Verification & Validation for Autonomous Systems “Safety Assurance”

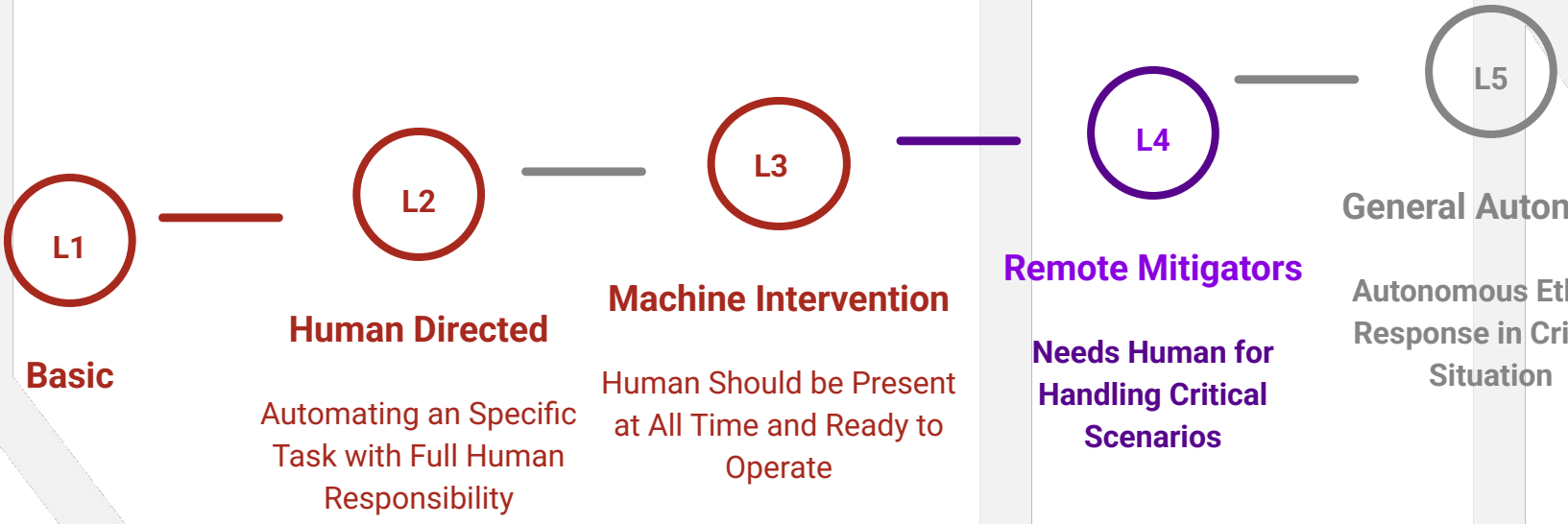
Aliasghar (Ali) Arab
Assistant Adjunct Professor
NYU Mechanical & Aerospace Engineering

December 2025 – Autonomous Mobile Robots and Autonomous Vehicles ROB-7863

Outline

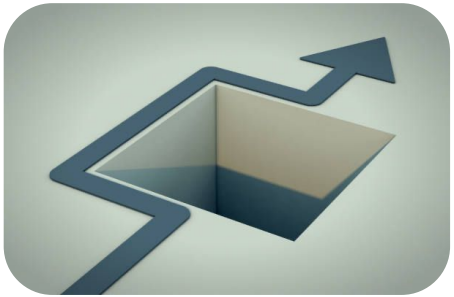
- Safety Concepts
- Systems Thinking
- V&V Methodologies
- System Engineering vs Safety Engineer
- Introduction to Standards
- Requirements, architecture design and testing

Autonomy Levels



Autonomous Systems

Safety Concept

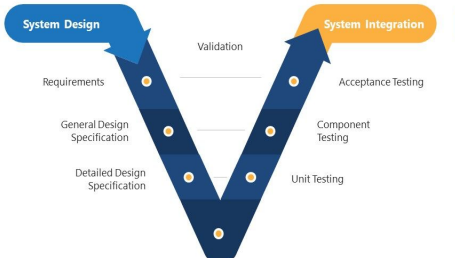


Identify and Analyze Hazards

Assess and Mitigate Risks

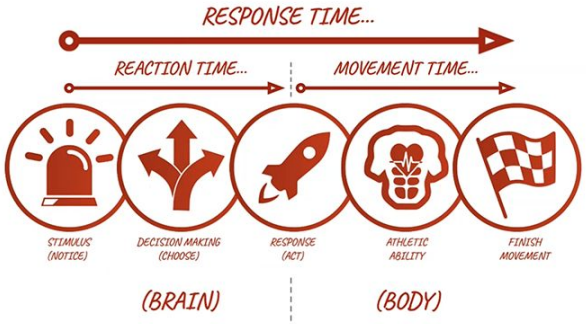
Iteratively enhance the reliability

Real-Time Safety Feature



Verification & Validation

Proper safety assessment framework for such a system is extremely challenging



Limited Response Time

The reaction time from detecting critical situation to act is very small.

Ethics and Legals

There will be always an ethical and legal challenge for deciding what should be prioritized.

Lack of Real Examples

Safety Concept

01

NYU

1.1 What is Safety?

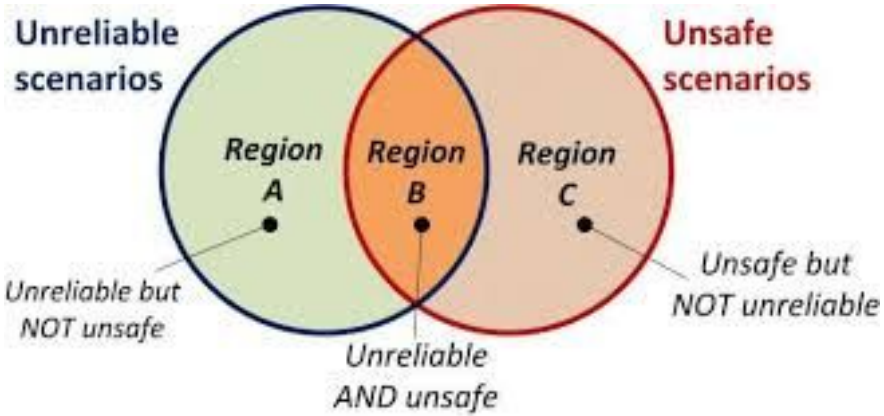
Safety is the condition where the risk of harm to people, property, or the environment is reduced to an acceptable level and kept there throughout the system's lifecycle.

Autonomous Mobility's Safety is Harder

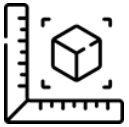
- Complex, dynamic, unpredictable environments
- Novel scenarios not seen in training
- Conditions with incomplete, noisy, or misleading sensor data
- Partial system failures Interactions with humans
- Shifting operational design domains

Reliability vs Safety

Safety is not just reliability or performance — a system may be reliable but unsafe.

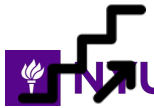


Why Measurement?



Measures, informs and accelerates model improvement and progress.`

Informs users and builders on trade-offs and progress.



Scientifically measuring “levels of intelligence”

AI Safety Measurements

The earlier you know, the earlier you mitigate.

AI has reached millions of practitioners. Scaling education and safe development.

Safety and reliability are capabilities!

1.2 Hazards | Risks | Failure

HAZARD

A Hazard is a situation that has the potential for a loss

RISK

Risk is the likelihood of a hazard causing loss

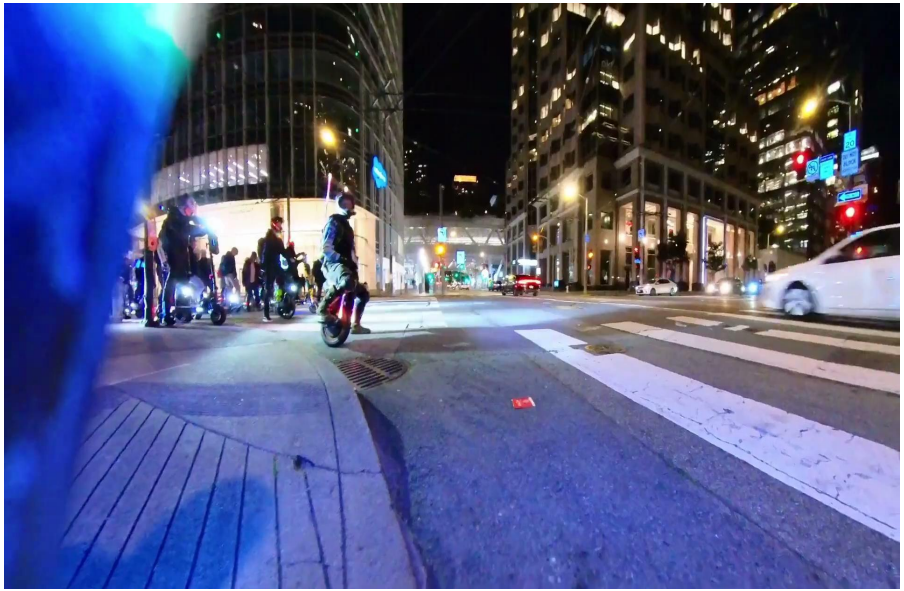
Loss

Accident is an event that results in a Loss

Type of Hazards

1. Hazard exists because something failed (FuSa)
2. Hazard exists without a failure — due to design limitations or unknown unsafe scenarios (SOTIF)

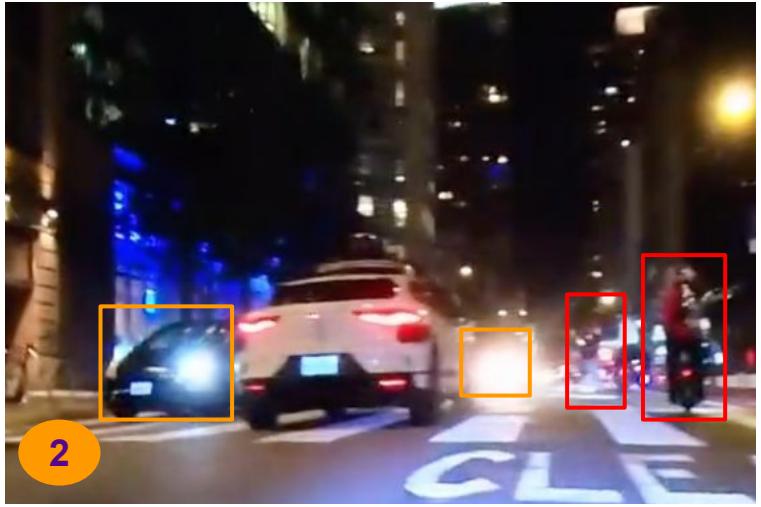
What kind of hazard is this?

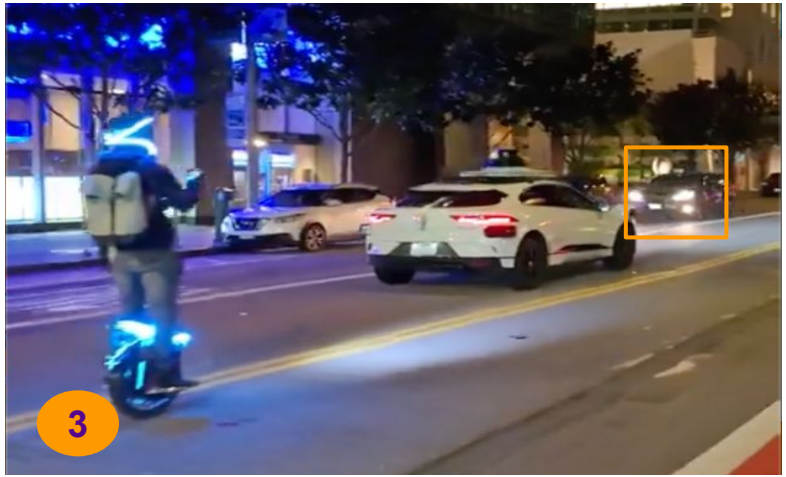


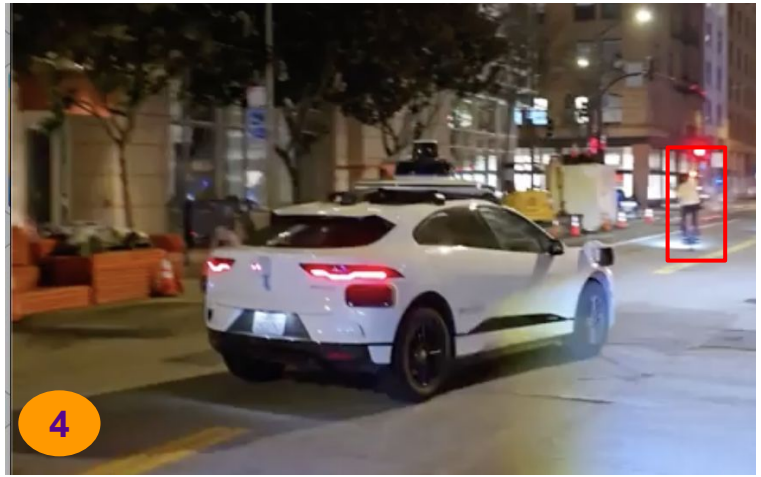
A risk to a hazard with similar or higher severity is caused while trying to avoid one hazard

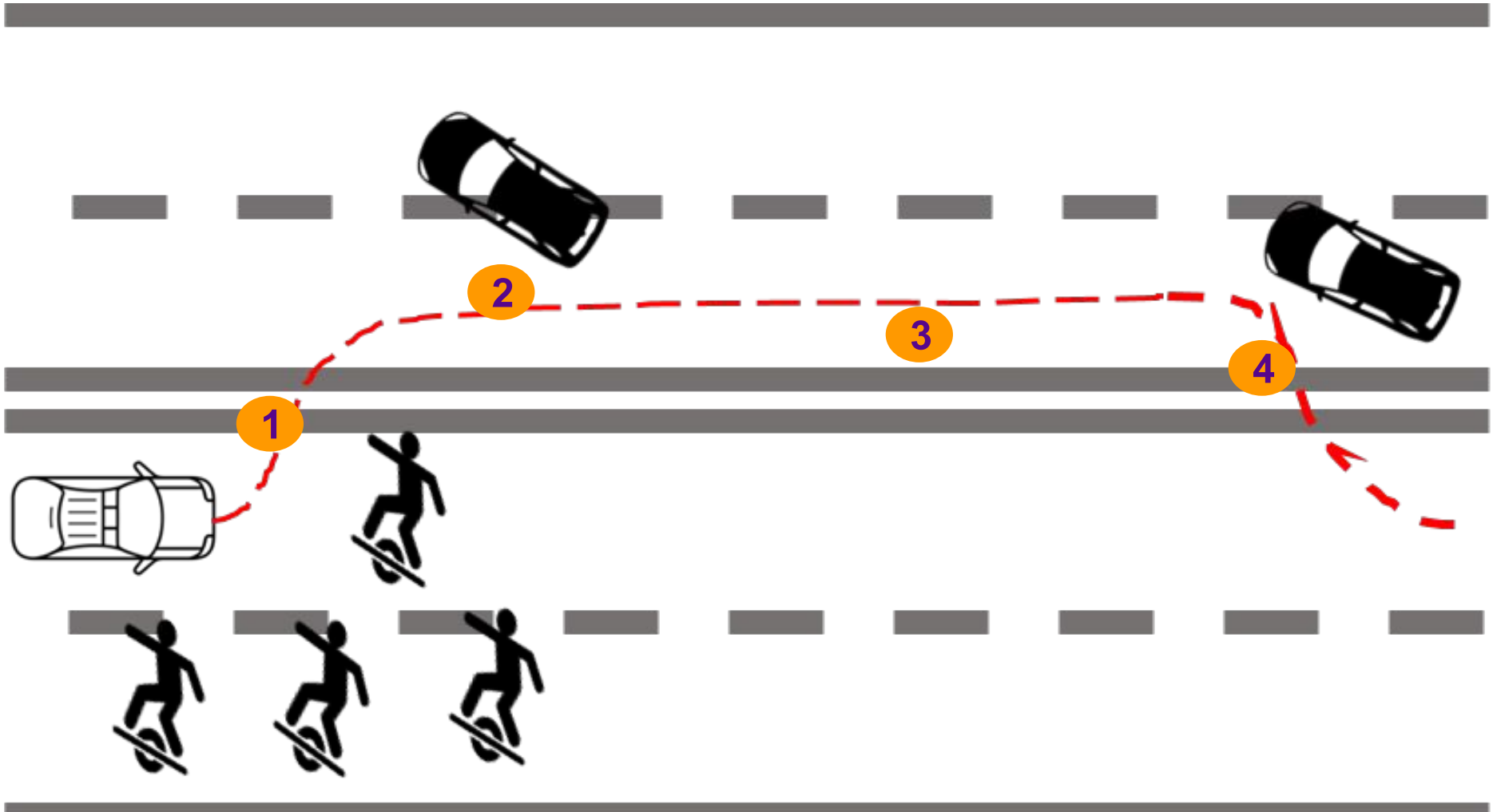
NYU


1


2

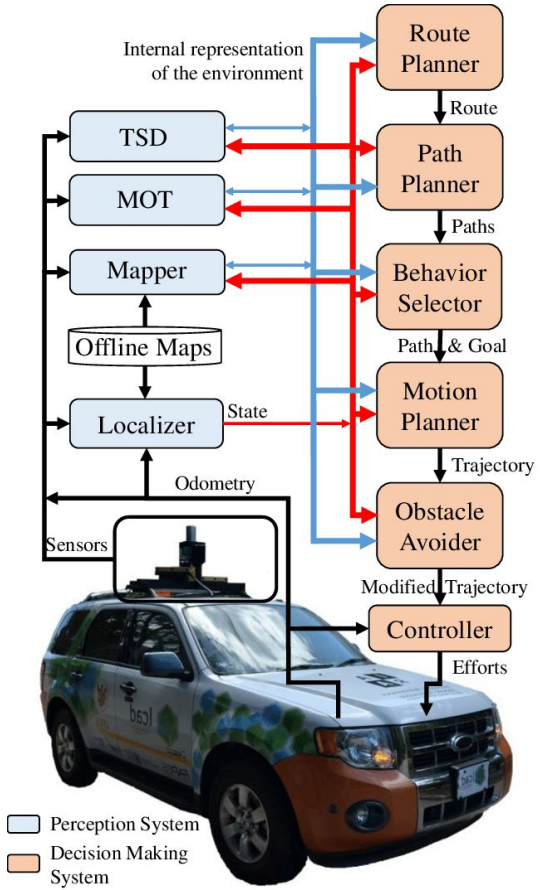

3


4


1
2
3
4

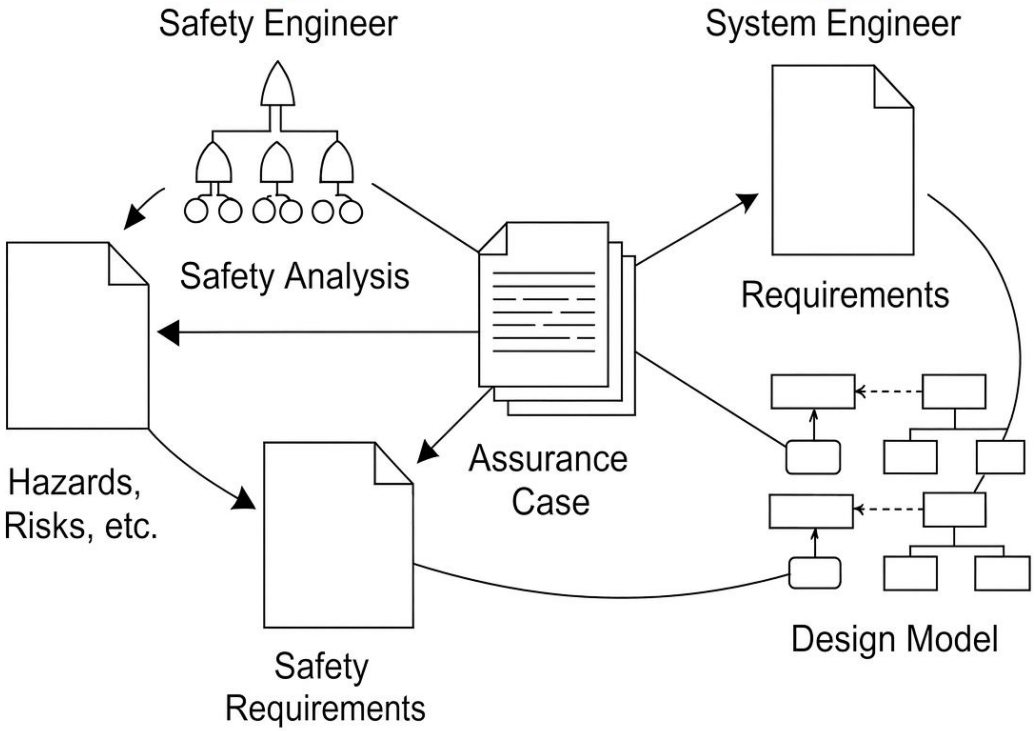
Source of Failure

- Hardware failures (sensor outage, actuator faults)
- Software/algorithmic failures
- ML/AI failures (bias, distribution shift, hallucination)
- Human interaction errors
- Environmental hazards



1.3) Operational Design Domain (ODD)

- Test design
- Scenario
- Coverage


Safety Engineer
System Engineer
Safety Analysis
Requirements
Hazards,
Risks, etc.
Assurance
Case
Safety
Requirements
Design Model

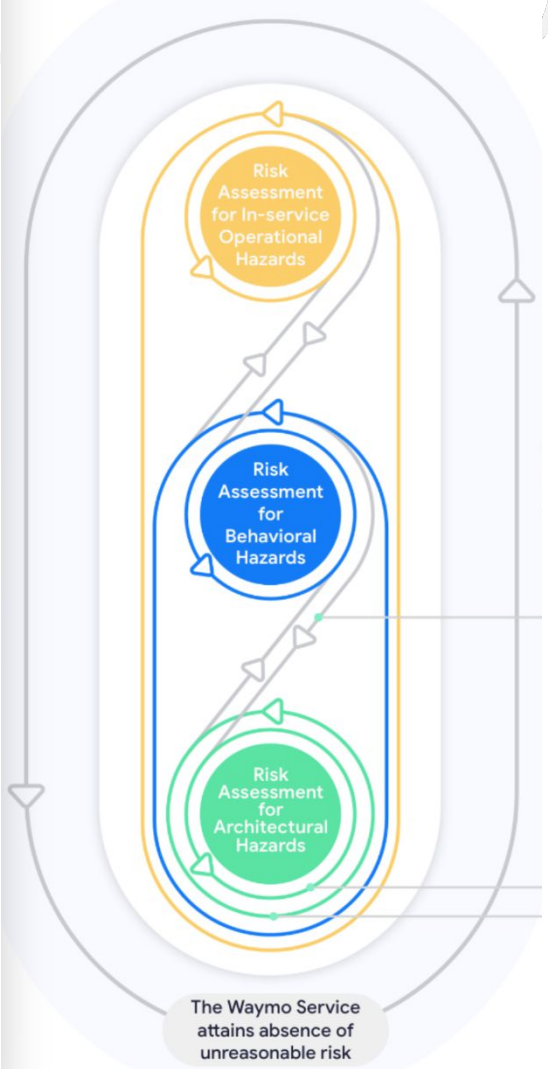
Verification & Validation

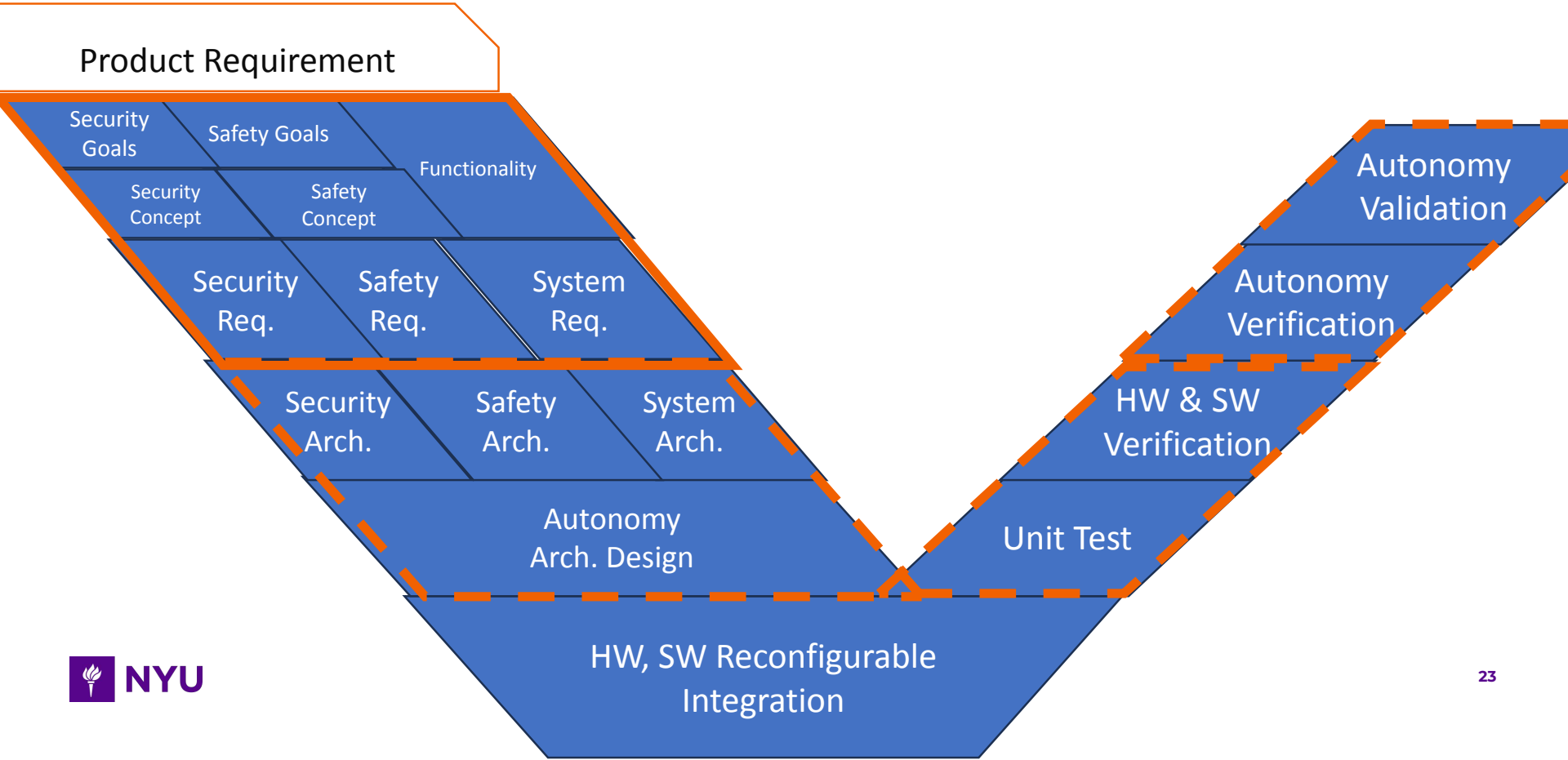
02

NYU

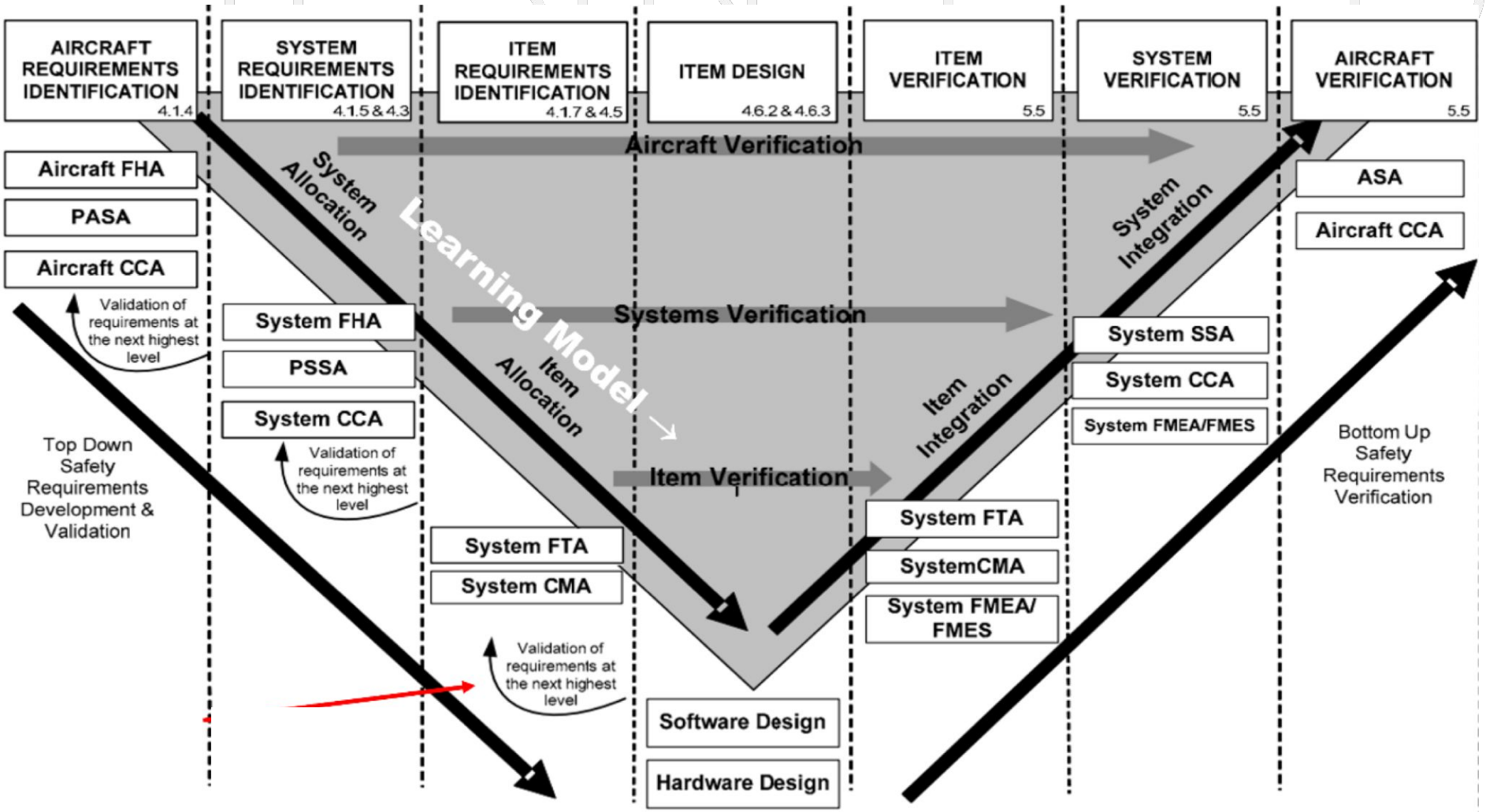
2.1 Verification Approach

1. Identify available tools & techniques
2. Plan how different tools will be combined
3. Choose verification approaches early




Product Requirement
Security Goals
Safety Goals
Functionality
Security Concept
Safety Concept
Security Req.
Safety Req.
System Req.
Security Arch.
Safety Arch.
System Arch.
Autonomy Arch. Design
HW, SW Reconfigurable Integration
Unit Test
HW & SW Verification
Autonomy Verification
Autonomy Validation

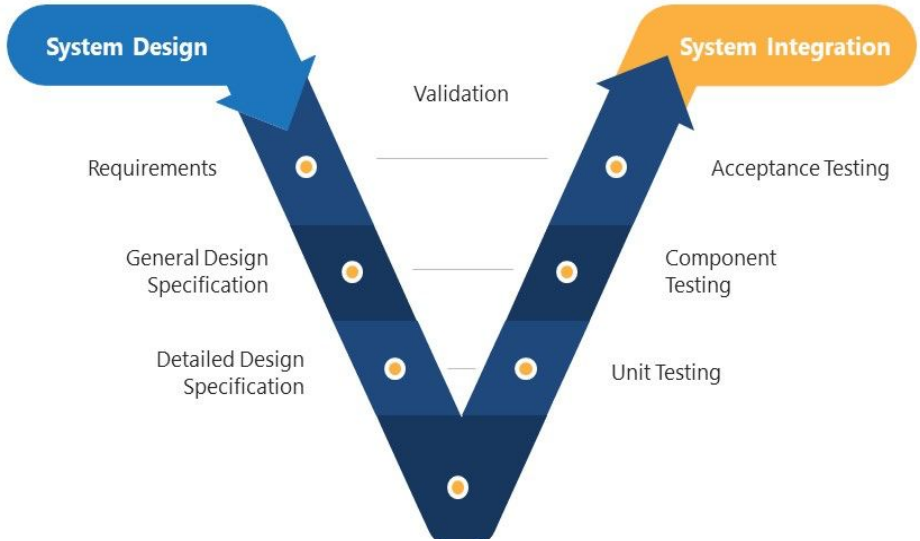
NYU



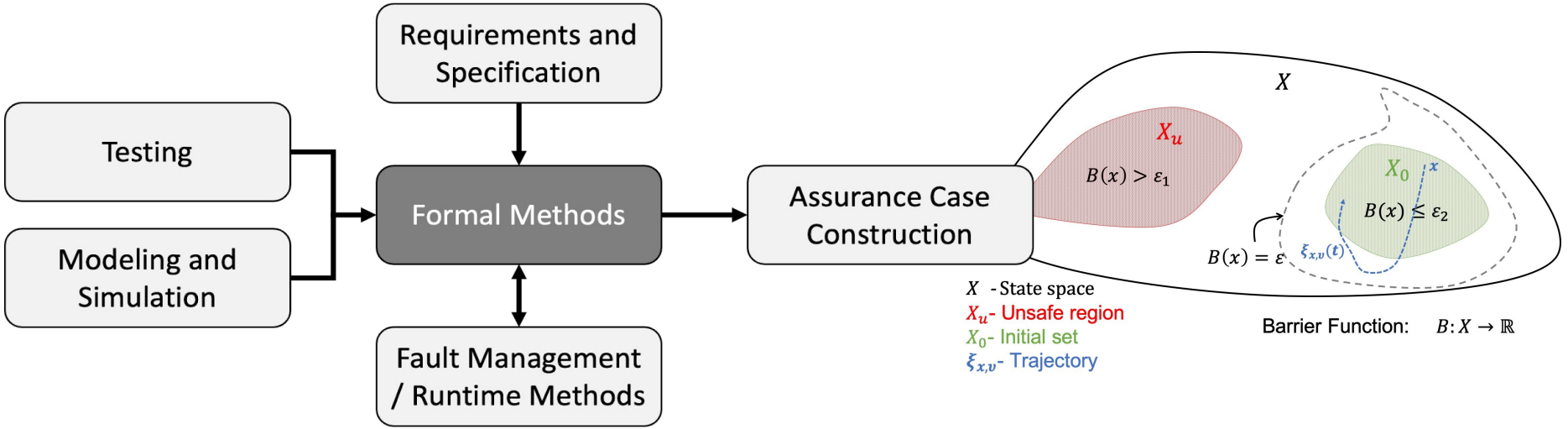
https://www.faa.gov/aircraft/air_cert/step/roadmap_for_AI_safety_assurance

2.2 Verification Process

1. **Establish a Verification Process**
2. Requirements
3. Specifications
4. Testing

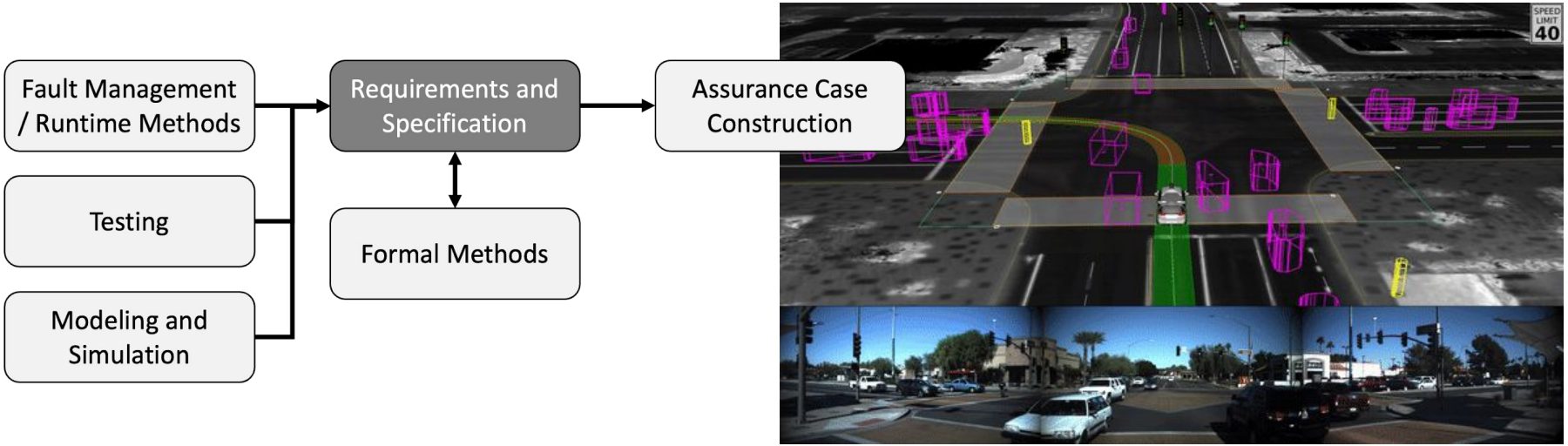


Formal Approach



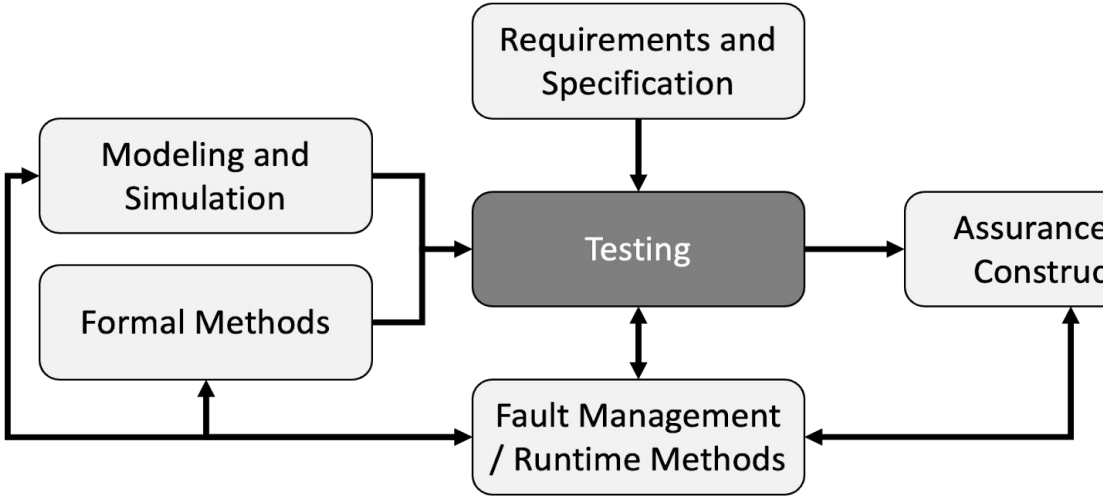
Formal methods are effective when it is possible to simplify an autonomous behavior's interaction with its environment to the point where its analysis is tractable.

Simulations



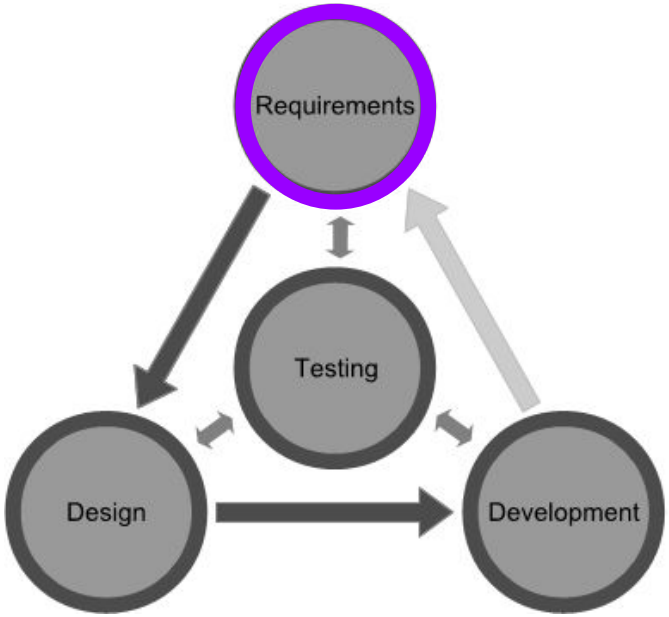
NYU

Physical


Requirements and Specification
Modeling and Simulation
Formal Methods
Testing
Assurance Case Construction
Fault Management / Runtime Methods

2.3 Recruitment Development

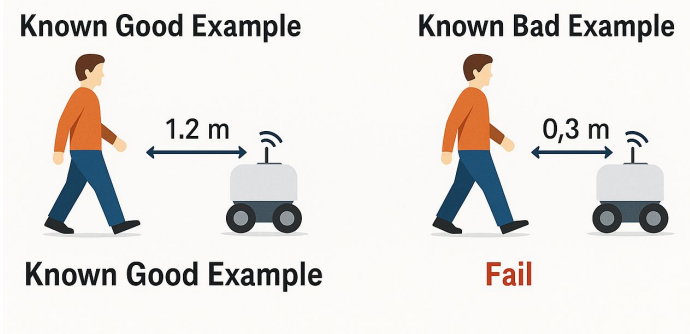
1. stakeholder needs and system-level objectives
2. Allow flexibility and expect revisions
3. Validate requirements with known good and bad examples



Validating Requirement

- **Drafted Requirement:**
  “The AMR shall maintain a safe distance from human.”

- **Validated:**
  “The robot shall maintain a minimum separation distance of ≥ 1.0 m from all detected humans under all operational conditions.”



2.4 Specification

- **Must** clearly describe what the autonomous system should achieve
- Should remain flexible and revisited throughout development
- Separates acceptable from unacceptable system behavior

Description	$h^{(j)}(\boldsymbol{z}, \boldsymbol{q_{hi}}) \geq 0$ condition
Keep distance from humans.	$\rho_{hi} - \rho_0 \geq 0$, if $\lvert\theta_{hi}\rvert < \theta_0$ $\rho_{hi} - \rho_1 \geq 0$, otherwise where $\rho_0 \geq \rho_1$
Yield to humans.	$\frac{d\rho_{hi}}{dt} > 0$, if $\lvert\theta_{hi}\rvert < \theta_0$ and $\rho_{hi} \leq \rho_0$ $\frac{d\rho_{hi}}{dt} > 0$, if $\lvert\theta_{hi}\rvert \geq \theta_0$ and $\rho_{hi} \leq \rho_1$ where $\rho_0 \geq \rho_1$
Limit speed near humans.	$\nu_{max}(\rho_{hi}) - \lvert\boldsymbol{v}\rvert \geq 0$
Limit acceleration near humans.	$a_{max}(\rho_{hi}) - \frac{d\lvert\boldsymbol{v}\rvert}{dt} \geq 0$

Validating Specification

- **Drafted Specification:**
  “The autonomous robot shall maintain a safe stopping distance from obstacles during forward motion.”

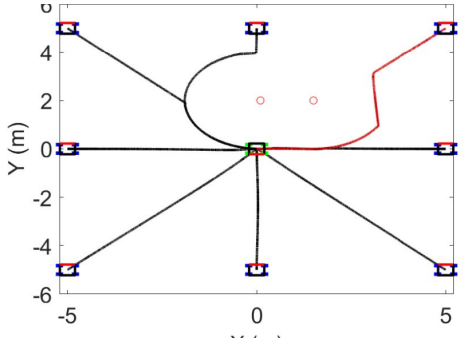
- **Validated:**
  “The robot shall maintain a forward stopping distance ≥ required stopping distance calculated from speed, friction, and braking capability, with a minimum margin of 10%.”

2.4 Testing

- Must focus on identifying erroneous or unexpected behavior
- Sample-based testing is essential for revealing failure cases
- Define stopping criteria before generating test data

“The robot shall maintain ≥ 1.0 m separation from humans and ≥ 0.5 m from other obstacles at all times while moving.”

1. *Static Obstacle*
2. *Human Walking Ahead*
3. *Human Sudden Stop*
4. *Crossing Pedestrian*
5. *Blind Corner / Occlusion*
6. *Multiple Humans / Crowded Area*
7. *Degraded Sensor*
8. *Edge Case – Small Object*



Standards

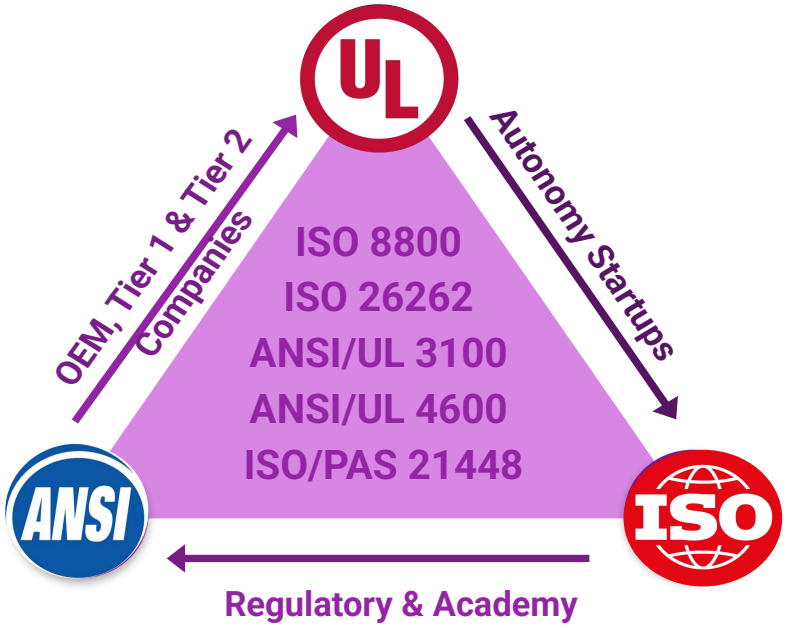
These standards act as a system of "checks and balances"

03

NYU

Standards are Guidelines

No single standard covers everything due to massive complexity of autonomous systems



Key Principles of Safety

Core frameworks as defined in modern autonomy standards

Functional Safety ISO 26262

Ensuring the system behaves safely even when **components fail**.

Example Redundant braking engages if the main controller fails.

:

SOTIF / Intended Functionality ISO 21448

Ensuring safety when there is **no failure**, but the environment or model is uncertain.

Example A camera misclassifies a pedestrian due to unusual lighting.

:

Autonomy Safety UL 4600

Ensuring the entire autonomous “item” (software, hardware, ML, data, ODD) is acceptably safe for deployment.

- ML model uncertainty
- Sensor degradation
- Human–robot interaction
- Out-of-distribution data
- Runtime monitoring

Operational Safety UL 3100 (for AMRs)

Ensuring the **physical equipment** (mobile platforms) is safe under normal and abnormal operation.

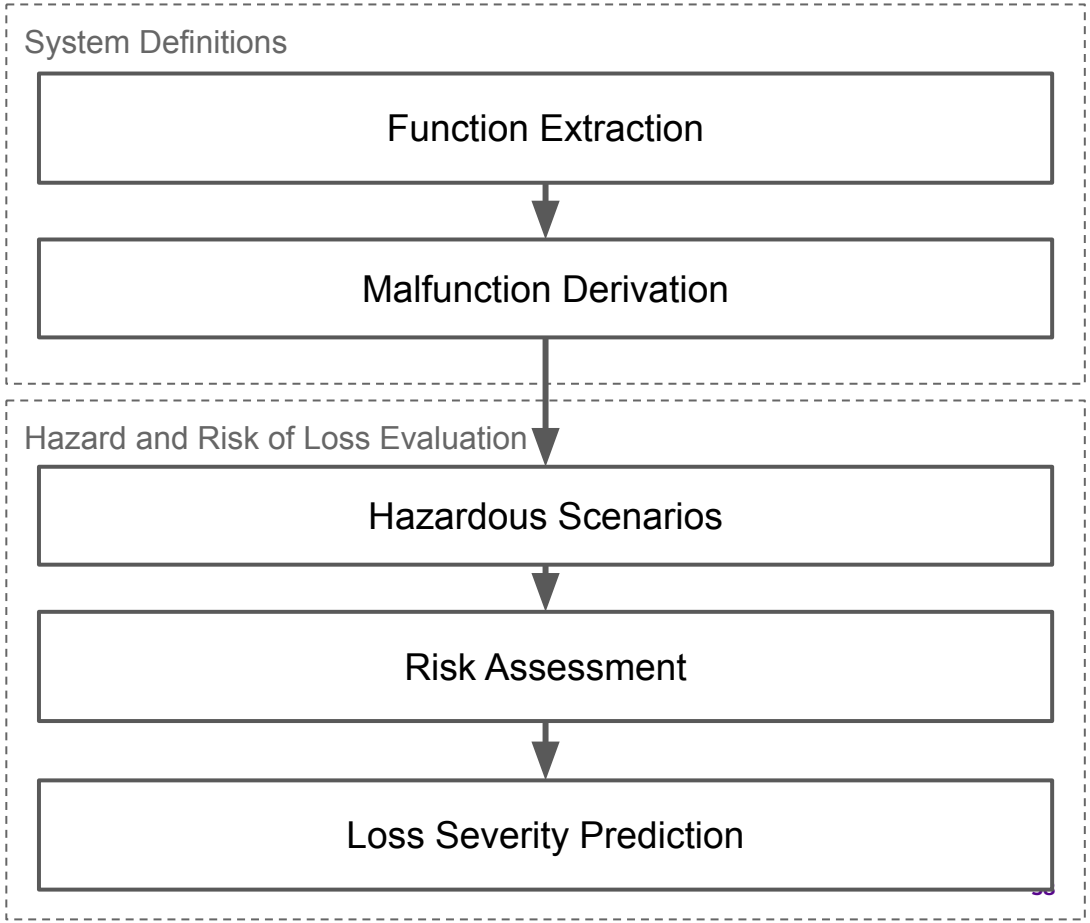
Focus: Mechanical stability, battery safety, and emergency stops.

Safety Frameworks

04

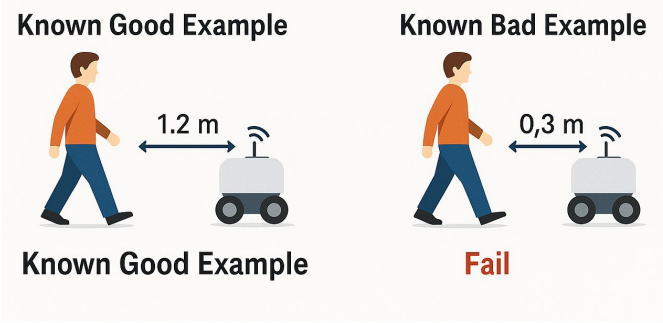
NYU

Hazard Analysis & Risk Assessment



NYU

Example

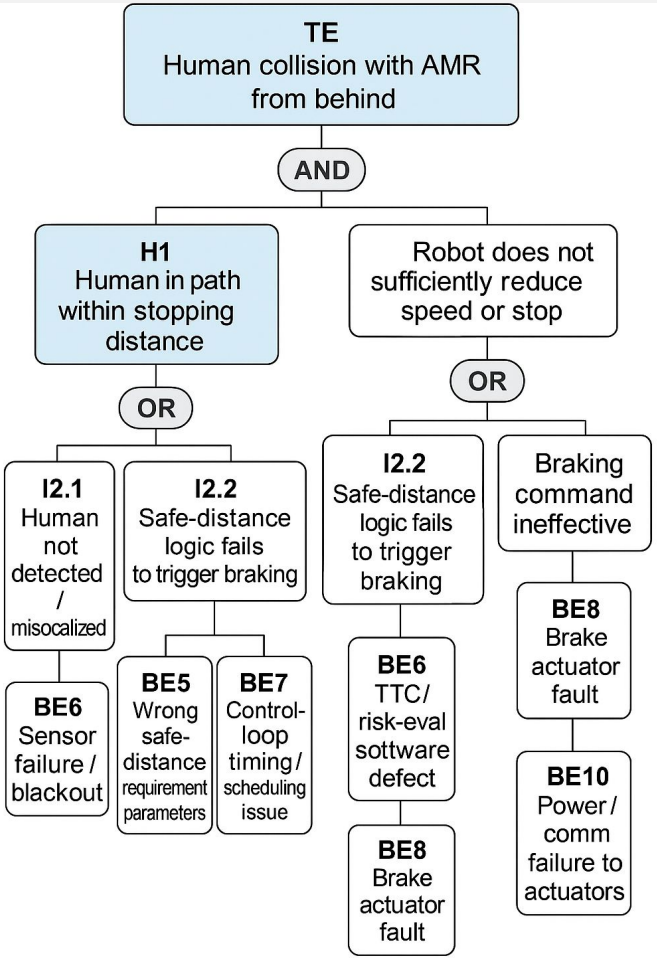


HARA Table – Safe Distance (Autonomous Mobile Robot)

HARA Element	Description
Item / Function	The MR shall mainntain a safe distance fri humans anod obsctacies during forward motion by detecting objects, estimating TTC), and braking or limiting speed to avoid breaching minimum
Malfunction	Underdetected distance or late braking causes the robo̅t reduce rebror below the required threshold. Possible causes, sensor degradation, occlusion, calibration error, software defect, timing /latency issue
Hazardous Scenario	Human walks ahead of the AMR in a narrow aisle, duco ininsufficient safe distance (e.g., robot closes gap to 0.3 m at 0.8 m)
Hazardous Event	S2 – Medium to severe injury (impact may cause falling, ankle/leg injuries, possible fractures)
	E4 – High exposure (following humans in aisles is a common, continuous operational scenario)
Safety Goal (SG)	SG1: Prevent collisions with humans due to insufficient sate distance
	SG2: Maintain ≥ 1,0 m separation from humans, or transition to *safe state* if safe distance cannot be guaranteed

NYU

Fault Tree Analysis (FTA)



NYU

Failure Mode and Effects Analysis (FMEA)

- Looks for consequences of component failures (forward chaining technique)
- Limitation: requires expert analysis to decide what to analyze

System-Theoretic Process Analysis (STPA)

- It analyzes how a system's control structure can lead to accidents by identifying "unsafe control actions," which are inadequate or missing controls.
  a. Not providing a required control action
  b. Providing a control action that causes a hazard
  c. Providing a correct action but at the wrong time or in the wrong order
  d. Providing a control action for too long or stopping it too soon

HARA Workshop

1) Operation
2) Behaviour
3) Architecture

05

NYU

20 Minutes Brain storming session

- Write 3-5 Requirements
  a. Formalize functions
  b. Extract malfunctions
  c. Layout hazardous (clarify worst-case) scenarios
  d. Discuss risks based on
     - Severity
     - Exposure
     - Controllability

Constraint Priotrization

Prioritizing risks on a public road with all different perspectives is really challenging.

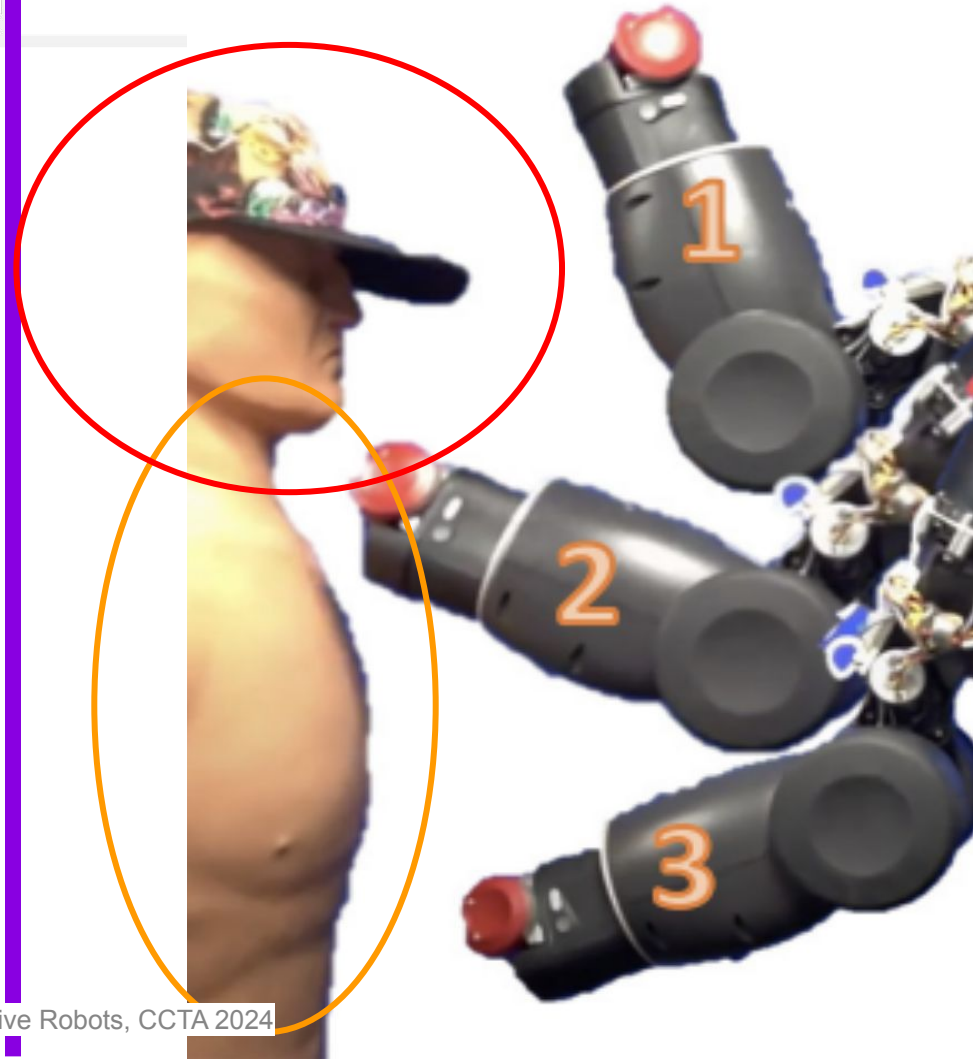
06

NYU

Identifying Priorities

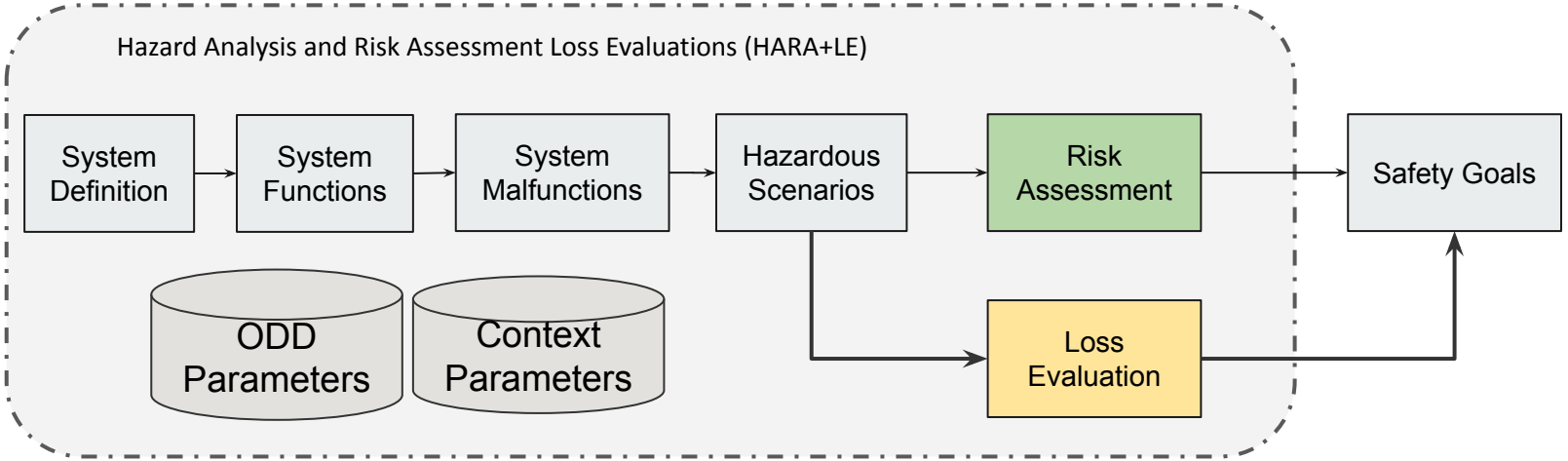
Identifying Safety Priorities is Challenging and makes it a Socio-Technological Problem

- Safety is very Subjective and can very from a Scenario to another
- Prioritization is Necessary but Needs Proper Safety Indicators and Mechanisms to Address Them
- Addressing Ethical Aspects w



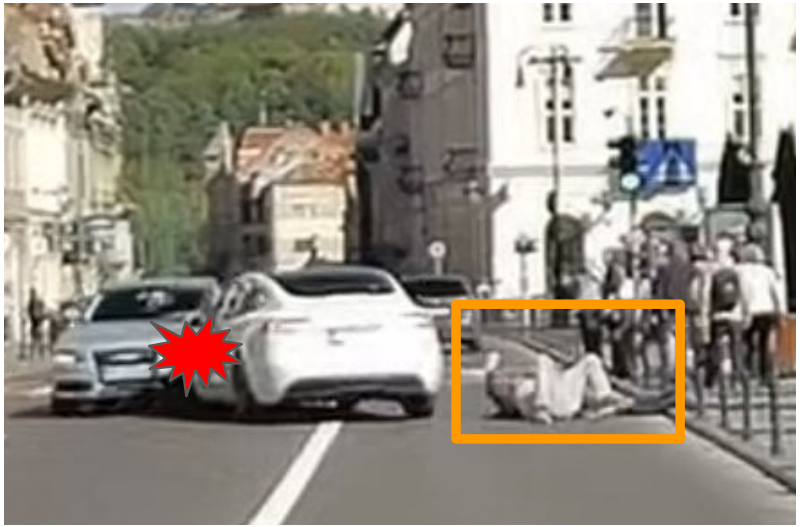
Safety Prioritization by Iterative Feedback Linearization Control for Collaborative Robots, CCTA 2024

Identifying Priorities



High-resolution safety verification for evasive obstacle avoidance in autonomous vehicles, IEEE Open Journal of Vehicular Technology

Probability of Something Goes Wrong?



Safety Goals


“
Designing a Verification & Validation Strategy is as Challenging as System Design is.
NYU
50

Thanks!

Aliasghar Arab, PhD

Research Professor @ NYU | Safe AI & Robotics

Email	aliasghar.arab@nyu.edu
ASAS Labs	www.asas-labs.com
Personal Page	www.aaarab.com
LinkedIn	linkedin.com/in/aliasghar-arab
Company Page	www.genauto.ai

IEEE Senior Member

SAE

UL Standards & Engagement

IEEE Robotics & Automation Society

TRB

IEEE VTS Connecting the Mobile World

Association for the Advancement of Artificial Intelligence

ASME