Evaluation of intelligent collaborative robots: a review
Miguel Da Silva, Remi Regnier, Maria Makarov, Guillaume Avrin, Didier Dumur

To cite this version:
Miguel Da Silva, Remi Regnier, Maria Makarov, Guillaume Avrin, Didier Dumur. Evaluation of intelligent collaborative robots: a review. IEEE/SICE International Symposium on System Integration (SII) 2023, Jan 2023, Atlanta, United States. pp.1-7, 10.1109/SII55687.2023.10039365 . hal-04095001

HAL Id: hal-04095001
https://hal.science/hal-04095001
Submitted on 11 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Evaluation of intelligent collaborative robots: a review

Miguel Da Silva¹, Remi Regnier², Maria Makarov³, Guillaume Avrin⁴, Didier Dumur⁵

Abstract—Intelligent collaborative robots or smart cobots can achieve high levels of flexibility by combining the human ability to adapt to new tasks with the performance of automated robots (precision, repeatability, etc.). This is a major innovation for Industry 4.0. Nevertheless, at present, cobots are not widely deployed in industry because they are difficult to evaluate and current standards for them are limited. The evaluation of collaborative tasks is difficult due to their specificities such as the variability of human behavior, artificial intelligence systems (e.g., smart cameras), and advanced control laws. This paper is a short review that aims to identify the methods and material resources needed to address this need for evaluation of intelligent cobots to foster their development and acceptability.

I. INTRODUCTION

Market research [1] shows that global sales of collaborative robots (or cobots) nearly doubled between 2017 and 2019. However, they are still a minority compared to conventional industrial robots, which accounted for 95% of global sales of industrial robot in 2019. The lack of standards to assess cobots is holding back their deployment. This statement is confirmed in the paper [2], it explains that robot manipulators are difficult to assess because of the "lack of guidelines for a realistic experimental setup". The evolution of the number of research papers available on the Science Direct database that mentions the terms “Human Robot Collaboration” and “Machine learning” in their title, abstract or keywords per year from 1996 to 2020 shows that the interest for intelligent collaborative robots is growing exponentially [3]. Nevertheless, it is important to be aware that the term "collaborative robot" can have different meanings. Most papers about cobots give their own definition of the word "collaborative" in the context of human-robot collaboration, sometimes through examples or sometimes more clearly, but there is no consensus on the definition of this term. This problem is also identified in [4]: “Several definitions exist in literature distinguishing collaboration from cooperation or interaction. For example, [5] notes that cooperation robots work with people step by step for a common goal, while collaborative robots work with people hand in hand on a common task. With a more lenient definition corresponding to the definition of cobot manufacturers, any robot operating alongside a human without a fence can be characterized as a collaborative robot [6]”. Several papers try to tackle this issue by proposing a global definition of "collaborative robots" or related concepts such as human-robot interaction (HRI). For example, [3] classifies robots according to their level of interaction. It proposes six levels of interaction by making an analogy with the standard levels of autonomy of autonomous vehicles created by the Society of Automotive Engineers (SAE): fully programmed, co-existence, assistance, co-operation, collaboration and fully autonomous. In this case, collaborative robots are those who belong to the level of interaction number 4 called “collaboration”. It means that a collaborative robot shares its workspace and works simultaneously with the human to achieve a common goal. Therefore, according to this definition, a robot that shares its workspace with a human and works on the same part at different times (e.g. not simultaneously) is not a collaborative robot but a cooperative robot. Mukherjee’s definition is incompatible with the one given by the ISO/TS 15066:2016 standard which does not make any difference between collaborative robots and cooperative robots. In this paper, the definition of cobot that is used, is the one the proposed by the ISO/TS 15066:2016 standard. It includes a wide variety of robots including automated guided vehicles (AGV) designed to perform fetching tasks [7], drones that are controlled by humans without any remote [8], exoskeletons [9] and robotic arms that work alongside humans [10]. We propose that the words collaboration and cooperation have the same meaning and we chose two terms to distinguish the different types of collaboration between humans and cobots: direct collaboration and indirect collaboration. Both of these terms are explained in the following section, which presents a taxonomy of cobots. The third section is about the technical requirements for collaborative robots that are considered in the selected research papers. To verify that these requirements are satisfied, the authors have applied evaluation methods that are discussed in section 4. To contribute to the creation of new standards, we compare a wide variety of methods used by different researchers to identify hindrances and promising methods. In the last section, thoughts on the prospects for the evolution of testing tools for intelligent cobot that comes from the literature analysis are presented.
II. TAXONOMY OF INTELLIGENT COBOTS

The purpose of the following taxonomy (Figure 1, 2, 3) is to define the terms used to characterize intelligent collaborative robotic systems. The words chosen for this taxonomy can have different meanings. To avoid ambiguities we propose the following glossary that details as much as possible the ambiguous terms of the taxonomy.

A. Glossary

**AI type**: type of AI algorithm that is characterised by its learning method (supervised, unsupervised, etc.) or by its architecture (symbolic, etc.).

**AI application**: task performed by the AI system.

**Collaboration type**: how the human collaborates with the robot to perform a collaborative task.

**Direct collaboration**: collaboration where the human and the robot interact on the same part at different times.

**Indirect collaboration**: collaboration where the human and the robot interact on the same part at the same time.

**Task**: common goal that human and robot try to reach in collaboration with each other.

**Co-carrying**: the robot and the human are carrying the same object at the same time.

**Gesture assistance**: hand-guided tasks where the robot reduces the effort required by the user to perform a task or constraints the user’s movements to help the user maintain a proper trajectory to perform the task.

**Handover**: the robot places an object in the human’s hand or holds it while the human interacts with it.

**Subsystem**: system that can complete a subtask that contributes to a collaborative task (ex: perception of obstacles with a 3D camera [11]).

**System**: hardware component that is necessary to complete a collaborative task: robot, camera or computer.

**Rapidity**: time necessary to complete a given task.

**Quality of human-robot interaction (HRI)**: global concept that can be described by several other concepts such as fluency, helpfulness, trust, flexibility, ergonomics, etc.

III. TECHNICAL REQUIREMENTS FOR INTELLIGENT COBOTS

In most of research papers about intelligent cobots, the requirements are implicit and there are no clear specifications. For example, [12], [13] and [14] aim to demonstrate that their system or algorithm work in a controlled environment. This approach highlights the relevance of using a new technology for a given collaborative robotics application, but it does not give enough guarantees on its performance to allow its commercialization. However, by observing what is tested during the evaluation, several high level requirements can
focused on the success of the task or the subtask that can be called “accuracy”. In this context, the accuracy of the visual motion detection is the ability of the robotic system to locate the pieces on the board and successfully detect the movements of each player. The success of all subtasks and their association is demonstrated by experimentation but this method is limited. Nevertheless, several papers such as [18] define precise metrics to explicit and measure the accuracy of their system. Their paper presents an image recognition AI algorithm that can recognize and locate an object on the cobot workspace regardless of how it is placed. This algorithm is designed to be integrated into collaborative production tasks in which the human deposits parts randomly in the robot’s workspace. To evaluate the accuracy of the AI algorithm, they use an accuracy score metric, calculated as the ratio of the number of error-free detections for each image to the total number of images. Other classical metrics in the evaluation of AI algorithms, such as F1-score, Jaccard and precision are also used in this paper. These metrics only guarantee the accuracy of the AI algorithm under certain conditions (parameters such as lighting) but are not sufficient to ensure the accuracy of the entire collaborative system. Evaluating the accuracy of several subtasks does not guarantee the accuracy of the whole system [2]. Moreover, making various tests varying the influence factor is necessary to guarantee the accuracy of the system in the largest number of use cases. This is the notion of robustness.

B. Robustness and resilience

Evaluating the robustness of an intelligent collaborative robotic system allows guaranteeing its performance on a large operating domain that takes into account the possible variations of the environment and system parameters. The robustness of all the cobot subsystems does not guarantee the robustness of the whole cobot. It still allows to identify influence factors that can disturb the robot and to make sure that critical subsystems are able to work under certain constraints. Consequently, there is a need to evaluate both the robustness of the subsystems and the robustness of the complete cobot. Nevertheless, in the literature, most papers in which collaborative robotic systems or subsystems are tested do not test their robustness. For example, [19] present a collaborative robot that sorts different objects based on their color and shape is tested, but the parameters of the test such as lighting conditions are fixed in advance and do not vary. Thus, if one those parameters vary, the behavior of the system is unpredictable. Some robots are designed to operate in controlled environments where most of the influence factors are controlled. For these robots, the required level of robustness is lower than the one required for autonomous robots that operates in unknown environments. Testing the resilience of a system takes the evaluation of the system one step further than a simple robustness test. It is the test of the system’s ability to continue operating outside its operating range (ex: in case of a component failure or a cyber-attack) [20]. Although robustness and resilience are not often evaluated in the case of collaborative robots,
they are evaluated a lot in the field of autonomous vehicles, which today is one of the most advanced fields in terms of evaluation techniques for complex and intelligent systems. That is why we believe that these requirements will be evaluated more and more for collaborative robots. Moreover, new evaluation tools and methods are being developed for collaborative robots, they are presented in the next section.

C. Rapidity

Rapidity is another important requirement. [21], [22] measure task completion time to see if human-robot collaboration has an impact on production rate. We noticed that this requirement concerns mainly pick and place collaborative tasks. Currently, for safety reasons, the velocity of cobots is limited by the ISO/TS 15066:2016 standard. This standard links the rapidity to safety requirements. However, safety requirements are currently very restrictive because of the lack of standards to test robustness of the systems. Thus, guaranteeing the robustness of a cobot can make the speed limitations imposed by the safety standard less severe and it could lead to an increase of the rapidity of the cobot. This statement shows once again the importance of the evaluation of the cobot’s robustness.

D. Quality of HRI

The quality of HRI is an essential yet difficult-to-assess requirement because it covers various concepts that are often subjective. Several papers try to evaluate this interaction in an explicit way [23], [22]. Human-robot interaction is a concept that cannot be measured directly but these papers have tried to demystify this concept and to quantify it. For example, [23] define human-robot interaction throughout several related concepts like robot’s trust, system usability, frustration, pleasure, satisfaction, and perceived physical workload. Moreover, [23], [24] take into account the quality of HRI through the ergonomics of the selected collaborative tasks. They are measuring the human health benefit of the collaborative robot to see if it improves human posture or reduces the loads carried by the human. Other similar concepts such as fluidity of human-robot interaction have been developed to describe HRI [25]. Thus, the demystification of the concept of quality of HRI made possible its evaluation even though it cannot be perfect because the related concepts are often subjective and the reproducibility of experiments is limited by the human factor.

IV. EVALUATION METHODS

In the scientific literature, most of the papers about the evaluation of cobots focus on the safety of cobots or on the HRI. However, there are also a few papers that use evaluation methods to test the whole intelligent collaborative robotic system or the subsystem.

A. Evaluation of safety of cobots

The safety of collaborative robots is one of the main barriers to the deployment of collaborative robots. This is the reason why the researchers are focusing mainly on this aspect. [26], [16] refer to the only safety standard that was designed specifically for cobots, ISO/TS 15066:2016, in order to guarantee the safety of their system. They perform physical tests to make sure that the system is compliant with this standard (i.e. that respects the constraints of speed and pressure on the human body stated by this standard). These constraints are expressed in the form of maximum values not to be exceeded according to the mass of the robot and the potential human/robot contact zones. However, as shown by [3], this standard is limited to robot related constraints but does not take into account the tool of the robot that can be dangerous (e.g. soldering iron) and the human related factors. Moreover, it does not take into account the danger related to the AI systems. For example, [27] present a collaborative robot that offers tools to a human to help with an assembly task and if an error in the AI system that plans the robot’s handover motion occurs then the human can be injured by a screw driver. That is why the evaluation of AI systems is essential to guarantee the safety of an intelligent cobot.

Methods for evaluating the performance of intelligent cobots can be used to evaluate several aspects of cobot safety, but not necessarily in accordance with ISO/TS 15066:2016.

B. Evaluation of HRI

HRI is one of the most complex aspects of cobot evaluation due to its subjectivity. Nelles et al. (2019) selected 30 relevant papers on HRI evaluation and described their evaluation method. These methods can be distinguished into two categories: subjective and objective.

1) Subjective evaluation methods: Subjective methods do not rely on measured data. They mainly rely on surveys that are answered by groups of users of the cobot or by experts who watch and interpret videos of users using the cobot. Besides, the result of these questionnaires are scores that correspond to the evaluation metrics. The scoring method depend on the type of questionnaire that is used, it is either personal with for example, scores from 1 to 5 for each question [28] or standard such as the NASA-TLX which aims a evaluating the perceived workload for the cobot user [29], [30].

2) Objective evaluation methods: Contrary to subjective evaluation methods, objective methods rely on physical data that is provided by sensors or measuring instruments. This data is mainly physiological, for example: the heart rate, the skin temperature, human energy consumption (kcal) etc. Such measures are often used to build metrics that quantify subcategories of HRI like anxiety [31], [32]. Instead of using scores from questionnaire results to evaluate the quality of HRI, they use physical measurements and try to correlate these measures with concepts related to HRI. For instance, [33] measured the trust of the human coworker in the cobot through the heart rate because they stated that there is a correlation between the stress level of the human and the heart rate. This kind of correlation allows to evaluate HRI but they are not exact and do not take into account several influencing factors and biases. Consequently, objective and subjective methods are not opposed but rather complemen-
C. Evaluation of cobot’s performances

The cobot performance is the ability of the robot and the human to perform a given collaborative task according to a set of specifications. If the specifications are guaranteed every time the robot and the human work collaboratively then the performance of the cobot is acceptable. This definition of performance includes safety and HRI, but as they were already discussed in the previous sections, this section focuses on task-related performance evaluation. There are various existing methods to assess the ability of a cobot and a coworker to achieve a common goal. These methods are sorted in four main categories: analytical methods, virtual tests, mixed tests and physical tests.

1) Analytical methods: In general, analytical methods do not concern the test of the entire cobot because modeling the behavior of such a complex system with all the constraints associated to the task that it is supposed to accomplish is impossible without the use of numerical methods. However, analytical methods are useful to evaluate elementary functionalities of cobots using simplified models and to prove the stability of the cobot in a given test configuration. Thus, it allows the cobot to pass a first validation step. For example, [9] proved the Lyapunov stability of the impedance control law under certain conditions. Nevertheless, when a cobot is equipped with an AI system, this approach is limited because most of them act like black boxes whose behavior cannot be predicted analytically.

2) Virtual tests: In contrast to analytical methods, virtual tests allows the prediction of cobot behavior in complex scenarios. They allow to numerically solve problems that are analytically difficult or even impossible to solve including those that use artificial intelligence. Thus, there are two types of virtual tests: a first one, closer to the analytical method, which consists in solving equations numerically without including a 3D model of the robot [35], and a second one which includes a 3D model of the robot and which aims at reproducing the scene of the cobot performing its task in a virtual environment [36]). Simulations without computer-aided design (CAD) models are used to assess specific cobot functionalities but they are not relevant to assess a cobot in complex scenarios. Nevertheless, they allow to evaluate several complex functionalities that involves AI such as task planning [37].

Virtual tests with CAD models allow to simulate realistic use cases of the cobots thanks to simulation softwares like Gazebo [19], RobWorkStudio [17], Tecnomatix [24], etc. This type of simulation is often called “virtual commissioning”. Currently, it is mainly used to validate prototypes but it can be unrepresentative of the reality, especially if the task performed by the human in the simulation is complex. Modeling humans for simulation purposes is still an active research area. However, several papers such as [19] show that while modeling and simulating a human interacting with a cobot is possible, it is still difficult to obtain reliable results with such models if the level of collaboration between the human and the robot is high. This puts into question the reliability of the simulation. Consequently, physical tests are sometimes required in addition to simulation in order to validate the system. The fact that the collaborative robotic system reaches the expected performance level in a simulation does not always mean that it will reach the same performances in the real life. However, if the system does not achieve the expected level of performance in the simulation, it is unlikely to achieve that level of performance in a physical test. In short, virtual testing is a powerful tool that can be used to test almost any cobot functionality and that is very efficient for prototyping but to validate an intelligent cobot in the case where the collaboration between the cobot and the human is direct and complex, this approach is limited. Its lack of realism is its main flaw.

3) Physical tests: There are two categories of physical testing: those performed in a laboratory in a controlled environment and those performed directly in the final use environment. Both have the advantage of being more realistic than simulation. Nevertheless, the number of use scenarios that can be assessed is limited because of the costs, the test time and the safety problems of several use cases. Actually, safety issues are particularly present in the physical evaluation of collaborative robots because their direct interaction with humans increases the risk associated with the cobot task. Therefore, evaluation of cobots in dangerous situations or near misses is difficult to implement. Consequently, several use cases cannot be evaluated properly with physical tests and that is why in the majority of cases, the robustness cannot be evaluated with physical tests. To overcome that issue, researchers often combine several testing methods. This allows to take advantage of the benefits of each method. Among the papers we analysed, 80% of those which use physical methods, such as [9], also use analytical or simulation to evaluate their system.

4) Mixed tests: Mixed tests are the least used among the sample of papers we analysed. They are still recent and sometimes difficult to implement. However, they have the advantage of offering a good compromise between the number of applicable test scenarios and the realism of the results. There are three main type of mixed tests for cobots: hardware in the loop (HIL), robot in the loop (RIL) and human in the loop (HuIL).

Hardware in the loop: [38] presents a simulation of a cobot arm close to the ones presented in the ”Virtual tests” section but instead of being only virtual, it integrates the real single-board computer module of the cobot that is used to process the control algorithm. This approach takes into account all the errors linked to the single-board computer module and makes the simulation much more realistic than a traditional simulation. It allows to test real algorithms in various use cases by taking into account the potential errors or disturbances related to the properties of the hardware that will run it.
**Robot in the loop:** RIL testing is still rare in the field of intelligent collaborative robotic systems. Nevertheless, this type of testing is developing in the field of autonomous vehicles. The latter is a critical application field in which the slightest error can lead to dramatic consequences with fatal accidents. Thus, for such a technology to be available one day on the general public market, its safety must be guaranteed and for that, it is necessary to demonstrate the robustness and resilience of these systems. This is often done through virtual or HIL tests, but these lack realism, so RIL or VIL (vehicle in the loop) tests have been developed. The principle is to immerse the real vehicle or robot in a virtual environment. It is the same principle as virtual reality for humans but applied to robots or autonomous vehicles. Only the operating environment is simulated and the robot or vehicle interacts with it via gateways that link its sensors to the virtual environment. In the case of the autonomous vehicle, screens are placed in front of the vehicle’s cameras, a computer that broadcasts the images of the virtual environment on these screens is placed in the trunk of the vehicle and by making the car drive in a large empty space, one can test the behavior of the vehicle on any virtually generated road (intersections, traffic circles, etc.). This is a promising evaluation method but it is still underdeveloped in the field of collaborative robots and might be unaffordable for most evaluators.

**Human in the loop:** During the HuIL assessment, the human is immersed in a simulated virtual environment using a virtual reality headset, a human motion detection system and sometimes wearing haptic devices. The human can interact with the virtual environment, for example pick a part and place it on a table near the cobot in the virtual world. HuIL evaluations overcome one of the main obstacles to the simulation of collaborative tasks, which is the modeling of human behavior. A significant advantage over physical testing is the ability to test dangerous or near miss use cases and its reproducibility is better than 100% virtual simulations. This is the evaluation method that was selected in [24] to test different safety devices to protect the human of a cobot in two different cases. The first one concerns a mechanical parts cleaning task in which the human’s role is to replenish the cobot’s cleaning tool, which is disposable. The second one concerns a printed circuit board (PCB) handling task in which the human has no role but for which the robot has to stop when he approaches to avoid injuring him with sharp PCBs. In both cases, dangerous situations could be tested and new sources of danger could be discovered without the human taking any risk. Thanks to the information provided by this simulation on the robot’s behavior, the algorithms that define the robot’s behavior in the presence of a human could be adjusted to make the HRI safer. However, even if this method does not model the human’s behavior, it still requires modeling the interaction between the human and the cobot in the virtual environment. This means that for cases where the collaborative task is complex, the simulation is likely to be unrealistic. This method is therefore optimal for evaluating systems where the human behavior is not predictable and difficult to model and where the human-robot interaction is not complex. This is often the case for autonomous vehicles, the equivalent in this domain is the driving simulator, which allows to acquire data on human and to adapt the vehicle behavior to each driver.

**V. CONCLUSIONS**

In this paper, a taxonomy was developed to clarify the notion of collaborative robotics and to introduce different notions related to evaluation. Through this literature review, we have identified several gaps and areas of progress for the evaluation of cobots. We noticed that robustness and resilience of a full collaborative robotic system are poorly evaluated. This is partly due to the fact that most of the cobots mentioned in the literature are proofs of concepts or prototypes that are not intended to be used in an industrial setting. It is therefore an interesting research area to explore. The progress of simulation software and the computing power of computers allows to set up numerical models which are more and more faithful to reality and which can provide information on the robustness of the intelligent cobot. Moreover, they give rise to new types of tests that we have called ”mixed tests”. These tests are still at an early stage development and are sometimes difficult to implement. They require a lot of equipment as well as direct access to the sensors of the cobots which are not always possible. However, we believe that they represent the future of simulation because they have the advantage of having the best compromise between exhaustiveness and realism of test scenarios. Through our analysis, we also identified the lack of linkages between the different evaluation methods. We believe that the different tests (physical, virtual, etc.) can be correlated. For example, we can imagine an approach where a collaborative task would be evaluated through a mixed test and then reproduced in a real environment in order to validate the mixed testbed which could then be used without physical testing afterwards.

**REFERENCES**


