big data and
cognitive computing

MDPI

Article

Engineering Human–Machine Teams for Trusted Collaboration

Basel Alhaji [1], Janine Beecken [1], Rüdiger Ehlers [2], Jan Gertheiss [3], Felix Merz [1], Jörg P. Müller [4,*], Michael Prilla [4], Andreas Rausch [2], Andreas Reinhardt [4], Delphine Reinhardt [5], Christian Rembe [6], Niels-Ole Rohweder [1], Christoph Schwindt [7], Stephan Westphal [8] and Jürgen Zimmermann [7]

[1] Simulation Science Center Clausthal-Göttingen, Technische Universität Clausthal, 38678 Clausthal-Zellerfeld, Germany; basel.alhaji@tu-clausthal.de (B.A.); janine.beecken@tu-clausthal.de (J.B.); felix.merz@tu-clausthal.de (F.M.); rohweder@iei.tu-clausthal.de (N.-O.R.)

[2] Institute for Software and Systems Engineering, Technische Universität Clausthal, 38678 Clausthal-Zellerfeld, Germany; ruediger.ehlers@tu-clausthal.de (R.E.); andreas.rausch@tu-clausthal.de (A.R.)

[3] School of Economics and Social Sciences, Helmut-Schmidt-Universität Hamburg, 22043 Hamburg, Germany; jan.gertheiss@hsu-hh.de

[4] Department of Informatics, Technische Universität Clausthal, 38678 Clausthal-Zellerfeld, Germany; michael.prilla@tu-clausthal.de (M.P.); andreas.reinhardt@tu-clausthal.de (A.R.)

[5] Institute of Computer Science and Campus Institute Data Science, Georg-August-Universität Göttingen, 37077 Göttingen, Germany; reinhardt@cs.uni-goettingen.de

[6] Institute for Electrical Information Technology, Technische Universität Clausthal, 38678 Clausthal-Zellerfeld, Germany; christian.rembe@tu-clausthal.de

[7] Institute of Management and Economics, Technische Universität Clausthal, 38678 Clausthal-Zellerfeld, Germany; christoph.schwindt@tu-clausthal.de (C.S.); juergen.zimmermann@tu-clausthal.de (J.Z.)

[8] Institute of Mathematics, Technische Universität Clausthal, 38678 Clausthal-Zellerfeld, Germany; stephan.westphal@tu-clausthal.de

* Correspondence: joerg.mueller@tu-clausthal.de

Received: 9 October 2020; Accepted: 18 November 2020; Published: 23 November 2020

check for updates

Abstract: The way humans and artificially intelligent machines interact is undergoing a dramatic change. This change becomes particularly apparent in domains where humans and machines collaboratively work on joint tasks or objects in teams, such as in industrial assembly or disassembly processes. While there is intensive research work on human–machine collaboration in different research disciplines, systematic and interdisciplinary approaches towards engineering systems that consist of or comprise human–machine teams are still rare. In this paper, we review and analyze the state of the art, and derive and discuss core requirements and concepts by means of an illustrating scenario. In terms of methods, we focus on how reciprocal trust between humans and intelligent machines is defined, built, measured, and maintained from a systems engineering and planning perspective in literature. Based on our analysis, we propose and outline three important areas of future research on engineering and operating human–machine teams for trusted collaboration. For each area, we describe exemplary research opportunities.

Keywords: human–machine collaboration; human–machine teams; human-in-the-loop; trust within teams; sensor and data analysis technologies

1. Introduction

With the rapid growth of autonomous systems, ubiquitous sensing and Artificial Intelligence technologies, over the years to come we shall be witnessing a dramatic change in how and to what ends humans and machines interact and work together. In many fields, such as industry, traffic and healthcare, where the environment is very complex and dynamic, these interactions are expected to yield huge benefits in improving productivity and reducing strain of humans. As per today, machines can perform repetitive, high-speed tasks with high accuracy. Humans, on the other hand, display higher flexibility and adaptability levels as well as high perception capabilities.

Combining the different but complementary abilities of humans and machines opens up the possibility of harnessing the strengths of both the human and the machine, in an increasing number of applications. The way humans and machines work together in industrial assembly and disassembly but also in office environments will change from a static hierarchical relationship to flexible collaboration on shared objects in team structures.

Teamwork requires trust between team members. Trust in human teams is a very well researched topic [1,2]. In addition, over the past few years, intensive work has been carried out on how humans can trust "Artificially Intelligent Agents" [3,4]. However, in the case of human–machine teams, there are new facets of trust to be considered, owing to the peer-to-peer relationships in teams. First, human workers must trust the machines. Think about yourself sitting in an autonomously driving car! Delegating choice of speed as well as longitudinal and latitudinal direction requires you to trust the vehicle. Second, and reciprocally, the machines must trust the human team members. For instance, for your autonomous car to hand over control over the vehicle back to you, you need to be sure that you are attentive, not distracted, and cognitively and emotionally capable to take over.

Hence, teams that consist of humans and artificially intelligent machines require mechanisms to bring about and maintain trust relationship from the human members to the machine members, and vice versa. Note that, by "trusted collaboration", we include both "trustfulness" as an internal quality denoting mutual trust between human and machine team members, as well as "trustworthiness" denoting a quality of the resulting human–machine system as perceived by an external entity.

While there has been intensive research on the human-to-machine direction of trusted collaboration [3], there is only little work on the machine-to-human direction.

Trust is an intrinsically social concept. However, the notion of trusted team collaboration in technical systems touches upon a wide range of heterogeneous and interdisciplinary aspects. It includes vulnerability and safety, reliability and dependability (which are well-researched concepts in Software Engineering), as well as the role of formal verification and validation. In addition, relevant topics are trust in sensors and actuators, transparency, interpretability, and explainability of AI systems and algorithms, but also notions of value alignment [3] between humans and AI systems, touching upon issues of preference elicitation, and on optimality, utility, and fairness of decisions made by algorithms [5].

While there is a plethora of discipline-specific models and methods that might be applicable [3], there is a blatant lack of methods for the systematic engineering of human–machine teams that bring about trusted collaboration. The first goal of this paper is to review and define fundamental concepts and approaches related to trusted collaborative human–machine teamwork. The second goal is—based on a review of the state of the art—to identify core research challenges to be tackled as we shall move towards overarching methodological concepts, theories, and tools for engineering future collaborative human–machine teams. Note that, within this paper, the term *machine* refers to any type of computerized agents equipped with some autonomy or intelligence, including robots, artificially intelligent software agents and smart software services.

This paper is organized as follows: In Section 2, we present and discuss a simple human–machine scenario, which will be deployed throughout the paper to illustrate aspects of human–machine cooperation. From this scenario, we then derive some basic engineering-related requirements, including the concept of mutual trust. The state of the art in relevant research areas is introduced and analyzed in Section 3, motivating the opportunities for future research, which we identify and discuss in Section 4. The paper ends with a short conclusion and outlook in Section 5.

2. Example Scenario: Disassembling a (Simple) Complex Product

In this section, we introduce a simple initial scenario in order to illustrate some general engineering challenges that arise in designing and operating human–machine teams. As shown in Figure 1, a human–machine team, composed of *Alice*, a human, and *Bob*, a robot, have the task to jointly disassemble a device into its components. In our example, the device is a box with a lid; the lid is fixed using different possible methods (e.g., screwed, glued, or pressed) and can be correspondingly opened with different tools, e.g., with a screwdriver, cutter, or crowbar. Note that, while this scenario is simple, the challenges we observe are very similar to those occurring, e.g., in complex (and often dangerous) processes for disassembling complex products such as electric vehicles including their batteries.

Once opened, the device can contain further sub-devices with lids, again fixed using different technologies. Connected sub-devices can be separated again using different methods and tools. The device structure can be recursive.

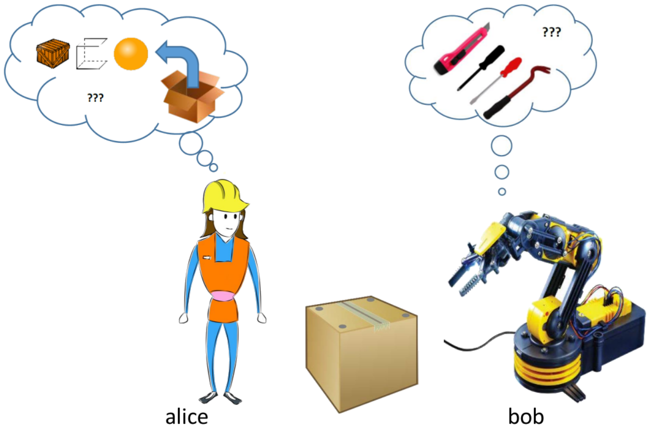

Figure 1. Introductory scenario: A human–machine team disassembling a product.

Even in this simple example, we note a number of interesting and highly non-trivial requirements regarding how the human–machine team can go about achieving the overall task.

First, the team needs to find an allocation of tasks and subtasks to team members. While this can be static, machines and humans may have different capabilities. For example, *Bob* may be very proficient in cutting, but performs less well with loosening or tightening screws. *Alice*, on the other hand, often has good ideas if an unexpected situation occurs. Often, *Bob* can assist *Alice* by handing her the tools she needs to open the box—or a cup of coffee when *Bob* realizes *Alice* becoming inattentive.

Second, note that tasks include domain-specific operational tasks (e.g., opening lids or removing sub-components) but also decision-making and planning tasks. This includes dynamically negotiating the allocation of the tasks to the team members, simple choices between alternatives in the process, but also more complex re-planning operations, e.g., if the device contains an unexpected component for which the selected opening method fails.

Third, since *Alice* and *Bob* work closely together and either of them may take the lead in some tasks in an alternating manner, the ways of safely transferring control among the two are required. e.g., in a somewhat delicate lid opening task, *Alice* may actually do the opening, using different tools in turn, while *Bob* will watch her and hand over the tool which she will need next. During this collaborative process, *Bob* should monitor *Alice*'s level of attentiveness (or distractions) to avoid harmful interactions. As another example, after *Alice* has forcefully removed the lid of the device using a crowbar, *Bob* should start inspecting the internals of the device together with *Alice*, getting close to her—while making sure she is aware of *Bob* approaching and will not make sudden arm movements that interfere with *Bob*'s operations and can be harmful to *Alice*.

Control transfer in both directions provides interesting challenges. In transferring control from the human to the machine, the machine must only accept control if it is capable of doing it. In traditional human–machine systems, it is assumed that the human will judge and decide on whether and when to transfer control to a machine. In a human–machine team, we shall expect the machine to take a more active part in this decision process. This requires the machine to be equipped with the capability to reflect and decide on its ability of taking over control in a certain situation. Two key issues here are (1) the ability of the machine to correctly recognize and categorize the situation, and (2) the ability of the machine to reflect upon its capability of taking control given a certain situation has been recognized.

In transferring control from the machine to the human, the machine needs to ensure that the human is capable of (and willing to) taking control over a certain task in a certain situation. Control transfer usually takes place in two phases: First, the process needs to be initiated, either by the human or by the machine. The latter case is more critical as the human may not be attentive, willing, or capable of taking over at a specific time; second, the transfer needs to be executed using some handshake process. From the perspective of the machine, two principal capabilities are required for the machine-initiated variant: (1) Decide whether (and when) to request a human to take over control for a certain task in a certain situation; (2) Decide under which conditions, when and how the machine should actually release control and hand it over to the human.

Fourth, the example reveals the need for the human–machine team to be capable of decision-making and planning under uncertainty, and, in particular, facing different levels of unexpected situations. This entails the capability of robustly recognizing and reacting to different types of situations that can occur in a non-deterministic process. This includes foreseeable failure situations (such as the device being fastened with the wrong type of screws), but also largely new situations, e.g., finding a hitherto unknown sub-component in a device, or *Bob* having broken the screwdriver. The latter type of situations is particularly difficult to deal with for a machine, as it requires reflection on its own capabilities, as well as the ability to conceptualize new situation types by analyzing deltas from known situations. In particular, based on this reflection, the machine will need to consult the human teammate for help, and possibly initiate a transfer of control as discussed above.

Fifth, human–machine teams are usually operating in and are influenced by the larger-scale context of a socio-technical system. Imagine that *Alice* and *Bob* are operating a single cell of a larger disassembly system together with many other human and machine agents, sharing some physical and digital infrastructure. In such a scenario, the compositions and the activities of the teams need be coordinated on a "macro" level to secure overall system effectiveness and efficiency. The coordination mechanism must allow for teams having private knowledge and pursuing individual objectives. In addition, establishing long-term trusted relationships among all agents of the system requires planning methods taking fairness aspects into account. For example, while Alice likes to alternate between disassembly and refurbishing jobs with Bob and Bert, respectively, her colleague Ann clearly prefers to steadily work with her favorite teammate Bert on refurbishing. An appropriate approach for team formation and job allocation balances the conflicting goals of Alice and Ann over time. In engineering human–machine teams, it is crucial to understand how planning and operational decision-making at the macro system level affects performance, stability, and user satisfaction of the "micro" level and vice versa. This will enable designers to define appropriate design objectives and utility functions, maximizing value alignment between human and machine actors, and thus contributing to sustainable overall process and system designs.

Sixth and last, we note that mutual (and reciprocal) trust between human and machine is a necessary basis for effective collaboration, including decision-making and planning, transfer of control, and increasing efficiency of and reducing uncertainty in interactions between human and machine. Issues such as over- and under-reliance on the human side, the inherent unpredictability of human actions, and verifying modes of safe behavior on the machine side can be addressed using the concept of trust (see Section 4.1).

In the following section, we provide an overview of basic concepts and notions and review the state of the art in the research areas, which are of highest importance for engineering human–machine teams for trusted collaboration.

3. Background and State of the Art

3.1. Collaborative Human Machine Teams

3.1.1. Team Collaboration

Groups and teams are the unit of analysis and support in collaborative work. The term "team" can be best defined by distinguishing it from the term "group". Both describe ensembles of two or more individuals working together in consent to achieve a certain goal [1]. While a group builds its structure and values in processes of grounding and group finding [2,6]), teams often contain "well-defined positions" ([7] p. 470) for individuals with often "highly-specialized functions or jobs" ([1] p. 34). While roles and structures in groups may develop along negotiations among members [6], in teams, usually "group structure, problem difficulty, leadership roles and similar variables cannot be varied beyond very narrow limits" ([1] p. 34).

Similar to understanding what a team is, the term "collaboration" can be best understood by distinguishing between the terms of cooperation and collaboration. Both describe work done together in groups or teams, which is directed towards a goal and coordinated among team or group members. Dillenbourg [8] has differentiated cooperation and collaboration by the criteria of symmetry, common vs. shared goals and labor division:

- Symmetry in collaboration is higher than in cooperation: In collaborative work, all actors are allowed to perform the same activities (symmetry of action), they possess (roughly) similar knowledge regarding the execution of the activity, and all individuals have a similar status regarding the collaborative process. On the contrary, in cooperation, individuals may perform different actions (e.g., actions that follow each other), they may have different knowledge (e.g., performing parts of their activity based on their knowledge), and there may be hierarchies for the overall process or parts of it. Regarding our example, the difference between a collaborative or cooperative disassembly process would be whether all participants are allowed to replace the lid, whether they possess enough knowledge to do all tasks and whether they decide collectively on sharing work or based on some hierarchy.
- In collaboration, it can be assumed that collaborators follow an overarching common goal in their collaboration, which goes beyond their (shared) individual goals. For cooperative processes, it would be enough if the individuals share similar goals (e.g., fulfilling their part of the processes with high quality).

- The division of labor is potentially the most obvious difference between cooperation and collaboration: "In cooperation, partners split the work, solve sub-tasks individually and then assemble the partial results into the final output. In collaboration, partners do the work together" [8] (p. 8). As described above, in our example, collaboration would describe a process in which the robot and human work conduct joint activities such as one teammate holding the object and the other dismantling it, or whether they perform them in sequential steps that happen after each other.

In addition, Dillenbourg [8] states that collaboration is more interactive and negotiable than cooperation. Therefore, collaboration implies synchronicity (working at the same time), while cooperation may also happen asynchronously.

3.1.2. Human–Machine Teams

One useful approach for structuring scenarios where humans are working together with an autonomous machine is the time-space classification. This classification from Computer-Supported Cooperative Work (CSCW) research can be adopted and applied to human–robot interaction [9]. This time-space classification known as Johansen matrix [10] classifies group work into four different categories: same time–same place, same time–different places, different times–same place, and different times–different places, see Figure 2.

The human in a human–machine team can have different roles depending on a pre-defined form of interaction (e.g., supervisor, operator, collaborator, cooperator).

In the example scenario introduced in Section 2, the focus is on designing human–machine teams with peer-like relationships between a team's partners. Therefore, it falls under the *same time, same place* cell of the Johansen matrix, synchronous and co-located cooperative or collaborative teamwork in which the human's role can be a collaborator and/or cooperator depending on the sub-task description.

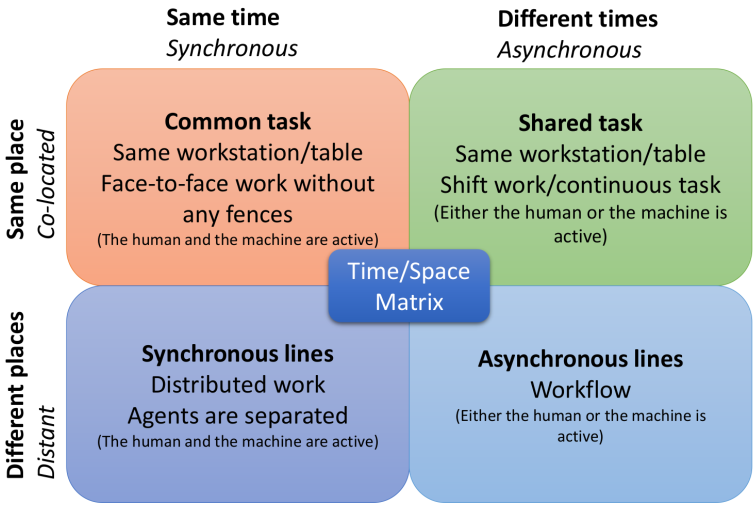

Figure 2. Johansen time-space matrix for robot-supported cooperative work.

There is ample research on teams consisting of automated robots, see [11,12] for two recent surveys. A classic example of robotic teams is the RoboCup initiative [13,14]. In addition, models for software agent teams and organizations have been investigated in the area of multi-agent systems, starting with early work on team-oriented programming [15,16], and in the Intelligent Virtual Agents community, see [17] for a survey. These approaches mostly focus on scenarios in which humans are sometimes simulated, sometimes users, but rarely part of the team.

Human–machine teams have been investigated in multi-agent systems, particularly aiming at planning and coordination tasks [18,19] and studying non-functional qualities such as explainability [20]. However, these approaches do hardly address the challenges of humans collaborating with cyber-physical systems, where physical interaction and safety concerns are prevalent. A promising conceptual approach highlighting human–robot collaboration (HRC) is Shared Autonomy, as proposed in [21,22], the authors propose a multidimensional view of human–machine collaboration, considering the freedom of intentions (normative), plans (strategic), and actions (operational), which require/bring about different types of interaction patterns between humans and machines. Trust is considered as an important component.

In the remainder of this section, we look into related work on this research area, elaborating on the crucial notion of trust Section 3.2 as well as on issues of engineering systems consisting of or containing human–machine teams Section 3.3.

3.2. Trust between Humans and Technology

3.2.1. What Is Trust?

There are many definitions of trust in the literature. Among these, trust in human relationships is understood as a "psychological state comprising the intention to accept vulnerabilities based upon positive expectations of the intentions or behavior of another" [23]. This does not only include trust towards individuals, but also the trust of teams in the team and its individual members [24].

Trust plays a particular role in human interaction and for human behavior, as it influences people's actions: "the extent to which a person is confident in, and willing to act on the basis of, the words, actions, and decisions of another" [25]. Therefore, trust among individuals or in the team has been described as a prerequisite for cooperation between actors in teamwork [26]. However, the necessity of trust in cooperation varies according to the risk that members of the cooperation process are taking [24]: If there is such a risk, (the quality of) cooperation depends on the trust members have in each other.

Trust is built over time and through interaction between those to trust each other [27,28]. The well-known phrase that "trust needs touch" [28] emphasizes that trust between humans is built through social interactions, which is also mirrored in group formation models [6].

A well-known understanding of trust is its differentiation into cognition-based trust and affect-based trust [29–31]. Corresponding trust models describe cognition-based trust as "Rational judgment of the partner‘s knowledge, competence and dependability" [29] and "A customer‘s confidence or willingness to rely on a service provider‘s competence and reliability" [30]. Affect-based trust is described as "Emotional bond between individuals or the confidence in the other that he or she is protective with respect to our interests and shows genuine care and concern for our welfare" [29] and "the confidence one places in a partner on the basis of feelings generated by the level of care and concern the partner demonstrates" [30]. This differentiation is important for the understanding of human trust, as it emphasizes that the rational (cognitive) aspect is only one aspect of human trust, and that it is accompanied by and emotional (affect) aspect that is much harder to formalize.

3.2.2. Human–Machine Trust Models

Building on the definition of trust by [23], we understand trust of humans towards autonomous machines as "the attitude that an agent [machine] will help [to] achieve an individual's goal in a situation characterized by uncertainty and vulnerability" [32] (p. 51). In a recent study [33] found that willingness, competence, benevolence, and reciprocity are the main (statistically significant) attributes that directly affect trust in machines.

Ref. [3] is a recent survey of human–machine trust, which focuses on algorithmic assurances as programmed components of machine operation, which are engineered explicitly to bring about user trust in machines. These assurances shall have an influence on user trust in interacting with the machine. The authors synthesize a trust model with four trust categories from the literature: dispositional, institutional, belief, and intention. They also identify different algorithmic assurances affecting human trust, and assess them according to how integral and essential they are to machine performance. In this assessment, the categories of value alignment between human and machine, interpretable models and processes (transparency, explainability), displaying human-like behavior, and putting the human in the loop are considered as the most essential assurance types. Note that [3] solely focus on human-to-machine trust and do not study reciprocity in terms of machine-to-human trust.

Another review of the state of art in research on trust in robots is presented in [34]. The authors address the dynamical fluctuation of trust and organize existing strategies in four groups [34]: (1) *Heuristics* in which trust calibration strategies follow the "rule-of-thumb" based on empirical evidence to handle over-trust and trust repair, (2) *Exploiting the process*, which concentrates more on the trustworthiness of the robot behavior, (3) *Computational trust models*, which require appropriate modelling of human trust dynamics, and (4) *Endowing robots with a theory of mind*, which enables robots to reason about their human users.

However, similar to [3], this work focuses only on unidirectional trust of humans toward robots.

A meta-analysis over the existing literature on human–robot trust has been conducted by Hancock et.al. [35]. They classified the identified factors that affect trust into three categories: human-related (ability-based), robot-related (performance-based and attribute-based), and environmental (team collaboration and tasking). The result of their analysis indicates that robot-related performance-based factors are strongly associated with trust [35]. These factors include dependability, reliability, predictability, and others.

Some researchers simplified the trust construct based on results of the meta-analysis provided by Hancock et al. and implemented a computational model of trust based on performance (originally identified by [36]) with different measures of performance depending on the task [37,38]. However, these computational models do not consider all trust dimensions relevant to their use case.

Ref. [39] surveyed the literature on human trust in robots. They also divided trust into two categories: performance-based trust and relational-based trust. By performance-based trust, they refer to the case where the robot does not interact with people but is spatially separated. Relational-based trust, however, is more about social activity. In the context of human–machine collaboration, these two notions of trust cannot be separated, since working with a robot as a team partner even in industrial settings embraces many social aspects.

Finally, trust in autonomous machines may also influence the relationships between human actors: As [40] shows, feelings of unfair service by a robot may also affect the relationships of those perceived to be preferred by the robot.

Compared to the abundant literature on trust in general and even trust in machines in particular, usage of the concept for machines in relation to humans is sparse. This direction of trust was considered by, e.g., [37,38,41], where the authors apply identical computational models for both partners (the human and the robot). Consequently, the differences between the human and the machine were not taken into account. An artificial cognitive architecture is proposed in [42], which can estimate the trustworthiness of the human partner. In this work, a probabilistic theory of mind along with an episodic memory system are used and implemented on a humanoid robot to decide whether to trust its source of information (the human) or not without any physical interactions. Ref. [43] considers a more narrow framework of machine trust in an operator's instructions, using it as a metric for the quality of a perturbation in an optimal control algorithm, in order to decide whether a given human instruction should be accepted.

In a physical collaboration setting, however, as depicted in our example with *Alice* and *Bob*, trust models should also include many additional (passive) inputs, e.g., human movements and actions. These are essential for the machine to form a justifiably correct (re)action.

3.2.3. Mutual Trust and Trust Cycles

Additionally, a collaborative setting leads to the necessity of maintaining a joint model of human and machine, encompassing *both* human trust and machine trust. In their review [44], Basu and Singhal note both the need for and the complete lack of such models describing mutual trust: when human and machine are equals and can intervene without prompting in the ongoing task, mutual trust is required to minimize interventions. The authors therefore propose the use of a framework providing for bi-directional trust dynamics, leading to a personalized state of trust between the machine and the human, as both get accustomed to one another. Similarly, Alhaji et al. [45] argue for the need of a joint model of mutual trust, in particular taking into account the respective differences of human and machine. Only such models describing the human–machine team as a unit (i.e., the human-in-the-loop) allow for the recursive questions required in collaborative settings: Do I, *Bob*, trust *Alice* to trust me, and how will my action affect her action, and as such impact the way I can execute my task?

Figure 3 shows this mutuality by extending the model proposed in [3] (Figure 1). The original one-way cycle between human and intelligent agents (solid blue) shows how the machine provides assurances to the human and tries to detect trust-related behaviour of the human. The new modification by [45] (dashed green) adds the reverse direction, taking human reactions to trust-related behaviour of the machine into account, and allowing for improved predictions of the state of the team as a whole.

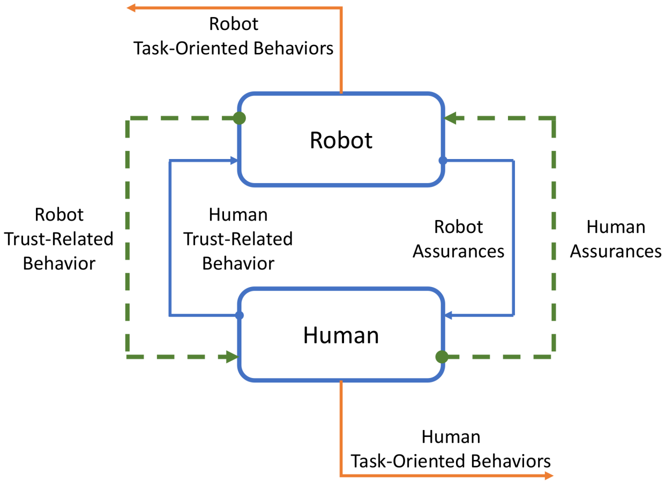

Figure 3. Mutual trust cycles between a human and a machine in a collaborative setting [45].

3.3. *Engineering AI-Based Cyber-Physical Systems*

Many state-of-the art approaches for interaction between technical systems and their environment use artifacts derived using approaches from the field of artificial intelligence (AI). For instance, for optical object recognition in robotics and autonomous driving, AI-based approaches have prevailed over classical approaches [46–49]. Integrating AI-enabled components into a complex system is a challenge in classical engineering approaches when the system to be constructed is to be trusted later. Trust requires not only the certification that a system should be safe, but also that the system exhibits demonstrably safe behavior.

Safety proofs for system architectures of complex cyber-physical systems rely on rigorous testing and verification of the components. While AI-based components are state of the art in terms of performance in practice, they frequently score poorly on traditional quality metrics. For instance, learned artificial neural networks for object recognition are prone to being fooled by so-called adversarial examples, i.e., input especially crafted to let the AI component misclassify or make a wrong decision.

However, even when brushing this problem aside, such networks are particularly difficult to test and verify, as their internal structure is not optimized towards testability and verifiability. Consequently, safety proofs for AI-based systems are hard to obtain, and, given the research on adversarial examples, the question of whether an AI-based system can ever be safe is not far-fetched.

Formally demonstrating trustable behavior of a technical system or its components starts from a specification of the behavior that should be satisfied. However, for important AI applications in autonomous systems, such as signal processing of raw sensor data (e.g., such as images and radar information) and, for subsequent processing steps, a complete specification of the functionality does not exist. Instead, extensive AI training data are required. For example, extensive image data with pedestrians are needed to train an AI-based pedestrian detection component. However, common standards and approaches to validation, such as ISO 26262 [50] and ISO 21448 [51], are based on the existence of corresponding requirement specifications. Instead of a complete requirement specification and associated standard development process, a new argument is needed that the AI training data are sufficiently relevant—in the sense of representative for the later series production. However, a method or indicator for the proof of relevance of the AI training data does not yet exist.

While the field of artificial intelligence has taken up this challenge, the problem with trusting a technical system is that it not only needs to be safe, but also that it should be obvious from its behavior that it is safe. Hence, for autonomous systems that interact with humans, research challenges that are already difficult on their own, namely explainability and safe learning, need to be solved in combination. From an engineering perspective, splitting off the trust considerations from the core AI techniques would be very appealing. Consequently, the first steps into this direction have been taken. For instance, it is shown in [52] how to augment a reinforcement learning with a safety monitor that enforces safe behavior, and Ref. [53] shows how to automatically compute a so-called shield, serving a similar role.

The authors of [54] introduce a holistic software systems engineering approach based on hierarchical Dependability Cages. It aims at autonomous systems, which include functionality partially realized via AI-based functions. The approach combines development time methods, in which monitors derived from system requirements specification are trained at development time via testing and simulation, and run-time techniques. At run time, qualitative monitors run in parallel to the system during operation time, checking the correctness of its behavior with respect to the requirements specification. Quantitative monitors gather data from the operational environment and check whether the current environmental situation has already been tested or it is a completely new situation. In this way, it supports that a system detects situations in which it has not been tested, which is particularly interesting in contexts in which this information can be communicated to humans to ask for a human hand-over of a task. However, a comprehensive engineering process for building autonomous systems that are not only trustworthy, but also demonstrate this to human collaborators appears to be missing.

4. Research Opportunities

In this section, we shall address three important research areas related to engineering human–machine teams for trustworthy collaboration. The first research area deals with methods and models for engineering and validation of human–machine teams, including formal methods, software engineering issues such as safety, reliability and dependability, and issues related to human factors and human–machine cooperation. The second area addresses technologies and methods for reliable sensing in Internet of Things (IoT) environments, and, for acquisition, modeling and prediction of human state, intentions, and behavior.

The third research area takes a macro-perspective on trusted collaboration in human–machine teams, scaling up and addressing problems related to preference elicitation, value alignment and optimization in heterogeneous collaborative human–machine teams. For each of these research areas, we summarize research gaps, state and resulting research opportunities.

4.1. Engineering and Validation of Collaborative Human–Machine Teams

The members of human–machine teams are tightly coupled during their collaboration. This needs mutual understanding of the respective collaboration partner's actions and trust in the behavior of the partner. More than that, the trust of humans in machines and trust of machines in humans reciprocally influence the behavior of both partners. Using the example of Section 2, if *Alice* (the human) does not trust *Bob* (the robot) to hand her the right tools, she will double-check or not rely on the robot at all, which will in turn slow down the collaboration process (this is also known as "under-reliance"). On the other hand, if *Alice*'s trust in *Bob* is higher than it should be, *Alice* might become less attentive to *Bob*'s actions and behave in a less trustworthy way (over-reliance). In this case, *Bob* must detect *Alice*'s untrustworthy behavior, and it should adapt its behavior accordingly (e.g., reducing its speed or maintaining a safe distance), which will negatively affect the overall task performance. This tight coupling of human and machine actions and trust leads to a plethora of engineering challenges that need to be solved.

First, models of trust and reliance, which are highly correlated, as well as models of capabilities of the human and the machine are needed. Using these models, trust levels can be estimated and validated continuously by all actors. These models must be implementable on the machine [55] for machine controllers to base on them, which requires that the machine has the capabilities needed to perceive the information related to the models. The machine should also provide feedback "assurances" [3] to the human to maintain an adequate level of trust. Validation and verification activities, which are usually referred to as providing "hard assurances" [3], focus on proving design requirements to be satisfied, such as safety, which is a prerequisite for trust. Using the example from Section 2, if we want to prevent *Bob*'s robot arm from colliding with the *Alice* under all circumstances, arm movement needs to be conducted in an overly conservative way, which prevents good performance. Rather, a model of trust towards human behavior should be employed by the robot to adapt its motion. Safety only needs to be verified against this model of trust, and the model needs to be validated to capture the important aspects of human–machine interaction.

Based on the previous discussion, the process of engineering controllers for human–machine collaboration deviates substantially from that of classical safety-critical systems. Architectures, development methodologies, and planning approaches that are common in the aviation or automotive domains are rarely applicable, as the arguments of component safety are more tightly coupled in human–machine interaction and reasoning about correct system behavior requires taking the trust between machine and human into consideration. The safety of robot behavior depends on whether it detects the human intention correctly. The correctness of the recognition of the human intention depends on what behaviors the robot can exhibit. The robot does not need to detect intentions that the human can only have in reaction to robot behavior that the robot cannot actually exhibit. For instance, highly complex evasion schemes by a human do not need to be detected correctly if the robot makes sure that the robot never forces the human to evade. A major challenge and research opportunity in engineering human–machine team collaboration based on trust is to interlink human and robot trust models in order to account for the reciprocity in the team, and to integrate them in an overall interaction design that allows all team members to form trust toward each other and act upon this trust.

The rest of this section lays out three research opportunities that arise from on the discussion above. The first one, Section 4.1.1, addresses the identification of the relevant dimensions of trust that are necessary and useful to capture trust dynamics in physical collaboration settings where partners' actions are interdependent. Section 4.1.2 concentrates on the interaction design within the context of such teamwork, including mutual adaptation and the associated uncertainty. Finally, issues that need to be tackled in deriving the system specifications, and the difficulty associated with the verification and validation processes of the system against these specifications, are discussed in Section 4.1.3.

4.1.1. Trust Dimensions for Human–Machine Hand-in-Hand Collaboration

Central to the Engineering process of human–machine interaction based on trust is the development of a model of trust that is useful in an engineering context. A collaborative robot in a team should have the capability of recognizing its human partner's trust in it and make sure that this human partner does not overly trust its skills and deploy it to tasks that it was not designed to perform. It should also dissuade the partner from placing itself or the overall task at risk [55]. This reflects the necessity of dynamic trust calibration during human collaboration with an autonomous machine (e.g., robot). Researchers who study trust in machines are usually concerned with detecting the current level of trust the human has in the machine and with keeping it within certain limits because inappropriate reliance may lead to serious undesired consequences. In order to develop a method that calibrates trust during the interaction, a model of human trust in autonomous machines is required. Moreover, since the robot in this setting is a teammate, its trust in its human partners should also be included in the trust model. A robot may fail in the collaboration, but failure of the human side is likely as well. This aspect of the model manifests a different motivation: Its goal is to ensure human safety rather than have a machine judge human behavior. The latter part of the model, addressing the machine's trust in the human partner, is further discussed in Section 4.2.

Physical human–machine collaboration in a synchronous co-located teamwork setting requires more investigation because of the revolution of intelligent system and autonomous machines that makes them part of our daily life. Especially in industrial settings, where robots should work as teammates alongside humans, knowing how and when these team partners trust each other is crucial for fruitful collaboration. To this end, research should be conducted in order to identify the relevant dimensions of trust in this specific collaboration scenario. For this model to capture trust properly, it should include all identified trust dimensions and be able to estimate trust in real-time.

4.1.2. Human–Machine Interaction Engineering

The design of Interaction schemes for human–machine collaboration requires taking into account the awareness of the human w.r.t. the machine state and vice versa. Coming back to our example: for *Bob* to hand over a tool to *Alice*, she needs to be aware of *bob's* intention. Vice versa, if *Bob* is busy with some other task, *Alice* may need to interrupt or replan her dissassembly process. At the same time, there is a high degree of uncertainty w.r.t. the state of both robot and human at runtime. Neither *Alice* nor *Bob* can directly observe each other's level attentiveness, but need to infer it from other observations. This makes the engineering of interaction schemes difficult, especially when they have to be certified later to be safe for deployment.

Tackling this problem asks for an interdisciplinary approach: not only do we need to integrate trust and intention models, which are inherently human–machine interaction topics, but we also need to design interaction in an adaptable way and with a high level of automation in the engineering process.

The former challenge is concerned with mutual adaptation and appropriation of humans and machines during the collaboration process. Just as humans adapt to each other over time when they cooperate, actors in human–machine teams need to develop ways to work together. As human behavior and trust (both of the human in the machine and of the machine in the human) will most likely differ between teams, there is a need to continuously build mutual trust from the beginning of the collaboration between humans and autonomous machines.

The latter of these challenges is tackled in machine learning (in particular reinforcement learning). Furthermore, the interaction between human and machine can be analyzed using game-theoretic models, in order to find suitable machine behavior that adapts to the behavior of the human. For instance, Ref. [56] uses game-based reactive synthesis to automatically construct a controller that meets all system specifications and productivity needs. However, to employ game solving, a specification of the interaction and trust is crucial, and current work can only be seen as a first step. The role that self-adaptation based on machine learning can play in this context without the resulting non-determinism in the machine behavior prevent trust in its behavior is also an important question, whose answer drives how engineering processes for human–machine interaction can look like in the future.

The research opportunity here lies in combining human machine interaction topics such as adoption and appropriation with algorithmic and machine learning topics, in particular in relation to computational models of human behavior. If both complement each other, this opens new perspectives for engineering trustful collaboration between humans and machines.

4.1.3. Validation and Verification for Self-Adapting Systems under Trust and Intention Models

Humans can only trust a machine if the machine is known to be well-tested, just like cooperation between humans is difficult when one of the participants has a bad reputation. In safety-critical contexts, this means that the verifiable safety of the system should be provably as good as possible. Showing this is a multi-step process: first, a specification is to be designed; then, the specification needs to be validated to capture the important safety aspects of the scenario, and finally the system needs to be verified to conform to the specification.

Recent works have started employing formal verification methods in systems that interact with humans. Askarpour et al. introduce a methodology called SAFER-HRC [57]. This method translates the informal and goal-oriented description of HRC application into a logic model. However, their model considers only the mechanical hazards (e.g., crushing, stabbing, etc.) [58]. We note that complex behavior schemes common to human–machine interaction are inherently difficult to verify, imposing a limit on such approaches, as does the fact that the collaborative processes under consideration are rarely well-defined routines, but rather occur in a dynamic and ad-hoc manner.

From an engineering perspective, full verification before system development has some further limits. First, it ignores that deriving a specification and validating it for a system with mutual trust between its actors is already a key difficulty. Then, verification of systems with AI components and with complex interaction schemes is computationally difficult. Finally, some properties lend itself to monitoring at runtime rather than up-front verification. This approach also tackles the problem that hardware failures at runtime also need to be accounted for as the real system behavior can be monitored.

For mutual trust between humans and machines, performing validation and verification for the state of the art is crucial. Hence, further improvements on the integration of up-front verification of machine-learned components with runtime monitoring is required along with new system architectures that simplify the engineering process of systems that promote trust, including the verification and validation aspects of this task.

4.2. Perceiving, Modeling, and Anticipating Human Condition and Behavior in Human–Machine Collaboration

Modeling of trust from the perspective of the machine is the measurable aggregate result of the evaluation of variables describing human state and behavior. Coming back to our example, typical variables include both physiological states (e.g., *Alice's* heart rate or her currently executed action) and mental states (such as her attentiveness). In this sense, it requires gauging levels of human ability to perform tasks, control processes or objects. In particular, the notion of a human being ready to assume control over a technical process from a machine (usually called "take-over-readiness" or TOR) is crucial in the considered scenario.

Deriving any such model of the human teammate has been identified by [59,60] as one of the three principal challenges in cyber-physical systems. A wide range of sensing modalities is required for the collection of contextual information to derive indicators about human state and behavior, from which indicators about trust can be derived. Generic models to describe the human have been variously considered in the literature. From simple control theoretical origins—modeling human behavior as an input/output relation, e.g., [61–63]—to behavior and cognitive models, e.g., [64] to a full Theory of Mind, e.g., [42,65,66], they draw from an increasingly diverse body of research, notably in sociology and psychology.

While human-in-the-loop models, where human–machine trust is considered as a reciprocal phenomenon, are well-suited for the collaborative situations as described in the scenario, application and concrete implementations of such models are still sparse, as noted before (Section 3.2.3). This is true not only for models specifically relating to trust [44], but also for the human-in-the-loop approach in general [67]. Notable examples of the latter include the opportunity-willingness-capability model by Eskins and Sanders [68], and second-order (recursive) considerations of knowledge by Jacq et al. [69].

Even in autonomous driving, where determining the human TOR is of high practical relevance, the question only recently attracted broader interest [70–72]. The focus so far has been on determining single, individual parameters such as tiredness, or on identifying specific disruptive tasks, such as talking on the phone [73–77]. A broader framework for acquisition, modeling, and predicting human state, intentions, and behavior is still missing, capturing physical features (such as hand/arm/feet/head/eye positions and movements), as well as mental features (including attention to/focus on the task at hand), and situation awareness.

In the remainder of this section, we sketch three promising research opportunities in this overall task. Section 4.2.1 deals with the identification of the basic features and parameters that influence and describe human actions (motions), plans, and intentions in human–machine teams. Methods for combining these parameters and features into robust models of human state and behavior based on ubiquitous sensing infrastructures and novel methods for sensor fusion are discussed in Section 4.2.2. The third research opportunity Section 4.2.3 explores models and methods suitable for simulating and predicting human (motion-related) behavior and intentions in collaborative human–machine teamwork scenarios.

4.2.1. Physiological Features to Infer the Level of Trust in Humans

Research results from pervasive computing [78] confirm that a plethora of physiological aspects can be captured through ambient and body-worn sensors [79]. Besides monitoring humans through sensors external to the body, the use of brain–computer interfaces [80] and swallowable computers [81] have been investigated to provide further extensive insights into user context, actions and intentions.

In particular, sensing physiological parameters have been proven to be well-suited to derive information about the human cognitive conditions, and thus proven highly relevant for attributing a level of trust. Capturing such phenomena is, however, not only limited by the acceptable degree of intrusiveness and potential interference with human actions, but also confined by the monetary cost of instrumenting workplaces with the sensing infrastructure to accurately gather the required data. Currently used devices to collect physiological parameters include wired body-worn sensors, occasionally complemented by remote sensors, e.g., cameras [82] or Laser-Doppler-Vibrometry, e.g., [83–85]. Mental parameters such as situation awareness can be requested by the machine from the user (e.g., SAGAT [86,87]). This, however, causes a pause in the ongoing process, with the corresponding loss of productivity [88]. The continuous collection of mental parameters is almost exclusively based on electroencephalogram (EEG) devices, see, e.g., [89–91]. While in our example, this impacts the workflow much less than *Bob* actively prompting *Alice* about her mental state, it nevertheless relies on fitting *Alice* with numerous complicated sensors, limiting the usefulness. Ways to measure EEG data remotely are only in experimental stages [92,93].

To the best of our knowledge, no universally applicable combination of unobtrusive stationary (i.e., workplace-based), ambient, and mobile (e.g., body-worn) sensing infrastructure to directly and comprehensively capture trust (or indicators to unambiguously derive the level of trust a machine can have in a human) has been presented in related work. This gives rise to the need for a systematic investigation into the relation of detectable time-dependent variations on the human body of any physical or medical property that can be used for predictions of human state, intentions, and behavior. On the one hand, this includes the identification of the parameters of greatest relevance, including physiological features, motion trajectories and unconsciously performed actions, and, in addition, finding the best sensing means to obtain such parameters. It remains an open research challenge to design the sensor devices and parameter configurations to accomplish this objective, and to maximize the resulting 'trust' or TOR levels while remaining economically viable at high levels of user acceptance. On the other hand, research is also needed to explore efficient and real-time capable means to network the sensor devices, cater to their seamless interoperability, and accomplish unified data representations to simplify processing and minimize the negative impact of sensor heterogeneity through using adequate middleware solutions for trust/TOR sensing.

A methodological investigation into the type of data required to determine *Alice*'s degree of TOR should be conducted by empirically verifying the efficacy of the most promising approaches proposed in literature. By way of example, *Alice* could be requested to wear a smartwatch, configured to determine her state of attention. In a subsequent experiment, her facial expressions could be captured by a different sensor, e.g., a camera system. Through an evaluation of the corresponding information content of all captured data streams, the optimum sensing strategy can be derived to enable *Bob* to determine an estimation of the trust into *Alice* and assess her TOR.

4.2.2. Novel Privacy-Preserving Data Fusion Methods to Extract Trust Levels

The individual consideration of sensor data from specific modalities is generally insufficient to capture a complete situational picture and infer trust levels with a high level of accuracy. Instead, the instrumentation of the environment with user- and environment-centric sensors and their networking (e.g., to form wireless sensor networks [94]) or body area networks [95] is required to create the technological foundation for capturing holistic user models that accurately put into place all available sensor data. The resulting combination of multiple sensor types and the fusion of the collected data using novel data processing mechanisms is required in order to cater to the semantically correct interpretation of available data. This makes it possible to detect user intentions [96], activities [97], and even further contextual features [98]. As a combination of these factors is unarguably indicative of the user's current state and behavior, their correct detection and availability is vital to enable the assessment of the level of trust autonomous machines can put in the human. To foster the acceptance of such solutions, however, data collected from involved users must be adequately protected. For example, these data should be protected against external threat in order to, e.g., avoid disclosure to unauthorized parties. In addition to secure the collected data, the employees' privacy must be respected to be compliant with the General Data Protection Regulation (GDPR) across the different steps including data collection, processing, and storage.

The required sensing infrastructure is heterogeneous by design. It is not only expected to contain devices simultaneously capturing multiple parameters of interest, but these often deliver data at different sampling rates, with different inherent degrees of inaccuracy [99], and application-specific semantics. The presence of these real-world aspects has a direct impact on the required post processing methods to extract relevant information pertaining to trust from raw data and necessitates novel data fusion methods to extract higher-level information from raw context and physiological data. In particular, solutions suited to measure the time-dependent variations on the human body of physical or medical properties must be identified in order to classify trust in a human partner. At the same time, the collection of data about *Alice* can allow the inference of sensitive information about herself. For example, changes in her health condition can be monitored based on the collection of her heart rate or her emotions can be inferred based on her facial expressions. As a result, a detailed analysis of threats related to *Alice*'s privacy should be conducted. Based on this analysis, the applicability of privacy-preserving solutions should be evaluated and appropriate protection methods applied. As a result, sensitive information about *Alice* will not be disclosed to *Bob* or any further parties, thus protecting *Alice*'s privacy. This may further contribute to increase her overall acceptance of such technical solution.

For modeling and predicting those individual parameters, literature typically has assumed the use of functional data, even though this terminology is not necessarily used (for an introduction to functional data analysis, see, e.g., [100]. In the applications considered, this means that data observed over a fixed time interval preceding the current point in time are used to draw conclusions about the current state. From those functional data, specific features are extracted using methods such as Wavelets, e.g., [89], Independent Component Analysis [90] or simple summary statistics [77]. Afterwards, those features are used as input for machine learning with techniques such as Support Vector Machines [77], Neural Networks [76,90], K-Nearest-Neighbors [89], Linear Discriminant Analysis [89] or Random Forests [77].

From the perspective of data analytics and machine learning, enhanced methods for judging physical and mental state of the human partner should be studied, e.g., embedded methods where feature extraction is part of the learning algorithm may provide more specific and targeted information. This may, for instance, be done by functional regression/classification, e.g., [101] or suitably designed neural networks, cf. [102,103]. In addition, learning algorithms for jointly handling multivariate and polytomous aspects of the human–machine relationship are needed. Multimodal sensor data must be combined and analyzed adequately to infer multi-dimensional parameters that all together determine whether or not the human is ready and capable to take over.

4.2.3. Real-Time Models and Anticipation of Human Behavior in Team Collaboration

Machine decisions in a human–machine collaboration, such as *Bob* handing over a tool to *Alice* in our illustrating scenario, require taking the current situation into account, but also considering longer-term aspects of *Alice's* performance on the specific task in the past, the history of the collaboration process, and *Bob's* state and capability. For instance, even in situations in which *Alice* is distracted or exhausted, *Bob* could make the decision to hand over in case *Bob* performed poorly on the specific task in the past, but *Alice* was typically able to do the job in acceptable time. It is also an important aspect of trust to make an informed decision to whom to pass control, in case multiple human users with different levels of trust are present.

Making such informed decisions typically requires the machine to not only model the current state of the human teammate, but also to be able to carry out reliable short-term predictions of the teammate's actions, plans, and even intentions [22]. In this context, the ability of modeling (and subsequently anticipating) human body movements and trajectories is crucial. Next, hand [104] and head [105] movements, arm trajectories and occupancy models [106–108] are of importance in the context of safe human–machine collaboration in assembly or disassembly processes as illustrated in the scenario, as are human walking trajectories. For the latter, different approaches exist, including outdoor applications such as traffic modeling [109], crowd modeling [110] and indoor applications, e.g., railway stations [111], industrial plant environments [112,113], and building evacuation [114]. Data-driven learning of these models has been discussed in Section 4.2.2.

These models form the basis for the recognition and, more importantly, the anticipation of human movement actions, plans and intentions. As Ref. [115] points out, making accurate predictions of human motion is difficult because human behavior is influenced by a large number of internal and external stimuli. Prediction approaches can be subdivided in physics-based (Newtonian) methods [116–118], pattern-based methods that learn motion patterns from trajectory data [119,120], and plan-based methods [109,115,121] that perform reasoning on motion intentions (mostly assuming rationality of the actors).

Ref. [122] investigates a variant of the social force model to study safe motion control of personal service robots. They propose a reciprocal human–robot mutual intention estimation model. However, they use this for simulating scenarios rather than for trying to predict human behavior online. Ref. [123] proposes an approach to design "natural" movement patterns of autonomous robots by executing trajectories a human would follow. Using such a model to predict human motion in a generic and scalable way remains a challenging task, however.

Three interconnected challenges that we see in this area are scalability, anticipation, and value assignment. Scalability deals with the question of how we can achieve real-time decision and reaction capability of the machine in large-scale systems comprising many humans and machines.

Anticipation addresses the challenge of deriving reliable short-term predictions of human condition, and actions, plans or intentions that may negatively influence the human's teamwork ability. Can we predict TOR for the future? What are future intentions (instead of simply assessing trust for the current point in time)?

Finally, value alignment addresses the longer-term challenge of constructing and maintaining effective and efficient human–machine teams: Can we extend predictive models to recognize and mend differences between robot expectations and human actions, plans and intentions [124]? In our scenario, how can we achieve that *Bob* has an accurate conception of *Alice's* goals and task-related preferences? In addition, more generally, can we thus improve the value alignment of human and machine utility or satisfaction functions [3] in a collaborative process?

4.3. Optimization and Simulation of Collaborative Human–Machine Teams

Real-time control of human–machine teams in manufacturing and services is mainly concerned with synchronizing the team members' actions during the collaborative handling and processing of individual operations to assure safe and effective working conditions. Human–machine teams combine the flexibility and polyvalence of human workforce with the higher precision, power, and endurance of machines. They are deployed in such diverse areas like manufacturing [125], maintenance and repair of complex machinery [126,127], disassembly of equipment and consumer goods [128], geriatric and health care [129], or search and rescue [130]. From the macro perspective of operations management, there is a need for coordinating the team formation, the assignment of tasks to the different teams, and the sequencing of the tasks allotted to the same team or to intersecting teams. This coordination should be organized in such a way that the potentials offered by the specific capabilities of machines and individual humans are fully exploited, operational goals like short order cycle times, adherence to due dates, and low operating costs are met, and the preferences of the human team members are adequately taken into account. Compared to traditional systems like production facilities with human operators or repairmen using toolkits, human–machine teams feature a heterarchic relation between humans and machines. This is why optimization approaches for the coordination of human–machine teams should consider the human and the machine agents jointly on a single, integrated planning level rather than performing machine and workforce scheduling separately.

A task results from the aggregation of all microscopic steps that are performed by a given team when processing the operations of some production or service process. In general, such a process consecutively runs through several teams, each of which provides the skills and facilities required for the respective operations. In the example scenario of Section 2, the recycling process may consist of the disassembly, the inspection, and the refurbishing steps, executed by the teams respectively formed by *Alice* and *Bob*, *Amy* and *Ben*, as well as *Ann* and *Bert*. The composition of a team is assumed to remain fixed during the entire execution of the task. In addition, it is supposed that, while the task is being performed, no team member can carry out further tasks in parallel even when he or she is not occupied in all individual steps. For example, *Alice* is expected to stay with *Bob* during the entire disassembly task and not to switch to the inspection team during intermediate idle times.

A common approach to cope with the short-term planning of production and service processes is resource-constrained scheduling [131] (Chap. 6). In a centralized, single-criteria, and purely deterministic problem setting under perfect information, the interdependent problems of team formation, task assignment, and task sequencing can be addressed within the framework of multi-modal scheduling problems [132] (pt. VII). The multi-mode dimension of the problem allows for including different types of trade-offs like complementary or substitutional relations between humans and machines (resource-resource trade-offs) or the acceleration of tasks by augmenting the team size (time-resource trade-offs). In such a scheduling problem, facilities of the same type and humans of comparable qualifications are aggregated to different renewable resources. Precedence constraints between tasks can be used to model the given sequence of process steps. Each task can be performed in alternative execution modes, differing in resource requirements and processing times. In this way, a team can be identified with the resource requirements of a specific execution mode, and mode selection is comprised of the team formation and task assignment problems. The scheduling part of the problem consists of allocating resource units to the tasks over time such that precedence constraints among the tasks are satisfied and some objective function in the mode assignment and task completion times is optimized. More specific variants of multi-modal scheduling problems are staffing and scheduling problems [132] (pt. VIII) focusing on the assignment and scheduling of heterogeneous and multi-skilled workforce to precedence-related tasks.

In real-world applications involving a high labor share of value added like project manufacturing, dismantling of facilities, health care and service operations, the centralized and deterministic approach may often prove to be considerably oversimplified. In contrast to machines, human behavior is characterized by individual and time-varying factors like specific preferences [133,134], fatigue [135], and personality traits [136] such as free will, risk aversion, moral values and opportunistic attitudes. These properties significantly add to the complexity of task scheduling given the less predictable human performance and comportment. In addition, individual preferences and opportunistic behavior can typically cause undesirable effects and issues that are addressed as moral hazard and adverse selection problems in microeconomic theory [137] (Chapters 13 and 14).

In the field of scheduling, there exist various paradigms addressing parts of the challenges caused by the human factor. Time-dependence as well as private knowledge and control of human performance can be captured by assuming the task processing times to be uncertain or vague. Depending on the type of uncertainty or vagueness and the risk attitude of the decision makers, different types of stochastic scheduling approaches [138,139], robust scheduling [140], or fuzzy scheduling [141] can be used. Alternatively, uncertainty can also be reduced by following decentralized approaches in which scheduling is managed by staff or team members disposing of exclusive domain knowledge.

Even more importantly, the decentralized perspective allows for accounting for human preferences and to cope with the resulting goal conflicts and self-interested behavior in a systematic way. Examples of respective scheduling approaches include multi-criteria scheduling [142], multi-agent scheduling [143], selfish scheduling [144] and decentralized project scheduling models [145]. Selfish and decentralized scheduling relies on game-theoretical mechanisms such as auctions [146] or negotiations among agents [147]. These mechanisms are well-suited to cover scenarios with conflicting objectives and asymmetric information distribution, which are particularly relevant to the short-term planning of production and service processes that are run by heterogeneous human–machine teams.

For a single-stage manufacturing setting, Ref. [148] proposes a descending multi-round auction with capacity constraints where the shop floor personnel compete for tasks while guaranteeing certain processing times. Each worker can only get assigned to a maximum number of tasks and is not allowed to increase his or her bid on any task compared to the preceding round. The actual minimum processing times are assumed to be private knowledge and the preferences of the workers are supposed proportional to the lag between the time announced and the minimum processing time. In every round, solving the auctioneer's winner determination problem provides an assignment of tasks to the workers that for the current bids achieves a minimum total per-unit processing time. Following a myopic best-response approach, the bidding strategies of the workers are based on the bids of their competitors in the preceding round. The auction terminates and the assignment chosen by the auctioneer becomes final when no worker changed his bid any more. Particular emphasis is placed on the efficient solution of the bidders' bilevel bid creation problems and on the way to deal with the multiplicity of optimal assignments, which typically comes along with the myopic best response scheme. To this end, the bid creation problem contains an optimism parameter, which can be negotiated between the personnel and the firm. Computational experience shows that, compared to the two other auction schemes, the mechanism is able to balance the interests of both the factory and the workers, while still ensuring a good coordination performance.

This myopic best response mechanism for task assignment is open to variations and expansions in multiple respects, which are briefly sketched as four research opportunities in the remainder of this section.

4.3.1. Integrated Team Formation and Task Sequencing/Planning

To fully exploit the power of decentralized coordination schemes, the team formation and task sequencing problems should be integrated into the mechanism. In principle, team formation could be approached from the perspective of cooperative game theory and be interpreted as the problem of forming stable coalitions. Following a hierarchical planning approach, the problem then would be decomposed into two consecutive decision levels, where the team formation precedes the task assignment and sequencing. For given teams, model [148] can readily be generalized to cover the assignment of tasks to work shifts, which provides the sequence in which the tasks assigned to the same team are executed. An alternative approach consists of incorporating the team formation into the task assignment auction. In this case, workers could bid on combinations of tasks and machines or on combinations of tasks and team configurations. The latter variant would lead to a combinatorial auction scheme, necessitating the prior construction of an appropriate set of alternative team configurations in a bundle-generation step.

As illustrated in the Alice-and-Bob scenario of Section 2, real-life processes are often subject to uncertainty with respect to processing times, resource availabilities, or even the set of tasks to be scheduled. Combining predictive and reactive planning can substantially curtail the impacts of encountering unexpected situations during process execution. Whereas predictive planning provides an initial robust baseline plan absorbing a large part of the data's variance, reactive planning dynamically re-plans the tasks each time the current schedule becomes infeasible. In the context of a decentralized approach, re-planning amounts to apply simple yet effective coordination mechanisms that can be carried out in real time during the implementation of the plan.

4.3.2. Heterogeneous Preferences, Fairness, and Trust-Building

The human team members may have preferences for certain tasks, shifts, team partners, or equipment. When they collaborate in a team, the question arises how individual preferences can be aggregated to joint team decisions. This problem is at the heart of computational social choice theory [149] and can be addressed using voting methods or concepts from game theory like auctions or negotiations. In the context of coordinating the activities of heterogeneous human–machine teams, the preference aggregation has to be embedded into the mechanisms serving to optimize the team formation, task assignment, and task sequencing. To ensure enduring acceptance of the aggregation and allocation mechanisms, fairness among the members of a team and among the different teams is a crucial point. Obviously, fairness is a precondition to establishing trusting relationships within and among the human–machine teams. Moreover, implementing game-theoretical coordination schemes in practice generally requires transferring large parts of the human team members' optimization calculus to software systems like bidding machines [150]. Since the computations are based on private data, providing truthful information will only be secured if there is confidence in the incentive-compatibility of the coordination mechanisms and in the privacy protection concept. Trust-related questions raised in this context can be treated within the framework of multi-agent systems [151]. The validity of findings from multi-agent simulations should then be tested empirically in laboratory experiments. In behavioral economics, such experiments are typically based on trust games, where trust is expressed by making some initial sacrifice [152].

4.3.3. Decentralized Scheduling Models with Complex Real-World Constraints

In the scheduling literature, different game-theoretic coordination mechanisms have been devised for machine and basic resource-constrained scheduling models. On the other hand, there exists a large body of literature dealing with more general scheduling models including various constraints that arise in practical production and service operations management [153]. Important features include material-availability and storage-capacity constraints [154], prescribed minimum and maximum time lags between the execution of individual tasks [155], and transfer times of facilities and personnel between consecutive tasks [156]. For example, material-availability constraints arise in divergent process structures that include operations feeding concurrent sub-processes. Minimum and maximum time lags can serve to model technical or organizational synchronization requirements, and transfer times arise as sequence-dependent setup times in production or transportation times in multi-site scheduling and routing applications. To promote the dissemination of decentralized scheduling approaches in practice, the coordination mechanisms must be combined with more complex scheduling models that cover the practical requirements often encountered in industrial production and service systems.

4.3.4. Methods for Reliable Performance Assessment

The effectiveness and efficiency of the different building blocks sketched in the previous paragraphs should be numerically compared against more traditional approaches from operations planning and control. A centralized planning method can provide benchmark results for single-criteria scenarios under perfect information, which allow for evaluating the price of anarchy incurred by adopting a decentralized multi-agent perspective. In practice, manufacturing control based on production authorization cards and simple decision rules are often a valuable alternative to the costly centralized short-term planning. Popular examples of such card-based production control are Kanban, ConWIP, or 19 POLCA systems. The underpinning principles can also be applied to service process control and be enhanced by simple rules governing local ad-hoc team formation and task assignment. To ensure practically meaningful and statistically significant insights into the pros and cons of the different paradigms, the experimental design of the performance analysis must be based on a systematic variation of problem parameters that are supposed to largely impact on the general difficulty of the problem instances and on the relative suitability of the different approaches. To generate a sufficiently large and representative set of problem instances, the test bed can be created, e.g., using generative methods from machine learning based on real-world samples.

A promising methodology for performance evaluation of optimization large-scale human–machine team scenarios with heterogeneous human satisfaction functions is multi-agent-based modeling and simulation [157]. This microscopic modeling approach has been successfully applied to a number of large scale sociotechnical systems, including traffic modeling and simulation [158,159], urban shared spaces exploration [109,160], platooning [161], or product lifecycle management [162]. An interesting problem in this context is making the recommendations or decisions found by optimization algorithms transparent to the humans affected. Kraus et al. [163] provide an overview of main problems and some promising approaches towards this problem.

5. Conclusions

In this paper, we carried out an analysis of the state of the art of an important class of interaction between humans and artificially intelligent machines. We considered situations of collaborative hand-in-hand human–machine teamwork on joint tasks or objects. Already today, this form of collaboration can be observed in industrial assembly processes; in the future, they are very likely to play important roles in other domains, including service robotics, office collaboration, or healthcare. For this class of systems, the focus of our analysis has been on how trust between humans and machines can be defined, built, measured, and maintained from a systems engineering perspective. Compared to a rich body of literature on trust in general and the trustworthiness of machines in particular, literature and concrete applications of models of trust about the human partner are sparse. A holistic approach of considering the human–machine team as a unit, stringently required for addressing the challenge of being equal partners in collaborative settings, is virtually absent.

An important outcome based on the analysis of the state of the art therefore has been that, for the class of collaborative applications under consideration, there is a need for research on systematic engineering methods for collaborative human–machine teams, fueled by a reciprocal understanding of trust relationships between humans and machines. We propose and describe three important areas of future research on engineering human–machine teams for trusted collaboration: methods and models for engineering and validation of human–machine teams, identifying sensing technologies and methods to measure and describe human conditions from the perspective of the machine, and coordination of teams from the macro perspective. For each area, we described exemplary research opportunities. The first includes developing holistic models of trust and reliance and using these to engineer safe human–machine interaction. The second is concerned with sensors and hardware to unobtrusively collect physiological features, data fusion methods to extract trust levels, and models to anticipate human behaviour. The final research field is devoted to the formation of teams and scheduling tasks, heterogeneous preferences, fairness and trust-building, decentralized scheduling models, and performance analysis. Figure 4 illustrates the resulting research agenda.

A final point the authors wish to make is that successfully grasping and tackling the research challenges discussed in this paper requires an interdisciplinary effort of various research fields. These include but are not restricted to human factors and human–machine interaction, software engineering, formal verification and validation, Industrial Internet of Things (IIoT) sensing technologies, computational modeling of sociotechnical systems and multi-agent systems, robotics and control, AI and data science, optimization, as well as expertise in the application areas under consideration.

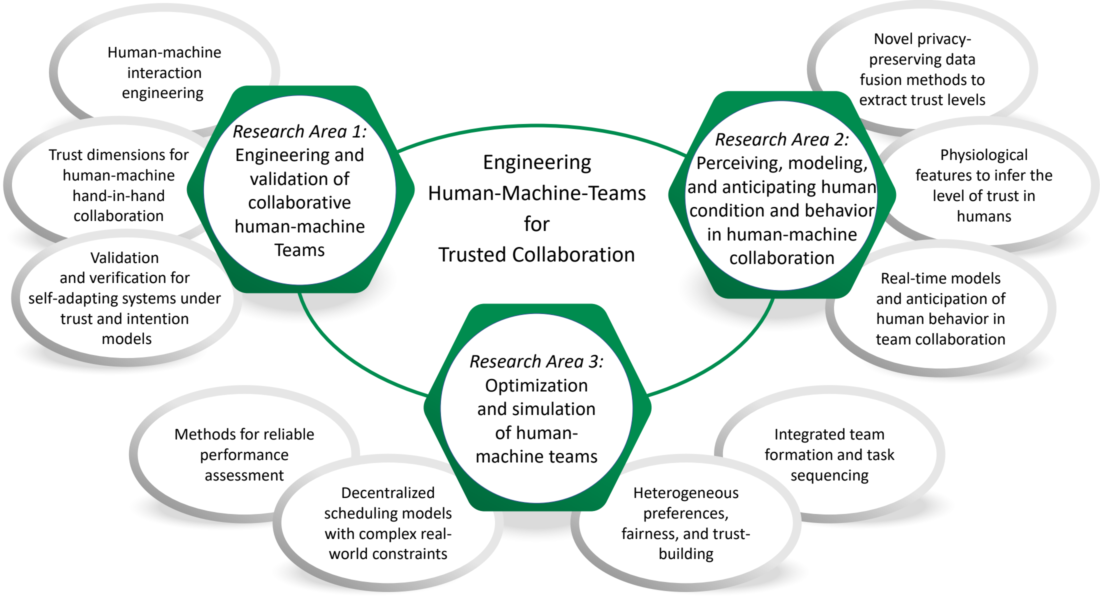
Human-machine interaction engineering
Trust dimensions for human-machine hand-in-hand collaboration
Validation and verification for self-adapting systems under trust and intention models
Research Area 1: Engineering and validation of collaborative human-machine Teams
Engineering Human-Machine-Teams for Trusted Collaboration
Research Area 2: Perceiving, modeling, and anticipating human condition and behavior in human-machine collaboration
Novel privacy-preserving data fusion methods to extract trust levels
Physiological features to infer the level of trust in humans
Real-time models and anticipation of human behavior in team collaboration
Research Area 3: Optimization and simulation of human-machine teams
Methods for reliable performance assessment
Decentralized scheduling models with complex real-world constraints
Heterogeneous preferences, fairness, and trust-building
Integrated team formation and task sequencing

Figure 4. Research areas and research opportunities.

Author Contributions: Conceptualization, R.E., J.P.M., M.P. and A.R. (Andreas Reinhardt); Funding acquisition, J.P.M. and A.R. (Andreas Rausch); Investigation, J.B., F.M. and N.-O.R.; Methodology, M.P., C.S. and J.Z.; Resources, A.R. (Andreas Rausch) and C.R.; Software, B.A., F.M. and N.-O.R.; Supervision, J.G., M.P., A.R. (Andreas Rausch), C.R. and S.W.; Validation, J.P.M., D.R. and S.W.; Visualization, J.B.; Writing—original draft, B.A., R.E., J.G., F.M., J.P.M., M.P., A.R. (Andreas Reinhardt), D.R., N.-O.R., C.S. and J.Z.; Writing—review & editing, J.B., J.G., A.R., C.S. and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This publication is based upon work supported and financed by the Clausthal University of Technology and by the Simulation Science Center Clausthal-Göttingen (SWZ), project HERMES: Heterogeneous Human–Machine Teams.

Acknowledgments: We gratefully acknowledge Alexander Herzog, Carsten Hesselmann and Marc Schlegel for providing administrative support of the HERMES project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Klaus, D.J.; Glaser, R. Reinforcement determinants of team proficiency. *Organ. Behav. Hum. Perform.* **1970**, *5*, 33–67.
2. Schmidt, K.; Bannon, L. Taking CSCW Seriously: Supporting Articulation Work. *Comput. Support. Coop. Work (CSCW) Int. J.* **1992**, *1*, 7–40.
3. Israelsen, B.W.; Ahmed, N.R. "Dave...I can assure you...that it's going to be all right..." A Definition, Case for, and Survey of Algorithmic Assurances in Human-Autonomy Trust Relationships. *ACM Comput. Surv.* **2019**, *51*, 1–37. [CrossRef]
4. Falcone, R.; Castelfranchi, C. Social trust: A cognitive approach. In *Trust and Deception in Virtual Societies*; Springer: Dordrecht, The Netherlands, 2001; pp. 55–90.
5. Schleibaum, S.; Greve, M.; Lembcke, T.B.; Azaria, A.; Fiosina, J.; Hazon, N.; Kolbe, L.; Kraus, S.; Müller, J.P.; Vollrath, M. How Did You Like This Ride? An Analysis of User Preferences in Ridesharing Assignments. In Proceedings of the 6th International Conference on Vehicle Technology and Intelligent Transport Systems, VEHITS 2020, Prague, Czech Repulic, 2–4 May 2020; pp. 157–168. [CrossRef]
6. McGrath, J. Groups and Human Behavior (Excerpt). In *Groups: Interaction and Performance*; Prentice Hall: Englewood Cliffs, NJ, USA, 1984; pp. 113–115.

7. Baker, D.P.; Salas, E. Principles for Measuring Teamwork Skills. *Hum. Factors* **1992**, *34*, 469–475. [CrossRef]
8. Dillenbourg, P. What do you mean by 'collaborative learning'? In *Collaborative Learning: Cognitive and Computational Approaches*; Elsevier: Oxford, UK, 1999; pp. 1–19.
9. Surdilovic, D.; Bastidas-Cruz, A.; Radojicic, J.; Heyne, P. Interaktionsfähige intrinsisch sichere Roboter für vielseitige Zusammenarbeit mit dem Menschen. In *baua: Fokus*; Bundesanstalt für Arbeitsschutz und Arbeitsmedizin: Dortmund, Germany, 2018.
10. Johansen, R. *Groupware: Computer Support for Business Teams*; The Free Press: New York, NY, USA, 1988.
11. Cortés, J.; Egerstedt, M. Coordinated control of multi-robot systems: A survey. *SICE J. Control Meas. Syst. Integr.* **2017**, *10*, 495–503.
12. Rizk, Y.; Awad, M.; Tunstel, E.W. Cooperative heterogeneous multi-robot systems: A survey. *ACM Comput. Surv. (CSUR)* **2019**, *52*, 1–31.
13. Candea, C.; Hu, H.; Iocchi, L.; Nardi, D.; Piaggio, M. Coordination in multi-agent RoboCup teams. *Robot. Auton. Syst.* **2001**, *36*, 67–86.
14. Marsella, S.; Adibi, J.; Al-Onaizan, Y.; Kaminka, G.A.; Muslea, I.; Tambe, M. On being a teammate: Experiences acquired in the design of RoboCup teams. In Proceedings of the Third Annual Conference on Autonomous Agents, Seattle, WA, USA, 19–21 April 1999; pp. 221–227.
15. Kinny, D.; Sonenberg, E.; Ljungberg, M.; Tidhar, G.; Rao, A.; Werner, E. Planned team activity. In *European Workshop on Modelling Autonomous Agents in A Multi-Agent World*; Springer: Heidelberg, Germany, 1992; pp. 227–256.
16. Pynadath, D.V.; Tambe, M.; Chauvat, N.; Cavedon, L. Toward team-oriented programming. In *International Workshop on Agent Theories, Architectures, and Languages*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 233–247.
17. Norouzi, N.; Kim, K.; Hochreiter, J.; Lee, M.; Daher, S.; Bruder, G.; Welch, G. A systematic survey of 15 years of user studies published in the intelligent virtual agents conference. In Proceedings of the 18th International Conference on Intelligent Virtual Agents, Sydney, Australia, 5–8 November 2018; pp. 17–22.
18. Rosenfeld, A.; Agmon, N.; Maksimov, O.; Kraus, S. Intelligent agent supporting human–multi-robot team collaboration. *Artif. Intell.* **2017**, *252*, 211–231. [CrossRef]
19. Ramchurn, S.D.; Wu, F.; Jiang, W.; Fischer, J.E.; Reece, S.; Roberts, S.; Rodden, T.; Greenhalgh, C.; Jennings, N.R. Human–agent collaboration for disaster response. *Auton. Agents Multi-Agent Syst.* **2016**, *30*, 82–111.
20. Rosenfeld, A.; Richardson, A. Why, Who, What, When and How about Explainability in Human-Agent Systems. In Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, Auckland, New Zealand, 9–13 May 2020; pp. 2161–2164.
21. Schilling, M.; Kopp, S.; Wachsmuth, S.; Wrede, B.; Ritter, H.; Brox, T.; Nebel, B.; Burgard, W. Towards A Multidimensional Perspective on Shared Autonomy. In *The 2016 AAAI Fall Symposium Series: Shared Autonomy in Research and Practice*; Technical Report FS-16-05; The AAAI Press: Palo Alto, CA, USA, 2016; pp. 338–344.
22. Schilling, M.; Burgard, W.; Muelling, K.; Wrede, B.; Ritter, H. Editorial: Shared Autonomy— Learning of Joint Action and Human-Robot Collaboration. *Front. Neurorobot.* **2019**, *13*. [CrossRef]
23. Rousseau, D.M.; Sitkin, S.B.; Burt, R.S.; Camerer, C. Not so different after all: A cross-discipline view of trust. *Acad. Manag. Rev.* **1998**, *23*, 393–404.
24. Mayer, R.C.; Davis, J.H.; Schoorman, F.D. An Integrative Model of Organizational Trust. *Acad. Manag. Rev.* **1995**, *20*, 709–734, [CrossRef]
25. McAllister, D.J. Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations. *Acad. Manag. J.* **1995**, *38*, 24–59. [CrossRef]
26. Schwaninger, I.; Fitzpatrick, G.; Weiss, A. Exploring Trust in Human-Agent Collaboration. In Proceedings of the 17th European Conference on Computer-Supported Cooperative Work, Salzburg, Austria, 8–12 June 2019; p. 12.
27. Gefen, D.; Straub, D.W. Consumer trust in B2C e-Commerce and the importance of social presence: Experiments in e-Products and e-Services. *Omega* **2004**, *32*, 407–424.
28. Handy, C. Trust and the virtual organization. *Long Range Plan.* **1995**, *28*, 126.
29. Bente, G.; Rüggenberg, S.; Krämer, N.C.; Eschenburg, F. Avatar-Mediated Networking: Increasing Social Presence and Interpersonal Trust in Net-Based Collaborations. *Hum. Commun. Res.* **2008**, *34*, 287–318. [CrossRef]

30. Johnson, D.; Grayson, K. Cognitive and affective trust in service relationships. *J. Bus. Res.* **2005**, *58*, 500–507. [CrossRef]
31. Lewis, J.D.; Weigert, A. Trust as a social reality. *Soc. Forces* **1985**, *63*, 967–985.
32. Lee, J.D.; See, K.A. Trust in Automation: Designing for Appropriate Reliance. *Hum. Factors* **2004**, *46*, 50–80.
33. Gulati, S.; Sousa, S.; Lamas, D. Modelling Trust: An Empirical Assessment. In *Human-Computer Interaction—INTERACT 2017*; Lecture Notes in Computer Science; Bernhaupt, R., Dalvi, G., Joshi, A.K., Balkrishan, D., O'Neill, J., Winckler, M., Eds.; Springer International Publishing: Cham, Switzerland, 2017; Volume 10516, pp. 40–61. [CrossRef]
34. Kok, B.C.; Soh, H. Trust in Robots: Challenges and Opportunities. *Curr. Robot. Rep.* **2020**. [CrossRef]
35. Hancock, P.A.; Billings, D.R.; Schaefer, K.E.; Chen, J.Y.C.; de Visser, E.J.; Parasuraman, R. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Hum. Factors* **2011**, *53*, 517–527. [CrossRef] [PubMed]
36. Lee, J.; Moray, N. Trust, control strategies and allocation of function in human–machine systems. *Ergonomics* **1992**, *35*, 1243–1270. [CrossRef] [PubMed]
37. Rahman, S.M.M.; Sadrfaridpour, B.; Wang, Y. Trust-Based Optimal Subtask Allocation and Model Predictive Control for Human-Robot Collaborative Assembly in Manufacturing. In *Volume 2: Diagnostics and Detection; Drilling; Dynamics and Control of Wind Energy Systems; Energy Harvesting; Estimation and Identification; Flexible and Smart Structure Control; Fuels Cells/Energy Storage; Human Robot Interaction; HVAC Building Energy Management; Industrial Applications; Intelligent Transportation Systems; Manufacturing; Mechatronics; Modelling and Validation; Motion and Vibration Control Applications*; American Society of Mechanical Engineers: Columbus, OH, USA, 2015; p. V002T32A004. [CrossRef]
38. Rahman, S.M.M.; Wang, Y.; Walker, I.D.; Mears, L.; Pak, R.; Remy, S. Trust-based compliant robot-human handovers of payloads in collaborative assembly in flexible manufacturing. In Proceedings of the 2016 IEEE International Conference on Automation Science and Engineering (CASE), Fort Worth, TX, USA, 21–25 August 2016; pp. 355–360. [CrossRef]
39. Law, T. Measuring Relational Trust in Human-Robot Interactions. In Proceedings of the Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, L'Aquila, Italy, 7–9 October 2020; pp. 579–581. [CrossRef]
40. Jung, M.F. Affective Grounding in Human-Robot Interaction. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction—HRI '17, Vienna, Austria, 6–9 March 2017; pp. 263–273. [CrossRef]
41. Walker, I.D.; Mears, L.; Mizanoor, R.S.M.; Pak, R.; Remy, S.; Wang, Y. Robot-Human Handovers Based on Trust. In Proceedings of the 2015 Second International Conference on Mathematics and Computers in Sciences and in Industry (MCSI), Sliema, Malta, 17 August 2015; pp. 119–124. [CrossRef]
42. Vinanzi, S.; Patacchiola, M.; Chella, A.; Cangelosi, A. Would a robot trust you? Developmental robotics model of trust and theory of mind. *Philos. Trans. R. Soc. B Biol. Sci.* **2019**, *374*, 20180032. [CrossRef]
43. Argall, B.D.; Murphy, T.D. Computable Trust in Human Instruction. In Proceedings of the AAAI Fall Symposia, Arlington, VA, USA, 13–15 November 2014; p. 2.
44. Basu, C.; Singhal, M. *Trust Dynamics in Human Autonomous Vehicle Interaction: A Review of Trust Models*; 2016 AAAI Spring Symposium Series; The AAAI Press: Palo Alto, CA, USA, 2016; p. 7.
45. Alhaji, B.; Rausch, A.; Prilla, M. Toward Mutual Trust Modeling in Human-Robot Collaboration. *arXiv* **2020**, arXiv:2011.01056.
46. Alam, M.; Samad, M.D.; Vidyaratne, L.; Glandon, A.; Iftekharuddin, K.M. Survey on Deep Neural Networks in Speech and Vision Systems. *arXiv* **2019**, arXiv:1908.07656.
47. Grigorescu, S.; Trasnea, B.; Cocias, T.; Macesanu, G. A survey of deep learning techniques for autonomous driving. *J. Field Robot.* **2020**, *37*, 362–386.
48. Li, J.; Cheng, H.; Guo, H.; Qiu, S. Survey on artificial intelligence for vehicles. *Automot. Innov.* **2018**, *1*, 2–14.
49. Singh, H.P.; Dimri, P.; Tiwari, S.; Saraswat, M. Segmentation Techniques through Machine Based Learning for Latent Fingerprint Indexing and Identification. *JSIR* **2020**, *79*, 201–208.
50. ISO. *26262: Road Vehicles—Functional Safety*; Standard, International Organization for Standardization: Geneva, Switzerland, 2011.

51. ISO. *21448: Road Vehicles—Safety of the Intended Functionality*; Standard, International Organization for Standardization: Geneva, Switzerland, 2019.
52. Fulton, N.; Platzer, A. Safe Reinforcement Learning via Formal Methods: Toward Safe Control through Proof and Learning. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, LA, USA, 2–7 February 2018; pp. 6485–6492.
53. Alshiekh, M.; Bloem, R.; Ehlers, R.; Könighofer, B.; Niekum, S.; Topcu, U. Safe Reinforcement Learning via Shielding. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, LA, USA, 2–7 February 2018; pp. 2669–2678.
54. Aniculaesei, A.; Grieser, J.; Rausch, A.; Rehfeldt, K.; Warnecke, T. Towards a holistic software systems engineering approach for dependable autonomous systems. In Proceedings of the 1st International Workshop on Software Engineering for AI in Autonomous Systems—SEFAIS '18, Madrid, Spain, 22–30 May 2018; pp. 23–30. [CrossRef]
55. Wagner, A.R.; Robinette, P.; Howard, A. Modeling the Human-Robot Trust Phenomenon: A Conceptual Framework based on Risk. *ACM Trans. Interact. Intell. Syst.* **2018**, *8*, 1–24. [CrossRef]
56. Schlossman, R.; Kim, M.; Topcu, U.; Sentis, L. Toward Achieving Formal Guarantees for Human-Aware Controllers in Human-Robot Interactions. *arXiv* **2019**, arXiv:1903.01350.
57. Askarpour, M.; Mandrioli, D.; Rossi, M.; Vicentini, F. SAFER-HRC: Safety Analysis through Formal vERification in Human-Robot Collaboration. In *Computer Safety, Reliability, and Security*; Lecture Notes in Computer Science; Skavhaug, A., Guiochet, J., Bitsch, F., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 283–295. [CrossRef]
58. Askarpour, M.; Mandrioli, D.; Rossi, M.; Vicentini, F. Formal model of human erroneous behavior for safety analysis in collaborative robotics. *Robot. Comput. Integr. Manuf.* **2019**, *57*, 465–476. [CrossRef]
59. Stankovic, J. Research directions for the internet of things. *IEEE Internet Things J.* **2014**, *1*, 3–9.
60. Stankovic, J.; Munir, S.; Liang, C.; Lin, S. Cyber physical system challenges for human-in-the-loop control. In Proceedings of the 8th International Workshop on Feedback Computing, San Jose, CA, USA, 24–28 June 2013.
61. McRuer, D.; Jex, H. A review of quasi-linear pilot models. *IEEE Trans. Hum. Factors Electron.* **1967**, *3*, 231–249.
62. McRuer, D.; Krendel, E. The human operator as a servo system element. *J. Frankl. Inst.* **1959**, *267*, 381–403.
63. McRuer, D.; Krendel, E. *Mathematical Models of Human Pilot Behavior*; Advisory Group for Aerospace Research and Development Neuilly-Sur-Seine: Neuilly-Sur-Seine, France, 1974.
64. Wray, R.E.; Chong, R.S. Comparing Cognitive Models and Human Behavior Models: Two Computational Tools for Expressing Human Behavior. *J. Aerosp. Comput., Inf. Commun.* **2007**, *4*, 836–852. [CrossRef]
65. Shanahan, M. Perception as abduction: Turning sensor data into meaningful representation. *Cogn. Sci.* **2005**, *29*, 103–134. [PubMed]
66. Winfield, A. Experiments in artificial theory of mind: From safety to story-telling. *Front. Robot. AI* **2018**, *5*, 75.
67. Nunes, D.S.; Zhang, P.; Sa Silva, J. A Survey on Human-in-the-Loop Applications Towards an Internet of All. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 944–965. [CrossRef]
68. Eskins, D.; Sanders, W.H. The Multiple-Asymmetric-Utility System Model: A Framework for Modeling Cyber-Human Systems. In Proceedings of the 2011 Eighth International Conference on Quantitative Evaluation of SysTems, Aachen, Germany, 5–8 September 2011; pp. 233–242. [CrossRef]
69. Jacq, A.; Johal, W.; Dillenbourg, P.; Paiva, A. Cognitive Architecture for Mutual Modelling. *arXiv* **2016**, arXiv:1602.06703.
70. Braunagel, C.; Rosenstiel, W.; Kasneci, E. Ready for take-over? A new driver assistance system for an automated classification of driver take-over readiness. *IEEE Intell. Transp. Syst. Mag.* **2017**, *9*, 10–22.
71. Mioch, T.; Kroon, L.; Neerincx, M. Driver readiness model for regulating the transfer from automation to human control. In Proceedings of the 22nd International Conference on Intelligent User Interfaces, Limassol, Cyprus, 13–16 March 2017; pp. 205–213.
72. Deo, N.; Trivedi, M. Looking at the driver/rider in autonomous vehicles to predict take-over readiness. *IEEE Trans. Intell. Veh.* **2019**, *5*, 41–52.
73. Watson, J.M.; Memmott, M.G.; Moffitt, C.C.; Coleman, J.; Turrill, J.; Fernández, Á.; Strayer, D.L. On working memory and a productivity illusion in distracted driving. *J. Appl. Res. Mem. Cogn.* **2016**, *5*, 445–453.

74. Lenskiy, A.; Lee, J.S. Driver's eye blinking detection using novel color and texture segmentation algorithms. *Int. J. Control Autom. Syst.* **2012**, *10*, 317–327.
75. Liu, A.; Li, Z.; Wang, L.; Zhao, Y. A practical driver fatigue detection algorithm based on eye state. In Proceedings of the 2010 Asia Pacific Conference on Postgraduate Research in Microelectronics and Electronics (PrimeAsia), Shanghai, China, 22–24 September 2010; pp. 235–238.
76. Yan, C.; Coenen, F.; Zhang, B. Driving posture recognition by convolutional neural networks. *IET Comput. Vis.* **2016**, *10*, 103–114.
77. Yüce, A.; Gao, H.; Cuendet, G.; Thiran, J.P. Action Units and Their Cross-Correlations for Prediction of Cognitive Load during Driving. *IEEE Trans. Affect. Comput.* **2017**, *8*, 161–175.
78. Satyanarayanan, M. Pervasive computing: Vision and challenges. *IEEE Pers. Commun.* **2001**, *8*, 10–17.
79. Schmidt, A.; Beigl, M.; Gellersen, H.W. There is more to context than location. *Comput. Graph.* **1999**, *23*, 893–901. [CrossRef]
80. Lance, B.J.; Kerick, S.E.; Ries, A.J.; Oie, K.S.; McDowell, K. Brain–Computer Interface Technologies in the Coming Decades. *Proc. IEEE* **2012**, *100*, 1585–1599. [CrossRef]
81. McCaffrey, C.; Chevalerias, O.; O'Mathuna, C.; Twomey, K. Swallowable-capsule technology. *IEEE Pervasive Comput.* **2008**, *7*, 23–29.
82. Scalise, L.; Bernacchia, N.; Ercoli, I.; Marchionni, P. Heart rate measurement in neonatal patients using a webcamera. In Proceedings of the 2012 IEEE International Symposium on Medical Measurements and Applications Proceedings, Budapest, Hungary, 18–19 May 2012; pp. 1–4.
83. Morbiducci, U.; Scalise, L.; De Melis, M.; Grigioni, M. Optical vibrocardiography: A novel tool for the optical monitoring of cardiac activity. *Ann. Biomed. Eng.* **2007**, *35*, 45–58.
84. Chen, M.; O'Sullivan, J.A.; Singla, N.; Sirevaag, E.J.; Kristjansson, S.D.; Lai, P.H.; Kaplan, A.D.; Rohrbaugh, J.W. Laser doppler vibrometry measures of physiological function: evaluation of biometric capabilities. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 449–460.
85. Mignanelli, L.; Rembe, C. Non-contact Health Monitoring with LDV. In *Laser Doppler Vibrometry for Non-Contact Diagnostics*; Kroschel, K., Ed.; Springer International Publishing: Cham, Switzerland, 2020; Volume 9, pp. 1–8. [CrossRef]
86. Endsley, M. Direct measurement of situation awareness: Validity and use of SAGAT. In *Situational Awareness*; Routledge: London, UK, 2017; pp. 129–156.
87. Endsley, M. Situation awareness global assessment technique (SAGAT). In Proceedings of the IEEE 1988 National Aerospace and Electronics Conference, Dayton, OH, USA, 23–27 May 1988; pp. 789–795.
88. Czerwinski, M.; Cutrell, E.; Horvitz, E. Instant messaging and interruption: Influence of task type on performance. In Proceedings of the OZCHI 2000 Conference Proceedings, Sydney, Australia, 4–8 December 2000; Volume 356, pp. 361–367.
89. Khushaba, R.; Kodagoda, S.; Lal, S.; Dissanayake, G. Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 121–131. [PubMed]
90. Lin, F.C.; Ko, L.W.; Chuang, C.H.; Su, T.P.; Lin, C.T. Generalized EEG-Based Drowsiness Prediction System by Using a Self-Organizing Neural Fuzzy System. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2012**, *59*, 2044–2055. [CrossRef]
91. Borghini, G.; Astolfi, L.; Vecchiato, G.; Mattia, D.; Babiloni, F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* **2014**, *44*, 58–75. [PubMed]
92. Park, S.; Whang, M. Infrared camera-based non-contact measurement of brain activity from pupillary rhythms. *Front. Physiol.* **2018**, *9*, 1400. [PubMed]
93. Rohweder, N.O.; Gertheiss, J.; Rembe, C. Towards a remote EEG for human–machine-interfaces. In *Forum Bildverarbeitung 2020*; Längle, T., Ed.; KIT Scientific Publishing: Karlsruhe, Germany, 2020.
94. Akyildiz, I.F.; Su, W.; Sankarasubramaniam, Y.; Cayirci, E. Wireless sensor networks: A survey. *Comput. Netw.* **2002**, *38*, 393–422. [CrossRef]

95. Chen, M.; Gonzalez, S.; Vasilakos, A.; Cao, H.; Leung, V.C.M. Body Area Networks: A Survey. *Mobile Netw. Appl.* **2011**, *16*, 171–193. [CrossRef]
96. Xu, B.; Li, J.; Wong, Y.; Zhao, Q.; Kankanhalli, M.S. Interact as You Intend: Intention-Driven Human-Object Interaction Detection. *IEEE Trans. Multimed.* **2020**, *22*, 1423–1432. [CrossRef]
97. Van Laerhoven, K.; Gellersen, H.W. Spine versus Porcupine: A Study in Distributed Wearable Activity Recognition. In Proceedings of the Eighth International Symposium on Wearable Computers, Arlington, VA, USA, 31 October–3 November 2004; pp. 142–149. [CrossRef]
98. Schilit, B.; Adams, N.; Want, R. Context-aware computing applications. In Proceedings of the 1994 First Workshop on Mobile Computing Systems and Applications, Santa Cruz, CA, USA, 8–9 December 1994; pp. 85–90.
99. Paganelli, F.; Giuli, D. An Evaluation of Context-Aware Infomobility Systems. In *Context-Aware Mobile and Ubiquitous Computing for Enhanced Usability*; IGI Global: Hershey, PA, USA, 2009. [CrossRef]
100. Sørensen, H.; Goldsmith, J.; Sangalli, L. An introduction with medical applications to functional data analysis. *Stat. Med.* **2013**, *32*, 5222–5240.
101. Scheipl, F.; Gertheiss, J.; Greven, S. Generalized functional additive mixed models. *Electron. J. Stat.* **2016**, *10*, 1455–1492. [CrossRef]
102. Xue, H.; Huynh, D.Q.; Reynolds, M. SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1186–1194.
103. Karim, F.; Majumdar, S.; Darabi, H.; Chen, S. LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access* **2018**, *6*, 1662–1669. [CrossRef]
104. Mummadi, C.K.; Leo, F.P.P.; Verma, K.D.; Kasireddy, S.; Scholl, P.M.; Kempfle, J.; Laerhoven, K.V. Real-time and embedded detection of hand gestures with an IMU-based glove. *Inform. Multidiscip. Digit. Publ. Inst.* **2018**, *5*, 28.
105. Carlson, T.; Demiris, Y. Collaborative control for a robotic wheelchair: Evaluation of performance, attention, and workload. *IEEE Trans. Syst., Man Cybern. Part B (Cybern.)* **2012**, *42*, 876–888.
106. Ajoudani, A.; Zanchettin, A.M.; Ivaldi, S.; Albu-Schäffer, A.; Kosuge, K.; Khatib, O. Progress and prospects of the human–robot collaboration. *Auton. Robots* **2018**, *42*, 957–975. [CrossRef]
107. Pereira, A.; Althoff, M. Calculating human reachable occupancy for guaranteed collision-free planning. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 4473–4480. [CrossRef]
108. Zanchettin, A.M.; Rocco, P. Probabilistic inference of human arm reaching target for effective human–robot collaboration. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 6595–6600.
109. Johora, F.T.; Müller, J.P. Zone-Specific Interaction Modeling of Pedestrians and Cars in Shared Spaces. *Transp. Res. Procedia* **2020**, *47*, 251–258. [CrossRef]
110. Vizzari, G.; Manenti, L.; Ohtsuka, K.; Shimura, K. An agent-based pedestrian and group dynamics model applied to experimental and real-world scenarios. *J. Intell. Transp. Syst.* **2015**, *19*, 32–45.
111. Schöbel, A.; Pätzold, J.; Müller, J.P. The Trickle-In Effect: Modeling Passenger Behavior in Delay Management. In Proceedings of the 19th Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2019), Munich, Germany, 12–13,September 2019; pp. 1–15. [CrossRef]
112. Zhang, J.; Liu, H.; Chang, Q.; Wang, L.; Gao, R.X. Recurrent neural network for motion trajectory prediction in human–robot collaborative assembly. *CIRP Ann.* **2020**, *69*, 9–12.
113. Richter, A.; Reinhardt, A.; Reinhardt, D. Privacy-Preserving Human-Machine Co-existence on Smart Factory Shop Floors. In *International Workshop on Simulation Science*; Springer: Cham, Germany, 2019; pp. 3–20.
114. Li, Y.; Chen, M.; Dou, Z.; Zheng, X.; Cheng, Y.; Mebarki, A. A review of cellular automata models for crowd evacuation. *Phys. A Stat. Mech. Appl.* **2019**, *526*, 120752.
115. Rudenko, A.; Palmieri, L.; Lilienthal, A.J.; Arras, K.O. Human motion prediction under social grouping constraints. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 3358–3364.

116. Coscia, P.; Castaldo, F.; Palmieri, F.A.; Alahi, A.; Savarese, S.; Ballan, L. Long-term path prediction in urban scenarios using circular distributions. *Image Vis. Comput.* **2018**, *69*, 81–91.
117. Kooij, J.F.; Flohr, F.; Pool, E.A.; Gavrila, D.M. Context-based path prediction for targets with switching dynamics. *Int. J. Comput. Vis.* **2019**, *127*, 239–262.
118. Chen, X.; Treiber, M.; Kanagaraj, V.; Li, H. Social force models for pedestrian traffic–state of the art. *Transp. Rev.* **2018**, *38*, 625–653.
119. Johora, F.T.; Cheng, H.; Müller, J.P.; Sester, M. An Agent-Based Model for Trajectory Modelling in Shared Spaces: A Combination of Expert-Based and Deep Learning Approaches. In Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems, AAMAS '20, Auckland, New Zealand, 9–13 May 2020; pp. 1878–1880.
120. Kucner, T.P.; Magnusson, M.; Schaffernicht, E.; Bennetts, V.H.; Lilienthal, A.J. Enabling flow awareness for mobile robots in partially observable environments. *IEEE Robot. Autom. Lett.* **2017**, *2*, 1093–1100.
121. Rehder, E.; Wirth, F.; Lauer, M.; Stiller, C. Pedestrian prediction by planning using deep neural networks. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; pp. 1–5.
122. Kato, Y.; Nagano, Y.; Yokoyama, H. A pedestrian model in human–robot coexisting environment for mobile robot navigation. In Proceedings of the 2017 IEEE/SICE International Symposium on System Integration (SII), Taipei, Taiwan, 11–14 December 2017; pp. 992–997. [CrossRef]
123. Antonucci, A.; Fontanelli, D. Towards a Predictive Behavioural Model for Service Robots in Shared Environments. In Proceedings of the 2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), Genova, Italy, 27–29 September 2018; pp. 9–14.
124. Medina, J.R.; Lorenz, T.; Hirche, S. Considering Human Behavior Uncertainty and Disagreements in Human–Robot Cooperative Manipulation. In *Trends in Control and Decision-Making for Human–Robot Collaboration Systems*; Wang, Y., Zhang, F., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 207–240. [CrossRef]
125. Krüger, J.; Lien, T.K.; Verl, A. Cooperation of human and machines in assembly lines. *CIRP Ann.* **2009**, *58*, 628–646. [CrossRef]
126. Donadio, F.; Frejaville, J.; Larnier, S.; Vetault, S. Human-robot collaboration to perform aircraft inspection in working environment. In Proceedings of the 5th International Conference on Machine Control and Guidance (MCG 2016), Vichy, France, 5–6,October 2016; p. 9.
127. Gely, C.; Trentesaux, D.; Le Mortellec, A. Maintenance of the Autonomous Train: A Human-Machine Cooperation Framework. In *Towards User-Centric Transport in Europe 2*; Müller, B., Meyer, G., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 135–148. [CrossRef]
128. Bogue, R. Robots in recycling and disassembly. *Ind. Robot* **2019**, *46*, 461–466. [CrossRef]
129. Bogue, R. Robots in healthcare. *Ind. Robot* **2011**, *38*, 218–223. [CrossRef]
130. Delmerico, J.; Mintchev, S.; Giusti, A.; Gromov, B.; Melo, K.; Horvat, T.; Cadena, C.; Hutter, M.; Ijspeert, A.; Floreano, D.; et al. The current state and future outlook of rescue robotics. *J. Field Robot.* **2019**, *36*, 1171–1191. [CrossRef]
131. Schwindt, C. *Resource Allocation in Project Management*; GOR-Publications, Springer: Berlin/Heidelberg, Germany, 2005.
132. Schwindt, C.; Zimmermann, J. (Eds.) *Handbook on Project Management and Scheduling Vol. 2*; International Handbooks on Information Systems; Springer International Publishing: Cham, Switzerland, 2015. [CrossRef]
133. Gombolay, M.; Bair, A.; Huang, C.; Shah, J. Computational design of mixed-initiative human–robot teaming that considers human factors: Situational awareness, workload, and workflow preferences. *Int. J. Robot. Res.* **2017**, *36*, 597–617. [CrossRef]
134. Mohan, S. Scheduling part-time personnel with availability restrictions and preferences to maximize employee satisfaction. *Math. Comput. Model.* **2008**, *48*, 1806–1813. [CrossRef]
135. Längle, T.; Wörn, H. Human-Robot Cooperation Using Multi-Agent-Systems. *J. Intell. Robot. Syst.* **2001**, *32*, 143–160. [CrossRef]

136. Damacharla, P.; Javaid, A.Y.; Gallimore, J.J.; Devabhaktuni, V.K. Common metrics to benchmark human–machine teams (HMT): A review. *IEEE Access* **2018**, *6*, 38637–38655. [CrossRef]
137. Mas-Colell, A.; Whinston, M.D.; Green, J.R. *Microeconomic Theory*; Oxford University Press: Oxford, UK, 1995.
138. Bruni, M.E.; Beraldi, P.; Guerriero, F. The Stochastic Resource-Constrained Project Scheduling Problem. In *Handbook on Project Management and Scheduling Vol. 2*; Schwindt, C., Zimmermann, J., Eds.; International Handbooks on Information Systems; Springer International Publishing: Cham, Switzerland, 2015; pp. 811–835. [CrossRef]
139. Möhring, R.H. Scheduling under Uncertainty: Optimizing against a Randomizing Adversary. In *Approximation Algorithms for Combinatorial Optimization*; Goos, G., Hartmanis, J., van Leeuwen, J., Jansen, K., Khuller, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2000; Volume 1913, pp. 15–26. [CrossRef]
140. Artigues, C.; Leus, R.; Nobibon, F.T. Robust Optimization for the Resource-Constrained Project Scheduling Problem with Duration Uncertainty. In *Handbook on Project Management and Scheduling Vol. 2*; Schwindt, C., Zimmermann, J., Eds.; International Handbooks on Information Systems; Springer International Publishing: Cham, Switzerland, 2015; pp. 875–908. [CrossRef]
141. Özdamar, L.; Alanya, E. Uncertainty modelling in software development projects (with case study). *Ann. Oper. Res.* **2001**, *102*, 157–178. [CrossRef]
142. T'kindt, V.; Billaut, J.C. *Multicriteria Scheduling: Theory, Models and Algorithms*; Springer: Berlin/Heidelberg, Germany, 2006.
143. Agnetis, A.; Billaut, J.C.; Gawiejnowicz, S.; Pacciarelli, D.; Soukhal, A. *Multiagent Scheduling*; Springer: Berlin/Heidelberg, Germany, 2014. [CrossRef]
144. Immorlica, N.; Li, L.E.; Mirrokni, V.S.; Schulz, A.S. Coordination mechanisms for selfish scheduling. *Theor. Comput. Sci.* **2009**, *410*, 1589–1598. [CrossRef]
145. Fink, A.; Homberger, J. Decentralized Multi-Project Scheduling. In *Handbook on Project Management and Scheduling Vol. 2*; Schwindt, C., Zimmermann, J., Eds.; International Handbooks on Information Systems; Springer International Publishing: Cham, Switzerland, 2015; pp. 685–706. [CrossRef]
146. Wellman, M.P.; Walsh, W.E.; Wurman, P.R.; MacKie-Mason, J.K. Auction protocols for decentralized scheduling. *Games Econ. Behav.* **2001**, *35*, 271–303. [CrossRef]
147. Homberger, J. A (μ, λ)-coordination mechanism for agent-based multi-project scheduling. *OR Spectr.* **2012**, *34*, 107–132. [CrossRef]
148. Merz, F.; Schwindt, C.; Westphal, S.; Zimmermann, J. A multi-round auction for staff to job assignment under myopic best response dynamics. In Proceedings of the International Conference of Industrial Engineering and Engineering Management (IEEM 2020), Singapore, 14–17 December 2020.
149. Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; Procaccia, A.D. *Handbook of Computational Social Choice*; Cambridge University Press: Cambridge, UK, 2016.
150. Gallien, J.; Wein, L.M. A smart market for industrial procurement with capacity constraints. *Manag. Sci.* **2005**, *51*, 76–91. [CrossRef]
151. Sabater-Mir, J.; Vercounter, L. Chapter 9: Trust and reputation in multiagent systems. In *Multiagent Systems*; Weiss, G., Ed.; MIT Press: Cambridge, CA, USA, 2013.
152. Alos-Ferrer, C.; Farolfi, F. Trust games and beyond. *Front. Neurosci.* **2019**, *13*, 865–889. [CrossRef]
153. Hartmann, S.; Briskorn, D. A survey of variants and extensions of the resource-constrained project scheduling problem. *Eur. J. Oper. Res.* **2010**, *207*, 1–14. [CrossRef]
154. Neumann, K.; Schwindt, C. Project scheduling with inventory constraints. *Math. Methods Oper. Res. (ZOR)* **2002**, *56*, 513–533, [CrossRef]
155. Neumann, K.; Schwindt, C.; Zimmermann, J. *Project Scheduling with Time Windows and Scarce Resources*, 2nd ed.; Lecture Notes in Economics and Mathematical Systems; Springer: Berlin/Heidelberg, Germany, 2003.
156. Weiss, I. *The Resource Transfer Problem: A Framework for Integrated Scheduling and Routing Problems*; Springer International Publishing: Basel, Switzerland, 2019.
157. Macal, C.M.; North, M.J. Tutorial on agent-based modelling and simulation. *J. Simul.* **2010**, *4*, 151–162. [CrossRef]

158. Briem, L.; Mallig, N.; Vortisch, P. Creating an integrated agent-based travel demand model by combining mobiTopp and MATSim. *Procedia Comput. Sci.* **2019**, *151*, 776–781.
159. Ziemke, D.; Kaddoura, I.; Nagel, K. The MATSim Open Berlin Scenario: A multimodal agent-based transport simulation scenario based on synthetic demand modeling and open data. *Procedia Comput. Sci.* **2019**, *151*, 870–877.
160. Ahmed, S.; Johora, F.T.; Müller, J.P. *Investigating the Role of Pedestrian Groups in Shared Spaces through Simulation Modeling*; Simulation, S., Gunkelmann, N., Baum, M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 52–69.
161. Sebe, S.M.; Kraus, P.; Müller, J.P.; Westphal, S. Cross-provider Platoons for Same-day Delivery. In Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems, VEHITS 2019, Heraklion, Crete, Greece, 3–5 May 2019; pp. 106–116. [CrossRef]
162. Hesselmann, C.; Kehl, S.; Stiefel, P.; Müller, J.P. Decentralized handling of conflicts in multi-brand engineering change management. In Proceedings of the 21st International Conference on Engineering Design (ICED 17), Vancouver, BC, Canada, 21–25 August 2017; Volume 4, pp. 683–692.
163. Kraus, S.; Azaria, A.; Fiosina, J.; Greve, M.; Hazon, N.; Kolbe, L.; Lembcke, T.B.; Müller, J.P.; Schleibaum, S.; Vollrath, M. AI for Explaining Decisions in Multi-Agent Environment. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, New York, NY, USA, 7–12 February 2020.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.