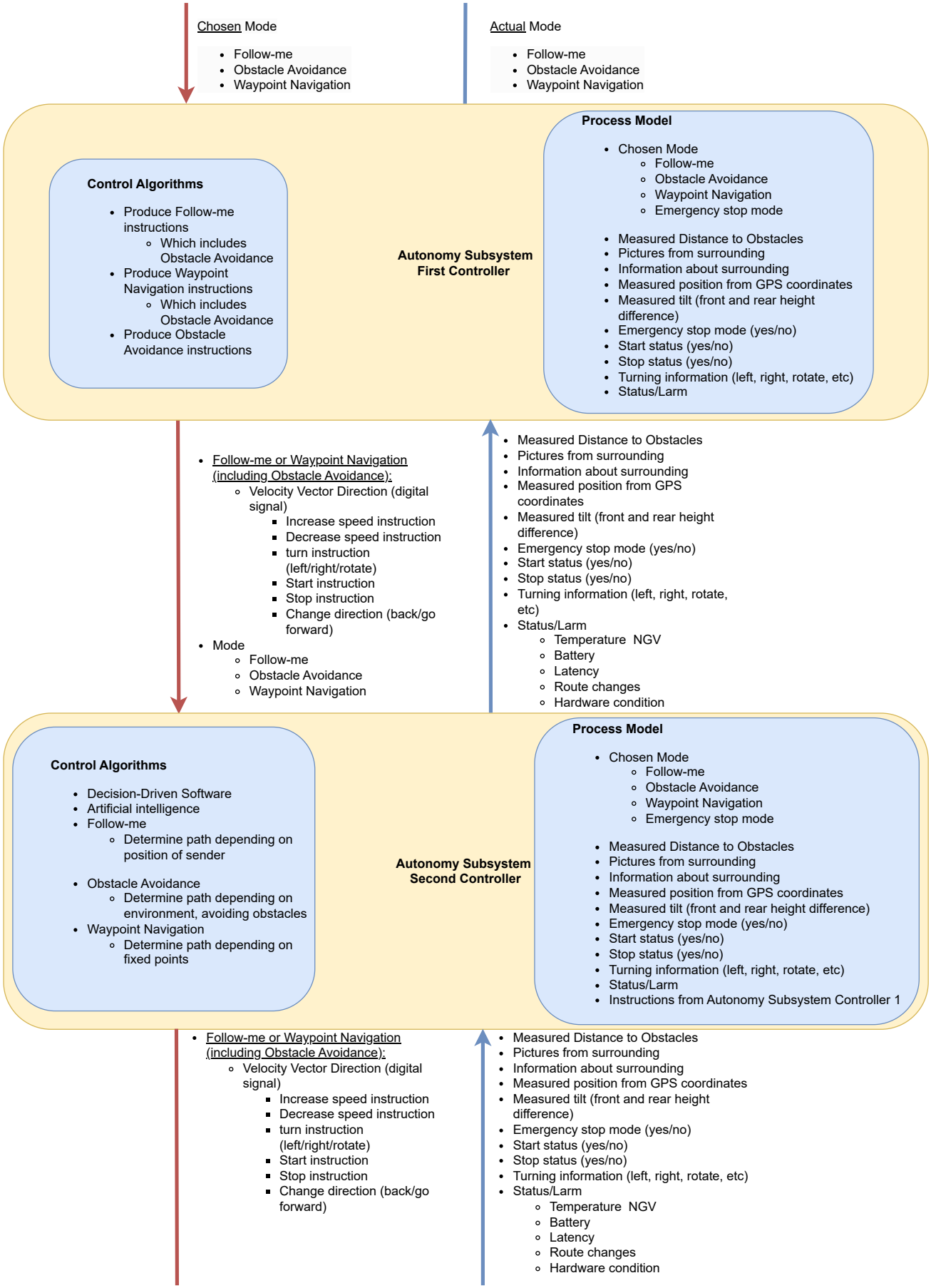


Figure 2.7: Autonomy Subsystem block (First and Second Controller).

Here, the focus shifts to the autonomous subsystem that the Operator can activate. This subsystem receives information regarding the chosen autonomous mode among the available three. It also processes data from the ECU, detailing the movements of the NGV. This information enables the autonomous subsystem to send control signals back to the ECU, based on the ongoing actions of the NGV. Within this subsystem, all intelligent features, artificial intelligence, and decision-driven software operate, allowing for autonomous decision-making regarding the NGV's operations.

As depicted in Figure 2.7, the autonomous subsystem comprises two blocks. The first block, known as the First Controller, functions locally and serves as a controller for the subsequent block, named the Second Controller. This illustrates a system where instructions for autonomy's decisions and operations are transmitted and executed. In the broader system context, the Second Controller operates as a controller over the ECU.

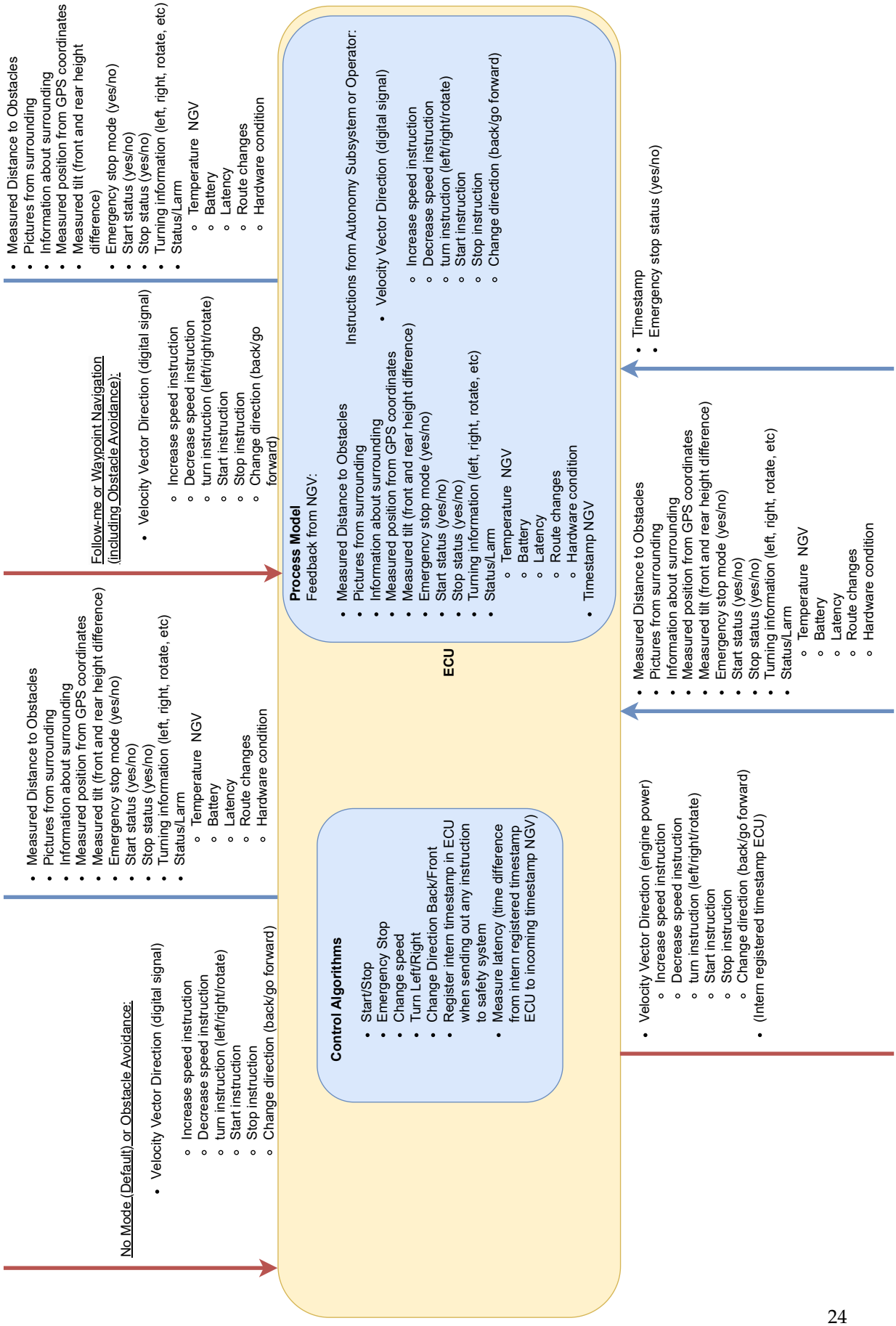


Figure 2.8: ECU block.

In Figure 2.8, the block representing the ECU is presented, this is essentially serving as the central program for the entire system. Most of the system's functionalities are contained within this component. The ECU's primary responsibility lies in interpreting control signals and converting them into directives capable of steering the NGV. It receives instructions either from the Operator or the autonomous subsystem, transforming these into signals interpretable by the NGV, thereby facilitating the NGV's movement.

The output from this block primarily consists of power, while it predominantly receives digital commands. There should likely be a form of DAC (Digital-to-Analog Converter) within this block. The ECU does not differentiate between signals received from the Operator or the autonomous subsystem; for instance, if it receives a 'turn right' signal, it will execute the action regardless of the source. However, it transmits data back to the controlling entities—Operator or autonomous system—providing feedback in the form of speed, position, potential alarms, etc. This feedback is crucial for steering the NGV effectively based on the circumstances.

Moreover, this block timestamps the moment signals are dispatched to the NGV and receives another timestamp upon completion of the operation by the NGV. This mechanism facilitates the measurement of potential latency, which is relayed as information to the controlling entity for monitoring purposes.

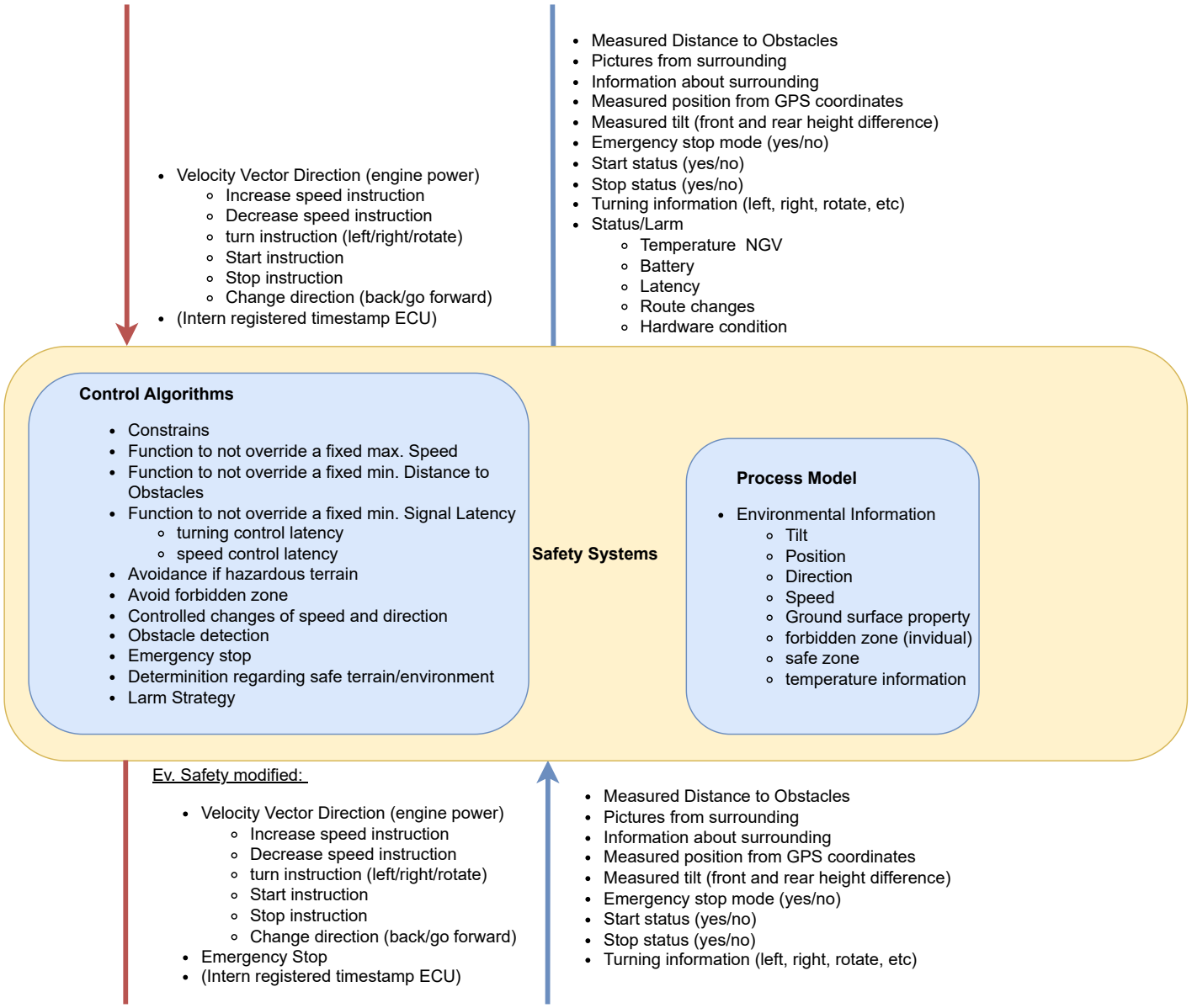


Figure 2.9: Safety Systems block.

Between the ECU and the NGV lies a block called the Safety Systems. This is presented in Figure 2.9. This component acts as a filter positioned between the NGV and the ECU. Its primary role is to monitor all potentially hazardous speeds and activities, imposing limitations and functionalities to ensure that the NGV does not engage in any unsafe actions. For instance, it prevents surpassing specific speeds, getting too close to obstacles, or entering hazardous terrain.

Additionally, it serves to alert about any hardware malfunctions, excessively high temperatures, battery health concerns, and more. It then proceeds to forward potentially modified signals to the NGV that are deemed safe or halts the NGV's operations altogether when

necessary.

The Safety Systems gather information about the NGV's operations and relay the exact data, along with status updates and potential alarms, to the ECU. This ensures that the ECU receives comprehensive information regarding the NGV's activities and any associated safety concerns.

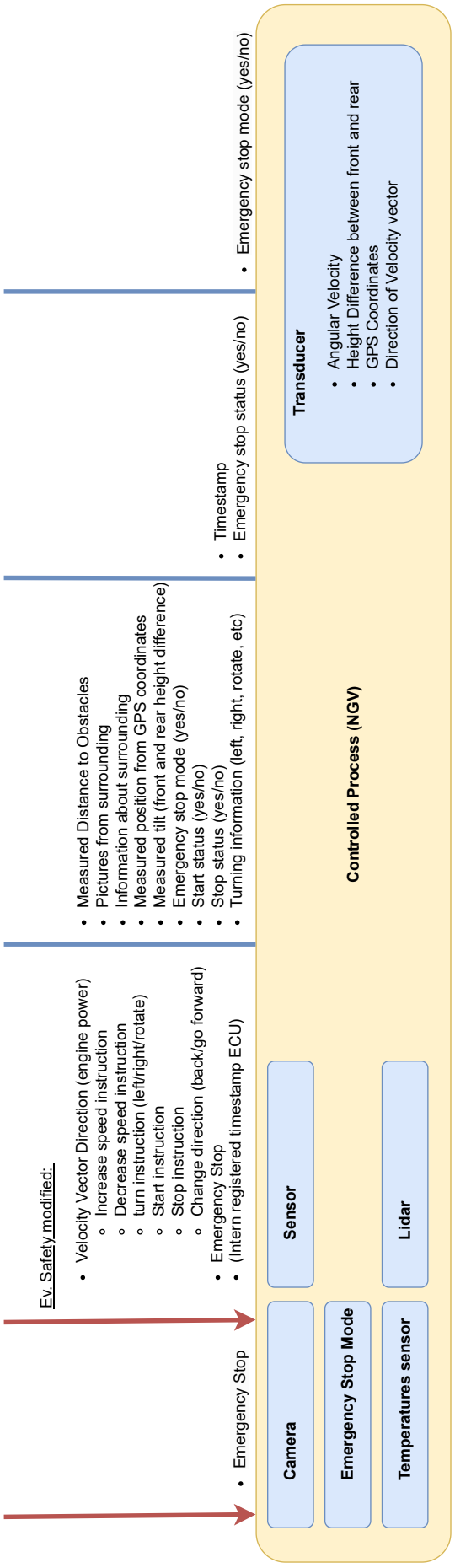


Figure 2.10: Controlled Process (NGV) block.

Finally, the NGV represents the ultimate Controlled Process, situated at the lowest level of the hierarchy. This is presented in Figure 2.10. It receives control signals from the ECU, filtered through the Safety Systems block, and proceeds to execute tasks accordingly. It also responds to emergency stop commands issued by the Supervisor, denoted as the arrow on the far left in Figure 2.4. The NGV communicates information regarding emergency stop activations to the system and relays data about its movement, encompassing speed, position, internal temperature, camera imagery, and other relevant details.

Equipped with cameras, sensors, LiDAR technology, and various measurement instruments such as angular velocity sensors for speed determination, height differential sensors for tilt estimation, GPS for coordinates, and velocity vectors, the NGV gathers comprehensive information about its environment and operational parameters.

2.2.4 Further explanation the definitive Control Structure

The development of the definite control structure was done by systematically examining all system constraints (SC) from Table 3.3 one by one and sequentially constructing the control structure. It evolved from its initial simplicity by iteratively adding functionalities. By addressing SC-1 and evaluating the essential components required, a loop was established. This process was repeated with SC-2, adding the necessary elements specific to that constraint, and so on. In this manner, the control model has been assembled, aligning each system constraint with the corresponding components within the model.

The developed control structure rigorously adheres to the methodology outlined in the STPA handbook, considering each identified system constraint and its connection to related system events and risks.

An upcoming explanation of each distinct system constraint will elucidate how they are all traceable within the model. This process aims to clarify the linkage between each constraint and its representation within the control structure. The upcoming explination ensures that each system constraint is appropriately captured and integrated into the overall control framework before proceeding to the next stage where each control action, the red arrows, are going to be analysed.

Additionally, the complete, unreduced control structure, inclusive of all details, is attached as a separate document to this report. This standalone attachment allows for direct access to the complete control structure for a comprehensive understanding of the system's safety framework, and with that the analysis can be easier to follow.

This analysis will involve assessing all control actions (depicted by the red arrows) in the subsequent analysis phase. Additionally, the examination will cover all feedback arrows (the blue ones) essential for the logic in the structure to ensure comprehensive information flow. These feedback arrows may also be scrutinized in the final stage of the whole analysis.

2.2.4.1 Development from system constraints

The evolution of the control structure, depicted in Figure 2.11, is a result of the analysis of each system constraint (SC). The figures presented here are intentionally incomplete to enhance clarity in following the development process. Notably, certain aspects, such as the NGV's capability for steering and braking, are assumed throughout the analysis, even if not explicitly mentioned in every step. Figure 2.11 shows a visual representation illustrating the

progression of the control structure from step 1 to step 4. The following explanations will detail how the control structure evolved by examining each SC from start to finish.

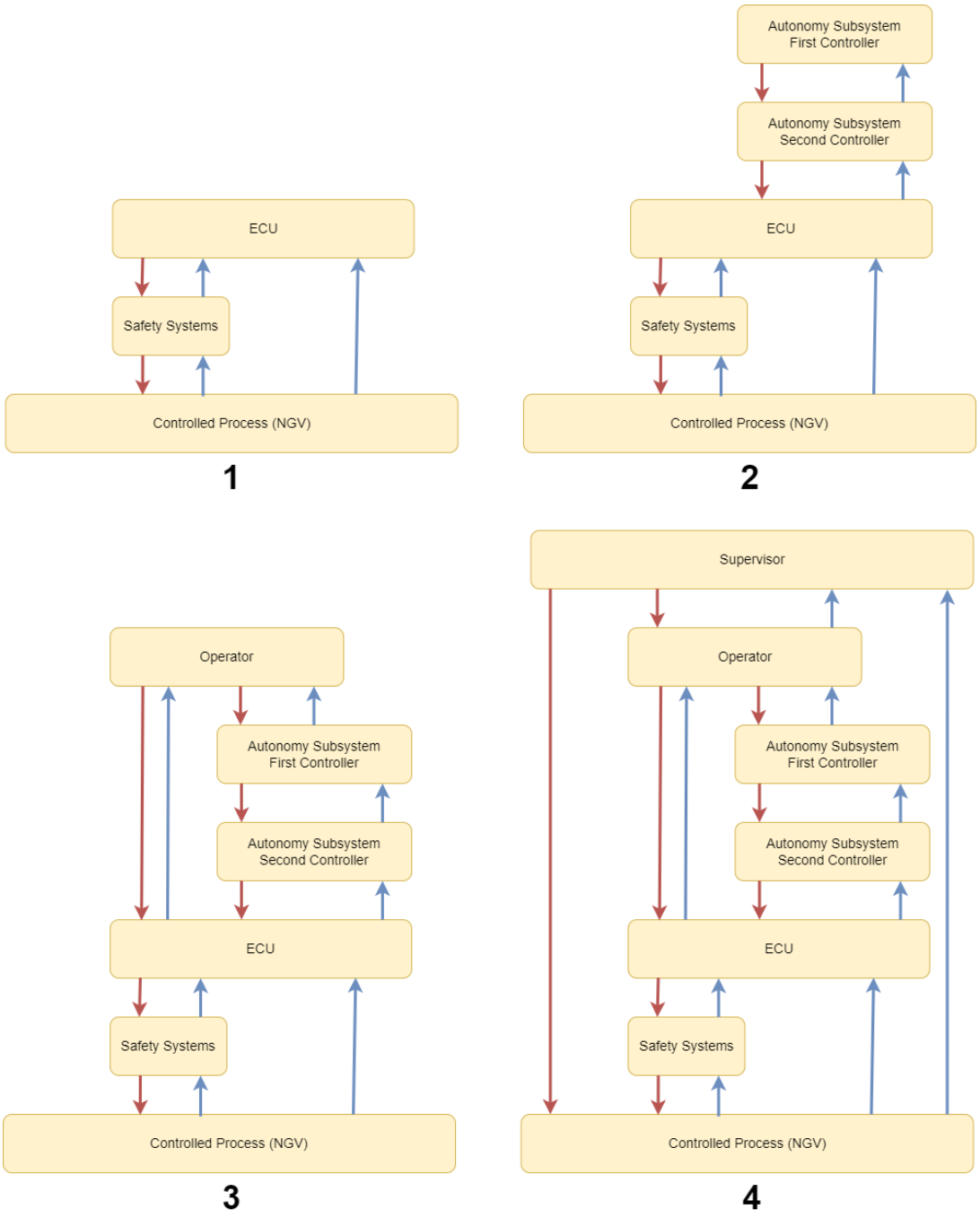


Figure 2.11: Process of developing the control structure.

SC-1 must not fail to maintain controlled movement

The initial loop implemented aimed to fulfill SC-1 requirements. At this stage, the system comprised solely of the necessity for controlled motion of the NGV. Consequently, an ECU was mandated to govern the NGV, incorporating safety processes within the Safety Systems blocks and the NGV. Essential feedback arrows were present. The contents within all blocks

were somewhat condensed from Figure 2.8- 2.10, yet the control algorithms and process models remained fundamentally unchanged, along with the similarity of the feedback arrows. Something directed the NGV, and feedback was received concerning its direction. A safety system ensured that the control adhered to defined parameters of 'controlled' movement. Further limitations were introduced in subsequent stages. In Figure 2.11, step 1, there is an initial draft of the system model.

SC-2 must not exceed safe speed

Within SC-2, the "must not exceed safe speed" requirement was essentially an additional specification incorporated into the Safety Systems. It represents a limiting function ensuring that a certain speed limit is not surpassed. Figure 2.9 shows this constraint within the Control Algorithms. The overall structure remains consistent with that depicted in
Figure 2.11, step 1.

SC-3 must not fail to maintain safe distance to obstacles

With the introduction of requirement SC-3, autonomy becomes a factor, and the autonomous subsystem now governs the NGV. An autonomous system is required to make decisions regarding its driving behavior, particularly in situations where obstacles are in proximity. It must possess the capability to interpret and avoid these obstacles rather than collide with them. Figure 2.11, step 2, is refered for this expansion.

SC-4 must not misinterprets the authorized operator's commands

An operator is required, positioned at the top of the hierarchy to fulfill SC-4. Figure 2.11, step 3, is refered for this expansion.

SC-5 must not fail to detect hazardous internal conditions

In the context of SC-5, the relevant reference remains in Figure 2.11, step 3. An alarm/status is now transmitted from the Safety Systems block to the ECU and further elevated to the highest level within the hierarchy. This involves a monitoring system for alarm/status concerning various factors such as Temperature, Battery, Latency, Route changes, and Hardware condition within the NGV. Refer to the feedback arrow from the Safety Systems block in Figure 2.9 for this integration.

SC-6 must not fail to maintain structural integrity when driving in rough terrain

For SC-6, Figure 2.11, step 3, remains pertinent. Additional features have been incorporated into the Process Model within various blocks and in the Control Algorithms. Information regarding tilt and position has been included to enhance environmental awareness for improved navigation in rough terrain. Furthermore, an alarm/status related to hardware condition has been introduced to address its relevance in such circumstances.

SC-7 must not fail to authenticate operator, safety officer, or both

Arriving at the final generic level of the control structure, it becomes apparent that this structure is the same as Figure 2.4. To ensure cybersecurity, an additional level is required to verify the identity of the individual operating the NGV. This addition is illustrated by introducing another entity into the system model with the highest hierarchy. Figure 2.11, step 4, is refered for this expansion.

SC-8 must not fail to maintain a safe distance to hazardous terrain

Once again, in fulfilling SC-8, functionalities are appended to both the autonomous and Safety Systems. These augmentations aim to ensure the maintenance of a safe distance from hazardous terrain.

SC-9 must not fail to accurately identify soft obstacles in terrain, such as tall grass or dense undergrowth

Similarly, within SC-9, enhancements are incorporated into both the autonomous and Safety Systems. The objective is to guarantee the accurate identification of soft obstacles in terrain, encompassing features like tall grass or dense undergrowth.

SC-10 must not misinterpret information displayed on the user interface

Regarding SC-10, a refinement involves integrating the UI as a component within the Operator block. Refer to Figure 2.6 for this integration.

2.3 Identify unsafe control actions

The subsequent phase of the analysis entails scrutinizing the Control Actions within the system framework. These actions are represented by the red arrows in the definitive control model illustrated in Figure 2.4, with detailed Control Actions outlined in Figure 2.5- 2.10. In the handbook it is generally explained as in Figure 2.12.

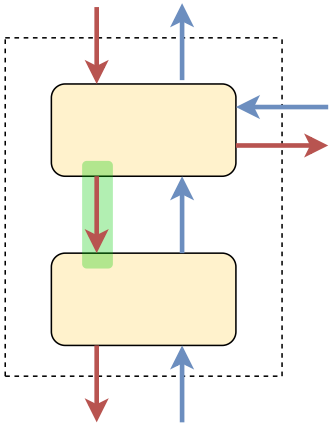


Figure 2.12: Identify Unsafe Control Actions.

Definition: An Unsafe Control Action (UCA) is defined as a control action that, within a specific context and worst-case scenario, could lead to a hazard.

As per the handbook, the following categories are examined when identifying various UCAs:

- Not providing causes hazard
- Providing causes hazard too early, too late, out of order
- Stopped too soon, applied too long

Therefore, these categories serve as the headings in the subsequent tables. "Unsafe" in this context refers to the hazards identified in STPA. The Control Actions being investigated are delineated in the corresponding figures within the generic control model to ensure clear traceability and analysis.

The Control Action arrows under investigation are highlighted in green in the left-hand side upcoming figures, and the specific Control Actions examined are aligned on the right side of the figures for easy correlation in Figures 2.13- 2.20.

After defining the Unsafe Control Actions (UCAs), the subsequent step involves establishing Controller Constraints derived from these actions. Each UCA is utilized to outline specific constraints governing the behavior of controllers.

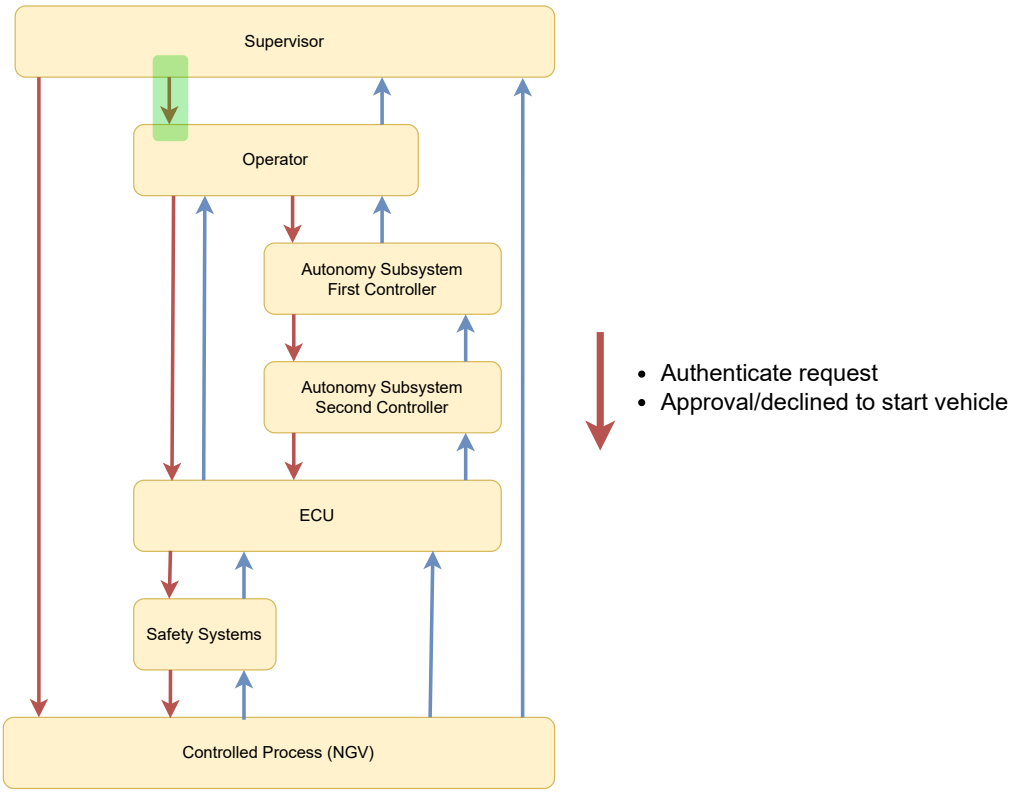


Figure 2.13: Control Action between Supervisor and Operator.

Table 2.4: UCA 1 - UCA 7.

Control Action	**Not providing causes hazard**	**Providing causes hazard too early, too late, out of order**	**Stopped too soon, applied too long**
Authenticate request	UCA-1 if the supervisor fails to send an authenticate request to the operator, it results in the absence of verification by a higher authority or someone in a superior hierarchy that allows the designated operator to drive the vehicle [H-7]	UCA-2 if the authentication fails, the operator cannot drive the vehicle since it is neccecary for the authentication to pass [H-7]	UCA-3 if the authenticate request occurs too quickly, the operator might not have enough time to respond [H-7][H-10]
Approval/declined to start vehicle	UCA-4 if the supervisor fails to provide an approval or decline to the operator, the authorization process cannot be completed [H-7]	UCA-5 if the approval or decline doesn't occur, the operator's access to control the vehicle remains restricted [H-7] UCA-6 if it grants approval without proper authentication, there's no higher authority authenticating the operator's driving access [H-7]	UCA-7 if the signal occurs too swiftly on the UI, the operator might miss the indication that they shouldn't proceed, leading to continuous attempts to seek authorization despite potential errors [H-10]

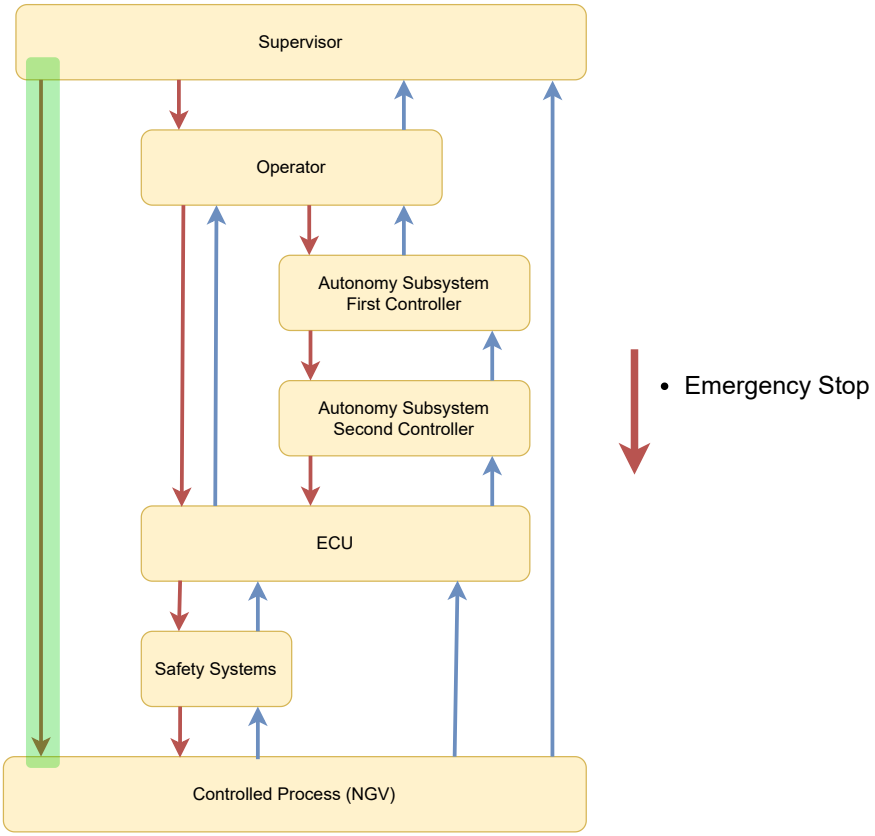


Figure 2.14: Control Action between Supervisor and NGV.

Table 2.5: UCA 8 - UCA 11.

Control Action	**Not providing causes hazard**	**Providing causes hazard too early, too late, out of order**	**Stopped too soon, applied too long**
Emergency Stop	UCA-8 supervisor cannot halt the vehicle and thus holds no ultimate authority over the system [H-4][H-7]	UCA-9 an untimely emergency stop or an erroneous signal triggers an abrupt vehicle halt when unnecessary [H-1] UCA-10 if the vehicle stops late, it risks allowing potential intruders in the system to cause further damage before halting [H-4][H-7]	UCA-11 if the supervisor accidentally triggers an emergency stop, recovering from that state becomes time-consuming and complicated, making it challenging to resume vehicle operation [H-1]

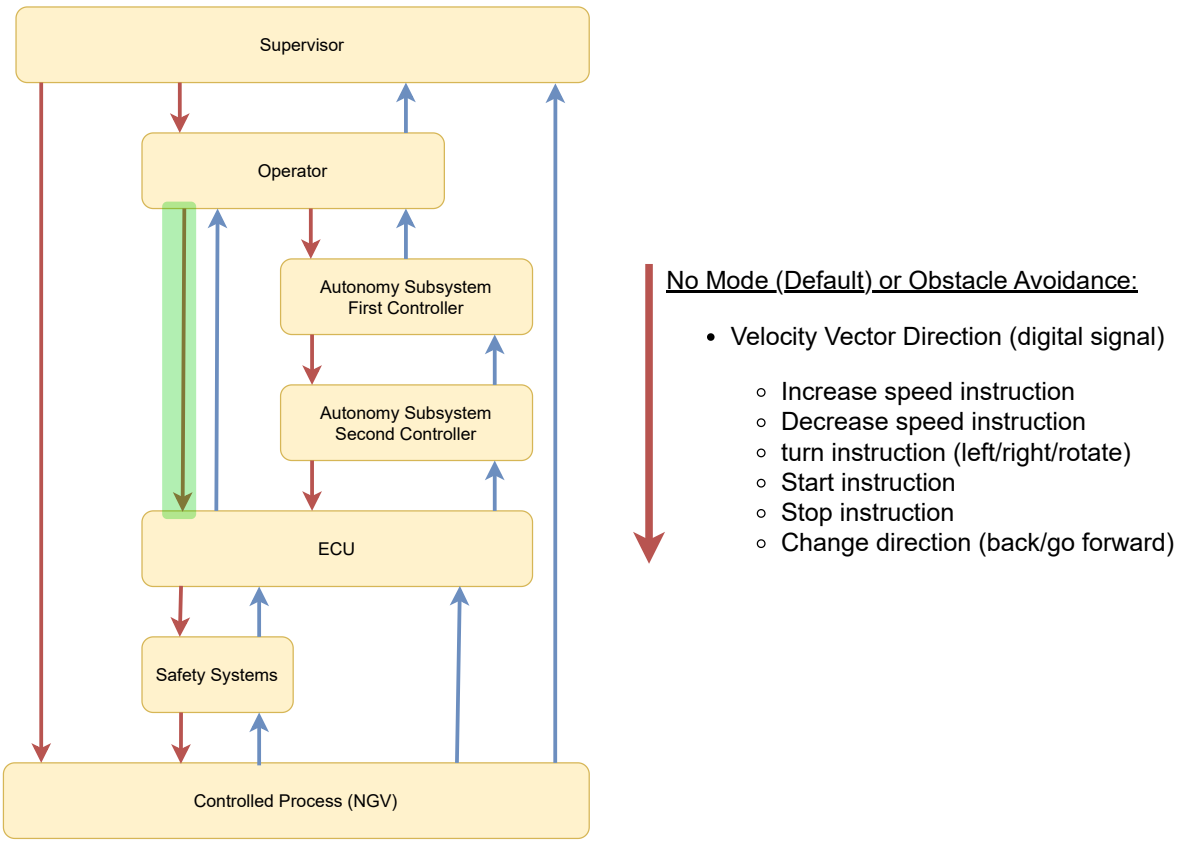


Figure 2.15: Control Action between Operator and ECU.

Table 2.6: UCA 12 - UCA 14.

Control Action	**Not providing causes hazard**	**Providing causes hazard too early, too late, out of order**	**Stopped too soon, applied too long**
Velocity Vector Direction	UCA-12 if the velocity vector direction is not transmitted, no instructions are relayed to the ECU regarding the vehicle's movement, resulting in its non-operation [H-1]	UCA-13 if the signal is sent too early or too late, it leads to incorrect driving behavior, causing the vehicle to behave erratically [H-1]	UCA-14 if the signals aren't read in time or remain applied for too long, it results in an unintended, uncontrolled operation [H-1]

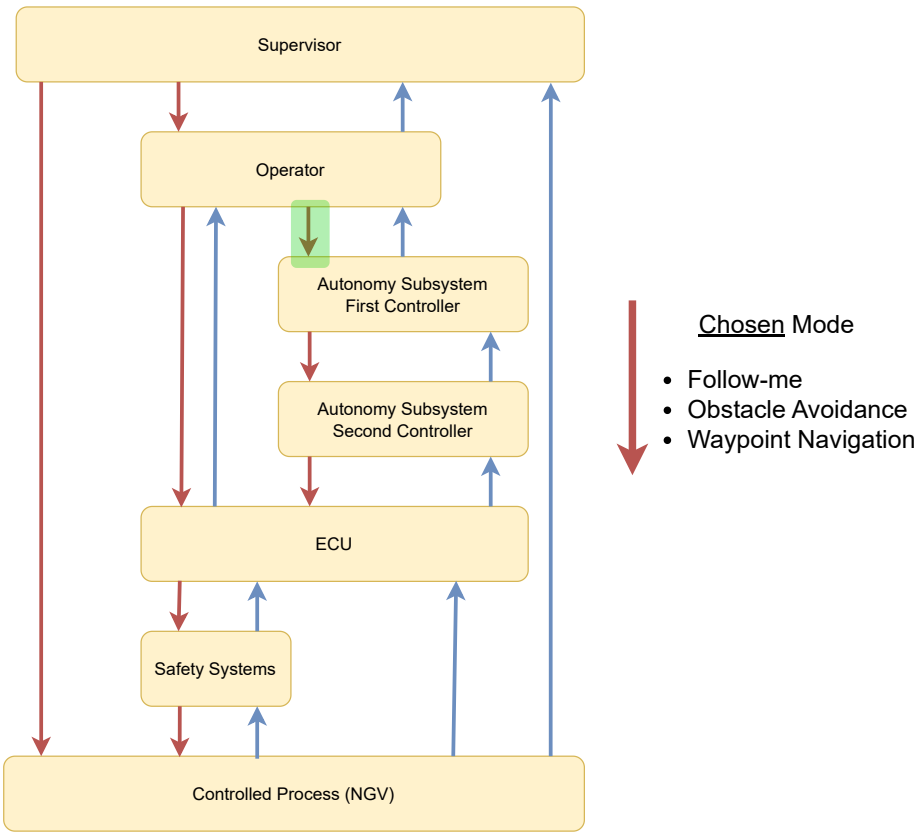


Figure 2.16: Control Action between Operator and Autonomy Subsystem First Controller.

Table 2.7: UCA 15 - UCA 18.

Control Action	**Not providing causes hazard**	**Providing causes hazard too early, too late, out of order**	**Stopped too soon, applied too long**
Chosen Mode	UCA-15 if the autonomous subsystem doesn't receive information about the specific autonomous mode to be engaged, it cannot operate autonomously [H-1]	UCA-16 if the information regarding which mode to operate arrives at an unexpected time, it may result in the vehicle operating uncontrollably [H-1] UCA-17 an operator might believe they're still controlling the vehicle, unaware that autonomy has taken over [H-1]	UCA-18 if the autonomous mode ends prematurely and the system anticipates operator intervention, it could result in an unintended halt or uncontrolled movement of the vehicle [H-1][H-10]

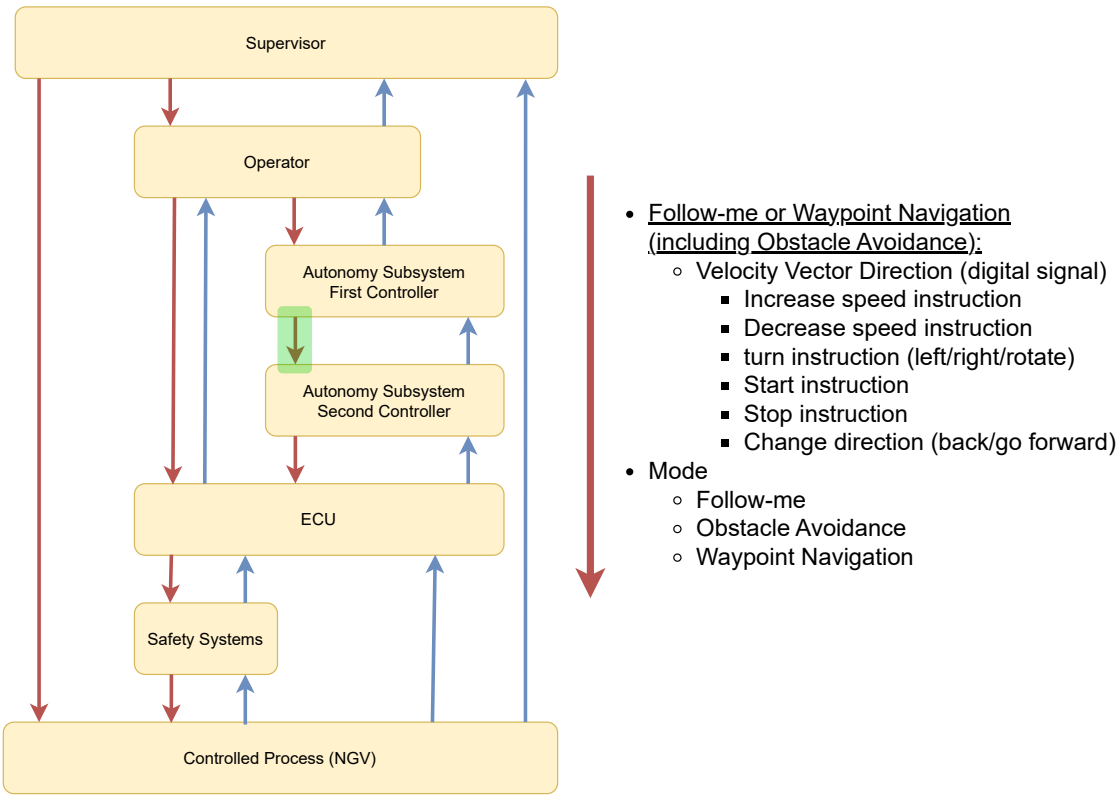


Figure 2.17: Control Action between Autonomy Subsystem First Controller and Autonomy Subsystem Second Controller.

Table 2.8: UCA 13 - UCA 14 & UCA 19 - UCA 22.

Control Action	**Not providing causes hazard**	**Providing causes hazard too early, too late, out of order**	**Stopped too soon, applied too long**
Mode	UCA-19 if the Autonomy Subsystem Second Controller doesn't receive information about the running mode from the Autonomy Subsystem First Controller, the vehicle cannot operate. This crucial information is necessary for making informed decisions, causing the vehicle to remain stationary [H-1]	UCA-20 if the signal arrives too late or too early, this could result in undesired behavior [H-1]	UCA-21 if the signal is applied for too long, it also results in undesired behavior and potential internal conflicts within the ECU. It might receive instructions both from the Autonomy Subsystem to perform a certain action and from the operator, creating conflicting directives [H-1][H-5]
Velocity Vector Direction	UCA-22 if no instructions are sent regarding how the vehicle should operate, it cannot move or make autonomous decisions [H-1]	UCA-13 if the signal is sent too early or too late, it leads to incorrect driving behav- ior, causing the vehicle to behave erratically [H-1]	UCA-14 if the signals aren't read in time or remain ap- plied for too long, it results in an unintended, uncontrolled operation [H-1]

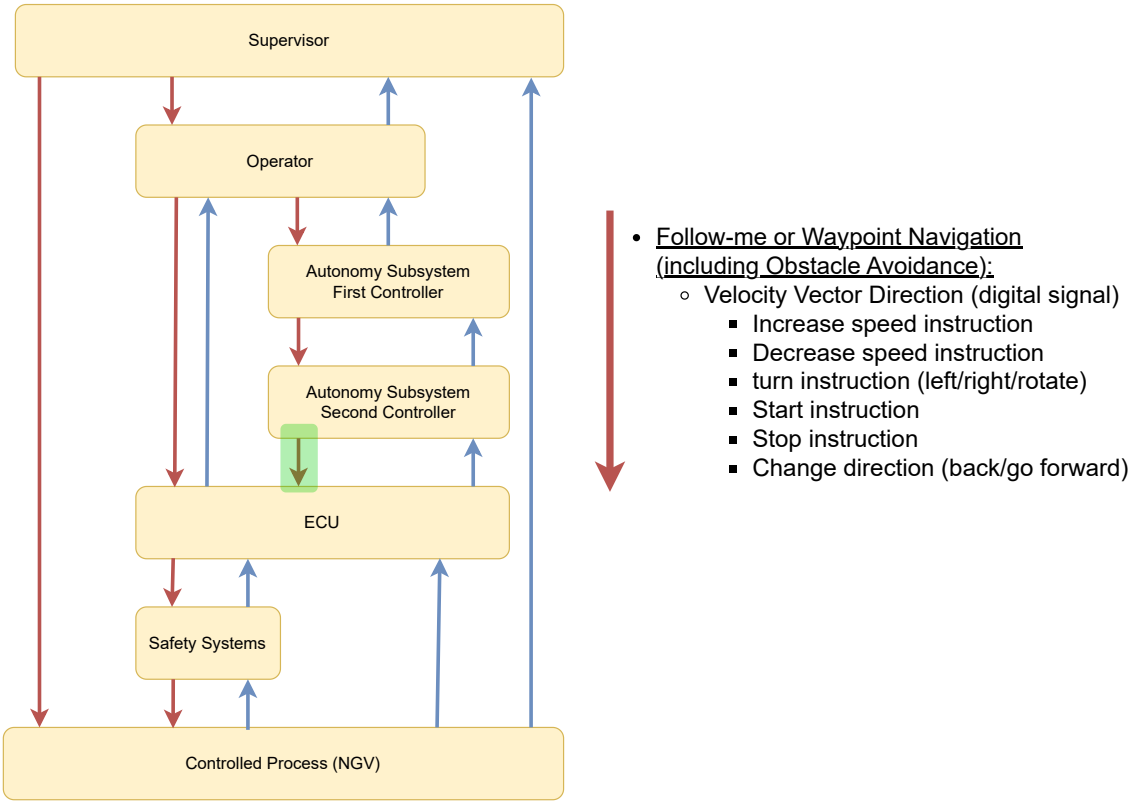


Figure 2.18: Control Action between Autonomy Subsystem Second Controller and ECU.

Table 2.9: UCA 12 - UCA 14.

Control Action	**Not providing causes hazard**	**Providing causes hazard too early, too late, out of order**	**Stopped too soon, applied too long**
Velocity Vector Direction	UCA-12 if the velocity vector direction is not transmitted, no instructions are relayed to the ECU regarding the vehicle's movement, resulting in its non-operation [H-1]	UCA-13 if the signal is sent too early or too late, it leads to incorrect driving behavior, causing the vehicle to behave erratically [H-1]	UCA-14 if the signals aren't read in time or remain applied for too long, it results in an unintended, uncontrolled operation [H-1]

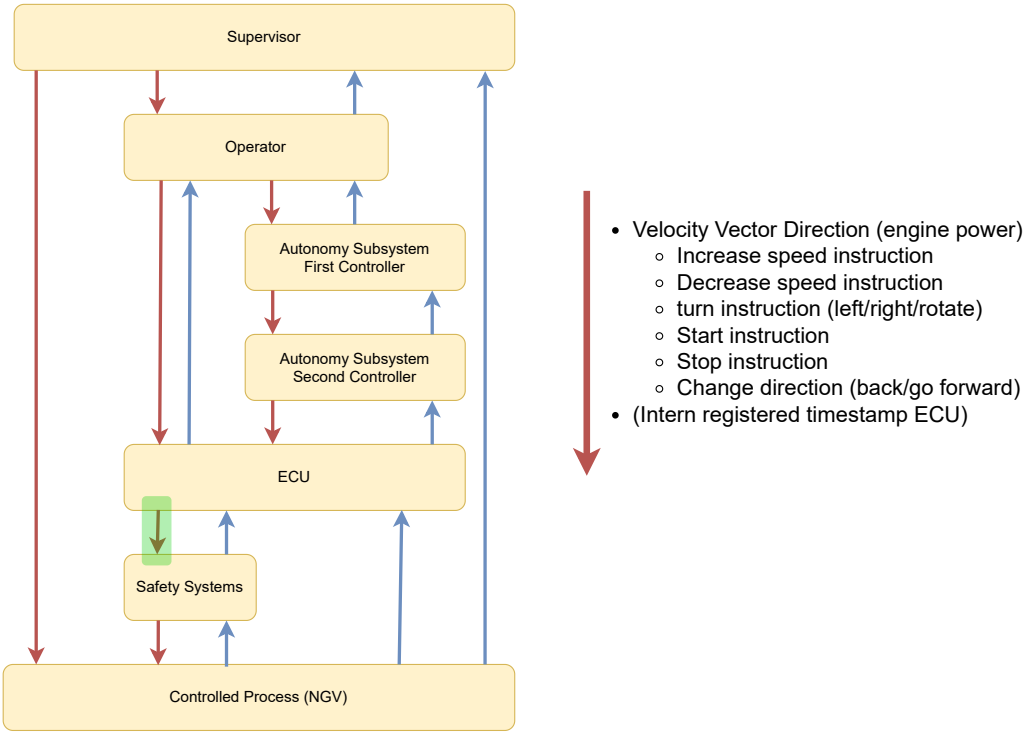


Figure 2.19: Control Action between ECU and Safety Systems.

Table 2.10: UCA 13 - UCA 14 & UCA 23 - UCA 27.

Control Action	**Not providing causes hazard**	**Providing causes hazard too early, too late, out of order**	**Stopped too soon, applied too long**
Velocity Vector Direction	UCA-23 if the safety system doesn't receive instructions from the ECU about how the vehicle should operate, it can't proceed because the safety system is responsible for transmitting instructions on how the vehicle should function [H-1]	UCA-24 erroneous instructions regarding the vehicle's operation, whether delayed or premature, may lead the safety system to make inaccurate decisions and impose incorrect limitations on the vehicle's operation [H-1][H-4] UCA-13 if the signal is sent too early or too late, it leads to incorrect driving behavior, causing the vehicle to behave erratically [H-1]	UCA-25 if the signals determining the vehicle's movement duration are either too long or too short, this could also prompt unnecessary safety measures or limitations that should not have been imposed [H-1][H-4] UCA-14 if the signals aren't read in time or remain applied for too long, it results in an unintended, uncontrolled operation [H-1]
Intern registered timestamp	UCA-26 without a registered timestamp here, it's impossible to measure the latency from when an instruction is sent from the ECU to when it's executed by the vehicle [H-1][H-5]	UCA-27 if the timestamp occurs too early or too late, it results in an incorrect latency measurement, leading to unnecessary alarms or failing to trigger alarms when necessary [H-1][H-5]	N/A

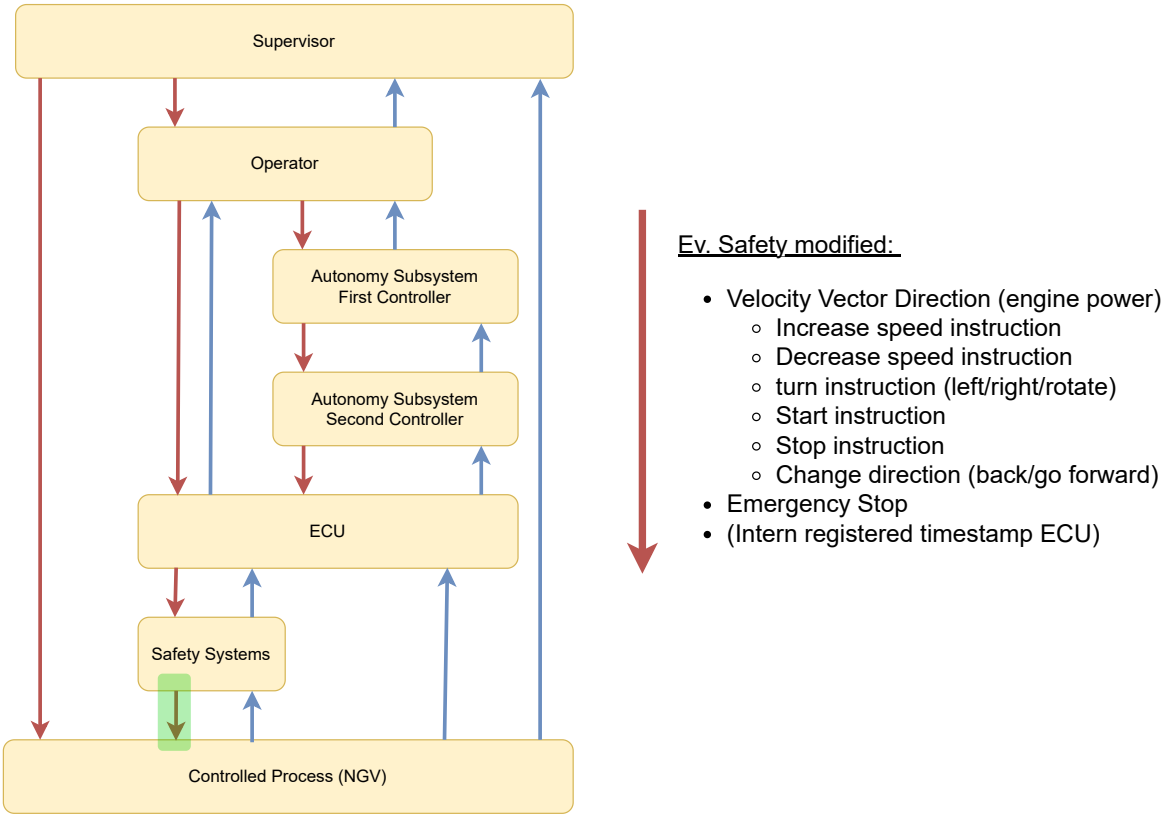


Figure 2.20: Control Action between Safety Systems and NGV.

Table 2.11: UCA 13 - UCA 14 & UCA 23 & UCA 27 - UCA 35.

Control Action	**Not providing causes hazard**	**Providing causes hazard too early, too late, out of order**	**Stopped too soon, applied too long**
Ev. Safety modified Velocity Vector Direction	UCA-28 if no information regarding how the vehicle should operate is sent from the safety systems, the vehicle won't be able to operate [H-1]	UCA-29 if an incorrect signal is sent or if it fails, the vehicle might not maintain the safe speed regulated by the safety systems [H-2] UCA-30 if there's an error in signaling or a failure, the vehicle might not maintain a safe distance from obstacles, a parameter also regulated by the safety systems [H-3] UCA-31 if the signals from the safety systems are incorrect or if the appropriate regulations haven't been implemented, the vehicle might fail to maintain structural integrity while driving in rough terrain [H-6] UCA-32 if the correct regulations aren't communicated by the velocity vector from the safety systems to the vehicle, it may fail to maintain a safe distance from hazardous terrain [H-8] UCA-33 if the safety systems fail to accurately identify soft obstacles in terrain, such as tall grass or dense undergrowths, there's a risk that the vehicle might not identify these obstacles accurately [H-9] UCA-13 if the signal is sent too early or too late, it leads to incorrect driving behavior, causing the vehicle to behave erratically [H-1]	UCA-14 if the signals aren't read in time or remain applied for too long, it results in an unintended, uncontrolled operation [H-1]

Emergency Stop	UCA-34 if the information regarding an emergency stop fails to transmit from the safety systems to the vehicle, the safety systems cannot initiate an emergency halt in hazardous situations when it's necessary to stop [H-1][H-3][H-8]	UCA-34 if the emergency stop occurs too late, an accident such as a collision may have already happened. If the emergency stop function fails to work, the situation remains the same [H-1][H-3][H-8] UCA-35 if the emergency stop happens unexpectedly or at an inappropriate moment, it leads to the vehicle halting abruptly [H-1][H-3][H-8]	N/A
Intern registered timestamp	UCA-23 without a regis- tered timestamp here, it's impossible to measure the latency from when an in- struction is sent from the ECU to when it's exe- cuted by the vehicle [H- 1][H-5]	UCA-27 if the timestamp occurs too early or too late, it results in an incorrect latency measurement, lead- ing to unnecessary alarms or failing to trigger alarms when necessary [H-1][H-5]	N/A

2.3.1 Defining Controller Constraints

The next step in this part of the analysis is to define Controller Constraints. These constraints specify the necessary behaviors for controllers to avoid UCAs. Once UCAs are identified, they can be translated into guidelines for each controller's behavior. For instance, each UCA can be reversed to establish constraints for individual controllers. Each UCA and its associated Controller Constraint are detailed in Table 2.12) below.

Table 2.12: Unsafe Control Actions and Controller Constraints.

UCA	**Unsafe Control Actions**	**Controller Constraints**
UCA-1	If the supervisor fails to send an authenticate request to the operator, it results in the absence of verification by a higher authority or someone in a superior hierarchy that allows the designated operator to drive the vehicle [H-7]	C-1: the supervisor must send an authenticate request to ensure verification by a higher authority or someone in a superior hierarchy that allows the designated operator to drive the vehicle [UCA-1]
UCA-2	If the authentication fails, the operator cannot drive the vehicle since it is necessary for the authentication to pass [H-7]	C-2: the operator cannot drive the vehicle unless the authentication passes [UCA-2]
UCA-3	If the authenticate request occurs too quickly, the operator might not have enough time to respond [H-7][H-10]	C-3: the authenticate request should allow enough time for the operator to respond [UCA-3]
UCA-4	If the supervisor fails to provide an approval or decline to the operator, the authorization process cannot be completed [H-7]	C-4: the supervisor must be able to provide approval or decline to the operator to complete the authorization process [UCA-4]
UCA-5	If the approval or decline doesn't occur, the operator's access to control the vehicle remains restricted [H-7]	C-5: the signal for approval or decline must be present for the operator to gain access to control the vehicle [UCA-5]
UCA-6	If it grants approval without proper authentication, there's no higher authority authenticating the operator's driving access [H-7]	C-6: granting approval without proper authentication should not allow the operator's driving access [UCA-6]
UCA-7	If the signal occurs too swiftly on the UI, the operator might miss the indication that they shouldn't proceed, leading to continuous attempts to seek authorization despite potential errors [H-10]	C-7: the signal on the UI should not occur too swiftly, allowing the operator to proceed without indications, leading to continuous authorization attempts despite potential errors [UCA-7]
UCA-8	Supervisor cannot halt the vehicle and thus holds no ultimate authority over the system [H-4][H-7]	C-8: the supervisor must have the capability to halt the vehicle, establishing ultimate authority over the system [UCA-8]
UCA-9	An untimely emergency stop or an erroneous signal triggers an abrupt vehicle halt when unnecessary [H-1]	C-9: the emergency stop should only be triggered corecctly to avoid unnecessary abrupt halts [UCA-9]
UCA-10	If the vehicle stops late, it risks allowing potential intruders in the system to cause further damage before halting [H-4][H-7]	C-10: the vehicle should stop within an appropriate short timeframe to prevent potential intruders from causing further damage [UCA-10]

UCA-11	If the supervisor accidentally triggers an emergency stop, recovering from that state becomes time-consuming and complicated, making it challenging to resume vehicle operation [H-1]	C-11: it should be challenging to trigger the emergency stop accidentally [UCA-11]
UCA-12	If the velocity vector direction is not transmitted, no instructions are relayed to the ECU regarding the vehicle's movement, resulting in its non-operation [H-1]	C-12: the transmission of velocity vector direction must be ensured to relay accurate instructions to the ECU for vehicle operation [UCA-12]
UCA-13	If the signal is sent too early or too late, it leads to incorrect driving behavior, causing the vehicle to behave erratically [H-1]	C-13: the signal transmission timing should be precise to avoid incorrect driving behavior and erratic vehicle operation [UCA-13]
UCA-14	If the signals aren't read in time or remain applied for too long, it results in unintended, uncontrolled operation [H-1]	C-14: the system should promptly read signals and avoid extended application to prevent unintended, uncontrolled operation [UCA-14]
UCA-15	If the autonomous subsystem doesn't receive information about the specific autonomous mode to be engaged, it cannot operate autonomously [H-1]	C-15: the information has to be received for the autonomous subsystem to operate [UCA-15]
UCA-16	If the information regarding which mode to operate arrives at an unexpected time, it may result in the vehicle operating uncontrollably [H-1]	C-16: information must arrive at the expected time to ensure timely and expected mode operation, preventing uncontrollable vehicle operation [UCA-16]
UCA-17	An operator might believe they're still controlling the vehicle, unaware that autonomy has taken over [H-1][H-10]	C-17: clear indications and communication must be established to inform the operator when autonomy takes over control [UCA-17]
UCA-18	If the autonomous mode ends prematurely and the system anticipates operator intervention, it could result in an unintended halt or uncontrolled movement of the vehicle [H-1]	C-18: proper autonomous mode duration management is necessary to prevent unintended halts or uncontrolled vehicle movements [UCA-18]
UCA-19	If the Autonomy Subsystem Second Controller doesn't receive information about the running mode from the Autonomy Subsystem First Controller, the vehicle cannot operate. This crucial information is necessary for making informed decisions, causing the vehicle to remain stationary [H-1]	C-19: the second controller within the autonomy subsystem must receive running mode information about the running mode from the first controller for vehicle operation to be ensured [UCA-19]
UCA-20	If the signal arrives too late or too early, this could result in undesired behavior [H-1]	C-20: timely signal arrival must happen to prevent undesired vehicle behavior [UCA-20]

UCA-21	If the signal is applied for too long, it results in undesired behavior and potential internal conflicts within the ECU. It might receive instructions both from the Autonomy Subsystem to perform a certain action and from the operator, creating conflicting directives [H-1][H-5]	C-21: proper duration of signal application must be necessary to prevent conflicts within the ECU and undesired vehicle behavior [UCA-21]
UCA-22	If no instructions are sent regarding how the vehicle should operate, it cannot move or make autonomous decisions [H-1]	C-22: transmission of instructions regarding vehicle operation must happen for autonomous decision-making [UCA-22]
UCA-23	If the safety system doesn't receive instructions from the ECU about how the vehicle should operate, it can't proceed because the safety system is responsible for transmitting instructions on how the vehicle should function [H-1]	C-23: transmission of operational instructions to the safety system from the ECU must happen for vehicle operation [UCA-23]
UCA-24	Erroneous instructions regarding the vehicle's operation, whether delayed or premature, may lead the safety system to make inaccurate decisions and impose incorrect limitations on the vehicle's operation [H-1][H-4]	C-24: accurate timing of instructions to the safety system must happen to prevent incorrect vehicle operation [UCA-24]
UCA-25	If the signals determining the vehicle's movement duration are either too long or too short, this could also prompt unnecessary safety measures or limitations that should not have been imposed [H-1][H-4]	C-25: proper duration of signals affecting vehicle movement must happen to prevent unnecessary safety measures or limitations [UCA-25]
UCA-26	Without a registered timestamp here, it's impossible to measure the latency from when an instruction is sent from the ECU to when it's executed by the vehicle [H-1][H-5]	C-26: accurate timestamping must happen for measuring latency between ECU instructions and vehicle execution [UCA-26]
UCA-27	If the timestamp occurs too early or too late, it results in an incorrect latency measurement, leading to unnecessary alarms or failing to trigger alarms when necessary [H-1][H-5]	C-27: timely timestamping must happen for accurate latency measurement and alarm triggering [UCA-27]
UCA-28	If no information regarding how the vehicle should operate is sent from the safety systems, the vehicle won't be able to operate [H-1]	C-28: the information regarding how the vehicle should operate must be sent from the safety systems [UCA-28]
UCA-29	If an incorrect sig nal is sent or if it fails, the vehicle might not maintain the safe speed regulated by the safety systems [H-2]	C-29: correct signal transmission must be ensured for maintaining safe speed [UCA-29]

UCA-30	If there's an error in signaling or a fail ure, the vehicle might not maintain a safe distance from obstacles, a parame- ter also regulated by the safety systems [H-3]	C-30: error-free signaling must be ensured for maintaining safe distance from obstacles [UCA-30]
UCA-31	If the signals from the safety systems are incorrect or if the appro priate reg- ulations haven't been implemented, the ve- hicle might fail to maintain structural integrity while driving in rough terrain [H-6]	C-31: proper safety signal must be implemented for structural integrity [UCA-31]
UCA-32	If the correct regulations aren't com municated by the velocity vector from the safety systems to the vehicle, it may fail to maintain a safe distance from hazardous terrain [H-8]	C-32: accurate communication of safety regulations crucial for hazardous terrain safety [UCA-32]
UCA-33	If the safety systems fail to accurately identify soft obstacles in terrain, such as tall grass or dense undergrowths, there's a risk that the vehicle might not identify these obstacles accurately [H-9]	C-33: accurate identification of soft obstacles must be ensured for obstacle response [UCA-33]
UCA-34	If the emergency stop occurs too late, an accident such as a collision may have already happened. If the emer- gency stop function fails to work, same [H-1][H-3][H-8]	C-34: appropriate timing of emergency stop to avoid collisions must be ensured [UCA-34]
UCA-35	If the emergency stop happens unex- pectedly or at an inappropriate mo- ment, it leads to the vehicle halting abruptly [H-1][H-3][H-8]	C-36: appropriate timing of emergency stop to avoid abrupt halts myst ve ensyred [UCA-35]

2.3.2 Further explanation of the Unsafe Control Actions (UCAs)

The process of identifying UCAs involved analyzing all control-action arrows to detect potential risks. It was relatively straightforward to draw conclusions about potential risks for each arrow. However, at this stage of analysis, it is clear that the STPA method delves deeper, prompting a more profound examination and contemplation of risks. Despite limited knowledge about the system, this approach has in this step revealed 35 fairly detailed potential specific unsafe control actions simply by following the analysis steps! A discernible pattern emerges where the closer we get to the safety system, the more UCAs associated with multiple system hazards appear. This pattern is logical as the control structure was constructed following all System Constraints, many of which required a safety system. The construction of the control structure being analyzed can be observed in Section 2.

2.4 Identify loss scenarios

The analysis has reached its final step which is identifying loss scenarios.

Definition: A loss scenario describes the causal factors leading to unsafe control actions and hazards. Two types of loss scenarios require consideration.

The analysis now progresses to its last phase, aimed at finding loss scenarios from the control structure's distinct loops and components see Figure 2.19. These scenarios explicitly outline the factors culminating in unsafe control actions and potential hazards.

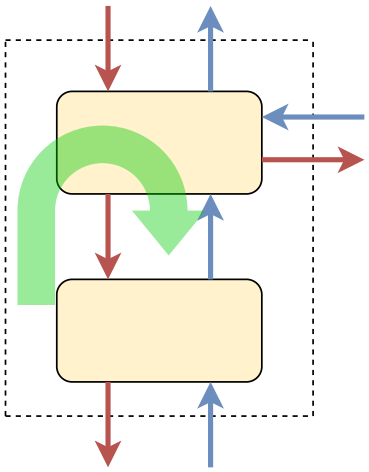


Figure 2.21: Identify Loss Scenarios.

Returning to previous steps involves a detailed examination of feedback loops within the control structure. This critical analysis aims to reveal potentially hazardous scenarios resulting from the systemic outcomes.

This final phase specifically focuses on delineating precise scenarios triggering control actions and delineating reasons behind potential execution failures or non-execution, ultimately leading to hazards. It is centered around extracting highly specific scenarios stemming from instances of Unsafe Control Actions (UCAs). Which can be illustrated in Figure 2.21.

This will be accomplished by examining each specific UCA identified in an earlier step and endeavoring to derive as many specific scenarios as possible from these instances.

UCA-1	If the supervisor fails to send an authenticate request to the operator, it results in the absence of verification by a higher authority or someone in a superior hierarchy that allows the designated operator to drive the vehicle [H-7]

- **Scenario 1.1:** An unwanted outsider gains control of the car and attempts to run over or harm the people around [H-7].
- **Scenario 1.2:** An unwanted outsider gains access to the vehicle's controls and potentially sensitive data about its destination [H-7].
- **Scenario 1.3:** An unauthorized individual hacks into the vehicle's control system, manipulating its route or inducing uncontrollable movements, endangering the vehicle's safe operation [H-7].
- **Scenario 1.4:** An unwanted outsider gains remote access, manipulates navigation settings, and directs the vehicle to an unintended location [H-7].

UCA-2	If the authentication fails, the operator cannot drive the vehicle since it is necessary for the authentication to pass [H-7]

- **Scenario 2.1:** System bug prevents operator startup due to authentication failure [H-7].

UCA-3	If the authenticate request occurs too quickly, the operator might not have enough time to respond [H-7][H-10]

- **Scenario 3.1:** In a critical emergency, an too quickly authentication request restricts the operator's ability to swiftly initiate the vehicle for urgent evacuation or assistance [H-7][H-10].
- **Scenario 3.2:** During heavy load transfer to the troops, rapid authentication procedures prevent the operator from promptly starting the vehicle, delaying essential supplies or equipment deployment [H-7][H-10].

UCA-4	If the supervisor fails to provide an approval or decline to the operator, the authorization process cannot be completed [H-7]

- **Scenario 4.1:** Communication glitch prevents supervisor approval, immobilizing the vehicle [H-7].
- **Scenario 4.2:** Software bug blocks supervisor's authorization, stopping vehicle operations [H-7].
- **Scenario 4.3:** The operator attempts to start the vehicle to escape a hazardous situation or deliver urgent supplies. However, the operator cannot initiate the vehicle because verification procedures fail to authenticate, leading to a critical delay in response or delivery [H-7].

UCA-5	If the approval or decline doesn't occur, the operator's access to control the vehicle remains restricted [H-7]

- **Scenario 5.1:** The operator is unable to start the vehicle and initiate its mission or escape from a hazardous situation [H-7].

UCA-6	If it grants approval without proper authentication, there's no higher authority authenticating the operator's driving access [H-7]

- **Scenario 6.1:** Improper authentication grants approval without higher authority verification, allowing unauthorized driving access [H-7].
- **Scenario 6.2:** System glitch mistakenly authenticates without proper protocol, providing unauthorized driving access [H-7].
- **Scenario 6.3:** An unauthorized individual gains access to the vehicle's controls and assumes command [H-7].

UCA-7	If the signal occurs too swiftly on the UI, the operator might miss the indication that they shouldn't proceed, leading to continuous attempts to seek authorization despite potential errors [H-10]

- **Scenario 7.1:** Swift signal on the UI causes operator to miss indications, leading to repeated attempts despite potential errors [H-10].
- **Scenario 7.2:** UI displays an unclear indication, prompting continuous authorization attempts due to operator confusion [H-10].

- **Scenario 7.3:** The mission initiation is stalled as the operator misinterprets UI information, leading to a failure in troubleshooting before commencement [H-10].

UCA-8	Supervisor cannot halt the vehicle and thus holds no ultimate authority over the system [H-4][H-7]

- **Scenario 8.1:** Supervisor's inability to halt the vehicle prevents ultimate authority over the system, leading to potential risks [H-4][H-7].
- **Scenario 8.2:** System malfunction prevents supervisor's intervention, resulting in a lack of control authority [H-4][H-7].
- **Scenario 8.3:** An unforeseen event occurs, and the autonomous system lacks specific guidelines for decision-making. Consequently, the system behaves dangerously and unpredictably, and cannot be halted [H-4][H-7].

UCA-9	An untimely emergency stop or an erroneous signal triggers an abrupt vehicle halt when unnecessary [H-1]

- **Scenario 9.1:** Erroneous signal triggers abrupt vehicle halt during normal operation, causing inconvenience [H-1].
- **Scenario 9.2:** System glitch misinterprets normal operation as an emergency, initiating an unnecessary halt [H-1].
- **Scenario 9.3:** The vehicle abruptly stops, causing damage to crucial, life-sustaining supplies needed by the troops [H-1].
- **Scenario 9.4:** During an urgent evacuation from a hazardous area, the vehicle faces unexpected delays due to an emergency stop, impeding its swift departure [H-1].

UCA-10	If the vehicle stops late, it risks allowing potential intruders in the system to cause further damage before halting [H-4][H-7]

- **Scenario 10.1:** An unauthorized individual has in some way system access and attempts to harm or run over nearby individuals before the stop signal is completely executed [H-4][H-7].

UCA-11	If the supervisor accidentally triggers an emergency stop, recovering from that state becomes time-consuming and complicated, making it challenging to resume vehicle operation [H-1]

- **Scenario 11.1:** Supervisor's accidental intervention halts the vehicle, leading to operational complications [H-1].
- **Scenario 11.2:** In a time-critical mission, an unexpected emergency stop occurs, and the recovery process takes too long, impacting the mission's deadline [H-1].
- **Scenario 11.3:** During an urgent evacuation from a hazardous area, the vehicle faces unexpected delays due to an emergency stop, impeding its swift departure [H-1].

UCA-12	If the velocity vector direction is not transmitted, no instructions are relayed to the ECU regarding the vehicle's movement, resulting in its non-operation [H-1]

- **Scenario 12.1:** Failure to transmit velocity vector direction leads to vehicle non-operation, causing standstill [H-1].
- **Scenario 12.2:** In a time-sensitive mission, the vehicle fails to receive crucial instructions, resulting in a delay in its departure, potentially impacting the timely delivery of essential materials [H-1].
- **Scenario 12.3:** In an urgent evacuation scenario from a hazardous area, the vehicle faces delays due to the ECU not receiving necessary instructions, hindering its swift departure [H-1].

UCA-13	If the signal is sent too early or too late, it leads to incorrect driving behavior, causing the vehicle to behave erratically [H-1]

- **Scenario 13.1:** Signal sent too early or too late results in incorrect driving behavior, causing erratic vehicle performance [H-1].
- **Scenario 13.2:** Timing error in signal transmission leads to incorrect vehicle behavior, impacting safety [H-1].
- **Scenario 13.3:** The vehicle operates uncontrollably, colliding with a tree, resulting in substantial damage to the vehicle's structure [H-1].
- **Scenario 13.4:** The vehicle exhibits erratic driving behavior caused by signal inaccuracies, leading to incorrect or unexpected movements, ultimately resulting in a collision with a person [H-1].
- **Scenario 13.5:** The vehicle encounters difficulty exiting a dangerous zone as it operates uncontrollably, rendering steering ineffective [H-1].

UCA-14	If the signals aren't read in time or remain applied for too long, it results in unintended, uncontrolled operation [H-1]

- **Scenario 14.1:** Signals not read in time or applied for too long result in unintended, uncontrolled operation, posing safety risks [H-1].
- **Scenario 14.2:** System failure to interpret signals leads to unanticipated, hazardous vehicle movements [H-1].
- **Scenario 14.3:** The vehicle, operating uncontrollably due to technical malfunctions or signal errors, collides with an obstacle or object [H-1].
- **Scenario 14.4:** The vehicle encounters difficulty exiting a dangerous zone as it operates uncontrollably, rendering steering ineffective [H-1].

UCA-15	If the autonomous subsystem doesn't receive information about the specific autonomous mode to be engaged, it cannot operate autonomously [H-1]

- **Scenario 15.1:** The vehicle remains stationary because the autonomous system fails to initiate without a specific mode selection, impeding the vehicle's operational start [H-1].

- **Scenario 15.2:** In an instance where the autonomous system malfunctions, forcing the operator to manually control the vehicle, the human limitations in observation could result in the vehicle colliding with obstacles or missing critical details [H-1].
- **Scenario 15.3:** In a scenario where the current team members accompanying the vehicle lack the time or capability to take control, attempting to initiate a "follow-me" mode becomes unfeasible, causing a delay in operation [H-1].
- **Scenario 15.4:** The vehicle is required to repeatedly traverse the same route, prompting the initiation of waypoint navigation mode. However, the system's failure to execute this function demands manual control, disrupting efficiency as the person operating manually could use their time more effectively elsewhere [H-1].
- **Scenario 15.5:** Manual control becomes necessary for the vehicle despite being less efficient, forcing an operator to take over [H-1].
- **Scenario 15.6:** The vehicle's inherent superiority lies in its integrated autonomy, granting it a significant edge over competitors. However, with the autonomous features malfunctioning, its advantage diminishes, making it less superior in comparison [H-1].

UCA-16	If the information regarding which mode to operate arrives at an unexpected time, it may result in the vehicle operating uncontrollably [H-1]

- **Scenario 16.1:** The unexpected activation of a mode confuses the supervisor, potentially leading to an trigger of the emergency stop due to confusion or misunderstanding [H-1].
- **Scenario 16.2:** Abrupt reception of mode operation instructions leads to erratic vehicle performance [H-1].
- **Scenario 16.3:** When a specific autonomous mode, such as waypoint navigation, is initiated too late, the vehicle follows an outdated route, causing inconvenience as the intended path is no longer relevant [H-1].
- **Scenario 16.4:** The vehicle's erratic behavior may lead the operator or supervisor to suspect a cyberattack or unauthorized access, causing them to hesitate or take immediate actions [H-7].

UCA-17	An operator might believe they're still controlling the vehicle, unaware that autonomy has taken over [H-1][H-10]

- **Scenario 17.1:** Operator remains unaware of autonomous takeover, assuming continued control [H-1][H-10].
- **Scenario 17.2:** Lack of clarity in autonomous operation indication leads to operator confusion about vehicle control [H-1][H-10].
- **Scenario 17.3:** The vehicle's erratic behavior may lead the operator or supervisor to suspect a cyberattack or unauthorized access, causing them to hesitate or take immediate actions [H-7].

UCA-18	If the autonomous mode ends prematurely and the system anticipates operator intervention, it could result in an unintended halt or uncontrolled movement of the vehicle [H-1]

- **Scenario 18.1:** Autonomous mode ends prematurely, anticipating operator intervention, leading to unintended halts or uncontrolled vehicle movement [H-1].
- **Scenario 18.2:** Premature termination of autonomous mode initiates abrupt vehicle halts or erratic movements due to expected operator intervention [H-1].
- **Scenario 18.3:** The vehicle abruptly stops, causing damage to crucial, life-sustaining supplies needed by the troops [H-1].

UCA-19	If the Autonomy Subsystem Second Controller doesn't receive information about the running mode from the Autonomy Subsystem First Controller, the vehicle cannot operate. This crucial information is necessary for making informed decisions, causing the vehicle to remain stationary [H-1]

- **Scenario 19.1:** Lack of communication between Autonomy Subsystem controllers results in immobilization due to missing mode information [H-1].
- **Scenario 19.2:** In an instance where the autonomous system malfunctions, forcing the operator to manually control the vehicle, the human limitations in observation could result in the vehicle colliding with obstacles or missing critical details [H-1].
- **Scenario 19.3:** In a scenario where the current team members accompanying the vehicle lack the time or capability to take control, attempting to initiate a "follow-me" mode becomes unfeasible, causing a delay in operation [H-1].
- **Scenario 19.4:** The vehicle is required to repeatedly traverse the same route, prompting the initiation of waypoint navigation mode. However, the system's failure to execute this function demands manual control, disrupting efficiency as the person operating manually could use their time more effectively elsewhere [H-1].
- **Scenario 19.5:** Manual control becomes necessary for the vehicle despite being less efficient, forcing an operator to take over [H-1].
- **Scenario 19.6:** The vehicle's inherent superiority lies in its integrated autonomy, granting it a significant edge over competitors. However, with the autonomous features malfunctioning, its advantage diminishes, making it less superior in comparison [H-1].

UCA-20	If the signal arrives too late or too early, this could result in undesired behavior [H-1]

- **Scenario 20.1:** Signal arrives too early or too late, causing undesired vehicle behavior [H-1].
- **Scenario 20.2:** Incorrect timing in signal transmission prompts unexpected vehicle behavior [H-1].
- **Scenario 20.3:** The autonomous system makes ambiguous decisions, causing confusion or uncertainty for the operator or supervisor regarding the vehicle's actions [H-1].
- **Scenario 20.4:** Uncontrolled vehicle movement leads to a collision, resulting in damage to the vehicle or collision with a person, causing potential harm or injury [H-1].
- **Scenario 20.5:** The vehicle's safety system triggers an internal alert to the operator indicating a discrepancy or irregularity, causing the operator to halt the vehicle to investigate the issue [H-7].

UCA-21	If the signal is applied for too long, it results in undesired behavior and potential internal conflicts within the ECU. It might receive instructions both from the Autonomy Subsystem to perform a certain action and from the operator, creating conflicting directives [H-1][H-5]

- **Scenario 21.1:** Internal conflicts within the ECU regarding whether the autonomous system or the operator is in control result in the vehicle abruptly stopping, interpreting the conflict as a malfunction, leading to an abrupt halt [H-1][H-5].
- **Scenario 21.2:** The vehicle's safety system triggers an internal alert to the operator indicating a discrepancy or irregularity, causing the operator to halt the vehicle to investigate the issue [H-7].

UCA-22	If no instructions are sent regarding how the vehicle should operate, it cannot move or make autonomous decisions [H-1]

- **Scenario 22.1:** Lack of operational instructions prevents vehicle movement and autonomous decision-making [H-1].
- **Scenario 22.2:** Missing operation guidelines inhibit vehicle mobility and autonomous functionality [H-1].
- **Scenario 22.3:** In critical situations where the ECU lacks instructions on vehicle control, the vehicle remains stationary, posing significant challenges if immediate relocation from a hazardous area is necessary [H-1].
- **Scenario 22.4:** An operator is compelled to take manual control of the vehicle due to the autonomous system's failure, especially in critical situations demanding immediate action [H-1].

UCA-23	If the safety system doesn't receive instructions from the ECU about how the vehicle should operate, it can't proceed because the safety system is responsible for transmitting instructions on how the vehicle should function [H-1]

- **Scenario 23.1:** Safety system lacking instructions from the ECU restricts vehicle operation, hindering functionality [H-1].
- **Scenario 23.2:** Missing guidance from the ECU impedes the safety system's role, leading to restricted vehicle operations [H-1].
- **Scenario 23.3:** The vehicle remains stationary and is unable to extricate itself from a hazardous situation or commence its designated mission [H-1].

UCA-24	Erroneous instructions regarding the vehicle's operation, whether delayed or premature, may lead the safety system to make inaccurate decisions and impose incorrect limitations on the vehicle's operation [H-1][H-4]

- **Scenario 24.1:** Erroneous instructions delay or arrive prematurely, causing the safety system to make inaccurate decisions and impose incorrect limitations on vehicle operation [H-1][H-4].
- **Scenario 24.2:** Timing errors in instructions lead to inaccurate decisions by the safety system, impacting vehicle operation [H-1][H-4].

- **Scenario 24.3:** The safety system receives false information indicating hazardous movement, triggering restrictive measures that limit the vehicle's speed unnecessarily during a critical situation [H-1][H-4].

UCA-25	If the signals determining the vehicle's movement duration are either too long or too short, this could also prompt unnecessary safety measures or limitations that should not have been imposed [H-1][H-4]

- **Scenario 25.1:** Signals determining vehicle movement duration are either too long or too short, prompting unnecessary safety measures or limitations [H-1][H-4].
- **Scenario 25.2:** Incorrect duration signals trigger improper safety measures, affecting vehicle operation [H-1][H-4].
- **Scenario 25.3:** The uncontrollable operation resulting from the signal leads the vehicle to collide with an obstacle or a person [H-1][H-4].

UCA-26	Without a registered timestamp here, it's impossible to measure the latency from when an instruction is sent from the ECU to when it's executed by the vehicle [H-1][H-5]

- **Scenario 26.1:** Lack of timestamp measurement impedes tracking latency between ECU instruction and vehicle execution, affecting system performance [H-1][H-5].
- **Scenario 26.2:** Absence of timestamp measurement complicates latency evaluation, impacting system efficiency [H-1][H-5].
- **Scenario 26.3:** In the absence of a registered timestamp, measuring the latency between issuing an instruction and its execution becomes impossible. This lack of timing data disorients the operator in controlling the vehicle, rendering it inoperable [H-1][H-5].

UCA-27	If the timestamp occurs too early or too late, it results in an incorrect latency measurement, leading to unnecessary alarms or failing to trigger alarms when necessary [H-1][H-5]

- **Scenario 27.1:** Incorrect timestamp timing causes faulty latency measurements, resulting in unnecessary alarms or failure to trigger alarms when needed [H-1][H-5].
- **Scenario 27.2:** Timing errors in timestamp measurement lead to inaccurate latency assessments, impacting alarm systems [H-1][H-5].
- **Scenario 27.3:** Misleading latency information provided to the operator leads to a misunderstanding of the vehicle's behavior. This misunderstanding results in an increased risk of collision due to incorrect perception and control lag [H-1][H-5].

UCA-28	If no information regarding how the vehicle should operate is sent from the safety systems, the vehicle won't be able to operate [H-1]

- **Scenario 28.1:** Lack of operation instructions from safety systems halts vehicle operation rapidly, causing damage to equipment or cargo on the vehicle [H-1].

- **Scenario 28.2:** The vehicle attempts to navigate away from a hazardous situation, whether controlled by an operator or the autonomous system. However, due to the safety systems's inability to provide instructions, it remains unable to escape, resulting in potential harm/destroyment to the vehicle[H-1].

UCA-29	If an incorrect signal is sent or if it fails, the vehicle might not maintain the safe speed regulated by the safety systems [H-2]

- **Scenario 29.1:** The vehicle exceeds the safe speed limit and fails to stop when a person suddenly appears in front, resulting in injury [H-2].
- **Scenario 29.2:** The vehicle fails to stop in time as an obstacle blocks its path, resulting in a collision and damage [H-2].
- **Scenario 29.3:** The vehicle rapidly moves away from the operator's control range, losing signal contact. The vehicle continues to drive away despite the operator's attempt to control it [H-2].

UCA-30	If there's an error in signaling or a failure, the vehicle might not maintain a safe distance from obstacles, a parameter also regulated by the safety systems [H-3]

- **Scenario 30.1:** The vehicle's autonomous mode fails to detect an obstacle that is too close, resulting in a collision damaging the vehicle [H-3].
- **Scenario 30.2:** The vehicle is in waypoint navigation mode and collides along the route due to a failure to execute a turning maneuver when an obstacle is encountered [H-3].

UCA-31	If the signals from the safety systems are incorrect or if the appropriate regulations haven't been implemented, the vehicle might fail to maintain structural integrity while driving in rough terrain [H-6]

- **Scenario 31.1:** The vehicle drives over rough terrain, lacking structural integrity due to erroneous signals, causing items critical for the troops or essential heavy equipment to be thrown off the vehicle [H-6].
- **Scenario 31.2:** The vehicle sustains damage due to uncontrolled bouncing [H-6].

UCA-32	If the correct regulations aren't com municated by the velocity vector from the safety systems to the vehicle, it may fail to maintain a safe distance from hazardous terrain [H-8]

- **Scenario 32.1:** The vehicle enters a hazardous terrain, such as thin ice, causing it to break through and submerge [H-8].
- **Scenario 32.2:** The vehicle becomes entangled among numerous trees, rendering it unable to maneuver out [H-8].

UCA-33	If the safety systems fail to accurately identify soft obstacles in terrain, such as tall grass or dense undergrowths, there's a risk that the vehicle might not identify these obstacles accurately [H-9]

- **Scenario 33.1:** The vehicle navigates through tall grass, encountering a large hidden stone. The collision with the stone results in damage to the vehicle [H-9].
- **Scenario 33.2:** While navigating through tall grass, the vehicle encounters an animal that gets injured in the collision [H-9].

UCA-34	If the emergency stop occurs too late, an accident such as a collision may have already happened. If the emergency stop function fails to work, same [H-1][H-3][H-8]

- **Scenario 34.1:** A disaster is imminent, and the supervisor initiates an emergency stop. However, due to the delayed signal, the accident has already occurred, resulting in the vehicle colliding with an obstacle or a person [H-1][H-3][H-8].

UCA-35	If the emergency stop happens unexpectedly or at an inappropriate moment, it leads to the vehicle halting abruptly [H-1][H-3][H-8]

- **Scenario 35.1:** An abrupt halt damages crucial equipment or items on the vehicle [H-1][H-3][H-8].
- **Scenario 35.2:** Restarting the vehicle takes time, becoming critical in situations requiring swift evacuation from a hazardous area [H-1][H-3][H-8].

2.4.1 Further Explanation on Loss Scenarios/General Conclusions

The requirements were organized into distinct sections, such as Sections 1.1, 1.2, etc., to enhance clarity. Some sections shared similarities or identical criteria but with minor differences. This arrangement simplifies the referencing of individual requirements while clearly associating them with their respective UCAs.

In certain cases, specific UCAs had identical requirements. Rather than repeatedly referring to the same requirement in some places and maintaining slightly varied versions in others, it was clearer to include all variations universally.

This structure also offers flexibility for adding or removing scenarios without necessitating alterations in multiple sections.

A comprehensive list of specific scenarios to avoid when dealing with an NGV is now available. The analysis's strength lies in deriving concrete information with minimal system knowledge. Each scenario addresses potential abstract faults within the control structure, aiding those formulating requirement specifications or initiating discussions on the system's design and related aspects.

3 Diverse Approaches and Analyses in Autonomous Vehicle Safety

This chapter aims to broaden the understanding of autonomous systems validation by exploring alternative methods and safety analyses. The objective is to offer comprehensive insights into the nuanced and effective validation of autonomous systems.

3.1 Relevant Methods and Standards in this Field

In the domain of autonomous systems safety analysis, several methods play a crucial role in ensuring the reliability and robustness of these complex systems. Some of the relevant methods include:

- **STPA:** The STPA method, as earlier mentioned, excels in providing a thorough understanding of system safety in autonomous vehicles. It offers actionable recommendations and is valuable in crafting system requirements. However, its effectiveness is context-dependent and influenced by operational data availability. A potential drawback is the interpretation of terms such as "simple," which may vary among analysts.

  Advantages:

  - Systematic analysis: STPA applies a systemic approach, focusing on analyzing the entire system, contributing to a more comprehensive understanding of potential risks and errors.
  - Process focus: By emphasizing processes, STPA helps identify and analyze potential inaccuracies and risks in the system's functions and processes.
  - In-depth understanding: STPA is designed to provide a profound understanding of the system and its dynamics, enabling the identification of potential flaws in design and operation.

  Disadvantages:

  - Complexity: Implementing STPA can be complicated and may require time and expertise to conduct a thorough analysis.
  - Subjectivity: Depending on the implementer, the interpretation of results may be subjective, potentially affecting the quality of the analysis [2].

- **Six-Step Model:** A comprehensive framework designed to integrate safety and security measures within autonomous vehicles. It operates across six critical hierarchies, ensuring consistency throughout the vehicle's life cycle. Challenges include intricate early alignment between safety and security measures and practical implementation complexities.

  Advantages:

  - Clear hierarchies: The Six-Step Model provides clear hierarchies, including functions, structure, faults, attacks, security measures, and safety measures, facilitating organization and understanding.
  - Lifecycle management: The model is designed to ensure consistency throughout the lifecycle of autonomous vehicles, crucial for managing risks across different phases.

  Disadvantages:

  - Potential limitation: Depending on the complexity of the autonomous vehicle, the model may be limited and may not cover all aspects of risk analysis.
  - General nature: The model can be generic and needs customization to specific vehicle systems to be most effective [3].

- **Threat Analysis and Risk Assessment (TARA):** TARA is used in the context of cybersecurity to identify potential threats and assess associated risks. Crucial for developing secure autonomous vehicles [4].

- **Process Hazard Analysis (PHA) Methods:**

  Conventional Process Hazard Analysis (PHA) methods are systematic approaches used to assess potential hazards in industrial processes. They rely on breaking down systems into components for analysis, focusing on identifying vulnerabilities and failure modes [5].

  - **Fault Tree Analysis (FTA):** FTA evaluates the combination of events leading to a specific undesired event, often visualized in a tree-like structure. It assesses system safety and reliability by identifying possible failure paths and their probabilities [6].
  - **Failure Mode and Effects Analysis (FMEA):** A systematic method for evaluating processes to identify potential failures and assess their relative impact. Used to enhance system reliability in various industries, including automotive [7].

- **Automotive Safety Standards:**

  - **ISO 26262:** ISO 26262 is an international standard specifically focused on functional safety in the automotive industry. It encompasses safety requirements across a spectrum from ASIL Level 0 (no automation) to Level 5 (full automation). The standard offers a systematic approach to identifying and managing safety risks throughout the entire automotive development life cycle, emphasizing safety measures across various automation levels [8].
  - **SAE J3016:** In contrast, SAE J3016, developed by the Society of Automotive Engineers (SAE), is a standard that primarily defines and categorizes the levels of driving automation in vehicles. These levels range from Level 0 (no automation) to Level 5 (full automation). Each level signifies a different degree of automation, with Level 5 indicating a vehicle's capability to perform all driving tasks under all conditions without human intervention. While ISO 26262 focuses on functional

safety, SAE J3016 provides a framework for understanding and classifying the extent of automation in driving tasks [9].

- **SAE J3061:** This standard developed by SAE focuses on the cybersecurity aspects of vehicle systems, including autonomous vehicles. It provides guidelines and practices for implementing cybersecurity in the design, development, and production of vehicle systems. Given the increasing integration of software and connectivity in modern vehicles, addressing cybersecurity is crucial to ensure the safety and security of autonomous vehicles [10].

3.2 Research Related to this Area, Methods and Standards.

Below are three research reports that, in one way or another, security analyses within this field. The reports include comments and conclusions regarding the identification of various methods advantages and disadvantages, according to the authors. Additionally, emphasis is placed on patterns and factors that play a role in the context of security analyses conducted for autonomous vehicles.

Safety Analysis of Driver-Vehicle Interaction (STPA-inspired) by *Max Stoltz-Sundnes* at Kungliga Tekniska Högskolan (KTH), 2019

This study, conducted as a master's thesis at Kungliga Tekniska Högskolan (KTH) 2019, focuses on the safety analysis of cooperative driving and human-machine interaction (HMI) in autonomous vehicles. Applying STPA to a case study involving cooperative driving functionality, the research identifies system-level safety constraints. Special emphasis is placed on the development of HMI-related aspects, leading to the enhancement of the driver-vehicle interaction module. This module addresses visual aspects, introduces new driver-centric risks, and proposes a strategy for a secure transition between autonomous and manual states. The analysis, which compares STPA with a new method for safe mode switching, underscores the significance of addressing accidental or faulty inputs from the driver as a major threat to mode confusion. The study contributes valuable insights to improve safety measures in autonomous vehicles and enhance driver-vehicle interaction.

In the conclusion of the study, the author highlights gaps in current safety standards, specifically ISO26262, when it comes to addressing the challenges posed by autonomous and cooperative functionalities in the automotive industry. The need for an extension or replacement of existing standards is emphasized to accommodate these new vehicle types and their interaction with others [11].

The author argues that ISO26262 is a significant international standard, but it dates back to 2011, making it outdated given the substantial developments that have occurred since then. In this study, one also gains insights into the specific relationship between the driver and the vehicle. It is intriguing and advantageous to observe that STPA can indeed adapt to various domains.

Integrating Autonomous Vehicle Safety and Security by *Giedre Sabaliauskaite and Jin Cui* at the Centre for Research in Cyber Security, Singapore University of Technology and Design, 2017

This paper is published by the Centre for Research in Cyber Security, Singapore University of Technology and Design and was presented at the Second International Conference on Cyber-Technologies and Cyber-Systems 2017. It discusses the integration of safety and security measures for autonomous vehicles (AVs) uses the Six-Step Model. The authors emphasize

the interdependence of safety and security in protecting AVs from accidents and intentional attacks. They highlight the challenges posed by the varying levels of driving automation defined by SAE J3016 and the absence of specific international standards for AV safety and security.

The Six-Step Model, previously proposed by the authors for Cyber-Physical Systems (CPS) safety and security, is extended to address the unique challenges of AVs. The model encompasses six hierarchies: functions, structure, failures, attacks, safety countermeasures, and security countermeasures. It is designed to ensure consistency across these hierarchies throughout the entire life-cycle of an AV.

The analysis points out that current vehicle safety standards, such as ISO 26262, do not consider driving automation levels. To address this gap, the authors propose an approach compliant with SAE J3016, SAE J3061, and ISO 26262. They advocate for the Six-Step Model as a comprehensive tool to integrate and align safety and security processes and artifacts for AVs.

The authors discuss the importance of early alignment between safety and security in AV development phases to ensure the necessary level of protection. They acknowledge the relative newness of the AV domain and the absence of specific international standards for AV safety and security. The proposed approach aims to fill this gap and offers a consistent method for analyzing AV safety and security throughout its life-cycle.

The safety analysis involves hazard analysis based on operational situation and hazard trees, considering different driving automation levels. The security analysis utilizes the SAE J3061 standard, incorporating TARA through attack tree analysis. The Six-Step Model is used to interconnect these analyses and ensure alignment between safety and security countermeasures [12].

This paper underscores the strategic use of a combination of safety standards, acknowledging the limitations of certain prominent ones. With a specific focus on cybersecurity, methods like SAE J3061 and TARA are deemed appropriate. However, the primary emphasis remains on early-stage safety evaluations in autonomous vehicle development. While the Six-Step Model provides clear hierarchies and lifecycle management advantages, STPA could be considered for a more detailed analysis. Both approaches offer valuable insights, with the Six-Step Model standing out for its clarity and adaptability, and STPA for its depth in analysis.

Methods for Assessing the Safety of Autonomous Vehicles by *David Robert Beachum* at the University of Texas at Austin, 2019

In the comprehensive study conducted by David Robert Beachum at the University of Texas at Austin, the exploration of AV technology's potential to revolutionize transportation and save lives takes center stage. The author underscores the critical need for stringent AV safety verification, given the substantial risks accompanying the widespread adoption of these vehicles. While automotive manufacturers race to lead the market with AVs, the lagging regulatory frameworks underscore the importance of alternative safety assessment methods.

The study delves into five distinct approaches employed in AV safety assessment: real-world testing, simulation, Failure Modes and Effects Analysis (FMEA), Fault Tree Analysis (FTA), and STPA. To gauge their relative effectiveness in quantifying and assessing AV safety, the author applies the latter four methods to a hypothetical AV system executing an unprotected left turn.

Real-world testing, deemed the most straightforward method, faces hurdles due to the infrequency of fatal collisions in human driving, demanding impractical amounts of AV driving data for statistically significant safety proof. On the other hand, simulation expedites data collection but raises concerns about the representativeness of fully simulated environments and the absence of interactions with other human drivers.

Analytical methods, such as FMEA, are acknowledged for systematic risk identification and mitigation. However, FMEA's limitations emerge from its assumption that multiple failure modes do not occur concurrently, limiting its ability to represent the interdependent subsystems of AVs. FTA, while capable of comparing reliability and failure rates, relies on challenging data acquisition for every component, including machine learning algorithms.

STPA, distinguished for its adept handling of complex interactions in dynamic systems, lacks a quantitative method for comparing the reliability of two systems.

Addressing the recurring challenge of characterizing hazards associated with machine learning algorithms in AV controllers, the author suggests promising approaches involving FTA or STPA coupled with secondary analyses to quantify the risk of encountering unclassified scenarios. Despite this, the author acknowledges the nascent nature of safety assessments in machine learning contexts, necessitating further exploration.

In the absence of a comprehensive analytical approach, the author advocates for simulation using real-world data as the most effective current method for assessing AV safety. This approach strikes a balance between speed, safety, and representation when compared to real-world testing and full simulation.

The author concludes that while each of the five methods—real-world testing, simulation, FMEA, FTA, and STPA—brings distinct advantages and disadvantages, the current state favors simulation using real-world data as the most effective approach. The regulatory landscape for AV technology remains uncertain, and if the current period of federal leniency persists, data from publicly sold vehicles could potentially simplify safety assessments as AV adoption becomes widespread [13].

4 Results

The results of the STPA analysis are evident at every step. The final step, which synthesized all previous stages, generated a plethora of diverse scenarios that are to be avoided. It is crucial to define and consider these scenarios before embarking on the implementation of autonomy in a vehicle. The analysis yields a substantial amount of data, minimizing the risk of actualizing any undesirable outcomes. However, the analysis result extends beyond the final step. In the beginning, Table 2.3 delineated the system constraints (SC) that should be imposed on the system, which can be incorporated into the vehicles requirement specification (in the next stage). In the fourth step, it was also obtained that the Controller Constraints (C) can form the basis for subsequent implementation, see Table 2.12. Analyzing all UCA provided insights into control signal flow and potential pitfalls between various components.

The conceptualization of a product now involves numerous requirements and constraints to prevent risks. Addressing the first research question:

1. What risks can be concretely identified in the early stages of developing a self-driving vehicle?

Starts from the top down in the analysis. In Chapter 3, specifically in Section 3.1.2, Table 2.1 delineates the primary general risks that serve as the basis for the analysis. Subsequently, in Chapter 3 and Section 3.1.4, Table 2.2 delves into system-level risks, and further in Tables 2.4- 2.11, Chapter 3, Section 3.3, detailed risks concerning control signals are identified. In the last step all specific scenarios where presented, and these ones can be interpreted as the clearest "risks" or scenarios that effectively answers the question. These scenarios are detailed in Chapter 3.4 of the report.

To draw generalizations from these scenarios, cybersecurity emerged as a central focus as it was presented in Table 2.1, at the beginning of the analysis. The vehicular network must not be compromised, as the consequences could be catastrophic. Scenarios such as unwanted outsider control, unauthorized hacking, and remote manipulation of the vehicle's navigation settings highlight the critical nature of this aspect. Other common scenarios involve the vehicle being unable to navigate certain areas or escape dangerous situations due to errors in the control structure.

Several scenarios emanated from control structure issues, where a signal governing the vehicle's operation was compromised, leading to critical situations. The scenarios outlined the risks of delayed response, failed authentication procedures, and unexpected delays during urgent evacuations. Additionally, issues with the operator's UI were identified, such as misinterpretation, unclear indications, and authorization problems, potentially resulting in repeated attempts and troubleshooting failures.

Authentication problems were also identified, such as system bugs preventing startup due to authentication failure. Timing issues with signals, whether they arrived too early, too late, or persisted for too long—often resulted in scenarios depicting unexpected behavior, including abrupt halts, unexpected swerves, collisions, and erroneous decisions by the autonomous system.

These examples provide a glimpse into the main areas that the scenarios predominantly addressed. For a comprehensive understanding, refer to Chapter 3.4 for the complete response to question 1.

And so on to the second research question:

2. How can the validation of an autonomous control system be ensured? This includes considering the use of the STPA method and other applicable approaches.

There are several methods for validating an autonomous system. STPA is here shown to be an effective, and flexible method that offers numerous advantages. By examining all the information obtained from the analysis of NGV, it becomes evident how much insight can be gained from a theoretical system that does not yet exist but is under investigation. The method allows for great flexibility, as demonstrated by a study at the Royal Institute of Technology (KTH) in 2019 [11]. The master's thesis specifically employs an STPA analysis to investigate the interaction between drivers and vehicles, highlighting its applicability to specific domains.

One can choose the number of iterations to perform with the method, providing the flexibility to tailor the level of detail. However, despite this flexibility, each iteration remains quite comprehensive, demanding time and effort.

Another method for system validation is the Six-Step Method, as demonstrated in an article published by the Centre for Research in Cyber Security at Singapore University of Technology and Design [3]. This approach is applicable and, perhaps, clearer due to its well-defined steps. It results in a less extensive analysis compared to STPA, making it a potentially better choice when time, resources, or the opportunity for a complex analysis is limited.

There are specific methods designed for security analyses within cybersecurity, such as TARA and SAE J3016. These methods are specifically crafted to validate the cybersecurity aspects of systems, and they excel in this domain. Therefore, when validating an autonomous system, it is beneficial to have a clear focus on the primary area of analysis, particularly in terms of cybersecurity.

in another study conducted by David Robert Beachum at the University of Texas at Austin [13], the author argues that analytical methods, including FMEA, are recognized for systematic risk identification and mitigation. However, FMEA's limitations stem from its assumption that multiple failure modes do not occur concurrently, restricting its capacity to

represent the interdependent subsystems of AVs. On the other hand, FTA, with its ability to compare reliability and failure rates, depends on challenging data acquisition for every component, including machine learning algorithms.

What he asserts, and what emerges from the discussed research, is that the optimal way to validate an autonomous system appears to be a combination of several methods to cover as many areas as possible. The prominent international standard ISO 26262, originating from 2011, might be suitable for integration with other methods as it remains globally recognized, but it is also considered outdated due to significant advancements in this field since 2011. The research suggests that the most effective approach to validate an autonomous system involves a combination of multiple methods, selecting those specifically tailored to one's own system rather than opting for a one-size-fits-all solution. Simulations may also be a powerful approach, especially when using real-world data/situations.

In conclusion, applying a precise and system-specific approach, and adeptly combining analyses such as STPA, TARA, FMEA, FTA, and ISO 26262, along with simulations derived from real-world situations, forms a robust strategy for ensuring the validation of an autonomous control system.

5 Discussion

Regarding the results of the STPA analysis, it is worth noting that the outcome essentially constitutes the entire analysis. All the information generated from each step in the analysis can serve as a foundation for further product development. The control structure can be utilized by designers as a basis for the actual programming and implementation of each step, and all the different constraints lists can be used as a foundation for a requirements specification. All the scenarios that emerged in the end can be kept in mind by all parties to prevent such incidents.

What becomes clear now that the analysis is complete is the multitude of risks associated with implementing autonomy in a vehicle, especially in terrains where directives may not be as clear as in traffic. There are many potential dangers that can arise. Even though many might agree that the vehicle should not collide, hit someone, get stuck, etc., it's easy to forget the various specific scenarios that can arise from different faults between different connections in the system. The connection between the operator and the ECU can lead to numerous faults, while the connection to the Safety System and the ECU can result in entirely different potential issues. Documenting all of this is essential before progressing with product development. Despite many parties having a clear vision of how a product should function, it is crucial, especially when there is much at stake and the risk of dangers is present, to have everything documented and outlined to anticipate as many potential issues as possible, thus reducing the risk of actual incidents when the product is completed.

By examining how to validate autonomous systems, it became evident that there are various approaches to accomplish this. Different standards have been established, such as ISO 26262, alongside methods like STPA and the Six-Step Method. Choosing the right method can be challenging, but it is crucial to tailor the method as much as possible to the system under examination. In this case, the analysis of NGV might have been enhanced if combined with a method specifically designed for cybersecurity, as cybersecurity was designated as a primary focus from the outset. Cybersecurity became a significant aspect in the STPA analysis, so exploring and incorporating TARA or SAE J3016 could have made it more comprehensive, possibly highlighting more specific risks related to cyber-attacks.

Another factor that may come into play is the stage of production of a product. In the

early stages, STPA proves to be beneficial. However, as one progresses in the development process and seeks to analyze the system, the Six-Step Method might offer greater strength. Its clarity and reliance on existing information about the system can save time and effort compared to the more theoretical and abstract nature of STPA.

Another aspect that became evident when exploring different methods and research is the importance of common standards. The international standard ISO 26262, despite being older and from 2011, demonstrates strength in its broad international recognition. A well-known standard becomes more valid and influential as more people and companies are familiar with its structure and requirements. The widespread adoption of a particular method enhances its validity, making it easier to compare different vehicles or systems when evaluated using the same criteria. This principle likely applies to all standards and methods.

Similarly, STPA also exhibits strength in this regard. Being a well-established and widely recognized method adds a layer of confidence and reliability, making it increasingly advantageous as more entities adopt and utilize it. The shared understanding and usage of a common framework contribute to a more robust and effective validation process across various domains and applications.

STPA, nonetheless, remains a substantial, effective, clear, and comprehensive method that provides a wealth of information. This analysis constituted only one iteration of the system. The concept with STPA is that one can choose how many iterations to perform, determining how deeply detailed the chosen area should be explored. In this analysis, only one iteration was completed, encompassing a substantial amount of data and control structures. This highlights a clear potential drawback of STPA—it demands time and effort. Although there was a hope to conduct multiple iterations, only one was completed in this analysis.

If more time had been available, another iteration specifically focusing on the autonomous subsystem would have been conducted. This would involve creating a new control model for that specific component, providing more information about control actions and specific elements within those blocks. Additional UCAs and scenarios would have been generated, leading to a more detailed understanding of the autonomy, which is the primary area of interest. With more time, there would have also been a discussion about the specific challenges arising when an autonomous vehicle operates in terrain, as opposed to traffic, where clear directives may be lacking. More research in these areas would have been explored.

Even though this analysis yielded a substantial amount of information and scenarios, it is a subject that has not been extensively researched. Further research in this domain is crucial, and additional studies on autonomy in off-road environments are needed. As we enter an era where autonomous vehicles play an increasingly prominent role in the market and the world, there are inherent risks. Trying to anticipate as much as possible about potential challenges minimizes the risk of unforeseen issues when such vehicles are deployed.

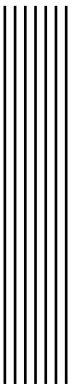
6 Conclusion

The primary objective of this study was to identify and analyze risks associated with the early-stage development of an autonomous vehicle, focusing on the Milrem Robotics NGV concept. By employing the STPA method, various risks and potentially hazardous scenarios were concretely identified. These risks encompass system-level issues within the vehicle and were analyzed to set fundamental requirements and establish clear system limitations.

In response to the first research question, "What risks can be concretely identified in the early stages of developing a self-driving vehicle?" the STPA analysis revealed requirements, limitations and risks. A control structure was devised for the vehicle, mitigating identified risks by defining necessary components to uphold the initially set requirements. The analysis further explored scenarios, highlighting risks related to autonomy failures, such as the vehicle being unable to navigate safely or colliding due to malfunction. Risks, especially those associated with cybersecurity, were also identified during the analysis.

Regarding the second research question, "How can the validation of an autonomous control system be ensured? This includes considering the use of the STPA method and other applicable approaches." The findings emphasized the importance of tailoring validation methods to the specific vehicle, development stage, and available resources. The STPA method proved effective in the early stages, offering flexibility and adaptability to diverse systems. However, considerations for later stages led to the exploration of alternatives, such as the Six-Step Model, which provides clearer steps and is less exhaustive. The study underscores the significance of selecting validation methods aligned with the purpose, system characteristics, and development stage. Combining multiple approaches and leveraging domain-specific analyses, like TARA for cybersecurity, enhances the comprehensiveness and effectiveness of validation. As autonomous vehicles evolve, adapting validation strategies becomes crucial, with early stages benefitting from the flexibility of STPA and later stages potentially favoring more structured methodologies like the Six-Step Model.

In conclusion, the study provides insights for ensuring the safety and reliability of autonomous control systems throughout their development lifecycle.



Bibliography

[1] *Milrem Robotics*. Accessed on 2023-09-24. URL: https://milremrobotics.com/news/ (visited on 09/24/2023).

[2] Nancy G. Leveson and John P. Thomas. *STPA Handbook*. Mar. 2018.

[3] Lin Shen Liew Giedre Sabaliauskaite and Jin Cui. *Integrating Autonomous Vehicle Safety and Security Analysis Using STPA Method and the Six-Step Model*. Accessed on 2024-01-7. Centre for Research in Cyber Security (iTrust). N/A. URL: https://pureportal.coventry.ac.uk/en/publications/integrating-autonomous-vehicle-safety-and-security-analysis-using.

[4] *Threat Assessment and Remediation Analysis (TARA)*. Accessed on 2023-09-24. MITRE Corporation. URL: https://www.mitre.org/news-insights/publication/threat-assessment-and-remediation-analysis-tara#:~:text=Threat%20Assessment%20and%20Remediation%20Analysis%20%28TARA%29%20is%20an,and%20select%20countermeasures%20effective%20at%20mitigating%20those%20vulnerabilities. (visited on 09/24/2023).

[5] *Process Hazard Analysis (PHA)*. Accessed on 2024-01-07. Wikipedia. URL: https://en.wikipedia.org/wiki/Process_hazard_analysis (visited on 01/07/2024).

[6] *Fault Tree Analysis (FTA)*. Accessed on 2024-01-07. IBM. URL: https://www.ibm.com/topics/fault-tree-analysis (visited on 01/07/2024).

[7] *Failure Modes and Effects Analysis (FMEA)*. Accessed on 2024-01-07. University of Cambridge. URL: https://www.ifm.eng.cam.ac.uk/research/dmg/tools-and-techniques/fmea-failure-modes-and-effects-analysis/ (visited on 01/07/2024).

[8] *ISO 26262 - Functional Safety Standard for Road Vehicles*. Accessed on 2024-01-07. ROHM Semiconductor. URL: https://fscdn.rohm.com/en/products/databook/white_paper/iso26262_wp-e.pdf (visited on 01/07/2024).

[9] *SAE J3016 Update*. Accessed on 2024-01-07. Society of Automotive Engineers (SAE). URL: https://www.sae.org/blog/sae-j3016-update (visited on 01/07/2024).

[10] *SAE J3061-2021: Cybersecurity Guidebook for Cyber-Physical Vehicle Systems*. Accessed on 2024-01-07. Society of Automotive Engineers (SAE). URL: https://img.antpedia.com/standard/files/pdfs_ora/20221211/sae/SAE%20J3061-2021.pdf (visited on 01/07/2024).

[11] Max Stoltz-Sundnes. "STPA-Inspired Safety Analysis of Driver-Vehicle Interaction in Cooperative Driving Automation". In: (2019). Master's Thesis at KTH Royal Institute of Technology, Accessed on 2024-01-07. URL: http://www.diva-portal.org/smash/get/diva2:1371216/FULLTEXT01.pdf.

[12] Giedre Sabaliauskaite and Jin Cui. "Integrating Autonomous Vehicle Safety and Security". In: *Journal of Autonomous Vehicles* (2022). URL: https://web.archive.org/web/20180409213006/http://www.thinkmind.org/download.php?articleid=cyber_2017_5_40_88003.

[13] David Robert Beachum. *Methods for Assessing the Safety of Autonomous Vehicles*. Master's Thesis at The University of Texas at Austin, Accessed on 2024-01-18. May 2019. URL: https://repositories.lib.utexas.edu/server/api/core/bitstreams/919f5ddc-4d3f-4475-8281-f09a9357e783/content.