Article

Smooth and Resilient Human–Machine Teamwork as an Industry 5.0 Design Challenge

Eija Kaasinen [1,*], Anu-Hanna Anttila [2], Päivi Heikkilä [1], Jari Laarni [1], Hanna Koskinen [1] and Antti Väätänen [1]

[1] VTT Technical Research Centre of Finland, 33101 Tampere, Finland; paivi.heikkila@vtt.fi (P.H.); jari.laarni@vtt.fi (J.L.); hanna.koskinen@vtt.fi (H.K.); antti.vaatanen@vtt.fi (A.V.)
[2] Finnish Industrial Union, 00531 Helsinki, Finland; anu-hanna.anttila@teollisuusliitto.fi
[*] Correspondence: eija.kaasinen@vtt.fi; Tel.: +358-40557-1005

Abstract: Smart machine companions such as artificial intelligence (AI) assistants and collaborative robots are rapidly populating the factory floor. Future factory floor workers will work in teams that include both human co-workers and smart machine actors. The visions of Industry 5.0 describe sustainable, resilient, and human-centered future factories that will require smart and resilient capabilities both from next-generation manufacturing systems and human operators. What kinds of approaches can help design these kinds of resilient human–machine teams and collaborations within them? In this paper, we analyze this design challenge, and we propose basing the design on the joint cognitive systems approach. The established joint cognitive systems approach can be complemented with approaches that support human centricity in the early phases of design, as well as in the development of continuously co-evolving human–machine teams. We propose approaches to observing and analyzing the collaboration in human–machine teams, developing the concept of operations with relevant stakeholders, and including ethical aspects in the design and development. We base our work on the joint cognitive systems approach and propose complementary approaches and methods, namely: actor–network theory, the concept of operations and ethically aware design. We identify their possibilities and challenges in designing and developing smooth human–machine teams for Industry 5.0 manufacturing systems.

Keywords: human–machine teams; Industry 5.0; joint cognitive systems; actor–network theory; concept of operations; ethics

1. Introduction

The European Commission has presented a vision for the future of European industry, coined 'Industry 5.0', in their policy brief [1]. Industry 5.0 complements the techno-economic vision of the Industry 4.0 paradigm [2] by emphasizing the societal role of industry. The policy brief calls for a transition in European industry to become sustainable, human-centric, and resilient. Industry 5.0 recognizes the power of industry to become a resilient provider of prosperity, by having a high degree of robustness, focusing on sustainable production, and placing the wellbeing of industry workers at the center of the production process.

Industry 4.0 is equipping the factory floor with cyber-physical systems that involve different human and smart-machine actors. These systems have much potential to improve the production process and to provide more versatile and interesting jobs for employees, as called for in the Industry 5.0 concept [1]. Reaping the full benefits of this potential requires that the new solutions are designed as sociotechnical systems, as suggested by Cagliano et al. [3], Neumann et al. [4] and Stern and Becker [5]. As illustrated in Figure 1, human–machine teams can include human actors with different skills, collaborative robots, and artificial intelligence (AI)-based systems for assistance or supervision. The design should cover the technical solutions and work organization in human–machine teams in

parallel. Task allocation in the human–machine teams should be designed to utilize the best capabilities of all actors, to ensure fair task allocation and to keep human work meaningful and manageable [6].
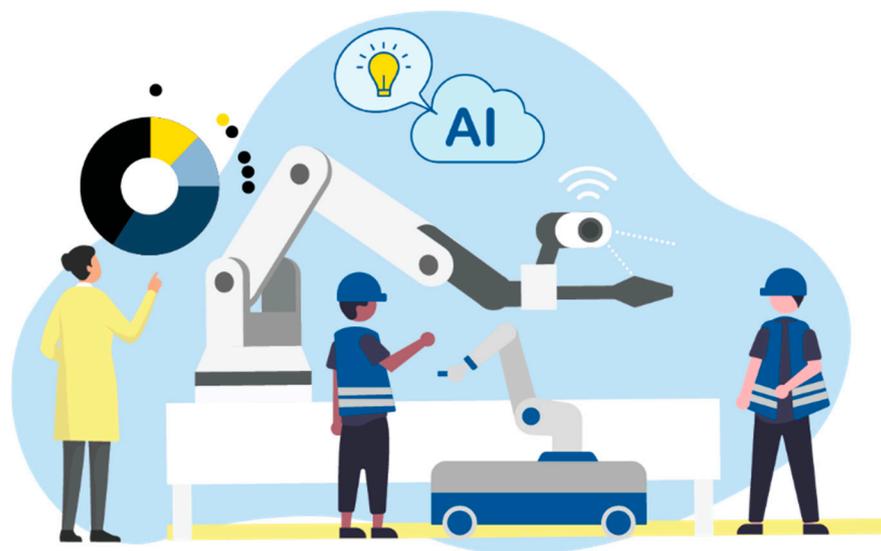

Figure 1. Human–machine teams optimally utilize the skills of each team member and the capabilities of smart machines.

Smart solutions based on collaborative robots and AI have much potential in easing physically and mentally demanding tasks. However, nowadays, these kinds of solutions tend to provide a two-fold experience to the workers: on the one hand, workers may feel that the work has become a passive and boring monitoring job, while on the other hand, they describe overwhelming challenges when they need to solve smart machine malfunctions [6]. There is clearly a need to develop work allocation and teamwork in human–machine teams so that human workers feel they are in the loop and that human jobs remain meaningful and manageable.

There are still many tasks where humans are superior to machines. Some manual tasks are far too challenging to automate, and human practical knowledge and intuition cannot be replaced with the analytical capabilities of AI. AI is good at learning from past data, while humans can think 'out of the box' and provide creative solutions. Integrating these two complementary viewpoints can bring about an ensemble that is greater than the sum of both human and AI capabilities [7]. Introducing smart solutions provides an opportunity to consider the roles of all actors, raising questions such as: What are the new tasks that each team member is responsible for? How should the team members collaborate and communicate? How can the competence development plans of each team member be supported?

Worker roles and related skills demands should be considered in the design of human–machine systems. This is in line with the initial value system in sociotechnical design [8]: while technology and organizational structures may change in the industry, the rights and needs of all employees must always be given a high priority. These rights and needs include varied and challenging work, good working conditions, learning opportunities, scope for making decisions, good training and supervision, and the potential for making progress. Moreover, in the dynamic Industry 5.0 environment, workers should be able to change their roles flexibly on the fly, based on learning new skills or simply for change [6].

Industry 5.0 includes three core elements: human-centricity, sustainability and resilience [1]. Resilience refers to the need to develop a higher degree of robustness in industrial production, better preparing it for disruptions. A central part of resilience is adaptability in the teams of humans and smart machines, where responsibilities and work division must be smoothly altered to react to changing situations.

In this paper, we investigate the challenge of smooth and resilient human–machine teamwork from a human-centric design point of view. The long research tradition in joint cognitive systems [9,10] provides a good basis for understanding the focus of the design: the overall system with common goals and various human and machine actors. Cognitive systems engineering is an established practice for designing complex human–machine systems [11,12]. However, complementary approaches and methods are needed to understand and develop the dynamic nature of the overall system where the capabilities and skills of all the actors change over time, and where the task allocation must be revised resiliently to adapt to the changes in the environment. In this paper our aim is to address the following research question: What kinds of approaches and methods could support human-centric, early-phase design and the continuous development of human–machine teams on the factory floor as smoothly working, resilient and continuously evolving joint cognitive systems?

In Section 2, we give an overview of related research regarding industrial human–machine systems and their design. In Section 3, we provide an overview of the research on joint cognitive systems and cognitive systems engineering, and we assess how these approaches could be utilized in designing human–machine teams on the factory floor. In this section we also identify the need for complementary approaches and methods. In Section 4, we propose three complementary approaches and methods to support human-centric early-phase design and the continuous development of human–machine teams on factory floor. The proposed complementary approaches and methods include: actor–network theory to observe dynamic human machine teams and their behavior; the concept of operations as a flexible co-design and development tool; and ethically aware design to foresee and solve ethical concerns. Finally, in Section 5, we discuss the proposed approach and methods, comparing them to other researchers' views.

2. Related Research

Several researchers have addressed the need to extend the focus of the design of industrial systems to the whole sociotechnical system (e.g., [3–5]). Cagliano et al. [3] suggest that technology and work organization should be studied both at the micro (work design) and the macro level (organizational structure) to implement successful smart manufacturing systems. Neumann et al. [4] propose a framework to systematically consider human factors in Industry 4.0 designs and implementations, integrating technical and social foci in the multidisciplinary design process. Their five-step framework includes (1) defining the technology to be introduced, (2) identifying humans in the system, (3) identifying for each human in the system the new and removed tasks, (4) assessing the human impact of the task changes, and (5) an outcome analysis to consider the possible implications on training needs, as well as the probability of errors, quality, and work well-being. The micro-level design of the sociotechnical system should focus on the sensory, cognitive and physical demands for the workers ([3,4]), human needs and abilities [5], change of work ([3–5]), job autonomy [3], supervision [4], support for co-workers [4] and interacting with semi-automated systems [5]. The macro-level design of the sociotechnical system should focus on organizing the work [5], decision making [5], social environment at the workplace [4] and the division of labor [5].

Stern and Becker [5] point out that many future industrial workplaces present themselves as human–machine systems, where tasks are carried out via interactions with several workers, production machines and cyber-physical assistance systems. An important design decision in these kinds of systems is the distribution of tasks between the various human and machine actors. Stern and Becker [5] propose design principles for these kinds of human–machine systems focusing on usability, feedback, and assistance systems. Grässler et al. [13] discuss integrating human factors in the model-based development of cyber-physical production systems. They claim that human actors are often greatly simplified in the model-based design, thus disregarding individual personality and skill profiles. They propose a human system integration (HSI) discipline as a promising approach that aims to

design the interaction between humans and technology in such a way that the needs and abilities of humans are appropriately implemented in the system design. In their approach, HSI concepts allow for the integration of individual capabilities of the workers as a fundamental part of a cyber-physical production system to enable the successful development of production systems [13].

Pacaux-Lemoine et al. [14] claim that the techno-centered design of intelligent manufacturing systems tends to demand extreme skills from the human operators as they are expected to handle any unexpected situations efficiently. Sgarbossa et al. [15] describe how the perceptual, cognitive, emotional and motoric demands on the user are determined in the design. If these demands are higher that the capacity of an individual, most probably there will be negative impacts both on system performance and worker well-being. Designers should understand employee diversity and the human factors related to new technologies [15]. Pacaux-Lemoine et al. [14] propose that human–machine collaboration should be based on a shared situational view at a plan level, a plan application level (triggering tasks) and at the level of directly controlling the process.

Kadir and Broberg [16] recommend following a systematic approach for (re)designing work systems with new digital solutions as well as involving all affected workers early in the design of new solutions. They also emphasize follow-up and standardizing the new ways of working after the implementation of new digital solutions.

In current research, human teamwork with collaborative robots and with autonomous agents has been studied separately. O'Neill et al. [17] argue that collaboration with agents and with robots is fundamentally different due to the embodiment of the robots. For example, the physical presence and design of a robot can impact the interactions with humans, the extent to which the robot engages and interests the human, and trust development.

Most human–robot interaction research focuses on a single human interacting with a single robot [18]. As robotics technology has advanced, robots have increasingly become capable of working together with humans to accomplish joint work, and thus raising design challenges related to interdependencies and teamwork [19]. The influence of robots on work in teams has been studied, focusing on task-specific aspects such as situational awareness, common ground, and task coordination [20]. Robots also affect the group's social functioning: robots evoke emotions, can redirect attention, shift communication patterns and can require changes to organizational routines [20]. Communication, coordination, and collaboration can be seen as cornerstones for human–robot teamwork [19]. Humans may face difficulty in creating mental models of robots and managing their expectations for the behavior and performance of robots, while robots struggle to recognize human intent [19].

Van Diggelen et al. [21] suggest design patterns to support the development of effective and resilient human–machine teamwork. They claim that traditional design methodologies do not sufficiently address the autonomous capabilities of agents, and thus often result in applications where the human becomes a supervisor rather than a teammate. The four types of design patterns suggested by van Diggelen et al. [21] support the description of dynamic team behavior in a team where human and machine actors each have a particular role, and the team together has a goal, which may change over time.

O'Neill et al. [17] present a literature review on empirical research related to human-autonomy teams (HAT) (teams of autonomous agents and humans, excluding human–robot teams). They state that the 40-year-long research tradition of human–automation interaction provides a vital foundation of knowledge for HAT research. What is different is that HAT involves autonomous agents, i.e., computer-based entities that are recognized to occupy a distinct role in the team and that have at least a partial level of autonomy [17]. Empirical HAT research is still in its infancy. From the 76 empirical papers that O'Neill et al. [17] found, 45 dealt with a single human working with a single autonomous agent, and only one paper studied a team with multiple humans and multiple agents.

The impacts of Industry 4.0 on the factory floor work have been studied under the theme "Operator 4.0", first introduced by Romero et al. [22]. Operator 4.0 describes human–automation symbiosis (or 'human cyber-physical systems') as characterized by

the cooperation of machines with humans in work systems, designed not to replace the skills and abilities of humans, but rather to co-exist with and assist humans in being more efficient and effective. Operator 4.0 studies introduce different operator types, based on utilizing different Industry 4.0 technologies. These operator types include, for example, a smarter operator with an intelligent personal assistant and a collaborative operator working with a collaborative robot [23]. The Operator 4.0 typology supports these changes in industrial work, but the focus is mainly on the viewpoint of individual operators rather than teamwork. Kaasinen et al. [6] further extended the Operator 4.0 typology by introducing a vision of future industrial work with personalized job roles, smooth teamwork in human–machine teams, and support for well-being at work.

Romero and Stahre [24] suggest that engineering smart resilient Industry 5.0 manufacturing systems implies adaptive automation and adjustable autonomous human and machine agents. Both human and machine capabilities are needed to avoid disruption, withstand disruption, adapt to unexpected changes and recover from unprecedented situations. The predictive capabilities of both humans and machines are needed to alert each other of potential disruptions. Adjustable autonomy supports modifying task allocation and shifting control to manage disruptions. With unexpected changes, human operators provide the highest contribution to resilience due to their agility, flexibility, and ingenuity. When recovering from unprecedented situations, human–machine mutual learning supports gradually improving self-healing properties.

Future manufacturing systems will be complex cyber-physical systems, where human actors and smart technologies operate collaboratively towards common goals. Human actors and technology may be physically separated, but not in their operation, where from a functional viewpoint, they are coupled in a higher-order organization, where they share knowledge about each other and their roles in the system. Human operators are generally acknowledged to use a model of the system (machine) with which they work. Similarly, the machine has an image of the operator. The designer of a human–machine system must recognize this and strive to obtain a match between the machine's image and the user characteristics on a cognitive level, rather than just on the level of physical functions. Different research and design approaches have been proposed to manage the challenge of designing these kinds of complex socio-technical systems.

In the design of socio-technical systems, technical, contextual, and human factors viewpoints should be considered. The socio-technical systems are dynamic and resilient, and the design should support designing adaptive systems with adaptive actors. These needs are becoming increasingly important as the systems include AI elements that facilitate continuous learning. The established design tradition of joint cognitive systems [9,10] is focused on designing complex human–machine systems, and thus provides a good basis for designing factory floor human–machine teams for Industry 5.0. In the following section, we describe the joint cognitive systems approach and related research. We also analyze how the joint cognitive systems approach could be complemented to support human centricity in the early-phase design and continuous development of dynamic and resilient human–machine systems for Industry 5.0.

3. Designing and Studying Joint Cognitive Systems

Hollnagel and Woods [9] argue that the disintegrated conception of the human–machine relationship has led to postulating an "interface" that connects the two separate elements. They point out that adopting an integrated approach would "change the emphasis from the interaction between humans and machines to human–technology co-agency" (p. 19). Understanding human–machine co-agency as a functional unity, led them to introduce a new design concept called a joint cognitive system (JCS). According to them, a "JCS is not defined by what it is but what it does" (p. 22). Consequently, the emphasis is on considering the functions of the JCS rather than of its parts. These authors see that the shift in perspective denotes a change in the research paradigm regarding human–technology interactions. This change in the research paradigm is widely adopted in the field of cognitive

system engineering (CSE), which JCS can also be viewed as being part of. They represent an alternative stream of research on human factors and the design of human-technology interaction ([11,12]), which earlier, to a large extent, was based on information-processing theories. That is to say, the human actor is seen as a creative factor, and in design the aim is to maximize the benefits of involving the human operators in the system operation by supporting collaboration and coordination between the system elements. Hollnagel and Woods [9] characterize JCSs according to three principles: (a) goal orientation, (b) control to minimize entropy (i.e., disorder in the system), and (c) co-agency in the service of objectives.

A cognitive system can be seen as an adaptive system, which functions in a future-oriented manner and uses knowledge about itself and the environment in the planning and modification of the system's actions [25]. Thus, the ability to maintain control of a situation, despite disrupting influences from the process itself or from the environment, is central. This requires considering the dynamics of the situation (i.e., creating an appropriate situational understanding) and accepting the fact that capabilities and needs depend on the situation and may vary over time [26]. Thus, according to Woods and Hollnagel [10] the focus of the analysis should be on the functioning of the human–technology–work system as a whole. They are especially interested in understanding how mutual adaptation takes place in a JCS and have recognized three types of adaptive mechanisms that they also labelled as patterns, that is, patterns of coordinated activity, patterns of resilience and patterns in affordance.

From the point of view of design, these patterns can be seen as empirical generalizations abstracted from the specific case study and situation and they are used to describe the functioning of the joint system. Thus, in design, the analysis of the dynamics and generic patterns of the JCS is proposed to be of central importance. For example, in his analysis of the role of automation in JCS, Hollnagel [27] emphasizes the importance of ensuring redundancy among functions assigned to the various parts of the system. He recognizes four distinct conditions related to various degrees of responsibility of automation in monitoring and control functions that all have their advantages and disadvantages. The opposite ends of the spectrum are conventional manual control and full automation take-over. The two other conditions in between are operation by delegation (human monitoring, automation controlling) and automation amplified attention (automation monitoring, with human control). When considering these four different conditions, it is easy to understand that none alone can be superior to the others in all operative situations. Instead, Hollnagel emphasizes that a balanced approach is necessary for considering the consequences of the defined automation strategy. According to him, the choice of an automation strategy is always a compromise between the efficiency and flexibility of the JCS since an increase in automation can positively affect efficiency but can also cause a certain inflexibility in the system and vice versa. Therefore, he concludes that "human centeredness is a reminder of the need to consider how a joint system can remain in control, rather than a goal. Each case requires its own considerations of which type of automation is the most appropriate, and how control can be enhanced and facilitated via a proper design of the automation involved" [27] (p. 11).

Norros and Salo [28] build partly on the ideas of the JCS (e.g., the notion of a pattern to describe joint system functioning) when developing their approach called the joint intelligent systems approach (JIS). The JIS approach acknowledges that technology is becoming ubiquitous and part of a new type of "intelligent environment", where the focus is no longer on interaction with an individual technological tool/solution. Thus, in JIS, the unit of analysis is extended to describe the structure of the joint system—a human, technology and environment system that is, by its very nature, coevolving. One influential researcher and author, Timo Järvilehto, has made particular contributions to how the relationship between humans and the environment is understood in JIS, that is to say, from a functional point of view, humans and their environment form a functional unity [29]. Here, the environment refers to the part of world that may potentially be useful for a particular organism. This understanding is also close to how, for example,

Gibson [30] defines human activity in an environment through affordances. Thus, a central idea is that the system is organized by its purpose and shaped by the constraints and possibilities of the environment, which must be considered when maintaining adaptive behavior. Moreover, in the JIS approach, "intelligence" is not an attribute of technology or the human element as such but, instead, it refers to the appropriate functioning and adaptation of a system. The whole adaptive system, or "intelligent environment," becomes the object of design [28]. Therefore, it follows that a JIS approach advocates methodologies that aid the understanding of how the system behaves, that is, the patterns of behavior or way of working that express an internal regularity in the behavior of the system. For instance, a semiotically grounded method for an empirical analysis of people's usages of their tools and technology [31] is suggested.

The joint cognitive systems approach has been proposed for the modeling of smart manufacturing systems by Jones et al. [32] and Chacon et al. [33]. Jones et al. [32] present smart manufacturing systems as collaborative agent systems that can include software and hardware, as well as machine, human, and organizational agents, among many others, in the service of jointly held goals. They propose modeling these agents as joint cognitive systems. They point out that cyber technologies have dramatically increased the cognitive capabilities of machines, changing machines from being reactive to self-aware. This allows humans and machines to work more collaboratively, as joint partners to execute cognitive functions. Humans and machines execute tasks as an integrated team, working on the same tasks and in the same temporal and physical spaces. This integrated view changes the emphasis from the interaction between humans and machines to human–machine co-agency. The behavior of these agents can be highly unpredictable due to the diverse technologies they use and behaviors they display. Jones et al. [32] propose that, since smart manufacturing systems represent joint cognitive systems, these systems should be engineered and managed according to the principles of cognitive systems engineering (CSE).

Chacon et al. [33] point out that the Industry 4.0 paradigm shift from doing to thinking has allowed for the emergence of cognition as a new perspective for intelligent systems. They see joint cognitive systems in manufacturing as the synergistic combination of different technologies such as artificial intelligence (AI), the Internet of Things (IoT) and multi-agent systems (MAS), which allow the operator and process to provide the necessary conditions to complete their work effectively and efficiently. Cognition emerges as goal-oriented interactions of people and artefacts to produce work in a specific context. In this situation, complexity emerges because neither goals, resources or constraints remain constant, creating dynamic couplings between artefacts, operators, and organizations. Chacon et al. [33] also refer to a CSE approach in analyzing how people manage complexity, understanding how artefacts are used and understanding how people and artefacts work together to create and organize joint cognitive systems, which constitute the basic unit of analysis in CSE.

The main challenges in the design of joint cognitive systems are recognized to be human–machine interactions and the allocation of functions between humans and autonomous systems. Therefore, methodologies have been developed to empirically approach and analyze the functioning of complex systems, including methods such as cognitive work analysis [12] and applied cognitive task analysis [34]. Regarding the allocation of functions in the design of joint cognitive systems, a key challenge is that the optimal allocation of functions depends on the operating conditions. In dynamic function allocation, several levels of autonomy are provided, and a decision procedure is applied to switch between these levels. Dynamic function scheduling is a special case of dynamic function allocation in which the re-allocation of a particular function is performed along the agents' timeline (e.g., [35]). Functional situation modelling (FSM), combining both chronological and functional views, is suitable for making considerations regarding dynamic function allocation ([36,37]). Functional situation modelling is based on Jen Rasmussen's abstraction hierarchy concept [38] and Kim Vicente's cognitive work analysis methodology [12]. The chronological and functional views determine a two-dimensional space in which important

human–machine actions and operational conditions are mapped. Actions are described under the functions that they are related to. The main function of the domain gives each action a contextual meaning. In the chronological view, the scenario is divided into several phases based on the goal of the activities; in the functional view, the critical functions that are endangered in a specific situation are illustrated [36].

Regarding interaction design, the main challenge is how to promote the mutual interpretation of the functional state of the joint system and the capability of adaptive actions to augment system resilience. On the one hand, this means assuring that the relevant human intentions, activities, and responsibilities are properly communicated to and understood by machines, and on the other hand, that the means by which humans interact with the machines meet their goals and needs. In particular, to ensure the smart integration of humans into the system, the focus should be on understanding how to improve the representation of system behavior and operative limits, in order to achieve its operative goals. One useful approach is to apply the ecological interface design (EID) approach, which is based on functional modelling [12] and the principles of James Gibson's ecological psychology [30]. EID offers a number of visual patterns for representing system states and opportunities for control; thus, it promotes the perception of functional states of the joint cognitive system in relation to the overall objectives of the activity.

Our research question is: What kinds of approaches and methods could support human-centric, early-phase design and the continuous development of human–machine teams on the factory floor as smoothly working, resilient and continuously evolving joint cognitive systems? The JCS approach and CSE design approach provide good bases for the design. However, to support the early-phase co-design of joint cognitive systems with relevant stakeholders, methods are needed for the planned allocation of functions, responsibilities, collaboration, and interaction to be described and shared with all stakeholders so that they can comment and contribute the design. Furthermore, as the human–machine system is continuously evolving, and due to the planned allocation of functions, responsibilities, collaboration and interaction change, there is a need for approaches with which the current status of the joint cognitive system and its actors can be studied. The increasing autonomy of machines raises many ethical concerns, and such concerns should be studied in the early design phases.

Human factors engineering (HFE) is a scientific approach to the application of knowledge regarding human factors and ergonomics to the design of complex technical systems. It can be considered as one engineering framework among several others under the CSE discipline [39]. It can be divided into tasks and activities in terms of which stages of the design process they are related to. Typically, HFE activities are classified into four groups: analysis, design, assessment, and implementation/operation. One of the first tasks in the analysis stage is to develop a Concept of Operations (ConOps) for the new system. The ConOps method was introduced by Fairley and Thayer in 1990s [40] as a bridge from operational requirements to technical specifications. The key task in the development of a ConOps is the allocation of functions and stakeholder requirements to the different elements of the proposed system. For example, it can be used to describe and organize the interaction between human operators and a swarm of autonomous or semi-autonomous robots. Typically, ConOps is considered to be a transitional design artefact, that plays a role in the requirements specification during the early stages of the design and involves various stakeholders. Thus, ConOps could support the early phase co-design of joint cognitive systems.

Actor–network theory (ANT) was developed in the early 1980s by the French scholars of science and technology studies (STS) Michel Callon and Bruno Latour, as well as British sociologist John Law [41]. ANT is a theory but also a methodology, which is typically used to describe the relations between humans and non-humans and ideas of technology. The theory argues that technology and the social environment interact with each other, forming complex networks. These sociotechnical networks [42] consist of multiple relationships between the social, the technological or material, and the semiotic. Digitalization has made

networks more complex and varied. In networks, the use of digital devices as actors can be just as important as human actors. They may be even more important, with smart machine actors that are not simply instruments for human interaction. To understand these unorthodox networks of human and non-human actors, more specific analysis tools are required. One option is to perceive them as actor–networks and name the actors as 'actants' when executing actor–network analyses. Thus, ANT could support the analysis of a continuously evolving human–machine system and the relations within it.

When designing a smooth human–machine teamwork, striving for ethically sound solutions is an important design goal. While the importance of ethics is widely acknowledged in design, whether it concerns human–machine teamwork or other fields of applying new technologies, embedding ethics in the practical work of designers may be challenging. Several approaches have been proposed to support considering ethics in the design process, and most of them highlight the need for already considering ethics in the early phases of design. Proactive ethical thinking to address ethical issues is particularly highlighted in the Ethics by Design approach [43] that aims to create a positive ethically aware mindset for the project group to actively work towards designing ethically sound solutions.

In the next section, we focus more on these three complementary approaches to support human-centric early phase design and the continuous development of human–machine teams on the factory floor. We will describe the approaches and the related methods, and we will analyze their opportunities and risks.

4. Approaches and Methods for Human-Centric Early-Phase Design and Continuous Development of Dynamic and Resilient Industry 5.0 Human–Machine Teams

Visions of the future present a factory floor where humans, collaborative robots, and autonomous agents form dynamic teams capable of reacting to changing needs in the production environment. The human actors have personally defined roles based on their current skills, but they can continuously develop their skills, and thus can also change their role in the team. Similarly, autonomous agents and collaborative robots are able to learn based on their artificial intelligence capabilities. All the actors are in continuous interaction by communicating, collaborating, and coordinating their responsibilities. How can this kind of a dynamic entity be designed? Clearly, after it has been initially designed and implemented, it starts developing on its own as both human actors and smart machine actors learn new skills and gain new capabilities. Based on the literature analysis and our own experiences, we propose that a joint cognitive systems approach can be complemented with three promising approaches and methods:

1. Actor–network theory is focused on describing sociotechnical network and interaction via empirical, evidence-based analyses. It is promising as it treats human and non-human actors equally. Actor–network theory can help understand how a dynamic human–machine team works and how it evolves over time. This can support design and development activities.

2. Concept of operations is a promising method and design tool that provides means to describe different actors and interdependencies between them. Compared to earlier methods based on modelling, it better supports both the dynamic nature of the overall system and co-design and development activities with relevant stakeholders.

3. Human–machine teams raise ethical concerns related to aspects such as task allocation, machine-based decision making, and human roles. The gradually evolving joint system may introduce new ethical issues that were not identified in the initial design. Thus, an ethical viewpoint should be an integral part of not only the initial design but also the continuous development of human–machine teams.

Figure 2 illustrates how the proposed three approaches incorporate different viewpoints for analyzing and designing Industry 5.0 human–machine teams as joint cognitive systems. In this section, we discuss actor–network theory, the concept of operations and ethically aware design in separate sub sections, analyzing what kinds of concrete support these

approaches and methods could bring to the design and development of human–machine teams. We also identify potential challenges and risks.
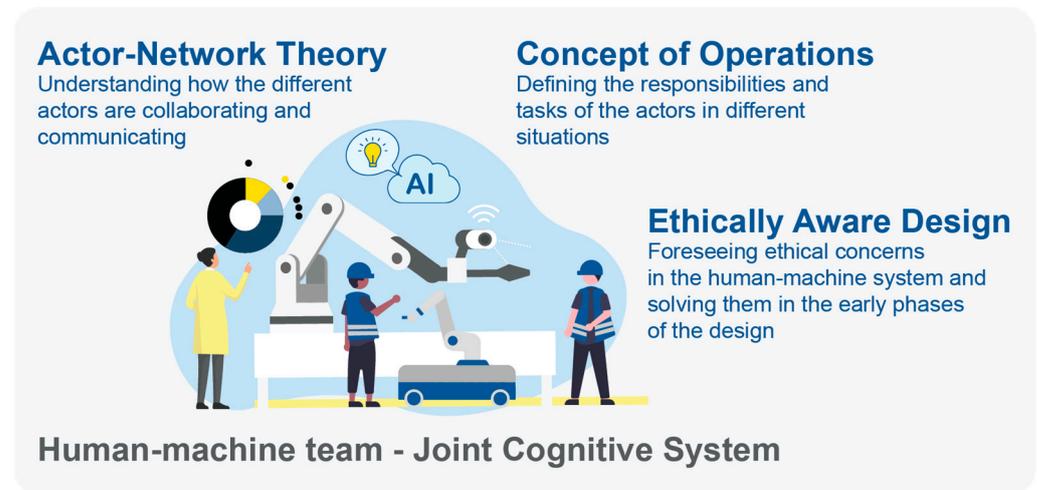

Figure 2. The proposed approaches and methods each bring a different viewpoint to the analysis and design of human–machine teams.

4.1. Actor–Network Theory (ANT)

Technology and the social environment interact with each other, forming complex networks between humans and non-humans. These networks consist of multiple relationships between the social, technological, and semiotic [41,42]. At best, actor–network analyses will clarify the roles and positions of actors, their relations, sociotechnical networks, and processes.

Industry 5.0 will create new networks, relations, and alliances, which include humans and non-humans. On the factory floor, non-humans are typically artefacts: machines, devices, and their human-created components. At many workplaces, networks are digitalized, workers need to act and interact with different actors: other humans, various machines, and new technologies. Three types of interaction in digitalized networks can be found: social interaction between humans, human–machine interactions, and interaction between machines.

Interaction between humans is social [44], and the oldest example of this is face-to-face interaction (F2F). In the analyses of social networks, attention is focused on human or organizational actors and their nodes (or hubs) or on the relations between these actors. It requires cultural knowledge to understand social relationships. Successful social networking requires understanding, trust, and commitment between human actors.

In other types of digitalized networks, at least one actor is non-human, and the interaction is characterized as a human–machine interaction (HMI) or, when interacting with digital devices, human–computer interaction (HCI). There is also a type of purely non-human interaction between machines and computers, better known as the Internet of Things (IoT).

In empirical analyses, actor–networks are usually approached ethnographically, as Bruno Latour and Steve Wooglar demonstrated in their classic research Laboratory Life: The Construction of Scientific Facts [45]; they showed the complexity of laboratory work. Latour and Wooglar wrote 'a thick description' with detailed descriptions and interpretations of the human and non-human activities they observed in the laboratory. As a result of their ethnographic field work, these descriptions were multi-dimensional, material, social, and semiotic.

ANT stresses the capacity of technology to be 'an actor' (sociology) or 'an actant' (semiotic) in and of itself, which influences and shapes interactions in material–semiotic networks. Here, the strict division of human and non-human (nature and culture) becomes

useless and all actors are treated with the same vocabulary. In an actor–network, actions can be blurred and mixed. For this reason, it can be hard to analyze causal relations or even to say which results different actions will produce in the network processes.

On the factory floor, non-human actors can be simple tools such as axes or hammers, or more complex items such as mechanical or automated machines, robotic hands or intelligent robots, in addition to elements such as electricity, software, microchips, or AI. Many of these actors are material, which makes it possible to observe their action and interaction. AI and software may obtain their appearance from intelligent robots, chatbots, digital logistic systems or augmented reality.

There are already examples of complex actor–networks in healthcare, where social and care robots are interacting with patients and seniors, registering their attitudes and habits, and collecting sensitive individual data. Christoph Lutz and Aurelia Tamò [46] analyzed privacy and healthcare robots with ANT. As a method, they argued that "ANT is a fruitful lens through which to study the interaction between humans and (social) robots because of its richness of useful concepts, its balanced perspective which evades both technology-determinism and social-determinism, and its openness towards new connected technology" [46].

Another essential key concept of ANT is 'translation'. According to Callon [41], translation has four stages: the problematization, interessement, enrollment and mobilization of network allies. John Shiga [47] has utilized translation in his analyses of how to distinguish human and social relations from technical and material artifacts (MP3s, iPods, and iTunes). His study shows how the non-human agency of artifacts constitutes the social world in translations of "industry statements, news reports, and technical papers on iPods, data compression codecs, and copy protection techniques". Shiga [47] argues that ANT "provides a useful framework for engaging with . . . foundational questions regarding the role of artifacts in contemporary life".

Even if robots or other AI-based actors can make decisions and enact them in the world, they are not sovereign actors (cf. [41,48]) or legal subjects in society. All these artefacts are not only non-humans but concurrently created by humans. Their presence can be observed with the senses, or it can take a more implicit form. One can say that the most extreme version of a net of non-human actants is the Internet of Things (IoT), but that is not correct. Even in this case, the influence of humans can be seen in the ideas and execution of technology. IoT-based services that utilize sensor data, clouds and Big Data networks are examples of the most sophisticated technologies where non-humans and humans are both present.

Actor–network methodology and analyses, for example, will show how things happen, how interaction works, and which actors will support sociotechnical network interaction. In general, it is a descriptive framework that aims to illustrate relationships between actants. Still, ANT does not explain social activity nor the causality of action. It does not give answers to the question of why.

The power of ANT lies in asking how and understanding the nature of agency and multiple interactions in sociotechnical networks. By writing an ethnographical, rich description, it becomes possible to understand more precisely how the interaction process works. ANT makes the active, passive, or neutral roles intrinsically visible, which 'actants' have taken, mistaken, or obtained in the network during the process. Additionally, understanding observed disorders and problems between human and non-human agency will underline the need for better design. The results of actor–network analyses can be utilized in designing smooth human–machine teams.

In the ANT method the purpose of action and the intention of agency are written in rich descriptions. With careful ethnographic fieldwork, the observed relations in the network, agency of humans and artifacts, and position of actors/actants can be determined in the analyses. The challenge of ANT is that the evaluation of human–machine teams becomes difficult if the executed fieldwork is poor, descriptions incomplete, and analyses

too shallow or blurry. Then, there is a risk that the ANT approach will be unable to produce useful information for the early phases of design.

4.2. Concept of Operations

IEEE standard 1362 [45] defines a ConOps document as a user-oriented document that describes a system's operational characteristics from the end user's viewpoint. It is used to communicate overall quantitative and qualitative system characteristics among the main stakeholders. ConOps documents have been developed in many domains, such as the military, health care, traffic control, space exploration, financial services, as well as different industries, such as nuclear power, pharmaceuticals, and medicine. At a surface glance, ConOps documents come in variety of forms reflecting the fact that they are developed for different purposes in different domains. Typically, ConOps documents are based on textual descriptions, but they may include informal graphics that aim to portray the key features of the proposed system, for example its objectives, operating processes, and main system elements [46].

ConOps has the potential to be used at all stages of the development process, and a general ConOps is a kind of template that can be modified and updated regarding specific needs and use cases [47]. A ConOps document typically describes the main system elements, stakeholders, tasks, and explanations of how the system works [49]. A typical ConOps document contains, in a suitable level of detail, the following kinds of information ([40,49]): the operational goals and constraints of the system, interfaces with external entities, main system elements and functions, operational environment and states, operating modes, allocation of responsibilities and tasks between humans and autonomous system elements, operating scenarios, and high-level user requirements. Väätänen et al. [50] pointed out that the ConOps specification should consider three main actors: the autonomous system, human operators, and other stakeholders. Initial ConOps diagrams can be created based on these three main parts of the ConOps structure. During the ConOps diagram development process, each actor can be described in more detail and their relationships with each other can be illustrated. On the other hand, the initial or case-specific ConOps diagrams can be used in co-design activities when defining, e.g., operator roles and requirements. The concept of operations approach illustrates how an autonomous system should function and how it can be operated in different tasks. Operators supervise the progress of the tasks, and they can react to possible changes. Robots can be fully autonomous in planned tasks and conditions, but in some situations an operator's actions may be required.

Laarni et al. [51] presented a classification of robot swarm management complexity. In their classification, the simplest operation mode is when one operator supervises a particular swarm. The most complicated situation was seen when human teams from different units co-operated with swarms from several classes. This scenario also considers that the human teams work together, and autonomous or semi-autonomous swarms engage in machine–machine interactions. The following numbered list indicates in more detail the six classes of robot swarm management complexity presented by Laarni et al. [51]:

1. An operator designates a task to one swarm belonging to one particular class and supervises it;
2. An operator designates a task to several swarms belonging to one particular class and supervises its progress;
3. An operator designates tasks to several swarms belonging to different classes and supervises their progress;
4. Several operators supervise several swarms belonging to different classes;
5. Human teams and swarms belonging to one particular class are designated;
6. Human teams and swarms belonging to several classes are designated.

A ConOps can be presented in different forms, and we have identified three types of ConOps that can be derived from the source: syntactic, interpretative, and practical.

We propose that since the uncertainties and novelties between actors become larger when moving from the syntactic to the practical level, "stronger" ConOps tools are required.

ConOps can be presented at different levels of detail so that, by zooming in and out of the ConOps hierarchy, different elements of the system come into focus. So, even though a ConOps is a high-level document by definition, this kind of high-level description can be outlined at different levels of system hierarchy.

Regarding collaborative robots and human–machine teams, these can be demonstrated at the level of factory infrastructure within a particular domain of manufacturing (i.e., system of systems level), at the level of a swarm of collaborative robots in a single factory (i.e., system level), and in the context of human-AI teams controlling a single robot (i.e., sub-system level).

Even though ConOps provides a good way to co-design human–machine systems with relevant stakeholders and to illustrate design decisions, the approach also has challenges and risks. Mostashari et al. [52] noted that ConOps can be perceived as burdensome rather than beneficial due to the laborious documentation requirements. Therefore, it may be difficult to justify its value to the design team and other stakeholders. Moreover, the contents and quality of ConOps documents can vary greatly, especially because ConOps standards and guidelines are poorly adhered to.

Since the development of ConOps is often a lengthy and laborious process, it may be challenging to estimate the required resources. ConOps work should be designed and tailored to match the available resources and time in development or research projects. This may be challenging if the proposed concept should address a wide range of actors and relationships and dependencies between them in a comprehensive and reliable way.

4.3. Ethically Aware Design

The successful design of human–machine collaboration is not only a matter of effective operations, but also a matter of ethical design choices. We encourage considering ethical aspects already in the early design phases, as well as later, in an iterative manner when socio-technical system and work practices evolve. For example, respecting workers' autonomy, privacy and dignity are important values to consider when allocating tasks between humans and machines and when designing how to maintain or increase the sense of meaningfulness at work, utilize the strengths of humans and machines and create work practices in which humans and machines complement each other.

To ensure that ethical aspects are considered in the design of smooth human–machine teamwork, several approaches can be applied. Ethical questions can be identified and addressed by identifying values of the target users and responding to them [53], by assessing the ethical impacts of design [54] or by creating and following ethical guidelines (see, e.g., [55]). In each of these approaches, different methods can be utilized.

One of the first and most established approaches for ethically aware design is value-sensitive design [53]. This is a theoretically grounded approach, which accounts for human values throughout the design process. The approach consists of three iterative parts: conceptual, empirical, and technical investigations. The conceptual part focuses on the identification of direct and indirect stakeholders affected by the design, definition of relevant values and identification of conflicts between competing values. Empirical investigations refer to studies conducted in usage contexts that focus on the values of individuals and groups affected by the technology and the evaluation of designs or prototypes. Technical investigations focus on the technology and design itself—how the technical properties support or hinder human values, and the proactive design of the system to support values identified in the conceptual investigation.

While value-sensitive design focuses on identifying the values of stakeholders, ethics can also be approached from the perspective of assessing the potential ethical impacts of adopting new technologies. A framework for ethical impact assessment [54] is based on four key principles: respect for autonomy, nonmaleficence, beneficence and justice. Each

of these principles includes several sub-sections and example questions for a designer, in order to consider the impacts of technology from the perspective of each principle.

More straightforwardly, ethics can be considered in design by following guidelines that can be used as guidance throughout the design process. For example, Kaasinen et al. [56] define five ethical guidelines to support ethics-aware design in the context of Operator 4.0 solutions, applying the earlier work by Ikonen et al. [55]. The guidelines are based on the ethical themes of privacy, autonomy, integration, dignity, reliability, and inclusion, and they describe the guidance related to each theme with one guiding sentence. For example, to support workers' autonomy, the guidelines suggest that the designed solutions should allow the operators to choose their own way of working, and to support inclusion, the designed solutions should be accessible to operators with different capabilities and skills.

Although established methods for ethically aware design exist, ethical aspects are not always considered in the design process. Even though ethics may be felt to be an important area of design, the lack of concrete tools or integration in the design process may prevent designers from considering ethical aspects in design. Ethical values or principles may sound either too abstract to be interpreted for design decisions or they may include an exhausting list of principles, which does not encourage inexperienced designers to study these principles. Moreover, the first impression of the design team may be that the new work processes would not include any ethical aspects, even though there may be several significant ethical questions when the process is regarded from a wider perspective of several stakeholders or from the perspective of selected ethical themes. In addition, new work processes may evoke unexpected ethical issues while new ways of working are adopted and gradually evolve.

What would the key means be to overcome the challenges of embedding ethics into existing design practices? In line with the ethics by design approach (Niemelä et al., 2014), one of the most important starting points is to already elicit ethical thinking in the early design process as a joint action to the design team and relevant stakeholders. This approach has a number of benefits: it supports proactive ethical thinking, engages diverse ethical perspectives on the design, and encourages commitment to the identified values or principles. Ethics awareness is more likely to be maintained throughout the design process when the ethical aspects are not brought from the outside of the project but co-identified within the project team. Co-created guidelines or principles are easier to understand and to concretize than a list of ready-made aspects. Furthermore, they are probably more relevant to the existing organization as they include aspects identified as particularly pertinent to the work process being designed.

An ethical design approach is particularly important in phases when value conflicts may occur and design compromises must be made. For example, pursuing both the efficiency of operations and workers' well-being may require making design choices that balance these impacts and design thinking that does not only consider short-term objectives, but also long-term sustainability. As all of the impacts of changing work practices and co-working with machines cannot be foreseen beforehand, an ethically aware mindset is also needed later to analyze the impacts and guide the process in the desired direction.

Ethically aware design should reduce the risks of planning human–machine teamwork that does not meet the needs and values of the stakeholders. Yet, encouraging ethical thinking carries a risk of overanalyzing the potential consequences of work changes and excessive thinking about exceptional situations or risks of misuse. This may hinder or slow down the adoption of new ways of working. To avoid this, ethics should be seen as one perspective of design among other important aspects, such as usability, safety or ergonomics. A framework that includes all such viewpoints, such as a design and evaluation framework for Operator 4.0 solutions [56] can serve as a tool to create a joint ethically aware mindset in the design group as well as serving as a reminder to include ethics in the design activities throughout the design process.

Ethics can be discussed as one theme in user studies or stakeholder workshops, and it can be considered when defining user experience goals for design. For example, a goal

to make workers feel encouraged and empowered at work [57] supports ethical design that does not concentrate on problems or exceptional situations, but drives the design towards the desired, sustainable direction. In addition, ethical questions can be addressed through dedicated design activities that fit the design process of the organization. In practice, for example, creating future scenarios and walking them through with workers, experts or the design team may help to identify the potential ethical impacts of the design and may guide the design activities further (see, e.g., [58]). As the system and the work practices evolve, scenarios can be re-visited to review whether new ethical questions have occurred or can be identified based on the experience of the new ways of co-working. This supports the continuous pursuit towards an ethically sustainable work community and human–machine teamwork.

4.4. Summary

In this section we have presented three complementary approaches to support the design of Industry 5.0 human–machine teams as joint cognitive systems. In Table 1, we summarize and compare the approaches regarding their background, aims, results as well as benefits and challenges. These approaches support human centricity in the early phases of the design, and thus support the more technical cognitive systems engineering approaches.

Table 1. Summary and comparison of the three approaches.

Approach	Actor–Network Theory	Concept of Operations	Ethically Aware Design
Background	Science and technology studies (STS) and sociology.	Systems engineering, especially human factors engineering (HFE).	Multidisciplinary, especially design research and social sciences.
The nature of the approach	Theory-based method.	Activity/tool in the planning and analysis phase of the HFE process.	Integrating ethical values/principles in the design process.
Main questions	What are the actors/actants of network and what roles do they have/take in the human–machine teamwork? How do humans and artifacts work together and how are they connected to each other? What kind of sociotechnical interaction and networking is happening?	What/who are the main system elements, stakeholders, tasks, and explanations that determine how the system works? What are the characteristics of the operational environment? How are responsibilities and tasks allocated? What are the main operating scenarios? What are the high-level user requirements?	What ethical aspects are relevant and should be considered in the design? What are the key stakeholder groups and their values? What kinds of ethical impacts may the adoption of technology have? How can the design be made ethically sustainable?
Study context	Ethnographic fieldwork on the shop floor.	Early phase design, re-design.	Early phase design, re-design.
Data collection	Field notes, observation, interviews, and mapping networks.	Stakeholder interviews and workshops, workshops within the design team.	Stakeholder interviews and workshops, workshops within the design team.
Research approach	Qualitative, empirical research.	Systems engineering.	Design research.
Results	Actors in the human–machine team, their roles, connections, and interactions.	Operational concept of the human–machine team, the responsibilities, and tasks of the actors.	Ethical contribution to the system design, ethical guidelines.
Form of results	A rich description with network maps.	Initial and use-case specific ConOps diagrams supplemented by user requirements tables and a high-level textual description of the system.	Ethical principles and aspects to be considered in the design, potential ethical impacts, contribution to design decisions.

Table 1. *Cont.*

Approach	Actor–Network Theory	Concept of Operations	Ethically Aware Design
Benefits	Makes complex reality visible and studies humans and non-humans equally. Describes the action of actors/actants in sociotechnical networks. At its best, shows where co-operation is smooth and where there are problems.	Illustrates complex systems from the end-users' points of view concentrating specifically on the relationships between stakeholders and the technical system and use cases.	Provides methods for identifying and considering ethical aspects in design, supports proactive ethical thinking, engaging of diverse perspectives to design and commitment towards ethically sustainable design.
Risks and challenges	Overly shallow or blurry analyses, misinterpreted purposes, and problems clarifying the responsibilities of the actors.	Laborious to draft and allocate a suitable amount of resources to its development; difficulties in convincing stakeholders of its value.	Overanalyzing potential consequences of technology adoption or work changes, excessive focus on exceptional situations or risks of misuse.

5. Discussion

Industry 5.0 visions describe sustainable, resilient, and human-centered future factories, where human operators and smart machines work in teams. These teams need to be resilient to be able to respond to changes in the environment. The resilience needs to be reflected in dynamic task allocation, and the flexibility of the team members to change roles. After a joint cognitive system has been initially designed and implemented, it continuously co-evolves as both human and machine actors learn, thus gaining new skills and capabilities. Additionally, changes in the production environment require the system to be resilient and adapt to respond to changing expectations and requirements.

We have studied the kinds of approaches and methods that could support a human-centric early design phase and the continuous development of Industry 5.0 human–machine teams on the factory floor as smoothly working, resilient and continuously evolving joint cognitive systems. It is not sufficient simply to have methods for the design of a joint cognitive system; there is also a need for methods that support the monitoring and observing of a joint cognitive system to understand how the system is currently functioning, what the current skills and capabilities of different human and machine actors are and how tasks are allocated among the actors. Moreover, there is a need for co-design methods and tools to allow relevant stakeholders to participate in designing the joint human–machine system and influence its co-evolvement. Autonomous and intelligent machines raise different ethical concerns, and thus ethical issues should already be studied at the early phases of the design and during co-evolvement.

The joint cognitive systems approach was introduced already in the 1980s and has changed the design focus from the mere interaction between humans and machines to human–technology co-agency [9]. This is important when designing human–machine teams as in these teams, human operators are not merely "using" the machines, rather, both humans and machines have their own tasks and roles, targeting common goals. Another characteristic feature of joint cognitive systems is the dynamic and flexible allocation of functions between humans and machines. The joint cognitive systems approach has been utilized in designing automated systems where the ability to maintain control of a situation, despite disrupting influences from the process itself or from the environment, is central. This requires managing the dynamics of the situation by maintaining and sharing situational awareness, and considering how the capabilities and needs depend on the situation and how they may vary over time [26]. Resilience to changes in the system itself and the environment are also central in designing human–machine teams. The design focus should be on supporting co-agency through shared situational awareness, collaboration, and communication.

Jones et al. [32] point out that cyber technologies have dramatically increased the cognitive capabilities of machines, changing machines from being reactive to self-aware.

Jones et al. identify the various actors in Industry 4.0/5.0 manufacturing systems as human, organizational and technology-based agents. Chacon et al. [33] point out how these agents are based on a synergistic combination of different technologies such as artificial intelligence (AI), the Internet of Things (IoT) and multi-agent systems (MAS). Manufacturing systems consisting of various human and machine agents can be seen as joint cognitive systems. Both Jones et al. [32] and Chacon et al. [33] propose a cognitive systems engineering (CSE) approach to designing these multi-agent-based manufacturing systems. CSE focuses on cognitive functions and analyzes how people manage complexity, understanding how artefacts are used and how people and artefacts work together to create and organize joint cognitive systems [33]. In CSE, the design focus is on the mission that the joint cognitive system will perform. The joint cognitive system works via cognitive functions such as communicating, deciding, planning, and problem-solving. These cognitive functions are supported by cognitive processes such as perceiving, analyzing, exchanging information and manipulating [33].

Cognitive systems engineering is based on various forms of modelling, such as cognitive task modelling, which is often a quite laborious process. As Grässler et al. point out [13], human actors are often greatly simplified in model-based design. Human performance modelling may then be hampered as it does not properly take into account the variability in personal characteristics, skills, and preferences. Another design challenge raised by van Diggelen et al. [21] is that traditional design methods do not sufficiently address the autonomous capabilities of machine agents, resulting in systems where the human becomes a supervisor rather than a teammate. The concept of operations method supports focusing on the teamwork in the design. The concept of operations is a high-level document providing a starting point for the modelling of a joint cognitive system by laying the basis for the requirements specification activity [40]. A ConOps can also be considered as a boundary object promoting communication and knowledge sharing among stakeholders, especially in the beginning phase of the design process. This supports the involvement of all affected workers early in the design, as suggested by Kadir and Broberg [16]. ConOps can also serve the continuous co-development of the system.

As the overall joint cognitive system is dynamically evolving over time, it becomes important to understand the entity and different actors, as well as their roles and responsibilities. For that purpose, we propose actor–network theory (ANT) [41] as an old but promising approach. Actor–network analyses require ethnographic field work with observation and other evidence-based analytical methods that aim to describe the interaction between actors and make network processes visible. The ANT approach is unbiased: the interest is for all, both of human and non-human agency. This is important because, in human–machine teams, the available capabilities should be utilized similarly regardless of which actor is providing the capability. Actor–network analyses support the understanding of how something happens, how interaction works, and which actors support sociotechnical network interactions. However, ANT does not explain social activity, the causality of action or explain the simple reasons why things happen. Despite these limitations, it can support the understanding of how the dynamic joint cognitive system works, what the current roles of different actors are, and what their relationship is with the other actors. All of this supports the further development of the overall joint cognitive system as well as developing the individual actors.

An important aspect when designing sustainable human–machine teams are ethical considerations. As Pacaux-Lemoine et al. [14] point out, techno-centered design tends to demand extreme skills from human operators. In the teams, human work should remain meaningful, and the design should not straightforwardly expect the human operators to take responsibility of all the situations where machine intelligence faces its limits. Still, the human actors should be in control and the machines should serve the humans rather than vice versa.

In this paper, we have investigated the challenge of designing resilient human–machine teams for Industry 5.0 smart manufacturing environments. We are convinced that

the long tradition of joint cognitive systems research provides a good basis also for the design of smart manufacturing systems. However, we suggest that the heavy cognitive systems engineering approach can be complemented with the concept of operations design approach, especially in the early phases of the design. The ConOps approach supports the involvement of different stakeholders in the design of the overall functionality and operation principles of the system. ConOps documents facilitate the sharing and co-developing of design ideas both for the overall system and different sub systems. The documents illustrate the allocation of responsibilities and tasks in different situations. Actor–network theory provides methods to analyze and observe a gradually evolving, dynamic team of various actors, thus understanding how the roles of the actors and the work allocation evolve over time. This supports a better understanding and the further development of the system. While traditional design methods tend to see humans as users or operators, actor–network theory sees human and machine actors equally, and thus can support analyzing actual collaborations to identify where it works well and where it could be improved. Smooth human–machine teamwork requires mutual understanding between humans and machines, and this easily leads to machines monitoring people and their behavior. The task allocation between humans and machines should be fair. Among many other issues, this raises the need to focus on ethical issues from the very beginning and throughout the design process. Ethically aware design supports proactive ethical thinking, integrating the perspectives of different worker groups and other stakeholders. Moreover, ethically aware design provides tools to commit the whole design team to ethically sustainable design. A challenge with the proposed approaches and methods is that a moderate amount of work is required from human factor experts, other design team members and factory stakeholders to efficiently apply the approaches. Another challenge is to make sure that the results actually influence the design.

Our future plans include applying the ideas presented in this paper in practice. This will provide a further understanding of the actual design challenges and how well the approaches and methods proposed here will work in practice. It will be especially interesting to study how to monitor and guide the continuous co-evolvement of a joint human–machine system.

References

1. Breque, M.; de Nul, L.; Petridis, A. *Industry 5.0: Towards a Sustainable, Human-Centric and Resilient European Industry*; European Commission, Directorate-General for Research and Innovation: Luxembourg, 2021.
2. Kagermann, H.; Wahlster, W.; Helbig, J. *Securing the Future of German Manufacturing Industry: Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0*; Final Report of the Industrie 4.0 Working Group; Secretariat of the Platform Industrie 4.0: Frankfurt, Germany, 2013.
3. Cagliano, R.; Canterino, F.; Longoni, A.; Bartezzaghi, E. The interplay between smart manufacturing technologies and work organization: The role of technological complexity. *Int. J. Oper. Prod. Manag.* **2019**, *39*, 913–934. [CrossRef]
4. Neumann, W.P.; Winkelhaus, S.; Grosse, E.H.; Glock, C.H. Industry 4.0 and the human factor—A systems framework and analysis methodology for successful development. *Int. J. Prod. Econ.* **2021**, *233*, 107992. [CrossRef]
5. Stern, H.; Becker, T. Concept and Evaluation of a Method for the Integration of Human Factors into Human-Oriented Work Design in Cyber-Physical Production Systems. *Sustainability* **2019**, *11*, 4508. [CrossRef]

6. Kaasinen, E.; Anttila, A.-H.; Heikkilä, P. New industrial work-personalised job roles, smooth human-machine teamwork and support for well-being at work. In *Human-Technology Interaction-Shaping the Future of Industrial User Interfaces*; Röcker, C., Büttner, S., Eds.; Springer Nature: Berlin/Heidelberg, Germany, 2022, *in press*.
7. Chakraborti, T.; Kambhampati, S. Algorithms for the greater good! on mental modeling and acceptable symbiosis in human-ai collaboration. *arXiv* **2018**, arXiv:1801.09854.
8. Mumford, E. *Socio-Technical Design: An Unfulfilled Promise or a Future Opportunity?* Springer: Boston, MA, USA, 2000; pp. 33–46.
9. Hollnagel, E.; Woods, D.D. *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*; CRC Press: Boca Raton, FL, USA, 2005.
10. Woods, D.D.; Hollnagel, E. *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*; CRC Press: Boca Raton, FL, USA, 2006.
11. Rasmussen, J. Merging paradigms: Decision making, management, and cognitive control. In Proceedings of the Third International NDM Conference, Aberdeen, UK, 1 September 1996; Ashgate: Farnham, UK, 1996.
12. Vicente, K.J. *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*; CRC Press: Boca Raton, FL, USA, 1999.
13. Gräßler, I.; Wiechel, D.; Roesmann, D. Integrating human factors in the model based development of cyber-physical production systems. *Procedia CIRP* **2021**, *100*, 518–523. [CrossRef]
14. Pacaux-Lemoine, M.P.; Trentesaux, D.; Rey, G.Z.; Millot, P. Designing intelligent manufacturing systems through Human-Machine Cooperation principles: A human-centered approach. *Comput. Ind. Eng.* **2017**, *111*, 581–595. [CrossRef]
15. Sgarbossa, F.; Grosse, E.H.; Neumann, W.P.; Battini, D.; Glock, C.H. Human factors in production and logistics systems of the future. *Annu. Rev. Control* **2020**, *49*, 295–305. [CrossRef]
16. Kadir, B.A.; Broberg, O. Human-centered design of work systems in the transition to industry 4.0. *Appl. Ergon.* **2021**, *92*, 103334. [CrossRef]
17. O'Neill, T.; McNeese, N.; Barron, A.; Schelble, B. Human–autonomy teaming: A review and analysis of the empirical literature. *Hum. Factors* **2020**. [CrossRef]
18. Jung, M.F.; Šabanović, S.; Eyssel, F.; Fraune, M. Robots in groups and teams. In Proceedings of the Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, Portland, OR, USA, 25 February–1 March 2017.
19. Ma, L.M.; Fong, T.; Micire, M.J.; Kim, Y.K.; Feigh, K. Human-robot teaming: Concepts and components for design. In *Field and Service Robotics*; Springer: Berlin/Heidelberg, Germany, 2018.
20. Jung, M.F.; Beane, M.; Forlizzi, J.; Murphy, R.; Vertesi, J. Robots in group context: Rethinking design, development and deployment. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017.
21. Van Diggelen, J.; Neerincx, M.; Peeters, M.; Schraagen, J.M. Developing effective and resilient human-agent teamwork using team design patterns. *IEEE Intell. Syst.* **2018**, *34*, 15–24. [CrossRef]
22. Romero, D.; Bernus, P.; Noran, O.; Stahre, J.; Berglund, Å.F. The operator 4.0: Human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems. In *IFIP Advances in Information and Communication Technology*; Springer: Cham, Switzerland, 2016; pp. 677–686.
23. Romero, D.; Stahre, J.; Wuest, T.; Noran, O.; Bernus, P.; Fast-Berglund, Å.; Gorecky, D. Towards an operator 4.0 typology: A human-centric perspective on the fourth industrial revolution technologies. In Proceedings of the International Conference on Computers and Inttrial Engineering (CIE46), Tianjin, China, 29–31 October 2016.
24. Romero, D.; Stahre, J. Towards The Resilient Operator 5.0: The Future of Work in Smart Resilient Manufacturing Systems. In Proceedings of the 54th CIRP Conference on Manufacturing Systems, Virtual Conference, 22–24 September 2021.
25. Hollnagel, E.; Woods, D.D. Cognitive systems engineering: New wine in new bottles. *Int. J. Hum.-Comput. Stud.* **1999**, *51*, 339–356. [CrossRef] [PubMed]
26. Hollnagel, E.; Woods, D.D. Cognitive systems engineering: New wine in new bottles. *Int. J. Man-Mach. Stud.* **1983**, *18*, 583–600. [CrossRef]
27. Hollnagel, E. *Handbook of Cognitive Task Design*; CRC Press: Boca Raton, FL, USA, 2003.
28. Norros, L.; Salo, L. Design of joint systems: A theoretical challenge for cognitive systems engineering. *Cogn. Technol. Work.* **2009**, *11*, 43–56. [CrossRef]
29. Järvilehto, T. The theory of the organism-environment system: I. Description of the theory. *Integr. Physiol. Behav. Sci.* **1998**, *33*, 321–334. [CrossRef] [PubMed]
30. Gibson, J.J. *The Theory of Affordances. The Ecological Approach to Visual Perception*; Houghton Mifflin: Boston, MA, USA, 1979.
31. Norros, L. *Acting under Uncertainty. The Core-Task Analysis in Ecological Study of Work*; VTT: Espoo, Finland, 2004.
32. Jones, A.T.; Romero, D.; Wuest, T. Modeling agents as joint cognitive systems in smart manufacturing systems. *Manuf. Lett.* **2018**, *17*, 6–8. [CrossRef]
33. Chacón, A.; Angulo, C.; Ponsa, P. Developing Cognitive Advisor Agents for Operators in Industry 4.0. In *New Trends in the Use of Artificial Intelligence for the Industry 4.0*; Intech Open: The Hague, The Netherlands, 2020; p. 127.
34. Elm, W.C.; Potter, S.S.; Gualtieri, J.W.; Roth, E.M.; Easter, J.R. Applied cognitive work analysis: A pragmatic methodology for designing revolutionary cognitive affordances. In *Handbook of Cognitive Task Design*; CRC Press: Boca Raton, FL, USA, 2003; pp. 357–382.

35. Hildebrandt, M.; Harrison, M. Putting time (back) into dynamic function allocation. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Denver, CO, USA, 13–17 October 2003; SAGE Publications Sage CA: Los Angeles, CA, USA, 2003.

36. Laarni, J. Descriptive modelling of team troubleshooting in nuclear domain. In Proceedings of the 10th International Conference on e-Health 2018, the 11th International Conference on ICT, Society, and Human Beings 2018 and of the 15th International Conference Web Based Communities and Social Media 2018, Part of the Multi Conference on Computer Science and Information Systems 2018, MCCSIS 2018, Madrid, Spain, 17–19 July 2018.

37. Savioja, P. *Evaluating Systems Usability in Complex Work-Development of a Systemic Usability Concept to Benefit Control Room Design*; Aalto University: Espoo, Finland, 2014.

38. Rasmussen, J. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Trans. Syst. Man Cybern.* **1983**, *3*, 257–266. [CrossRef]

39. Hugo, J.V.; Gertman, D.I. Development of Operational Concepts for Advanced SMRs: The Role of Cognitive Systems Engineering. In *Small Modular Reactors Symposium*; American Society of Mechanical Engineers: New York, NY, USA, 2014.

40. Fairley, R.E.; Thayer, R.H. The concept of operations: The bridge from operational requirements to technical specifications. *Ann. Softw. Eng.* **1997**, *3*, 417–432. [CrossRef]

41. Callon, M. The sociology of an actor-network: The case of the electric vehicle. In *Mapping the Dynamics of Science and Technology*; Springer: Berlin/Heidelberg, Germany, 1986; pp. 19–34.

42. Callon, M. The role of hybrid communities and socio-technical arrangements in the participatory design. *J. Cent. Inf. Stud.* **2004**, *5*, 3–10.

43. Niemelä, M.; Kaasinen, E.; Ikonen, V. Ethics by design-an experience-based proposal for introducing ethics to R&D of emerging ICTs. In *ETHICOMP 2014-Liberty and Security in an Age of ICTs*; CERNA: Paris, France, 2014.

44. Goffman, E. *Interaction Ritual (New Brunswick, NJ)*; Transaction Publishers: Piscataway, HJ, USA, 1967; p. 3.

45. Latour, B.; Woolgar, S. *Laboratory Life*; Princeton University Press: Princeton, NJ, USA, 2013.

46. Lutz, C.; Tamò, A. Privacy and Healthcare Robots—An Ant Analysis. In *We Robot Conference Proceedings*; University of Miami: Miami, FL, USA, 2016.

47. Shiga, J. Translations: Artifacts from an actor-network perspective. *Artifact J. Des. Pract.* **2007**, *1*, 40–55. [CrossRef]

48. Law, J. Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Syst. Pract.* **1992**, *5*, 379–393. [CrossRef]

49. Tommila, T.; Laarni, J.; Savioja, P. Concept of Operations (ConOps) in the Design of Nuclear Power Plant Instrumentation & Control Systems. In Proceedings of the 31st European Conference on Cognitive Ergonomics, Toulouse, France, 26 August 2013; A working report of the SAREMAN project.

50. Väätänen, A.; Laarni, J.; Höyhtyä, M. Development of a Concept of Operations for Autonomous Systems. In Proceedings of the International Conference on Applied Human Factors and Ergonomics, Washington, DC, USA, 24–28 July 2019; Springer: Berlin/Heidelberg, Germany, 2019.

51. Laarni, J.; Koskinen, H.; Väätänen, A. Concept of operations development for autonomous and semi-autonomous swarm of robotic vehicles. In Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017.

52. Mostashari, A.; McComb, S.A.; Kennedy, D.M.; Cloutier, R.; Korfiatis, P. Developing a stakeholder-assisted agile CONOPS development process. *Syst. Eng.* **2012**, *15*, 1–13. [CrossRef]

53. Friedman, B.; Kahn, P.H.; Borning, A.; Huldtgren, A. Value Sensitive Design and Information Systems. In *Early Engagement and New Technologies: Opening Up the Laboratory*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 55–95.

54. Wright, D. A framework for the ethical impact assessment of information technology. *Ethics Inf. Technol.* **2011**, *13*, 199–226. [CrossRef]

55. Ikonen, V.; Kaasinen, E.; Niemelä, M. Defining Ethical Guidelines for Ambient Intelligence Applications on a Mobile Phone. In *Intelligent Environments (Workshops)*; IOS Press: Amsterdam, The Netherlands, 2009.

56. Kaasinen, E.; Liinasuo, M.; Schmalfuß, F.; Koskinen, H.; Aromaa, S.; Heikkilä, P.; Honka, A.; Mach, S.; Malm, T. A worker-centric design and evaluation framework for operator 4.0 solutions that support work well-being. In Proceedings of the IFIP Working Conference on Human Work Interaction Design, Espoo, Finland, 20–21 August 2018; Springer: Berlin/Heidelberg, Germany, 2018.

57. Heikkilä, P.; Honka, A.; Kaasinen, E.; Väänänen, K. Quantified Factory Worker: Field Study of a Web Application Supporting Work Well-Being and Productivity. *Cogn. Technol. Work* **2021**, *23*, 831–846. [CrossRef]

58. Heikkilä, P.; Honka, A.; Mach, S.; Schmalfuß, F.; Kaasinen, E.; Väänänen, K. Quantified factory worker-expert evaluation and ethical considerations of wearable self-tracking devices. In Proceedings of the 22nd International Academic Mindtrek Conference, Tampere, Finland, 10–11 October 2018.