Table of contents

Abstract

In this thesis, I present my research on artificial intelligence (AI) and predictions associated with it. I also examine one of its current forms – Watson, cognitive technology developed by IBM, which has gained worldwide attention after winning 'Joepardy!' over human contestants.

Further, I utilize my analysis of current development in AI – Watson, in order to determine if AI is capable of developing expertise. For that purpose I analyze what constitutes expertise in humans and examine if the necessary constituents are present in Watson, presuming that current AI is generally build modeling human cognition, considering that it is a cognitive technology.

1. Introduction

1.1 AI

It is not an easy task to define what artificial intelligence (AI) is, since it incorporates in itself vast amount of different things and is expressed through various practices in various heterogeneous networks that consist both of human and non-human actors. In order to define AI in general terms I will use Frankish's and Ramsey's (2014) definition of AI:

> "Very generally, artificial intelligence (AI) is a cross-disciplinary approach to understanding, modeling, and replicating intelligence and cognitive processes by invoking various computational, mathematical, logical, mechanical, and even biological principles and devices."(Frankish and Ramsey, 2014:1)

There have been expressed many different opinions in the scientific community across various disciplines about the emergence of artificial intelligence (AI). Many have tried to speculate and foretell what this emergence will entail for humanity, among them such well known authors as Vinge (1993) and Drexler (1984). In their contributions excitement and fear have been intertwined in an anticipation of something big that will take place as a result of AI evolution. According to them, AI has potential to elevate entire human race to yet unknown and not entirely imaginable technological heights. However, it may also contribute to the total extinction of humanity, or, as a less apocalyptic scenario – AI may consider us being inferior to it and thus end up treating us the way we are treating animals: as pets or as a supply of proteins (movie "Matrix" would be a perfect example of such scenario). However, they also predict that AI might have the capacity to contribute to prosperity, longevity, ecological sustainability, creativity and overall improvement of human lives. In the further sections of this thesis, I will analyze some of the major points from predictions of Vinge and Drexler, made decades ago about AI's evolution: opportunities and threats that they saw in AI, as well as some of the suggestions of possible ways to deal with potential threats. I will further compare them with the current capabilities of AI and with predictions made by those who are actually working very close to current AI technology and have direct knowledge of its capabilities. Through this comparison I want to find out whether some of the predictions did come true, and if so, what could this mean with regards to the rest of the predictions: should we be excited or concerned, or perhaps both. Based on the results of the comparison, I want to come up

with suggestions on how to address the issues and embrace the opportunities expressed in those predictions in a more efficient way, instead of disregarding them or blindly embracing them without considering the consequences.

Human intelligence is a driving factor in technological evolution, however, just as we influence technological evolution, technological evolution influences us, thus, creating an intelligent technology - artificial intelligence, may potentially influence our own intelligence. Going in line with this idea, we can refer to the discussion carried out by Birkbak (2013) in "Why all Anthropology should be called Techno-anthropology", where he highlights arguments first presented by John Dewey, a philosopher – pragmatist, whose statement with regards to technology could be summarized as following: technology creates our understanding about the world by giving us the means by which we can make sense of it, by treating it systematically and by understanding the connections between actions and consequences. Thus, it is possible to say that technology, being a non-human, gives us means to exist as human (Birkbak, 2013:122-127). From this we can extract two main aspects: the existence of mutual influence – we influence technology and technology influences us, and the fact that we establish the connections between actions and consequences through the help of technology, or in other words, we use technological perspective for that purpose. With this I am going to follow Dewey's notion of absence of any distinction between physical and psychological technologies (Birkbak, 2013:127). I will try to utilize this approach by applying it to human cognition in order to view it as a set of "technological processes" which are taking place in our minds and shape our behavior, contributing to our skill acquisition and our ability to acquire expertise. In a similar manner, I will attempt to dissect cognitive processes that were employed in AI in order to make it a cognitive technology. I will try to draw a parallel between human and AI cognition in order to compare the processes that are taking place, and see what similarities and what differences I may encounter in human and AI cognitions, and how it may contribute to answering the question about expertise: how AI may affect the current notion of expertise?

Technology creates and shapes who we are on an individual level, it has the power to reshapes society, influence our psychological and physical states, thus, it is important to discuss new technologies at an early stage of their development, in order not to overlook what potential changes they may bring along, and try to determine whether those changes are desirable or not, and if not, what could be done in order to deal with those unwanted changes, in case the process of development is too far advance and it is impossible to consider complete obliteration (Collingridge, 1980:16).

It is necessary to continue the discourse about where a breakthrough in AI technology might lead us, since AI evolution is inevitable, and our future depends on our ability to predict and utilize these predictions: will we be victorious or will we be defeated, will we be the masters or will we be the slaves, or maybe something in between - will humans need to constantly compete in the infinite race against AI through such means as an augmentation, or perhaps there will emerge a third species – a hybrid consisting of human and AI merged in one body and in one consciousness? There are many possible outcomes,

but let's examine some of them through examining some of the interesting aspects discussed by some prominent authors in AI theory, and let's compare them with what the current version of AI, namely Watson, has to offer, and what creators of Watson anticipate Watson to be in the near future. This, hopefully, will enlighten us about where we may be positioned currently with regards to AI evolution and where we might be heading next.

1.2 Watson or Watsons?

Watson is a cognitive technology, which takes its inspiration from human mind, and which is a product of many scientific minds in IBM, who were collaborating with universities and other partners in a joint cross disciplinary research. Watson is an AI that learns from data. Watson is capable to extract knowledge from a large corpus of unstructured data with the help of algorithms. According to Kelly and Hamm (2013), "Watson learns the way people learn, by experiencing a lot of things and drawing inferences and lessons from those encounters" (Kelly and Hamm, 2013:24).

Next, I want to introduce my interviewees and their view of what is Watson. Each of them presents and thus enacts a certain reality of Watson from the perspective of their own practices. Thus, the same Watson is expressed in multiple different versions with each interviewee's description of it. This happens due to the fact that each interviewee has a different set of practices related to Watson, so they enact different aspect of it through their practices. Moreover, practices often change over time, thus realities change as well, which leads to *ontological multiplicity.* Thus, it is possible to say that Watson is a set of different things in different practices, because it is being done and being enacted differently in those different practices (Law and Singleton, 2014:7).

Each of the interviewees was asked the same question – what is Watson? The following are their answers reflecting what Watson is for each of them.

Watson, according to X1

X1 is innovation executive and chief technologist at IBM. He is the expert in and advocate of Watson in Denmark. He has a master's degree in financial accounting and taxation. He has been a management consultant for approximately seventeen years. And here is how he defines Watson:

> *X1: "So, Watson is a, it's a technology, it's a, I think most of it, I would say, it's a software technology, it's an artificial intelligence that, you can say, thinks like we, as humans think or resonate [reason], I would say, not think, but resonate [reason]. So, it's a cognitive technology that is capable of analyzing very big amounts of data, unstructured data, that understand natural language, that is able to answer questions by generating a lot of hypothesis and evaluating these, evaluating the resources, and that can give you an answer and confidence level on how much it's, it believes it's the right answer, and it's a system that learns and improves all the time by itself and by the knowledge that you make available to it. So that's, I think, that's the core of Watson. And then, of course, this core capability we have, we work with companies and the organizations to develop specific kinds of use cases and different kind of flavors of using these capabilities, these basic capabilities, yeah." (X1, 2015)*

Watson, according to X2

X2 has been working in IBM on various sites of the organization for around four years. Recently he has spent a lot of time talking to pharmaceutical companies about Watson. X2 has a BSc in physics, MSc in biotechnology, and PhD in molecular diagnostics. Previously X2 has worked for a couple of years at a consulting organization for the pharmaceutical industry. He has also worked for about twelve years at 'GlaxoSmithKline' on the sales and marketing side of pharmaceuticals and he has worked for two years at a company called 'IMS Health' which specializes around data for the pharmaceutical industry. X2's definition of Watson is:

> *X2: "So, Watson is the flagship output of IBM research facility at the moment. It has recently, has been made into its own business unit, so IBM is now putting the whole capability together in such a way that it can be used by other companies, but it was born, Watson came out of an IBM Grand Challenge. Now that, typically every ten years IBM looks at something which is significant in terms of technology capability, and addressed it as a challenge to result. The most famous of those is the computer that IBM built to beat Garry Kasparov at chess. Now, by beating Garry Kasparov at chess, it received a lot of profile and publicity about what it could actually do, but the thinking behind creating the computer at that time to beat the chess master was speed of calculation. The kind of hardware that's required to play chess at human speed and be able to go through the calculations were all very relevant to other parts of technology and that computer was called 'Blue Gene', and that was used to calculate protein folding, and crystallography, and to really understand nanomaterials, and that is a, that was the application of the technology that came out of that Grand Challenge. So, Watson came out of the next Grand Challenge after the Garry Kasparov, which was: can a computer, can a machine understand and respond to natural language questions in a way that it could beat or compete at least with humans in that kind of scenario. So, we applied Watson in a game show called 'Jeopardy!, and it's the challenge there is clearly to understand the questions, to understand the context, to extract the right information, to have confidence that the answers to what that information is pointing to are the right ones, and then to respond with the answer in a time scale that beats it's human competitors. So, there are number of key elements around winning this challenge that computing, would take computing to the next level, which is what happened, it was successful." (X2, 2015)*

Interesting to note that X2's version of Watson is very close to the one described by Kelly and Hamm (2013) despite the fact that he has not read the book (at the point when the interview took place) (Kelly and Hamm, 2013:27-31).

Watson, according to Y1

Currently Y1 is a PhD student at DTU and is writing his thesis about intelligent text algorithms for search engines. He is also working on developing search engine for rare diseases, called 'Find Zebra'. He was invited as a teaching assistant to give an introduction to students about Watson during the lecture dedicated to Watson, as a part of Advanced Machine Learning course. Previously Y1 has studied computer science at the University of

Copenhagen, where he wrote his thesis in applied mathematics. He has also spent some time in the industry, where he worked as an analyst in Danske Markets, where he was creating risk models for use in pension funds and large firms. Here is Y1's definition of Watson:

> *Y1: "Watson is like the state of the art in text search, I think. I think they've tried almost all the models that you can use up till now, on this Watson project, but it's not that very well known as last - Watson has very poor social skills. They have not been that good at networking in, for example, in the industry and I haven't really seen it used a lot of places. I know they are used in some hospitals, but... For example, they have built this medical Watson."(Y1, 2015)*

Watson, according to Y2

Y2 is a senior researcher at DTU. He is attached to research called "Neuro twenty four seven" - a research project related to continuously monitoring information about the brain from EEG devices. He has a background in civil engineering from DTU, from where he also has his PhD. Following is Y2's definition of Watson:

> *Y2: "So, for us, it's a cloud service with some natural language processing that allows you to do, what it's called, passage retrieval based on natural language formulated questions, so it's not directly, yes… you might have seen TV, YouTube, where the 'Jeopardy!' version allows you to specific questions with specific, with short answers. What we are returning from cloud version is a passage from the document retrieval. So it is more related to ordinary search engine, internet search engine, or search engine that you install on your own computer and index your own databases. " (Y2, 2015)*

Each of the interviewees enacted different realities of Watson, some of those realities appear similar while others not. The reason for this is that each of them, even though they all talk about Watson, is unlikely talking about one and the same thing, because Watson is enacted differently in their different practices. The difference would be even more obvious if I would have interviewed IBM's clients who would use Watson in their practices, such as doctors, customer support operators, or bankers, each of them would enact Watson into yet another reality, opening up the complexity of Watson and its *ontological multiplicity*. With my research and my thesis, through the practices that I have performed, I am enacting yet another reality of Watson. All the reasoning above is a result of Actor Network Theory (ANT), which I intent to use in order to figure out how different realities of Watson enacted through different practices relate together in heterogeneous practices (Law and Singleton, 2014:9).

1.3 Problem formulation

Do predictions made about AI decades ago resonate in today's AI – IBM's Watson? If yes, then what significance do they hold for the overall evolution of AI? How could they be utilized in a beneficial manner in order to minimize the risks and maximize the benefits which may come as a consequence of AI evolution? What cognitive technology such as Watson can offer to experts and how it may influence current notion of expertise?

1.4 Getting assess

Getting access to the phenomenon under study is a very important step in the research process. Well known and classical anthropological example that illustrate the importance of getting access is Geertz's (1973) account of how he and his wife finally become insiders and have been accepted by Balinese villagers after a period of alienation. During one of the illegal cockfights, when a police showed up, Geertz and his wife have expressed their solidarity with the locals by running away from the police together with them, instead of simply using their distinguished visitor status to get away from the trouble. This was a turning point - Geertz and his wife got access, which influenced their entire research:

> "It was a turning point so far as our relationship to the community was concerned, and we were quite literally "in". The whole village opened up to us, probably more than it ever would otherwise (I might actually never have gotten to that priest, and our accidental host became one of my best informants), and certainly much faster."(Geertz, 1973:556)

In the following I want to present somewhat detail and rich description of how I got access to Watson in order to illustrate the anthropological side of the research and give an inside into my "methodology" of getting access. For anonymity reasons I will not mention the names of my interviewees and use the code names for them instead.

Initially I did not have a well-defined research question, and I knew that its definition would greatly depend on where and whether I would get access. Perhaps I could have done purely theoretical research, however, I wanted to work with actors that are directly involved with AI, and after making an inquiry into possibilities of getting close to AI in Denmark, I have discovered that IBM was entering into collaboration with Technical University of Denmark (DTU) on having a course covering self-learning IT program – Watson:

> "As the only university in the Nordic countries, DTU has been selected by IBM to offer courses in self-learning IT program Watson…From spring 2015, MSc students at DTU can specialize in a new form of cognitive computing… Watson is an advance computer program which can understand questions while creating huge data volumes in a flexible way, just like a human brain. But much faster." (DTU, 2015).

This sounded exciting and I have attempted to establish contact with DTU and those responsible for the course, with hopes to get access to Watson. However, getting access turned out to be not that easy. I have not received a response to my email, where I have asked for possibility to collaborate with DTU on my thesis through making observations of the course and conducting interviews with lecturers and students. Not knowing what

to do, I returned once again to the article on DTU website, and this time, after reading it I decided to try contacting IBM directly. I guess one can simply be lucky, or so I felt, when, after making a phone call and leaving a voice mail with a brief description of who I am and what I want with Watson for my thesis, after few days I have received a phone call back. It was X1, head of innovation department in IBM, Denmark. After giving a bit more detail explanation of what I was interested in and a brief, friendly talk about AI and new technologies, X1 assured me that he would be willing to help. Later I sent him an email with my ideas on possible directions for my thesis with regards to Watson, and asked him for an interview.

Few days went by and after not hearing back from him I decided to send him a reminder email and also gave him a call. Once again, I was greeted by his answering machine, where once again I have left a message, asking if he still was willing to help me with my research. This time I did not have to wait long, X1 called back after a short time, saying that he thought one of the ideas for my research direction was interesting (he thought it would be interesting to do a research on Watson's application in medical area) and that we could have an interview. This was my initial way in.

Later on X1 gave me the names and contact details of the few people from DTU, who were working with Watson, as well as the name and contact details for the person who is working with Watson at IBM, United Kingdom. X1 promised to try to figure out if I could interview some of IBM's clients in the medical field (I am not entirely certain whether it was pharmaceutical or medical clients). However, knowing the sensitivity and secrecy often prevailing in these kind of fields, due to high competitiveness, I was not surprised when his attempts were fruitless.

I have written emails to all the people, whom X1 suggested me to contact, presenting myself and my thesis and asking for the interviews. I have also mentioned that I had an interview with the X1, in order not to sound as a complete outsider. I got a response from one of the people at DTU, whom I shall call Y1 and also from the person from IBM, UK – X2.

New contact in DTU has suggested to me that I could come to the lecture, where, as a part of Advanced Machine learning course (the one I have originally been striving to observe), he was supposed to present Watson to the students. I have accepted his suggestion and made an observation of that lecture, and I have also interviewed him few days after the lecture.

In the lecture I met the person, whom I have tried to access initially for possibility of collaboration with DTU on my thesis. We had a brief dialog during the break in the lecture and days after I made additional attempts to get an interview with him, however, my efforts were fruitless.

After participating and observing the lecture I felt I was somewhat an "insider", so when I sent an email to one of the people who presented the theoretical part behind Watson to the students, and told him that I participated in that lecture and wanted to have an interview with him for my thesis, I was not surprised when after a few days I received a

positive response which resulted in an interview. Hence, I did not get discouraged by my previous fruitless attempts to have an interview with another person from DTU. I mention this, because being a single researcher, breaking the ice in the unknown waters may appear to be a challenging endeavor. Nevertheless, I felt victorious and an email response from the person in IBM UK with acceptance towards my request for the Skype interview added to my happiness.

I have embarked on this project lead by my curiosity and love for new and exciting technologies, of which AI seemed to be the most exciting for me for many years. And I have managed to get access to Watson through getting access to people who are involved with Watson. This resulted in a decent amount of empirical data, analysis of which I will present in further sections.

2. Theoretical approaches

2.1 Triangulation

In my research I have decided to use theoretical triangulation, which can be defined as the use of different theoretical approaches in the same research. My incentive for doing that was the fact that triangulation allows to conduct the study using different theoretical lenses and questions, in order to test the findings. Different theories used in theoretical triangulation do not necessarily have to be complementary in their viewpoint, and their selection depends on the choice of the researcher, which in its turn depends on his/her intended goal (Denzin, 1970 in Thurmond, 2001:254). One of the benefits of theoretical triangulation is that it can provide broader and deeper analysis of findings. Another potential benefit is that looking through the lens of different theories researcher is challenged to look beyond the obvious explanation. Triangulation helps to rule out competing hypothesis and increase confidence in developing concepts or constructs in theory development (Banik, 1993 in Thurmond, 2001:256). There have been expressed various critiques in regards to theoretical triangulation. Issue with it may be that it may cause confusion if theoretical frameworks are not adequately defined initially, especially in case when opposing theories are used (Banik, 1993 in Thurmond, 2001:257). Another issue could be that the use of various theories in the same study may be epistemologically unsound and faulty (Lincoln and Guba, 1985 in Thurmond, 2001:257). However, I believe if this critique is being considered and accounted for, theoretical triangulation can be successfully used. One more benefits of it is that it can raise the researcher above the personal biases that stem from single method or theory use (Denzin, 1989:236 in Flick, 2007:179). Thus, I intend to combine theories and methods carefully and purposefully, with the intention to add breadth and depth to my research and not for the purpose of finding the objective truth (Fielding and Fielding, 1986:33 in Flick, 2007:179).

For the epistemological base of my research I have used constructivism position within Science and Technology Studies (STS), which concentrates its attention on ontology instead of epistemology. Originally STS has elements of both - postmodernism and constructivism. It states “that knowledges are socially constructed”, which is the belief of the Social Construction of Technology (SCOT) that belongs to postmodernism, however,

it also states that realities are being enacted, and this belief belongs to Actor Network Theory and feminist material semiotics, which are representatives of constructivism (Law, 2009b:3). SCOT and ANT, though both belong to the branch of STS, believe and preach different things. SCOT, just as its name suggests, believe in the social construction of technology, giving the supremacy to social over technological, and suggesting that different social groups are involved in the process of stabilization of technological artifacts through constant negotiations, till it finally reaches the point of stability (Pinch and Bijker, 1987). On the other hand, ANT does not have the distinction between human and non-human actors (which includes technology), meaning that both human and technology influence each other, thought different heterogeneous networks and through a variety of processes and practice that take place in those networks (Law, 2009a:141).

2.2 ANT –analytical toolkit

As one of the analytical tools in my research I have used Actor Network Theory (ANT). ANT name can be confusing, since it implies that it is a theory, however, it is not:

> "In 1999, Bruno Latour argued that, despite the name, it does not count as a theory. Certainly it is not predictive, and it does not offer social laws. But if it is not a theory, then what is it? One plausible response is that it is a tool-kit for thinking about and studying the social; more particularly, that it is a tool-kit for charting practices of association." (Law and Singleton, 2014:3)

As I have mentioned in the previous section, ANT does not accept the distinction between human and non-human actors, and it believes that they all are connected in heterogeneous networks:

> ".. the actor network approach describes thus the enactment of materially and discursively heterogeneous relations that produce and reshuffle all kinds of actors, including objects, subjects, human beings, machines, animals, "nature", ideas, organizations, inequalities, scale and sizes, and geographical arrangements". (Law, 2009a:141)

The analysis of literature related to Watson greatly expanded my initial overview of the actors that are related to Watson and AI in general. It also led me to the discovery of new areas of interest for my research. Initially I have created a table with the list of actors/key-words I came across with in order to keep an overview of them all (Table 1). As my research progressed, the networks of actors, that I've come across within my inquiry into the background of Watson and AI, were expanding: one network lead to another, and so on.

Table 1 – List of actors/key-words related to Watson and AI in general

Actor/key-word	Actor/key-word	Actor/key-word	ctor/key-word
"Brain box"	Data analytics	Logistic regression	Scanning tunnelling microscope/STM
Algorithms	Data-centric computers	Machine learning	Science
Algorithms	DC	Mapping	Scientists
Algorithms and Machines	Discovery machines/computer-aided discovery	Microchip	Semantic mirroring
Analytics software	Doctors	Microelectronics	Semantics
ASTRON	DOME	Microprocessors	Silicon
Atomic-scale computing	DRAM-memory chip	Moore's law	Simulations
Atoms	Dual-process model/theory	Nanocomputer/living cell	SKA
Automated engineering systems	Energy efficiency	Nanophotonics	Smart grid
Banks	Entity analytics	Nanotechnology	Software
Big data	EU human brain project	Neural simulation	Spin effect/spin
Bit	EURISKO	NORA	Stochastic optimization
City apps	Experts	systems/urban operating	Stream analytics
Cloud	Fiber-optics	Passage retrieval	Students
CMOS integrated	Field-programmable	Patients	Supercomputer
cognitive and social psychology	First Of A Kind/FOAK	Pervasive analytics	SyNAPSE
Cognitive chip	Government	Photonics	Technology
Cognitive chip	Graph database	Physics	The hybrid "memory cube"
Cognitive computing	Hardware	Physics of computing	Transistors
Cognitive Enterprise Lab	Heuristics	Politics	Translation
Cognitive systems	Holistic approach to data	Probability theory	TrueNorth
Cognitive-Experiential Self-Theory (CEST)	IBM	Public	Turing test
Companies	Insight in Motion	Quantum bit/qubit	Uncertainty
Computer simulation	Integrated circuits	Quantum computers	Universities
Context	Intelligent operations	Quantum computing	working memory
Context engines	Internet of things	Quantum machines	
Contextual analytics	Jobs	Quantum mechanics	
Cyc	Language	Quantum physics	
DARPA	Living laboratories	Quantum-mechanical devices	

Thus, my list of actors/key-words was growing as well, and even at the end of my research it would be false to call it complete. However, even incomplete list is particularly interesting and useful, since it gives an overview of the actors involved in AI subject, it also visualizes the extent and complexity of the AI related networks. One of the main sources for the actor/key-word list in Table 1, among the others sources, is a book titled “Smart Machines: IBM’s Watson and the Era of Cognitive Computing”, written in 2013 by John E. Kelly lll – the director of IBM research and Steve Hamm – a former technology journalist and a current writer at IBM.

I have chosen to use both actor and key-word notions, since each of them refers to certain reality of the particular entity entered in the list: each entity is an actor situated in a particular network, but also a key-word, which turns results when it is googled or searched for in the literature. So, term actor identifies the entity, while term key-word signifies that

it also a mean to identify where it is situated in the network. If I were to explore the list further, I could have specified the relations between actors, however, it would be a cumbersome work, since relations in the networks are changing, thus enacting new practices and new realities, resulting in *ontological variability and multiplicity* (Law and Singleton, 2014:7).

In the Introduction section I have stated that human and technology, both influence each other, and even language or thought processes can be seen as technologies, which means cognition and cognitive processes can also be seen as technologies and thus can be treated as actors (Birkbak, 2013:127).

As I have mentioned earlier, using ANT I have explored different actors and some of them lead me to the discovery of yet new actors. For instance, reading about early AI systems I have come across the fact that heuristics were used in them, which lead me to look into heuristics, which in its turn lead me to look into cognitive psychology and dual processing, which then lead me to CEST, from where I further established connection between rational system, experiential systems and expertise, heuristics, and intuition. I have applied the same rationale to interview analysis, where certain elements of Watson lead to other elements and when united in networks enacted other actors mentioned previously, such as expertise, intuition, heuristics, etc.

So, it is safe to state that each of the actors in the Table 1 influence and enact each other into being in one or another way, and are connected in heterogeneous networks. In the following sections, staying true to admitting my own biases, I will try to enact my version of AI and Watson's reality and show how I experience and perceive it, and I can only hope that it will illuminate important aspects which should be taken into consideration, since elements of reality enacted by me might be shared by other realities enacted by others experiencing Watson and AI, now and in the future.

2.3 "Five stages of skill acquisition", heuristics, intuition, CEST and TempT – toolbox to understand expertise

Since Watson is a cognitive technology and my research question is related to expertise, which in humans is acquired through cognitive processes and experience, it is interesting to explore some elements of cognitive psychology with these regards. When making an analysis of some literature related to AI, I have stumbled upon the fact that in order for computer systems, who were the predecessors to current AI, not to rely entirely on human knowledge and instead gain a level of independence and flexibility, heuristics were applied in a system called EURISKO:

> "..EURISKO is designed to explore new areas of knowledge. Guided by heuristics – pieces of knowledge that suggest plausible actions to follow or impossible ones to avoid; in effect, various rules of thumb. It uses heuristics to suggest topics to work on, and further heuristics to suggest what approaches to try and how to judge the results. Other heuristics look for patterns in results, proposes new heuristics, and rate the value of both new and old heuristics. " (Drexler, 1986:62).

It is interesting to note that heuristics is a well-known phenomenon in cognitive psychology and can be defined as subjective probability judgments which "are relied upon in uncertain decision making situations where there is no certain answer immediately obvious, and where all possible answers cannot be identified"(Cioffi, 1997:203). Some argue that heuristics directly affect intuition in decision-making situations, which could be true due to the fact that intuition is a sign of expertise and expertise is ability to eliminate many steps of thinking process, and welding whole sequences together, which could be summarized as ability to simplify complexity, which leads back to nature of heuristics (Cioffi, 1997:207). Dreyfus and Dreyfus (1988) elaborate on intuition with regards to expertise even more in their famous "five stages of skill acquisition":

> "When we speak of intuition or know-how, we are referring to the understanding that effortlessly occurs due to discrimination resulting from previous experiences." (Dreyfus and Dreyfus, 1988:28)

This leads to heuristics being an important part of expertise. But to expand overview of the network of cognitive concepts from looking into heuristic, intuition and expertise I propose to examine yet another topic in cognitive psychology – dual processing. Dual processing is differentiation between two types of cognitive process: those that are fast, automatic, and unconscious, and those that are slow, deliberative, and conscious (Evans, 2008:255).

I have chosen to look at dual-processing through the perspective of Cognitive-Experiential Self-Theory of Personality (CEST). Originally dual-processing, as explained by Evans (2008) infers the existence of two types of cognitive processes: System 1 - fast, automatic and unconscious, and System 2 - slow, deliberative and conscious (Evans, 2008:255). Epstein (2003) presents CEST as an integrative theory that is compatible with many other theories among which are learning theories and modern cognitive scientific views on information processing. CEST follows the dual processing model and states that there are two different, independent, but also interacting and parallel to each other cognitive systems: preconscious *experiential system* and conscious *rational system* (Epstein, 2003:159). Thus, we can establish the link between the terminology of the dual-processing theory and CEST: *System 1* corresponds with *experiential system* and *System 2* with *rational system.*

Higher order animals have adapted to their environment over millions of years by using *experiential system*. Just as them, we use *experiential system* in precautious, effortless, automatic, holistic, minimally demanding of cognitive recourses way. We do that through encoding information in the form of memories of individual events, especially those that were emotionally arousing, and by more abstract general ways. In fact, *experiential system*, though being a cognitive system, is based on directing the behavior in a manner that would facilitate positive and avoid negative affect, without using any logical inferences (Epstein, 2003:160). *Experiential system*, which regulates the everyday behavior is dynamic and emotionally driven - the affect determines what is attended to and what is reinforced, thus, affect plays a key role in *the experiential system* (Epstein, 2003:161).

Rational system is unique to humans and has a brief evolutionary history. It operates in conscious, analytical, affect-free, effortful, highly demanding of cognitive recourses manner, and it depends on the level of understanding of rules of reasoning and of evidence. *Rational system* employs the use of language, whereas experiential system does not depend on it (Epstein, 2003:161).

When looking at those two systems, *experiential* and *rational*, it is possible to recognize elements of them in each stage of Dreyfus and Dreyfus' (1986) "Five stages of skill acquisition"- detail description of expertise. It is interesting to note that when learner or practitioner reaches a higher level of expertise he/she mostly utilizes *experiential system*, whilst in the initial stages of skill acquisition *rational system* is being the one that is used more. But let's look at it a bit closer. According Dreyfus and Dreyfus (1988) there are five stages of skill acquisition: *novice, advanced beginner, competence, proficiency* and *expertise*. Each level has its own characteristics. For instance, *novice* stage is defined by the application of rules to the elements of the situation, which are clearly and objectively defined in order for the novice to operate in a context-free manner - without the need of referencing to the overall situation. Thus, operating in such manner, through application of precise rules to the unambiguously defined context-free elements, can be referred to as *information processing* (Dreyfus and Dreyfus, 1988:21). This stage can be thus considered to be driven by *rational* cognitive system.

As novice advances to *the advanced beginner* stage, he/she gains experience, which becomes more important than any form of verbal descriptions. Experience allows him/her to recognize meaningful elements in certain situations, which then can be defined as situational elements that are no longer only objective and context free elements (Dreyfus and Dreyfus, 1988:22-23).

Once *advance beginner* gains even more experience, he/she meets with an overwhelming amount of context-free and situational elements, which present difficulty of prioritizing those elements. Thus, once an advance beginner reaches *competence* stage, he/she is able to implement hierarchical decision-making procedures – "by first choosing a plan to organize the situation, and by examining only a small set of factors that are most important given the chosen plan, a person can both simplify and improve his performance" (Dreyfus and Dreyfus, 1988:23-24).

Competence stage deals with complexity and the ability to simplify it, thus, it is possible to identify elements of heuristics developing in this particular stage, which as discussed earlier could be seen as a contributing factor or a prerequisite for intuition. At this point I can further add that heuristics –the probability assessment, estimates the subjective likelihood, and is the mode of reasoning, which does not use analytical methods or deliberate calculations (Kahneman and Tversky, 1984 cited in Cioffi, 1997:206). This leads heuristics further away from the rational system towards the experiential system, which means that during the advanced beginner's stage experiential system is slowly coming into play. In a *competence* stage learner reaches the level where he/she makes decisions and choices after reflecting on various alternatives, rather just following rules, he/she is also able to make plans and have goals.

Moving further in skill acquisition stages and reaching *proficiency,* the learner feels deep involvement in the task and experience it from a certain perspective as a result of previous experiences. He/she is able to associate the present situation with previous situations through the number of salient elements, which will stand out or recede as a result of past experiences. Thus, proficient learner is able to utilize plans that were used then and which worked, in order to achieve wanted outcomes (Dreyfus and Dreyfus, 1988:27-28). However, the process of recognition of elements which are similar to previous situations does not happen by applying rational thinking or rules - it happens automatically. Dreyfus and Dreyfus (2003) give an example of how this happens:

> "A boxer seems to begin an attack, not by combining by rule various facts about his body position and that of his opponent, but when the whole visual scene in front of him and sensations within him trigger behavior which was successful in an earlier similar situation. We call the ability to intuitively respond to patterns without decomposing them into component features "holistic discrimination and association." (Dreyfus and Dreyfus, 1988:28)

Returning back to cognition, we can see that proficient learner exhibits behavior which is almost identical to behavior that is a result of the *experiential system* at work. As noted previously, *experimental system* is driven higher by affect, which means that emotionally significant events (and we presume that the boxer, from the example above, is highly emotionally involved in achieving the victory over his opponent, which also goes along with the fact that in *the proficiency* stage learner is deeply involved in the task) trigger an automatic and instantaneous searches in the memory banks for related events and all the memories and feelings associated with them. Here is where *heuristics* come into play even more prominently than in the previous stage. Just as in *heuristics,* the *experiential system* learns from past experiences and thus it tends to avoid actions or situations which evoke unpleasant or negative feelings from the past, as well as it tends to helps to reproduce positive and pleasant feelings by acting in the same way that contributed to experiencing positive and pleasant feelings in the past (Epstein, 2003:162). This means that automatic and subconscious pattern recognition is taking place, and when specific pattern has been recognized, it then triggers specific actions which worked best in in the past in relation to that specific pattern.

Next and final stage of skill acquisition is *expertise* stage. According to Dreyfus and Dreyfus (1988), expertise can be described as the ability to perform tasks without problem solving or decision making, but by simply doing what works relying on the experience. The skill at this stage is no longer an external element depending on external factors, it becomes internally incorporated in the expert (Dreyfus and Dreyfus, 1988:30-31). Fluid performance, which does not involve any deliberative cognitive processes and is relying on intuition, is another sign of expertise:

> "There is no choosing. It happens unconsciously, automatically, naturally. There can be no thought, because if there is thought, there is a time of thought and that means a flaw…Tennis players "react" when expert, and, a surprising amount of time, so do

> business managers and experienced doctors and nurses when deeply involved in their professional activities." (Dreyfus and Dreyfus, 1988:32)

From the above quote we can see that expertise can be directly linked to *experiential system*, *heuristics* and intuition, while initial stages of skill acquisition are based on *rational system*. This also resonates with Dreyfus and Dreyfus' (2003) statement:

> "Although intuition is the final fruit of skill acquisition, analytical thinking is necessary for beginners to learn a new skill." (Dreyfus and Dreyfus, 1988:xx)

At this point I want to briefly introduce one more theory called Template Theory of Expert Intuition (TempT) described by Gobert and Chassy (2008). TempT states that experts as well as novice have cognitive limits, such as limits of short term visual memory or inability to focus attention on more things simultaneously. It also suggests that on the way to expertise *novices* have to learn a large number of perceptual patterns called *chunks*. Moreover, *chunks* are elements of a more complex data structures called *templates. Templates* possess both *core* (encodes stable information) and *slots* (encodes variable information). It takes about 8 s to store information in *chunk*, which could be considered relatively slow, while it takes about 250 ms to store it in a slot. *Chunks* and *templates* belong to long term memory, and thus "according to TempT expertise is made possible by acquisition of a large number of *chunks* and *templates* that are linked to possible actions" (Gobert and Chassy, 2008:133). Looking from TempT perspective, human cognitive processes sound much like computer processes, however, reflecting on Dreyfus and Dreyfus (1988) description of expertise, *experiential system* and intuition, there is one important element that is making human expertise distinct – emotions. I will return to this point later.

According to Gobert and Chassy (2008), perceptual nature of expertise can be explained by pattern recognition, which can be explained by chunks and templates giving access to relevant links in long-term memory:

> "..pattern similar to one met during previous experience is recognized, and thus an action, possibly a solution to the problem the problem at hand, is automatically elicited." (Gobert and Chassy, 2008:134)

TempT also suggests that "mechanisms enabling the access to long-term memory are unconscious; only the end product of recognition, which is placed in short term memory, is conscious" (Gobert and Chassy, 2008:134). This leads to the fact that pattern recognition in expertise takes place in *experiential system*, and in general expertise can thus be defined by the fact that just like in any skill acquisition, or in any learning activity, with time skill becomes part of the learner, it is no longer separate, defined by the rules reality – expert becomes one with his/her skill and this oneness is expressed through presence of unconscious, effortless, automatic execution of task, which can be explained by *experiential system* being at work. This also goes in line with dual-processing theory claiming that *System 2* thinking requires access to a central *working memory* system of limited capacity, while *System 1* does not. *Working memory* has short-term capacity memory, it also has executive and inhibitory functions and it could be described as

awareness at any given time, through which conscious thinking flows in a sequential manner (Evans, 2008:259). Thus, when talking about expertise, the ability to perform in an effortless and automatic manner could be explained by the fact that while *novice* has to relay on *the rational system* (*System 2*) which is slow, sequential and has low-capacity nature, due to its need to access *working memory*, experts do not need such access, since they got automatic, unconscious access to long term memory through pattern recognition (Evans, 2008:259).

Returning back to *heuristics*, already back in late 50's Herbert A. Simon and Allen Newell (1958) in "Heuristic Problem Solving: The Next Advance in Operation Research" have stated that humans were no longer the only ones who possessed learning, intuition and insight:

> "..we now have the elements of a theory of heuristics (as contrasted with algorithmic) problem solving; and we can use this theory both to understand human heuristic processes and to simulate such processes with digital computers. Intuition, insight, and learning are no longer exclusive possessions of human beings: any large high-speed computer can be programed to exhibit them also."(Simon and Newell, 1958:6)

However, such claims were slow at yielding in any tangible results in AI development, what led many to doubt the success of AI in ever achieving the ability to have intuition and expertise. Among them was Dreyfus (1988), who stated that:

> "Hunches and intuitions, and even systematic illusions, are the very core of expert decision-making, so whether one seeks to use digital computer to model the heuristic rules behind actual problem solving, as Newell and Simon did, or whether one tries, like Stuart and Richard Bellman, to find optimal algorithms, the result fails to capture the insight and ability of the expert decision-making." (Dreyfus and Dreyfus, 1988:10)

After presenting my analysis in the later sections of this thesis, I want to address and, perhaps, to challenge Dreyfus and Dreyfus' (1986) statement about human and machine expertise:

> "..our intuitive expertise, irreducible to rules, casts the weight on the side of the human mind as we try to establish a new balance between ourselves and our ever more powerful, yet perhaps perpetually limited, machines." (Dreyfus and Dreyfus, 1988:xxi)

3. Methodology

3.1 Grounded Theory

I have used Grounded Theory (GT) as a main methodological approach in this research. Despite the fact that I have not used GT's theoretical assumptions as the only epistemological guidelines (as a result of triangulation), and mainly used its methodology for data sampling and analysis, I believe I owe GT an introduction. The main perspective of GT is that reality is constantly changing and is being negotiated. This is exactly the case with AI due to it constantly evolving and due to the changes that take place in society meanwhile, thus active inquiry proposed by GT seems reasonable (Richards & Morse, 2013:191). Even though GT is generally used for generation of theory which is grounded in the data, it can also be used for elaborating or modifying the existing theories that are relevant to the area of investigation through them being played against empirical data (Strauss and Corbin, 1994:273).

GT has an emphasis on detail knowledge, constant comparison, and trajectory of the event (Richards & Morse, 2013:191). GT method also "promotes a stance that refuses to accept a report at face value, a sort of methodological restlessness that leads the researcher to seek characteristics, conditions, causes, antecedents, and consequences of event or responses as ways of drawing them together in an integrated theory"(Richards & Morse, 2013:192). These main elements of GT are in line with my intent to follow the trajectory of AI development through reflecting on predictions associated with it and made decades ago and comparing them with current AI. I will also take a look at expertise development in humans and try to determine through comparison what position current AI has achieved in relation to human expertise and in comparison to its predecessor AIs. This should allow me to make assumptions of what could be the future potentials of AI expertise. Both objectives require constant comparison of collected empirical data, constant inquiry into epistemological resources related to the research area, while following the trajectory of the event in the research, which in my case is the AI evolution.

With this research I don't necessarily have the ambitions of developing an entirely new theory, and I don't necessarily intend to elaborate or test particular theories which I have come across during this research as part of an inquiry into the field of AI. However, playing those theories against my empirical data might give me an insight and necessary ground for creation of potentially new concepts or theoretical ideas, which could further be developed and solidified into theories.

GT methodology fit well with my use of theoretical triangulation, since it does not require any particular commitment to any particular theory:

> "..methodological trust of grounded theory approach to qualitative data is towards development of theory, without any particular commitment to specific kinds of data, lines of research or theoretical interests." (Strauss, 1987:5 in Richards & Morse, 2013:192)

3.2 Practical part

The empirical data of this research was gathered from inquiry into the background of the AI field through analysis of articles and books, conducting semi-structured interviews with actors who are working with AI. I have also performed an observation.

During the research process research diary was kept and field notes and memos were made. In GT another level of data is always produced, usually through memos and other documentation forms of insights gained during the process of analysis (Richards & Morse, 2013:193). These additional data sources have also helped me in writing this thesis.

It is necessary to mention that initially the direction of my research was oriented towards investigation of knowledge production between DTU and IBM through collaboration on Watson, however, since I could not get access to DTU, I reached out to IBM, where I was asked what direction my thesis would have.

My new contact at IBM expressed interest in one of my ideas which was investigating the application of Watson in the medical field, as a result my research direction has shifted. This new direction sounded exciting to me, since I had previous experience doing research about the application of new technology in pharmaceutical field (and medical field felt somewhat related), looking at its implications with regards to ethics, responsibility and privacy. However, knowing the difficulty of getting access in pharmaceutical field due to high secrecy, I was skeptical over my chances, thus, I have decided to keep my options open, and when an unexpected opportunity to observe the course on Watson in DTU (precisely the one, that I was striving to observe initially, but could not get access) presented itself, I gladly took it.

At that point I had no idea that entire Advanced Machine Learning course in DTU had just one lecture about Watson (since I knew from the article on the DTU homepage that the course was supposed to last for 13 weeks). This single lecture was supposed to give students an introduction to Watson, and give them the opportunity to choose Watson as a subject for their final project work, if they wanted to (DTU, 2014). So, I was excited about finally getting a chance to see Watson. Later on, when I was informed that it was just this one lecture dedicated to Watson, I decided that one observation which I have performed was insufficient for me to build my research upon. Moreover, after an interview with Anders, I knew there were more exciting topics I wanted to explore with relation to AI. Nevertheless, observation gave me an opportunity to experience Watson on a theoretical as well as on a practical level, even if only in a very limited way. It also broadened my perspective and lead me to new contacts.

Eventually I realized that getting access to medical area and possibility to research Watson's application there was not going to happen, since my contact informed me that his attempts to reach the ones responsible for Watson in the medical field, asking them for the possibility to be interviewed by me for my thesis, were unsuccessful.

I was left with empirical data from four interviews, which was collected with focus on possible opportunities, risks and dangers that Watson's implementation in the medical field, among various other fields, could carry in it. Thus, the final version of my research

direction and problem formulation was born from data, which perhaps could be seen as bias of this research, nevertheless, it yielded in some interesting findings which will be presented in further sections of this thesis.

3.3 Interviews

I have conducted four semi-structured interviews. Two of the interviews were conducted with people working in IBM, while two others were conducted with those working or studying at DTU.

I have used initially prepared interview guides (refer to Appendix A) with a list of open-ended questions in combination with additional questions which were generated on location following the response from the interviewees, which at times lead to discussion. The interviews were audio recorded and transcribed. Transcriptions were coded, abstracted and categorized. For anonymity purpose each interviewee was assigned a code name.

3.4 Observation

Picture 1 - Advanced Machine Learning course, DTU, lecture about Watson, 05.03.2015

I have performed one observation, during which an audio recording was made, field notes were written and photos were taken. The observation took place during the lecture on Watson that was held as a part of the MSc course in "Advanced machine learning"[1] that is held by Department of Applied Mathematics and Computer Science at Danish Technical University (DTU). The observation gave me an inside on how Watson operates, since I was able to follow the lecture and the explanations that lecturers gave to the student about how Watson functions. However, my ability to follow those explanations was limited, due to me being unfamiliar with the technicalities of the course. I also got the opportunity to

[1] Advanced Machine Learning course, DTU http://www.kurser.dtu.dk/02460.aspx?menulanguage=en-GB [accessed 25.05.2015.]

observe an instance of how IBM's collaboration with the university is taking shape, how knowledge is shared and created. The observation of this lecture also gave me an opportunity for snowballing sampling, since I got the opportunity to establish connection with one of my future interviewees.

Eventually, I decided not to put my efforts into the analysis of empirical data collected during this observation, since it did not fit with my new research direction, and would not contribute to my research in a substantial manner.

4. Analysis

4.1 AI - anticipations

When creating or designing new technology, it is important to consider the fact that this new technology will have the capacity to impact our reality in many different ways. However, it is often very difficult to predict or consider all of the possible scenarios, due to the fact that we simply don't have experience with this new technology and there may be too many unknown factors that may emerge only after technology have matured. Thus, the factor of uncertainty is often unavoidable. Collingridge (1980) in her paper titled 'The Dilemma of Control' states that it is difficult to predict social consequences of fully developed technology during its infancy (Collingridge, 1980:16). However, the inevitable presence of uncertainty should not discourage those responsible for the development of new technology from creating elaborated and detail scenarios, trying to consider all possibilities and all the actors involved in the process. Questions such as how technology would be initially accepted, how its performance after its implementation would influence society, stakeholder and various actors in general, should be asked. Analysis of how the process of stabilization may occur, and speculations with regards to what consequences there may be in the long term using this technology, should be made. No guesses should be overlooked and must be considered in detail, along with elaborated planning of how to address them, since uncertainty entails the possibility of wildest guesses having the potential of coming true. Despite that this may seem as unreasonable, irrational, time consuming and costly process, it may prove highly valuable in case if one of previous speculations might materialize. These guidelines have been inspired by real-time Technology Assessment (TA), which will be further elaborated a bit later in this thesis and which, I believe, should be implemented as necessary measure in the development of all new technologies (Guston and Sarewitz, 2002:98).

As stated above, it is difficult to predict the future of new technology, and often only when the stabilization of it occurs it is possible to evaluate its consequences. In order to look into the process of stabilization, I will refer to SCOT. As Pinch and Bijker (1987) explains, it takes time for a new technology to be stabilized and different closure mechanisms have to be employed for stabilization to occur. For instance, an example of such closure mechanisms could be solving the problems that usually have the potential to spark the controversy around new technology among different social groups, who are using the same technology, but at times in different ways (Pinch and Bijker, 1987:). Thus, it could be said that different social groups or single individuals create different use cases for new

technology, which are not always predetermined or considered by the designers of technology. In order to clarify the notion of various use cases for new technology, I shall turn to the ANT and use Law and Singleton (2005) discussion about the character of complex objects in Science and Technology Studies (STS). I want to use their reference to Star and Griesemer (1989), where the notion of *boundary objects* is explained. Due to the fact that different people may see or experience new technology differently and use it or *enact* it in different ways, whether in a single use cases or in larger social groups with different cultures, we can consider this new technology to be a *boundary object*. Thus, this new technology may be *enacted into being* differently by different social group or individual users (Law and Singleton, 2005:333,334).

The fact that the same technology can be enacted, and thus experienced differently in different *multiple* ways leads to the notion of *multiplicity,* which is discussed by Mol (1999) in her paper where she presents the idea of reality as *multiple*. If applying her rationale to new technology, we can say that the same technology can be differently enacted, resulting in different realities of it, or, in other words, *multiplicity*. These different realities at times may clash, while other times they may collaborate or even depend on one another (Mol, 1999:83). What this entails with regards to evaluating possible consequences of this new technology - AI, is that *multiplicity* adds complexity to the whole process, and this complexity may be overlooked, resulting in a misconception. *Multiplicity* adds the whole new dimension to the pursuit of predicting the future when creating new technology.

It is very important to consider the future of the new technology as early as possible by evaluating all possible issues, dangers and threats, along with benefits and opportunities, what changes new technology will entail and how it will affect society, because in the early stage of its development it is still easy to implement amendments:

> "In the early days of technology's development it is usually very easy to change the technology. Its rate of development and diffusion can be reduced, or stimulated, it can be hedged around with all kinds of controls, and it may be possible to ban the technology all together." (Collingridge, 1980:16)

There may be several methods that could be used in order to cope with uncertainty which is present in new technology. One of them is the scenario-based design. As Carroll (2000) put it in his article titled "Five reasons for scenario-based design", those who are involved in design process of the new technology often have to face "convoluted networks of tradeoff and interdependency, the potential of untoward impacts on people and their social institutions, and the likelihood that changing cultural and technological circumstances will obviate any solution before it can be deployed" (Carroll, 2000:43). This is where scenarios can come in handy. Brief description of scenarios is that scenarios are stories, which have a certain setting, actors with their goals or objectives and a plot which consists of sequences of events or actions (Carroll, 2000:45). How exactly scenarios could be beneficial in AI development? Some of the potential benefits are:

> "..scenarios evoke refection in the content of design work, helping developers coordinate design action and reflection. Scenarios are at once concrete and flexible, helping developers manage the fluidity of design situations. Scenarios afford multiple views of an interaction, diverse kinds and amounts of detailing, helping developers manage the many consequences entailed by any given design move. Scenarios can also be abstracted and categorized, helping designers to recognize, capture and reuse generalizations and to address the challenge that technical knowledge often lags the needs of technical design. Finally, scenarios promote work-oriented communication among stakeholders, helping to make design activities more accessible to the great variety of expertise that can contribute to the design, and addressing the challenge that external constraints designers and clients face often distract attention from the needs and concerns of the people who will use the technology." (Carroll, 2000:43)

I believe that scenario use could be extended to include the reflections on such important topics as ethics, responsibility, security, safety and privacy. Thus, those who are developing the AI technology should not overlook this tool.

AI has not yet reached its full capacity, thus it is still possible, if necessary, to make amendments in its design as well as create the necessary framework for its management and regulation, as yet another method to cope with uncertainty. But for that it would be beneficial to have an overview of predictions made with regards to AI and its possible effects on society. For that purpose I am going to refer to two authors. One of them is Kim Eric Drexler, an engineer who holds a PhD in molecular nanotechnology and who has written a book called "Engines of Creation: The Coming Era of Nanotechnology" (1986), which, according Google Scholar has been cited whole 2137 times[2]. In this book he explores different possibilities how nanotechnology and artificial intelligence may influence humanity. Naturally, I concentrated on his ideas regarding AI and its possible future consequences.

A second author that I will refer to is Vernor Vinge, who is a retired professor in mathematics and computer science as well as a science fiction author. In 1993 he wrote a paper called "The coming technological singularity: how to survive in post-human era", which according to Google Scholar has been cited 421 times[3]. In this publication, which was originally presented at the VISION-21 Symposium, sponsored by NASA Lewis Research Center and the Ohio Aerospace Institute, 30th -31st of March, 1993, Vinge discusses his idea about AI and how it may affect humanity. So, both authors are concerned with how AI will affect humanity and try to make a series of predictions of what may happen as a

[2] Number of publications where "Engines of Creation" book was cited along with the list of publications where it was cited https://scholar.google.dk/scholar?cites=6079284407963947148&as_sdt=2005&sciodt=0,5&hl=en accessed on 22.05.2015.

[3] Number of publications where "The coming technological singularity: how to survive in post-human era" book was cited along with the list of publications where it was cited https://scholar.google.dk/scholar?cites=3688559863916355898&as_sdt=2005&sciodt=0,5&hl=en accessed on 22.05.2015.

result of AI evolution. I have extracted some of the major points from both publications and placed them in a Table 2.

Table 2 – Opportunities, Threats and Solutions – predictions made by Drexler (1986) and Vinge (1993)

	Opportunities **O**	Threats **T**	Solutions **S**
1	Fast AI systems due to AI parts working faster than human brain parts (neurons)	Progress towards AI is unstoppable, even if the dangers are real	Embed rules or physical confinement for AI
2	Fast engineering due to fast, automated engineering systems	AI competitive power so great that prohibition to create it will only make sure others will create it or it will be created in secret	Asimov's law - build rules into the mind of the created AI
3	Quick simulations resulting in rapid progress; quick experimentations; quick design	AI replaces higher level jobs (AI replacing experts), contributing to unemployment	Meta-Golden Rule embedded in AI - "Treat your inferiors as you would be treated by your superiors."
4	No need for human labour in engineering and production	Social AI could pose the same threats as terrorists or demagogues while technical AI could destabilize the world's military balance	Foresight and careful strategy for safe use; take dangers seriously
5	An ultimate tool to create all other tools	The breakthrough will be so abrupt and destabilizing, country lagging behind could face fatalities	Developing institutions for AI management and dealing with questions of threats and opportunities it poses
6	Promises advantages in commercial and military areas	AI gives ultimate power and assures the global rule	If necessary, suppression of AI technology (Local/unilateral, global by agreement, or global by force)
7	AI increases competitive power in all the areas	Human less intelligent than superintelligent AI	Cooperative development
8	Great upward leap in technology advancement	Superintelligent AI will not be humanities tool	Unilateral advance
9	AI will have technical and social abilities; understanding human speech and wishes	AI treating us as we treat animals	Make AI law-abiding citizens
10	AI creating more intelligent AI	Out of control evolution of AI	
11	AI controlling AI	AI contributing to the extinction of humanity	
12	Power to extend life and to liberate it		
13	Help build a new and better world		

Table 2 consists of three columns. The first column is called “Opportunities” (O) and is dedicated to the possible opportunities that AI may create. The second column is called “Threats” (T) and includes possible threats that AI evolution may pose. The third column is called “Solutions” (S) and include suggestions for possible solutions how to avoid threats

that are listed in the second column. Each column has a letter assigned to it which comes from the first letter of the name of the column, thus three columns are O, T and S, correspondingly. Each row of the table has a number assigned to it for quick reference in the further sections. Thus, for instance, T11 would be a key to the statement "AI contributing to extinction of humanity". The rest of the statements in the table can be referred to in a similar manner.

The table was generated by making a thorough analysis of Drexler's (1986) and Vinge's (1993) texts. After reading the texts, the number of themes were selected: Risks/dangers/issues, Benefits, Suggestions for solutions, Issues with particular suggestions, AI misconceptions, Factors influencing AI, Results of AI. The quotes from both texts that fit under these themes were selected and using excel spread sheet distributed under these themes. The main points from one or several quotes under each theme were extracted and written down in the form of short statements. These short statements then were distributed under the same themes again. Finally, some of the themes were joined together, while others were left out in order to make an easily readable, but yet informative table.

The purpose of the table is not only to give a quick overview of predictions made by these two authors, but also to give the possibility to compare them with the actual current level of development in AI technology through the case of IBM's Watson. In the next section I will present the analysis of the interviews, from which the final table representing current AI will be built and presented in the 4.3 "AI - current reality compared to anticipations" section.

4.2 Analysis of the interviews

In this section I will present the analysis of four interviews (for transcriptions refer to Appendix B). All the interviewees got their names anonymized in order to protect their privacy. Data from each interview is divided in categories, which I have found to be important in relation to Table 2.

4.2.1 Interview with X1

Following is the analysis of the interview with the X1, which took place in IBM Client Center Copenhagen, located in Holte.

X1 is an expert in Watson in Denmark:

> *X1: "..I'm innovation executive, chief technologist at IBM and also Watson, the Watson, you can say, the expert and advocate in IBM Denmark for the time being, and my background is - I am a master in finance accounting and taxation. I've been a management consultant for about nine years and then I've been at IBM for nine years, no, it's more, it's from 1997.." (X1, 2015)*

- <u>Data</u>

Watson is able to process and analyze big amounts of both structured and unstructured data:

X1: "..core analytics, so being able to provide analytic capabilities, and visualization, could be unstructured, could be structured and then combine that with this cognitive intelligence, so you actually use the intelligence of Watson to make you aware when you doing some kind of analytics, you are sitting with graphs and you have tables of databases of data that you looking at. What is interesting in these figures, in these amounts of also structured data, that is what we are doing with IBM, with Watson analytics, it's actually one of the new things around this, this is, it's thinking for you, so, it's your data scientist, you can say." (X1, 2015)

X1 says that Watson has capability to make use of all the data, or big data, which is the product of digitalisation as not being used optimally. Watson can actually use it as an experience:

X1: "..healthcare we are quite far in Denmark with the digitalization, so we have a lot of data here, we have some very big historians [phonetic], so, databases where we have collected information about people, so, you know, sickness, and the treatments they have had, and the outcomes that have been. So, we have a lot of material to actually use as experience, which is basically not used today. It's collected and some analysis, and some reports are being made, but it's not used on an everyday basis, we don't have tools for that today in Denmark. So, I think that's, basically, that's the one of them. And public sector, I think with all the regulation and procedures, and we have in the public administration…there is an information overload in that area.." (X1, 2015)

However, according to X1, there might be different concerns such as concerns about privacy, data security, and responsibility with regards to Watson having access to the data:

X1: "..there are a number of dilemmas, that, you know, you have to, you know, or some resistances that you may have to overcome, you know, by…, because this is as with other technology that you infuse into some business processes, or that you give to humans, then is going to change something for them, so there could be some resistance to this: can you rely on this, and, you know, who is taking responsibility, and this kind of, of things. There could be concerns about privacy, security in this, so, you know, if you use this for medical treatment, medical advice, then Watson would like to know your personal history, and, you know, whatever has happened to you, and the measurements, etc. over time, so what about these data, where are they, and can you make sure that they, these data are safe." (X1, 2015)

- <u>Capability to understand natural language</u>

Watson is capable to understand natural language:

X1: "..it's a cognitive technology that is capable of analyzing very big amounts of data, unstructured data, that understands natural language.." (X1, 2015)

X1: "..you have the whole human-computer interface, so this first thing is that you have the near language processing which I think we have come quite far." (X1, 2015)

However, according to X1, currently Watson is not ready to communicate through spoken language, and interaction is possible through written text:

> *X1: "We haven't implemented speech to text, so, you know, so, when we are asking or interacting with Watson today, it's through use of typing. So, we type in questions, we are not speaking to it. But of course we are working with the, with the speech also and, and, you know, we can demonstrate it today, but we just wanted to work really good before we, we release it.." (X1, 2015)*

IBM is trying to handle this limitation by having third party software for voice recognition, but the issue is that currently there is no software available that would be working well, especially when Danish language is used:

> *X1: "We are working on it, and, you know, we also moving into having a third party software, but I don't think I have met any device today where you say, you know, if you try to speak to Siri yourself, or if you try to, especially if you try to speak it to it in Danish it won't work." (X1, 2015)*

The precision level of current voice recognition software does not fit the purpose for which Watson is made:

> *X1: "There is too many errors, it, you just find it irritating. We have for a long time in Denmark, actually that is, that is being used, but it also requires some training, sold system where to doctors, and also to case workers in public, where they actually speak to the computer, and computer will translate that into text, and then they will have to do some editing afterwards, and there we have a precision around, I think around eighty, eighty five percent is actually max…it's ok to do it there, because you don't have to have this to be precise immediately. " (X1, 2015)*

- <u>Capability to learn</u>

Watson is a cognitive system that learns, so when it is implemented in a company or organization, it has to be taught and trained in order to learn:

> *X1: "..it's a system that learns and improves all the time by itself and by the knowledge that you make available to it, so that's, I think, that's the core of Watson." (X1, 2015)*

However, it takes time to bring Watson to the optimal operational level – Watson need to be taught and building knowledge takes time:

> *X1: "..have to start building the knowledge around specific domains. So if you want to use this within health care, then you would have to understand the human biology.." (X1, 2015)*

- <u>Capability to reason</u>

Watson is capable to reason the way humans do, and according to X1 the ambition of IBM is to make Watson work on a human level:

X1: "..it's an artificial intelligence, that you can say thinks like we, as humans think or reason, I would say, not think, but reason." (X1, 2015)

X1: "..we want to have, you know, promise of Watson is actually to work on human terms.." (X1, 2015)

- Capability to recognize patterns

Watson is capable of recognizing patterns within data, which gives possibility for it to be used for discovering things:

X1: "And then is discovery, so you can add some analytics capabilities to the, to this capability of handling of this big unstructured data to actually find new patterns, relationships that you couldn't, could not see before, and that's for instance what we are doing in life sciences for drug discovery." (X1, 2015)

- Validity, precision and being up-to-date

When talking about the medical field, Watson is able to perform a quick search using validated data sources, resulting in a quick suggestion for a better informed diagnoses, benefiting both doctor and patient, since normally doctors might have difficulty to staying up-to-date with regards to newest discoveries in the medical field which might influence their performance level (same goes for other fields where Watson might be used):

X1: "..when we look at the precision of the Watson technology in, you know, in making, or in, yeah, in making diagnosis for certain sicknesses or illnesses, and Watson is more precise today than the best specialists.." (X1, 2015)

X1: "..Watson is relying on validated knowledge recourses, and some of these knowledge recourses may be on the internet. So, there are certain, you know, patent databases, which we access via the internet, but then we have situations where it may be a public database, but the use case requires that you have the response within, you know, a short period of time, so for instance if it is a doctor who has to go to research some using Watson as advice, and, is and the Watson adviser here is relying on some scientific literature, then maybe the doctor has got seven-eight seconds, you know, siting with the patient inform of him, or her, or, and needs to get this suggestion back from Watson that quickly and there are, we have then, you can say, uploaded databases in our servers, and, you know, indexed and made them ready for very quick researches...core strength would be that you can base your treatment of the patient on the latest research and the latest knowledge." (X1, 2015)

- Hypothesis generation along with a confidence level

Watson is capable to express a confidence level which it has with regards to the hypothesis that it generates about specific question. This might be helpful when human expert will have to evaluate the advice given by Watson, as well as choose from several advices, if that's the case, based on the confidence level each of the advices is assigned by Watson:

X1: "..that is able to answer questions by generating a lot of hypothesis and evaluating these, evaluating the resources and that can give you an answer and confidence level on how much it believes it's the right answer...it's actually making this, you know, using this capability of questioning and answering, perhaps also foreseeing what could a question be that you have not discovered yourself, and then providing answers for that.." (X1, 2015)

- Capability to understand context:

Watson is capable to understand context, which is an important factor in health care, as well as in other fields, and that is one more factor which sets this technology apart from all the other existing technologies:

X1: "Watson already today understands, you know, the international terminology...we want to have, you know, promise of Watson is actually to work on human terms, so we would have to teach Watson all these deviations. So Watson, and that's why you cannot just, you know, make a machine translation, so. There will be different ways of telling you, for instance, that a patient is not a smoker. So, you know, and there is one of the classical examples that, you know, if doctor is writing in the journal that that the patient hasn't smokes for thirty days, then ok, he is a non-smoker, but in terms of the patient history he is a smoker still, because there are some effects on, from that.." (X1, 2015)

Loreta: "So it has to understand that within the context." (Loreta, 2015)

X1: "Yes, exactly, yeah" (X1, 2015)

- Capability to interpret images and video:

X1 says that IBM is working on Watson's capability to interpret data from sensors as well as images and video. Currently, Watson is capable to interpret MRA scans. IBM is continuingly working on Watson's capability to have situation awareness, which means it could understand the context of videos and images:

X1: "..could be inputs from video, we also working with that right now, so you can actually, so you can, Watson would be able to, to interpret video feed…within the cancer area MRA scanning, we just released that. So, if there are images in a patient journal, or attached to journal, we can actually scan those images and come up with the recommendation on what's on these pictures, and, I think, right now they are being, it's being tested by radiologists as a kind of second opinion on, you know, analyzing these pictures. But it's something that we are continuing to work on, so that you can actually interpret not only these images, but also videos, to say what's on the video, what is the situation, so you have the situation awareness and have the context...like you and I sitting here, that Watson could understand this is a meeting, you know, I would have to take some kind of memo of meeting, this is not a party or something else, that Watson can actually understand the situation and then do some actions according to that awareness of the situation." (X1, 2015)

- Languages

Watson originally was developed in English language, and recently Spanish, Portuguese and Japanese languages were added:

> *X1: "..for the last three years Watson has been only been able to understand and process English, now it can also, it's just been released on Spanish and Brazilian Portuguese and Japanese.." (X1, 2015)*

Factors like market size and presence of customers who are actually interested in Watson has determined why precisely these additional languages have been chosen among all the other languages:

> *X1: "..it's something to do with the, it is with the market size, it is one factor, another factor, there is a customer. So, for instance in Spain, and there is no secret, there is a big bank, who said: ok, you know, IBM make this work in Spanish and we will sing up for a project in the size for around 50-80 million dollars to get Watson into the front of our interaction with our customers, and then we do it." (X1, 2015)*

- Flexibility - fluidity

Watson is flexible technology and IBM takes time and puts effort to develop specific variations of Watson according to customers' needs and requirements:

> *X1: "..*we work with companies and the organizations to develop specific kinds of use cases and different kind of flavors of using these capabilities, these basic capabilities, yeah." *(X1, 2015)*

Watson's flexibility is also expressed by the fact that it can be used by experts and non-experts, as well as by ordinary people, as a part of some particular service:

> *X1: "Adviser role for other experts, human experts, or for non-experts in the organization who wants access to this expertise, or for, you know, consumers, like you and I." (X1, 2015)*

There is a benefit for those who want to use Watson – ready tools available on the cloud and clear time schedule for how long it will that take and what is necessary to start using Watson:

> *X1: "..we can make, we can have them make a quick assessment which will take them in calendar time between six or eight weeks. …Watson has to learn about this business process and this knowledge would take six to eight months, and then they would have a pilot that can be scaled up. …We do have the tools ready, they are in the cloud, they can get access.." (X1, 2015)*

Easy access to Watson and easy operation of it through the cloud, whether by the client themselves or by IBM specialists adds to flexibility of Watson:

> *X1: "They can get access to it, you know, or we can do it for them, but it's in the cloud, so it's just the matter of, you know, logging in, having your credentials approved and then logging in and then you can start." (X1, 2015)*

- Application opportunities

The best opportunities for Watson are in big companies and organizations which have a lot of expertise and knowledge:

> *X1: "There is big potential, you know, the more knowledge, the more expertise that is required in business or organization, the bigger is potential for Watson.." (X1, 2015)*

However, smaller business would not be really benefiting from Watson, unless they have a lot of knowledge accumulated in them:

> *X1: "..so, you don't find potential in small manufacturing companies, unless there is a lot of knowledge you have to have around the product then you can start using Watson." (X1, 2015)*

Watson intelligence can be applied in robotics and used for robots, so that if a robot would be the body, Watson could be its brain, this way adding a new dimension to the robot-human interaction:

> *X1: "..applying Watson to robots, so a robot, the intelligence in the robot would, to a certain extent, come from a Watson, and that's because we see that the human-computer interface, having a robot that can actually express perhaps some feelings or some awareness and, who can react to what you are, what you are doing, is giving another interaction, a new interaction between the human and the computer." (X1, 2015)*

Watson has good prospects in Denmark - IBM already has clients who are interested in Watson. IBM's current clients in Denmark are from banking and pharmaceutical sectors, as well as life science and software companies that want to add intelligence to their applications:

> *X1: "..we do have within banking, we do have within research pharmaceuticals, we also have customer who is actually building a prototype right now in the life sciences space, we do have software companies that are developing applications where they can see that, you know, where they want to sign up for actually using Watson to add intelligence to some of the functionality they have already in the application." (X1, 2015)*

The fact that Danish authorities show big interest in Watson, present big potential for this technology to be implemented in Denmark:

> *X1: "They are very interested, we have had minister in New York for, for Watson briefing, what is it, four weeks ago we had Bjarne Corydon and I think it was, you know, finance minister and we have Morten, what is it [inaudible], Østergaard [checked online], our economy minister, over there for a session on Watson, so they could see the potential…we want to be on the leading action in Denmark, we do want to say that we are a knowledge society, and we do want to say we have the*

best patient care and all these things. And then maybe they can see this technology as a very important, enabling." (X1, 2015)

However, the fact that Watson is the cloud service might present some security concerns for some of the companies:

Loreta: "Do you think many companies would be worried about security, data security?" (Loreta, 2015)

X1: "Oh yeah! They already are, so, basically, you know, as a starting point Watson is in the cloud… So, it's a cloud service, but we can actually, if it's required, will be, it's expensive, but companies are doing that – install it locally." (X1, 2015)

Some of these companies might be the ones, where data security is of great importance, for instance, pharmaceutical companies, who are in fierce competition among each other and thus are very apprehensive about data security and thus may be reluctant with regards to Watson:

X1: "..for instance the pharmaceutical companies are very cautious about this, and really have strict requirements. It could be for data privacy purposes, because they do calculations and it's on patient data that they have been granted access to from public authorities, or it could be that they are looking into some components that they don't want their competitors to know about, there could be really something about.." (X1, 2015)

When talking about Watson application in non-English speaking countries, the fact that the core knowledge in those countries' industries is in their own languages, which is also the case in Danish industries, the current inability of Watson to operate in other languages (besides Spanish, Portuguese and Japanese) could be a limitation for Watson:

X1: "..the limitation we have today for parts of the Watson technology is that we cannot understand Danish language today, so Watson has to, because it's so core in understanding the corpus of knowledge that maybe in the country that Watson has to understand that knowledge or the language of that knowledge, and basically if you look into the, in healthcare system today it's in Danish, and the same thing goes for, well, in banking or in most industries you will find that the core knowledge will be in Danish, that's the limitation we have today." (X1, 2015)

- Threats

Jobs?

According to X1, Watson might be perceived as a threat by unions with regards to fears of Watson taking over jobs which are currently performed by human employees. X1 draws parallel between digitalization and Watson, in the sense that both are affecting the status quo and both have destabilizing effects. Digitalization made some well stabilized jobs disappear, and Watson has the same potential. However, Watson threatens not only routine jobs, but also some expert level jobs. X1 mentions that there is a level of skepticism with regards to Watson entering areas of work, where usually a lot of experience is needed in order to perform work tasks. High level of experience leads to

high level of proficiency or expertise of particular work, which can signify jobs that have a more advance level of expertise and not only routine jobs, which digitalization has eliminated when it began:

> *X1: "I don't know, unions, perhaps would see this as a threat? …when we had digitalization of the paper printing, you have these clashes here in Denmark and typograf [phonetic], I don't know what's the English word, you know, the people who was normally kind off designing the papers, and the printing, you know, that kind of job has totally vanished now. And there you had a lot of clashes between unions and the owners or the publishers, because, you know, it is an area, Watson is going into an area, where, you know, people generally think this is not something that can be taken all by machines, because this is, you know, complex, this is where we are building experience over a long period of time to be very good and recognized about this. So, and this is where, you know, we have to train the next young person coming to the organization over long time before he or she can do this himself, and I think that picture is going to be changed and that is something people have to realize, and that's, I think that will take time. So, and then there will be a resistance to that, but on the other hand, it's a progress, so you can't stop that. " (X1, 2015)*

By saying that there might be "*some resistances that you may have to overcome*", X1 refers to the process which usually takes place around all new technologies and is known in SCOT as stabilization process:

> *Loreta: "So you think like any new technology, like, Watson just have to go through this period?" (Loreta, 2015)*
>
> *X1: "Yeah, so there always will be anxiety and concerns when new technologies come again, but I think when you look at the good sides, and what this will mean, and I think it's worth it, yeah." (X1, 2015)*

X1 draws a parallel with stabilization process that took place with cars and adds that education of the public is necessary in order to make them see the potentials that new technology brings:

> *X1: "I think at some point of time it's just like with the cars, you know, we have so many people killed on the road. When we have a car that is really self-driving and really intelligent, we would start, you know, looking at it as being more secure than actually having people driving the car. That's the same, I think the same thing will happen with this kind of technology, so...I would say it's a concern, and it's concern around new technologies that has always been there...like, you know, any technology, new technology, there is kind of education, you can say, of the public and awareness to make them aware of the capabilities, but also, of course also to remove any anxiety that might be around this." (X1, 2015)*

X1 admits that the concerns mentioned in the interview might slow down the process of establishing Watson in the market:

X1: "that's the kind of concerns that will be and some of the anxieties that can, could kind of slow the penetration down." (X1, 2015)

- Expertise

X1 presents Watson as an adviser to human experts and non-experts. With its capability to fast analyze big amounts of structured and unstructured data, find patterns, make links, come up with a hypothesis and make discoveries, Watson sound like an excellent tool for experts, providing them with up-to-date and quick support in their daily work:

X1: "..it's the adviser role. Adviser role for other experts, human experts, or for non-experts in the organization who wants access to this expertise, or for, you know, consumers, like you and I. And then is discovery, so you can, you can add some analytics capabilities to the, to this capability of handling of this big unstructured data to actually find new patterns, relationships that you couldn't, could not see before, and that's for instance what we are doing in life sciences for drug discovery…you actually use the intelligence of Watson to, to make you aware when you doing some kind of analytics, you are sitting with graphs and you have tables of databases of data that you looking at. What is interesting in these figures, in these amounts of also structured data, that is what we are doing with IBM, with Watson analytics, it's actually one of the new things around this is, it's thinking for you, so, it's your data scientist you can say." (X1, 2015)

Above X1 says that Watson in a way might be doing the thinking for the expert. In many cases precious time is at stake: time may cost money, it may also cost lives (considering time used by medical experts in coming up with diagnoses in life threatening situations). Watson could radically cut the time spent in cases, when experts have to come up with important decisions fast.

Watson can contribute to rise in the confidence level of experts when they are making decisions, since they would know that their decision was supported by an intelligent adviser, who have just accessed all possible databases and analyzed all available data using its expertise which has been built over time and to some extent may even surpass human expertise. Thus, an expert might be more confident about the choice he/she has made or decision he/she has taken in an unknown situation, which he/she has not experienced previously, and thus does not possess the adequate expertise level to solve. X1 refers to Watson's role, saying that *"it's a decision or support for a specialist of some kind" (X1, 2015).*

According to X1, doctors often do not possess the optimal level of expertise due to inability to stay updated with the latest scientific knowledge and instead of relying on facts they are relying on their intuition:

X1: "..normally, you know, doctors, if a doctor or even if a specialist has to just be reasonably up to date on the knowledge on which they are basing their, you know, their diagnoses and treatments of patients, then they have to study around hundred and sixty hours, read papers, a week, and how many hours do you have a

week?...you cannot do it, you cannot do it. And reality shows, you know, there has been surveys around this, you know, they perhaps spend two or three hours a week on that. So, the doctor you are visiting is outdated, you can, you know, you can count on that, and that they are relying on their gut feeling." (X1, 2015)

However, from description of expertise by Dreyfus and Dreyfus (1988) we know that intuition, or "gut feeling", how it is often called, is an important element of human expertise. However, in the era of cognitive machines this key-element of human expertise might be seriously challenged.

It is questionable whether there is a chance that experts could develop dependency on external helper such as Watson, and if that could potentially lead to them stopping to rely on their own intuition in decision making. If so, what would this do to the expertise in general?

X1 brings forward an interesting question - whether the emergence of AI could influence our cognitive processes in a manner that we would become dependent on AI doing all the thinking for us, and if as a result our cognitive processes would start to deteriorate:

X1: "..what it is this going to do to humans, if humans are not required, need to challenge themselves on knowledge and seek knowledge them self, and think them self, because, certainly you have the machine here, doing and thinking, and you'll just pull on your autopilot in your head and say: ok, this is what we are doing. Then would we as humans deteriorate? Our ability to think, and to resonate our self, could be, you know, concern, I am not concerned about that personally." (X1, 2015)

Even if an AI would be taking a role of an adviser in decision making or would be the source of second opinion, it is still uncertain how would it influence expert in his/her decision making. For instance, in the situation where expert's intuition would contradict with Watson's suggestion, which would be rated by Watson with highest confidence level (meaning that Watson "believes" it is the most proper suggestion out of all the other suggestions), what would be the choice of expert? Would he/she follow his/her intuition or would he/she follow the suggestion of Watson? And what would be the most responsible choice, considering the fact that Watson will be always up-to-date with scientific knowledge, but also considering the fact that just as humans Watson can make mistakes. Would situation like this present the ground for conflicts of conducts, ethics, responsibility or morals? It certainly should be looked into closer, however, not in this research.

How long would it take before human experts would be outdated and outsmarted by AI? This may seem as a radical question to ask, however, it is worth considering, since human experts might have difficulties to keep up with the speed and quality of the way expertise is being acquired by artificial intelligence such as Watson and what it has to offer:

X1: "But already now, when we are, you know, when we look at the precision of the Watson technology in, you know, in making, or in, yeah, in making diagnosis for

certain sicknesses or illnesses, and Watson is more precise today than the best specialists." (X1, 2015)

When talking about expertise in the medical field, it is clear that experts have direct impact on human lives. In order to consider possibility that one day Watson could become medical technology and have the right to directly impact human lives through its expertise, without human expert, would require an entire procedure of Watson's validation by various instances, one of which would be Food and Drug Administration (FDA):

X1: "as soon as you start having something like robot, or some interaction which, actually, you know, impacts people's health or security without intervention or having an intermediary human being, an expert, then you are into the FDA field and you have to be very dated and that's not our core competence." (X1, 2015)

Despite the fact that currently Watson is being marketed as an adviser to human experts, X1 does not exclude possibility that in the future Watson might be used directly as medical technology:

X1: "Well, no, I wouldn't rule that out, but then we have to be so sure by ourselves and we would have to have probably somebody to take the responsibility.." (X1, 2015)

If Watson would become medical technology, responsibility would be the key factor, since Watson's direct interaction with patients would involve decision making which could be lifesaving or deadly, just as in the case with human experts. Human experts make errors sometimes, and they must take responsibility for it, however, if Watson would make a fatal error, who would take responsibility? If technology is gaining the same level of recognition in expertise as human experts, and if technology is able to reason on the same level as humans, a thorough analysis of how a non-human expert may contribute to safety or risk, and what does this entails to the notion of error in expertise, is necessary, however, this is not in scope of this thesis.

The fact that Watson is targeted at companies and organizations which possess higher level of expertise and have more knowledge, could indicate that Watson might be capable to embody the expertise, which is currently expressed in the form of multiple human experts:

X1: "..there is big potential, you know, the more knowledge, the more expertise that is required in business or organization, the bigger is potential for Watson, so you don't find potential in small manufacturing companies, unless there is a lot of knowledge you have to have around the product then you can start using Watson." (X1, 2015)

Currently, Watson is advocated as an adviser, however, having a smart non-human adviser that is capable of containing vast amounts of knowledge and operating it in a very fast and efficient manner might influence the way companies and organizations would distribute their recourses for employment of human experts. When talking about experts,

X1 seems not to assign knowledge possession in jobs to the sign of expertise, and thus it is not entirely clear if his statement is misinterpreted by me or he has a different view on what comprises expertise. However, it is worth looking into with regards to concerns over the possibility of Watson replacing some knowledge-based jobs:

> *X1: "..society like Denmark, I think vast number of the jobs that we have here are knowledge based, and, and the vast number of those knowledge based jobs are, you know, not really expertise jobs, it's jobs where you could say you could put that knowledge into technology like Watson and you could eliminate a lot of these routine knowledge jobs that you have today and use technology, so, something around employment." (X1, 2015)*

In contrast to previous statement about Watson being oriented towards routine knowledge jobs, when asked about what practical steps it would take for a company to start using Watson, X1 gives a response, which could point towards the fact that after all Watson may be oriented towards expert job positions, since complex decisions and knowledge are present in expertise jobs:

> *X1: "We call it a cognitive value assessment. So, companies who saying, you know, this is really interesting for us, because we have knowledge, we have a lot of expertise and complex decisions, etc., and there, you know, to find out, we choose cases, you can say the most profitable for them to go into and most, and also where do they have the, the prerequisites for this, do they have the knowledge within reach and in machine reach, all these practical things, so that we can make, we can have them make a quick assessment.." (X1, 2015)*

Watson may change the way expertise is going to be developed within the companies. Currently it takes a reasonably long time for a novice to acquire the level of expertise required for specific work occupation, and it is done through training and other ways, which is a costly and time consuming process for the company. According to X1, this is the area where Watson is going to enter and bring changes:

> *X1: "Watson is going into an area, where, you know, people generally think this is not something that can be taken all by machines, because this is, you know, complex, this is where we are building experience over long period of time to be very good and recognized about this. So, and this is where, you know, we have to train the next young person coming to the organization over long time before he or she can do this himself, and, I think, that picture is going to be changed, and that is something people have to realize, and that's, I think, that will take time. So, and then there will be a resistance to that, but on the other hand, it's a progress, so you can't stop that." (X1, 2015)*

It might require much less time and less resources to train Watson and Watsons expertise can be scaled out, so it could substitute several human experts, which might be considered to be beneficial by the companies or organizations, who will go along with the "progress", as the X1 refers to it.

4.2.2 Interview with X2

Interview with X2 took place over the Skype, with his location being in the United Kingdom and my location being in Copenhagen.

In the following X2 presents his expertise area:

> *X2: "I did BSc in physics and MSc in biotechnology, a PhD in molecular diagnostics… I've worked for a couple of years at a consulting organization for pharmaceutical industry, then I worked for about twelve-thirteen years at 'GlaxoSmithKline'…on the sales and marketing side of that, of GSK…I've worked for two years at a company called 'IMS Health' which is around data for the pharmaceutical industry and then I've been in IBM for around four and a halve years, working on the working on various different sites of the organization, but most, most recently I've been spending a lot of time talking to pharmaceutical companies about Watson." (X2, 2015)*

- Data

When asked about Watson in relation to big data, X2 said that Watson is able to handle the data and make connections fast, which is an important factor when talking about big data:

> *X2: "So, Watson's analytics, its ability to make connections and the speed at which it works, I think, are very important. Now, you have all this information streaming in, what does it actually mean? You need a cognitive system that can understand whether linkages are happening very, very fast, and present the findings as answers, as insights, as decisions that need to be made." (X2, 2015)*

- Capability to learn

X2 explains that Watson is a learning cognitive system, which is capable to make connections in unrelated data and learn from its mistakes:

> *X2: "..we classify Watson as a cognitive system…when IBM uses this phrase, means to learn. Ok, so it, it's, what is what, there are two aspects: cognitive means that it can bring, it can connect fact's information data that was previously unconnected and identify that this is something you may be interested in, right? So that's one thing that's cognitive, the other bit about cognitive is that it is iterative… if it knows why it got something wrong, then the chances of getting the same thing wrong again are very, are reduced, right? And so it becomes learning, it can, and it improves." (X2, 2015)*

However, the fact that Watson learns means that Watson also needs to be taught and X2 consider it to be a possible weakness, since it takes time for it to be taught and trained before it can be operational at necessary capacity:

> *X2: "..when you look at any kind of software system, any software program that's available today, it operates at its best possible level on day one, and at exactly the same level on day one thousand, right? Watson is at its poorest level on day one and*

will be incredibly good on day thousand, because it keeps learning, it gets feedback, so you have to think of Watson as being an additional resource in your team, not a computer. It needs training, it needs feedback, it needs to learn and it gets better and better, and better, ok? So, that's how it works. So, in terms of weaknesses Watson needs to be fed information, right? So the weakness here is that you need to spend time teaching Watson what it needs to do, ok? But other than that, I think it's difficult to identify the weaknesses at this stage, simply because it hasn't really been maximized, we don't know what it would do when it is at full scale." (X2, 2015)

This means that companies or organizations that will want to use Watson will have to invest their own time and put efforts in teaching and training Watson before they can receive benefits from having it. However, considering that X2 says that Watson should not be seen as a computer, but as an additional recourse in the team, which brings him closer to the level of human team members, who, before they could achieve a certain level of expertise, also had to be taught and trained, the fact that it is necessary to put additional effort in Watson does not seem as problematic. Besides, if Watson will be additional recourse in the team and considering that it would come to company without having any prior experience in the field or business area in which they operate, unlike his human counterparts, who all would have one or another form or education, often supplemented by practical experience, the teaching and training could be viewed as an investment rather than weakness.

- Capability to understand natural language

The capability to understand natural language, plus the capability to understand the context within written text enables Watson to make links and make a hypothesis based on examination of related data:

X2: "..so when you think about medical records and you think about doctors writing down what they've learned when they've taken a patients history, they might be saying 'The child's nose is running', 'he was feeling very hot', and there were various differe-, and 'there was a rash', right, now Watson can understand everything that it just written, so that if you then wanted to find, now suppose Watson does that for five thousand medical records, you can then ask Watson 'well, was there a rash every time the nose was running?', and then it can come back and answer that question. It can come back and say 'the correlation between the rash and smelly feet is x percent', right? So it understands what it is you are asking, it understands what is written down, and then it can come back and actually make the links." (X2, 2015)

When asked about Watson's inability to understand spoken language, X2 does not think that to be an issue. According to him, any process of digitizing speech could be used for speech recognition in order to have it ready for Watsons use:

X2: "Oh, but, I mean speech recognition is something that a lot of companies have got very good systems, right? So, the way IBM approaches this is we got speech recognition, but we could use any speech recognition program, right? It could use Siri, right? To understand text and then as soon as speech become digitized then

Watson can extract that. So, it just needs language to be digitized and that can be through any process." (X2, 2015)

- Capability to understand context

To better understand how Watson operates, we can take a look at the comparison which X2 draws between Google and Watson:

X2: "Watson is a system that can answer questions, ok? Google is not a system that answers questions. Google is a great system for finding facts from the internet, and fact finding isn't always answering a question. Fact finding won't answer questions that begin with 'Why?', right? Or 'How?', or 'What if?', right? And what you need to answer those kinds of questions is a system that can understand the context within which your question is being asked, and then be able to go and look for facts, that are relevant to the question that is being asked and then put them together in such a way that it addressed the underlines question." (X2, 2015)

As X2 continues, he gives examples of how Watson makes use of the context, which is radically different from how other fact-finding technologies operate:

X2: "So, Watson can understand the difference between the phrases 'noses run' and 'feet smell', right? Now this is, that's a big deal, it can understand when you say 'My nose is running', it's not thinking, well, ok, your nose is now grown feet and is running down the street, right? It knows that you got a cold. If someone says 'My feet smell', again, it doesn't think, ok, your feet have got a nose growing on them and that's capturing odor, it understands that there is something that is causing a bad smell coming from someone's feet. And that's very different from any kind of fact-finding technology that is out there, such as Google, right?" (X2, 2015)

- Languages

X2 says that Watson has some components which operate in multiple languages, while the cognitive side of Watson currently operates only in a few languages:

X2: "Oh yeah, it's able to do Portuguese, Spanish and Japanese I think. We have a bunch of languages on the road map, but those were all driven by client demand. Now, Watson is composed of different components, and some components have like seventeen languages. So, some, so the content analytics piece, for example, has I think seventeen or eighteen languages that it could work on, but the cognitive side of Watson is only currently in Portuguese, Spanish, English and Japanese, I think." (X2, 2015)

- Flexibility

IBM can build a prototype of Watson in a very quick period of time, so that the interested organization could start using it or testing it pretty fast:

> *X2: "..we can build a prototype for most organizations within about two to four months, right? So, for organizations to start using or testing ideas for cognitive is very quick." (X2, 2015)*

- Opportunities

X2 explained that Watson is its own business unit:

> *X2: "So Watson is the flagship output of IBM research facility at the moment. It has recently has been made into its own business unit, so IBM now putting the whole capability together in such a way that it can be used by other companies." (X2, 2015)*

Watson has big potential for application in the medical field. X2 tells, that seeing Watson win the 'Jeopardy!' brought it to the attention of the medical field, and eventually lead IBM to close collaboration with one of the leading oncology hospitals in the US:

> *X2: "Well, right after the game show was televised in the US, we started getting calls from medical professionals about how this kind of technology could be applied to health care. And so, IBM started working with various different partners in the health care space to define particular use case, and something that would be meaningful and something that would affect people's lives on the profound level. And, so, we started working with one of the US leading oncology hospitals – 'Memorial Sloan Kettering' to apply the Watson technologies to helping doctors to treat cancer." (X2, 2015)*

However, X2 confirms that the fact that Watson is a cloud service might be an issue for some companies, for instance, pharmaceutical companies, for which data security is of very high importance due to secrecy. However, he also believes it is not the biggest issue there is with Watson, and IBM can work on this issue in collaboration with their clients:

> *X2: "…Yeah, I think that is a problem, right? And, it's going to be something that IBM and its clients have to work through, ok?...But I doubt it's the biggest problem." (X2, 2015)*

According to X2, Watson has really good prospects in Denmark due to several factors, one of which is the fact that in Denmark there is a very well developed social sense:

> *X2: "..I think Denmark will adopt cognitive computing very, very quickly. Very quickly! The population is small, the infrastructure is extremely good, the appetite to innovate, to improve quality of life is huge in Denmark, right? And the most important thing there, the right scene is that the people have a very, very high social sense, right? So, they don't want something that just good for them, which is the kind of American way of going about it, right? So, 'what do I get, how do, how does it then affect me?'. I think the Danes tend to think more about 'Ok, this might be*

good for me, but is it good for my neighbor?', right? And 'Will the guy across the street get upset if I did this?', right? And, and so I think that, if cognitive was able to improve healthcare in a way that people get better quality health care at a lower cost too, you know, that would be that initial way of getting cognitive in. And then you have more of the consumer type things coming in, where, you know, it can be more ecological about food, it can be more cost effective about travel, it can help, you know, various different aspects of your, your life. I think that Denmark is going to be open to that." (X2, 2015)

Among other important factors influencing Watson's success X2 mentions good infrastructure and high interest in new technologies:

X2: "The infrastructure is fantastic, the appetite for gadgets is huge, right? So, internet of things, quantified self and mobility, and social are extremely high in Denmark. Those are all the fundamental requirements for something like Watson." (X2, 2015)

X2 also tells that currently IBM has many customers in Denmark in different sectors:

X2: "..we have many customers in healthcare, in consumer goods, in public sector, so, we are very present in Denmark." (X2, 2015)

When asked if there could be any barrier in Denmark for Watson, X2 was very optimistic:

X2: "No, I, I don't think there are any barriers specific to Denmark, right? I mean, there is, there is the general view of: well, until I understand what it actually does I am not going to use it." (X2, 2015)

- Threats

Uncertainty?

According to X2, in order to know full operational capacity of Watson, it should be used in practice by many practitioners. This would help to open up and discover the true potentials of Watson, as well as figure what weaknesses and limitations it may have:

X2: "So it has capability, it's just its scale, so it hasn't been applied to fifty hospitals, right? You haven't got ten thousand doctors using it yet. You don't have it being used in fifty research and development departments in the pharmaceutical industry, there is just no scale to it yet. Once people start using Watson at its maximum scale, then you will start to see what the weaknesses are and what its limitations are. At this stage we can only say that, you know, it takes time to teach it what to do. We just don't know what the limits of what it can do is yet." (X2, 2015)

However, it may be potentially risky to release technology in the market without knowing what "the limits of what it can do" are. This could mean that there is a need for a closer look into Watson and a thorough assessment by the experts in order to determine what potential implication this new and existing technology can have not only on such things as expertise and our understanding and use of it, but also on a broader, social aspects of

our lives, considering the fact that it has the potential capability to influence job market, ethics, privacy and many other areas of our lives.

X2 states that no one really knows what will be the implications of Watson and where this new technology will lead us:

> *X2: ".. if anybody thinks that they have a compelling idea of how this is going to be – they are wrong. The one thing we know for sure is we have no idea what is going to happen, ok? And that's the thing, we have to remember about this. Any kind of fundamental technology, any kind of fundamental technology, when it becomes mature is totally different from before it was mature. And, so, we have no idea what it will actually be. Look at mobile phones. In the mid-nineties nobody expected smart phones to be where they are ten years later. It was missed by everyone, right? And it has changed the world. So, no one can forecast these things." (X2, 2015)*

This could go in line with SCOT, considering that Watson is not yet a stabilized technology and it will take time before it will be stabilized. However, this does not change the fact that we are accepting the risk of developing new technology without fully understanding what consequences it will have on our lives. Performing real-time TA would be one of the possible ways to addressing this issue, initiating development of a regulatory framework to regulate and manage AI technologies would be another (Guston and Sarewitz, 2002).

Jobs?

X2 says that Watson will affect jobs: it will create new jobs, but it also will overtake other jobs. X2 gives several examples of jobs operated by humans have been replaced by technologies, and, according to X2, this is how it should be:

> *X2: "There, there is, I mean, think about any digital, sorry, not digital, any innovation – it changes the way people work. That's the point, right? I mean, if you think about what happened with mobiles phones and telephone operation companies, right? Mobile phone has for all intense and purposes reduces the number, well, it's, it eliminated telephone operators – you don't need people plugin in the, the cables or to connect landlines, or anything like that anymore. If you think about video cameras, right? For surveillance you're saying? So, don't you know a lot of security jobs have now gone, changed, right? You don't need guards patrolling building anymore, you need them looking at video. So, we think that Watson will change the way people do jobs. It will create a lot of jobs, but certain jobs, of course, just will get eliminated, because that's the whole point of technology progress…So, there've always been changes taking place in the economy – Watson will drive more changes. This, it's the way it should be." (X2, 2015)*

X2 also mentions some of the jobs which he thinks will be affected, also stating that Watson will contribute to less people in the working places, due to Watson overtaking their job positions and changing the way these jobs are performed:

> *X2: "And that just been a constant part of human civilization and we think that Watson is going to drive a lot of change in society, and we can imagine some of the*

tasks around, so, the transactional processing, service centers, call centers, all these kinds of things are going to change because of Watson - you'll have less people." (X2, 2015)

Ethics, responsibility and privacy?

X2 says that with Watson there is a potential for ethical issues as well as issues related to privacy. However, he also states that it is hard to predict the dangers that may come along with Watson:

X2: "Ethics? Definitely, yes, of course!...People, the, you know, complete lack of privacy can come as a result of this. Do you remember some of the discussions when 'Google Glass' came out? People are wearing 'Google Glass' and they walk down the street – everybody they see has their picture taken! What if they don't want a picture taken?! Just because this guy has got, is wearing 'Google Glass', he can do that…We, again, like any cutting edge technology, the power of the opportunity for damage is huge, because we just don't know the extent of its capability." (X2, 2015)

X2 draws analogy between Watson and research on radioactivity, as well as genetic engineering, in order to illustrate that potential implications of Watson are still unknown:

X2: "When people first researched radioactivity they didn't know it could cause cancer, right? When we're on the verge of a large scale genetic engineering, we don't know what can happen to ecosystem as a result of that. So, I think that there is a huge requirement for a deep thinking about ethics and the application of cognitive at this infinitely scalable cognitive systems, and what they mean for the way that we live and work, right? So, you know, there are problems." (X2, 2015)

So, according to X2 there are problems related to Watson, however, he suggests we will have "to address them as we come across them" *(X2, 2015)*:

X2: "Of course, there are huge problems and we going to need to address those as we come across them. And we going to need to think about what they potentially could be and come up with ways of managing those." (X2, 2015)

However, drawing from the experience with the same dangerous technologies as radioactivity research, it is dangerous to wait till we actually come across the problems, since then it simply may be too late to do anything. Following Collingridge (1980), I want to bring to the attention fact that in the early days of technology, it is still easy to take set of actions, such as implement changes, controls or make amendments, despite the fact that it is difficult to predict the consequences of that particular technology in order to justify the necessity of taking those actions. Here is where the real-time Technology Assessment could play an important role. This issue should be taken seriously, since once the technology gets more mature and becomes diffused it is very hard and costly to change anything, and often it is impossible (Collingridge, 1980:16).

Danger of one global system?

X2 does not think there would emerge one unified AI system, which may be able to do different tasks. X2 bases his assertion on the idea that in order to perform different task, there would be a need to possess a different set of skills:

> *X2: "No. I don't think that you would even, I don't think you would pay for one system to do everything for you, right? Because that's you, right? So, a self-driving car need a very different set of skills and an inside from a cognitive fridge, right? Now, the cognitive fridge and the car can talk to each other. So, you can say to the fridge 'Ok, this is my shopping list that I want'. It says to the car 'Go to the shop at this time', and the fridge tells the shop 'Have this grocery stuff ready, because the car is going to be there at that particular time', right?" (X2, 2015)*

However, considering that Watson is a technology that consist of many different components, and does not operate on human terms, in theory, it is not unrealistic to imagine that it could be taught and trained to possess different and not necessary complementary skills.

- <u>Suggestions for solutions</u>

When asked if in his opinion a thorough technology assessment and employment of precautionary principle could be seen as a possible way to address the uncertainty that Watson carries with it, X2 suggests that there might be necessary to create a regulatory framework in order to manage Watson. To illustrate why this could be a good idea, X2 draws parallel between Watson and medicines. He explains that eventually there emerged a necessity to manage medicines, in order to derive the benefits they create and avoid issues and even dangers, which they have the power to carry as well. For that purpose the Food and Drug Administration (FDA) was established in the United States. The FDA takes the role of managing and regulating medicines:

> *X2: "I think we need to look, and it might be interesting for you to draw this as a parallel: you need to look at the evolution of the Food and Drug Administration – the FDA in the US, right? So, when medicines first became more than simply some guy walking up and saying 'Look, drink this bottle of elixir, it will change..', you know, so, when did apothecaries change and became chemists? Right? And, and how did the regulatory framework that the current drug industry has to work on that, how did that come about? The reason I think that that's the important parallel is because medicines have extended human life span, right… And that, they've changed, you know, medicine changed the way in which the economy, in which humans work. And I think that cognitive computing is going to have a similar effect: it will make people's lives very different, it might, if Watson becomes very good at medicine, it might actually become, if Watson becomes good at medicine, it might actually become a way in which people's life span gets expanded even further. Now, if that's the case, how do we regulate that? How do we manage that? How do we ensure that the benefits outweigh the risks, ok?… if we are at the place where antibiotics just got*

invented, so cognitive computing was to be seen as antibiotics, what could the next path be for the evolution of a regulatory framework for cognitive?" (X2, 2015)

X2 believes that for adoption of cognitive technology to happen, it should be used by more customers, this way it would create a chain reaction, which would contribute to its adoption and thus its stabilization process:

"I think the real, the way in which adoption of cognitive computing will accelerate is when one or two organizations start showing benefits from using it, right? So, we need to get beyond the proof of concept or prototype phase and into actual implementation for business, for people to then say 'Well I want one of those', right? And then everyone starts to get on the band wagon. So, it's only because it is a new technology that is taking a bit more time." (X2, 2015)

- Expertise

Due to the fact that Watson is capable of learning, it is constantly improving, which means it is getting more experienced with time. This fact differentiates Watson from any other current technology and, X2 suggests that that enables it to be seen as an additional resource in the team – which could be paraphrased as Watson being part of the team:

X2: "Now, when you look at any kind of software system, any software program that's available today, it operates at its best possible level on day one, and at exactly the same level on day one thousand, right? Watson is at its poorest level on day one and will be incredibly good on day thousand, because it keeps learning, it gets feedback, so you have to think of Watson as being an additional resource in your team, not a computer." (X2, 2015)

These capabilities enable Watson to gain experience, and thus gain expertise just as our ability to learn through experience gives us possibility to become experts in certain practices.

X2 says that unlike human, who can decide on his/her own to go and acquire certain information, Watson needs to be told what to do, it needs to be taught:

X2: "..It needs training, it needs feedback, it needs to learn and it gets better and better, and better, ok? So, that's how it works. So, in terms of weaknesses Watson needs to be fed information, right? Unlike a human it won't just go out and say 'Oh, actually I got some free time, let me go and learn this' alright? 'Let me apply this new back of knowledge to.. ' It needs constant feedback, it needs to be taught how to think and how to do things for particular tasks." (X2, 2015)

However, on a closer look this fact places Watson on the stage of novice in Dreyfus and Dreyfus' (1988) "Five stages of skill acquisition", and considering that Watson is constantly learning and getting better, we can assume that this is just the initial stage of Watson and with time it has potential to go through all five stages and reach the top stage – become an expert.

In the following comment X2 differentiates between human and artificial intelligence:

> *"This is artificial intelligence, it's not human intelligence and so it doesn't work in the same way human intelligence does, ok? It's very good at doing one particular, very narrow task: that might be winning 'Jeopardy!', it might be advising doctors on oncology treatment, but the intelligence and the learning required for 'Jeopardy!' is very different from the intelligence and the learning required for the oncology, and so they are not the same, right?" (X2, 2015)*

However, looking at human intelligence, skills required for being a doctor are not the same as the ones that are required to be a good bus driver, which in fact weakens the argument presented by X2 and minimizes the gap between human and AI intelligence, since we can see from previous sections of this thesis that basically the same elements that constitute to the skill acquisition are present both in human and in AI intelligence: ability to learn, recognize patterns and ability to make links. Besides, just as in human mind different parts of it are responsible for different cognitive processed, AI in theory could be composed of different elements responsible for different cognitive processes and contribute to different skill and expertise acquisition.

Previously stating that Watson can be considered as an additional resource in a team, X2 considers it to be a weakness that it takes time to teach Watson:

> *X2: "So, the weakness here is that you need to spend time teaching Watson what it needs to do, ok? But other than that, I think, it's difficult to identify the weaknesses at this stage, simply because it hasn't really been maximized, we don't know what it would do when it is at full scale." (X2, 2015)*

However, considering the fact that it takes considerably longer time for its human colleagues to reach the stage of expert, regardless of which practice we are talking about (yet assuming we are talking about practices involving reasonable amount of education, such as doctors, for instance), a time which takes to teach Watson could be considered minimal, also keeping in mind the fact that Watson is capable to pass a medical exam at its current development stage:

> *X2: "..we trained it to the level where it can now pass a medical exam, ok? So, that's a really big deal. I mean, a medical, when you think about students, what the students have to go through and how exam questions are asked, right? You know, that starts with 'In your opinion..' right? (X2, 2015)*

We shall also consider that Watson won 'Jeopardy!' in 2011, which means that Watson in theory should have reached even higher level of cognition since then. I was having the interview with X2 in 2015, so, Watson's cognitive abilities on a reasonably high level were under development for roughly 4 years, and in these 4 years Watson reached the level where it can pass medical exam. Considering the fact that it takes as minimum as 7 years, according to X2, to train a doctor, Watson's advancement is very impressive:

> *X2: "It, it takes seven years to train a doctor, alright? Six years in some countries, seven years in most. That seven years is spent in building up knowledge, being,*

getting practical experience with patients, and then becoming experienced enough to address all kinds of tests, if a test is telling me something, is giving me a number, then based on a history I took I can make this diagnosis. That's exactly the kind of process that Watson goes through. It looks for facts, it understands what's being said by medical professional and then it makes the link to say: ok, if those two things are coming together, then this is what potentially could be happening." (X2, 2015)

When talking about how Watson handles big data, which would apply how he would handle knowledge corpus in the specific area in which it would be employed, X2 mentions such factors as the speed at which Watson is able to access and process information, make linkages and able create meanings:

X2: "Watson's analytics, its ability to make connections and the speed at which it works I think are very important. Now, you have all this information streaming in, what does it actually mean? You need a cognitive system that can understand whether linkages are happening very, very fast, and present the findings, as answers, as insights, as decisions that need to be made." (X2, 2015)

Following from that we can assume that the time necessary to train Watson to the same level of its human counterpart is reasonably shorter. This could be due to such factors as Watson not being limited by memory factors (humans have *working memory,* and long term memory, which they can't access easily (refer to section 2.3)), the speed with which it is capable to access and process information and possibly other factors.

From X2's words, it is possible to state that any current assumptions about Watson's inability to compare to human expertise are premature:

X2: "..I think it's difficult to identify the weaknesses at this stage, simply because it hasn't really been maximized, we don't know what it would do when it is at full scale." (X2, 2015)

4.2.3 Interview with Y1

Interview with Y1 took place in DTU, Lyngby.

In the following Y1 presents his background and current occupation:

Y1: "I've been studying computer science at the university of Copenhagen, and I wrote my thesis in applied mathematics in here, at DTU, and afterwards I have some time in the industry, where I have worked as an analyst…I've created risk models for use in pension funds and large firms…I am a PhD student here in DTU, and I am writing about intelligent text algorithms for search engines… specifically I am developing this new search engine for rear diseases, called 'Find Zebra', which is online and functioning.." (Y1, 2015)

The experience in computer science, applied mathematics and intelligent text algorithms qualifies Y1 as an interesting interviewee despite the fact that he is not an expert in Watson and only been learning about Watson on his own, in order to give a lecture about Watson for DTU students:

Y1: "..I have given this small presentation about Watson, about what's behind Watson, but I think that's it. So, that's basically all I know about Watson. So, I am not totally a specialist in that, I am just a, I've read the articles behind it, so…The course had a lot of people and they had the possibility of getting teaching assistant, and that was me, so, and I'm concerned about Watson and that's what I'm focusing on. This and some other projects about text mining, I think we also have some, two projects in text mining, and two projects in yeah." (Y1, 2015)

- Data

Y1 says that in short time Watson performs many processes, which make him "heavy", thus he is not sure if Watson would be helpful with regards to big data:

Y1: "Watson is very computer attentive just to ask, eller [foreign] just to answer a single question in like a feasible time, like, three seconds it needs, like, almost three thousand processes, and, yeah, and that would not really do if you have…perhaps it will, I am not that certain about it. But I think it's, when we have tried to upload a corpus for example for this test version, I think it's very difficult to…, takes a lot of time just to upload a little corpus, so…, imagine if you have one hundred gigabytes of data then it will not really fit into to that framework." (Y1, 2015)

- Simplicity

Y1 states that Watson is based on some basic and not that advanced technique. He also suggests that simplicity and simple models used in Watson might actually be the "state of the art", since more complex models turned out to be not useful:

Y1: "..Watson is like the state of the art in text search, I think. I think they've tried almost all the models that you can use up till now on this Watson project...the technique used in Watson is not very advanced…they use some very basic stuff, but that also a good thing in a way, because they've tried all kinds of advanced stuff and it didn't work. Now they are back to basics, and that actually shows, they show what actually works. So, I think that's a really good thing, this shows that some of the more simple models actually work really well and is actually the simple models are perhaps state of the art right now, until we get some more advanced models, and that's a good thing, but I think the methods used are very basic, it's not very hard core deep learning stuff.." (Y1, 2015)

- Learning

In order to be able to use Watson, knowledge corpus has to be built, and Y1 tells that it took long time to upload just some documents to it, which could mean that in practice building knowledge corpus might be a time consuming process:

Y1: "..I think that Watson is very heavy too, you heard that, Finn talked about that it took long time to upload just a little bit of a document corpus, if you have a lot of documents then it will become very cumbersome *to use Watson.." (Y1, 2015)*

- Languages

The fact that Watson is very rule based when it approaches sentences could present some difficulty when using Watson for a different language than English:

> *Y1: "Watson is kind of rule based, a lot of Watson is rule based, that means that they look for sentences that…, if they, for example, want to discover the focus of sentence, that is, what is the answer, what is that, what part of the question is referring to the answer, then they use this rule, for example: he, she, it. If that occurs in the sentence, then that is the focus of the question. So, if you want to do that in Danish, you have to find the corresponding words, or, perhaps, it's different in Danish, so, of course, moving to another query on another language would be quite difficult, you also need a more, a custom part for Danish language, it could be also, could be difficult, yeah." (Y1, 2015)*

- Application opportunities

Y1 commented that he sees two major areas where Watson could be used – medical field and financial industry:

> *Y1: "Yeah, medical healthcare is a.., …in the financial industry also I think that they are actually been, it has been used in financial industry, but other areas… No, I can't really say." (Y1, 2015)*

Y1 tells that he is aware of Watson being used in the medical field, more precisely, in hospitals:

> *Y1: "I know they are used in some hospitals…they have built this medical Watson…I think it's available everywhere…and that is a system, where you can, for example, I think, where you can speak to Watson, and then you can input some patient history and stuff like that and then you get some proposed diagnosis and stuff like that." (Y1, 2015)*

However, Y1 thinks that IBM has not been very good at promoting Watson:

> *Y1: "Watson has very poor social skills, they have not been that good at networking in, for example, in the industry and I haven't really seen it used a lot of places.." (Y1, 2015)*

Nevertheless, Y1 mentions knowing a person who has been visiting doctors who are using Watson:

> *Y1: "I think you can find some hospitals, perhaps…I know a guy who have just been visiting some doctors, who actually use this Watson, but it's in United States." (Y1, 2015)*

The main benefit of Watson in the medical field, according to Y1, is the extra knowledge that Watson incorporates in the proposed diagnoses, by accessing external databases:

Y1: "Well, if they can find the right diagnoses, or not the right diagnoses, but they might, when a doctor tries to find a, what, to find what diagnoses for a patient, they usually just write this, these, for example, ten possible hypothesis of diagnoses and then they find out which one is most likely of those. That's basically what Watson is doing, because, they find the ten most, perhaps, likely diagnoses and then they, then Watson ranks the different diagnoses and say: this has this probability of being the right diagnoses based on what we know. Watson also has the ability to easily incorporate, for example, external knowledge, like is find in these UMLS databases 'Unified Medical Language' something, where you have, for example, list of synonym diseases and stuff like that.." (Y1, 2015)

Loreta: "So it's like extra…" (Loreta, 2015)

Y1: "Extra knowledge, yes, so you can use.." (Y1, 2015)

However, Y1 is a bit skeptical about how much doctors would actually use Watson in practice. In this case Watson could be seen as an extra tool for experts to be used in uncertain situations.

Y1: "I don't think that doctor will use system like Watson, except when he doesn't really know what can be the cause of this disease, and then he only will use Watson, I think, to find a list of candidate diseases. Then he will decide which of these actually makes sense, and in that way he really doesn't make his decision based on Watson, but he makes his own decision based on the suggestion from Watson, so." (Y1, 2015)

Y1 thinks, that IBM might not have that many customers in Denmark, despite that fact that Denmark is ready for Watson:

Y1: "I think… I am not quite sure, because you have to focus on the domain, because if you look at the medical domain, for example, there might be a possibility for Watson, but again, people will only use this tool rarely, I think. So, perhaps, they cannot make that much money of Watson in Denmark, but I'm, I don't think that there are anyone who is not really ready for Watson. It's like, I'm more uncertain if IBM has a business case here in Denmark, but I don't know. I don't know about.." (Y1, 2015)

- Expertise

Y1 thinks that Watson will only be used in unknown situation, when human experts won't know what to do:

Y1: "I don't think that doctor will use system like Watson, except when he doesn't really know what can be the cause of this disease, and then he only will use Watson, I think, to find a list of candidate diseases. Then he will decide which of these actually makes sense, and in that way he really doesn't make his decision based on Watson, but he makes his own decision based on the suggestion from Watson, so." (Y1, 2015)

Y1's view of Watson's application is similar to the adviser role of Watson which is promoted by IBM. However, if Watson would only be used in the unknown situations, the

potential for its commercial future could decrease. Nevertheless, it does not diminish its value, since have an adviser in an unknown situation could prove to be very valuable.

4.2.4 Interview with Y2

Interview with Y2 took place in DTU, Lyngby.

Y2 describes his expertise area as follows:

> *Y2: "So, I am Y2, senior researcher here at DTU – Technical University of Denmark, and I have a background as a civil engineer from this university and also PhD…Senior researcher, so I am attached to research projects at the moment, it is neuro twenty four syv [foreign], what you call, 'Neuro twenty four seven' – a research project related to monitoring continuously information about the brain from EEG devices."(Y2, 2015)*

- Data

Y2 explains that the commercial version of Watson is cloud based, which means that the documents, which are used to build the corpus are transferred through the internet to the site which will host them, which means that they will be stored on the hard drive, which belongs to IBM, somewhere in the world. Y2 also assumes that this means that IBM will have reading access to the document files, which will be part of technical costumer support service:

> *Y2: "The commercial version is cloud based, the service is running on computers on somewhere in the world, I actually don't know where it's hosted. So, it's not sort of computer that IBM provide and then put somewhere in DTU. They just give us internet address and account information, so we can log in and use that. So, when documents are, need to be indexed, or they need to be transferred across the internet to the site that host the files, so, somewhere the documents that I've uploaded are stored and on some hard drive that IBM controls, and… I suppose that the developers and some of the technical people have access to looking up the documents that I've uploaded, so, to support, for example, also. If they sell it, I suppose that they, the support people would need to access the documents so that they can support the customer with problems that the customer might have." (Y2, 2015)*

When asked if in his opinion, there could be some issues with data security and privacy, Y2 commented that companies would need to consider how they would expose their internal documents to IBM. However, he also mentioned that majority of ordinary policy intranet does not contain private information, which points to the fact that the data privacy issue depends on the data content and on the measures of data protection which IBM will employ to secure their customers:

> *Y2: "Yeah, that could presumably have some problems, yes, but as, if you, on the other hand, you'll sort of ordinary policy documents intranet, where there is no personality, private information, I suppose that would not be a problem, but the*

company that want to gage in IBM would then think about how should we then expose our internal documents to IBM." (Y2, 2015)

When asked about Watson with regards to possible concerns of pharmaceutical companies due to the high level of secrecy because of high competitiveness, Y2 commented that companies could create a corpus by uploading public documents to it, thus not risking any sensitive data:

Y2: "But I suppose they can also upload specific public documents and use that to query." (Y2, 2015)

However, it is questionable if Watson would be used in its full capacity if for creating its corpus only selected data, which could be labeled not sensitive data, would be used. So, the issue of data security when talking about cloud services still remains open and needs to be solved.

Y2 also mentioned the fact that when researchers are querying in Google or PubMed[4], those search engines then know the content of the query, thus, the simple act of performing Google or PubMed search can be seen as potentially unsafe, considering secrecy concerns in the pharmaceutical industry:

Y2: "They already, when they do a patent search or internet search using PubMed, they already using public facilities, so it is, so if a researcher uses google, if biomedical researcher is using google, then google already sort of knows that the researcher is, what researcher is using. So, it's just another facility or, unless you want to download and sort of have the entire internet and have, and also the entire PubMed, and, and other facilities, you need to have some kind of library, internet library access." (Y2, 2015)

Considering this comment, it is possible to assume that Y2 suggests that Watson could be compared to existing search engines, among which there are medical search engines, and the risks that Watson may pose don't differ much from the risks undertaken by researchers while using existing search engines.

According to Y2, the issue with privacy is met each time a search is performed on Google, and performing it using Watson would not be much different:

Y2: "If you're using a standard internet search engine today and you have a disease and you google for the symptoms, then the search engine will know by your query terms, probably, that you or one of your friends, or family are interested in disease, so that will also be the, perhaps, the case…depending on how Watson internally operates, I suppose, that could be also an issue." (Y2, 2015)

[4] PubMed - free search engine, that is accessing primarily the MEDLINE database of references and abstracts on life sciences and biomedical topics http://en.wikipedia.org/wiki/PubMed [accessed 06.06.2015]

Y2 also assumes that in the case of Watson, the entity, which has the diagnostic corpus would be the one who would receive the queries, and then send it to Watson service, however, this would then indicate the presence of third party, which could in theory complicate the issue of privacy and data security even further. However, Y2 suggest that this actually could be beneficial in terms of queries getting received by the service provider and then, from the pool of such queries sent to Watson service:

> *Y2: "..in the Watson case, I would imagine that the service provider, that the entity that has the diagnostic corpus would be the ones that receive each individual's query and then send it on to the Watson service. So, it would be some of a pool, it would be harder for IBM, as such, to so see any, to link it to any particular individuals. It would be for example the entity that has the actual service." (Y2, 2015)*

When asked what he thinks with regards to feeding personal medical history to Watson when using it in the medical field and if this could create some privacy issues, Y2 admits that in his opinion this could be an issue:

> *Y2: "So, you don't necessarily need to put the patient history, but, of course, that would boost if you can personalize by putting in the medical history, that would presumably boost the performance of the system when doing diagnoses, so there, in that case there is, there will be an issue." (Y2, 2015)*

With regards to data privacy, Y2 speculates that Watson may become similar to Google data monopoly, in the way Google is accessing and using information, which is freely given away to it by its users in return for convenience:

> *Y2: "..at the moment Google has quite large share of search, of internet search engine, also reading emails, if they are in Gmail, and if you are…if Watson is passing, crunching data, it might be along the same line as Google, it's as data monopoly of information…perhaps it's better to say "Big Mother" or something like that, that you use a service because of convenience, on the other hand you give away some of your information and some control." (Y2, 2015)*

- <u>Capability to understand natural language</u>

According to Y2, the cloud service form of Watson that they are working with in the university is based on passage retrieval, which is a natural language processing mode that allows to retrieve passages from heterogeneous texts, which are matched with the user's information needs (IMS, 2015):

> *Y2: "So, for us, it's a cloud service with some natural language processing that allows you to do, what it's called, passage retrieval based on natural language formulated questions, so it's not directly, yes…you might have seen TV, YouTube, where the 'Jeopardy!' version allows you to specific questions with specific, with short answers. What we are returning from cloud version is a passage from the document retrieval. So it is more related to ordinary search engine, internet search engine, or search engine that you install on your own computer and index your own databases." (Y2, 2015)*

According to Y2, one of the benefits of Watson is the fact that it has a better understanding of synonyms and relations between words than standard search engines do, which makes Watson better at search retrieval:

> *Y2: "..that would be to have a better search retrieval performance, what I regard at least at the moment as the cloud service is sort of the search engine, like where you can build up a special relation between words, so it has better understanding in synonyms and stuff like that, and issue that you sort of rarely find in standard search engines, perhaps. There is, some of the search engine has the ability to do spelling corrections, I suppose, but not the real synonyms, I believe, it's a bit rare." (Y2, 2015)*

Y2 tells that according to the documentation about Watson there should be possible to use speech to text way of communicating with Watson, however, the instance of Watson cloud service which he has access to does not have it:

> *Y2: "Yes, but the IBM developer cloud has also some other services and unfortunately I am not familiar with that…there is some speech to text and text to speech conversion, so, it might, if you have access to these part of the IBM cloud services, you might be able to do that. But I am, unfortunately, I am not into that part and I haven't tried it. But as far as I can see on the documentation that there is, there might be these possibilities.." (Y2, 2015)*

- Capability to learn

For cloud based Watson to function, the knowledge corpus has to be built by providing the data for it, and from this corpus Watson will return the answers to the questions asked by the user in the form of the relevant passage, retrieved from the documents or whatever is the knowledge corpus is built of:

> *Y2: "..you actually supposed to provide your own data and the answers you are getting out is not short questions to specific fact in the document, but rather the default approach to this return a page with the, with the passage of relevance to you." (Y2, 2015)*

According to Y2, the 'Jeopardy!' Watson's main objective was to answer specific 'Jeopardy!' questions. In his opinion, this could have been achieved by running different algorithms, such as rhyming algorithm and by the presence of a corpus which, for example, could have been based on Wikipedia:

> *Y2: "There is the Watson, the 'Jeopardy!' Watson version and there is the cloud based commercialized version of Watson. So not clear to me at least how…, what's the, how much overlap there is. But the one, one difference is that in the, in the 'Jeopardy!' version, they, the task was to answer specific 'Jeopardy!' questions. They had algorithms, for example, for rhyming, and they had fixed corpus with, for example, based on Wikipedia and related over.." (Y2, 2015)*

In a similar vein, cloud version of Watson is based on corpus, however, in this case the user is building it, by uploading the documents he/she want to be in the corpus:

Y2: ".. in the cloud-based version you can upload your own document... the cloud version is not going to answer, sort of 'Jeopardy!'-like questions out of the box. You need to get the corpus and think about how much should there be and how long does it take to process the documents, and how many do we need for Watson to be trained at…one has to find a suitable corpus." (Y2, 2015)

According to Y2, the instance with which DTU is working has limited amount of space and is taking a considerable amount of time for document processing, which lead them to limit the amount of documents to upload in order to create a corpus:

Y2: "..at the moment there is a limitation how much memory each instance can have, at least in the instance that we have. So, I believe it's on some gigabytes, something like that, but there is also sort of the instance how much processing time each document takes, so there again working with Wikipedia scale or web scale document collection is…, at the moment will take quite a considerable time. So at the moment we have to sort of scale down to fairly few documents loaded up to the instance." (Y2, 2015)

This could be due to the fact that the university is given the version of Watson which has scaled down capabilities compared to the fully scaled up version of Watson, however, it is uncertain at this point. Another issue that Y2 mentions is the issue of interaction. In order to teach Watson, the questions for queries have to be entered manually, which might appear as a difficult task, considering that Watson should be presented with as many questions as possible in order to account for possible questions in the future:

Y2: "..other issue is related to the interaction. The default interface, there you will need to specify the questions by hand, so to speak, in the weird interface, so you got a lot of questions, there is the issue that you'll have, either you'll need, you will need a lot of people to enter the questions to allow Watson to train. You can type in the question and the answer on the webpage, but if you got the three thousand questions it will take considerable time to do that." (Y2, 2015)

According to Y2, there is always a limit to performance, and Watson may not be performing 100%, since even his optimized 'Jeopardy!' version failed to answer some questions:

Y2: "There is of course also limitation on the performance in some way or another, even in the 'Jeopardy!' version that they optimized specific for that, that they couldn't answer every question, that they only answer with the certain probability and we are also at the moment trying to evaluate what is the limitation of Watson, and in that respect we are sort of learning what's possible and what's not possible. So, this obvious in any information retrieval system, there is this obvious limitation, it's not completely hundred percent perfect." (Y2, 2015)

- Languages

Y2 thinks that Watson may have some algorithms that are indifferent to language, however, the main language used for Watson is English, so it could be seen as limitation to use Watson in Denmark, for instance:

> *Y2: "Oh yes, that's also a problem. At the moment it's only working on English, and I believe they are working, perhaps in Japanese, perhaps in some…I am not sure about that. The language is, presumably, I don't know whether, perhaps some of the algorithms might be indifferent to the language, but I suppose that there're quite a number of the algorithms within Watson English, it's been on, that it's actually in English language." (Y2, 2015)*

- Application opportunities

Y2 gives an example of how Watson could be used in DTU for intranet and extranet pages:

> *Y2: "So, if you, for example, have, applying here in DTU, one obvious thing would be to index the, for example, the English extranet pages or, perhaps, also the intranet pages of DTU, and then use the Watson service to let other users, other DTU employees or prospect DTU students be able to query the service and then get the documents back that are related to the query, hopefully to better performance than the standard search interface that is on the DTU homepage." (Y2, 2015)*

Y2 also shares his more practical idea for future use of Watson in DTU's administration or library. In his opinion Watson could contribute to a better experience through better interface when performing queries:

> *Y2: "So, I was thinking of asking central administration, whether they had good corpora and the good questions for bots, and, so imagine that's a Watson service on central DTU servers to back up ordinary internet based search, could be future interest for, or at the administration or perhaps the library…I suppose that when we use the intranet of DTU and search the, search experience using that optimally, you have to find relevant documents, so you need to perhaps reformulate the query and down the list and read, so, perhaps, another interface, better and Watson based interface could perhaps in the future be relevant." (Y2, 2015)*

- Threats

Y2 thinks that one of the jobs, where human workers will be substituted by Watson will be support jobs:

> *Y2: "..one issue could be that they are automating the support functions in companies, so that people that call in and want to speak with some support, that they have problem, and instead of getting a human that would try to solve your problem, you getting a Watson voice, as something like that that try to guide you through. So, I suppose, many support, it's possible that some kind of support job, telephone support, people answering emails about customers having problems, some kind of, I don't know…That could be effected, yes." (Y2, 2015)*

Another job area which may be affected by Watson, in Y2's opinion, could be medical patients' support, where, in case of emergency, instead of talking to human operator you would get to talk to Watson:

> Y2: "*And also, I don't know whether, I think in the long term we'll also see that electronic based patient's collaboration would take place, so you would instead of calling in and your symptoms, you typing in and call in and that's the speech recognition, so the first line speaking to the medical system is electronic system that does some recording, perhaps, also some kind of, trying to understand the issue of the patient: can it be solved automatically in some way, or should it go through, should we call an ambulance home." (Y2, 2015)*

When asked about whether he thinks that the medical field in Denmark was ready for Watson, Y2 expressed his concern over the issue of obtaining medical records, which would be necessary to build knowledge corpus, if Watson should be used in the medical field to its optimal potential:

> *Y2: "I suppose the medical system, it would be, I think that the problem might be the issue of that it's difficult to obtain data and to work with data, you need to have special permissions. For example the Deep neural network doing well at image classification, and they are using public images from large databases with million documents and you don't usually see five million medical, Danish medical records floating around in systems. I suppose, there has been a few research groups that have access to the medical, Danish medical electronic medical records and can make some kind of core morbidity studies on it.." (Y2, 2015)*

- <u>Expertise</u>

When asked about what is his connection to Watson, Y2 commented as follows:

> *Y2: "So, as I've been setting for so phenotetical [phonetic] lead, or so, I suppose you could call it, I've been setting it up with some corpus corpora and testing with the few question collections we have." (Y2, 2015)*

When asked in a follow-up email to specify the meaning of "phenotetical lead", Y2 returned with following comment:

> *Y2: "I think what I meant was that: I have been the technical lead on the project, setting Watson up with corpora and testing it with the few question collections we have." (Y2, follow-up email, 13.05.2015)*

With regards to partnership between doctors and Watson, Y2 has commented that IBM was not alone in pursuit to assist doctors in making diagnosis:

> *Y2: "I think it's obvious that a lot of people, not just IBM Watson team are working on some systems that can help doctors and perhaps even patients also to do diagnoses. For example also the 'Find zebra' the system that comes from here, we set up, is one system in that domain." (Y2, 2015)*

When asked what in his opinion patients would think about doctors using Watson for assistance in making diagnoses, Y2 drew parallel to the fact that already now doctors are using search engines for that purpose, and some patients could be surprised to see that. However, even if doctors are using such assistance, they are the ones who in the end use their expertise to choose from suggested diagnoses the one that is relevant in particular case and formulate it properly in order to present it to the patient:

> *Y2: "So, doctors are already using google search engine for, to do diagnoses. I don't know, perhaps some patient would be surprised that they see the doctor is googling for diagnoses…I suppose it could be that some patients would find it bit odd. But, of course, the doctor has the ability to formulate regard the medical, doctors are able to formulate better the words for specific syndromes and better than perhaps a patient would have and also evaluate the response you get from a search engine, see which one are relevant and which one are not relevant." (Y2, 2015)*

Drawing from that, the same would potentially go for using Watson for assistance in making diagnoses.

Y2 also points out that the doctors are familiar with the tendency of patients to over-diagnosing themselves when using search engines for self-diagnoses:

> *Y2: "So, in terms of using the search engines, internet, ordinary internet search engines on this system, there are some doctors have said the patients tend to perhaps over diagnose themselves. If they google on, with, on Google for some of the symptoms they have and find out they have dangerous disease and go down to medical doctor and say I have this dreadful disease and then the doctor, the medical doctor with his experience can sort of immediately see that it's not the case. (Y2, 2015)*

So, the expertise of doctor assists him/her to recognize the symptoms and assign them to particular diagnoses, based on previous experiences. In comparison, when thinking about the automated system as an assistant for making diagnoses, Y2 is a bit skeptical considering the fact that it may not give direct diagnoses:

> *Y2: "And then, I suppose, that the problem might be that because of, systems are not, there are some limitations on these systems, so that they might not have, might not give the direct diagnosis…if this will benefit the medical doctor when he is doing the diagnoses." (Y2, 2015)*

It is questionable how this would affect doctors' performance, in case if instead of using his own expertise or intuition, he/she would reach out for an assistance of an automated cognitive system, and the latter would not meet his/her expectation of presenting relevant suggestion.

When asked if he thinks that one day Watson could be on the same level of expertise as human or perhaps even superior to human expert, Y2 expressed the opinion, which supports the idea that it is quite possible:

Y2: "I think that medical domain is becoming so difficult that it is difficult to educate doctors to sufficient high levels of the, to understand and remember all the things and this can be possible to do this more efficiently with electronic systems…I think that will be a matter of time before electronic system can do a really good job on at least some part of medical cases.." (Y2, 2015)

Where does this put human expert, who is limited by its human factor in comparison to AI expert, who has huge capacity memory-wise, has capability to access and process vast amounts of information in a blink of an eye, and thus always be up-to-date with the latest medical news and discoveries, who has ability to present hypothesis based on huge amount of data (both patients personal history and histories of others patients whose symptoms match, along with all the medical literature which has mentions of those symptoms) related to the symptoms that it has processed, and who, due to capability to learn, has the same capacity to develop expertise, though perhaps in a much faster time, than its human counterpart? Would human expertise seize to exist due to inability to compete with AI expertise? Those are few of the questions that are worth looking into.

However, Y2 sees the issue that electronic system will be missing the human element, which is, for instance present when the doctor is giving the diagnoses to the patient:

Y2: "I suppose that the issue is that some…, when a medical doctor confronts the patients, there is sort of a human, a more a human contact, a medical doctor trying to sort of comfort or so I suppose, that these aspects that are.." (Y2, 2015)

Electronic system, which does not possess emotions and feelings, and is disembodied could hardly compare to the comfort level which human expert may possess when communicating with patients. However, even with humans, the comfort element is not always given and perhaps depend on an individual ability or trait of an expert. However, when referring back to the patient, it may be more acceptable to receive diagnoses from fellow human than computer.

4.3 AI - current reality compared to anticipations

In this section I want to present the comparison between predictions about possible consequences of AI development made by a number of scientists decades ago, presented in "AI - anticipations" section in Table 2, and current AI, more specifically - Watson, seen through the eyes of four interviewees, whom I have interviewed. I have extracted main point from the interviews and placed them in Table 3.

By this comparison attempt, to some extent, to illustrate how the knowledge, perception, and values are evolving over time with regards to AI. This also partially goes in line with real-time Technology Assessment, which I believe is a necessary mean to provide mechanisms for observing, critiquing, and influencing social values as they become embedded in new technology, through policy analytics, economic, ethical, and other social science research that attempts to anticipate how new technology will interact with social systems, maximize the benefits and minimize risks of this new technology, along with making sure to be responsive to public interests and concerns with regards to it (Guston and Sarewitz, 2002:94-95).

Table 3 – Watson and its opportunities, threats and solutions to handle the threats

	Watson - Opportunities **WO**	Watson - Threats **WT**	Watson - Solutions **WS**
1	Capable to fast analyze big amounts of data (unstructured and structured) - "Data scientist"	Data security (cloud service might present some security concerns)	Regulatory framework for management of AI (inspired by the FDA regulatory framework for drug management)
2	Capable to understand natural language	Privacy	
3	Capable to reason the way humans do; work on a human level	Taking over human jobs	
4	Accesses databases with scientific articles, etc., which contributes to its precision and being up to date	Taking over expertise requiring jobs	
5	Capable to learn	Uncertainty – not knowing what are the limits	
6	Capable to recognize patterns and make links	Ethics	
7	Capable to learn from mistakes	Responsibility	
8	Hypothesis generation and confidence level		
9	Capable to understand context		
10	Capable to interpret pictures and video		
11	Adviser to human experts; decision support		
12	Flexible and customizable		

Table 3 consists of three columns. The first column is called "Watson - Opportunities" (WO) and is dedicated to Watson's abilities and capabilities and opportunities that arise from them. The second column is called "Watson - Threats" (WT) and includes possible threats that may come as a result of using Watson. The third column is called "Watson - Solutions" (WS) and include suggestions for possible solutions how to deal with some of the threats mentioned in the second column. Each column has two letters assigned to it. Each row of the table has a number assigned to it for quicker referencing. Thus, for instance, WO2 would be the key to the statement "Capable to understand natural language". The rest of the statements in the table can be referred to in a similar manner.

The table was generated by making an analysis of the interviews and by no means is the final version, since to make it more complete, more interviews (or other methods) and more research is necessary. It is important to mention that the table should be constantly updated since AI is constantly evolving – Watson is "growing" and so does its capabilities, which affect the opportunities it creates and threats it may pose.

The purpose of the table is to give a quick overview of Watson's capabilities and opportunities which arise from them, as well as threats that Watson may pose and possible ways to address those threats. Using Table 3 it is possible to compare it to Table 2 (section 4.1) and see how far or close current AI is from the predictions made by Drexler (1986) and Vinge (1993) decades ago. The benefit of this comparison could be expressed in ability to see whether some of the predictions did come true and if so, is there a valid ground to believe that the rest or some of the rest predictions may also come true, based on the presence of necessary factors which may contribute to them actually coming true.

Back in 1986 Drexler has defined what would genuine and flexible AI should look like, and his definition seems to be the mirror image of IBM's Watson:

> "A genuine, flexible AI system must evolve ideas. To do this, it must find or form hypotheses, generate variations, test them, and then modify or discard those found inadequate…To avoid becoming trapped by initial misconceptions, it will have to consider conflicting views, seeing how well each explains the data, and seeing whether one view can explain another." (Drexler, 1986:153)

Keeping in mind the quote above, it is possible to speculate that Watson has all the necessary components which may contribute to some of the predictions listed in Table 2 coming to reality. Looking at IBM's Watson, already now it is possible to state that predictions such as O1, O2, T3 (Table 2) did come true to some extent, and other predictions, such as O3, O6, O7, O12, O13, T1 (Table 2) may come true, as a result of the first row of the prediction coming true. Thus, it is possible to state that all predictions are interconnected and depend on each other, along with dependency on many other factors. From this leads that it is not unrealistic to suppose that even such dramatic predictions as T11 in theory are not impossible.

Watson is just one actor, or rather a huge network of many smaller actors, and it is important to acknowledge that in order to speculate whether the rest of the predictions in Table 2 might come true, all those many other actors and networks of actors have to be taken into consideration. However, it is still possible to recognize the underlying patterns and recognize the factors which may pave the road for ultimate AI, which in its turn may contribute to predictions coming true.

When talking about suggestions of how to deal with some of the threats that AI pose, S5 from Table 2 and WS1 from Table 3 are similar in nature, they both propose the development of a regulatory framework which could be expressed in the establishment of an international institution, which would take care of management and regulation of responsible and ethical development and use of AI technology. Keeping in mind that interviewees were not asked specifically how in their opinion AI threats could be

prevented or dealt with, and, keeping in mind that interviewees in general did not perceive current AI to be a threat to the extent Drexler and Vinge did few decades ago, it is not surprising that the number of suggestions for solutions in Table 3 is limited to only one.

In addition to above analysis I want to add a couple of other predictions made relatively recently by Stephen Hawking, British theoretical physicist, cosmologist and author of many books, and Ray Kurzweil, American computer scientist, inventor, futurist and author. Hawking has expressed his concern numerous times about the capacity of intelligent machines to eventually surpass humans and take over the world. His thoughts expressed in his recent interview to some extent resonate with Drexler's (1986) and Vinge's (1993) predictions (T4, T6 and T10) listed in Table 2:

> "Whereas the short-term impact of AI depends on who controls it, the long-term impact depends on whether it can be controlled at all." (Kolodny, 2014)

In the list of solutions presented in Table 2 and Table 3 we could add Hawking's suggestions, quoted by Ray Kurzweil (2001), of two possible ways to prevent AI surpassing humans in intelligence:

> "..Improve human intelligence with genetic engineering to "raise the complexity of … the DNA" and (ii) develop technologies that make possible "a direct connection between brain and computer, so that artificial brains contribute to human intelligence rather than opposing it." (Kurzweil, 2001)

However, Kurzweil is rather sceptical about Hawking's suggestion to enhance human intelligence through genetic engineering, due to it being a lengthy process at the time when intelligent machines are accelerating in the evolution of their intelligence. In Kurzweil's opinion direct connection between brain and computer, though, is a more realistic solution (Kurzweil, 2001). His latest prediction states that by 2030 there will exist "hybrid brain":

> "Our thinking then will be a hybrid of biological and non-biological thinking. We'll be able to extend our limitations and think in the cloud. We're going to put gateways to the cloud in our brains."(RT, 2015)

Both suggestions may appear extreme in their nature. However, they could signify the presence of urgency, which statements of such renowned scientists as Stephen Hawking and Ray Kurzweil may imply with regards to the inevitability of AI evolution, and uncertainty to where this will lead humanity.

It is clear that it is difficult to predict the consequences of technological evolution due to presence of uncertainty and due to the fact that "consequences emerge not from the static attributes of a fully formed technology, but from the complex co-production that simultaneously and continually moulds both technology and social context" (Guston and Sarewitz, 2002:98). However, by following the actors involved in the predictions which did come true and analysing the networks they are connected in, analysing how they influence each other, we can identify the core elements or the key-actors that played the

major role in those predictions coming true. In a similar manner, we can attempt to identify the core elements of the key-actors who would render indispensable in order for particular prediction to come true. Thus, by being aware of their presence, absence, or what would constitute to their presence or absence, we would be able to a certain degree know how realistic the certain prediction is and how far we are from it actually coming true.

5. Discussion

5.1 AI and humanity

Taking a reference point in Collingridge (1980) discussion, I want to bring attention to the question of objective. When talking about the development of a particular technology, it is important to ask ourselves: what is the objective with current technology? Is the objective to develop the technology which will be seen as a technological breakthrough and will move our human capabilities to a new level? An example of such technologies could be a moon landing program, or Manhattan Project, which resulted in technological success, of course, the success of these technologies on a purely human level could be questionable. Or is it a human-oriented objective, such as the case with Green Revolution, which, while being a success technologically, could not be seen as a successful on the human level (Collingridge, 1980:15). Thinking from the perspective of objective, we can take a new look at technology under question – artificial intelligence. If our objective is to bring humanity on the verge of technological breakthrough of unprecedented levels in comparison to any other technological breakthroughs in human history, then we are on the right path towards achieving this objective, for instance considering achievements of IBM with Watson. However, choosing this objective, we leave out any questions which may regard the human level. The question could arise though, how responsible it is to take this, purely technological, objective upon ourselves. However, if we choose human objective, for instance, with help of AI improving quality of life and contributing to longevity, sustainability and thus betterment of the ecosystem of our planet, we should also ask a whole set of other questions regarding ethics, responsibility, privacy, security, justice, autonomy, wellbeing, as well as many other questions. From the interviews we can see that Watson is promoted as the technology that has the capacity to improve our lives, through different means, such as contributing to better informed decision making for human experts, contributing to discoveries in medical and other fields, helping us to cope with big data and make sense out of it as well as using it as an experience. Thus, when assessing Watson, the above questions (ethics, responsibility, privacy, security, justice, autonomy, wellbeing) have a key role and should not be neglected, on the contrary, they should be evaluated, elaborated and asked as soon as possible, before it is too late.

We can refer to the case of Green Revolution – efforts to provide food to poor population in developing countries through the introduction of high yielding varieties of wheat, maize and rice into their agriculture, as possible example. I shall attempt to draw parallel between Green Revolution technology and AI technology. Looking at Green Revolution, it is clear that while it was a technological success, the main objecting – feeding poor, could

not be reached “without changes in the social conditions which determine how the benefits from this are divided” (Collingridge, 1980:15). So, the fact that human level objective was met purely from a technological perspective resulted in failure. Simple assumptions, such as more food, less hungry poor people, were simply insufficient, while such questions like how exactly this food should guarantee less hungry poor people were not asked. Insufficient attention to practical questions and, as a result, not taking practical measures how to achieve the objective resulted in failure of achieving it.

Looking at AI technology with regards to its availability worldwide and what the lack of access to it would mean to developing countries on the global scale, or, if looking on a micro level, what the lack of access to it would mean to poor or middle classes in the developed countries (X2, 2015), we can see that failing to ask important questions regarding this may result in serious issues on a global scale. AI may have the capacity to widen the gap between social classes drastically, or, it may even contribute to the development of two classes, poor and those with access to AI. The possibility of AI impacting *status quo* and totally destabilize it were discussed by Drexler back in 1986, where he looks into what possible supremacy in AI may entail for military balance worldwide:

> "...the coming breakthroughs will be too abrupt and destabilizing for the old balance to continue. In the past, a country could suffer a technological lag of several years and yet maintain a rough military balance. With swift replicators or advanced AI, though, a lag of a single day could be fatal." (Drexler, 1986:166)

In sum, it is possible to state that the lack of understanding how technology interacts with society may lead to failure of reaching human objective, even if it will prove to be technological success. Thus, “competent social scientists should work hand in hand with natural scientists, so that problems may be solved as they arise, and so that many of them may not arise in the first instance.” (Steelman, 1947 in Guston and Sarewitz, 2002:95)

For that reason I want to stress the importance of real-time TA which should be performed in close collaboration and co-production between social scientists, natural sciences and potential stakeholders (Guston and Sarewitz, 2002:101)

5.2 Human expertise vs. AI expertise

Keeping in mind that human cognition processes operate through the use of dual-processing, we can try to apply the same rationale to AI. From the previous sections of this thesis, we can state that AI possess *rational system* (*System 2*), it follows logic and analytical processes, however, comparing to humans using rational system, AI is not restricted by limitations of *working memory*, moreover, it does not possess any other types of memory, but rather long-term working memory, to which it has constant and effortless, instantaneous access through access to various databases. Let’s imagine for a second what would that mean for a human to be able to have instantaneous and constant access to our long-term memory and be able to use it as our *working memory* – would that be a blessing or a curse, and would it contribute to enlightenment or to chaos? Hard

to tell when thinking about humans, considering our emotions and feeling. But when it comes to AI, such “powers” might set its performance above our expectations.

Continuing on the same line, compared to humans, whose *rational system* is defined by slow and effortful cognitive processes, AI will be able to perform the same cognitive processes in an instance, which will also contribute to shorter time necessary for it to achieve expertise, presuming it would follow the human skill acquisition rationale.

When talking about *an experiential system*, we know now that it takes no time to access it for experts, once pattern recognition is enacted as a natural constituent of expertise, but we also know that it takes time for us to learn the skill to the point it is “transferred” from *rational* to *experiential system* in our cognition. Considering that the AI has constant access to long-term working memory, the time necessary for achieving an expertise in a particular area would be drastically shorter in AI.

What set us apart from AI before was our ability to learn and recognize patterns. AI was delimited by its rule-based, preprogramed behavior, which was not defined by any progression in performance through the time. However, Watson has proved that learning and pattern recognition is no longer the human prerogative. New cognitive systems are able to learn from experience just as we humans do, and, learning from experience as well as pattern recognition, which they are able to do as well, is the road to expertise. Intuition is an important signifier and constituent of expertise, and with pattern recognition and heuristics AI has necessary prerequisites for intuition to develop.

Language is the key to *rational system*, and rational system leads to *the experiential system* through skill acquisition in expertise. Now that AI has mastered natural language, it has presumably gained the access to the world of human expertise, only without its delimitations.

At this point it is safe to state that AI has all necessary constituents for an AI expertise to develop. However, it is important to keep in mind that AI expertise will not be the same as human expertise, nether AI intuition will be the same as human intuition, keeping in mind the speed with which cognitive systems will be able to operate, the unlimited access to memory banks – databases, as well as constant access to new information, and other elements of AI superiority might set AI expertise at a level, compared to which human expertise will become inferior. Assuming that AI will be humanities tool, this might sound like an exciting development, however, this leads to a question of whether human expert with inferior expertise and being prone to human factor erroneous behavior will be morally and ethically in a position to have the right of the first opinion. Of course, who is to say that AI won’t make mistakes, and it probably will, at least initially, but with speed in which it will be able to reach the level of expertise and ability to maintain its expertise by constantly learning in speed which is impossible to humans (even the smartest humans need to sleep while AI never sleeps!), learning from other AI experts and being constantly connected to heterogeneous networks of knowledge production and sharing, AI will have all necessary tools to make its errors marginal.

However, on a final note, I want to draw attention to the fact that human expertise possesses one important constituent which current AI does not possess – emotions. The importance of emotions in intuition, which is a key-component of expertise, is illustrated in following quote from Gobet and Chassy (2008):

> "..the key role of emotion in intuition has recently been buttressed by investigation with neurological patients showing how the lack of emotion negatively affects intuitive decision-making." (Gobet and Chassy, 2008:130)

Since AI does not possess the ability to have emotions, one can wonder how that would reflect on its expertise development and the AI expertise itself, what shape an AI intuition would take without emotions. Would it be impaired compared to human intuition? Or, would it require for a new definition of AI intuition and expertise, and how would this reflect on human expertise? Another set of questions comes in mind. What ethical implications AI expertise would have and what would an AI expertise be guided by? It opens a need for a new discussion and here is where suggestions made by Drexler (1986) and some of the interviewees in this research could come in handy and should be asserted with proper attention in the dawning of the new area of expertise – an AI expertise.

6. Conclusion

My hypothesis is that AI has all the necessary prerequisites for expertise development and through the use of multiple theories and empirical data interplay I believe I have managed to prove it. My secondary hypothesis is that AI expertise will challenge human expertise.

> *X1: "But already now, when we are, you know, when we look at the precision of the Watson technology in, you know, in making, or in, yeah, in making diagnosis for certain sicknesses or illnesses, and Watson is more precise today than the best specialists." (X1, 2015)*

Following X1 comment, I want to bring to the attention the idea about exponential improvement in computer technology, commonly being referred to as Moor's law[5]. It is possible to assume that this law also applies to artificial intelligence's evolution, due to AI being dependant on hardware and software. From this follows that eventually human expertise might be surpassed by AI expertise and the question would then be whether it would be ethical or responsible to let humans to maintain the superior status of expertise while not being able to live up to it.

How long would it take before the roles of these two experts trade places - an AI would no longer be just an adviser to human experts, but a decision making expert? Where would this leave human experts? Would human expertise be a thing of the past in parallel to some routine knowledge jobs disappearing with the coming of digitalisation, as well as many physical labour jobs disappearing with the coming of industrialisation? This shows

[5] "In a narrow sense, "Moor's law" refers to the observation that the number of transistors on integrated circuit have for several decades doubled approximately every two years. However, the term is often used to refer to the more general observation that many performance metric in computing technology have followed a similarly fast exponential trend." (Bostrom, 2014:27)

the importance of continuing analysis, dissection and discourse among all the stakeholders (scientific community, industry, public and others) about the notion of expertise and what makes one expert, whether it is human or AI.

For that purpose I suggest a necessity for a thorough study of AI expertise with practical study of AI application in expertise-driven fields through a field study or case study with one of the aims, among many others, being to determine how AI expertise is formed.

Despite acknowledging the necessity for a further study, I dare to conclude that new cognitive systems, such as Watson, will change the world of expertise and it is up to us, humans, at this stage to determine how this new world will be.

In addition to my previously established direct links between AI expertise formation and its ability to understand natural language, which leads to cognition, pattern recognition, ability to learn from experience, I want to use words of Ray Kurzweil (2001) in support to my conclusion that current AI expertise is in the position to evolve in a very near future through practical application to the extent that it might surpass human expertise and redefine the notion of expertise in general:

> “Once our machines can master human powers of pattern recognition and cognition, they will be in a position to combine these human talents with inherent advantages that machines already possess: speed (contemporary electronic circuits are already 100 million times faster than the electrochemical circuits in our interneuronal connections), accuracy (a computer can remember billions of facts accurately, whereas we’re hard pressed to remember a handful of phone numbers), and, most importantly, the ability to instantly share knowledge.” (Kurzweil, 2001)

I want to conclude my thesis with two important quotes. The first quote is by Stephen Hawking, which I believe is the best advice we all can and should follow with regards to AI:

> "All of us should ask ourselves what we can do now to improve the chances of reaping the benefits and avoiding the risks."(Kolodny, 2014)

The second quote is by Ray Kurzweil, which he said with regards to AI and technology in general, and which places responsibility on humans with regards to making sure to develop means for responsible and ethical development and use of AI:

> “As I wrote starting 20 years ago, technology is a double-edged sword. Fire kept us warm and cooked our food, but also burnt down our houses. Every technology has had its promise and peril.” (RT, 2015)

Bibliography

1. Birkbak, A. (2013), "Why all anthropology should be called techno-anthropology: On the consequences of a pragmatist understanding of technology", In Børsen & Botin (eds.): *What is Techno-Anthropology?* Aalborg: Aalborg University Press, pp 117-134

2. Bostrom, N. (2014), "Paths to Superintelligence", In *Superintelligence: Paths, Dangers, Strategies*, United Kingdom: Oxford University Press, pp 22-51

3. Carroll, J. M. (2000), "Five reasons for scenario-based design", In *Interacting with Computers* 13 (2000), pp 43-60

4. Cioffi, J. (1997), "Heuristics, servants to intuition, in clinical decision-making", In *Journal of Advanced Nursing* (26) pp 2013-208

5. Collingridge, D. (1980) "The dilemma of control", In *The Social Control of Technology*, Milton Keynes, UK: Open University Press, pp 13-22

6. Drexler (1987), "Thinking Machines", In *Engines of Creation: The Coming Era of Nanotechnology*, p.p. 55-70, accessed online http://xaonon.dyndns.org/misc/engines_of_creation.pdf

7. Drexler (1987), "Engines of Destruction", In *Engines of Creation: The Coming Era of Nanotechnology*, pp 145-161, accessed online http://xaonon.dyndns.org/misc/engines_of_creation.pdf

8. Drexler (1987), "Strategies and Survival", In *Engines of Creation: The Coming Era of Nanotechnology*, pp 162-171, accessed online http://xaonon.dyndns.org/misc/engines_of_creation.pdf

9. Dreyfus, H. L. and Dreyfus, S. E. (1988), "Preface", In *Mind over Machine: The Power of Human Intuition and Expertise in the Era of Computer*, New York: The Free Press, pp xv-xxi

10. Dreyfus, H. L. and Dreyfus, S. E. (1988), "Prologue: The Heart Has Its Reasons That Reason Does Not Know", In *Mind over Machine: The Power of Human Intuition and Expertise in the Era of Computer*, New York: The Free Press, pp 1-15

11. Dreyfus, H. L. and Dreyfus, S. E. (1988), "Five Steps from Novice to Expert", In *Mind over Machine: The Power of Human Intuition and Expertise in the Era of Computer*, New York: The Free Press, pp 16-51

12. DTU (2014), “DTU to teach the artificial intelligence 'Watson'”, Available at: http://www.dtu.dk/english/News/2014/11/DTU-to-teach-the-artificial-intelligence-Watson [accessed online 25.05.2015.]

13. Epstein, S. (2003), “Cognitive-Experiential Self-Theory of Personality”, In Millon, T. and Lerner, M.J. (eds.) *Personality and Social Psychology* 5, Handbook of Psychology, Hoboken, New Jersey: John Wiley & Sons, Inc, pp 159-184

14. Evans, J. St. B. T. (2008), “Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition”, In *Annual Review Psychology* 2008. (59) pp 255-78

15. Flick, U. (2007), “Triangulation Revisited: Strategy of Validation or Alternative?”, In *Journal for the Theory of Social Behavior*, Volume 22, Issue 2, Article first published online: 1 JUL 2007

16. Frankish, K. and Ramsey, W.M. (2014), “Introduction”, In *The Cambridge Handbook of Artificial Intelligence*, United Kingdom: Cambridge University Press, pp 1-11

17. Geertz C. (1973), “Deep play: Notes on the Balinese Cockfight”, In McGee & Warms *Anthropological Theory*, pp 553-574

18. Gobet, F. and Chassy, P. (2008), “Towards an alternative to Benner’s theory of expert intuition in nursing: A discussion paper”, In *International Journal of Nursing Studies* (45), pp 129-139

19. Guston, D. H. and Sarewitz, D. (2005) “Real-time Technology Assessment”, In *Technology in Society* 2002(24), pp 93-109

20. IMS (2015), “Passage retrieval”, Information Management Systems, Available at: http://ims.dei.unipd.it/websites/cms/past-research-subjects/passage-retrieval.html [accessed on 13.05.2015]

21. Kelly, J. and Hamm, S. (2013), “Building learning systems”, In *Smart Machines: IBM’s Watson and the Era of Cognitive Computing*, United States of America: Columbia University Press, pp 23-42

22. Kolodny, C. (2014), “Stephen Hawking Is Terrified Of Artificial Intelligence”, The Huffington Post, 5May. Available at: http://www.huffingtonpost.com/2014/05/05/stephen-hawking-artificial-intelligence_n_5267481.html [accessed 05.06.2015]

23. Kurzweil, R. (2001), “Response to Stephen Hawking”, Kurzweil Accelerating Intelligence, Available at http://www.kurzweilai.net/response-to-stephen-hawking [accessed online 05.06.2015.]

24. Law, J. (2009a), Actor Network Theory and Material Semiotics, In *The New Blackwell Companion to Social Theory*, Blackwell Publishing, 2nd edition, pp 141-158

25. Law, J. (2009b), “The Greer-Bush test: On politics in STS”, version of December 23rd 2009 Department of Sociology, Lancaster University, Lancaster LA1 4YN, UK, pp 1-13 Can be downloaded from http://heterogeneities.net/publications/Law2009TheGreer-BushTest.pdf [accessed online 25.05.2015.]

26. Law, J. and Singleton, V. (2014), “ANT, Multiplicity and Policy”, pp 1-21. This version was published by heterogeneities.net on 21 February 2015 at http://www.heterogeneities.net/publications/LawSingleton2014ANTMultiplicityPolicy.pdf [accessed online 26.05.2015.]

27. Law, J., & Singleton, V. (2005), “Object Lessons”, In *Organization*,12(3), pp 331-355

28. Mol, A. (1999), “Ontological politics: A word and some questions”, In J. Law, & J. Hassard (Eds.), *Actor network theory and after*. Oxford: Blackwell Publishers, pp 74-89

29. Pinch, T. J. and W. E. Bijker (1987), “The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other”, In Bijker, W. E., T. P. Hughes, and T. J. Pinch (Eds.) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, Boston, MA: MIT Press, pp 17-50

30. Richards, L. and Morse, J. (2013), “From Method to Analysis”, in *Readme first for a user's guide to Qualitative Methods*, Sage, pp 185-211

31. RT (2015), “Google’s Ray Kurzweil says humans will have 'hybrid' cloud-powered brains by 2030”, RT, 4 Jun, Available at: http://rt.com/usa/265029-kurzweil-google-hybrid-brain/ [accessed 05.06.2015]

32. Simon, H.S. and Newell, A. (1958), “Heuristic Problem Solving: The Next Advance in Operation Research”, In *Operations Research*, vol.6, pp 1-10

33. Strauss, A. L., and Corbin, J. (1994), “Grounded theory methodology: An overview”, In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research, Newbury Park*, California: Sage, pp 273-285 (13)

34. Thurmond, A. V. (2001), “The point of triangulation”, In *Journal of Nursing Scholarship*, 33 (3), pp 253-258

35. Vinge, V. (1993), "The coming technological singularity: how to survive in post-human era", In Vision -21: *Interdisciplinary Science and Engineering in the Era of Cyberspace*, pp 11-22. NASA Conference Publication 10129. NASA Lewis Research Center

List of figures