INDUSTRIAL HUMAN-ROBOT COLLABORATION

Verneri Hakkarainen

Aalto University
School of Electrical Engineering
Automation and robotics

Industrial Human-Robot Collaboration

Verneri Hakkarainen

Bachelor’s thesis

Espoo 30.5.2025

Supervisor

Prof. Pekka Forsman

Advisor

Dr. Tsvetomila Mihaylova

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

TEKNIIKAN KANDIDAATINTYÖN
TIIVISTELMÄ

Tekijä Verneri Hakkarainen

Työn nimi Industrial Human-Robot Collaboration

Koulutusohjelma Automaatio ja robotiikka

Työn valvoja Prof. Pekka Forsman

Työn ohjaaja Dr. Tsvetomila Mihaylova

Päivämäärä 30.5.2025

Kieli Englanti **Sivumäärä** 20 + 5

Tiivistelmä

Ihmisen ja robotin yhteistyöllä (human-robot collaboration, HRC) tarkoitetaan tilannetta, jossa robotti suorittaa työtä samanaikaisesti ihmisen kanssa. Robottikäsivarsia voisi käyttää erilaisten kokoonpanotöiden avustajana. Jotta tämänkaltainen yhteistyö olisi mahdollinen, täytyy robotin pystyä tulkitsemaan ja mallintamaan ihmisen käytöstä. Sopivilla mallinnusmenetelmillä robotti kykenee havaitsemaan ihmisen käytöstä ja eleitä ja tulkitsemaan niiden perusteella ihmisen aikeita sekä tulevia liikkeitä. Ihmisen seuraavia eleitä voidaan arvioida havaitsemalla ympäristöä. Esimerkiksi katsomalla mitä ihminen pitää kädessään. Ihmisten käsiä seuraamalla voidaan määrittää mihin pisteeseen käsi on kurottamassa ja ennakoiden ojentaa siihen pisteeseen työkalu. Ihmisen katsetta seuraamalla voidaan päätellä, minkä kokoonpanon osan hän seuraavaksi tarvitsee.

Robotin ja ihmisen sujuva yhteistyö edellyttää myös sujuvaa kommunikointia ja työn suunnittelua. Kommunikointi voi tapahtua ihmiselle luontaisesti verbaalisti tai muilla tavoin kuten käsieleillä tai visuaalisesti lisätyn todellisuuden laseilla. Robotin on tärkeä osata työn vaiheet ja suunnitella suunnitelma työnsuoritusta varten ottaen mukaan ihmistyöntekijän rajoitteet, vahvuudet sekä mieltymykset. Robotti kykenee mukautumaan ja muuttamaan suunnitelmaa työn edetessä.

Tämä työ on kirjallisuuskatsaus ihmisen ja robotin yhteistyöstä. Työssä tarkastellaan ihmisen mallinnuksen menetelmiä sekä ihmisen aikeiden ennakoimista. Työn aihealueeseen kuuluu myös kommunikointi ihmisen ja robotin välillä, joka mahdollistaa ohjeiden antamisen robotille sekä työvaiheista keskustelemisen. Tässä työssä tarkastellaan tutkimuksia, joissa ihmisen mallinnusmenetelmiä sekä ihmisen ja robotin kommunikaatiota on tutkittu. Työn tavoitteena on tutkia, millä menetelmillä ihmisen ja robotin yhteistyö on mahdollista ja miten sitä voi soveltaa teollisuudessa. Työssä selvitetään, mitä etuja ja haasteita ihmisen ja robotin yhteistyöllä on teollisuuden sovelluksissa. Työn lopussa esitellään kuvitteellinen esimerkkitapausta ihmisen ja robotin yhteistyön hyödyntämismahdollisuudesta.

Kirjallisuusselvityksen perusteella ihmisen ja robotin yhteistyö tarjoaa monia etuja. Kokoonpanotöiden tuotantoa voidaan lisätä samalla vähentäen ihmisen fyysistä taakkaa, edistäen ihmisen jaksamista ja hyvinvointia. Ihmisen ja robotin yhteistyö mahdollistaa sellaisen työn automatisoinnin, joka vaatii ihmisen luovuutta ja sopeutuvuutta. Kuitenkin haasteena ihmisen ja robotin yhteistyössä on sen vaatima suuri laskentatehon määrä. Robotin on laskettava ihmisen liikerata tarpeeksi nopeasti, jotta se voi välttää törmäyksen. Suurin haaste on kehittää tarpeeksi turvallinen ja luotettava yhteistyörobottijärjestelmä. Ihmisen ja robotin yhteistyön lisääntyvä tutkimuksen määrä sekä nykyisistä tutkimuksista löydetyt teollisen käytön hyödyt osoittavat ihmisen ja robotin yhteistyön olevan lupaava ratkaisu moneen teolliseen työhön.

Avainsanat: ihmisen ja robotin yhteistyö, teollisuuden robotiikka, ihmisen mallinnus, yhteistyörobotti, ihmisen ja robotin kommunikaatio

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
BACHELOR'S THESIS

Author Verneri Hakkarainen

Title Industrial Human-Robot Collaboration

Degree program Automation and robotics

Supervisor Prof. Pekka Forsman

Advisor Dr. Tsvetomila Mihaylova

Date 30.5.2025

Language English **Number of pages** 20 + 5

Abstract

Human-robot collaboration can be a powerful tool when applied in industrial settings, such as manufacturing. The robot co-operates with the human on the same task and uses intention prediction and communication methods to make the collaboration effective. The robot can benefit from the same social cues that humans use when collaborating with other people. The robot predicts human actions by using human movement modeling, gaze, gestures and surroundings as a guide. Using human intention prediction techniques increases fluency and makes collaboration seem more natural. Additionally, effective human-robot collaboration requires communication between participants. The robot can use speech or visual communication like augmented reality, or many communication methods simultaneously. Effective communication enables the human and the robot to discuss crucial task information like a task completion plan that the robot can generate. The robot can create human centered task plans that reduce human physical effort and emphasize human skills. This thesis is a literature review that examines human modeling and intention prediction methods and communication methods. This thesis also reviews human-robot collaboration applied in industrial settings. The review found that human-robot collaboration offers many benefits in industrial applications by combining human creativity and flexibility with robot efficiency. The robot can increase the throughput of manufacturing by decreasing work task completion times and reducing human fatigue. A major difficulty in human-robot collaboration is implementing a safe collaboration robot to use near humans. The study of this field is gaining interest, and many studies indicate that human-robot collaboration offers benefits in industrial applications, making the potential look promising.

Keywords: human-robot collaboration, industrial robotics, human modeling, intention prediction, collaborative robot, human-robot communication

Table of Contents

Abbreviations

HRC	Human-robot collaboration
VR	Virtual reality
AR	Augmented reality
LLM	Large language model
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
STGAIN	Spatial-temporal Graph Attention Informer Neural Network

1. Introduction

Humans can predict the actions of other people by observing them and use effective communication to coordinate work. These fundamental skills allow humans to effectively collaborate on tasks such as building tables, assembling industrial parts or assisting in everyday tasks. Such human-to-human interactions have been found useful in robotics when considering human-robot collaboration [1] [2]. Human-robot collaboration can be used to enhance manufacturing processes by combining the adaptability of a human and the efficiency of a robot. [3].

There are many aspects to consider in implementing human-robot collaboration systems. Depending on the scenario and the given task, the robot might need to predict human actions to better adapt to humans. This is called human intention prediction or human modeling. Intention prediction can be used to avoid collisions and to increase task fluency [4]. Additionally, the robot can model and predict human actions using gaze, gestures and the surroundings. In human-human collaboration, communication plays a major role. The same applies to human-robot collaboration where communication can be achieved with speech, augmented reality and combining different communication methods. Having a clear and shared understanding of a task plan is important before collaborating with anyone. The completion of a shared human-robot collaboration task requires communication between the human and the robot on the task plan and the current task process state [5]. The robot can design a task plan favoring human skills and minimizing physically demanding task work from the human.

This thesis examines human-robot collaboration in industrial applications. The study is conducted as a literature review using recent papers to examine human-robot collaboration. Specifically, the research focuses on human intention prediction and communication methods used in human-robot collaboration. Furthermore, this paper aims to get a broad overview of using collaborative robots in industrial applications and to point out study gaps and new directions in which the research could proceed. This paper looks at different human-robot collaboration systems from different papers to construct an idea of what human modeling and communication methods are needed for successful human-robot collaboration. This paper will not inspect the methods in major detail. In chapter two, this paper examines human modeling and human intention prediction. The third chapter looks at communication methods used in human-robot collaboration. The third chapter also includes generation and communication of shared task plans. The fourth chapter considers human-robot collaboration in industrial applications. The benefits and drawbacks of using a collaborative robot are evaluated. The fourth chapter additionally includes an example of a human-robot collaboration system that uses multiple communication and human modeling methods to achieve successful collaboration. Chapter five summarizes the study and points out directions in which the study of human-robot collaboration could proceed.

2. Human modeling in human-robot collaboration

Accomplishing effective human-robot collaboration requires a robot understanding a human partner's intention from many social cues such as posture, movement path, gaze and gestures. A common human intention prediction application can be seen used in self-driving cars. Self-driving cars use human intention prediction methods to predict whether a human is going to pass the street or will come into the trajectory of the car [6]. This chapter divides physical human intention prediction into categories of features that are modeled. First, we will look at human motion prediction and then move on to gestures, context recognition and gaze recognition

2.1. Motion prediction

Motion prediction is an obvious application of human intention prediction. The robot could model the human's path or limb movements to predict intended walking path or next hand position. It has applications in safety when considering mobile robots. For example, a study by N. Zapata et al. (2024) [7] applied human intention prediction in caregiving robots. The robot estimated a patient's walking path and when deriving that the patient will collide with an object in the room, the robot pointed out the object by moving near it.

Y. Tao et al. (2025) [4] conducted a study where they built a model for safe mobile robot manipulation. The robot is a mobile robot with wheels and a robot arm attached to it. The robot works in the same space as humans and thus needs to predict human movement to avoid collisions. Figure 1 shows the robot delivering a tool to a worker. The robot monitors its surroundings with a laser scanner used for determining position and an RGB camera which captures information of the surrounding scene. Additionally, the work area is monitored with depth RGB cameras. These cameras monitor obstacles and human movement. The sensory data is then interpreted using Long Short-Term Memory (LSTM) network. LSTM network retains movement data and adjusts attention to important parts of the human movement using data history. LSTM network can utilize scene data, further enhancing its prediction accuracy. LSTM network was used in the study to predict human movement up to one second in the future. This further enhanced the robot's ability to avoid collisions. The model was trained by incorporating data from two humans performing repeatable tasks in the workspace. Data from one individual was kept as test data. The implemented method accurately predicted human movement with an accuracy of 83.1% and simulated work scenario showed a success rate of 90% which is significantly higher than the methods the study compared it to. When the robot approaches a human, it drops the speed it moves at and when the robot is working, it monitors humans near it to avoid collisions with the robot arm. Four real-world experiments were conducted where the implemented method allowed the robot to safely complete given tasks. Y. Tao et al. state in their paper that the implemented framework could be used in workshops where workstations are clearly defined, such as electronic product assembly. However, the study states that for successful implementation of the framework, proper configuration is needed, and further study is required.

Figure 1 The mobile robot delivering a tool to a worker [4] (modified)

A study conducted by Y. M. Halim et al. (2025) [8] implemented model to predict human upper limb movements. The method focuses on handover tasks of robot arms in collaborative task scenarios. The handover task using the right arm can be seen in figure 2. The model utilizes physics in its predictions, opposed to only learned movement data which might lead to the model not taking physics into account. The model is taught by applying inverse kinematics to human arms. The model utilizes motion capture data with neural networks. Long short-term memory is used to predict joint pose coordinates alongside with recurrent neural network. The experiment used an industrial arm robot to give objects to human hands. A camera captures human hand joint coordinates and velocity. The robot predicts human hand motion and determines whether the human is reaching, handing an object or retracting hand. After predicting and planning, the robot moves accordingly to make the handout more natural and smoother. Utilizing the predictive model decreased handover time from 5.5 s to 5.1 s. Furthermore, users reported the interaction using the model as more natural and efficient opposed to without the model. The model showed an accuracy of 90.17% in predicting human arm position.

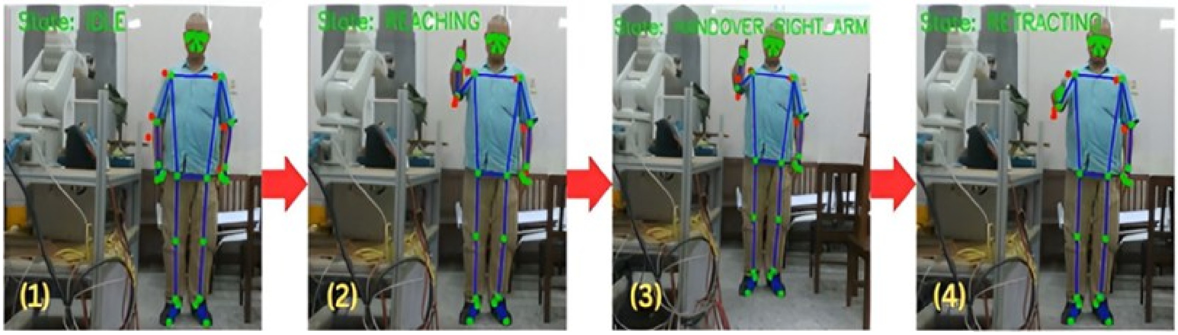


Figure 2 States of interacting with the robot using the right arm [8]

2.2. Gesture recognition

Gestures are an important aspect of human communication. Utilizing gestures in human-robot collaboration has been researched increasingly. Allowing humans to use hand gestures to point at objects or head nods to communicate simple ideas makes the collaboration experience more natural. There are many ways to gather sensory information of gestures such as, depth camera, wearable devices and an RGB camera. In general, gesture recognition consists of collecting sensory information of the gesture, recognizing what the gesture is and then classifying it to infer the meaning of the gesture. After gesture recognition the robot can use this information in the human-robot interaction. [9]

Q. Gao et al. (2017) [10] conducted a study on static hand gesture recognition for space human-robot interaction. The study utilizes parallel convolutional neural networks CNN. CNN translates images of gestures into gesture classifications that can be semantically identified. The gesture classification method utilizes two sets of sensory data: an RGB camera and a depth camera. These two sensory inputs are classified parallel using CNN and then fused, leading to increased gesture recognition accuracy. Figure 3 shows the parallel CNNs. The accuracy of this gesture recognition method is 93.3%. The experiment was conducted as a simulation with a gesture dataset containing 120000 hand gesture images. The paper stated that the given recognition method could be utilized by an astronaut to control a space robot.

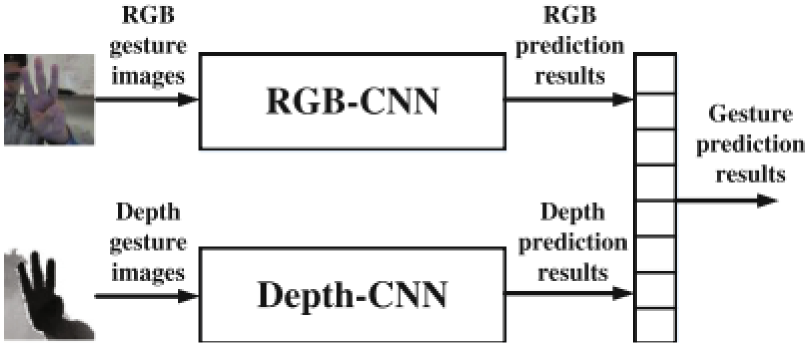


Figure 3 The structure of parallel CNNS [10]

Gestures could also be made by touching the robot. D. Jung. et al. (2024) conducted a study [11] where they incorporated touch controls in a HRC assembly task. In the experiment a deep learning model called gated recurrent unit was introduced and combined with convolutional neural network to learn and detect touch gestures. The model was trained with 5 different gestures for all the robots' 7 links. The gestures are dial, shake and tap. Tap has 3 variations: One tap, two taps and taps more times than two. The setting with many taps is used as a signal for emergency stopping. Taps are designed to be used for simple controls like start or pause. Dial gesture is done by rotating one joint to any direction by the joint. It could be used to input a value from a certain range. Shake gesture is done by grabbing a robot link and shaking it. The shake gesture could be used for example to signal task completion. The touch gestures can be seen in figure 4. In the experiment the robot and the user assembled a tiny chair. At the start of a task the user would tap to indicate to start the robot to start the task. Taps were used to move the chair when the robot was holding it. Shake was used to finish the task. The experiments showed an accuracy of 96.94% for the touch gesture recognition model.

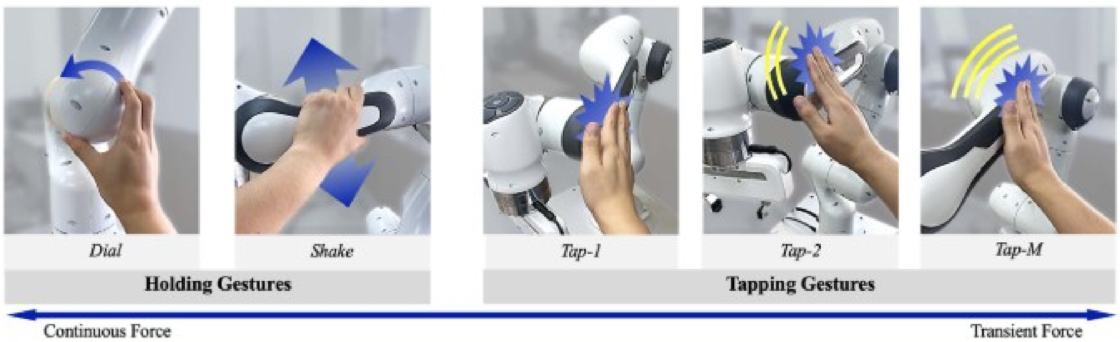


Figure 4 Touch gestures of a collaborative robot [11]

2.3. Utilizing context recognition in human intention prediction

Context recognition offers significant improvement in human intention prediction. In HRC a contextual image of the surroundings could be formed by recognizing objects and how they relate to humans. For example, holding a hammer near wood material could indicate that the user is going to hammer a nail. Although, context recognition might be inadequate in scenarios where objects and surroundings have many purposes, context recognition can be utilized to reduce uncertainty. [12]

Tong et al. (2025) [13] conducted a study where they built a model to predict humans in more common household situations. The model is based on scene graphs, and the study focuses on human intention prediction in complex scenes. The presented model was trained with videos of humans doing everyday such as watching television, eating, cooking, drinking and other common tasks. The videos were cut before the human did a certain action to test the predictive ability of the model. To represent the world, the model extracts objects and humans and then creates a dynamic scene graph of them. This dynamic scene graph uses nodes to represent objects and humans and edges to represent relationships. The relationships were categorized into human gaze, positional relationships and contact relationships. These relationships were used to determine whether the human is interacting or might interact with an object. After constructing a view of the world, the conclusions of human intentions must be made. For this purpose, Tong et al. utilize Spatial-temporal Graph Attention Informer Neural Network (STGAIN). STGAIN is used to analyze the dynamic scene graph. The STGAIN model learns the graph nodes and calculates attention coefficients from the relationships which are used to extract different likelihoods of prediction outcomes. Figure 5 shows a frame of the experiment and the corresponding graph. Tong et al. state in the paper that this method was more accurate in human intention prediction than similar counterparts. The model presented 81% accuracy rates in predicting what the human is going to do next. The model had difficulty telling apart tasks that were similar and had the same objects in them such as food when cooking or eating. Overall, the method was capable of interpreting complex scenarios of human actions and predicting human intentions.

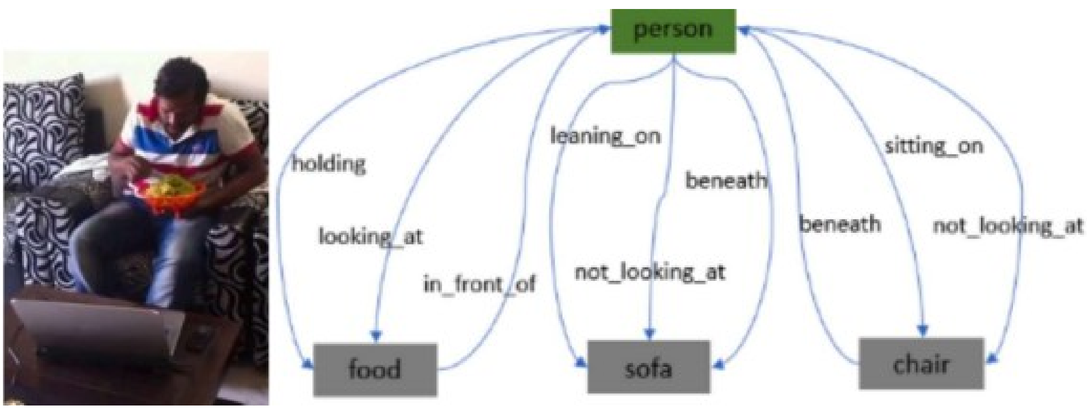


Figure 5 A frame from the experiment and the corresponding graph [13]

S. Li et al. (2024) conducted a study [14] in which they utilized a similar context recognition model. They demonstrated the model in a disassembly task of electronic vehicle batteries. Disassembly of the batteries includes difficult subtasks that require a human operator, such as cutting wires and removing glue. However, using a robot in certain subtasks that require picking and placing battery cells increases safety. The robot is a typical robot arm. The knowledge graphs recognize the human, tools, actions, commands and parts and then learns relationships between them. The method was able to recognize objects and relationships in 11 different tests with an accuracy over 90%. The paper states that the same approach could be additionally used for other HRC scenarios.

2.4. Gaze recognition

When collaborating with others, gaze plays a major part in anticipating the actions of others. From gaze humans can see where the other person is looking and anticipate predict actions. A study conducted by V. Belcamino (2024) [15] utilized eye movement tracking in assembly. The eye tracking was conducted with wearable eye tracking equipment and the gaze point was determined in an online simulation in unreal engine. A human had to assemble a table, and a robot handed parts for the human. The robot predicted human intentions using human gaze. When the humans gaze was focused on the outside of the workspace the robot knew to retrieve more pieces and when the focus was on the human's workspace, the robot didn't have to retrieve more pieces. The eye tracking approach was compared to an approach utilizing wearable motion sensors. Both approaches performed similarly and the user reports state that experiences are very similar. The paper states that human intention prediction using gaze tracking could be utilized as an enhancement along already existing human intention prediction methods.

M. Rakovi´ c et al. (2024) [16] conducted a study where they analyzed gazes of two collaborating agents. In the study the agents had to construct a tower from pieces in turn. Roles were assigned to the agents and one of them was a leader and one follower. The roles switched for subsequent actions. The goal is for both agents to construct their tower from pieces in front of them. The leader picks a piece and puts it in their tower. When the piece belongs to the follower's tower the leader gives it to them. Figure 6 shows the collaboration setting. The study found that in human-human collaboration the leader's action can be predicted from gaze fixations while followers gaze was fixating on the leaders' actions. A Gaze Dialogue Model was constructed based the human-human collaboration observations. The model utilizes hidden Markov models and predicts human intention based on gaze. The model was able to correctly predict actions. However, the correct classification happened at 60% task completion for giving and 80% completion for placing. The study additionally conducted experiments with human-robot collaboration. The used robot is a humanoid robot, and it utilizes the gaze dialogue model to predict intentions from human gaze and mimic gaze with artificial eyes. The model was able to predict intentions in human-robot collaboration at an average of 75% completion of the task. However, this implementation requires humans to wear eye tracking glasses and the robot's arm speed couldn't keep up with the robot's gaze leading to less natural feeling interaction.

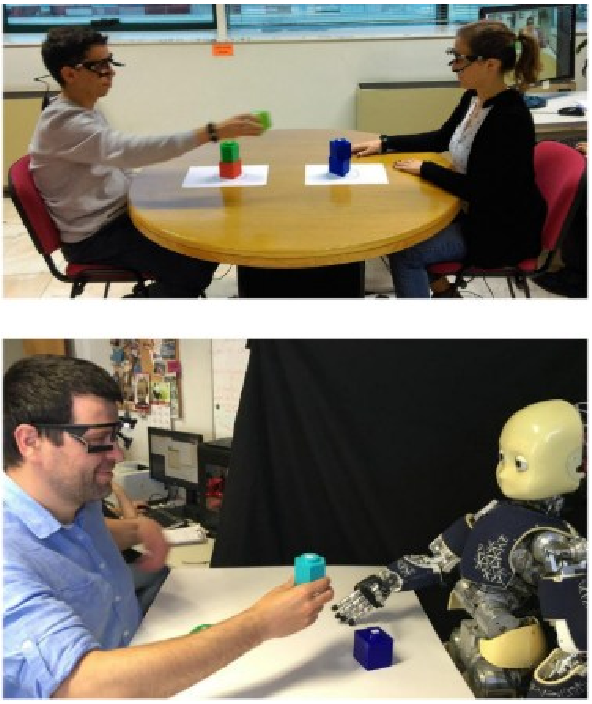


Figure 6 Collaborative setting of a tower building experiment utilizing eye gaze tracking. [16]

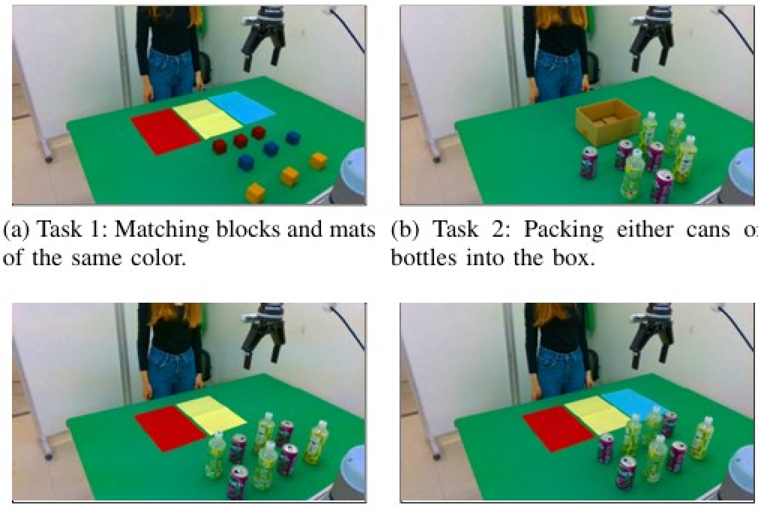
3. Information sharing in human-robot collaboration

Effective collaboration requires a shared understanding of what needs to be achieved among participants. In this chapter, sharing information between humans and collaborative robots is discussed. In the first subchapter, different ways of communication among humans and robots are examined. The second subchapter talks about communicating and generating shared task plans between humans and robots.

3.1. Human-robot communication

Apart from human intention prediction and modeling, straightforward communication can offer benefits to human robot collaboration. One possible solution is using human language as a means of communication. Using speech, a human could give instructions to the robot, and the robot could inform the user about the current and the next robot task. This kind of communication frees up the human workers' gaze and hands to continue the task while communicating. However, speech communication is ineffective in noisy environments. [17] Communication effectiveness can be increased by combining different communication methods. For example, clarity of communication can be increased using gesture commands alongside speech. When the user asks the robot to give an object, the user can point at it and say: "give me that". [18]

Martin et al. (2025) [19] conducted a study in which they designed a system for multimodal communication and task planning. The system uses a large language model (LLM) to interpret task information and to construct task plans. Furthermore, the system recognizes speech and gestures as means of communication. The human can interrupt and instruct the robot during false operation. The human can talk to the robot and point at objects. In the experiment the robot must perform tasks without prior knowledge based on human instructions and interpreting images captured by a camera. The robot used is a typical arm robot with a gripper and a camera. Robot system performance is evaluated using simple sorting and packing tasks that can be seen in figure 7. The evaluation of effectiveness of multimodal communication was conducted with task 3 seen in figure 7. In the task the human-robot team must sort cans based on color. After the experiment the participants filled out a survey regarding the experience of the experiment. The experiment showed that using multimodal communication improved collaboration on trusting the robot, willingness to use the robot again, feeling of control over the robot actions and feeling of successful collaboration. The use of multimodal communication also showed a significant increase in correct task completion and correct object placement.



(a) Task 1: Matching blocks and mats of the same color.

(b) Task 2: Packing either cans or bottles into the box.

(c) Task 3: Sorting bottles on the yellow mat and cans on the red mat.

(d) Task 4: Sorting bottles on the yellow mat, cans on the red mat, and leaving the blue mat empty.

Figure 7 The experiments used to evaluate the performance of the multimodal communication system [19]

Large language models (LLM) could be used to interpret human commands and to give robots a commonsense reasoning, such as in the previous multimodal communication experiment. However, further research needs to be conducted for utilizing LLMs in human-robot collaboration to increase safety and reliability. [20] This recent study [21] used a LLM for communication between human and a robot. The goal of the experiment was to determine the effects of a collaborative robot being able to converse in small talk while working. The experiment mimicked an industrial assembly scenario with a robotic arm that assists a human worker in assembling PVC pipes. The human worker verbally asked the robot to deliver different assembly pieces. The experiment was conducted in two ways. One implementation only communicated functional information and other communicated functional information and small talk. Small talk responses were generated with LLM model. The study found that including small talk in HRC increased task duration time. However, the paper states that the lengthened duration might be because the participants were surprised by a talking robot. Furthermore, some participants said that small talk made the experiment more enjoyable.

Augmented reality (AR) can also be utilized as a means of communication between humans and robots. AR has its benefits in providing visual information in the real-world environment. Wearable AR head mounted displays can also provide hand and eye tracking. [22] A study conducted by R. S. Lunding et al. (2025) [22] experimented on incorporating AR in an industrial HRC scenario. The experiment was conducted in a real factory task. In the experiment, workers who were familiar with the work task collaborated with a robot arm on the task. The task was completed in two ways: a robot helping with some of the subtasks and one where the robot acted as a mount for the assembly part. Assembly instructions and a list of parts and tools were imported to the AR interface. The interface shows the worker information about the workers' current task, robots planned movement, place where the next part will be inserted and the robot's current task state. The view with the AR gear can be seen in figure 8. The overall experience of workers with the robot was positive and all study participants were able to complete the tasks with the robot. The study also found that AR assistance can be beneficial in training or giving instructions in rare assembly tasks and quality control. However, prolonged use of AR head mounted display can be uncomfortable. Longer and more broad study must be conducted to determine overall effectiveness of AR assisted HRC.

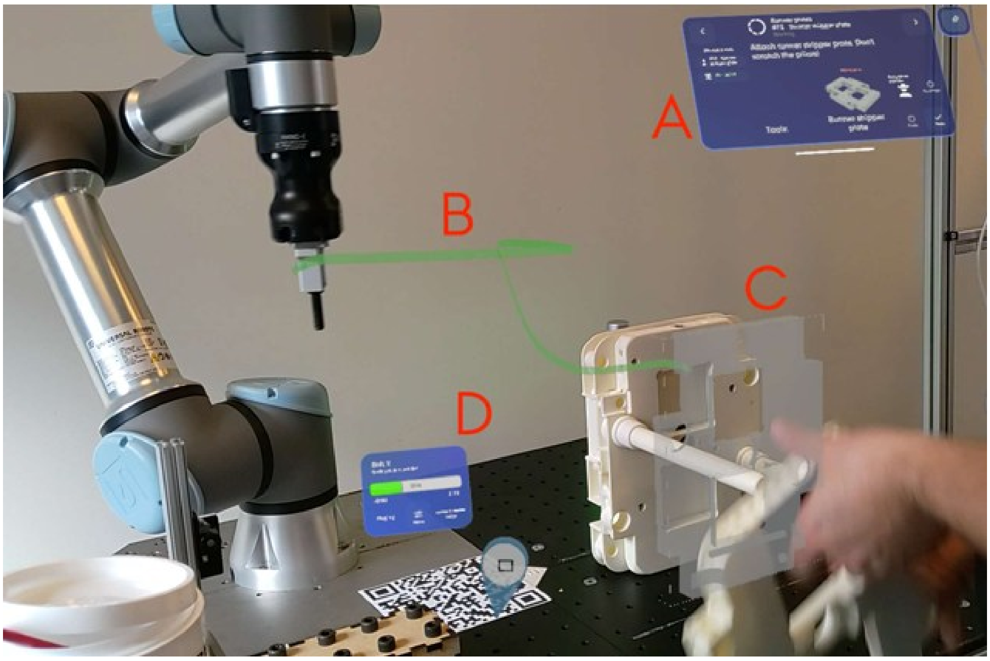


Figure 8 View of the AR interface. A shows a step overview of the current task. B shows the robot's path. C is an animated hologram of the next part to be placed. D is the robot's status that displays the progress of the robot's task. [22]

3.2. Collaborative task execution planning

Collaborating with humans is not straightforward. Something will likely go differently than in the general task plan. Different people like to do the same kind of work tasks differently. Completing tasks with humans requires dynamic task planning from the robot during the task completion [23]. This way the robot can adjust task plan and sequencing during work based on human errors and human working preferences.

A recent study conducted by Zhao et al (2025) [24] presents an implementation of safety aware task planning. The method aims to balance safety and efficiency while planning task execution sequences. The algorithm plans the robot and the human to execute separate tasks simultaneously or in turn. Some tasks are executed collaboratively. The algorithm tries to optimize efficiency of these task sequences while taking safety into consideration. Safety is determined by how close the human and the robot operate when executing tasks. When closer to the human, the robot must move slower to decrease the risk of injury upon collision which slows down the task completion. The paper conducts an experiment using the method. In the experiment, a human and a single robot arm assembled a desktop computer. Figure 9 shows the experiment setup. They experimented with different task sequences where the human and the robot installed different sets of parts every time. They found that the use of the task planning method to optimize safety and efficiency effectively reduces total assembly completion time. The most effective assembly was achieved by planning the robot to operate at tasks farther away from the human, allowing the robot to operate at maximum speed.

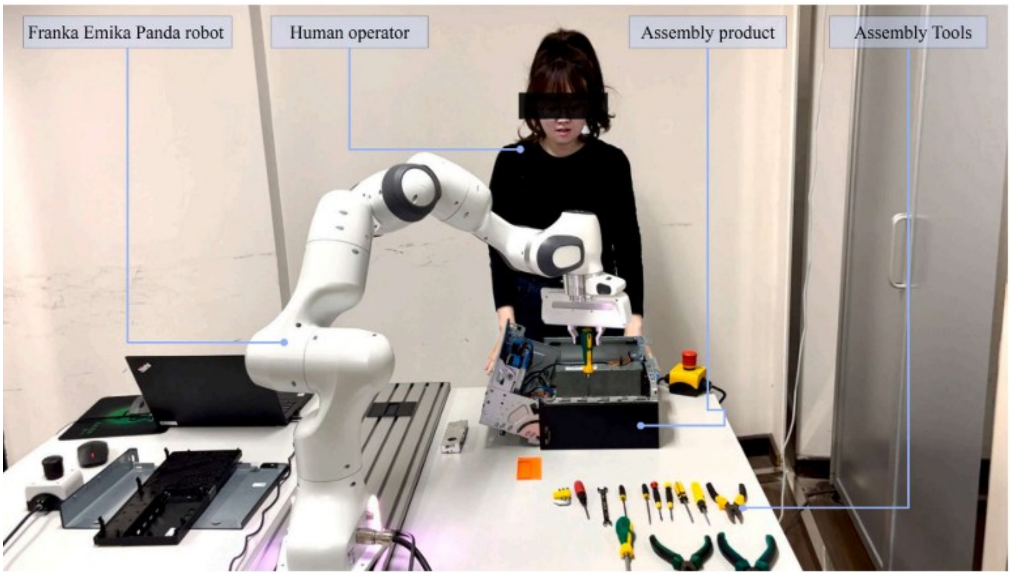


Figure 9 The setup of task sequencing experiment. The human and the robot collaboratively assemble a desktop computer. [24]

There are also other aspects to consider in co-task completion such as human task completion time and human preferences. For example this paper [25] introduces a method which plans robot execution times based on human task execution ability to make collaboration more fluent and effective while maintaining safety. The method decreased execution time and increased safety distances between the robot and the human. Another related paper by C. J. Conti et al. (2022) [26] introduced an experiment in which a robot collaborated with a human in vehicle assembly. The goal of the study was to demonstrate how common reasoning in robots would benefit human-robot collaboration. The knowledge of the robot is used in selecting parts based on their properties to make collaborative assembly more effective and more enjoyable for the human. The robot knows that humans prefer carrying lighter parts and picking up parts that are closer to them. The robot fulfills human preferences by carrying heavier objects and parts that are farther away. The robot also handles more fragile parts that are hard to handle for humans. Objects have attributes that are used with human preferences to calculate the next part to be retrieved. The algorithm was tested in a simulation and in a small-scale real-world experiment with a toy car. The experiment setup can be seen in figure 10. The experiments concluded that the robot working alongside the human will decrease completion time since the human experiences less fatigue with the robot's help. The solution is stated to be viable in long and repetitive assembly tasks. However, wider studies must be conducted to get a deeper understanding of the benefits of task planning while taking human preferences into account.

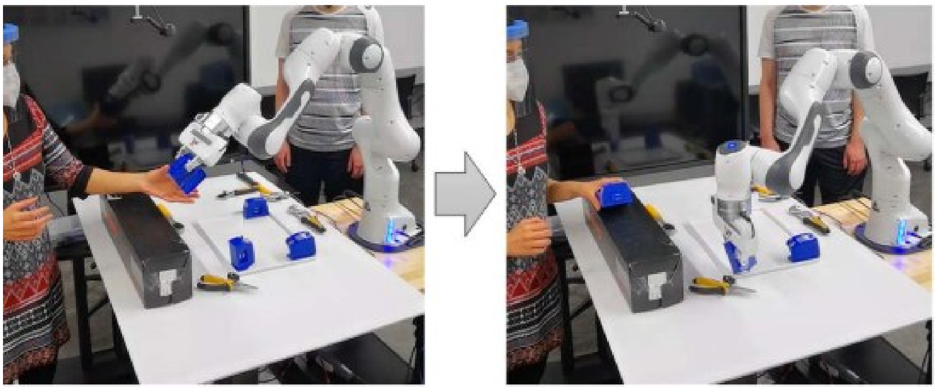


Figure 10 Experimental setup of assembling a toy car using task planning that considers human preferences [26]

4. Human-robot collaboration in industrial robotics

Human-robot collaboration has many benefits in industrial robotics and manufacturing. Utilizing collaborative robots in manufacturing can improve the adaptability of manufacturing due to a human being able to conduct some of the tasks [27]. Collaborative robots could be used in welding where the robot could assist by doing the welding itself or holding and handling the weldable part while a human does the welding. Collaborative robots could additionally be utilized in assembly for products that require customization [3]. In the paper regarding AR assistance in HRC [22], the test subjects are factory workers and stated that the collaborative robot could be used in assembly training and quality assurance. The robot verifies correct assembly steps while the assembly is in progress, saving time. Collaborative robots can increase human workers' physical wellbeing by doing physically hard and demanding tasks in place of the human worker. This is achieved with human centric task planning. [28] This kind of planning could also enable workers to work longer without getting tired [26]. This could increase the throughput of the manufacturing line. Furthermore, opposed to traditional industrial robots, collaborative robots do not need to be put in safety cages, saving up space [3].

One of the key issues with collaborative robots is the increased need for computing power which can be a problem considering scalability of these systems [29]. Furthermore, the demanding calculations need to be done within real-time constraints to ensure safety and avoid collisions with the human [28]. The robot must calculate human intentions and act quickly to achieve smooth and natural feeling collaboration. A current major limitation of collaborative robot usage is the concern for safety of the human [3]. The experiments reviewed in this paper are conducted with a relatively small arm robot to avoid hurting humans upon collisions. Furthermore, the collaborative robots reviewed in this paper only utilize only one or a few of the previously presented methods. Additionally, many of the experiments test the quality of the prediction and communication methods using only one type of work task and are designed for only one user. Further study is required to implement a full system which uses many human modeling and communication methods. Many of the modeling methods reviewed in chapter 2 showed promising accuracies in recognition. However, in a real-life scenario, accuracy might be worse due to the lack of model training material and change in scenery and personnel. Humans are hard to predict, and something might happen that the robot cannot predict, making it hard to implement a robust and safe collaborative robot system.

This bachelor's thesis presents a possible human-robot scenario utilizing previously examined collaboration-enabling human modeling and communication methods. A possible implementation for collaborative robots is in customized desktop computer assembly. The company assembles desktop computers where the parts are chosen by the customer. The robot in question is a mobile robot with a robot arm. The robot uses human motion prediction to predict human movement paths and limb movements to navigate safely in the workspace while avoiding collisions. The computers are customized so they are different every time, which makes the work require human adaptability. The robot uses context recognition to recognize components that need to be worked with. A worker can communicate with the robot using speech to order it to retrieve parts for the assembly or to give an object to the worker. The worker can point at objects to clarify which object needs to be given. The gestures are interpreted with gesture recognition methods. If the part is visible, the robot could use gaze recognition to predict which object is in question. The robot could have a small screen attached to it to show the current robot work state to clarify that orders have been received correctly. The confirmation can also be done with speech. The robot uses task planning to create task plans that emphasize human skills and safety. The human does the subtasks which the human is good at, and the robot does the repetitive tasks like retrieving parts from storage. While the robot is used to co-assemble computers, the robot inserts parts into the computer. The robot can additionally act as a mount, holding parts in place while the human screws them in. The robot can be ordered to give nearby parts to the worker or to retrieve items and components from the storage. The robot could be used by many workers to do small helpful tasks like retrieving items. The robot can speed up the assembly process by cutting component retrieval from the workers. The robot could also continue to do some tasks which don’t require a human while the human is taking a break. Overall, a collaborative robot can be implemented in many ways and for many purposes. This was an example of an imaginary scenario using a collaborative robot which utilizes methods examined in this thesis. Conducting real world experiments with different collaborative robot solutions will reveal overall benefits and feasibility of these systems.

5. Conclusions

This bachelor’s thesis examined human-robot collaboration in industrial applications. Chapter two focused on examining human modeling and intention prediction while chapter three reviewed the usage of communication and task planning. The goal was to achieve an understanding of how a collaborative robot can be effectively implemented and used. Furthermore, this paper aimed to evaluate the current use state of collaborative robots and problems that need to be addressed in further studies.

First, this thesis examined human modeling and human intention prediction methods. A collaborative robot can benefit from predicting human movement. The trajectory of a human walking is predicted to avoid collisions with humans with mobile robots, or human arm movement could be modeled and predicted to seamlessly give objects to human hand. Gesture recognition can be used to quickly communicate simple ideas to a robot. The user could effortlessly point at objects to communicate to the robot, what the robot should fetch or where to place objects. Touch gestures could be used to guide the robot or to communicate simple ideas. Furthermore, hand and touch gestures could be used to quickly stop the robot in case of faulty task execution. Human eyes reveal a lot of information on human intentions. Shifting gaze in the work area could indicate where the human is going to operate next. Context recognition is used to derive meaning of the environment to see how humans interact with it. Knowing what object the human interacts with can offer a major increase in intention prediction accuracy. For example, holding a hammer reveals that a human is likely going to do hammer something.

Secondly, this thesis reviewed ways of communication between a human and a robot. An efficient way of communication gives humans the ability to quickly give instructions and commands to a collaborative robot. Speech can be used as a means of communication. However, noisy sound input can lead to misinterpretation by the robot which can lead to critical errors. Speech could be used alongside many other modeling and communication methods. Visual communication such as augmented reality is a way to show useful information about the current task. However, augmented reality requires the usage of a VR headset during work and the communication is mainly from the robot to the human. Multimodal communication uses different types of communication simultaneously to make collaboration easier. Humans could tell the robot to pick something up and then point to it to clarify which object is meant. Multimodal communication increases robustness of the collaboration scenario. The collaboration between a human and a robot is dynamic, and the work task steps must be communicated before and during communication. For this purpose, the robot uses task planning to divide the main tasks into subtasks and divide them between the human and the robot. The robot can give tasks for the human in which the human is good at. In turn, the robot can choose to do physically demanding tasks to lessen the fatigue of the human.

Using collaborative robots in industrial applications like manufacturing offers many benefits. Effective utilization of collaborative robots can increase the throughput of assembly by decreasing work task completion times and human fatigue. Collaborative robots also enable automation of work that requires human creativity and flexible problem solving. Effective use of predictive human modeling leads to workers viewing human-robot collaboration as natural and effective. However, human modeling and intention prediction methods require a lot of computing power, which can be a problem when considering safety. The robot needs to think and act quickly, and the computationally expensive calculations need to be done fast enough to avoid collisions and to achieve sufficient collaboration.

The study of human-robot collaboration is an emerging field starting to gain popularity. More research is required for robust and safe human-robot collaboration. Safety is the biggest concern in collaborative robots. Especially with bigger robot arms, collision avoidance and safe task planning need to be reliable. Furthermore, most of the studies reviewed in this paper only experiment with collaborative robot systems that are designed for one work task. More studies need to be conducted on collaborative robots that are designed to perform many work tasks. Additionally, the studies reviewed in this paper are experimented with one user only. The use of collaborative robots with more than one user present at a time must be studied. To achieve a deeper and more concrete understanding of the benefits of collaborative robots in industry, wider research must be conducted with real work tasks and larger test groups. However, there is research evidence indicating that collaborative robots offer benefits in industrial applications. With the current evidence and continuous study in this field, the potential of collaborative robots is promising.

References

[1] N. C. Krämer, A. Von Der Pütten, and S. Eimler, "Human-agent and human-robot interaction theory: Similarities to and differences from human-human interaction," *Studies in Computational Intelligence,* Article vol. 396, pp. 215-240, 2012, doi: 10.1007/978-3-642-25691-2_9.

[2] N. Jarrassé, V. Sanguineti, and E. Burdet, "Slaves no longer: Review on role assignment for human-robot joint motor action," *Adaptive Behavior,* Article vol. 22, no. 1, pp. 70-82, 2014, doi: 10.1177/1059712313481044.

[3] A. K. Inkulu, M. V. A. R. Bahubalendruni, A. Dara, and K. SankaranarayanaSamy, "Challenges and opportunities in human robot collaboration context of Industry 4.0 - a state of the art review," *Industrial Robot,* Review vol. 49, no. 2, pp. 226-239, 2022, doi: 10.1108/IR-04-2021-0077.

[4] Y. Tao *et al.*, "A safety posture field framework for mobile manipulators based on human–robot interaction trend and platform-arm coupling motion," *Robotics and Computer-Integrated Manufacturing,* Article vol. 93, 2025, Art no. 102903, doi: 10.1016/j.rcim.2024.102903.

[5] S. Devin and R. Alami, "An implemented theory of mind to improve human-robot shared plans execution," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2016, vol. 2016-April, pp. 319-326, doi: 10.1109/HRI.2016.7451768.

[6] N. Muscholl, A. Poibrenski, M. Klusch, and P. Gebhard, "SIMP3: Social Interaction-Based Multi-Pedestrian Path Prediction by Self-Driving Cars," in *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020*, 2020, pp. 2731-2738, doi: 10.1109/SSCI47803.2020.9308130.

[7] N. Zapata, G. Pérez, L. Bonilla, P. Núñez, P. Bachiller, and P. Bustos, "Guessing Human Intentions to Avoid Dangerous Situations in Caregiving Robots," *Applied Sciences (Switzerland),* Article vol. 14, no. 17, 2024, Art no. 8057, doi: 10.3390/app14178057.

[8] M. Y. Halim, M. I. Awad, and S. A. Maged, "Hybrid Physics-Infused Deep Learning for Enhanced Real-Time Prediction of Human Upper Limb Movements in Collaborative Robotics," *Journal of Intelligent and Robotic Systems: Theory and Applications,* Article vol. 111, no. 1, 2025, Art no. 38, doi: 10.1007/s10846-025-02237-0.

[9] H. Liu and L. Wang, "Gesture recognition for human-robot collaboration: A review," *International Journal of Industrial Ergonomics,* Article vol. 68, pp. 355-367, 2018, doi: 10.1016/j.ergon.2017.02.004.

[10] Q. Gao, J. Liu, Z. Ju, Y. Li, T. Zhang, and L. Zhang, "Static hand gesture recognition with parallel CNNs for space human-robot interaction," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10462 LNAI, pp. 462-473, doi: 10.1007/978-3-319-65289-4_44.

[11] D. Jung, C. Gu, J. Park, and J. Cheong, "Touch Gesture Recognition-Based Physical Human-Robot Interaction for Collaborative Tasks," *IEEE Transactions on Cognitive and Developmental Systems,* Article 2024, doi: 10.1109/TCDS.2024.3466553.

[12] T. Tong, R. Setchi, and Y. Hicks, "Context change and triggers for human intention recognition," in *Procedia Computer Science*, 2022, vol. 207, pp. 3820-3829, doi: 10.1016/j.procs.2022.09.444.
[13] T. Tong, R. Setchi, and Y. Hicks, "Human intention recognition using context relationships in complex scenes," *Expert Systems with Applications,* Article vol. 266, 2025, Art no. 126147, doi: 10.1016/j.eswa.2024.126147.
[14] S. Li, P. Zheng, Z. Wang, J. Fan, and L. Wang, "Dynamic Scene Graph for Mutual-Cognition Generation in Proactive Human-Robot Collaboration," in *Procedia CIRP*, 2022, vol. 107, pp. 943-948, doi: 10.1016/j.procir.2022.05.089.
[15] V. Belcamino *et al.*, "Gaze-Based Intention Recognition for Human-Robot Collaboration," in *ACM International Conference Proceeding Series*, 2024, doi: 10.1145/3656650.3656675.
[16] M. Rakovic, N. Ferreira Duarte, J. Marques, A. Billard, and J. Santos-Victor, "The Gaze Dialogue Model: Nonverbal Communication in HHI and HRI," *IEEE Transactions on Cybernetics,* Article vol. 54, no. 4, pp. 2026-2039, 2024, doi: 10.1109/TCYB.2022.3222077.
[17] Y. Cohen, M. Faccio, and S. Rozenes, "Vocal Communication Between Cobots and Humans to Enhance Productivity and Safety: Review and Discussion," *Applied Sciences (Switzerland),* Review vol. 15, no. 2, 2025, Art no. 726, doi: 10.3390/app15020726.
[18] S. Park, C. C. Menassa, and V. R. Kamat, "Integrating Large Language Models with Multimodal Virtual Reality Interfaces to Support Collaborative Human-Robot Construction Work," *Journal of Computing in Civil Engineering,* Article vol. 39, no. 1, 2025, Art no. 04024053, doi: 10.1061/JCCEE5.CPENG-6106.
[19] E. Martin *et al.*, "Integrating Multimodal Communication and Comprehension Evaluation during Human-Robot Collaboration for Increased Reliability of Foundation Model-based Task Planning Systems," in *2025 IEEE/SICE International Symposium on System Integration, SII 2025*, 2025, pp. 1053-1059, doi: 10.1109/SII59315.2025.10871045.
[20] C. Zhang, J. Chen, J. Li, Y. Peng, and Z. Mao, "Large language models for human–robot interaction: A review," *Biomimetic Intelligence and Robotics,* Review vol. 3, no. 4, 2023, Art no. 100131, doi: 10.1016/j.birob.2023.100131.
[21] K. T. Pineda, E. Brown, and C.-M. Huang, ""See You Later, Alligator": Impacts of Robot Small Talk on Task, Rapport, and Interaction Dynamics in Human-Robot Collaboration," presented at the Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction, Melbourne, Australia, 2025.
[22] R. S. Lunding, T. Feuchtner, and K. Grønbæk, "Investigating AR Assistance for Human-Robot Collaboration in Mould Assembly "in the Wild"," presented at the Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction, Melbourne, Australia, 2025.
[23] A. Gottardi *et al.*, "Dynamic Human-Aware Task Planner for Human-Robot Collaboration in Industrial Scenario," in *Proceedings of the 11th European Conference on Mobile Robots, ECMR 2023*, 2023, doi: 10.1109/ECMR59166.2023.10256268.
[24] R. Zhao, S. Tao, and P. Li, "Safety-efficiency integrated assembly: The next-stage adaptive task allocation and planning framework for human–robot collaboration,"

Robotics and Computer-Integrated Manufacturing, Article vol. 94, 2025, Art no. 102942, doi: 10.1016/j.rcim.2024.102942.

[25] S. Sandrini, M. Faroni, and N. Pedrocchi, "Learning and planning for optimal synergistic human–robot coordination in manufacturing contexts," *Robotics and Computer-Integrated Manufacturing,* Article vol. 95, 2025, Art no. 103006, doi: 10.1016/j.rcim.2025.103006.

[26] C. J. Conti, A. S. Varde, and W. Wang, "Human-Robot Collaboration With Commonsense Reasoning in Smart Manufacturing Contexts," *IEEE Transactions on Automation Science and Engineering,* Article vol. 19, no. 3, pp. 1784-1797, 2022, doi: 10.1109/TASE.2022.3159595.

[27] M. Dhanda, B. A. Rogers, S. Hall, E. Dekoninck, and V. Dhokia, "Reviewing human-robot collaboration in manufacturing: Opportunities and challenges in the context of industry 5.0," *Robotics and Computer-Integrated Manufacturing,* Review vol. 93, 2025, Art no. 102937, doi: 10.1016/j.rcim.2024.102937.

[28] S. Patil, V. Vasu, and K. V. S. Srinadh, "Advances and perspectives in collaborative robotics: a review of key technologies and emerging trends," *Discover Mechanical Engineering,* Review vol. 2, no. 1, 2023, Art no. 13, doi: 10.1007/s44245-023-00021-8.

[29] C. Urrea and J. Kern, "Recent Advances and Challenges in Industrial Robotics: A Systematic Review of Technological Trends and Emerging Applications," *Processes,* Review vol. 13, no. 3, 2025, Art no. 832, doi: 10.3390/pr13030832.