Toward a Holistic Human-Robot Interaction Model for Society 5.0 and Industry 5.0, a Systematic Review and Taxonomy of Human-Centered and Performance Metrics

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[b,a]

[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*

Abstract

Robots are starting to take relevant and complex roles in real-world scenarios. However, society's long-term adoption of these machines will depend on the capacity of robotics systems to satisfy not only performance-centered goals but also human-centered. Unfortunately, most projects outside the social robotics community ignore or minimize the importance of human-centered aspects. This article contributes to the robotics community by presenting: i) a performance-centered taxonomy of measures and metrics for Human-Robot Interaction (HRI) and ii) a holistic model for HRI that puts human well-being at the center. We built this taxonomy and model based on the results of a systematic literature review of research articles focused on human-robot collaboration. For this, we performed a systematic search in relevant databases for robotics (Science Direct, IEEE Xplore, ACM digital library, and Springer Link). The results of this search were 75 peer-reviewed research articles published until 2020. To help practitioners and new researchers in the area, we also briefly explain complex

*Corresponding author
Email address: `enriquecoronadozu@gmail.com` (Enrique Coronado)

and overlapped terms in many cases misused in different disciplines. Finally, we identify five emergent research topics and open challenges in the area. The HRI model and taxonomy presented in this article can help researchers and practitioners to select suitable tools or methods for evaluating performance-centered and human-centered aspects in applications composed of teams of robots and humans.

Keywords: Human-Robot Interaction, Human-Robot Collaboration, Metrics, Robotics, Industry 5.0, Society 5.0

2010 MSC: 00-01, 99-00

---

1. Introduction

Nowadays, many industries adopt the Industry 4.0 paradigm, also referred to as Smart Manufacturing or Industrial Internet of Things. Industry 4.0 focuses on the digital transformation of manufacturing and production processes empowered by emergent technologies such as Virtual Reality (VR), autonomous robots, the Internet of Things (IoT), Big Data, and Cloud Computing [1, 2]. The goal of Industry 4.0 is analogous to previous revolutions: "to increase productivity and achieve mass production using innovative technology" [3]. 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) [3, 1]. Therefore, Industry 4.0 and previous revolutions can be considered as *technology-driven* [1]. 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 [4]. For example, and contrary to the optimistic predictions often done in academia, reports, such as [5, 6], 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 [7, 8]. 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 [8]. Consequently, some social experts and futurists argue that "robots are taking the human jobs and are moving society towards more inequality" [9, 10]. 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 [11, 12]. Therefore, the creation of counter-measures to affront current sustainability and social fairness problems caused by industrial activity and directions will be one of the most relevant research topics for the next decades.

1.1. Industry 5.0 and Society 5.0

Futurists and governments are starting to discuss new paradigms for solving relevant social and planetary problems. In this context, two of the most relevant paradigms are Industry 5.0 [4] and Society 5.0 [13]. On the one hand, Industry 5.0 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 able 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 [4, 1]. 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 [1]. Finally, the *resilient* principle focuses on the creation of more agile, flexible, and adaptable industries [4]. On the other hand, Society 5.0 is a related concept adopted and promoted in Japan. While Industry 5.0 focuses on the manufacturing sector, Society 5.0 considers a larger variety of scenarios. Similar to Industry 5.0, Society 5.0 is encouraged within the *human-centered* and *sustainability* principles. 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 [13]. While some works in the literature consider *human-centered* approaches and Human-Robot Collaboration as an extension or emergent trend of Industry 4.0, this article makes the distinctions between Industry 4.0 and Industry 5.0 according to [4, 3, 1, 14]. These differences are summarized in Table 1. 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" [1]. Nahavandi [14] 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 [15] 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. In these human-robot collaborative scenarios promoted by Industry 5.0, the repetitive, unsafe, physically demanding tasks are assigned to robots, while humans will be in charge of critical thinking and customization [15, 14].

1.2. Measures and metrics for Human-Robot Interaction

Human-Robot Interaction (HRI) is one of the core technologies of Industry 4.0 and Industry 5.0. When implementing an HRI system, developers must evaluate how well the proposed system meets individual, collective, and production needs or objectives. In this context, measures and metrics take a keystone role not only to validate the suitability of robotics systems but to build indicators

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

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

that can guide future implementations or development cycles. Therefore, the identification, definition, and analysis of measures and metrics is an essential issue not only for the progress of the HRI discipline but also any technological and scientific area [16]. In this article, we affront the challenges of identifying and classifying measures and metrics enabling the evaluation of smart environments where humans and robots work together. For this, we performed a systematic review of relevant and novel research articles using and proposing measures, evaluation methods, and metrics for HRI with special attention to industrial and collaborative scenarios. Unlike previous works where the concepts such as measures, metrics, and indicators are often confused or used interchangeably, we start by presenting standard definitions and relevant models explaining these terms' meaning. We present the results of the systematic search from two points of view. On the one hand, we present relevant measures and metrics that better adapt to the classical *performance-oriented* objectives of Industry 4.0 and previous paradigms. For this, we classify measures and metrics using as inspiration more general performance models described in the literature of related areas. We also explain the difference between the often confused terms of *efficiency*, *effectiveness*, *productivity* and *profitability*. On the other hand, we propose a novel holistic quality model for HRI that includes both *performance-oriented* and *human-centered* attributes. For this, we start by presenting the meaning, interpretations and limits of *human-centered* areas, such as *usability*, *user experience*, *hedonomics* and *ergonomics*. Then, we introduce a set of common measures and metrics that have been used in the robotics community to measure quality elements in these areas. The proposed HRI quality model is adapted from the Human-Computer Interaction (HCI) literature and summarizes relevant attributes used in literature to evaluate interactive robotic systems in industrial and collaborative contexts. Finally, we identify emergent approaches, challenges, and research gaps towards evaluating Industry 5.0 scenarios.

1.3. 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 presents a taxonomy of objective and quantitative measures and metrics oriented to measure different performance aspects in an HRI system. Section 6 presents the proposed holistic model of HRI. Section 7 presents a set of common *human-centered* metrics and quality factors according to the results of the performed systematic review. Section 8 presents emergent approaches, challenges, and research gaps. Conclusions follow.

2. Background and related work

2.1. Concept of quality

Quality is an ambiguous and multidimensional concept that can vary according to different interests and points of view [16]. As described in [16, 17],

Figure 1: Quality in use model from ISO/IEC 25010

the different interpretations of quality can vary from intangible (i.e., that can be judged but not measured) and philosophical to professional and objective perspectives. From the engineering and professional points of view, the concept of quality usually refers to the degree to which a system, service, product, component, or process is in conformance to specified requirements [16]. Moreover, different approaches exist describing quality from the engineering point of view. Some of the most popular are product-based quality (which defines a set of desired attributes for a product), process-based quality (in which the objective is to achieve continuous process improvement), and user/customer based quality (in which the objective is to build products or services that satisfy needs and expectations) [17].

2.2. Quality models in software engineering

The main idea behind the definition of a quality model is to break down the complex and ambiguous concept of "quality" into a set of attributes, which can be further broken down to build a hierarchy or taxonomy of factors, concepts or metrics [17]. Relevant examples of quality models are described in ISO/IEC 25010 [18]. This standard presents two quality models for human-computer systems. On the one hand, the *quality in use* model described in ISO/IEC 25010 is composed of five main characteristics: *satisfaction*, *efficiency*, *freedom from risk*, *effectiveness* and *context coverage*. Some of these characteristics are divided into sub-characteristics as shown in figure 1. On the other hand, the *product quality* model defined in ISO/IEC 25010 is composed of eight main

Figure 2: Product quality from ISO/IEC 25010

characteristics: *Functional suitability*, *Performance efficiency*, *Compatibility*, *Usability*, *Reliability*, *Security*, *Maintainability* and *Portability*. In this model, each category is divided into sub-characteristics or concepts, as shown in figure 2. Characteristics of quality models presented in [18] are defined to be applicable to both computer systems and software products. Other popular quality models for software systems are the McCall Model [19], the Boehm Model [20] and the FURPS model [21, 22].

2.3. Quality models in Human-Robot Interaction

Unlike software and computer systems, the literature reports few attempts to put quality factors, concepts, and metrics together for interactive robotics systems. 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. Common metrics proposed in [24] inspired many posterior works, such as [25, 26, 27]. For example, [26] 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 seven elements), robots-related (composed of six elements), and system-related (composed of 28 elements). In 2018, a review of common metrics for Human-Machine Teams (HMT) was presented in [25]. 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 [25] 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 [27] an overview of challenges in the design of human-machine-interfaces (HMI) and HRI in collaborative manufacturing applications. Many of the metrics listed in [27] were obtained from [23, 24] and from ISO/IEC 25010 quality models presented in section 2.2. They also identify a set of 41 subjective measurements for HRI. For this, the authors performed an analysis of 290 articles from the 2015 and 2019 Association of Computing Machinery (ACM) and Institute of Electrical and Electronics Engineers (IEEE) International Conference of HRI. Marvel et al. determine the final set of measures and metrics presented in [27] as performance metrics. Some of these previous works identify evaluation methods and metrics grasping the human perspectives and some hedonomics factors (e.g., pleasure and emotions) and recognize their importance in social interactions with robots. However, they also tend to underestimate the importance of these quality attributes in professional and industrial settings; therefore, contrasting with more recent and holistic efforts in Human-Robot Collaboration. Some examples recently presented in [28, 29] highlight the importance and effects of hedonic attributes (e.g., emotions) in Human-Robot Collaboration. Even when their authors do not explicitly indicate it, it is possible to consider the previous works presented in this subsection as initial efforts to create performance-oriented quality models for HRI. In fact, the first step toward creating a quality model is to discover all possible and relevant quality factors, concepts, and metrics for the aimed product, service, and system. While this work recognizes the efforts and arguments done in previous works extending classical performance-oriented models for HRI, we also explore a novel and holistic perspective beyond the traditional considerations in robotics (being social robotics an exception). This approach recognizes the importance of multi-disciplinary research tasks not only focused on optimizing task performance but also considering *human-centered* and hedonics paradigms.

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 propose a holistic quality model for HRI that not only considers performance factors but also puts the human emotional, cognitive and physical well-being at the center; and (iv) to discover emergent approaches, open issues, research gaps and challenges in the context of manufacturing.

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 2.3, as follows:

- Initial works, such as [23, 24], have identified measures and metrics using the experience of their authors. In this article, the process used for identifying metrics and quality factors from the literature follows a systematic literature review approach.

- Some previous works have collected measures and metrics performing a search in the literature, such as [26, 27]. However, the search performed in this article spans over a broader period and more databases. Unlike [26], the search methodology is presented. We also provide relevant references defining or using the identified measures and metrics. Unlike [27], the search process in digital databases also includes objective measures and metrics.

- Unlike [25], the focus of this article is HRI systems and excludes other types of machines or interfaces (e.g., software interfaces, autonomous cars, and digital assistants); this enables the presentation of metrics that can be suitable and applicable for different types of HRI systems.

The second contribution of this article is defined as:

Through the analysis of the results obtained from the systematic search, we propose a holistic model of quality factors that not only considers those aspects used to evaluate task performance but also puts human well-being at the center.

As described in section 2.3, 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 propose a holistic model of HRI quality factors and metrics inspired by recent advances and new paradigms in some related areas such as ergonomics, usability engineering, and HCI.

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. Moreover, many measures and metrics presented in previous works are ambiguous and lack suitable references that help in their understating. In this article, we provide references to theoretical frameworks or practical cases in HCI and HRI of common *human-centered* metrics that can be useful for new researchers.

4. Methodology

Systematic studies are objective and strict research processes designed to give a broad overview of current trends, gaps, and challenges in a specific discipline [30]. They can also be used to structure a research area, synthesize evidence, and help in the position of research directions and activities [30, 31, 32]. We followed the methodology suggested in [31, 33] and updated in [30] for performing systematic literature searches in software engineering. The main stages for conducting a systematic review according to [30, 33] are: (S1) identification of the need for systematic review and development of a review protocol, (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.

4.1. Identification of the need for systematic review and development of a review protocol

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 [34]. The review protocol was developed and approved through online meetings between a postdoc student, one Assistant Professor, two Senior Researchers, an Associate Professor, and a Distinguished Professor. As described in [30], the review protocol is composed of all the stages or elements of the review plus some additional planning information (e.g., project timetable).

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?*

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

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 taxonomy and holistic model presented in sections 5 and 6, respectively. 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 7. 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

We used the PICO (Population, Intervention, Comparison, and Outcomes) method suggested in [31] 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 [30], *intervention* can refer to the technology or procedure

Database	Search result	Results of step 1	Results of step 2
IEEE Xplore	3,753	55	21
ACM Digital Library	11,811	27	5
Springer Link	9,963	87	17
Science Direct	9,598	46	34

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

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*. We refined this search string through different iterations, in which we discarded the keywords "validation," "measures", and "agent". 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 during the entirety of 2020 and before and sorted them by relevance. We searched and collected research articles for their review in January and February 2021. Reading and selection of articles by applying the inclusion and exclusion criteria were performed in March and May, 2021. We performed data collection, analysis of results, and classification of metrics and measures for the proposed models between June and November 2021. Moreover, every two weeks, all authors of this article discussed the collected data and proposed classifications in online meetings. Table 3 shows the results obtained from each database.

4.4. Study selection, quality assessment and data extraction

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.
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, 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 3, we conducted a quality assessment of the 77 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?
- Has the article 2 or more pages and is peer-reviewed?
- Has the article been cited by other articles?

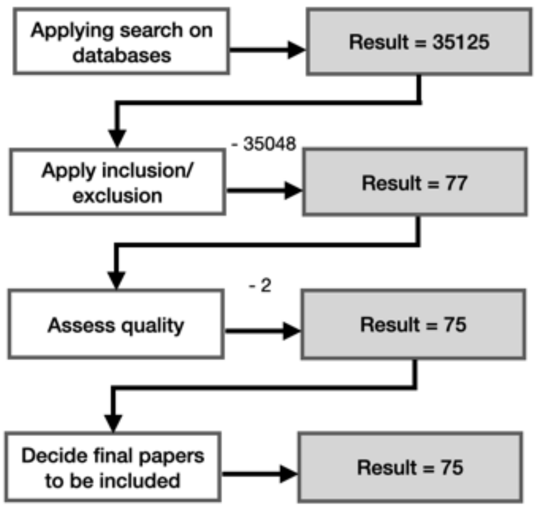

Figure 3: Number of included articles during the study selection process

Figure 3 shows the number of articles processed in each of the steps mentioned before. The search and data extraction processes were performed by authors 1 to 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.

4.5. Limitations of the study and validity evaluation

[30, 35] 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" [32]. As explained in [36, 30] 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 [30]. The *interpretive validity* "is achieved when the conclusions drawn are reason-

able given the data" [30]. 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. Development of a performance-oriented model for Human-Robot Interaction

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. [25] 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, 26, 25] these agents are selected as *human* (or operator), *robot* (or machine), and *team* (or system). Marvel et al. [27] 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. [25] 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 [26], the *Human* and *Robot* categories do not present sub-categories. Table 4 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 4 only shows the number of metrics composing each main category.

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. [26]	System	Productivity, Efficiency, Reliability, Safety, Coactivity	28
	Human		7
	Robot		6
Damacharla et al. [25]	Team		3
	Human		4
	Machine		3
Marvel et al. [27]	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

Table 4: Comparison between main categories and attributes proposed for taxonomies of performance-oriented metric in HRI

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 4 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.

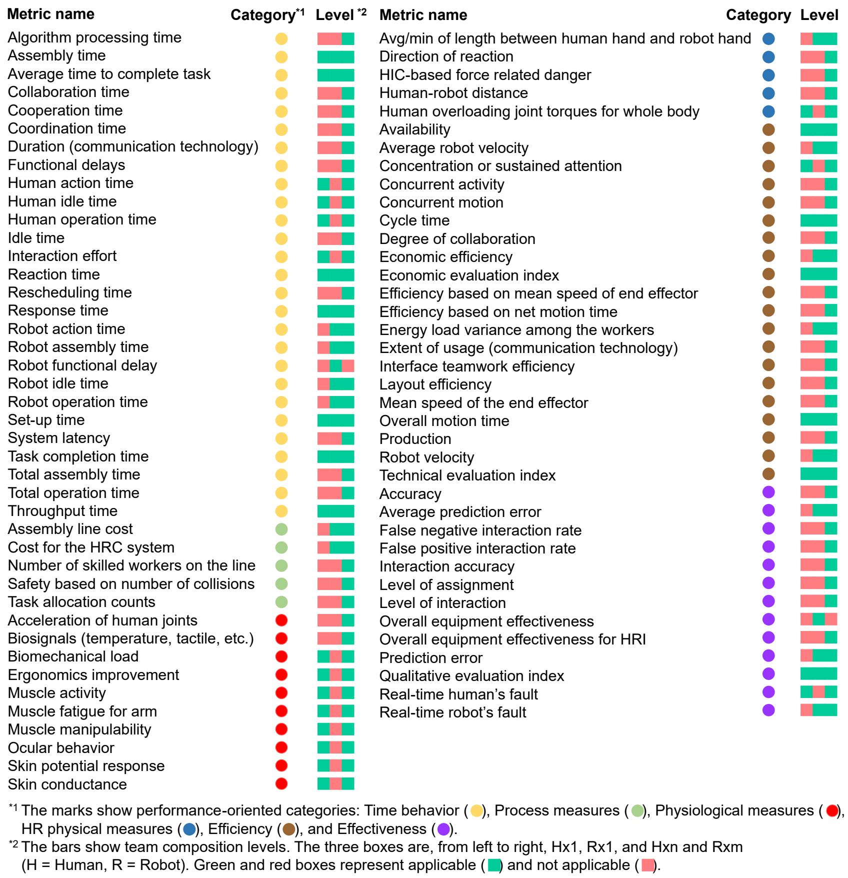

Figure 4: Performance-oriented categorization for the metrics obtained in the systematic search performed in this article.

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* [37]. These concepts are often vaguely defined and poorly understood in the literature of several disciplines [38]. Moreover, due to the subtle differences and mutual dependencies between these terms, they are in many cases used interchangeably [39, 37]. As described in Wagner et al. [40], 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 5 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 [38].

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* [41, 42, 43]. ISO/IEC/IEEE 24765 [44] 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 [43]. 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" [44]. 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

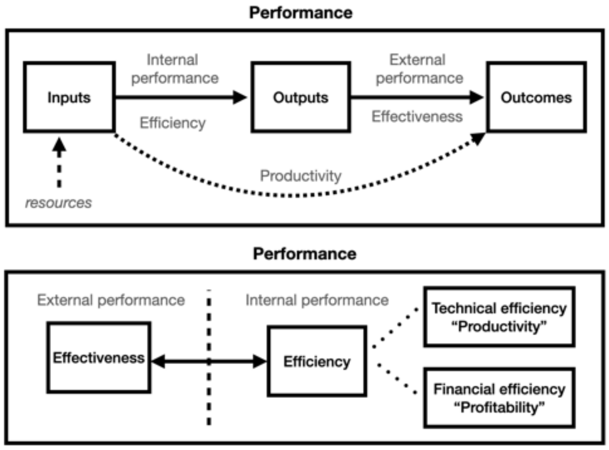

Figure 5: Relationship between performance, efficiency, profitability, effectiveness and productivity according [47, 48, 49]

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, [43] 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 [45, 46] 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 [44] and recent reviews of [45, 43, 43]. 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 [50] we identified the following groups of metrics in this category:

- *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." 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. Development of a holistic and human-centered model for Human-Robot Interaction

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 [51]. 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 [52, 53]. 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 [54]. Therefore, the main objective of this section is to present a human-centered and holistic taxonomy of metrics and quality factors for HRI. The procedure we followed to build this model is:

1. Define human-centered quality for HRI and identify the high-level quality attributes
2. Identify if exists an overlap or disagreement in the HCI and HRI community between the elements composing these high-level quality attributes and summarize the different points of view.
3. Define a model that presents and classifies those quality factors obtained as a result of the systematic review (described in section 4). We presented this model as a Venn diagram, which shows the limits between human-centered areas and identified quality attributes.

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) [55] 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* [67]. On the one hand, usability is in some cases related to "ease-of-use". However, its concept is more comprehensive.

Usability models	Usability attributes
ISO 9241-11:2018 [55]	Effectiveness, efficiency, satisfaction
ISO/IEC 9126-1:2001[56]	Understandability, learnability, operability, attractiveness
ISO/IEC 25010 [18]	Accessibility, flexibility, reliability, maintainability, compatibility
Nielsen [57]	Learnability, efficiency, memorability, errors, satisfaction
Ríos et al. [58]	Knowability, operability, efficiency, robustness, safety, satisfaction
Shackel et al. [59]	Effectiveness, learnability, flexibility, subjectively pleasing
Gupta et al. [60]	Efficiency, effectiveness, satisfaction, memorability, security, universality, productivity

Table 5: Usability attributes in the Human-Computer interaction literature adapted from [34, 60]

UX models	UX attributes
ISO 9241-11:2018 [55]	Emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors and accomplishments
UX honeycomb [61]	Usefulness, usability, desirability (i.e., emotional appreciation), findability, accessibility, credibility
Zarour et.al [62]	Hedonic (emotional, trustworthiness, aesthetics, fun, privacy, sensual), pragmatic (usability, functionality, usefulness)
Lachner et.al [63]	Look (aesthetics/design, interface, information value), feel (control, learnability, pleasure, satisfaction, ease of use), usability (efficiency, utility, effectiveness, functionality)

Table 6: Relevant user experience (UX) attributes in the Human-Computer Interaction literature

Source	Ergonomics area	Domains and attributes
Kadir et.al [64]	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. [65]	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

Table 7: Cognitive and physical ergonomics attributes and domains found in recent surveys on ergonomics applied on industrial environments

Source	Domains and attributes
Zarour et.al [62]	Emotional, trustworthiness, aesthetics, fun, privacy, sensual
Diefenbach et.al [66]	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 8: Hedonomics

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*) [67]. Despite this standardized definition, there is no consensus in the HCI and HRI communities about the definition of usability [68]. Therefore, several authors propose different attributes composing the definition of usability. Examples of review articles summarizing the different definitions of usability are [69, 68]. Table 5 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 [62] 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 6 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*, [67] proposed a holistic model designed to be consistent with the ISO standards' definitions. This model integrates the

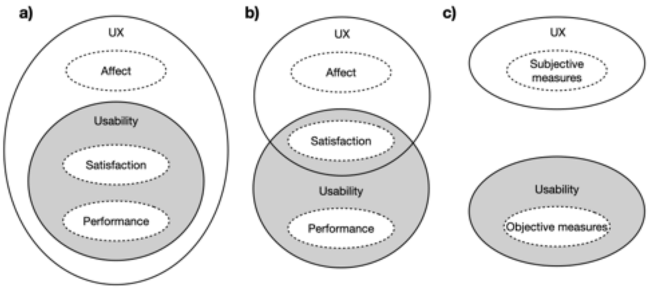

Figure 6: Different interpretations and relationships of User Experience (UX) and usability found in the literature. Adapted from [67] and [62]

holistic approach of *user experience* 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 [67] as a starting point for the development of the HRI model presented in this article. Figure 6 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 [71] 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 [72].

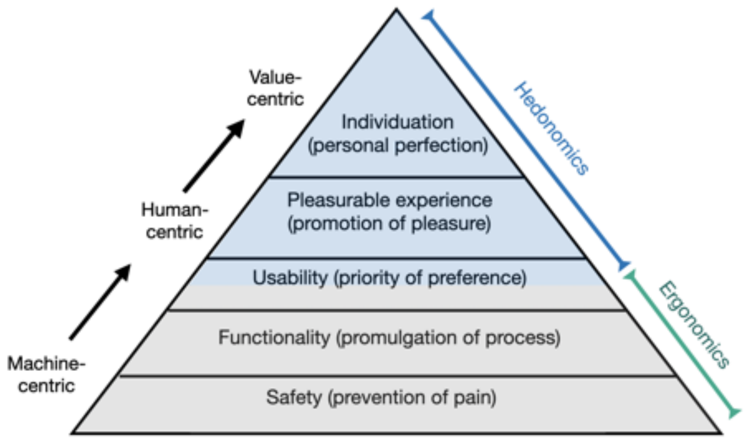

Figure 7: Hancock's Hedonomic Pyramid adapted from [70]. This pyramid shows the limits between *hedonomics* (in blue) and *ergonomics* (in gray)

Most works in the literature identify two main areas of ergonomics: *physical* and *cognitive ergonomics*. These areas are explained in section 6.3.2. Relevant factors and domains described in the literature for *physical* and *cognitive ergonomics* are shown in table 7.

Hedonomics represents a conceptual companion of *ergonomics* focused on "the pleasant or enjoyable aspects of human-technology interaction" [70]. As explained in [72], 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" [72]. Conversely, *hedonomics* focus on more positive aspects of work interactions by "promoting the occurrence of satisfying interactions, which can be proved or observed" [72]. Areas related to *hedonomics* are *user experience*, *kansei engineering* [51] and *pleasurable design* [73]. These satisfaction and affective focused paradigms proposed in hedonomics disciplines contrast with the predominant safety and productivity-oriented focus of traditional research in *ergonomics* [72]. The Hancock's Hedonomic Pyramid

Type	Measurable dimension/quality aspect
Affect	Emotional responses (1)
Beliefs	Attitudes/Acceptance (9), Anxiety (3), and Trust (11), Perceived robot ability (1), Perceived robot intelligence (1), Social Presence (1), Human-likeness (1)
Cognitive ergonomics	Mental workload (5), Concentration/Attention (3), Mental models and Awareness (6)
Physical ergonomics	Physical workload (3), Safety (15), Physical fatigue (1), Physical comfort (1), Workplace design (1)

Table 9: 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.

proposed in [70] (shown in figure 7), 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).

6.3. Definition of a human-centered model from the results of the systematic review

Taking as inspiration the works, concepts, and models proposed in the HCI literature, specifically [58, 67, 62], as well as the results of the systematic search proposed in this article, we propose a holistic quality model adapted for Human-Robot Interaction. Figure 8 shows the relationships between more relevant attributes found in the literature. This models shows existing relationship and limits between *usability*, *user experience*, *hedonomics* and *ergonomics* using the concepts explained in section 6.2.

6.3.1. 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 [74]. Moreover, 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) [75]. Our quality 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 [76]. 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 [77]. 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 [78, 79]. According to [80], 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.

6.3.2. 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,

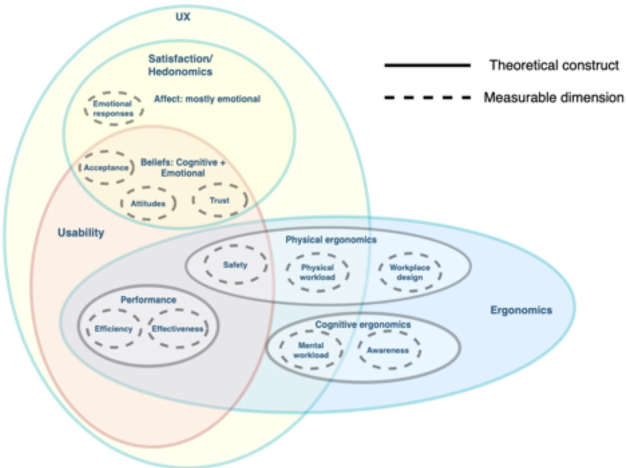

Figure 8: Most representative quality factors analyzed in the HRI literature according to the results of the systematic review. This diagram is adapted to HRI from the Interaction Experience model proposed in [67]

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 [81]. 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 [82]. 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 [82]. Unlike the factors presented in section 6.3.1 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. Common human-centered factors for Human-Robot Interaction

From the results of the systematic search, we identified those factors where the robotics community has put most of the attention. We briefly present those factors below.

7.1. Safety

Safety is a critical quality aspect in *ergonomics*. As shown in figure 7, 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 10 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) [95] ISO/TS 15066:2016 (collaborative robots) [96], ISO 13855:2010 (positioning of safeguards with respect to the approach speeds of parts of the human body) [97] , 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 [88]. Other popular methods used in the industry to evaluate physical ergonomic risks at assembly lines are summarized in [94] and displayed in Table 11. Unlike most of the metrics presented in Table 10, which can be specific to HRC, methods displayed in Table 11 are more general. Therefore, they are applicable in environments where workers have some risk of presenting musculoskeletal disorders.

Article	Context	Methods/Metrics
Marvel et.al [27]	Safety features for collaborative robots	Speed and separation monitoring (SSM) Power and force limiting (PFL).
Matsas et.al. [83]	Standards for Human-Robot Collaboration	Velocity of the end-effector [84], Maximum dynamic power [84], Maximum static force [84]
	Process safety output	Number of collisions between human and robot, mean velocity of the end-effector
Gualtier et. al. [85]	Evaluation if an activity can provide physical stress or if it could be dangerous for humans	Safety and Ergonomic evaluation index (SEEI)
Vemula et.al. [86]	Assessment of the severity of a transient physical contact between a robot manipulator and a human body region	Safety design metric based on power flux density
Zhao et.al. [87]	Human-Robot Collaboration safety metrics	Safety index (safety as function of the distance between human and robot)
Kumar et.al [88]	Human-Robot Collaboration safety metrics	Number of conflicts between human and robot, Average separation distance between human and robot
Saenz et.al [89]	Safety when mobile robots work in close proximity to human operators	Protective separation distance between the tool and a human operator
Hippertt et.al [90]	Assign levels of safety that allow a robot to perform a collaborative activity	Hazard Rating Number
Oyekana et.al [91]	Calculate the effect on the human if a robot were to hit the human	Head Injury Criteria (HCI)-based force related danger
Avanzini et.al [92]	Assess how dangerous a particular robot configuration could be for a human standing in the robot's workspace	Danger field [93]

Table 10: Resulting articles proposing, gathering or using metrics for physical safety

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)

Table 11: Most common risk assessment methods according to [94]

As described in [94], 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).

7.2. *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 [98]. 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* [99]. However, there is not a consensus on a single definition of *trust* in the HRI community [100]. 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 [98, 101], 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" [101], 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" [100, 102]. Another human-centered and 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" [103]. 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 [104], 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 [77, 98]. Charalambous et al. [98] and Yagoda et al. [105] additionally present scales enabling the evaluation of trust in industrial HRC and HRI respectively. Relevant articles surveying factors affecting trust in HRI contexts are [100, 106]. 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* [107, 108], discussed in section 8.3. 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 [109], 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. Authors of [109] 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).

7.3. 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 [110]. 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 [111] 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 [112, 113] 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) [114]. 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 [115] for measuring *technofobia*. The recent survey proposed in [80] 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 [114], the Godspeed Questionnaire [116] (particularly in the likability dimension) and self-report measured based in semantic differential scales, such as those proposed in Kansei Engineering [51]. For *cognitive attitudes*, [80] 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 [117] and Unified Theory of Acceptance and Use of Technology [118]. Finally, *general attitudes* are identified as a mix of *affective* and *cognitive* measures. For this, [80] reveals the use of self-report and the Implicit Association Test [119] in social robotics. Additionally, we identified the Multi-dimensional Robot Attitude Scale [120] as an recent method focused on measuring attitudes towards robot in domestic scenarios and the Robot Perception Scale [121], 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 [80]. Methods identified for measuring attitudes and acceptance are the Frankenstein Syndrome Questionnaire [122], the Technology Acceptance Model (TAM) [123], and their major upgrades TAM 2 [124] and TAM 3 [125]. However, the suitability of methods for evaluating attitudes in industrial and collaborative scenarios is still uncertain. An exception is the TAM reloaded [126], 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.

7.4. 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 [127, 128]. *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" [129, 130]. The literature presents two main classifications of workload. One of the initial classifications of workload, proposed in [131], 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 [132], *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. [133] 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" [134, 133]. As described in [127], the methods and metrics considered under *mental workload* are from numerous and task-specific research activities about the limitations and capacities of information processing systems in humans. These methods are classified in [134] 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) [134]. 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" [134]. The metric of with-me-ness was introduced in [135] 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 [136]. *Subjective reports* are the most popular way to measure *mental workload*. Traditional methods such as NASA Task Load indeX [137], the Subjective Workload Assessment Technique (SWAT) (Reid and Nygren 1988) and the simple and fast Rating Scale Mental Effort (RSME) (Zijlstra 1993) are known to be complicated and time-consuming as well as to present retrospective/recall bias (i.e., incorrect recall due memory effects) [134]. Results from the systematic review performed in this article show that the self-reporting method, particularly the NASA-TLX [137], 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) [132]. Other physiological measurements correlated with an increase in mental workload are Skin Conductance Activity (SCA) and breathing rate.

7.5. Physical workload

The overall workload can be decomposed into seven components: cognitive, gross motor, fine motor, tactile, visual, speech, and auditory [130, 138]. According [130], *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" [139]. Works reporting the evaluation of physical workload use the NASA-TLX. Objective metrics able to measure *physical workload* have been classified in [130]. 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

[140], the Nordic Body Discomfort questionnaire [141] and The McGill Pain Questionnaire (MPQ) [142].

7.6. 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 [143]. 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 [144]. 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 [145, 143]. Stanton et al. [133] 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 [146, 147, 143]. This conception is defined by Endsley [147] 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" [146, 147]. 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 [143], 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 [146] 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 [146] 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 [148, 144]. Therefore, *mental models* can be used to build and maintain *situation awareness*, especially in the levels of comprehension and projection [148]. 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 [148]. Tabrez et al. [149] 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 [149, 150]. Metrics to quantitative evaluate mental convergence and similarity of shared mental models in Human-Robot Interaction are described in [151]. 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 [152] *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, 27] are considered on the robot-side. According [153], 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 [153, 24]. On the other, *human-awareness* is defined in [24] as "the degree to which a robot is aware of humans." *Context Awareness* [154] is another related concept used in HCI and robotics

[155]. Nikolas et.al. [155] recently presented a framework that integrates context and situation awareness under the less known theory of Smith and Hancock of situation awareness [156].

8. Emergent approaches and open challenges toward Industry 5.0

8.1. Individualized Human-Robot Interaction

Due to practical reasons, applications enabling interactions between humans and robots are generally short and static [157]. 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) [157]. Individualized machine interaction is defined [158] 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 [1], endorsing interaction quality and engagement across long-term interactions, increasing intention to use and actual usage, and maintaining trust [157, 159]. Technologies enabling individualized human-machine interaction are identified in [158] 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 [159, 160, 157]. However, personalized HRI systems could be not universally accepted due to possible privacy concerns of users [161, 162]. As described in section 6.3.1, *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. 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, [1]. 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 [163], 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.

8.2. Creation of transparent robotics systems

Many Industry 4.0 applications rely on black-box Artificial Intelligence (AI) methods to enhance the level of autonomy [164, 15]. However, Industry 5.0 systems able to interact and cooperate with humans must be able to display transparent behaviors [14, 15]. Transparency in human-robot interaction can be used as an umbrella term to cover other overlapped concepts, such as predictability, legibility, and explainability [164]. 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" [164]. 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 [164]. A related research topic is the generation of legible robot movements which can help humans to anticipate the robot intentions [165]. Busch et al. [166] 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 [167]. 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" [167]. Examples of works focused on the creation of legible motions for handover tasks are presented in [168, 169]. Examples of works using self-reports and physiological methods to evaluate legibility are presented in [170, 108]. 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 [166]. On the other, eXplainable Artificial Intelligence (XAI) has presented rapid growth and increase in academic attention in the last years [171, 172]. According [172] 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 [171] 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 [171] 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 [172].

8.3. Evaluating fluency

Rather than be considered a metric, *fluency* is described in [107] 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 [173], 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” [107]. 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 [107]. A recent review of metrics used by the robotics community to evaluate *fluency* is presented by Hoffman [107]. 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” [107]. 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.

8.4. Development of adaptive workload systems

As described in section 7.4, 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 [130, 174]. 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 [174]. Systems capable of those actions can be denoted as adaptive workload or adaptive teaming systems [130]. 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 [175]. The use of these algorithms in other human-robot teaming paradigms and scenarios is still an open challenge [175].

8.5. 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 [176]. 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 [177], obstacles traversed [178, 179], items correctly placed [180]). Few competitions, such as the Future Convenience Store Challenge [181], 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.

9. 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 five 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 individualized HRI, transparent robotic systems, fluency, protocols and 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 focused on the performance and human-centered aspects presented in Industry 4.0 and Industry/Society 5.0. We 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. 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.

Acknowledgment

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

Compliance with Ethical Standards

Conflict of interest. The authors declares that they have no conflict of interest.

References

[1] X. Xu, Y. Lu, B. Vogel-Heuser, L. Wang, Industry 4.0 and Industry 5.0—inception, conception and perception, Journal of Manufacturing Sys-

tems 61 (2021) 530–535. doi:https://doi.org/10.1016/j.jmsy.2021.10.006.

[2] A. Gilchrist, Industry 4.0: the industrial internet of things, Springer, 2016.

[3] K. A. Demir, G. Döven, B. Sezen, Industry 5.0 and human-robot co-working, Procedia computer science 158 (2019) 688–695. doi:https://doi.org/10.1016/j.procs.2019.09.104.

[4] L. De Nul, M. Breque, A. Petridis, Industry 5.0, towards a sustainable, human-centric and resilient european industry, Tech. rep., General for Research and Innovation (European Commission) (2021).

[5] D. Acemoglu, P. Restrepo, Tasks, automation, and the rise in US wage inequality, Tech. rep., National Bureau of Economic Research (2021). doi:https://doi.org/10.3386/w28920.

[6] P. Beaudry, D. A. Green, B. M. Sand, The great reversal in the demand for skill and cognitive tasks, Journal of Labor Economics 34 (S1) (2016) S199–S247. doi:https://doi.org/10.1086/682347.

[7] H. Schwabe, F. Castellacci, Automation, workers' skills and job satisfaction, PLOS ONE 15 (11) (2020) e0242929. doi:https://doi.org/10.1371/journal.pone.0242929.

[8] J. Złotowski, K. Yogeeswaran, C. Bartneck, Can we control it? autonomous robots threaten human identity, uniqueness, safety, and resources, International Journal of Human-Computer Studies 100 (2017) 48–54. doi:https://doi.org/10.1016/j.ijhcs.2016.12.008.

[9] D. Brougham, J. Haar, Smart technology, artificial intelligence, robotics, and algorithms (STARA): Employees' perceptions of our future workplace, Journal of Management & Organization 24 (2) (2018) 239–257. doi:https://doi.org/10.1017/jmo.2016.55.

[10] D. Susskind, A world without work: Technology, automation and how we should respond, Penguin Books UK, 2020.

[11] A. A. Almetwally, M. Bin-Jumah, A. A. Allam, Ambient air pollution and its influence on human health and welfare: an overview, Environmental Science and Pollution Research International 27 (20) (2020) 24815–24830. doi:https://doi.org/10.1007/s11356-020-09042-2.

[12] A. Bouchikhi, W. Maherzi, M. Benzerzour, Y. Mamindy-Pajany, A. Peys, N.-E. Abriak, Manufacturing of low-carbon binders using waste glass and dredged sediments: Formulation and performance assessment at laboratory scale, Sustainability 13 (9) (2021) 4960. doi:https://doi.org/10.3390/su13094960.

[13] M. Fukuyama, Society 5.0: Aiming for a new human-centered society, Japan Spotlight 27 (2018) 47–50.

[14] S. Nahavandi, Industry 5.0—a human-centric solution, Sustainability 11 (16) (2019) 4371. doi:https://doi.org/10.3390/su11164371.

[15] P. K. R. Maddikunta, Q.-V. Pham, B. Prabadevi, N. Deepa, K. Dev, T. R. Gadekallu, R. Ruby, M. Liyanage, Industry 5.0: a survey on enabling technologies and potential applications, Journal of Industrial Information Integration (2021) 100257doi:https://doi.org/10.1016/j.jii.2021.100257.

[16] S. H. Kan, Metrics and models in software quality engineering, Addison-Wesley Longman Publishing Co., Inc., 2002.

[17] S. Wagner, Software product quality control, Springer, 2013.

[18] Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models (2011).

[19] J. P. Cavano, J. A. McCall, A framework for the measurement of software quality, in: Software quality assurance workshop on Functional and performance issues, Association for Computing Machinery, 1978, pp. 133–139. doi:https://doi.org/10.1145/953579.811113.

[20] B. W. Boehm, J. R. Brown, M. Lipow, Quantitative evaluation of software quality, in: Proceedings of the 2nd international conference on Software engineering, Association for Computing Machinery, 1976, pp. 592–605.

[21] A. Basu, Software quality assurance, testing and metrics, PHI Learning Pvt. Ltd., 2015.

[22] A. AL-Badareen, M. Selamat, M. A. Jabar, J. Din, S. Turaev, Software quality models: A comparative study, in: International Conference on Software Engineering and Computer Systems, Springer, 2011, pp. 46–55. doi:https://doi.org/10.1007/978-3-642-22170-5_4.

[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, pp. 33–40. doi:https://doi.org/10.1145/1121241.1121249.

[25] P. Damacharla, A. Y. Javaid, J. J. Gallimore, V. K. Devabhaktuni, Common metrics to benchmark human-machine teams (HMT): A review, IEEE Access 6 (2018) 38637–38655. doi:https://doi.org/10.1109/ACCESS.2018.2853560.

[26] R. R. Murphy, D. Schreckenghost, Survey of metrics for human-robot interaction, in: ACM/IEEE International Conference on Human-Robot Interaction, IEEE, 2013, pp. 197–198. doi:https://doi.org/10.1109/HRI.2013.6483569.

[27] J. A. Marvel, S. Bagchi, M. Zimmerman, B. Antonishek, Towards effective interface designs for collaborative HRI in manufacturing: metrics and measures, ACM Transactions on Human-Robot Interaction 9 (4) (2020) 1–55. doi:https://doi.org/10.1145/3385009.

[28] C. Bröhl, J. Nelles, C. Brandl, A. Mertens, V. Nitsch, Human–robot collaboration acceptance model: development and comparison for germany, japan, china and the usa, International Journal of Social Robotics 11 (5) (2019) 709–726. doi:https://doi.org/10.1007/s12369-019-00593-0.

[29] A. Toichoa Eyam, W. M. Mohammed, J. L. Martinez Lastra, Emotion-driven analysis and control of human-robot interactions in collaborative applications, Sensors 21 (14) (2021) 4626. doi:https://doi.org/10.3390/s21144626.

[30] K. Petersen, S. Vakkalanka, L. Kuzniarz, Guidelines for conducting systematic mapping studies in software engineering: An update, Information and Software Technology 64 (2015) 1–18. doi:https://doi.org/10.1016/j.infsof.2015.03.007.

[31] D. Budgen, P. Brereton, Performing systematic literature reviews in software engineering, in: International conference on Software engineering, Association for Computing Machinery, 2006, pp. 1051–1052. doi:https://doi.org/10.1145/1134285.1134500.

[32] E. Coronado, F. Mastrogiovanni, B. Indurkhya, G. Venture, Visual programming environments for end-user development of intelligent and social robots, a systematic review, Journal of Computer Languages 58 (2020) 100970. doi:https://doi.org/10.1016/j.cola.2020.100970.

[33] B. Kitchenham, Procedures for performing systematic reviews, Tech. rep., Keele University (2004).

[34] K. Sagar, A. Saha, A systematic review of software usability studies, In-

ternational Journal of Information Technology (2017) 1–24doi:https://doi.org/10.1007/s41870-017-0048-1.

[35] S. Keele, et al., Guidelines for performing systematic literature reviews in software engineering, Tech. rep., EBSE Technical Report (2007).

[36] C. Wohlin, P. Runeson, P. A. d. M. S. Neto, E. Engström, I. do Carmo Machado, E. S. De Almeida, On the reliability of mapping studies in software engineering, Journal of Systems and Software 86 (10) (2013) 2594–2610. doi:https://doi.org/10.1016/j.jss.2013.04.076.

[37] R. Markey, C. Harris, F. Lamm, S. Kesting, K. Ravenswood, G. Simpkin, D. Williamson, Improving productivity through enhancing employee wellbeing and participation, Labour, Employment and Work in New Zealanddoi:https://doi.org/10.26686/lew.v0i0.1666.

[38] S. Tangen, Demystifying productivity and performance, International Journal of Productivity and performance management 54 (1) (2005) 34–46. doi:https://doi.org/10.1108/17410400510571437.

[39] J. Forth, R. McNabb, Workplace performance: a comparison of subjective and objective measures in the 2004 workplace employment relations survey, Industrial Relations Journal 39 (2) (2008) 104–123. doi:https://doi.org/10.1111/j.1468-2338.2007.00480.x.

[40] S. Wagner, F. Deissenboeck, Defining productivity in software engineering, in: Rethinking Productivity in Software Engineering, Springer, 2019, pp. 29–38. doi:https://doi.org/10.1007/978-1-4842-4221-6_4.

[41] J. Mainz, Defining and classifying clinical indicators for quality improvement, International journal for quality in health care 15 (6) (2003) 523–530. doi:https://doi.org/10.1093/intqhc/mzg081.

[42] 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, pp. 821–826. doi:https://doi.org/10.1109/UBMK.2017.8093539.

[43] P. P. Texel, Measure, metric, and indicator: An object-oriented approach for consistent terminology, in: IEEE SoutheastCon, IEEE, 2013, pp. 1–5. doi:https://doi.org/10.1109/SECON.2013.6567438.

[44] Systems and software engineering — vocabulary (2017).

[45] K. K. Choong, Understanding the features of performance measurement system: a literature review, Measuring Business Excellence 17 (4) (2013) 102–121. doi:https://doi.org/10.1108/MBE-05-2012-0031.

[46] K. K. Choong, Use of mathematical measurement in improving the accuracy (reliability) & meaningfulness of performance measurement in businesses & organizations, Measurement 129 (2018) 184–205. doi:https://doi.org/10.1016/j.measurement.2018.04.008.

[47] D. M. Mihaiu, A. Opreana, M. P. Cristescu, et al., Efficiency, effectiveness and performance of the public sector, Romanian journal of economic forecasting 4 (1) (2010) 132–147.

[48] V. Eiriz, N. Barbosa, J. Figueiredo, A conceptual framework to analyse hospital competitiveness, The Service Industries Journal 30 (3) (2010) 437–448. doi:https://doi.org/10.1080/02642060802236137.

[49] L. Berntzen, Electronic government service efficiency: how to measure efficiency of electronic services, in: Measuring E-government Efficiency, Springer, 2014, pp. 75–92. doi:https://doi.org/10.1007/978-1-4614-9982-4_5.

[50] A. Van Looy, A. Shafagatova, Business process performance measurement: a structured literature review of indicators, measures and metrics, SpringerPlus 5 (1) (2016) 1–24. doi:https://doi.org/10.1186/s40064-016-3498-1.

[51] 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. doi:https://doi.org/10.1007/s12369-020-00709-x.

[52] H. Shiizuka, A. Hashizume, The role of kansei/affective engineering and its expected in aging society, in: Intelligent Decision Technologies, Springer, 2011, pp. 329–339. doi:https://doi.org/10.1007/978-3-642-22194-1_33.

[53] L. Bannon, Reimagining HCI: toward a more human-centered perspective, Interactions 18 (4) (2011) 50–57. doi:https://doi.org/10.1145/1978822.1978833.

[54] M. Nagamachi, A. M. Lokman, Innovations of Kansei engineering, CRC Press, 2010. doi:https://doi.org/10.1201/EBK1439818664.

[55] Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts (2018).

[56] Software engineering — Product quality — Part 1: Quality model (2001).

[57] J. Nielsen, Usability 101: Introduction to usability (2012).
URL https://www.nngroup.com/articles/usability-101-introduction-to-usability/

[58] D. Alonso-Ríos, A. Vázquez-García, E. Mosqueira-Rey, V. Moret-Bonillo, Usability: a critical analysis and a taxonomy, International journal of human-computer interaction 26 (1) (2009) 53–74. doi:https://doi.org/10.1080/10447310903025552.

[59] B. Shackel, Usability–context, framework, definition, design and evaluation, Interacting with computers 21 (5–6) (2009) 339–346. doi:https://doi.org/10.1016/j.intcom.2009.04.007.

[60] D. Gupta, A. K. Ahlawat, K. Sagar, Usability prediction & ranking of SDLC models using fuzzy hierarchical usability model, Open Engineering 7 (1) (2017) 161–168. doi:https://doi.org/10.1515/eng-2017-0021.

[61] P. Morville, Experience design unplugged, in: ACM SIGGRAPH, Association for Computing Machinery, 2005, pp. 10–es. doi:https://doi.org/10.1145/1187335.1187347.

[62] M. Zarour, M. Alharbi, User experience framework that combines aspects, dimensions, and measurement methods, Cogent Engineering 4 (1) (2017) 1421006. doi:https://doi.org/10.1080/23311916.2017.1421006.

[63] 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, pp. 1–10. doi:https://doi.org/10.1145/2971485.2971501.

[64] B. A. Kadir, O. Broberg, C. S. da Conceicao, Current research and future perspectives on human factors and ergonomics in Industry 4.0, Computers & Industrial Engineering 137 (2019) 106004. doi:https://doi.org/10.1016/j.cie.2019.106004.

[65] W. P. Neumann, S. Winkelhaus, E. H. Grosse, C. H. Glock, Industry 4.0 and the human factor–a systems framework and analysis methodology for successful development, International Journal of Production Economics 233 (2021) 107992. doi:https://doi.org/10.1016/j.ijpe.2020.107992.

[66] 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, pp. 305–314. doi:https://doi.org/10.1145/2598510.2598549.

[67] J. Sauer, A. Sonderegger, S. Schmutz, Usability, user experience and accessibility: towards an integrative model, Ergonomics 63 (10) (2020) 1207–1220. doi:https://doi.org/10.1080/00140139.2020.1774080.

[68] D. Rajanen, T. Clemmensen, N. Iivari, Y. Inal, K. Rızvanoğlu, 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, pp. 218–239. doi:https://doi.org/10.1007/978-3-319-68059-0_14.

[69] S. K. Dubey, A. Rana, Analytical roadmap to usability definitions and decompositions, International Journal of Engineering Science and Technology 2 (9) (2010) 4723–4729.

[70] P. A. Hancock, A. A. Pepe, L. L. Murphy, Hedonomics: The power of positive and pleasurable ergonomics, Ergonomics in Design 13 (1) (2005) 8–14. doi:https://doi.org/10.1177/106480460501300104

[71] Ergonomics principles in the design of work systems (2016).

[72] T. Oron-Gilad, P. A. Hancock, From ergonomics to hedonomics: Trends in human factors and technology—the role of hedonomics revisited, in: Emotions and affect in human factors and human-computer interaction, Elsevier, 2017, pp. 185–194. doi:https://doi.org/10.1016/B978-0-12-801851-4.00007-0.

[73] M. G. Helander, H. M. Khalid, Underlying theories of hedonomics for affective and pleasurable design, Human factors and ergonomics society annual meeting 49 (18) (2005) 1691–1695. doi:https://doi.org/10.1177/154193120504901803.

[74] E. Hudlicka, To feel or not to feel: The role of affect in human–computer interaction, International journal of human-computer studies 59 (1–2) (2003) 1–32. doi:https://doi.org/10.1016/S1071-5819(03)00047-8.

[75] K. Ruggeri, E. Garcia-Garzon, Á. Maguire, S. Matz, F. A. Huppert, Well-being is more than happiness and life satisfaction: a multidimensional analysis of 21 countries, Health and quality of life outcomes 18 (1) (2020) 1–16. doi:https://doi.org/10.1186/s12955-020-01423-y.

[76] P. Zhang, The affective response model: A theoretical framework of affective concepts and their relationships in the ICT context, MIS Quarterly (2013) 247–274.

[77] W. Kim, N. Kim, J. B. Lyons, C. S. Nam, Factors affecting trust in high-vulnerability human-robot interaction contexts: A structural equation modelling approach, Applied ergonomics 85 (2020) 103056. doi:https://doi.org/10.1016/j.apergo.2020.103056

[78] T. Bosse, M. Hoogendoorn, M. C. Klein, J. Treur, C. N. Van Der Wal, A. Van Wissen, Modelling collective decision making in groups and crowds: Integrating social contagion and interacting emotions, beliefs and intentions, Autonomous Agents and Multi-Agent Systems 27 (1) (2013) 52–84. doi:https://doi.org/10.1007/s10458-012-9201-1.

[79] N. H. Frijda, A. S. Manstead, S. E. Bem, Emotions and belief: How feelings influence thoughts, Cambridge University Press, 2000.

[80] S. Naneva, M. Sarda Gou, T. L. Webb, T. J. Prescott, A systematic review of attitudes, anxiety, acceptance, and trust towards social robots, International Journal of Social Robotics 12 (2020) 1179–1201. doi:https://doi.org/10.1007/s12369-020-00659-4.

[81] T. Z. Ahram, W. Karwowski, Advances in physical ergonomics and safety, CRC Press, 2012. doi:https://doi.org/10.1201/b12323.

[82] R. W. McLeod, Designing for Human Reliability, Gulf Professional Publishing, 2015. doi:https://doi.org/10.1016/B978-0-12-802421-8.00001-1

[83] E. Matsas, G.-C. Vosniakos, D. Batras, Prototyping proactive and adaptive techniques for human-robot collaboration in manufacturing using virtual reality, Robotics and Computer-Integrated Manufacturing 50 (2018) 168–180. doi:https://doi.org/10.1016/j.rcim.2017.09.005

[84] Robots and robotic devices — safety requirements for industrial robots — part 1: Robots (2011).

[85] L. Gualtieri, E. Rauch, R. Vidoni, D. T. Matt, An evaluation methodology for the conversion of manual assembly systems into human-robot collaborative workcells, Procedia Manufacturing 38 (2019) 358–366. doi:https://doi.org/10.1016/j.promfg.2020.01.046

[86] B. Vemula, B. Matthias, A. Ahmad, A design metric for safety assessment of industrial robot design suitable for power- and force-limited collaborative operation, International journal of intelligent robotics and applications 2 (2) (2018) 226–234. doi:https://doi.org/10.1007/s41315-018-0055-9.

[87] 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, pp. 4349–4354. doi:https://doi.org/10.23919/ACC45564.2020.9147277.

[88] 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, pp. 1–6. doi:https://doi.org/10.1109/SYSOSE.2017.7994967.

[89] J. Saenz, C. Vogel, F. Penzlin, N. Elkmann, Safeguarding collaborative mobile manipulators - evaluation of the VALERI workspace monitoring system, Procedia Manufacturing 11 (2017) 47–54. doi:https://doi.org/10.1016/j.promfg.2017.07.129.

[90] 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. doi:https://doi.org/10.1016/j.promfg.2020.01.139

[91] J. O. Oyekan, W. Hutabarat, A. Tiwari, R. Grech, M. H. Aung, M. P. Mariani, L. López-Dávalos, T. Ricaud, S. Singh, C. Dupuis, The effectiveness of virtual environments in developing collaborative strategies between industrial robots and humans, Robotics and Computer-Integrated Manufacturing 55 (2019) 41–54. doi:https://doi.org/10.1016/j.rcim.2018.07.006.

[92] G. B. Avanzini, N. M. Ceriani, A. M. Zanchettin, P. Rocco, L. Bascetta, Safety control of industrial robots based on a distributed distance sensor, IEEE Transactions on Control Systems Technology 22 (6) (2014) 2127–2140. doi:https://doi.org/10.1109/TCST.2014.2300696.

[93] B. Lacevic, P. Rocco, Kinetostatic danger field-a novel safety assessment for human-robot interaction, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2010, pp. 2169–2174. doi:https://doi.org/10.1109/IROS.2010.5649124.

[94] A. Otto, O. Battaïa, Reducing physical ergonomic risks at assembly lines by line balancing and job rotation: A survey, Computers & Industrial Engineering 111 (2017) 467–480. doi:https://doi.org/10.1016/j.cie.2017.04.011.

[95] Robots and robotic devices – safety requirements for personal care robots (2014).

[96] Robots and robotic devices — collaborative robots (2016).

[97] Safety of machinery — positioning of safeguards with respect to the approach speeds of parts of the human body (2010).

[98] G. Charalambous, S. Fletcher, P. Webb, The development of a scale to evaluate trust in industrial human-robot collaboration, International Journal of Social Robotics 8 (2) (2016) 193–209. doi:https://doi.org/10.1007/s12369-015-0333-8

[99] K. A. Hoff, M. Bashir, Trust in automation: Integrating empirical evidence on factors that influence trust, Human factors 57 (3) (2015) 407–434. doi:https://doi.org/10.1177/0018720814547570

[100] Z. R. Khavas, S. R. Ahmadzadeh, P. Robinette, Modeling trust in human-robot interaction: A survey, in: International Conference on Social Robotics, Springer, 2020, pp. 529–541. doi:https://doi.org/10.1007/978-3-030-62056-1_44.

[101] R. Gervasi, L. Mastrogiacomo, F. Franceschini, A conceptual framework to evaluate human-robot collaboration, The International Journal of Advanced Manufacturing Technology 108 (2020) 841–865. doi:https://doi.org/10.1007/s00170-020-05363-1.

[102] 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, pp. 175–178. doi:https://doi.org/10.1109/COGSIMA.2011.5753439.

[103] A. R. Wagner, R. C. Arkin, Recognizing situations that demand trust, in: IEEE International Conference on Robot and Human Interactive Communication, IEEE, 2011, pp. 7–14. doi:https://doi.org/10.1109/ROMAN.2011.6005228

[104] R. C. Mayer, J. H. Davis, F. D. Schoorman, An integrative model of organizational trust, The academy of management review 20 (3) (1995) 709–734.

[105] R. E. Yagoda, D. J. Gillan, You want me to trust a ROBOT? the development of a human–robot interaction trust scale, International Journal of Social Robotics 4 (3) (2012) 235–248. doi:https://doi.org/10.1007/s12369-012-0144-0.

[106] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, R. Parasuraman, A meta-analysis of factors affecting trust in human-robot interaction, Human factors 53 (5) (2011) 517–527. doi:https://doi.org/10.1177/0018720811417254.

[107] G. Hoffman, Evaluating fluency in human–robot collaboration, IEEE Transactions on Human-Machine Systems 49 (3) (2019) 209–218. doi:https://doi.org/10.1109/THMS.2019.2904558.

[108] 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, pp. 51–58.

[109] S. M. Rahman, Y. Wang, Mutual trust-based subtask allocation for human–robot collaboration in flexible lightweight assembly in manufacturing, Mechatronics 54 (2018) 94–109. doi:https://doi.org/10.1016/j.mechatronics.2018.07.007.

[110] A. Meissner, A. Trübswetter, A. S. Conti-Kufner, J. Schmidtler, Friend or Foe? understanding assembly workers' acceptance of human-robot collaboration, ACM Transactions on Human-Robot Interaction 10 (1) (2020) 1–30. doi:https://doi.org/10.1145/3399433.

[111] A. H. Eagly, S. Chaiken, The psychology of attitudes, Harcourt brace Jovanovich college publishers, 1993.

[112] I. Ajzen, Nature and operation of attitudes, Annual review of psychology 52 (1) (2001) 27–58. doi:https://doi.org/10.1146/annurev.psych.52.1.27.

[113] S. W. Edison, G. L. Geissler, Measuring attitudes towards general technology: Antecedents, hypotheses and scale development, Journal of targeting, Measurement and Analysis for Marketing 12 (2) (2003) 137–156. doi:https://doi.org/10.1057/palgrave.jt.5740104.

[114] T. Nomura, T. Suzuki, T. Kanda, K. Kato, Measurement of negative attitudes toward robots, Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems 7 (3) (2006) 437–454. doi:https://doi.org/10.1075/is.7.3.14nom.

[115] L. Rosen, M. Weil, Measuring technophobia. a manual for the administration and scoring of the computer anxiety rating scale, the computer thoughts survey and the general attitude toward computer scale (1992).

[116] C. Bartneck, D. Kulić, E. Croft, S. Zoghbi, Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots, International journal of social robotics 1 (1) (2009) 71–81. doi:https://doi.org/10.1007/s12369-008-0001-3.

[117] M. Heerink, B. Kröse, V. Evers, B. Wielinga, Assessing acceptance of assistive social agent technology by older adults: the almere model, International journal of social robotics 2 (4) (2010) 361–375. doi:https://doi.org/10.1007/s12369-010-0068-5.

[118] V. Venkatesh, M. G. Morris, G. B. Davis, F. D. Davis, User acceptance of information technology: Toward a unified view, MIS Quarterly (2003) 425–478doi:https://doi.org/10.2307/30036540.

[119] A. G. Greenwald, D. E. McGhee, J. L. Schwartz, Measuring individual differences in implicit cognition: the implicit association test, Journal of personality and social psychology 74 (6) (1998) 1464. doi:https://doi.org/10.1037//0022-3514.74.6.1464

[120] T. Ninomiya, A. Fujita, D. Suzuki, H. Umemuro, Development of the multi-dimensional robot attitude scale: constructs of people's atti-

tudes towards domestic robots, in: International Conference on Social Robotics, Springer, 2015, pp. 482–491. doi:https://doi.org/10.1007/978-3-319-25554-5_48.

[121] 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.

[122] 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, pp. 242–247. doi:https://doi.org/10.1109/HUMANOIDS.2012.6651527.

[123] F. D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS Quarterly 13 (3) (1989) 319–340.

[124] V. Venkatesh, F. D. Davis, A theoretical extension of the technology acceptance model: Four longitudinal field studies, Management Science 46 (2) (2000) 186–204. doi:https://doi.org/10.1287/mnsc.46.2.186.11926.

[125] V. Venkatesh, H. Bala, Technology acceptance model 3 and a research agenda on interventions, Decision Sciences 39 (2) (2008) 273–315. doi:https://doi.org/10.1111/j.1540-5915.2008.00192.x.

[126] 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, pp. 97–103. doi:https://doi.org/10.1007/978-3-319-40548-3_16.

[127] W. MacDonald, The impact of job demands and workload on stress and fatigue, Australian Psychologist 38 (2) (2003) 102–117. doi:https://doi.org/10.1080/00050060310001707107.

[128] P. A. Hancock, Human factors psychology, Elsevier, 1987.

[129] C. D. Wickens, S. E. Gordon, Y. Liu, J. Lee, An introduction to human factors engineering, Vol. 2, Pearson Prentice Hall Upper Saddle River, NJ, 2004.

[130] J. Heard, C. E. Harriott, J. A. Adams, A survey of workload assessment algorithms, IEEE Transactions on Human-Machine Systems 48 (5) (2018) 434–451. doi:https://doi.org/10.1109/THMS.2017.2782483.

[131] K. Lindström, Psychosocial criteria for good work organization, Scandinavian journal of work, environment & health 20 (1994) 123–133.

[132] R. Parasuraman, D. Caggiano, Mental workload, in: V. Ramachandran (Ed.), Encyclopedia of the Human Brain, Academic Press, New York, 2002, pp. 17–27. doi:https://doi.org/10.1016/B0-12-227210-2/00206-5.

[133] N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, H. W. Hendrick, Handbook of human factors and ergonomics methods, CRC press, 2004.

[134] M. S. Young, K. A. Brookhuis, C. D. Wickens, P. A. Hancock, State of science: mental workload in ergonomics, Ergonomics 58 (1) (2015) 1–17. doi:https://doi.org/10.1080/00140139.2014.956151.

[135] 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, pp. 157–164. doi:https://doi.org/10.1109/HRI.2016.7451747.

[136] L. Paletta, M. Pszeida, H. Ganster, F. Fuhrmann, W. Weiss, S. Ladstätter, A. Dini, S. Murg, H. Mayer, I. Brijacak, et al., Gaze-based human factors measurements for the evaluation of intuitive human-robot collaboration in real-time, in: IEEE International Conference on Emerging Technologies

and Factory Automation, IEEE, 2019, pp. 1528–1531. doi:https://doi.org/10.1109/ETFA.2019.8869270.

[137] S. G. Hart, L. E. Staveland, Development of NASA-TLX (task load index): Results of empirical and theoretical research, in: Advances in psychology, Vol. 52, Elsevier, 1988, pp. 139–183. doi:https://doi.org/10.1016/S0166-4115(08)62386-9.

[138] D. K. Mitchell, Mental workload and ARL workload modeling tools, Tech. rep., Army Research Lab Aberdeen Proving Ground MD (2000).

[139] T. Chihara, N. Fukuchi, A. Seo, Optimal product design method with digital human modeling for physical workload reduction: A case study illustrating its application to handrail position design, The Japanese Journal of Ergonomics 53 (2) (2017) 25–35. doi:https://doi.org/10.5100/jje.53.25.

[140] N. Williams, The Borg rating of perceived exertion (RPE) scale, Occupational Medicine 67 (5) (2017) 404–405. doi:https://doi.org/10.1093/occmed/kqx063.

[141] J. Deakin, J. Stevenson, G. Vail, J. Nelson, The use of the nordic questionnaire in an industrial setting: a case study, Applied ergonomics 25 (3) (1994) 182–185. doi:https://doi.org/10.1016/0003-6870(94)90017-5.

[142] R. Melzack, S. N. Raja, The McGill pain questionnaire: from description to measurement, Anesthesiology 103 (1) (2005) 199–202. doi:https://doi.org/10.1097/00000542-200507000-00028.

[143] N. A. Stanton, P. R. Chambers, J. Piggott, Situational awareness and safety, Safety science 39 (3) (2001) 189–204. doi:https://doi.org/10.1016/S0925-7535(01)00010-8.

[144] M. R. Endsley, Situation awareness misconceptions and misunderstandings, Journal of Cognitive Engineering and Decision Making 9 (1) (2015) 4–32. doi:https://doi.org/10.1177/1555343415572631.

[145] D. B. Kaber, M. R. Endsley, Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety, Process Safety Progress 16 (3) (1997) 126–131. doi:https://doi.org/10.1002/prs.680160304.

[146] W. Karwowski, A. Szopa, M. M. Soares, Handbook of standards and guidelines in human factors and ergonomics, CRC Press, 2021.

[147] M. R. Endsley, Design and evaluation for situation awareness enhancement, Human Factors Society Annual Meeting 32 (2) (1988) 97–101. doi:https://doi.org/10.1177/154193128803200221.

[148] G. Salvendy, Handbook of human factors and ergonomics, John Wiley & Sons, 2012.

[149] A. Tabrez, M. B. Luebbers, B. Hayes, A survey of mental modeling techniques in human–robot teaming, Current Robotics Reports (2020) 259–267doi:https://doi.org/10.1007/s43154-020-00019-0.

[150] J. E. Mathieu, T. S. Heffner, G. F. Goodwin, E. Salas, J. A. Cannon-Bowers, The influence of shared mental models on team process and performance, Journal of applied psychology 85 (2) (2000) 273–283. doi:https://doi.org/10.1037/0021-9010.85.2.273.

[151] 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, pp. 33–40.

[152] S. Opiyo, J. Zhou, E. Mwangi, W. Kai, I. Sunusi, A review on teleoperation of mobile ground robots: Architecture and situation awareness, Inter-

national Journal of Control, Automation and Systems 19 (2021) 1384—-1407. doi:https://doi.org/10.1007/s12555-019-0999-z.

[153] A. Gorbenko, V. Popov, A. Sheka, Robot self-awareness: Exploration of internal states, Applied Mathematical Sciences 6 (14) (2012) 675–688.

[154] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, P. Steggles, Towards a better understanding of context and context-awareness, in: International symposium on handheld and ubiquitous computing, Springer, 1999, pp. 304–307. doi:https://doi.org/10.1007/3-540-48157-5_29.

[155] N. Dahn, S. Fuchs, H.-M. Gross, Situation awareness for autonomous agents, in: IEEE International Symposium on Robot and Human Interactive Communication, IEEE, 2018, pp. 666–671. doi:https://doi.org/10.1109/ROMAN.2018.8525511.

[156] K. Smith, P. A. Hancock, Situation awareness is adaptive, externally directed consciousness, Human Factors 37 (1) (1995) 137–148. doi:https://doi.org/10.1518/001872095779049444.

[157] 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, pp. 685–686. doi:https://doi.org/10.1109/HRI.2019.8673076.

[158] J. Muller, Enabling technologies for Industry 5.0, results of a workshop with europe's technology leaders, Tech. rep., General for Research and Innovation (European Commission) (2020).

[159] M. Funk, P. Rosen, S. Wischniewski, Human-centered HRI design - the more individual the better?. chances and risks of individualization, in: Workshop on behavioral patterns and interaction modeling for personalized Human-Robot Interaction, Fraunhofer IAO, 2020, pp. 8–11. doi:https://doi.org/10.1145/3371382.3374846.

[160] A. Y. Gao, W. Barendregt, G. Castellano, Personalised human-robot co-adaptation in instructional settings using reinforcement learning, in: IVA Workshop on Persuasive Embodied Agents for Behavior Change, 2017.

[161] T. Schürmann, P. Beckerle, Personalizing human-agent interaction through cognitive models, Frontiers in Psychology 11. doi:https://doi.org/10.3389/fpsyg.2020.561510.

[162] Y.-C. Ku, P.-Y. Li, Y.-L. Lee, Are you worried about personalized service? an empirical study of the personalization-privacy paradox, in: International Conference on HCI in Business, Government, and Organizations, Springer, 2018, pp. 351–360. doi:https://doi.org/10.1007/978-3-319-91716-0_27.

[163] S. Drolshagen, M. Pfingsthorn, P. Gliesche, A. Hein, Acceptance of industrial collaborative robots by people with disabilities in sheltered workshops, Frontiers in Robotics and AI 7 (2021) 173. doi:https://doi.org/10.3389/frobt.2020.541741.

[164] V. Alonso, P. De La Puente, System transparency in shared autonomy: A mini review, Frontiers in Neurorobotics 12 (2018) 83. doi:https://doi.org/10.3389/fnbot.2018.00083.

[165] S. El Zaatari, M. Marei, W. Li, Z. Usman, Cobot programming for collaborative industrial tasks: An overview, Robotics and Autonomous Systems 116 (2019) 162–180. doi:https://doi.org/10.1016/j.robot.2019.03.003.

[166] B. Busch, J. Grizou, M. Lopes, F. Stulp, Learning legible motion from human–robot interactions, International Journal of Social Robotics 9 (5) (2017) 765–779. doi:Learninglegiblemotionfromhuman--robotinteractions.

[167] 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, pp. 301–308. doi:https://doi.org/10.1109/HRI.2013.6483603.

[168] 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, pp. 39–46.

[169] C. Lichtenthäler, T. Lorenzy, A. Kirsch, Influence of legibility on perceived safety in a virtual human-robot path crossing task, in: IEEE International Symposium on Robot and Human Interactive Communication, IEEE, 2012, pp. 676–681. doi:https://doi.org/10.1109/ROMAN.2012.6343829.

[170] F. Dehais, E. A. Sisbot, R. Alami, M. Causse, Physiological and subjective evaluation of a human–robot object hand-over task, Applied ergonomics 42 (6) (2011) 785–791. doi:https://doi.org/10.1016/j.apergo.2010.12.005.

[171] R. Setchi, M. B. Dehkordi, J. S. Khan, Explainable robotics in human-robot interactions, Procedia Computer Science 176 (2020) 3057–3066. doi:https://doi.org/10.1016/j.procs.2020.09.198.

[172] 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, pp. 1078–1088.

[173] G. Hoffman, C. Breazeal, Cost-based anticipatory action selection for human–robot fluency, IEEE transactions on robotics 23 (5) (2007) 952–961. doi:https://doi.org/10.1109/TRO.2007.907483.

[174] J. Heard, C. E. Harriott, J. A. Adams, A human workload assessment algorithm for collaborative human-machine teams, in: IEEE International Symposium on Robot and Human Interactive Communication,

IEEE, 2017, pp. 366–371. doi:https://doi.org/10.1109/ROMAN.2017.8172328.

[175] J. Heard, R. Heald, C. E. Harriott, J. A. Adams, A diagnostic human workload assessment algorithm for collaborative and supervisory human–robot teams, ACM Transactions on Human-Robot Interaction 8 (2) (2019) 1–30. doi:https://doi.org/10.1145/3314387.

[176] M. Fujita, Y. Domae, A. Noda, G. A. Garcia Ricardez, T. Nagatani, A. Zeng, S. Song, A. Rodriguez, A. Causo, I.-M. Chen, T. Ogasawara, What are the important technologies for bin picking? technology analysis of robots in competitions based on a set of performance metrics, Advanced Robotics 34 (7-8) (2020) 560–574. doi:https://doi.org/10.1080/01691864.2019.1698463.

[177] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., ImageNet large scale visual recognition challenge, International journal of computer vision 115 (3) (2015) 211–252. doi:https://doi.org/10.1007/s11263-015-0816-y.

[178] M. Buehler, K. Iagnemma, S. Singh, The DARPA urban challenge: autonomous vehicles in city traffic, Vol. 56, Springer, 2009.

[179] E. Krotkov, D. Hackett, L. Jackel, M. Perschbacher, J. Pippine, J. Strauss, G. Pratt, C. Orlowski, The darpa robotics challenge finals: Results and perspectives, Journal of Field Robotics 34 (2) (2017) 229–240. doi:https://doi.org/10.1002/rob.21683.

[180] A. Causo, J. Durham, K. Hauser, K. Okada, A. Rodriguez, Advances on Robotic Item Picking: Applications in Warehousing & E-Commerce Fulfillment, Springer, 2020.

[181] K. Wada, New robot technology challenge for convenience store, in: IEEE/SICE International Symposium on System Integration, IEEE, 2017, pp. 1086–1091. doi:https://doi.org/10.1109/SII.2017.8279367.