Journal of Manufacturing Systems 63 (2022) 392–410

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

Evaluating quality in human-robot interaction: A systematic search and classification of performance and human-centered factors, measures and metrics towards an industry 5.0

Enrique Coronado [a,b,*], Takuya Kiyokawa [c,d], Gustavo A. Garcia Ricardez [d,e], Ixchel G. Ramirez-Alpizar [b], Gentiane Venture [a,b], Natsuki Yamanobe [a,b]

[a] *Department of Mechanical Systems Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan*
[b] *National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan*
[c] *Department of Systems Innovation, Graduate School of Engineering Science, Osaka University, Osaka, Japan*
[d] *Division of Information Science, Graduate School of Science and Technology, Nara Institute of Science and Technology (NAIST), Nara, Japan*
[e] *Ritsumeikan University, Shiga, Japan*

A R T I C L E I N F O

Keywords:
Human-robot interaction
Human-robot collaboration
Metrics
Robotics
Industry 5.0
Society 5.0 2010 MSC: 00–01
99–00

A B S T R A C T

Industry 5.0 constitutes a change of paradigm where the increase of economic benefits caused by a never-ending increment of production is no longer the only priority. Instead, Industry 5.0 addresses social and planetary challenges caused or neglected in Industry 4.0 and below. One relevant the most relevant challenges of Industry 5.0 is the design of human-centered smart environments (i.e., that prioritize human well-being while maintaining production performance). In these environments, robots and humans will share the same space and collaborate to reach common objectives. This article presents a literature review of the different aspects concerning the problem of quality measurement in Human-Robot Interaction (HRI) applications for manufacturing environments. To help practitioners and new researchers in the area, this article presents an overview of factors, metrics, and measures used in the robotics community to evaluate performance and human well-being quality aspects in HRI applications. For this, we performed a systematic search in relevant databases for robotics (Science Direct, IEEE Xplore, ACM digital library, and Springer Link). We summarize and classify results extracted from 102 peer-reviewed research articles published until March 2022 in two definition models: 1) a taxonomy of performance aspects and 2) a Venn Diagram of common human factors in HRI. Additionally, we briefly explain the differences between often confusing or overlapped concepts in the area. We also introduce common human factors evaluated by the robotics community and identify seven emergent research topics which can have a relevant impact on Industry 5.0.

1. Introduction

Human-Robot Interaction (HRI) is one of the most promising technologies in service, social and industrial contexts. When implementing an HRI system, developers must evaluate how well the proposed system meets individual, collective, and production needs or objectives (i.e., it's quality). From the engineering point of view, the concept of quality usually refers to the degree to which a system, service, product, or process is in conformance to specified requirements and under specified conditions [1,2]. In this context, quality models are well-accepted means to support, evaluate and manage quality [3]. The most basic type of quality model in engineering is denoted as conceptual or definition model. The purpose of a definition model is to describe a set of quality factors or characteristics and the possible relationships between them [3,4]. This type of quality models generally do not provide ways for constructive quality assurance. However, they can be used as a reference when it is required to select a set of relevant factors enabling the experimental validation of applications, services or systems. These models can also be used as the theoretical or conceptual core of more complex and analytical quality models, toolboxes, or frameworks [3]. A relevant step previous the creation of a definition model is the identification of those relevant factors able to describe quality in the specified context as well as the metrics and measures that can be used to evaluate these factors. While the identification and classification of factors,

* Corresponding author at: Department of Mechanical Systems Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan.
E-mail address: enriquecoronadozu@gmail.com (E. Coronado).

https://doi.org/10.1016/j.jmsy.2022.04.007
Received 19 January 2022; Received in revised form 18 March 2022; Accepted 8 April 2022
Available online 28 April 2022

measures, and metrics describing quality have been received significant attention in areas such as software engineering [5,6], these efforts are still rare in HRI [7]. Moreover, factors that can be considered valuable for describing quality or robotics systems can vary depending on different interests or points of view in the robotics community. From the industrial point of view, it is possible to identify two different types of interests: *performance-centered*) and *human-centered*. On the one hand, *performance-centered* or *machine-centered* paradigms consider robots as means enabling the optimization of the production process with the main objectives of reaching mass production and improving economic benefits. This point of view, in many cases, implies the substitution of humans with more efficient machines to reach full automation. On the other hand, the main objective of *human-centered* paradigms is to improve general human-well being by respecting their role, needs, job, talents, and rights. In this point of view, robotics systems must complement or enhance human work and capabilities rather than machines designed to substitute them [8,9].

Industry 4.0 is maybe the most clear example of a *performance-centered* paradigm. As described in [9], the goal of Industry 4.0 was constructed under economic and *technology-driven* interests. This goal is analogous to previous revolutions: "to increase productivity and achieve mass production using innovative technology" [10]. To reach this goal, previous revolutions used machines powered by steam (Industry 1.0), electricity (Industry 2.0), as well as electronics and Information Technology (IT) artifacts, such as Programmable Logic Controllers (PLC) (Industry 3.0) [9,10]. In the case of Industry 4.0, digital transformation of manufacturing and production processes is empowered by emergent technologies such as Virtual Reality (VR), autonomous robots, the Internet of Things (IoT), Big Data, and Cloud Computing [9,11]. While these technological transitions have been a valuable source of economic growth for decades, the continuous increase of social and planetary problems related to the existing industrial activities are starting to push for a change of paradigms [8]. For example, and contrary to the optimistic predictions often done in academia, reports, such as [12,13], argue that automation technology has played a major role in wage inequality over the last decades. Due to this, there exists a low inclination to accept and trust automation technology [14,15]. This inclination is mostly present among low-skilled and middle-skilled workers (i.e., those carrying out routine-based tasks), who can see machines as possible threats to their jobs, identity, uniqueness, and safety [15]. Consequently, some social experts and futurists argue that "robots are taking the human jobs and are moving society towards more inequality" [16,17]. Another consequence of the increasing industrial activity is the rise in pollution-related chronic diseases, as well as contamination of air, water, and soil, and the over-exploitation of natural resources [18,19].

Industry 5.0 [8] is a very recent concept adopted by the European Commission whose vision is to reach *human-centered*, *sustainable* and *resilient* industries. This approach contrasts with the *machine-centered* or full-automation principle of past industrial revolutions, where the main motivation is to reach mass production, therefore underestimating planetary and human costs. The *human-centered* principle aims to respect the role, talents, and rights of humans by putting their general well-being at the same level of importance as the optimization of industrial processes. This principle proposes the introduction of technologies and tools ables to empower and promote the talents and diversity of industrial workers. Systems developed with these technologies must also safeguard fundamental human rights (e.g., autonomy, dignity, and privacy), create inclusive work environments, prioritize human mental and physical health as well as enhance job efficiency, safety, and satisfaction [8,9]. The *sustainable* principle focuses on the creation of production processes able to respect the planetary boundaries through the re-use and recycling of natural resources, as well as the reduction of industrial waste [9]. Finally, the *resilient* principle focuses on the creation of more agile, flexible, and adaptable industries [8]. Society 5.0 is another example of *human-centered* paradigm promoted in Japan. While Industry 5.0 focuses on the manufacturing sector, Society 5.0 considers a larger variety of scenarios. For this, Society 5.0 promotes the integration of cyberspaces (i.e., the virtual world) with physical spaces (i.e., the real world) as a key solution to enable both economic advancement and solve social issues [20]. Table 1 summarizes the differences between Industry 4.0 and Industry 5.0 according to [8,10,9,21]. Unlike Industry 4.0 predecessors, which are *technology-driven*, Industry 5.0 is identified as a *value-driven* paradigm that "requires the industry to re-think its position and role in society" [9]. Nahavandi [21] provides a more energetic distinction and states that the biggest problem of Industry 4.0 is that "its sole focus is to improve the efficiency of the process, and it thereby inadvertently ignores the human cost resulting from the optimization of processes." Maddikunta et al. in [22] describe that while the main priority of Industry 4.0 is process automation, which intrinsically produces a reduction of human intervention in the manufacturing processes, Industry 5.0 can bring back the human force to factories and promote more skilled jobs compared to Industry 4.0. This article focuses on one of the most relevant technologies for Industry 5.0, Human-Robot Interaction. Specifically in those scenarios where humans and robots share the same working space to reach a common objective in a synchronized, cooperative or collaborative way. In these Human-Robot Collaboration (HRC) scenarios, the repetitive, unsafe, physically demanding tasks are often assigned to robots, while humans will be in charge of critical thinking and customization [21,22].

The few efforts focused on identifying and classifying metrics for HRI have been developed based on their author's experience and from the performance-centered point of view. However, due to the lacking efforts and interests in the human-centered perspective by most of the robotics community (excepting social robotics), previous works generally presented little attention to human factors. Furthermore, initial efforts proposing taxonomies of metrics for HRI focused on teleoperation scenarios rather than environments where humans share the same working environment working in synchronized, cooperative or collaborative ways. In this article, we address the challenges of identifying and classifying factors, measures, and metrics enabling the evaluation of smart environments where humans and robots work together to reach a common objective. For this, we performed a systematic review of relevant and novel research articles using and proposing measures, evaluation methods, and metrics for HRI with particular attention to industrial and collaborative scenarios. We summarized results in different tables and two definition quality models, a taxonomy of performance-centered metrics, and a Venn diagram showing relationships between human-centered disciplines and common human factors. As we mentioned before, prioritizing human well-being (i.e., the human-centered perspective) is one of the three main elements of Industry 5.0. Therefore, we identify emergent approaches and challenges from a

Table 1
Differences between the general vision presented in Industry 4.0 and the keystone aspects required to reach a Society/Industry 5.0.

Feature	Industry 4.0	Industry 5.0 and Society 5.0
Motto	Smart Manufacturing	Human-Robot co-working and Bioeconomy
Motivation	Reach mass-production and increase economic benefits	Smart society, Social fairness, Resilient industries, Human well-being and Sustainability
Role of humans	Humans are substituted by machines	Bring back the human force to factories by respecting the talents, rights, needs, and identity of humans
Core technologies	Internet of Things, Cloud Computing, Big Data, Robotics and Artificial Intelligence	Human-Robot Collaboration, Renewable Resources, Bionics, Bio-inspired technologies and Smart Materials
Typical scenario in robotics	Interaction between humans and machines/robots is limited to offline programming and monitoring	Highly adaptable and personalized scenarios, where humans and robots can cooperate or collaborate to reach common goals

human-centered standpoint, which can become relevant in the following years to develop Industry 5.0 scenarios. The objective of this article is not to propose a qualitative assessment model or framework for HRI. Instead, it provides an overview of those quality factors used by the robotics community to assets robotics systems in industrial settings. Therefore, our results can be helpful for future efforts in robotics towards more complex quality models, such as assessment and predictive models [3,4] and evaluation frameworks or toolbox in HRI.

1.1. Paper organization

This paper is structured as follows. Section 2 presents the theoretical background and related works. Section 3 clarifies the contributions of this article. Section 4 presents the methodology followed to perform the systematic search of relevant research articles in the area of industrial and collaborative robotics. Section 5 and 6 report results of the study. On the one hand, section 5 presents a taxonomy of objective and quantitative measures and metrics oriented to measure different performance aspects in an HRI system. On the other hand, section 6 presents an overview of human-centered disciplines, such as usability, user experience, and ergonomics, introduce common human factors used to evaluate HRI systems in manufacturing environments, and present a Venn diagram showing the relationships between human-centered disciplines and human factors for HRI. Section 7 presents emergent approaches, challenges, and research gaps. Conclusions follow.

2. Related work

The literature reports few attempts to put quality factors, concepts, and metrics together for interactive robotics systems in a comprehensive or hierarchical way. Moreover, there is no standard of a widely adopted metrics toolkit or a quality model enabling researchers and practitioners to benchmark HRI systems. In this context, one of the first attempts was made by Olsen & Goodrich [23]. They present a list of six quality measures and metrics (task effectiveness, neglect tolerance, robot attention demand, free time, fan-out, and interaction effort). Olsen & Goodrich highlight that these factors were selected to evaluate the effectiveness of robotics systems controlled by humans (such as remote control of mobile robots). Subsequently, Goodrich et al. extended this list in [24]. Measures and metrics presented in [24] are divided into two groups: *task-oriented metrics* and *common metrics*. On the one hand, the *task-oriented metrics* group defines a set of tasks traditionally performed by mobile robots. These tasks include navigation (i.e., the action of moving robots from a point A to B), perception (i.e., enable robots to understand the environment), management (i.e., enable the coordination of humans and robots), manipulation (i.e., enable robots to interact with the environment) and social skills (i.e., enable robots to exhibit social competencies). On the other hand, the *common metrics* group evaluates the overall performance of HRI systems. This group of metrics has three sub-groups: i) *system performance* or *team performance*, which describes how well the robots and humans perform in a team composition; ii) *robot performance*, which describes the degree of awareness that robots have about humans and the environment, as well as their autonomy; and iii) *operator performance*, which lists a set of factors that can impact how well humans perform when using HRI systems. Metrics proposed in [24] were source of inspiration for posterior works, such as [7,25,26]. For example, [25] extended the classification by presenting a tree-structured taxonomy of HRI metrics and measures. Their taxonomy displays a set of 42 elements classified into three main types: human-related (composed of 7 elements), robots-related (composed of 6 elements), and system-related (composed of 28 elements). In 2018, a review of common metrics for Human-Machine Teams (HMT) was presented in [7]. The focus of this review included a broad type of machines, such as unmanned aerial vehicles, autonomous cars, robotic medical assistants, digital assistants, and cloud assistants, among others. The main outcome of [7] was the proposal of 10 common metrics for specific application areas (search and identification, navigation, ordnance disposal, geology, surveillance, and healthcare). According to their authors, a key limitation of these metrics is that many of the proposed metrics can be machine- or application-dependent and can have multiple interpretations for different types of applications, machines, or contexts. Most recently, Marvel et al. presented in [26] an overview of challenges in the design of human-machine-interfaces (HMI) and HRI in collaborative manufacturing applications. They collected in a hierarchical way objective a subjective metrics and measures proposed in some previous works, such as [23,24], and from ISO/IEC 25010 definition models for quality assessment [27].

For years, the vision of the robotic community was analogous to Industry 4.0 and previous revolutions, considering the improvement of performance as the most relevant HRI aspect [28]. However, new value-driven paradigms such as Industry 5.0 and Society 5.0 require that the robotics community puts more and more attention to more holistic approaches that prioritize the different elements enabling the general well-being of humans. This work tries to present a more holistic overview of HRI, introducing and classifying human-centered factors, which can be valuable to develop Industry 5.0 applications. However, it is relevant to highlight that both Industry 4.0 and Industry 5.0 can coexist. Therefore, we also recognize the efforts and arguments done in previous works by summarizing and classifying those metrics and measures used in the robotics literature to assess HRI applications' performance.

3. Objectives and contributions

In order to contribute to the HRI community in the creation of usable and comprehensive quality models in HRI, the goals of this article are: (i) to identify relevant measures, metrics and quality aspects enabling the evaluation and analysis of HRI systems in a systematic way; (ii) to propose a performance-oriented taxonomy that considers objective and qualitative aspects for HRI; (iii) to provide an overview of human-centered disciplines and identify human factors considered by the robotics community to assets robotics systems on industrial settings; and (iv) to discover emergent approaches, open issues, research gaps and challenges in the context of manufacturing. Therefore, the first contribution of this article is:

Through a systematic study, we identify common and relevant metrics for HRI, focusing on robotics systems operating in co-existence, cooperation and collaboration scenarios with humans.

This article presents three main differences/novelties in comparison with previous works described in section 2, as follows:

- Damacharla et al. [7] argue that most of the works identifying common metrics in HMI propose taxonomies or models that are mainly based on the experience of their authors. In this context, systematic reviews are relevant alternatives to reduce research bias.
- Few works, such as [25] and [26] have identified metrics by reviewing research articles published in the Institute of Electrical and Electronics Engineers (IEEE) HRI conference. Unlike these works, we performed a systematic search of scientific papers in 4 different databases until March 2022, including Journal and conferences papers.
- Except for [7,26], none of the related articles mentioned in section 2 reported a search protocol. Unlike [7], this work focus on Human-Robot Interaction in manufacturing environments and excludes other types of machines or interfaces (e.g., software interfaces, autonomous cars, and digital assistants).

The second contribution of this article is defined as:

Through the analysis of the results obtained from the systematic search, we present an overview of relevant human-centered disciplines, as well as human factors and evaluation tools used by the robotics community to assess HRI systems in industrial settings.

As described in section 2, the main focus of related works was to identify those metrics or factors that objectively evaluate performance-related aspects. This is due to the conventional vision often observed in Industry 4.0 (and previous paradigms), where the primary motivation is to reach mass production. In section 6, we introduce and briefly explain the relationship between relevant human-centered disciplines. We also identify common human factors recently used in robotics projects using industrial or collaborative robots. In this context, we identified that a common source of misunderstanding in related works, reviewed articles, and literature of different areas is the different interpretations of some multidimensional, overlapped, or complex concepts. Examples are the difference between a) *usability* and *user experience*, b) *performance* and *efficiency*, and c) *measure* and *metric*. This issue can produce misconceptions or confusion for new researchers. Therefore, in this article, we collect and present the different meanings used in the literature and relevant models used to differentiate them.

4. Methodology

Systematic literature review studies are objective and strict research processes designed to give a broad overview of current trends, gaps, and challenges in a specific discipline [29]. They can also be used to structure a research area, synthesize evidence, and help in the position of research directions and activities [29–31]. A relevant aspect of systematic literature studies is the so-called *review protocol*. According to [32], the review protocol "specifies the research question being addressed and the methods that will be used to perform the review". Therefore, the review protocol is composed of all the stages required to perform a systematic review plus additional planning information (e.g., project timetable) [29]. The review protocol we use to perform this literature review is based in [30,32], which proposed guidelines for performing systematic literature review studies, nowadays used in many software engineering and related areas, such as computer and robot programming [31,33], internet of the things [34], smart manufacturing [35] and robotics [36]. These guidelines propose a suggested review protocol composed of the next steps: (S1) identification of the need for systematic review (S2) definition of research questions, (S3) definition of the search strategy, (S4) study selection of criteria and procedures, (S5) study quality assessment, (S6) data extraction and synthesis, and (S7) results' report. We present each step of the research protocol as follows: step S1 in section 4.1, step S2 in section 4.2, step S3 in section 4.3, and steps S4, S5 and S6 in section 4.4. The analysis of results is reported in sections 5 and 6.

4.1. Identification of the need for systematic review

As described in section 2, previous works presented HRI taxonomies biased by the experience of researchers as well as the conventional needs of previous technological-driven paradigms. Moreover, many of them lack a detailed review protocol and documentation of the search process. Systematic reviews are suitable alternatives to reduce the risk of research bias as well as to provide more comprehensive studies [37].

4.2. Research questions

The research questions (RQs) guiding this article are:

1. **RQ1:** *What metrics and measures have been used or proposed in the literature to evaluate performance-related aspects in HRI and industrial environments and how they are applied?*
2. **RQ2:** *Which human-centered factors are commonly evaluated in industrial environments?*
3. **RQ3:** *Which are the emergent approaches and possible research directions toward the development of Industry 5.0 applications?*

We used the results of this systematic search to build the taxonomies, diagrams, and tables presented in sections 5 and 6. In this search, we put special attention to those approaches and research articles in industrial and collaborative robotics. RQ1 aims to identify relevant and well-defined qualitative and objective measures and metrics for assessing performance-related aspects in HRI. To classify and understand how they are applied, we propose the dimensions defined in Table 2. RQ2 aims to identify frequently addressed human-centered quality aspects in industrial environments. Therefore, we registered the number of articles evaluating each identified quality factor to answer this question. We introduce those human-centered factors classified as commonly evaluated in selected primary studies in section 6.3. Finally, RQ3 aims to determine emergent aspects or methods in HRI. We present these challenges from the point of view of the human-centered principles of Industry 5.0 and Society 5.0.

4.3. Definition of search strategy

The search strategy "aims to detect as much of the relevant literature as possible" [32]. We used the PICO (Population, Intervention, Comparison, and Outcomes) method suggested in [30] to select the keywords for the systematic search. For this work, *population* may refer to the main entities of this study: "humans" and "robots." A related word to "robot" is "agent." In the context of this article and as suggested in [29], *intervention* can refer to the technology or procedure performed between humans and robots. In this case "interaction" and "collaboration." In this study, we do not perform a *comparison* with alternative interventions. Finally, expected outcomes are "metrics" for HRI. We consider "evaluation," "validation", and "measurements" as related concepts to "metrics" and "measures." After contrasting the keywords obtained from the PICO criteria with our general objective and our proposed research questions, we defined the search string as *(Metric OR Evaluation OR Measurement) AND (Collaboration OR Interaction) AND Robot AND Human* AND (Industrial OR Manufacturing). We refined this search string through different iterations, in which we discarded the keywords "validation," "measures", and "agent". To improve the filtering process (i.e., to avoid the introduction of too many unrelated search results) we added the words "industrial" and "manufacturing". We used the final string to search research articles in relevant databases for robotics, namely, IEEE Xplore, ACM Digital Library, Springer Link, and Science Direct. For this search, we considered articles published until March of 2022 and sorted them by relevance. Table 3 shows the results obtained from each database. Table 4 shows the advanced setting configuration used for each database.

4.4. Study selection, quality assessment and data extraction

The study selection process implies selecting and applying inclusion/exclusion criteria [32]. The selection of articles for their review was composed of three steps. In step 1 we excluded papers based on their abstract and title. In case of doubt, we proceed to read the whole paper. In this step we applied the following inclusion criteria:

1. The focus of the article is to present an HRI/HRC framework or system for industrial tasks and not in purely social or medical scenarios (e.g., assistive therapy, rehabilitation, surgical) and not other interactive machines such as smart speakers, autonomous vehicles.

Table 2
Dimensions used to obtain general information of measures and metrics.

Label	Dimension	Objective
RQ1-D1	Name	Identify each measure/metric
RQ1-D2	Category	Identify the main aspect to evaluate of each measure/metric
RQ1-D3	Target	Identify where each measure/metric is applied (human, robot or team)
RQ1-D4	Team composition	Identify the HRI configuration

Table 3
Number of studies per database and results after applying inclusion (step 1) and exclusion (step 2) criteria.

Database	Search result	Results of step 1	Results of step 2
IEEE Xplore	577	66	29
ACM Digital Library	493	57	16
Springer Link	1197	102	20
Science Direct	154	58	39

Table 4
Advanced setting used for each database.

Database	Search string	Other settings
IEEE Xplore	(Metric OR Evaluation OR Measurement) AND (Collaboration OR Interaction) AND Robot AND Human AND (Manufacturing OR Industrial)	
ACM Digital Library	[[Title: collaboration] OR [Title: interaction]] AND [[Full Text: metric] OR [Full Text: evaluation] OR [Full Text: measurement]] AND [Title: robot] AND [Title: human]	
Springer Link	“Human Robot" AND (Metric OR Measurement OR Evaluation OR Collaboration OR Industrial OR Collaborative OR Interaction)	Content Type: Article
Science Direct	Title and abstract: (Metric OR Evaluation) AND (Collaboration OR Interaction) AND Robot AND Humans	Article Type: Research articles, Publication title: Robotics and Autonomous Systems, Procedia CIRP, Procedia Manufacturing, Robotics and Computer-Integrated Manufacturing, International Journal of Human-Computer Studies, Journal of Manufacturing Systems

2. The article gathers or proposes tools or metrics for evaluating human-centered or performance-related aspects of HRI/HRC applications.

For each database, the search process finished if after 50 consecutive articles none of them met some inclusion criteria. In step 2, results from step 1 are used to apply the following exclusion criteria. In this step, we process to read the full papers.

1. The article does not propose an HRI/HRC task and only evaluates the technological suitability of some specific hardware (e.g., sensor, actuator) or algorithm (e.g., perception, decision-making, and control).
2. The article does not present or use measures, evaluation methods, or metrics for assessing its framework or application.
3. The article is not accessible in full-text, has less than 2 pages, is not written in English, or is a duplicate or extension of other previous studies of the same authors (i.e., presenting the same or similar results or frameworks in different conferences or Journals).

In step 2, we conducted a quality assessment of the 104 resulting primary studies in step 2. The next questions were used to assess the quality of the identified primary studies:

- Are the measurements, metrics, evaluation methods and methodology clearly stated?
- Is the article peer-reviewed?

Figure 1 shows the number of articles processed in each of the steps mentioned before. The search and data extraction processes were performed by authors 1–5. All the authors of this article reviewed the results. Papers where differences in the grasped or interpreted data occurred were discussed to have a consensus between the authors on this article.

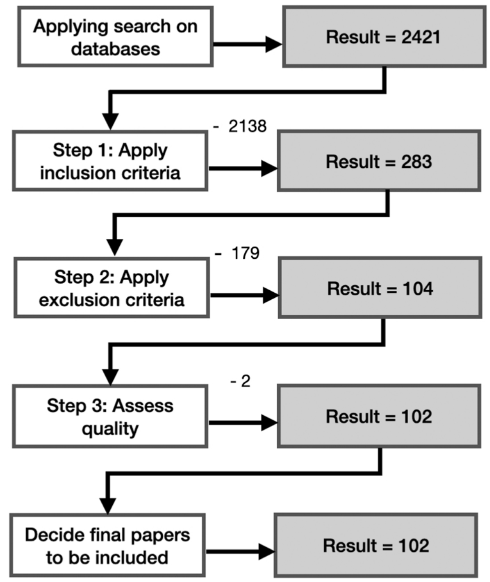

Fig. 1. Number of included articles during the study selection process.

4.5. *Limitations of the study and validity evaluation*

[29,38] describes the most common factors that can limit the validity of a systematic review. Factors that can be applied to this article are *theoretical validity* and *interpretive validity*. The *theoretical validity* “is determined by the ability that researchers have to grasp the intended data" [31]. As explained in [29,39] two systematic searches of the same topic can end up with different sets of articles. Therefore, some studies might have been missed. There is also a potential threat during data extraction due to researcher bias. However, this step is difficult to eliminate completely, as it involves human judgment [29]. The *interpretive validity* “is achieved when the conclusions drawn are reasonable given the data" [29]. Threats when interpreting data can be present due to researcher bias. To reduce *interpretive validity* and *theoretical validity* threats, researchers experienced in different areas of robotics, such as HRC/HRI, Industrial Robotics, Social Robotics, Software Architectures for Robotics, Human-Centered Design, Safety in Human-Robot Collaboration, Motion Planning and Artificial Intelligence, were involved in the validation of extracted data and conclusions.

5. Performance-centered aspects in human-robot Interaction for manufacturing purposes

This section reports and classifies performance-centered aspects used in the robotics community to evaluate HRI systems with manufacturing/industrial purposes. Reported factors, measures, and metrics were extracted from the articles selected for their review (steps S1 to S6 of the review protocol). Therefore, this section is part of step S7 of the review protocol described in section 4. Classification of human-centered aspects is reported in section 6.

As described in section 2, most taxonomies and classifications of metrics and measures for HRI put process optimization at the center. In this context, Damacharla et al. [7] describes a methodology to build these type taxonomies, which is composed of two basic steps. The first step identifies the agents involved in the HRI task. These agents compose the taxonomy’s main categories (or first level). In [24,25,7] these agents are selected as *human* (or operator), *robot* (or machine), and *team* (or system). Marvel et al. [26] additionally include the category of *process* that includes economic and process performance indicators. The second step is to identify high-level attributes that cluster a set of related metrics. These attributes compose the sub-categories (or second level) in the taxonomy. It is relevant to highlight that the final taxonomy proposed by Damacharla et al. [7] does not have sub-categories. Instead, their taxonomy only includes ten common metrics distributed in the three main categories, four metrics in the category of *human*, three metrics in the category of *machine* and three metrics in the category of *team*. In the case of [25], the *Human* and *Robot* categories do not present sub-categories. Table 5 shows the main categories and sub-categories defined in previous works. The final level of the taxonomy (i.e., the leaf nodes in a tree structure) displays the corresponding metric for each category or sub-category. Due to the different structures of these taxonomies, Table 5 only shows the number of metrics composing each main category.

In this article, the steps used to build this taxonomy are: 1) present a clear vocabulary for avoiding misunderstandings presented in the literature and many previous works between complex and overlapped concepts; 2) identify the main attributes composing the definition of performance used in the literature; 3) identify the different types of measures and metrics used to evaluate performance attributes from the results of the systematic review; and 4) identify in which agent and scenarios these performance measures and metrics are applied.

Figure 2 shows the taxonomy built by the proposed methodology. The category classification are shown as marks colored according to the six performance-oriented categories described in the following sections. The adjacent bars are colored according to the corresponding team composition level.

5.1. Definition of performance and main attributes

Performance is a multi-faceted concept which, according to the Merriam-Webster dictionary, and in the context of system implementation, can be defined as “the fulfillment of a claim, promise, or request." In the organizational and workplace context, there exist a huge degree of slippage and confusion between different terms related to performance, such as *productivity*, *effectiveness*, *efficiency*, and *profitability* [40]. These concepts are often vaguely defined and poorly understood in the literature of several disciplines [41]. Moreover, due to the subtle differences and mutual dependencies between these terms, they are in many cases used interchangeably [40,42]. As described in Wagner et al. [43], this issue has been a topic of discussion for more than four decades. They also highlighted the importance of having an established, clearly defined terminology that can serve as a basis for further discussions. Literature also provides comprehensive frameworks that help in the understating of these concepts. Figure 3 shows the main elements used to differentiate performance-related terms in different areas. Moreover, there exists a general agreement that *performance* is an umbrella term that includes almost any objective of competition and manufacturing excellence [41].

Table 5
Comparison between main categories and attributes proposed for taxonomies of performance-oriented metric in HRI.

Authors	Category	Sub-category/Attributes	# of metrics
Steinfeld et al. [24]	System	Quantitative performance, Subjective rating, Utility of mixed initiative	7
	Operator	Accuracy of mental models, Workload, Situation awareness	5
	Robot	Self-awareness, Human awareness, Autonomy	5
Murphy et al. [25]	System	Productivity, Efficiency, Reliability, Safety, Coactivity	28
	Human		7
	Robot		6
Damacharla et al.[7]	Team		3
	Human		4
	Machine		3
Marvel et al. [26]	Team	Quantitative performance, Utility of mixed initiative, Qualitative performance, Team composition	11
	Operator	Situation awareness, Workload, Qualitative operator performance	8
	Robot	Self awareness, Human awareness, Features, Safety, Qualitative Robot performance	11
	Process	Return on investment (ROI), Equipment effectiveness (OEE), Interface, Timing, Interface, Diagnostics and feedback	24

5.2. Definition of measures, metrics and indicators

Another common source of misunderstanding that is widespread in the literature of different knowledge areas is the concepts of *measures*, *metrics* and *indicators* [44–46]. ISO/IEC/IEEE 24765 [47] defines a *measure* as “a variable to which value is assigned as the result of measurement" and a *metric* as “a combination of two or more measures or attributes." However, some authors provide opposite definitions [46]. Finally, an *indicator* according to ISO/IEC/IEEE 24765 is a “measure that provides an estimate or evaluation of specified attributes derived from a model with respect to defined information needs" [47]. ISO/IEC/IEEE 24765 also defines a *direct metric* as a “metric that does not depend upon a measure of any other attribute." Examples of direct metrics are the duration of a process (elapsed time) and the number of errors or defects. ISO/IEC/IEEE 24765 also defines a *indirect metric* as a “ metric that is derived from one or more other metrics." Finally, [46] provides an object-oriented approach of consistent terminology between *measures* (simple numerical values with little or no context), *metrics* (collection of measures with context), and *indicators* (comparison of metric to baseline). Most recent review in [48,49] defines *measure* as a “quantitative whole number, either in monetary (financial) form, dimension form (e.g. square meter) or unit form (e.g. production output)," *metric* as a “quantitative standard in fraction form," and indicators as “quantitative or qualitative form for measuring things more generally." It is possible to see a general agreement between standard definitions in [47] and recent reviews of [48,46,46]. From the information theory point of view, “measures" can be classified as data (i.e., collection of facts with no contextual relationships and little or no meaning) and information (i.e., useful signals or knowledge that result from the understood or interpretation of facts in some specific context) [50]. In contexts such as workplace training and business, metrics defined under the goal-setting theory [51], as relational measures that provide valuable information on the progress towards desired goals. Examples of goals are motivating employees, facilitating communication with stakeholders, preventing problems, and improving performance [52]. In the next sections, we adopt these terminologies by considering measures as simple and direct values, and metrics as composite values composed of one or more measures or other metrics generally resulting from some mathematical function (often a fraction).

5.3. Performance measures for Human-Robot Interaction

They are rudimentary, accurate, or simple variables obtained from an aggregate of facts (e.g., total cost and the number of errors) or direct physical measurements in either the robots or the humans (e.g., time for completing some action and joint acceleration). They are used to clarify the current or final state of the human, robot, process, or interaction. From the results of the systematic search as well as the performance measurement models reviewed in [56] we identified the following groups of metrics in this category:

Fig. 2. Performance-oriented categorization for the metrics obtained in the systematic search performed in this article.

- *Time behavior measures* indicates the response and processing times that a human, robot, or a combination of humans and robots requires to perform its functions, a sub-task, or a complete task. Examples of these metrics are human idle time, algorithm processing time, collaboration time, and task completion time.
- *Process measures* are an aggregation of facts generated from the start to the end of a task or sub-task as well as cost-related, workspace design, safety, or product quality-related elements. Examples of these metrics are the number of errors and the number of assembles reached.
- *Physiological measures* are values obtained from body measures that help to understand the current state of the human (e.g., acceleration of human joints and heart rate)
- *Human-Robot physical measures* are values obtained from sensors that indicate the current state of the interaction (e.g., the distance between the human and the robot)

5.4. Performance metrics for human robot Interaction

We define performance metrics for HRI as a combination of direct measures using a mathematical expression (usually a division) with other measures or metrics to express a rate, an average, or an input/output relationship. In this work, we consider *efficiency* (internal performance) and *effectiveness* (external performance) as the main attributes used to evaluate task *performance*.

Efficiency metrics. According to ISO 9241, efficiency is the "relation between the resources (inputs) used, and the results (outputs) achieved."

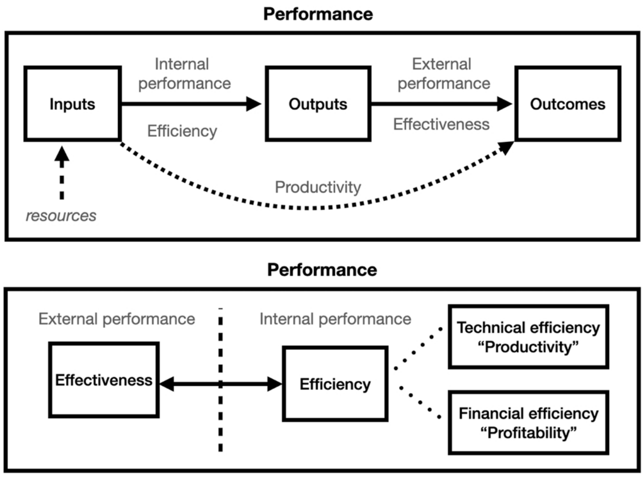

Fig. 3. Relationship between performance, efficiency, profitability, effectiveness and productivity according to [53–55].

In this article, metrics evaluating efficiency are defined as input/output relationships. The main idea behind efficiency metrics is to evaluate if HRI systems are "doing things right." Therefore, these metrics evaluate the progress toward completing defined objectives. Consequently, the typical question they try to answer is how well resources (time, costs, materials) are used.

Effectiveness metrics express the ratio between the actual or obtained results and the programmed, wanted, or intended results to achieve. The main idea behind effectiveness metrics is to evaluate if HRI systems are "doing the right things." Therefore, these metrics evaluate the accuracy and completeness with which HRI systems achieve specified goals. Consequently, the typical question they try to answer is which is the success or failure rate?

6. Human-centered aspects in human-robot Interaction for manufacturing purposes

Year by year, holistic and multidisciplinary paradigms, such as human-centered design, have gained more importance in different disciplines. This contrasts with the traditional performance-oriented vision generally presented in the initial stages of many emergent technologies. Research teams with technical backgrounds predominantly conduct the design and development cycles in these initial stages. The primary motivation that often guides these researchers is to build interactive systems able to meet a set of functional requirements as well as to prove the superiority of the proposed architectures and algorithms against previous solutions [57]. However, many mature technologies nowadays accepted and adopted by the general public have historically switched their design approaches from *performance-oriented* to a more holistic point of view [58,59]. Smartphones and web interfaces are examples of mature technologies that people widely adopt these days. In these technologies, non-functional aspects, such as emotional responses, comfort, social value, and aesthetics, play essential roles not only to reach commercial success but also to be appreciated-by-users [60]. Therefore, the main objective of this section is to present an overview of human-centered disciplines and human factors used by the robotics community to assess robotics systems in industrial settings. The organization of this section is as follows: We define human-centered quality for HRI and identify the high-level quality attributes in section 6.1. Then, in 6.2 we identify if there exists an overlap or disagreement in the scientific community between the elements composing these high-level quality attributes and summarize the different points of view. Then, in section 6.3 we introduce a set of common human factors used in the works finally selected for its review. Finally, in section 6.4, we propose a classification of these common human factors and present a Venn diagram that shows the limits between human-centered areas and identified quality attributes for HRI.

6.1. Definition of human-centered quality for HRI

We extended the definition of human-centered quality detailed in the ISO 9241–11:2018 (ergonomics of human-system interaction) [61] to HRI systems. This international standard provides a set of definitions, requirements, and recommendations designing human-centered products, systems, and services. Therefore, in this work we consider that an HRI system presents human-centered quality when is able to met requirements of *usability*, *accessibility*, *user experience*, and *avoidance of harm from use*. These requirements can be considered top-level quality concepts.

6.2. Relationships between usability, user experience, ergonomics and hedonomics

Quality factors given in the ISO 9241–11:2018 present a significant overlap and different conceptualizations. The two concepts that present more overlap are *usability* and *user experience* [73]. On the one hand, usability is in some cases related to "ease-of-use". However, its concept is more comprehensive. According to ISO 9241–11:2018, usability is "the extent to which a system, product or service can be used by specified users to achieve specified goals with *effectiveness*, *efficiency*, and *satisfaction* in a specified context of use". In this definition, two different elements can be identified: those related to objective and performance-oriented factors (*effectiveness* and *efficiency*) and those related to subjective aspects (*satisfaction*) [73]. Despite this standardized definition, there is no consensus in the HCI and HRI communities about the definition of usability [74]. Therefore, several authors propose different attributes composing the definition of usability. Examples of review articles summarizing the different definitions of usability are [74,75]. Table 6 shows some of the common attributes of usability presented in the literature. On the other hand, ISO 9241–11:2018 defines user experience as "the person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service." This standard also indicates that "user experience includes all the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors and accomplishments that occur before, during and after use." As described in [68] user experience is considered by some authors as a subset of the satisfaction component of usability. In contrast, others can consider usability a subset of the user experience. Moreover, a third perspective considers that usability emphasizes objective measures and user experience emphasizes subjective measures. Table 7 shows the different quality attributes of user experience presented in the HCI literature. To reduce the confusion presented between the concepts of *usability* and *user experience*, [73] proposed a holistic model designed to be consistent with the ISO standards' definitions. This model integrates the holistic approach of *user experience*

Table 6
Usability attributes in the Human-Computer interaction literature adapted from [37,66].

Usability models	Usability attributes
ISO 9241–11:2018 [61]	Effectiveness, efficiency, satisfaction
ISO/IEC 9126–1:2001 [62]	Understandability, learnability, operability, attractiveness
ISO/IEC 25010 [27]	Accessibility, flexibility, reliability, maintainability, compatibility
Nielsen [63]	Learnability, efficiency, memorability, errors, satisfaction
Rios et al. [64]	Knowability, operability, efficiency, robustness, safety, satisfaction
Shackel et al. [65]	Effectiveness, learnability, flexibility, subjectively pleasing
Gupta et al. [66]	Efficiency, effectiveness, satisfaction, memorability, security, universality, productivity

Table 7
Relevant user experience (UX) attributes in the Human-Computer Interaction literature.

UX models	UX attributes
ISO 9241–11:2018 [61]	Emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors and accomplishments
UX honeycomb [67]	Usefulness, usability, desirability (i.e., emotional appreciation), findability, accessibility, credibility
Zarour et.al [68]	Hedonic (emotional, trustworthiness, aesthetics, fun, privacy, sensual), pragmatic (usability, functionality, usefulness)
Lachner et.al [69]	Look (aesthetics/design, interface, information value), feel (control, learnability, pleasure, satisfaction, ease of use), usability (efficiency, utility, effectiveness, functionality)

and the mixed formulation often presented in *usability* definitions, which considers both subjective and objective elements. Moreover, emotion-related elements, such as pleasure, acceptance, trust, and aesthetics are considered out of the scope of *usability*, which is an approach in many cases accepted by practitioners. We use the approach proposed in [73] as a starting point for the development of the HRI model presented in this article. Figure 4 shows a summary of the main interpretations and relationships between the concepts of *user experience* and *usability* in the HCI literature, as well as the main factors used to differentiate them (satisfaction, performance, affect, subjective measures, and objective measures).

Ergonomics (also denoted as human factors) is also a human-centered discipline which goals and tools in many cases overlap with those presented in *usability* and *user experience* design. The ISO 6385:2016 [77] defines *ergonomics* as a "scientific discipline concerned with the understanding of interactions among human and other elements of a system, and the profession that applies theory, principles, data, and methods to design in order to optimize human well-being and overall system performance." However, the most common conception of *ergonomics* refers to how companies design tasks, scenarios, and interfaces able to maximize the efficiency and working condition of their employees' work [78]. Most works in the literature identify two main areas of ergonomics: *physical* and *cognitive ergonomics*. These areas are explained in section 6.4.2. Relevant factors and domains described in the literature for *physical* and *cognitive ergonomics* are shown in table 8.

Hedonomics represents a conceptual companion of *ergonomics* focused on "the pleasant or enjoyable aspects of human-technology interaction" [76]. As explained in [78], the moral foundation or main core of *ergonomics* is focused into reduce pain, injuries, and suffering in the workplace. However, this discipline is often limited to show the importance of preventing negative events that "eventually do not happen" [78]. Conversely, *hedonomics* focus on more positive aspects of work interactions by "promoting the occurrence of satisfying interactions, which can be proved or observed" [78]. Areas related to *hedonomics* are *user experience*, *kansei engineering* [57] and *pleasurable design* [79]. These satisfaction and affective focused paradigms proposed in hedonomics disciplines contrast with the predominant safety and productivity-oriented focus of traditional research in *ergonomics* [78]. The Hancock's Hedonomic Pyramid proposed in [76] (shown in figure 5), which is based on the Maslow's psychological hierarchy of needs, clarify the limits of both *hedonomics* and *ergonomics*. This pyramid starts in the bottom by defining aspects that are able to meet collective and functional goals. Each higher level of the pyramid focuses more and more on individual and non-functional aspects. Moreover, *usability* factors are divided in those closer to the definition of *hedonomics* (mostly subjective) and those traditionally presented in *ergonomics* (mostly objective).

Table 8
Cognitive and physical ergonomics attributes and domains found in recent surveys on ergonomics applied on industrial environments.

Source	Ergonomics area	Domains and attributes
Kadir et.al [70]	Physical	Working postures, materials handling, repetitive movements, musculoskeletal disorders, workplace layout, safety and health
	Cognitive	Perception, memory, reasoning, motor response, mental workload, decision-making, skilled performance, human reliability, work stress, training
Neumann et. al.[71]	Physical	Safety, fatigue, posture, gesture, musculoskeletal disorder
	Cognitive	Learn, knowledge, training, capabilities, skills, experiences, education, teaching, talent, competencies, creativity, confusion, e-learning, forgetting, memory, reasoning

6.3. Common human-centered factors for Human-Robot Interaction

From the results of the systematic search,table 9 we identified human factors used by the robotics community to assess HRI systems in industrial settings. We classify those factors in table 10 and briefly present those factors that have received more attention from the robotics community.

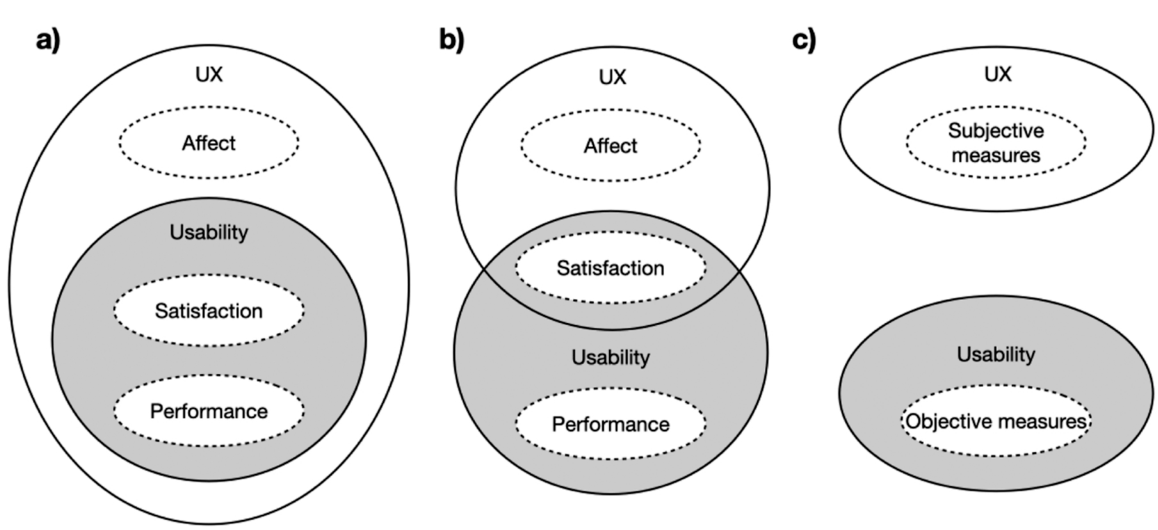

Fig. 4. Different interpretations and relationships of User Experience (UX) and usability found in the literature. Adapted from [73] and [68].

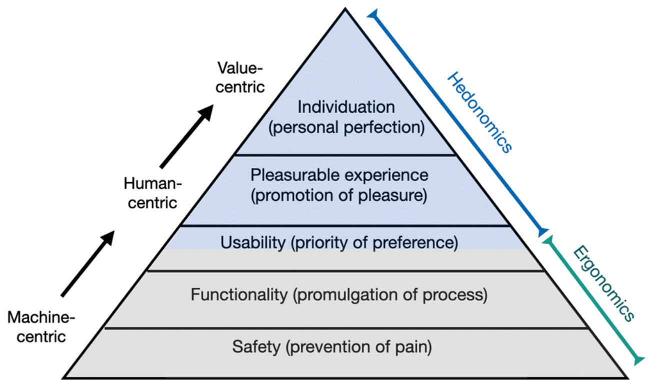

Fig. 5. This pyramid shows the limits between *hedonomics* (in blue) and *ergonomics* (in gray).
Hancock's Hedonomic Pyramid adapted from [76].

Table 9
Hedonomics.

Source	Domains and attributes
Zarour et.al[68]	Emotional, trustworthiness, aesthetics, fun, privacy, sensual
Diefenbach et.al [72]	Stimulation, fun, entertainment, affect, emotion, pleasure, enjoyment, happiness, identification, self-expression, psychological needs, end in itself, be-goals, beyond the instrumental, beauty, aesthetics, visual appeal, social value, social interaction, relatedness, imagination, fantasy, memories, long-term use, trust

Table 10
Most relevant human-centred quality attributes used in Human-Robot Interaction systems with industrial and collaborative purposes. Inside parentheses is indicated the number of articles using each quality factor for its analysis.

Type	Measurable dimension/quality aspect
Affect	Emotional responses (3)
Beliefs	Attitudes/Acceptance (11), Anxiety (3), and Trust (16), Perceived robot ability (1), Perceived robot intelligence (3), Social Presence (1), Human-likeness (3)
Cognitive ergonomics	Mental workload (13), Concentration/Attention (3), Mental models and Awareness (6)
Physical ergonomics	Physical workload (11), Safety (18), Physical fatigue (1), Physical comfort (2), Workplace design (1)

6.3.1. Safety

Safety is a critical quality aspect in *ergonomics*. As shown in figure 5, this aspect is located at the base of the functional requirements of any technological system. Results of the systematic review presented in this article indicate that *physical safety* is the most common quality aspect evaluated in the context of industrial and collaborative robotics. Table 11 shows the most relevant articles resulting of the systematic search that propose or use metrics for safety in the area of collaborative robotics. Some of these metrics are based on international standards for industrial robotics and HRC. Standards mentioned in these articles are: ISO 10218–2:2011 (safety requirements for industrial robots), ISO 13482 (personal care robots) [92] ISO/TS 15066:2016 (collaborative robots) [93], ISO 13855:2010 (positioning of safeguards with respect to the approach speeds of parts of the human body) [94], and NSI/RIA R15.06- 2012 (robot systems safety requirements). Most of these metrics can assist in the development of systems that reduce the possibility of presenting dangerous or fatal situations, such as the collision between a robot and a human co-worker. Others, such as the number of conflicts between human and robot and mean velocity of the end-effector, can be used to measure both *safety* and robot performance [87]. Other popular methods used in the industry to evaluate physical ergonomic risks at

Table 11
Summary and classification of safety metrics used and proposed in the reviewed articles.

Type	Methods/Metrics
Assessment of Robot-Human Collisions	Safety design metric based on power flux density [80], Head Injury Criteria (HCI)-based force related danger [81], Injury indices for head, neck and chest area [82]
Evaluation of dangerous situations	Safety and Ergonomic evaluation index (SEEI) [83], Safety index (safety as function of the distance between human and robot) [84], Hazard Rating Number[85], Danger field [86]
Safety and robot performance	Number of conflicts between human and robot [87], Average separation distance between human and robot [87], Protective separation distance between the tool and a human operator[88], Safety features for collaborative robots [26], Speed and separation monitoring (SSM) [26] Power and force limiting (PFL)[26]
Safety standards for Human-Robot Collaboration	Velocity of the end-effector [89,90], Maximum dynamic power[89,90], Maximum static force [89,90]

assembly lines are summarized in [91] and displayed in Table 12. Unlike most of the metrics presented in Table 11, which can be specific to HRC, methods displayed in Table 12 are more general. Therefore, they are applicable in environments where workers have some risk of presenting musculoskeletal disorders. As described in [91], the level of physical ergonomic risks will depend on the frequency, intensity, and duration of physical workload factors (e.g., repetitive movements and awkward postures) and environmental factors (e.g., temperature and noise).

6.3.2. Trust

Results from the systematic search performed in this article suggest that *trust* is the second most common human-centered quality aspect evaluated in the context of industrial and collaborative robotics. *Trust* is a broad and multidimensional concept which is highly-depended of the context [95]. Examples are trust in social media, interpersonal relationships, organizations and governments. In robotics, *trust* is mostly described from the technological point of view and under the concept of *trust in automation* [96]. However, there is not a consensus on a single definition of *trust* in the HRI community [97]. Additionally, *trust* towards robots can be defined from two perspectives: performance-oriented and human-centered. An example of a performance-oriented definition of trust is given by [95,98], where *trust* is defined as "the attitude that an agent will help to achieve an individual's goals in a situation characterized by uncertainty and vulnerability." In this perspective *trust* is identified as an important factor able to influence the performance under certain tasks and conditions. The main idea behind this approach is that "if people do not believe in the collaborative capabilities of a robot, they will tend to underutilize or not use it at all" [98], which consequently can produce a drop in the task performance. An example of a human-centered definition of *trust* is described as "the reliance by one agent that actions prejudicial to the well-being of that agent will not be undertaken by influential others" [97,99]. Another human-centered and

Table 12
Most common risk assessment methods according to [91].

Context/Objective	Methods/Metrics
Lifting task	National Institute for Occupational Safety and Health lifting equation (NIOSH-Eq)
Assessment of postures	Rapid Upper Limb Assessment (RULA), Rapid Entire Body Assessment (REBA)
Risk assessment of upper extremities	OCcupational Repetitive Action tool (OCRA) and the Job Strain Index (JSI)
Noisy workplaces	Daily Noise Dosage (DND)
General risk assessment tools	The Ergonomic Assessment Work Sheet (EAWS) and the energy expenditure method (EnerExp)

comprehensive definition of trust is described in “a belief, held by the trustor, that the trustee will act in a manner that mitigates the trustor’s risk in a situation in which the trustor has put its outcomes at risk" [100]. We observed that one of the most relevant trust-related research topics inside the robotics community is the identification of factors affecting *trust* towards robots and human-robot interaction. While these articles propose a set of different attributes affecting trust, many of them considers the bases set by [101], which establishes the three main attributes of *trust* as: *ability*, *integrity*, and *benevolence*. Articles dealing with this topic discovered in the performed literature review are [95,102]. Charalambous et al. [95] and Yagoda et al. [103] additionally present scales enabling the evaluation of trust in industrial HRC and HRI respectively. Relevant articles surveying factors affecting trust in HRI contexts are [97,104]. They classified factors affecting the development of trust in HRI in *performance-related* (e.g., proximity, apology for failure and feedback), *human-related* (e.g. personality, culture and experience with robots) and *task/environment-related* (e.g., workload, duration of interaction and physical presence of the robot in task site). In some of the articles reviewed, *trust* is also considered as one of the most relevant subjective factor composing the attribute of *fluency* [105,106], discussed in section 7.4. We also observed that *trust* is mostly evaluated using subjective methods such as questionnaires, which are often applied after humans have interacted or worked together with robots. Moreover, this evaluation is generally unidirectional (i.e., it measures the level of trust that human has towards the robot but not the other way around). A relevant exception is [107], which proposes a bidirectional computational model that evaluates human’s *trust* in robot and robot’s *trust* in human. Additionally, *trust* is measured in real-time during collaboration. Similar to other human-psychological terms, such as emotions, trust is a complex process that is challenging to simulate in robots, as well as there is not a consensus of the best way to do it. To calculate “human’s trust in the robot", [107] proposed a trust model based on time series that consider factors strongly correlated to trust in automation (the robot performance and the robot failures). Similarly, computation of “robot’s trust in the human" requires to know actual the human performance and robot failures. More details of this model are described in [107]. Then the assembly is performed based on the mutual trust model by influencing subtask re-allocation. Authors of [107] claim that bilateral *trust* models can help to increase the performance of industrial tasks, such as assembly, that those only considering one-way *trust* (from humans to robots).

6.3.3. Attitudes and acceptance

Robotics is an emergent technology able to produce both positive and negative impacts on society and individuals. There exists a consensus that HRC can only be successful if human workers and society are willing to use and adopt this novel technology [108]. In this context, ethical and social issues such as fears towards robots replacing human workers, disinformation and false expectations given by social media and science fiction movies, and even the individual resilience in the adoption of uncertain technologies can affect people’s thoughts and feelings towards using robots. Results from the literature review identify *attitudes* and *acceptance* as popular aspects used to understand the level of adoption or resistance towards the robots in factories. Additionally, we also observed that many researchers in the HRI community use these highly coupled concepts in an interchanged way. On the one hand, the Cambridge dictionary defines *attitudes* as “a feeling or opinion about something or someone, or a way of behaving that is caused by this". This concept is also defined in [109] as “a psychological tendency that is expressed by evaluating a particular entity with some degree of favour or disfavour". Similar to *trust*, the identification of factors able to influence the attitudes that certain groups have towards technological devices is an active research topic. However, according to [110,111] there exist an agreement in the psychological community that *attitudes* can be described as a summary of semantic dimensions, such as pleasant-unpleasant, harmful-beneficial, good-bad, and likeable-dislikeable. Results from the systematic review indicate that the most popular tool for measuring attitudes in industrial and collaborative contexts is the Negative Attitudes Towards Robotics Scale (NARS) [112]. Other methods used in the articles reviewed are the Computer Thoughts Survey, and General Attitudes Towards Computers Scale, which together with the Computer Anxiety Rating Scale constitute the methods defined by Rosen and Weil [113] for measuring *technofobia*. The recent survey proposed in [36] summarizes common methods and results from articles evaluating attitudes, anxiety, acceptance, and trust in the social robotics context. This article identifies three distinct components of attitude *affect*, *cognition* and *behavior/general*. Methods used to measure *affective attitudes* are the NARS-S1 (interaction with robots) and NARS-S3 (emotions in interaction with robots) subscales [112], the Godspeed Questionnaire [114] (particularly in the likability dimension) and self-report measured based in semantic differential scales, such as those proposed in Kansei Engineering [57]. For *cognitive attitudes*, [36] reports the use of the NARS-S2 subscale (beliefs about the social influence of robots) as well as sub-scales of the Almere Model of robot acceptance [115] and Unified Theory of Acceptance and Use of Technology [116]. Finally, *general attitudes* are identified as a mix of *affective* and *cognitive* measures. For this, [36] reveals the use of self-report and the Implicit Association Test [117] in social robotics. Additionally, we identified the Multi-dimensional Robot Attitude Scale [118] as an recent method focused on measuring attitudes towards robot in domestic scenarios and the Robot Perception Scale [119], which enables to measure general attitudes toward robots and attitudes toward human-robot similarity and attractiveness. On the other hand, *acceptance* is generally defined in terms of the intention to use or the actual use of robots [36]. Methods identified for measuring attitudes and acceptance are the Frankenstein Syndrome Questionnaire [120], the Technology Acceptance Model (TAM) [121], and their major upgrades TAM 2 [122] and TAM 3 [123]. However, the suitability of methods for evaluating attitudes in industrial and collaborative scenarios is still uncertain. An exception is the TAM reloaded [124], which main focus of its authors is the development of an acceptance model that enables the assessment of human-robot cooperation tasks in production systems.

6.3.4. Mental workload and attention

Workload is one of the most extensively studied factors in the domain of *ergonomics*. This quality aspect is strongly related to other human factors such as stress, fatigue, motivation, the difficulty of tasks performed, job satisfaction, and success in meeting requirements [125,126]. *Workload* can be defined as “the ratio of resources required to achieve tasks to the resources the human has available to dedicate to the task" [127,128]. The literature presents two main classifications of workload. One of the initial classifications of workload, proposed in [129], distinguishes between quantitative and qualitative workload. While quantitative workload affects biomechanical and stress factors, qualitative workload affects mental overload and overall physical well-being. However, the most common classification distinguishes between mental and physical workload. According to [130], *mental workload* or *cognitive workload* is “a composite brain state or set of states that mediate human performance of perceptual, cognitive, and motor tasks." Stanton et.al. [131] propose a definition of *mental workload* as “the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support, and past experience" [131,132]. As described in [125], the methods and metrics considered under *mental workload* have been proposed from numerous and task-specific research activities about the limitations and capacities of information processing systems in humans. These methods are classified in [132] as: task performance measure, subjective reports, and physiological metrics. Human performance can create a cause and effect relationship with *mental workload*. An example happens when there is a drop in the effectiveness and efficiency of the tasks, which can increase the human perception of workload. In order to avoid errors and accidents, one of the main objectives in *ergonomics* is to identify and reduce sub-optimal levels of *mental workload* (i.e., when an excessive load or low engagement in the task) [132]. A common activity in *cognitive ergonomics* is the registration of the operator's capability to perform high tasks priority at acceptable levels. In this context, peripheral detection tasks (PDT) emerge as a suitable tool to evaluate cognitive workload from a high-priority task. The main idea behind PDT is that "visual attention narrows as workload increases" [132]. The metric of with-me-ness was introduced in [133] to measure "how much the user is with the robot during a task." An example of systems able to measure the concentration or sustained attention in the area of HRC is presented in [134]. *Subjective reports* are the most popular way to measure *mental workload*. Traditional methods such as NASA Task Load indeX [135], the Subjective Workload Assessment Technique (SWAT) [200] and the simple and fast Rating Scale Mental Effort (RSME) [201] are known to be complicated and time-consuming as well as to present retrospective/recall bias (i.e., incorrect recall due memory effects) [132]. Results from the systematic review performed in this article show that the self-reporting method, particularly the NASA-TLX [135], is the most common approach used to measure *mental workload* in industrial settings. Finally, *physiological metrics* enable the objective evaluation of *workload* by collecting real-time data (e.g., heart, brain, and muscle activity) in many cases collected by wearable devices attached to the human body. However, these methods often require the use of intrusive devices, which can reduce the comfort of human subjects and workers. Examples of quantitative methods to measure mental workload based on brain activity are electroencephalography (EEG), event-related potentials (ERPs), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI) [130]. Other physiological measurements correlated with an increase in mental workload are Skin Conductance Activity (SCA) and breathing rate.

6.3.5. Physical workload

The overall workload can be decomposed into seven components: cognitive, gross motor, fine motor, tactile, visual, speech, and auditory [128,136]. According to [128], *physical workload* can be defined as the "amount of physical demands placed on a human when performing a task" and is composed of gross motor, fine motor, and tactile components. Chihara et al. define *physical workload* as "mechanical load acting on the musculoskeletal system of human" [137]. Works reporting the evaluation of physical workload use the NASA-TLX. Objective metrics able to measure *physical workload* have been classified in [128]. Examples of these metrics are Variance in Posture, Postural Load, Vector Magnitude, Heart Rate, Respiration Rate, Galvanic Skin Response, and Skin Temperature. Other subjective approaches include the Borg Rating of Perceived Exertion [138], the Nordic Body Discomfort questionnaire [139] and The McGill Pain Questionnaire (MPQ) [140].

6.3.6. Situation awareness and mental models

Initially identified during World War I, the concept of *situation awareness* started to gain technical and academic importance until the late 1980's in the aviation industry [141]. During the next years, research in *situation awareness* constituted a substantive portion in the area of *ergonomics* and applied in the design of advanced information displays and automated systems [142]. In particular, this area gained importance in those applications requiring the supervision, monitor, or control of automated systems where multiple and simultaneous tasks or goals compete for the attention of the operator [141,143]. Stanton et al. [131] present a colloquial definition of *situation awareness* as "the understanding and use of information about what's happening during dynamic tasks." However, the most referenced conception of *situation awareness* is modeled as an information processing framework [144,145, 141]. This conception is defined by Endsley [145] as "the perception of the elements of the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [144,145]. This definition suggests that *situation awareness* is mostly composed of three levels: 1) noticing or perception of the elements of the environment (denoted as Level 1 SA); 2) understanding or comprehension of the current situation (denoted as Level 2 SA), and 3) prediction or projection in the near future (denoted as Level 3 SA). According to [141], most of the theoretical approaches of *situation awareness* considers *mental models* (i.e., drawing on knowledge, experience and skills) as of its main elements. A mental model is defined in [144] as a "dynamic representation of an event or scenario that reflects the person's understanding of the situation and can promote accurate situation awareness." According to [144] mental models are "cognitive mechanisms that embody information about system form and function as well as how components of a particular system interact to produce various states and events." They can be used to: direct the comprehension of new information, make decisions under uncertainty, direct attention to relevant information, tell the agent or people how to combine and interpret the significance of disparate pieces of information as well as how to create suitable projections of what will happen in near future [142,146]. Therefore, *mental models* can be used to build and maintain *situation awareness*, especially in the levels of comprehension and projection [146]. Therefore, an incomplete or wrong mental model can result in poor comprehension and projection of the information. A particular case of a wrong mental model is *mode errors*, in which people mistakenly believe to be in one mode or state, but is in another [146]. Tabrez et al. [147] presented a recent review of mental models in Human-Robot Teaming. They identify three categories of mental modeling in human-robot teaming as first-order mental models, second-order mental models, and shared mental models, being shared mental models strongly correlated to team performance [147,148]. Metrics to quantitative evaluate mental convergence and similarity of shared mental models in Human-Robot Interaction are described in [149]. In robotics, tools and frameworks enable to increase *situation awareness* was initially applied for the teleoperation of robots in applications, such as search and rescue, agriculture, and surveillance. According to [150] *situation awareness* can be improved in this type of robotics system through the use of maps, the fusion of sensory information, the minimization of multiple windows, and by providing spatial information to the operator. While the concept of *situation awareness* is generally considered to be a process presented on the human side (comprehension of the robot's states and the working environment), the concepts of *self-awareness* and *human-awareness* identified in [24,26] are considered on the robot-side. According to [151], self-aware robots are able to "attend to their own internal states, thus providing a means of generating introspection and self-modification capabilities." Examples of these internal states are emotions, beliefs, desires, intentions, expectations, mobility and sensors limitations, task progress, faults, perceptions, and actions [24,151]. On the other, *human-awareness* is defined in [24] as "the degree to which a robot is aware of humans." *Context Awareness* [152] is another related concept used in HCI and robotics [153]. Nikolas et.al. [153] recently presented a framework that integrates context and situation awareness under the less known theory of Smith and Hancock of situation awareness [154].

6.4. Definition of a human-centered conceptual model for HRI

Taking as inspiration the works, concepts, and conceptual models proposed in the HCI literature, specifically [64,73,68], as well as the results of the systematic search proposed in this article (summarized in table 10), we propose a holistic conceptual model adapted for Human-Robot Interaction. Figure 6 shows the relationships between more relevant attributes found in the literature. This conceptual model shows existing relationship and limits between *usability*, *user experience*, *hedonomics* and *ergonomics* using the concepts explained in section 6.2.

6.4.1. Hedonomics quality factors

The creation of interactive experiences able to maintain optimal emotional levels are important for the reduction of stress levels, avoiding disastrous errors and increasing task performance [155]. Moreover,

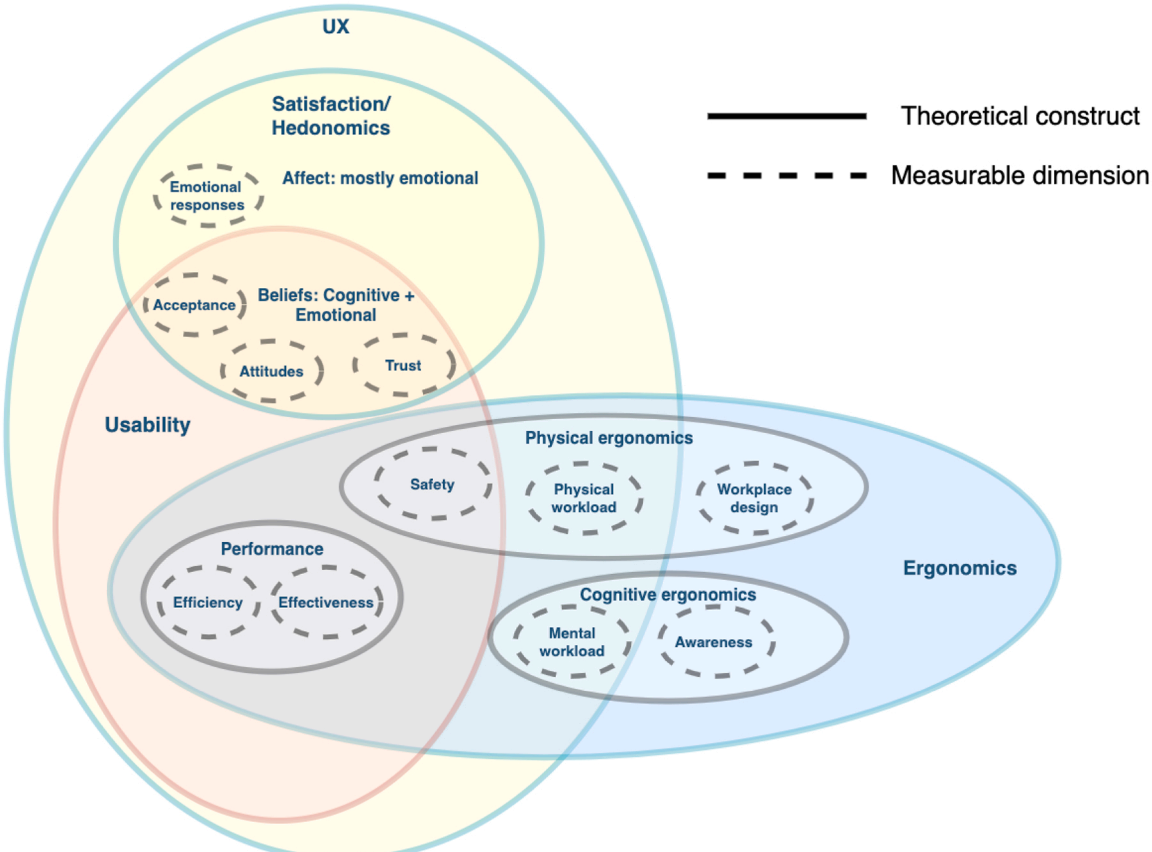

Fig. 6. Most representative quality factors analyzed in the HRI literature according to the results of the systematic review. This diagram is adapted to HRI from the Interaction Experience model proposed in [73].

hedonic-related factors such as happiness, emotional stability, and positive emotions are often considered as relevant dimensions reflecting the people's *well-being* (a concept defined as a combination of functioning well and feeling good) [156]. The proposed conceptual model classifies hedonic factors into two groups. On the one hand, the first group considers those factors predominantly influenced by emotional aspects. In this group, the top-level concept is *affect*, which is often used to include emotional-related terms [157]. According to the results obtained in the systematic review performed in this article, very few articles have considered affective factors when developing HRI systems for industrial scenarios. A relevant exceptions is [102]. However, they only consider the affective response as one of the factors affecting trust. On the other hand, the second group considers those factors where both emotional and cognitive aspects take relevance. In this group, the top-level concept is *beliefs*. In the affective computing literature *beliefs* are often associated with cognitive responses able to trigger emotions (i.e., affective response). Furthermore, emotions can influence the strength, resistance to modification, and content of the people *beliefs*. This influence is denoted as affective biasing [158,159]. According to [36], relevant HRI factors associated to *beliefs* are *attitudes*, *anxiety*, *acceptance* and *trust*. Unlike purely emotional factors, *beliefs* are taking more attention in the HRI community with industrial focus, being *trust* the most common hedonomic aspect evaluated or discussed.

6.4.2. *Ergonomic quality factors*

As described in section 6.2 ergonomics commonly focus on two main objectives. The first objective is to optimize human mental and physical well-being by preventing pain and risk situations when interacting or working with machines. The second objective is to optimize the system's performance by improving its objective usability and functionality. We divide ergonomics factors in three main classes: *performance*, *physical ergonomics* and *cognitive ergonomics*. Section 5 discussed performance metrics for HRI. *Physical ergonomics* and *cognitive ergonomics* are the most used classifications in *ergonomics*. On the one hand, physical ergonomics deals with the potential negative effects or consequences on the human body produced by working situations, such as postures, heavy work, repetitive movements, or forces [160]. In this context, the main goal is to build interactive systems and working environments that are compatible with the size, strength, and physical capabilities of users, and that at the same time does not create additional health or injuries risks [161]. On the other hand, factors in *cognitive ergonomics* focus on the creation of systems that matches the perceptual and psychological capabilities of users; therefore, enabling users to understand the state of the environment and reasoning about it [161]. Unlike the factors presented in section 6.4.1 where emotions can present a considerable influence, this class includes those factors where mostly cognitive and rational capabilities are required and where cognitive and perceptual elements can be potentially influenced in a negative way. Another difference done in this article is that *beliefs* and *affective factors* can be measured, changed or influenced before, during and after the interaction with robots, while the factors included in *cognitive ergonomics* and *physical ergonomics* are predominantly measured or relevant during interaction with robots.

7. Emergent approaches and open challenges toward Industry 5.0

7.1. *No-invasive monitoring and online analysis of human factors*

The Industry 5.0 paradigm promotes technologies enabling the optimization of human well-being by designing a new generation of human-centered smart environments. Therefore, evaluating human factors before, during, and after the interaction with robots is an essential process that can enable developers and researchers to improve working conditions. In this context, most of the tools for assessing human factors require offline or “pen-and-paper" observational techniques [162]. Ajoudani et al. [163] argues that many traditional and offline tools for ergonomics assessment “are mostly based on heuristic algorithms and limits the design of effective technologies personalized to individual workers". Lorenzini et al. [162] additionally argues that classical ergonomic tools, such as RULA and EAWS, mainly focus on kinematic aspects and omit the impact of dynamic elements in Human-Robot Interaction. Many alternative and online solutions often present complex and advanced biomechanical systems that require intrusive and computational expensive sensor systems(e.g., marker-based motion capture systems and EMG). Moreover, their results are often be obtained weeks after the tasks are performed by the users, which makes these systems not viably transferable outside laboratory settings [162,163]. Therefore, creating accurate, noninvasive, and online ergonomic assessment tools or frameworks that require short preparation represents a relevant challenge in HRI for manufacturing settings.

7.2. Individualized human-robot Interaction

Due to practical reasons, applications enabling interactions between humans and robots are generally short and static [164]. In factories, robots are often used to follow collective goals (such as the promulgation of system progress and functionality) over the human’s individual goals (i.e., adaptable and personal perfection) [164]. Individualized machine interaction is defined [165] as one of the five main categories for Industry 5.0. This factor is essential for reaching the interconnection and combination of humans and robots strengths [9], endorsing interaction quality and engagement across long-term interactions, increasing intention to use and actual usage, and maintaining trust [164,166]. Technologies enabling individualized human-machine interaction are identified in [165] as human action recognition, intention prediction, augmented, virtual or mixed reality for training and inclusiveness, exoskeletons, and collaborative robots. In HCI and HRI, individualized user-adaptive or personalized systems are able to continuously collect and processes personal and physiological data for monitoring and safety purposes, adapt to the individuals’ needs, emotions, and preferences, learn to interact with humans, and maintain long-term interactions [166,167,164]. However, personalized HRI systems could be not universally accepted due to possible privacy concerns of users [168,169]. As described in section 6.4.1, *hedonomics* factors mostly focus on individual goals. Many of these factors are often underestimated in previous works and Industry 4.0 applications. However, *hedonomics* factors will require more research attention on applications for Industry 5.0. In this context, recently, Hu et al. [170] explored the relationships between physical and physiological data with personalities and perceptions during physical HRI. In addition, Diamantopoulos et al. [171], argue that *hedonomics* factors, such as emotional response, can be used to improve safety and effectively assist human co-workers by adapting robot behaviors to human emotions. Aside from human-machine cooperation and operator assistant technologies, human-centered initiatives need also to consider technologies enabling job satisfaction, work-life balance, as well as up-skilling and re-skilling of workers, [9]. We believe that the creation of inclusive HRI environments that prioritize health, autonomy, dignity, and privacy of people with different mental and physical abilities, such as [172], as well as background and cultures, will be a relevant research topic for the next years for the Industry 5.0 and Society 5.0.

7.3. Creation of transparent robotics systems

Many Industry 4.0 applications rely on black-box Artificial Intelligence (AI) methods to enhance the level of autonomy [22,173]. However, Industry 5.0 systems able to interact and cooperate with humans must be able to display transparent behaviors [21,22]. Transparency in human-robot interaction can be used as an umbrella term to cover other overlapped concepts, such as predictability, legibility, and explainability [173]. Transparent AI systems under concepts of observability and predictability of system behavior follow the user-centered design principle of: “keep the user aware of the state of the system" [173]. In this context, to provide a good level of transparency, the human must be able to know what the robot is doing and why, what the robot will do next, why and when there is a failure in the system and possible solutions to solve errors [173]. A related research topic is the generation of legible robot movements which can help humans to anticipate the robot intentions [174]. Busch et al. [175] consider that a behavior can be considered to be legible when “an observer is able to quickly and correctly infer the intention of the agent generating the behavior." This HRI quality, denoted as *legibility* or *readability*, is generally applied in the context of robot motions. A formal definition of *legibility* is presented in [176]. They also highlight the differences between *legibility* and *predictability*, which can be considered contradictory properties of the robot motion. While a legible motion “enables an observer to quickly and confidently infer the correct goal *G*," a predictable motion “matches what an observer would expect, given the goal *G*" [176]. Examples of works focused on the creation of legible motions for handover tasks are presented in [177,178]. Examples of works using self-reports and physiological methods to evaluate legibility are presented in [106,179]. In this context, the creation of trajectories universally legible (i.e., with different cultural backgrounds) is one of the main open issues in this topic [175]. Humans use redundant signals from multiple modalities, such as vision, hearing, and touch, to anticipate robots and other humans’ intent. The reverse problem consists of developing robots to understand and anticipate human actions. An example of a research work dealing with this issue is presented in [180]. A possible approach for addressing this problem is to create multimodal systems. However, redundancy in multimodal systems often produces data presenting high dimensionality [181], which often deteriorates the efficiency of the machine learning algorithms. The most intuitive solution to this problem is the use of dimensionality reduction techniques such as Principal Component Analysis (PCA) [182]. A review of solutions addressing dimensionality issues in multimodal systems is presented in [183]. On the other, eXplainable Artificial Intelligence (XAI) has presented rapid growth and increase in academic attention in the last years [184,185]. According to [185] XAI methods can be data-driven (focused on the understanding and overcoming of the opaqueness of black-box algorithms) or goal-driven (agents and robots capable of explaining their behavior to users). Explainable Robotics is a goal-driven approach in the context of HRI [184] that focuses on developing cognitive models and algorithms that enable the generation of explanations, work in different levels of autonomy, and improve trust and situational awareness. Some of the challenges of goal-driven XAI for HRI are: the creation of methods enabling explanations using past experiences [184] and the creation of metric able to evaluate how efficient and effective explanations given by the robot are and how humans react to these explanations [185].

7.4. Evaluating fluency

Rather than be considered a metric, *fluency* is described in [105] as a quality of interaction presented when a team (e.g., a human and a robot) collaborate on a shared activity. Guy Hoffman, who first introduced the term of *fluency* in [186], considers that a team is fluent when they reach “a high level of coordination, resulting in a well-synchronized meshing of actions or joint activities, which timing is precise and efficient" [105]. Moreover, they must to dynamically adapt their plans and actions when needed. However, research in human-robot collaboration *fluency* is still in their initial stages. Moreover, many frameworks proposing metrics of *fluency* are task-specific, making other of the metrics more suitable for different scenarios [105]. A recent review of metrics used by the robotics community to evaluate *fluency* is presented by Hoffman [105]. Hoffman classifies metrics for fluency as subjective (grasping the human perception of fluency) and objective (quantitatively estimating the degree of fluency). Hu also concludes that "fluency in human-robot collaboration is not a well-defined construct and is inherently somewhat vague and ephemeral" [105]. Therefore, we consider that the factors affecting or composing *fluency* as well as the design of metric able to assess *fluency* for different types of collaborative settings will still be a topic of discussion in the robotics community for the next years.

7.5. Development of adaptive workload systems

As described in section 6.3.4, maintaining optimal workload levels in humans (i.e., avoid situations of excessive load or low engagement) is relevant for reducing accidents and tasks errors as well as improving the general task performance. For this, a robotic system must be able to accurately estimate in real-time the level workload in humans via a workload assessment algorithm [128,187]. Inputs of a workload assessment algorithm are generally physiological measures, such as heart rate, neurophysiological signals, and skin temperature. Results of the workload assessment algorithm can be used to change interaction mediums, the level of autonomy and reallocate roles, tasks, and responsibilities between the human and the robot [187]. Systems capable of those actions can be denoted as adaptive workload or adaptive teaming systems [128]. A recent example of a human-robot adaptive teaming system where the team is required to follow a set of steps that simulate a response to a disaster event is presented in [188]. The use of these algorithms in other human-robot teaming paradigms and scenarios is still an open challenge [188].

7.6. Privacy in data-driven human-robot Interaction

Data-driven technologies such as Big Data, Machine Learning, Cloud computing, and the Internet of the Things provide can provide relevant benefits not only for improving production performance but also improve the working environment of humans [189]. However, because these technologies require the acquisition, communication, storage, and processing of personalized and potentially sensitive data of human workers, privacy concerns are becoming more and more relevant [189, 190]. As described by Mannhardt et al. [191], Industry 4.0 has been focused on sensing the environment and improving the manufacturing performance (through the use of equipment, such as robots) "without much regard to the human operator". Mannhardt et al. also argue that "the paradigm of the cyber-physical system does not adequately cover the human factor" and its design principles often neglect privacy and trust-related issues [191]. Privacy is defined as the "right to be left alone" [192]. In human-centered manufacturing, the focus of privacy effort must be centered in the protection of personal information of human workers as well as frameworks and methods ensuring data security [190]. In this context, cybersecurity assessment criteria have recently been proposed in [193] for HRI in automobile manufacturing. However, more effort towards the development of more comprehensive cybersecurity metrics for HRI and HRC will be required [193].

7.7. Benchmarks

In recent years, international robotics competitions have become a powerful tool to evaluate the performance of robotics systems. While fostering innovation and pushing the state of the art, competitions also constitute a particular form of reproducibility. Besides the evident applicability to the competing teams, the publicly available information about the tasks, rules, results, videos, and sometimes even code enable the evaluation of non-competing systems.

The competition framework makes heterogeneous systems perform the same tasks under a commonly shared set of rules and, typically, in near-real-world conditions. Once the common ground is set, the scoring system becomes key to evaluate the competitors' performance. Since the competition scores tend to hide underlying characteristics of the systems that lead to a given performance, it is also necessary to use existing or propose new sets of metrics that unveil the hidden features [194]. The competitions facilitate the analyses by enabling the comparison of the competitors' systems, linking the relevant metrics to the score, and elucidating what features influenced the score and in which way.

Most commonly, the score is an objective evaluation of the performance based on the task completion (e.g., accuracy of image classification [195], obstacles traversed [196,197], items correctly placed [198]). Few competitions, such as the Future Convenience Store Challenge [199], also evaluate the safety in HRI. Such safety score is awarded if all the following subtasks are completed: the robot stops upon a customer incursion in its workspace, announces its intentions to withdraw from the shelf targeted by the customer, withdraws, and, finally, comes back and resumes the task. As highly simplified to fit in the format of the competition as it may be, this score signals for a shift toward a more human-centered objective evaluation.

8. Conclusions

In order to move toward a more human-centered society and industry, HRI researchers require to broaden their focus from mere task-fulfillment to more holistic approaches enabling robotics systems to meet collective and individual goals. In this article, we identified measures, metrics, and quality factors adopted or applied in the HRI literature using a systematic approach; therefore answering research question RQ1. We proposed two models that classify performance-related and human-centered aspects of robotics systems. While these models are mainly constructed under the needs and concepts in industrial and collaborative robotics, they can also be applicable to other robotics disciplines. We also present those human-centered quality factors that have received more attention in the robotics literature; therefore answering research question RQ2. These factors are attitude, acceptance, trust, mental and physical workload, awareness, mental models, and safety. Finally, we also identified seven emergent research areas, which can be relevant in the next years to build Industry 5.0 applications; therefore answering research question RQ3. These areas are no-invasive monitoring and online analysis of human factors, individualized HRI, transparent robotic systems, fluency, privacy in data-driven HRI, benchmarks, and adaptive workload systems. Additionally, we summarize theoretical frameworks presented in the literature to help researchers and practitioners understand and differentiate between complex and often confusing terms in the area.

This article proposed a taxonomy of performance metrics and measures based on current trends in robotics and previous works and a holistic model for HRI based on recent frameworks in HCI. Researchers and practitioners that would like to build human-centered applications for Industry 5.0 and Society 5.0 will require to be trained in human factors and select suitable metrics and measurements to evaluate their applications. Results summarized in the proposed models can be used as a reference for these researchers. However, more efforts must be performed to identify or propose measures and metrics able to assess hedonomics (e.g., fun, pleasure, and emotional reactions) and sustainability (e.g., carbon footprint, energy consumption, waste reduction). Therefore, future work will expand our holistic model in these directions. Many of the human-centered concepts and tools we present in this article are inspired by other disciplines, such as Human-Computer Interaction, workplace ergonomics, and social robotics. Therefore, the suitability of these tools for specific HRC scenarios will require extensive research efforts.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This paper is based on results obtained from a project, JPNP20006, commissioned by the New Energy and Industrial Technology Development Organization (NEDO), Japan.

Compliance with ethical standards

The authors declares that they have no conflict of interest.

References

[1] Kan SH. Metrics and models in software quality engineering. Addison-Wesley Longman Publ Co, Inc 2002.
[2] Zubrow D. Software quality requirements and evaluation,the iso 25000 series. Softw Eng 2020.
[3] Deissenboeck F, Juergens E, Lochmann K, Wagner S. Software quality models: purposes, usage scenarios and requirements. 2009 ICSE Workshop Softw Qual, IEEE 2009:9–14.
[4] Systems and software engineering-systems and software quality requirements and evaluation(square)-evaluation process (2011).
[5] P. Nistala, K.V. Nori, R. Reddy, Software quality models: A systematic mapping study, in: 2019 IEEE/ACM International Conference on Software and System Processes (ICSSP), IEEE, 2019, 125–134.
[6] Yan M, Xia X, Zhang X, Xu L, Yang D, Li S. Software quality assessment model: a systematic mapping study, Science China. Inf Sci 2019;62(9):1–18.
[7] Damacharla P, Javaid AY, Gallimore JJ, Devabhaktuni VK. Common metrics to benchmark human-machine teams (HMT): a review. IEEE Access 6 2018: 38637–55. https://doi.org/10.1109/ACCESS.2018.2853560.
[8] De Nul L, Breque M, Petridis A. Industry 5.0, towards a sustainable, human-centric and resilient european industry. Tech Rep Gen Res Innov Eur Comm 2021.
[9] Xu X, Lu Y, Vogel-Heuser B, Wang L. Industry 4.0 and Industry 5.0 - inception, conception and perception. J Manuf Syst 2021;61:530–5. https://doi.org/10.1016/j.jmsy.2021.10.006.
[10] Demir KA, Döven G, Sezen B. Industry 5.0 and human-robot co-working. Procedia Comput Sci 2019;158:688–95. https://doi.org/10.1016/j.procs.2019.09.104.
[11] Gilchrist A. Industry 4.0: the industrial internet of things. Springer; 2016.
[12] Acemoglu D, Restrepo P. Tasks, automation, and the rise in US wage inequality. Tech Rep Natl Bur Econ Res 2021. https://doi.org/10.3386/w28920.
[13] Beaudry P, Green DA, Sand BM. The great reversal in the demand for skill and cognitive tasks. J Labor Econ 2016;34(S1):S199–247. https://doi.org/10.1086/682347.
[14] Schwabe H, Castellacci F. Automation, workers' skills and job satisfaction. PLOS ONE 2020;15(11):e0242929. https://doi.org/10.1371/journal.pone.0242929.
[15] Złotowski J, Yogeeswaran K, Bartneck C. Can we control it? autonomous robots threaten human identity, uniqueness, safety, and resources. Int J Hum-Comput Stud 2017;100:48–54. https://doi.org/10.1016/j.ijhcs.2016.12.008.
[16] Brougham D, Haar J. Smart technology, artificial intelligence, robotics, and algorithms (STARA): employees' perceptions of our future workplace. J Manag Organ 2018;24(2):239–57. https://doi.org/10.1017/jmo.2016.55.
[17] Susskind D. A world without work: Technology, automation and how we should respond. Penguin Books UK 2020.
[18] Almetwally AA, Bin-Jumah M, Allam AA. Ambient air pollution and its influence on human health and welfare: an overview. Environ Sci Pollut Res Int 2020;27 (20):24815–30. https://doi.org/10.1007/s11356-020-09042-2.
[19] Bouchikhi A, Maherzi W, Benzerzour M, Mamindy-Pajany Y, Peys A, Abriak N-E. Manufacturing of low-carbon binders using waste glass and dredged sediments: formulation and performance assessment at laboratory scale. Sustainability 2021; 13(9):4960. https://doi.org/10.3390/su13094960.
[20] Fukuyama M. Society 5.0: Aiming for a new human-centered society. Jpn Spotlight 2018;27:47–50.
[21] Nahavandi S. Industry 5.0-a human-centric solution. Sustainability 2019;11(16): 4371. https://doi.org/10.3390/su11164371.
[22] Maddikunta PKR, Pham Q-V, Prabadevi B, Deepa N, Dev K, Gadekallu TR, et al. Industry 5.0: a survey on enabling technologies and potential applications. J Ind Inf Integr 2021:100257. https://doi.org/10.1016/j.jii.2021.100257.
[23] D.R. Olsen, M.A. Goodrich, Metrics for evaluating human-robot interactions,in: PERMIS Conference, 2003.
[24] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, M. Goodrich, Common metrics for human-robot interaction, in: ACM SIGCHI/SIGART conference on Human-robot interaction, Association for Computing Machinery, 2006, 33–40.10.1145/1121241.1121249.
[25] R.R. Murphy, D. Schreckenghost, Survey of metrics for human-robot interaction, in: ACM/IEEE International Conference on Human-Robot Interaction, IEEE, 2013, 197–198.10.1109/HRI.2013.6483569.
[26] Marvel JA, Bagchi S, Zimmerman M, Antonishek B. Towards effective interface designs for collaborative HRI in manufacturing: metrics and measures. ACM Trans Hum-Robot Interact 2020;9(4):1–55. https://doi.org/10.1145/3385009.
[27] Systems and software engineering-Systems and software Quality Requirements and Evaluation(SQuaRE)-System and software quality models (2011).
[28] Coronado E, Venture G, Yamanobe N. Applying kansei/affective engineering methodologies in the design of social and service robots: a systematic review. Int J Soc Robot 2021;13(5):1161–71.
[29] Petersen K, Vakkalanka S, Kuzniarz L. Guidelines for conducting systematic mapping studies in software engineering: an update. Inf Softw Technol 2015;64: 1–18. https://doi.org/10.1016/j.infsof.2015.03.007.
[30] D. Budgen, P. Brereton, Performing systematic literature reviews in software engineering,in: International conference on Software engineering, Association for Computing Machinery, 2006, 1051–1052.10.1145/1134285.1134500.
[31] Coronado E, Mastrogiovanni F, Indurkhya B, Venture G. Visual programming environments for end-user development of intelligent and social robots, a systematic review. J Comput Lang 2020;58:100970. https://doi.org/10.1016/j.cola.2020.100970.
[32] Kitchenham B. Procedures for performing systematic reviews. Tech Rep, Keele Univ 2004.
[33] Noone M, Mooney A. Visual and textual programming languages: a systematic review of the literature. J Comput Educ 2018;5(2):149–74.
[34] Al-Turjman F, Nawaz MH, Ulusar UD. Intelligence in the internet of medical things era: a systematic review of current and future trends. Comput Commun 2020;150:644–60.
[35] Baroroh DK, Chu C-H, Wang L. Systematic literature review on augmented reality in smart manufacturing: collaboration between human and computational intelligence. J Manuf Syst 2021;61:696–711.
[36] Naneva S, SardaGou M, Webb TL, Prescott TJ. A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. Int J Soc Robot 2020;12: 1179–201. https://doi.org/10.1007/s12369-020-00659-4.
[37] Sagar K, Saha A. A systematic review of software usability studies. Int J Inf Technol 2017:1–24. https://doi.org/10.1007/s41870-017-0048-1.
[38] S. Keele, et al., Guidelines for performing systematic literature reviews in software engineering, Tech. rep., EBSE Technical Report(2007).
[39] Wohlin C, Runeson P, Neto PA dMS, Engström E, do Carmo Machado I, De Almeida ES. On the reliability of mapping studies in software engineering. J Syst Softw 2013;86(10):2594–626. https://doi.org/10.1016/j.jss.2013.04.076.
[40] Markey R, Harris C, Lamm F, Kesting S, Ravenswood K, Simpkin G, et al. Improving productivity through enhancing employee wellbeing and participation. Labr Employ Work NZ 2019. https://doi.org/10.26686/lew.v0i0.1666.
[41] Tangen S. Demystifying productivity and performance. Int J Product Perform Manag 2005;54(1):34–46. https://doi.org/10.1108/17410400510571437.
[42] Forth J, McNabb R. Workplace performance: a comparison of subjective and objective measures in the 2004 workplace employment relations survey. Ind Relat J 2008;39(2):104–23. https://doi.org/10.1111/j.1468-2338.2007.00480.x.
[43] Wagner S, Deissenboeck F. Defining productivity in software engineering. In: Rethinking productivity in software engineering. Springer; 2019. p. 29–38. https://doi.org/10.1007/978-1-4842-4221-6_4.
[44] Mainz J. Defining and classifying clinical indicators for quality improvement. Int J Qual Health Care 2003;15(6):523–30. https://doi.org/10.1093/intqhc/mzg081.
[45] G. Canbek, S. Sagiroglu, T.T. Temizel, N. Baykal, Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights,in:International Conference on Computer Science and Engineering, IEEE, 2017, 821–826.10.1109/UBMK.2017.8093539.
[46] Texel PP. Measure, metric, and indicator: an object-oriented approach for consistent terminology. in:IEEE SoutheastCon,IEEE. 2013. p. 1–5. https://doi.org/10.1109/SECON.2013.6567438.
[47] Systems and software engineering-vocabulary 2017.
[48] Choong KK. Understanding the features of performance measurement system: a literature review. Meas Bus Excell 2013;17(4):102–21. https://doi.org/10.1108/MBE-05-2012-0031.
[49] Choong KK. Use of mathematical measurement in improving the accuracy (reliability) & meaningfulness of performance measurement in businesses & organizations. Measurement 2018;129:184–205. https://doi.org/10.1016/j.measurement.2018.04.008.
[50] J.V. Stone, Information theory: a tutorial introduction.
[51] E.A. Locke, G.P. Latham, New developments in goal setting and task performance.
[52] Huwe RA. Metrics 2.0: creating scorecards for high-performance work teams and organizations: creating scorecards for high-performance work teams and organizations. Abc-clio 2010.
[53] Mihaiu DM, Opreana A, Cristescu MP, et al. Efficiency, effectiveness and performance of the public sector. Rom J Econ Forecast 2010;4(1):132–47.
[54] Eiriz V, Barbosa N, Figueiredo J. A conceptual framework to analyse hospital competitiveness. Serv Ind J 2010;30(3):437–48. https://doi.org/10.1080/02642060802236137.
[55] Berntzen L. Electronic government service efficiency: how to measure efficiency of electronic services. In: Measuring E-government Efficiency. Springer; 2014. p. 75–92. https://doi.org/10.1007/978-1-4614-9982-4_5.
[56] Van Looy A, Shafagatova A. Business process performance measurement: a structured literature review of indicators. Meas Metr, Springe 2016;5(1):1–24. https://doi.org/10.1186/s40064-016-3498-1.
[57] E. Coronado, G. Venture, N. Yamanobe, Applying kansei/affective engineering methodologies in the design of social and service robots: A systematic review, International Journal of Social Robotics 13 (2020)1161–1171.10.1007/s12369–020-00709-x.
[58] Shiizuka H, Hashizume A. The role of kansei/affective engineering and its expected in aging society. In: Intelligent Decision Technologies. Springer; 2011. p. 329–39. https://doi.org/10.1007/978-3-642-22194-1_33.

[59] Bannon L. Reimagining HCI: toward a more human-centered perspective. Interactions 2011;18(4):50–7. https://doi.org/10.1145/1978822.1978833.
[60] Nagamachi M, Lokman AM. Innovations of Kansei engineering. CRC Press; 2010. https://doi.org/10.1201/EBK1439818664.
[61] Ergonomics of human-system interaction-Part11:Usability:Definitions and concepts 2018.
[62] Software engineering-Productquality-Part1:Quality model 2001.
[63] Nielsen J. usability 101: Introd usability 2012. ⟨https://www.nngroup.com/articles/usability-101-introduction-to-usability/⟩.
[64] Alonso-Ríos D, Vázquez-García A, Mosqueira-Rey E, Moret-Bonillo V. Usability: a critical analysis and a taxonomy. Int J Hum-Comput Interact 2009;26(1):53–74. https://doi.org/10.1080/10447310903025552.
[65] Shackel B. Usability-context, framework, definition, design and evaluation. Interact Comput 2009;21(5–6):339–46. https://doi.org/10.1016/j.intcom.2009.04.007.
[66] Gupta D, Ahlawat AK, Sagar K. Usability prediction & ranking of SDLC models using fuzzy hierarchical usability model. Open Eng 2017;7(1):161–8. https://doi.org/10.1515/eng-2017-0021.
[67] Morville P. Experience design unplugged,in: ACM SIGGRAPH. Association for computing machinery. 2005. https://doi.org/10.1145/1187335.1187347.
[68] Zarour M, Alharbi M. User experience framework that combines aspects, dimensions, and measurement methods. Cogent Eng 2017;4(1):1421006. https://doi.org/10.1080/23311916.2017.1421006.
[69] F. Lachner, P. Naegelein, R. Kowalski, M. Spann, A. Butz, Quantified UX: Towards a common organizational understanding of user experience,in: Nordic conference on human-computer interaction,Association for Computing Machinery, 2016, 1–10.10.1145/2971485.2971501.
[70] Kadir BA, Broberg O, daConceicao CS. Current research and future perspectives on human factors and ergonomics in Industry 4.0. Comput Ind Eng 2019;137:106004. https://doi.org/10.1016/j.cie.2019.106004.
[71] Neumann WP, Winkelhaus S, Grosse EH, Glock CH. Industry 4.0 and the human factor-a systems framework and analysis methodology for successful development. Int J Prod Econ 2021;233:107992. https://doi.org/10.1016/j.ijpe.2020.107992.
[72] S. Diefenbach, N. Kolb, M. Hassenzahl, The 'hedonic' in human-computer interaction: history, contributions, and future research directions, in: Conference on Designing Interactive Systems, 2014, 305–314.10.1145/2598510.2598549.
[73] Sauer J, Sonderegger A, Schmutz S. Usability, user experience and accessibility: towards an integrative model. Ergonomics 2020;63(10):1207–20. https://doi.org/10.1080/00140139.2020.1774080.
[74] D. Rajanen, T. Clemmensen, N. Iivari, Y. Inal, K. Rılzvanoğlu, A. Sivaji, A. Roche, UX professionals' definitions of usability and UX-a comparison between Turkey, Finland, Denmark, France and Malaysia, in: IFIP Conference on Human-Computer Interaction, Springer, 2017, 218–239.10.1007/978–3-319–68059-0_14.
[75] Dubey SK, Rana A. Analytical roadmap to usability definitions and decompositions. Int J Eng Sci Technol 2010;2(9):4723–9.
[76] Hancock PA, Pepe AA, Murphy LL. Hedonomics: The power of positive and pleasurable ergonomics. Ergon Des 2005;13(1):8–14. https://doi.org/10.1177/106480460501300104.
[77] Ergonomics principles in the design of work systems 2016.
[78] Oron-Gilad T, Hancock PA. From ergonomics to hedonomics: Trends in human factors and technology-the role of hedonomics revisited. Emot Affect Hum Factors Hum-Comput Interact, Elsevier 2017:185–94. https://doi.org/10.1016/B978-0-12-801851-4.00007-0.
[79] Helander MG, Khalid HM. Underlying theories of hedonomics for affective and pleasurable design. Hum Factors Ergon Soc Annu Meet 2005;49(18):1691–5. https://doi.org/10.1177/154193120504901803.
[80] Vemula B, Matthias B, Ahmad A. A design metric for safety assessment of industrial robot design suitable for power- and force-limited collaborative operation. Int J Intell Robot Appl 2018;2(2):226–34. https://doi.org/10.1007/s41315-018-0055-9.
[81] Oyekan JO, Hutabarat W, Tiwari A, Grech R, Aung MH, Mariani MP, et al. The effectiveness of virtual environments in developing collaborative strategies between industrial robots and humans. Robot Comput-Integr Manuf 2019;55:41–54. https://doi.org/10.1016/j.rcim.2018.07.006.
[82] Robla-Gómez S, Becerra VM, Llata JR, Gonzalez-Sarabia E, Torre-Ferrero C, Perez-Oria J. Working together: a review on safe human-robot collaboration in industrial environments. IEEE Access 2017;5:26754–73.
[83] Gualtieri L, Rauch E, Vidoni R, Matt DT. An evaluation methodology for the conversion of manual assembly systems into human-robot collaborative workcells. Procedia Manuf 2019;38:358–66. https://doi.org/10.1016/j.promfg.2020.01.046.
[84] W. Zhao, L. Sun, C. Liu, M. Tomizuka, Experimental evaluation of human motion prediction toward safe and efficient human robot collaboration, in: American Control Conference, IEEE, 2020, 4349–4354.10.23919/ACC45564.2020.9147277.
[85] M.P. Hippertt, M.L. Junior, A.L. Szejka, O.C. Junior, E.R. Loures, E.A. P. Santos, Towards safety level definition based on the HRN approach for industrial robots in collaborative activities, Procedia manufacturing 38(2019)1481–1490.10.1016/j.promfg.2020.01.139.
[86] B. Lacevic, P. Rocco, Kinetostatic danger field-a novel safety assessment for human-robot interaction,in: IEEE/RS J International Conference on Intelligent Robots and Systems, IEEE, 2010, 2169–2174.10.1109/IROS.2010.5649124.
[87] S. Kumar, F. Sahin, A framework for an adaptive human-robot collaboration approach through perception-based real-time adjustments of robot behavior in industry, in: System of Systems Engineering Conference, IEEE, 2017, 1–6.10.1109/SYSOSE.2017.7994967.
[88] Saenz J, Vogel C, Penzlin F, Elkmann N. Safeguarding collaborative mobile manipulators - evaluation of the VALERI workspace monitoring system. Procedia Manuf 2017;11:47–54. https://doi.org/10.1016/j.promfg.2017.07.129.
[89] Robots and robotic devices - safety requirements for industrial robots - part 1: Robots 2011.
[90] Matsas E, Vosniakos G-C, Batras D. Prototyping proactive and adaptive techniques for human-robot collaboration in manufacturing using virtual reality. Robot Comput-Integr Manuf 2018;50:168–80. https://doi.org/10.1016/j.rcim.2017.09.005.
[91] Otto A, Battaïa O. Reducing physical ergonomic risks at assembly lines by line balancing and job rotation: a survey. Comput Ind Eng 2017;111:467–80. https://doi.org/10.1016/j.cie.2017.04.011.
[92] Robots and robotic devices-safety requirements for personal care robots 2014.
[93] Robots and robotic devices-collaborative robots 2016.
[94] Safety of machinery - positioning of safe guards with respect to the approach speeds of parts of the human body 2010.
[95] Charalambous G, Fletcher S, Webb P. The development of a scale to evaluate trust in industrial human-robot collaboration. Int J Soc Robot 2016;8(2):193–209. https://doi.org/10.1007/s12369-015-0333-8.
[96] Hoff KA, Bashir M. Trust in automation: Integrating empirical evidence on factors that influence trust. Hum Factors 2015;57(3):407–34. https://doi.org/10.1177/0018720814547570.
[97] Z.R. Khavas, S.R. Ahmadzadeh, P. Robinette, Modeling trust in human-robot interaction: A survey, in:International Conference on Social Robotics, Springer, 2020, 529–541.10.1007/978–3-030–62056-1_44.
[98] Gervasi R, Mastrogiacomo L, Franceschini F. A conceptual framework to evaluate human-robot collaboration. Int J Adv Manuf Technol 2020;108:841–65. https://doi.org/10.1007/s00170-020-05363-1.
[99] K.E. Oleson, D.R. Billings, V. Kocsis, J.Y. Chen, P.A. Hancock, Antecedents of trust in human-robot collaborations,in: IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, IEEE, 2011, 175–178.10.1109/COGSIMA.2011.5753439.
[100] A.R. Wagner, R.C. Arkin, Recognizing situations that demand trust, in: IEEE International Conference on Robot and Human Interactive Communication, IEEE, 2011, 7–14.10.1109/ROMAN.2011.6005228.
[101] Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. Acad Manag Rev 1995;20(3):709–34.
[102] Kim W, Kim N, Lyons JB, Nam CS. Factors affecting trust in high-vulnerability human-robot interaction contexts: a structural equation modelling approach. Appl Ergon 2020;85:103056. https://doi.org/10.1016/j.apergo.2020.103056.
[103] Yagoda RE, Gillan DJ. You want me to trust a ROBOT? the development of a human-robot interaction trust scale. Int J Soc Robot 2012;4(3):235–48. https://doi.org/10.1007/s12369-012-0144-0.
[104] Hancock PA, Billings DR, Schaefer KE, Chen JY, De Visser EJ, Parasuraman R. A meta-analysis of factors affecting trust in human-robot interaction. Hum Factors 2011;53(5):517–27. https://doi.org/10.1177/0018720811417254.
[105] Hoffman G. Evaluating fluency in human-robot collaboration. IEEE Trans Hum-Mach Syst 2019;49(3):209–18. https://doi.org/10.1109/THMS.2019.2904558.
[106] A.D. Dragan, S. Bauman, J. Forlizzi, S.S. Srinivasa, Effects of robot motion on human-robot collaboration, in: ACM/IEEE International Conference on Human-Robot Interaction,IEEE, 2015, 51–58.
[107] Rahman SM, Wang Y. Mutual trust-based subtask allocation for human-robot collaboration in flexible lightweight assembly in manufacturing. Mechatronics 2018;54:94–109. https://doi.org/10.1016/j.mechatronics.2018.07.007.
[108] Meissner A, Trübswetter A, Conti-Kufner AS, Schmidtler J. Friend or Foe? understanding assembly workers' acceptance of human-robot collaboration. ACM Trans Hum-Robot Interact 2020;10(1):1–30. https://doi.org/10.1145/3399433.
[109] Eagly AH, Chaiken S. The psychology of attitudes. Harcourt Brace Jovanovich Coll Publ 1993.
[110] Ajzen I. Nature and operation of attitudes. Annu Rev Psychol 2001;52(1):27–58. https://doi.org/10.1146/annurev.psych.52.1.27.
[111] Edison SW, Geissler GL. Measuring attitudes towards general technology: Antecedents, hypotheses and scale development. J Target, Meas Anal Mark 2003;12(2):137–56. https://doi.org/10.1057/palgrave.jt.5740104.
[112] Nomura T, Suzuki T, Kanda T, Kato K. Measurement of negative attitudes toward robots. Interact Stud: Soc Behav Commun Biol Artif Syst 2006;7(3):437–54. https://doi.org/10.1075/is.7.3.14nom.
[113] Rosen L, Weil M. Measuring technophobia. a manual for the administration and scoring of the computer anxiety rating scale. Comput Thoughts Surv Gen Attitude Comput Scale 1992.
[114] Bartneck C, Kulić D, Croft E, Zoghbi S. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. Int J Soc Robot 2009;1(1):71–81. https://doi.org/10.1007/s12369-008-0001-3.
[115] Heerink M, Kröse B, Evers V, Wielinga B. Assessing acceptance of assistive social agent technology by older adults: the almere model. Int J Soc Robot 2010;2(4):361–75. https://doi.org/10.1007/s12369-010-0068-5.
[116] Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. MIS Q 2003:425–78. https://doi.org/10.2307/30036540.
[117] Greenwald AG, McGhee DE, Schwartz JL. Measuring individual differences in implicit cognition: the implicit association test. J Personal Soc Psychol 1998;74(6):1464. https://doi.org/10.1037//0022-3514.74.6.1464.

[118] T. Ninomiya, A. Fujita, D. Suzuki, H. Umemuro, Development of the multi-dimensional robot attitude scale: constructs of people's attitudes towards domestic robots, in: International Conference on Social Robotics, Springer, 2015, 482–491.10.1007/978–3-319–25554-5_48.
[119] S.F. Warta, I don't always have positive attitudes, but when i do it is usually about a robot: Development of the robot perception scale, in: FLAIRS Conference, 2015.
[120] T. Nomura, K. Sugimoto, D.S. Syrdal, K. Dautenhahn, Social acceptance of humanoid robots in japan: A survey for development of the frankenstein syndorome questionnaire, in: IEEE-RAS International Conference on Humanoid Robots, IEEE, 2012, 242–247.10.1109/HUMANOIDS.2012.6651527.
[121] Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319–40.
[122] Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000;46(2):186–204. https://doi.org/10.1287/mnsc.46.2.186.11926.
[123] Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decis Sci 2008;39(2):273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x.
[124] C. Bröhl, J. Nelles, C. Brandl, A. Mertens, C.M. Schlick, TAM reloaded: a technology acceptance model for human-robot cooperation in production systems, in: International conference on human-computer interaction, Springer, 2016, 97–103.10.1007/978–3-319–40548-3_16.
[125] MacDonald W. The impact of job demands and workload on stress and fatigue. Aust Psychol 2003;38(2):102–17. https://doi.org/10.1080/00050060310001707107.
[126] Hancock PA. Human Factors Psychology. Elsevier; 1987.
[127] Wickens CD, Gordon SE, Liu Y, Lee J. An introduction to human factors engineering. Pearson Prentice Hall Saddle River, NJ 2004;vol. 2.
[128] Heard J, Harriott CE, Adams JA. A survey of workload assessment algorithms. IEEE Trans Hum-Mach Syst 2018;48(5):434–51. https://doi.org/10.1109/THMS.2017.2782483.
[129] Lindström K. Psychosocial criteria for good work organization. Scand J Work, Environ Health 1994;20:123–33.
[130] Parasuraman R, Caggiano D. Mental workload. In: Ramachandran V, editor. Encyclopedia of the human brain. New York: Academic Press; 2002. p. 17–27. https://doi.org/10.1016/B0-12-227210-2/00206-5.
[131] Stanton NA, Hedge A, Brookhuis K, Salas E, Hendrick HW. Handbook of human factors and ergonomics methods. CRC Press,; 2004.
[132] Young MS, Brookhuis KA, Wickens CD, Hancock PA. State of science: mental workload in ergonomics. Ergonomics 2015;58(1):1–17. https://doi.org/10.1080/00140139.2014.956151.
[133] S. Lemaignan, F. Garcia, A. Jacq, P. Dillenbourg, From real-time attention assessment to "with-me-ness" in human-robot interaction, in: ACM/IEEE International Conference on Human-Robot Interaction, IEEE, 2016, 157–164.10.1109/HRI.2016.7451747.
[134] Paletta L, Pszeida M, Ganster H, Fuhrmann F, Weiss W, Ladstätter S, et al. Gaze-based human factors measurements for the evaluation of intuitive human-robot collaboration in real-time. IEEE Int Conf Emerg Technol Fact Autom, IEEE 2019: 1528–31. https://doi.org/10.1109/ETFA.2019.8869270.
[135] Hart SG, Staveland LE. Development of NASA-TLX (task load index): Results of empirical and theoretical research. : Adv Psychol 1988;52:139–83. https://doi.org/10.1016/S0166-4115(08)62386-9.
[136] Mitchell DK. Mental workload and ARL workload modeling tools. Tech Rep Army Res Lab Aberd Proving Ground MD 2000.
[137] Chihara T, Fukuchi N, Seo A. Optimal product design method with digital human modeling for physical workload reduction: a case study illustrating its application to handrail position design. Jpn J Ergon 2017;53(2):25–35. https://doi.org/10.5100/jje.53.25.
[138] Williams N. The Borg rating of perceived exertion (RPE) scale. Occup Med 2017; 67(5):404–5. https://doi.org/10.1093/occmed/kqx063.
[139] Deakin J, Stevenson J, Vail G, Nelson J. The use of the nordic questionnaire in an industrial setting: a case study. Appl Ergon 1994;25(3):182–5. https://doi.org/10.1016/0003-6870(94)90017-5.
[140] Melzack R, Raja SN. The McGill pain questionnaire: from description to measurement. Anesthesiology 2005;103(1):199–202. https://doi.org/10.1097/00000542-200507000-00028.
[141] Stanton NA, Chambers PR, Piggott J. Situational awareness and safety. Saf Sci 2001;39(3):189–204. https://doi.org/10.1016/S0925-7535(01)00010-8.
[142] Endsley MR. Situation awareness misconceptions and misunderstandings. J Cogn Eng Decis Mak 2015;9(1):4–32. https://doi.org/10.1177/1555343415572631.
[143] Kaber DB, Endsley MR. Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. Process Saf Prog 1997;16(3):126–31. https://doi.org/10.1002/prs.680160304.
[144] Karwowski W, Szopa A, Soares MM. Handbook of standards and guidelines in human factors and ergonomics. CRC Press; 2021.
[145] Endsley MR. Design and evaluation for situation awareness enhancement. Human Fact Soc Annu Meet 1988;32(2):97–101. https://doi.org/10.1177/154193128803200221.
[146] Salvendy G. Handbook of human factors and ergonomics. John Wiley & Sons; 2012.
[147] A. Tabrez, M.B. Luebbers, B. Hayes, A survey of mental modeling techniques in human-robot teaming, Current Robotics Reports (2020)259–267.10.1007/s43154–020-00019–0.
[148] Mathieu JE, Heffner TS, Goodwin GF, Salas E, Cannon-Bowers JA. The influence of shared mental models on team process and performance. J Appl Psychol 2000; 85(2):273–83. https://doi.org/10.1037/0021-9010.85.2.273.
[149] S. Nikolaidis, J. Shah, Human-robot cross-training: computational formulation, modeling and evaluation of a human team training strategy, in: ACM/IEEE International Conference on Human-Robot Interaction, IEEE, 2013, 33–40.
[150] Opiyo S, Zhou J, Mwangi E, Kai W, Sunusi I. A review on teleoperation of mobile ground robots: architecture and situation awareness. Int J Control, Autom Syst 2021;19:1384–407. https://doi.org/10.1007/s12555-019-0999-z.
[151] Gorbenko A, Popov V, Sheka A. Robot self-awareness: exploration of internal states. Appl Math Sci 2012;6(14):675–88.
[152] Abowd GD, Dey AK, Brown PJ, Davies N, Smith M, Steggles P. Towards a better understanding of context and context-awareness. Int Symp handheld ubiquitous Comput, Springe 1999:304–7. https://doi.org/10.1007/3-540-48157-5_29.
[153] Dahn N, Fuchs S, Gross H-M. Situation awareness for autonomous agents. IEEE Int Symp Robot Hum Interact Commun, IEEE 2018:666–71. https://doi.org/10.1109/ROMAN.2018.8525511.
[154] Smith K, Hancock PA. Situation awareness is adaptive, externally directed consciousness. Hum Factors 1995;37(1):137–48. https://doi.org/10.1518/001872095779049444.
[155] Hudlicka E. To feel or not to feel: The role of affect in human-computer interaction. Int J Hum-Comput Stud 2003;59(1–2):1–32. https://doi.org/10.1016/S1071-5819(03)00047-8.
[156] Ruggeri K, Garcia-Garzon E, Maguire Á, Matz S, Huppert FA. Well-being is more than happiness and life satisfaction: a multidimensional analysis of 21 countries. Health Qual life Outcomes 2020;18(1):1–16. https://doi.org/10.1186/s12955-020-01423-y.
[157] Zhang P. The affective response model: a theoretical framework of affective concepts and their relationships in the ICT context. MIS Q 2013:247–74.
[158] Bosse T, Hoogendoorn M, Klein MC, Treur J, Van Der Wal CN, Wissen AVan. Modelling collective decision making in groups and crowds: Integrating social contagion and interacting emotions, beliefs and intentions. Auton Agents Multi-Agent Syst 2013;27(1):52–84. https://doi.org/10.1007/s10458-012-9201-1.
[159] Frijda NH, Manstead AS, Bem SE. Emotions and belief: How feelings influence thoughts. Cambridge University Press; 2000.
[160] Ahram TZ, Karwowski W. Advances in physical ergonomics and safety. CRC Press; 2012. https://doi.org/10.1201/b12323.
[161] McLeod RW. Designing for human reliability. Gulf Prof Publ 2015. https://doi.org/10.1016/B978-0-12-802421-8.00001-1.
[162] Lorenzini M, Kim W, Ajoudani A. An online multi-index approach to human ergonomics assessment in the workplace. IEEE Trans Hum-Mach Syst 2021.
[163] Ajoudani A, Albrecht P, Bianchi M, Cherubini A, DelFerraro S, Fraisse P, et al. Smart collaborative systems for enabling flexible and ergonomic work practices [industry activities]. Autom Mag 2020;27(2):169–76.
[164] B. Irfan, A. Ramachandran, S. Spaulding, D.F. Glas, I. Leite, K.L. Koay, Personalization in long-term human-robot interaction, in: ACM/IEEE International Conference on Human-Robot Interaction, IEEE, 2019, 685–686.10.1109/HRI.2019.8673076.
[165] Muller J. Enabling technologies for Industry 5.0, results of a workshop with europeas technology leaders. Tech Rep Gen Res Innov(Eur Comm) 2020.
[166] Funk M, Rosen P, Wischniewski S. Human-centered HRI design - the more individual the better? chances and risks of individualization. Workshop Behav Patterns Interact Model Pers Hum-Robot Interact, Fraunhofer IAO 2020:8–11. https://doi.org/10.1145/3371382.3374846.
[167] Gao AY, Barendregt W, Castellano G. Personalised human-robot co-adaptation in instructional settings using reinforcement learning. :IVA Workshop Persuas Embodied Agents Behav Change 2017.
[168] Schürmann T, Beckerle P. Personalizing human-agent interaction through cognitive models. Front Psychol 2021;11. https://doi.org/10.3389/fpsyg.2020.561510.
[169] Ku Y-C, Li P-Y, Lee Y-L. Are you worried about personalized service? an empirical study of the personalization-privacy paradox. :International Conf HCI Bus, Gov, Organ, Springe 2018:351–60. https://doi.org/10.1007/978-3-319-91716-0_27.
[170] Hu Y, Abe N, Benallegue M, Yamanobe N, Venture G, Yoshida E. Toward active physical human-robot interaction: quantifying the human state during interactions. IEEE Trans Hum-Mach Syst 2020.
[171] Diamantopoulos H, Wang W. Accommodating and assisting human partners in human-robot collaborative tasks through emotion understanding. : 2021 12th Int Conf Mech Aerosp Eng (ICMAE), IEEE 2021:523–8.
[172] Drolshagen S, Pfingsthorn M, Gliesche P, Hein A. Acceptance of industrial collaborative robots by people with disabilities in sheltered workshops. Front Robot AI 2021;7:173. https://doi.org/10.3389/frobt.2020.541741.
[173] Alonso V, De La Puente P. System transparency in shared autonomy: a mini review. Front Neurorobotics 2018;12:83. https://doi.org/10.3389/fnbot.2018.00083.
[174] ElZaatari S, Marei M, Li W, Usman Z. Cobot programming for collaborative industrial tasks: an overview. Robot Auton Syst 2019;116:162–80. https://doi.org/10.1016/j.robot.2019.03.003.
[175] Busch B, Grizou J, Lopes M, Stulp F. Learning legible motion from human-robot interactions. Int J Soc Robot 2017;9(5):765–79 (doi:Learning legible motion from human-robot interactions).
[176] A.D. Dragan, K.C. Lee, S.S. Srinivasa, Legibility and predictability of robot motion, in:ACM/IEEE International Conference on Human-Robot Interaction, IEEE, 2013, 301–308.10.1109/HRI.2013.6483603.

[177] R. Alami, A. Clodic, V. Montreuil, E.A. Sisbot, R. Chatila, Toward human-aware robot task planning, in: AAAI spring symposium: to boldly go where no human-robot team has gone before, 2006, 39–46.
[178] Lichtenthäler C, Lorenzy T, Kirsch A. Influence of legibility on perceived safety in a virtual human-robot path crossing task. IEEE Int Symp Robot Hum Interact Commun, IEEE 2012:676–81. https://doi.org/10.1109/ROMAN.2012.6343829.
[179] Dehais F, Sisbot EA, Alami R, Causse M. Physiological and subjective evaluation of a human-robot object hand-over task. Appl Ergon 2011;42(6):785–91. https://doi.org/10.1016/j.apergo.2010.12.005.
[180] W. Wang, R. Li, Y. Chen, Y. Jia, Human intention prediction in human-robot collaborative tasks, in:Companion of the 2018 ACM/IEEE international conference on human-robot interaction, 2018, 279–280.
[181] Thiran J-P, Marques F, Bourlard H. Multimodal. Signal processing: theory and applications for human-computer interaction. Academic Press,; 2009.
[182] Wold S, Esbensen K, Geladi P. Principal component analysis. Chemom Intell Lab Syst 1987;2(1–3):37–52.
[183] Gao L, Song J, Liu X, Shao J, Liu J, Shao J. Learning in high-dimensional multimedia data: the state of the art. Multimed Syst 2017;23(3):303–13.
[184] Setchi R, Dehkordi MB, Khan JS. Explainable robotics in human-robot interactions. Procedia Comput Sci 2020;176:3057–66. https://doi.org/10.1016/j.procs.2020.09.198.
[185] S. Anjomshoae, A. Najjar, D. Calvaresi, K. Främling, Explainable agents and robots: Results from a systematic literature review, in: International Conference on Autonomous Agents and Multiagent Systems, International Foundation for Autonomous Agents and Multiagent Systems, Association for Computing Machinery, 2019, 1078–1088.
[186] Hoffman G, Breazeal C. Cost-based anticipatory action selection for human-robot fluency. IEEE Trans Robot 2007;23(5):952–61. https://doi.org/10.1109/TRO.2007.907483.
[187] Heard J, Harriott CE, Adams JA. A human workload assessment algorithm for collaborative human-machine teams. IEEE Int Symp Robot Hum Interact Commun, IEEE 2017:366–71. https://doi.org/10.1109/ROMAN.2017.8172328.
[188] Heard J, Heald R, Harriott CE, Adams JA. A diagnostic human workload assessment algorithm for collaborative and supervisory human-robot teams. ACM Trans Hum-Robot Interact 2019;8(2):1–30. https://doi.org/10.1145/3314387.
[189] Mannhardt F, Petersen SA, Oliveira MF. Privacy challenges for process mining in human-centered industrial environments. 2018 14th Int Conf Intell Environ (IE), IEEE 2018:64–71.
[190] Petersen SA, Mannhardt F, Oliveira M, Torvatn H. A framework to navigate the privacy trade-offs for human-centred manufacturing. Work Conf Virtual Enterp, Springe 2018:85–97.
[191] Mannhardt F, Petersen SA, Oliveira MF. A trust and privacy framework for smart manufacturing environments. J Ambient Intell Smart Environ 2019;11(3):201–19.
[192] Warren SD, Brandeis LD. Right to privacy. Harv L Rev 1890;4:193.
[193] Rahman SM. Cybersecurity metrics for human-robot collaborative automotive manufacturing. 2021 IEEE Int Workshop Metrol Automot (MetroAutomotive), IEEE 2021:254–9.
[194] Fujita M, Domae Y, Noda A, GarciaRicardez GA, Nagatani T, Zeng A, et al. What are the important technologies for bin picking? technology analysis of robots in competitions based on a set of performance metrics. Adv Robot 2020;34(7–8):560–74. https://doi.org/10.1080/01691864.2019.1698463.
[195] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015;115(3):211–52. https://doi.org/10.1007/s11263-015-0816-y.
[196] Buehler M, Iagnemma K, Singh S. The DARPA urban challenge: autonomous vehicles in city traffic. Springer; 2009.
[197] Krotkov E, Hackett D, Jackel L, Perschbacher M, Pippine J, Strauss J, et al. The darpa robotics challenge finals: results and perspectives. J Field Robot 2017;34(2):229–40. https://doi.org/10.1002/rob.21683.
[198] Causo A, Durham J, Hauser K, Okada K, Rodriguez A. Advances on robotic item picking: applications in warehousing & e-commerce fulfillment. Springer; 2020.
[199] K. Wada, New robot technology challenge for convenience store,in: IEEE/SICE International Symposium on System Integration, IEEE, 2017, 1086–1091.10.1109/SII.2017.8279367.
[200] Reid Gary B, Nygren Thomas E. The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload. Advances in Psychology. Hancock Peter A, Meshkati Najmedin, editors. . Human Mental Workload, 52. North-Holland; 1988. p. 185–218.
[201] Zijlstra FRH. Efficiency in Work Behavior: a Design Approach for Modern Tools. Doctoral thesis. Delft, The Netherlands: Delft University of Technology; 1993.