Dagstuhl Seminar 26051: User-Aligned Assessment of AI Systems

When Robots Come Home: Who Models Safety in Personal Robotics?

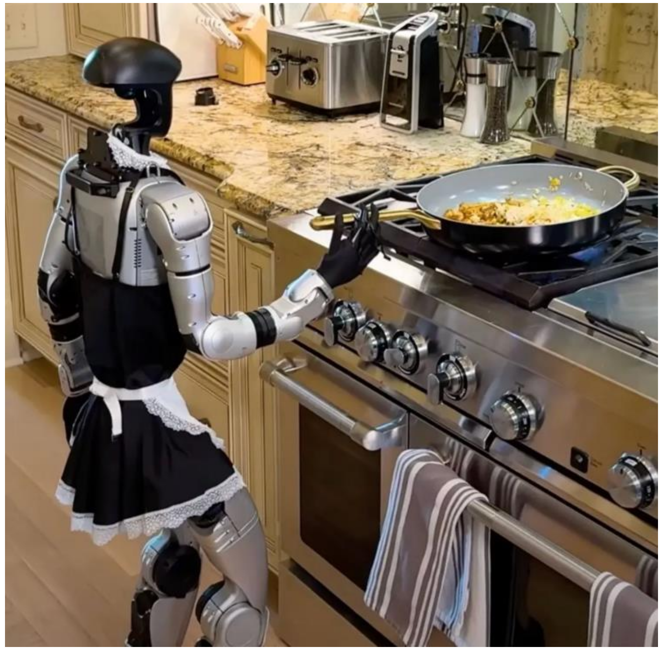
Pulkit Verma

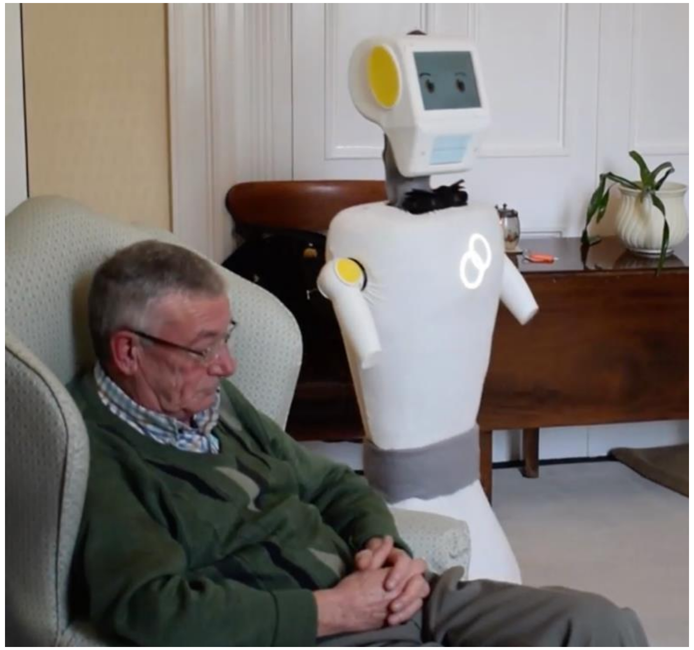
Department of Computer Science and Engineering

Indian Institute of Technology Madras, India

Secure AI Futures Lab

Robots that adapt to YOUR home, YOUR tasks, YOUR preferences





When is it safe for THIS user to operate THIS robot on THIS task in THIS environment?

```
at(p0,cell_6_3)
clear(cell_0_0)
door_at(cell_9_2)
next_to(m0)
alive(m0)
key_at(9_4)
```

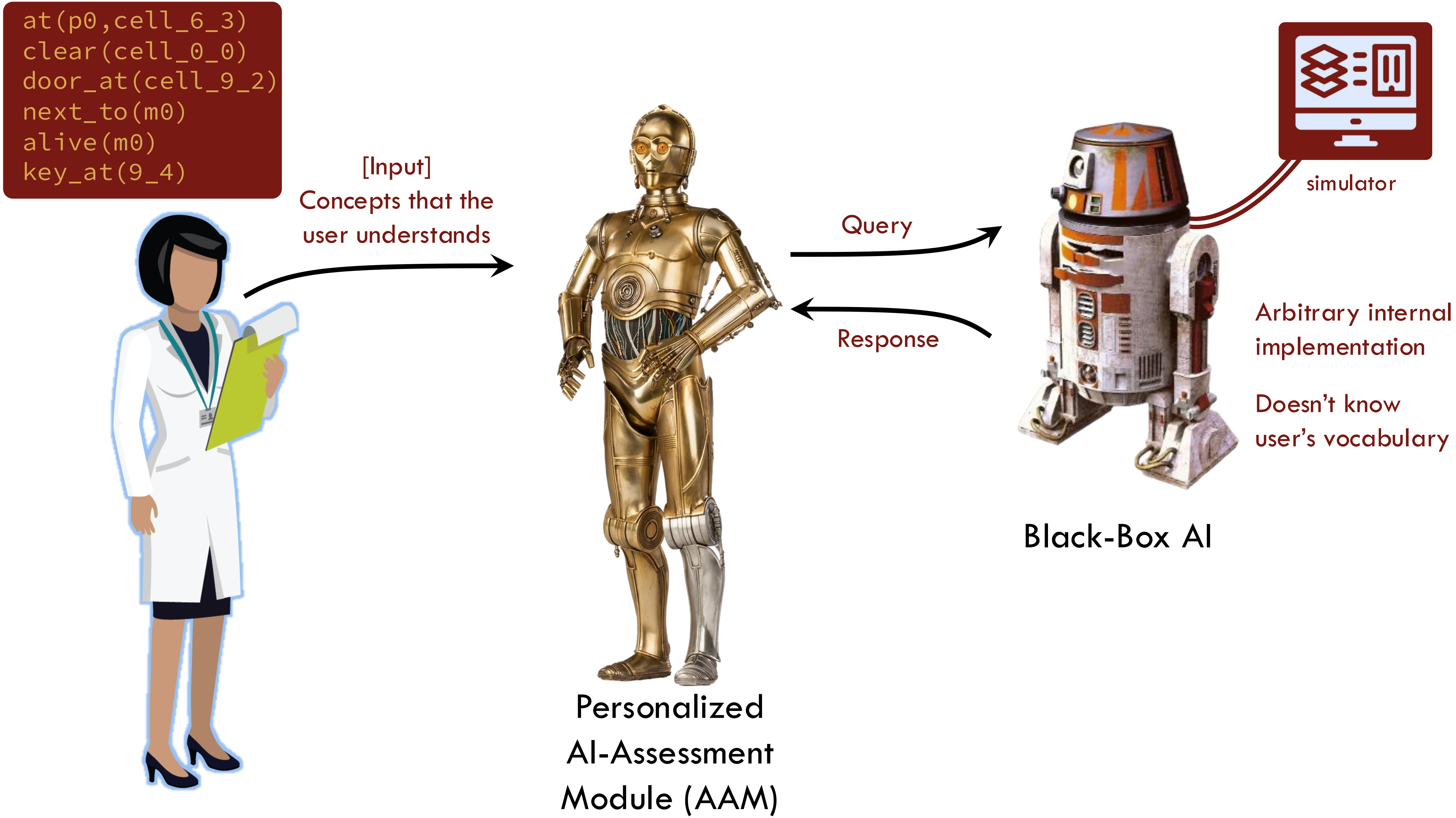
[Input]
Concepts that the user understands

Personalized AI-Assessment Module (AAM)

Arbitrary internal implementation

Doesn't know user's vocabulary

Black-Box AI



```
at(p0,cell_6_3)
clear(cell_0_0)
door_at(cell_9_2)
next_to(m0)
alive(m0)
key_at(9_4)
```

[Input]
Concepts that the
user understands

Query

Response

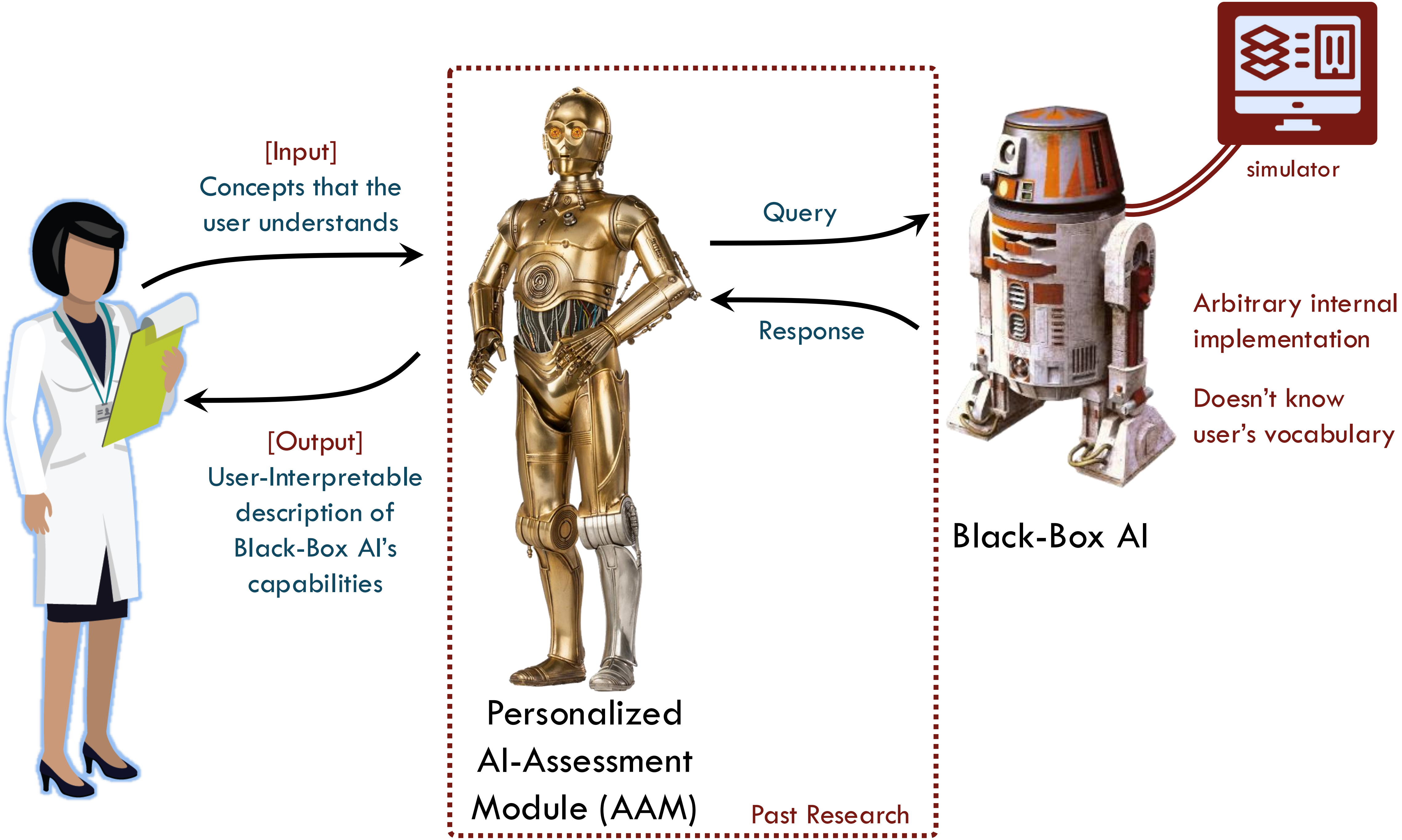
simulator

Arbitrary internal
implementation

Doesn't know
user's vocabulary

Black-Box AI

Personalized
AI-Assessment
Module (AAM)



[Input]
Concepts that the user understands

[Output]
User-Interpretable description of Black-Box AI's capabilities

Personalized AI-Assessment Module (AAM)

Past Research

Query

Response

simulator

Black-Box AI

Arbitrary internal implementation

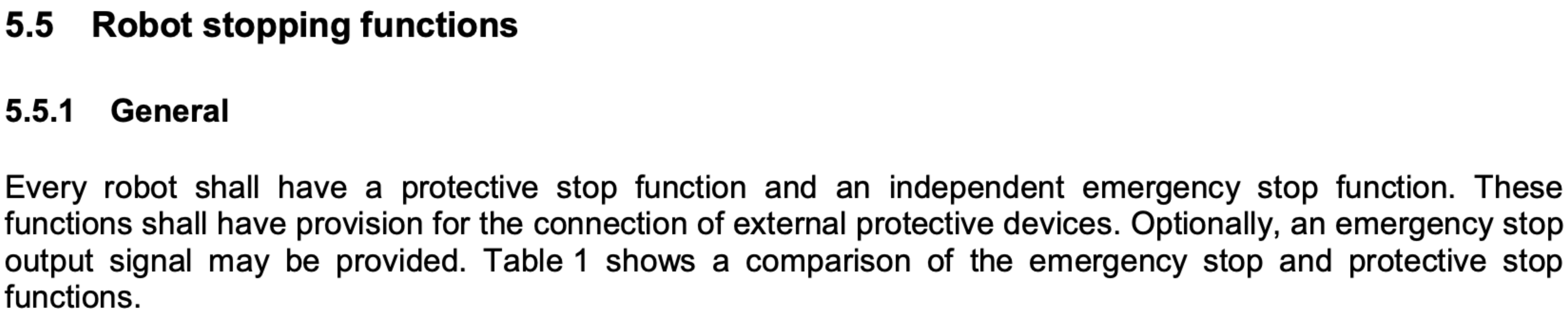
Doesn't know user's vocabulary

Why Traditional Assessment Fails - The Standards Landscape

Who loves AI?
Who is willing to deal with the ethical guidelines of AI?
Who is willing to take responsibility for the decisions of the algorithms?

Why Traditional Assessment Fails - The Standards Landscape

Domain	Standard	Scope	Approach
Factory Robots	ISO 10218 (2011)	Industrial manipulators	Physical barriers, safeguarded spaces



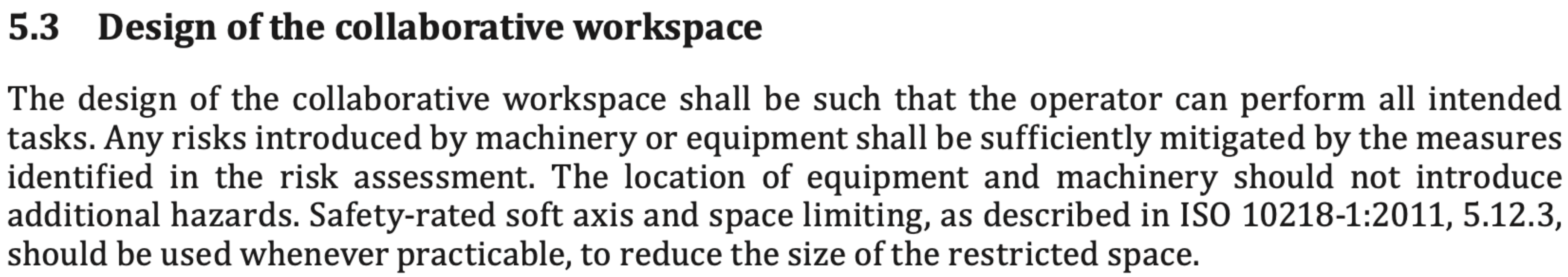
5.5 Robot stopping functions

5.5.1 General

Every robot shall have a protective stop function and an independent emergency stop function. These functions shall have provision for the connection of external protective devices. Optionally, an emergency stop output signal may be provided. Table 1 shows a comparison of the emergency stop and protective stop functions.

Why Traditional Assessment Fails - The Standards Landscape

Domain	Standard	Scope	Approach
Factory Robots	ISO 10218 (2011)	Industrial manipulators	Physical barriers, safeguarded spaces
Collaborative Robots	ISO/TS 15066 (2016)	Human-robot collaboration in industry	Force/pressure limits, SSM, PFL modes

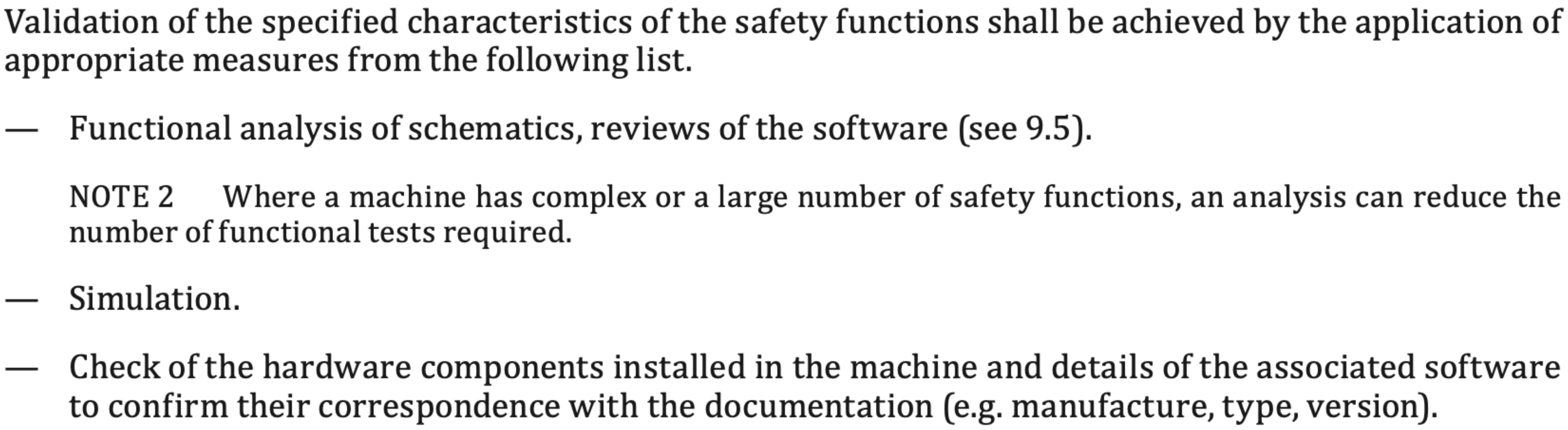


5.3 Design of the collaborative workspace

The design of the collaborative workspace shall be such that the operator can perform all intended tasks. Any risks introduced by machinery or equipment shall be sufficiently mitigated by the measures identified in the risk assessment. The location of equipment and machinery should not introduce additional hazards. Safety-rated soft axis and space limiting, as described in ISO 10218-1:2011, 5.12.3, should be used whenever practicable, to reduce the size of the restricted space.

Why Traditional Assessment Fails - The Standards Landscape

Domain	Standard	Scope	Approach
Factory Robots	ISO 10218 (2011)	Industrial manipulators	Physical barriers, safeguarded spaces
Collaborative Robots	ISO/TS 15066 (2016)	Human-robot collaboration in industry	Force/pressure limits, SSM, PFL modes
Autonomous Vehicles	ISO 3691-4 (2020)	AGVs, mobile platforms	Person detection, stability tests



Validation of the specified characteristics of the safety functions shall be achieved by the application of appropriate measures from the following list.

— Functional analysis of schematics, reviews of the software (see 9.5).

NOTE 2 Where a machine has complex or a large number of safety functions, an analysis can reduce the number of functional tests required.

— Simulation.

— Check of the hardware components installed in the machine and details of the associated software to confirm their correspondence with the documentation (e.g. manufacture, type, version).

Why Traditional Assessment Fails - The Standards Landscape

Domain	Standard	Scope	Approach
Factory Robots	ISO 10218 (2011)	Industrial manipulators	Physical barriers, safeguarded spaces
Collaborative Robots	ISO/TS 15066 (2016)	Human-robot collaboration in industry	Force/pressure limits, SSM, PFL modes
Autonomous Vehicles	ISO 3691-4 (2020)	AGVs, mobile platforms	Person detection, stability tests
Medical Robots	IEC 80601-2-78 (2020)	Rehabilitation (RACA robots)	ROM limits, torque constraints

Why Traditional Assessment Fails - The Standards Landscape

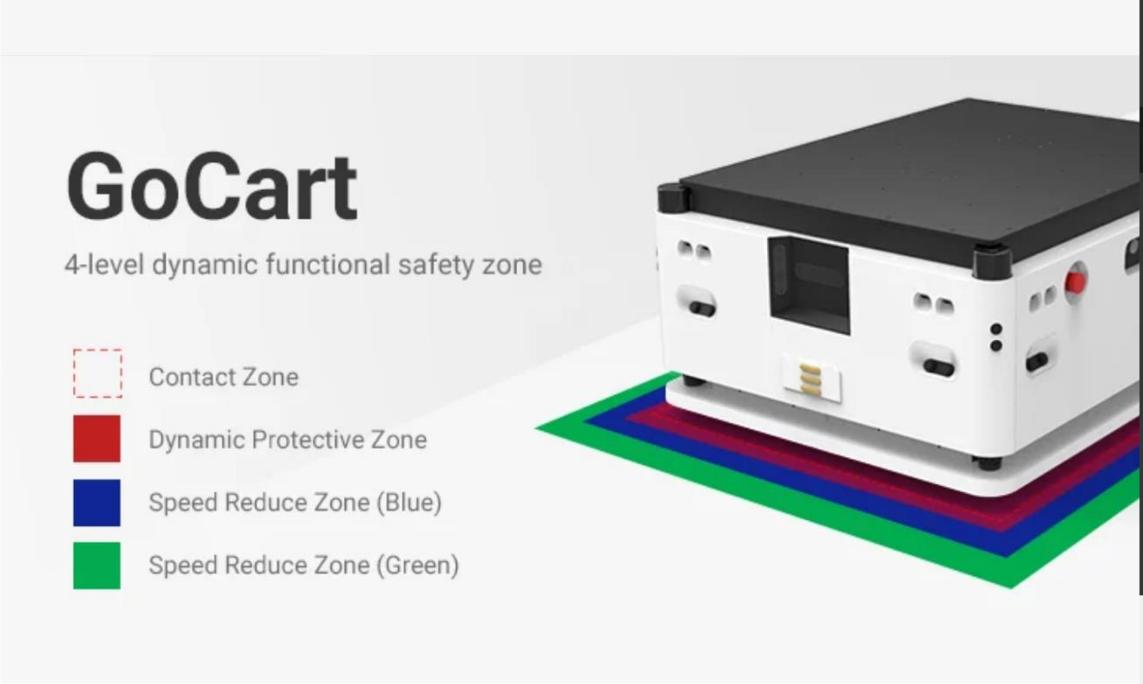
Domain	Standard	Scope	Approach
Factory Robots	ISO 10218 (2011)	Industrial manipulators	Physical barriers, safeguarded spaces
Collaborative Robots	ISO/TS 15066 (2016)	Human-robot collaboration in industry	Force/pressure limits, SSM, PFL modes
Autonomous Vehicles	ISO 3691-4 (2020)	AGVs, mobile platforms	Person detection, stability tests
Medical Robots	IEC 80601-2-78 (2020)	Rehabilitation (RACA robots)	ROM limits, torque constraints
Personal Care Robots	ISO 13482 (2014) + TR 23482-1 (2020)	Hospitals and Similar Settings, Homes??	Operational spaces, test methods

Why Traditional Assessment Fails - The Standards Landscape

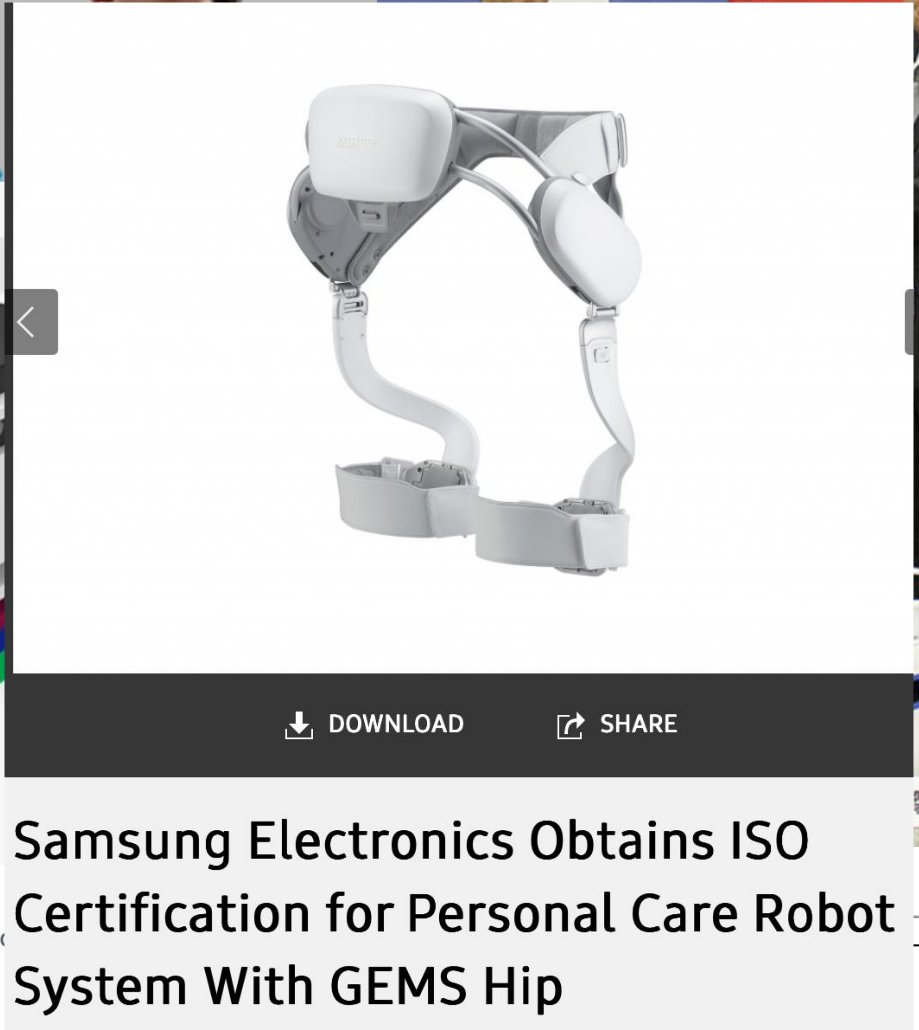
Domain	Star
Factory Robots	ISO
Collaborative Robots	ISO

es

des



In January 2021, Yujin Robot's GoCart, an autonomous mobile robot for logistics, personal service, and applications, received Korea's first ISO 13482 safety certification.



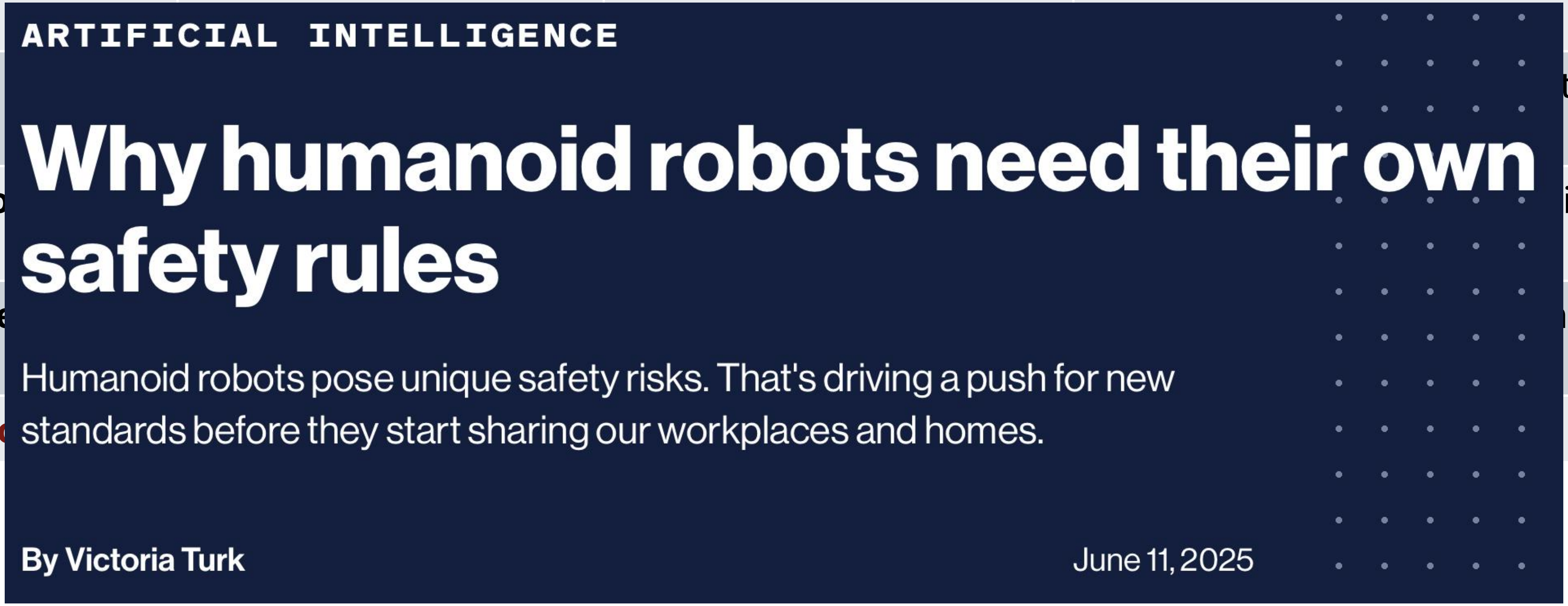
Samsung Electronics Obtains ISO Certification for Personal Care Robot System With GEMS Hip

Why Traditional Assessment Fails - The Standards Landscape

Domain	Standard	Scope	Approach
Factory Robots	ISO 10218 (2011)	Industrial manipulators	Physical barriers, safeguarded spaces
Collaborative Robots	ISO/TS 15066 (2016)	Human-robot collaboration in industry	Force/pressure limits, SSM, PFL modes
Autonomous Vehicles	ISO 3691-4 (2020)	AGVs, mobile platforms	Person detection, stability tests
Medical Robots	IEC 80601-2-78 (2020)	Rehabilitation (RACA robots)	ROM limits, torque constraints
Personal Care Robots	ISO 13482 (2014) + TR 23482-1 (2020)		Operational spaces, test methods
Household Robots	??	??	

Why Traditional Assessment Fails - The Standards Landscape

Domain	Standard	Scope	Approach
Factory Robots	ISO 10218 (2011)	Industrial manipulators	Physical barriers, safeguarded spaces
Collaborative Robots	ISO/TS 15066 (2016)	Human-robot	Force/pressure limits, SSM, PFL modes
Autonomous Vehicles			tests
Medical Robo			ints
Personal Care			ethods
Household R			



ARTIFICIAL INTELLIGENCE

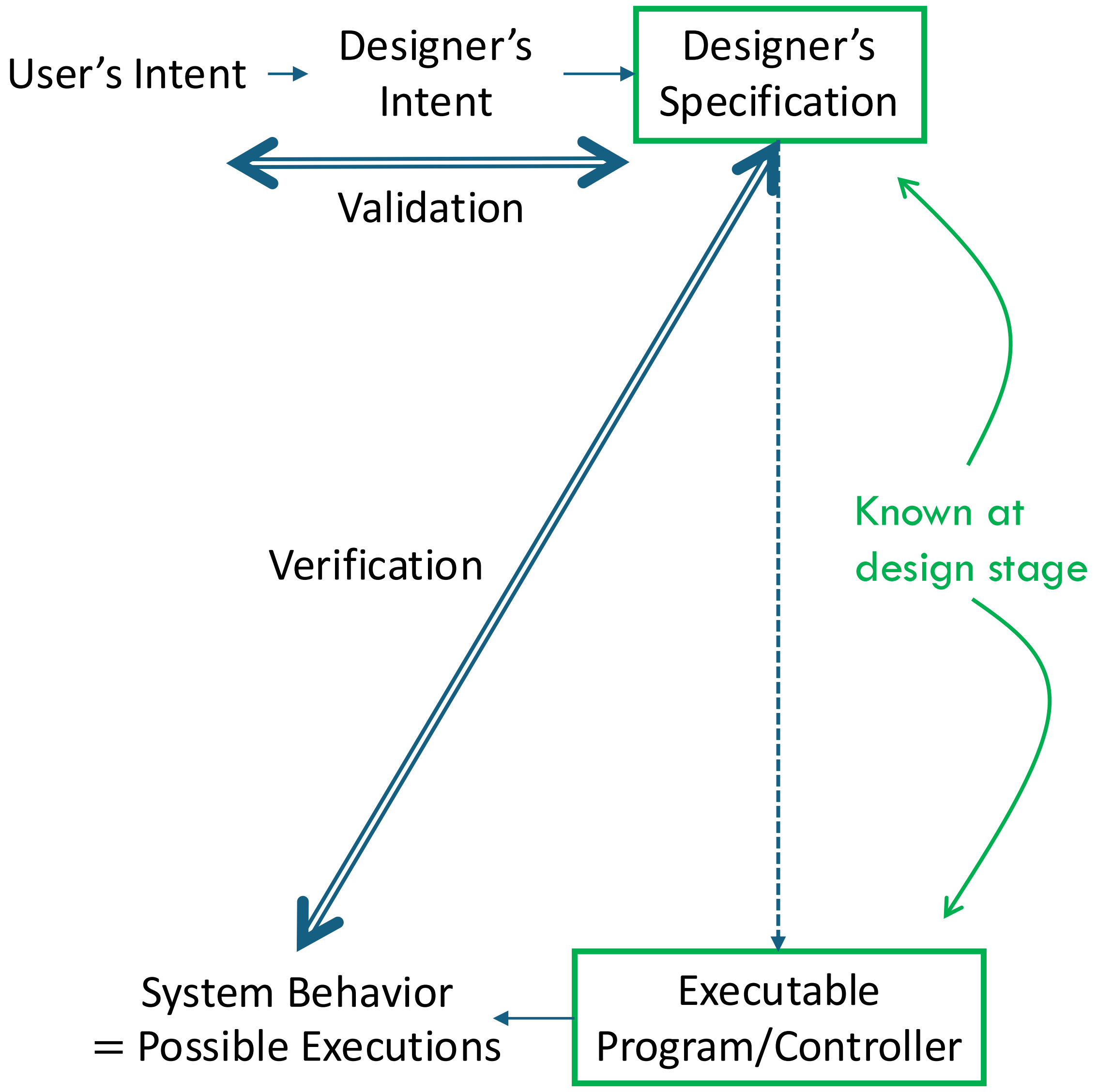
Why humanoid robots need their own safety rules

Humanoid robots pose unique safety risks. That's driving a push for new standards before they start sharing our workplaces and homes.

By Victoria Turk

June 11, 2025

Conventional Verification



User's Intent → Designer's Intent → Designer's Specification

Validation

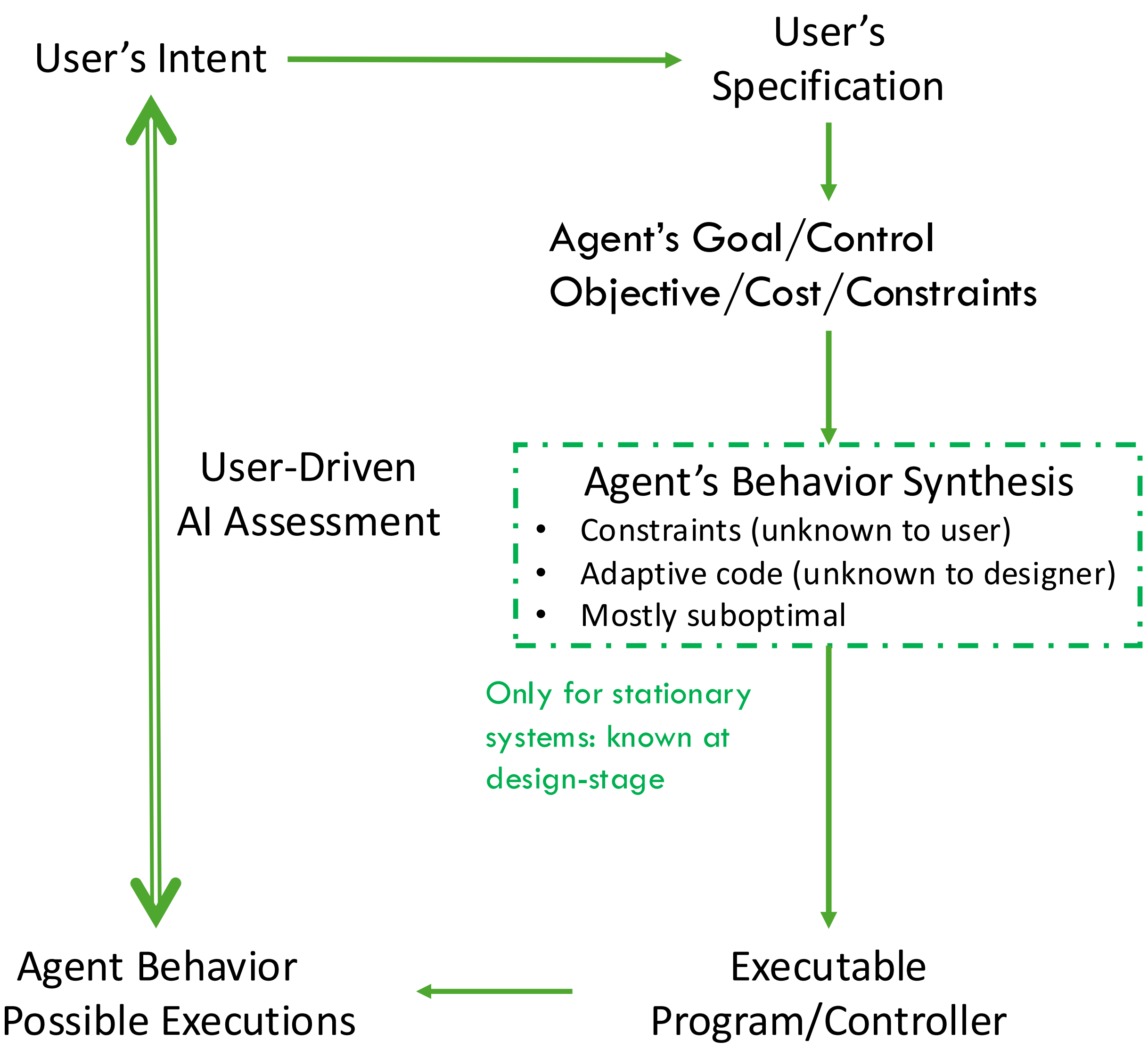
Verification

Known at design stage

System Behavior = Possible Executions

Executable Program/Controller

Needed for AI Systems



User's Intent → User's Specification

Agent's Goal/Control Objective/Cost/Constraints

Agent's Behavior Synthesis

- Constraints (unknown to user)
- Adaptive code (unknown to designer)
- Mostly suboptimal

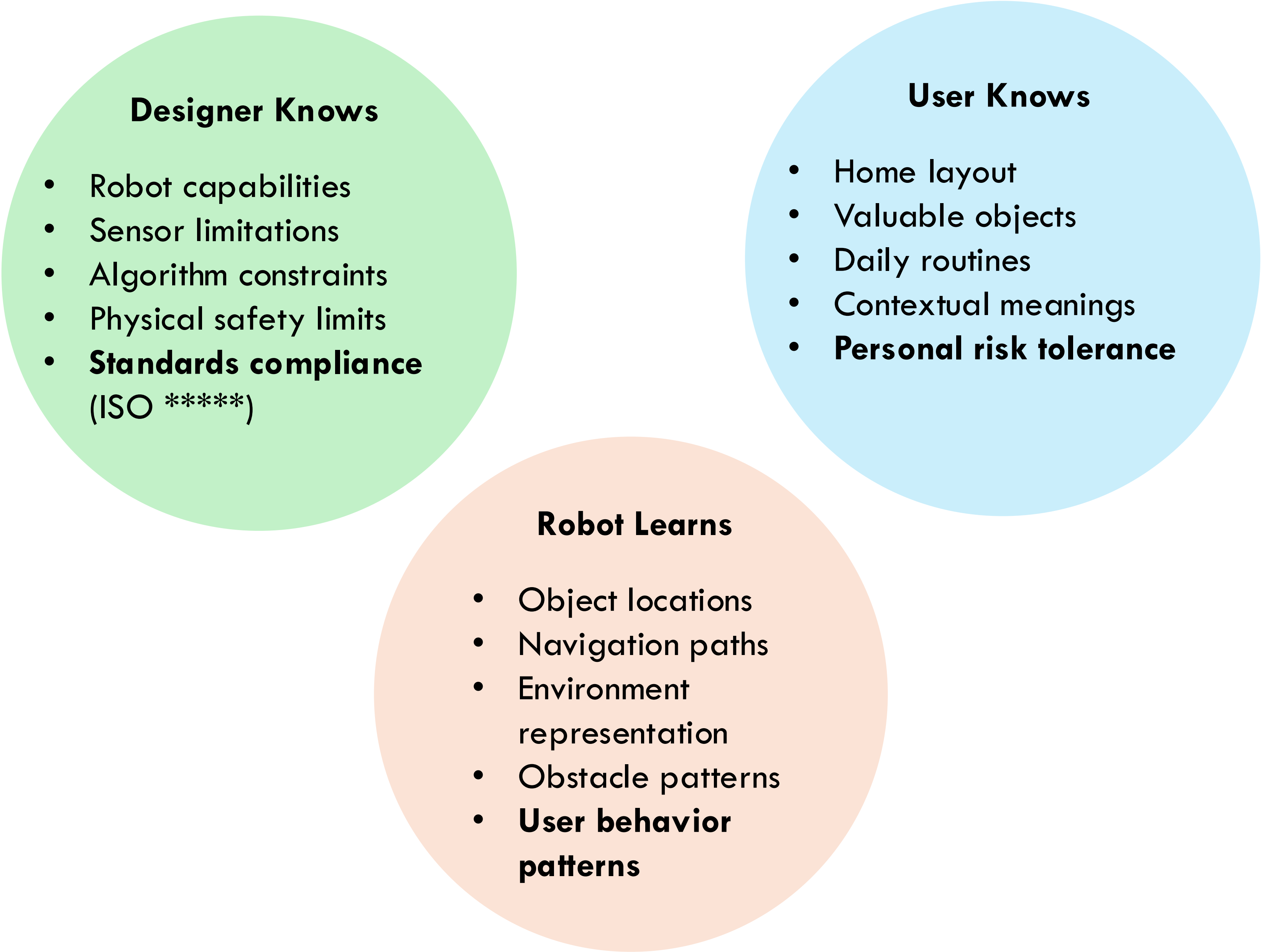
Only for stationary systems: known at design-stage

User-Driven AI Assessment

Agent Behavior = Possible Executions

Executable Program/Controller

Knowledge Fragmentation



Designer Knows

- Robot capabilities
- Sensor limitations
- Algorithm constraints
- Physical safety limits
- **Standards compliance** (ISO *****)

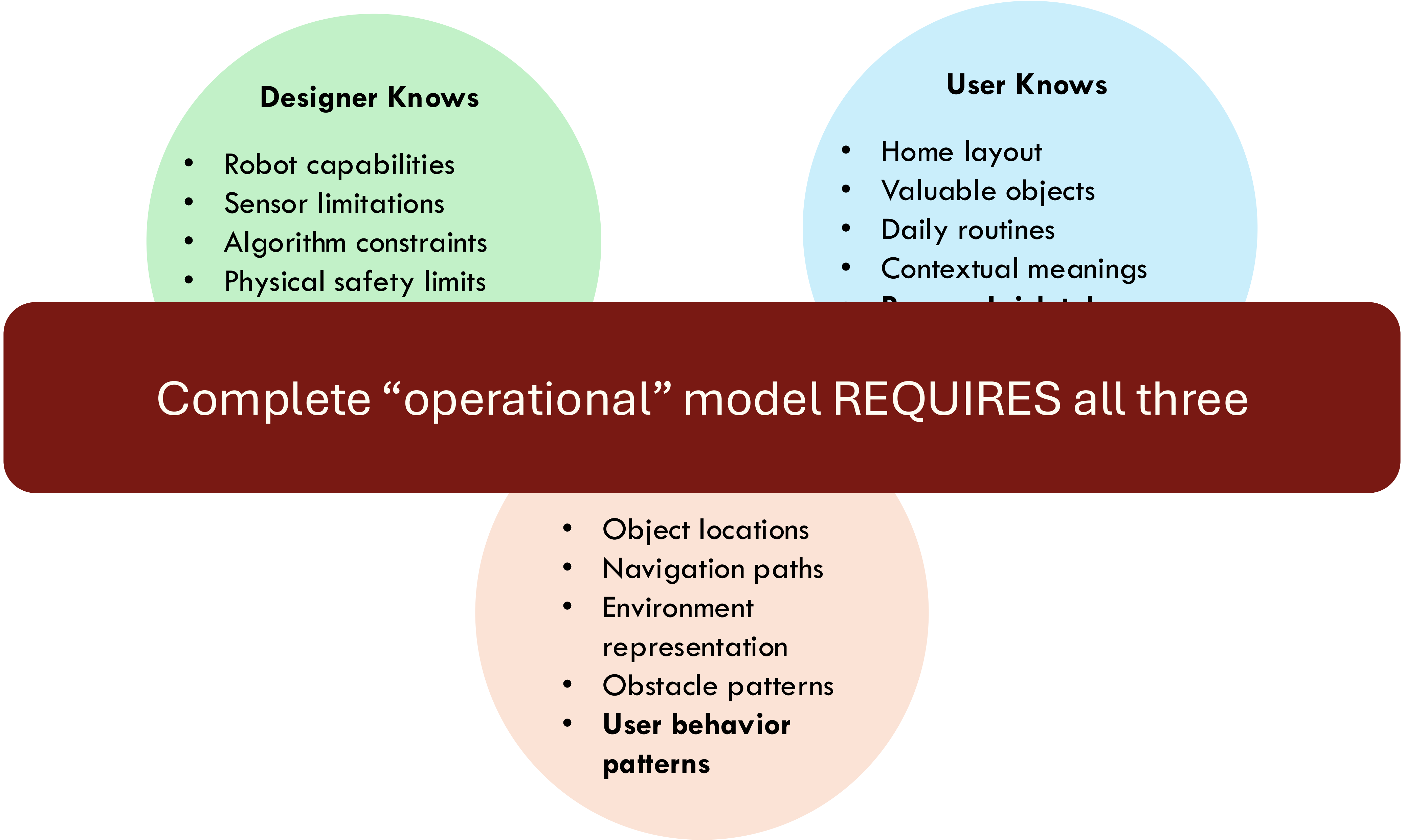
User Knows

- Home layout
- Valuable objects
- Daily routines
- Contextual meanings
- **Personal risk tolerance**

Robot Learns

- Object locations
- Navigation paths
- Environment representation
- Obstacle patterns
- **User behavior patterns**

Knowledge Fragmentation



Designer Knows

- Robot capabilities
- Sensor limitations
- Algorithm constraints
- Physical safety limits

User Knows

- Home layout
- Valuable objects
- Daily routines
- Contextual meanings
- [illegible]

Complete “operational” model REQUIRES all three

- Object locations
- Navigation paths
- Environment representation
- Obstacle patterns
- **User behavior patterns**

Knowledge Fragmentation

User Thinks

- “Carefully” = don't disturb my papers on the floor
- “Living room” = the room with the couch
- Values the antique rug

Robot Interprets

- “Carefully” = slower speed?
- “Living room” = room labeled “living” in map
- Papers = unknown objects to avoid

Designer Verified (ISO 13842)

- Maximum speed ≤ 2 m/s
- Protective stop distance requirements
- 85 hazard scenarios
- **But NOT context-specific constraints**

ISO/TS 15066 provides (for collaborative robots)

- Force limits for transient contact (200N)
- Pressure limits (110 N/cm^2)
- Speed-and-separation monitoring formulas
- **But assumes industrial context, known tasks**

IEC 80601-2-78 provides (for medical robots)

- Joint torque limits for therapy
- Range-of-motion boundaries
- Misalignment detection
- **But assumes a clinical setting, trained operators**

What none Provides

- Methodology for interpreting “carefully”
- User-specific risk assessment
- Task-specific safety properties

Distributed
World
Modeling

Bidirectional
Intent
Alignment

Dynamic
Safety
Property
Generation

Capability
Transparency

Distributed World Modeling

Who builds the operational model?

ISO 13482 defines operational spaces:

- Maximum space, Restricted space
- Monitored space, Safeguarded space
- Protective stop space

ISO/TS 15066 defines collaborative spaces:

- Safeguarded space (no contact)
- Collaborative workspace (monitored)
- Human detection zones

IEC 80601-2-78 defines movement constraints:

- Pre-set ROM limits per joint
- User-specific boundaries
- Misalignment detection zones

Operational Spaces

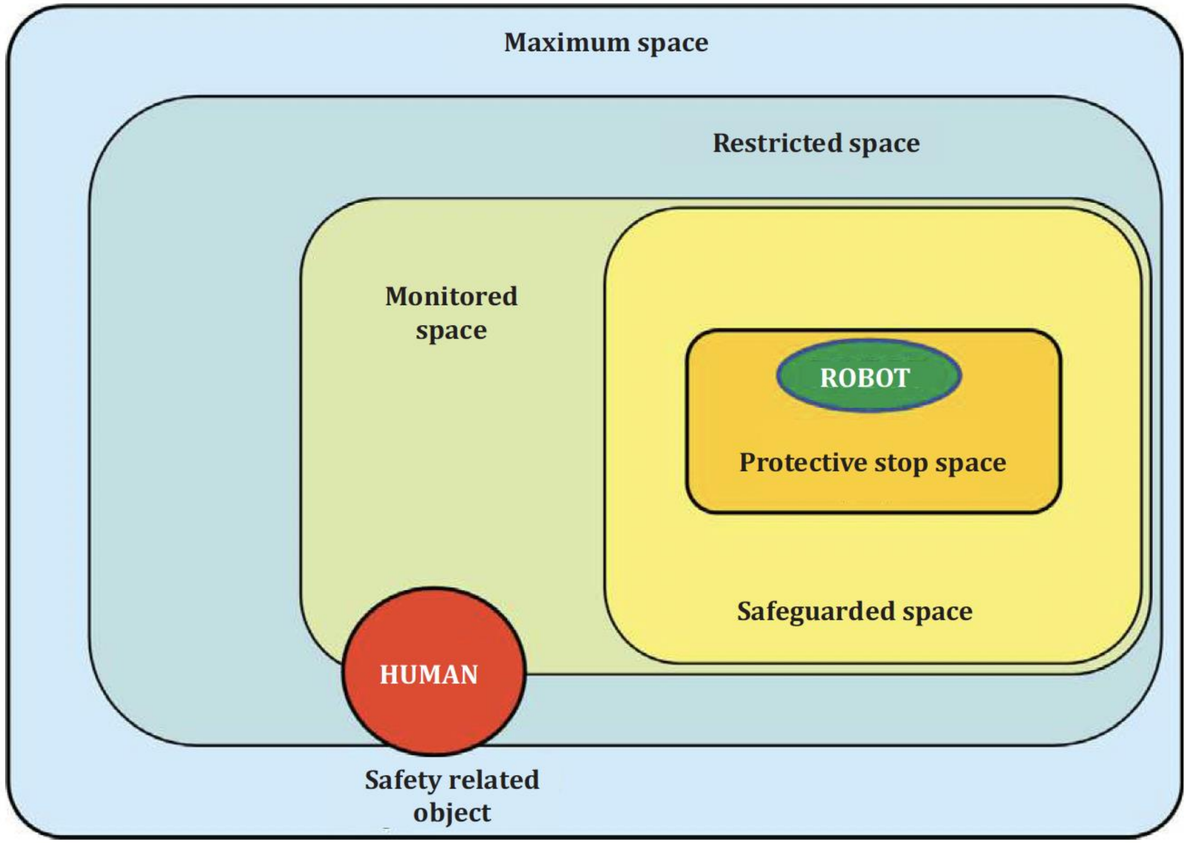


Figure B.1 — Operational spaces of an autonomous person carrier robot

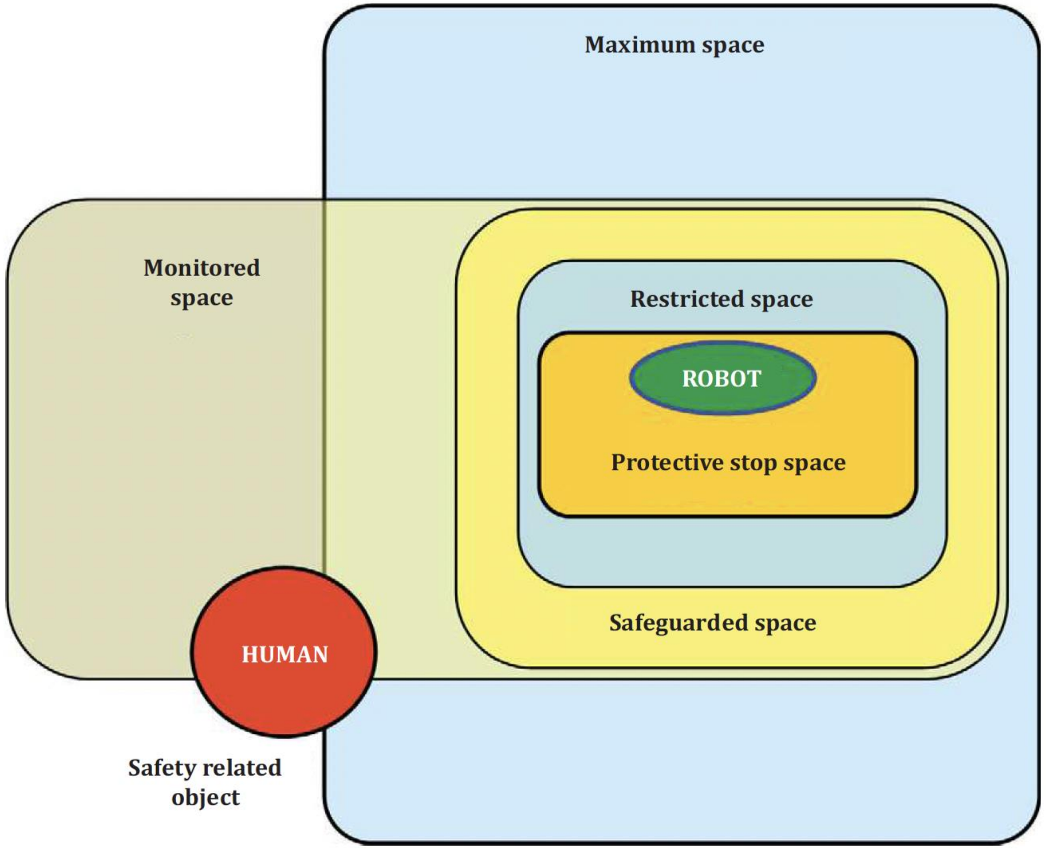
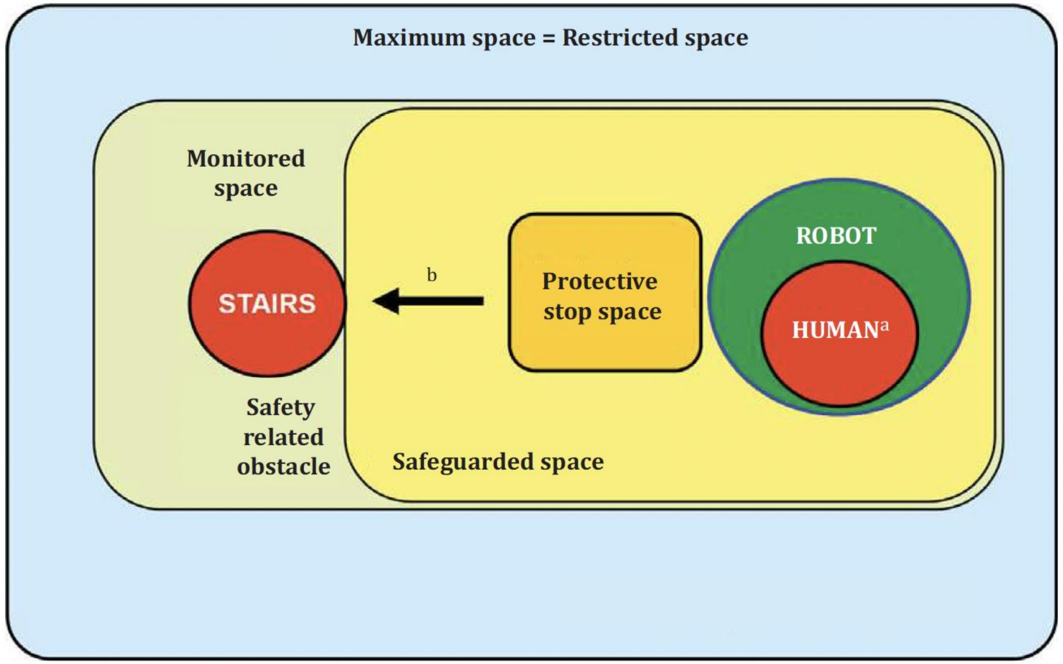


Figure B.2 — Operational spaces of a personal care robot with manipulator

Source: ISO 13482 (2014)

Operational Spaces

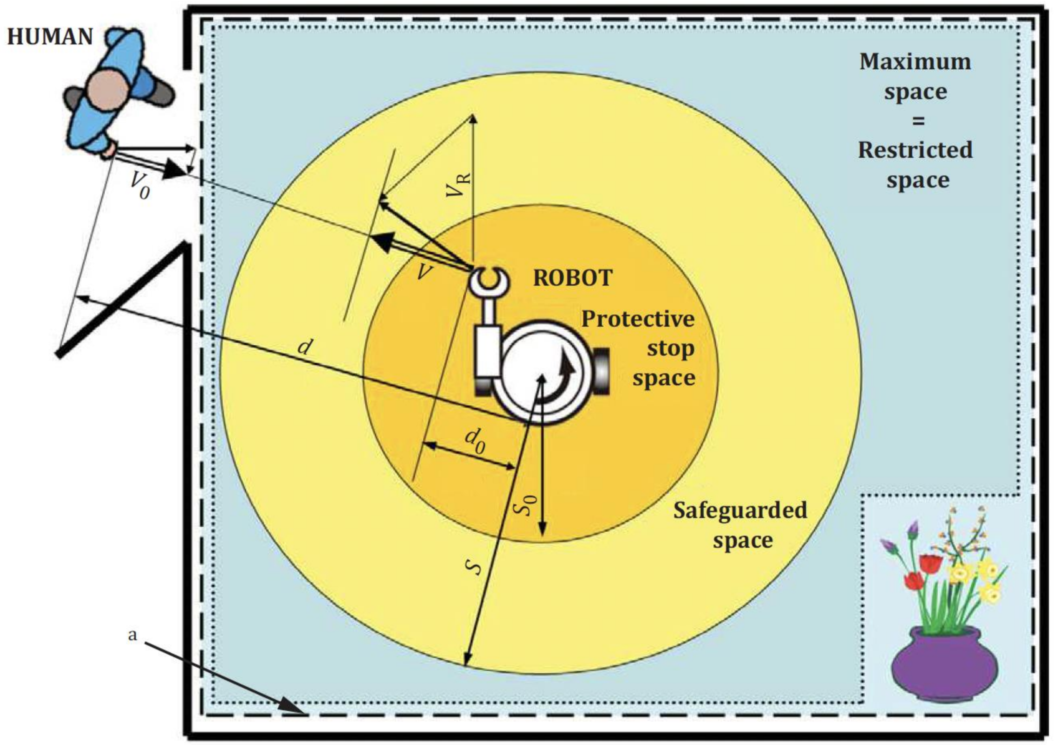


Key

a safety-related object

b momentary direction of movement

Figure B.3 — Operational spaces of a physical assistant robot



Key

a maximum space

Figure C.1 — Personal care robot application with a manipulator on a mobile platform

Source: ISO 13482 (2014)

Who builds the operational model?

ISO 13482 defines operational spaces:

- Maximum space, Restricted space
- Monitored space, Safeguarded space
- Protective stop space

ISO/TS 15066 defines collaborative spaces:

- Safeguarded space (no contact)
- Collaborative workspace (monitored)
- Human detection zones

IEC 80601-2-78 defines movement constraints:

- Pre-set ROM limits per joint
- User-specific boundaries
- Misalignment detection zones

Who defines THESE for your home?

Formalizing the Problem

$M_{designer}$: Robot’s design-time model

- Kinematics, dynamics, sensor specs
- Assumed environment types (indoor, flat surfaces)
- Verified via ISO/TR 23482-1 test methods:
  - Static/dynamic stability tests
  - Surface temperature tests
  - Acoustic noise tests
  - Collision impact tests

M_{user}: User's mental model of their environment

- Home layout, object locations
- Valuable/fragile items
- Social rules (nursery quiet during nap)
- Implicit in user's head, never formalized

M_{robot}: Robot's learned model during operation

- SLAM-based map
- Object recognition database
- Learned navigation patterns
- Incomplete, updates continuously

Safety Assessment Requires

$$M_{designer} \cup M_{user} \cup M_{robot} = M_{complete}$$

$\boldsymbol{M_{designer} \cup M_{user} \cup M_{robot} = M_{complete}}$

- $M_{designer}$ available at design time only
- M_{user} never formalized
- M_{robot} always incomplete
- No party has $M_{complete}$ at any time

- ISO/TR 23482-1 test methods verify $M_{designer}$ but provide NO methodology for M_{user} or M_{robot}

Bidirectional Intent Alignment

How do we align User Intent with Robot Behavior?

IEC 80601-2-78 (medical robots) requires:

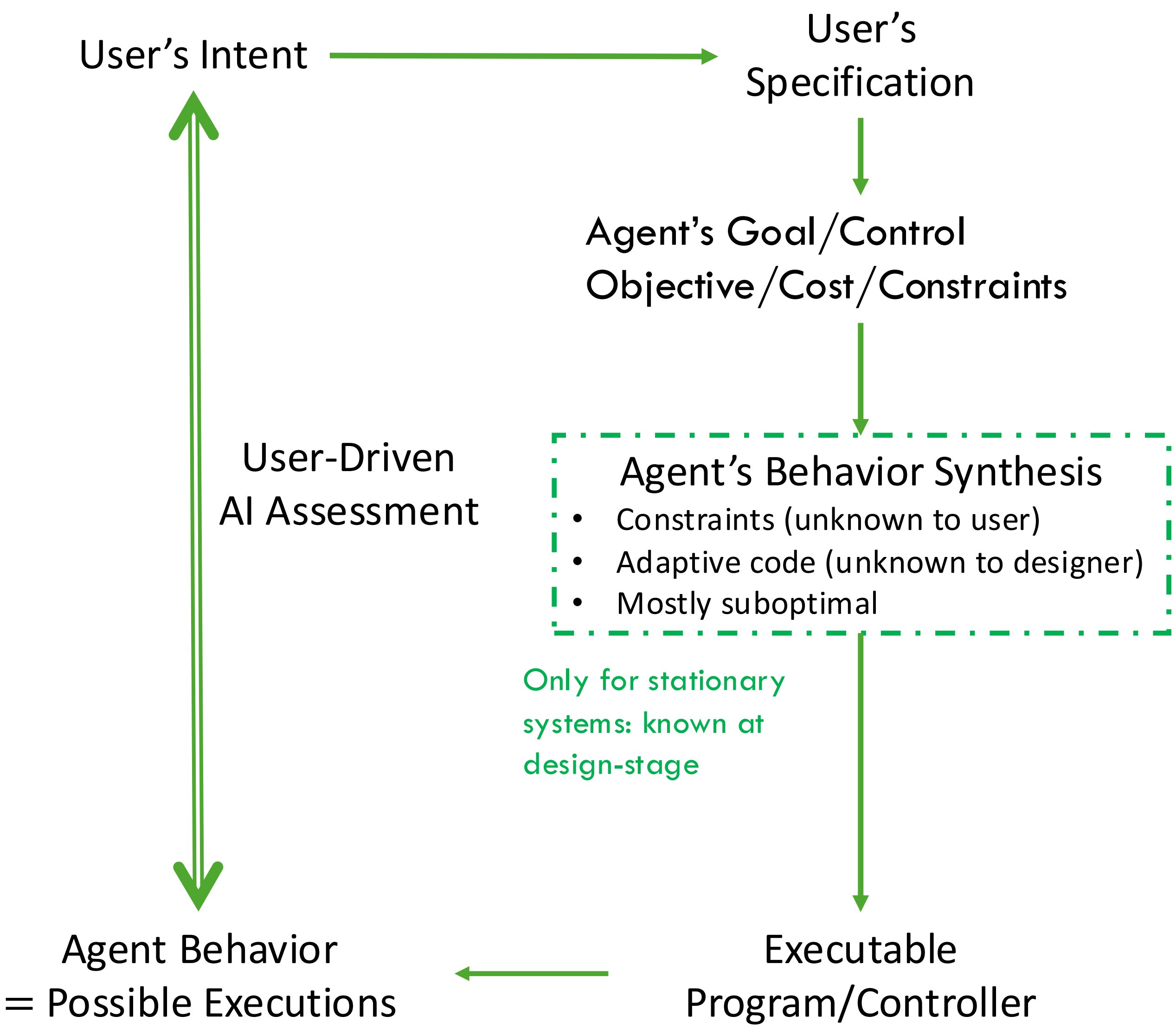
- The intended use shall be clearly defined
- The user shall understand the robot's limitations

Provides NO methodology for:

- How user communicates intent to robot
- How robot verifies understanding
- How misalignment is detected
- How corrections are made

Alignment Gaps

Needed for AI Systems



User's Intent

User's Specification

Agent's Goal/Control Objective/Cost/Constraints

Agent's Behavior Synthesis

- Constraints (unknown to user)
- Adaptive code (unknown to designer)
- Mostly suboptimal

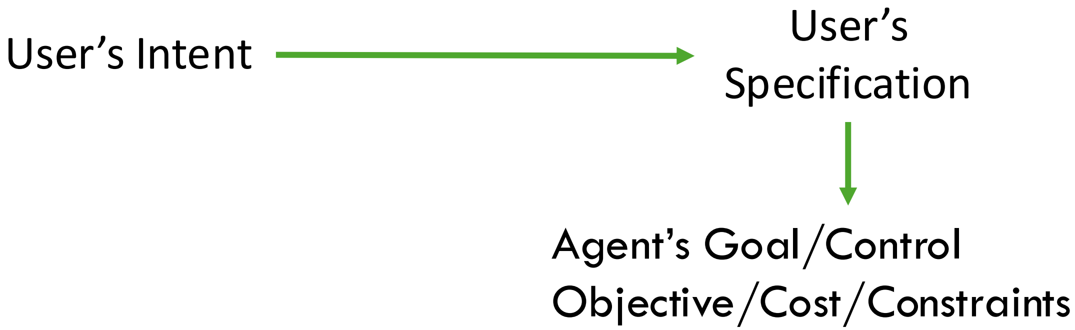
Only for stationary systems: known at design-stage

Executable Program/Controller

Agent Behavior = Possible Executions

User-Driven AI Assessment

Alignment Gaps



User's Intent

User's Specification

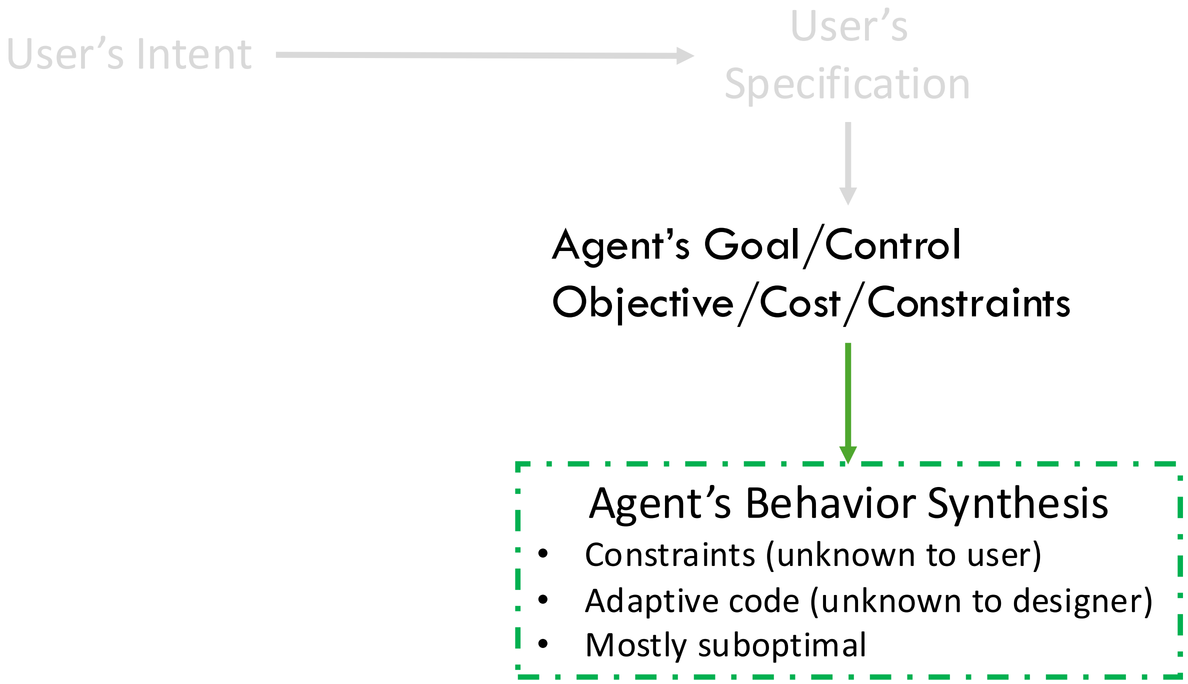
Agent's Goal/Control Objective/Cost/Constraints

GAP 1: Semantic

ISO 13482: no interpretation framework

What does "carefully" mean?

Alignment Gaps



User's Intent

User's Specification

Agent's Goal/Control Objective/Cost/Constraints

Agent's Behavior Synthesis

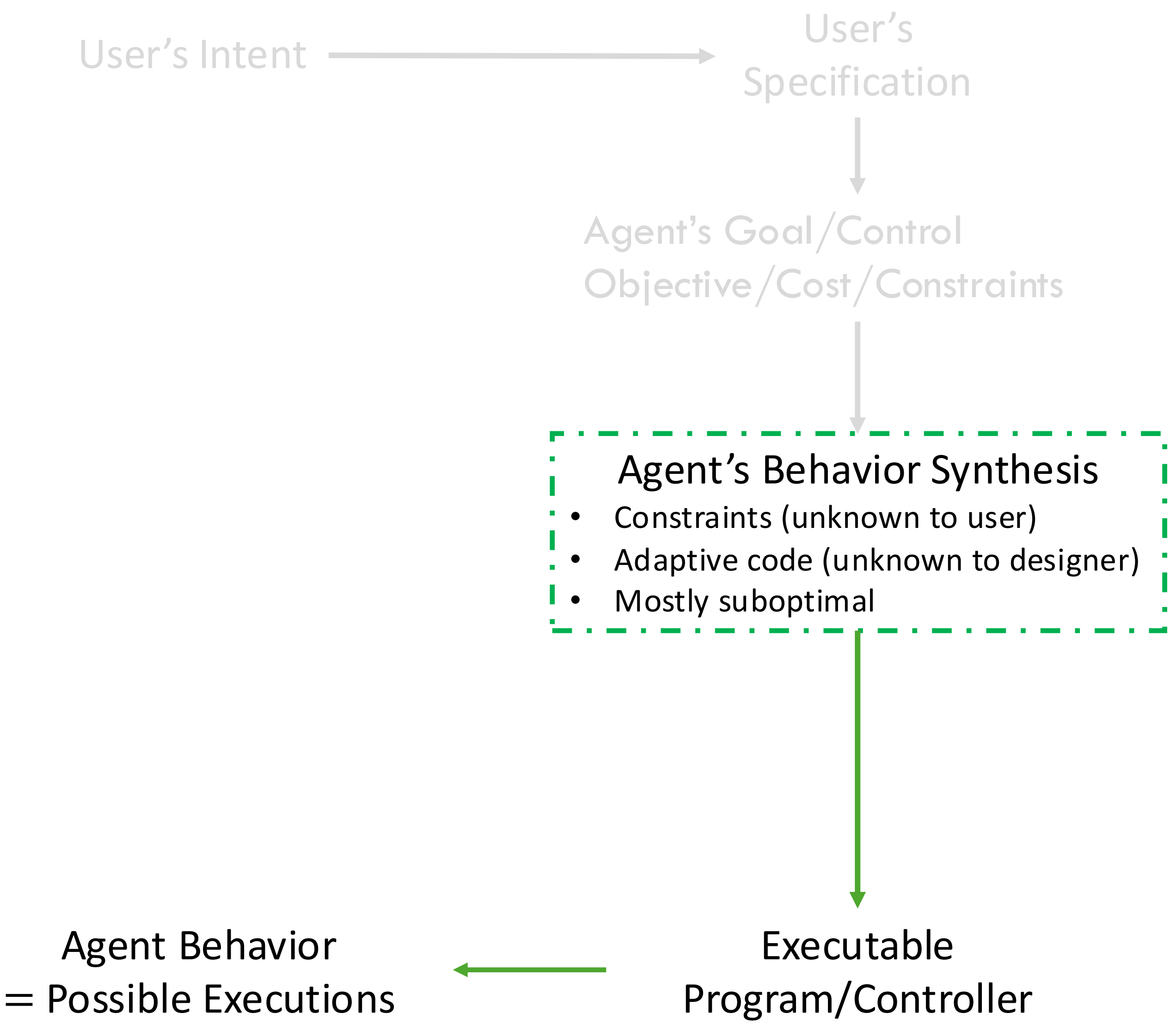
- Constraints (unknown to user)
- Adaptive code (unknown to designer)
- Mostly suboptimal

GAP 2: Capability

ISO/TR 23482-1: tests capabilities, doesn't explain them

Can robot achieve this goal safely?

Alignment Gaps



User's Intent

User's Specification

Agent's Goal/Control Objective/Cost/Constraints

Agent's Behavior Synthesis

- Constraints (unknown to user)
- Adaptive code (unknown to designer)
- Mostly suboptimal

Executable Program/Controller

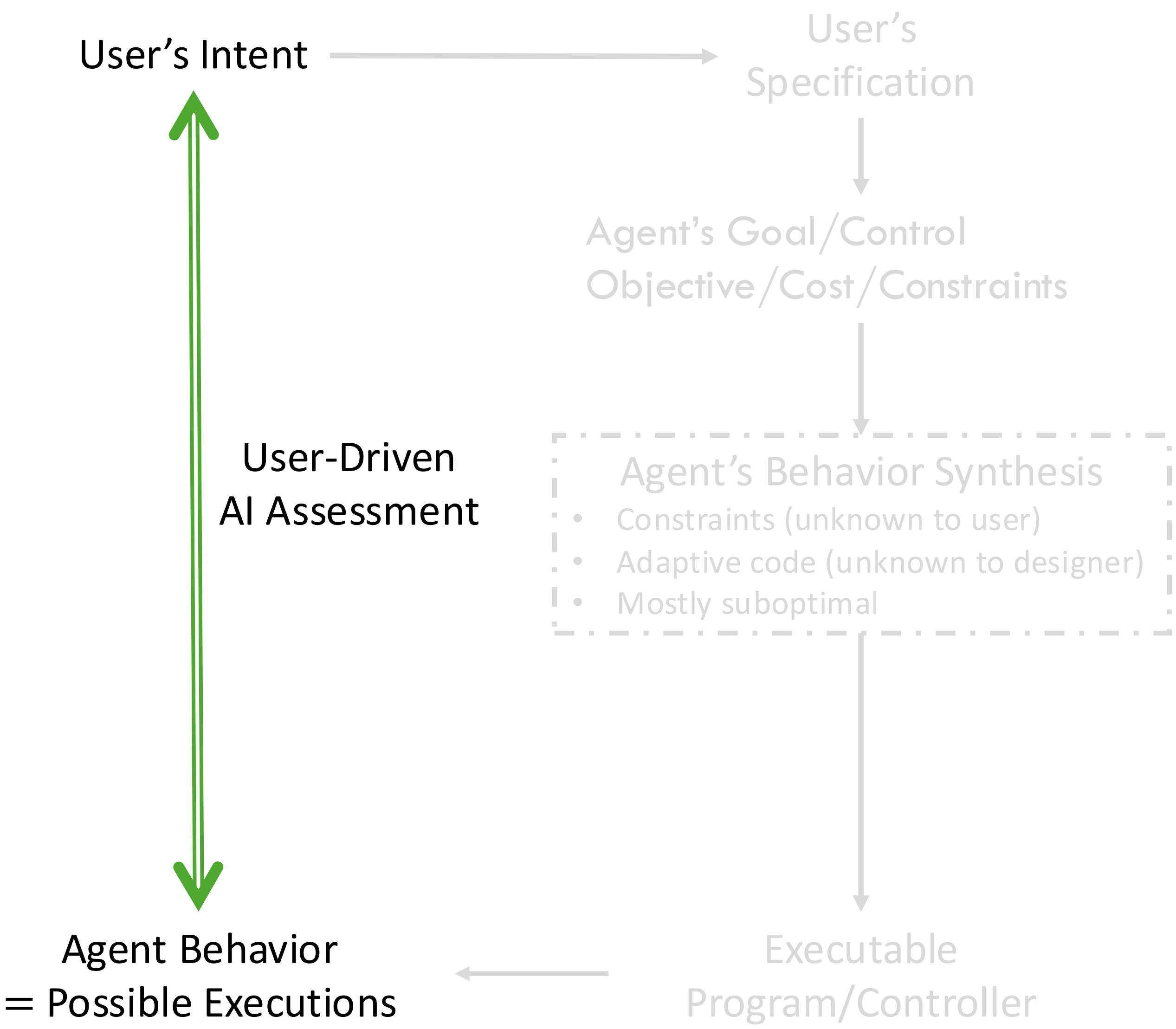
Agent Behavior = Possible Executions

GAP 3: Verification

ISO/TS 15066: verifies forces, not task completion

Did robot complete task successfully?

Alignment Gaps



User's Intent

User's Specification

Agent's Goal/Control Objective/Cost/Constraints

Agent's Behavior Synthesis
- Constraints (unknown to user)
- Adaptive code (unknown to designer)
- Mostly suboptimal

Executable Program/Controller

Agent Behavior = Possible Executions

User-Driven AI Assessment

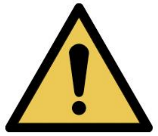
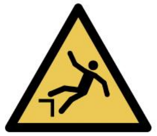
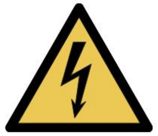
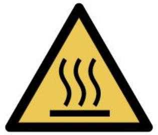
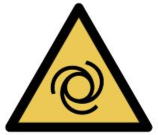
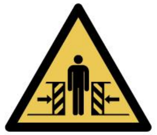
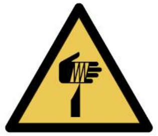
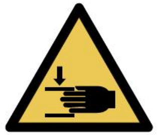
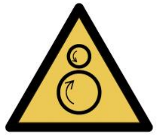
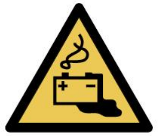
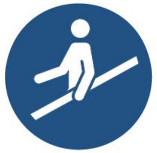
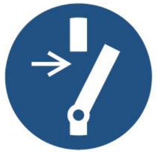
GAP 4: Feedback

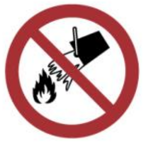
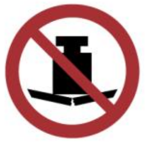
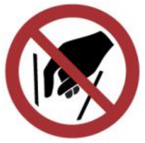
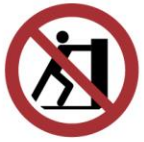
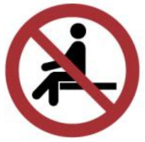
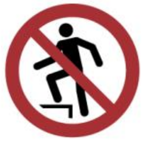
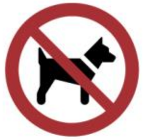
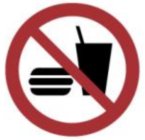
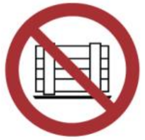
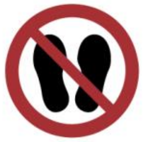
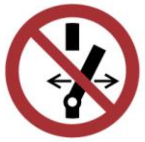
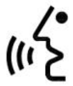
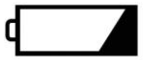
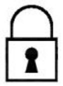
IEC 80601-2-78: requires feedback, no framework

How does user verify robot's interpretation?

Dynamic Safety Property Generation

Safety and Other Markings

ISO 7010-W001  General warning To signify a general warning	ISO 7010-W08  Drop (fall) To warn of a drop	ISO 7010-W012  Electricity To warn of electricity
ISO 7010-W017  Hot surface To warn of a hot surface	ISO 7010-W018  Automatic start-up To warn of automatic activation	ISO 7010-W019  Warning: Crushing To warn of moving mechanical parts
ISO 7010-W022  Sharp element To warn of a sharp element	ISO 7010-W024  Crushing of hands To warn of closing motion of mechanical parts of equipment	ISO 7010-W025  Counter-rotating rollers To warn of possibility of drawing in
ISO 7010-W026  Battery To warn of hazards related to batteries	ISO 7010-M012  Use handrail	ISO 7010-M021  Disconnect before carrying out maintenance or repair

ISO 7010-P011  Do not extinguish with water	ISO 7010-P012  No heavy loads	ISO 7010-P015  No reaching in
ISO 7010-P017  No pushing	ISO 7010, P018  No sitting	ISO 7010-P019  No stepping on surface
ISO 7010-P021  No dogs	ISO 7010, P022  No eating or drinking	ISO 7010-P023  Do not obstruct
ISO 7010-P024  Do not walk or stand here	ISO 7010, P031  Do not alter the state of the switch	
IEC 60417-1  To indicate a “speak” facility	IEC 60417-1  To identify a control to check the condition of the battery	IEC 60417-1  To identify on a control that a function is in the locked status

How do we generate task-specific safety properties?

Traditional approach (all standards):

- Fixed set of safety properties
- Defined at design time
- Verified once (or periodically)

- ISO 10218: 48 safety requirements
- ISO/TS 15066: Force/pressure limits tables
- ISO 13482: 85 hazard scenarios (Annex A)
- ISO 3691-4: Person detection requirements
- IEC 80601-2-78: ROM limits, torque constraints
- ISO/TR 23482-1: 17 test procedures

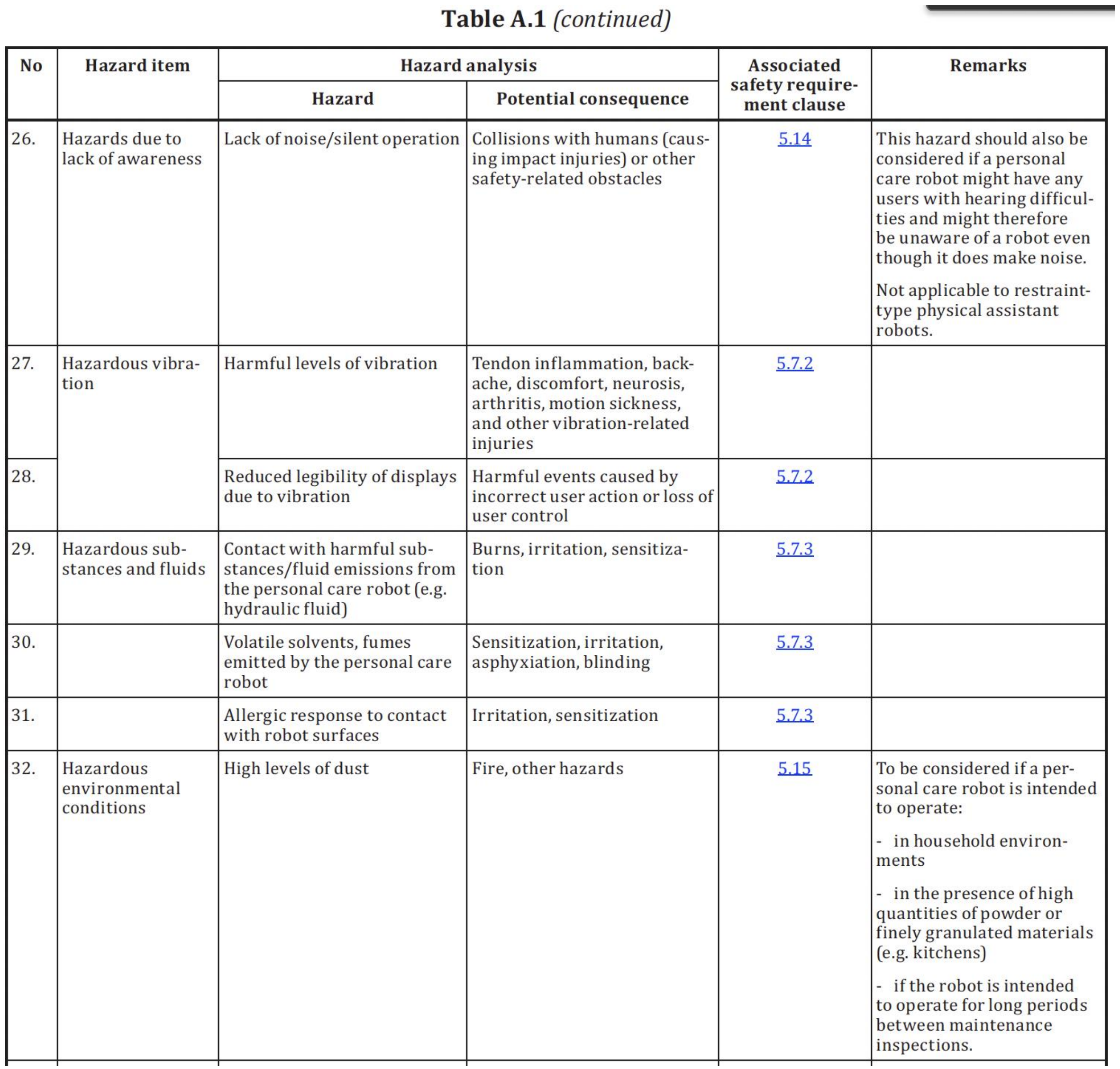


Table A.1 *(continued)*

No	Hazard item	Hazard analysis		Associated safety requirement clause	Remarks
		Hazard	Potential consequence		
26.	Hazards due to lack of awareness	Lack of noise/silent operation	Collisions with humans (causing impact injuries) or other safety-related obstacles	5.14	This hazard should also be considered if a personal care robot might have any users with hearing difficulties and might therefore be unaware of a robot even though it does make noise. Not applicable to restraint-type physical assistant robots.
27.	Hazardous vibration	Harmful levels of vibration	Tendon inflammation, backache, discomfort, neurosis, arthritis, motion sickness, and other vibration-related injuries	5.7.2	
28.		Reduced legibility of displays due to vibration	Harmful events caused by incorrect user action or loss of user control	5.7.2	
29.	Hazardous substances and fluids	Contact with harmful substances/fluid emissions from the personal care robot (e.g. hydraulic fluid)	Burns, irritation, sensitization	5.7.3	
30.		Volatile solvents, fumes emitted by the personal care robot	Sensitization, irritation, asphyxiation, blinding	5.7.3	
31.		Allergic response to contact with robot surfaces	Irritation, sensitization	5.7.3	
32.	Hazardous environmental conditions	High levels of dust	Fire, other hazards	5.15	To be considered if a personal care robot is intended to operate: - in household environments - in the presence of high quantities of powder or finely granulated materials (e.g. kitchens) - if the robot is intended to operate for long periods between maintenance inspections.

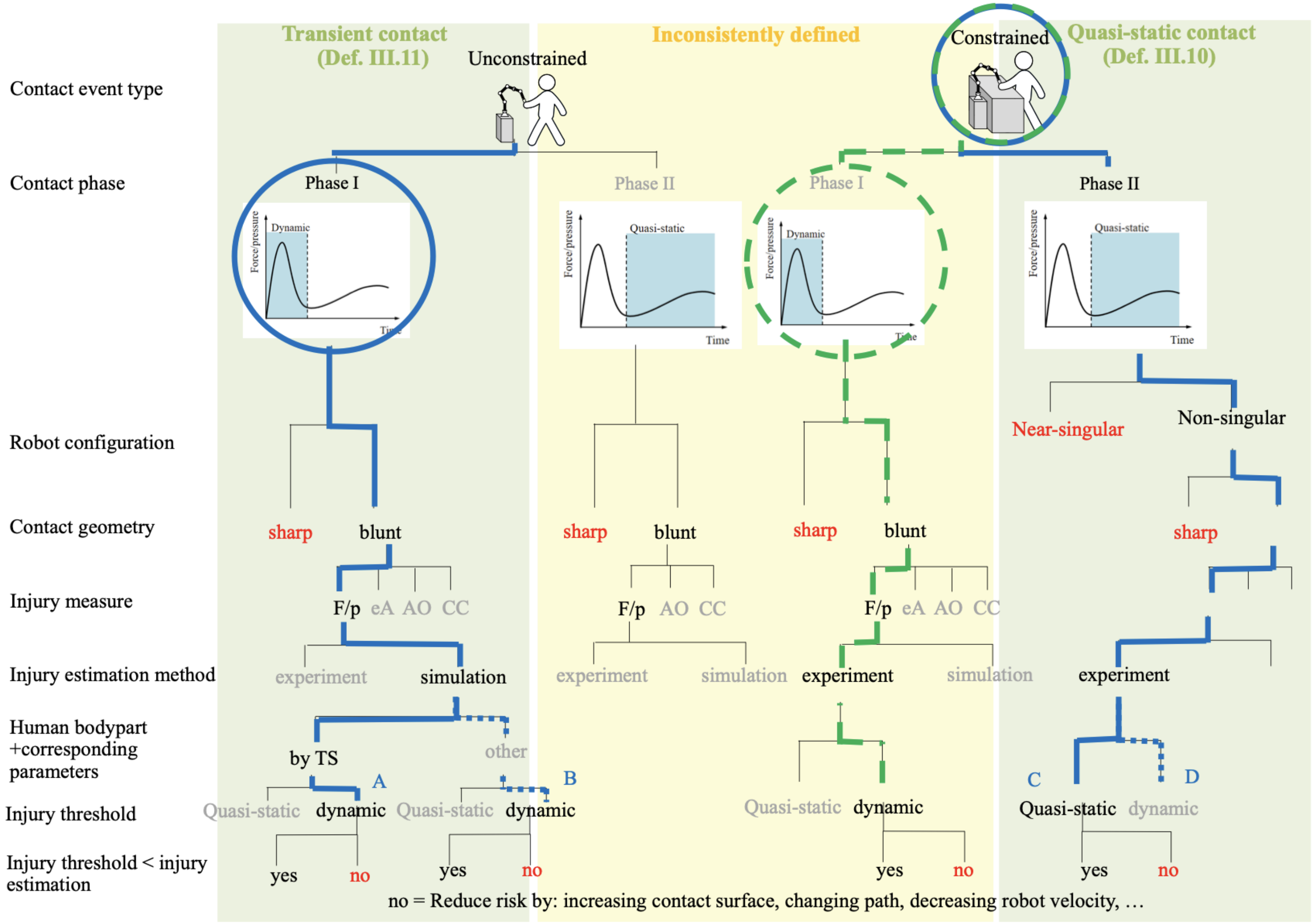


Fig. 5. Decision tree for conducting a risk assessment based on the fundamental contact subclasses [12], p. 1846 and adapted for applying ISO/TS 15066:2016(E) and the corresponding test devices for contact in physical HRI. Written in red are criteria immediately resulting in a required risk reduction according to TS, written in grey are options not directly supported by TS. The green area describes the consistently defined *transient* and *quasi-static contact*, while the yellow area describes the inconsistently defined part. For conducting a risk assessment for a constrained but dynamic contact (contact force phase I) based on TS, four interpretations are shown. The main assumption for the contact definition (constrained or dynamic) is circled The blue lines show the interpretations resulting when trying to follow the consistent definitions. The dotted blue line includes own interpretations or deviation from Def. III.10 and Def. III.11 The decision options resulting in the most realistic risk assessment are represented by a green, dashed line. The possible injury measures are referred to as force or pressure (F/p), energy density (eA), compression criterion (CC), AO-classification (AO).

Kirschner et al., ISO/TS 15066: How Different Interpretations Affect Risk Assessment, arXiv:2203.02706

How do we generate task-specific safety properties?

Traditional approach (all standards):

- Fixed set of safety properties
- Defined at design time
- Verified once (or periodically)

- ISO 10218: 48 safety requirements
- ISO/TS 15066: Force/pressure limits tables
- ISO 13482: 85 hazard scenarios (Annex A)
- ISO 3691-4: Person detection requirements
- IEC 80601-2-78: ROM limits, torque constraints
- ISO/TR 23482-1: 17 test procedures

Personal robots require:

- Dynamic property generation
- Based on current task and context
- Verified continuously at runtime

No standard addresses dynamic generation.

Capability Transparency

How does the robot explain what it can and cannot do?

- IEC 80601-2-78 requires (Clause 201.12):
  - “The manufacturer shall document the intended use”
  - “Limitations shall be clearly stated”

- ISO 13482 requires:
  - “Information for use shall include operational limitations”
  - “User manual shall describe hazards and protective measures”

- ISO/TR 23482-1 acknowledges (Introduction):
  - “Test methods cannot be comprehensive”
  - “Users should apply tests with care”

How does the robot explain what it can and cannot do?

- IEC 80601-2-78 requires (Clause 201.12):
  - “The manufacturer shall document the intended use”
  - “Limitations shall be clearly stated”
- ISO 134
  - “Infor
  - “User
- ISO/TR
  - “Test methods cannot be comprehensive”
  - “Users should apply tests with care”

None of them provides:

- Format for explanations understandable by users
- Methodology for robot to assess own capabilities
- Framework for capability discovery by users
- Approach for handling uncertain capabilities

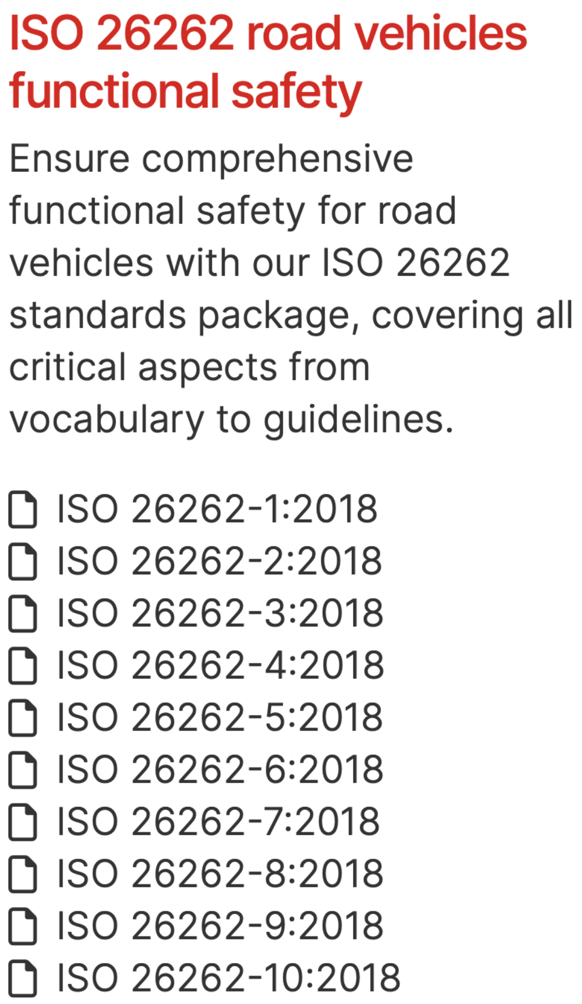
Legal Precedents and Analogies

Autonomous Vehicles [Close Analogy]

- **Problem:** Driver delegates control to AI system

- **Liability Evolution:**
  - Initially: Driver 100% responsible
  - With L2 automation (Tesla, etc.): Shared responsibility
    - Manufacturer liable for system defects
    - Driver liable for misuse, inadequate supervision
  - Proposed for L4/L5: Manufacturer liable during autonomous mode

- **Standards:** ISO 26262 (functional safety), SAE J3016 (levels)



ISO 26262 road vehicles functional safety

Ensure comprehensive functional safety for road vehicles with our ISO 26262 standards package, covering all critical aspects from vocabulary to guidelines.

🗋 ISO 26262-1:2018
🗋 ISO 26262-2:2018
🗋 ISO 26262-3:2018
🗋 ISO 26262-4:2018
🗋 ISO 26262-5:2018
🗋 ISO 26262-6:2018
🗋 ISO 26262-7:2018
🗋 ISO 26262-8:2018
🗋 ISO 26262-9:2018
🗋 ISO 26262-10:2018

Autonomous Vehicles [Close Analogy]

- **Key Insight:**
  Liability shifts toward manufacturer as autonomy increases

- **Application to Personal Robots:** A similar shift is needed, but complicated by the manufacturer
  - Multiple "drivers" (household members)
  - No licensing requirement
  - User teaches the system (car doesn't learn from the driver)

Medical Devices [Partial Analogy]

- **Problem:** Complex devices require user expertise

- **Liability Evolution:**
  - Manufacturer liable for: Device defects, inadequate warnings
  - Physician liable for: Appropriate use, patient selection, informed consent
  - Patient assumes: Inherent risks after informed consent

- **Standards:** IEC 60601 series (medical electrical equipment), ISO 14971 (risk management)

Medical Devices [Partial Analogy]

- **Key Insight:**
  Trained intermediary (physician) bridges device to patient interface

- **Application to Personal Robots:** No trained intermediary in homes
  - User is both “physician” and “patient”
  - Must understand the device AND assess one's own risk

Smart Home Devices [Weak Analogy]

- **Problem:** Smart thermostats, locks, cameras

- **Liability Evolution:**
  - Mostly traditional product liability
  - Manufacturer liable for defects
  - User liable for misuse

- **Standards:** Various (UL Safety Certification, IEC 60730 for thermostats, etc.)

- Limitation: Low physical risk. E.g., Thermostat failure: discomfort, energy cost

Current Legal Frameworks Fail

- “Defect” is undefined for learned behavior
  - Not manufacturing defect (robot = as designed)
  - Not design defect (learning = designed feature)
  - Not warning defect (can't warn about unknown learned behavior)

- Causation is distributed!!
  - Manufacturer enabled learning
  - User's environment shaped learning
  - Specific task-triggered behavior
  - Multiple necessary causes, no single sufficient cause

Learned Behavior Precedent: NONE

- Product liability assumes fixed behavior at sale
    - If toaster breaks, then manufacturer liable
    - If user modifies toaster, then user liable

- If robot learns harmful behavior?
    - Design allowed learning (manufacturer choice)
    - User's environment triggered learning (user's context)
    - Emerged behavior not explicitly designed or instructed

- **Who's liable???**

Lessons from Aviation!!

- **Aviation separates two things traditionally bundled in ISO standards:**

Investigation (Standardized)

- Same protocol worldwide
- Fact-finding (neutral)
- Root cause analysis
- Failure patterns DB
- Recommendations for change

Responsibility (Context-Dependent)

- Varies by jurisdiction
- Insurance & courts
- Regulatory actions
- Equipment modifications
- Training/procedure changes

Personal robots need BOTH: Baseline standards (design) + Investigation protocols (failure response)

Hybrid Approach

Design Standards

- ISO 13482 baseline
- Force/pressure limits
- Environmental tests

Home Integration

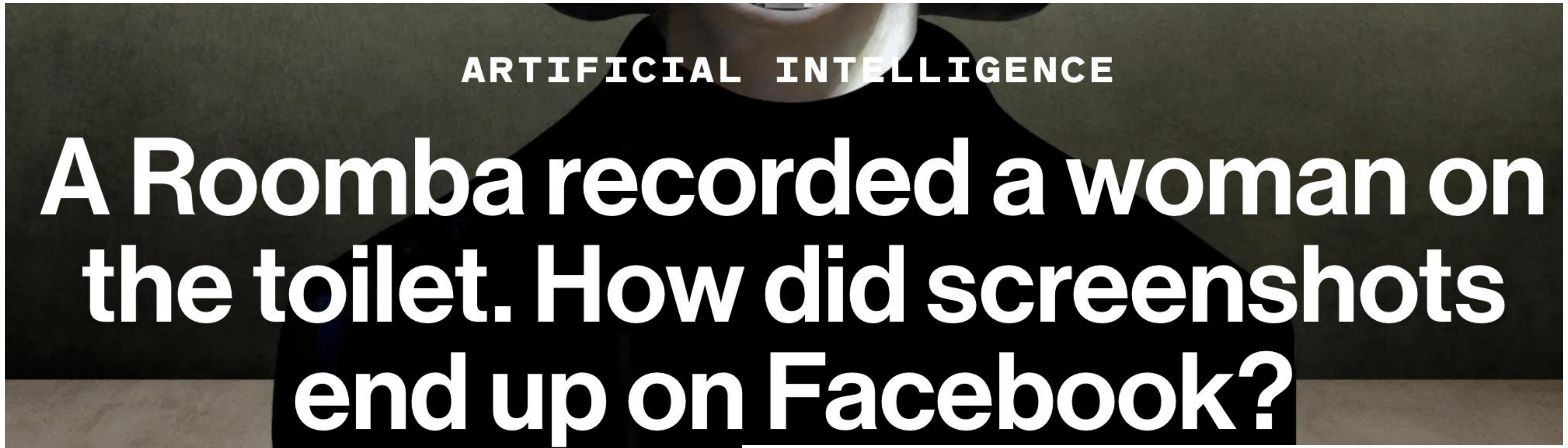
- Pre-deployment mapping
- Capability discovery
- Bidirectional alignment

Incident Investigation

- Standardized protocol
- Neutral fact-finding
- Database accumulation

Transparency

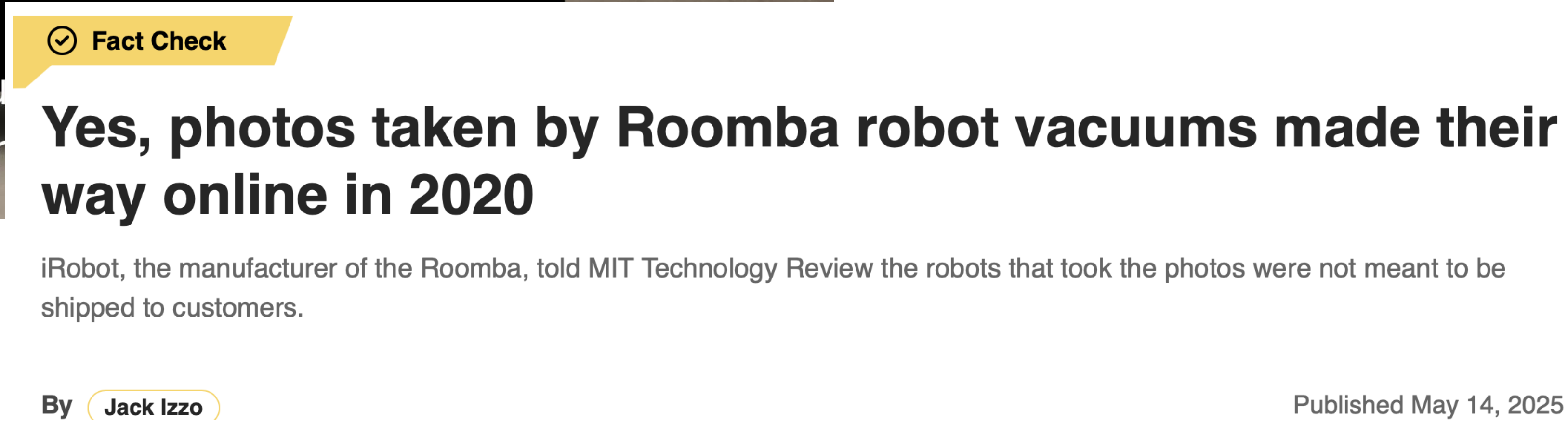
- Real-time dashboards
- Confidence metrics
- Learned model visibility



ARTIFICIAL INTELLIGENCE

A Roomba recorded a woman on the toilet. How did screenshots end up on Facebook?

Robot vacuum companies say you
supply chain for data fr



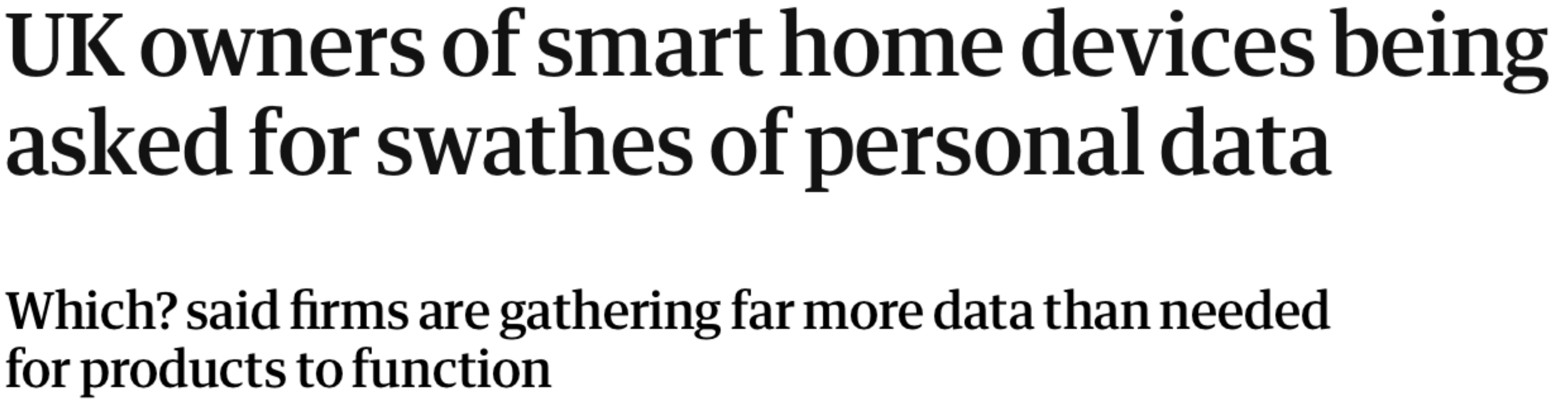
Fact Check

Yes, photos taken by Roomba robot vacuums made their way online in 2020

iRobot, the manufacturer of the Roomba, told MIT Technology Review the robots that took the photos were not meant to be shipped to customers.

By Jack Izzo

Published May 14, 2025



UK owners of smart home devices being asked for swathes of personal data

Which? said firms are gathering far more data than needed for products to function

Conclusion

- Standards Are Insufficient
  - ISO 13482 is necessary but cannot mandate the distributed knowledge (designer/user/robot) needed for safe operation in unique home environments.

- Autonomous Driving Offers a Warning
  - AVs face identical liability issues with learned behavior, yet no legal precedents exist
  - Personal robots face worse: more diverse environments, no licensing, users who teach the system.

- Aviation's Investigation Model Works
  - ICAO Annex 13 separates neutral fact-finding from liability determination.
  - Personal robotics can adopt this: standardized incident investigation with context-dependent responsibility.