Evaluation of User Experience in Human–Robot Interaction: A Systematic Literature Review

Ainhoa Apraiz [0000-0002-9426-4546], Ganix Lasa [0000-0002-2424-5526] and Maitane Mazmela [0000-0003-0848-2131]

Mondragon Unibertsitatea. Faculty of Engineering, Mechanical and Industrial Production, Loramendi 4, Mondragon 20500 Gipuzkoa (Spain)
aapraizi@mondragon.edu

Abstract. Industry 4.0 has ushered in a new era of process automation, thus redefining the role of people and altering existing workplaces into unknown formats. The number of robots in the manufacturing industry has been steadily increasing for several decades and in recent years the number and variety of industries using robots have also increased. For robots to become allies in the day-to-day lives of operators, they need to provide positive and fit-for-purpose experiences through smooth and satisfying interactions. In this sense, user experience (UX) serves as the greatest link between persons and robots. Essential to the study of UX is its evaluation. Therefore, the aim of this study is to identify methodologies that evaluate the human–robot interaction (HRI) from a human-centred approach. A systematic literature review has been carried out, in which 24 articles have been identified. Among these, are 15 experimental studies, in addition to theoretical frameworks and tools. The review has provided insight into how evaluations are conducted in HRI. The results show the most evaluated factors and how they are measured considering different types of measurements: qualitative and quantitative, objective and subjective. Research gaps and future directions are correspondingly identified.

Keywords: Human–robot Collaboration (HRC), Human–robot Interaction (HRI), User Experience (UX), Systematic Literature Review (SLR), Human factors.

1 Introduction

Industry 4.0 has ushered in a new era of process automation, thus redefining the role of people and changing existing workplaces into unknown formats [1]. The number of robots in the manufacturing industry has been steadily increasing for several decades and in recent years the number and variety of industries using robots have also increased [2]–[4]. In this context, operators will continue to be of great importance, so optimising the interactions between persons and robots will be crucial. In contrast to standard automation, collaborative robots (cobots) [5] enable close and safe interactions between humans and machines, taking advantage of the benefits of both sides.

ISO 8373 [6] defined a robot as a powered mechanism controlled via an interface, it is programmable in two or more axes with a degree of autonomy and moves within its environment to perform intended tasks. Dautenhahn [7, sec. 38.2] defined human–robot interaction (HRI) as 'the science that studies people's behaviour and attitudes towards robots in relation to the physical, technological, and interactive characteristics of robots, with the aim of developing robots that facilitate the generation of human–robot interactions that are at the same time efficient (in accordance with the original requirements of their intended area of use), acceptable to people, meet the social and emotional needs of their individual users, and respect human values'.

For robots to become allies in the day-to-day lives of operators, they need to provide positive and fit-for-purpose experiences through smooth and satisfying interactions [8]–[11]. In this sense, the user experience (UX) serves as the greatest link between persons and robots. ISO 9241-210 [12, sec. 2.15] defined UX as 'a person's perceptions and responses resulting from the use or anticipated use of a product, system or service'. This includes user emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviours, and achievements that occur before, during, and after use [12]. This means that humans must experience robots as fulfilling existing goals, and as entities that act efficiently and make people feel confident, safe, and comfortable while they are working together [13]. A clearer understanding of social cognitive constructs (such as determining intentionality, which suggests an intimate connection between social cues and the perception of robots as social agents) is required to fully optimise HRI [14]. This statement emerges from a shift in the perception of robots as tools that extend human capabilities to teammates that collaborate with people [15]–[18].

Over the next few years, the coexistence between people and robots will increase [19]. This will take place in technologically enriched environments, where information will be exchanged "*naturally*" between humans and robots, giving rise to hybrid environments in which people move between the digital and real worlds [19]. The combination of human and robotic skills is becoming increasingly important [20]. While certain routine tasks or specific skills can be effectively supported by automation, local decisions or exceptional interventions often require human input. This could arise from the extraordinary characteristics of the given situation or the complexity or the implicit nature of the knowledge required to find a feasible solution within a limited period. To date, the combination of human and artificial resources has not been part of standard automation practice, in which i) robots and people are usually kept at arm's length from each other, and ii) people must adhere to work procedures that are as rigid as the rest of the automated production environment. Symbiotic human–robot collaboration (HRC) goes beyond these constraints and requires a more responsive, transparent, and accessible environment. Thus, for the improvement of HRI, the skills and expertise of humans must be combined with the accuracy and automation of robots, which work not as passive tools but as active partners [21].

To this end, it is important to optimise the UX between human and robot. The evaluation of the UX will enable the continuous improvement of the industry's workplaces.

1.1 Research background on HRI design and evaluation

Numerous contributions have been written on HRI design and evaluation. The recent adoption of the concept industry 5.0 by the European Commission [22], increased the interest to incorporate human factors. Nevertheless, the literature reports few attempts to put human factors metrics in a comprehensive way in order to evaluate the UX on HRI.

A method for performing detailed ergonomic assessments of co-manipulation activities exists, and this could be applied to optimise the design of collaborative robots [23]. Maurice et al. [23] defined multiple ergonomic indicators to estimate different biomechanical demands (muscle force, tendon deformation, muscle fibre length…) that occur during whole-body activities (*e.g.,* joint loads, joint dynamics, mechanical energy…). These indicators are measured through virtual human simulations.

Amoretti et al. [24] stated that understanding the characteristics, advantages and disadvantages of different technical architecture paradigms and software strategies for their use in the robotics domain is crucial for the design, implementation, and successful use of cobotic software architectures.

There are several literature reviews in the context of HRI. The work by Hentout et al. [25] proposes a rough classification of the content of works in HRI into several categories and subcategories, such as hardware and software design of collaborative robotic systems, safety in industrial robotics and cognitive HRI. They stated that the goal of HRI is to provide robots with three fundamental requirements: i) human intention should be easy to infer by the robot, ii) the control should be intuitive from the human viewpoint, and iii) the designed controller should be safe for both humans and robots. Simões et al. [21] listed a number of guidelines broadly classified into: i) human operator and technology, ii) human–robot team performance, and iii) an integrated approach to design HRC. As a generic conclusion, they highlighted the importance of feedback in improving trust and blame attribution. They presented recommendations for the design of safe, ergonomic, sustainable, and healthy human-centred workplaces where not only technical but also social and psychophysical aspects of collaboration are considered. Savela et al. [26] examined how the social acceptance of robots in different occupational fields had been studied and what kinds of attitudes the studies had discovered regarding robots as workers. Their results imply that attitudes toward robots are positive in many fields of work. Nevertheless, they indicated that there is a need for validated measures. Veling et al. [27] analysed the use of qualitative methods and approaches in the HRI literature to contribute to the development of a foundation of approaches and methodologies in the research area. Their review revealed six predominant qualitative data gathering methods in the HRI literature: qualitative observations, semi-structured interviews, focus groups, generative activities, reflective and narrative accounts, and textual/content analysis.

According to Moulières-Seban et al. [28], focusing on humans, tasks, robots, and system interactions when designing a cobotic system is necessary. These authors introduced a method of designing cobotic systems that is composed of four stages: i) activity analysis, ii) basic analysis, iii) detailed design, and iv) realisation, setup, validation and putting into service.

Numerous studies on robots in industries have been published, but most of them focus on safety and security aspects [25], [29], [30]. Other researchers have studied standardisations to improve workplaces [31], [32]. In this sense, the robotics industry is growing to a level where people and robots will be able to collaborate [33]. However, as Harriot et al. [33] pointed out, there is still no universal model that assesses the effect of this collaboration on people's performance.

Furthermore, it is noted that no attention has been paid in the literature to the human factors resulting from the human-robot interaction. Emotional factors such as trust, satisfaction or mental workload have been poorly studied for the optimisation of collaborative robotic systems. The assessment of these factors is beneficial to know how people feel before, during and after the interaction. In this way, robot actions could be adapted to people's needs, in line with the human-centred design approach.

The direction indicated by European Commission's framework Horizon 2020 establishes that gender equality must be promoted through changes in the culture of scientific institutions, changes in the composition of research teams in order to achieve equality, and changes in the content and design of research activities [34]. Nevertheless, the gender perspective has not been included in any of the identified studies. The integration of gender perspective in research is necessary to avoid biases, where sometimes the realities, experiences, and expectations of a group of people (considering men as a reference) are constructed as the norm, thus producing partial and non-universal results. The gender perspective in research means integrating sex and gender variables in the scientific process, which will have implications when considering gender norms, identities, and relations as explanatory variables of the analysed phenomenon. On the other hand, many people face accessibility barriers when interacting with robots, mainly people who do not usually interact with new technologies, elderly people, and users with disabilities [35]. Designing and developing robotic systems which ensure accessibility to all users with different abilities and needs is essential to make HRI systems more inclusive. Nevertheless, none of the identified reviews considers the inclusivity as a necessary aspect to be approach.

Therefore, the objective of this systematic literature review (SLR) is to identify evaluations of HRI that include a human-centred design perspective. Furthermore, we aim to understand the human factors that affect HRI in an industrial environment.

The organisation of this review paper is as follows. Section 2 provides information about HRI and interaction types. Additionally, it presents information about UX evaluation. Section 3 explains the search methodology used in this article. Section 4 details the results, which are the literature characterisations and the answers to the research questions (RQs). Section 5 presents the discussion, and Section 6 the research gaps and future research directions. In Section 7 the limitations are set out, and, finally, Section 8 presents the summary and conclusions.

2 Human–Robot Collaboration and UX evaluation

HRC implies a deeper interaction between the two entities involved (i.e., humans and robots). In this context, interfaces play central roles as the main communication

channels. A key aspect of collaboration is interaction and talking about interactions also means talking about interfaces. High-quality HRI requires intuitive user interfaces [36]. On the one hand, operators can give robots simple inputs without any distraction from their main tasks. On the other hand, robots provide clear information to users, resulting in an immediate understanding and interpretation of data [19]. The adoption of intuitive interfaces becomes even more important in the case of closer collaborations between robots and humans. Humans naturally interact with the world using multiple resources simultaneously [37]. Consequently, interacting with cobotic systems should be easy for them [25].

Establishing what effective communication entails and determining the interfaces through which humans and robots can communicate are necessary. In this regard, we should define i) the intended interactions between persons and robots, and ii) the purpose of the information exchange. Both elements are largely outlined by the scope of the application and the functions of humans and robots [38], and they need to be adapted to different contexts.

Interfaces can generate different types of interactions [39]. For example, graphical communication can take place using specific devices (e.g., a monitor or a touch screen), voice-based communication can use natural language interfaces and gesture-based communication can use cameras suitable for tracking human hands. Depending on the typology of communication, human–robot interfaces can be classified into four categories: i) visual displays (e.g., graphical user interfaces and augmented reality [AR] interfaces), ii) gestural (e.g., hand and face movements), iii) voice and natural language (e.g., auditory and text-based responses) and iv) physical and haptic interactions [40].

HRI has been classified into different areas depending on the authors. Prati et al. [39] used the classification by Schmidtler et al. [41], who categorised HRI into: i) human–robot coexistence, ii) human–robot cooperation and iii) human–robot collaboration (HRC). According to Prati et al. [39], these interfaces can also be related to the level of interaction provided. In particular, the first level of interaction (coexistence) is usually satisfied with graphical interfaces. The second level (cooperation) often requires more advanced interfaces, such as voice and gestures. Finally, the third level (collaboration) may require direct physical or haptic interaction to be both effective and natural.

On the other hand, Wang et al. [42] studied the relationship between interaction and risk. At the first level (coexistence), commitment and complexity are low, and safety is easy to guarantee because the operator is protected by physical boundaries. In the second level (interaction), the interaction is not significantly higher, but the safety risk is much higher, as the person and the robot start to share the same space. At the human–robot cooperation level, the interaction and the safety risk increase significantly, as direct contact between the operator and the robot is high. In a fully symbiotic human–robot partnership, the task is carried out by both parties in a collaborative manner, and it is inevitable that the human operator comes into direct contact with the robot. Therefore, the level of safety risk is higher than that in cooperation.

As indicated by Wang et al. [20], the development of solutions for HRC requires an analysis and synthesis framework containing i) means to classify and characterise the problem and ii) solution templates and guidelines for developing a solution that fits seamlessly into existing production requirements. The authors identified the

fundamental elements in an HRC scenario as i) agents (robots and humans actively participating in the production process), ii) the working environment, which includes resources that are necessary for production but do not play an active role in conditions (ambient light, noise, etc.), and iii) parts and operations.

UX is a term that has become established in human-computer interaction research and practice. It denotes that the interaction with a contemporary technological system goes beyond usability and extends to the emotions before, during and after using the system. UX cannot be defined solely by studying the fundamental attributes of usability, such as effectiveness, efficiency, and user satisfaction. Measuring UX becomes a more complicated task when the target of the interaction is not just a technology system or an application but an entire environment.

UX is a key factor in the quality of a product, service or system [43]–[45]. Essential to its study is its evaluation, which refers to the application of a set of methods and instruments whose objective is to determine the perception of the use of a system or product, allowing the identification of aspects to improve or maintain [46].

In a study on "User eXperience Evaluation Methods" by Väänänen-Vainio-Mattila et al. [47], they proposed a set of requirements for good UX evaluation in industrial environments. Although they stated that it is not possible to have a single method that meets all the requirements because some of them may be contradictory or even unrealistic, it would be interesting to identify the different evaluation methods to assess HRI in the literature.

3   Research methodology

As stated above, this literature review is about identifying evaluations of HRI that include a human-centred design perspective and understanding the human factors that affect HRI in an industrial environment. This can be achieved by performing a SLR and identifying all the available research papers within a specific duration.

In this literature review method, the guidelines proposed by Kitchenham et al. [48] were used to carry out the SLR, which included three phases: planning, organising, and documenting. These phases have their own components: (1) research questions, (2) data/information sources, (3) criteria for the inclusion/exclusion of selected papers, (4) quality assessment (QA), (5) systematic review strategy and, (6) extraction of data and synthesis. The following sections describe these phases.

3.1   Research questions

In this SLR, two RQs were formulated which are given as follows: The questions must be clearly answered to complete the SLR successfully. Table 1 presents the RQs.

Table 1. Defined Research Questions.

ID	Research Questions
RQ-1	Is there a human–robot interaction assessment model that includes human factors in industrial settings?
RQ-2	What human factors does it include, and how does it assess them?

3.2 Data/Information sources

There are three main groups of keywords. Scholars use varying terms to describe the concepts. Therefore, a range of keywords were identified and combined to discover different studies comprehensively and objectively. The following terms were used in the search for information:

"Human–robot Interaction" OR "Human–robot Collaboration" OR "Human–robot Coexistence" OR "Human–robot Workstation"

"Human factors" OR "user experience" OR "ux"

Evaluation OR Assessment

In this way, the following main search equation was created:

("Human–robot Interaction" OR "Human–robot Collaboration" OR "Human–robot Coexistence" OR "Human–robot Workstation") AND ("Human factors" OR "user experience" OR "ux") AND (Evaluation OR Assessment)

The electronic databases used for the search are shown in Table 2.

Table 2. Description of research databases.

Data bases	Type of data base	Description
ACM Digital Library	Research database	A scientific database on subjects related to informatics and computer science.
Engineering Village	Research database	A specialised engineering database.
IEEE Xplore Digital Library	Research database	A scientific database on subjects related to computer science, electrical engineering, and electronics.
Inspec	Specialised database	A database specialising in physics, electrical and electronic engineering, computer and control, information technology, and mechanical and production engineering.
Science Direct	Publisher's database	Offers its own full-text scientific journals. Covers most disciplines, but mainly focuses on science, technology and social sciences and related publishers, Elsevier
Scopus	Citation database	Scientific citation indexing service for citation searching of peer-reviewed journal articles. It is mainly used bibliometric calculations, Elsevier.
Web of Science	Citation database	Scientific citation indexing service for citation searching of peer-reviewed journal articles. It is mainly used in bibliometric calculations, Thomsson/Reuter.

3.3 Literature Search

Each database was searched, adapting the equation as required by the database. One problem with this breadth of databases is the noticeable difference in their search functionality, which require adjustment according to each database, as detailed in Table 3. All articles must meet these general requirements: peer-reviewed journal articles dated between January 2011 and the date this search was conducted (November 2021). Using peer-reviewed journal articles ensured validated knowledge [49],while the publication year limit was set to reduce the number of inappropriate hits. We did not expect to find any articles before 2011 that were significant for the review because it is an emerging field and the interest in human factors is also recent. Moreover, we assumed that the latest work builds on that of previous years. We also excluded papers that were not mainly written in English or Spanish.

Table 3. Adopted search syntax for each database and number of results obtained.

Nº	Database	Date search was carried out	Search equation	No. of results
1	ACM	05/11/2021	[[All: "human-robot interaction"] OR [All: "human-robot collaboration"] OR [All: "human-robot coexistence"] OR [All: "human-robot cooperation"] OR [All: "human-robot workstation"]] AND [[All: "human factors"] OR [All: "user experience"] OR [All: ux]] AND [[All: "evaluation"] OR [All: "assessment"]] AND [Publication Date: (01/01/2011 TO 12/31/2021)]nt")	67
2	Engineering Village	12/11/2021	(((((("Human-Robot Interaction" OR "Human-Robot Collaboration" OR "Human-Robot Coexistence" OR "Human-Robot Cooperation" OR "Human-Robot Workstation") AND ("Human factors" OR "user experience" OR ux) AND ("Evaluation" OR "Assessment")) WN ALL)) AND (({ja} WN DT) AND ((2021 OR 2020 OR 2019 OR 2018 OR 2017 OR 2016 OR 2015 OR 2014 OR 2013 OR 2012 OR 2011) WN YR)))	29
3	IEEE Xplore	12/11/2021	("Human-Robot Interaction" OR "Human-Robot Collaboration" OR "Human-Robot Coexistence" OR "Human-Robot Cooperation" OR "Human-Robot Workstation") AND ("Human factors" OR "user experience" OR ux) AND ("Evaluation" OR "Assessment")	20
4	Inspec	13/11/2021	("Human-Robot Interaction" OR "Human-Robot Collaboration" OR "Human-Robot Coexistence" OR "Human-Robot Cooperation" OR "Human-Robot Workstation") AND ("Human factors" OR "user experience" OR ux) AND ("Evaluation" OR "Assessment")	53

5	Sci-enceDi-rect	13/11/2021	5.1 ("Human-Robot Interaction" OR "Human-Robot Collaboration") AND ("Human factors" OR "user experience") AND (evaluation OR assessment) 5.2 ("Human-Robot Coexistence" OR "Human-Robot Cooperation" OR "Human-Robot Workstation") AND ("Human factors" OR "user experience" OR ux) AND ("Evaluation" OR "Assessment")	296
6	Scopus	13/11/2021	TITLE-ABS-KEY (("Human-Robot Interaction" OR "Human-Robot Collaboration" OR "Human-Robot Coexistence" OR "Human-Robot Cooperation" OR "Human-Robot Workstation") AND ("Human factors" OR "user experience" OR ux) AND ("Evaluation" OR "Assessment")) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012) OR LIMIT-TO (PUBYEAR , 2011)) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English"))	51
7	Web of Science	13/11/2021	("Human-Robot Interaction" OR "Human-Robot Collaboration" OR "Human-Robot Coexistence" OR "Human-Robot Cooperation" OR "Human-Robot Workstation") AND ("Human factors" OR "user experience" OR ux) AND ("Evaluation" OR "Assessment")	39

There are 555 identified papers, nearly the half of which were found in the ScienceDirect database, whose disciplines focus on science, technology, and the social sciences.

3.4 Selection of Literature

The next step continued with the review protocol. The main motivation for applying inclusion and exclusion criteria was to ensure that the studies selected for the systematic review were related to the evaluation of HRI taking into account human factors.

Criteria for inclusion/exclusion. Table 4 shows the criteria used in this review process. In addition to the language limitation (LL), we also ensured the credibility of the published papers by excluding journal articles that were not peer-reviewed (LP1). Another limitation was the publication year, which was set to reduce the number of inappropriate hits (LP3). Therefore, the first step was to exclude duplicate articles (LP2). A total of 117 duplicate articles were identified.

The next step consisted of literature selection based on the article title and abstract and taking into account the directly related (DR), partially related (PR) and loosely related criteria (LR). A total of 331 articles were excluded.

Table 4. Inclusion and Exclusion Criteria.

I/E	Criteria	Code	Description
Inclusion	Directly Related	DR	A paper responds to both research questions.
	Partially Related	PR	A paper responds to at least one of the research questions.
Exclusion	Limited language	LL	A full text is not mainly written in English or Spanish
	Limited publication	LP1	A paper is not published as a journal article in the databases studied.
		LP2	A paper is duplicated on the different databases.
		LP3	A paper is not published between 2011 and 2021.
	Loosely related	LR	A paper does not respond to any research question.

Quality assessment (QA). The next step is to conduct a QA. This process allowed us to identify whether the articles were related to the specific topic being reviewed and whether they were useful when considering an evaluation of the UX in an industrial robotic environment. For this purpose, five QA questions were formulated, reviewed in detail and scored based on the analysis. Because of the nature of this type of experimentation, sample size or replicability has not been established as a crucial aspect in the QA. There is a large variability in the sample size used in the different experiments because of the different application protocols that may exist [50]. Considering that QA does not exist in isolation, but directly or indirectly serves to answer RQs and support conclusions in this study [51], our RQs are merely associated with the human factors in HRI, ways to evaluate it and the consideration of gender and inclusiveness. Therefore, we defined the QA questions by following the recommendations by Yang et al. [51], resulting in the following ones:

- QA-1 - *Does the proposed topic relate to human factors in a human-robot interaction in an industrial environment?* QA-1 aims to give higher scores to articles related to human factors in HRI and specifically in industrial settings.
- QA-2 – *Does this research help identify human factors that affect human–robot interaction?* QA-2 aims to value papers which at least assist in the identification of human factors influencing HRI, those that are most aligned with the human-centred design approach, and therefore with the purpose of this research.
- QA-3 – *Is the proposed topic adequately described?* QA-3 has been established to assess quality and rigorous articles, which are well written, and the subject is described correctly.
- QA-4 – *Does the proposed theme consider a gender or inclusive perspective?* Directly aligned with the scope of this research, QA-4 aims to value papers that consider the gender perspective and inclusiveness.

- QA-5 – *Does the research describe how to evaluate the user experience of human-robot interaction?* QA-5 aims to assess those papers that present experiments where HRI evaluations are carried out from a UX perspective.

The five QA questions mentioned help in evaluating the selected studies in terms of their contributions to the present literature review. The aim of the QA was to facilitate the understanding of the studies' appropriateness and usefulness to the current study. Nidhra et al. [52] proposed high-level quality criteria by providing specific scores for the findings, which consisted of three types of ratings for the assessment: high, medium, and low. These ratings are given by answering QA questions.

A score of 2 was given to studies that fully met the quality standard, a score of 1 was given to studies that partially met the quality standard, and a score of 0 was given to the studies that did not meet the quality standard. Therefore, the maximum score for each study is 10 (i.e., $5 \times 2 = 10$), and the lowest possible score is 0 (i.e., $5 \times 0 = 0$).

In this SLR, we considered those articles that obtained a score higher than 7 which is a quite reasonable result to answer the RQs of this study and to ensure a high quality and reliable findings (Table 5), which was a total of 24 articles.

Table 5. Papers that obtained a quality score higher than 7.

Score	Nº of articles	References
9	8	[11], [53]–[56], [57], [58], [59]
8	16	[60], [39], [33], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73]

Systematic Process Review. Fig. 1 describes the process carried out during the literature review. In the first phase, a total of 555 articles were identified, of which 117 were duplicates and were therefore discarded. In the second phase, screening was carried out by reading the titles and abstracts based on the previously defined inclusion and exclusion criteria (Table 4). A total of 438 titles and abstracts were reviewed, of which 331 were discarded. The next phase consisted of a complete reading of the 107 remaining articles, which were evaluated one by one based on the QA questions. After evaluation, only articles scoring more than 7 were considered, of which 24 articles were finally analysed in depth for data extraction and synthesis.
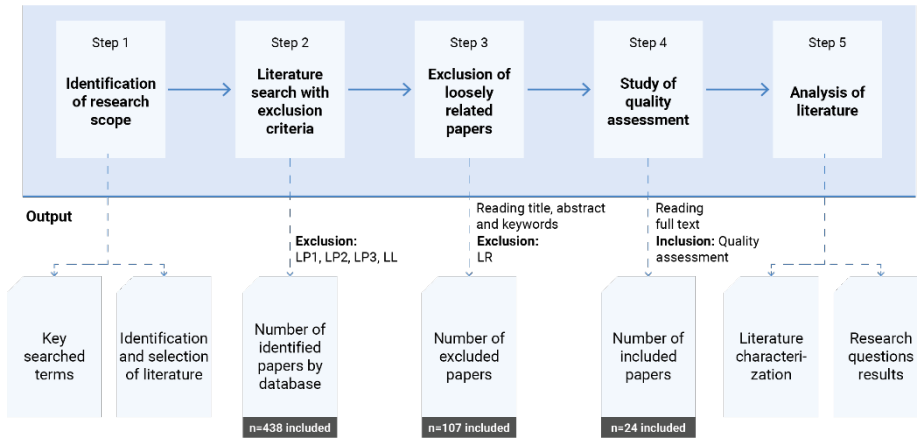
Fig. 1. Diagram of the data collection process, according to the guidelines [48].

4   Results

4.1   Literature Characterisation

Evolution in the field. The first article identified dates from 2013. Three articles have been identified for that year: they concern the studies of [33], [53] and [54]. It is only in 2017 that publications studying the evaluation of human factors in industrial human–robot environments begin to increase. In fact, more than 80% of the identified publications date from 2017 to 2021 (Fig. 2).
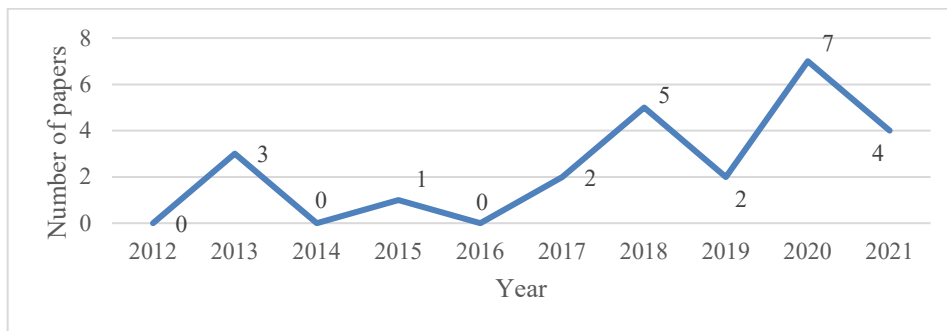
Fig. 2. Number of articles identified per year. Note that the search only includes articles published before mid-November 2021 and thus does not include all 2021 publications.

Nature of journals. The journals in which the most articles have been identified, with three articles in each, are Procedia CIRP, a journal focused on publishing high quality proceedings of CIRP conferences and, ACM Transactions on Human–Robot Interaction (THRI). Two articles have been identified in the IFAC PapersOnLine journal, two

in Procedia Manufacturing, and two in Robotics and Computer Integrated Manufacturing (Fig. 3). Journals in which most articles have been identified

Fig. 3. Journals in which most articles have been identified.

As for the impact of the publications, Table 6 shows that only one journal is not indexed. In other words, 96% of the articles belong to indexed journals. Of the 17 journals identified, 5 of them (29%) are classified in the first quartile, and 9 (53%) in the second quartile.

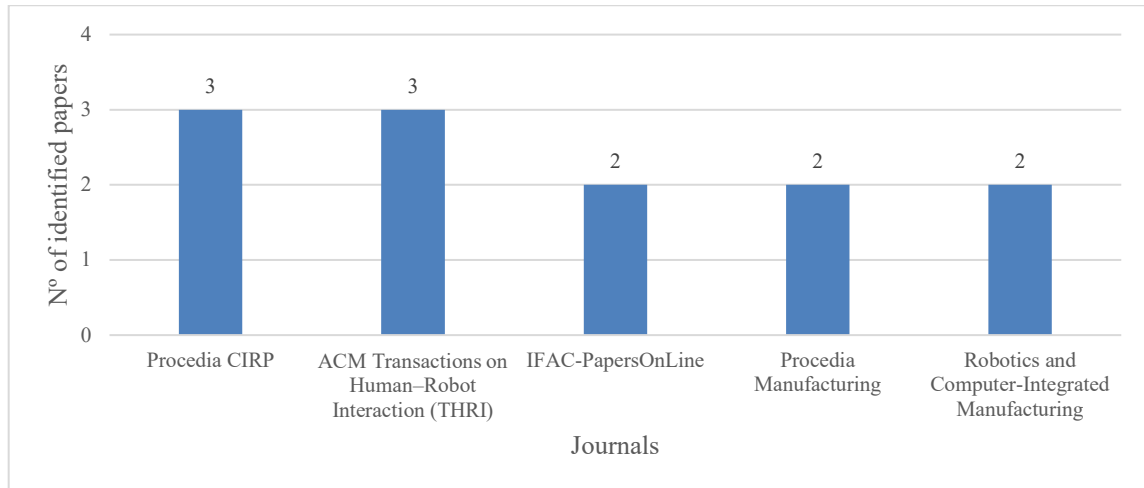**Table 6.** Impact of the identified journals (nd=no data, SJR=Scimago Journal Rank).

Journals	Quartile	SJR 2020
ACM Transactions on Human-Robot Interaction (THRI)	Q2	0,6
Applied Sciences	Q2	0,44
Electronics	Q2	0,36
Human Factors	Q1	0,82
IFAC-PapersOnLine	Q3	0,31
International Journal of Advanced Manufacturing Technology	Q1	0,95
International Journal of Human-Computer Studies	Q2	0,73
International Journal of Social Robotics	Q1	0,66
Journal of Robotics	Q2	0,3
Mechatronics	Q1	0,94
Modeling, Identification and Control	Q4	0,22
Procedia CIRP	nd	0,68
Procedia Manufacturing	Q2	0,5
Proceedings of the ACM on Human-Computer Interaction	Q2	0,38
Robotics and Computer-Integrated Manufacturing	Q1	1,56
Sensors	Q2	0,64
Sustainability	Q2	0,61

Number of citations per article. The number of citations in the articles is relatively low compared with that in broader fields. The most cited article is that by Villani et al. [63] with 246 citations. The paper by Lasota et al. [55] is in second place with 97 citations. With 31 citations, the third most cited article is that by Hietanen et al. [68]. Table 7 lists the 10 most cited articles included in this review.

Table 7: Papers by citations, retrieved from December 2021.

Ref.	Title	Authors	Year	Nº of citations (December 2021)
[63]	Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications	Villani, V., Pini, F., Leali, F., Secchi, C.	2018	246
[55]	Analyzing the Effects of Human-Aware Motion Planning on Close-Proximity Human–Robot Collaboration	Lasota, P., Shah, J.	2015	97
[68]	AR-based interaction for human–robot collaborative manufacturing	Hietanen, A., Pieters, R., Lanz, M., Latokartano, J., Kämäräinen, J.	2020	31
[53]	Evaluating the Effect of Saliency Detection and Attention Manipulation in Human–robot Interaction	Schillaci, G., Bodiroza, S., Hafner, V.	2013	22
[61]	The development of a Human Factors Readiness Level tool for implementing industrial human–robot collaboration	Charalambous, G., Fletcher, S., Webb, P.	2017	19
[56]	Assessing Instructions in Augmented Reality for Human–robot Collaborative Assembly by Using Demonstrators	Danielsson, O., Syberfeldt, A., Brewster, R., Wang, L.	2017	14
[65]	The Design and Evaluation of an Ergonomic Contactless Gesture Control System for Industrial Robots	Tang, G., Webb, P.	2018	10
[58]	Empowering assembly workers with cognitive disabilities by working with collaborative robots: a study to capture design requirements	Kildal, J., Martín, M., Ipiña, I., Maurtua, I.	2019	9
[66]	Digital Human and Robot Simulation in Automotive Assembly using Siemens Process Simulate: A Feasibility Study	Baskaran, S., Niaki, F., Tomaszewski, M., Gill, J., Chen, Y., Jia, Y., Mears, L., Krovi, V.	2019	9
[33]	Assessing physical workload for human–robot peer-based teams	Harriott, C., Zhang, T., Adams, J.	2013	8

4.2 RQ-1: Is there a HRI assessment model that includes human factors in industrial settings?

Although no widely validated assessment model has been identified, 24 papers in which assessment of HRI in industrial settings is carried out have been determined. Five of these studies are theoretical frameworks, 5 are tools and 15 are experimental studies (Table 8). The study by Prati et al. [39] first showed the theoretical framework and then the tools. This is why the article appears in both groups, making a total of 25; however, there are only 24 articles.

Table 8: Classification of the identified papers taking into account the type of study.

Type of study	Nº of articles	References
Theoretical frameworks	5	[11], [39], [59], [63], [64]
Tools	5	[39], [57], [61], [69], [73]
Experimental studies	15	[33], [53]–[56], [58], [60], [62], [65]–[68], [70]–[72]

Theoretical frameworks. This SLR has identified five recently created theoretical frameworks. The studies by Cohen et al. [64] and Villani et al. [63] date from 2018. The studies by Meissner et al. [59] and Lindblom et al. [11] are from 2020, and that by Prati et al. [39] is from 2021. As these are recent theoretical frameworks, it could be said that interest in the field is growing. In addition, the high impact of the journals included shows the acceptance and interest of the scientific community of HRI. Specifically, the study by Villani et al. [63] has 246 citations (as of December 2021).

In the study by Cohen et al. [64], they proposed a theoretical framework for analysing and improving workplaces, the framework focuses on three phases: observation, analysis, and reaction. They emphasised the importance of examining the inputs, whether from the operator or the workplace itself, analysing them and selecting how the reaction must take place on that basis. According to Schillaci et al. [53] and Kildal et al. [58], providing the right feedback is important for the interaction to be perceived satisfactorily by users. In this regard, Cohen et al. [64] underlined the selection of the reaction mode (i.e., the channel, frequency and intensity by which feedback should be given) based on observation and analysis of the different elements involved.

Villani et al. [63], on the other hand, placed safety at the centre of the system. According to Bo et al. [74] and Hentout et al. [25], safety is the foremost consideration in HRI. Industrial cobots interact and perform tasks with humans, creating close ties between the two. However, this close relationship changes the current paradigm regarding safety procedures and separation in workspaces between humans and robots [75]. The safety of working with cobots is a challenge today. Reducing the weight of their moving parts is one of the main factors to be considered when designing intrinsically safe cobots [76]. The sensorial apparatuses of robots could be improved [25], such as, using proximity sensors, to reduce risk during interactions.

To this end, the importance of designing intuitive interfaces has been emphasised. The information provided by robots should be adequate for users to be aware of the

situation, understand the behaviour of the system and thus intervene in dynamic and unexpected situations. Villani et al. [63] stated that affective robotics could be suitable for guaranteeing an intuitive interface, alleviating the cognitive load of the user, as the robot would adapt to the person's situation. Cohen et al. [64], Schillaci et al. [53] and Kildal et al. [58] reported that providing adequate feedback is indispensable in establishing bidirectional person-robot communication. Furthermore, Villani et al. [63] stressed the importance of having adequate design methods and introducing adaptive solutions for inclusive robotics.

Beyond feedback, the theoretical framework of Meissner et al. [59] showed the factors that influence worker acceptance in HRC contexts. They indicated that the most influential factors (the primary ones) are perceived risk, perceived benefits, and positive and negative emotions. They also pointed out that a number of secondary factors affect the acceptance process, such as object-related, subject-related and context-related factors. All these factors influence people's attitudes towards the acceptance of the system [59].

In relation to system acceptance, the study by Lindblom et al. [11] was based on Donald Norman's seven-stage action model [77], [78]. In the model, the person starts from an intention, which is executed and subsequently has consequences, such as perception, interpretation, and evaluation of the context. In this sense, designers develop systems with specific intentions for users. However, this does not always happen, and people's resulting emotions may differ from the intention with which the system is designed. Along these lines, Lindblom et al. [11] constructed the ANEMONE theoretical framework. It consists of a phased, iterative procedure that focuses on i) determining whether people can perceive, understand, and predict robots' intentions and actions and on ii) providing relevant insight into why something works or does not work in a particular use situation. The goals are to provide guidance on how UX evaluation can be conducted and to facilitate an understanding of why something does or does not work by identifying UX issues.

Prati et al. [39] proposed a structured UX-oriented method to investigate human–robot dialogue. The method aims to introduce a set of UX techniques that support interface design. In accordance with the human centred design approach, the method places the user at the centre, it is also an iterative process. The first step of the method consists of requirements gathering, this involves a multidisciplinary team, user analysis (for which a set of tools is proposed), activity analysis and interaction visualisation. The second step consists of interface design, subsequent prototyping and, finally, UX evaluation. For the latter, the authors proposed user testing, but they did not present any process, technique, or tool to carry out the evaluation.

Operational tools. This SLR has also identified five tools. The five articles showing the tools are fairly recent. The study by Charalambous et al. [61] dates from 2017; Von Der Pütten et al. [57], from 2018; Gualtieri et al. [69], from 2020; and Qbilat et al. [73] and Prati et al. [39] from 2021.

Charalambous et al. [61] proposed a system to determine industrial maturity level. The goal is to develop a new human actor readiness level tool for system design practitioners and thus optimise the successful implementation of industrial HRC.

Von Der Pütten et al. [57] developed and validated a new measure of self-efficacy in HRI. After conducting several experimental studies, they proposed a questionnaire consisting of 18 items. Participants have to rate the items on a six-point Likert scale [79].

In the study by Gualtieri et al. [69], a collection and classification of prerequisites and design guidelines were developed. These guidelines could help application designers properly develop and evaluate safe, people-centred, and efficient collaborative assembly workstations. Qbilat et al. [73] proposed HRI accessibility guidelines. These guidelines were evaluated by 17 HRI designers and/or developers. The authors developed a questionnaire consisting of nine five-point Likert-scale questions and six open-ended questions to evaluate the proposed guidelines for developers and designers in terms of four main factors: usability, social acceptance, UX and social impact.

Prati et al. [39] presented two design tools in addition to the theoretical framework. The first is the user/task matrix, which is used to synthesise in a chart all the information about users and tasks, as well as operational conditions, this tool helps designers define suitable interfaces. The second is experience maps, which represent a synthetic visualisation of the entire end-to-end experience that a "*generic*" user goes through to achieve a given goal. These maps are used to understand general human behaviour, as opposed to journey maps, which are more specific and focused on aspects related to a specific business.

Experimental studies. The SLR has also identified 15 experimental studies evaluating HRI. Interest in experimental studies is growing, as can be seen in the results, more than 50% of the studies were carried out in the last three years (2019-2021). Furthermore, the high impact of the journals shows the scientific community's interest in the field.

In general terms, the experiments are divided into three phases: i) prior to the execution of the task, ii) during the execution of the task and iii) after the execution of the task. However, no validated evaluation model has been identified, as each of the experimental studies uses a different process. A comparison is shown in Table 10.

Regarding the phase prior to task execution, none of the studies have conducted an expert evaluation using tools, such as heuristics. Nor are data collected from expert's or user's perspectives. It is important to note that experience consists of the emotions before, during and after an interaction, as stated in the definition of UX in the ISO 9241-210 [12].

In the phase during execution, most studies have used robots of various kinds, commonly robotic arms (defined in Table 9). However, the study by Baskaran et al. [66] was an evaluation carried out using Siemens Process software, so the authors did not use any robots. Similarly, Colim et al. [60] performed the experimentation on a workstation, did not use any robots, and the same was the case in the study by Almeida et al. [67], who focused on interfaces.

Table 9: Type of robot used in each experimental study.

Type of Robot	Robot	References
Humanoid Robot	Robot NAO	[53]
Robotic arm	Nachi MC70	[54]
	ABB IRB-120	[55]
	YumiIRB 14000	[58]
	Universal RobotsUR5	[65] [68]
	UniversalRobotsUR10	[71]
	KUKA LBR IIWA 14 R820	[70]
Integral robot	Robot consisting of: (1) a four-wheel drive (4WD) remote-controlled vehicle, (2) a robotic arm, a Lynx al5d, and (3) a pan and tilt camera.	[62]
	Nomad Scout	[72]
	Magabot	[72]
	Giraff Robot	[72]

Performance is measured in seven experiments. Psychophysiological measures are only used in one of the experiments. In seven of the experiments, observation during the task is also carried out, from which qualitative information about the interaction is obtained.

Table 10: Summary of the reviewed experimental studies.

Ref.	PRE – Expert Assessment		PRE – Collection of the participant's perception – Questionnaire		PRE – Collection of the participant's perception – Interview		DURING – Task execution – Use of Robot		DURING – Quantitative objective measures – Performance		DURING – Quantitative objective measures – Psychophysiological measurements		DURING – Observation		POST – Questionnaire		POST – Interview	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
[53]		░		██		██	██			░		██	██		██			░
[54]		░		██	██		██			░		██		██	██		░	
[55]		░		██		██	██			░		██	██		██			░
[56]		░		██		██	██			░		██	██		██			░
[58]		░		██		██	██			░		██	██			██		░
[33]		░		██		██	██			░	██		██		██			░
[62]		░		██		██	██			░		██			██			░
[65]		░		██		██	██			░					██			░
[66]		░		██		██		██		░			██			██		░
[67]		░		██		██		██		░			██		██			░
[68]		░		██		██	██			░			██		██			░
[70]		░		██		██	██			░		██	██		██			░
[71]		░		██		██	██			░		██	██		██			░
[60]		░		██		██		██		░		██	██		██			░
[72]		░		██		██	██			░		██	██		██			░

As for the phase after execution, data collection of the participants' perceptions is carried out mainly through questionnaires (on 13 occasions). In one of the experiments, an interview is also carried out.

Sample size and gender perspective. The sizes of the samples used in the different case studies and the number of men and women have also been collected. Table 11 shows the number of people who participated in each of the case studies, and the distribution between genders. Of the total sample, 30% are women and 65% men. There were 450 participants (134 women and 291 men). Only two studies (one of the case studies of Aromaa et al. [62] and the study by Tang et al. [65]) use equal samples. Furthermore, the studies do not show the results obtained disaggregated by gender, preventing us from determining whether there are differences between genders when interacting with robots.

Table 11. Sample size in each experimental study (nd=no data).

Ref.	No. of people	Women	Men	Others
[53]	28	8	20	0
[54]	16	4	12	0
[55]	20	7	13	0
[56]	25	7	17	1
	27	7	19	1
	25	7	18	0
	21	4	14	1
[62]	10	2	8	0
	9	4	5	0
[65]	8	4	4	0
[67]	25	2	23	0
[68]	20	nd	nd	nd
[70]	19	4	15	0
[71]	32	4	28	0
[72]	85	35	50	0
	80	35	45	0
Total	**450**	**134**	**291**	**3**

4.3 RQ-2: What human factors does it include and how does it assess them?

Human factors is the scientific discipline concerned with the interaction between humans and artifacts and design of systems where people participate [80]. The purpose is to match systems, jobs, products and environments to the physical and mental abilities and limitations of people [80]. According to Beith [81] human factors focus on system usability and designing system interfaces to optimize the users' ability to accomplish their tasks error-free in a reasonable time and, therefore, to accept the system as a useful tool. Considering applying human factors principles leads to designs that are safer, more acceptable, more comfortable, and more effective for accomplishing their given tasks [81]. Table 12 shows the factors evaluated in the experimental studies and

the ways in which they were evaluated. Four groups of measures have been identified: i) performance, ii) posture, iii) robot-related factors and iv) emotion-related factors.

Performance. Performance is the most evaluated factor; it refers to how people perform their task. Users' performance is shaped by their capabilities (e.g., memory, attention, flexibility), and it is the consequence of the human factors displayed in the system. Therefore, a human centred approach involves taking into account also performance consideration. The indicators in this group are indicators that are directly or indirectly reflected in human performance, and hence provide insights into the human factors and UX. Eight indicators to evaluate it have been identified in seven of the fifteen studies.

i) **Task execution time** is measured in seven of the fifteen studies (47%). Thus, it is the most evaluated factor among the studies identified, in fact, it is measured by all the studies that assess performance.

ii) **The number of interactions performed** is measured once, i.e., in the study by Daniel et al. [54]. As the authors stated, this variable shows the quality of the user interface and offers insight into the possibilities for incorrect data input [54].

iii) **Errors** are measured once, i.e., in the study by Almeida et al. [67].

iv) **Robot idle time** is measured twice, i.e., in the studies by Lasota et al. [55] and Hietanen et al. [68]**.**

v) **Person idle time** is measured once, i.e., in the study by Lasota et al. [55].

vi) **Variability in production times** is measured once, i.e., in the study by Colim et al. [60].

vii) **Production rate** is measured once, i.e., in the study by Colim et al. [60]. According to the authors, this is a key indicator measuring performance in terms of pieces produced within a specified time interval (e.g., number of preforms per hour).

viii) **The ratio between the time required to complete the task with and without the robot** is measured once, i.e., in the study by Beschi et al. [71]. As they stated, this indicator verifies whether human productivity is also affected by robot movement during unsynchronized tasks [71].

Posture. Related to anthropometrics and biomechanics, this focuses on eliminating harmful and unsafe work practices and aims to study human capabilities and limitations in order to adapt the task to the person while minimizing fatigue [82]. Four of the fifteen studies analyse the posture of the person. Six indicators have been identified from the studies.

i) **Postural load** is measured once, i.e., in the study by Harriott et al. [33]. It measures the percentage of time the participants spent with the flexion of their trunks at an angle of more than 45º from the vertical [33]. The longer a participant spent with severe trunk flexion, the higher the physical workload [83].

ii) **Variance in posture** is measured once, i.e., in the study by Harriott et al. [33].

iii) **Total movement** is measured once, i.e., in the study by Harriott et al. [33]. It is presented as the total number of times the participant stood up and crouched down.

iv) **Vector magnitude** is measured once, i.e., in the study by Harriott et al. [33]. As stated by the authors, it is a measure of overall physical activity and combines acceleration from the three axes of movement. Vector magnitude measures participants' physical movement in the evaluation area.

v) **Rapid upper limb assessment (RULA)** is an observational method [84] to evaluate physical work-related upper limb disorders [85]. Its application involves the assessment of a worker's posture, as well as the exerted forces, the repetitiveness of movements, and external loads (e.g., handling heavy materials) [85]. It is measured in three studies: those by Aromaa et al. [62], Tang et al. [65] and Colim et al. [60].

vi) **The revised strain index (RSI)** is also measured in the study of Colim et al. [60]. The RSI consists of a five-variable model using continuous multipliers. The five variables/risk factors measured are the intensity of exertion (force), exertions per minute (frequency), duration per exertion, hand–wrist posture and duration of a task per day [86].

Robot-related factors. This includes the characteristics that describe the nature of the system and, therefore, influence human perception. Factors related to the robot are measured in five of the fifteen studies. In total, seven factors have been identified.

i) **Anthropomorphism** is measured using the Godspeed questionnaire [87] in two studies. As Bartneck et al. [87] stated, anthropomorphism refers to the attribution of a human form, human characteristics, or human behaviour to nonhuman things, such as robots.

ii) **Animacy** is measured in the same two studies using the Godspeed questionnaire [87]. As Bartneck et al. [87] stated, the goal of many robotics researchers is to make their robots lifelike.

iii) **Likeability** is also measured in the same two studies using the Godspeed questionnaire [87]. As Bartneck et al. [87] stated, the way in which people form positive impressions of others is, to some degree, dependent on the visual and vocal behaviour of the targets, positive first impressions of a person often lead to more positive evaluations of that person.

iv) **Perceived intelligence** is also measured in the same two studies using the Godspeed questionnaire [87]. As Bartneck et al. [87] stated, interactive robots face tremendous challenges in acting intelligently. The reasons can be traced back to the field of artificial intelligence (AI). Robots' behaviours are based on methods and knowledge developed with AI.

v) **Perceived safety** is measured in three studies. In two of them, it is measured using the Godspeed questionnaire [87]. A key issue for robots interacting with humans is safety [87]. This topic has received considerable attention in

the robotics literature, particularly in terms of the systems and standards established for both industrial robots and service robots intended for use in the home. In the study by Lasota et al. [55], they used a self-generated four-item questionnaire to measure perceived safety.

vi) **Usability** is measured twice. In the study by Danielsson et al. [56], they used the system usability scale (SUS) questionnaire [88], and in the study by Almeida et al. [67], they used a questionnaire based on IBM Computer Usability Satisfaction Questionnaire. The ISO 9241-11 defines usability as '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' [89].

vii) **Learnability** is measured once, i.e., in the study by Danielsson et al. [56], using the SUS questionnaire [88]. According to Joyce [90], learnability considers how easy it is for users to accomplish a task the first time they encounter the interface and how many repetitions it takes for them to become efficient at that task.

In summary, the seven robot-related factors have been measured through four questionnaires, which are as follows:

i) **Godspeed questionnaire** [87] – to measure anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. It was used twice, i.e., in the studies by Schillaci et al. [53] and Joosse et al. [72].

ii) **SUS questionnaire** [88] – to measure usability and likeability. It was used once, i.e., in the study by Danielsson et al. [56].

iii) **A questionnaire based on the IBM Computer Usability Satisfaction Questionnaire,** was used once, i.e., in the study by Almeida et al. [67], to measure usability and satisfaction.

iv) Another **questionnaire** was also used in the study by Lasota et al. [55] to measure satisfaction with robots as teammates and to determine perceived safety and comfort.

Emotion-related factors. This includes emotional responses resulting from human-robot interaction which evaluate the hedonic quality [91] of the system. Eight studies of the fifteen measured factors related to emotions. A total of five factors have been identified.

i) **Trust** is measured once, i.e., in the study by Daniel et al. [54]. The authors asked participants some questions adapted from the web accessibility initiative (WAI) Site Usability Testing Questions [92].

ii) **Satisfaction** is measured in two studies, i.e., those by Lasota et al. [55] and Almeida et al. [67]. In the former [55], the authors used a questionnaire that measured satisfaction with robots as teammates; in the latter [67], the authors used a questionnaire based on the IBM Computer Usability Satisfaction Questionnaire.

iii) **Mental workload** is measured in three studies. All of these studies, i.e., those by Aromaa et al. [62], Pantano et al. [70] and Harriott et al. [33], used NASA-Task Load Index (NASA-TLX) [93] questionnaire. Harriott et al. [33] also used physiological measures, such as heart rate and heart rate variability.

iv) **Physical and mental stress** is measured once, i.e., in the study by Hietanen et al. [68], using a self-generated questionnaire.

v) **Perceived risk** is also assessed once, i.e., in the study by Beschi et al. [71], using a self-generated questionnaire.

Table 12. Factors and techniques evaluated in each experimental study.

Factor	Technique		Reference
Performance	Task execution time		[54] [55] [33] [67] [68] [60] [71]
	Number of interactions		[54]
	Errors		[67]
	Robot idle time		[55] [68]
	Person's idle time		[55]
	Variability in production times		[60]
	Production rate		[60]
	Ratio between the time needed to complete the task with and without the robot		[71]
Posture	Postural load		[33]
	Variance in posture		[33]
	Total movement		[33]
	Vector magnitude		[33]
	RULA		[62] [65] [60]
	RSI		[60]
Robot-related factors	Anthropomorphism	Godspeed questionnaire	[53] [72]
	Animacy	Godspeed questionnaire	[53] [72]
	Likeability	Godspeed questionnaire	[53] [72]
	Perceived Intelligence	Godspeed questionnaire	[53] [72]
	Perceived Safety	Godspeed questionnaire	[53] [72]
		Self- generated questionnaire	[55]
	Usability	SUS questionnaire	[56]
		IBM Computer Usability Satisfaction Questionnaire based questionnaire	[67]
	Learnability	SUS Questionnaire	[56]
Emotion-related factors	Trust	Self-generated questionnaire about Trust	[54]
	Satisfaction	Self-generated questionnaire	[55]
		IBM Computer Usability Satisfaction Questionnaire based questionnaire	[67]

Mental workload	NASA-TLX	[62] [70] [33]
	Heart Rate	[33]
	Heart rate Variability	[33]
Physical and mental stress	Physical and mental stress questionnaire (self-generated)	[68]
Perceived risk	Perceived Risk Questionnaire	[71]

Types of measurements. A classification of the measurements used is made to understand the types of measurements applied and the interest in them. In addition to those previously mentioned, other measurements have been identified. The classification involves objective, subjective, qualitative, and quantitative measures.

Table 13 presents that the subjective and quantitative measures are basically questionnaires. Various questionnaires have been identified according to the indicator to be measured.

i) **Path following precision** is measured in the study by Almeida et al. [67]. According to the authors, this indicator refers to a 3D path following precision with the tip of a stick.

ii) **The ratio between the touch screen and keyboard interactions** is measured once, i.e., in the study by Daniel et al. [54]. The ratio between the interactions with the touch screen and the keys on the teach pendant or the robot controller may indicate the tendency of divided attention caused by the user interface [54].

iii) **Percentage of concurrent motion** is measured once, i.e., in the study by Lasota et al. [55].

iv) **Distance between the human and the robot** verifies whether human motion is affected by robot movement during unsynchronized tasks. It has been measured twice, i.e., in the studies by Lasota et al. [55] and Beschi et al. [71].

v) **The Siemens process** is capable of virtually validating manufacturing concepts up front. Through this platform, an evaluation of the interactions between associates working on the assembly line, equipment and materials flow can be performed [66].

Regarding subjective measurements, we can classify them into: i) quantitative and ii) qualitative, as shown in Table 13. Questionnaires are particularly efficient methods of application and analysis that are commonly used for user-driven assessments [94]. They allow for efficient quantitative measurements of product characteristics, as they are usually measured using Likert scales [79] or semantic pairs [95]. Questionnaires measure user's perspectives and do not necessarily require any kind of monitoring.

i) **A general interest questionnaire** is used in the study by Danielsson et al. [56]. Participants filled out questionnaires with six questions regarding general interest and five questions regarding the information displayed on the screen.

ii) **Observation** is used in two studies, i.e., those by Danielsson et al. [56] and Kildal et al. [58].

Table 13. Type of measurements and tools used in each study.

		Tools	References
Objective measure-ments	Quantitative measurements	Task execution time	[54] [55] [33] [67] [68] [60] [71]
		Number of interactions	[54]
		Errors	[67]
		Path following precision	[67]
		Ratio between touch screen and keyboard interactions	[54]
		Percentage of concurrent motion	[55]
		Average separation distance between the human and the robot	[55] [71]
		Robot idle time	[55] [68]
		Human idle time	[55]
		Variability in production times	[60]
		Production rate	[60]
		Ratio between the time needed to complete the task with and without the robot	[71]
		BioHarness heart rate monitor	[33]
		Siemens process	[66]
	Postural measurements	Postural load	[33]
		Variance in posture	[33]
		Total movement	[33]
		Vector magnitude	[33]
		RULA	[62] [65] [60]
		RSI	[60]
Subjective measure-ments	Quantitative measurements	Godspeed questionnaire	[53] [72]
		Questionnaire Trust in Automation	[54]
		Questionnaire (Satisfaction with the robot as a team-mate and perceived safety and comfort)	[55]
		SUS questionnaire	[56] [65]
		General interest questionnaire	[56]
		NASA-TLX	[33] [62] [70]
		IBM Computer Usability Satisfaction Questionnaire based questionnaire	[67]
		Physical and mental stress questionnaire	[68]
		Perceived Risk Questionnaire	[71]
		Questionnaire on changes in planning configuration and what they are most comfortable with	[71]
		Questionnaire on worker perception (impact of robot-ics in the work context, perceived effort associated with the tasks and overall evaluation of the job)	[60]
	Qualitative measurements	Observation	[56] [58]
		Verbal rating of workload (auditory, visual, speech, tactile and motor)	[33]

5 Discussion

In the context of HRC, the role of people remains central. This mutual relationship between people and robots results in a powerful collaboration framework with a positive impact on productivity and flexibility. Using a human-centred approach is essential to knowing people's perceptions and thus bringing out the best in them during interactions. Therefore, analysing UX in these environments is crucial. Users must perceive robots as allies so that they can leverage the strengths of both for common goals. Collaborative robots enable closer and safer interactions between humans and machines, so that both sides can benefit from each other's strengths. An SLR was conducted to learn how evaluations of HRI occur. The review identified a total of twenty-four articles, of which five were theoretical frameworks, five were evaluation techniques or tools and fifteen were experimental studies.

The theoretical frameworks identified are consistent, and they present similarities. Although the articles are recent, their similarities make us understand that there is a common line within this field of research.

In general terms, the importance of safety is emphasised to ensure an effective HRC. However, the safety perceived by the person should be considered equally important because if the user does not perceive it as such, the interaction will not be satisfactory and the UX will not be evaluated positively.

According to Norman [77], [78], when a person interacts with any object, in this case a robot, they start from an intention, which is executed and subsequently has consequences, such as the perception, interpretation and evaluation of the context. Along this line, Lindblom et al. [11] built a theoretical framework consisting of a phased and iterative procedure with the aims of providing guidance on how to carry out a UX evaluation and facilitating an understanding of why something works or does not work identifying UX problems.

The human factors that influence HRI have also been identified, and those described by Meissner et al. [59] stand out. The most influential factors affecting the acceptance of workers in HRC contexts are perceived risk, perceived benefits, and positive and negative emotions. These factors influence attitudes towards system acceptance.

The present paper reveals how HRI evaluations have been performed to date. There is a lack of experiments that evaluate UX before, during and after interactions. It would be appropriate to include the evaluation in these three phases and with a comprehensive approach, i.e., by using different measures (qualitative and quantitative, objective and subjective), to better interpret the data obtained from each measurement.

In the experiments on HRI assessment, only a single study using physiological monitoring was identified. The use of physiological monitoring could be beneficial in obtaining objective data on user emotions. Compared with traditional methods, including physiological monitoring in UX testing has limitations in terms of price, complexity and time required to ensure that the assessment is done properly. Moreover, physiological signals require some degree of interpretation, as the output must be processed to move from the raw data to actionable insights [50].

It can be concluded from this study that the evaluation of UX by combining methods, tools and physiological devices could be beneficial, as the interaction would be

evaluated considering different types of information. Physiological devices provide quantitative and objective data about the user at the moment of the interaction, questionnaires help obtain quantitative and subjective data, useful to understand user perceptions of a system.

Finally, this study highlights the need to consider the gender perspective. A marked difference was observed in the samples used in the experimental studies, which consisted of 65% men and only 30% women. To contribute to the reduction of the gender digital gap, researchers should use equal samples and disaggregate the data obtained by gender to determine whether differences between genders in exist HRI.

6   Research gaps and future research directions

6.1   Research on the Correlation between Dynamics of Robots and User Perceptions

Dynamic variables have had significant influence on perception studies. An important factor in HRI is the speed at which robots act. Several works in the literature have examined the appropriate speeds for robotic actions [58], [80], [81], determining that robots should act slower than people. According to Joosse et al. [72], a possible reason for people's preference for a slower robot speed may be that not all robots give a clear indication of when they are going to stop, i.e., they do not provide feedback on their intentions. One way to overcome this is to equip robots with functional feedback systems so that they can convey their intentions. Therefore, there is an opportunity for research into dynamic variables and the correlation between user perceptions and performance. One further opportunity is evaluating people's abilities to understand robots, the degree of accuracy in predicting the robots' actions, and whether the sequence of actions performed is appropriate.

According to Lindblom et al. [11], UX is not absolute, which means that each person may perceive their experience differently. Helping a person understand a robot, perceive its intention, predict the sequence of actions that will take place, evaluate its actions, determine whether any action is necessary, specify a sequence of actions and perform these actions is necessary to optimally perceive UX in HRI. To this end, and according to Cohen et al. [64] and Villani et al. [63], robots must provide adequate feedback so that users can understand and predict their actions. There is a research opportunity to analyse how this feedback can be presented according to the robots' actions and the people's emotions. In this sense, affective robotics could be used to provide useful feedback by determining the appropriate channel, frequency, and intensity of interaction.

6.2   Research on the Evaluation of HRI in Design and Interaction

Operators need to have positive and fit-for-purpose experiences through trust-based, smooth, safe, and satisfying interactions in order to integrate robots as natural parts of their daily lives. Therefore, evaluation is a key aspect of ensuring a good UX. According to Gammieri et al. [98], virtual reality (VR) and AR are effective tools that are

capable of simulating industrial cobotic systems with a high level of immersion. These can simulate HRI safely and economically with a digital twin, even in earlier design phases in which the workplace is still under development.

In conclusion, the lack of a structured evaluation method that is adapted to different HRI contexts and stages of the design process has been identified in this SLR. Such a model would need to function in a virtual context, since the HRI system might be designed virtually in the early stages of the design process. This model would also need to be operational at later stages when the system is, for example, in a laboratory or even in a real environment. The framework proposed by Prati et al. [39], called the UX cycle in HRI, is the closest approach, however, it has shortcomings in the evaluation phase.

Meissner et al. [59] identified several influencing factors on the individual in the context of collaborative robotics in industry. There is an opportunity to investigate how these factors can be evaluated, and more specifically the correlation between the influencing factors and already validated assessment questionnaires. For example, perceived risk could be identified using questionnaires measuring perceived safety (Godspeed questionnaire [87] or the one that was used in the study by Lasota et al. [55]) or confidence (SUPR-Q [67] or UEQ+ [100]). Physical and mental relief could be related to ease of use, which could be measured by the Perceived Usefulness and Ease of Use Questionnaire [101], PSQ [102], ASQ [103], and USE [104]. It could also be related to intuitive use (measured by UEQ+ [100]) or cognitive effort (measured by the DEEP questionnaire [105]). As for the perception of progress, it could be measured by efficiency (SUMI [106], WAMMI [107], UEQ [94], UMUX [108], UEQ+ [100]), and effectiveness (UMUX [108]). In this sense, another line of research could study positive and negative emotions and the use of physiological tools.

The lack of experiments evaluating the temporal nature of UX, i.e., before, during and after interactions, provides an opportunity to investigate how it can be included in these three phases. In particular, emphasising the phase before the interaction is necessary because only in one experiment was an interview conducted prior to task execution (Table 12).

It has also been observed that, in the phase during the interaction, performance-related aspects are mainly evaluated, whereas emotions are hardly examined. The retrospective and subjective evaluation facilitated by questionnaires and interviews is, in all cases, not an optimal approach, because it does not measure UX at the moment when the interaction between a person and a robot occurs. Thus, it can be prone to human error because of inaccurate recall. Using physiological monitoring to assess UX during interactions is essential. According to Neumann et al. [109], psychophysiological measures are more objective than self-recording measures, such as questionnaires. The opportunities offered by physiological signals are increasing because of the evolution of sensors and signal processing [110].

6.3 Research on the Differences between Genders when Interacting with Robots

In the experimental studies, the gender variable was not studied. Therefore, investigating the gender digital gap and to addressing it are considered important.

Women continue to be underrepresented in technology compared with their number in the overall workforce [111]. As indicated by Holtzblatt et al. [112], research shows that a diverse and inclusive workforce correlates with higher innovation, creativity, revenue, and profit [113], [114]. Bala et al. [115] stated that today's workforce needs to be filled in a holistic manner that ensures a diverse group of people creates the technology of the future. From the perspective of the analysis of HRI, focusing on the differences in the behavioural modes, perceptions and emotional effects experienced by women and men when faced with different designs of these machines is appropriate.

For all these reasons, evaluating UX during HRI in industrial environments using an equal sample and disaggregating the results obtained by gender are necessary. Doing so will help identify whether there are differences in HRI between genders. In this way, workplaces can be designed in a gender-inclusive way to mitigate the gender digital divide.

7   Limitations

Our review has the following limitations:

i)   The SLR methodology does not guarantee that all the publications related to a given research area will be identified [116].

ii) Having limited the search to peer-reviewed articles, we may have missed case studies published at conferences that could have been relevant to the study.

iii) The reviewer bias: Despite having attempted to objectify the review, we may have introduced bias in some cases.

iv) The choice of databases used. Although we strategically selected the databases to ensure appropriate coverage of this research area and designed a search strategy to ensure that as many publications as possible were captured, it is possible that if we had used any additional databases, we could have identified more significant articles for the research.

v)   The QA criteria. If we had defined other QA questions, the result of the SLR would have been completely different. However, we wanted to focus on human factors, gender, and accessibility. The same with having set the cut-off in the QA at 7. If we had set a lower value, may be other relevant papers would have been identified. Despite our efforts to avoid bias, according to Yang et al. [51] QA could result from factors that potentially bias the findings of the study. Nevertheless, both the used QA questions and the cut-off value have provided us with quality papers, and this is supported by the literature characterisation. Since 96% of the articles correspond to articles published in indexed journals, 29% in the first quartile and 56% in the second quartile.

vi) Another limitation is restricting results to English and Spanish languages only.

8 Conclusions

The number of robots in the manufacturing industry has been steadily increasing for several decades and in recent years the number and variety of industries using robots have also increased. As stated by Hentout et al. [25], HRI can effectively contribute to developing future factories in which humans and robots can share tasks and work shoulder to shoulder. Therefore, operators need to have positive and fit-for-purpose experiences through trust-based, smooth, safe, and satisfying interactions in order to integrate robots as natural parts of their daily lives. According to the HCD approach, placing the human at the centre of the system is important to guarantee fluid, safe and satisfactory interactions.

This article reviewed existing works on HRI evaluation that considered human factors in industrial environments and were published between 2011 and 2021. A total of twenty-four full-text articles that provided a summary of i) theoretical frameworks, ii) operational and evaluation tools and iii) experimental studies were analysed.

i) The theoretical frameworks identified emphasised safety and provided insight into the human factors that could influence HRC. In the context of interaction with a robot, safety and perceived risk are determining factors, as they directly impact the person's performance and emotions. The theoretical framework proposed by Villani et al. [63] placed safety at the centre, and the framework proposed by Meissner et al. [59] indicated that one of the primary influential factors was perceived risk. Emotions, such as loyalty, stimulation, and trust, must also be considered, and even appearance must be taken into account, because little attention has been given to the emotional effects of aesthetic impressions on users [117]. A theoretical framework of HRI from the UX perspective was also identified [39], but it had shortcomings in the evaluation phase. Nevertheless, evaluation is a key aspect to optimising UX.

ii) The operational tools identified through SLR are recent. They are tools of different nature that determine industrial maturity, assess the measure of self-efficacy in HRI, propose design guidelines for collaborative assembly workstations, or allow synthesising HRI experiences. In addition to the development of these tools, other tools have been identified which have been applied in experimental HRI evaluation studies. On the other hand, physiological tools allow objective assessment because they provide information without retrospective bias. Given the lack of experiments using these tools to evaluate UX in HRI, integrating user monitoring using physiological tools would be essential in future experiments. Including different tools in a new evaluation model would also be essential, allowing the evaluation of UX in different phases of the design process, such as VR or AR, which would, in turn, enable assessment of UX in the design phases of the workplace. The combination of different tools, with different characteristics, at different times could help ensure the accuracy and reliability of the

      results, as it would provide a better understanding of the context in which the interaction takes place

iii)   The present article summarised how evaluations of HRI were performed in the literature. No validated model to assess UX in HRI was identified. In general terms, the experiments are composed of three phases: i) prior to the execution of the task, ii) during the execution of the task and iii) after the execution of the task. Performance is the most evaluated factor, mainly through task execution time. As for subjective evaluation, questionnaires were the most frequently used tools, although different questionnaires were identified in the different case studies. Complementing the study with traditional tools, such as questionnaires or interviews, that provide subjective insight into users' perceptions would be interesting

Future experimentations must integrate a holistic approach to capture people's perceptions at all times, i.e., before, during and after an interaction. Including different methods of measurement at different moments of the interaction—quantitative and qualitative, objective and subjective—and integrating gender aspects to guarantee the interpretation of the data and the understanding of the entire flow of interaction.

Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article.

Compliance with ethical standards

Ethical approval is not applicable for this article. This article does not contain any studies with human or animal subjects, and there are no human subjects in this article and informed consent is not applicable.

Declaration of competing interest

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

References

[1]   S. Waschull, J. A. C. Bokhorst, E. Molleman, and J. C. Wortmann, "Work design in future industrial production: Transforming towards cyber-physical systems," *Comput Ind Eng*, vol. 139, p. 105679, 2020, doi: https://doi.org/10.1016/j.cie.2019.01.053.

[2]   J. A. Marvel, S. Bagchi, M. Zimmerman, and B. Antonishek, "Towards effective interface designs for collaborative HRI in manufacturing: metrics and measures," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 9, no. 4, pp. 1–55, 2020, doi: https://doi.org/10.1145/3385009.

[3]   I. Karabegović, R. Turmanidze, and P. Dašić, "Global trend of implementation of industrial robots relating to industry 4.0," in *Design, Simulation, Manufacturing: The Innovation Exchange*, 2020, pp. 147–155.

[4]   I. Karabegović, R. Turmanidze, and P. Dašić, "Structural Network for the Implementation of 'Industry 4.0' in Production Processes," *Industry 4.0*, vol. 7, no. 1, pp. 3–6, 2022.

[5]   J. Van, "Mechanical Advantage: Two Northwestern University engineers are developing cobots -- machines that, unlike robots, cooperate with workers without displacing them," Dec. 11, 1996. https://peshkin.mech.northwestern.edu/cobot/chitrib/jonvan.html (accessed Mar. 09, 2022).

[6]   ISO 8373, "Robots and robotic devices," *ISO Online Browsing Platform*, no. 30, 2012.

[7]   K. Dautenhahn, "Human-Robot Interaction | The Encyclopedia of Human-Computer Interaction, 2nd Ed.," 2013. https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-robot-interaction (accessed Mar. 05, 2021).

[8]   M. Chen, S. Nikolaidis, H. Soh, D. Hsu, and S. Srinivasa, "Trust-aware decision making for human-robot collaboration: Model learning and planning," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 9, no. 2, pp. 1–23, 2020, doi: https://doi.org/10.1145/3359616.

[9]   M. Boden *et al.*, "Principles of robotics: regulating robots in the real world," *https://doi.org/10.1080/09540091.2016.1271400*, vol. 29, no. 2, pp. 124–129, Apr. 2017, doi: 10.1080/09540091.2016.1271400.

[10]   P. H. Kahn Jr *et al.*, "What is a Human?: Toward psychological benchmarks in the field of human–robot interaction," *Interact Stud*, vol. 8, no. 3, pp. 363–390, 2007, doi: https://doi.org/10.1075/is.8.3.04kah.

[11]   J. Lindblom and B. Alenljung, "The ANEMONE: Theoretical Foundations for UX Evaluation of Action and Intention Recognition in Human-Robot Interaction," *Sensors*, vol. 20, no. 15, p. 4284, 2020, doi: https://doi.org/10.3390/s20154284.

[12]   ISO 9241-210, "Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems," *International Standard*, vol. 2.Udgave. 2019.

[13]   D. Benyon, "Designing User Experience: a guide to HCI," *UX and interaction design*, 2019.

[14]   S. F. Warta, K. A. Kapalo, A. Best, and S. M. Fiore, "Similarity, Complementarity, and Agency in HRI: Theoretical Issues in Shifting the Perception of Robots from Tools to Teammates; Similarity, Complementarity, and Agency in HRI: Theoretical Issues in Shifting the Perception of Robots from Tools to Teammates," 2016, doi: 10.1177/1541931213601287.

[15]   T. Wiltshire, "Towards Modeling Social-Cognitive Mechanisms in Robots to Facilitate Human-Robot Teaming Related papers," 2013, doi: 10.1177/1541931213571283.

[16]   T. J. Wiltshire, S. F. Warta, D. Barber, and S. M. Fiore, "Enabling robotic social intelligence by engineering human social-cognitive mechanisms," *Cogn Syst Res*, vol. 43, pp. 190–207, Jun. 2017, doi: 10.1016/J.COGSYS.2016.09.005.

[17]   K. E. Schaefer, E. R. Straub, J. Y. C. Chen, J. Putney, and A. W. Evans, "Communicating intent to develop shared situation awareness and engender trust in human-agent teams," *Cogn Syst Res*, vol. 46, pp. 26–39, Dec. 2017, doi: 10.1016/J.COGSYS.2017.02.002.

[18]   P. B. Morrow and S. M. Fiore, "Supporting Human-Robot Teams in Social Dynamicism: an overview of the Metaphoric Inference Framework; Supporting Human-Robot Teams in Social Dynamicism: an overview of the Metaphoric Inference Framework," 2012. doi: 10.1177/1071181312561344.

[19]   O. Sebastián García and M. A. del Hoyo Delgado, "La carga mental de trabajo," Madrid.

[20]   X. V. Wang, Z. Kemény, J. Váncza, and L. Wang, "Human–robot collaborative assembly in cyber-physical production: Classification framework and implementation," *CIRP Ann Manuf Technol*, vol. 66, no. 1, pp. 5–8, Jan. 2017, doi: 10.1016/j.cirp.2017.04.101.

[21]   A. C. Simões, A. Pinto, J. Santos, S. Pinheiro, and D. Romero, "Designing human-robot collaboration (HRC) workspaces in industrial settings: A systematic literature review," *J Manuf Syst*, vol. 62, pp. 28–43, Jan. 2022, doi: 10.1016/J.JMSY.2021.11.007.

[22]   M. Breque, L. de Nul, and A. Petridis, "Industry 5.0: towards a sustainable, human-centric and resilient European industry," *Luxembourg, LU: European Commission, Directorate-General for Research and Innovation*, 2021.

[23]   P. Maurice, V. Padois, Y. Measson, and P. Bidaud, "Human-oriented design of collaborative robots," *Int J Ind Ergon*, vol. 57, pp. 88–102, Jan. 2017, doi: 10.1016/J.ERGON.2016.11.011.

[24]   M. Amoretti and M. Reggiani, "Architectural paradigms for robotics applications," *Advanced Engineering Informatics*, vol. 24, no. 1, pp. 4–13, Jan. 2010, doi: 10.1016/J.AEI.2009.08.004.

[25]   A. Hentout, M. Aouache, A. Maoudj, and I. Akli, "Human–robot interaction in industrial collaborative robotics: a literature review of the decade 2008–2017," *Advanced Robotics*, vol. 33, no. 15–16, pp. 764–799, 2019, doi: https://doi.org/10.1080/01691864.2019.1636714.

[26]   N. Savela, T. Turja, and A. Oksanen, "Social Acceptance of Robots in Different Occupational Fields: A Systematic Literature Review," *International Journal of Social Robotics 2017 10:4*, vol. 10, no. 4, pp. 493–502, Dec. 2017, doi: 10.1007/S12369-017-0452-5.

[27]   L. Veling and C. McGinn, "Qualitative Research in HRI: A Review and Taxonomy," *Int J Soc Robot*, vol. 13, no. 7, pp. 1689–1709, Nov. 2021, doi: 10.1007/S12369-020-00723-Z/FIGURES/3.

[28]   T. Moulières-Seban, D. Bitonneau, J.-M. Salotti, J.-F. Thibault, and B. Claverie, "Human factors issues for the design of a cobotic system," in *Advances in human factors in robots and unmanned systems*, Springer, 2017, pp. 375–385.

34

[29]   P. Tsarouchi, S. Makris, and G. Chryssolouris, "Human–robot interaction review and challenges on task planning and programming," *Int J Comput Integr Manuf*, vol. 29, no. 8, pp. 916–931, 2016, doi: https://doi.org/10.1080/0951192X.2015.1130251.

[30]   V. Gopinath, F. Ore, and K. Johansen, "Safe Assembly Cell Layout through Risk Assessment – An Application with Hand Guided Industrial Robot," *Procedia CIRP*, vol. 63, pp. 430–435, Jan. 2017, doi: 10.1016/J.PROCIR.2017.03.160.

[31]   C.-S. Tsai, J.-S. Hu, and M. Tomizuka, "Ensuring safety in human-robot coexistence environment," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 4191–4196. doi: 10.1109/IROS.2014.6943153.

[32]   L. Gualtieri, G. P. Monizza, E. Rauch, R. Vidoni, and D. T. Matt, "From Design for Assembly to Design for Collaborative Assembly - Product Design Principles for Enhancing Safety, Ergonomics and Efficiency in Human-Robot Collaboration," *Procedia CIRP*, vol. 91, pp. 546–552, Jan. 2020, doi: 10.1016/J.PROCIR.2020.02.212.

[33]   C. E. Harriott, T. Zhang, and J. A. Adams, "Assessing physical workload for human–robot peer-based teams," *Int J Hum Comput Stud*, vol. 71, no. 7–8, pp. 821–837, Jul. 2013, doi: 10.1016/J.IJHCS.2013.04.005.

[34]   C. Díaz, "SheScience. Science from a gender perspective," *Metode Sci Stud J*, no. 7, pp. 94–95, 2021.

[35]   M. Qbilat and A. Iglesias, "Accessibility guidelines for tactile displays in human-robot interaction. A comparative study and proposal," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10897 LNCS, pp. 217–220, Jun. 2018, doi: 10.1007/978-3-319-94274-2_29/COVER.

[36]   B. Dániel, P. Korondi, G. Sziebig, and T. Thomessen, "Evaluation of flexible graphical user interface for intuitive human robot interactions," *Acta Polytechnica Hungarica*, vol. 11, no. 1, pp. 135–151, 2014.

[37]   J. Marcos Correia Marques, "A POINT-AND-COMMAND INTERFACE FOR GRASPING UNKOWN OBJECTS WITH ROBOTIC MANIPULATORS," 2017.

[38]   F. Driewer, M. Sauer, and K. Schilling, "Discussion of Challenges for User Interfaces in Human-Robot Teams.," 2007. Accessed: Nov. 10, 2022. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.1755&rep=rep1&type=pdf

[39]   E. Prati, M. Peruzzini, M. Pellicciari, and R. Raffaeli, "How to include User eXperience in the design of Human-Robot Interaction," *Robot Comput Integr Manuf*, vol. 68, p. 102072, Apr. 2021, doi: 10.1016/J.RCIM.2020.102072.

[40]   M. A. Goodrich and A. C. Schultz, *Human-robot interaction: a survey*. Now Publishers Inc, 2008.

[41]   J. Schmidtler, V. Knott, C. Hölzel, and K. Bengler, "Human Centered Assistance Applications for the working environment of the future," *Occupational Ergonomics*, vol. 12, no. 3, pp. 83–95, 2015, doi: 10.3233/OER-150226.

[42] X. V. Wang, A. Seira, and L. Wang, "Classification, personalised safety framework and strategy for human-robot collaboration," 2018.

[43] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Applying the user experience questionnaire (UEQ) in different evaluation scenarios," in *International Conference of Design, User Experience, and Usability*, 2014, pp. 383–392. doi: https://doi.org/10.1007/978-3-319-07668-3_37.

[44] A. P. O. S. Vermeeren, E. L.-C. Law, V. Roto, M. Obrist, J. Hoonhout, and K. Väänänen-Vainio-Mattila, "User experience evaluation methods: current state and development needs," in *Proceedings of the 6th Nordic conference on human-computer interaction: Extending boundaries*, 2010, pp. 521–530. doi: https://doi.org/10.1145/1868914.1868973.

[45] D. Wallach, J. Conrad, and T. Steimle, "The UX metrics table: A missing artifact," in *International Conference of Design, User Experience, and Usability*, 2017, pp. 507–517. doi: https://doi.org/10.1007/978-3-319-58634-2_37.

[46] I. Díaz-Oreiro, G. López, L. Quesada, and L. A. Guerrero, "Standardized questionnaires for user experience evaluation: A systematic literature review," in *Multidisciplinary Digital Publishing Institute Proceedings*, 2019, vol. 31, no. 1, p. undefined. doi: https://doi.org/10.3390/proceedings2019031014.

[47] K. Väänänen-Vainio-Mattila, V. Roto, and M. Hassenzahl, "Towards practical user experience evaluation methods," *Meaningful measures: Valid useful user experience measurement (VUUM)*, pp. 19–22, 2008.

[48] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," 2007.

[49] J. Kerr and N. Coviello, "Formation and Constitution of Effectual Networks: A Systematic Review and Synthesis," *International Journal of Management Reviews*, vol. 21, no. 3, pp. 370–397, Jul. 2019, doi: https://doi.org/10.1111/ijmr.12194.

[50] A. Apraiz Iriarte, G. Lasa Erle, and M. Mazmela Etxabe, "Evaluating User Experience with Physiological monitoring: A Systematic Literature Review," *DYNA New Technologies*, vol. 8, no. 1, pp. 21-undefined, 2021, doi: https://doi.org/10.6036/NT10072.

[51] L. Yang *et al.*, "Quality Assessment in Systematic Literature Reviews: A Software Engineering Perspective," *Inf Softw Technol*, vol. 130, p. 106397, Feb. 2021, doi: 10.1016/J.INFSOF.2020.106397.

[52] S. Nidhra, M. Yanamadala, W. Afzal, and R. Torkar, "Knowledge transfer challenges and mitigation strategies in global software development—A systematic literature review and industrial validation," *Int J Inf Manage*, vol. 33, no. 2, pp. 333–355, Apr. 2013, doi: 10.1016/J.IJINFOMGT.2012.11.004.

[53] G. Schillaci, S. Bodiroža, and V. V. Hafner, "Evaluating the Effect of Saliency Detection and Attention Manipulation in Human-Robot Interaction," *International Journal of Social Robotics 2013 5:1*, vol. 5, no. 1, pp. 139–152, Nov. 2013, doi: 10.1007/S12369-012-0174-7.

[54] B. Daniel, T. Thomessen, and P. Korondi, "Simplified Human-Robot Interaction: Modeling and Evaluation," *Modeling, Identification and Control: A*

Norwegian Research Bulletin, vol. 34, no. 4, pp. 199–211, 2013, doi: 10.4173/MIC.2013.4.4.

[55]   P. A. Lasota and J. A. Shah, "Analyzing the Effects of Human-Aware Motion Planning on Close-Proximity Human–Robot Collaboration," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 57, no. 1, pp. 21–33, Jan. 2015, doi: 10.1177/0018720814565188.

[56]   O. Danielsson, A. Syberfeldt, R. Brewster, and L. Wang, "Assessing Instructions in Augmented Reality for Human-robot Collaborative Assembly by Using Demonstrators," *Procedia CIRP*, vol. 63, pp. 89–94, Jan. 2017, doi: 10.1016/J.PROCIR.2017.02.038.

[57]   A. R. von der Pütten and N. Bock, "Development and Validation of the Self-Efficacy in Human-Robot-Interaction Scale (SE-HRI)," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 7, no. 3, Dec. 2018, doi: 10.1145/3139352.

[58]   J. Kildal, M. Martín, I. Ipiña, and I. Maurtua, "Empowering assembly workers with cognitive disabilities by working with collaborative robots: a study to capture design requirements," *Procedia CIRP*, vol. 81, pp. 797–802, Jan. 2019, doi: 10.1016/J.PROCIR.2019.03.202.

[59]   A. Meissner, A. Trübswetter, A. S. Conti-Kufner, and J. Schmidtler, "Friend or Foe? Understanding Assembly Workers' Acceptance of Human-robot Collaboration," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 10, no. 1, Jul. 2020, doi: 10.1145/3399433.

[60]   A. Colim *et al.*, "Lean Manufacturing and Ergonomics Integration: Defining Productivity and Wellbeing Indicators in a Human–Robot Workstation," *Sustainability 2021, Vol. 13, Page 1931*, vol. 13, no. 4, p. 1931, Feb. 2021, doi: 10.3390/SU13041931.

[61]   G. Charalambous, S. R. Fletcher, and P. Webb, "The development of a Human Factors Readiness Level tool for implementing industrial human-robot collaboration," *The International Journal of Advanced Manufacturing Technology*, vol. 91, no. 5, pp. 2465–2475, Jan. 2017, doi: 10.1007/S00170-016-9876-6.

[62]   S. Aromaa, N. Frangakis, D. Tedone, J. Viitaniemi, and I. Aaltonen, "Digital Human Models in Human Factors and Ergonomics Evaluation of Gesture Interfaces," *Proc ACM Hum Comput Interact*, vol. 2, no. EICS, Jun. 2018, doi: 10.1145/3229088.

[63]   V. Villani, F. Pini, F. Leali, and C. Secchi, "Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications," *Mechatronics*, vol. 55, pp. 248–266, Nov. 2018, doi: 10.1016/J.MECHATRONICS.2018.02.009.

[64]   Y. Cohen, M. Golan, G. Singer, and M. Faccio, "Workstation–Operator Interaction in 4.0 Era: WOI 4.0," *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 399–404, Jan. 2018, doi: 10.1016/J.IFACOL.2018.08.327.

[65]   G. Tang and P. Webb, "The Design and Evaluation of an Ergonomic Contactless Gesture Control System for Industrial Robots," *Journal of Robotics*, vol. 2018, 2018, doi: 10.1155/2018/9791286.

[66] S. Baskaran *et al.*, "Digital Human and Robot Simulation in Automotive Assembly using Siemens Process Simulate: A Feasibility Study," *Procedia Manuf*, vol. 34, pp. 986–994, Jan. 2019, doi: 10.1016/J.PROMFG.2019.06.097.

[67] L. Almeida, P. Menezes, and J. Dias, "Interface Transparency Issues in Teleoperation," *Applied Sciences*, vol. 10, no. 18, p. 6232, Sep. 2020, doi: 10.3390/APP10186232.

[68] A. Hietanen, R. Pieters, M. Lanz, J. Latokartano, and J. K. Kämäräinen, "AR-based interaction for human-robot collaborative manufacturing," *Robot Comput Integr Manuf*, vol. 63, p. 101891, Jun. 2020, doi: 10.1016/J.RCIM.2019.101891.

[69] L. Gualtieri, E. Rauch, R. Vidoni, and D. T. Matt, "Safety, Ergonomics and Efficiency in Human-Robot Collaborative Assembly: Design Guidelines and Requirements," *Procedia CIRP*, vol. 91, pp. 367–372, Jan. 2020, doi: 10.1016/J.PROCIR.2020.02.188.

[70] M. Pantano, D. Regulin, B. Lutz, and D. Lee, "A human-cyber-physical system approach to lean automation using an industrie 4.0 reference architecture," *Procedia Manuf*, vol. 51, pp. 1082–1090, Jan. 2020, doi: 10.1016/J.PROMFG.2020.10.152.

[71] M. Beschi, M. Faroni, C. Copot, and N. Pedrocchi, "How motion planning affects human factors in human-robot collaboration," *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 744–749, Jan. 2020, doi: 10.1016/J.IFACOL.2021.04.167.

[72] M. Joosse, M. Lohse, N. van Berkel, A. Sardar, and V. Evers, "Making Appearances: How Robots Should Approach People," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 10, no. 1, Jan. 2021, doi: 10.1145/3385121.

[73] M. Qbilat, A. Iglesias, and T. Belpaeme, "A Proposal of Accessibility Guidelines for Human-Robot Interaction," *Electronics 2021, Vol. 10, Page 561*, vol. 10, no. 5, p. 561, Feb. 2021, doi: 10.3390/ELECTRONICS10050561.

[74] H. Bo, D. M. Mohan, M. Azhar, K. Sreekanth, and D. Campolo, "Human-robot collaboration for tooling path guidance," *Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics*, vol. 2016-July, pp. 1340–1345, Jul. 2016, doi: 10.1109/BIOROB.2016.7523818.

[75] S. Robla-Gomez, V. M. Becerra, J. R. Llata, E. Gonzalez-Sarabia, C. Torre-Ferrero, and J. Perez-Oria, "Working Together: A Review on Safe Human-Robot Collaboration in Industrial Environments," *IEEE Access*, vol. 5, pp. 26754–26773, 2017, doi: 10.1109/ACCESS.2017.2773127.

[76] G. Hirzinger, A. Albu-Schäffer, M. Hähnle, I. Schaefer, and N. Sporer, "On a new generation of torque controlled light-weight robots," *Proc IEEE Int Conf Robot Autom*, vol. 4, pp. 3356–3363, 2001, doi: 10.1109/ROBOT.2001.933136.

[77] D. Norman, *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[78] D. A. Norman, "Cognitive engineering," *User centered system design*, vol. 31, p. 61, 1986.

[79]   R. Likert, "A technique for the measurement of attitudes.," *Archives of psychology*, 1932.

[80]   M. G. Helander, "The human factors profession (Chapter 1)," in *Handbook of Human Factors and Ergonomics*, New York, 1997.

[81]   B. Beith, "Human Factors and the Future of Telemedicine," *Medical Device & Diagnostic Industry Magazine*, Jun. 01, 1999. https://www.mddi-online.com/news/human-factors-and-future-telemedicine (accessed Oct. 05, 2022).

[82]   A. Cardoso, A. Colim, E. Bicho, A. C. Braga, M. Menozzi, and P. Arezes, "Ergonomics and human factors as a requirement to implement safer collaborative robotic workstations: A literature review," *Safety*, vol. 7, no. 4, p. 71, 2021.

[83]   P. P. F. M. Kuijer, B. Visser, and H. C. G. Kemper, "Job rotation as a factor in reducing physical workload at a refuse collecting department," *Ergonomics*, vol. 42, no. 9, pp. 1167–1178, 1999.

[84]   G. C. David, "Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders," *Occup Med (Chic Ill)*, vol. 55, no. 3, pp. 190–199, 2005.

[85]   L. McAtamney and E. N. Corlett, "RULA: a survey method for the investigation of work-related upper limb disorders," *Appl Ergon*, vol. 24, no. 2, pp. 91–99, 1993, doi: 10.1016/0003-6870(93)90080-S.

[86]   A. Garg, J. S. Moore, and J. M. Kapellusch, "The Revised Strain Index: an improved upper extremity exposure assessment model," *Ergonomics*, vol. 60, no. 7, pp. 912–922, 2017, doi: 10.1080/00140139.2016.1237678.

[87]   C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots," *International Journal of Social Robotics 2008 1:1*, vol. 1, no. 1, pp. 71–81, Nov. 2008, doi: 10.1007/S12369-008-0001-3.

[88]   J. Brooke, "SUS-A quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[89]   N. Bevan, J. Carter, and S. Harker, "ISO 9241-11 revised: What have we learnt about usability since 1998?," in *International conference on human-computer interaction*, 2015, pp. 143–151.

[90]   A. Joyce, "How to Measure Learnability of a User Interface," Oct. 20, 2019. https://www.nngroup.com/articles/measure-learnability/ (accessed Sep. 30, 2022).

[91]   S. Diefenbach, N. Kolb, and M. Hassenzahl, "The 'hedonic' in human-computer interaction - History, contributions, and future research directions," *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS*, pp. 305–314, Jun. 2014, doi: 10.1145/2598510.2598549.

[92]   "WAI Site Usability Testing Questions." https://www.w3.org/WAI/EO/Drafts/UCD/questions.html (accessed Mar. 02, 2022).

[93]   S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Advances in Psychology*,

vol. 52, no. C, pp. 139–183, Jan. 1988, doi: https://doi.org/10.1016/S0166-4115(08)62386-9.

[94]   B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *Symposium of the Austrian HCI and usability engineering group*, 2008, pp. 63–76.

[95]   C. E. Osgood, "The nature and measurement of meaning.," *Psychol Bull*, vol. 49, no. 3, p. 197, 1952.

[96]   K. Dautenhahn *et al.*, "How may I serve you?," p. 172, 2006, doi: 10.1145/1121241.1121272.

[97]   J. T. Butler and A. Agah, "Psychological Effects of Behavior Patterns of a Mobile Personal Robot," *Autonomous Robots 2001 10:2*, vol. 10, no. 2, pp. 185–202, Mar. 2001, doi: 10.1023/A:1008986004181.

[98]   L. Gammieri, M. Schumann, L. Pelliccia, G. di Gironimo, and P. Klimant, "Coupling of a Redundant Manipulator with a Virtual Reality Environment to Enhance Human-robot Cooperation," *Procedia CIRP*, vol. 62, pp. 618–623, Jan. 2017, doi: 10.1016/J.PROCIR.2016.06.056.

[99]   J. Sauro, "SUPR-Q: A comprehensive measure of the quality of the website user experience.," *J Usability Stud*, vol. 10, no. 2, 2015.

[100] M. Schrepp and J. Thomaschewski, "Design and Validation of a Framework for the Creation of User Experience Questionnaires.," *International Journal of Interactive Multimedia & Artificial Intelligence*, vol. 5, no. 7, 2019, doi: 10.9781/ijimai.2019.06.006.

[101] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS quarterly*, pp. 319–340, 1989.

[102] J. R. Lewis, "An after-scenario questionnaire for usability studies: psychometric evaluation over three trials," *ACM SIGCHI Bulletin*, vol. 23, no. 4, p. 79, 1991.

[103] J. R. Lewis, "Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ," *ACM Sigchi Bulletin*, vol. 23, no. 1, pp. 78–81, 1991.

[104] A. M. Lund, "Measuring usability with the use questionnaire12," *Usability interface*, vol. 8, no. 2, pp. 3–6, 2001.

[105] T. Yang, J. Linder, and D. Bolchini, "DEEP: design-oriented evaluation of perceived usability," *Int J Hum Comput Interact*, vol. 28, no. 5, pp. 308–346, 2012, doi: 10.1080/10447318.2011.586320.

[106] J. Kirakowski and M. Corbett, "SUMI: The software usability measurement inventory," *British journal of educational technology*, vol. 24, no. 3, pp. 210–212, 1993.

[107] J. Kirakowski and B. Cierlik, "Measuring the usability of web sites," in *Proceedings of the Human Factors and Ergonomics Society annual meeting*, 1998, vol. 42, no. 4, pp. 424–428. doi: 10.1177/154193129804200405.

[108] K. Finstad, "The usability metric for user experience," *Interact Comput*, vol. 22, no. 5, pp. 323–327, 2010, doi: 10.1016/j.intcom.2010.04.004.

[109] D. L. Neumann and H. R. Westbury, "The psychophysiological measurement of empathy," *Psychology of empathy*, pp. 119–142, 2011.

[110] I. I. P. Oliveira and N. M. Guimarães, "Practical Neurophysiological Analysis of Readability as a Usability Dimension," in *International Conference on Human Factors in Computing and Informatics*, 2013, pp. 194–211.

[111] L. C. Landivar, "Disparities in STEM employment by sex, race, and Hispanic origin," *Education Review*, vol. 29, no. 6, pp. 911–922, 2013.

[112] K. Holtzblatt and N. Marsden, "Retaining women in technology," in *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 2018, pp. 1–8.

[113] V. Hunt, D. Layton, and S. Prince, "Diversity matters," *McKinsey & Company*, vol. 1, no. 1, pp. 15–29, 2015.

[114] K. W. Phillips, D. Medin, C. D. Lee, M. Bang, S. Bishop, and D. N. Lee, "How diversity works," *Sci Am*, vol. 311, no. 4, pp. 42–47, 2014.

[115] S. Bala and N. Khan, "Gender, Leadership, and ICT," in *Gender Perspectives on Industry 4.0 and the Impact of Technology on Mainstreaming Female Employment*, IGI Global, 2022, pp. 149–173.

[116] F. G. Aleu and E. M. van Aken, "Systematic literature review of critical success factors for continuous improvement projects," *International Journal of Lean Six Sigma*, 2016.

[117] A. Apraiz Iriarte and G. Lasa Erle, "Questionnaires for the evaluation of user experience and usability in industrial environments: A bibliographic review," *DYNA New Technologies*, vol. 7, no. 1, p. [21 p.]-[21 p.], 2020, doi: 10.6036/nt9845.