A Pathway Study for Future Humanoid Standards

Technical Report · September 2025

DOI: 10.13140/RG.2.2.27892.21122

16 authors, including:

Aaron Prather
ASTM International

Ingo Keller
Heriot-Watt University

A Pathway Study For Future Humanoid Standards

IEEE Humanoid Study Group

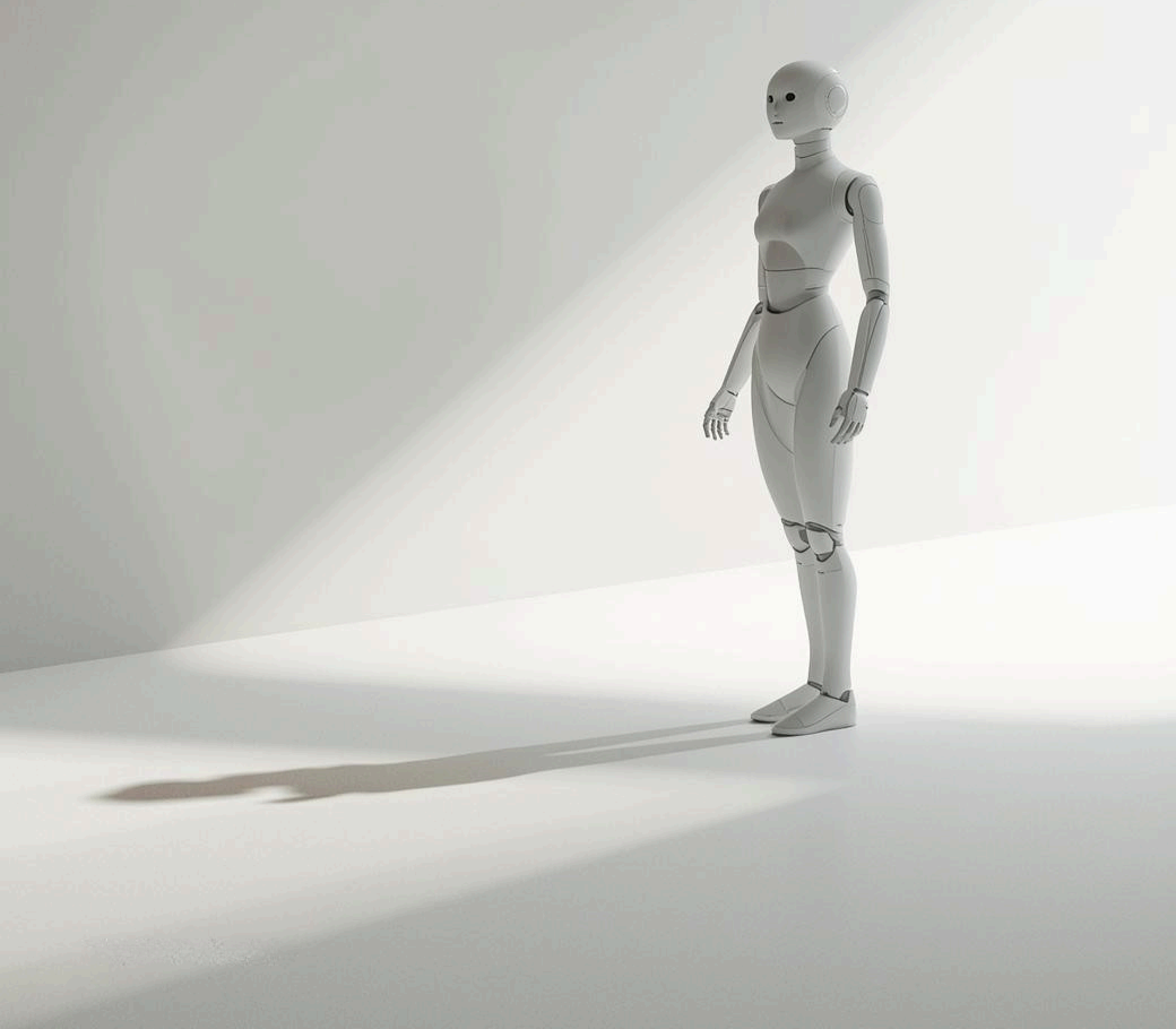

TABLE OF CONTENTS

Introduction: Building the Framework for Humanoid Robotics Standards

Author: Aaron Prather, IEEE Humanoid Study Group Lead

Humanoid robots occupy a unique and highly anticipated space in the robotics landscape. Unlike other automation systems, they promise to operate in environments designed for humans, performing tasks as diverse as industrial assembly, healthcare support, and public-facing services. Their appeal is obvious: a single robotic form that can, in theory, adapt to almost any setting.

Yet, that promise largely remains unfulfilled. The reality is that humanoids face much greater challenges than most robotic systems, not only technically but also in how they are evaluated, certified, and trusted. The current standards framework is not designed for them. Most existing robot standards assume fixed or statically stable systems and do not consider the dynamic, inherently unstable nature of a humanoid's locomotion. Nor do they fully address the complex ways these machines interact with people, not just physically, but socially and psychologically.

This report seeks to bridge that gap. It is not a set of final answers, but rather a framework of findings and recommendations that can guide the next phase of development for humanoid standards. The analysis draws on three critical themes—classification, stability, and human-robot interaction—each of which is deeply interconnected and essential to moving beyond pilot programs toward widespread deployment.

Why This Matters Now

For those developing humanoids, these findings are a roadmap to broader acceptance. Manufacturers need clear criteria to prove that their robots are safe and effective. Customers need assurance that the robots they adopt will function reliably in their intended environments. Regulators require standards that are precise enough to be applied consistently, yet flexible enough to support innovation.

For Standards Development Organizations (SDOs), the need is even more pressing. Humanoids are not just another class of mobile robots; they combine characteristics from nearly every existing category. Without a unified approach to classification and risk assessment, different SDOs risk producing fragmented or conflicting requirements that slow adoption and erode public trust.

The timing is critical. Investment in humanoids is accelerating globally, with manufacturers already piloting systems in factories, logistics centers, and healthcare facilities. Governments are beginning to explore policies for robots operating in public spaces. Without dedicated standards, the market will either move forward without adequate safety assurances or remain stagnant due to uncertainty.

A Coordinated Path Forward

This report is written not only for engineers and researchers, but also for SDO members who will turn these ideas into actionable standards. It is designed to show where existing standards can be extended, where entirely new ones are required, and how organizations can collaborate.

- Classification is presented as the foundational shared taxonomy to identify how humanoids relate to other robotic systems, what capabilities it should be expected to have, and which standards apply.
- Stability is identified as the critical bottleneck—the area where performance metrics and safety validation are most urgently needed.
- Human-Robot Interaction is explored as both a safety and societal issue, requiring new guidance for physical interaction, workflow impacts, and even perceived safety.

Taken together, these elements provide a structured approach to answering the questions regulators, customers, and the public are already asking: Which humanoids can we trust, and in which environments?

The Value of This Report

The findings outlined here are valuable to two groups in particular:

- For innovators and manufacturers, they provide insight into how to design humanoids that can be certified and deployed in diverse human environments. The recommendations on classification, stability testing, and interaction guidelines will help align engineering priorities with future regulatory expectations.
- For Standards Development Organizations, this report offers a starting point for coordination. It highlights where ASTM, IEEE, and ISO efforts can intersect—ASTM leading on test methods, IEEE on performance metrics, ISO on safety thresholds—and why these must evolve together rather than in isolation.

If humanoids are to progress from prototypes to mainstream tools, their success will depend as much on shared standards as on technical breakthroughs. Without clear, harmonized guidelines, deployment will remain slow, inconsistent, and limited to tightly controlled environments. With them, manufacturers and regulators can move forward with confidence, knowing that safety, performance, and trust have been addressed in a systematic, evidence-based way.

Looking Ahead

The chapters that follow dive deeper into each theme, presenting the details behind these findings and offering concrete recommendations for moving forward. While the path to full standardization will take time and require close collaboration among multiple SDOs, the framework is ready to be built. For those shaping the future of robotics, including engineers, researchers, and standards professionals, this report serves as both a guide and a call to action. The decisions made today about classification, stability, and human-robot interaction will determine whether humanoids remain a niche technology or become trusted, integrated tools in the spaces where we live and work.

The Humanoid Robot Market

Authors: Mike Oitzman, Aaron Prather, and Robert Little

The humanoid robot market presents a fascinating enigma. Despite the optimistic projections of venture capitalists and the fervent visions of founders, it's a space that is simultaneously immense, stubbornly difficult to categorize, and always seemingly five years in the future. The automation market is also segmented into three distinct spaces: industrial, public/commercial, and home. Each of these spaces has unique safety and feature requirements. The bottom line is that a robot should never harm a human.

Science fiction has long been captivated by humanoid robots – machines designed and built to emulate human appearance, setting them apart from robots with specific industrial functions. Humanoid robots, by their nature, are general-purpose machines, distinct from single-purpose robots deployed into commercial and industrial settings. Industrial robots are often optimized for speed and/or accuracy, able to perform tasks that humans can not.

It could be said that humanoids are designed to "displace" rather than "replace" humans. People can deploy general-purpose humanoid robots to perform menial tasks while we attend to other, "higher-value" tasks. By contrast, industrial robots are designed to move faster, move more precisely, and lift heavier payloads than humans can. Industrial automation has been positioned for dull, dirty, and dangerous applications, tasks that humans don't want to do.

One core measure that has made industrial robots so successful is that there is a clear return on investment (ROI) and a measurable payback period for this autonomous equipment. For any automation investment, the system must return greater value than the cost of the solution.

The ROI for humanoid robots remains unclear. Humanoid robots will be able to perform a variety of tasks, with each task having a different value to the robot operator. Compare this to an industrial robot deployed into a specific task like spray painting a car body or assembling a circuit board for an 8-hour shift - this work is measurable and quantifiable.

The difficulty in determining the value of a humanoid robot comes when you realize that a dynamically balancing humanoid robot is an order of magnitude more complex than the state-of-the-art industrial systems or wheeled autonomous mobile robots (AMRs) in production today. This makes it difficult to size the market, and results in wildly varying market sizes and growth rates.

Several factors have sparked a “Cambrian explosion” of humanoid innovation in the past few years. The cause is a unique nexus of factors:

- Computing power has reached a threshold and become small enough that the computing power of the first Cray Supercomputers now fits inside your smartphone.
- Battery technology has achieved a level of power density that can offer reasonable runtimes for dynamic machines like humanoid robots.
- Artificial intelligence has evolved at an unprecedented rate over the past five years.

Until now, the limits of computing, power, and AI have hindered the realization of a humanoid robot form factor.

In researching this project, we collected data for over 160 different humanoid robot models that are being developed around the globe by over 120 companies.[5] China and the Asia region in general lead the world with the number of models and companies, and the Chinese government is pouring billions of dollars into its domestic robotics industry. The U.S. and the Americas are No. 2. Europe, the Middle East, and Africa (EMEA), and the rest of the world (ROW) are in a distant third place. See Figure 1.

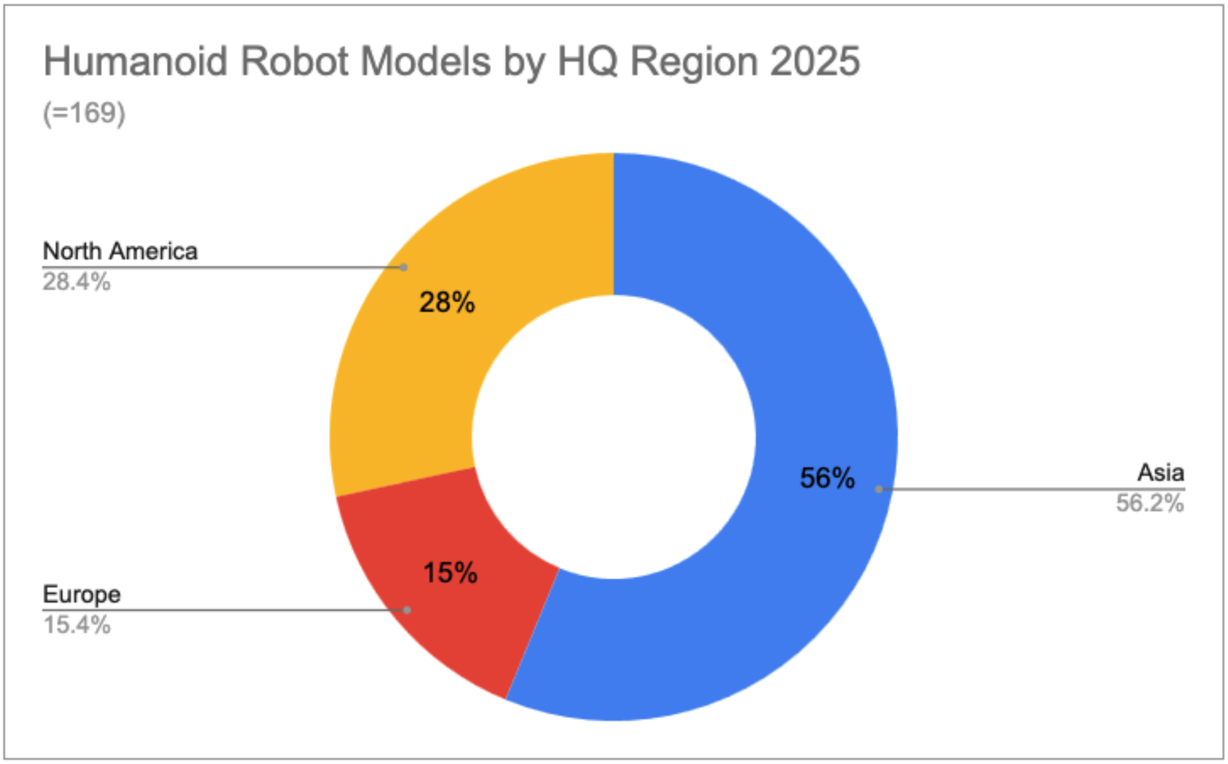

Figure 1 - Humanoid robot models by headquarter region, 2025. (n=169) Source: ASTM Humanoid project database.

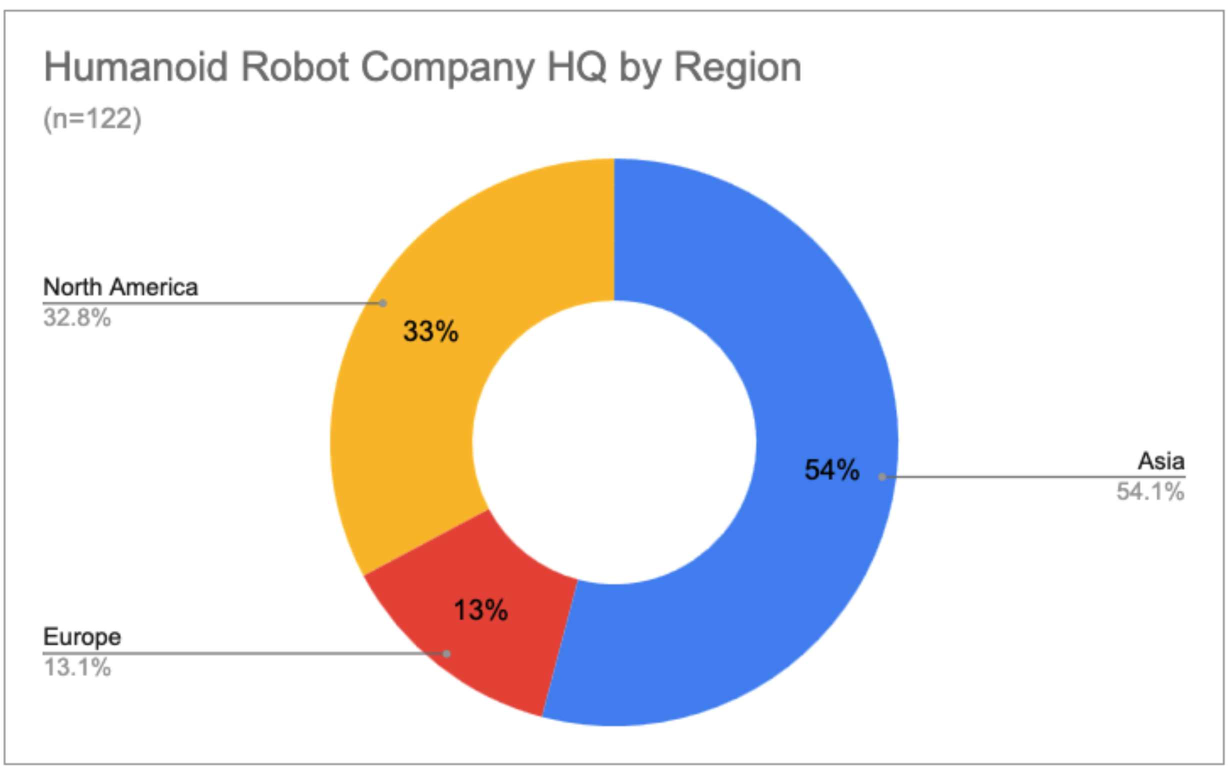

Figure 2 - Humanoid robot companies by headquarter region, 2025.[5] Source: ASTM Humanoid project database.

There is currently no believable market-sizing estimate for humanoid robots. However, all of the research consistently predicts a multibillion-dollar valuation in the next five to 10 years. The most conservative and best informed is USD 2 billion by 2032 (Interact Analysis).[5]

Some projections include:

- USD 13.25 billion by 2029 (MarketsandMarkets),[1] USD 3.83 billion in the US market alone by 2029 (MarketsandMarkets).[2]
- USD 76.97 billion by 2032 (SNS Insider).[3]
- USD 6.72 billion by 2034 (Cervicorn Consulting).[4]

Humanoid robots are unlikely to replace traditional automation systems, which excel at the high-speed, high-precision, and heavy-lifting tasks that humanoids currently cannot match. For a widespread shift to occur, the performance-to-cost ratio of humanoids must surpass that of purpose-built machines, a potential outcome driven by the scalability of their general-purpose nature.

Mobile robots have also evolved quickly in the past decade, and in the process, many of the early-to-market companies have either been acquired or gone out of business as these systems commoditize.

As stated earlier, autonomy deployed into industrial and logistics applications requires provable ROIs. Industrial robot applications require that the robot operate in a guarded work cell, separated from humans. Safety regulations and protocols for these applications are mature and well-defined.

Collaborative robots can work near humans but must work to avoid contact. If contact occurs, they must be force-limited to ensure humans are unlikely to be hurt. Humanoids have similar characteristics to collaborative robots, with a key difference: they can tip over and potentially harm a person nearby, even if not in direct contact. Future humanoids will be developed to have the ability to touch and hold humans, e.g., helping an elderly person out of bed. This is currently beyond the collaborative standard.

The very nature of humanoid design is that these robots are likely to end up sharing workspaces with humans. The majority of existing models have been engineered to mimic the physical characteristics of humans. The average height of the current crop of humanoids is 163 cm (64 in),[5] and the weight is 66 kg (145 lbs).[5] With two exceptions — 1X and Clone – the humanoid robots are all covered in hard metal alloys, carbon fiber, or hard plastic.

Humanoid mobility is enabled by either bipedal walking or via a wheeled base. Bipedal robots require dynamic stability, which means that there is always a danger of the system tripping or falling to the ground. If there is a catastrophic failure, the 60+ kg machine is likely to tumble or slump to the ground, trapping anything underneath, including pets, toddlers, or the elderly.

Because of this inherent instability and catastrophic failure modes, humanoid robots require an accepted safety standard for all machines before deployment around the public or untrained persons.

Sources:

- https://www.marketsandmarkets.com/Market-Reports/humanoid-robot-market-99567653.html
- https://www.marketsandmarkets.com/PressReleases/usa-humanoid-robot.asp
- https://www.snsinsider.com/reports/humanoid-robot-market-1616
- https://www.cervicornconsulting.com/humanoid-robot-market
- ASTM humanoid robot database
- https://interactanalysis.com/insight/humanoid-robots-large-opportunity-but-limited-uptake-in-the-short-to-mid-term/

The Unique Risks Humanoids Bring and Why Standards Must Evolve

Authors: Ingo Keller, Aaron Prather, and Rahul Ramakrishnan

Humanoid robots hold tremendous promise because of their human-centric design. By mirroring our form and movement, they can utilize our tools, navigate our spaces, and undertake tasks that once required human hands. This very similarity, however, is what makes them uniquely complex and uniquely risky.

Unlike industrial robots operating behind cages or mobile robots confined to controlled routes, humanoids are meant to work with us, often in close physical and social proximity. They blur the line between machine and partner, shifting from simple cooperation (working alongside us on separate tasks) to deep collaboration, where robots and humans share workflows, tools, and real-time decision-making. This shift amplifies existing safety concerns and introduces entirely new ones, from unpredictable physical behaviors to psychological and ethical challenges.

The risks are multi-dimensional:

- Physical and Functional – Bipedal robots, by design, operate in states of managed instability. A single loss of balance can have serious consequences in shared spaces. Their dexterity still lags far behind the human hand, creating hazards in tasks that require fine motor control.
- Psychosocial and Ethical – Their lifelike form encourages overtrust, leading people to expect intelligence and emotional understanding far beyond their true capability. In sensitive roles—such as healthcare, eldercare, and childcare—this can erode trust, create emotional dependencies, or even cause harm if performance falls short.
- Privacy and Cybersecurity – Constant sensing and networked connectivity make humanoids potential surveillance tools and targets for malicious control, raising concerns not just for safety, but for data security and personal autonomy.

- Reliability and Predictability – Unlike traditional robots, humanoids must adapt to constantly changing human environments. A sensor glitch or software fault isn't just a technical failure—it can directly impact human safety.

These challenges are not insurmountable, but they demand a new level of rigor. Existing Robotics standards, designed for fixed, wheeled, or cooperative systems, do not account for the dynamic balance, high-stakes collaboration, and human-like interaction that humanoids bring. Simply extending current safety requirements will not be enough.

How This Report Addresses These Challenges

The chapters that follow tackle these issues directly, offering a structured way forward for manufacturers, researchers, and Standards Development Organizations (SDOs):

- The first section examines classification, outlining why a shared classification system is essential to defining humanoids, distinguishing them from other robots, and mapping risks to specific applications.
- The second focuses on stability, identifying it as the most significant technical and safety barrier to deployment, and proposing a roadmap for performance metrics and safety validation tailored to actively balancing robots.
- The third explores human-robot interaction, highlighting psychosocial, ethical, and functional considerations that must guide how humanoids are designed, tested, and introduced into shared spaces.

Together, these sections establish a framework for creating the standards necessary to evolve humanoids from experimental prototypes into reliable, certifiable tools. The specific risks mentioned above are not reasons to oppose humanoids but are instead a call to address them differently through a coordinated, evidence-based standards effort that reflects the complexity of the technology.

Bridging the Gap: From Risks to Standards

The risks humanoids introduce are no longer theoretical. As these robots move out of controlled labs and into warehouses, hospitals, schools, and homes, the challenges outlined above are already emerging in real deployments. The question is not whether humanoids can perform the tasks, as they can increasingly do so, but whether they can do so safely, predictably, and in ways that humans will accept and trust.

Current standards only partially address this reality. Most were designed for fixed or wheeled robots operating in either isolated industrial cells or highly structured service roles. By contrast, they are generalists by design, capable of working in environments that are not engineered for automation and interacting directly with untrained users. This mismatch creates standards gaps in three critical areas:

- Defining what kind of humanoid is being deployed (and for what level of risk),
- Evaluating its stability and performance in dynamic, unstructured settings, and
- Guiding human-robot interaction, physical and psychological, in diverse populations.

Bridging these gaps requires more than simply adding new safety rules; it demands insight tailored to the specific application. Not every humanoid use case carries the same level of risk, and not every risk requires an entirely new standard. Some scenarios can be managed with existing guidance, while others represent critical barriers to deployment unless addressed through targeted innovation, validation, or policy updates.

The following use case analysis examines how these risks manifest across key sectors—manufacturing, healthcare, public services, and home environments—and assesses where standards are sufficient, where they require adaptation, and where entirely new frameworks may be necessary.

The table that follows provides this sector-by-sector risk view, setting the stage for deeper discussions on how classification, stability, and human-robot interaction standards can close these gaps.

RISK RATIONALES FOR SELECT HUMANOID USE CASE

USE CASE	TOP RISKS	KEY STANDARDS GAPS / NEEDS
WAREHOUSING OPERATIONS	Physical safety in crowded aisles Load handling ergonomics	Extend ISO 10218 & ISO/TS 15066; ASTM ergonomic guidance for repetitive material handling.
MANUFACTURING SUPPORT	Physical safety near tools Reliability in precision tasks Psychosocial impact on skilled labor	Extend ISO 10218 & ISO/TS 15066; IEEE psychosocial and task-sharing guidelines.
FACILITY MAINTENANCE	Reliability in handoffs Functional adaptability in unstructured spaces	ISO 10218 & ISO 13482; UL/ISO guidance for ergonomic tool handoffs and adaptive maintenance
CUSTOMER SERVICE & RE-CEPTION	Psychosocial overtrust Privacy & data use; Physical safety is manageable	ISO 13482, IEEE 7001; expand ISO/IEC 24029 for AI trust; UL public HRI guidelines
SECURITY & MONITORING (PUBLIC)	Physical safety in crowds Privacy & surveillance ethics Cybersecurity vulnerabilities	IEEE 7010 & NIST IR 8269; new IEC frameworks for surveillance ethics & bias mitigation
DELIVERY & LOGISTICS (OUT-DOOR)	Physical safety on roads Reliability in dynamic terrain Cybersecurity threats	ISO 13482 & NIST CSF; ASTM protocols for environmental adaptability; ISO road-interaction safety
GENERAL PURPOSE HOME ASSISTANTS	Privacy & cybersecurity Psychosocial overreliance Functional limits in cluttered homes	ISO 13482 & IEEE 7001; UL consumer certifications for residential robotics
ELDERLY & DISABILITY SUP-PORT	Physical safety in close care Emotional dependence Privacy Ergonomic adaptability	ISO 13482 & IEC 80601-2-77; ISO care robotics standards; ASTM behavioral compliance metrics
CHILDCARE & EDUCATION SUPPORT	Psychosocial & ethical sensitivity Privacy & data governance Predictability in interaction	ISO 13482, ISO/IEC 29134, IEEE 7004; new ISO/IEEE developmental safety & ethical guidelines

The table highlights where humanoid deployment faces manageable versus critical risks; however, context is crucial. Below is a brief overview of how these risks appear across key use cases and what that means for standards development.

Warehousing Operations

Humanoids in warehouses primarily handle picking, packing, and inventory transport. Most risks are manageable, provided existing safety and functional standards are applied. However, collaborative scenarios—such as handoffs or navigation in crowded aisles—require additional ergonomic and stability guidance to prevent collisions and improve load handling.

Standards Needs: Leverage ISO 10218, ISO/TS 15066, and IEC 61508; expand ergonomic guidance under ISO or ASTM for repetitive material handling.

Manufacturing Support

Assembly lines and machine-tending roles require humanoids to work closely with humans, where precision and timing are crucial. Physical safety, reliability, and ergonomic limitations are significant barriers, compounded by the psychosocial impacts of robots supplementing skilled labor. Fine motor control remains a major limit.

Standards Needs: Extend ISO 10218 and ISO/TS 15066 for collaborative industrial tasks; develop IEEE guidance for psychosocial impact and adaptive task-sharing behaviors.

Facility Maintenance

Inspection and minor repair tasks generally carry moderate risks. Collaborative handoffs—such as tool delivery—highlight the need for reliability and better functional adaptability in unstructured environments, though psychosocial and ethical risks remain low.

Standards Needs: Apply ISO 10218 and ISO 13482; create UL or ISO guidance for adaptive maintenance behaviors and handoff ergonomics.

Customer Service & Reception

Public-facing indoor roles pose increased psychosocial and ethical risks due to overtrust, unrealistic expectations, and concerns regarding data privacy. Cybersecurity is essential to protect sensitive interactions, while physical risks remain modest.

Standards Needs: Build on ISO 13482 and IEEE 7001; expand ISO/IEC 24029 for public trust in decision-making and UL guidelines for public-facing HRI.

Security & Monitoring (Public Spaces)

Humanoids in security roles face some of the highest barriers. Operating in crowded, unpredictable environments amplifies physical, ethical, and privacy risks. Networked systems are particularly vulnerable to cyberattacks, raising the stakes for both safety and trust.

Standards Needs: Integrate IEEE 7010 and NIST IR 8269; develop new IEC frameworks for robotic surveillance ethics, bias mitigation, and active threat-response reliability.

Delivery & Logistics (Outdoor)

Outdoor navigation introduces the most severe combination of risks—unstable terrain, unpredictable human interaction, weather, and cyber threats. Functional limitations and reliability issues are critical barriers to deployment.

Standards Needs: Expand ISO 13482 and NIST CSF; create ASTM protocols for environmental adaptability and ISO standards for robot-road interaction safety.

General Purpose Home Assistants

Residential assistants face moderate but layered risks: privacy, cybersecurity, and psychosocial impacts from overreliance or unmet expectations. Ergonomic and functional limits are less critical but still require attention in cluttered home environments.

Standards Needs: Strengthen ISO 13482 and IEEE 7001; develop UL consumer safety certifications for residential humanoids.

Elderly & Disability Support

This is one of the highest-stakes use cases. Close physical assistance, health monitoring, and emergency support demand rigorous safety, reliability, and ethical oversight. Emotional dependence and privacy concerns are significant, and ergonomic adaptability is often limited.

Standards Needs: Build on ISO 13482 and IEC 80601-2-77; introduce ISO standards for human-centered care robotics and ASTM behavioral compliance metrics.

Childcare & Education Support

Working with children raises heightened psychosocial and ethical concerns, including data privacy and developmental impacts. Physical safety and predictable behavior are non-negotiable, especially during play or collaborative learning tasks.

Standards Needs: Extend ISO 13482, ISO/IEC 29134, and IEEE 7004; create new ISO/IEEE standards for developmental appropriateness, safe interaction, and responsible data handling.

The Classification of Humanoid Robots

Authors: Benjamin Beiter, Brandon J. DeHart, William Harrison, Syrine Mansour, and Thomas Mather

Introduction

One of the first recommendations regarding standards for humanoid robots is to create a classification. Such a standard would classify humanoid robots into different groups, and differentiate between individual humanoid robots and between humanoids and other classes of robots. Safety is a primary concern for both manufacturers and potential users of humanoid robots, and so it is a primary factor that dictates the need for standards that provide the tools needed to meet safety requirements. Classification is an important component of standardization as well, as it could be used across all humanoid-related standards, setting metrics, test methods, guides, certifications, and definitions for specific robot classes.

A well-formed classification could be used to convey robot features, exclude other features, identify additional relevant standards for a particular robot, as well as understand design requirements, needed performance capabilities, and safety restrictions for a specific class of robots. This chapter is not intended to be a complete robot classification, but will discuss the challenges related to classification, summarize prior standardized classifications of robots, and highlight several potential approaches that could be used to build a complete classification.

Defining a "Humanoid Robot" … Or Not?

The first question when considering the creation of standards for humanoid robots is: "What is a humanoid robot?" While seemingly simple when considered colloquially, there arise distinct discrepancies in any definition when attempting to precisely define the categorical bounds of robots to which the name applies. Does a robot need to have two arms and two legs to be a humanoid? Does it need a head? Does that head need to have sensors just like humans? How human-like does it need to look? What about size, weight, shape, communication, mobility, strength, behaviors, identity………?

Innumerable characteristics could be used to define humanoids. This variance in characteristics means that using any particular one in a definition would result in overly restrictive categories that are not general enough to be useful to manufacturers, users, or standards creators. The alternative is to define a humanoid as a robot with any human-like characteristic at all, regardless of application, functionality, or any other feature.

While not excluding any robots, this definition is also not useful, and would require many subclassifications to sort humanoids into useful categories, leading back into the first problem. As such, this chapter avoids defining humanoid robots entirely, and instead recommends classifying robots by their physical structure and capabilities, executive functionality, and use case/application, just as is done in some current robotics standards.

This approach results in an overall classification system that can be used to sort all robots, inclusive of humanoid robots. Relative to this report, in the same effort to fully classify humanoid robots relative to other robotic applications, a classification for all robots can be presented. This decoupling of the term "humanoid" from classification will lead to a more useful classification of robots by their function, capabilities, and use cases. Some of these classes may involve, but will not wholly depend on, the robot's appearance or anthropomorphic structure, which are most often given as the qualitative defining factors of humanoid robots..

Note: Any use of the term humanoid for the rest of this chapter will refer broadly to any robot that someone may consider to be even partially anthropomorphic.

Current Classifications

In general, the field of robotics is far too broad to facilitate a low-dimensional classification, as robots are, at their core, an integration of many different types of hardware and software that come together into an emergent system with an almost infinite variety of possible designs, paired with an endless variety of potential applications. Furthermore, because humanoid robots are general-purpose robots, they are harder to classify compared to the more narrow scope of classical robotic systems. This section will give examples of prior and current robotic classification efforts, setting up a basis on which the classification of humanoids could be built.

Prior Classification Efforts

Currently, there is no broadly accepted cohesive classification of robotic systems by which to guide humanoid robot classification efforts in general. Furthermore, there are only a few examples of industry-facing classification efforts with robotics applications. ANSI R15.08 is effectively the only published standard with any real classification coverage for robots, though it only covers mobile robots. There have been various categorization, taxonomy, and ontology efforts originating from academia and other places (Kirschner et al., 2025; Prestes et al., 2013; Kunze et al., 2011; Kim et al., 2024); however, they often lack the industry perspective.

R15.08

ANSI R15.08 provides a classification for industrial mobile manipulators, shown in Figure 1. In general, humanoid robots that can move like a human can be considered to be Type C Mobile Manipulators. Humanoids that do not move like a human (are non-legged or are not mobile at all, just anthropomorphic) can either be classified as industrial robots, if they have a manipulator, or none of the above ("END" on the flowchart). However, this classification only focuses on industrial robots, not any other application domain. It also only considers the physical structure of the robot (i.e., possessing a manipulator arm or not), and leaves out other relevant details such as the functional capability or interactability of the robot. As such, additional classifications should be created to deal with the uses of humanoid robots with varying capabilities, intended use cases, and needed interaction capabilities.

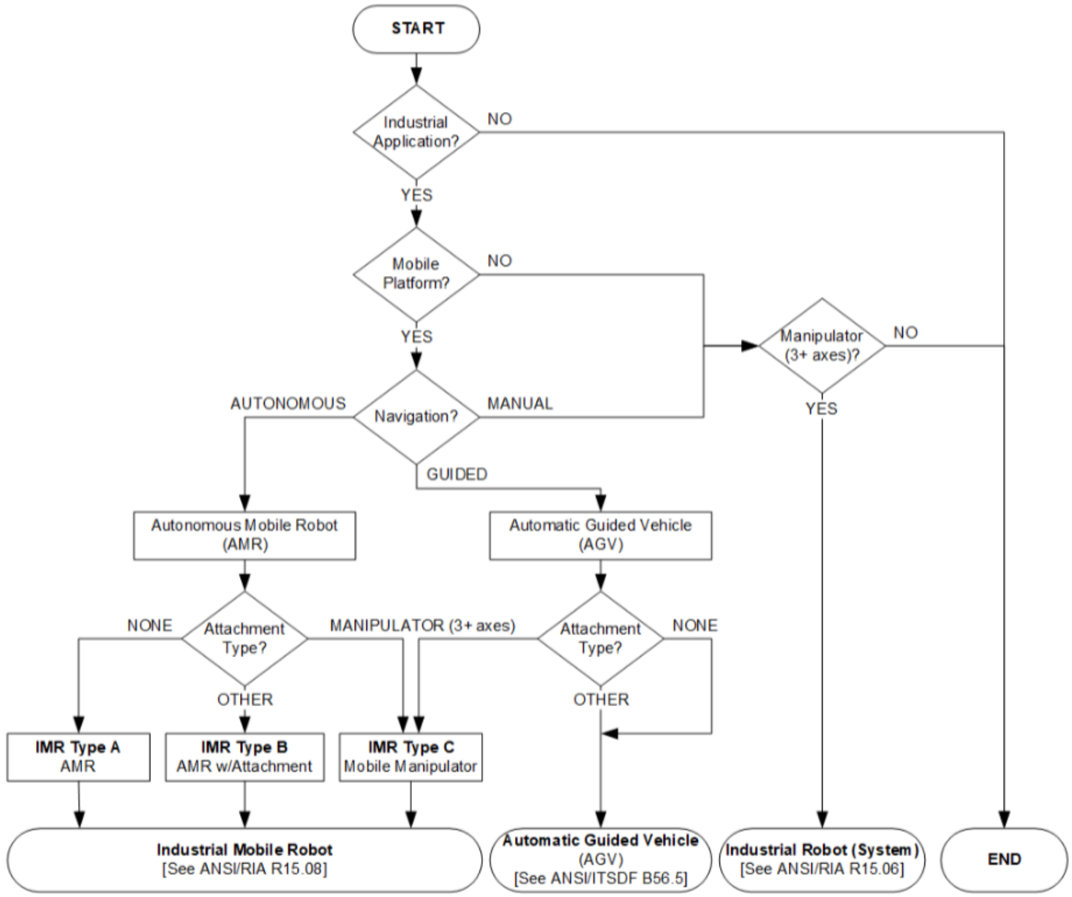

Figure 1: Classification of Industrial Mobile Manipulators and applicable standards from R15.08.

ASTM F48 (FORTHCOMING)

ASTM Committee F48 is currently working to create a classification for exoskeletons. While not published yet, the effort to classify such a broad type of robotic system is relevant to the effort with humanoid robots, as it also must balance complex design, variable use cases, and performance measures (which are hard to fully quantify).

SHANGHAI UNIVERSITY HUMANOID EMBODIED INTELLIGENCE STANDARD (FORTHCOMING)

Finally, there is a forthcoming standard being developed as part of a national initiative in China entitled: "Shanghai University Humanoid Embodied Intelligence Standard". This standard presents a classification of humanoid robot 'intelligence', which is based on several different capability factors as detailed in the standard.

Example of Possible Classifications

As can be seen from prior robotics classification approaches, there are varieties of features upon which humanoids could be classified (a feature here being any characteristic of the robot that can be used for differentiation). Prior classifications are restricted in scope to specific applications or forms of robots. Looking forward, however, humanoid robots are intended to be general-purpose robots, and so can fill many roles and come in many forms, and as such, most potential classifications that apply to any robots can and will apply to humanoids as well.

The most prominent capabilities of humanoid robots (e.g., manipulation, mobility, human interaction) are important features of almost all robots. Therefore, it is recommended that instead of specifically creating a classification of humanoid robots, effort be put into the general classification of all robots as a necessary tool for both academics and industry to talk about specific kinds of robots. Then humanoids can be classified by the same features as all robots to best understand the use-cases, performance, and capabilities of individual robots. An additional classification based on features unique to humanoids, or at least common to humanoids but rare in other robots, may be useful, but should fit within a larger robotic classification.

Based on the types of features that are often used to qualitatively describe humanoids, there are many possible ways to classify them, and robots in general. When creating classifications, the goal is to balance simplicity with usefulness, often making a single scale of classes (e.g., Class 1 to Class 5). However, a scalar approach inherently limits the generality and usefulness of the classification. Without a coherent classification approach for robotics at the moment, this chapter presents many possible features by which a classification of robots could be made, focusing on features that also apply to humanoid robots. The variety of features illustrates the complex considerations that go into designing and implementing humanoid robots, as well as the challenge presented by the prospect of creating a clear classification of robots.

In this section, we present a non-exhaustive list of example features by which the classification of robots could be implemented in general. These features are sorted into three categories, described in Table 1. This specific list of categories is the most convenient to describe the features chosen here, but other categories could readily be used (or added to this list).

TABLE 1: EXAMPLES OF CATEGORIES OF FEATURES BY WHICH ROBOTS CAN BE CLASSIFIED.

PHYSICAL CAPABILITIES	HIGH-LEVEL BEHAVIORS	APPLICATION: INTERACTIONS AND RISK
FEATURES THAT ARE INHERENT TO THE DESIGN AND MOTION OF THE ROBOT. THIS INCLUDES DYNAMIC MOTION CAPABILITIES, SUCH AS LOCOMOTION AND MANIPULATION, AS WELL AS THE LOW-LEVEL MOTOR AND MOTION CONTROL CAPABILITIES THE ROBOT POSSESSES.	Ways the robot acts in certain situations, including with safety-critical responses to environmental inputs and other observations. These behaviors could be entirely modified via the software and controls of the robot.	Tasks the robot is expected to perform, and in what environment. These features are not dependent on the physical design of the robot itself, but instead on what it will be expected to do.

Each of the classifications within these categories is defined by a set of integer levels starting at 0 (indicating the system cannot perform the associated action) and moving up, with higher values roughly translating to the increasing complexity of controlling and using the physical system to successfully perform the action associated with the design aspect. Classification levels may be inclusive, so for example a level 3 robot could also have the capabilities from levels 1 and 2, but would still be classified as level 3. It should be noted that future classifications can make use of sub-levels as demonstrated by the E-stop behavior feature. The different levels for each of these design aspects are described in the following sections and are meant only as examples of possible approaches to classifying robots; there are many viable alternative approaches that a future standards committee could consider.

Physical Capability Features

These classification approaches are about the features that are inherent to the design and motion of the robot. The features describe active capabilities that the robot has, including features of its physical structure and design. The features will change if the physical characteristics of the robot change, especially for any robot that can transition between levels.

ENVIRONMENTAL MANIPULATION

This classification is based on what level of environmental manipulation the robot is capable of. Specifically, it can be summarized as a measure of the degrees of freedom (DOF) of the system's manipulator(s), how many manipulators it has, and whether cooperative manipulation is possible or not with multiple manipulators.

Level 0: No environmental manipulation capability.

Level 1: Single manipulator with less than 3 DOF (e.g., shelf lifting attachment)

Level 2: Single manipulator with at least 3 DOF (e.g., typical industrial arm)

Level 3: Multiple 3+ DOF manipulators with fully independent workspaces

Level 4: Multiple 3+ DOF manipulators with partially shared workspaces

LOCOMOTION/MOBILITY

With locomotion, or mobility, it is difficult to define discrete transition points between one level of capability and the next. For this classification, we will be using the support polygon of the system, defined based on the locations and types of contact points the system has with the environment. The support polygon is a key indicator of static stability for any system being acted on by gravity: provided the centre of mass of the system is "above" the support polygon (in the direction opposite the force of gravity), then the system can be maintained in static equilibrium using its current contact points with the environment.

Level 0: Fixed-base robots, which are rigidly attached to their environment

Level 1: Mobile robots, which are statically stable (turning off leads to a stable pose), as the support polygon is effectively constant and always under the center of mass

Level 2: Balancing robots, where upright balance is only achieved with active control due to the underactuated nature of their contact with the environment

Level 3: Stepping/Jumping robots, which are balancing robots that can dynamically change their support polygon's shape, size, and position by moving one or more limbs

Level 4: Climbing/grasping robots, where the limbs used for mobility (legs, arms, or otherwise) can make fully actuated environmental contacts (e.g., using handrails)

HUMAN INTERACTION CAPABILITY

To evaluate the interaction capabilities of a system, without needing to also incorporate an evaluation of the software and/or controllers generating any related interactive behaviors, the focus must be on the physical attributes of the system which are used to interact with people rather than how those physical elements are used in any particular application or behavior. Therefore, the levels defined for this category are predicated on evaluating what level of interaction is possible given the design of a particular system without regard for whether or not any interactive behaviors have been developed.

Level 0: No interaction capabilities

Level 1: No direct interaction capabilities, but capable of interacting with people indirectly via a wired or wireless interface device (e.g., teach pendant)

Level 2: Includes lights, speakers, a screen, and/or other features to enable basic one-way interaction with people in its local environment (e.g., turn signals)

Level 3: Includes a camera, microphone, touchscreen, and/or other features to enable two-way interaction with people in its local environment (e.g., verbal communication)

Level 4: Includes an active head, human-friendly hands, and/or other features to enable two-way physical interaction with people in its local environment

Behavior Features

Here, a behavior is defined as a series of coordinated actions to achieve a goal. These types of behaviors are still dependent on the physical capabilities of the robot, but are mostly related to and focused on how the robot acts in the real world, based on the controllers, sensors, and other combinations of software and hardware functionality and features that the robot has.

For this form of classification, categories can be defined related to how dynamically the robot can adapt to human interaction, variations in payload, adapting its locomotion based on sensing its environment, how intelligent it appears to be to an observer, and even what its behavior might be when its e-stop is pressed or other safety-related inputs are engaged.

DYNAMIC BEHAVIOR GENERATION

This outlines the level of interaction that exists for a person to affect how the robot will perform its task(s) while the task, or behavior, is underway.

Level 0: Autonomous completion of pre-planned tasks, limited human-robot interaction for task scripting only, before operation.

Level 1: Fixed workstation collaboration. Humans can continuously communicate, command, and interact with the robot through an external UI.

Level 2: Robot can interact with humans directly (e.g., verbally, gestures) and continuously through prescripted menus or other interface structures.

Level 3: Robot can interact with humans directly (e.g, verbally, gestures) and continuously, not requiring any user interface. (human-equivalent interaction)

DYNAMIC PAYLOAD COMPENSATION

How capable the robot is at adjusting its planned motions based on the payload that it is carrying. Here, the term "Hard-coded" refers to pre-identified, assumed inertial parameters of the payload, while "Adaptive" refers to online identification and/or compensation of measured inertial parameters of the payload.

Level 0: No dynamic compensation

Level 1: Hard-coded gravity compensation of known payloads

Level 2: Hard-coded gravity and inertial compensation of known payloads

Level 3: Adaptive gravity compensation of measured payload parameters

Level 4: Adaptive gravity and inertial compensation of measured payload parameters

APPLIED LEARNING

This captures the extent models learned from data determine the behaviors of the robot.

Level 0: Basic autonomy, following pre-programmed controls and discrete logic for behavior generation. No learning.

Level 1: Models have been trained and employed for specific uses/processing, but overall behavior generation is still explicitly coded.

Level 2: Underlying behaviors are still traditionally controlled, but high-level action planning is done by a pre-learned policy.

Level 3: Executive action planning is done by a policy that continues to learn in operation.

Level 4: End-to-end learning, or at least all processes being effected and governed by an integrated, continuously learning policy.

LOCOMOTION PLANNING

This capability captures the extent to which a robot can plan and then follow a path plan in a variety of environments. More complex environments require more complex sensing, planning, and motion control algorithms.

Level 0: Can follow a pre-planned set of footsteps.

Level 1: Periodically replan footsteps towards locomotion goal on simple terrain (flat, no obstacles)

Level 2: Can replan footsteps based on static, complex terrain (knowledge of terrain can come from onboard vision/mapping or an oracle)

Level 3: Can replan footsteps based on changing, complex terrain, including added obstacles, fragile ground / infeasible footstep locations, or physical disturbances.

Level 4: Can plan whole-body environment interactions, utilizing both hand and foot interactions for optimal plans.

E-STOP, FAULT, OR POWER-LOSS BEHAVIOR

The fault-response behavior of the robot. Risk cannot be completely removed, but more mitigation factors can reduce the risk posed to nearby users and the environment. An individual robot would be classified based on the highest level of behavior it is capable of, even though it may be able to demonstrate lower-level behaviors.

Level 0: Immediate power cut to all components.

Level 1: Brakes (either powered or unpowered) are employed to enforce passivity/dissipation of total energy in the system.

Level 2: Robot actively reduces its footprint / size / effective endangered area by curling up, leading it to cover pressure points and other dangerous surfaces on the robot. (Action is irrespective of surroundings)

Level 3: Robot views surroundings, then plans and takes action to quickly and as safely as possible move to a safe robot state. This specifically addresses the hazard, e-stop, or other fault that initiated the hazard-response state.

Level 3-1: Robot additionally maintains power enough in joints, even after a main power source loss, to enact these safety actions.

Level 4: Robot is always operating in a way such that failure of components or power loss will not endanger nearby humans or other designated protected objects. (e.g., maintain a pose at all times such that a power-loss collapse will occur away from a human)

Application: Interactions and Risk Features

An interaction and risk-based classification addresses many of the key safety concerns surrounding the environments in which a humanoid robot may exist. Robots' interactions in various types of environments induce different levels of risk. With the following features, risks are described qualitatively with a comparative scale based on how the robot is expected to interact with the environment.

When a robot does not have to interact with humans in an environment, it can act autonomously. Multiple robots can communicate individually or with a centralized control. When in a mixed environment, meaning both humans and robots are present, the risk depends not only on the robots' control but also on the background of the present humans. Having only trained humans in the environment means they know how to behave in a shared space, are aware of the instructions, and are adept at handling different scenarios.

Mediated human surroundings refer to a controlled human environment with a mix of trained and untrained participants; the trained participants would take the lead or supervise whenever an interactive situation occurs. A public environment is completely uncontrolled, where participants do not take specific actions and may be unaware of how to respond across a diverse range of interactions. The type of environment, and thus expected interactions, is also dependent on the application the robot is being used for.

INTERACTION COMPLEXITY

Refers to the variety of other actors that the robot must be able to interact with, including both other robots and humans of various amounts of experience.

Level 0: Fully Separated - inline with current industrial standards (e.g., traditional industrial robots in fenced zones with light curtains, safety PLCs triggering shutdowns). Interaction is prevented.

Level 1: Interaction with other robots (e.g., fleets of AMRs in a warehouse).

Level 2: Interaction with trained humans (e.g., collaborative robots on assembly lines working alongside knowledgeable staff, potentially involving remote operation).

Level 3: Interaction with untrained humans (e.g., service robots in public spaces, delivery robots on sidewalks, potential future home assistants).

Level 4: Interaction with Vulnerable Populations (children, seniors, medical patient, etc.)

INTERACTION TYPE

Refers to the mode of interaction that the robot must be capable of, from digital communication with a central server, to physical communication and interaction with humans.

Level 0: Interact with no other robots or humans

Level 1: No direct interaction, but communicates and receives commands from a central fleet manager system.

Level 2: Direct Communication with other robots

Level 3: Direct Physical Interaction with other robots

Level 4: Direct Communication with humans

Level 5: Direct Physical Interaction with humans

APPLICATION DOMAIN

This refers to the general application the robot will be used for. It implies the environment, interactions, expected actions, and capabilities the robot must have, as well as the most common disturbances, safety considerations, and design requirements that are expected from a robot in this application domain. This classification breaks with the numeric, performance-based metric structure.

- Industrial - The robot is used for job-related applications, often in structured environments with set tasks.
- Residential - The robot is for personal use in homes. The environment is unstructured and/or changing, but the tasks may remain constant.

- Service - The robot is implemented and used by an owner to interface and interact with the public.
- Public - The robot must autonomously act in an unstructured environment, and adjust plans and actions based on observations of the surroundings.

  .....

Note: Many other application domains could be listed, such as construction, medical, underwater, etc. The above is only a sample of possible application domains.

Classification of Humanoids

Humanoid robots, or rather “robots with human features”, are not necessarily novel in their capabilities or features, as many other robotic systems can share those details.. Humanoids do, however, indicate a clear expansion of the types of environments robots will be deployed in. Deployable applications will move from isolated cages into shared spaces, necessitating layers of new safety and performance requirements. Currently, robots existing in shared spaces are typically either small or purely animatronic. Robots capable of the speed and strength of industrial robots are usually completely isolated in cages or at least only deployed in industrial environments around trained personnel. Humanoid robots will potentially have the characteristics of today’s industrial robots while also existing in shared spaces around untrained people. A classification of humanoids could thus be used to understand the safety expectations and performance requirements associated with the robot and its behaviors in these environments. Consumers want to know what to expect when products come to market, and manufacturers want to know what kind of capabilities and functions need to be built into their robots.

Even though humanoid robots do share characteristics with other robotic systems, it can still be useful to have a classification of humanoid robots, based on features common to all humanoids, or features that are unique to humanoids. One such unique feature is humanoid robots’ humanlike appearance. A classification based on visual features is hard to define, especially when those features may not provide any function or have questionable use. However, one potential method of classification, as outlined below, is based on the purpose of the human form, which could be used to sort the different humanoid robots that are currently being designed and manufactured. It is useful for capturing the intent behind the creation of the robot, describing why the robot that is used for an application should be a humanoid rather than a more classical robot structure. Note again that this is an example of a classification approach for humanoid robots, and is meant only as a starting point from which future classification efforts can begin.

REASON FOR HUMAN FORM

For what reason has this robot been made to look like a human?

Class 0: Has no human form

Class 1: Purely visual appeal (appearance does not convey the expected function or capability of the robot)

Class 2: Leveraging form to affect the reaction (for example: physical, emotional, psychological, etc.) users have to the robot

Class 3: Acting in environments built for humans

Class 4: Replicating human capability for general action

Conclusion

The effort to develop a classification for humanoids uncovers the larger need for a robotic classification in general. The classifications presented above are example strategies for classifying all robots. However, because the classifications are based on features that humanoid robots also possess, the same classifications can be used to sort humanoids by capabilities and features. Once a classification for all robots has been created, then a classification for humanoids can be created within that context, adding features that are unique to humanoids, such as the similarity to human form, as discussed above. Furthermore, in the interest of engaging a broad range of participants in the standards creation process, if any readers had a strong reaction, opinion, critique, or idea about these example classifications, they should join the on follow-up committee that is creating a standard on robot classification after this report.

Chapter Bibliography

- Kirschner, Robin Jeanne, et al. "Categorizing robots by performance fitness into the tree of robots." Nature Machine Intelligence (2025): 1-12.

- Prestes, Edson, et al. "Towards a core ontology for robotics and automation." Robotics and Autonomous Systems 61.11 (2013): 1193-1204.
- Kunze, Lars, Tobias Roehm, and Michael Beetz. "Towards semantic robot description languages." 2011 IEEE International Conference on Robotics and Automation. IEEE, 2011.
- Kim, Stephanie, Jacy Reese Anthis, and Sarah Sebo. "A taxonomy of robot autonomy for human-robot interaction." Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. 2024.

The Role of Stability

Authors: Ben Beiter, Brandon J. DeHart, Federico Vicentini, Francisco Andrade, and Dieter Volpert

For stability, in this chapter, we are considering classic humanoid robots that are bipedal, powered, and actively-balancing. Such robots differ from other types of robots for which we have standards by the fact that they can only stand through powered balance and can change the shape and position of their support polygon relative to the rest of their structure during normal operation (i.e., they can take steps). These mobility features provide the agility, range of possible motions, and responsiveness that give humanoids the highly varied potential use cases they have. However, these same capabilities are also the main source of hazards caused by humanoid-type robots. At all times during operation, a loss of power or too large an unexpected disturbance could result in the robot toppling over. This could harm the robot itself, any parts of its environment, and any nearby people. As a result of this risk, stability-related safety concerns are a main barrier to the adoption of humanoids in any space shared with humans.

The proper way to address these safety concerns would be through a risk assessment that identifies all potential hazards and the appropriate safety measures that must be taken in response to those hazards. Ideally, technology- and application-specific standards would be available to guide readers (standards developers, policy makers, robot researchers, etc.) through the risk assessment process. Currently, however, there is a lack of standards that provide the tools needed to quantify the level of risk or validate the effectiveness of safety functions on actively-balancing robots. The entire burden of proving that the robots are safe enough is thus on the manufacturer of each robot. Creating standards for evaluating stability will ease this burden on manufacturers. Such standards will be used to measure and prove stability during key applications, allowing customers to have confidence in the safety of the robot, as well as allowing manufacturers to measure performance on tasks and prove utility for the applications that customers want to use them for. Creating practically useful standards is the necessary first step towards creating certifications for humanoid robots.

An example first step is to address the fact that standards currently do not have a cohesive definition of stability. The word itself means slightly different things in different contexts. The technical engineering definition of a system state settling to a defined equilibrium, the colloquial sense of a system able to act and interact with the environment without harming anything, and the specific humanoid sense of being able to walk and balance without falling over, or at least managing the risk of toppling in the same way humans do. In addition to the varying types of stability, there are two aspects of stability to be considered: safety and performance. Safety considers the avoidance of harm to people in the vicinity. This can be considered a minimum requirement and more of a pass/fail behavior. Performance considers evaluating capabilities to both perform intended tasks and respond to changing conditions, which can also include measuring potential risks posed to the robot itself and the environment as a desirable byproduct.

Despite not being inherently stable, humanoid robots are expected to be able to perform a wide variety of tasks while also being robust to a wide variety of environments and external disturbances. On the other hand, the complex structure and walking capability of these robots give them many options in how to respond to events and disturbances (e.g., navigating variable terrains or stepping to avoid falling rather than just leaning to balance). Current robots have displayed increasingly impressive capabilities as they approach applications that live up to the high expectations of humanoids. However, the variation in how a robot might respond to certain situations, as well as the variety of control approaches that drive such behaviors, makes evaluating the stability of humanoid robots more challenging. Additionally, because the functioning of the robot is so dependent on the performance of the control algorithm, if a specific combination of task, robot, environment, and inputs has not been seen and tested before, any standardized validation space would be very broad, at the limits of practicability.

There is a need for metrics, test methods, design procedures, and implementation guidelines for humanoid robots that can guide the use of humanoid robots in workplaces. This chapter outlines a recommended multi-layered approach for creating standards that help quantify the high-performing potential applications of humanoids and provide requirements to verify the safeguards implemented to protect humans, equipment, and the environment in which the robot is deployed. Section 1 reviews current standards and prior research related to humanoids. Section 2 outlines the recommended roadmap of stability-related standards that will be detailed. Within that roadmap, Section 3 discusses the need for quantifiable performance metrics and methods to aid in understanding humanoid robot motions, and Section 4 discusses how the stability standards are closely related to the development of safety standards as well. Section 5 provides concrete recommendations for Standards Development Organizations (SDOs), and Section 6 concludes the chapter.

Review of Current Safety Standards

For robotics, considerations of safety begin with safety standards for machinery. Safety for machinery is implemented in two steps: risk assessment and risk mitigation. Risk assessment for machines is guided by ISO 12100, with other standards covering individual classes of risks, such as those derived from electrical hazards (IEC 60204-1). The fundamental hazards for machinery in ISO 12100 are extended in product-specific safety standards like industrial robots (ISO 10218-1 and -2, recently revised), personal care robots (ISO 13482), and others. Risk mitigation is often done in one of two ways: either avoiding a particular risk entirely through safety requirements or actively avoiding it during operation through safety functions. Some safety requirements inside the product-specific safety standards point to technology-specific or narrow standards, for instance, the ISO 13850 on Emergency Stop or the ISO 13855 on safe distances from static and moving hazards. Product-specific standards address vertically the risk assessment for robots, the selection of equipment, both electrical and mechanical, the properties of controls and power sources, the mandatory safety-related stops, mechanical and non-mechanical physical or virtual safety guarding to protect people, motion limits such as avoiding kinematic singularities or speed and separation monitoring, labelling, instructions, and more. When these protective measures are implemented with devices dedicated to the reduction of risks via the monitoring and active control of reaching or maintaining safe states, then Functional Safety is implemented following IEC 61508 (or its derivation for machinery ISO 13849). Altogether, these standards establish general requirements for the safety of machinery, and robots specifically.

Beyond these general safety standards, ISO and other SDOs have produced several safety standards that address more subcategories of robots. However, these standards make an unwritten assumption that the base of the robots being considered is either fixed or has a statically stable base. Due to this, none of these specialized safety standards apply to humanoids. For humanoids, even with all the joints braked and powered off, which is a very conventional safe state for fixed manipulators, the robot is not necessarily in a safe state. In general, it must be assumed that a locked pose is unstable. This potential instability, present in normal conditions, means that the robot can become unbalanced and fall over even with high-performing controllers. Current standards do not adequately address stability requirements, and so a dedicated series of standards is now recognized as a market need.

The only standard that does not exclude robots with actively controlled stability, including legged robots, from its scope is ANSI/RIA R15.08-1. This US national standard provides safety standards for industrial mobile robots (IMRs), including their navigation and control. R15.08-1 is uniquely silent concerning the mobility principle of robots, and legged robots with manipulators could fit in the scope. The standard defines “Type C IMRs” when manipulators are integrated with a generic mobile base, which is the mobility representation closest to humanoids. While R15.08-1 states that “The IMR shall maintain stable operation during travel” and includes some methods to validate this stability condition, it only considers that instability could be caused by a payload that is too heavy and/or held too far away from the base of the robot, causing the IMR to tip over. These requirements are derived from the risks analyzed for single manipulators mounted on top of moving bases.

This limited approach to stability is mirrored in other available safety standards. The ANSI/ITSDF B56.5 for industrial trucks (e.g., AGVs) includes dynamic stability considerations, but they only account for the inertia of the vehicle and do not cover any balancing robots with a changeable base of support. B56.5 also states that “The user shall be responsible for the load stability and retention. When deemed necessary by the user, verification shall be required,” though it does not present any test methods or requirements for controlled stability.

For manipulation, no safety standards directly address the effects of manipulation on stability. All requirements for robot arms are in ISO 10218, while ISO/TR 20218-1 further illustrates hazards for end effectors and their interfaces, stating, for instance, that robots should ensure “that loss of power does not lead to loss of load, unexpected motion, or other hazards”. This might be straightforward for fixed-base manipulator robot arms or whenever the reaction energy is completely absorbed or balanced by a stable base, but humanoids do not follow this rule, as losing power at any time could result in falling over.

Humanoids are far more complex in the aspects of dynamic stability and the use of limbs for and during balance. A dedicated family of standards that address stability for actively-balancing and legged robots would greatly contribute to the clarity of specifications and requirements about manipulation and balance.

In the domain of Service Robotics, IEC 63310 on Active Assisted Living Robots states, “AAL robots with assisted mobility functions other than wearable or wheeled ones should have the ability to finish their intended tasks.” ISO 13482 on Personal Care Robots simply states that robots should have sufficient stability: “The personal care robot shall be designed to minimize mechanical instability (e.g. overturning, falling, or excessive leaning when in motion) due to failure or reasonable foreseeable misuse.” It references ISO 7176-1 and ISO 7176-2, which are about static and dynamic stability for wheelchairs. Again, these standards only determine that the device should not be destabilized while in normal use. It does not address robots with dynamic self-balancing or the ability to alter their base of support during operation.

Even further from industrial and service robotics safety standards, a potential use of humanoid robots appears in the Technical Report ISO 4448-1 about Public-area Mobile Robots (PMRs). This report is form-agnostic and explicitly includes legged robots, as well as the more common wheeled mobile robots, and includes safety considerations for the application of automated pick-up and drop-off (PUDO) of goods and people at the interface between roadways and sidewalks (the “kerb”), in addition to the behavioral requirements for PMRs operating in pedestrian spaces, including sidewalks and roadways. The core goal of this report is to provide a framework for municipal, provincial/state, and federal governments to establish regulations surrounding the operation and use of automated systems performing PUDO tasks. Interestingly, the report introduces the use of a “shy distance” which dictates how far away a PMR should attempt to stay at a minimum from any “inattentive, uninvolved, unprotected, and untrained bystanders”. For comparison, all existing mobile robot standards do not dictate any predetermined separation distance, but refer to a more sophisticated computation of sufficient space to come to a safe stop or prevent hazardous collisions. Additionally, the problems of exposure of the general population and outdoor conditions are very challenging.

Overall, most standards require that robots, whether mobile manipulators or personal care robots, must not cause undue hazards due to instability during normal operations, nor should maintaining stability increase the risks of already existing hazards. First and foremost, “preventing hazards” due to loss of stability is not an option for actively controlled robots. By the very nature of actively controlled stability, hazards are not eliminated. Only risk reduction and the evaluation of residual risks of instability are eligible considerations. Second, and challenging from both technical and regulatory standpoints, balance-related actions are part of the safety system in the scope of functional safety. A safety function is the capability to detect a particular hazardous effect (i.e., a quantity) through sensing of internal states (e.g., a fault) or external phenomena (e.g., an obstacle), then enact corrective actions that successfully either avoid any hazards or reduce the consequences of the hazardous event to acceptable levels. Conventionally, the primary approaches to establishing confidence that safety-related control actions are sufficient to reduce risks are by measuring their ability to control or avoid failures according to standardized levels of residual failures. This metric is established in IEC 61508 with Safety Integrity Levels (SIL), and inherited with variants by the derived sector-specific functional safety standards like ISO 13849 for machinery with Performance Levels (PL). Both SIL and PL measures are based on intervals of Probability of Dangerous Failures per Hour that are suitable for random failures and are paired to a corresponding level of control or avoidance of systematic failures (Systematic Capability). These levels are used to create requirements for designing increasingly robust safety functions intended to reduce risks that are proportional to the frequency and intensity of the associated hazards. The higher the risk, the more demanding the protective function must be.

However, these functional safety standards are progressively not fully adequate when hardware and software failures are intertwined. This is the case of actively controlled stability and perception of the surrounding environment that is instrumental in achieving successful locomotion and navigation.

For example, typical top-risk safety functions for industrial robots are mandated in ISO 10218 to achieve a PL d range of residual dangerous failures, specifying the required range of those probabilities depending on the hardware architecture (structure category 2 or 3), when expressed per ISO 13849 - or to reach SIL 2 with a hardware fault tolerance of 1, when expressed in accordance to IEC 61508 or IEC 62061. This approach is deeply understood and consolidated for relatively simple functions. However, the dynamically balanced nature of humanoid motion means that any safety function related to stability will have varying results and a chance of success depending on both the state of the robot and any external factors that might be present. Recovering, or just maintaining balance, requires a complicated series of sensing, internal modelling, motion planning, and control capabilities that are harder to characterize as SIL or PL safety performance levels. Additionally, there are currently no test methods or procedures to measure performance and validate the capability of balance safety functions. To have full confidence in the performance of stability safety functions, more than just the standard SIL and PL process is needed.

Hazards related to instability are, in fact, not well expressed in terms of rates of failures per hour, but rather as events that depend on the execution of actual behaviors and the conditions of the environment. Hardware and software components implementing safety functions can be characterized by distributions of failures with time-based rates, but aggregated behaviors are better expressed in terms of the rate of success, and risks may still arise even without system failures. This scenario is common in the domain of autonomous vehicles, where methodology for the analysis of functional insufficiencies is a preliminary step to instruct safety functions to address the ultimate execution of a safety action (see ISO 21448, in combination with the functional safety in ISO 26262). Similarly, assessing the safety of complex systems like humanoid robots requires a comparable evolution in approach, focusing not just on preventing/detecting malfunctions, but on ensuring an acceptable level of safety of their intended actions within unpredictable environments.

Finally, an important consideration for dynamically stable robots is failure handling, even during the execution of a safety function. What does the robot do when something goes wrong? Most safety functions measure success by the capability of the system to return to a safe state. Most safety standards have the phrasing that “No new hazards shall be introduced during a stop”.

However, if it is already moving, a humanoid can require additional motion planning and/or intermediary actions to reach a stable, safe state. These intermediary actions (e.g., recoiling, balance compensation by parts of the body) could add new hazards or concerns, even if they help the robot avoid falling over. Guidance from safety standards for application-specific environments will be needed to determine the proper behavior in certain situations, and thus be able to properly consider humanoids in risk assessments.

In conclusion, current standards created requirements for safety with the underlying application domains and their risks in mind (collaborative applications, service robots, AGVs, AMRs, autonomous vehicles, etc.). Those standards cannot be applied directly to humanoids, where the problem of stability affects all safety-critical functions and the degree of successful resolution to safe states. Dedicated risk assessment considerations and analysis of functional insufficiencies are not common in the domain of robotics but are concluded to be necessary for complex dynamic behaviors and highly variable autonomous functions. On top of this, even the ability of humanoids to attain balance and/or recovery implemented as safety-related parts of control systems, there is a lack of standardized test methods to validate such safety capabilities.

All of this should be addressed early in the standards creation timeline. Currently, it is part of the mission for the newly created working group in the ISO Technical Committee 299 tasked to develop a safety standard (ISO/AWI 25785-1) for industrial mobile robots with actively controlled stability, notably including humanoids.

Review of Current Test Methods and Performance Standards

Beyond safety, there is a need to be able to measure the performance of a humanoid. Safety is the avoidance of hazards, and especially doing harm to humans, while measuring performance is how capability and productivity can be quantified and validated. Standardized metrics and test methods are the tools to measure performance. While there are some ongoing efforts to create performance test methods for humanoid and similar systems, such as through ASTM International Subcommittee F45.06 on Legged Robot Systems, there are, so far, no published standard test methods that consider humanoids specifically. However, because humanoids are general-purpose robots, meant to perform a wide variety of tasks, current task and application-specific standards could be used to measure certain aspects of a humanoid's capabilities, although users of standards would need to make an extra effort to interpret and adapt them.

For example, ISO 9283 focuses on establishing the accuracy and repeatability of industrial robots applied to manipulation tasks. These tests are mostly applied to fixed-base robot arms, with accuracies on the order of millimeters. Humanoids can also perform manipulation tasks; however, this particular standard is limited in that it only considers open-loop performance, not covering the integration of the large number of sensing systems that humanoids have. It also only focuses on end-effector positioning, not addressing robots without a fixed base. This makes the test methods not fully applicable as is, because unmeasured error in the positioning of the base will interfere with manipulator performance. Similarly, ISO 18646 is a series of performance measurements and test methods for service robots. The tests do not consider stability beyond static loads offset from the COM or inertia from dynamic motions, which cause tip over, with relevant work undertaken in the ISO/DIS 18646-5 project (which is in final development stages at the date of writing). Overall, for all robotics test methods that only consider fixed-base robots, while the method itself may be able to be accomplished by the humanoid, balance and stability performance will affect the overall results of all of those test methods and must be considered.

Another source of applicable performance standards is the set of test methods created for other systems that are not humanoids, but have some common aspects in mobility-agnostic application scenarios. For example, the ASTM International Subcommittee E54.09 on Response Robots published 20+ test methods that define many apparatuses, tasks, and procedures that do not depend on locomotion principles or capabilities. Other legged robots (primarily quadrupeds) are regularly tested with these methods, which cover many elemental behaviors included in navigation, coarse manipulation, and inspection. Many of these methods, with some possible alterations for scale and difficulty, would be directly applicable to humanoids, some even being useful for testing stability and balance. For most standard test methods considering mobile-base robots, there is nothing keeping them from being used to evaluate humanoids as well; there is just a need for more and more targeted methods and metrics to evaluate the specific template motions that a humanoid employs. Metrics such as these will need to evaluate the performance of both humanoids as a whole and their individual subsystems and capabilities.

As inspiration for new standard test methods, there have also been several academic projects dedicated to evaluating humanoid robots in the past. There are many papers that discuss humanoid robot benchmarking, though most focus on finite aspects of humanoid robots such as WHole-Body-Control approaches, learned control policies, specific tasks, or individual robots (as a non-exhaustive list). Of note is the EUROBench Project, which was a 5-year, multi-institution project that set out to create a unified benchmarking framework for robotic systems in Europe. Various sub-projects generated test methods for many aspects of robotics applicable to humanoids, from whole-body manipulation to balance. Of all the research mentioned here, however, none has yet to make it into any standards as metrics, test methods, or practices. These projects are starting points for future committees to begin forming standards for evaluating human performance.

Stability Standards Roadmap

To address the challenges that stability presents for the adoption of humanoid technology, we recommend an approach that:

- Keeps accruing experience in the use of humanoid robots in the presence of humans, without entirely blocking the adoption of technology due to the lack of dedicated standards.
- Coordinates and fosters the definition of practical metrics, test methods, and validation conditions to fully evaluate humanoid stability. This element is very important for the quantifiable and objective evaluation of stability and for establishing acceptable limits for safety.
- Creates dedicated safety standards. Safety standards have a deep influence on regulatory matters, so clarity and limited room for interpretation would be beneficial.

To establish both measurement and safety requirements, the to-be-proposed standards may share key information currently existing for other robotics applications. In particular, performance standards elaborating testing criteria and methods with efforts undertaken by various SDOs. This will allow the definition of minimum stability performance thresholds. The same metrics will allow for potential live monitoring of stability, with the capability to enter safety-critical control modes if certain conditions are met.

Path Forward

Across all prior research and standards, we have found that despite balance/stability control being done on all humanoid robots, balance/stability measurement has not been published in any standards yet, though several programs for standard development are ongoing. Given the body of knowledge, however, stability criteria should become part of accessible standards with a high potential for benefit to the humanoid industry. Despite this clear goal, implementing a codified set of criteria in technical standards is still not simple.

Therefore, we recommend a two-part process for creating new standards focused on humanoid stability performance. The first is on measuring and quantifying stability. This first effort is focused on creating test methods to test stability performance in a variety of tasks and conditions. These tests will become the foundation for overall humanoid performance testing in the future. This effort will also focus on creating usable stability metrics. The second part is on developing safety standards for humanoids. This will involve building upon the current approaches to robot safety to extend to the particular hazards presented by humanoids. The design of safety standards will come from two directions. One will build on the test method and metric creation, where each test method can establish safety thresholds for each task, and the other will be a new approach to standards, looking at integrating safety requirements into the controller itself at the design stage.

TABLE 1: A SUMMARY OF CONTRIBUTIONS NEEDED IN NEW STANDARDS, AS WELL AS RECOMMENDATIONS FOR ASSOCIATED NEXT STEPS TO DEVELOP THOSE STANDARDS

NEEDED STANDARDS	RECOMMENDATIONS
METRICS AND TEST METHODS THAT CAN VERIFY REQUIRED PERFORMANCE CAPABILITIES WHEN A ROBOT ENTERS AN UNSAFE STATE.	Consolidate metrics and test methods to evaluate stability performance Identify the subset of metrics and stability tests that are safety-critical Create specific validation configurations and scenarios
AGREED UPON STABILITY METRICS FOR ONLINE MEASUREMENT OR ESTIMATION OF SAFETY-CRITICAL STABILITY	Consolidate metrics and test methods to evaluate stability performance Identify the subset of metrics and stability tests that are safety-critical Create specific validation configurations and scenarios
AGREED UPON CONDITIONS FOR SAFETY-RELEVANT LOSS OF STABILITY	Discuss safety standards for potentially unstable mobile robots Establishment of minimum thresholds of task performance required for safety, depending on environmental and application hazards Online measurement of stability that a robot can react to, depending on the safety requirements

NEEDED STANDARDS	RECOMMENDATIONS
RISK EVALUATION OR SAFETY TARGETS FOR THE IMPLEMENTATION OF ROBOTS, SUB-JECT TO INSTABILITY IN A SPACE THAT HUMANS MIGHT ALSO OCCUPY	Discuss safety standards for potentially unstable mobile robots Establishment of minimum thresholds of task performance required for safety, depending on environmental and application hazards Online measurement of stability that a robot can react to, depending on the safety requirements
ACCEPTABLE BEHAVIORAL REQUIREMENTS TO AVOID SITUATIONS THAT COULD LEAD TO HARM TO HUMANS FROM A ROBOT TOPPLING OVER.	Discuss safety standards for potentially unstable mobile robots Establishment of minimum thresholds of task performance required for safety, depending on environmental and application hazards Online measurement of stability that a robot can react to, depending on the safety requirements

Quantifiable Performances: Test Methods and Metrics

A key need to be addressed for humanoid adoption is the capability to measure and validate stability performance, which captures the stability of the robot while interacting with any environment. This scope is broader than human safety. We need to consider the risk of losing stability posed to the robot itself and other surrounding equipment within the environment. Like safety for humans, we cannot practically guarantee no risk to the environment from an underactuated robot moving through it, but demonstrated performance on a set of representative tasks can establish confidence that future performance will be both safe and high-performing enough. To quantify this performance, there are two necessary parts: (i) test methods and guides for global attributes that a robot is to be evaluated on; (ii) objective, observation-based metrics to score the stability of a robot during specific tasks.

The goal of these performance standards is to provide a way for manufacturers, customers, and regulators to understand, evaluate, and compare the performance of robots and the subsystems that make them up. Customers wish to purchase robots that meet their application requirements, manufacturers wish to shape designs to meet these requirements and provable improvements in performance capability, and regulators want to be able to set thresholds for performance in different environments that can eventually enable guaranteed safe implementation of humanoid robots in the highest complexity, but also highest value, areas (e.g., humanoids walking in complex, crowded spaces alongside humans, able to collaborate, etc.).

Test Methods and Task-Based Metrics

Standardized test methods present repeatable robot tasks and metrology procedures for measuring performance. Such test methods detail a procedure to follow during a test, as well as any physical apparatuses that must be created to conduct the test. These methods typically focus on a particular task to be performed, and include any repeatable disturbances or variations in the test that should be used in the evaluation.

To apply a similar strategy to evaluating the balance/stability of humanoid robots, we recommend the creation of a suite of stability-related test methods. The tasks that these methods evaluate should be of significant variety to cover all the most probable situations that a humanoid robot will face in the workplace. Variations belong to three categories:

- behaviors to be evaluated, including templates of operations or stereotypical situations;
- changes in the environment in which these tasks would have to be performed;
- robustness to common disturbances that the robot might be exposed to in such configurations.

Table 2 below lists a sample set of test methods that could be developed for a humanoid robot that would be implemented in simple load pick-up, move, and place tasks in a standard industrial environment.

TABLE 2: EXAMPLE TEST METHODS TO EVALUATE STABILITY IN HUMANOID ROBOTS.

TEST METHOD	DESCRIPTION	METRICS
STATIC BALANCE TEST	Robot maintains a static pose on various surfaces (flat, inclined, uneven) under external disturbances (pushes, pulls).	Maximum disturbance force/ torque, sway angle, time to recover balance.
DYNAMIC WALKING TEST	Robot walks on flat, inclined, and uneven terrain, navigating obstacles and maintaining balance.	Walking speed, step length/ height variability, number of stumbles/falls, energy consumption.
LOAD CARRYING STABILITY TEST	Robot carries varying payloads (weight, size, center of mass) while performing static and dynamic tasks.	Maximum payload, stability margin, task completion rate, fall rate with load.
MANIPULATION STABILITY TEST	Robot performs manipulation tasks (picking, placing, assembling) while maintaining balance and resisting interaction forces.	Manipulation accuracy, task completion time, stability margin during interaction, force limits.
FALL RECOVERY TEST	Robot is intentionally destabilized or falls, and its ability to recover to a safe state (e.g., upright, curled) is measured.	Recovery time, success rate of recovery, impact forces during fall, final safe pose.
HUMAN INTERACTION STA-BILITY TEST	Robot interacts physically with a human (e.g., handoff, guiding) while maintaining balance and ensuring human safety.	Interaction force limits, stability margin during interaction, human comfort/ safety ratings.

Task	Balance (no stepping) Walking Forward, Backward, and Sideways (at varying speeds) Turning in Place Carrying a Load (of varying weights and in various poses) Manipulation (e.g., door opening)
Environment	Flat Terrain Sloped Terrain (steep or shallow, facing uphill or downhill) With obstacles (of varying size) Uneven terrain (varying slopes, steps, and obstacles) Narrow walking path (e.g., balance beam) Narrow walking area - confined space (corridor walking) Step-under tasks
Disturbance	Push/Pull to Pelvis/COM* Push/Pull to Torso* Push/Pull to Lower Body* Payload suddenly changing Encountering a person or high-value equipment * Vary by magnitude of disturbance and direction of disturbance relative to the robot and task

When creating these validation methods, the procedures should be flexible enough to allow robots of different motion capabilities and morphologies. The tests should be practical and rigorous, easy enough to set up and conduct so that they can be widely used, but still difficult enough to accurately differentiate robots of various capabilities.

From each test, numerous evaluation data points can be taken, such as success rate, speed of execution, accuracy, etc. Each of the individual example tests that could be created can be considered to be a basic capability. Since a humanoid is a general-purpose robot, more complex behaviors can also be tested by combining multiple of the basic capability tests outlined in Table 2. Such combination tests should be performed if the application requires specific use cases that can be evaluated.

One limitation of this approach tied to test method performance is that the total number of possible permutations of tests is very high, more than the number of test methods that can realistically be written. This is a limitation of all demonstrated behavior-based measurements of performance. Especially for humanoid robots, which are performing multiple tasks at once, it has not been explicitly proven that just because the robot can perform the pre-defined test methods well, it can also perform well on the tasks in a real work environment.

The correlation between performance on test methods and performance in real work is still meaningful, and more research should be done to understand the uncertainty in this relationship.

Beyond template operations to be accomplished, there is also a need to evaluate intrinsic capabilities of the robot that are not directly measured by the statistics of application-related scenarios. For example, capabilities like swinging arms to aid in balance while walking, state estimation of loads when picked up for manipulation, or replanning footsteps to adjust to disturbances. All of these properties affect stability, but they are only indirectly captured in basic test methods. Specific test methods can be created for any special capability of the robot and should be used to understand how that feature impacts the overall robot, possibly without compromising any safety constraints.

Control-Based Metrics

While test method performance is the first and most straightforward way of evaluating humanoid capability, completion of those tasks requires a combination of control, modelling, planning, and behavior generation functions. Walking stability is an inherent part of that system, but it can be hard to isolate the specific stability performance when only evaluating by task performance. As such, targeted metrics for stability performance need to be identified or produced that can be implemented during tests. To be standardized, such metrics should be repeatable, robust, and accurate enough without requiring too expensive equipment to measure.

Stability and balance control of humanoid robots has been a well-researched topic for several decades now, and so there are various metrics that have been previously presented that a standardization process can start with. A few examples are: the instantaneous capture point (ICP), zero-step-capturability, the margin of stability, sway angle, or maximum allowable angular momentum. Most measures of stability, including these, are instantaneous measures of stability, able to interpret the current state of the robot and identify if it is currently unstable, or how close to unstable it is. This type of metric is useful for evaluating the performance of a robot after it has completed an action, as the instantaneous stability could be evaluated throughout the motion. However, to meet requirements laid out in current standards, that no motion or action should destabilize the robot, some amount of prediction of immediate future performance of the robot, given a planned set of motions, is required, especially for more dynamic actions such as walking. Many state-of-the-art humanoid robot control approaches already incorporate some amount of prediction, such as with optimal control or model predictive control (MPC) based approaches. To fully evaluate stability performance, metrics will need to incorporate predictions of possible future dynamic states and what is acceptable as “stable enough”.

The closest current works to forward-looking stability metrics are control approaches that include “safety” constraints, or dynamic stability constraints in the motion planner, and in the executing controller. Examples of such constraints are one-step-capturability and stability regions [2]. These constraints are based on calculating the regions around the robot during a foot swing phase in which a robot would be able to stabilize itself if it places its foot down within the region, and then to always ensure the robot can reach this region. However, these constraints end up as simply necessary criteria for stability, rather than a true metric of stability that could be maximized in a motion planner, or reported as a metric of performance. These stability constraints could be useful in creating a standard that establishes the minimum criteria that must be met to be considered a stable action. Such a standard would only be a first step towards establishing standards for stability quantification and measurement.

Beyond knowing when a robot is unstable, there is a need to know how stable a robot is, such that other motions and interactions can be planned accordingly. For industrial and other applications in human spaces, reliability and safety are primary concerns for which measures of stability and an understanding of how stability determines behaviors are needed. Higher stability and more robust motions may be required, or desired in some situations, while lower stability, but higher performance actions may be acceptable in others. While a standard measure of stability is not needed to design the structure or capability of a robot, it would be extremely beneficial for risk assessment and proper implementation planning. Additionally, just like humans walking, humanoid robots in real-world spaces would be subject to disturbances, interruptions, and other unpredictable events. As such, the goal of these metrics is to have a way of continually evaluating the performance of the robot, not only during validation tests, but also during real-world applications as a monitor of performance.

Finally, an additional need for stability metrics in the creation of future standards is for realizing safety requirements in non-standard tasks. The simplest way to prove that a robot can safely perform a task is to establish a standard test method evaluating the performance of that task, and then conduct trials validating the required capability. However, it is impossible to test every action that it is desired for humanoids to accomplish, so there must be measures of stability that can be applied to any generalized task. Showing performance with these metrics will then allow for meeting the guarantees of stability and performance that controllers must deliver, as required by safety standards for various classes of robots and applications.

Development of Safety Standards

Closely intertwined with the quantification of stability performance is the development of safety requirements for bipedal robots in various application domains. Current safety standards are insufficient for fully addressing the capabilities and operations of bipedal (active-balancing, stepping) robots. Implementation following current standards would result in a severe limitation of the scope of potential applications that these robots could have, not from a lack of capability, but from an inability to determine that the task could be performed safely. This stems from a lack of safety definitions and requirements, as well as a lack of methods to validate safety-related control functions of bipedal-type robots. Such validation can be initially performed simply by defining performance thresholds on key test methods, which together may establish confidence in safe operation on known tasks. However, to achieve safe general-purpose robots, additional safety methods and stability guarantee requirements will need to be created that can verify the performance of robot subsystems such as controllers, sensing systems, and internal robot models. Establishment of clear, well-defined, objective, and easily verifiable safety requirements is the key for addressing stability concerns arising from dynamic, legged robots, and thus is a major step towards supporting the successful development and implementation of humanoid robot technology.

Limitations of Existing Safety Standards

In the current absence of a safety standard for humanoids, current deployment solutions would need to temporarily reference some elements in existing standards. This is a common practice in contingent conditions, but it requires careful consideration of the scope and degree of interpretation of standardized requirements. In general, safety technical standards must not be used à la carte. Potential options to consider are the (obvious) complete physical separation between robot space and human spaces, and the residual exchange of energy in case of physical contact. In both cases, the impact of the active control of stability has a direct effect on the size of separation to either maintain from humans or to evaluate when defining a safe state based on residual contacts.

The most relevant reference is then ISO 13855:2024 "Safety of machinery — Positioning of safeguards with respect to the approach of the human body", which illustrates the methods for computing the safe distances concerning static and moving sources of hazards, to set or configure protective devices. The standard is valid across products. In the case of mobile robots, the source of hazard is moving, together with humans, so all components for safe separation (relative velocity, time to come to a safe state, distance between potentially colliding objects) shall be adapted and calculated together. The implications of maintaining distances propagate to the components that measure distances during the execution of safety functions and behaviors.

The second reference is the standard for collaborative operations using industrial robots: ISO/TS 15066:2016 "Robots and robotic devices — Collaborative robots". The main clear mismatch for humanoids is that the scope of the standard is limited to robots fixed on the ground, or mounted on constrained tracks, or part of an integration onto statically stable bases. If motion is stopped in actively controlled joints as a result of safety limits, then the robot will not lose stability, result in falls, or extend beyond the pose at stopping time. ISO/TS 15066:2016 enforces safety limits (a) on the (dynamic) safeguarded space around a robot to maintain separation or (b) on the contact conditions that may occur whenever a human is in proximity or shares workspace with a robot. The modalities are known as Speed and Separation Monitoring (SSM) and Power/Force Limitation (PFL), respectively. In both cases, limitations on a humanoid's motion will have to deal with guaranteeing that the possibility of falling will reduce the risk to the human to the greatest possible extent. Again, active control of stability may be high-performing, but not perfect, as robots can still topple if inputs or disturbances are too large (just as humans can sometimes fall over).

For SSM-like approaches, mobility and active control of stability to remain upright will require the introduction of additional space to create a collaborative non-contact zone. If a human approaches this safe region around the robot, then the robot could, for instance, enter a stance and control mode that is practically motionless. For a robot with two legs, that would mean with both feet on the ground and not taking any steps or at least not any step that could generate large or unpredictable displacement. Common sense would indicate that such a robot pose should be such that if power were to be lost (or an e-stop hit), the robot would fall away from the human, and not towards it. In this basic, yet robust, implementation of the SSM criteria, the exact definition of the size of the stable safe region is determined as a design solution (not from the standard's requirement). For instance, it could be a region with a radius matching the height of the robot, and then shaped based on the current centroidal velocity such that the region represents the possible space a robot could fall into if it were to topple in any direction at any moment. Note that we are not necessarily recommending that this is the exact adoption of a standard, nor should it be the standardized requirements. It is instead an example of a solution to illustrate what type of framework and requirements (i.e., the targets for such solutions) are needed in safety standards dedicated to humanoids.

For PFL-like approaches, it's currently completely unknown what the potential configurations or limits are for residual contacts within the limits of pain onset that are specified for collaborative applications. ISO/TS 15066 (resp. ISO 10218-x: 2025) is, in fact, completely dedicated to single manipulators, while the complex effects of energy (or power) flux density, effect of exposed contact surfaces, and distribution of forces are unknown for humanoids. Standard methodologies for recommending the settings of limits and their verifications are heavily restricted to the hypothesis proper of manipulators. Still, the principle of quantifying the effects of physical interaction illustrated in the PFL mode of ISO/TS 15066 is becoming a solid foundation for the extension of such standardized limits to general machinery when contacts are part of the intended application. Important to be reminded, the PFL mode involves a purposeful collaboration between humans and robots in the same shared space. Accidental contacts have a distinct risk profile that can be studied by the specific workflow of such applications. An extension to occasional contacts for non-task-related operations (e.g., random contacts with bystanders) is, in general, out of the scope. It is indeed tempting to consider collaborative limits generic because a mobile robot can be anywhere in a working space. However, this shortcoming would bypass a due risk assessment with obvious consequences in missing proper estimation and evaluation of risks.

In conclusion, references to existing safety standards that give requirements for human proximity or potential exposure may be found, but remain largely insufficient or, at the very least, subject to extensive discussion on the conditions of applicability. The verification of safety functions dedicated to implementing the separation or physical interaction principles remain very challenging because of the lack of clear indication from a functional safety standpoint. Partial or total removal of humanoids from human spaces will severely limit the potential capabilities and uses that customers and manufacturers want to implement using humanoids. Note that this observation would apply to any robot actively controlling stability, which includes, but is not limited to, legged robots. There is a need for a safety standard that specifically details the behavioral and design requirements for a humanoid robot to be safely implemented.

Safety as Performance Thresholds

To create new standards for safety, especially as it is related to stability, the most straightforward approach would be to establish performance thresholds on tasks that the robot needs to be able to perform, and then require robots to demonstrate that level of capability on those tasks through validation test methods. The test methods and metrics discussed in section 3 could be used for this purpose, with the exact threshold that must be met for any particular metric-task pair determined by the user and application. The set of required tasks and the level of the safety thresholds can be altered for different applications and environments. For example, a humanoid designed to go out into the public may require a higher level of stability performance and robustness to disturbances, but a lower level of manipulation accuracy, compared to a robot that is to be used in an industrial environment with only infrequent interaction with fully trained professionals.

In this approach, a robot would demonstrate the capability to perform a task safely, and then, as long asmost of the tasks a humanoid is expected to do fall within the scope of those capabilities, it can be expected that the robot can safely perform its required tasks under a minimally varying set of conditions. For example, if safety requires that the humanoid robot be able to fall away from any nearby humans in the case of a fall, the capability can be shown via a disturbance-based test method that represents the tasks in question, that a robot can fall in a predetermined direction if a fall occurs. If, during operation, a robot enters a state or environment for which it does not have as rigorously proven stability, then a different, fallback safe state could be utilized. While an approach like this is useful and would be an accessible first step to establishing stability and safety requirements, the lack of generalizability of safety to any particular task is a limitation of this approach. More sophisticated safety requirements will need to be defined that can handle any arbitrary task a robot might need to perform.

Going Beyond Current Standards

As stated above, a performance threshold method of proving that a robot can perform a certain task safely does work well to establish robot capability. However, it still does not offer guarantees of safety or stability when a robot is implemented in a real-life environment. A robot may be capable of performing a task safely, but to meet standard safety requirements, the robot must be guaranteed to perform the task safely, or it needs to alter its task plan to one that it can.

As a result, we can state that there are two parts to meeting the safety requirements to be set out in future standards: first is demonstrating the robot's capability to safely perform a task (as measured by specific metrics and test methods), second is implementing guarantees that the robot will perform the task in a real-world implementation, subject to any foreseeable disturbances. Such guarantees could look like stability constraints in the controller, high-accuracy path tracking, or predictive control that models the robot throughout an entire dynamic motion. Several such approaches have been presented in academic research, but to be implemented in practice, new standards are needed with the details and procedures necessary to validate such controllers. Additionally, if meeting these safety requirements involves state estimation of an internal model or external sensing of the environment, then validation of sensor and modelling reliability will also need to be presented.

These details and procedures are needed to guarantee the stability, and thus safety, of a bipedal robot within foreseeable and reasonable circumstances. However, it is recognized that 100% guaranteed stability is not always possible. Just as humans sometimes will fall over due to unforeseen circumstances or too large of disturbance, a bipedal robot may fall over as well. This means that standards for fault handling, fall response, and handling of all other errors must also be presented.

Finally, a better understanding of the relationship between possible instability and risk is needed. The risk of a robot falling over scales up many factors, including the weight of the robot, the mass of the payload, the maximum energy/momentum that it is allowed to move with, and the environment it is in. The risk will also change depending on the robot's robustness to disturbances and the capability of the balance controller. Standards will need to define minimum levels of acceptable risk for various application spaces, including what minimum levels of stability will be necessary to achieve stability with dynamic interactions. The key point to operate within the same workspace together with humans is to set and measure stability thresholds that are deemed safe enough.

Posed in more procedural terms, the tried and true approach to managing risks inherent to machines in human spaces is to institute appropriate controls that reduce the risk. However, the behaviors of balance and stability that bipeds exhibit are produced by increasingly complex algorithms (optimal control, multi-layer MPC, learned policies) that make it harder to understand what risk the robot poses from possible instability, and therefore harder to understand what the appropriate controls are. As the controllers that govern robot behavior increase in mathematical and algorithmic complexity, the standards that are used to measure and evaluate these controllers will need to be developed to match the capability. This does not replace the need and use for control-independent, task-performance test methods, but should be in addition.

Of note is that requirements listed in future safety and stability standards should be specific and descriptive, connected to how such robots are designed. Most current safety requirements for robotics are very loosely defined. For example, the required capability for performing SLAM on autonomous vehicles is generally "fit for the safety goal"; the requirement to avoid instability on mobile robots is open-ended; the ability to identify nearby humans for collaborative robots is not even close to defining a target success rate. General requirements are not detailed enough to be useful for humanoid robots and other complex robotic applications beyond simply requiring "safe" operation near humans. Definitions of sufficient safety (sufficient levels of risk reduction) must be agreed upon and set. Additionally, approaches for mitigating risks can go beyond behavior generation constraints and safety control functions, to include physical design specifications such as soft contact points or low centers of gravity and modular safety devices such as air bags or overhead gantries in hazardous areas.

Overall, future standards will need to do the following:

- reference performance metrics to address the specific capabilities of the robot in a quantitative form;
- establish targeted quantities that define the level of safety;
- consider foreseeable interactions with the environment and other actors (both human and robotic) that can modify the threshold quantities;
- define and set the amount of prediction up to a defined time-horizon over such quantities (thus, standards must be developed to consider internal aspects of controllers and algorithms);
- define and provide safety validation criteria directly for control, sensing, and motion planning capabilities, rather than only defining a set of tasks that a humanoid should be expected to safely perform without losing stability.

Reaching consensus on the establishment of such standards will necessitate a highly coordinated effort between manufacturers, regulators, researchers, and end users. The coordinated efforts of multiple standards development organizations will be needed to accomplish this goal. As such, the recommendation is to establish a group dedicated to specific requirements for humanoids. The overarching standardization goal is to ensure that standards evolve towards the ideal of objective safety guarantees and test methods, but at a pace that design and implementation can keep up.

Future of the Study Group

Given that a large standardization effort over an extended period will be required to achieve needed standards for humanoids, cooperation and collaboration will be needed between all SDOs, and with manufacturers, researchers, and customers. To manage such an effort, a central working group is needed to keep track of the ongoing efforts in multiple SDOs and ensure research gaps are being addressed. Another recommendation is that this IEEE Study Group on humanoids continues beyond the publishing of this report to fulfill this role. However, because a humanoid is a ‘general-purpose’ robot, its intended functionality extends into all aspects of robotics, as well as the standards that shape them. Therefore, this group will need to successfully work with other groups and SDOs, addressing all other aspects of robotics as well.

At the time of the publication of this report, a new project from ISO/AWI for developing a safety standard that will cover industrial bipedal robots has begun: ISO/AWI 25785 Part 1: Safety requirements for industrial mobile robots with actively controlled stability (legged, wheeled, or other forms of locomotion)--Part 1: Robots. This is the first, and a significant, step towards creating updated safety standards for bipedal robots.

Conclusion

In this chapter, humanoid robots are characterized by bipedal, powered, self-balancing operation. This legged nature enables dynamic active control of stability combined with dynamic reshaping of the robot’s support polygon, which allows the robot to theoretically navigate any potential terrain while completing arbitrary tasks. This capability, however, also increases the risk of all motions due to the potentially unstable status of bipedal legged locomotion. Humanoids (including both humans and bipedal robots) are always one mistake or unexpectedly large disturbance away from a hazardous falling or other hazardous limb motion. The high level of risk means a high validation threshold to be achieved for stability-related safety control functions. However, no current standards account for robots that have unstable states as part of regular functioning. Standards must be extended and/or newly created to address the unique capabilities of humanoid-type robots. This chapter outlines a multi-directional approach to the improvement of standards for measuring, evaluating, and guaranteeing stability, together with standards that lay out the practical requirements for maintaining the safety of bipedal robots. Specifically needed are: safety standards that lay out the requirements for maintaining safe operation in various application domains, standard metrics for the measurement of stability and performance of bipedal robots given their complex motions and multi-objective control systems, and standard test methods to form a common understanding of performance across varying robots and physical capabilities. Within these standards, measurements of stability will need to be consolidated in objective metrics, verification criteria, and validation scenarios. Safety standards will need to provide a dedicated illustration of risk assessment regarding stability and clear requirements for safety functions and safety behaviors that would include several contributions to risk reduction. Safety standards will benefit from testing standards, as most safety behaviors will require quantifiable conditions to establish minimum acceptance thresholds of performance. Additionally, as new subsystem performance requirements are defined (balance, state estimation, navigation, multi-objective coordination), new standards will need to be created that can verify the capability and reliability of these subsystems.

Humanoid robots, like all machines, will never be perfectly risk-free. However, with the proper safety controls implemented, there can still be a safe and effective implementation of the technology across many intended application domains. Creating the standards, as described above, will provide the tools and common understanding necessary to achieve these successes. Finally, the same standards created with humanoid robots in mind will apply to a wide variety of robots that share the targeted features with humanoids. Standard efforts resulting from this report will improve the implementation and applicability of many robotics technologies throughout industry.

Sources:

https://www.marketsandmarkets.com/Market-Reports/humanoid-robot-market-99567653.html

https://www.marketsandmarkets.com/PressReleases/usa-humanoid-robot.asp

https://www.snsinsider.com/reports/humanoid-robot-market-1616

https://www.cervicornconsulting.com/humanoid-robot-market

ASTM humanoid robot database

https://interactanalysis.com/insight/humanoid-robots-large-opportunity-but-limited-uptake-in-the-short-to-mid-term/

Human-Robot Interaction for Humanoids

Authors: Francisco Andrade Chavez, Benjamin Beiter, Marie Charbonneau, Brandon J. DeHart, Greta Hilburn, Jeremy Marvel, Kartik Sachdev, and Dieter Volpert

Humanoid robots operating in human spaces will typically be highly interactive. In particular, humanoid robots are envisioned as valuable tools for task-oriented assistance in a variety of applications that will involve interactions with humans of various backgrounds and abilities, including but not limited to household chores, caregiving, or healthcare. Therefore, in this chapter, we will focus specifically on the interaction between humanoid robots and humans.

Although human-robot interaction doesn't always have clear safety implications, the collected knowledge and experience that have been distilled into this chapter will demonstrate that it is crucial to include a consideration of the interaction aspects of humanoids in any discussion of safety and standards development, as humanoid robots are inherently "interaction machines".

Human-robot interaction (HRI) is a broad field of interdisciplinary research being investigated by a diverse swath of academia, industry, and government organizations. The primary goal of this research is to better understand how to develop, measure, and improve interactions between humans and robots as they become more and more prevalent in our society. Human-robot interaction begins with the first impression that a robot inspires in a human. Similar to how humans interact, this initial first impression evolves through further interaction. These first impressions are primarily based on the appearance of the robot, including the robot's physical features, how it moves, how it sounds, its mannerisms, and how it behaves.

How does a humanoid robot designer know what details to include or change, and how much? There is much research on these topics, and more needs to be done as humanoids (as highly interactive robots working in human spaces) are used more broadly. To start to understand the various aspects of this problem, however, we can lean on the expertise of those in the field and the expectations of prospective users, as described in the next section.

Qualitative Data Collection on HRI and Humanoids

To facilitate a deeper look into this topic, a qualitative survey was developed and shared with stakeholders across many different fields, both within HRI itself and beyond. The intent of distributing this survey widely was to collect responses from experts on the topic, experts on related topics, and laypeople to generate an indicative cross-section of opinions on which to base the discussions and recommendations found later in this chapter.

At the start of the survey, respondents were informed that its purpose was to help us determine what aspects of HRI might impact the development and application of standards for humanoid robots. The survey consisted of the following 12 questions, divided into three thematic sections:

Robot Appearance and Communication Methods

- How does a humanoid robot's appearance/presence/pose/motion impact how unfamiliar people react to it?
- How can the appearance/presence/pose/motion of a humanoid robot lead to incorrect assumptions about its capabilities?
- What should the minimum communication capability/capacity be based on for humanoid robots (e.g., application)?
- What key elements are required for people to recognize the intent of a robot's actions or behaviors?

Communication Behaviors

- How should humanoid robots communicate their intentions (upcoming motions/behaviors) to people in their vicinity?
- How should humanoid robots communicate their awareness of people's presence/actions/intentions in their vicinity?
- How should humanoid robots communicate potential accidental/intentional physical contact with the people around them?

- How should humanoid robots respond to accidental/intentional physical contact with people to ensure both safety and utility?

Safety and Uncertainty

- What does a 'safe' or ‘trustworthy’ humanoid robot mean to you?
- What should the minimum safety-specific behaviours be of a humanoid robot depending on its application and environment?
- How should a robot balance risk to itself and other equipment with risk to nearby people when planning dynamic motions and tasks in its environment?
- What should the role of a human supervisor be in applications and environments with different amounts of uncertainty?

Over a few months, we collected fifty responses to this survey, with responses to each question from each respondent ranging from a few words to several paragraphs. Based on the inductive method of qualitative content analysis process introduced by Elo and Kyngas in 2008, we have summarized and distilled the responses to inform the discussions and recommendations below.

User Centered Design Supportive Research

This chapter has also been informed by a set of 12 user interview participant responses collected regarding the potential integration of robots into daily life, as part of a study conducted in the USA by one of our team members. This consisted of an in-depth analysis of 12 user interviews on humanoid robot integration that included a diverse range of participants (all non-engineers), which reveals a complexity of expectations and ethical considerations that are critical for future development.

The user interview questions focused on user expectations as well as current general perceptions concerning humanoid robots, to support the HRI survey and find any potential missed opportunities in design, safety, and future user expectations. Participants articulated viewpoints ranging from profound concerns about the ethical implications of robot "slavery" to optimistic visions of robots as compassionate caregivers and assistants.

The findings consistently underscore the importance of emotional intelligence and trust in human-robot interactions, with users emphasizing robots' need to comprehend and respond to human emotions, especially when engaging with children, older adults, and other vulnerable populations. The user interview responses regarding accessibility, data privacy, and potential social implications highlighted an urgent demand for ethical design and equitable deployment.

Regarding humanoids in human spaces, there's a clear desire for robots to be task-oriented and functionally reliable, but users also expect personalized interaction with the robots, including meaningful communication. A prominent concern emerged around robots becoming a luxury status symbol, advocating for business models that prioritize affordability and equitable access through mechanisms like rental programs or insurance.

Ultimately, this user feedback points to the necessity of a human-centered approach to robot development, one that prioritizes safety, simplicity, and the cultivation of positive human-robot interactions based on positive connections. The core findings emphasize that for humanoid robots to be safely and effectively integrated into daily life, particularly for vulnerable populations like children and older adults, they must prioritize emotional intelligence, trustworthiness, and seamless accessibility. Users expressed a strong desire for robots that demonstrate empathy, understand human emotions, and build genuine rapport.

This user interview feedback informs standards by highlighting the need for humanoid robots to not only avoid physical harm but also prevent emotional distress, ensure data privacy, and mitigate potential social disruptions. It also outlines that the end user will be expecting: **(i)** a focus on accessibility and inclusivity, **(ii)** emotional intelligence and understanding to be prioritized, and **(iii)** seamless human-robot communication.

Similarly to the survey feedback, users wanted a sense of genuine understanding between themselves and the robot, although the user interview feedback included a need for more empathetic interactions, including:

- **Human-like communication:** beyond simple speech, the robot should mimic natural conversation patterns, including pauses, response anticipation, and empathetic tone.
- **Non-verbal communication:** eye contact and realistic hand gestures are crucial for conveying warmth and attentiveness.

- **Accessibility and inclusivity:** the robot must be capable of understanding and responding to users with communication impairments, such as those caused by strokes or other health conditions, as well asl children who have suffered trauma with trust issues.
- **Emotional intelligence:** the robot should be able to interpret subtle cues, including minimal facial expressions (micro-expressions) and vocal tones, to understand the user's emotional state and needs. Examples given: fear, pain, confusion, and trust.

Supporting the HRI Framework

The user-centered design recommendations based on the user interviews, particularly those on advanced emotional connections, are foundational to establishing trust and perceived safety. A robot that can interpret and respond empathetically is inherently perceived as safer and more reliable, reducing anxiety and increasing acceptance.

Furthermore, the proposed real-time trust measurement and renewable licensing model that were discussed in the user interviews offer innovative approaches to continuously monitor robot performance against established safety and ethical benchmarks. This dynamic feedback loop can inform regulatory adjustments, incentivize manufacturers to prioritize human well-being, and provide a concrete mechanism for ensuring robots meet evolving safety and social standards.

In essence, this qualitative data argues that effective safety and standards for humanoid robots cannot be developed in a vacuum. They must be deeply informed by real human needs, fears, and aspirations, ensuring that technology serves humanity in an inclusive, empathetic, and ultimately, safe manner.

The remainder of this chapter consists of 3 main sections, each dedicated to one of the thematic topics included in our survey within the field of HRI as it relates to humanoid robots, followed by a section outlining the lessons learned via the user interviews mentioned above, and concluding with a section outlining our overall recommendations for standard makers and robot designers, based on both internal expertise and the results of the survey and user interviews.

Robot Appearance and Communication Methods

To optimize the safety and efficacy of human-robot interactions, designers must consciously prioritize a robot's appearance (along with its dynamic motions and physical capabilities), as it dictates how users will likely perceive and interact with a given robot. On this topic, the survey asked how respondents think a robot's appearance can impact communication between a user and the robot: how it shapes first impressions, the assumptions it can lead to, and what specific features and capabilities they think humanoids should, and will, have in the future.

Broadly, the responses we collected confirm that appearance is indeed important for shaping how humans and humanoid robots communicate. Users set their initial expectations of robot capability based on the appearance of the robot, and especially on how similar it looks to a human. This appearance-based first impression can also determine the levels of comfort and familiarity a user has with the robot, even after interacting with the robot regularly.

While both factors can be used to convey necessary information quickly and intuitively, mismatches between the impressions given by appearance and the real capabilities of the robot can cause safety concerns. People who feel too familiar with the robot, or think it has a higher cognitive capacity than it does, will be more likely to not show proper caution and to not follow proper procedures to keep themselves and their surroundings safe. Such concerns can be mitigated by proper training, signage, and other controls, but the appearance, actions, and direct communication of the robot with the people around it will continually reinforce certain impressions, so designers should be deliberate about conveying accurate information.

Based on user responses, three types of communication may be needed and/or desired from the robot: Explicit, Implicit, and External, as described in Communication types. While users have varying opinions about what the minimum threshold of technical communication capability or modality should be, Communication Purpose and Means describes how the purpose of communication with a humanoid robot was considered to have three components. The first is to convey the robot's intent, allowing nearby humans to accurately predict the robot's next actions, while the second part should convey the general state of the robot. Finally, the third component conveys details about the environment a robot is working in, the obstacles it needs to work around, and tasks that the robot is performing.

Building on this 3-part perspective on human-robot communication, we will next explore how the relationships between robot appearance, communication types, capabilities, and human expectations need to be accounted for in the design of a humanoid robot.

Appearance and Capability Alignment (Emotional Expectations)

Robot appearance significantly influences user expectations and trust. Accurate alignment between appearance and actual capabilities is crucial to avoid misinterpretations that could lead to safety hazards or diminished trust.

With one exception, all survey respondents agreed that appearance does have an impact on communication. While this is potentially helpful, it can also lead to miscommunications when humans make false assumptions about the capabilities of a robot based on its appearance. Many noted that the more human-like a robot looks, the more people will expect it to be able to perform actions (both in mobility and communication) equal to what a human can do. Based on user feedback from the user interviews, the majority of individuals who were interviewed agreed that a robot's appearance should be determined by the robot's intended function.

As some survey responses noted, anthropomorphism is inherent to being human, as we apply it to everything, but humanoid robots convey anthropomorphism on another level, even prompting character design, giving the robot an identity to interact with rather than just base functionality. Conversely to anthropomorphism, if a robot has more ‘robot-like’ features (flat surfaces, hard shapes, points, jerky motions, etc.), then people will tend to treat it more like an autonomous robot, expecting less interactability or intelligence and more machine-like behaviors.

When evaluating the effect of appearance, the most common response was that, currently, the anthropomorphic appearance of humanoids across the board causes people to overestimate the robot’s true capabilities. People assume it has mobility beyond what is possible at the moment, a level of sensing and awareness beyond reality, higher intelligence, reasoning, and emotional understanding than it does, higher reliability in task performance than currently has been proven, and a higher level of safety in the robot's immediate vicinity than is guaranteed. A less common, but still possible problem is that a less human-looking robot could lead to users underestimating what the robot is capable of, though this poses less of a safety concern.

Further complicating the matter, everyone comes to an interaction with different expectations, impacted by prior experience, knowledge, identity, physical size (if a robot is larger/smaller or heavier/lighter than they are), and context. These prior expectations then combine in complex ways with the appearance-based expectations in a variety of ways, meaning that it is impossible to perfectly define public expectations of robots (which are not the same as the expectations of trained personnel), though appearance can attempt to guide it in certain directions. As a result, the physical features and mannerisms of a robot need to be carefully considered based on the expected users, and should not just be an afterthought of design wrapped around a machine.

User feedback indicates a desire for humanoid robots with the ability to be personalized by user preference and adapted to a diverse range of users by age and capabilities. The majority of the groups that expressed a desire for a more detailed appearance in specific aspects of the robot were young children and older adults. Facial features that could be expressive and could emote directly were favored for both groups. Options for both digital and more humanlike features were considered best for the desired features in terms of interaction based on individual needs.

In terms of capability, interviewed individuals also expressed that the appearance of the robot when in a caregiver capacity should have more humanlike qualities, with its arms and hands showing empathy in its movements. Social robots were also indicated as requiring more humanlike features or having the option to be upgraded or expanded in their capabilities and appearance.

Many of the interview respondents gave emotionally driven responses based on real-life situations related to the socioemotional components of empathetic interactions in specific environments. Home, school, and healthcare facilities were all specified as places respondents would expect to see a humanoid robot that had features that would signal their abilities to users. These specific features would be indicators included in its design deliberately to visually display its purpose, functionality, and capabilities to all humans who see it.

Expectations of the robots' mobility and physical gestures were also emotionally driven. If the robot was intended for basic domestic assistance, individuals wanted to have the ability to personalize the robot’s appearance to match the environment, and have the option to “put it away”. Other interview respondents wanted a high-performing surrogate human companion robot that served as a caregiver with the ability to interact and move with human precision.

Trust measurement based on how interview respondents explained a connection to a robot’s appearance was based on Attractors and Detractors. Attractors were in the comfort that respondents felt from the visible understanding of a simple robot design. Detractors were a respondent's emotional conflict or confusion created by a complicated design in the robot's appearance. All related to how one might communicate or interact with the robot.

This also implies that as a robot is continually developed and its capabilities are improved, the appearance should also change, although implementing this is not a simple task. Design and movement remain critical in managing expectations throughout the development of a robot.

Appearance Determines Comfort and Trust

As discussed above, the appearance and behaviors of a robot are a form of technical communication, implying details of the robot’s capabilities and competencies. The same features of the robot also evoke an emotional response from the human, shaping feelings of familiarity, comfort, and trust within the user towards the robot.

Any feature of the robot can have these effects, including size, color, shape, voice, mannerisms, motion style, similarity to a human, etc. Many survey responses associated certain types of features with whether they would evoke positive or negative reactions in users, and what they believe those reactions might be, as summarized in Table X. While these features and reactions seem to be common, the magnitude of the reactions, as well as which features will lead to which reactions varies greatly from person to person. As with prior expectations, how an individual reacts is heavily dependent on background, culture, experience, and many other factors.

TABLE X: SUMMARY OF AFFECTIVE REACTIONS TO THE ROBOT AND WHICH FEATURES COULD BE THE CAUSE.

	POSITIVE	NEGATIVE
FEATURES THAT EVOKE THIS REACTION:	Calmness, smooth and predictable motions, cuteness, familiar features such as eyes, friendliness, open/natural poses	Aggressiveness, jerky or unnatural motions, overly fast motions, intimidating form (size, shape, and hardness), or actions
THIS REACTION RESULTS IN:	Familiarity, willingness to interact, and being accepting	Fear, disinterest, discomfort, avoidance, negative associations, uncertainty/ wariness, surprise

Communication Purpose and Means

While the appearance and other features of the robot may primarily shape the first impression a user has of the robot, other forms of voluntary communication also shape the human-robot interaction. Based on the survey responses, there were three general reasons identified for why communication between the robot and human is necessary. Communicating the robot's intent to the human is the most commonly stated purpose for communication. A secondary purpose of communication is to always make a human aware of the state of the robot. This includes the task it is performing, as well as its operating status, power status, controller state, health, environmental concerns, and other errors or faults. Similarly, a third purpose is to communicate the details of the environment and task that the robot is operating in, including awareness of the presence and actions of humans in its surroundings.

The form of communication can vary, with communication behaviours that humans may expect from humanoid robots including visual cues (lights, gaze direction, facial expressions, gestures or display screens), audio cues (verbal speech or electronic sounds such as buzzes, beeps and boops), or physical cues (touch, modulating movements). Overall, given their sophisticated nature, people may expect humanoid robots to display communication behaviours that are at the intersection of machine and human capabilities. Further discussion on this theme can be found in the Communication Behaviours section.

The strategy to generate the communication also varies, from the ability to hold a basic conversation, to just confirmation of commands when received or interjection when safety is a concern, such as shouting “STOP”. The overall purpose is for nearby humans to have a clear understanding of what the robot is about to do. Essentially, every situation has an associated type and level of communication that is required to safely perform the task, and the robot must meet this requirement. Additionally, the communication capability changes further based on whether the task is in a domestic, medical, industrial, service, or public environment. Small details such as color, surface finish, lights, etc., can indicate the type of application that the robot is probably going to be used for (and, ideally, has been designed specifically to fulfill).

Communication Types

To achieve the goals of the three purposes of communication stated above, responses detailed many ways of communication projected to be useful for human-robot interaction. These have been categorized into three main types of communication: Explicit communication through words or motions, implicit communication through the shaping of all behaviors, and external communication through user interfaces (UIs) not located on or run by the robot. The details of these types are given in the table below.

TABLE X2: SUMMARY OF HUMAN-ROBOT COMMUNICATION TYPES

COMMUNICATION TYPE	DESCRIPTION
EXPLICIT:	Capabilities or actions that are done exclusively for communication. This includes lights or screens, voice and other sounds, facial expressions, or simple motions such as pointing with hands, looking in a direction with the head, or the pose of the body.
IMPLICIT:	This involves shaping other behaviors to convey a certain impression. This includes movement profiles, patterns of motion, gaze and eye position, body language, posture, and other nonverbal behaviors that are ancillary to the functional purpose of the motion.
EXTERNAL:	These are forms of communication that are not on board the robot itself. This includes user interfaces, external displays of sensor information, network communications, guides, and other training materials to understand the robot.

Minimum Communication and Emotional Intelligence Capability Requirement

Within these types of communication, survey responses also detailed the desired minimum levels of competency believed to be necessary for implementing humanoids. The most commonly stated requirement was that the robot look similar to a human and have a human-like communication capability. Specifically noted was the capability to understand human speech, gestures, and mannerisms, be able to communicate back in the same way by holding a basic conversation, and be able to remember individuals to personalize interactions. Varying levels of technical competency were also mentioned, from “Better than Alexa” and “The communication capability of a dog” up to “able to utilize a small task-based vocabulary of interpretable commands”, “able to understand natural language via NLP or ChatGPT”, and “able to interact with a central warehouse management system”.

While most of the survey responses specify how the robot should communicate with humans, several people also state that the robot should be able to interpret commands from humans as well and be able to respond to them. Robots should also have an accessible way for nearby humans to halt or stop undesirable behavior. Based on these responses, users generally expect humanoid robots to exhibit emotional intelligence, including recognizing and responding to human emotions, and to engage in meaningful, human-like conversations. Therefore, advanced multimodal communication (visual, audio, physical) is essential for conveying robot intent and status, and for facilitating natural interaction with surrounding humans in the environment.

This is confirmed by user interview feedback, with an expectation of a high level of emotional intelligence and advanced abilities to communicate and converse with a robot via voice activation and an expectation of intuitive language adoption. This was introduced by users both in the preprogramming of the robot and in the trainable language and task orientation stages of robot learning, based on individual preferences and needs.

A portion of the communication would likely be performed via a dashboard that can take in data and share it with individuals and others based on approval by the user. A therapist, teacher, doctor, or other healthcare provider could have access to both the data and the potential modifications in behavior required to make adjustments to the operation of the robotic system.

In interview responses, the desired communication via interaction was often based on the type of robot in question. If it were a social robot, it would be dependent on the individual, age, and cognitive abilities, designed to be used in tandem with a support system of humans. The communication was primarily verbal and included digital facial expressions and sounds to communicate to the user, as well as voice and tablet or an alternate device to communicate feedback.

Robot emotional understanding was also an expected language adaptation between the user and the robot, especially with extended or repeated usage. As an example, cognitive barriers were brought up by respondents as an example of how a robot would have to be trained to provide personalized interactions with vulnerable populations like children or older adults.

Discussion and Recommendations

Overall, there seems to be a consensus that the appearance of a humanoid robot has a large impact on shaping its interaction with humans, which starts at the first impression and continues to grow and change throughout all communication. The appearance of a robot also shapes the expectations of what that robot is capable of.

When manufacturing a robot, it is critically important that the capabilities implied by a robot's appearance match its actual capabilities. Mismatches dramatically increase the risk of harm during interactions. However, due to the inter-individual variability of reactions to a robot's appearance, designing a proper appearance is a complex, difficult problem. A feature that makes one person trust a robot might make another mistrust the robot, and both could be wrong, possibly unsafe expectations relative to what the robot's actual capabilities are.

Additionally, over- or under-estimation of capabilities is based on the current public perception of the capabilities of humanoids, and the relationship between expectations and reality will change as robots' capabilities grow. Therefore, manufacturers should carefully consider the appearance of any robot they design to accurately represent the current capabilities of that specific robot, with a need to balance features that make others more comfortable and not scared of the robot, while remaining cautious enough to maintain safety procedures.

Building on this, it is worth noting that many responses set extremely high expectations of the minimum levels of required communication capability. People expect humanoid robots to be able to communicate at the same level of competency as humans, including being able to understand natural language and gestures, and then be able to reproduce both by themselves.

Even more so, some responses specified that robots should be able to know everything about what actions are being performed, be able to reason at a high level about why it is doing it with "sound logic", and then be able to communicate that reasoning if asked. While it is true these capabilities would be ideal for clear communication, they may be beyond where the minimum requirements should be set. Even the same responses that detail how people have too high expectations for humanoid robots still set high expectations for standards for the robots.

Though responses agree that appearance heavily influences levels of comfort and trust, people disagree on whether certain appearances have a positive or negative impact. Some say that human-like appearances increase comfort and intuitive interaction, while others say that human-like appearance heightens discomfort because of the uncanny valley and should be avoided. This question of whether or not a robot should look more or less like a human has no consensus, so at minimum, manufacturers should consider their prospective users and which effect is a priority, or the lowest risk, for their application.

All communication and appearance functionality should be defined and/or guided by consensus standards, with validation by extensive test methods. To achieve these standards, however, there needs to be an industry-wide push to identify what features cause what reactions, and how to create humanoid morphology and the capability to optimally interact with human users. This includes defining minimum communication capabilities, methods, standard signalling and recording, safety requirements, and minimum performance thresholds.

In summary, appearance should be shaped to accurately convey the capability of the robot, balancing the need to evoke comfort and trust in humans nearby (or at least avoiding discomfort) while still encouraging proper safety procedures. Similarly, for communication, the quantity, quality, and mode of communication should be tuned to efficiently and effectively convey the necessary information without overcomplicating the resulting interactions.

Communication Behaviours in Social and Physical Human-Robot Interaction

Communication with nearby humans may not necessarily be critical in settings where a robot follows a predetermined work sequence. However, projected uses of humanoid robots will certainly fall beyond strictly predictable contexts. As introduced in Communication Purpose, humanoid robots are expected to communicate with human users in different ways to inform users of the robot's intent, state, task, and perception of the environment. This indicates that the majority of users want to be aware of, or able to provide consent for, upcoming robot actions.

When robots and humans share physical space, communication abilities may be especially critical for humanoid robots, for example, when communicating robot intentions (such as upcoming motions and behaviours) and a robot's awareness of people's presence in its vicinity, as well as its understanding of their actions or intentions.

It is generally understood that accidental contact between robots and humans should be avoided as much as possible. For humanoid robots operating in settings where maintaining physical separation from humans is not always feasible, there will be a possibility of contact between people and robots. In these cases, a robot may be expected to communicate potential accidental or intentional contact with people, as well as to respond to contact.

In settings such as healthcare, service, or home environments, humanoid robots will often need to operate in proximity to untrained individuals. In these cases, some form of generally understandable communication of intent would be advisable. It would need to be as clear as possible, so that the robot does not cause unintended confusion, surprise, or stress in people, but instead that its movements and behaviors can be easily anticipated by people.

The level of communication required will depend on context and risk assessments, such as an understanding of who the robot is working in the vicinity of, what they are doing, where their attention is, and what they can effectively perceive. Additionally, the communication behaviors and emotional intelligence expected from a humanoid robot will likely differ depending on the application and level of proximity between a human and a robot, such as in situations where physical distance can be maintained versus those where physical contact may occur.

In all cases, however, to convey meaning intuitively, humanoid robot designers should aim for communication that is as human-like as possible. Communication with untrained individuals could also be facilitated by a robot's physical appearance if it is designed to indicate the robot's intended task. However, the use of multiple modalities (ideally, as many as possible) is critical to ensure inclusive communication with individuals with diverse sensory and cognitive (dis)abilities.

Given how human-to-human communication can sometimes be confusing, how a humanoid robot is designed to communicate robot intent, awareness of humans, potential contact, and to respond to contacts (as focused on in the survey) needs to be carefully considered to ensure clear communication. To that point, survey respondents suggested a simultaneous combination of visual, audio, and physical communication modalities. In effect, using multiple types of signals in coordination may help produce clear communication.

Survey respondents also indicated a strong expectation that the way humanoid robots are made to communicate with humans would borrow from typical human-to-human communication behaviors. However, they did not exclude the use of machine-like communication behaviors. For example, colored lights or beeps from a robot may not always clearly convey the intended information to untrained people on their own, but they could be part of a robust and successful communication system when combined with other behaviors.

The following sections break down survey respondents' feedback related to each of the four communication behaviour scenarios identified in the survey questions, which included an increasing degree of physical proximity and contact with people in the environment.

Communicating Robot Intent

When it comes to communicating robot intentions (including upcoming motions and behaviours), respondents suggested the use of various visual, audio, and physical cues.

Visually, lightsignalling may be useful (perhaps in a similar way as car lights indicate the intended direction of motion, or in a similar way to position lighting, projecting lights where the robot is intending to move), but standards may need to establish clear meanings for light signals. Borrowing from human-to-human communication behaviours, predictable robot motions, as well as human-like gestures, gaze direction, and facial expressions, may help to intuitively communicate intended robot actions. Examples may include pointing, leaning, or turning the head of the robot in the direction of intended motion. In a more machine-like approach, visual display screens may be used to display arrows or descriptive text indicating current and intended robot actions. However, visual cues assume that humans in the vicinity of the robot can see it, and that the robot has their visual attention.

Aurally, speech (potentially in conjunction with motions) may be used to provide awareness of upcoming robot actions to nearby humans (e.g., “on your left”), but it could also be used to seek consent or to negotiate actions with humans, given the context (e.g., “May I move ahead?”). However, communication through speech may not always function in loud environments or when communicating with humans who have hearing impairments; additional communication modalities would be beneficial in these cases. Survey respondents also suggested that beeps or tones could be used before each movement to provide awareness of upcoming motions, provided they do not negatively impact or cause fatigue in the people the robots interact with.

Physically, slowing down robot motion around humans is often done to ensure safety, but may come with a tradeoff on productivity. As suggested by some respondents, perhaps beyond communicating intent, robots could also communicate to surrounding humans how to move around them, to minimize slowdowns. Haptic wearables, for example, may be used to enhance human awareness of robot actions (e.g., vibrating when a robot is nearby).

Communicating Robot Awareness of Nearby Humans

Communicating a robot’s awareness of people’s presence, actions, or intentions in the robot’s vicinity requires more complex communication than just communicating robot intent: people may want to confirm whether a robot is aware of them when nearby. Intuitive, human-like behaviour is likely to be expected: robot designers should take inspiration from how humans negotiate personal space and turn-taking to inform their designs’ behavior. Survey respondents suggested that awareness may be communicated in the form of a greeting or short message (verbal or nonverbal). However, they also indicated that the amount of information communicated should be appropriate, as over-communication may cause confusion or discomfort.

Lights can be used as a visual indicator, but just as for robot intent, standards may need to be defined. Lights may change color when a person is detected in the vicinity; they may be used to indicate that the robot is listening and processing, or they may be used in other ways. Body language, including facial expressions, gaze direction, and gestures, may be used to indicate awareness of the presence of humans in the form of a greeting, for example, with a smile, a brief look toward a person, a wave, or a head nod. Communication may be more involved, for example, directing the robot's head and gaze towards people, to show awareness of their presence, and then tracking their movements to communicate awareness of people's actions. A display screen could also be used to inform people of the robot's awareness. Robot behaviours may, however, be adjusted to avoid generating feelings of unease in humans who may feel observed if a robot is "staring" at them.

Speech in natural language could be used to provide verbal acknowledgement or confirmation of awareness to humans (e.g., saying "Hello" or "Hi" as a greeting, or "I see you"), as well as to confirm a robot's perception of a person's actions or intentions (e.g., "Are you picking up that box in the corner?"). However, respondents indicated that just as for robot gaze, the use of voice could potentially cause unease, and thus might need to be socially adjusted. None of the respondents indicated that beeps or tones would convey the required information in this case, but this could be investigated as a potential new type of standardized robot "language".

Robot motions can also be adjusted to indicate awareness of humans in the vicinity, for example by pausing or stopping when someone enters the robot's "personal space", moving more gently or safely around humans, and reactively adjusting proximity to people.

Communicating Imminent Contact

A humanoid robot's communication of imminent contact with a human and response to a contact that has occurred is critical to both safety and utility. However, before physical contact may be permissible, robots must first be verified to have the capacity to: detect the proximity of a human, sense contacts, apply physical avoidance approaches, and physically respond to contacts (e.g., by following collaborative robot standards). However, by their nature and typical use cases, humanoid robots may be expected to approach and respond to contacts differently than collaborative robots, especially when being used in a wider range of environments.

In such scenarios, communication should be clear, loud, and obvious to ensure humans are aware of the situation. The more dangerous a robot is, the louder and more attention-grabbing this communication should be. Multimodal communication should be emphasized to ensure that humans get the message, regardless of their sensory (dis)abilities. In cases where humanoid robots are interacting with untrained individuals (or trained individuals who may not be relied upon to remember the training received), human-like robot behaviours may be more conducive to providing effective communication. This would help ensure that communication signals are designed to clearly and intuitively convey an intended message (as opposed, for example, to having several warning lights and buzzers going off without an obvious meaning).

When interacting with humanoid robots, people may unconsciously expect a minimum level of human-like deference from these machines. Local cultural norms and specific contexts may thus need to be considered. For example, it may be acceptable in certain cultures for a robot to make contact with a human to prevent potential injury or to navigate a crowded environment, but perhaps not for other reasons. In some cultures, it may be customary to apologize or ask for consent before making intentional contact, or to apologize after making accidental contact with someone. These cultural norms must be integrated into general-purpose humanoid robots.

While it is becoming well established that humanoids need reliable communication abilities, physical human-robot interaction is still in its early research stage. Additional research is needed to appropriately understand physical and psychological safety needs in these situations. For this reason, recommendations can be expected to evolve as new findings emerge.

If a humanoid robot is programmed with the ability to detect that accidental or intentional contact with a human in its proximity is about to occur, it may also be programmed to alert and provide adequate warning to the concerned human(s), as a way to mitigate physical and psychological safety risks arising from the interaction. A combination of visual and audio signals may be used to get a human's attention while ensuring that the trajectory of the robot is predictable.

Visual signals may include the use of light indicators, perhaps following industry-standard safety color codes for ease of interpretation. Intensity, color, and flashing frequency of lights may be modulated to convey urgency. Additionally, robot gestures may be used, such as deliberate movements that convey the need for caution (e.g., raising the hands, turning the head to point the gaze towards a potential contact location) or that convey the intended robot motions and pathway (as discussed in Communicating Robot Intent). Robot facial expressions could help convey the need for caution (e.g., moving eyebrows, eyes, and mouth to display surprise). A screen display may also be used to communicate an intended contact via text.

Audio signals may include the use of speech or electronic sounds (such as beeps, boops, and buzzes) to warn humans of an impending contact or to communicate an intended contact. In both cases, volume, pitch, and speed may be modulated to convey urgency.

Communicating After a Contact has Occurred

An appropriate robot response to contact is critical to ensure safety and utility, as well as to maintain trust in humanoid robots. Humans may expect a humanoid robot to respond to accidental or intentional contact in different ways, including **(i)** acknowledgingthe physical contact, **(ii)** assessing the situation, before **(iii)** responding. The robot may respond either by stopping motion, disengaging, or moving back, mitigating the effects of the intended contact, safely continuing a previous action, or asking for human assistance, reporting accidents, or calling for backup (be it a person in charge or emergency services).

In this perspective, survey respondents expressed that communication after physical contact may be used to acknowledge the contact, apologize for the contact, gain context from a human, or provide context to a human. For example, communication to gain context may be used to confirm: **(i)** the safety of those involved, **(ii)** whether there was perceived or actual harm resulting from the contact, **(iii)** the intention of someonemaking contact with the robot, or **(iv)** confirm the next appropriate action for the robot. Communication to provide context may for example be used to: **(i)** confirm whether the contact was intentional or not, **(ii)** describe the intent of the robot, **(iii)** describe the actions of the robot in response to the contact, or **(iv)** communicate the robot's intent to correct, adapt, or avoid future contacts. To communicate such information, a combination of visual, audio, and physical signals may be used.

Visual signals may include the use of lights, although they may have a limited ability to convey a required meaning when used on their own, as discussed in previous subsections. Robot gestures and facial expressions may be used, such as deliberate movements that convey acknowledgment of or apologies for the contact (e.g., turning the head towards the location of the contact, moving the hands apologetically). A display screen could also be envisioned to provide acknowledgments, apologies, or context to a human through text.

Audio signals may include speech, such that a contact may be acknowledged or apologized for, and such that context may be gained or provided through verbal communication. Electronic sounds may also be used, although they may have limited ability to convey the required communication when used on their own.

Physical signals may include communicative robot motion, such as modulating motion direction, velocity, or compliance (e.g., stopping, slowing down, moving away, increasing compliance while in contact) to help ensure safety and haptically convey acknowledgment of the contact. Some survey respondents also indicated that a physical interface, with physical interaction capabilities (e.g., a button, a computer, or making use of haptic robot capabilities), could be used to gain context, for example, allowing a human to adjust robot motions.

Discussion

Among all scenarios discussed above, a large proportion of respondents suggested having the robot communicate through speech. This could indicate that humans intuitively consider speech as an effective means to convey information, potentially due to its ubiquitous use in human interactions. Lights, body language, display screens, and electronic sounds may be considered as complementary tools to enhance the clarity and effectiveness of robot communication. The selection of communication modes should be context-dependent, for instance, considering who may be interacting with the robot, environmental conditions, and tasks being accomplished. Additional research may be needed to improve understanding of personal space, comfort, trust, and perceived safety in social and physical human-robot interaction, but existing literature on user experience design and human-robot interaction may provide a starting point.

Nonetheless, it remains clear that expectations of the interactive abilities of humanoid robots are highly sophisticated. Survey responses indicate underlying assumptions that humanoid robots have sensing, perception, reasoning, and communication abilities similar to those of a human. User interviews add to that, expressing a need for intuitive and simple user interfaces that facilitate clear communication and that can be adapted to different users (offline and online), while including options for plain language explanations through speech and/or digital display.

In addition, communication mores may differ from one application to another, for example, with specific jargon, sounds, and gestures used to communicate between individuals of a given workplace. This will need to be considered in the design of interactive humanoid robot behaviors. As communication heavily depends on human detection, it will also be critical to define how the reliability of mechanisms for human detection can be validated, and how potential failures may be handled. Additionally, as clear and effective communication may not always be guaranteed to prevent incidents during humanoid robot operation, the next section covers safety considerations for human-robot interaction.

Safety & Uncertainty in Humanoid Robot Interaction

Any given robotic system intended to be used by or around people must exhibit safe behaviors while in operation, and humanoid robots may come with their own set of challenges due to their interactive nature, along with their high complexity, mobility, and power.

The survey invited input on factors impacting the safety (or the perception thereof) of robots performing tasks with or around people. The survey did not provide prompts that encouraged respondents to consider safety in any particular way, but instead encouraged feedback based on their expectations of safe interactions. Responses were largely focused on physical safety and touched upon topics including the definitions of safe behaviors, the characteristics of robots that promote trust, and the reduction of risk due to the design, functionality, and operating environment of robots. That is not to say that psychological safety in these interactions should not be considered. Rather, it indicates a blind spot in how most people think about safety around robots, and calls for deliberate attention to be brought to this aspect of safety, including in the topics covered below: how to define a safe and trustworthy humanoid robot, what behaviours make a humanoid safe for HRI, and the implications of human supervision.

Defining a “Safe” Humanoid in the HRI Context

Regardless of whether a humanoid robot is public-facing or merely working in an environment that is (or was previously) intended for human presence, the robot must adhere to certain safety guidelines and design considerations. As emphasized by survey respondents, a safe humanoid robot must be designed to minimize the risk of physical harm to humans, whether nearby humans behave predictably or not, either directly or indirectly through its actions and presence in the environment. This is achieved through a combination of elements.

For humans to be safe in the proximity of a humanoid robot, the robot needs awareness of its surrounding environment and the people in it, such that it can perceive humans and the objects that can be involved in human-robot interaction, anticipate human actions and potential chain of events in a dynamic environment, and react appropriately to prevent hazardous situations.

Just as importantly, individuals need to be able to predict the robot’s behavior. A humanoid robot is composed of a large number of moving parts compared to other robots, which could affect predictability when the robot moves in ways that are not quite human-like. Predictability can be facilitated with predictable robot movements (for example, smooth trajectories with easily interpretable goals) that are communicated by the robot – see the above sections for details on communication types and behaviors. Maintaining predictability and communication is especially critical during HRI, where changes in robot motions may be required to dynamically accommodate hazard avoidance.

Just as with any other mobile manipulator, a humanoid’sphysical design is also critical for safe HRI: lightweight components, soft materials, rounded edges, and compliant joints (intrinsically or through control) can reduce the potential for injury. Additional considerations in the control systems may be needed to account for the safety hazards that could stem from the many moving parts on the robot, the behavior of which a human may not be fully aware, and which may result in unpredictable motions when the humanoid’s balance is affected through HRI.

To remain mobile while carrying loads, a humanoid robot’s motors may also be more powerful than those in traditional cobots. Maintaining safety in this case may call for the implementation of robust safety systems, such as fail-safe mechanisms, error detection, safe stop, and error recovery mechanisms that are designed with the assumption that humans may be nearby. The presence of humans may also need to be taken into account for the robot to safely respond to power loss/fluctuation, localization errors, and uncertainty in dynamic situational awareness.

While the most critical concerns relate to physical safety, psychological safety may need further consideration. In particular, how safe a robot is may not always directly correlate with how much humans trust a robot, which will affect human-robot interactions as described next.

Defining a “Trustworthy” Humanoid Robot

If humans are expected to exist or move through environments in which humanoid robots are active, a form of social contract is implied between the humans and the robots’ manufacturers, integrators, and owners. Specifically, humans are extending a degree of trust to the robots and the entities that they represent; trust that the robots will not intentionally harm people, or indirectly contribute to conditions that could lead to harm. To build a trustworthy humanoid robot, survey respondents indicated that focus should be placed on several key elements.

As described in Robot Appearance and Communication Methods, from the start, a humanoid robot must communicate its abilities and limitations for humans to adjust their expectations and avoid over- or under-trust. Throughout individuals’ interactions with a humanoid robot, it must demonstrate consistent reliability, capability, and predictability in performing its tasks. While clear communication is already critical for safety, transparency reinforces trust. Given the complexity of humanoid robots, having a robot convey and explain its intentions, actions, and decision-making can be beneficial. Additionally, given the large data gathering capacity of humanoid robots, transparency on how user data is collected and used, as well as evidence of robust privacy protection measures and ethical decision-making processes, would affect how much trust individuals place in their interactions with humanoid robots.

Minimum Safety-Specific Behaviors for Human Environments

When asked what they would consider as minimum safety-specific behaviors for a humanoid robot, survey respondents further emphasized the following aspects. To ensure safety, humanoid robots must be equipped with reliable emergency stop mechanisms, offering various levels of halting, from immediate power removal to controlled stops, including moving back to a safe state. Stop mechanisms should includeboth autonomous and human-activated ones, provided that the latter is accessible to the humans interacting with the robot. This may require careful consideration, given humanoid robots' high mobility.

A physical emergency stop button installed on the body of a humanoid robot would ensure it is always in the proximity of the robot, but requires that attempting physical human-robot interaction be safe (and feasible for individuals with different (dis)abilities) at the moment where an emergency stop may need to be activated. Novel, accessible approaches to remote emergency stop activation may be necessary. When humans are in the proximity of a humanoid, they may also need the ability to reason about the safest courses of action, especially in the case of a loss of balance. Humanoids need to be equipped with robust collision detection, avoidance, and reaction systems that can detect and react to impacts with humans and objects before, during, and after they happen, while also managing the robot's stability and the stability of anything it is handling.

Given humanoid robots modelled on the human body, whether front-facing cameras mounted on the head are sufficient, or if additional sensing is required for the robot to maintain 360-degree awareness, may need to be carefully assessed. Ensuring robot movements are safe for interactions with humans is paramount, requiring smooth, controlled motions with adjustable speed and force limits. Clear, bi-directional communication is essential, with multimodal signals conveying the robot's intentions and alerting nearby humans of any issues or hazards, and mechanisms allowing the robot to perceive and interpret human communication (for example, confirmation that the robot may proceed with a task).

Internal fault detection and diagnostic capabilities enable timely maintenance and, if necessary, safe shutdowns. Humans must retain the ability to easily override the robot's actions, and any object handling must be performed with secure grasping and releasing to prevent hazards. Rigorous testing and validation of all safety features in realistic scenarios will be crucial to guarantee safe operation.

Humanoid robot safety requirements can vary significantly depending on the environment in which they operate, which will affect the kinds of interactions a humanoid robot is likely to encounter and prioritize as part of its task. While some hazards are intrinsic to the environment itself, others manifest through the interactions between agents and the environment. Movable objects can become dislodged and impact people, robots, and other environmental features. People and robots moving around can unintentionally close routes of ingress and egress.

Pinching and crushing hazards can manifest when movable objects are put near immovable features. Changes in operational conditions (e.g., lighting, surface textures, and clutter), some of which may be dynamically introduced by human activities, can make it difficult–if not impossible–to safely and reliably move through a given environment.

When balancing risks to humans, to itself, or objects and beings in the environment, a common narrative is that a humanoid robot must operate under the unwavering principle that human safety is paramount, superseding all other concerns. However, when it comes to risk to itself or objects and other beings in the environment, beyond what may be hardcoded, contextual information and interactions with humans may influence priorities. For instance, a human could specify what the robot should be careful with and what should be prioritized over the integrity of the humanoid robot itself. To satisfy these requirements when operating within a dynamic environment, whether and how a robot should be equipped with real-time risk assessment and lowest-risk planning capabilities would need to be examined.

Human Supervision for Safe HRI

In some situations, a human supervisor may assist a humanoid robot's operation, which introduces a distinct mode of human-robot interaction through teleoperation and remote communication between the supervisor and the robot. In some cases, a human supervisor may also need to communicate with nearby people through the intermediary of the robot. The human supervisor's role will be fundamentally shaped by the level of environmental uncertainty.

In low-uncertainty environments, characterized by predictable routines and well-defined tasks, and a low probability of human presence disrupting robot operation, would allow for a more passive supervisory approach. Here, the supervisor would act primarily as a monitor and overseer, requiring timely checks to ensure optimal performance and anticipate maintenance needs, consistent monitoring to detect deviations from expected robot behavior and environmental conditions, and quick intervention when unexpected events occur.

In contrast, in high-uncertainty environments, where humans are likely to be present, unexpected events are frequent, and the robot's tasks are complex or novel, a human supervisor would be called for to assume a highly active, "hands-on" role. This is a demanding role for a human supervisor, requiring them to remotely maintain awareness of the robot’s dynamic surroundings, make rapid diagnoses and decisions to mitigate risks, operate the robot, potentially overriding the robot’s autonomous actions, and coordinate with nearby humans.

For a system as complex as a humanoid robot, the design of the user interface is critical to **(i)** adequately communicate the robot’s capabilities and limitations, but most importantly, to **(ii)** ensure the supervisor is not overloaded (mentally, physically, temporally, …) and that their involvement does not introduce further hazards.

Conclusion and HRI-Related Recommendations

As outlined across this chapter, individuals expect human-like abilities from humanoid robots; for some, this goes to the point of expecting that a robot can recognize and respond to human emotions. Some users may desire humanoid robots to engage in meaningful conversations and emotional connections. These high expectations, going well beyond functional task completion, are likely influenced by the human-like shape of a humanoid robot. As a result, robot designers may need to carefully consider the design of a robot to accurately convey its abilities to the humans it will interact with.

Humanoid robots also need the ability to communicate their intentions, state, and awareness of the environment. A variety of signals may be considered in communication implementation, involving various modalities including audio (verbal or nonverbal), visual, or physical signals, whether used explicitly, implicitly, or through external devices. However, the interactive capabilities of a humanoid robot need to be adapted to the needs and abilities of the individuals with whom it will interact, and the environment in which it will do so. Special attention to accessibility is necessary when working with individuals with disabilities or members of vulnerable populations.

When humanoid robots share space and interact with humans, functionality and reliability are critical, as are the inclusion of robust safety mechanisms, including human supervision for uncertain situations. In some scenarios, the benefits of robotic assistance may need to be balanced with potential harms, physical or emotional, that may occur during human-humanoid robot interaction. As part of risk assessment, robot malfunctions ranging from minor issues to severe physical or emotional harm will need to be considered. How to define acceptable rates of humanoid robot malfunctions within different use cases remains an open question.

Concerns relating to privacy, data collection, and transparency were only briefly touched upon, but will affect the trust people place in their interactions with humanoid robots. It is also worth reinforcing that the themes covered in this chapter were those that the team considered the most critical. However, additional concerns about interacting with humanoids may have been missed and may surface as humanoid robots are being developed, tested, and deployed.

While the standards remain unfinished, manufacturers should be aware of intended standards and begin building capabilities into their robots (i.e., lights for communication, shaping of motions to be faster/slower, smoother/more robotic) such that standards can be met when they are implemented. Organizing this could be another job for what the IEEE study group becomes in the future.

Report Summary: Building a Standards Framework for Humanoid Robots – Classification, Stability, and Human-Robot Interaction

Humanoid robots are approaching a tipping point in development, promising general-purpose functionality across industrial, service, and public applications. However, widespread deployment is limited by one overriding challenge: the lack of standards designed for the unique risks and capabilities of humanoids.

A coordinated effort among Standards Development Organizations (SDOs) is urgently needed to establish a structured framework. This framework should be built on three interconnected pillars—Classification, Stability, and Human-Robot Interaction (HRI), each informing the others to form a comprehensive standards pathway.

Classification: The Foundation of All Standards

The first step in any standardization effort is a clear, agreed-upon classification system. Current robotics standards were designed for fixed-base or statically stable robots, not for humanoids that combine bipedal locomotion, dynamic balance, and manipulative dexterity. Without a common taxonomy, regulators, manufacturers, and end users cannot consistently determine which standards apply or what performance expectations are realistic.

A multi-layered classification framework is recommended, encompassing:

- Physical Capabilities – locomotion type (bipedal, hybrid), dexterity level, and sensory systems.
- Behavioral Complexity – degree of autonomy, manipulation skills, and adaptive behaviors.
- Application Domains – industrial, healthcare, public service, and other specialized use cases.
- Humanoid-Specific Traits – anthropomorphic resemblance, naturalistic motion, and interaction modalities.

For SDOs, this classification system can serve as the "table of contents" for future standards. It allows committees to map which standards are broadly applicable (e.g., functional safety from ISO 13849) and which need humanoid-specific extensions (e.g., balance safety, fall-response behaviors).

Stability: The Critical Bottleneck for Safety and Performance

If classification is the foundation, then stability becomes the obstacle that must be overcome for humanoids to operate effectively in shared human spaces. Unlike wheeled or fixed robots, humanoids constantly deal with managed instability; even powered-down robots can fall, which creates inherent hazards.

Key gaps identified in current standards include:

- No quantifiable stability metrics tailored for actively balancing robots. Existing measures, such as those in industrial mobile robot standards, only consider tip-over from payload shifts, not active balance control.
- No standardized test methods exist to evaluate stability performance under realistic tasks, disturbances, or environmental variability.
- Functional safety models that do not account for dynamic balance—current SIL and PL measures assume deterministic systems, whereas humanoids require predictive, probabilistic risk modeling.

A two-phase standards roadmap is recommended:

- Performance Standards (ASTM, IEEE) – Develop stability metrics (e.g., margin of stability, capture point, predictive stability regions) and task-based test methods (walking on uneven terrain, carrying loads, recovering from pushes).
- Safety Standards (ISO, IEC) – Translate those performance metrics into minimum safety thresholds, incorporating predictive control requirements, fall-handling behaviors, and residual risk limits.

For SDO members, stability testing and safety validation must be treated as intertwined efforts. ASTM and IEEE can lead the development of repeatable test methods and quantifiable metrics, while ISO and IEC integrate these into regulatory-grade safety standards.

Human-Robot Interaction: Managing Risk and Perception

Humanoids occupy a unique space in human-robot interaction (HRI) because of their Anthropomorphic form. People naturally project human-like expectations onto them, resulting in two major risks: overtrust and psychosocial discomfort.

From a safety perspective, humanoids introduce indirect risks that are not covered by traditional robot safety standards. Automating a process with humanoids can alter workflow pacing and repetition, thereby increasing musculoskeletal and cognitive risks for human workers in both pre- and post-automation tasks.

From a perception standpoint, technically safe motions can still feel unsafe; fast limb swings, sudden steps, or a robot standing too close can cause discomfort, especially in public environments where bystanders are untrained.

Future HRI standards should therefore address:

- Collaborative Task Safety Thresholds – Extending ISO/TS 15066 principles for humanoid mobility and manipulation, including safe-fall zones and controlled stance modes when humans are nearby.
- Interpretable Behavior and Body Language – Defining motion guidelines to signal intent.
- User Training and Expectation Management – Standardizing how capabilities are communicated to avoid overestimation of performance.

Classification and stability standards directly inform these HRI standards. For example, a humanoid classified for public use would require higher stability thresholds, stricter fall-response requirements, and interaction-specific body language guidelines than one confined to a closed industrial setting.

Constructing the Standards Framework

The interdependence of classification, stability, and HRI highlights the need for a coordinated, multi-SDO approach rather than piecemeal adaptations of existing standards. A suggested pathway includes:

- Classification as the Organizing Principle

- Adopt a shared taxonomy across SDOs to define humanoid types, guiding which existing standards can apply and where gaps remain.
- Parallel Development of Stability Metrics and Test Methods
- ASTM and IEEE can lead to task-based performance test method creation (e.g., walking, manipulation under disturbances).
- IEEE can standardize stability metrics for predictive and instantaneous balance assessment.
- Integration into Safety Standards
- ISO and IEC can develop application-specific safety thresholds, incorporating test methods and metrics into regulatory safety validation.
- HRI and Perception-Based Standards
- Build on classification and stability results to create interaction guidelines addressing both physical and psychosocial safety.
- Centralized Coordination
- Establish a joint working group spanning ISO, IEEE, and ASTM to ensure alignment. A shared roadmap would prioritize stability-related standards first, as they are the primary barrier to safe adoption.

Toward Trustworthy Humanoids

The three pillars are not isolated; they are mutually reinforcing. Classification clarifies what type of humanoid is being evaluated, stability standards prove that it can function safely, and HRI standards ensure it does so in ways that humans find acceptable and trustworthy.

Until these elements are developed in tandem, humanoids will remain limited to controlled environments and pilot programs. But with a coordinated standards effort, SDOs have the opportunity to build the framework that will make humanoids reliable, certifiable, and ultimately, deployable in the diverse human spaces for which they are designed.

Special Thanks To The Following Contributors Who Provided Insights Into The Development Of This Report

Francisco Andrade-Chavez	Tanya Grinberg	Aleshin Nikita
Chioma Aso	William Harrison	Mike Oitzman
Fahmi Badreddine	Matthew Howard	Rob Paolozzi
Christian Becker-Asano	Michael Jensen	Jim Pippine
Jan Butz	Salim Karaer	Deb Prince
Marie Charbonneau	Sagar Kasrung	Maximo Roa
Francisco Chavez	Andra Keay	Ted Rozier
Vincent Clerc	Niamh Kingsley	Kamel Saidi
Parker Conroy	Jacques Lemire	Kartik Sachdev
Alvaro Costa	Shang-Ching Liu	Joan-Wilhelm Scharze
Maharshi Digumarthi	Pablo Lopez Garcia	Werner Varro
Claude Dismoor	Matthew Marge	Tom Vardon
Nikolai Ensslen	Riccardo Mariani	Dieter Volpert
Carlos Felix	Syrine Mansour	Axel Wechsler
Toni Garcia Rodriguez	Carlos Mastalli	Ruoshi Wen
Animesh Garg	Bill Mather	Yuke Zhu
Julien Ghaye	Jonathan McCormick	Zhijie Zhao
Mona Ghassemian	Nicholas Milton	

Bibliography (Extracted from Footnotes)

1. In this document, ‘self-balancing’ and ‘actively-balancing’ are used as interchangeable terms when referring to a robot. They both describe a robot that only remains upright through active, powered balance.

2. ISO 12100:2010 Safety of machinery — General principles for design — Risk assessment and risk reduction

3. **IEC 60204-1 (Consolidated version 2016+AMD1:**2021) Safety of machinery - Electrical equipment of machines - Part 1: General requirements

4. ISO 10218-1:2025 Robotics — Safety requirements Part 1: Industrial robots

5. ISO 10218-2:2025 Robotics — Safety requirements Part 2: Industrial robot applications and robot cells

6. ISO 13482:2014 Robots and robotic devices — Safety requirements for personal care robots

7. ISO 13850:2015 Safety of machinery — Emergency stop function — Principles for design

8. ISO 13855:2024 Safety of machinery — Positioning of safeguards with respect to the approach of the human body

9. IEC 61508 Functional safety of electrical/electronic/programmable electronic safety-related systems - (Part 1 to 7)

10. ISO 13849-1:2023 Safety of machinery — Safety-related parts of control systems Part 1: General principles for design, and ISO 13849-2:2012 Safety of machinery — Safety-related parts of control systems Part 2: Validation

11. ANSI/RIA R15.08-1-2020 Industrial Mobile Robots - Safety Requirements - Part 1: Requirements for the Industrial Mobile Robot

12. ANSI/ITSDF B56.5-2024 Safety Standard for Driverless, Automatic Guided Industrial Vehicles and Automated Functions of Manned Industrial Vehicles

13. ISO/TR 20218-1:2018 Robotics — Safety design for industrial robot systems Part 1: End-effectors

14. IEC 63310:2025 Functional performance criteria for AAL robots used in connected home environment

15. ISO 7176-1:2014 Wheelchairs Part 1: Determination of static stability

16. ISO 7176-2:2017 Wheelchairs Part 2: Determination of dynamic stability of electrically powered wheelchairs

17. ISO/TR 4448-1:2024(en) Intelligent transport systems — Public-area mobile robots (PMR) — Part 1: Overview of paradigm

18. See full details and options for PFHd values in 5.3.3 of ISO 10218-1:2025

19. ISO 21448:2022 Road vehicles — Safety of the intended functionality

20. ISO 26262 - Road vehicles — Functional safety (parts 1 to 10)

21. ISO/AWI 25785-1 - Safety requirements for industrial mobile robots with actively controlled stability (legged, wheeled, or other forms of locomotion) - Part 1: Robots

22. ISO 9283:1998 Manipulating industrial robots — Performance criteria and related test methods

23. ISO 18646 Robotics — Performance criteria and related test methods for service robots (Part 1 to 8)

24. ISO/DIS 18646-5 Robotics — Performance criteria and related test methods for service robots — Part 5: Locomotion for legged robots

25. F. Aller, D. Pinto-Fernandez, D. Torricelli, J. L. Pons and K. Mombaur, "From the State of the Art of Assessment Metrics Toward Novel Concepts for Humanoid Robot Locomotion Benchmarking," in IEEE Robotics and Automation Letters, vol. 5, no. 2, pp. 914-920, April 2020

26. Ramuzat N, Stasse O, Boria S. Benchmarking Whole-Body Controllers on the TALOS Humanoid Robot. Front Robot AI. 2022 Mar 4;9:826491.

27. C. Sferrazza, D.-M. Huang, X. Lin, Y. Lee, and P. Abbeel, "HumanoidBench: Simulated Humanoid Benchmark for Whole-Body Locomotion and Manipulation." Robotics: Science and Systems Conference 2024

28. J. Baltes and S. Saeedvand, "Rock Climbing Benchmark for Humanoid Robots," 2022 International Conference on Advanced Robotics and Intelligent Systems (ARIS), Taipei, Taiwan, 2022, pp. 1-4,

29. Stasse O, Giraud-Esclasse K, Brousse E, Naveau M, Régnier R, Avrin G, Souères P. Benchmarking the HRP-2 Humanoid Robot During Locomotion. Front Robot AI. 2018 Nov 8;5:122.

30. "Eurobench project." [Online]. Available: https://eurobench2020.eu/

31. W. Thibault, F. J. A. Chavez and K. Mombaur, "A Standardized Benchmark for Humanoid Whole-Body Manipulation," 2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids), Ginowan, Japan, 2022, pp. 608-615

32. V. Lippi, T. Mergner, T. Seel and C. Maurer, "COMTEST Project: A Complete Modular Test Stand for Human and Humanoid Posture Control and Balance," 2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids), Toronto, ON, Canada, 2019, pp. 630-635

33. Peng, W. Z., Song, H., and Kim, J. H. (March 12, 2021). "Stability Region-Based Analysis of Walking and Push Recovery Control." ASME. J. Mechanisms Robotics. June 2021; 13(3): 031005.

34. J. A. Castano, E. M. Espuela, J. R. Rojas, E. M. Hoffman, and C. Zhou, “Benchmarking standing stability for bipedal robots,” IEEE Access, p. 1, Jan. 2025, doi: 10.1109/access.2025.3529191.

35. C. Curtze, T. J. W. Buurke, and C. McCrum, “Notes on the margin of stability,” Journal of Biomechanics, vol. 166, p. 112045, Mar. 2024, doi: 10.1016/j.jbiomech.2024.112045.

36. T. Mergner and V. Lippi, “Posture Control—Human-Inspired Approaches for Humanoid Robot benchmarking: conceptualizing tests, protocols and analyses,” Frontiers in Neurorobotics, vol. 12, May 2018, doi: 10.3389/fnbot.2018.00021.

37. ISO 13482:2014 “Robots and robotic devices — Safety requirements for personal care robots” has some relevant references for human-robot interactions but it is specifically dedicated to service applications and under deep revision at the time of publication of this Report, so omitted as an option.

38. This Technical Specification is soon to be withdrawn as it is incorporated in the revised series ISO 10218-1 (and 2):2025 “Robotics — Safety requirements - Part 1 (and 2)”.

39. S. Kajita et al., "Impact acceleration of falling humanoid robot with an airbag," 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids), Cancun, Mexico, 2016, pp. 637-643,

40. Elo, S. and Kyngäs, H. (2008), The qualitative content analysis process. Journal of Advanced Nursing, 62: 107-115. https://doi.org/10.1111/j.1365-2648.2007.04569.x https://doi.org/10.1111/j.1365-2648.2007.04569.x

41. Greta Hilburn, UX Researcher and Designer, US Defense Acquisition University