Accountable system design architecture for embodied AI: a focus on physical human support robots.

著者	Mizuki Takeda, Yasuhisa Hirata, Yueh Hsuan Weng, Takahiro Katayama, Yasuhide Mizuta, Atsushi Koujina
journal or publication title	Advanced robotics : the international journal of the Robotics Society of Japan
volume	33
number	23
page range	1248-1263
year	2019-11-11
URL	http://hdl.handle.net/10097/00130924

doi: 10.1080/01691864.2019.1689168

To appear in *Advanced Robotics*
Vol. 00, No. 00, Month 201X, 1–18

FULL PAPER

Accountable System Design Architecture for Embodied AI: A Focus on Physical Human Support Robots

Mizuki Takeda[a]*, Yasuhisa Hirata[a], Yueh-Hsuan Weng[b],
Takahiro Katayama[c], Yasuhide Mizuta[c], and Atsushi Koujina[c]

[a] *Department of Robotics, Tohoku University, 6-6-01 Aramaki-Aoba, Aoba-ku, Sendai 980-8579, Japan*; [b] *The Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, 6-3 Aramaki-Aoba, Aoba-ku, Sendai 980-8579, Japan*; [c] *RT.WORKS CO., LTD., 1-10-26, Nakamichi, Higashinari-ku, Osaka 537-0025, Japan*

(*Received 00 Month 201X; accepted 00 Month 201X*)

Although the development of robot-based support systems for elderly people has become more popular, it is difficult for humans to understand the actions, plans, and behavior of autonomous robots and the reasons behind them, particularly when the robots include learning algorithms. Learning-based autonomous systems which are called AI are treated as an inherently untrustworthy 'black box,' because machine learning or deep learning algorithms are difficult for humans to understand. Robot systems such as assistive robots, which work closely with humans, however, should be trusted. Systems should therefore achieve accountability for all stakeholders. However, most research in this field has focused on particular systems and situations, and no general design architecture exists. In this study, we propose a new design method, focused on accountability and transparency, for learning-based robot systems. Describing the entire system is a necessary first step, and transcribing the described system for each stakeholder based on several principles is effective for achieving accountability. The method improves transparency for systems, including learning algorithms. A standing assistive robot is used as an example of the entire system to clarify which system parts require greater transparency. This study adopted the Systems Modeling Language (SysML) to describe the system and the described system is used for the information representation. Information should be represented considering the relationships between stakeholders, information, and the system interface. Because of their complexity, it is difficult for humans to understand the complete set of information available in robot systems. Systems should therefore present only the information required, depending on the situation. The stakeholder-interface relationship is also important because it is more beneficial for professionals to view information relevant to their specialized field, which would be difficult for others to understand. By contrast, the interface should be intuitive for general users. Visualization and sound are very useful means of transmitting information, with advantages and disadvantages for different circumstances. These relationships are important for achieving accountability. Finally, we show an example of implementation with a developed support system. It is confirmed that accountable systems can be designed based on the proposed design architecture.

Keywords: Health Care Management, Physically Assistive Devices, Robot Ethics, Black Box

1. Introduction

An aging population increases the demand for support systems, and various robotic systems have been developed to meet this demand. Robotic support systems are expected to not only become alternatives to human caregivers, but also to provide better support, owing to features including mechanical strength, estimation algorithms, and AI technology.

*Corresponding author. Email: m.takeda@srd.mech.tohoku.ac.jp

It is possible to comfortably use home electronics or machines without knowing the internal processes of the systems. These machines have limited operations and are controlled by humans. The machines function by following simple conditional decision-making logic, which humans can easily understand. Learning-based systems that include character recognition and recommendation systems also do not require transparency. AI systems, including autonomous robots that use learning algorithms, however, are difficult for humans to use because they are difficult to understand. Physical human-robot interaction comes with safety risks, and therefore humans become anxious when they do not understand them. It is difficult for humans to cooperate with or rely on such robots to support human actions. When a person is carrying out a task that involves cooperation with other people, the task cannot be completed well if they cannot communicate with each other. Communicating is much more difficult in human-robot interactions than in human-human interaction. Humans become anxious if they cannot understand the actions of robots. People feel uncomfortable consigning health care tasks to robots that are perceived as unpredictable. Learning-based robots that interact with humans need to clearly present their safety-critical actions, states, plans, and reasons for acting.

Adopting robots that use learning algorithms raises additional problems. It is difficult to investigate and fix system failures in systems with 'black boxes.' It is also difficult to decide who is responsible in such cases. Knowledge representation is adopted to make systems transparent. Visualization and auditory display of knowledge representation improves system interpretability. Appropriate interfaces make systems transparent and interpretable to achieve accountability. However, almost all research is carried out for specific stakeholders of specific systems and there are generally several stakeholders for each system as shown in Figure 1. In the case of a walking assistive system with sit-to-stand functionality, a user uses the system, while caregivers or family members support the user in using the system correctly and safely. Engineers should maintain the system, and if problems arise, they need to investigate and repair it. If an accident occurs, accident investigators are responsible for the investigation. Although there are different stakeholders according to the specific system properties, these people generally represent the stakeholders of support robot systems. However, there is no general design architecture for a learning-based support robot system that is accountable to all stakeholders. Almost all research in this area focuses on specific situations for specific robots, aiming to improve system efficiency. Some researchers have discussed ethical design of robot systems; however, almost all research indicates general design principles or provides several examples. We therefore propose a new design architecture for accountable learning-based support robot systems. Considering accountability in the design process will be important for realization of accountable systems.

This paper proposes a design architecture to achieve accountability for learning-based support robot systems. The design architecture is explained using a standing assistive robot as an example. First, the entire system should be described, then the described system should be transcribed for each stakeholder based on several principles to effectively achieve accountability. Because each stakeholder requires different information, the entire system should be described to clarify the internal information of robot systems. In this study, we adopt the Systems Modeling Language (SysML) to describe the entire system; the language was created to describe systems and is popular in the systems engineering field.

Describing the system as a whole also contributes to AI transparency. It is difficult to achieve transparency in machine learning or deep learning algorithms and models. However, general systems consist of more than learning algorithms used for recognition or estimation in robot systems. Thus, the input-estimation relationship is opaque for humans. By contrast, the relationship between the decided action of the robot and the output of the actuators can be transparent because learning is not used for this function. The relationship between the estimated information and the action of the robot, which is the main interest of robot system stakeholders, can also be described. AI robot systems can be made transparent by describing the systems, even if the learning algorithms are opaque.

Relationships between stakeholders, information, and interfaces are important for providing re-

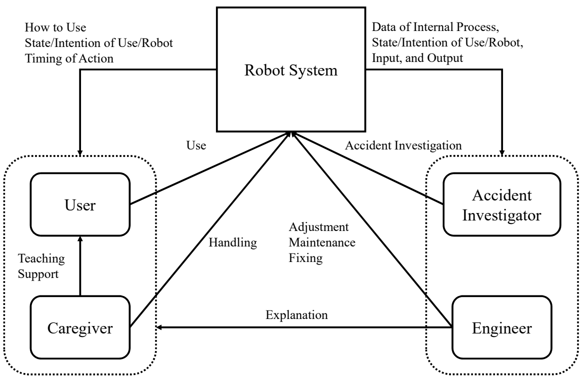


Figure 1. Example of System and Stakeholders.

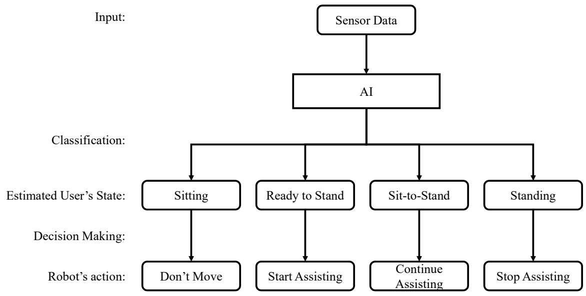


Figure 2. AI Robot System.

quired information to each stakeholder in the most appropriate form. Systems should represent all information, but because humans are unable to understand so much information at once, information should be represented in an appropriate way depending on the stakeholder, case, time, and other relevant factors. We considered system use cases as examples of stakeholder-information relationships. The professional ability of stakeholders should be considered as stakeholder-interface relationships, while information-interface relationships should be considered by examining the features of the information. This paper summarizes these relationships.

Finally, a support robot system following the proposed design method was developed as an example of the effectiveness of the proposed design architecture.

2. Related Works

2.1 *User's State Estimation*

It is important for support systems to provide appropriate support depending on the situation. Therefore, many researchers are focusing on real-time user state estimations.

Some researchers focus on the center of gravity (CoG) or center of mass (CoM) for human state estimation. These parameters are particularly useful for support robot systems, as they can be simply measured. Human activity such as gait can be analyzed by placing inertial measurement units (IMUs) on the human body in locations related to CoG acceleration [1, 2]. CoG position can also be calculated from human link models, using link parameters obtained with motion capture systems, laser range finders (LRFs), position sensitive detector sensors (PSDs), IMUs, and force sensors [3–6].

There has been interest in machine learning and deep learning algorithms for state estimation. On-body sensors such as accelerometers are frequently used for human activity estimation [7, 8]. Vision-based estimation has received a lot of attention. Convolutional Neural Network is one of the most famous methods for human pose estimation [9, 10]. Neural networks are also useful for robot control [11, 12].

The user state is generally evaluated for anomaly detection or robot function changes [13, 14]. User state, action, and intent can be used for motion control [15–17]. State estimation is also useful for improving safety. Anomaly detection can be used for accident prevention. Accumulation of estimated data is also useful for care monitoring and deep learning.

2.2 *Knowledge Representation*

If robots detect an anomaly, they usually stop their operations and alert users. Such alerts are useful to draw attention to the anomaly, and are effective for letting users know why robots have stopped operating. Representation of a robot's actions or plans is effective under both normal and abnormal operating conditions. If a robot has many functions, humans are unable to understand the robot's action and plan without representation, even if the system does not include learning algorithms. Displaying the robot's plan helps humans understand the robot's future actions [18]. Sound is one effective means of knowledge representation [19] and simple LED lights can also represent the robot's state [20]. These methods are considered useful for learning-based robot systems.

The ability of robots to correctly recognize visual and auditory inputs is not always reliable. Some studies in the computer vision field have addressed the reasoning behind learning-based classifications [21, 22]. For robots, both the input-classification relationship as well as the relationship between the classification result and the robot's action are important. Representation of recognition or estimation results, as well as asking humans for confirmation, facilitate the robot's tasks [23]. Confirmation of tasks which are ordered by humans can reduce the number of mistakes [24, 25].

When robots interact with humans, both the robot and the user actions are important. If human and robot are cooperating on a task, the robot will work more effectively if there is an understanding of what the human should do. Teaching robots also enables users to learn which actions are required of them [26]. Some researchers also study representation of artificial emotions in human-robot interaction [27].

These studies show that real-time knowledge representation is effective for using robot-based systems, as humans can understand and predict robots' actions via knowledge representations. Describing the systems in this way has advantages for designing and investigating the systems. Some researchers create original modeling languages to describe their specific systems [28, 29].

Some studies evaluated the construction of accountable robot systems by making their systems transparent; however, almost all studies have focused on the stakeholders for their specific systems. Systems design should follow some sort of guideline.

Transparency of learning-based robot system is less frequently discussed, although AI transparency has been discussed in the computer vision and machine learning fields.

For physical systems such as robots with AI, the surrounding AI transparency poses other issues as well as those related to the learning algorithm. However, the general design architecture for assistive AI robot systems has not been widely discussed.

2.3 *Ethically Aligned Design*

In situations where autonomous systems such as AI robots interact with humans, the systems should deal with ethical issues, as well as with safety concerns related to physical human-robot interactions. Machine and deep learning make systems unpredictable for humans, which makes them anxious. Complicated autonomous systems should also clarify the boundaries of responsibility. Learning-based robot systems should therefore be designed based on ethical principles.

Ethical design has been discussed in various fields such as telehealth [30] and the Internet of Things (IoT) [31]. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems has published Ethically Aligned Design, First Edition [32], which discusses general principles for autonomous and intelligent systems, such as human rights, data agency, transparency, and accountability Guidance implementation methodologies is provided; however, there is no specific architecture for designing robot systems. Ethically aligned design for autonomous and intelligent systems has been discussed elsewhere [33], and its importance for assistive robots is also suggested [34]. However, concrete and technical methods for implementation and design have not been discussed in detail. By contrast, engineers who are technical design specialists often focus on safety and don't consider ethics. The Japan Agency for Medical Research and Development (AMED) encourages development and introduction of nursing-care robots and released a guideline [35] that focuses on safety and ethical review, but did not discuss ethically-aligned design.

Transparency is one of the most important robot ethics-related issues for AI systems. Transparency of learning-based reasoning and the actions of autonomous robots has been studied; as explained in section 2.2, knowledge representation is a possible solution. However, learning-based robot systems are different from learning algorithms alone or autonomous robots with no learning. We therefore consider that a general design architecture is necessary for accountable learning-based robot systems. In the following sections, we propose a design architecture focusing on physical human support robot systems.

3. System Design Concept

This section explains the main concept of design architecture of support robot systems to achieve accountability. We used a walking assistive robot with sit-to-stand function as an example. The robot has an armrest with a linear actuator for sit-to-stand assist, and its wheels have attached motors for walking assistance. The robot can estimate the user's state by means of several sensors based on Support Vector Machine (SVM), a machine learning algorithm. The design architecture can be used for other autonomous robot systems that interact with humans, such as robot sports instructors and communication AI robots.

Robot systems which include learning algorithms and interactions with humans should achieve accountability. The actions and intentions of such systems tend to be opaque for humans, making it difficult for humans to cooperate with such robots. When the systems carry out an action, they should make stakeholders understand what they are doing and why. Humans relate to such systems not only during use, but also for maintenance, repair, investigation, and other functions. Therefore, describing an entire system and transcribing it for all stakeholders is important for achieving accountability.

System description is required to determine necessary information depending on the stakeholder and situation. Machine learning algorithms are difficult for humans to understand, thus robot systems with learning algorithms are also opaque. Our method for describing the entire system can contribute to the transparency of learning-based robot systems. The method is explained in detail in section 3.1.

System stakeholders include different types of people, such as users, caregivers, and engineers. The information required varies according to the situations. Interfaces should also be determined according to the stakeholder. It is important to appropriately transmit the required information. The concept of information ontology is explained in section 3.2.

3.1 *AI Transparency*

Systems that adopt types of learning algorithms, such as Random Forest, Support Vector Machine (SVM), Logistic Regression, and Neural Network, are increasing. Learning algorithms include 'black box' components that are difficult for humans to understand, and the opacity of learning algorithms makes the systems seem untrustworthy [36]. Some researchers focus on transparency of learning algorithms [37–40]. Studies on making AI transparent by representing the reasons for decisions [21, 22] can provide some understanding of learning algorithms, however, these methods cannot make all algorithms transparent for ordinary people. Even though the AI included in the system is opaque, systems should be transparent.

Systems usually use learning algorithms for some types of estimation, the results of which are used for system decision making. For example, user state estimation of a standing assistive robot can be realized with data from learning sensors. Effective operation of the robot requires that assistance motion begins when the user is ready to stand up. For state estimation, teaching data is collected and models are learned before the robot is used. During use, the robot measures and collects user data such as foot position and upper body angle using sensor input, as shown in Figure 2. Sometimes the data is processed, and the learned model uses the raw or processed data as input and classifies it as state estimation. The action determined by the system depends on the estimation result, then the actuators provide an output to the external environment. This system can be fully described separately from the learned model itself.

Transparency of classification reasons is studied in the computer vision and AI fields [21, 22]. For the systems addressed in this research, input data and classification results are transparent and classification reasoning is unclear. For example, clinical image recognition systems detect diagnostic signs based on deep learning, and report to a doctor. Thus, humans can understand classification results and decision making, while conducting the action remains the humans role. This type of research focuses on algorithmic transparency, however there are various parts of learning-based robots that are unclear, as shown in Figure 2. Learning-based systems are unclear for estimation results, decision making, and robot action, which are the usual stakeholders' usual interests. Users generally need to know the robot's current actions, action plans, and decision making for efficient operations; the user's required actions are also important information. Engineers who investigate or maintain robot systems should understand the internal processes of the systems for maintenance, investigation, or repair. Sensor data and actuator output torques are also required. By contrast, the relationship between input data and classification results, which is the role of a learning algorithm, is not relevant for these tasks. By using the representation of classification reasons, the system will become more transparent, however, we consider that representing the relationships between data, decisions, and actions is more important. Representation can effectively improve systems interpretability; therefore, we propose a method for describing robot systems that adopts machine or deep learning.

3.2 *Information Ontology*

Describing the entire system is essential not only for AI transparency but also for information ontology. Defining the internal processes of the system allows us to know which parts of the system are not transparent, which enables us to determine what information should be represented. For example, if a problem occurs where an actuator moves without the user's intent, engineers and accident investigators will want to investigate it. If there are no representations of internal process information, investigators can only estimate based on the system outputs. Investigators would then have to extract the required information from a large quantity of stored data, much of which is unnecessary. Meanwhile, if the whole system is described, internal processes and data relationships can be understood. The system can represent related information such as decided actions, estimation results, and inputs.

To determine the best representation of information, we should consider relationships between stakeholders, information, and interfaces. If the relationships between stakeholders and information, known as use cases, are known, the system can represent information easily according to the situation.

Stakeholder-interface relationships are also important because specialized interfaces are useful and efficient. However, it is generally difficult for ordinary people to use such specialized tools. It is therefore important to change interfaces according to the stakeholder as shown in Figure 3. An intuitive interface should be provided for general users, while specialized interfaces should be created according to professional standards.

Systems should provide a variety of different types of information. For example, the user state is spatial information, and does not change rapidly. By contrast, the timing of the robot's support is temporal information, and is not continuous but event-driven. There are several media by which to transmit information such as vision and sound. The medium should be selected according to the features of transmitted information.

All information can be extracted by describing the entire system. Interfaces representing the required information for specialists are created based on professional standards. By transcribing the interfaces into intuitive formats, an interface for general users can be constructed.

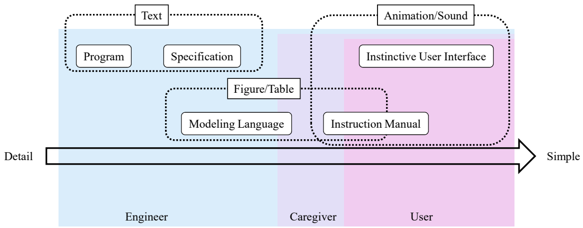


Figure 3. Stakeholder-Interface Relationship.

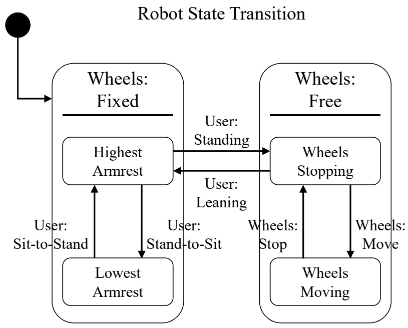


Figure 4. Example of State Machine Diagram.

4. Describing Whole System

This section explains a method for describing the entire system. AI transparency can be achieved by describing the entire system; we can then determine which information from the described system should be represented.

There are various ways to describe algorithms, processes, and systems. Flowcharts are one of the most famous diagrams used for representing algorithms. They can represent not only processes but also conditional decisions. For describing state transitions, a Finite State Machine is used and applied to several modeling languages. Each of these have different advantages, and several modeling languages are composed by applying these models.

For this study, we adopted the modeling language SysML [41] to describe system states, internal processes, and external interactions. SysML has several types of diagrams for describing systems. Some diagrams are suitable for real-time representation, while others are useful for investigation. They can be selected depending on the intended use.

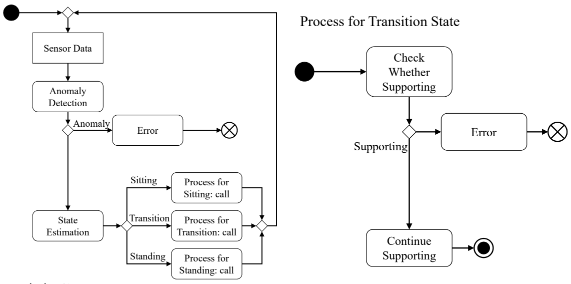
SysML itself and the features of its diagrams are explained in section 4.1. The uses and advantages of SysML diagrams are also explained, using an assistive robot as an example. We then explain how to describe systems that include learning algorithms by using SysML in section 4.2.

4.1 *SysML*

SysML is a modeling language, developed for systems engineering based on Unified Modeling Language (UML). UML is useful in software engineering; however, systems with hardware cannot be sufficiently described with UML. SysML was developed to enable experts from different fields, such as programmers, designers, and electrical engineers, to collaborate on system development. SysML consists of nine types of diagrams: Requirement Diagrams, Activity Diagrams, Sequence Diagrams, State Machine Diagrams, Use Case Diagrams, Block Definition Diagrams, Internal Block Diagrams, Parametric Diagrams, and Package Diagrams. It is unnecessary to adopt all the diagram types, and it is possible to make more than one diagram for each diagram type.

State Machine Diagrams are for describing a system's state based on a Finite State Machine (FSM). By using the diagrams, the system's state or behavior can be described as shown in Figure 4. The example diagram is for a support robot which can assist with standing, sitting, and walking. The robot has two basic states, in which wheels are fixed or free, and assists the user's standing and sitting by moving the armrests vertically. It is assumed unsafe if the wheels are fixed or the armrest moves when the wheels are moving; thus, a state transition between the wheels fixed state and the wheels free state can occur only when the wheels are stopped, as shown in Figure 4. If the system estimates user states, surrounding people, or the environment, they can also be described with State Machine Diagrams.

Activity, Sequence, and Internal Block Diagrams are useful for describing a system's internal processes, input-output relationships, and actions. Internal Block Diagrams can clearly describe input-output relationships and are especially useful in cases where the system contains many



(a) State Estimation and Conditional Branching

(b) Activity for User in Transition State

Figure 5. Examples of Activity Diagram.

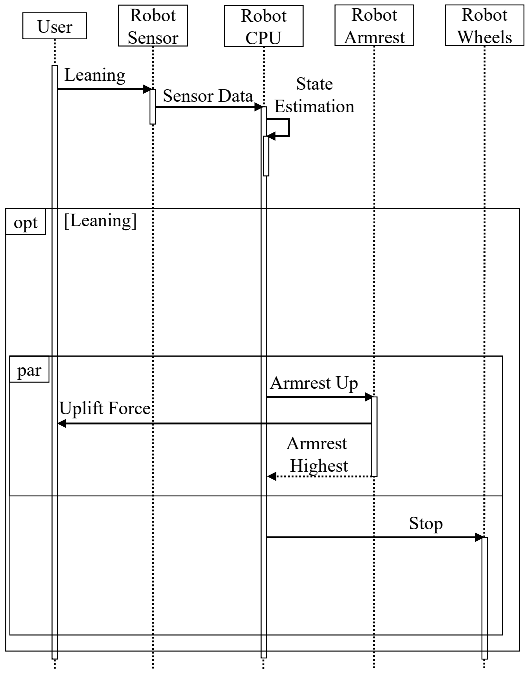


Figure 6. Example of Sequence Diagram.

parts. Activity Diagrams are good for describing internal processes included in some kinds of algorithms, such as learning and estimation. They have advantages if there is some conditional branching. Internal processes of an example robot system are shown in Figure 5. The system has anomaly detection and state estimation as shown in Figure 5(a). First, the robot checks for anomalies using the sensor data. If no anomaly is detected, the robot classifies the user's state, whether sitting, standing, or transitioning between them. The next processes for each state are described in other diagrams; an example is shown in Figure 5(b). This diagram shows the process when the user is in a transition state. The robot is meant to be conducting a support action when the user's state is transitioning between sitting and standing. If no support is conducted when the user is in a transition state, the system determines that an error has occurred, as shown in Figure 5(b). If support is provided and the user is still in a transition state, then the support is not yet complete and the robot continues to provide support. Sequence Diagrams are most useful for describing a time sequence for internal or external systems interaction, as shown in Figure 6. If the user is sitting and leaning forward, the sensor data is sent to the computer embedded in the robot and the computer estimates the user's leaning from sensor data in the same way as the state estimation described above. The robot's computer then sends a command to the linear actuator of the armrest and the robot's wheels, and the armrest applies lifting force to the user to assist with the sit-to-stand motion.

Defining stakeholders and their use cases is valuable for achieving accountability. Use Case Diagrams make it possible to define the systems stakeholders and explain the purpose of their relationship with the system as shown in Figure 7. The example shows that the robot supports three activities for a user: sit-to-stand, stand-to-sit, and walking. Engineers interact with the robot for maintenance and repair. By using the Use Case Diagrams, the relationship between stakeholders and information, which is necessary for representation of information, can be obtained, as explained in section 5 in detail.

4.2 *Describing Systems That Include Learning Algorithms*

The process for a system that includes a learning algorithm is described in SysML as shown in Figure 5. Sensor data is first used as input for anomaly detection, as shown in Figure 5(a). Next, the raw sensor data are processed for state estimation based on a machine learning algorithm. The next process is determined based on the result of the state estimation. The process that occurs during user transition is shown in Figure 5(b). It can be determined that the output action maintains support from Figure 5(b). Although the relationship between the input data and estimated state cannot be fully understood, we can know the decision making reason.

By describing information required in various situations based on these figures, stakeholders can more fully understand the robot's actions and plans than they would without representation. Stakeholders need to understand the robot's actions and reasons to use, maintain, and repair it. For example, users can understand the robot's actions from Figure 5(b). It is also clear that the robot decides the appropriate action based on the estimated user state from Figure 5(a). A system that includes a learning algorithm can therefore be described transparently enough for internal processes to be understood in this way. Stakeholders can also investigate accumulated data based on these figures, however, the figures themselves are difficult to understand for non-engineer stakeholders, as other stakeholders rarely have SysML knowledge. Appropriate interfaces are required for each stakeholder, as explained in section 5.

5. Representation of Information

This section explains the importance of a system's accountability of the relationships between stakeholders, information, and interfaces.

To achieve accountability, systems should provide the required information to stakeholders. It is necessary to determine which stakeholders need which information. This stakeholder-information relationship can be treated as a use case of the system, which is explained in detail in section 5.1.

It is less than transparent if the provided information is difficult to understand. Any existing standards for presentation of the information should be followed. People who have professional knowledge can easily understand information written in specialized language, however, people who have no specialized knowledge cannot. This stakeholder-interface relationship is discussed in section 5.2.

It is also important to select a transfer method based on the features of the information. For example, spatial information is easily expressed visually, while temporal information can be expressed through sound. This information-interface relationship is explained in section 5.3.

5.1 *Use Case: Stakeholder-Information Relationship*

Use cases mainly depend on the types and functions of the system. Developers of systems invariably consider use cases of the systems, and SysML Use Case Diagrams can inform their consideration and discussions. Developers tend to focus only on use cases of users; however, there are other stakeholders for the systems. Almost all systems have users, engineers, and accident investigators as stakeholders. For assistive robots, stakeholders may include people such as caregivers, family, and certification authorities.

Who interacts with the system and how they interact is integral to the question of who needs which information; thus, use cases are important for information representation. For example, users often want to know about a system's behavior, state, or action and people become anxious if there is an autonomous robot doing something incomprehensible. Users require understanding of current behaviors and actions, as well as future plans, and they may want to know the reason for the system's actions. If there is a difference between the intended action of the user and that of the system, people want to know the reason and a solution. Therefore, 'what should I do' is one of the most important pieces of information for systems that include interaction with humans.

Engineers relate to systems for development as well as for maintenance and repair. They can conduct their work more efficiently if they understand the system. To check whether a system is behaving normally, information about a system's internal and external behaviors are required. This knowledge is also useful for accident investigators.

5.2 *Professional Standard: Stakeholder-Interface Relationship*

As described in section 4, SysML has been popular among engineers who are involved in development. Engineers can easily understand systems by using SysML, however, it is not easy for others; therefore electing the interface according to stakeholders is therefore important.

If there are standards for required information, these should be followed. Accident investigation is governed by laws and there are standards for home electronics. Laws for robots and AI systems are now being discussed, therefore we consider other professional standards, such as engineering standards, in this paper. If standards for accident investigation for robots is determined, we intend to adapt our interface to the standard.

Generally, it should be assumed that users have no professional knowledge, and a user interface should be intuitive and usable.

5.3 *Media: Information-Interface Relationship*

Humans receive information mainly through eyes and ears. Vision and sound are therefore good options for information transfer media.

Each of these media has advantages and disadvantages. For example, visualization is useful for representing multiple or spatial information, however it is impossible to transmit information by sight if the user's focus is not where the information is being presented. In this case, sound is a more effective means of information transfer.

These features have been discussed in the interface design and feedback in motor learning fields [42–44]. Following these studies and considering the information that needs to be represented, we summarize the features of vision and sound as follows:

- Vision
  - Spatial Information
  - Multi-Information
  - Steady Information
- Sound
  - Temporal Information
  - Event-Driven Information
  - Information Which Should Be Transmitted Even If Human Is Unnoticed

Selecting media based on the features of transmitted information is especially effective for general users because it makes the interface more instinctive.

6. Implementation

In this section, the assistive robot system developed based on the proposed design method is explained. The robot has been developed to assist the user's walking, sit-to-stand, and stand-to-sit motions. It estimates the user's state by using a machine learning algorithm and decides on its action based on the result of the estimation.

Two types of interfaces are explained in section 6.3 and section 6.4. The investigation interface should be effective for accident investigation, especially for an autonomous robot, as it makes it easier to determine responsibility for accidents as well as to detect and fix failures. The user interface makes the system transparent for users and helps them use the system. The system uses the user's estimated state to decide the robot's action. Misunderstanding the user's state is a safety critical issue, so the user interface representing the estimation result is important.

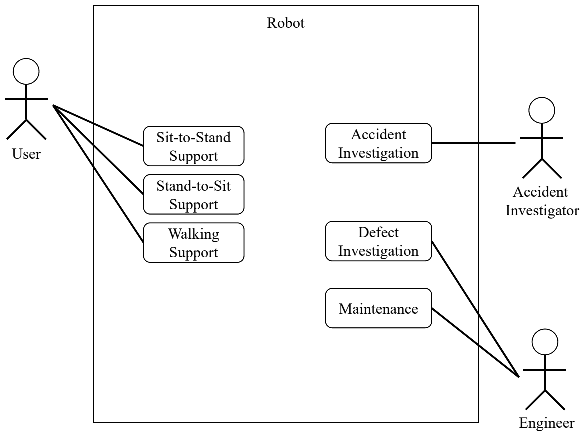


Figure 7. Example of Use Case Diagram.

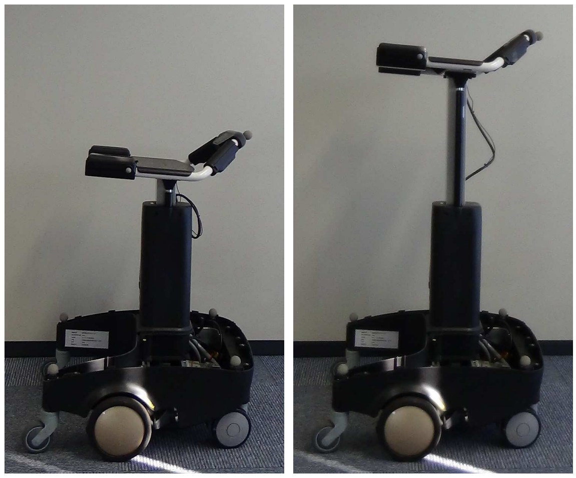
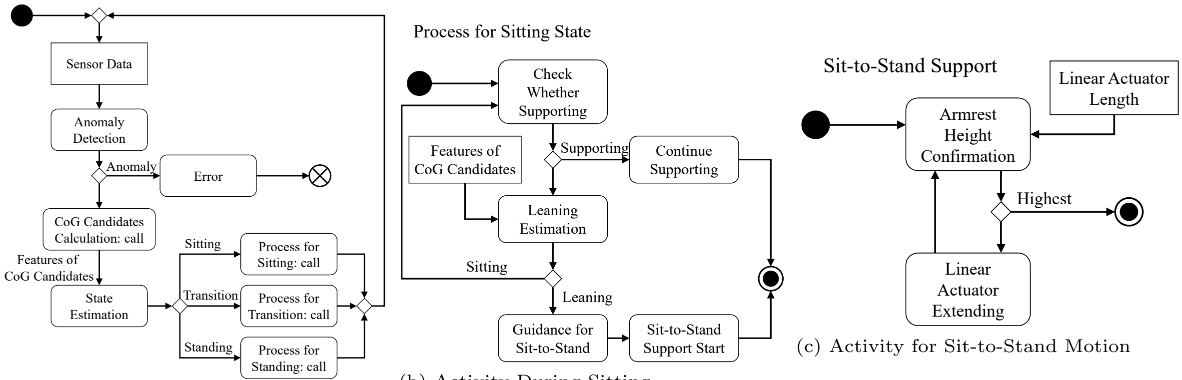


Figure 8. Developed Robot: (Left) Lowest Armrest; (Right) Highest Armrest [47].



(a) Rough State Estimation

(b) Activity During Sitting

(c) Activity for Sit-to-Stand Motion

Figure 9. Activity Diagrams.

6.1 *Developed Robot System*

The developed support robot is shown in Figure 8. Its 2 center wheels can be controlled with motors and 4 casters to allow the robot to assist a user's walking. The armrest can be moved vertically by a linear actuator, and it can support lifting the user's body upward during a sit-to-stand motion. Such armrest rising can assist user's standing up motion [45, 46]. The armrest can also be used for a stand-to-sit motion. When the linear actuator is moving, the wheels are braked for safety reasons. Hardware specifications are detailed in [46].

Pressure sensors are installed on the robot's gripper and armrest to recognize whether the user is touching them. A Position Sensitive Detector (PSD) on the armrest can measure the distance between the robot and its user. The system can use these sensors to estimate the user's state based on the method proposed in [47]. The detail of sit-to-stand support is explained in section 6.2.

6.2 *Describing in SysML*

Data from the robot's sensors is used for anomaly detection as shown in Figure 9(a). If the user's hands and elbows are not touching the gripper and armrest, the robot does not move for security reasons. CoG candidates are calculated by using sensor data if no anomaly is detected. The CoG candidates calculation method is explained in detail in [47, 48]. The features of CoG candidates are used for SVM and the system estimates the user's state from three possibilities:

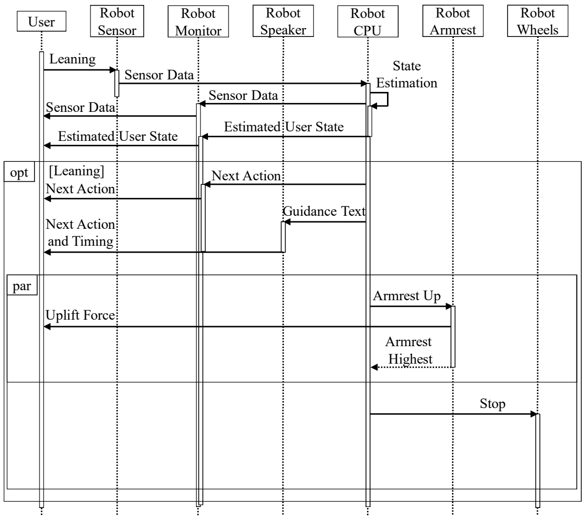


Figure 10. Sequence Diagram.

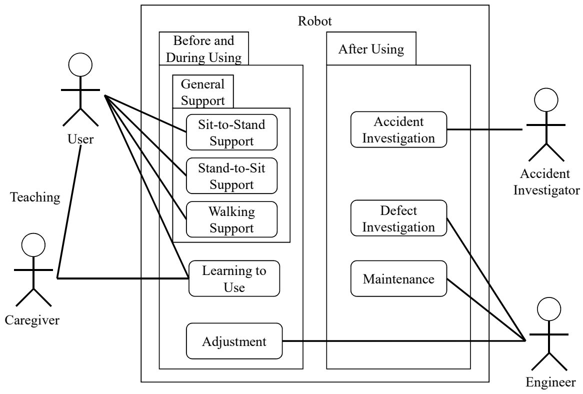


Figure 11. Use Case Diagram.

sitting, standing, or the transition between them. The state estimation method is explained in [47]. The system then carries out functions depending on the estimated user's state. Functions for each state are described in other diagrams, and an example is shown in Figure 9(b).

The process for the case where the user is sitting is shown in Figure 9(b). If the user is estimated to be sitting, the system checks whether the robot is carrying out the sit-to-stand support action. If no support action is being conducted, the system estimates whether the user is only sitting or sitting with a leaning upper body using the same method as the state estimation described above. The system begins supporting the user's sit-to-stand motion if the user is leaning on the armrest of the robot. The output of the system for sit-to-stand support is extending the linear actuator and moving the armrest upwards as shown in Figure 9(c). Extension of the linear actuator continues at constant speed until the armrest reaches its maximum height. These processes can be described in SysML as shown in Figure 9. Sequence diagrams can also describe the processes as shown in Figure 10.

Use cases are shown in Figure 11. Users and caregivers of the system primarily interact with the system before and during its use. The most important pre- and during-use information are the instructions on how to use the system.

The system is autonomous and includes learning algorithms, hence users need to know what the system does and why. Engineers usually interact with the system before and after its use. Before use, engineers adjust the system for the user in cases where the user's personal parameters are required for adjustment. After using, engineers interact with the system for maintenance and trouble or accident handling, and accident investigators relate to it for investigating accidents.

It is confirmed by Figure 9, Figure 10, and Figure 11 that the internal processes of the robot system can be described by SysML. Effectiveness of modeling languages for describing systems is also presented in [28, 29]. SysML was created to describe systems and is becoming popular in the system engineering field; thus, it is thought to be effective. We can detect the required information from the system as described. Examples of transcription of the information to the appropriate interface for each stakeholder are explained in section 6.3 and section 6.4.

6.3 *Interface for Investigation*

When a failure or accident happens, engineers and accident investigators need to investigate or repair it. They want to know the system architecture, reasons for the accident, solutions to faults, and other information. In this section, an example of representation for investigation is explained.

As explained in section 6.2, when a user leans forward while sitting, the armrest moves up. If the linear actuator does not move, the user believes that it is defective and wants an engineer to solve the problem. During investigation, the engineer enters their professional title, required time, and known information, and the system extracts and presents related information. Internal processes are also presented based on described system architecture.

An example of the interface is shown in Figure 12. In this case, the engineer knows only the robot's action (failure to move the armrest) and the time. The engineer first wants to compare the robot's actual action with its decided action. Related information is required for investigation, namely: raw input data (PSD, pressure), processed input data (CoG candidates and their features), estimation (user's basic state: sitting, standing, or transition; user's state for sitting: normal sitting or leaning forward), decided action (none or moving the armrest up), and actuator output (linear actuator torque). The system shows the internal process for that time as SysML diagrams, as shown in the right part of Figure 12. The related information is marked using different colors according to the type of information.

A list of all information is represented in the left part of the interface. Related pieces of information are marked with the same color as the SysML diagrams. By selecting the information, a graph of the data is displayed as shown in Figure 12. From the graph on the right, the system estimation appears to be correct, and from the graph on the left, we can see that the defect was caused by abnormal gripper sensor data. In this way, the interface make it possible to investigate defects and detect the causes of them.

6.4 *Interface for Users in Use*

An intuitive interface is required for system users. Required information and features are shown in TABLE 1. The user's state of contact with the gripper and armrest is used for anomaly detection, while the user's estimated state is used for deciding on the robot's action. The user can check whether there is a difference between the user's actual state and that recognized by the robot. The user's intended action, such as the sit-to-stand motion, is also useful. The user's required action such as leaning forward, the robot's intended action such as moving the armrest up, and the support timing are required information, as they are useful for assisting the user's action.

Contact information is for both the gripper and the armrest; thus, it contains two data streams. The contact situation, user's state, and robot's armrest height change infrequently and are all comprised of spatial information. The actions of the user and the robot are also spatial information, and are event-driven, while the timing of support is temporal information. When the robot carries out an action, it should make the user aware of the action.

From TABLE 1, the best means of communication, whether vision or sound, can be determined. Vision is selected for the user's state, armrest height, and user's required action, while sound is adopted for transmitting support timing. The robot's intended action is transmitted using both vision and sound.

An example of the displayed user interface is shown in Figure 13. The user's body is represented with black lines and a black circle. On the upper part of the interface, the user's state and robot's armrest height are displayed. Current status is shown in black, while other statuses are in grey as shown in Figure 14. The armrest movement is described by a red arrow, drawn according to the armrest height as shown in Figure 14(d) and (e). The robot's intended action, in this case moving the armrest up, is displayed with a flashing red arrow as shown in Figure 14(b) and (c), and is also denoted by sound from a speaker. Timing of the support is represented by sound as a countdown.

The lower half of the display represents the user's required action. Black lines and circles represent the user's body in the same manner as on the upper part of the display. This represents the user's leaning action by animation as shown in Figure 15. The four figures are displayed in sequence to represent a leaning motion.

Table 1. Selecting Media for Each Information

	Vision			Sound		
	Spatial	Multi	Steady	Temporal	Event-Driven	Make Aware
Contact Situation	○	○	○			
User's State	○		○			
Armrest Height	○		○			
User's Intended Action	○				○	
User's Required Action	○				○	
Robot's Intended Action	○				○	○
Support Timing				○	○	

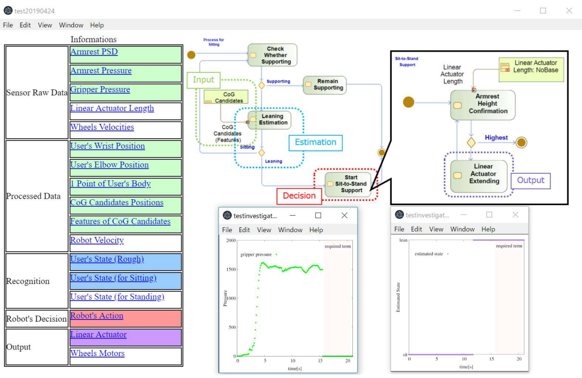


Figure 12. Investigation Interface.

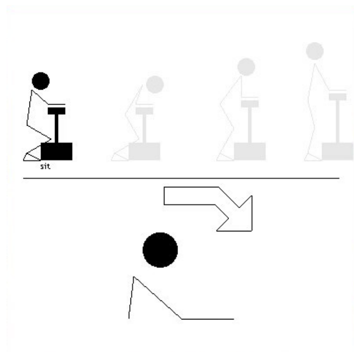
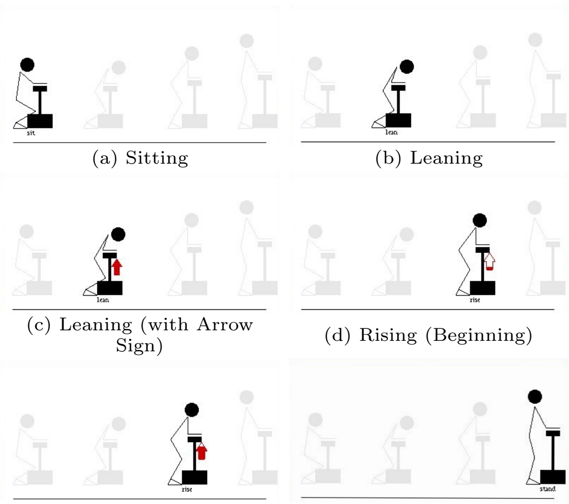
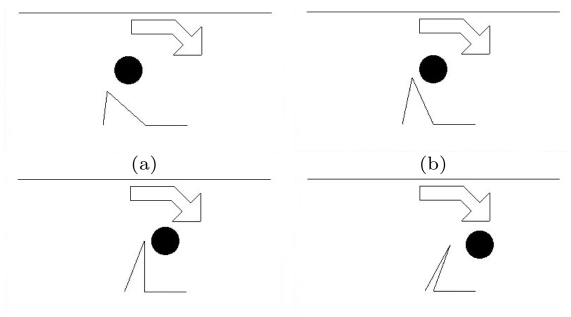


Figure 13. Overall View of User Interface.



(a) Sitting
(b) Leaning
(c) Leaning (with Arrow Sign)
(d) Rising (Beginning)
(e) Rising (Ending)
(f) Standing

Figure 14. Upper Part of User Interface.



(a) (b) (c) (d)

Figure 15. Lower Part of User Interface. The animation consist of four figures which appear in order of (a) to (d).

Several participants tested the user interface during robot use to confirm the usability of the interface and system. The participants were all males in their 20s with no physical disabilities. Informed consent was obtained from all participants prior to the experiments. The participants conducted sit-to-stand motions several times, using the robot user interface. From the completed questionnaires, the effectiveness of the interface is confirmed.

From the experimental results and the developed interfaces, we confirmed that systems can be accountable for various stakeholders and their situations by designing based on the proposed design architecture. We intend to conduct experiments that simulate daily activity with elderly people as validation of the proposed method. The final interface should be designed using interface design principles, although appropriate information can be confirmed by using the interface prototype.

7. Conclusion

In this study, we proposed general design architecture for accountable learning-based support robot systems. Robot systems, especially support robots that include AI and p-HRI, should be able to account for their actions. Describing systems and representing the required information for all stakeholders can make the system transparent.

It is difficult to make a learning algorithm transparent, although a system that includes a learning algorithm can be explained. Robot systems usually adopt learning algorithms for recognition or estimation of some parameter. Input sensor data, estimation results, decided action, and actuator output can all be explained, and thus humans can understand the systems. SysML is adopted in this paper for describing the robot system.

Information should be transmitted to each stakeholder in an appropriate way, hence relationships between stakeholders, information, and interfaces are important. Use cases show the required information, which varies according to stakeholders and situations. SysML use case diagrams have particular advantages for this application. Specific factors influence the stakeholder-interface relationship. An intuitive user interface should be adopted for ordinary users, and appropriate interfaces will vary based on the different features and uses of each type of information. An implementation example is shown in section 6. The effectiveness of the proposed design architecture is validated to make accountable learning-based support robot systems.

A completely implemented system should be developed for validation in future work. An interface for accident investigators can be designed by expanding the interface for engineers. Legal standards for robot accidents are still under discussion because robots using AI are not yet common. The accident investigation interface should be designed following the relevant standards. Tests with elderly people can validate the proposed design method and improve interfaces.

ACKNOWLEDGMENT

We would like to thank Editage (www.editage.com) for English language editing.

References

[1] R. Moe-Nilssen. A new method for evaluating motor control in gait under real-life environmental conditions. part 2: Gait analysis. *Clinical Biomechanics*, 13(4):328 – 335, 1998.

[2] Xinan Li, Hongyuan Xu, and Jeffrey T. Cheung. Gait-force model and inertial measurement unit-based measurements: A new approach for gait analysis and balance monitoring. *Journal of Exercise Science & Fitness*, 14(2):60 – 66, 2016.

[3] Y. Hirata, S. Komatsuda, and K. Kosuge. Fall prevention control of passive intelligent walker based on human model. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1222–1228, Sep. 2008.

[4] P. Di, J. Huang, S. Nakagawa, K. Sekiyama, and T. Fukuda. Fall detection and prevention in the elderly based on the zmp stability control. In *2013 IEEE Workshop on Advanced Robotics and its Social Impacts*, pages 82–87, Nov 2013.

[5] M. Miezal, B. Taetz, and G. Bleser. Real-time inertial lower body kinematics and ground contact estimation at anatomical foot points for agile human locomotion. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3256–3263, May 2017.

[6] K. Zabjek, C. Coillard, C. Rivard, and F. Prince. Estimation of the centre of mass for the study of postural control in idiopathic scoliosis patients: A comparison of two techniques. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society*, 17:355–60, 04 2008.

[7] Andrea Mannini and Angelo Maria Sabatini. Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors*, 10(2):1154–1175, 2010.

[8] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L. Reyes-Ortiz. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In José Bravo, Ramón Hervás, and Marcela Rodríguez, editors, *Ambient Assisted Living and Home Care*, pages 216–223, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[9] Denis Tome, Chris Russell, and Lourdes Agapito. Lifting from the deep: Convolutional 3d pose estimation from a single image. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul 2017.

[10] Jonathan J Tompson, Arjun Jain, Yann LeCun, and Christoph Bregler. Joint training of a convolutional network and a graphical model for human pose estimation. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 1799–1807. Curran Associates, Inc., 2014.

[11] Xiuxing Yin and Li Pan. Enhancing trajectory tracking accuracy for industrial robot with robust adaptive control. *Robotics and Computer-Integrated Manufacturing*, 51:97 – 102, 2018.

[12] Xiuxing Yin and Li Pan. Direct adaptive robust tracking control for 6 dof industrial robot with enhanced accuracy. *ISA Transactions*, 72:178 – 184, 2018.

[13] K. Kiguchi and Y. Hayashi. An emg-based control for an upper-limb power-assist exoskeleton robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(4):1064–1071, Aug 2012.

[14] Y. Hirata, H. Yamaya, K. Kosuge, A. Koujina, T. Shirakawa, and T. Katayama. Anomaly state assessing of human using walker-type support system based on statistical analysis. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 146–152, Aug 2015.

[15] T. Takeda, K. Kosuge, and Y. Hirata. Hmm-based dance step estimation for dance partner robot -ms dancer-. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3245–3250, Aug 2005.

[16] Y. Hirata, A. Muraki, and K. Kosuge. Motion control of intelligent walker based on renew of estimation parameters for user state. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1050–1055, Oct 2006.

[17] O. Chuy, Y. Hirata, Zhidong Wang, and K. Kosuge. Motion control algorithms for a new intelligent robotic walker in emulating ambulatory device function. In *IEEE International Conference Mechatronics and Automation, 2005*, volume 3, pages 1509–1514 Vol. 3, July 2005.

[18] R. H. Wortham, A. Theodorou, and J. J. Bryson. Improving robot transparency: Real-time visualisation of robot ai substantially improves understanding in naive observers. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 1424–1431, Aug 2017.

[19] Robert H. Wortham, Andreas Theodorou, and Joanna J. Bryson. Robot transparency: Improving understanding of intelligent behaviour for designers and users. In Yang Gao, Saber Fallah, Yaochu Jin, and Constantina Lekakou, editors, *Towards Autonomous Robotic Systems*, pages 274–289, Cham, 2017. Springer International Publishing.

[20] Sichao Song and Seiji Yamada. Designing led lights for a robot to communicate gaze. *Advanced Robotics*, 33(7-8):360–368, 2019.

[21] Dong Huk Park, Lisa Anne Hendricks, Zeynep Akata, Bernt Schiele, Trevor Darrell, and Marcus Rohrbach. Attentive explanations: Justifying decisions and pointing to the evidence. *CoRR*, abs/1612.04757, 2016.

[22] Jeffrey De Fauw, Joseph R. Ledsam, Bernardino Romera-Paredes, Stanislav Nikolov, Nenad Tomasev, Sam Blackwell, Harry Askham, Xavier Glorot, Brendan O ' Donoghue, Daniel Visentin, George van den Driessche, Balaji Lakshminarayanan, Clemens Meyer, Faith Mackinder, Simon Bouton, Kareem Ayoub, Reena Chopra, Dominic King, Alan Karthikesalingam, and Olaf Ronneberger. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24, 08 2018.

[23] Rahmadi Kurnia, Md. Altab Hossain, Akio Nakamura, and Yoshinori Kuno. Generation of efficient

and user-friendly queries for helper robots to detect target objects. *Advanced Robotics*, 20(5):499–517, 2006.

[24] A. Green and K. S. Eklundh. Designing for learnability in human-robot communication. *IEEE Transactions on Industrial Electronics*, 50(4):644–650, Aug 2003.

[25] Kazuhiko Shinozawa, Takahiro Miyashita, Masayuki Kakio, and N. Hagita. User specification method and humanoid confirmation behavior. In *2007 7th IEEE-RAS International Conference on Humanoid Robots*, pages 366–370, Nov 2007.

[26] D. F. Paez Granados, B. A. Yamamoto, H. Kamide, J. Kinugawa, and K. Kosuge. Dance teaching by a robot: Combining cognitive and physical human-robot interaction for supporting the skill learning process. *IEEE Robotics and Automation Letters*, 2(3):1452–1459, July 2017.

[27] Jekaterina Novikova and Leon Watts. Towards artificial emotions to assist social coordination in hri. *International Journal of Social Robotics*, 7(1):77–88, Feb 2015.

[28] Mahmood Hosseini, Alimohammad Shahri, Keith Phalp, and Raian Ali. A modelling language for transparency requirements in business information systems. In Selmin Nurcan, Pnina Soffer, Marko Bajec, and Johann Eder, editors, *Advanced Information Systems Engineering*, pages 239–254, Cham, 2016. Springer International Publishing.

[29] Robert H Wortham, Swen E Gaudl, and Joanna J Bryson. Instinct: A biologically inspired reactive planner for embedded environments. In *Proceedings of ICAPS 2016 PlanRob Workshop, London, UK*, 2016.

[30] S. Pakrasi, Oliver Burmeister, J.F. Coppola, T. J. McCallum, and G. Loeb. Ethical telehealth design for users with dementia. *Gerontechnology*, 13(4):383–387, 2015. Imported on 12 Apr 2017 - DigiTool details were: Journal title (773t) = Gerontechnology. ISSNs: 1569-1101;.

[31] Gianmarco Baldini, Maarten Botterman, Ricardo Neisse, and Mariachiara Tallacchini. Ethical design in the internet of things. *Science and Engineering Ethics*, 24(3):905–925, Jun 2018.

[32] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, first edition. https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html, 2019.

[33] J. Bryson and A. Winfield. Standardizing ethical design for artificial intelligence and autonomous systems. *Computer*, 50(5):116–119, May 2017.

[34] Y. Weng and Y. Hirata. Ethically aligned design for assistive robotics. In *2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR)*, pages 286–290, Aug 2018.

[35] AMED. Nursing-care robots development guidline. `http://robotcare.jp/?page_id=5731`, 2018.

[36] Bernhard Waltl and Roland Vogl. Explainable artificial intelligence—the new frontier in legal informatics. *Jusletter IT*, 4:1–10, 2018.

[37] Guido Bologna and Yoichi Hayashi. Characterization of symbolic rules embedded in deep dimlp networks: A challenge to transparency of deep learning. *Journal of Artificial Intelligence and Soft Computing Research*, 7:265, 07 2017.

[38] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.*, 267:1–38, 2019.

[39] Adrian Weller. Challenges for transparency. *CoRR*, abs/1708.01870, 2017.

[40] Bernadette Boscoe. Creating transparency in algorithmic processes. *Delphi - Interdisciplinary Review of Emerging Technologies*, 2(1), 2019.

[41] OMG. OMG Systems Modeling Language (OMG SysML), Version 1.3. `http://www.omg.org/spec/SysML/1.3/`, 2012.

[42] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin & Review*, 20(1):21–53, Feb 2013.

[43] Jenny Burke, Matthew Prewett, Ashley A. Gray, Liuquin Yang, Frederick Stilson, Michael Coovert, Linda Elliott, and Elizabeth S. Redden. Comparing the effects of visual-auditory and visual-tactile feedback on user performance: A meta-analysis. In *ICMI'06: 8th International Conference on Multimodal Interfaces, Conference Proceeding*, pages 108–117, 01 2006.

[44] Markus Funk, Juana Heusler, Elif Akcay, Klaus Weiland, and Albrecht Schmidt. Haptic, auditory, or visual?: Towards optimal error feedback at manual assembly workplaces. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '16, pages 43:1–43:6, New York, NY, USA, 2016. ACM.

[45] T. Hatsukari, S. Kuroko, N. Miyake, J. Higuchi, R. Kawazoe, Y. Hirata, and K. Kosuge. Self-

help standing-up method based on quasi-static motion. In *2008 IEEE International Conference on Robotics and Biomimetics*, pages 342–347, Feb 2009.

[46] S. Kawazoe, D. Chugo, S. Yokota, H. Hashimoto, T. Katayama, Y. Mizuta, and A. Koujina. Development of standing assistive walker for domestic use. In *2017 IEEE International Conference on Industrial Technology (ICIT)*, pages 1455–1460, March 2017.

[47] M. Takeda, Y. Hirata, T. Katayama, Y. Mizuta, and A. Koujina. State estimation using the cog candidates for sit-to-stand support system user. *IEEE Robotics and Automation Letters*, 3(4):3011–3018, Oct 2018.

[48] M. Takeda, Y. Hirata, K. Kosuge, T. Katayama, Y. Mizuta, and A. Koujina. Human cog estimation for assistive robots using a small number of sensors. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6052–6057, May 2017.