Cover sheet

Conference: 23rd International Command and Control Research and Technology Symposium (ICCRTS)

Date: 6-9 November 2018, Pensacola, Florida, USA

Track: 3: Battlefields of the Future and the Internet of Intelligent Things

Title: How to Cooperate with Intelligent Machines: Lessons for Defence Operations from the Integration of AI and Robotics across Multiple Domains

Keywords: artificial intelligence, robotics, human-machine collaboration

Authors: Tim Lamballais Tessensohn, tim.lamballaistessensohn@tno.nl
Bob van der Vecht, bob.vandervecht@tno.nl
Aletta R. Eikelboom, aletta.eikelboom@tno.nl

Affiliation: TNO, The Netherlands

Address: PO box 96864, 2509 JG, Den Haag, The Netherlands

How to Cooperate with Intelligent Machines: Lessons for Defence Operations from the Integration of AI and Robotics across Multiple Domains

T. Lamballais Tessensohn, B. van der Vecht, and A. R. Eikelboom

TNO, the Netherlands

Abstract

In the early years of Artificial Intelligence (AI) research, scientists envisioned AI as embodied in a robot. Since then, the fields of AI and robotics diverged and mostly developed separately. Thanks to rapid developments and recent scientific breakthroughs in both fields, there are many upcoming applications where the integration of AI and robotics leads to great added value.

In this article, we investigate the opportunities of integrated AI and robotics for the Defence domain. Based on recent scientific developments, we introduce the concept of a Non-human Intelligent Collaborator (NIC), an intelligent artificial team member with social and collaborative skills.

We explore new applications of integrated AI and robotics across multiple domains, from agriculture to warehousing. The domains reveal a variety of problem characteristics, offering valuable lessons from the solutions presented, which can be translated to the Defence domain. We focus on the behavioural impact of deploying NIC-human teams on conducting future operations.

1. Introduction

Artificial Intelligence (AI) has historically seen periods of rapid progress, accompanied by exaggerated expectations and hype. These periods of progress were followed by "*AI winters*", long periods of stagnation, that disappointed those who had believed the grand claims of AI possibly achieving the end goal - reaching the same intelligence level as human beings. Currently, we are witnessing another period of great progress. AI has managed to beat humans at Jeopardy, is capable of recognizing objects in images with a greater accuracy than humans, has beaten humans at the incredible complex game of Go, can provide high quality translations, can play video games, recognize voices, and generate near-perfect fake images. Surprising and impressive feats occur regularly.

This current progress of AI also influences the Defence organisation. Unmanned autonomous systems are increasingly used in military operations, e.g., (De Spiegeleire, Maas, & Sweijs, 2017; Endsley, 2015), and AI developments promise to enhance the intelligence of these systems, allowing a diverse use beyond the dull, dangerous, and dirty tasks. It is therefore expected that the role of unmanned (intelligent) systems will shift from support tool to team mate and will be fully integrated in essential operational processes and task execution. This shift affects future concepts of operations and collaboration between soldiers and unmanned (intelligent) systems. The Defence organisation needs to gain knowledge on the potential psychological, social, and organisational impact of their deployment. For example, Explosive Ordnance Disposal (EOD) has already been working for more than a decade with unmanned systems. Some teams have become so emotionally attached to their unmanned team buddy that they hesitate to let it execute dangerous tasks (Hsu, 2018). This is but one example of a psychological effect of human-unmanned system teaming.

We believe that future operational effectiveness is highly influenced by the way artificial intelligent systems are incorporated in the Defence organization. Such systems would include robots or software agents with social capabilities that are to some extend comparable to human team members. To scope the research, we introduce the concept of a Non-human Intelligent Collaborator (NIC). The aim of this article is to investigate the consequences of deploying NICs for (1) the soldiers that collaborate with them in future concepts of operations, and (2) the Defence organisation, and to identify open issues that may serve as input for future research. First, we describe recent developments in AI and robotics that may lead to the development of NICs. We continue by establishing the concept of a NIC. This is followed by an overview of the state of the art by examples of applications with NICs in the international Defence domain and in other domains such as manufacturing. From these we conclude with possibilities for Defence and provide further research topics to cope with the psychological and social impact of deploying NICs on the humans working with NICs.

2. Trends in AI and Robotics

In recent years, developments in AI research have made the discussion about artificial systems as collaborators in a team with humans more relevant. Collaboration is no longer purely the domain of science fiction scenarios, but becomes relevant when one must prepare for scenarios of fifteen, twenty years from now, which may be necessary in Military research and development (R&D) and procurement. Some aspects of human-NIC collaboration can already be found today. We first provide our vision on AI, follow up by exploring recent AI trends in more detail, before concluding this chapter with our vision on human-machine teaming.

Vision on Artificial Intelligence

The alteration of periods of AI progress and AI periods of stagnation makes it difficult to predict what the effect of current AI techniques will be on society. Historically, when a technology advances, we have typically tended to overestimate the effect of the technology in the short term and underestimate the effect in the long run (Brooks, 2017).

With AI surpassing humans in a growing number of areas, it is natural to ask: Will this time be different? Will the current period of rapid progress see AI achieve General Human Intelligence (GHI)? Or will we witness another period of stagnation, pricking the bubble of society's grand expectations once more? The opinions on this vary. Some visionary reports claim that GHI will be reached within 30 years from now (De Spiegeleire, Maas, & Sweijs, 2017), whereas others are more sceptical (Brooks, 2017).

On the one hand, one can make the point that this time will be the same, and stagnation is inevitable. The parallels of AI's achievements in chess in the 90s with the recent achievements in Go (Silver, et al., 2017) are striking: now, as then, AI beat humankind at a complex game that humans had thought was at a high intellectual level, so high it was out of reach for computers and so complex it was thought to require uniquely human reasoning powers. Now, as then, the breakthrough leading to this achievement was more a matter of high computation power, than of an intellectual breakthrough.

In the case of the game Go, the breakthrough was the transition from neural networks to "*Deep Learning*", which simply refers to neural networks with many more layers than before. Although there was scientific progress, such as the step towards convolutional neural networks, or the invention of Generative Adversarial Networks, it was mainly the step from Central Processing Unit (CPU) to

Graphics Processing Unit (GPU) computations, accompanied by specialized, purpose-built hardware and the availability of large datasets, that allows the current breakthrough of neural networks.

We tend to overestimate the effect of AI breakthroughs for several reasons. When we see an exponential trend, we tend to extrapolate it to the future, without realizing the exponential growth may end for technical or economic reasons (Brooks, 2017). The progress in deep learning, for example, has depended on the (exponential) growth of data, but at some point in the near future it may no longer make economic sense to keep storing so much data[1].

Another reason we overestimate, is that buzzwords and catchphrases, i.e., *"Deep Learning"*, obfuscate the technical details (Brooks, 2017). The obfuscation gives the impression that the scope of the breakthrough is far wider than it is. For someone unfamiliar with the techniques, the results of the breakthrough may *"seem like magic"*. Moreover, people are used to infer skills and competences from performance and results (Brooks, 2017). When an AI makes a breakthrough, we tend to assume, implicitly, that the AI has certain competences, and hence should be able to carry out other tasks that require the same competences, but this is not the case with AI. AI has learned to play old games like *"space invaders"*, but why are modern games out of reach? IBM's Watson has beaten Jeopardy, so it may seem reasonable to assume that Watson has the competence of understanding the context of the questions posed, which is simply not the case.

On the other hand, one can also make the point that this time will be different, and there will be no stagnation. First, it is not merely the field of AI that is making leaps: the field of robotics is progressing with leaps and bounds as well, as components for robotics have become significantly and exponentially cheaper over the past decades. The two trends reinforce each other. Robotics can shift from controlled indoor settings such as factories and warehouses, to outdoor settings such as agriculture, due, among others, to advances in image recognition. Robotics in turn gives rise to new kinds of *"embodied"* AI that learn from real-time interaction with the environment. Secondly, open source software such as the deep learning frameworks of Google and Facebook, and the Robot Operating System (ROS), allow anyone with a computer to experiment with these advances in AI and (to some extent) in robotics. The more people become acquainted with these techniques and experiments, the more novel applications may be found, which creates and sustains momentum for developing new techniques in both fields.

We tend to underestimate the effects of a technological breakthrough in the long run, because we tend to predict the effects of that one technological breakthrough in isolation (Brooks, 2017). The long-term effects, however, are reinforced by advances in other fields and in turn reinforce future advances. This may be especially true for AI and robotics, because the speed of deployment is starkly different. New AI software and frameworks can be deployed nearly instantly after they have been coded, whereas hardware such as robotics tends to be long term investments with renewal cycles spanning years or even decades. The different deployment speeds facilitate our tendency to predict the effects of the AI breakthroughs in isolation.

Even if developments in the field of AI itself stall, integrating existing AI techniques with robotics is likely to prove fruitful, and may in the near future lead to the precursors of NICs. One of the earliest goals of AI was to build embodied intelligent systems but the fields of AI and robotics have evolved mostly separately. In recent years, in many domains, breakthroughs have been possible due to the

[1] For example, in the healthcare industry data storage requirements have been growing exponentially, and it is not sure that this pace can be sustained in the long run. See for example http://www.nextech.com/blog/healthcare-data-growth-an-exponential-problem

integration of AI and robotics. These include drones and autonomous tractors involved in precision agriculture, space robots autonomously exploring other planets in harsh conditions and conducting experiments, robots in warehouses and container terminals navigating their surroundings, autonomous cars in urban areas, drones and robots employed for search and rescue missions together with human operators, cyber-physical systems in industry, a wide range of healthcare applications, and, potentially, many different innovations with regards to smart cities. In other words, practical applications abound. However, has the integration of AI and robotics brought us further along the path the GHI?

The above discussions about AI developments are present in public debate about societal impact of AI, where prominent researchers raise concerns about unwanted consequences of future AI. For example, the next step after GHI would be Artificial Superintelligence (ASI): machine intelligence that exceeds human intelligence across any task, which could lead to the creation of uncontrollable systems that would threaten humanity (Future of Life Institute, 2015).

Practical outcomes of this debate are calls for more attention to ethics in AI research. The discussion on meaningful human control addresses a growing concern about system autonomy potentially leading to conflicts with humanitarian law and ethics (Lin, Bekey, & Abney, 2009). Topics of interest are the balance between human control and system autonomy, and accountability and explainability of behaviour of autonomous systems. These aspects strongly relate to proper interaction and collaboration between humans and autonomous systems. Therefore, both the technical developments and the social and ethical developments in AI are relevant for the operational deployment of unmanned (intelligent) systems in the military context.

Recent Trends in AI: Deep Learning and Explainable AI

The increased possibilities and use of deep learning contributes to the demand for explainable AI. The Defense Advanced Research Projects Agency (DARPA) distinguishes three waves of AI (DARPA, 2017): *"Describe"* (handcrafted knowledge), *"Categorize"* (statistical learning), *"Explaining"* (contextual adaptation). The biggest breakthrough of recent years has been deep learning, which falls under the second wave of AI, statistical learning. This technique is so advanced that it has become impossible for a human to reconstruct the outcome, given the input variables. Therefore, the breakthrough in deep learning created momentum for further developing *"explainable AI"*, which refers to the capability of explaining the results of, among others, deep learning. We will mention current development and provide predictions and challenges for deep learning in combination with robotics.

Deep learning is based on Deep Neural Networks (DNN). For the coming 5 to 20 years, deep learning is poised to accelerate the invention of intelligent autonomous systems by offering the robotics community a new toolkit for overcoming numerous challenges. DNNs are used in robotics, four common DNN structures are: (1) function approximating models, (2) autoencoders, (3) recurrent networks, and (4) policy models (Pierson & Gashler, 2017). These DNN structures have different characteristics and are used for different types of problems or problem domains. Together, these DNN structures have the potential to enable robots to learn novel dynamics and movements, to learn control policies in dynamic environments, to acquire advanced object recognition, to learn how to grasp and manipulate physical objects, to fuse information from many sensors into compact representations of the environment, to interpret and anticipate human actions, and to assist in high-level task planning (Pierson & Gashler, 2017).

However, explainability of DNNs is a drawback. Even when the structure of a DNN and all parameters are known, it typically is nearly impossible to predict how a DNN will transform the input into an output. Therefore, deep learning effectively operates as a black box (Samek, Wiegand, & Müller, 2017). Yet, as society finds more applications for deep learning and it becomes more ingrained in the systems that society uses, demands grow for explanations of how DNN reached decisions, verdicts and conclusions.

Recently, DARPA started the "*explainable AI*" project (Gunning, 2017), which focusses in particular on explaining (the results of) deep learning. The field of explainable AI that is emerging as a result focuses mainly on neural networks, as neural networks has created by far the most applications for which explainability has not been explored yet. Explainability is not only useful for neural networks, but also for other statistical learning techniques such as Support Vector Machines (SVM)s, or (sub)symbolic approaches to AI (Samek, Wiegand, & Müller, 2017). Explainability is a desirable characteristic of a model as it allows the discovery of errors and flaws in the model, and of biases in the data. Explainability also allows the user to verify that the model works as intended and complies with the law. Moreover, explainability may give practitioners information relevant to improving the system. Therewith, explainability has strong links with usability of a system and human factors, such as trust (Samek, Wiegand, & Müller, 2017). It is a relevant aspect in the context of NICs as collaborative team members.

Trends in Human-Machine Collaboration

Machines are increasingly designed to be more autonomous, which changes the way humans work with or use them. This raises some questions: how to design machines to be good collaborators? What is required to optimize human machine collaboration?

The relevance of collaboration in human-machine teaming has become evident in recent years, as autonomous systems are becoming more integrated in daily life, of which autonomous vehicles are one example. The complexity of interaction between human and machine is bound to increase further in the coming years. As a result, the dominant type of interaction between human and machine, namely the human directing and fully controlling the machine, may steadily lose ground to a different type of interaction, namely human and machine cooperating interdependently on a more equal footing as with less human control. An interdependency is broadly speaking a set of complementary skills, resources or knowledge, across multiple parties, where the whole set is needed in a joint activity, see also Johnson M., et al. (2014). Interdependency does not necessarily mean that the machine operates independently, although it also does not exclude it.

The topic of balancing human control and system autonomy is addressed in research on human-machine collaboration. However, as argued by Klein, et al. (2004), it is a long way until artificial entities and systems will be equal team members to human beings. The authors describe ten challenges that need to be solved to reach this, which are based upon requirements for joint activity among people. For example, mutually predictable and directable. Johnson, et al. (2014) and Klein, et al. (2004) state that the crucial capabilities of autonomous systems to function as a team member can be summarized as *Observability*, *Predictability*, and *Directability*. Observability means making pertinent aspects of one's status, as well as one's knowledge of the team, task, and environment observable to others. Predictability means one's actions should be predictable enough that others can reasonably rely on them when considering their own actions. The complementary relationship is considering others' actions when developing one's own. Directability means one's ability to direct the behaviour of others and be directed by others. Directability includes explicit commands such as task allocation and role assignment as well as subtler influences, such as providing guidance or suggestions.

The idea of *"Coactive design "* (Johnson, et al., 2014) is to use the concepts of observability, predictability, and directability when designing human-robot systems to accommodate interdependency between human and robotic team members. A dynamic and adaptive function allocation is necessary for the coordination of activity in order to flexibly and resiliently manage the interdependencies. Software frameworks are proposed to facilitate collaboration in human-machine teams. Example are: KAoS[2], (Uszok, et al., 2003) which provides flexible policy management for task coordination; and, Social Artificial Intelligence Layer (SAIL) (van der Vecht, et al., 2018), which creates a software layer with social teaming functions between human and autonomous team members. Recently, Mioch, et al. (2018) have experimented with adaptive human-robot team behaviour using the principles of work agreements in a urban search and rescue context.

Interdependency across hybrid teams of humans and robots requires the humans to trust the robots sufficiently (Lohani, et al., 2017). Trust is facilitated when the robots also have non-technical skills, such as various communication skills, to build rapport. Interdependency across hybrid teams also requires communication channels and clear signals across such channels. Evans, et al. (2017) discuss the importance of gestures, eye gaze, language, and other factors that the human(s) can use to direct the robots. In addition, following eye gaze may enable a robot and human to develop a shared mental model or joint attention model, which increases their observability towards each other.

The DARPA Robotic Challenge in the years 2012 to 2015 offered an opportunity to test the performance of the contemporary state of the art human-machine teams. Johnson, et al. (2014) describe the main lessons they learned at this challenge. Their first lesson is that maximizing several types of performance should have a higher priority than maximizing the autonomy of the robotic system. Sometimes tele-operation rather than autonomous behaviour promises the best chance for a mission's success. Also, a trade-off exists between development effort and autonomy. Creating an 80% solution where the operators are actively involved, can therefore be preferable to a 100% solution with little operator involvement. To build in resilience, there should however, always be multiple ways to success. In addition, rather than having a human in the loop or on the loop, it is better to assess where and how humans and machines best complement each other, so they can perform team work efficiently. Enabling efficient human-machine teamwork in turn requires that the algorithms and the user interfaces evolve in lockstep.

State of the art human-machine teams do not take a machine with a "*personality*" into account. However, scientists have pursued the possibility of endowing AI with personality. This begs the question: would it facilitate collaboration, would it be superfluous, or would it prove detrimental? And if it facilitates collaboration, will this also be the case in a military setting?

Existing examples are virtual assistants, and chatbots . Such entities are no longer merely tools or equipment, as computers are, but since they interact with humans while taking (to some extent) the context of the interactions into consideration, they have become, for all appearances and for certain purposes, collaborators. For example, van den Bosch et al. (2012) provide a framework for developing virtual characters and show that their test subjects could correctly classify the personality traits of the virtual characters they had interacted with.

As Kim, et al. (2017) show, shame and embarrassment may play a role in efficient teamwork. People who believed the IQ to be fixed and unamenable to change, were more reluctant to ask a human-like AI for assistance than a non-humanised AI, whereas people who believed the IQ to be flexible were not. It seems that the former were afraid to *"lose face"* to their human-like AI assistants. More directly focused on teamwork is the work by Hanna and Richards (2015), which studies the development of a shared mental model across team members. The *"Intelligent Virtual Agents"* (IVA) are either introvert or extravert and are also either agreeable or disagreeable. They show that humans prefer the

[2] http://ontology.ihmc.us/kaos.html

personality of an IVA to be consistent across both verbal and non-verbal communication, and to complement rather than resemble their own personality. The extraversion of an IVA did not seem to influence teamwork or the development of a shared mental model, but the agreeableness of an IVA did, with more agreeable IVAs leading to more strongly developed shared mental models. However, since personality types influence collaboration in a relatively simple setting, it seems plausible that personality could also influence collaboration in more complex settings, such as collaboration between agents and humans. As Ahrndt, et al. (2014) argue, giving AI a personality may improve collaboration between a human and an AI by making the decision making of the AI more predictable to the human.

3. The NIC Concept

Chapter 2 showed several trends on the intersection of AI and robotics. The future will bring a broad range of semi-intelligent, semi-autonomous systems with hugely varying skillsets and capabilities into many areas of society. This chapter discusses what type of systems fall within that scope. To do so, however, we first clarify one element that has been missing so far: autonomy.

Autonomy

The concept of autonomy is important for NICs, as they are assumed to have capabilities for performing their tasks independently or interdependently and to have capabilities for reasoning and interaction that are needed for collaboration. The term "*autonomy*", however, needs more clarification as it may be used in multiple ways.

Autonomy in relation to robotics is sometimes conflated with automation. An autonomous system, then, "*performs its actions without human intervention"*. It can be fully pre-programmed and may have no choices about its action execution. However, from a more philosophical perspective, this interpretation of autonomy as automation is debatable. Influential philosophers (e.g., Kant), start from the position that an autonomous entity can be held accountable for its actions. Autonomy, then, is related to the notion of free will.

Based on this more fundamental perspective on autonomy, AI researchers have imposed requirements on autonomous systems regarding their internal reasoning process and decision-making process (Castelfranchi, 1994; van der Vecht, 2009; Endsley, 2015). Furthermore, an autonomous system is not necessarily independent, which was also briefly mentioned above in our discussion of interdependency. It may allow external influences (e.g., human guidance), as long as it explicitly accepts these influences. This notion is important in the context of NICs, as it combines social and collaborative capabilities in autonomous systems.

NICs and Semi-NICs

These elements of autonomy also return in our definition of a NIC. A NIC is defined as: "*an artificial entity which is able to autonomously engage with its environment in direct interaction, involvement and/or interdependency with humans and other artificial entities in order to meet a certain objective. Besides deciding and acting on an individual basis, both the human and the artificial entity complement each other's decision-making process and actions. To do so, they must be able to understand complex ideas (relative to the activity), to adapt effectively to the environment, and to combine task related- with social and team related skills that enable effective and efficient collaboration*".

This definition leaves several dimensions open along which NICs can vary. A NIC can vary in its appearance, in the type of tasks it is developed for, and in the manner of its interaction with humans. However, not all non-human intelligent beings or machines that collaborate with humans fall in the category of NIC according to our definition. We will clarify our definition of a NIC by explaining the differences between NICs and non-human intelligent collaborators that fall outside of our definition. We label the latter "*semi-NICs*".

A NIC has the followings skills and abilities that a semi-NIC does not. A NIC can take the team and the environment into account when taking decisions, with the purpose of reaching a goal it understands and shares with human team members. The NIC can adapt its task(s) when the environment changes, can deal with noise, and can explain the motivation behind its actions and decisions. A NIC can cooperate closely with human team members and may even advise or lead them. This list is not exhaustive, but gives an indication of the criteria for being a NIC rather than a semi-NIC. Examples of operational NICs in future military operations could be a robot swarm that enhances a commanders situational awareness and supports decision making, or a UAV system that autonomously executes logistic operations such as munition distribution for military in the field. Benda, et al., (2018) describes three future operational scenarios in which a military team collaborates with NICs.

For further clarification, we introduce several non-human collaborators that are potential candidates for falling in the NIC category, but are semi-NICs nonetheless. Certain species of animals, for example, could arguably be categorized as a NIC. Animals are non-human and intelligent, and there are numerous examples of animals and humans cooperating in the Defence, Security and Safety domain. Rats have been trained to hunt for land mines, the Dutch police has employed hawks to attack drones, and dogs have assisted the police and airport security for decades. However, animals are not NICs but semi-NICs, because they do not understand the goal of their human team members and cannot explain the motivation behind their actions to their human team members.

Another possible NIC candidate would be advanced *"smart"* technology and equipment, such as a *"mule"*. A *"mule"* is a transport robot consisting of a rump with four legs that can carry inventory. It resembles the animal mule and, human owners / collaborators tend to develop an emotional attachment. However, they are not NICs but semi-NICs, because they cannot explain the motivation behind their actions, nor can they adapt their tasks when their environment changes.

Another potential NIC candidate would be a smart exo-skeleton. The communication between the human and its exo-skeleton is intense and sophisticated. However, a smart exo-skeleton does not autonomously engage with its environment, but require human involvement in order to function.

As these examples suggest, the current state of technology is not at a level where we are capable of creating NICs. At present, precursors to NICs do exist in some domains, as the next chapter shows.

4. Lessons Learned from Use Cases in Other Domains

This section explores the lessons that can be learned from the development in AI and robotics outside of the Defence and Security domain. We examine four use cases: (1) autonomous cars, (2) search and rescue, (3) care for children and the elderly, and (4) Industry 4.0 and Agriculture.

The subsections below each discuss a use case separately in depth. As Schwendner and Kirchner (2014) mention, "*robots [are] the manifestation of AI in the physical world*". Therefore, we opted to use the terms AI, robot(ic)s and (semi-)NICs interchangeably in the descriptions of the use cases. We studied three factors: (1) the motivation for employing AI and robotics, (2) the intelligence required from the semi-NICs, and (3) the way humans and semi-NICs collaborate. For each factor, we focuses on several aspect. Table 1 below provides an overview of these factors including the several aspects and foreshadows the use cases to these factors. The factors "*intelligence"* and "*collaboration*" were chosen as they are an essential part of the NIC. The factor ‘motivation’ is added as this factor puts the use case into perspective and is relevant to draw conclusions for the Defence and Security domain.

Table 1: overview of use case characteristics

	Autonomous cars	Search and rescue	Care for children and the elderly	Industry 4.0 and Agriculture
Motivation				
Cost / time savings			X	X
New functionality			X	X
Higher quality output				X
Avoiding risk / danger		X		
Speed / adaptability		X	X	X
Faster communication		X		X
Gaining new knowledge	X			
Intelligence				
Forecasting				X
Self-configuration				X
Runtime adaptive	X	X	X	X
Decision making	X			X
Continual monitoring	X	X	X	X
Intent recognition		X	X	
Collaboration				
Remote control		X		
Machine-to-machine				X
Fleet of robots		X		X
Separate virtual agent(s)			X	
Value tension	X			
Central computer				X

The first factor, the motivation, for most use cases included multiple aspects, except for autonomous cars. Interestingly, the main motivation for pursuing the development of autonomous cars currently seems to be to gain new knowledge, or rather: we invent them because we can. Since drivers still need to sit behind the wheel in case of emergencies, and autonomous cars are designed to behave similarly to human-driven cars, there do not appear to be time or cost savings. Perhaps in the future this will change. For the other use cases, the motivations are a mix of cost / time savings, new functionality, higher quality, lower risk, improved adaptability or faster communication.

As for the second factor, the intelligence required from the AI and robots, we focused on the capabilities that the literature explicitly mentioned. For example, although the ability to predict the near future is probably relevant in multiple use cases, it was only explicitly mentioned and emphasized in the Industry 4.0 and Agriculture use case. Besides this capability, other capabilities that appeared to be important are: the ability of a NIC to (re)configure and finetune itself, for example by swapping components or changing internal parameters in its software, the ability to adapt at runtime to unpredicted, unforeseen events and circumstances, and the ability to make decisions and to choose the best option from the ones available. The ability to continually monitor the environment and have continuous situational awareness was emphasized in all use cases, and the ability to recognize the intention of human operation is especially important both for search and rescue missions and for healthcare applications.

The third factor, the collaboration between robots and humans, differs strongly across use cases. Remote control of robots by human operators is especially important in search and rescue missions, due to a strong risk aversion in this domain. However, in industry 4.0 or agriculture, instead of human operators, a central computer tends to orchestrate the fleet(s) of robots, if central coordination is present at all. The robots can also collaborate in a decentralized manner with humans, or with each other in a machine-to-machine manner. In addition to these methods of collaboration, it is also possible that virtual agents, that exist separately from the robots, facilitate the collaboration between human(s) and machine(s). Finally, the last noteworthy element of collaboration is so-called *"value tension"*, which means that different values and priorities of human users clash with regard to pre-programmed behaviour of the autonomous system.

Autonomous Cars

DARPA held several grand challenges for autonomous driving, culminating in the DARPA Urban Challenge in 2007. Teams from across the world competed and one team succeeded in navigating the urban environment. Since then the technology has advanced at a rapid pace (Berger & Rumpe, 2014).

Autonomous cars are an interesting topic from a legal and ethical perspective. As their behaviour is coded at design time / production time, but they take decisions at runtime, the question of responsibility in the case of accidents is nontrivial. Moreover, ethical dilemmas may arise during driving, but the decision the car makes must be explicitly or implicitly coded at design or production time. Manufacturers have to hard-code answers to ethically dilemmas into the behaviour of the car, meaning that the values used to weigh different options must be made explicit, and choices have to be made for ethical dilemmas that are essentially unsolvable in general. Users may therefore drive in cars that may make different ethical decisions than the users would make themselves, . This leads to so-called *"value tension"*, especially since customers may not agree with the built-in values or choices of their car.

In general, the public has a positive opinion of autonomous cars, and high expectations of the benefits, while also worrying about the general driving safety and the cybersecurity of the vehicles (Schoettle & Sivak, 2014).

Nevertheless, besides the value tension aspect, autonomous cars seem to offer surprisingly few benefits or lessons for the Defence domain. Autonomous cars are engineered to behave similarly to human drivers, offering few novel features or behaviour, except that autonomous cars may drive more cautiously and with more respect for traffic rules than human drivers. The main reason that autonomous cars need to behave similarly to human drivers, is because other drivers need to be able to predict the behaviour of the car and cannot see from the outside that the car is autonomous. For example, an autonomous car may be able to brake much more quickly and effectively than a human driver, meaning that an autonomous car could drive closer to the car in front of it than a human driven car could safely do. However, if the autonomous car would drive closely behind the car in front of it, the human driver in that car in the front may consider this a dangerous situation since the human driver cannot see whether the car behind him is autonomous. The interaction with drivers is very limited, with a human mainly present as an emergency fall-back. The intelligence in the software seems to be mostly advanced image recognition combined with hard-coded decision trees. That said, new driving behaviour that is specific for autonomous cars may be developed in future, as sensor and communication capabilities allow to anticipate earlier. Furthermore, new ways of cooperation and interaction with humans may arise in new concepts of use, such as collaborative driving, car sharing and possibilities for traffic control.

Search and Rescue (Civilian)

Search and rescue provides interesting use cases for AI and robots. The environments may contain collapsed buildings and landslides, may suffer from severe weather conditions, and be subject to flooding or radiation. Moreover, a response time is crucial for finding survivors. A heterogenous team of robots is desirable in such circumstances, e.g., drones can perform reconnaissance, while other robots can remove debris, prop up structures, and transport survivors (Ventura & Lima, 2012).

Despite the obvious benefits of deploying robots in such hazardous circumstances, there are compelling reasons for limiting the use of robots (Murphy, 2014). Due to legal issues and the fact that human lives hang in the balance, operators tend to be suspicious of their robots and do not trust them. Crucial decisions are therefore always made by humans on the team. If the operators do not find it easy and intuitive to deploy the robots, they will not use them. Moreover, due to the lack of trust, operators want to be in frequent contact with the robots, which is not always possible, for example due to severe weather conditions. This lack of trust is exacerbated by the fact that, due to rapid progress in the robotics field, training and expectations of operators tend be outdated. Hence operators may not know how to operate their robots effectively and may not know the (full extent of the) capabilities of their robots. Acceptability also plays a strong role here: if human operators do not accept robot autonomy, for example because they cannot understand or predict a robot's actions, then despite the possible benefits of autonomous robot behaviour, the operators will revert to teleoperation.

The key concepts for robots in a search and rescue setting are transparency and adjustable / adaptive autonomy (Ventura and Lima, 2012 and Kruijf, et al., 2014). Transparency makes it easier for operators to accept the robots and their behaviour. Operators work under stressful, rapidly changing conditions and will revert to tele-operation if the abilities, behaviour or possible achievements of an autonomously functioning robot are not transparent. How a robot adapts its behaviour to continually changing circumstances needs to be clear as well. One way to improve transparency is for the robot(s) to understand the spoken dialogue of the operators, including nuances in tone, and in turn for the robots to produce contextually appropriate verbal feedback for the operator (Kruijff, et al., 2014).

Adjustable autonomy means that, while the robot is still supervised and (tele)operated by a human, for certain tasks the autonomy of the robot increases according to the context of the situation. Such tasks can include stair climbing, tether docking, aerial navigation and tracking of land robots by aerial robots (Ventura & Lima, 2012). Another example of adjustable autonomy is *"implicit robot selection"* (Cacace, Finzi, & Lippiello, 2016): if an operator controls multiple, similar robots, it may be difficult to distinguish the individual robots accurately, especially in adverse weather conditions. Therefore, when an operator issues a command, it is beneficial if the groups of robots can accurately estimate which robot the operator meant to give the command to.

However, to improve search and rescue operations, it is not sufficient to focus solely on the operators and robots. Other stakeholders need to be prepared as well. Currently, local government officials are not sufficiently aware of the robots and their capabilities, and robotics companies tend not to be prepared for joining search and rescue operations. Data about the local environment may be insufficient. Success depends not only on the team of humans and NICs, but also on the support around that team (Murphy, 2014).

Care for Children and the Elderly

Robotics in healthcare take various forms and provide various services. Some provide therapy and counselling, such as the animal type robots that have proven popular for alleviating feelings of loneliness. They can even provide so-called *"animal-type"* therapy (Ahn, et al., 2017). Chatbot counsellors can prove surprisingly effective for therapy as well. In addition, another experiment showed that an autonomous virtual avatar on a smartphone that provides personalized positive psychological interventions to users had an immediate positive effect on the patients' wellbeing (Jeong, et al., 2017). An advanced form of interaction between human and robot is achieved with co-learning, where both actors simultaneously adapt their behaviour to one another, in order to reach a common training goal (Saunders, et al., 2013).

Robots can also provide other services. Surgical robots assist doctors and may be used remotely, while other types of robots can safely lift and transport patients, remind patients to take medication, measure vital signs or simply provide a terminal from which to launch skype to talk to family members (Ahn, et al., 2017). Moreover, several studies have found positive effects of robots interacting with children. Van der Drift, et al. (2014) found that children shared more personal stories and detail in their diaries when they were interacting with a robot. Jeong, et al. (2017) found that, compared to a plush toy or virtual character, a robot lead to more verbal and physical engagement from the children, talking and playing longer. Family members were also more likely to join if the child interacted with a robot rather than a virtual character or plush toy.

From a NIC perspective, two findings stand out. First, the acceptance of robots and their services by patients and family members depends on their gender, age, cultural background, and job type. Secondly, in the study by Ahn, et al. (2017) patients requested that the robot not only take a measurement of the vital signs, but also provide an interpretation of the numerical results, as the patients did not understand them.

Industry 4.0 and Agriculture

The applications of AI and robotics have been advancing rapidly in the fields of manufacturing and agriculture. Since recent years, the notion of *cobots* is being used for robots that need to collaborate with humans in manufacturing, where simple support evolves to intelligent interaction (Pittman, 2016). The advances in manufacturing are also known as "*Industry 4.0*", a popular term in Germany that refers to *"smart factories"*, i.e., complex networks of humans, (smart) machines and resources (Dopico, et al., 2016). In agriculture, the introduction of fleets of specialized agricultural drones and vehicles has enabled the automation of harvesting, ground preparation, seeding, fertilization, pruning, weeding, and spraying (Bogue, 2016). This has reduced the need for human labour, improved productivity and led to *"precision agriculture"*, i.e., only watering or spraying, etc, where needed (Ibid, and Emmi, et al. (2014)). In both manufacturing and agriculture, these applications of AI and robotics have developed along similar themes, namely reconfigurability, robustness, and decision making.

Reconfigurability for an Industry 4.0 system mean that it can configure and optimize itself, can swap components with other Industry 4.0 systems, and can temporarily form an *"ensemble of systems"* together with other Industry 4.0 systems in order to achieve a task (Mosterman & Zander, 2016). Due to the wide range of possible Industry 4.0 systems, reconfigurability is broad term in Industry 4.0 and can occur through central or decentralized organization, and possibly between dissimilar systems (Jazdi, 2014). For agriculture, reconfigurability is more narrowly defined as the plug and play nature of the implements that the fleets of drones and vehicles use; these implements include different sensors, sprays, fertilizer spreaders etc. Different operations may require different implements, and agricultural robots such as drones, driverless tractors, or dairy robots need to be able to change their implements (Bogue, 2016).

Adaptability refers to how quickly a system or ensemble of systems can change its functions and capabilities. Industry 4.0 systems communicate and collaborate with each other and with humans, to form production lines with novel features. Industry 4.0 systems can adapt in real time to the latest customer demands. Such a system may not only transform production lines but may transform products as well. In addition to manufacturing a product, it can also integrate services with the product, so-called *"servitization"* (Dopico, et al., 2016). The ideal system is also capable of *"individualized mass production"* (Ibid), i.e., producing vast numbers of the product while simultaneously tailoring each individual copy to the tastes and demands of the customer that purchases that copy of the product. Whereas adaptability in manufacturing is mainly about the product, adaptability in agriculture is mainly concern with the speed and robustness of changing the implements, as this determines which type of operations can be done at any time. Both the robot fleets and the implements may be bought from a wide range of vendors, meaning that most agricultural robot operations employ a group of heterogenous systems, equipped with different and possibly non-compatible sensors, actuators, and controllers. The robots carry sensors, but the implements may also carry sensors, creating redundancies in the operation and the possibility of conflicting signals. The mean time between equipment / robot failure tends to decrease exponentially as the number of robots and implements increases. Well-functioning agricultural robot fleets therefore tend to rely on one central, powerful computer system that orchestrates the robot fleets and assigns implements to robots (Emmi, et al. 2014).

This central computer also does most or all of the decision making in an agricultural context, whereas in an Industry 4.0 context decision making can be more decentralized. Moreover, in Industry 4.0, the system(s) use information from the production environment to predict the next steps in production with barely any user interaction and can launch preventive actions to minimize stopping times during production (Li, et al., 2017).

5. AI and Robotics in a Military Context

In this chapter we first describe the current developments surrounding AI and robotics in a military context, and secondly, we summarize the lessons of the use cases for the Defence domain.

The Anticipated Impact of AI and Robotics for Defence

In this section we discuss two reports about AI and robotics in a military context, that together provide insights from a broad perspective, being The Hague Centre for Strategic Studies (HCSS) report *"Artificial Intelligence and the Future of Defence"* (De Spiegeleire, Maas, & Sweijs, 2017) and the Joint Concept Note (JCN) from the United Kingdom (UK) (UK Ministry of Defence, 2017). There are many recent studies on the domain of AI and robotics in the military context. The HCSS report collects and summarizes leading studies written in different languages and coming from various nations. The JCN summarizes in detail the impact of autonomous systems across a broad spectrum of military operations.

The HCSS report projects how developments in AI may impact the military and focusses on small to medium sizes forces, such as the Dutch military. According to this report, studies on the impact of AI on the military tend to describe it as an incremental impact and assume that future armed forces resemble current armed forces. The report argues that AI may have a more transformational impact on defence and security. The foreseen transformational impact is described along a dimension of layers of defence, consisting of: (1) armed forces, (2) the governmental defence organization, (3) defence and security organizations, and (4) the defence ecosystem, which consists of all actors that are somehow linked to achieve defence and security goals. The authors state that whereas most effort goes to enhancing the armed forces with AI developments, most impact will come from changes based on AI developments that change the defence organization itself, the network of partners and the way defence and security will be organized as a whole. For example, self-learning algorithms can be used for organization tasks (Human Resources (HR), policy analysis, etc.), and other applications. The report explicitly challenges to think about the use of AI in a defence context in a different way than extrapolating current practices.

The second report, JCN from the UK ministry of Defence on Human Machine Teaming, focusses more on how further technological developments in AI, Robotics, and autonomous systems will change the teamwork dynamic between man and machine within military operations. It discusses the relative strengths and weaknesses of different kinds of AI and robotics and how this may affect the roles that humans can play within the team. Therefore, the JCN examines the developments in AI and robotics and how these may impact the military. The JCN sees the following advantages that new forms of AI and robotics will allow us: to scale physical mass and battlefield points of presence increasingly independent of the numbers and locations of human combatants; to extend the reach and persistence of our ISR and weapon systems; and to exploit information for a decision-action cycle that makes better decisions more rapidly, enabling a higher tempo of operations.

Furthermore, new AI and robotics technology creates new types of operations. AI will be used to automate both defensive and offensive cyber operations. An overview of research on autonomous military systems efforts by prominent nations is given in (De Spiegeleire, Maas, & Sweijs, 2017).

Lessons from the use cases for Defence

The value tensions and moral dilemmas that plague autonomous cars may transfer to some extent to the Defence domain. What if a NIC can protect only one of two human team members, which one should it choose? Another reason for value tension is that human team members may find it difficult to understand or discover the built-in beliefs or values of their NIC team members. Over time, a human team member may come to understand the beliefs and values of another human team member through the words and actions of that team member. NICs, however, could be programmed to act "*out of character*". It could theoretically be possible, for example, to program a NIC to radically change its personality as the situation requires. If a NIC changes personality, its actions may thereafter run counter to the behaviour that its human team member may have come to expect from the NIC based on the beliefs and values the human believes to have inferred from the NIC's behaviour in the past. To some extent, this type of value tension is also present in the case of autonomous cars. From the outside, other drivers cannot see whether a car is an autonomous car or a human-driven car, therefore to prevent value tension, autonomous cars are programmed to drive like humans even in situations where they are capable to improve upon human driving. An autonomous car could for example keep a shorter driving distance to the car in front of it, since it can react faster than a human can, but such behaviour may make the human in the front car nervous. Another example would be the appearance of a NIC. A human can see in what direction another human team member is looking and hence warn his team member when there is danger behind him or her. Similarly, it may be beneficial to give a NIC something resembling a face, so that human team members can infer in which direction the NIC is looking.

The urban search and rescue use case has interesting parallels for the defence domain, because of the hazardous environments, where lives hang in the balance. The defence organisations are familiar with the task environment for combat search and rescue, where the aim is to rescue own troops in combat zones. In these operations, the presence of hostiles are an extra dimension to the environment. However, the urban search and rescue is relevant, as crucial decisions need to be taken by humans and humans must always be in the loop. The main lesson here is that transparency makes it easier for human team members to accept and trust the robots and their behaviour. Transparency can be achieved in part when the NICs has the ability to understand the spoken dialogue of the operators, including nuances in tone, and when the NIC in turn can produce contextually appropriate verbal feedback for the human team members.

Another lesson is that NICs may need to have built-in adjustable autonomy settings and be able to lower its own autonomy depending on the demands of the operational environment. NICs may have to be hard-coded to adaptively lower their level of autonomy in certain kinds of environments or situations until a human team member allows the NIC more autonomy. A swarm of NIC drones guarding a road may decrease their autonomy upon arriving at a city for example, where the probability of collateral damage in case of an erroneous identification of a target is higher.

A possible lesson from the care for children and the elderly use case is that whether a NIC is accepted by its human team members, may depend on the characteristics of the human team members, such as cultural background and job type. Future military training and education should be able to deal with these kind of individual differences in order to allow for the proper acceptance rate of NICs for all recruits.

The main observation for the Defence domain from the Industry 4.0 and Agriculture use case is that NICs are more dynamic entities than humans: a single NIC may fragment into multiple NICs, multiple NICs may combine or swarm to effectively become one NIC, multiple NICs can swap components and recombine to change their skills, capabilities and appearance. Such dynamic and flexible NICs create different psychological effects on their human team members than static NICs that do not reconfigure themselves. For example, will humans develop trust with reconfigurable NICs that continually change their appearance, partially or maybe even completely? In other words, the lesson for the Defence domain is that reconfigurable NICs may have different psychological effects on their human team members than more static NICs.

Larger numbers of NIC team members and larger numbers of components per NIC increase the probability that at least one is not (entirely/partly) functioning and needs maintenance or repair. Carrying redundant components would make reconfigurable NICs more robust against outage due to component failure. However, as with the agricultural robots, NICs who are reconfigurable and can swap components need to have compatible software, which can be an issue, especially if older / outdated NICs and more modern NICs are present in the same team.

Summarizing, the use cases show six lessons for the Defence organization: (1) value tension may hamper human-NIC collaboration, (2) transparency makes it easier for human team members to accept and trust the NICs and their behaviour, (3) NICs may need to have built-in adjustable autonomy settings, (4) future military training and education should anticipate individual differences in the background of new recruits that may affect their acceptance of NICs, (5) reconfigurable NICs may have different psychological effects on their human team members than more static NICs, and (6) redundancy and compatibility of components is desired to make NICs more robust.

6. Conclusions and Directions for Future Research

This study touched upon several aspects that are important for future deployments of NICs in Defence operations. First, technological trends and developments in AI are discussed to provides a base for the extrapolation of future technological developments in Defence operations. Secondly, we use the concept of autonomy in our vision on the composition of a NIC and explain the characteristics and skills that NICs may possess and along which they may vary. Thirdly, the existing applications of (semi) NICs in other domains than Defence is described. These new applications of integrated AI and robotics across multiple domains offer valuable lessons of the impact of AI and robotics on the humans operating in these domains. Some can be translated into lessons for the Defence domain. Finally, literature on the possibilities and impact of deploying AI and robotics for the Defence context are discussed that focus merely on the organizational and operational impact when adopting AI and robotics and less on the behavioural impact on the humans working with these systems.

Future operations will change due to the deployment of NICs. We believe that future operational effectiveness will be highly influenced by the way NICs are incorporated in Defence operations. We can conclude that the impact of deploying NICs in future operational context is diverse. First, there are new possibilities for operational tactics, new concepts of operations, new forms of manoeuvre.

Second, the impact on individuals, teams and organization. New types of task division may be applied in teams, and new capability requirements for human operators come to play. This affects individuals regarding collaborative skills, psychological effects, etc. Meaningful human control is highly relevant in the Defence context, which requires visions on responsibility and accountability, and lead to human-in-the-loop concepts, and capability requirements to autonomous systems.

Regarding the behavioural impact, three main topics of research have been defined that are considered highly relevant for deploying NICs in the future: (1) psychological effects, (2) human-machine collaboration and, (3) training and education. These three topics are the main focus of the TNO research program BIHUNT, "*Behavioural Impact of Human and Non-human intelligent collaborator Teaming*". The topic of psychological effects studies the social psychological impact on individuals working together with NICs. What are the implications on human team members, regarding trust, biases and emotions? Human-NIC collaboration discusses effectiveness of collaboration of humans and NIC within teams. Topics of interest are adaptable working arrangements, peer-to-peer collaboration versus hierarchical collaboration. To what extend do we allow delegation of decision rights to NICs and what are the consequences for C2? The topic training and education studies the impact of deploying NICs as operational team members on requirements for training and education of military personnel. Which (new) competences are needed when working with NICs? How do human tasks change cause of NIC cooperation, which new aspects should be trained? How does that impact the current training cycle and methods?

Research on the behavioural impact of working with NICs will improve the usability and acceptance of the promising applications that come available from the developments in AI and robotics in the military domain, and will enable a safe and effective operational deployment of NICs.

7. Acknowledgements

This research was done at the Netherlands Organisation for applied scientific research TNO (TNO) as part of the research program V1719 BIHUNT: Behavioural Impact of Human and Non-human intelligent collaborator (NIC) Teaming, sponsored by the Dutch Ministry of Defence. The research program BIHUNT will develop insight in the social psychological effects of deploying NICs as team members on military personnel in these teams. It will also provide insight in the organisational implications the Ministry of Defence should consider to enable an effective and safe operational deployment of NICs.

8. References

Ahn, H. S., Lee, M. H., Broadbent, E., & MacDonald, B. A. (2017). Gathering Healthcare Service Robot Requirements from Young People's Perceptions of an Older Care Robot. *IEEE International Conference on Robotic Computing (IRC)* (pp. 22-27). IEEE.

Ahrndt, S., Aria, A., Fähndrich, J., & Albayrak, S. (2014). Ants in the OCEAN: Modulating agents with personality for planning with humans. *European Conference on Multi-Agent Systems* (pp. 3-18). Springer.

Berger, C., & Rumpe, B. (2014). Autonomous driving-5 years after the urban challenge: The anticipatory vehicle as a cyber-physical system. *arXiv preprint arXiv:1409.0413*.

Bogue, R. (2016). Robots poised to revolutionise agriculture. *Industrial Robot: An International Journal*.

Brooks, R. (2017). The Seven Deadly Sins of AI Predictions. *Technology Review, https://www.technologyreview.com/s/609048/the-seven-deadly-sins-of-ai-predictions/*.

Cacace, J., Finzi, A., & Lippiello, V. (2016). Implicit Robot Selection for Human Multi-Robot Interaction in Search and Rescue Missions.

Castelfranchi, C. (1994). Guarantees for autonomy in cognitive agent architecture. *International Workshop on Agent Theories, Architectures, and Languages.* Springer.

DARPA. (2017). *A DARPA Perspective on Artificial Intelligence.*

De Spiegeleire, S., Maas, M., & Sweijs, T. (2017). *Artificial Intelligence and the future of defense.* HCSS.

Dopico, M., Gomez, A., De la Fuente, D., García, N., Rosillo, R., & Puche, J. (2016). A vision of industry 4.0 from an artificial intelligence point of view. *Proceedings on the International Conference on Artificial Intelligence (ICAI).*

Emmi, L., Gonzalez-de-Soto, M., Pajares, G., & Gonzalez-de-Santos, P. (2014). New Trends in Robotics for Agriculture: Integration and Assessment of a Real Fleet of Robots. *The Scientific World Journal*, 1-21.

Endsley, M. A. (2015). *Autonomous horizons: System autonomy in the Air Force: A path to the future.* Washington, DC: US Air Force Office of the Chief Scientist.

Evans, A. W., Marge, M., Stump, E., Warnell, G., Conroy, J., Summers-Stay, D., & Baran, D. (2017). The Future of Human Robot Teams in the Army: Factors Affecting a Model of Human-System Dialogue Towards Greater Team Collaboration. In *Advances in Human Factors in Robots and Unmanned Systems* (pp. 197-209). Springer.

Future of Life Institute. (2015). *Research Priorities for Robust and Beneficial Artificial Intelligence.* Retrieved from https://futureoflife.org/ai-open-letter/.

Gunning, D. (2017). *Explainable Artificial Intelligence (XAI).* DARPA/I20.

Hanna, N., & Richards, D. (2015). Do Birds of a Feather Work Better Together? The Impact of Virtual Agent Personality on a Shared Mental Model with Humans during Collaboration. *AAMAS*, (pp. 28-37).

Hsu, J. (2018, 05 15). Real Soldiers Love Their Robot Brethren. Retrieved from https://www.livescience.com/5432-real-soldiers-love-robot-brethren.html

Jazdi, N. (2014). Cyber physical systems in the context of Industry 4.0. *IEEE International Conference on Automation, Quality and Testing, Robotics* (pp. 1-4). IEEE.

Jeong, S., Breazeal, C., Logan, D., & Weinstock, P. (2017). Huggable: Impact of Embodiment on Promoting Verbal and Physical Engagement for Young Pediatric Inpatients.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., Van Riemsdijk, M. B., & Sierhuis, M. (2014). Coactive design: Designing support for interdependence in joint activity. *Journal of Human-Robot Interaction, 3*(1), 43-69.

Johnson, M., Bradshaw, J. M., Hoffman, R. R., Feltovich, P. J., & Woods, D. D. (2014). Seven cardinal virtues of human-machine teamwork: examples from the DARPA robotic challenge. *IEEE Intelligent Systems, 29*(6), 74-80.

Kim, S., Zhang, K., & Park, D. (2017). Don't Want to Look Dumb? The Role of Theories of Intelligence and Humanlike Features in Online Help Seeking . *Psychological science*.

Kruijff, G., Janíček, M., Keshavdas, S., Larochelle, B., Zender, H., Smets, N., . . . Liu, M. (2014). Experience in system design for human-robot teaming in urban search and rescue. In *Field and Service Robotics* (pp. 111-125). Berlin Heidelberg: Springer.

Li, B.-H., Hou, B.-C., Yu, W.-T., Lu, X.-B., & Yang, C.-W. (2017). Applications of artificial intelligence in intelligent manufacturing: a review. *Frontiers of Information Technology & Electronic Engineering, 18*(1), 86-96.

Lin, P., Bekey, G. A., & Abney, K. (2009). Robots in war: issues of risk and ethics. In *Ethics and Robotics* (pp. 49-67). IOS Press.

Lohani, M., Stokes, C. K., Oden, K., Frazier, S. J., Landers, K. J., Craven, P., . . . Macannuco, D. J. (2017). *The Impact of Non-Technical skills on Trust and Stress*. Retrieved 11 15, 2017, from http://dblp.uni-trier.de/db/conf/hri/hri2017c.html

Mioch, T., Marieke, P. M., & Mark, N. A. (2018). Improving Adaptive Human-Robot Cooperation through Work. *The 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).*

Mosterman, P. J., & Zander, J. (2016). Industry 4.0 as a cyber-physical system study. *Software & Systems Modeling, 15*(1), 17-29.

Murphy, R. R. (2014). International Cooperation in Deploying Robots for Disasters: Lessons for the Future from the Great East Japan Earthquake. *Journal of the Robotics Society Japan, 32*(2), 104-109.

Pierson, H. A., & Gashler, M. S. (2017). Deep learning in robotics: a review of recent research. *Advanced Robotics, 31*(16), 821-835.

Pittman, K. (2016, October 28). A History of Collaborative Robots: From Intelligent Lift Assists to Cobots. *Engineering.com*.

Salvit, J., & Sklar, E. (2012). Modulating agent behavior using human personality type. *Proceedings of the Workshop on Human-Agent Interaction Design and Models (HAIDM) at Autonomous Agents and MultiAgent Systems (AAMAS)*, (pp. 145-160).

Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv:1708.08296v1*.

Saunders, J., Burke, N., Koay, K., & Dautenhahn, K. (2013). A user friendly robot architecture for re-ablement and co-learning in a sensorised home. *Assistive Technology: From Research to Practice (Proceedings of AAATE), 33*, 49-58.

Schoettle, B., & Sivak, M. (2014). *A survey of public opinion about autonomous and self-driving vehicles in the US, the UK, and Australia.* University of Michigan, Ann Arbor, Transportation Research Institute.

Schwendner, J., & Kirchner, F. (2014). Space Robotics: An Overview of Challenges, Applications and Technologies. *Künstliche Intelligenz, 28*, 71-76.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., . . . Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature, 550*, 354-359.

UK Ministry of Defence. (2017). *Joint Concept Note 1/18 Human Machine Teaming.* Swindon: The Development, Concepts and Doctrine Center.

Uszok, A., Bradshaw, J., Jeffers, R., Suri, N., Hayes, P., Breedy, M., . . . Lott, J. (2003). KAoS policy and domain services: Toward a description-logic approach to policy representation, deconfliction, and enforcement. *Policies for Distributed Systems and Networks* (pp. 93-96). IEEE.

van den Bosch, K., Brandenburgh, A., Muller, T. J., & Heuvelink, A. (2012). Characters with Personality! *Intelligent Virtual Agents* (pp. 426-439). Heidelberg: Springer.

van der Drift, E. J., Beun, R. J., Looije, R., Blanson Henkemans, O. A., & Neerincx, M. A. (2014). A remote social robot to motivate and support diabetic children in keeping a diary. *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 463-470). ACM.

van der Vecht, B. (2009). *Adjustable autonomy: Controling influences on decision making.* Utrecht University.

van der Vecht, B., van Diggelen, J., Peeters, M., Barnhoorn, J., & van der Waa, J. (2018). SAIL: a Social Artificial Intelligence Layer for Human-Machine Teaming. *16th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS).* Toledo, Spain.

Ventura, R., & Lima, P. U. (2012). Search and Rescue Robots: the Civil Protection Teams of the Future. *Third International Conference on Emerging Security Technologies.*