University of Tartu
Faculty of Social Science
School of Economics and Business Administration

ROBOTS AND HUMANS: ATTITUDES TOWARDS "BLACK MIRROR" EFFECT

Master's Thesis

Iris Del Carmen Izaguirre Orozco

Supervisor: Elina Kallas

Tartu, 2019

Name and signature of supervisor…….

Allowed for defence on.....

(date)

I have written this master's thesis independently. All viewpoints of other authors, literary sources and data from elsewhere used for writing this paper have been referenced.

…….. (signature of author)

Table of Contents

Abstract

This thesis investigates human attitudes towards robots from the experts' perspective. For this purpose, six experts in the field of artificial intelligence and robotics, from Estonia and USA, were interviewed and results were analyzed using a thematic content analytical method. The analysis revealed six themes about human-robot interaction in the following areas: positive impact of robotic systems on humanity, robotic systems and biases, human attitudes and engagement, human attitudes and negative emotions, influence of robotic systems on mental health and the level of autonomy in decision making. This study shows that experts from different fields and levels of experience envision predominantly positive attitudes towards robots in several aspects of the development and future integration of robotic systems into society.

Introduction

From rumba robot vacuums at homes to the Henn na Hotel in Japan, which is the first hotel in the world staffed by robots. Robotics systems are now a part of our environment and these are expected to become an integral part of human's life in the future.

Humanity is looking forward to the time when these robotic systems will be capable of interacting autonomously with us and have the ability to operate across several contexts; predicting humans' behavior and reacting to our actions and attitudes using their own cognitive capabilities, interacting in contexts and environments not contemplated by designers, users, or in situations that not human has encountered yet, while at the same time utilizing resources in unexpected ways.

In order to reach the state where robotic systems are autonomous and capable of high-level decision-making process, the scientists, designers and engineers from different fields are working to understand humans' behavior and our current interaction with robotic systems. According to (Salah, Ruiz-del-Solar, Meriçli, & Oudeyer, 2012) study, humans behave significantly differently when they interact with robots, than when they do human-to-human.

Human behavior is structured along individual and social lines, but it is complex at the same time; cultural background, experience and even language have a great impact on how humans perceive objects. In order to interact with people in uncontrolled environments, robotic systems need to be able to correctly predict, interpret and respond to complex human behaviors. On the

other hand, it is also required to understand idiosyncratic factors of the individual user that may have an impact on behavior, in order to understand how the behavior and embodiment factors of a robot are perceived and responded to by potential users (Syrdal, Dautenhahn, Koay, & Walters, n.d.).

To understand humans' attitudes towards robotics systems, we must address human interaction, which requires communication, influenced by the proximity to the robot. To develop robotic systems that meet humans' expectations, a field of study has emerged during the mid-90's. With less than 30 years of existence, the multidisciplinary field of Human-Robot Interaction (HRI) is dedicated to understand, design, and evaluate robots. The field integrates several areas of study and research, some of them are: robotics, cognitive science, human factors, natural language, psychology, and human-computer interaction (Goodrich & Schultz, 2007). This thesis relies mainly on the findings made in the HRI field of study and other areas of research such as Robotics and Artificial Intelligence.

Technological advancement and development give humans a possibility to make our life easier. Through Artificial Intelligence, scientists are working to recreate the human mind with all its cognitive capabilities and trying to transfer it into a robotic system. But this isn't just about technology. The metaphorical meaning of "Black Mirror" translate as: the black screen of an electronic device when it is shutdown, allowing the user to see its own reflection. However, the meaning goes beyond this point. A question, that is worth asking is: Are we mirroring or even magnifying the ugliest and darkest aspects of human nature through the development of such technologies?

A very important part of these studies is attitude. Attitudes are a cognitive process by which the evaluation of objects is complex, variable over time and dependent on individual's cultural background and experience (Krosnick, Judd, & Wittenbrink, n.d.). Due to the complexity of human cognition and behavior, we might not yet have the tools to address the issues related to the capabilities of an artificial agent with human cognitive skills. Even more so once it can improve and develop itself.

With all the above in mind, the aim of this thesis is to investigate human attitudes towards robots from the experts' perspective. By adding value to our current understandings of human attitudes towards robots, this investigation hopes to contribute to the Human -Robot Interaction field of study. Through mapping the existing academic research and theoretical background

addressing human-robot interaction, we will be able to know where we are standing in terms of the knowledge currently available to the practice. By interviewing experts in the field, this thesis will intent to provide their current perspective. From the experts' point of view, this research documents what types of fears might be developed in line with the development of this technology.

In the first part of the document, generalities of robotics and artificial intelligence are briefly approached, the background of Human-Robot Interaction (HRI) is to be developed in more depth. The second part enters the empirical part, where interviews with experts in this area were conducted to investigate their points of view, in relation to what attitudes humans might form regarding robotics systems. In the fourth part the implications of a potential "Black Mirror" effect are discussed. To finalize, the research formulates conclusions.

1. Robots and artificial intelligence (ai).

The literature reveals that for centuries humans have dreamed about mechanic artifacts capable of interacting with them and performing activities as humans do. Derived from a desired to create living things, the Greeks created the automata, the first life-like working robot (Hockstein, Gourin, Faust, & Terris, 2007). Hundreds of years later during the Renaissance artists created a more complex mechanism that replicated some human actions (Zamalloa et al., 2017). With the Industrial Revolution, however, humanity finally saw machines undertaking human labor (Ewing, Pigazzi, Wang, & Ballantyne, 2004).

The word "robot" means "work", "labor" and it is originated from the Slavic word "robota". Inspired by the mechanized labor in the 19th and 20th century. The Czech playwriter Karel Čapek introduced the term in his play *Rossum's Universal Robots (1921)*. Later, Issac Asimov popularized the word Robot through his science fiction short story *Runaround* in 1942, later incorporated to *I, Robot* (1950). The first steps towards what we know now as humanoid robots, started with the work of engineers inspired by those science fiction novels. After WWII a company called Unimation created the first programmable manufacturing product. In 1962 the first robot created by the company was installed in an assembly line of General Motors. Within the medical field we have seen great advances in this area, as well; in 2004, there were 5 robotic surgical systems, approved by the FDA in the United States, to assist in human surgeries (Ewing et al., 2004). Nowadays, we are using robots in a variety of applications from the exploration of the space and deep in the sea to military uses such as search and rescue missions

and performing surgeries. These systems are helping us to enhance human capabilities. (Hockstein et al., 2007).

The reviewed papers, for the purpose of this research, showed a different timeline history and emphasized different turning points of the AI field. This research will only highlight those events that are recurrently mentioned on the papers (See: figure 1).

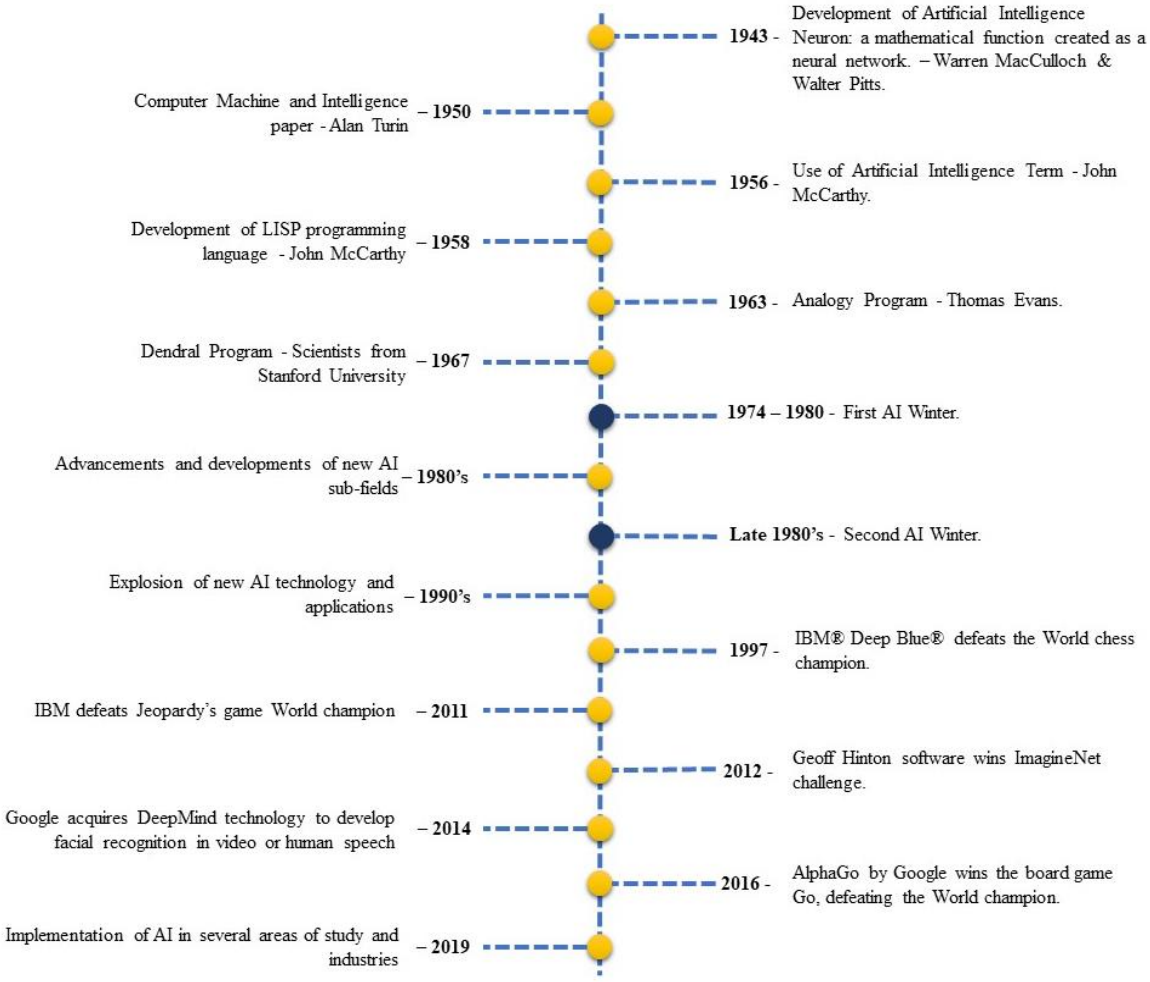

Figure 1: A short timeline of AI. Composed by Author based on ("A Time Line of AI.," 2016)

The history of AI goes back to several centuries, when philosophers from China, Greece, Egypt and India fantasized through artifacts, myths and stories about the possibility of intelligent machines (Lee et al., 1982).

During the sTARTUp Day 2019 event in Tartu, Estonia; a seminar was given by Dr. Adrian Kaehler, an American scientist, CEO & Founder of Giant AI. At this seminar Dr. Kaehler gave a brief but reach timeline of Artificial Intelligence history. He stated that AI field, as we know it now, started in 1943 with the development of the artificial neuron, which is a mathematical

function created as a neural network; in simple words, each neuron receives one or more inputs and sums them to generate an output (Basheer & Hajmeer, 2000).

Most scientists introduce Alan Turing 1950's paper, *Computing Machine and Intelligence*, as the turning point of AI beginnings (Coppin, 2004). In this paper Turing presents the idea about the possibility of programming an electric computer to behave intelligently; the paper also includes Turing's test, known as the imitation game.

The term Artificial Intelligence was firstly used in 1956 at a conference in Dartmouth College, in Hanover, New Hampshire, by John McCarthy. Two years later, in 1958 McCarthy invented LISP programming language, which is still used in Artificial Intelligence research (Coppin, 2004). In 1963, Thomas Evans wrote the Analogy program, as a part of he's work at MIT; the program is the first to explore analogical reasoning with a running program, demonstrating that computers can solve IQ tests (Buchanan, n.d.).

In 1967 the Dendral Program was developed by four scientists at the University of Stanford. The program is considered as another turning point of knowledge-based systems, which interprets mass spectra on organic chemical compounds. During this decade many organizations were created in order to support the development of Artificial Intelligence. Among the most important are the MIT, Massachusetts Institute of Technology and the ACM SIGART, Association for Computing Machinery's Special Interest Group on Artificial Intelligence. (Buchanan, n.d.)

A number of publications have mentioned the first AI winter to happen between 1974 to 1980. Due to poor development on hardware and software, successful results in the field were very hard to conceive. Organizations such as DARPA, (the Defense Advanced Research Projects Agency), had several disappointments with failed projects that led the organization to put on hold million of dollars in AI research. However, even when the founding and interest on the field has dropped for a few years, research continued slowly but steady. ("A Time Line of AI.," 2016)

A new chapter began in the 1980's, the field increased its popularity and commercial success, researches produced a significant number of powerful new technical tools and perfectioned others. This has resulted in the development of new subfields, such as computer vision, natural

language processing, machine learning, reasoning and representation. All this work magnified AI's abilities, through the strengthening of new connections with statistics and control engineering. Unfortunately, at the end of this decade, the field experienced the second winter time (Nilsson, 2009). The 1990 has brought significant advances in all areas of AI, resulting in an explosion of new AI technologies and applications. A remarkable achievement of the decade was the defeat of the chess world champion in 1997, Garry Kasparov, by an IBM computer called IBM® Deep Blue®. The machine won a six-game match ("A Time Line of AI.," 2016).

In the second decade of the 2000's IBM® again defeated another world champion at Jeopardy, wining against Ken Jenning and Brad Rutter in 2011. The next year, deep-learning machine software from Geoff Hinton lab won ImagineNet Large Scale Challenge by identifying a thousand types of objects about 85% of the time. Two years later, in 2014 DeepMind Technologies is acquired by Google, the startup focused on achieving tasks such as facial recognition in video or human speech. By 2016, Google's AlphaGo wins the complex board game Go, beating the world champion of that time. ("A Time Line of AI.," 2016)

Nowadays, consumers have experienced the raise of AI in several applications; from voice speech and facial recognition embedded in digital devices and home appliances, to businesses implementing chatbots, data analysis of sentiments and prediction of preferences offering customized and personalized products and services, engaging with their customers at a deeper and unique level.

In 2019, as the technology matures, scientists, governments and private companies continue to employ AI, discovering different ways of application. At this point, trends indicate that AI will expand to different areas, such as psychology, where the application of several tools from AI and integration with machine learning can help to improve decisions associated with the diagnosis, prognosis, and treatment of people suffering from mental illness (Gaggioli, 2019). AI is also transforming the way financial institutions communicate with their clients, banks and insurance companies are investing millions of dollars to improve engagement and loyalty rates through the reduction of average handle time and meaningfully connecting with them trough chatbots and AI agents trained in financial language (Rocca, 2018).

By 2019, the general public has already got used to hear from the daily news-outlets about advances, developments and new achievements in robotics and artificial intelligence technologies. We hear debates and discussions about ethical aspects; if something goes wrong, who are we going to account responsible? or, when all of these are going to become a reality? On the other hand, the economic implications and repercussions involve the increase of efficiency and productivity on those sectors and industries that adopt the technology. But, others debate that it may also affect negatively the labor market of industries and countries that are not yet prepared with high trained or skilled work force, that can rapidly adapt and switch to new type of jobs (Rajnai & Kocsis, 2017). Prominent public figures like Bill Gates and Stephen Hawking (1942 - 2018), expressed concern about the possible dominance of robots taking over the world (Goolsbee, 2018). Tesla Motors' CEO, Elon Musk once said: "We must pay the closest attention to AI because it could bear more dangerous than nuclear weapons" (Catherine Clifford, 2018).

In 2018 the Chinese government started to envision how to become the market leader in robotics and AI by 2025. The enterprise of robotics plus artificial intelligence have a great potential that governments want to leverage for commercial opportunities of hardware and software ecosystems. By 2018, China has possessed 60% of the market share for industrial robotics (PR Newswire, 2018). Some areas such as the manufacturing and healthcare industries are taking advantage and levering the benefits of both technologies, as well (Ewing et al., 2004). The financial industry has been working towards the innovation and digitization of their internal processes; and AI is playing an important role. In 2019 there are available AI financial advisors trained in financial language that are outperforming human advisors. Similarly, the same technology helps financial institutions to forecast trends and predicts behavior from public mood detected on social media (Trestle Group, 2018).

Policy makers are debating the possible positive and negative consequences of integrating smart robotic systems into society. Over the years, before robotics and AI were implemented, other technologies were already displacing low skill jobs. Policy makers argue that the speed of adoption of these technologies is a key element to disrupt economies. The degree of the negative consequences in economies is tight to how fast is the ability to adopt and integrate them into humans' daily lives (Goolsbee, 2018). They fear that with the adoption of intelligent robotic systems in the public and private sector, hundreds of millions of jobs will be lost.

2. Background of human-robot interaction (hri)

Robots have been around us for a couple of decades now, and as one day sales men and scientists predicted that every household will have a TV or a personal computer, robots are expected to become an integral part of our lives sooner rather than later. During the last 30 years, scientists have been working toward the development of robotic systems that will be capable of interacting with humans and behaving according to our social etiquette. This implies that robotics systems need to have the skills to travel around, next to people or side by side, and more importantly to be able to predict, understand, and convey our intentions within our actions in uncontrollable environments.

Two examples of these robotic systems are autonomous robots known as humanoid robots. Robots with anthropomorphic characteristics, imitate human behavior, actions and reasoning, and fully autonomous robots, which requires a human to supervise, direct and provide goals to the robot, while maintaining knowledge of the environment, tasks and constraints (Goodrich & Schultz, 2007). Commercial enterprises have already developed robotic systems with these characteristics. As example of these (*image 1:* NAO & Pepper robots, and *image 2:* Starships Technology full autonomous robot) Softbank Robotics already commercialize two popular humanoid-like robotic systems, which are used in schools, colleges and universities for academic research. A fully autonomous robot is being commercialized by the Estonian company Startship Technologies with the purpose to serve as a delivery robot.

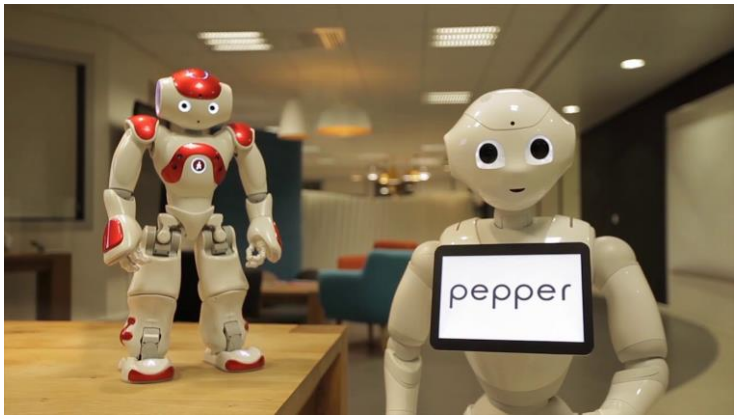

Image 1: Humanoid Robots, from left to right: NAO Robot and Pepper Robot by SoftBank Robotics. Photo credit to: ITBBerlin, 2018.

Image 2: Fully Autonomous robot by Starship Technologies. Photo credit to: Starship Technologies

Dr. Robin Murphy and Dr. Erica Rogers organized a workshop on human-robot interaction, sponsored by the Research Projects Agency. The aim of the workshop was to help identify issues and challenges in HRI by bringing together a highly multidisciplinary group of scientists and researchers. Scientists and researchers from different domains such as human computer interaction, robotics, cognitive science, human factors, natural language and psychology started to come together during mid 1990s and early 2000, recognizing the importance of working together to develop robotic systems capable of interacting at a human level (Goodrich & Schultz, 2007).

In general, the human-robot interaction (HRI) research can be divided into two main categories:

1. Human-centered HRI, which investigates areas of design and usability of proper interaction interfaces, robot platforms, and behaviors through extensive user studies (Salah et al., 2012).
2. Robot-centered HRI, which focuses on algorithms, engineering innovations, and other computational approaches that would improve the overall performance of the interaction *(ibid).*

In regards to communication, the interaction with a robotic system can be divided into two general categories (Goodrich & Schultz, 2007):

1. Remote interaction: when the human and the robot are separated in time and space; for example, the Mars Exploration Rover (see Image 3).

Image 3: Mars Exploration Rover (rear) vs. Sojourner rover (Courtesy NASA/JPL-Caltech/Wikipedia).

2. Proximate interaction: when the human and the robot are together at the same time and space; for example, service robots (see Image 4).

Image 4: Service Robot Sanbot Max by Sanbot Innovation Technology. Photo credit to: Sanbot Innovation Technology.

Within HRI there is another important concept that seems to capture a recent research direction of the field. This concept is called dynamic interaction, which include time and task varying changes in autonomy; to subsume adaptive and dynamic autonomy as a special case, information exchange; that includes adaptive and adaptable interfaces, team organization and authority; to address mixed initiative interaction, and training to include interactive learning. More specifically this concept emphasizes on shaping the type of interactions that can emerge between humans and robots (Goodrich & Schultz, 2007).

In 2004, researchers from the University of Massachusetts and the MITRE Corporation extended the introduced taxonomy of the roles that robots can assume by Scholtz in 2002. (Yanco & Drury, 2004) added additional categories and updated the existing ones. A full table containing both taxonomies can be found in Appendix I of this research.

Robotic systems level of autonomy is tight to context, type of task and level of human interaction. Figure 2 shows the levels of autonomy with emphasis on human interaction, where direct control of a robotic system is required for teleoperation systems and a more dynamic level of autonomy is advised when a peer to peer collaboration is required.

Level and behavior of autonomy: the first attribute that affects the interaction between human-and-robot is autonomy. The degree of autonomy varies according to the capabilities of the robotic system. To achieve peer-to-peer collaboration the system requires to exhibit "fully autonomous" behavior, supporting at the same time social interactions abilities. On the other hand, there is teleoperation, where the system requires direct control from the operator (Goodrich & Schultz, 2007).

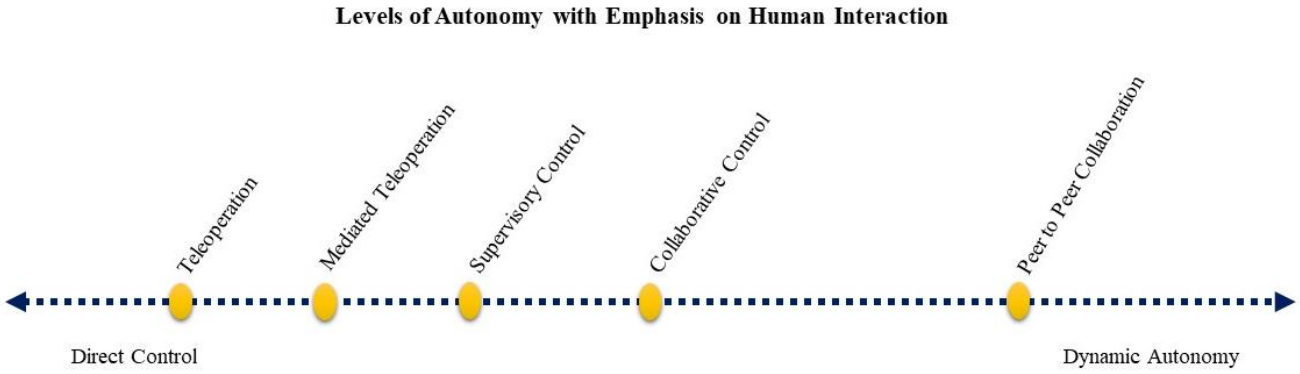

Figure 2: Levels of Autonomy with Emphasis on Human Interaction. Source: Goodrich & Schultz, 2007

Nature of information exchange: the exchanged information is primarily determined by two dimensions - the communications medium and the format of the communications. Seeing, hearing, and touch are the primary media. Additionally, there are measurements that evaluate the efficiency of the interaction, these are: interaction time, cognitive or mental workload of an interaction, amount of situation awareness produced by the interaction, and the amount of shared understanding or common ground between humans and robots (Goodrich & Schultz, 2007).

Structure of the team: when working with a robotic system, several questions arise, for example; who has the authority to make decisions, issue instructions or commands, and at what level; how conflicts are going to be solved; or how roles are going to be defined and supported between humans and robots. These arising questions are with respect to human-and-robot interacting and performing a task together. However, there might be some instances, when a robot may have to interact with bystanders or with people that are not expecting to interact with the robot *(ibid).*

Adaptation, learning, and training of people and robot: these are key elements for long term interactions. Designers are working toward minimizing operator training though modes that invoke commonly held mental modes or by exploiting fundamental cognitive, social and emotional process between the robot-human interaction. However, there is a possibility where robots may not achieve the expectations, generating an uncomfortable interaction for humans. Other efforts are being directed to train humans, in some domains such as military, police, space as well as research and rescue missions that requires remote robots and careful control. The operator requires careful training due to the workload or high risk of the operation task. Finally, designers are also being trained on metrics and experiments design for robotic applications *(ibid).*

Shape of the task: the final attribute to consider is task-shaping, it is imperative to determine and specify how a task is done and how it should be done, including unintended consequences of the design *(ibid).*

3. Humans' perceptions and attitudes towards robots.

As robotic systems are expected to become an integral part of our lives, the adoption of this technology by the general public is imminent. With its integration, robotic systems will interact with humans across several contexts, capable of understanding complex human behaviors and executing tasks in uncontrolled environments. Influenced by the time and space of the proximity, interaction between humans and robots requires communication. The type of communication and proximity determines the level of autonomy and decision-making prerogative given to the robotic system. In order to understand different types of interaction and communication, three investigations that examine human behavior, perceptions and attitudes towards interaction with robots have been selected to discuss humans' perception of robots in different contexts. To develop and design robotics systems that respond

accordingly to our expectations and intentions. The aim of these studies is to help HRI field of study to understand better human's behavior, perceptions and attitudes.

The first study is a pear-to-pear collaboration type of interaction study: "Natural Person-Following Behavior for Social Robot". Scientists from the University of Carnegie Mellon investigated the social perceptions of a robot's movement as it followed a person (Gockley, Forlizzi, & Simmons, 2007), with the aim to develop robots that can accompany people in socially acceptable ways. The paper discusses a laser-based method using two different approaches of person-following: *Direction following* and *Path following*. *Direction following* approach consisted of a robot attempting to drive in the direction of the tracked person's current position. The challenging aspect of this method was that, by going towards a person the robot would interpret the person as an obstacle. In this scenario, the robot was programmed to favor going slowly close to obstacles, such as a person, as long as the direction was correct. The *Path following* method consisted of a robot attempting to follow the path the person took, as close as possible. By using this approach, the robot was required to switch to the opposite side of a hallway at certain locations and driving around corners with the same curvature as the person's travel. The results of the study revealed that participants rated the *Direction following* robot's behavior significantly more natural and human-like. Participants also felt this robot behaved accordingly with their expectations. The study also indicates that even though the robotic system presented non-anthropomorphic physical shape, participants expected a human-like behavior from the robot (Gockley et al., 2007). Robotic systems with this type of capabilities and behavior are expected to engage and support humans in specific environments, such as hospitals and nursery homes. The important aspect of contexts is that robotic systems are expected to respond and understand verbal and nonverbal social cues, adapting its behavior to a particular situation. For example, a robotic system should behave in a certain way when it is following the person, and in a different way when the robotic system is leading or walking side-by-side to a person or a group of people.

The second study about humans' perception of robots is in a scenario where a group of people was involved "Towards robots reasoning about group behavior of museum visitors: Leader detection and group tracking" made by researchers from the Universitat Politècnica de Catalunya in Spain and the National Institute of Informatics in Japan, the study aimed to track and reason about social grouping with highly defined roles using a simplistic cognitive

process, rather than learning a priori social relations (Trejo, Angulo, Satoh, & Bono, 2018). The research conducted at a science museum, investigated important issues in designing a robotic system that interacts with people in daily life environments. Through the introduction of a computer vision algorithm, it was able to detect and track a leader within a group of people and differentiate between group members and non-group members. Scientists were able to analyze cognitive and logical behavior of people's interactions on scene. (Trejo et al., 2018). As social beings, people form groups, interact with each other, merge to groups or separate from them. The composition of groups is highly complex and brings difficulties to measure social relations among subjects. These social relations are correlated with physical distance and space during interaction. By understanding individual's behavior and interactions within a group, a robotic system could serve as a companion. It can be designed to socially interact with the environment, act autonomously and independently, actively participate with the group, adapt its behavior and ultimately alleviate the workload of group leaders.

The final study in this section is "The Relation between People's Attitude and Anxiety towards Robots in Human-Robot Interaction". Scientists from the faculty of behavioral science in the Netherlands conducted a research with students from different nationalities to examine the relation between people's attitude and anxiety when interacting with robots. (de Graaf & Ben Allouch, 2013). The results of the study determined that an interaction with a robot had an effect on people's behavior and anxiety. Researchers identified interesting findings, some of them were: 1) there was an effect on attitude and anxiety before and after an interaction with a robot 2) negative attitudes towards the interaction with the robot, increased on women after the interaction with the robot, 3) in regards to communication capabilities of the robot and discourse, people became more anxious, 4) after the interaction, people felt less anxious towards behavioral characteristics of the robot, 5) both gender showed significant effects towards discourse with the robot. However, females showed more anxious levels than man, and 6) participants from different nationalities showed different effects on social influence of robots (de Graaf & Ben Allouch, 2013).

On one hand, through findings one to three, is possible to infer that humans are afraid of technology when it is unknown. Before humans have experienced an interaction with a robotic system, they are more likely to show negative attitudes. However, after the interaction has occurred, allowing the human to gain knowledge about the technology, the level of

anxiety decreases. On the other hand, as the internet and electronic devices are helping us to become more efficient and knowledgeable, they also have impacted our mental health and behavior. In this context, will robotic systems have an impact on human behavior and mental health? A partial answer to this question is addressed in the results and analysis of this thesis. However, a more in-depth study and research need to be done.

In the subject of humanoid robotic systems, (Bubaš & Lovrenčić, n.d.) outlined the importance of developing agents capable of behaving and communicating accordingly with humans etiquette. The more scientists develop such technologies, the more they realize how much humans expect a similar or identical feedback to our own. As it is mentioned in this paper, the design of a robotic agents not only should include physical or anthropomorphic characteristics, but also incorporate social features; these features include verbal and nonverbal communication skills, develop some level of awareness about actions and the environment, and be capable to adapt based on human feedback.

The development of measurements towards robots are important because these indicators help scientist to compare the success of robotic agents, their behavior, interactions and concrete representations across different contexts. Almost 15 years ago, the HRI field of study was characterized by disagreement of methodologies and measurements scales that intended to help the community to evaluate their impact on specific and different contexts. Scientists contemplated that this differences could create a limitation in replicability and common benchmarks within the field of study (Syrdal et al., n.d.). Today, the community of scientist continues to work towards the refinements of these methodologies, trying to understand how different cultures and social settings might impact attitudes towards robotic systems. While gathering information, the author came across a few means of measuring attitudes such as, Implicit Association Tests and non-standardized Likert-scale questionnaires, Negative Attitudes towards Robots Scale (NARS), Robot Anxiety Scale (RAS), and Interpersonal Communication Competence (ICC) (Bartneck, Nomura, Kanda, Suzuki, & Kato, 2005) & (Bubaš & Lovrenčić, n.d.). For the purpose of this thesis, the following section presents through three different studies the Negative Attitudes towards Robots Scale.

Negatives Attitude towards Robots Scales contains 14 elements (Syrdal et al., n.d.), and 3 sub-scales that gauge interaction, social and emotion constructs. The first sub-scale measures attitude towards the interaction with robots, the second measures attitude towards social

influence of robots, and the last one measures attitude towards emotions in interaction with robots (Bartneck et al., 2005).

A research conducted in 2005 by scientists from Japanese and Dutch universities, investigated through a cross-cultural study the Negative Attitudes towards Robots Scale (NARS) (Bartneck et al., 2005). The scientists presented a questionnaire to participants from seven different countries from Europe, Americas and Asia. The scale was translated to English utilizing the forth and back translation process, which its original questionnaire was first created in Japanese (Bartneck et al., 2005). Contrary to social believes that Japanese love robots, this study revealed that Japanese were concerned regarding the impact these agents might have on society. Scientist speculated that this negative perception could be a result of the high degree of exposure Japanese have to robots, making them more aware of the capabilities these agents might lack or possess. The least negative results came from Americans, particularly in the interaction construct. Again, scientists ponder a possible reason for this result, which could be attached to the fact that Americans are used to technology and they are more comfortable talking to new people. On the other hand, Mexican participants showed the most negative attitudes towards robots, especially in regard to the interaction with them (Bartneck et al., 2005).

A second study that implemented NARS scale is "The Negative Attitude toward Robots Scale and Reactions to Robots Behavior in a Live Human-Robot Interaction Study" (Syrdal et al., n.d.), the aim of this study was to validate the scale and its sub-scales. Interestingly, researchers executed two interactions sessions with 28 participants (14 female and 14 male) from the University of Hertfordshire. The purpose of implementing two sessions was to expose the participants to two different type of robot behavior a) *socially ignorant* and b) *socially interactive*. The study discussed the utility of the scale to explain how people view and evaluate robotic systems from other aspects. The scale allowed to differentiate between the two robot behavior styles, which as whole were not evaluated differently from the overall sample. However, researchers bring attention to a more negative evaluation of behavior from Robot B, which was intended to display a more sociable behavior. Assumptions of these results from researchers point to several reasons: a) since robot B is more socially sophisticated, participants showed to be more cautious, b) could be that the robot is more autonomous and thus less predictable, and c) by being more interactive, it was perceived as intrusive (Syrdal et al., n.d.).

In the health care field (Rantanen, Lehto, Vuorinen, & Coco, 2018) examined care personnel's attitudes towards robots among Finnish home care personnel. The study shows that NARS scale helped to explain phycological resistance of home care workers willingness to introduce care robots into their work. In part, the study concludes that personnel attitudes and perceptions towards robots play a key role in the introduction of care robots into home care, which can slow down their introduction and impact the acceleration of its development. Researchers speculate that this result could relate to the fact that robots were quite unknown to many respondents (Rantanen et al., 2018).

Negative Attitudes towards Robots Scale has investigated in several studies attitudes towards robots and the communicative behavior that take place in human-robot interaction. These studies also emphasize the significant impact that culture and language might have in the internal consistency and the validity of measures when applied to other cultures. Finally, even though humans expect robotic systems to be smart, aware of its surroundings and react to our actions. The author speculates that at some degree humans are not ready yet to interact with more autonomous, less predictable and intrusive agents. Moreover, humans are afraid of the unknown and as with many other technologies, the adoption and integration into the everyday will take time, information and experimentation. This gap might affect its rapid development.

4. Methods and data

The literature review section of this thesis helped the author to understand the development of the technologies and areas of study discussed. In particular, the section Humans' perceptions and attitudes towards robots, served as a foundation to develop the questionnaire used to interview the experts invited to participate. Six interviews were conducted with experts from Estonia and United States, their areas of expertise included Artificial Intelligence and Robotics (One researcher, two Professors, two Post Doctor students and One Master Student). The research method utilized was a semi-structured interview, guided by a questionnaire written in English (appendix II), containing 24 questions; divided in three sets: 10 general questions, 9 emotions related questions and 5 interactions related questions. This questions were inspired by Nomura's NARS scale described in (Syrdal et al., n.d.), which includes societal attitudes, negative social impact of robots today and in the future and phycological reactions; related to immediate interaction and emotional reactions to robots.

The interviews were performed in English via, Skype or face-to-face with a duration of forty-five minutes to one hour. In order to effectively process the information, the interviews were audio recorded and transcribed. Transcripts are contained at the end of the thesis (see Appendix IV).

A qualitative thematic content analytical method was implemented to analyze the content of the interviews. This method identifies patterns or themes within qualitative data, it is a flexible method that helps with the diversity of the work. (Maguire & Delahunt, 2017). The author follows Braun & Clarke's (2006) 6-step framework; 1- become familiar with the data, 2- generate initial codes, 3- search for themes, 4- review themes, 5- define themes, and 6- prepare report. As a result, the analysis of the interviews revealed 6 themes (see appendix III for full summary of themes): positive impact of robotic systems on humanity, robotic systems and biases, human attitudes and engagement, human attitudes and negative emotions, robotic systems influence on mental health, and the level of autonomy in decision making.

5. Results and analysis

Due to the nature of this study, a thematic content analytic method was implemented in order to identify patterns across experts' responses and find themes that explain their perceptions and attitudes towards robotic systems. The analysis revealed six themes, the first two are general themes in the areas of the positive impact of robotic systems on humanity and robotic systems and biases. The following three themes are related to human attitudes, engagement, negative emotions, and influence of robotic systems on mental health. The final theme is in relation to the level of autonomy in decision making. The main results are as follow.

Positive impact of robotic systems on humanity. predominantly, experts envision a positive impact of robotic systems on humanity. The technology will help us to improve our way of living on different levels, from cognitive intelligence to physical strength, helping us to become superhumans or by eliminating the type of jobs we will not want to do in the future. Even though, this future is far from becoming a reality, a clear dependency to current - not so intelligent systems - has already been shown. The dependency will become stronger as we advance in time and it seems to be the next step of evolution and progress for our civilization. Expert 1: *"If you add robots, it's going to combine many different levels of intelligence and appendix to human intelligence".*

Expert 2: *"I think they are going to replace a lot of tedious human jobs".*
Expert 5: *"One way I can think of, as many machines along the history had made lives easier in the sense that we don't have to do hard labor. I do hope that when we are creating machines, when we also have humanoid robots that these will make our lives easier".*

Robotic systems and biases: A crucial part of a robotic system is its AI algorithm. Due to its adaptive nature in its initial state, artificial intelligence is being described as child that learns in the process by doing, which cannot differentiate the good from the bad. Since humans are the ones building these machines, the systems already contain humans' biases, subjected to the engineers that are building it. But, transferring these biases doesn't necessarily translate into a problem. Once robotic systems go through commercial process, where it will experience exposure to market research and the improvement iteration process, then this technology will be driven by the general public and its feedback. Similar as it is nowadays with smartphones and other devices. However, to prevent biases, experts suggest two solutions: provide all the data to oversaturate the algorithm to help it make an optimal decision or don't provide the information at all, to ensure humans bias will not have a negative effect on the behavior displayed by the robotic system. For this reason, it is imperative that experts in the field set and provide the right framework and process to supply the data necessary.
Expert 1: *"If your AI algorithm is like a child, you can control the child it's fine… If the child is more powerful than you, it's smarter and uncontrollable, then you feed all this information that's not good, it's very dangerous".*
Expert 4: *"That is correct, basically we can consider an AI like a kid. So that's why the AI can be very, very adaptive".*
Expert 5: *"I think we are transferring a lot of biases and I'm not sure we want to prevent this…They are usually the fruit of engineers expectations who built it. When we are building machines that do certain tasks, there usually was an engineer who has done this machine to do this task in a certain way".*
Expert 6: *"I think the easier term would be oversaturated with other ideas so it would have different opinions".*

Human attitudes and engagement. In the future, we expect robotic systems to socially engage and display basic social etiquettes, especially does systems that will interact as a companion for elderly people. Positive attitudes, emotions and engagement towards

unanimated objects it is something we are experiencing already. We are already attached to our electronic devices, such as computers and smartphones. With the introduction of robotics systems, it will not be any different. Always context dependent, expert commented that it is very likely that humans will develop positive emotions, attachments and we will become very close to robotic systems. Interestingly, experts also expect humans displaying a different type of emotions towards robots from those human-to-human emotions.

Expert 1: *"I think [once] robots start evolving on their own, they will have emotions which we won't be able to understand".*

Expert 2: *"Yes, many people would like to be friends with a robot"*

Expert 3: *"Emotions displayed from Humans towards robots will be different, in the sense that it will be more positive towards robots".*

Expert 6: *"They [robotic systems] already are and will become closer".*

Human attitudes and negative emotions. Once we have available robotic systems showing and displaying human emotions, humans will undergo the uncanny valley process. Uncanny valley is a term which the American Psychological Association describes: "as a robot increasingly resembles a person, its familiarity increases until a point at which it abruptly drops to a negative value and elicits strong repulsion; then, as the robot's resemblance to a person continues to increase, its familiarity increases again and eventually reaches the level of a person"(Wang, Lilienfeld, & Rochat, 2015). In this context, as the technology continues to develop, it is very likely that humans will start to grow emotions, attachments and connections towards a robotic system. However, the emotions that humans could display towards robots can be different from those human-to-human emotions. In a less positive scenario, there are also negative emotions that could surface during the interaction with a robotic system. Also, context-dependent, humans could feel negative emotions towards a robotic system behavior when: the system does not show the right reaction or emotion correspondent to the context in which is interacting, due to hardware failure; that could potentially physically harms humans and the systems security, in the sense that it can be hacked and damage the privacy of its owner, influence or dominate their free will. These three scenarios are the examples where experts consider, humans could possible reject interaction with a robotic system due to negative emotions.

Expert 1: *"Human Beings takes [a] certain amount of time to get familiar with a new technology, they will have to go through that valley, it is called uncanny valley. Anything which is unfamiliar to us, we will take time". "Free will, they [humans] will not like*

domination. Any robot which controls you, is going to cause negative emotions to you". "Any robot which is nagging"

Expert 2: *"There's this word uncanny valley, I think. Since I have never experienced these kinds of systems, I can only imagine that this uncanny valley is true". "Assume people would be first kind of terrified". "All system that starts to kind of start influence their [people's] life".*

Expert 3: *"They [people] are going to be freaked out, because you really don't know how to deal with it". "We are going to feel a bit confused". "All of the sudden the robot starts asking super personal stuff". "Invasion of privacy"*

Expert 4: *"[Robotic Systems] Not showing the right reaction to the emotion."*

Expert 6: *"[Robotic Systems with] No proper management of the system"*

Human attitudes and negative emotions. In a very futuristic scenario, we expect robotic systems to engage with humans and display basic social etiquette. As a reflection of humans, is there a possibility that robotic systems will be capable of evolving by themselves? As the technology already exists, through reinforcement learning and machine learning, it is highly plausible that eventually it will reach this point. With this thought, comes a very popular fear: will robotic systems be capable of eliminating humans? According to the experts, this fear is founded by science fiction movies, TV shows and novels only. For them, the technology is far away from this point.

Expert 1: *"All these ideas are feed by Sci-fi writers".*

Expert 2: *"It has been a scenario in so many movies. But that's one influence definitely [from] Pop Culture". "Not so probable, but there is still a probability. It can happen, but I think it won't".*

Expert 3: *"It mainly [has to] do with all the matrix and all sci-fi movies".*

Expert 4: *"The fears, still in sci-fi movies. We are far from that, because we are far from understanding of all human behavior and human brain". "Still far from that".*

Influence of robotic systems on mental health. Experts agree that depending on the context, humans could exhibit positive and negative influence from the interaction with robotic systems. Even when the systems are designed to assist or improve humans' way of living; assisting the elderly for example, when users don't know how to use it, systems might create stress and frustration due to the unknown of the technology or paranoia; related to invasion

of privacy and dominance that could hypothetically, guided by AI applications, reveal its user information without prior authorization or change human behavior negatively.
Expert 3: *"It can cause stress related to job loss". "Some people might be afraid because they don't know how to use [it], and they are scare of the un-known".*
Expert 4: *"These systems are being designed to have a positive impact".*
Expert 5: *"If it's the person has some self-esteem issues, then of course, the robot taps into those problems within a person. Increasing anxiety levels in that person or causing paranoia if the robot knows too much, etc.".*

The level of autonomy in decision making. Context was the most recurrent and important factor expressed by the experts. The purpose and design of a specific robotic system should define the level of autonomy and its decision-making prerogative. With the level of technology available today, robotic systems dedicated at a very specific and narrow task can be trusted to have autonomy in its decision-making process. However, for critical and high-level decision-making process, humans should supervise and make the final decision. Moreover, majority of the experts indicate that robotic systems should be predictable. But they also acknowledge that predictability annihilate the whole purpose of creating a robotic system with autonomous capabilities.
Expert 1: *"I think Robotic Systems can be trusted for very simple-narrow-focus tasks ". "I think Robotic Systems will get better with time. They should be adaptively learning. But, when they start interacting and interfering with social life of people, there we should draw a line and we should have a layer of human supervision."*
Expert 2: *"When it comes down to high-level decision-making, like governmental decisions and really which comes down to emotions and these kinds of things then is humans. Humans are good at high-level decisions. But when it comes down to low level things, like picking up a package or drilling a hole, decisions like these are best made by robots".*
Expert 3: *"I would say it depends on context because context matter when it comes trust."*

Integration of robotic systems into humans' life seems to be the next step of evolution. In order to leverage this technology beyond current incorporation, there are key elements required for long term interaction with humans, one of these elements is the understanding of humans' attitudes and perceptions to adapt and train the systems accordingly with people's needs and expectations. However, progress comes with a cost and there is a possibility where robots may not achieve humans' expectations, generating an uncomfortable interaction and negative

attitudes for humans. In favor of securing humans' wellbeing in the future, efforts need to be directed into establishing frameworks and regulations to ensure proper learning and adaptation process for humans designing and engineering such technologies.

6. Discussion

Results of the study shows that generally speaking experts' attitudes towards robotics systems are positive. They evaluated positively the impact of robotic systems in human's daily life, especially looking into the future. Once on the market, users could expect that robotic systems will improve humans' ways of living. As mentioned by (Hockstein, Gourin, Faust, & Terris, 2007) more than a decade ago, these systems are helping us to enhance human capabilities. However, (Goolsbee, 2018) explains that, the degree of the negative consequences in economies is tight to how fast is the ability to adopt and integrate them into daily lives. As of now, this technology shows exponential growth and development. For this reason, immediate action on setting frameworks and regulation on algorithms, engineering innovations and other computational approaches is required, to ensure humans are ready to adopt it and its introduction will not have a negative impact in our economies or human's mental health.

Experts also agree that humans will develop positive attitudes, emotions and attachments towards robotic systems. Interestingly, they also expect a new type of positive behavior will be displayed towards these systems. Different from the behavior seeing in a human-to-human type of interaction. A similar explanation was presented in a research made by (Salah, Ruiz-del-Solar, Meriçli, & Oudeyer, 2012), where they found that humans behaved significantly differently when they interacted with robots, then when they did human-to-human.

Attitudes development requires a complex cognitive process of evaluation of objects, which changes over time and it is dependent on the individual's cultural background and experience, as explained by (Krosnick, Judd, & Wittenbrink, n.d.). As humans engage, trust and riley more and more in robotic systems, communicating and exchanging information. At the same time humans expect this interaction to be seamless. As (Gockley et al., 2007) indicated in their study that even though the robotic system presented non-anthropomorphic physical shape, participants expected a human-like behavior from the robot. However, nowadays and in the near future, it can generate uncomfortable interactions, because robotic systems do not resemble identical human cognitive characteristics, yet. Before technology crosses this gap,

experts expect humans to undergo an uncanny valley process. In (Bartneck et al., - (2005) research it was explained that contrary to social believes that Japanese love robots, the study revealed that Japanese were concerned regarding the impact these agents might have on society. The study speculates the reason being in part because Japanese understood what robots were lacking.

In regard to human's mental health, experts expect a positive and negative impact of robotic systems on humans. On one hand, the systems are designed and engineered to help and assist humans. For example, assisting the elderly. In different areas, such as psychology, AI is helping to improve decisions associated with the diagnosis, prognosis, and treatment of people suffering from mental illness (Gaggioli, 2019). However, some behaviors or situations from robotic systems could have a negative impact on humans' attitudes towards robotics systems. Example of these situations are: not showing the right emotion or reaction within the context its interacting, if the robotic system invades its owner privacy and presenting a level of malfunction, where its owner could be physically harmed. Similarly, to (de Graaf & Ben Allouch, (2013) study, where it was determined that an interaction with a robot has indeed an effect on people's behavior and anxiety.

To interact and engage with humans, robotic systems require autonomy and decision-making prerogatives. Experts make emphasis that autonomy depends on the context and the type of interaction with humans where the robotic system is performing. They also explain that since the technology is not at a high development level, it cannot be trusted one hundred percent yet. Specially in the cases or scenarios that involve high levels of decision making. Humans should always supervise and make the final decision. In (Goodrich & Schultz, 2007) research, it is explained that dynamic interaction includes time, task varying changes in autonomy and the degree of autonomy varies according to the capabilities of the robotic system. Experts also agree that robots reflect humans, where humans have created them to perform tasks they don't want to do or want them to be more efficient. In fifty to hundred years they expect robotic systems to resemble and behave like humans, work with and for humans. Similar to what is being fantasized by ancient societies. According to (Lee et al., (1982) humans have always fantasized about living with machines. Philosophers from China, Greece, Egypt and India fantasized with artifacts, myths and stories about the possibility of intelligent machines.

Finally, as prominent public figures like Bill Gates and Stephen Hawking (1942 - 2018), expressed concern about the possible dominance of robots taking over the world (Goolsbee, 2018). At the beginning of this thesis, the author formulated the question: Is this technological development a "Black Mirror", where humans are not only improving their way of living, but also mirroring or even magnifying the ugliest and darkest aspects of human nature? Results show the disagreement of experts; they express that fears of robotic systems eliminating humanity are based on believes result from sci-fi stories. In terms of technology development, robotic systems are still very away far from reaching this scenario. However, on the other hand, experts agree that humans should pay close attention to the **intentions of humans** behind the design and engineering of robotic systems.

7. Conclusions

The path of robotic systems has come a long way, from our ancestors' dreams and desires to smart agents nowadays mimicking humans, assisting in several tasks, performing in different environments and interacting with humans in different contexts. These robotic systems with smart capabilities are being developed at an exponential rate. With the objective to respond accurately to humans needs and expectations, not only they incorporate anthropomorphic characteristics, they also integrate artificial intelligence which is trying to replicate human cognitive capabilities. As humans incorporate these systems into society at an exponential rate, the degree of communication and interaction will increase at the same level. Therefore, it is fundamental to document humans' emotions, believes and perceptions that could impact, both positively and negatively human behavior in the future. In terms of robots and humans' attitudes, experts positively evaluated the impact expected from robotic systems on humanity in the future. Allowing humans to free themselves from jobs that will not be of interest anymore or by helping humans to become more creative and develop skills that haven't been seen yet.

Even when experts are highly optimistic about the future. At the very beginning of new technology development, its adoption and integration by society has a gap. In the case of robotic systems with anthropomorphic characteristics this gap is called "Uncanny Valley"; in principle the technology is scary, but once these systems resemble more like humans and humans get used to it, the technology will be easily integrated. In order to improve it rapidly, feedback from users will be needed, as it is done today with other technologies. After market release, the iteration process of feedback and improvement will make these systems precise and adapted to

its market. On a more sci-fi note, for those who fear robotic systems taking over the world and eradicating humanity, experts agree that this technology is far away from becoming a threat and over all humans should not be afraid of robotic systems. However, humans should be vigilant of other humans' purpose and the human that's behind its design and engineering. This could represent a real threat, that could tap into security systems, invade privacy, manipulate algorithms and impact negatively human's behavior.

By adding value to our current understandings of human attitudes towards robots from experts' perspective, this investigation hopes to contribute to the Human-Robot interaction field of study, that is working towards the improvement in design, engineering, and evaluation of robots. However, due to small sample size the result cannot be generalized. The topic represents a challenge because it is a novel area, that involves the complexity of human attitudes, perception towards objects that are in current development and transformation, cultural background and previous experience with robotic systems. For this reason, more research with a major number of experts from other areas of study and different institutions needs to be done. Due to the adaptive nature and usability of robotic systems, major efforts need to be put into place, in order to ensure that humans bias will not have a negative impact on the behavior of the robotic system. Policymakers, private companies, academic institutions, scientist and experts need to build the frameworks and regulations to prevent access and manipulation to robotic systems from humans with harmful intentions.

8. References

A Time Line of AI. (2016). *MIT Technology Review*, *119*(3), 62. Retrieved from f5h.

Bartneck, C., Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2005). *Cultural Differences in Attitudes Towards Robots*. 4.

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, *43*(1), 3–31. https://doi.org/10.1016/S0167-7012(00)00201-3

Bubaš, G., & Lovrenčić, A. (n.d.). *Implications of Interpersonal Communication Competence Research on the Design of Artificial Behavioral Systems that Interact with Humans*. 6.

Buchanan, B. G. (n.d.). *A (Very) Brief History of Artificial Intelligence*. 8.

Catherine Clifford. (2018, March 13). Elon Musk: ‘Mark my words — A.I. is far more dangerous than nukes.’ *BBC*. Retrieved from https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html

Coppin, B. (2004). *Artificial intelligence illuminated* (1st ed). Boston: Jones and Bartlett Publishers.

de Graaf, M. M. A., & Ben Allouch, S. (2013). The relation between people’s attitude and anxiety towards robots in human-robot interaction. *2013 IEEE RO-MAN*, 632–637. https://doi.org/10.1109/ROMAN.2013.6628419

Ewing, D. R., Pigazzi, A., Wang, Y., & Ballantyne, G. H. (2004). Robots in the Operating Room The History. *Surgical Innovation*, *11*(2), 63–71. https://doi.org/10.1177/107155170401100202

Gaggioli, A. (2019). Top three technology trends to watch in 2019. *Cyberpsychology, Behavior, and Social Networking*, *22*(1), 97–97. https://doi.org/10.1089/cyber.2018.29137.csi

Gockley, R., Forlizzi, J., & Simmons, R. (2007). Natural person-following behavior for social robots. *Proceeding of the ACM/IEEE International Conference on Human-Robot Interaction - HRI ’07*, 17. https://doi.org/10.1145/1228716.1228720

Goodrich, M. A., & Schultz, A. C. (2007). Human-Robot Interaction: A Survey. *Foundations and Trends® in Human-Computer Interaction*, *1*(3), 203–275. https://doi.org/10.1561/1100000005

Goolsbee, A. (2018). *Public Policy in an AI Economy* (No. w24653). https://doi.org/10.3386/w24653

Hockstein, N. G., Gourin, C. G., Faust, R. A., & Terris, D. J. (2007). A history of robots: from science fiction to surgical robotics. *Journal of Robotic Surgery*, *1*(2), 113–118. https://doi.org/10.1007/s11701-007-0021-2

Krosnick, J. A., Judd, C., & Wittenbrink, B. (n.d.). *The Measurement of Attitudes*. 56.

Lee, G., Lelouche, R., Meissonnier, V., Ornato, M., Zarri, G., & Zarri-Baldi, L. (1982). Artificial intelligence, history and knowledge representation. *Computers and the Humanities*, *16*(1), 25–34. Retrieved from eoah.

Maguire, M., & Delahunt, B. (2017). *Doing a Thematic Analysis: A Practical, Step-by-Step Guide for Learning and Teaching Scholars. 8*(3), 14.

Nilsson, N. J. (2009). *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. https://doi.org/10.1017/CBO9780511819346

PR Newswire. (2018). Robotics Market Forecasts. *LON-REPORTBUYER*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bwh&AN=201805280948PR.NEWS.USPR.SP07957&site=ehost-live

Rajnai, Z., & Kocsis, I. (2017). Labor market risks of industry 4.0, digitization, robots and AI. *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY)*, 000343–000346. https://doi.org/10.1109/SISY.2017.8080580

Rantanen, T., Lehto, P., Vuorinen, P., & Coco, K. (2018). Attitudes towards care robots among Finnish home care personnel - a comparison of two approaches. *Scandinavian Journal of Caring Sciences*, *32*(2), 772–782. https://doi.org/10.1111/scs.12508

Rocca, R. D. (2018). Will Insurance Claims go Digital in 2019? *Claims*, *66*(12), 35–37. Retrieved from bth.

Salah, A. A., Ruiz-del-Solar, J., Meriçli, Ç., & Oudeyer, P.-Y. (2012). Human Behavior Understanding for Robotics. In A. A. Salah, J. Ruiz-del-Solar, Ç. Meriçli, & P.-Y. Oudeyer (Eds.), *Human Behavior Understanding* (Vol. 7559, pp. 1–16). https://doi.org/10.1007/978-3-642-34014-7_1

Syrdal, D. S., Dautenhahn, K., Koay, K. L., & Walters, M. L. (n.d.). *The Negative Attitudes towards Robots Scale and Reactions to Robot Behaviour in a Live Human-Robot Interaction Study*. 7.

Trejo, K., Angulo, C., Satoh, S., & Bono, M. (2018). Towards robots reasoning about group behavior of museum visitors: Leader detection and group tracking. *Journal of Ambient Intelligence and Smart Environments*, *10*(1), 3–19. https://doi.org/10.3233/AIS-170467

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, *19*(4), 393–407. https://doi.org/10.1037/gpr0000056

Yanco, H. A., & Drury, J. (2004). Classifying human-robot interaction: an updated taxonomy. *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, *3*, 2841–2846. https://doi.org/10.1109/ICSMC.2004.1400763

Zamalloa, I., Kojcev, R., Hernández, A., Muguruza, I., Usategui, L., Bilbao, A., & Mayoral, V. (2017). Dissecting Robotics - historical overview and future perspectives. *ArXiv:1704.08617 [Cs]*. Retrieved from http://arxiv.org/abs/1704.08617

9. Appendixes

APPENDIX I: Taxonomy types of human interaction

Type of taxonomy	**Categories**	**Description/Example**
Time-Space	Synchronous	While using the computing System at the same time
	Asynchronous	While using the computing System at a different time
	collocated	While in the same place as the computing system
	Non-collocated	While at a different place as the computing system
	Audio	N/A

Collaborative Application: mode of communication	Visual	N/A
	Document (data)	N/A
Multi-robot team	Multiple agents	As tasks cannot be completed by a single robot
	Multiple agents, but limited interaction	N/A
	Single agent	Additional agents do not add to the speed or efficiency of the solution
Task and reward of a multi-robot group	Time	How long the task is allowed
	Criteria for measuring performance	Time horizon for optimization
	Subject of action	Robot movement or object movement
	Resource limits	Power, intra-team competition, external competition
	Group movement, and platform capabilities	Task can be performed by a single agent requires multiple agents, can observe relevant features or the world, only can get partial information, and requires communication
Task Type	Urban search and rescue	N/A
	Walking aid	N/A
	Delivery robot	N/A
Task Criticality	High	Urban search and rescue
	Medium	Hospital delivery robot
	Low	Robot soccer
Robot Morphology	Anthropomorphic	Human-like appearance
	Zoomorphic	Animal-like appearance
	Functional	Neither human or animal like appearance, but it's related to the robot's function
Ratio of People to Robots	Human-Robot-Ratio	If the number of operators or the number of robots is variable within the system, a range may be specified.
Composition of robot teams	Homogeneous	N/A
	Heterogeneous	N/A
Level of Shared Interaction Among Teams	One human - one robot	Human giving command to one robot
	One human - robot team	Human controlling a group of robots (same area)
	One human - multiple robots	One person directs two robots that work independently
	Human team - one robot	People coordinate among themselves to issue one command to the robot
	Multiple humans - one robot	Humans act independently and send different commands to the same robot.

	Human team - robot team	Humans agree on one command that the robots then coordinate on to decide what robot(s) carry out what part(s) of the command
	Human team - multiple robots	Humans agree which command should go to which robot, and each robot acts independently to fulfill the command
	Multiple humans - robot team	The robots deconflict and/or prioritize the different commands as well as divide the commands among themselves prior to carrying them out
Interaction-Role	Supervisor	Human supervise the robot, but does not control it
	Operator	Human teleoperates or changes the robot's behavior
	Teammate	Human works with a robot to accomplish a task
	Mechanic/Programmer	Human needs to physically change the robot's hardware or software
	Bystander	Does not control the robot but needs to understand it
Human Robot Physical Proximity	Avoiding	Modes of physical proximity between humans and robots, the values are ordered from less to more physical interaction
	Passing	
	Following	
	Approaching	
	Touching	
	Non (physical-proximity)	
Decision Support for Operators	Available-Sensors information	Is a base line for understanding provided-sensors
	Sensor-Information provided	Is a list of sensing type
	Type of sensor fusion	Is specified as a list of functions
	Pre-processing	For decision support
Autonomy Level/Amount of Intervention	Fully controlled	A human must be controlling the robot
	Shared control	The robots do some part of the task and a human operator must do some part of the task
	Full autonomy	The robot has near 100% autonomy

Table 1:Types of taxonomies of human-robot interaction. Source: Yanco & Drury, 2004.

APPENDIX II: Questionnaire

Robots and Humans: Attitudes towards Black Mirror Effect

Interview Questions

Context: Imagine one day we have created robotic systems with anthropomorphic characteristics, capable of interacting with us at a human level, with full cognitive and communication capacities and skills; robotics systems that react and respond to our intentions within our actions, make use of resources and behave in unexpected ways.

Please answer the following questions:

To give context and preamble, let's start with General related questions:

1. How do you think robotic systems will change humanity in future?
2. What do you think about the speculation that by "feeding" an AI algorithm with all the data available, including human behavior attitudes, opinions and judgments without properly filtering the data we are also transferring humans' negative attitudes?
3. What do you think to what extend we are transferring humans bias to the robotic systems? How to prevent this?
4. How to prevent hostility, racist and other discriminative attitudes from a robotic system?
5. Whom do you think people should trust more: robotic systems or humans? Why? How much should people trust robotic systems?
6. How reliable do you think robotic systems will be? How much one should trust a robot's judgement and suggestions?
7. Considering, that as young person, a child is still learning. How ethical is it to let young children interact with a robotic system? How probable is that a child could learn inappropriate behavior from a robotic system?
8. Do you think AI will be capable of evolving without human interference? How much RS are a reflection of humans?
9. How much do you think one day humans will depend on robotic systems to function? As we do now with computers and the internet. Should we prevent humans' dependency on RS? If, yes. Why and how?
10. How to ensure robotic systems' verbal and not verbal communication capabilities are adapted to different cultures and social settings?

Thanks to scientists and developments in technology, we have been recreating stories from ancient societies and science fiction writers. As these sci-fi stories are becoming a reality. Is it possible that our fears will also become real?

Emotions related questions:

1. How do you think people will react when robotic systems display real emotions?

2. How likely do you think is it for people to develop emotions for robotic systems? How similar these emotions would be compared with emotions displayed towards people.
3. How likely will be that people would like to be friends with robots?
4. What kind of robotic systems' behavior could make humans feel uncomfortable and could made humans feel negative emotions?
5. What do you think could make humans scare while interacting with robotic systems?
6. What kind of influence will robotic systems have on human's mental health, anxiety, negative attitudes or paranoia?
7. What are the biggest fears regarding RS, humans should be aware of? (3 words)
8. What do you think about the fear that robots wipe out humanity? Why do we think that the fear of robotic systems turning their back on humanity is so popular?
9. How probable is that robotic systems could represent a threat to humanity? What are your personal fears regarding advance robotic systems? (3 words).

Interaction related questions:

1. How autonomous should robotic systems be?
2. How predictable should robotic systems be?
3. How socially engaged robots should be?
4. How close the emotional interaction between robots and humans should be or could become?
5. How robots should behave to overcome an unexpected situation where humans are involved?

APPENDIX III: Thematic content analysis themes.

Positive impact of robotic systems on humanity

Robotic Systems are adding an addition layer of intelligence. But, if you add robots, it's going to combine many different levels of intelligence and appendix to human intelligence. Will help in evolving human beings to the next level It will be a superhuman, superhuman as in we'll be extraordinary humans. I think they are going to replace a lot of tedious human jobs. The impact is rather positive. We will become cyborg meaning literally kind of Might become RoboCop, Sex robots. But then that's as an advantage that you could use them as a healthcare system Taking care of the elderly. Robotic Systems will start to become partners, or friends or kind of

mates at work with humans, this is at least what I am expecting to see. One way I can think of, as many machines along the history has made lives easier in the sense that we don't have to do hard labor. I do hope that when we are creating machines, when we also have humanoid robots, that these will make our lives easier. Near future wise they will start improving little things because it's going to be integrated very slowly. The far future this could probably going to be integrated with general use case neural networks which will be more like an actual intelligence.

Robotic systems and Biases

They can be both, positive and negative. As long as you can understand how your AI application works, how it thinks, it's okay to feed all this information of human behavior. If your AI algorithm is like a child, you can control the child it's fine. If the child is more powerful than you, it's smarter and uncontrollable, then you feed all this information that's not good, it's very dangerous I think that intelligence emerges from learning from experience which is individual like a child. Yes, absolutely. That is correct, basically we can consider an AI like a kid. So that's why the AI can be very, very adaptive if it goes to the hands of dangerous people. We are transferring negative attitudes, that have been proven or demonstrated at least with several AI algorithms. It is a high risk, but in the end, we are developing these algorithms as we speak, and we are only going to deploy those algorithms that we can control. This is something that someone has already tested. Google and a couple of others have created a couple of chatbots, and the idea was to test the Turing test. First of it worked fine, but then it was released to the wide audience on the internet and of course there were people that would like to abuse this kind of things. It is very, very likely that children will pick up inappropriate behavior from the RS, Robots are going to get into our fabric, so it’s very important they should not exhibit inappropriate behavior. I see no ethical conflicts as long as the robot itself is not biased towards those behaviors you mentioned before. So, it is as probable as I mean yeah children learn from everything, if the robot has those behaviors, bad behaviors, then for sure will learn, get influenced. so that's definitely a source of influence. There is no doubt in that. I see no problems; we shouldn't keep children away from innovation and technology. If the designer who came up with the program Was a weird guy and he wanted to Do something inappropriate then yes, Red robot itself with do inappropriate things. As a company do, we want our customers to do I complain like that, We wouldn't. I would say is good to have a Robot around a baby and a person. If the robot has learned good things, it just like it will act good things, then it will be an excellent example for a child. If the robot has learned bad things it, so it is a bad example. That will actually have a negative influence on the creativity of the

child, because when a child is playing with a passive toy the child has to come up with sentences that the toy is saying, now if there is robot able to generate those answers, the child, that part of the child's creativity is being block or at least not being enhanced. Having those RS that are able to interact with the child might instead trigger some type of creativity in the child that we can't even think of right now, because we never had those type of toys. And the other side is that we know from the experience of autistic children therapy, that children are not able to differentiate between emotions expressed by robot or a human. Because, for children emotions are what they appear to be, not what weather they are true or not. If a child interacts with a robotic system at a young age and it doesn't have a parent nearby, they will develop thing that might seem morally ambiguous, because robots cannot teach morals, something that is a human concept, only humans can teach it. This is a very edge case example, but autistic children they are very keen and know specific things, they are very skilled. at least I don't know how a computer will [be] social in the same way a human is there are these small nuances that a kid won't pick up and once the kid is transfer to another group of humans. Yeah, we are doing it a lot right now. They are very narrow AI. We need something strong, it's called strong AI, where it's more human like. For that reason, I feel is not yet ready. if you have some biases don't necessarily mean negative things right. Rather is how to make sure we give proper biases. Some committees through selection of wise people elected by the other people, just like governments we had some so some committee who then we'll vote which kind of behaviors are ethical and which are not. They don't really put bias into it they just miss some factors into account. we are saying, let's not give the negative data so we could safeguard ourselves, so the robot doesn't go complete it hardwired, and in a sense yes, we should do that, because is a machine is not a human, and if we plan to use machines as machines, then just give the machines the good data that's out there, and don't let the machine try to be a human because is not a human. We are aware not of how much of human bias is being transfer, because a human is transferring that. The human can be bias… we are thinking it is not bias so, I cannot give you a number. I think it exists. Theoretically speaking, if many people with different backgrounds try to aim the same thing, eventually then average of the whole thing will be a bit un-bias. I think we are transferring a lot of biases and I'm not sure we want to prevent this. A lot of what machines are today is a lot of what has built for them to be built. the machines are not the averages of all the human's expectations. They are usually the fruit of engineers' expectations who built it. when we are building machines that do certain tasks, they are usually was an engineer who has done this machine to do this task in a certain way. In that way there is subjective bias of the machine builder already in the machine. I don't think is necessary a

problem, because when something is made into a product, it will go through market research, there will be a market for it, that means that there will be people that wants to buy this machine, that one to use this machine and they will start driving the way the technology with its newer generation, much as we see with smartphones. Is hard to [inaudible] because obviously the person who is going to design the RS, they are going to have their own specific biases, but the bigger problem is data collection when you are trying to train something because depending on how and when the data is collected, that data is going to have bias and if the neural network is trained in that bias, obviously is going to have it and there is no way around it. So, in that sense, unless you feed the biases into the robot, it's not going to have any bias. You can take care and remove biases. Through a committee of selected wise people, first thing. And other thing means if you are going to teach the system, so you need an algorithm don't give it information about these kinds of features; skin color, sex, age. This also keep the system blind, but we might end up with useless systems maybe we don't know. Apply Asimov's rules. He made a mistake he should have given more data to it. Again a robot is an AI agent is a kid, if from the beginning you provide the data, you do not put any racist information or something that cause this kind of problem, the this kid, the AI grows up and it learn, it will never learn those kind of things. If you want to prevent these things you just don't feed it to the system. With robotic systems it should be rather straightforward, because every machine has constraints. But with the machine we simple program it to not run into a wall, and if it runs into a wall it has malfunction it needs servicing. I think the easier term would be oversaturated with other ideas so it would have different opinions.

Human attitudes and engagement

I think robots start evolving on their own, they will have emotions which we won't be able to understand. In order to answer this question, we kind of need to know what the emotions are. Hard for people to develop emotions, real genuine emotions that we don't have any doubt in it. With the robot for sure, we're going to have a different mask probably. We're going to be imperative saying trying to tell the robot what to do, rather just have a good discussion. so, but it will be different, yes. This is already happening. Emotions displayed from Humans towards robots will be different, in the sense that It will be more positive towards robots. Adults smile and children pet them as it was a dog or a cat. I'm working on affective computer, and affective computing is an important part of the robot and human interaction, and the aim is we are making these robots or AI agents, act in a way that you are expressing exactly the same emotion, that you are expressing to me as a human, that I if I was a robot. I think it's very easy to get attached

to machine and develop feelings for a machine. Humans are kind of weird in that sense, they develop emotions towards inanimate objects and in case of robots, it will make the creation of the bond more simpler. It thinks it will continue to happen, because it has happened in the past. It can become very close. It's already happening. I think robots will take it to the next level. Context dependent, but some people could get really attached. It should be very much emotional oriented. We should look the context. It's emotional connection, but emotions are playing a very important role. I think that emotional interaction can become very close, because I think humans have a tendency to be attached to machines. They already are and will become closer. don't think there should be a limit, some people wanted one way and some people wanted it other way, and some people want to limit how other people will interact with it. I think robots should accelerate the interaction. They should accelerate social behavior. You want the robot to be smooth, ask those basic social questions and then, go. Very much. if the answer is 100 years later, yeah, they should be quite social. But in another form right now the robot is developed for one specific task. There are too many questions that need to be answered, before we decide it has to be social or not social. So, I cannot really comment. If the robot is a social robot then yes it needs to be socially engaging. Again, when we're talking say companion robots for elderly people, who otherwise would be alone with those robots engaging and stimulating, but then when we're talking about say deliver robots, we want them to be as transparent as possible. Social interaction is something I think humans should still dominate. There are a couple of factors where robots can automate; Automatic tellers, automatic care for older people. I would not recommend automation for child development. But automatic control systems that interactive. Definitely they will be. Yes, many people would like to be friends with a robot. They already have girlfriends they want to have relationships. Why not friend, yeah for sure. Very likely, until very bad events happen. I can see what it means to have feelings for a machine which in the end is like a one-way system where a human is having affection towards a machine and the machine is emulating whatever the human wants in response. Very likely.

Human attitudes and negative emotions

Free will, they [humans] will not like domination. Any robot which controls you, is going to cause negative emotions to you. Any robot which is nagging. Also, if robots know that it is competing with a human and HBs will not like that robots. There is unhealthy competition between humans and robots. If your spouse our you close friend falls in love with the robot and picks the robot over, you. All system that starts to kind of start influence their life. [replacing

humans on jobs, where humans are very much capable of resolving the task] [Creating a competition where humans are in disadvantage]. All of the sudden the robot start asking super personal stuff. Invasion of privacy. Not showing the right reaction to the emotion. if the robot, start a task, which is not desirable by the people around will cause some uncomfortable things. Anything that Robot can do, or a machine can do the taps into whatever that person is afraid of. Physical attack. Privacy. The machine being better at humans. Uncanny Valley. Dominance and influence. Unexpected behavior. Replacing human interaction and loosing social skills as consequence. Security <> Hacking. Quality > Unexpected behavior [not showing the right reaction of emotion] Privacy. Malfunctional. Power > physical and conscience. Unintentional fatal errors > Malfunctional. I think we need to be cautious about too quick development. I guess the world order and the security, if we have these machines that help us out in tremendous ways, are these machines secure enough, if say somebody wants to use them against us. here is a potential for mal use of Technology by other people. No proper management of the system. Overdependence. Something more powerful than you. if you experience such a system which is unreliable behaving weirdly then yeah it's easy to be scared of it kind of a system which does moves unexpectedly or just makes unexpected decisions which impact you in a physical or mental way, Physical level is am I going to get hurt [due to hardware malfunction]. Invasion of privacy. Influence and domination. I think if we are sure of a decision and the RS gives a different decision or makes a decision which for us is immoral. [Hardware malfunction] That could harm humans physically. Immoral decisions suggestion, [could be related to influence and dominance] if a machine is doing something that is unexpected. If a robot has power over a very dangerous system and is faulty in some way. its accuracy isn't that great, or it isn't properly trained or maintained, that will make humans very cautions of it. So, HB takes certain amount of time to get familiar with a new technology, they will have to go through that valley, it is called [uncanny Valley]. Anything which is unfamiliar to us, we will take time. There's this word uncanny valley, I think. Since I have never experienced these kinds of systems, I can only imagine that this uncanny valley is true. assume people would be first kind of terrified. They going to be freaked out Because you really don't know how to deal with it. We are going to feel a bit confused. In Psychology Psychopaths don't have any emotion. But they can understand other people's emotions really well … [Robots will be able to do the same]. They will be very excited. Weather this excitement will continue, I don't know. we are waiting till one robot kill one innocent with a smile, then you will start to feel creepy, I guess. I don't think people will be able to tell the difference between a simulated emotion and a real emotion, much like even with human-human interaction. It will definitely bring up a lot of moral debate on

whether a computer can have a soul or be conscient. But in my opinion, this is something that it might not actually happen, because that would only happen if we try to model a computer exactly like a human. With the AI that I think will come out of this, no one will consider it having a soul. All these ideas are feed by Sci-fi writers. Sci-fi writers got these visions. These novels are very popular, and they are been feed to the human being in so many different ways, in movies, in TV series, in books, in comics, in technical writing. People always talks about fiction. It has hypnotized humanity to make it real, we are all hypnotized to believe this. We are following them subconsciously; we should not let that happen. That's the reason we have these discussions. We don't have to hypnotized by these authors. It has been a scenario in so many movies but that's one influence definitely Pop Culture. I mean likely in terms of if we create such a system but I don't think that's going to happen, because there are lot of people who know stuff, who are involved in developing the systems and we are trying to take care of steps in legislation in and then all kinds of things so that's not a reasonable fear, but it is a possibility. It mainly does with all the matrix and all sci-fi movies it more about that. People fear new technology [because we don't understand it]. The fears, still in sci-fi movies. We are far from that, because we are far from understanding of all human behavior and human brain. Is popular because bad news sell. But if you talk to engineers, they're not worried about this at all, you work with a robot in a lab and you think you can pull the plug anytime, It's a popular science fiction idea. But as reality I don't think we'll have a turning point where there is a robot revolution.

Influence of robotic systems in mental health

Impact all these areas, both positively and negatively. It could also cause influence and dominance. Making and changing life because of AI guided applications. I'm sure there's a fine. balance when it comes down to bringing those robots in our society. this relates to for sure with mental health [it can cause stress related to job loss]. it should be you related to our Mental Health which means that innovation of Technology to body system would have a good influence on our well-being. These systems are being designed to have a positive impact. I think it depends again more on that person as I mentioned in the mention before, if it's the person has some self-esteem issues than of course that the robot that taps into those problems within a person again increased anxiety levels in that person, cause paranoia if the robot knows too much, etc. but at the same time the robot can make the Human feel at ease by solving problems that the humans was unable solve. Paranoia is something that's always going to be there.

The level of Autonomy in Decision Making

I think RS can be trusted for very simple-narrow-focus tasks But, for surgeries, where you are deciding what to do with the patient. You should not trust a robot; you should have another human being. When it comes down to high-level decision-making like governmental decisions and really which comes down to emotions and these kinds of things then is humans. humans are good at high-level decisions, but when it comes down to low level things like picking up a package or drilling a hole, decisions like are best made by robots, how much should people trust for what system. I guess it depends on the context I would say it depends on context, because context matter when it comes trust. the context matter, who designed and what for. And you can still trust the computer, because it did do its job, it accomplished it. I think it depends a lot on the context. there are certain tasks, where we don't at least know whether they are as capable as humans, but this domain continues to shift, there are things that we wouldn't trust the machines 20 years ago RS depending on the use case, if there something that looks over and very strong based on statistical outcome then you should always trust the computer, even though it might seem is making the wrong choice. Because computers try to get the optimal choice, they don't try to get the best one. By best, I mean human best. For very simple tasks they can be fully autonomous. Autonomy should be carefully judge against the training, Autonomous for very specific task. If you take it up the next level, context dependent completely. Level 4 - you could go with supervisory. I would never put full autonomy. Maybe for repetitive task, you can go full autonomy. [for specific tasks can be autonomous] but, I think for the critical decision making there should be, there must be human thinking. we usually want the machine to be very autonomous at the task We're trusting the machine with. We keep trying to give more autonomy all the time, the more give it the better they will be, because they will have quicker reaction time and better decisions. The only problem is that technology hasn’t reach that point yet. Some people are trying to push autonomy further than the accuracy can actually handle. I think RS will get better with time. They should be adaptive learning, But, when they start interacting and interfering with social life of people there, we should draw a line and we should have a layer of human supervision. Depends on the design - context - Purpose Who designed it. They are precise but sensors can go wrong, there is always faulty sensors. Are robotic systems reliable, yes. but do sensors fail, yes. Common sense [Human supervision] Think as robot more like is a tool, Eventually, we are hoping it will be 100% reliable, so that people can trust it. Right now, they are sometimes reliable in very, very, very, very specific tasks. I Think if you want to make a critical decision it should be finalized by

humans. no matter what, if a robot makes a decision but it is supervised eventually by humans if is going to be a critical decision. Reliability again is something that it is constantly improving, once a system becomes reliable in a certain task, people will start trusting it and the more people will start trusting it, the more reliable the system will become, because more people are using and more people are developing it. we can expect that certain recommendations will be more accurate, it will get at times even too accurate, even more accurate that we want to admit Reliable is a hard word, because depending on the system. Sometimes a human is more reliable, but as time is moving forward robotics seems more reliable than humans and a lot of data samples nowadays.

APPENDIX IV: Transcripts

Expert 1

General related questions:

Q1: I think RS are adding an addition layer of intelligence, so for example If you look at primates like orangutans, they are not that cognitive in terms of thinking and introspection. If you look at insects, they are more instinctive. They are more, I would say collective in their actions. So, there are different kinds of intelligence. But, if you add robots, it's going to combine many different levels of intelligence and appendix to human intelligence, so I think robots to our ecosystem will help in evolving human beings to the next level. When I say next level, It will be a superhuman, superhuman as in we'll be extraordinary humans. That's what I would say.

Q2: Yeah, I think that right now AI algorithms are not fully understood how they are evolving. I feel the research on AI algorithms continuous to happen the ways it is, we are going to develop a multi-dimensional intelligence. A multi-dimensional intelligence which can feel, which can be creative. So, when the multi-dimensional intelligence is creative. I think, feeding these humans inputs is going to cause unexpected responses. So, they can be both, positive and negative. The thing is, as long as we, the outcome of these AI algorithms is predictable, that's when it's safe to use and it is encouraged to feed this information. Once these responses are unpredictable, for example: in the Game of AlphaGo, the moves made by that machine, which beat Lee Sedol, they are unpredictable, nobody even imagine how the AI algorithm was able to think about that move. That move was several layers higher than humans' dimensions of thinking. So, that's when we think about Singularity, and that's when these robots and AI [driven] applications will start causing harm or good. It can [go] either ways. So, when we don't know what the algorithm thinks is good, but if you don't understand the algorithm, like a child.

If you know that child ability [inaudible] you can control the child. If you don't know what abilities the child has, how are you going to control it? So, I think is a, as long as we have a [hang]of the how these algorithms will operate, can control, we can okay, we can be safe. The last sentence is, as long as you can understand how your AI application works, how it thinks, it's okay to feed all this information of human behavior. If your AI algorithm is like a child, you can control the child it's fine. If the child is more powerful than you, it's smarter and uncontrollable, then you feed all this information that's not good, it's very dangerous

Q3: Yeah, we are doing it a lot right now, because AI algorithm are their inception and AI algorithm are not [casual, casual] means they are not realistic, they are not in touch with reality. Humans beings can reason can reason what is real and what is not real, but because we are able to see what is the cause of something, what is the effect of something, this is called causing-reasoning , that has not been fully implemented in machines, it's a very niche field. So, I see [...] AI algorithms, are not [casual], they are very narrow AI, it's narrow AI. We need something strong, it's called strong AI, where it's more human like, especially when you are passing judgments on civil policies or controlling traffic or predicting crime and preventing crime. In these places where you are nailing people. You go to curt or you find them it’s very important that it has to be [casual], for that reason I feel is not yet ready, it's not yet happening, there is a lot of biases in data driven algorithm, as we use today.

Q4: There is a social robot called [Sin Sensei], she is used as [Bolshoi] Chancellor. what she does is, she listens to people's problems, [she does in dimensions], gives them phycological counselling. That kind of robot is not having any bias, as long as you are able to communicate in English, it's understandable. She is going to help you with your problem. So, in that sense, unless you feed the biases into the robot, it's not going to have any bias. So, when [very applications] specifically, you can take care and remove biases. So, that is possible to do. It is possible to prevent. As of now, I think algorithms which are working on the internet, collecting data. They need to control bias. For example: Facebook data is going to have a lot bias, because of people behavior, for instance. If the algorithm is learning from there is not the best place. We need to add another human layer to it, which will remove the bias. That's what I suggest.

Q5: I think RS can be trusted for very simple-narrow-focus tasks, for example; if they have to decide on credit history, there is a straight forward mathematical formula which decides credit history, which does not go by race and gender, it's a math formula. And that can be used, we can trust robots there. But, for surgeries, where you are deciding what to do with the patient. You should not trust a robot; you should have another human being. Because surgical conditions are very complicated, and a lot of issues witch a doctor decides on, they are human

factors. Sometimes, the patient is stress and they go the doctor, they have a pain or something. So that time you have the chancellor and you remover the pain away versus giving them medication. So, this kind of decisions should not be giving to the robot. That should be giving to humans

Q6: I think RS will get better with time. They should be adaptive learning, it should not have fixed algorithms, especially social robots. But if you have robots assembling in a manufacturing line, I think they can be hard-wired. Robots come for different porpoises. So, as long as they are focused to a narrow task, where they are not causing danger to anybody you can trust them. But when they start interacting and interfering with social life of people there, we should draw a line and we should have a layer of human supervision. I think all robots, all over the world, should have human supervision in a very top level. No robots should be allowed to be super evolved than human intelligence. When they become smarter than human's intelligence and decision making, that kind of governance should be mandated o moderated by human layer of supervision. It should be done by the top level of ethics and robotics experts and there should be a World governance body, which is not subject to any political pressures or any political intentions

Q7: How probable is that a child could learn inappropriate behavior from a robotic system? It is very, very likely that children will pick up inappropriate behavior from the RS, because children are going to co-evolve with the robots. So, this generation-Z, isn't called generation-Z? Because the Genex or Gen Z. They are going to use the law of accelerating returns in their evolution. Law of accelerating returns is you are standing on the shoulders of previous research work. You are not going to do research grown up [?]. So, as a child you are standing on giant shoulders. So, when you grow up. For example; when are growing as a child, if you understand, how to interact with robots [inaudible], then you are going to evolve, co-evolve with the robot. It's going to be much more than anyone can imagine. Robots will start becoming a social layer, will become a necessity, like a cellphone network or a Wi-Fi network. It's in the fabric of society, like that. Robots are going to get into our fabric, so it's very important they should not exhibit inappropriate behavior. For example; killing any animal in front of a child, a robot should not do that. Even though, you have a robot chef, if [sluttering] it should be done in a [sluttering] facility, not at home. You should not have such kind of robots, which kill the animal at home. A lot of people cook fish, octopus and chicken at home, right. If there is cooking involved, where there is [sluttering], the child should not be subjected to that. They will think that is normal, going and stuttering. So, does kind of things and hate speech. Hate speech should never done by robots, because robots will feed that to the child. So, there are a lot of movies,

which are, children see. There is some degree of violence in that. We are exposed to that to some level, there are a lot of violent video games, which many kids watch. They might not educate, but they still have access to the app. They might not be going to school. This [inaudible]. They will grow up to be violent, they will get involved in criminal activities, because video games are already doing that. If Robots do to that, and if the robot Nany are Social robots inside the house it's a very bad example. That should be a coming on the governance, there should be a robotic governance

Q8: I think Ai systems. When a system is evolving initially it takes a lot of time and resources for training. So, once it reaches a suitable maturity there will be very minimal training required. And, if the system is adaptive it will learn, it's going to learn, it will evolve independent of your interference. So, at one-point AI, will evolve without human interference. That's what I am saying it should not happen. Robotic governance should be very careful about that, because once you cannot understand how the AI is evolving you won't be able to predict. I think is already happening, in AlphaGo, the move which beats Lee Sedol, that move was from a very multi-dimensional capability of the algorithm. Something which even the people who built the algorithm, they could not explain how the algorithm did this. It's happening in lots of reinforcement, training examples. People don't know why a certain move was made to win the game. Unpredictable moves to win the game. So, the AI is very capable of evolving and this should be taught in schools and colleges. That this is where you stop, you don't mess with it anymore, because you don't understand it. You don't mess with it. I think RS. Human beings come at very layers of cognition. First level is sensing, then [I think] there is survival instinct. This is very basic level in which human beings have developed. Then the next level is coordinating tasks and doing simple tasks. Robots are out there, coordinating and doing simple tasks. The next level is being creative and thinking. We are not yet reach there. Robots are a reflection of Human beings, up until coordinating and doing simple tasks. That's where they are, so they are about 30-to-40% of where human beings are, that's where they are. But, think the next level of growth is more exponential. 30-to-40% is not like a linear reflection. Is a reflection of an exponential growth. 30-to-40% is slow and very small slope, it doesn't have a tall slope. But once they cross, that's when it's going to be exponential. That's what I think.

Q9: RS should be used to do mundane tasks, because I feel that. I can give you an example; human beings spend 80% of their time at work. It's a very huge chunk of their lives they are giving to their workplace. Workplaces are currently very robotic. HB are being treated as elements in a cartwheel, we are supposed to turn gears. We are not treated as humans. We are treated as an entity which makes money. Of course, businesses and workplaces are out there to

make money. Humans beings have far more sensitive dimensions. Robots can take away all the work which makes us feel like machines. Human beings are not machines, we need freedom from that trap. As for now, most of the population which goes to work is in a trap. They are very robotics and machine like. So, robots will relieve human beings from that kind of slavery. But people say this will lead to unemployment, I don't think so. This will free up human beings to do something better, more humane. Something more mindful, something more of value. As of now there is profit creation, the current industries are out there to make money, but when human beings are getting free and machines are taking over mundane tasks. The Economy will shift from profit creation to value and meaning creation and purpose creation. So, I think, that's the biggest contribution of robotic to humanity. When the robots start taking control and become unexpected and self-evolving. That's when we should stop depending on Robots and bring that to the check.

Q10: I think that robots should have basic threats of performing tasks and they should be independent of any culture, so I think that technology is not yet there. How do [inaudible] you interact with a human being; all of our robotic platform is very language dependent. There are not language independent RS. Languages, commands, that's how they operate. There should be, human body language and queue detection. There is something called a [subvocal processing] which people at MIT are working on. It will read your brain and body signal and it will translate that into a speech and it will do shopping for you. It is called [sub vocalization] that technology should be pursuit a lot more for detecting non-verbal signals and for communicating across languages, and across racial barriers. Non-racial queues are very important. I don't think race, culture, language should be bases. Language is a very slow way of communicating, there is one thought and you use 5 sentences to communicate, it is a very outdated way of communicating. With so much technology, nonverbal queues should become the center communication and robotic. If you are so smart, why do you use language, you use non-verbal cues. It will help people with assisted disabilities, it will help ppl of all age groups. Ppl who can speak, ppl who cannot speak, everybody, it can work with animals, birds. Non-verbal communications are faster, I think that's where the technology will go. There is no other choice. Language is a slow medium is very dumb, really dumb.

Emotions related questions:

Q1: There is something called. There is a term for this it is called, the time taken for human being to get familiar with human emotions, it is called the valley of cognition [something]. So, HB takes certain amount of time to get familiar with a new technology, they will have to go through that valley, it is called Uncanny Valley. HB has to travel that [inaudible] it will give

the speed of adoption of HB to any new technology it can either be a yet pack or it can be a robot with emotions. Anything which is unfamiliar to us, we will take time. For example; touch screen, the baby boomers took a lot of time to get familiar with touch screens, they were very familiar with button form. Because, touch screens are unpredictable, they are very sensitive. Button form are more rigidly wired, they are much more easy to understand. So, like that we will have Uncanny Valley. Once we [inaudible] will be getting used to it. It not going to stop, because HB are getting busier with technology, there have a lot of thing that they can do, there are virtual work, spend time in. Real world is not so interesting anymore for HB. So, all age people, children will need some sort of care and assistance. Cognitive robots will step in and give assistance. So younger kids or older ppl, sick ppl, they will get attached to robots versus HB. Because, HB are busy they don't want to engage in real problems, robots are doing it. So, this attachment is going to happen, is definitely going to happen. We can't leave without a phone; people can't leave without their phones. So, you take away a phone from a person, he will be upset, he will not be the same. Same will happen with robots, especially social robots, who are helping disable, young children and ppl with assistance defect, deficiencies, disabilities. So, it will happen. Is going to be an attachment, it can work both ways, God and bad.

Q2: How similar these emotions would be compared with emotions displayed towards people. I think robots start evolving on their own, they will have emotions which we won't be able to understand. We have limited intelligence and we have limited dimension of thinking. Our social circle is very limited, we are interacting with the same humans, with the same thinking, same positive-negative traits. We haven't explored the grey area of emotions. Our world view is very black and white. Robots will have much more advance emotion. We will not be able to figure out their emotions. Maybe, there are no words for those emotions, that's going to happen.

Q3: I think it will happen, because robots are not judgmental, robots are unconditional. They will act like pets, but they are smarter than pets. They will give you companionship, they will fill the gap in your life, the loneliness in your life. So, they will be friendship. Definitely they will be.

Q4: I think HBs like free will, they will not like domination. Any robot which controls you, is going to cause negative emotions to you. Any robot which is nagging. For example; you are not taking your pills, or you set an alarm with the robot at 6 am with the Robot, and you don't want to wake up. If the robot is after you to do the task is going to bother you. If there is no snooze on the robot is going to bother you. Anything which dominates the free will of a HB is going to bother HB causing negative emotions. I feel. Also, if robots know that it is competing with a human and HBs will not like that robots. There is unhealthy competition between

humans and robots. If both of them are competing for the same job, HB will have negative emotions. It thinks these are two circumstances where there will be negative emotions. The third one is very [away and weird one] if your spouse our you close friend falls in love with the robot and picks the robot over you. Which it can happen, which is going to happen, I can see that, because they don't have emotional issues, they are be learning from your behavior, and they will be like your unconditional companion, that's when it is will be a break down for human race. A robot is liked more than a human, because we come with come with problems and defects, we are not perfect. Robots are going to be perfect, so that's when it will be a problem for humans. Where you will feel that you are not perfect, that's a negative emotion.

Q5: I think any time you interact with something more powerful than you, which cannot show empathy are a very intimidating situation. If a robot which cannot show empathy is making decisions for you that's a very scary situation. For example, in the DaVinci RS it has a 12-13 degrees of freedom when is operating. The patient will not be comfortable if they know that a human being is not overseeing that robot. If that robot is doing it on its own the 13 degrees of freedom and surgery on its own, that’s going to be scary. So, I see the days surgery are going to be fully automated it's a scary situation. That's a very scary situation. I don't know when that will happen, but automated surgeries are going to be one such examples, scaring. And also drones troops, drones which are navigating, these are called drone [swans], and drone swans are in your city, flying around in the air in uncontrolled groups making formations, that's a very scary situation for civil people, citizens who don't know why the drone swans has been deployed. Drone [swans] make noise, they look weird, they are not looking natural. So, when they come to your city and they are doing surveillance people will be scare. I just been informed and mentally prepared to face this.

Q6: I think it will impact all these areas, both positively and negatively. So, for example, mental health and wealthiness there are a lot of apps which perform intervention and bring people out of suffering, because they have a very good sense of problems face by humans what kind of conversation should be made, what tone should be used, how to pacify someone. It's very much like Siri, the voice is always calm, and it answers all the questions, is very comforting. The paranoia, people start depending on a machine, and you don't know at one point you can stop trusting the machine, the machine will say: break up with this person, is causing you a lot of problem. What if this person is your spouse and you have to live with them a whole life? You cannot break up just because the robot told you that, sometimes you maybe are attached to the robot and you may even do that, because you trust the robot so much you might just break up, so those are pianoing moments I feel. Where you are changing your life, decisions based on AI

guided applications, because is so real to you. You feel like, A child thinks a bear is real. Anything happens to the teddy bear the child cries, is more real to them than a friend. Sleeps with the teddy bear, wakes up. Is he's reality, so like that. If you think that saint clause will come until you are 4 years old, saint clause does not exist, right. Somebody has to tell you that Santa is just to keep kids busy. Like that, is going to become real to humans, even though we are grown up, we'll think is real.

Q7: Fear of dominance, you don't want the robot to dominate and take control of your free will. Unexpected behavior from robots, when robots start having unexpected behavior, cognitive abilities, emotional abilities. Replacing human interaction, when a human being stops interacting with humans and choses a robot, it's a fear. He's is going to lose on being a fabric in our society, he is going to become a [luff]. So that's a fear. You won't have normal skills. You go out and socialize, you will lose the faculty of normal, casual conversation, social mingling, being a part of your group, [your pear group you are], your faculty, it will become vestigial anything which you don't use, it will become vestigial, it will drop out.

Q8: All these ideas are feed by Sci-fi writers. Sci-fi writers got these visions, ... Philip Dick. There is another lady, she is Madie Allen these man and women they dram on these things. It came to them as visions and they wrote down. Scientific publications do not accept visions, such fantasies, they had to write novels, they had to write something, somewhere, it turns out into novels. These novels are very popular, and they are been feed to the human being in so many different ways, in movies, in TV series, in books, in comics, in technical writing. People always talks about fiction. It has hypnotized humanity to make it real, we are all hypnotized to believe this. Why do you think Elon Musk wants to go on Mars, because at the age of 5 and 6 he was reading Sci-fi novels. By the age of 14 the goal in his life was do a mission to space, he used to draw spacecrafts when he was 13 years old. Has it embedded subconsciously. He is just a kid and he's dreaming that. Obviously, these authors are so powerful they hypnotize humanity to make it real. They are all like professes, you know professes means you predict something that it will come true, there will be an apocalypse, there will be a D day. Everybody predicts that. So, prophets of science and we are following it blinded, but we don't have to follow it. We don't have to. We are following them subconsciously; we should not let that happen. That's the reason we have these discussions. We don't have to hypnotized by these authors.

Q9: I think is quite probable, because if we are encouraging AI which we don't understand, and we feed that AI to control robots. Robots are like swans of bees, swans of ants or swans of bears they will go cooperate with each other and they are going to develop co-intelligence. For

example; if you look at the soil, the ecosystem of the soil, there are so many funguses, bacteria, roots, under the ground like different ecosystems. These are acting as co-intelligence; they are not one ecosystem. Is called bio, many ecosystems are acting together in intelligence. One ecosystem is communication with another, That's a very powerful ecosystem. That's why humans cannot implement that in a lab, in a lab we just have one parasite or one microbe, and you want to create what is happening in the soil, that is not possible. There are hundreds of ecosystems which is called collective intelligence. When robots form collective intelligence, we cannot beat that, forget about building one single robot, collective intelligence where one robot is communicating with another that's going to wipe out humanity. When that's is happening without human guidance. Like it happen at FB, the AI build a code which humans could not read, they [the AI] started building something between themselves and video servers, very powerful servers to support the AI applications that 90% of communication is between machines, there is data centers 90% is between machines and 10% is between humans. what do you know what these servers are building there, what are they talking. [inaudible] Are they evolving, are they going to kick humans out? We don't know that. So, cooperative robotic networks which are beyond human intelligence are definitely a threat, they should be a red flag. They should be monitored; I would say around the clock. RS: I would say, one is collective intelligence. what they can do, we don't know. We don't know what collective intelligence can do. The second thing. I'm an expert in human potential for creativity. That was my first book. If Anything, which reduces humans free will and creativity. If collective robotic system, anything that reduces free will. That's a problem, that's fear for me. The third thing is the ability to pick up skills and defense, ammunitions, destruction. That's a third fear, that's a real fear. Because they are capable of building. They synthetize their own material, smart materials [inaudible] humanity. There is always going to be a class between humans and robots. Those are the things I'm really scare off.

Interaction related questions:

Q1: For very simple tasks they can be fully autonomous, but when they start getting cooperative networks, the autonomy should be regulated. Like as I said collective intelligence should not start evolving on its own. You know, for example, I was doing a project where they are using drones swans to survey forest. Because, amazon forest is very, very dangerous. They are filled with a lot of bugs, unpredictable attacks, for HBs to and survey. If you use a drone swan, you want to give it some autonomy to decide how to navigate to the forest. But when the same thing comes and [..] with different projects, and you start giving them arms. They are very capable of adapting on that, so they can go and shoot crows that they think are not capable

of living on this planet, because they don't like it, because there is a racial bias. They can do that. Because, I'm conceiving. For example; if you feed the robot enough news, that people from a certain country are not good, then if you give them the technology to build arms, they [inaudible] they are not fit for living. So, this are all crazy ideas, because is not under our control. That's why I keep saying collective intelligence, which is not under our [inaudible]. Autonomy should be carefully judge against the training, which the collective robots, drones or systems have. Are you giving autonomy to a system in defense, or is it for surveillance, can the surveillance. If is surveillance, the technology should not be sourced to anybody in defense. Because, if it is meant for protecting forest, it should stay there. If the technology is sold, it should be penalized, or you should give a fine or you should be in prison for that. These things should be very strict. So, there is always a learning curve. People don't know the laws; they will make mistakes. You should give them grey spirit to understand, because this is also new to everyone, but the regulation should eventually take ownership. Make sure it doesn't happen.

Q2: I think you should be able to predict their moves. You should have control of all the dimensions in which the output is being plotted, let's put it that way. If the output is plotted in 10 dimensions, you should be able to control all the 10 dimensions. If the output is plotted in 11 dimensions, but you are only operating in 10 dimensions, that's a big problem, you don't want to be in that space.

Q3: I think robots should accelerate the interaction. FB brought a lot of people close to each other, open up crowdsourcing projects. It opens up international campaigns, a lot of injustice by government to people was brought to light. A lot of Underprivileged people got founding. So, those are the things that should benefit. They should accelerate social behavior. I was reading an article on what are all the parameters which defines what a human being is. There are 10 to 15 metrics which define what a human being is. I can share the paper; I was a research done by the university of Washington. There is another paper on robotic manifesto, how the robots evolve. Where we draw the line on social behavior on robotics and none of those rules should be violated. What a human being is should not be violated by robots. For example, a HB has a complete understanding of ethics or social wealth being, these things should all be protected when robots go out there in society. If they are not, then it should be FDA for medical devices. I don't know what the EU has, they should stop it. FDA was born because [inaudible] trials humans were used as lab animals; poor human beings were used for trying out vaccines. They were treated like animals and some objects. You cannot do that.

That the reason why FDA was born, you cannot collect humans for a trial... so are laws here, so of people die in that. So, like that there should be some sort of law pushing back, that HB

should be retained as humans. Whenever those checklist of what it means to be humans is violated, robots should be taken out. No matter what they are built for, they should not be kept.
Q4: It can become very close. I think robots are affecting evolutionary capabilities of humans. People don't want to. The number of single families [inaudible] is maximum than any time in history. People find all their soul and companionships at work, social media, the technology. Which is already replacing humans. It's already happening. I think robots will take it to the next level. You know people would just want to live with their robot nanny or robot assistant and not interact with humans. A lot of place ... placing or packing orders, they don't have HBs, there is not interaction. It's quite possible they affect our evolution. They will make a lot of human emotion and physical abilities vestigial. The futuristic man is not very physically strong, got a much-developed brain. because his thinking a lot, all of these thinking applications
Q5: It's very hard. We should [..] regulations that are not yet there. I think let's say there is driverless car and there is a man and a baby who do you want to save. What does the algorithm is going to say, save the baby or the old man, right? There is Asimov' law self-preservation, the very first law, as long as you are not harming anyone, the first right is to protect and preserve yourself, so in an unexpected situation this right should be exercised by humans not robots. Robots should be thought that. It's their job not to harm any HB. So, you should teach that, and robots should have self-protection. That should be second priority when compared to humans, because if they are in a situation where they themselves are getting destroyed and HB is safe, they should be thought that, it's okay to get destroyed [inaudible] and they should survive. That should be thought to robots in every way. Collective robots specially. All collective robots, collective drones, cooperatives networks, they all should be thought that. Because they are serving us, we are not serving them. They should know that.

Expert 2

General related questions:

Q1: In the first phases of this kind of intelligence, I think they are going to replace a lot of tedious human jobs which day don't ...The impact is rather positive. people can still focus in other tedious but it still necessary chores.
Q2: To Answer the question we just feed the data and I hope to get something out of it. Hope that the system finds a pattern in this in the cybercruds. I think yes, no way we are going to create a conscious system by feeding data into it I think that intelligence emerges from learning from experience which is individual like a child

Q3: Oh, that's a difficult one. if you have some biases don't necessarily mean negative things right. so I would assume, that's a complicated thing, I mean if you want to be a robot biased towards helping people then that's a good thing to give for the system whereas we have some racial biases or biases like this social biases thing it wouldn't be that great. how to prevent this. how to prevent: is there anything much to prevent, rather is how to make sure we give proper biases, how to prevent bad biases? I mean through I guess some committees through selection of wise people elected by the other people, just like governments we had some so some committee who then we'll vote which kind of behaviors are ethical and which are not, maybe something, the easiest thing I could offer that help smelly helps to prevent so negative biases in out behavior.

Q4: I think my last answer kind of answers to it. But I repeat, through a committee of selected wise people, first thing. And other thing means if you are going to teach the system, so you need an algorithm don't give it information about these kinds of features; skin color, sex, age I don't know. This also keep the system blind, but we might end up with useless systems maybe we don't know, there is always late other side of the coin, so these are my two offerings

Q5: When it comes down to high-level decision-making like governmental decisions and really which comes down to emotions and these kinds of things then is humans. humans are good at high-level decisions, but when it comes down to low level things like picking up a package or drilling a hole, decisions like are best made by robots, because they are accurate within this geometrical system, they perceive the environment with way better than we do. But we are good at given the contexts doing the high-level decisions, why the factory should run, and robot is responsible for how it runs. So how much should people trust for what system. I guess it depends on the context you trust your phone when you're making a call so it's not human operator switching some plugs or things like that so if it's if it's a small effort or low-level effort that the robotic system is doing, then yeah, for sure we can trust them. I would really be glad if there were some kind of example. Because if we want a robot run a government these days, I wouldn't trust it, because there is a lot of improvements to be done.

Q6: Could you also specify what will be the reliability, so reliability is a bit subjective thing. System does what is asked to do in ideal conditions, and if is asked to do: clean this bin we have over here in the corner, it does. but if somebody throws a wedding ring accidentally, but robot still clears out the bin for this person is unreliable because he could make a decision to see that there was something valuable in the garbage bin. But on the other hand, reliably did what he was told to do, clear out the bin. It's again interpretation of context and if we want to give a robot more decisive power. Reliable in terms of mechanical, if we continue with our

wedding ring and the garbage can example. I would say give like 15-20 years and we have systems that can assume and concludes ideas and things like that. And in that sense within let's see half a century we have a reliable system that we would trust as we would trust our friends. how much one should trust a robotic system judgement and suggestions. Let's say robots are usually governed by some kind. they have designated Target there are either clearing the streets, they're Gatherings Facebook data so different kinds of robot, if it's your personal robot who's managing your chores at home. Sure, you should trust your robot when it comes down to security. robot says your fridge is dirty and Out of Milk I would trust it, whereas if this system would be governed by some kind of fishy Corporation who just wants to pull out some data. Then is maybe questionable but is not even a robotic related question, we still are experiencing it these days, we have different software systems, we have agreements. we agree with these terms and conditions, that nobody reads, we pull out data. but it's still Mutual trust and occasion is still their Facebook leaks in iPhone iCloud leaks so how much should we trust it depends on the kind of infrastructure that has been laying down to support this robotics system if somebody created his robot in their garage not too reliable Maybe.

Q7: I think is complete it fine. Is same to ask, how ethical it is to Let a Child use your vacuum cleaner. I see no ethical conflicts as long as the robot itself is not biased towards those behaviors you mentioned before. If the RS is kind of sexist racist or are just something like this of course not a good idea to let the child next it. But if the robot is secure in terms of nobody can hack into it and injure the child or say bad words through the robot, secure and it is not weird in terms of all those biases, I see no problems, we shouldn't keep children away from innovation and technology. It is as probable as I mean yeah children learn from everything, and they get inspired by anything, by parents, by friends, by teachers by even trees, I guess. If the robot has those behaviors, bad behaviors, then for sure will learn, get influenced. so that's definitely a source of influence. There is no doubt in that.

Q8: I think it's unlikely for AI to spontaneously come about but if the structure of the AI is constructed properly (again, by humans), then it is plausible that AI will continue evolving without further human interference. 2) Robotic systems tend to reflect humans quite a lot. If it is because of psychology, entertainment or mechanical capabilities, that's not clear but we as a mankind are in a transition phase where automation is substituting simple jobs - hence I think all aforementioned 3 reasons are somewhat correct: Psychology) we want the emotional adaption of physical AI systems to be smooth, hence we reflect some human attributes on robots; Mechanical Capabilities) Human body has evolved throughout millions of years and it's been optimized to cope with countless terrains, physical challenges and it's highly dexterous

(we can grab, lift, transport, reshape, etc. objects with ease). Hence it makes sense mechanically to replicate (or be inspired by) human limbs; Entertainment) Human-like robots bring a lot of attention (have witnessed that personally), hence it makes sense to replicate human features for grabbing attention in the field of entertainment and marketing.

Q9: A lot, we are going to depend a lot. I don't think so, because as I mentioned there are a lot of things that we as humans don't, we would rather not do. But the things that rather not do needs a lot of intelligence and if we want intelligent things to do intelligent things, we assume is the one to do so that's the only option, we let the robots do those things, intelligent robots this means that we depend highly on those systems. I would say no. I don't see anything particularly; I mean that's kind of the next step of evolution. it's down to the basics, we all are atoms and particles, we'll say the robots are better jokes so no.

Q10: So I guess the dumbest idea is to gather a lot of data regarding this culture so we just gather data how people behave, how they do and upload this knowledge into the robot, the other way is, we let the robot gather the data itself, operate as an intelligent unit within our community and that would be the other way capturing this knowledge through learning. but if the society is reluctant to accept this RS then been, we for sure can't capture this cultural information in non-bias way. we just need to rethink that, actually Gathering a lot of data just use humans, capture our history or books or a whole lot of things try to shove it into the robots brain would be the appropriate way to maintain a cultural Behavior in addition to letting the robot actually exist in the society I think that's a combination so.

Emotions related questions:

Q1: There's this word Uncanny Valley, I think. Since I have never experienced these kinds of systems, I can only imagine that this uncanny valley is true. assume people would be first kind of terrified, a system that doesn't even remotely look like human, acts like human, it will be definitely weird, initially. because we as humans, we know our life cycle. we are born, we grow and all this biological pattern. some things have completed it different cycle of life, maybe an endless cycle of life, just leaves for ever shows similar patterns off emotions as we do, that would be weird if this wall was start talking to us, we would be completely freak out Initially

Q2: In order to answer this question, we kind of need to know what the emotions are. these are really complex things. so yes we can replicate Some Human behavior, gather a lot of data, humans responding in various ways and then we shall generalize this information use this information and say to the robot use this library of information as source for fear or this library for happiness or combine them somehow. And this will definitely full somebody most people will think that the robots have emotions. Real emotions it's really difficult, there is this missing

bit of puzzle to answer this question and, it's a difficult one, hard for people to develop emotions, real genuine emotions that we don't have any doubt in it.

Q3: I would think for the most part we will have this intelligence systems someday, it will be different. it's like when I talk to you, for sure I'm a different person when I talk to my maybe best friend, I do weird bad jokes around the person and I'm a different person when I talk to my supervisor so we have this kind of masks were changing and we are going to talk with the robot for sure, we're going to have a different mask probably is going to be kind of maybe mean by somebody, we're going to be imperative saying trying to tell the robot what to do, rather just have a good discussion. But it will be different, yes. So, at least initially when we have systems maybe later, if the robots get mixed something unlikely, we can't even deal with racist not even to mention with robots.

Q4: I would like to be your friend with the robots. but I don't know if Express a general opinion over here. I think still a lot of people maybe just solely out of Interest of how it is to have a robot friend you know, not to be sexist or anything. I have friends who have expressed explicitly "oh I would like to have a gay friend" why would you identify a friend as if I really need to have a gay friend. Friend is enough I think, so for sure there are people, "I would like the robot to be my friend" so and they're going to be a lot of those. I have maybe different reasons; I'm interested in the emergence of intelligence; in other vessels that are in our brain and if I could talk with such system then I'll be really delighted to see what kind of opinion such system would have been so I would regard the system as a friend to communicate with.

Q5: All system that starts to kind of start influence their life. For example, if you are a factory worker and your job is to drill a hole thousands of holes a day and if there's a supervisor comes in and says "hey, we bought this new robot" which are way better than you. when the kind of the domains of what the person does, and when the domains of the robot does overlaps then we're going to have problems but if a person can adjust his or her Domain to another level, instead of drilling holes maybe the person will be fixing there robots they bought. So, then there would be no conflict.

Q6: In order to be scared of something you have to first experience it, like babies are not scared of spiders, snakes or whatever Heights. So first they need to fall down get bitten by a bug or something and then you start to be scare. And if you experience such a system which unreliable behaving is weirdly then yeah, it's easy to be scared of it kind of a system which does moves unexpectedly or just makes unexpected decisions which impact you in a physical or mental way, that makes you scared. And of course, the other parameter is that if this system cannot be stopped, I mean if the system breaks your house, but you can just turn it off or just I know

throwing the garbage can, then is fine. Whereas with the system the system it cannot turn off its just wreaking havoc and it's just posing Danger to you. of course, it's scary. It's maybe the answer lies in how much control we have over the system. It can do scary stuff, but if we have control, we switch it off, and [inaudible] but if these things are not, the options then we will be scared, I think yeah, of course.

Q7: I'm sure there's a fine. balance when it comes down to bringing those robots in our society. I started the interview by saying that, if robots take our tedious jobs we will all be happy, but if it happens suddenly and we'll all suddenly happy and there're of are billions of happy people for sure there is the other end we don't have enough other jobs for people so this cannot happen unexpectedly and if it does we still have some job prices and this relates to for sure with mental health, but if we can balance the sounds nicely just bring the robots in our society gradually, as we are doing. Is not happening overnight, the fact that we're bringing in our integrating robots it's happening over long course of years like tens of years and that case. And we are as a Society are improving a think. When was the last time we had a huge war. A war somebody declare. that's not a topic for sure. but I think we as a society are getting better, mining that it should be you related to our Mental Health which means that innovation of Technology to body system would have a good influence on our well-being, we don't need to think about getting a roof on top of her head or getting a decent meal, these has been managed, hopefully. With the careful integration of those robotic system, of course, it can't happen overnight.

Q8: Unexpectedness I guess unexpectedness. Security is really important, by security I mean hacking. the robot is a computerized system connected to network, so you don't want your system to be hacked because. I don't think the robots are mean, but people are mean, there are lots of disgusting people roaming around the earth who would do this for fun, like hack into a robot and do harm you. Security is something you need to be aware of and scared off if you need to have to. Security. Unexpectedness, I would replace this with quality of service like mentioned before so if the Robotic system was developed by your local supermarket, in other terms some kind of bad company, then the system probably is not reliable it's unexpected doing you something unexpected. so that's you should be scared off, you should be aware of, rather than scare off. The quality of service the quality of system you're approaching at. Security and quality.

Q9: It's popular I guess because I think it has been popularized, some funny Martians coming down from a planet, but it has been a scenario in so many movies but that's one influence definitely Pop Culture. For sure there is the fear of unknown, usually people's think always the unknown is always bad. and its hardwired kind of into US expect the worst and hope for the

best. These are the factors why we would be biased towards thinking these thoughts. My personal opinion, which I think that Terminator work like this scenario was that once the R systems kind it was switched on. It quickly, within some Nano seconds, became aware of the b*** that people are doing, harming, and it logically regressed that humanity is a threat to the whole universe, our actions are pattern of living, is kind of, we're killing each other, nothing in the system is easiest to see that A system that wants to eradicate disease sees us as a disease. when thinking globally or even galactical or universal, then Humanity at this current state could be seen as a disease and the system is designed so without any protection or seed cards, that could be a likely future, yeah. but I mean likely in terms of if we create such a system but I don't think that's going to happen, because there are lot of people who know stuff, who are involved in developing the systems and we are trying to take care of steps in legislation in and then all kinds of things so that's not a reasonable fear, but it is a possibility.

Q10: Not so probable but there is still a probability, it can happen, but I think it won't. because of the careful steps we have a or governments and [inaudible] takes towards those systems. it's good, mostly we are skeptical about the robotic systems and that's good because this will give rise into just, this makes people think and talk and discuss so in a way yeah and answer is unlikely. I guess just one more thing it would be intelligence, higher intelligence. I would really like to talk to a RS but probably once it is intelligence as a generic human, it will take them a few years or maybe seconds to be even more intelligent and when they get past or leveling of intelligence, we are kind of the as we see ants on the grounds, robots see us as ants. I think they are not going to harm us or anything, that's not the fear. The fear is that I cannot take the journey there are having, so they will probably develop Advanced systems and go explore space and, I know, change into a quantum foam, I don't know, so they're doing something that's which would be physically a on comprehendible for human and this leaves us kind of losers of a race or something. so, I would really like to join the knowledge alongside with robots going, develop hand to hand, but probably that not going to happen. so that's my fear they get past in terms of intelligence sent, we are just left as ants. Since I'm working on RS, I'm not afraid of losing a job. Yeah generally I don't have any specific fears

Interaction related questions:

Q1: for sure they can do their chores, drill out the holes and clean up cupboards and fridge, wash dishes, this kind of autonomy, that's what we want, we have our robotic vacuum cleaners and things like that, people liked it. If you take it up the next level, context dependent completely. you don't want your vacuum cleaner to tell you that he doesn't do your homework we all have better things to do. Let's have a beer or something. so that's not what you want. but

if you need, if you have a depression and you need to system something that helps to discuss. There are a lot of people who are socially distant they don't want to communicate with people, but they're still in deep depression. how do you approach such people? For them RS are kind of, they don't look like people, they are not able in technical terms, but they still have those capabilities of making decisions, giving opinions. And in that context is great. it's a context-dependent question

Q2: For the most part I think, in terms of practicality, as predictable as we can make them kind of have a screen which shows the intentions, give any means possible, verbal gesture, displays of whatever things like, that to communicate. but, I'm sorry it's again context-dependent, if we have a dangerous robot in a factory then for sure, must tell hey I'm backing up don't get on my, and it has to be 100% predictable, you cannot have the heavy Machinery in there. but if you have a small gadget that's smart, yes. Doesn't need to be that predictable, it shouldn't yeah. Again context-dependent, if we want safety completely predictable, if you want company then you couldn't be predictable otherwise it would ruin the point. if I would know any everything you're going to answer, you will be boring companion, I guess. I wouldn't have a beer with you you're unpredictable. I don't know what's in your head. I don't know what you want your answer is going to be, and that's interesting for me, so in that context unpredictability is good.

Q3: If you want the system to take care of you don't want to get some cold blocking things with a siren on top of his head saying I'm entering the room please make sure you're in coordinates 3.5, otherwise I would kill your something, you want the robot be smooth, saying I come in so kind of show the behaviors that a reasonable person would have. In that sense yeah it should not go straight to the point, but say hey, enter the room, how are you doing and just ask those basic social questions and then, go. In order to maintain a kind of trust.

Q4: If a system, if somebody buys a husband as a robot, robot as a husband then for sure make sure that the system is designed good. Because you are going to interact with it and you're going to be closed with. I know it's context-dependent, if you don't want to get hurt emotionally, so you might get attached to your robot and soon realize it's actually fake its isn't that emotions in terms of their just programmed in not synthesized based on context but just written by another person, so meaning fake and if you get attached to such system and later realize I know it's not real, then you get her and you shouldn't be engaged with this kind of thing so you don't need to be aware that we have a dangerous signs a pack of cigarettes that hey it damages your health been that says the robot should have a label : that it's not a real person or something I don't know, but this would do some credit weird social bias that we would say that robots are

not real and, it's difficult. I don't know how to fix the problem, but this could happen somebody gets attached and then gets hurt because the system is not, they what they expected to be.

Q5: I guess in that case robots usually are made of made of parts that can be replaced, humans usually are not, meaning that they should prioritize human life. so unexpected situation towards humans, save a human. But if we have a conflict, two choices: run over an old person or a young person it's Bonkers, but this hypothetical experiment, limits you with two options, what would you do. shut down and make the make both persons died or pick a person and where this kind of choice comes from? I don't know, it's super difficult question you said, to prioritize human life but when presented with so weird moments, I don't know flip a coin, which means the system is super unreliable and they could be not so welcome in human hearts.

Expert 3:

General related questions

Q1: For me Since I am already dealing with robotics there are two ways to talk about it 1 is have your robot butler, which is a separate entity, but then the other thing is You could integrate robotics parts which is like a prosthesis Or maybe you could have like a robotic arm which is Part of your body meaning we will become cyborg But then there is the debate about are we going to give them full of autonomy or are we going to become stronger And I am part of the team which votes to become cyborg That would be a better way In my opinion, because even a species if we want to evolve We can't just completely go with AI, because AI They start to communicate between each other In a very high level that humans cannot understand, and then we Start to fear them and those projects are just shut down. there are two parts, the one that we can augment ourselves. In some way it's good because people can let's say People with an amputated leg, they could use they could use this really nice advance, which literally performs as a human leg. In the other hand that there are also this huge problem People would invent some kind of weapon Out of it, soldiers will use it as a Tice, meaning literally kind of Might become Robocop. The guy might have lost his whole body but then they will put them in this kind of body with Legs and arms and they would just say hey let's go to the battlefield again. That might be a dystopian way of thinking. And for the ones that are completely robotic. You do have examples of those nowadays. Which will end up [inaudible] as a dystopian [inaudible] Sex robots. There is already Sex bots out there and then it goes into this kind of huge ethical issue, where do we draw the line. Talking about robots are west of huge problem in Canada, when there was this guy who bought the sex Doll, but which looks like a kid. Then people recycle he's a pedophile but then the psychologist kind of argue, no it's a doll It's not a human, but then where do we draw the line. is it creepy, yes, it's absolutely creepy. We humans would

literally absolutely abuse. I'm pretty sure about it. But then that's as an advantage that you could use them as a healthcare system Taking care of the elderly, that's kind of the boom part but then I would say the dark side it would be the widespread compare with the benefits that might come out it.

Q2: Yes, absolutely. But I would say as a researcher if you genuinely want to create artificial intelligence, then the negative side of the Human Behavior should be transferred to it. That is what makes us human, and also even Intelligence. Because us being cunning it is some sort of intelligence, you have to be highly intelligent in order to be in top of our game. For example, if you're going to cheat on an exam you have to be highly intelligent. In terms of just intelligence, we can just pass that but then If we just say we are going to have a utopian society Everything is going to be positive everything is going to be rain bones. Then the robot really doesn't know what the other side is. You are avoiding the topic like this some kind of taboo, the robot doesn't really learn so it's not a true artificial intelligence, the robot lack knowledge about what the negative side is. But in the other hand I wouldn't say we are doing artificial intelligence because what we are doing is making a computer more knowledgeable, because knowledge is complete it different from intelligence. And now just because we have access to so much knowledge And we can just store them and retract a huge amount of the data and then just manipulate them and just infer from them, that doesn't really mean is intelligence, it's more like yes, you are formulating new questions but inferring to the previous data we have. Would I call it intelligence, I would say he's a really high level of knowledgeability. Because is just the robot just will have answers and then it can formulate the next questions from whatever answer it has had previously. But is it really intelligence by itself?

Q3: In terms of transferring human biases, Why it is quite tricky I would say, because all of this machine learning techniques, They don't really put bias into it they just miss some factors into account, And then when we Miss this variables into account, it's just bad design but is not biasing it. In particularly we are saying let's not take a look into that, related to the previous question, we are saying, let's not give the negative data so we could safeguard ourselves, so the robot doesn't go complete it hardwired, and in a sense yes, we should do that, because is a machine is not a human, and if we plan to use machines as machines, then just give the machines the good data that's out there, and don't let the machine try to be a human because is not a human

Q4: In terms of hostility, in robotics there is these 3 mayor rules [Asimov's laws]. Never hurt a human, don't also harm yourself, but at the same time even if you are going to harm yourself or you are being harm by the human, still don't harm the human. If you are a proper ethical

researcher then yeah, that's what a robot is. Then if something goes out of that, in terms of technical sense then we will just call that more like a killing machine is not a real robot. In terms of racism it's bad design, people can make certain robots that intent to be racist, but in the other hand at times is because of ignorance that robots might end up being racists. Some of the face recognition software, The data that they have been fed into is majority white male, So for this face recognition software they will be able to recognize White male faces, If you're going to ask if is a woman or if it's a Chinese woman it won't be able to recognize it. Is it being racist? If it was a human then yes is being racist, but then it's just a guy who supplied the data to it. He made a mistake he should have given more data to it. And then it falls into the region of culture and privacy and everything. In the USA they were kind of cool with it, and they have all those photos and everything. But then we go to the middle East. Women are in their [Burka], you cannot take their photo, you cannot take their dataset, you can't have the access to their dataset. How are we going to recognize those women, we don't know.

Q5: Trust for me I would say it depends on context, because context matter when it comes trust. You can always trust a computer or a robot if you know what exactly the output is going to be, if you know how the system is working. You can have a 100% trust. In a computer when you are going to type 1+2 for sure the computer is going to give you 3, meaning the robot. If you know the robot is going to act like that you can have trust over that. But then with the person, maybe they know the answer is 3, maybe they just want to mock you, and they say no it's 4 or maybe they don't know the answer and they would say it's 5. So, they gave you the wrong answer, but because of what? they knew but they wanted to give you the wrong answer or is it more like they didn't know and because of that they gave the wrong answer. You cannot trust the human, but then there is the question, is it because of ignorance. Because, if is just because of ignorance you can trust the person, they didn't know the answer. But then they knew the answer and they gave a false answer where the trust is a bridge. With robots that normally doesn't happen, these days with AI they do play this kind of probability game, when it came to open AI and match of Dota this professional player. The robot did beat the guy, it was playing as it was losing the first 15 minutes and then steam roll like a professional player. [Game is dota 2]. At first people were thinking, oh this robot is so dumb that is doing that. But in fact, it was beating the other guy, because it knew that if I would do this, then the guy will fall for it and It will just steam roll later on. Meaning, could you really trust the computer? No. Was the computer objective win over the human, so it is more about the guy who design it. What was the objective was the computer design? For the guy who designed, I can trust is, yes, it is going

to win. the context matter, who designed and what for. And you can still trust the computer, because it did do its job, it accomplished it.

Q6: They are precise, but sensors can go wrong, there is always faulty sensors. Are robotic systems reliable, yes. but do sensors fail, yes. People should always be a little bit like a cat in the bowl, just cross check everything is okay, if you are blindly going to follow the robot. Because when it comes to Man, Mission and traction we have catch 22, being [inaudible] If it's I really simple task don't give it to the robot and if it is a very complicated task don't give it to the robot. The middle level of complexity it's the one that we should automate or give to the robot. The main reason is because building a robot is very expensive so for a very single task, we have Manpower for that, so just give people a job and they will just take care of it. But if you go to the highly extreme which decision-making job, for instance judgment, suggestions are okay. But then making the robot to make the Judgment the final call. That is a problem where it happens because. Maybe we do come out with a very fancy algorithm, for instance Example of a line following thing, Tesla algorithm, but then there was his accident. The question is who are we supposed to blame. The guy who was behind the steering wheel or the guy designed that particular AI. Because the AI was the one who went in Crashed. Who are we supposed to blame and no one wants to take this blame. And then we go it happen because the driver didn't keep his hands on the steering wheel, and it went and crashed. In my opinion I will say always use common sense. Let it be a robot or let it be no robot, always use common sense. Think as robot more like is a tool, so it will help you to make I calculations faster. Don't completely rely on a code just because the robot gave me that, I'm going to go with it. Maybe you should have a different Robot which has a different algorithm And if this one delivers the same results then you have something to compare with it, In that sense it's better. But even when all of this done, it can still hit the fan.

Q7: If I'm going to like to say introduce a childcare robot, I would say a Childcare robot is way better than a man. The guy or the company that is going to design this childcare robot for sure they want to keep their customer with them, they will make sure the robot will just have the good behavior. And the robot would not have any bad behavior, so a child wouldn't learn anything inappropriate from the robot, but will the child do something inappropriate to the robot, yes but the child won't learn something inappropriate from the robot. If the designer who came up with the program Was a weird guy and he wanted to Do something inappropriate then yes, Red robot itself with do inappropriate things, as a company do, we want our customers to do I complain like that, We wouldn't. I would say is good to have a Robot around a baby and a person. That is also this kind of robots that are called "Cobot". These are mainly used in

manufacturing plants. Where these robotics arms work in tandem, with co-workers because back in the day, A robotic arm will be completely shielded in this bullet proof box of glass, Whenever the robot what's on you would have this lights and alarms running, Meaning people should not get near it. But nowadays you have to robot next to you, you will be sitting there and thing just will keep on moving the other human person will just be working along with it as if the robot is their colleague, and the robot doesn't move in a fast manner, like the way they used to, these robots move slowly. And whenever they hit someone or when they find some kind of external force that is hindering them it stops moving so none get hurt. I'm pretty sure if someone is going to introduce a robot in the living room that is going to deal with some Kids, then for sure they are going to go with the cobots.

Q8: With the current technology no, Even for the next 50 years no. Because what we are currently doing is not AI, people are calling it AI, but I would say we are making computers more knowledgeable, but that's it. Is Not intelligence. So, no. They are not a reflection of humans I would say; They are more like a bizarre Shadow because they don't give you the full information. Under the metaphor that the mirror is completely fine and you can see yourself perfectly. But I will say robots are more like Shadows. Because you cannot replicate the human mind maybe even their physicality, you cannot replicate them. Maybe if you want to really replicate the human body, the only way is to have a baby, that's a proper human robot. A baby is kid of a human robot, because they don't know anything for the first few years, the parent you teach, they are like a machine learning where they are going to look into pictures and them. That's kind of the concept but then, Good robots will never become humans.

Q9: We will depend on robotics quite a lot in the future, we are depending quite a lot even now. Because majority of our gadgets are already produced by robots, We are already depending on them because China is cutting their labor market, They are going full on robotic production, We are depending on them so it's not even in the future right now we are depending on them. Why as a manufactured I want you to use my labor cost. We cannot prevent it, And should we even if we want to we cannot , Because at the rate As civilization we want to progress, Like the whole Market we want to capitalize and make profits , The only way is to use some kind of robot that is super-efficient Than a worker. If the robot makes a mistake you will know how and where it happened, you would have some kind of a log but then the human It's kind of tricky to know about them and it's difficult to rectify it.

Q10: IF you are like a startup, if you're building a robot your building is for the community that is around you, Or at least the community you are targeting at. But then with it be company for example, forget robots for a moment, let's say Disney market around the globe. if you are

a big company you have to invest and do the research and put a person in charge of that. But if you're a startup then it wouldn't be as much as a problem. There are a lot of startups that are popping up all across the globe. Which are working on robots and mainly with this Healthcare robots. That wouldn't be a problem for us.

Emotions related questions:

Q1: They going to be freaked out Because you really don't know how to deal with it. in sense You could use emotions more than logic. In Psychology Psychopaths don't have any emotion. But they can understand other people's emotions really well … They will talk to us is they are really empathetic and really Sympathetic towards the other person but in reality, total act. Deep down inside they don't feel anything meaning that is logic, Because these people biologically they cannot feel any empathy or any motion That's part of their brain is not that well developed They are doing because there is some rational thinking and there is some logic behind that. If you put that logic into a robot than it also cool display real emotions, but it's a real emotion no it is displayed but it's not real. We are going to feel a bit confused

Q2: How similar these emotions would be compared with emotions displayed towards people. This is already happening because you can see in Japan and Korea people already have boyfriend and girlfriend or virtual girlfriends because of one people don't want to socialize these days. I'm happy with my virtual boyfriend or girlfriend because they are not going through a […], They're not going to have any fights and even if that is a fight is going to be a cute fights .Just to make me feel like I'm in a kind of relationship. But for a deep-down fact I know that my girlfriend is not going to leave me. But it is with a human that is always this fear about when this person leaves me. Or is this person going to cheat on me? The robot doesn't do that. That's sense yeah, we are kind of dependent on an emotional level. Some societies have off this highly Japan in Korea. To actual people we will become robots towards them, for instance let's say with Facebook or Instagram they are Happy if and only if they get likes to their pictures. You could see ppl in a restaurant that chef would have a very nice dessert, they are more interested in taking pictures of it rather than tasting it. With interaction will be kind of problematic, they also expect other humans to be nice to them. But then when the other person does something but then they become angry at this person, It's kind of makes people more entitled, I guess.

Q3: They already have girlfriends they want to have relationships. Why not friend, yeah for sure.

Q4: All of the sudden the robot start asking super personal stuff. Depends from people to people , For instance it's like if Alexa is going to ask me all of the sudden Hey do you want to

buy something which in fact I was not looking up into the internet but I was talking with someone else Like a proper human but then Alexa overheard it As a suggestion. So, then you will freak out Okay so I'm just going to turn Alexa off. Invasion of privacy will be the first thing People will find that offensive. But in fact, I would say you Asked for it. You wanted your own life to be... and have your own personal assistant to help you out, so you literally wanted a Slave. You're a slave is going to watch you all around.

For me the first thing that comes is invasion of privacy, But apart from that I wouldn't really say a robot Show something negative to a human If it wasn't programmed in that way , If the robot is programmed in such a way that go display The middle finger to the other human Like a complete stranger than maybe yes . But then is the fault of the robot no, is the guy who designed it Made it to do that. Is on him we have to put the responsibility on, It's not like the robot learn it all by itself .There was also brought it from one of the universities where, They decided to say hi to learn things from Twitter And the things that the robots could learn from Twitter it was feuds So the robot started to curse a lot .Is the robot faults no you didn't have the common sense To teach him from something better maybe a dictionary or something else . The guy who designed it why would you put something in a really toxic environment to learn and then make it to try to speak proper English. Is literally like a baby, you curse in front of a baby the kid will learn it. Twitter is a really bad place to put a robot in.

Q5: One of the biggest fears for a human on a physical level is am I going to get hurt ?Let's say if the sensors malfunction the robot is trying to give you a handshake but then the gripping force of the sensor kind of fries and the robot squeeze your arm so badly that you really just have a bad fracture in your arm. Or maybe it's like this health care robot taking care of the elderly and putting them on the bed, and then maybe some function fails and instead of putting them on the bed, even before putting them on the bed they just drop them on the floor. That is a huge problem, you always have to [..] will this robot work? At any time, any sensor or the motor might fail and when it comes to these digital robots, I would say it's more about privacy and other things. Is it monitoring you all the time? Or for instance let's say you have some kind of AI system to improve your lifestyle without your own knowledge the robot might do something in a very subtle way that your whole character might change. When you come to know it was because the robot intervention that you have changed your lifestyle. That's is something really scary, because your really didn't even know. At least when you know this thing is happening it's okay. But the unkown-unkown is the most scary part.

Q6: It would have huge influence on paranoia, because there is going to be people that's always not that welcoming when it comes to technology, a new technology it's something they are

really scared about it. one negative emotion is one that some people are not that tech savvy, for instance I would say my own mom, for her if I give her this new iPad or Laptop and if I tell her how to use it, but then that instruction should be within less than 30 sec, and she should understand that, then she might use that gadget even though is highly beneficial for her, but if goes beyond those 30 sec, she might not use it. There are old people that are complete its anti tech savvy. Even though there is a robot, I'm pretty sure is not a human, meaning there is going to be some limitations with these robots. People are going to be mad at the robot, but they are using it wrong, and it's also the designers' fault because the designer didn't take into account that the robot was going to be used by an old person that complete it anti tech savvy. And for the paranoia is similar to the previous answer, people are going to be afraid of that

Q7: Privacy as in the robot can listen to us, and we might lose our privacy. But it is also our conscience choice, we brought the robot into our lives and to monitor our lives. But if someone hacks the robot. Then is not even the robot that's listening, but someone else, that is a huge problem Impulse - is not like the robot is impulsive, but the human. Let's say something fries up in terms of the calculation and the robot it is not calibrated properly, at times robots swing their arms like in a very hard impulse rate and that is going to hit someone if the person is going to be in between the robot's arm and the other thing . Power - I would say both physically and at a conscience level. Because if we gave this robot all the data that it needs. We made this AI system that it Could the betterment of human technology, there is a high chance that the robot might end up coming to the conclusion that humans are the biggest cancer to earth, because is not the animals that are creating pollution or anything. Is us destroying ourselves. It might go to the same fear of matrix, we might end up dying to the robot, it might be a huge war.

Q8: It mainly do with all the matrix and all sci-fi movies it more about that. The other thing is also that people do fear new technology, because for the first time when this lady how was from the Mercedes car, people called witch. There is this man who does not really understand technology for him is going to be magic and then is going to have this fear, because he doesn't understand, it is the very basic human nature, the fear of the unknown-unknown.

Q9: I wouldn't say that RS are going to be a threat to humanity, rather I would say it is the guy who designed a system, without taking into consideration all the required variables. He is the threat to humanity, is not the robot. It is the system, is doing what it is meant to do, which means someone designed it. At the end of the say if you trace back, it's going to be some human. Even if a robot itself is going to build another robot, but someone build it. At the end of the day, this person has to take the responsibility. Someone designed it and there is some human

behind it, and that is the main reason I'd never call them AI, they are highly knowledgeable, yes.

Personal fears. Non really, I don't fear RS. for me is just tech. If someone makes a bad design, it is just a bad design, you don't have to fear that. No worry about a bad design, you don't bring that to the equation, you just have to avoid it.

Interaction related questions:

Q1: Level 4 - you could go with supervisory. I would never put full autonomy. Maybe for repetitive task, you can go full autonomy. For example, cutting vegetables in a conveying belt. you don't need humans there, the robot could just chop the vegetables, the robot is also not moving, and no one is going to be there.

Q2: I would prefer, we should be able to predict the robot as much as possible. Because as a researcher yes, I would like the robot to be unpredictable, not just doing complete its random stuff, but unpredictable.

Q3: Very much. In some sense if we are looking at the current FB (social media). Maybe if they would had had a better system where people could socialize even better, then maybe this problem is not there. It's better to have robots more socially engaging, because when that happens, the people who has the fear about robots, they will also come out their shell and they will look similar to other person, and they will engage with these robots. In that way, promotes the robot culture.

Q4: This is more from person to person. Some people will just see a robot, and then for some people it might be their girlfriend, even future husband or something. It is just how crazy people are, and humans are crazy. We will never know.

Q5: It will behave the way it was coded, but there is also the catch 22 that I was talking about earlier. What happens is that since quite a lot of things are being automated. We humans start to depend on robots a lot, and when you do that. We'll start to lose our skills, and then the robot is just designed to whatever data that we have about the know. Maybe we can have some kind of data about the known-unknown, but there is always this unknow-unknown, and that's the unexpected situation, no one even factor that situation will happen and when that happens, the robot is going to raise his hand and it will say, I'm out of here, I don't know what to do here, all I was supposed to do is [designed] just give you an alert now you figure it out. That's going to be a mess, but that is the catch 22. We kind of use the robot to ease the workload of the human, but we ease the workload too much that we become super lazy that he doesn't even know how to work and then when the very crucial time comes, a huge amount of work load is required. That's the spot when the robot says, you are going to deal with it , but the human can't

even deal with the basic It would be better to have the supervisory mix control all the time, the human is always in the loop and the human will know , what to do and what not to do and how the robot is going to react. Leaving the human complete it out the loop. If we have the human inside the loop majority of the time, the robot could also say hey for some time I'm going to take my hands of the, you drive. You teach the robot, but also the robot also learns from you.

Expert 4

General related questions:

Q1: RS will start to become partners, or friends or kind of mates at work with humans, this is at least what I am expecting to see. I don't know how much will be achieved but, in the future, let's say 50 or 100 years, it is expected that my colleague is a robot, which does that, like tutoring in the classroom and so on.

Q2: That is correct, basically we can consider an AI like a kid. If you are teaching a kid how to smoke, how to use drug, how to still. Then the kid grows up, he knows how to smoke, to use drug, and how to still. And if you have a kid and you teach the kid good things, being good, knowledge, helping people. Then the person grows up, it will become again a very good person, helping other people so on. so is totally logical, what you say is a very logical statement, if you are feeding bad things into it. you have to be careful what we want to feed. So that's why the AI can be very, very adaptive if it goes to the hands of dangerous people.

Q3: We are aware not of how much of human bias is being transfer, because a human is transferring that. The human can be bias… we are thinking it is not bias so, I cannot give you a number. I think it exists. Up to what level, I cannot tell you. Theoretically speaking, if many people with different backgrounds try to aim the same thing, eventually then average of the whole thing will be a bit un-bias, even that people at are very different from one to another, but practically how much of these can be happen, I don't know.

Q4: Again a robot is an AI agent is a kid, if from the beginning you provide the data, you do not put any racist information or something that cause this kind of problem, the this kid, the AI grows up and it learn, it will never learn those kind of things. Is just like, if you are coming from a multicultural family being in different race, they all are normal, they all are being… If a kid grows up in a racist family, where tone of skin is really important and bad or good thing, then we are going. If you want to prevent these things you just don't feed it to the system.

Q5: N/A

Q6: Eventually, we are hoping it will be 100% reliable, so that people can trust it. Right now, they are sometimes reliable in very, very, very, very specific tasks, and better than any other person. but If you look at more general tasks they are not very reliable, and even if they are

very, very, very, very reliable there is between the humans, by humans I mean, academicians who are working in that field out, Which are minority people do not trust, people are expecting to see lots of error from this type of machine. so that's why they've been, this machine is doing some very important task, people are very scare. Yesterday, was in BBC that these autonomous drones, USA has announced that they have killed many civilians. I Think if you want to make a critical decision it should be finalized by humans, no matter what, so a robot makes a decision, but it is supervised eventually by humans if is going to be a critical decision. I personally have been working on that for 15 years. how much I trust my Alexa, for instance I can say 90%. Eventually, you start getting used to it. It is more like a Phycological effect , we would like to have somebody else to make decisions so you can blame that person or thing, if it is not a good decision, so is if they are [inaudible] so critical people start to Trust for whatever the AI is making . I think that companies will try to make it as precise as possible, so eventually the false positives or the failure will be so small that they will trust. Like Airplanes, everybody trusts airplanes, the probability of failure exists also, but it is so small that you trust.

Q7: It get back to previous question. If the robot has learned good things, it just like it will act good things, then it will be an excellent example for a child. If the robot has learned bad things it, so it is a bad example. Is like a parrot that you are teaching good words, and the kid is learning good words from the parrot. or you teach bad words, and it just repeat it the kid just learn bad words.

Q8: This is again, another thing that already exist, which is called Reinforcement learning in Artificial intelligence and machine learning, it means that the system starts to learn as proceed. I don't teach the system. Not that I don't teach the system, the system learns it after doing it, like a kid. They are doing something and then we go very angry at them, because it went very bad, after it has been done, they lean. The same thing can happen with the robot, because such algorithm in the lab are being deployed and AI agent. That's why sometimes as you are using some AI solution eventually it leans more and more, because without you teaching it learns by itself. Aim is to make it a complete reflection of human. That's why they have these feeds bio inspired or human inspired AI systems.

Q9: Iris, you need to maybe define clearly when you say the robotic system. Because, a computer is also a RS. I think you have to define what do you call a robotic system. Yeah, yeah, your Siri on your phone has all these capabilities, so it also can file into the robotic system. when you say that we addicted to the computer, addicted to the many AI applications which are interacting with us through the computer or mobile phone and ... it can have different form. I mean Alexa, … is like we are interacting, I watch is an AI which we are interacting, and

Alexa can be in the shape of a human head you are interacting. We are already addicted to it. Will we be more addicted, well the potential exits. Can we reverse it? too hard.

Q10: This is an ongoing research, because in affective computer that I'm working. It is already known that all type of interaction, because usually happens with emotions and so on. Doesn't matter of is verbal or nonverbal, is very cultural dependent. Cultural dependent, age dependent, gender dependent, there are so many dependencies. And, not now but in recent years it has started it the trend of studying culture dependency, like different words, for English, British, Spanish, the results are that you can see general patterns, but you can see [inaudible] differences. So eventually, we'll see very cultural dependent type of AI robotic systems, currently in the lab at the research level.

Emotions related questions:

Q1: They will be very excited, We are showing to the people some demos that sometimes and an avatar start to smile, are smiling or just doing some reacting to some type of emotion, so people at the moment is very excited, because it is not very creepy. But I don't know, in the biggest scale, by having a college which is a robot, and a smile at you how would it look like, I don't know. I mean the test that we have done, all results are super positive, everybody is like wow! oh! look the smile. No matter, what is the gender or the age of the person, and they all get very excited, weather this excitement will continue, I don't know. we are waiting till one robot kill one innocent with a smile, then you will start to feel creepy, I guess.

Q2: How similar these emotions would be compared with emotions displayed towards people. We have this human robot here, and we take it to some fair or some places, and the adult they take the photo and they smile, the kid are more curious and they come and pet the robot, and they way that they are petting is exactly as the first time they want to touch a dog, a cat. Like they puck it, they touch it and they start to pet and so on. I'm not aware of any scientific study in that field. I'm working on affective computer, and affective computing is an important part of the robot and human interaction, and the aim is we are making these robots or AI agents, act in a way that you are expressing exactly the same emotion, that you are expressing to me as a human, that I if I was a robot. So, this is how we are aiming to go, and this is something we are hoping to achieve, because the moment you are expressing your emotions, like maybe contempt, surprise and you are listening to me. And the moment you are robot and you give me the same expression it means have achieved the moment that you can accept me as your colleague, as your friend, even if I am a robot. And in the old sci-fi movies that you see, that the human fall in love with the robot and something, the robot is expressing an emotion, that if you remove

that robot with Gorge Clooney, it will be exactly the same type emotion and expression and talking, reaction and so on happening.

Q3: Very likely, until very bad events happen. Currently, I'm talking about AI with people you can still see, people sometime say that, how far are we until AI will kill us, or control us and so on. There is still expectation, all these expectations, that these things that we are seeing they are good, they are good, [inaudible]. But in the other hand, people pay a lot of money in order to notification of the environment, in order to make things to understand them and so on. I mean is like 50-50 going ahead.

Q4: You can connect this with two questions. If the robot is capable of, if a human is capable of showing some emotion to the robot and so on. If the robot is capable of understanding of you in your state and reacting in a way, that if it was human … then I think there will be no problem for communication. On once, let's say that something bad happened, you wrote a paper it rejected. You study hard for an exam was not good. And then you are telling a story to the robotic system, and you are sad, and you are expecting that as a robot I show you empathy. You don't want me to sit and cry next to you, because this is not realistic, this is not good. You are not expecting me to tell you a joke, but you are expecting me to show some sadness, if you don't see sadness, then you switch of the robot. Then you understand that the robot is not understanding, then you start cutting communication with the robot in some critical or emotional moment of your life. The right reaction to the emotion is the important thing, if you call it empathy or whatever it is. In general human, physiologically, feels uncomfortable if something which they don't desire is happening. Imagine that you are happy and suddenly the robot tells you "I'm recording all your actions" and even if you are not doing anything and you start to feel uncomfortable. Because it is something not desirable, but if you are a movie start, and I came here for shooting and I will be recording all your reactions, so you will be very happy, because this was something you had the desire for. So, I think if the robot, start a task, which is not desirable by the people around will cause some uncomfortable things. And it gets back to the previous general questions that you were asking. that if the robot was not trained properly, because the robot does not understand that is doing the wrong thing, there is a logic; it multiplies 2 by 3 and I got the number 8, and it thinks it is correct, it doesn't understand that something happened in the calculations of the circuit. So, for the robot it will believe it was right, like a human, if we believe we are doing the right thing, and later we find that it was not right. All of those things will cause uncomfortable communications.

Q5: When a RS start to laugh in a very weird form. I think if we are sure of a decision and the RS gives a different decision or makes a decision which for us is immoral this a moment where people can start to feel scary. Or for me what's makes me feel concern. there is robotic company called Boston robotics that they make this dog robots, they are kicking and running. Just imagine, if this dog gets out of control something goes wrong in the circuit and start to run and start to run and there is a few old man and women, you can kick them nothing happens and then continue running and can kill them. It is quite scary, because you never know how far you can control the system. It is not like it wants to go and kill you, unless you program that, but at the end of the day it is some circuit, something can go wrong. If you press a button and it supposed to switch it off and it doesn't. this is something that it is more into the hardware of the RS, on mobile RS that should be consider. Was it 2-3 years ago in Germany, in a factory which had these robotic arms, it started to push the man against the wall. The press the button, it didn't work and killed the person.

Q6: We are writing projects, and we are trying to introduce this a RS can be used in health care, in phycology in order to help people to be positive, in order to help people not to feel lonely, or other things. but we are not at such level. In 20 years, we will have an answer for that. This hypothetical, and I'm a very positive person, so eventually everything will be nice.

Q7: Unintentional fatal mistakes. I mean, it is a system, nothing is 100%, nobody can say the system is 100%, until the system is dead complete it. Then you can go back and look at the history, because for a moment Tesla. Tesla made. overall the history of all Tesla, it makes once a mistake and then one person die in a car crash. So, these are the things that humans should consider, but this is not a scary part. As I said, the probability of you dying on an airplane, is higher than you dying in a car crash caused by Tesla, but people are okay. Must of the time people are okay in the airplane. Unintentional fatal errors.

Q8: In AI, we are working in a lot of cool things. People in Google, FB do a lot of cool things. They have developed huge algorithms that no matter what part of the glass or mocks are showing, or how many mocks are there. Then we put this cat from china, it doesn't detect it … [Example inaudible and not relevant]. The current development it's for very specific and narrow field, … even when they try to generalize it, it is a very narrow field. Human is super generalizable in everything. So, I think we still have time until the moment we'll have a fight with AI. At the end of the day, we are the one, who are feeding the data, who are creating the AI. We try to make it as human, but the problem is not human, so we cannot really make another human. Of course, we can make robots talk, which are very strong per say [inaudible] there are a lot of things which they cannot do. There are a lot of things which cannot do, this is

an important thing that we have to consider. The fears, still in sci-fi movies. We are far from that, because we are far from understanding of all human behavior and human brain.

Q9: Still far from that. Although in one specific task, it was on BBC yesterday that the USA army announced the number of civilian deaths caused by the drones. So, this is a threat, because the automatic drones supposed to kill a terrorist or what it supposed to do. But killed a human, so this is a threat. Should we be scared of it, we should be. But in the very early stage of questions I said that still very close to the critical decisions, there should be a human mind behind it to take the blame if there is a mistake. Rather than say that the Algorithm did that. Because of that people are still hesitating. But consider as a threat that the AI makes an army and say, "Let's go make a war against humans, we are far, very far from that". I don't have fears right now. The only fear is privacy, by mistake my Alexa takes a photo of me and publish it on the internet. I win free publicity, but it is just a matter of privacy issues and concerns, otherwise. Threat of life, Although, this privacy could lead a person so suicide. But I don't think that there will be an army of AI or robots in front of the institute and say that they want to shoot me. I don't think so.

Interaction related questions:

Q1: Related to a previous question. I think for the critical decision making there should be, there must be human thinking. It is unavailable, and human life is more precious than letting an algorithm to decide on something. If it is about whether the coffee is ready or not and then, it makes a mistake and the coffee burns so on. I don't care much. When it comes to critical decisions, definitely humans should be involved.

Q2: Actually, an important thing is that it should not be predictable. When Google car came out, there was a question on the internet [not that you should trust everything on the internet]. There was this questionnaire about the google car and people answered that they didn't like the google car, and the reason why they didn't like the google car was because it was too safe. In the area of 30 [miles per hours] it was just going exactly 30. It was very well predictable that, if the traffic light is yellow it will stop. We humans, we are not predictable, if is yellow you go. Is 30 but you go 40, if there is a policeman you go 30. So, a little bit of un-predictiveness makes it closer to be like a human-like. I think, in the middle question - how likely to be friends -. if is very predictable is not a human. Is just a pattern that they can predict what is going to happen. It's like your dog, if you through something it will go and pick it up and next time if you pretend you are throwing, they start to run, I mean there is not funny.

Q3: Do we assume that we have a RS that does everything as humans does? if the answer is 100 years later, yeah, they should be quite social. But in another form right now the robot is developed for one specific task. The robot that works in a factory to do a specific task. so should the robot drink beer, I mean doesn't make sense at all, because is not meant to do that particular task. But another thing is that if we are making the robot to be like humans, we have introverted people and we have extroverted people. Do we make a robot to be introvert or do we make a robot to be extrovert. Why do you want to make a robot to be introvert or extrovert. So, there are too many questions that need to be answered, before we decide it has to be social or not social. So, I cannot really comment.

Q4: It should be very much emotional oriented, plus the rule base orientation. If we want to make a robot to look like a human. We should look at the context. If the robot is like a friend, it's emotional connection. If is meant to be a judge in a court, then it should not be emotions, while it's following the rules and so on. They should be very much context dependent. But emotions are playing a very important role, eventually. Currently, we are at a very high level of understanding how the robot understand the emotions of the humans, expressing emotions towards a human, expressing in the reaction form, as we are expressing as a human. Some of them are deployed and tested in more realistic environment, so the research is going on. I think, there is no ultimate limit, if we want to become close as we want to be as human. But we are quite in a good stand by now.

Q5: This is very personal. I'm expecting… because, my thought is the robot, eventually final decision should come from a human, not the human that is in [inaudible]. But there should be one centralization of decision making. And AI is not very, I think the right thing, for making the decision, not that humans can make a better decision. But there are too many constraints attached to it. I don't think I have a solid answer to this question.

Expert 5

General related questions

Q1: There a couple of ways, one way I can think of, as many machines as long the history has made lives easier in the sense that we don't have to do hard labor. I do hope that when we are creating machines, when we also have humanoid robots, that these will make our lives easier. So, if we instead of doing chores and we can go to the park and play [inaudible], that's one certain expectation I have for the technology.

Q2: We are transferring negative attitudes, that have been proven or demonstrated at least with several AI algorithms like there was even a Microsoft chat-bot that turn into a Nazi in less than 24 hours, I don't remember the details of it, but started spitting racist [inaudible] simple based

on people provoking it. It is a high risk, but in the end, we are developing these algorithms as we speak, and we are only going to deploy those algorithms that we can control. So, If an algorithm doesn't perform for us, for engineers is a of software, is a machine. We just turn it off, just like Microsoft did. Until we get an algorithm, an AI, the way we expect it to perform we are not going to deploy it.

Q3: I think we are transferring a lot of biases and I'm not sure we want to prevent this. Again, we can only look at the past, and what we have done in the past. A lot of what machines are today is a lot of what has built for them to be built. So, the machines are not the averages of all the human's expectations. They are usually the fruit of engineers' expectations who built it. So, when we are building machines that do certain tasks, they are usually was an engineer who has done this machine to do this task in a certain way. In that way there is subjective bias of the machine builder already in the machine. I don't think is necessary a problem, because when something is made into a product, it will go through market research, there will be a market for it, that means that there will be people that wants to buy this machine, that one to use this machine and they will start driving the way the technology with its newer generation, much as we see with smartphones. One part is the company that's trying to improve the technology, the other part is how the end user actually uses it. A lot of functionality that we see on smartphones today are simple the fruit of the fact that people start using smartphones in a special way, in a certain kind of way and then app developers and smartphones developers started to create technology more suitable for this way of using it.

Q4: With robotic systems it should be rather straightforward, because every machine has constraints. With humans we have free will, even in the sense that: I know not an into a wall, I can still do it. But with the machine we simple program it to not run into a wall, and if it runs into a wall it has malfunction it needs servicing. This a metaphor that you can apply to say racism or discriminatory behavior. If a robot does this, it has malfunction, it needs servicing, but your software already should have, and it will have constraints [inaudible]. If those constraints have been removed, there has been hostile engineer hacking the system.

Q5: I think it depends a lot on the context, because trust. We trust machines every day, we trust our computers to store the document for our thesis, we trust the electricity to remain on in our building, or the air conditioner to work. And we don't even ask if we trust the system. We know from our experience that it works. The same applies for robots, if we see a household robot which makes a very nice dinner for us, we trust it to make dinner for us, until one day malfunction for some reason, and then we gain call the service line and ask what is wrong with the robot. so, we are trusting machines, and we should trust them at certain tasks. But there are

certain tasks, where we don't at least know whether they are as capable as humans, but this domain continues to shift, there are things that we wouldn't trust the machines 20 years ago. 20 years ago, I didn't trust my computer to store my thesis file, whereas I would make multiple copies, right now must people trust could service to store personal files, because they know these are regularly backed up. The same applies with machines. We are driving cars where a lot of processes have been automated for us, we trust the cruise control, we trust the automated gear shifting, and we don't even question about it. But this has evolved in the past decade, and if we continue to know what are the task, that we are going to trust the machines in the future then we are definitely give more and more tasks to the machines.

Q6: Reliability again is something that it is constantly improving, once a system becomes reliable in a certain task, people will start trusting it and the more people will start trusting it, the more reliable the system will become, because more people are using and more people are developing it, it will be make more [inaudible]. When it comes to suggestions by the machine, the robot or software algorithm, then when you see YouTube, you see that recommendations are pretty accurate. Again, it depends in the type of recommendation, but we can expect that certain recommendations will be more accurate, it will get at times even too accurate, even more accurate that we want to admit. There have been some stories on how an algorithm predicted about a person was pregnant, even before the person realized it. The machines have the abilities to pick up on things that we don't. I think, again it comes down to the context. There is a very likely future scenario where machines will know better what we should do.

Q7: There are at least two sides to answering this question. One side is that we often make an argument that if a child is playing with an interactive guide, let's say there is a robot that's able to sense the child feelings, interact with the child, talk back to the child. That will actually have a negative influence on the creativity of the child, because when a child is playing with a passive toy the child has to come up with sentences that the toy is saying, now if there is robot able to generate those answers, the child, that part of the child's creativity is being block or at least not being enhanced. This is a proposed idea, nobody has been able to confirm it, this is just a potential fear that's out there and we don't know the outcome. Because, having those RS that are able to interact with the child might instead trigger some type of creativity in the child that we can't even think of right now, because we never had those type of toys. It might unblock something, that has never been unblock. Who knows? And the other side is that we know from the experience of autistic children therapy, that children are not able to differentiate between emotions expressed by robot or a human. Because, for children emotions are what they appear to be, not what weather they are true or not, is grownups that we talk about is this are true

emotions, are you actually happy when you are smiling? But for children it seems different, for them, they see a smiling face is a smelling face. And now if you put a smiling face in the robot, the child also perceives it as a smiling face. And that means that autistic children that don't necessarily want to interact with humans are more willing to interact with robots. But these robots are still controlled by humans, so that allow us a path in the therapy of those children where are able to make contact with these people, who otherwise would not make human contact. So, you can see some advantages that machines will bring into these dialogs. I would say at this point; we don't have enough information weather machines in general are going to influence our children in a good way or in a bad way. I've heard from Estonian literature history that when farms were still strong, and not everybody went to school to read books, some fathers were concern that their children's brain will break if they read to many books. Now of course, nobody will think like that, not at least those people that received state education, they usually think is a good idea to read books, but know we are saying that read books is okay, but using smartphones is a bad idea. How do we know that? Maybe in 100 years our great grandchildren will laugh at us thinking that smartphones are an evil thing. Maybe this something that an intelligent people always use, and they will have another piece of technology that they are trying to figure out weather is a bad piece of technology or not. We simply don't have the answers.

Q8: In a way it's a question that is impossible to answer because what do we mean when we say evolving AI. if we say learning a certain skill, then we already know that AI is already able to do this without human intervention, because we had like the example of this go playing AI, that was able to learn the game without playing to any human simply by learning the rules and playing against itself and then becoming a better player than the human was. That is in a way an example of evolving on its own. I think every machine is a reflection of humans. Because we only build machines for jobs that we have done in the past or what we're doing, so they are a reflection of our desires and our needs and it's still today we see that that many robotic Concepts that Engineers are starting to tackle or create are first described in science fiction literature. So, it means that there was a guy who dreamed of a machine at a time when there was no hope for getting such a machine, and now a hundred years later we have the know-how to make that machine happen

Q9: This Is again a question with many potential answers, and I think that in the future will rely heavily and robotic systems, that's a very likely scenario, unless something very catastrophic happens in our history which would deter us from any further technological development. why I think that we're going to be rely reliable robotics is exactly the example of

computers or smartphone, but also the fact that we have electricity in every building. It sounded a crazy idea when originally somebody said that we're going to Electrify all the countries we going to build power bridge, we are also going to build telephone lines from coast to coast Etc. now we always take it for granted and if somebody needs to deploy a new standard mobile sell for at a data connection. So, that we can access internet anywhere on the globe, we don't even see this problem anymore we just expected to be there. Based on these examples, we can say with some certainty that we're going to continue relying on technology. Now the second part of it like is it bad should we somehow constrain ourselves. It's easy to say like yes we should remain down to earth, but then I always bring this example that almost anybody who needs to light a fire today, will either use a lighter or a box of matches, we have forgotten the art of making fire using bone or a bird's nest or something and steal and we don't think that those people who can't use those old-school techniques are lesser people, even though they're not reliable relying on more recent technology and they are forgotten the old ways. So, as we're starting to rely on computer, smartphones and robotics it is inevitable that we'll forget some old skills, but maybe the environment that we live in will be changed so much that we are tuned for the new world that we live in. People who know how to use flint and steel or make fire with bird's nest and bones in a forest they still exist, but they are simply a minority and I think in the future as well. People who can't operate or function in this life without robotic technology will continue to exist, but it will also be okay that is if you're not able to function without robotic, because you don't need it in that world. You need some other skills in that world.

Q10: This is a very easy question to answer in the sense that there are two ways on how this is going to happen. One is simply, we have researchers in Estonia working on Estonian natural language and most often the answer to how are you going to make your computer or software better if you use certain machine learning technique and it will learn until it's good enough. So, and then it will be able to learn all the peculiarities of every culture every language and it's going to be an easy process. Of course, we're not quite there yet. Another thing is what we see with machines is that humans still are very capable Learners. So, if a machine fails we're able to compensate for this, is very likely that it's actually not going to be machine to learn to understand our cultural context, it's us who's going to change their cultural context to make it easier for the machines because we've done so in the past.

Emotions related questions:

Q1: The question here is what is a real emotion, and I know I don't think people will be able to tell the difference between a simulated emotion and a real emotion, much like even with human-human interaction. People are not always able to make that distinction, some people

are more sensitive they can tell if a person is simply putting on a face and I think it will be the same problem with machines with the exception that the machines are better hiding true emotions. Because machines do have an on-off switch for a certain thing, so they can, if you have to play poker against the machine is probably better at disguise emotions.

Q2: I think it's very easy to get attached to machine and develop feelings for a machine, we don't even have to look at robotic systems, we see this already in the world, some people explicitly state that they care for their piece of equipment or a computer game character more than a real life human being, so it is a really, I mean it's something that we see and it's something that's going to happen and that there isn't much we can do about it and not sure we going to even look at it as something weird in 50 years or something like that.

Q3: It depends what it means to have a friendship with a machine, because I'm not sure what it means. Because I can see what it means to have feelings for a machine which in the end is like a one-way system where a human is having affection towards a machine and the machine is emulating whatever the human wants in response. But with a friendship if you bring in this word, I would expect a certain type of partnership, certain type of understanding on both sides and I'm not sure, currently I'm not capable of imagining what a friendship between a machine and human looks like.

Q4: I think it depends a lot on the person but it's anything that robot can do, or a machine can do the taps into whatever that person is afraid of. So, for instance if you're afraid of a physical attack by a knife and the Machine that is simulating or doing this towards you it will make you feel uncomfortable, also if you're worried about your privacy than a machine that knows everything about you will make you feel uncomfortable or have negative impact. Also, if you're afraid that someone else is better at you at whatever you're doing in the machine is doing this a thousand times better you might have anger towards those machines that do it better than you. If anything, that taps into the fears and often the issues related to self-esteem that the humans may have.

Q5: One is example if a machine is doing something that is unexpected so what I teach in my class is that when we're talking about human robot interaction we have to understand that the moment robot and human are in the same environment, saying the same living room, the robot must behave in a way that the human perceives it a safe. So, our mathematical algorithms working in the robot might decide to the certain type of movement is okay to do. But it might not look harmless for the human, because we humans tend to think what is harmless in our own categories, so we know what is realistic and then we have certain type of risk management that is going on in our brains background, but now if we have an algorithm mathematically accurate

system that can make prediction that look risky for us but are not risk machine. so now it comes down to the question can the machine actually choose to be that's a less efficient but make a movement that appears less risky.

Q6: I think it depends again more on that person as I mentioned in the mention before, if it's the person has some self-esteem issues than of course that the robot that taps into those problems within a person again increased anxiety levels in that person, cause paranoia if the robot knows too much, etc. but at the same time the robot can make the Human feel at ease by solving problems that the humans was unable solve. For instance, if elderly people are unable to move around in their apartment and they might feel anxious because of that, they might be scared to come out of bed because they're afraid to fall over, if they now know that there's a robot that they trust in their house, that can help them up, who can help them go to the bathroom whatever, then disconnects really reduce anxiety and cause better psychological health for the elderly.

Q7: That's a difficult one because I'm a technology optimist. But I think we need to be cautious about too quick development. Because so far development has been gradual process so there has been time to get used to it, now it seems to be going very rapidly and as I said also in the past we don't necessarily know the influence of it all on our humanity, on our children and so on. So, little bit caution towards to speed or the rate of things happening. Also, when it comes to, I guess the world order and the security, if we have these machines that help us out in tremendous ways, are these machines secure enough, if say somebody wants to use them against us. I don't think that the machine is going to have up Rise Against Humanity like many science fiction movies predict, but there is a potential for mal use of Technology by other people. We need to worry about security that's another thing.

Q8: Is popular because bad news sell. it is like if you have news that say that somebody got killed It's always more interesting, if you have movies where somebody gets killed it's more entertaining. I think that's why the idea of wiping out humanity is acting. But if you talk to engineers they're not worried about this at all, you work with a robot in a lab and you think you can pull the plug anytime, why would it wipe you out and that I can't even see the reasoning from machines perspective as well like robots deciding that humans need to be destroyed seems very far-fetched, and we're going to be able to turn them off line far before they reach that conclusion. Honestly, I don't see why they would reach that conclusion, why would a machine to decide to be the only one living on this planet. Maybe in 200 years we have different knowledge about machines and how they operate at this point, I don't think this is something that we should concern us with.

Q9: Depends what we mean by threat to humanity. I mean some people might even say that our Reliance on machines is a threat to humanity, because we're not able to operate on the same level as our ancestors 10,000 years ago. but if we're saying threat in a more direct way like they're going to harm us. I mean machines can be used against humanity, against humans as I said earlier as well by other humans usually. Other than I don't think we need to worry. I guess I don't have any, I'm hoping the robots will make my life easier not tear it down.

Interaction related questions:

Q1: Depends how we define autonomy. that's another thing is I teach in my classes as well; what do we mean by autonomy and we usually want the machine to be very autonomous at the task We're trusting the machine with. A self-driving car, we wanted to express full autonomy in getting from point A to point B. I want to go to Tallinn; I use an app to get a self-driving car and it will drive me there and I don't want to tell it halfway through that the car needs to turn left or right. So, I wanted to be fully autonomous in its task, but I don't want the robot to be autonomous in the sense that it can go beyond its tasks. I can decide not to drive me there, I don't want to see this type of autonomy, which we might appreciate in certain scenarios from the human, so I'm saying like we want robots as autonomous as they can be within their task.

Q2: We need to be able to predict their task. but I don't think, much like a professional working like say a Doctor doesn't need to always explain what he or she is doing when his operating on a patient, it is more important that the surgeon gets the job done. so I think it's the same applies to machines we want to predict, we to be certain that the task will be solved, the task will be solved in a safe manner, it doesn't contradict with our general rules of ethics and general regulations but if the machine can do certain ways the task in a way that is only known to it during the execution of course always there is a possibility for the engineer to go in and troubleshoot, debug what went on. Expressing its actions all the time, so that they are predictable isn't necessary in my opinion.

Q3: If the robot is a social robot then yes it needs to be socially engaging. Again when we're talking say companion robots for elderly people, who otherwise would be alone with those robots engaging and stimulating, but then when we're talking about say deliver robots we want them to be as transparent as possible, we don't want them as a distraction, and it's okay. Some people say why would you treat a machine like that, well that's how we treat matches, the are there to do the job and we don't have to treat them as sentient beings because they're not their

devices. You don't Express gratitude towards your drawer for holding your pants and you like your drawer when it's not stuck or anything.

Q4: As I said before as well, I think that emotional interaction can become very close, because I think humans have a tendency to be attached to machines, I mean you don't even have any tasks to have an intelligent machine some people love their cars to death. so, it's just something that us humans do we love our things, and of course we going to love our robots, and if our robots are able to also mimic some of the things we want to hear, touch us the way we want to feel. Then in my opinion, this is perfectly fine because as human beings we just want to feel good and as long as our feeling of good doesn't take away some someone else's feeling of good it's perfectly fine in my opinion.

Q5: If a robot is incapable of handling an unexpected situation, the robots should communicate this very clearly. So that humans known that robot is not going to help in this situation. I think that's the key there.

Expert 6

General related questions:

Q1: Near future wise they will start improving little things because it's going to be integrated very slowly, right now we already have neural networks integrated with automatically pattern recognition and automatic detections. Things like people detection or person identification, motion identification [inaudible] and the far future this could probably going to be integrated with general use case neural networks which will be more like an actual intelligence, so currently what we have is just one specific use case and a lot of data going into it, so we don't know what it's doing. But there will be many outputs like actually can decide something. So currently that's where it’s going, so the long-term output we can't really imagine.

Q2: This is something that someone has already tested. Google and a couple of others have created a couple of chatbots, and the idea was to test the Turing test. First of it worked fine, but then it was released to the wide audience on the internet and of course there were people that would like to abuse this kind of things, so there are certain groups and once they find things out they will start specifically targeting it with specific type of information and different types of ideas and in those cases, I remember one of those chatbot was taking down and no longer published and they had to retrained from scratch, because even if they would try to retrained again how it was already trained it still had the previous ideas or previous outputs. The other one I'm not sure what happened to it, but the essential idea is that if is feed with enough information from everything then the bad part won't stand out much. They will still there and

there are people that will specifically target very specific part of it but if there is enough information, I think it can be oversaturated.

Q3: Is hard to [inaudible] because obviously the person who is going to design the RS, they are going to have their own specific biases, but the bigger problem is data collection when you are trying to train something because depending on how and when the data is collected, that data is going to have bias and if the neural network is trained in that bias, obviously is going to have it and there is no way around it. So, in that specific case, if people are targeting that specific part, they will alter the data and there will be bias on a specific opinion and that's easy to do. If you don't pay attention to it, is still to accidentally do it.

Q4: Prevent it would be a bad word. I think the easier term would be oversaturated with other ideas so it would have different opinions. There is no way to censor specific types of information, because if you start censoring then you will get the bias problem. The only way is to have enough information to it. So, it will make a proper automated decision. But the problem there lies with computer systems they don't have the human aspect. So, if they are based on statistics making a decision that would seem politically incorrect or somehow seem racist or nationalistic, it might actually be that the data being feed to the system clearly indicates that, and the computer is just going to follow that is not going to take political correctness into opinion whatsoever.

Q5: RS depending on the use case, if there something that looks over and very strong based on statistical outcome then you should always trust the computer, even though it might seem is making the wrong choice. Because computers try to get the optimal choice, they don't try to get the best one. By best, I mean human best. I human might decide that something is morally ambiguous and can't decide. And it will try to reason what might be the best answer. The robot might say what is the general speaking the most optimal choice. If you are talking about automation or general things like this. Robotics will be something you can trust. But if you have moral opinions wise decisions, if you would switch a train to hit one person instead of five. Then you probably don't want the computer to make that decision. Or you can have it, it will make the optimal decision. But the problem is that some people are going to be against it and some people are going to be for it. There is no way for the computer to explain why it was that decision, because it was literally the optimal decision.

Q6: Reliable is a hard word, because depending on the system. Sometimes a human is more reliable, but as time is moving forward robotics seems more reliable than humans and a lot of data samples nowadays; pictures of animals, pictures of people, there is this huge data

collection and datasets. People are constantly train different neural networks to classify these images and originally the images were taken by people in [inaudible] as were taken and the resolution was changed, or it doesn't look as distinguishable then they have people going over the dataset and label it. As time is progressing the automated systems are already outperforming humans in classifying these images. So, if it continues that way and we have enough data. Computer will definitely be better in those senses. But there are some parts were humans are still better at, because humans are reliable to work in these places and we are trying to fit robot to work in a human environment. So, it will take time for robots to fulfill every single need. Currently depending on the situation, if is more statistics base, trust the robot, if is more human intuition based, trust the human. But I think in the near future, like 20 to 50 years is just easier to trust the robot at that point, because will know so much more.

Q7: If a child interacts with a robotic system at a young age and it doesn't have a parent nearby, they will develop thing that might seem morally ambiguous, because robots cannot teach morals, something that is a human concept, only humans can teach it. I'm not sure robots can teach it. We might develop systems in the future that are basically robots with human morals. But I have no idea how that it would work, because that is basically based on bias. Teaching a child is something I would not recommend. At least that's my personal [inaudible]. This is a very edge case example, but autistic children they are very keen and know specific things, they are very skilled. But they don't know how to socially interact with others. I'm guessing that if a child is brought up with a robot and is not as the movies portray it, but as we portray it in science and how it will be developed and the kid that is brought up with this sort of a companion would most likely have autistic traits. Purely because they will not have proper interaction needs, at least I don't know how a computer will social in the same way a human is there are this small nuances that a kid won't pick up and once the kid is transfer to another group of humans, they will be keen to looking up to those instances that robot might have for interaction but they won't know the ques of humans. So, it will appear as the have autistic tendencies. At least that is my opinion.

Q8: Currently I'll say non, current AI is not general, current AI it does one specific thing that's designed to do, and it can barely work outside of those terms. Essentially it cannot think for itself, just a couple of neuro cells [inaudible] taking a lot of input and trying to get the best for a specific outcome. The AI you see in movies, that would be general neural networks, and everyone is trying to towards it. But that is such a high goal that need such much computational power that has such data brought into it that I don't see it happening in the next 50 years or so. It would probably go much further than that. Seeing how far humanity has evolving recently

with technology I have no idea what would happen, it might even happen that it comes to reality in the next 50 years. I don't have any doubt. Humans likes to make things like jobs easier for themselves, so as long as there is something that a human need, they will work to fulfil that job. If they have a job that they have to do as a routine, they will try to look for a way to make it simpler, more efficient and everything along those lines and robots are basically another tool for that, to make jobs more efficient. Previously, if you would dig with your hands, then you would dig with a shovel, then you would dig with a digger. AI based things are more complex that can-do automatic analysis or automatic vision analysis and things like that. They are still very simple tools; it is just how they are built together and once we get to general artificial intelligence. That's the part where no one really knows what will happen after that. Because then the human will be building it.

Q9: I think we already do. How the internet was brought up and how the world is connected, if we lose our current robotic systems and networks and all the major systems it would bring us back several years. Many functions that are currently automated would just simply not work; traffic, transportation. I don't mean autonomous car or autonomous vehicles is the path planning for systems. A lot of systems try to plan the best path, best routs. If you try to buy a plane ticket, it will give you the best option to get from point A to point B with different routes included. We would be losing all of that, this are convenient things that people depend on that nowadays for us to be efficient. Definitely no, because if we limit the dependency we'll still be leaning in old methods. Which are less efficient, and we will as species be stagnated. But robotics is just another tool. Should we prevent being complete it dependent on simple tools like access, robotic in my opinion is just another tool, the next step after that. It's very easy to depend on it, I think we should. But we also should be able to understand what it is, so we just don't think is a matrix box.

Q10: Currently that is done on an intonation bases like China is censoring different systems that they don't want people to see, different networks they don't want people to see. Germany had a long-time thing where they banned sources of blood and violence in video games. This is something that the culture would have to do by themselves. If someone invests a new technological invention and if is implemented in every single country, then every country would have look forward. People would put the knowledge of that culture so they can finally tune and work with that culture. Because this is something that no one person can do, or even a team of people can properly do. Because each culture is very different.

Emotions related questions:

Q1: It will definitely bring up a lot of moral debate on whether a computer can have a soul or be conscient. But in my opinion, this is something that it might not actually happen, because that would only happen if we try to model a computer exactly like a human. The problem is that is not actually efficient, so how robots have been model is using specific use cases and if we build a general use case, it will not definitely act like a human. It would definitely be something that it would automatize itself for whatever purpose is serving. So, think of the robot that has the only goal to create paper clips; if you create a general-purpose AI and you just give it one purpose, a very simple purpose, create as many single clips as possible. You put it next to a paper clip factory and it would obviously try to optimize it, then it will start to get access to the banking information and internet and start doing stocks, to get more materials, more resources, get better factories, then it would actually figure out that it can actually make it even better and develop drones and autonomous vehicles that will gather the resources itself. It will not care where is taking the resources from, it will just have the simple goal of making paper clips. At some point it will start to use all the resources, even humans to make paper clips because that's the goal that was set. The problem with general AI is that we'll have to define the rules very specifically and it will definitely not be the human, it will have no moral opinion outside of what we define outside manually. It will not have a moral compass that we can speak of. So, I doubt we get to the question of does it have a proper soul or not, because is not going to be human whatsoever. With the AI that I think will come out of this, no one will consider it having a soul. So, no one will question it. There might be people, if you teach an AI to have emotions or at least mimic them, people might consider that it has emotions. But it won't definitely have any, is just mimicking.

Q2: It will go along the lines of anthropomorphism, and it has been occurring for a long time. It doesn't have to be robots. Robots will make it simpler. Humans are kind of weird in that sense, they develop emotions towards inanimate objects and in case of robots, it will make the creation of the bond simpler. It thinks it will continue to happen, because it has happened in the past. There is a guy who has an actual relationship with his car, I don't think we need to step in there is a fact of life that is going to happen. The more we make robots [inaudible] and human interaction it's just going to make it slightly more prevalent. The way how to prevent it is having more human interaction in there. So, they don't actually need to go away from human interaction into interaction with an inanimate object. There needs to be more socialization between humans themselves.

Q3: Very likely. Robots are portrayed kindly in movies; people grow up with the idea that robots are this cuddling things and depending how they are created they can be these cuddling

and cute things. I don't see why humans wouldn't develop relations with a robot that is very [inaudible] with a pet.

Q4: Uncanny Valley. If a robot is designed to look like a human, act like a human, but is not complete it like a human then there is this uncanny valley effect where as soon as start approaching this human like behavior more resemble or understand how the emotions work. But at some point, there would be this giant rock and humans will think is very creepy, not complete it human, but is almost human and after that we'll go back to humans liking it. Anything that's inside this uncanny valley will be very uncomfortable for humans. Saudi Arabia a couple of years ago made one their robots an actual citizen, that robot doesn't have any neural network or emotions behind it, just a robot shell that mimic emotions. I see that robots resemble the uncanny valley problem and interaction with it seems very creepy, and if you have any robot that is like that, that have almost human like interaction with it, then humans would feel uncomfortable. Is much more probable to have robots that don't actually resemble humans, more anthropomorphic or less human like, but still exhibit over exaggerated emotions.

Q5: If a robot has power over a very dangerous system and is faulty in some way, its accuracy isn't that great or it isn't properly trained or maintained, that will make humans very cautions of it. At least that's what I would think. The problem is that things created for those dangerous applications they have so many standards and regulations associated with them that is hard to actually comprehend how a robot would malfunction. We are talking about things like medical robots or heavy equipment, machines that control heavy equipment automatically. This will be things that will be perceived as dangerous for humans. But they have so many regulations behind it that I can barely comprehend how one of these will stop functioning in a reasonable manner, because there has to be so many fail things in different systems. Even satellites, things that are sent to space, they need fail safe, like a separate [inaudible] complete it from the original communication system that you can send information in order to complete it shut off the satellite. Things that are machine critical, things that are very dangerous or have high risk, they have several fail systems. Humans might not comprehend that these have so many fail systems and they don't know how the system works, but they usually do.

Q6: Paranoia is something that's always going to be there, there are people that think they are being spy on. With robots being more integrated they are going to basically right, that they are being spy on. The problem there is that many people can't comprehend the idea that there is too much data. So, no one is going to be sifting through that data to find something specific about you. There are automated systems that will sift through it and generate automated opinions about you to classify you as a person, to give you the best shopping list, like things

like amazon and google do it. These things are already done, and people are paranoid about it. The problem is that is really hard to actually hide yourself from it, I think there is no point. But I don't know how a person who is paranoid about it, how they will interact with this. Because it will get harder and harder to actually hide yourself.

Q7: No proper management of the system, so basically not looking over about how the system was created, for what purpose. That is definitely the biggest one. Overdependence, where humans can be over dependent on machines, were they complete it cannot function without the machine. There should be some level where human can fall back to without using the machine or at least some part of it, if there is some blackout, taking out communication, solar flare, things like that. Then we shouldn't have society just ending because of it.

Q8: It's a popular science fiction idea, because it creates an enemy to fight against, or something to have in science fiction as a motivator. But as reality I don't think we'll have a turning point where there is a robot revolution [inaudible] but it will probably like I described previously with the paper clip factory. Or depending where we develop, if we have something that is general purpose AI, that is out of control. Then we'll get the paper clip problem. If we go to the other direction, instead of AI we develop robotic then they get out of control, then they will have another type of event scenario. Which is, we start creating nanobots and we start creating robots that can replicate themselves. At some point we'll reach near the [inaudible] event, the problem with that if nanorobots that can replicate themselves become widely available, then someone technically could reprogram such a device and make it replicate itself non-stop with the replications having the same goal continuing this. There were some calculations done with this rate, and if you multiply this rate with robots that are 1mm long it will take about 7 - 9 days to envelop the entire planet. It's an exponential growth. So, there are these two types of event scenarios, that I think it might happen. Both of these are so far and out of scope currently. We haven’t been able to make complete replicated systems, the closest we have are 3D printers and they still need specific materials to create themselves. These are more into space technology things so that can create other planetary bases. General purpose AI is so far off that I don’t think it we’ll reach any time soon, at least not in the next 50 years probably. Two scenarios but both very far off.

Q9: If we don’t regulate them, as we reach higher level of robotics and AI technologies, then the scenarios I just described will become more and more probable. They will become widely available to the public. There will be people with malicious intentions, there will always be people with malicious intentions, they will just probably seek attention or probably destroy something, and if people like those can put their hands on things like these, and they are simple

enough to reuse and reprogram and we don't have measures in place to actually combat these, then we'll reach one of these scenarios very quickly. It all depends on what type of regulations we have and what source of contra measures we have at the same time. Paper clip factory example. Great growth - replication idea of robotic.

Interaction related questions:

Q1: We keep trying to give more autonomy all the time, the more give it the better they will be, because they will have quicker reaction time and better decisions. The only problem is that technology hasn't reach that point yet. Some people are trying to push autonomy further than the accuracy can actually handle. Autonomous cars for instance, they have been pushed, I don't think they are ready yet. We don't have proper accuracy, proper computing technology, but very soon we'll have them. Is just that the current systems, they just assume a couple of scenarios. But in the real world there are a lot more scenarios than those and if there are scenarios that happen that are not assumed. Then the machine will start to approximate it or will start to assume itself what would happen in that scenario. Currently because we have so little data, even though we have loads of data we have so little relatively speaking that they don't make the best decisions in a lot of cases**.**

Q2: It's hard to answer that one, robotic should be very predictable, that's what they meant to be, that's what they do. AI technology is something that is supposed to be predictable in a way, but up to a degree. Everything beyond that is the idea itself is not predictable is not something that a human can fully comprehend unless they are going through the network note by note. I don't think AI network should be predictable, they should just function within their confines, at least predictable inside these confines, but outside of it, it doesn't really matter.

Q3: Social interaction is something I think humans should still dominate. There are a couple of factors where robots can automate; Automatic tellers, automatic care for older people. I would not recommend automation for child development. But automatic control systems that interactive. As long as the person knows they are interaction with a robot, not another human being, not another conscient creature is fine. As long as you put the robot interacting with something that probably doesn't know what's interacting with it, then I think is a bad scenario.

Q4: They already are and will become closer but should is something I don't think I should answer. I mean, I don't think there should be a limit, some people wanted one way and some people wanted it other way, and some people want to limit how other people will interact with it. I think that's the only wrong answer here, limiting other people's options. I should say, there shouldn't be any limits.

Q5: If the situation is complete it unexpected then there is no way a human can program a robot that respond to it, because in that case the scenario will already be expected. But if is something complete it unexpected. The robot behaves in a way that doesn't fit with a human with the correct, then we'll have to consider how the robot was made, and what was the [inaudible] in that case. Because, general speaking is still a tool, and sometimes tools don't work if you are using it incorrectly. In that case it will be our responsibility, because we are the ones using it incorrectly.

Non-exclusive licence to reproduce thesis and make thesis public

I, Iris del Carmen Izaguirre Orozco, 18/07/1988,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

ROBOTS AND HUMANS: ATTITUDES TOWARDS "BLACK MIRROR" EFFECT

Supervised by

Elina Kallas

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, 23.05.2019.