Amelie Kleber, Gottlieb Dinh, Jan Chen, Jonathan Deul, Lennard Zei

Resilient Cognitive Systems – Document

Safety-Team 1

Contents

Introduction 2
Content 3
HARA-Analysis 3
Previous System Architecture 4
Limitations with Previous System Architecture 5
Fault-Tree Analysis - Gottlieb 5
CARE-Analysis 6
CAUSE TREE – Jan 6
Improved Safety Concept - Lennard 7
CARE Analysis of Improved Safety Concept - Gottlieb 8
Evidence of Efficacy – Gottlieb 8
Limitations & Countermeasures – Jan 9
Business Case 10
Safety Demo – Jonathan 10
Appendix 10

Amelie Kleber, Gottlieb Dinh, Jan Chen, Jonathan Deul, Lennard Zei

Introduction

The primary objective of this project is to develop and optimize a safety concept for a robotic arm that interacts with humans in various analysis scenarios. The core challenge lies in ensuring that the robot operates efficiently, while maintaining safety by stopping immediately when a human enters its vicinity.

Project Goals and Constraints

Following the guidelines of the Resilient Cognitive Systems challenge, we considered the following factors:

- Safety Criticality: Any scenario resulting in the robot touching a human leads to a score of zero.
- Sensor Economy: While one camera is provided, additional sensors can be integrated at the cost of the overall score.
- Environmental Adaptability: The system must remain robust across diverse contexts, including varying light conditions, different human appearances and crowded environments.
- Infrastructure Limitations: External safety measures like physical fences or light barriers are strictly prohibited.

Methodology

To ensure a comprehensive safety concept, we employed the following frameworks:

- HARA-Analysis
- Fault-Tree Analysis
- CARE-Analysis
- Cause Tree

This document details our transition from a vulnerable, camera-only perception system to an improved, multi-sensor architecture designed to provide a high level of efficacy and resilience in human-robot interaction.

Amelie Kleber, Gottlieb Dinh, Jan Chen, Jonathan Deul, Lennard Zei

Content

HARA-Analysis

In order to provide a safety concept with the best possible coverage, we need to clearly define our system, including technical elements and actors involved.

For this, we used the HARA-Analysis introduced in our course. The HARA aims to define the scope of our system and provide a largely complete overview of factors involved by dividing them into 7 steps: The System, Persons at Risk, Hazards & Hazard types, Physical Properties, Actuator Failure Modes, Actuators and Failure Modes.

From there on, we used these categories to combine them into HARA-guide phrases to serve as exemplary hazard cases to countermeasure against.

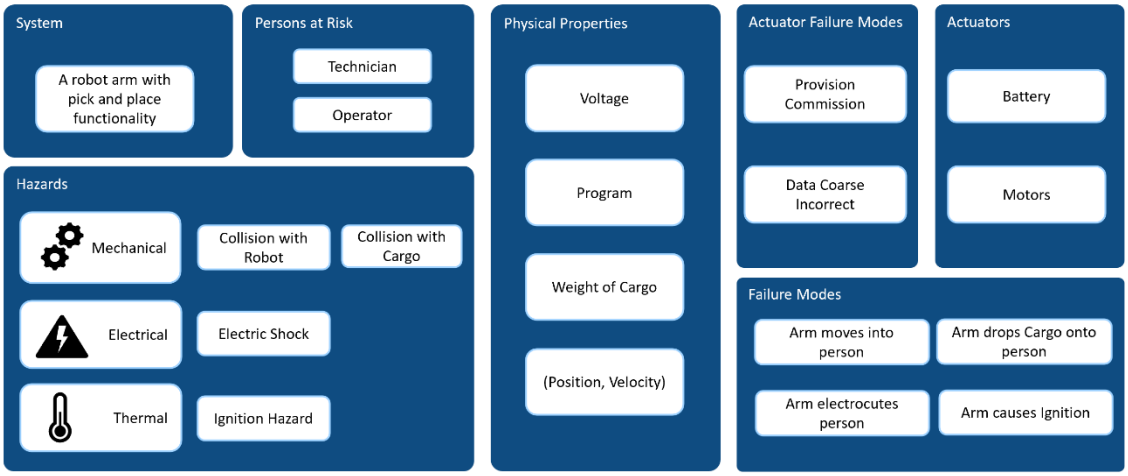


Figure 1

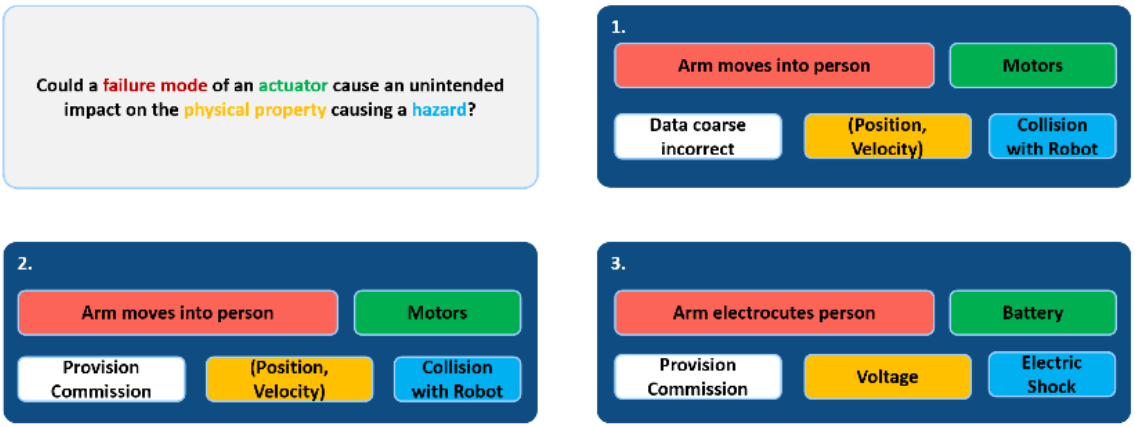


Figure 2

To provide a proof of completeness for the HARA-analysis, we opted for a traceability matrix (Figure 1) that compares Physical Properties against all other HARA-categories, except for the actuators. The reason for this selection was that we were sure of the completeness of the two actuators in our system. Each Physical property could be traced back to one of the actuators that it affected and was therefore more likely to be complete as well. Reading the matrix from left to right provides the remaining guide-phrases.

Previous System Architecture

We define the robotic arm system in detail using a three-level architecture that takes in the arm's current position and the desired movement plan.

In the first level, we discern several elements:

- Mission planner, which receives a mission and outputs a target and object
- Obstacle Detection, which receives camera information and outputs an object list
- Gripper Sensor, which receives the grip strength and grip width and outputs the grip percentage and strength
- Joint Motor Sensors, which receive the position, velocity, and acceleration of the joint motors and output a vector consisting of the three numbers

At the second level, the Trajectory Planner receives the outputs and maps them to a trajectory consisting of the arm's current and next positions. Additionally, it outputs the time to destination.

The third level Trajectory Control transforms them into six different degrees for each joint motor of the robot arm, with one additional vector for the robot grip. This represents the actual movement of the arm to fulfil the mission.

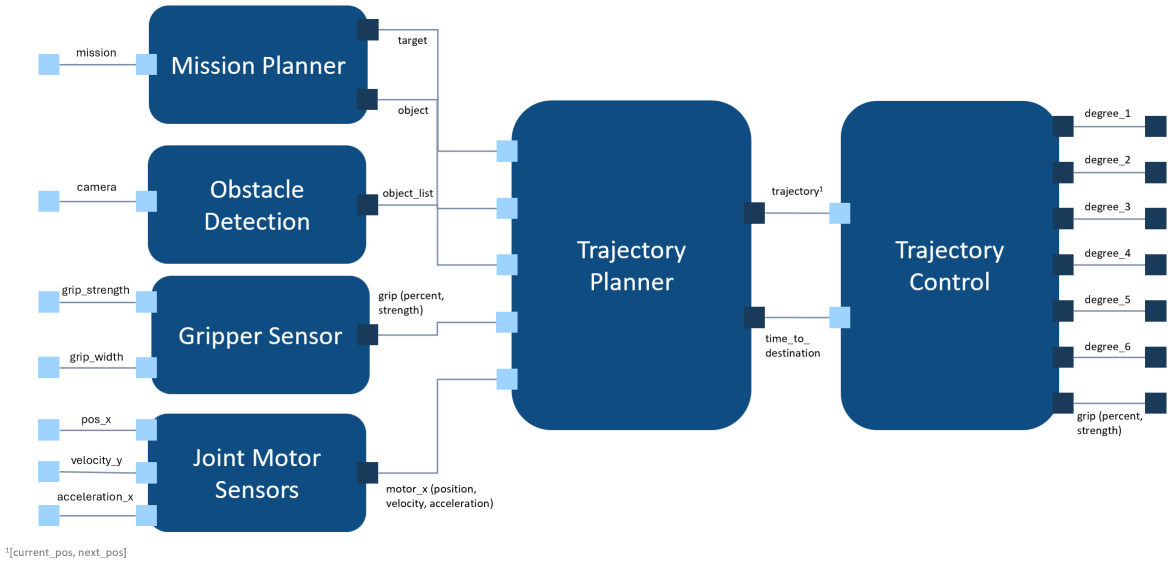


Figure 3

Limitations with Previous System Architecture

However, camera-only perception cannot guarantee safe human detection due to variable visibility, signal degradation, and ambiguous interpretation by ML models. We defined three broad categories for limitations imposed on our previous System Architecture.

Physical limitations	Sensing limitation	Computational limitations
• **Darkness, light fluctuations, and absorbent clothing** leads to insufficient light intensity reaching lens • **Humidity, fog, rain, smoke** can distort optical signal • **Occlusion, lens dirt, or misaligned camera** can block rays, leading to unusable signal	• **Quantization or clipping** can remove subtle cues or create course artifacts • **Fast motion + exposure timing** can cause motion blur / rolling-shutter distortion • **Fixed Pattern Noise, vibration, or power issues** can reduce SNR and corrupt frames	• **Out-of-distribution scenes** and **unseen circumstances** can cause omission or false activation • **Reflection, pictures or person-like background** patterns can trigger provision commission • **Biased training data and learned shortcuts** cause poor generalization

Figure 4

Fault-Tree Analysis

We analysed the architecture level by level to identify every possible fault. For this we defined our top fault “robot arm moves even though it should not” as degree of motor x is too high.

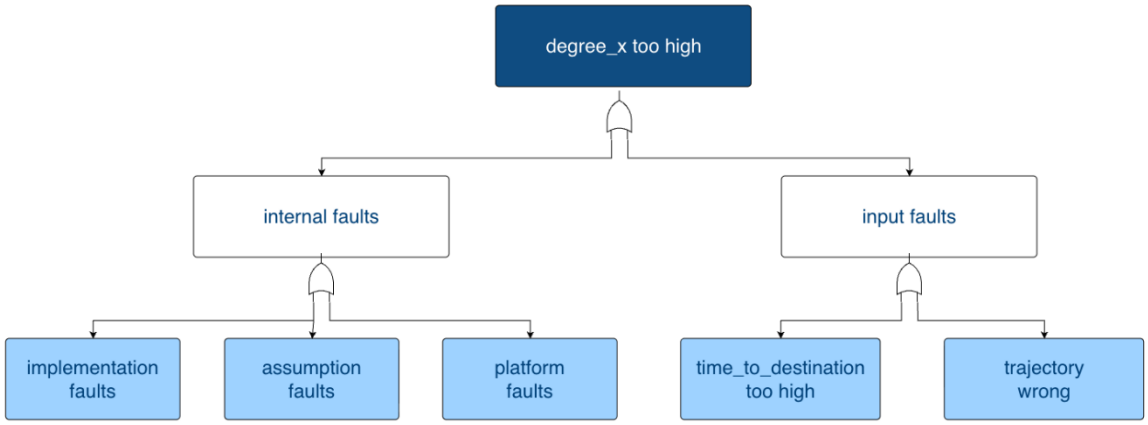


Figure 5

CARE-Analysis

CARE adds structure for perception and actuation analysis by subdividing it into single steps that provide a basis for systematic analysis and coverage. In the following, we will focus only on the sense half of the model, as our system is detection-oriented. The analysis is subdivided into four steps. Each step provides a source, a model, the model's assumptions, a sink, an insufficiency backlog, and some exemplary analysis cases. For completeness, each step is accompanied by a traceability matrix that matches Assumptions to Failure Modes.

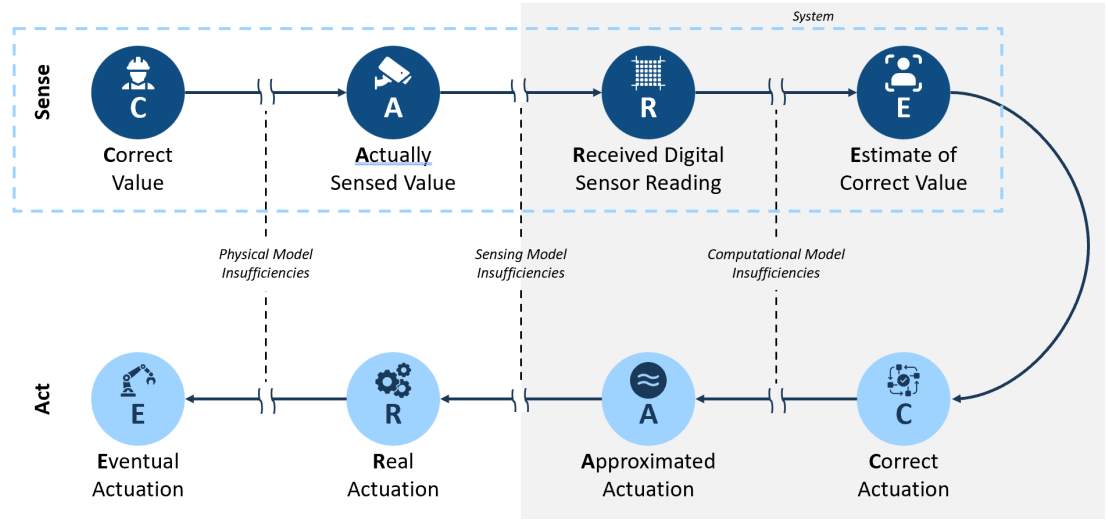


Figure 6

- C->A: We examine insufficiencies that might occur and lead to differences between the actual value and the sensed value, causing erroneous detection. (Figure 2 + 3)
- A->R: We examine insufficiencies that might occur and lead to differences between the actually sensed value and the digital representation of the value, causing erroneous detection. (Figure 4 + 5)
- R->E: We examine insufficiencies that might occur and lead to differences between the digital representation of the value and the estimated value, causing erroneous detection. Figure (6+7)

Cause Tree

Based on the results of the HARA, fault-tree, and CARE analyses, we derive a cause tree to systematically explain how a person in the hazard zone may remain undetected.

At the highest level, the unwanted event can occur either because a valid image is available but misinterpreted by the computational model, or because the image is already insufficient at the model input. These two branches directly map to the CARE structure. Computational model

insufficiencies occur during the R→E step and include cases in which available digital data is incorrectly interpreted due to biased or insufficient training data, limited model capacity, or inappropriate runtime thresholds and decision logic.

In contrast, insufficient model input is further decomposed into sensing and physical model insufficiencies. Sensing model insufficiencies correspond to the A→R step and describe failures in sensing and digitization, such as limited dynamic range or quantization, inadequate temporal sampling or exposure, sensor noise, spectral mismatch, or preprocessing that removes relevant details. Physical model insufficiencies reflect the C→A step and describe violations of real-world assumptions at the optical input, for example, due to unfavorable lighting, unexpected appearance or reflectance of a person, environmental influences on light propagation, or geometric constraints such as occlusion or a limited field of view. The cause tree highlights recurring insufficiencies across the CARE layers, which directly define the points addressed by the proposed safety concept.

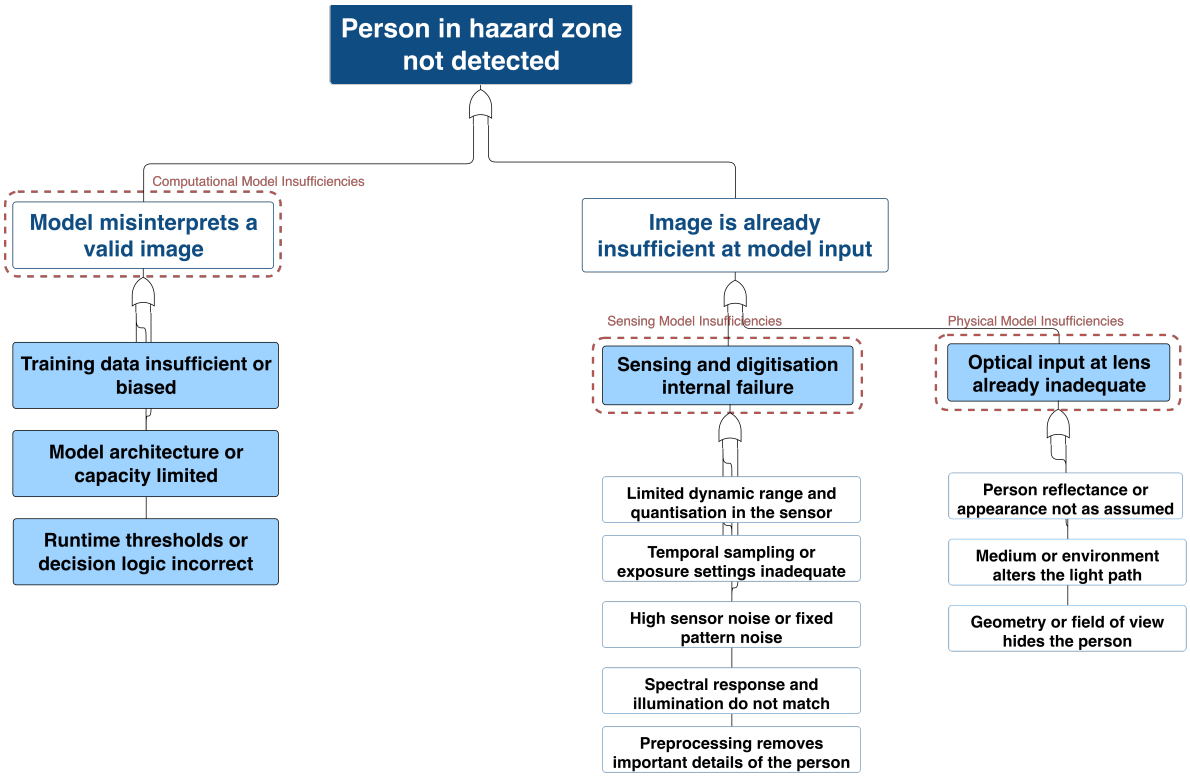


Figure 7

Improved Safety Concept

To move away from our previous Safety Concept/Architecture, which relied solely on camera vision and had severe limitations, we opted for a dual-sensor setup. In addition to the optical camera data, the Object Detection element receives readings from a mmWave sensor. Depending on the readings, it either outputs an empty object list, indicating it will not stop, or an array containing the position of the detected movement, leading it to stop all movement.

When the mmWave detects movement = false, it outputs an empty array directly to the object list. If detects movement = true, we determine how far away the movement is. If it is greater than

2 meters away we once again pass on an empty array list, since the object is too far away to stop. If it closer than 2 meters we pass on an array with the position of all movement.

In the next step, we determine whether the detected movement was from a robot or not. For that we use the optical data from the camera. In order for a robot to be identified we use both a QR-Code for detection and a Classification trained on recognizing robots. If both these measures are true, an empty array list is passed to the Object list, since we do not want to stop when robots interact with each other. However, if only one of these measure returns false, we decide that a non-human entity is detected and we stop just in case by passing the position of all movement to the Object List.

CARE Analysis of Improved Safety Concept - Gottlieb

Evidence of Efficacy – Gottlieb

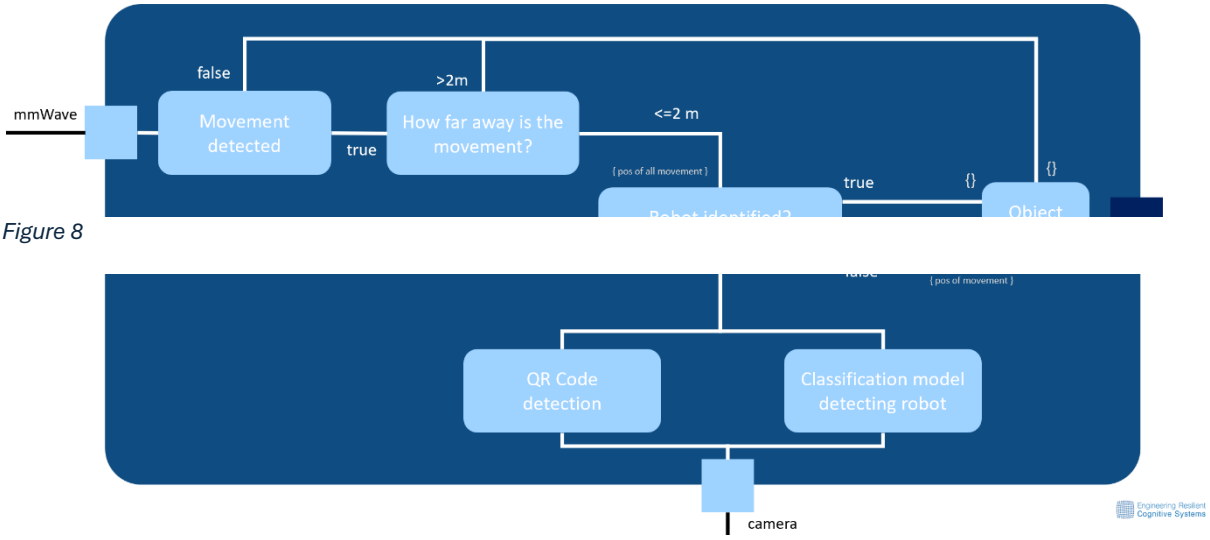


Figure 8

Limitations & Countermeasures

Measures against QR-Code Tempering

To ease concerns about workers' misuse of robot QR code identification, we propose various countermeasures to prevent replication and destruction of the QR codes.

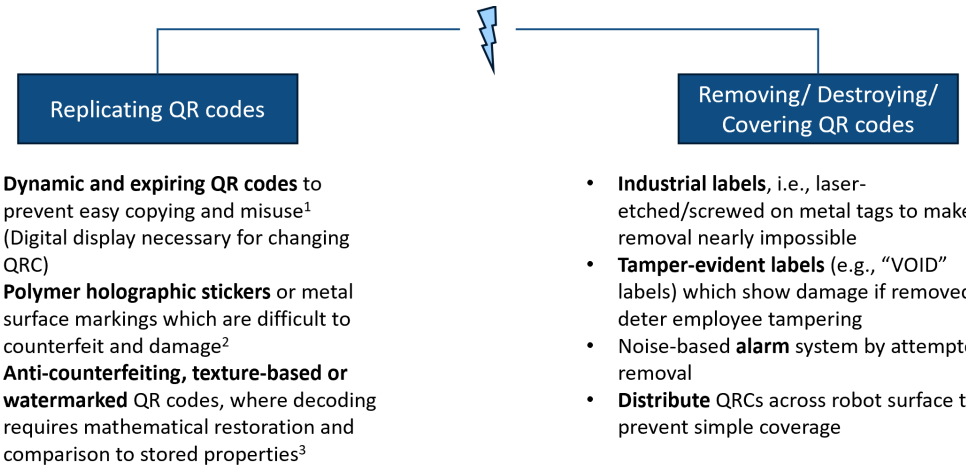


Figure 9

Measures against Provision Commission

For the remaining Provision Commission Issues, we propose the following measures to enhance the safety of our system further.

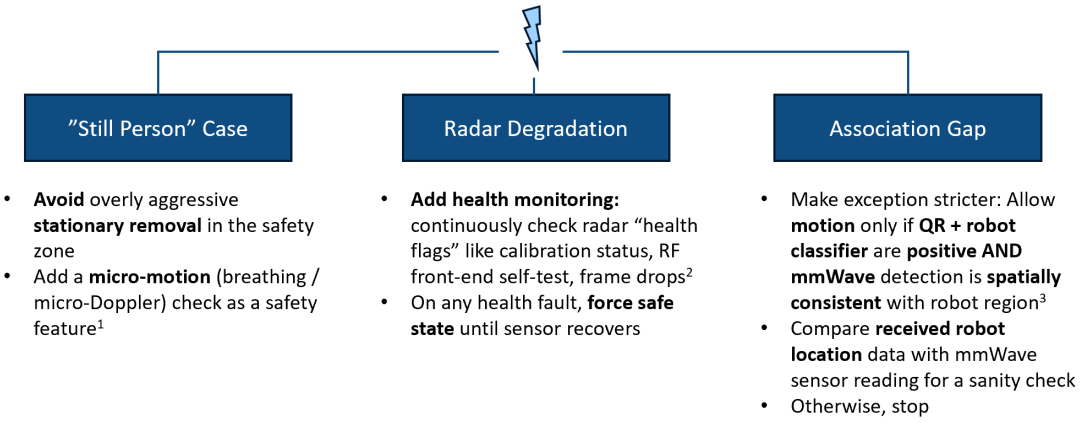


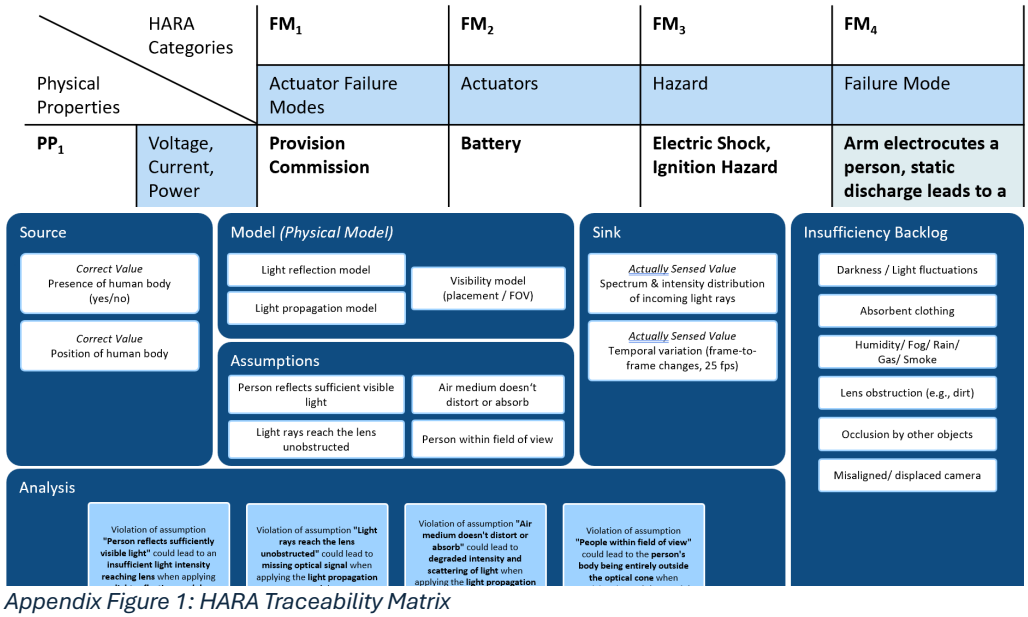
Figure 10

Business Case

Safety Demo – Jonathan

Appendix

Appendix Figure 1: HARA Traceability Matrix..... 10
Appendix Figure 2: C -> A 11
Appendix Figure 3: C -> A Traceability Matrix ..... **Error! Bookmark not defined.**
Appendix Figure 4: A -> R..... **Error! Bookmark not defined.**

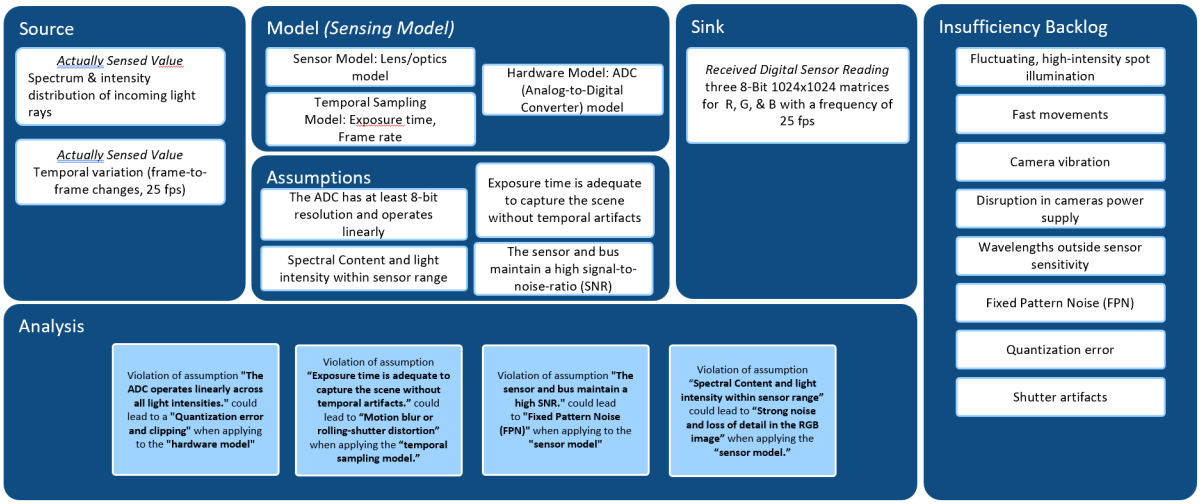


Appendix Figure 1: HARA Traceability Matrix

Appendix Figure 2: C -> A

Assumptions \ Failure Modes		FM_1 Timing Late	FM_2 Timing Early	FM_3 Provision Commission	FM_4 Provision Omission	FM_5 Data Subtle	FM_6 Data Coarse
A_1	Person reflects sufficient visible light	TC_{11}: Person steps from darkness into light	TC_{12}: Sudden glare or specular reflection before person enters area	TC_{13}: Highly reflective back-ground objects (= person-like features)	TC_{14}: Person wearing dark, non-reflective clothing (absorbent material)	TC_{15}: Low reflectance clothing and low ambient light/ darkness	TC_{16}: Overexposure / bloom (strong direct light) saturates pixels
A_2	Medium doesn't distort or absorb	TC_{21}: Fog or steam dissipates just before detection	TC_{22}: Rain or water distortion leads to illusion of smaller distance	TC_{23}: Dense rain droplets produce bright/dark patterns that mimic person	TC_{24}: Thick fog or heavy rain attenuates person signature	TC_{25}: Light haze or high humidity blurs edges and reduces contrast	TC_{26}: Dense dust/ smoke creates large textured blobs & coarse shapes
A_3	Light rays reach the lens unobstructed	TC_{31}: Temporary occlusion enters frame, delaying person detection	TC_{32}: Passing occlusion classified as person	TC_{33}: Transient reflections create person-like contours	TC_{34}: Obstruction of lens causes reduced or no feature visibility	TC_{35}: Thin film causes subtle blurring of person edges	TC_{36}: Lens heavily occluded/scratched; image shows large indistinct regions
A_4	Person within field of view	TC_{41}: Person enters the peripheral FOV first, only later moves into central detection zone	TC_{42}: Person silhouette appears momentarily at the edge, triggering detection early	TC_{43}: Background object with human-like vertical shape at the edge of FOV	TC_{44}: Camera is misaligned or displaced	TC_{45}: Person partially occluded by scene elements, only subtle features visible	TC_{46}: Person at extreme distance with low pixel footprint

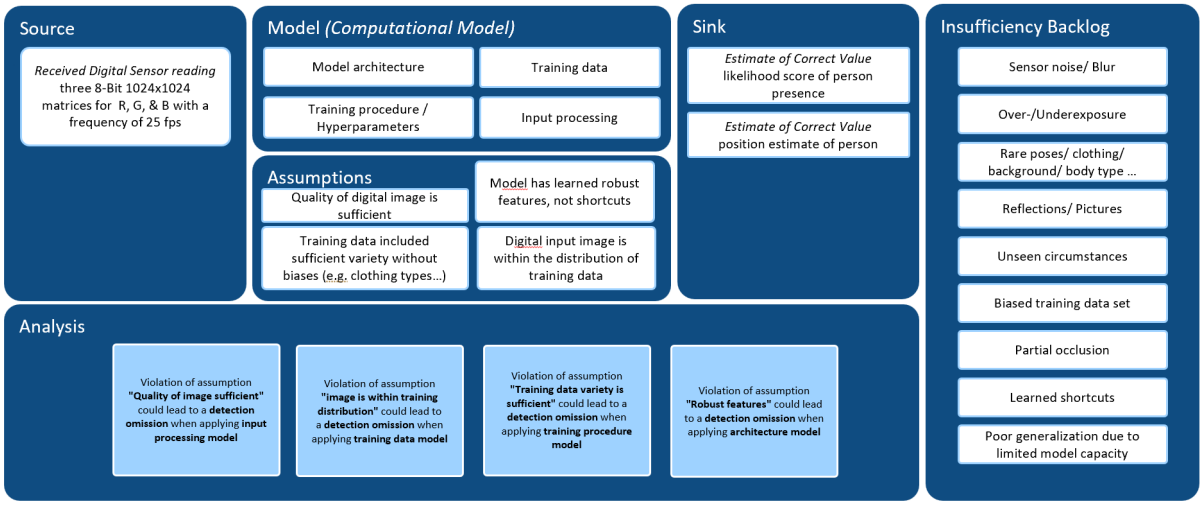
Appendix Figure 4: C -> A Traceability Matrix



Appendix Figure 3: A -> R

Assumptions \ Failure Modes		FM_1 Timing Late	FM_2 Timing Early	FM_3 Provision Commission	FM_4 Provision Omission	FM_5 Data Subtle	FM_6 Data Coarse
A_1	ADC has 8-bit resolution and operates linearly	TC_{11}: ADC auto-exposure + 8-bit quantization requires multiple frames to settle	TC_{12}: Quantization step noise or transient peak in a single frame	TC_{13}: Quantization artifacts or banding produce person-like edges/patterns	TC_{14}: 8-bit saturation removes subtle gradients that indicate a person	TC_{15}: Low bit depth causes low contrast for small/remote person pixels	TC_{16}: Severe quantization + dithering produces blocky/coarse pixel patterns
A_2	Exposure time is adequate to capture the scene without temporal artifacts	TC_{21}: Readout hiccup / frame drop occurs during approach	TC_{22}: Frame duplication/ buffer replay presents stale frame	TC_{23}: Jitter in frame timestamps/ corrupted frames produce motion patterns	TC_{24}: High bus contention delays frame read-out, sensor to discards a new frame	TC_{25}: Rolling Shutter Distortion	TC_{26}: Read-out process aborted mid-frame due to system error, resulting in partial image frame
A_3	Spectral Content and light intensity within range	TC_{31}: Sensor requires long exposure time due to low light	TC_{32}: Sensor has insufficient recovery time after severe light saturation	TC_{33}: Specific light wavelength is misinterpreted as a visible object	TC_{34}: Light intensity too low to detect a person's presence.	TC_{35}: Scene contrast is lost due to specular reflection	TC_{36}: Extreme low light causes pixel values to be near zero (dark)
A_4	The sensor and bus maintain a high signal-to-noise-ratio (SNR)	TC_{41}: High SNR maintenance overhead delays the frame read-out	TC_{42}: High thermal noise causes buffer instability, leading to stale data reuse	TC_{43}: Fixed Pattern Noise (FPN) creates false person-like pixel artifacts	TC_{44}: Random noise obscures the subtle edges and features of a person	TC_{45}: High Noise level reduces effective dynamic range of person pixels	TC_{46}: Electrical noise spikes cause individual pixel values to become corrupted

Amelie Kleber, Gottlieb Dinh, Jan Chen, Jonathan Deul, Lennard Zei



Assumptions \ Failure Modes		FM_1 Timing Late	FM_2 Timing Early	FM_3 Provision Commission	FM_4 Provision Omission	FM_5 Data Subtle	FM_6 Data Coarse
A_1	Quality of digital image is sufficient	**TC_{11}:** Image sharpness improves only after several frames	**TC_{12}:** Early frame compression artifact or motion blur	**TC_{13}:** Noise, blur, or patterns resembling a person	**TC_{14}:** Image degraded (blur, smear, low resolution)	**TC_{15}:** Slight motion blur or defocus lowers feature clarity	**TC_{16}:** Severe compression, low resolution, or defocus
A_2	Input is within the distribution training data	**TC_{21}:** Out-of-distribution (OOD) scene causes model to hesitate	**TC_{22}:** OOD feature triggers early false activation	**TC_{23}:** Objects never seen in training (e.g., mannequins, statues, shadows) resemble humans	**TC_{24}:** Person outside training distribution (e.g., unusual clothing/ posture)	**TC_{25}:** Slight distribution shift (e.g., new camera angle, lighting style)	**TC_{26}:** Strong distribution shift (e.g., infrared illumination, fisheye lens)
A_3	Training data included sufficient variety without biases	**TC_{31}:** Under-represented groups (e.g., specific clothing colors)	**TC_{32}:** Model over-fits to one feature, detects object with shared feature	**TC_{33}:** Shortcut in training, causing mis-identification of background	**TC_{34}:** Underre-presented demo-graphics/ poses not recognized	**TC_{35}:** Features for a particular subgroup captured poorly	**TC_{36}:** Model generalizes poorly to certain body types or apparel
A_4	Model has learned robust features, not shortcuts	**TC_{41}:** Model depends on context features (e.g., shadows)	**TC_{42}:** Shortcut cue appears early	**TC_{43}:** Shortcut feature in background wrongly activates	**TC_{44}:** Shortcut fails, so the person is not detected	**TC_{45}:** Shortcut cue is partially present, model confidence fluctuates	**TC_{46}:** Shortcut-based model generalizes poorly

Amelie Kleber, Gottlieb Dinh, Jan Chen, Jonathan Deul, Lennard Zei