PROCEEDINGS OF
PRAGMATIC
CONSTRUCTIVISM
journal homepage: www.ProPraCon.com

Human vs machine intelligence: How they differ and what this implies for our future society

Thomas Bolander

DTU COMPUTE, Department of Applied Mathematics and Computer Science
Technical University of Denmark
Richard Petersens Plads
2800 Kgs. Lyngby
tobo@dtu.dk

Abstract

To be able to predict the impact of artificial intelligence (AI) on the required human competences of the future, it is first and foremost necessary to get an overview of what AI at all is and how it differs from human intelligence. The main goal of this paper is to provide such an overview to readers who are not experts in the area. The focus of the paper is on the similarities and differences between human and machine intelligence, since understanding that is of essential importance to be able to predict which human tasks and jobs are likely to be automatised by AI - and what consequences it will have.

Keywords: Artificial intelligence (AI), connectionist AI, symbolic AI, explainability, trust, human competences.

1 Introduction

The last few years have seen an explosive increase in industrial-scale applications of artificial intelligence (AI), and an even higher increase in the expectations of the problems and tasks that AI will be able to solve in the near future. There is no doubt that AI will have a profound impact on many aspects of our lives, jobs, and society as a whole. However, it is much less clear what exactly the impact will be. Many human cognitive tasks can seemingly be automatised by AI, but we risk a loss in predictability and explainability when doing so. We can not yet communicate with AI systems the way we communicate with fellow humans, and AI systems cannot explain their own reasoning and behaviour the way humans can. Even though the goal of AI is to simulate aspects of human cognition, machine intelligence is still fundamentally different from human intelligence, and has a rather different set of strengths and weaknesses. Some tasks that are easy for humans to solve have turned out to be exceedingly dificult for machines, and vice versa.

Predicting the exact impact of AI on our future society is overwhelmingly difficult. Most predictions made about the future of AI in the last 60 years have turned to be wrong, independently of whether these predictions were made by laypersons or AI researchers (Armstrong & Sotala, 2015). This paper will not attempt to make any profound predictions or conjectures, but mainly focus on providing some essential insights into current AI techniques, and their strengths and weaknesses. Hopefully, this will then provide the reader with a clearer view of what AI is (and isn't), and which future perspectives of AI are the most likely. The paper will point to some of the challenges that AI methods are currently facing, in terms of robustness, explainability, and lack of human-level social and linguistic capabilities. These challenges of course at least give us an impression of the expected relative role of machines and humans in the near future.

2 What is artificial intelligence (AI)?

More than 60 years ago John McCarthy, the father of artificial intelligence (AI), defined the field as “the science and engineering of making intelligent machines, especially intelligent computer programs.” The complication of this definition is that we do not know exactly what intelligence is, and hence even less what it means for a computer program to be intelligent. In the 1950s and 1960s, AI was expected to develop very rapidly into computers and robots with human-level cognitive capabilities. This however did not happen, at least not yet.

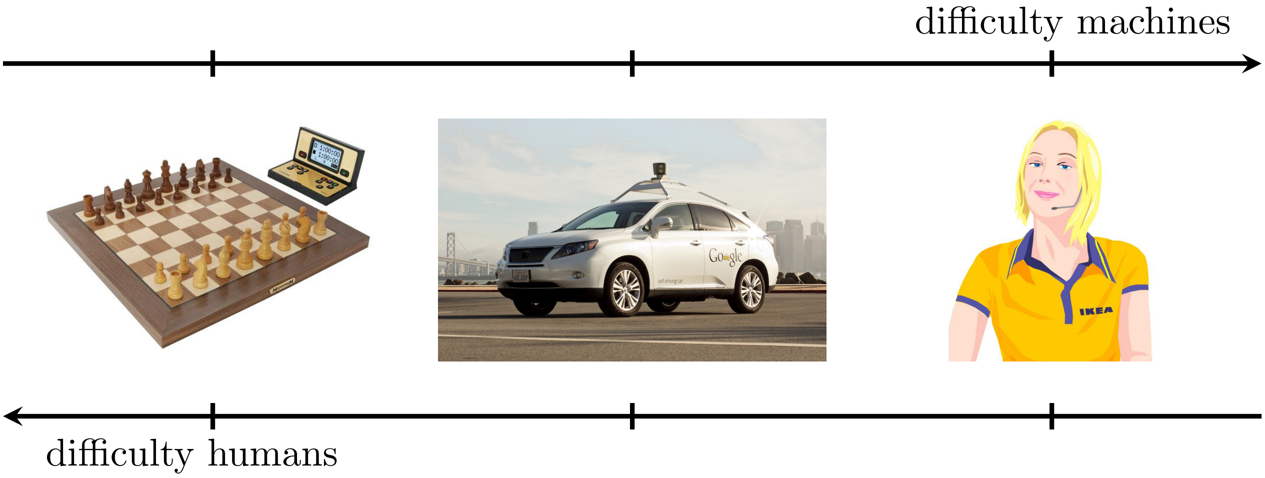
Lacking a precise definition of AI, we can still give an approximate characterisation. It is almost always about building machines – computers or robots – that can perform tasks that otherwise only humans have been able to, e.g. play chess, drive a car, do medical diagnosis, or engage in a dialogue. Furthermore, when such machines are built, AI researchers are almost always directly inspired by how humans solve the same tasks. It can be in terms of the machine directly trying to mimic some of the neurological processes of the human brain (see *Connectionist AI* in Section 7 below); or it can be via a more abstract model of human problem solving, e.g. an approximate model of the reasoning steps involved in a human deciding the next move in a game of chess (see *Symbolic AI* in Section 6 below).

AI today is a wide range of different techniques for simulating different aspects of human cognition. Computers can play chess, drive cars, recognise skin diseases and engage in dialogues, but all these applications are based on different techniques within AI, and require individual programming tailored for the specific application at hand. This makes current AI very different from human beings solving similar tasks. Human beings learn to master all of these different tasks during their lifetime without having to be preprogrammed specifically to solve them. One of the important emerging trends in AI is *Artificial General Intelligence* (*AGI*) (Gorttzel & Pennachin, 2007), where the goal is to make AI systems more human-like by giving them the ability to learn a range of different skills without having been specifically preprogrammed for them. However, the success of such systems is still rather limited.

3 Characteristics of current AI

AI systems tend to be tailored to specific types of applications, and often new types of applications requires new methods to be developed (or existing methods to be combined in a novel way). Developing a robust driverless car is not just about taking an existing AI system down from the shelf, plug it into the car and then let itself figure out how to drive. Given that AI systems need to be tailored to specific applications, the complexity of building such a system depends crucially on how well-defined and clearly delimited the problem to be solved is. As a rather robust rule of thumb, the more well-defined and clearly delimited a problem is, the easier it is to make AI that can solve it. With this in mind, it is not too surprising that already in 1997 it was possible to build a computer program, IBM Deep Blue, that was able to become world chess champion. Chess is extremely well-defined and clearly delimited: there is only a very few and strict rules to obey, and there is a very precisely formulated goal to achieve. For a human being, chess has an overwhelming combinatorial complexity in terms of possible move sequences, but modern computers are not easily overwhelmed by the need to consider an enormous number of options. Deep Blue could compute 200 million chess moves per second.

Figure 1: For many types of tasks, the axis of difficulty for machines is opposite the one for humans.



Modern computers are however much more easily overwhelmed by problems in which the rules or the goal – or both – are less clearly formulated and delimited. One such example is driverless cars. As in chess, there are also rules to obey in traffic, but there are many more rules than in chess, and they are much less formally specifiable. It is even more complicated for computers to successfully small talk with a human for a few minutes over a cup of coffee. Chatbots are computer systems for engaging in dialogue with humans, and the dialogue can either be in writing or through a voice interface. Building a chatbot that can engage successfully in small talk with humans is exceptionally difficult, as the rules of such dialogues are even much less clear than the rules in traffic. This is also why a lot of the chatbot technology that was in the early 2010s implemented on web pages to help customers, e.g. by IKEA, Scandinavian Airlines and Deutsche Post, has now been taken out of use. There seems to be a new wave of chatbots arising here on the edge to the 2020s, and they are probably better, but it does not change the fact that general natural language dialogue is everything but well-defined and clearly delimited, and hence extremely hard to bring to human level on a computer.

It is interesting to compare the difficulty for computers on the three tasks mentioned above – playing chess, driving cars, and chatting – with the difficulty for humans. For most humans, the relative difficulty of the three tasks is opposite the one for computers: It is easier for most humans to engage in small talk for a few minutes than it is to drive a car safely through downtown Rome on a Friday afternoon, which again is easier than becoming world chess champion. Figure 1 illustrates this. The fact that the two axes of difficulty are opposite one another illustrates that intelligence is not just one thing, and that different types of intelligence cannot always be compared along the same axis. Many people seem to have the view that computers are becoming more and more intelligent, and that it is just a matter of time before they become more intelligent than us humans. But that view assumes that we can directly compare human and machine intelligence on a single axis of intelligence. The figure illustrates that it might not be as simple as that. Humans and machines currently have very different strengths and weaknesses, and there is no simple way of comparing their "level of intelligence". Computers will forever be better than humans at board games with high combinatorial complexity, but no matter how explosive the development of AI is going to be, it is conceivable that we humans will forever be better natural language users (not the least since natural language was invented by us, and developed in a way that is to a large extend dependent on our culture, our brains and our bodies (Lakoff, 2006)).

4 Human-machine dualism

The difference between human and current level machine intelligence is so large that it is probably more relevant to talk about a duality. We humans have a very flexible intelligence, are good at abstract thinking and conceptualising the world. We are often good at solving problems that are not very clearly delimited and well-structured (but where the solutions do not have to be either). Conversely, machines are primarily good at clearly delimited and well-structured problems, but can then also provide solutions that are very precise and well-structured. They are much less competent at abstract thinking and conceptualising the world, though a lot of research is invested in developing AI systems that have these competences as well.

The human-machine duality can be illustrated by the case of the IBM Watson system that was originally developed for playing the game of Jeopardy (Ferrucci, 2012). Jeopardy is about answering questions concerning trivia knowledge, which is not a particularly well-defined and clearly delimited problem, but still much more well-defined than general dialogues (even when those dialogue are just small talk about much more down-to-earth subjects than considered in professional Jeopardy). In 2011, IBM Watson became (unoficial) world-champion in Jeopardy. It didn't do so by being better at understanding the questions, but by compensating somewhere else: The system had 200 million pages of text in memory and could process 1 million books per second. This is of course far beyond what any human can do. The point here is that humans are actually much better at understanding questions and finding answers from relatively small amounts of data (the small amounts we can keep in our brains), but computers can in some cases compensate by having access to enormous amounts of data and being able to process that data with exceptional speed. And in some cases, as with Watson, they can compensate so well that they actually outperform humans on certain tasks.

Roughly speaking, one can consider problem solving abilities to be a combination of 1) an ability to extract *information from data* (intuition, abstraction, conceptualisation), and 2) an ability to *process data quickly* (search). Humans are much better at 1 than 2, and for computers it is opposite. In cases where the primary task is to extract information from data, a deficiency in 1 can often be compensated by a sufficient corresponding increase in 2. This is exactly what we see with Watson. If the answer to a question is well hidden in a piece of text, Watson is not likely to find it. However, it compensates by having access to an enormous text library that it can browse through in seconds, and then it will probably find another source, where it is more trivial to extract the answer. When humans look for answers in texts, we can only read extremely slow compared to computers, but we have a much deeper understanding of what we read, and are very good at finding the deeper meanings and the well-hidden answers.

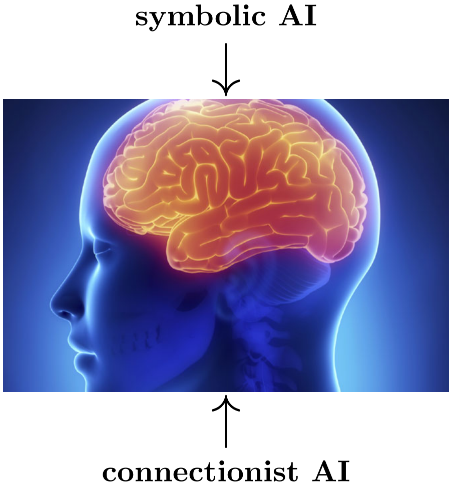
One of the important conclusions of the Watson example is that even when we succeed in constructing a computer program that achieve above human level on a certain task, it doesn't at all imply that it solves it in the same way as a human, and therefore we cannot use it to conclude anything about the relative intelligence between humans and machines. In contrast to Watson, humans can also answer questions that nobody has asked before, e.g. whether a crocodile can run a steeplechase (Levesque, 2014). In order to answer such a question, we need our human ability to create mental models of the content of the question, that is, to picture the poor crocodile with its extremely short legs trying to jump a high barrier. We use our rich existing models of the world to answer questions, whereas Watson simply tries to look the question up in its enormous library, and if the answer is not there because nobody considered that question before, it has no chance of answering.

5 Collaboration between humans and machines

Machines and humans will probably never be good at exactly the same things, so the ideal future perspective of AI is to ensure efficient collaboration between humans and AI systems, so we can each do what we are best at, and let the other support us for the things we are less good at. A nice example of this is the use of current IBM Watson technology to create an online chatbot for teaching assistance in the AI course at Georgia Tech (Goel & Polepeddi, 2016). The chatbot can answer simple questions, where the answer can easily be looked up in the existing course material. When a student asked whether they should be aiming for 1000 or 2000 words in an essay, the chatbot answered "There isn't a word limit, but we will grade on both depth and succinctness. It's important to explain your design in enough detail". The student then asked a follow-up question: "Can you please elaborate on *it's important to explain your design in enough detail*. What kind of design are you referring to?" That question was too hard for the chatbot, but the chatbot then passed the question on to a human who could elaborate. Many questions students ask are quite well-defined and easy to answer, and are the same year after year. Such questions can easily be answered by chatbots, and can save time that human teaching assistants would otherwise have to spend on answering the same question over and over again, year after year. New questions or questions where the answer cannot easily be looked up still have to be answered by humans.

In order to achieve efficient collaboration between machines and humans in general, it is required that both have a decent level of social intelligence, for instance that both have the ability to take the perspective of the other and to explain itself in a way that is comprehensible to the other. As with natural language comprehension, social intelligence is unfortunately one of the very hard problems in AI. For both of these cognitive abilities, one of the main challenges is that we still do not have a suficiently deep and precise understanding of how these cognitive abilities work in humans, and therefore we do not yet have any suficiently precise models that we can implement in machines.

Figure 2: Symbolic and connectionist AI have opposite approaches to simulating aspects of human cognition.



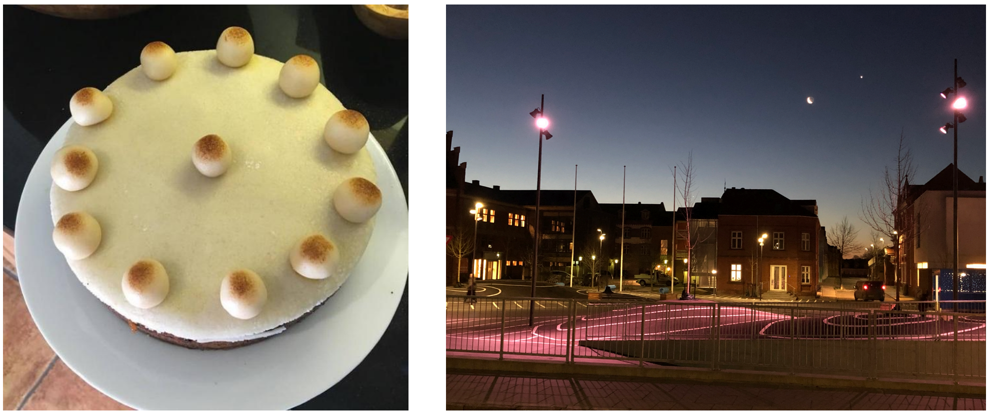
6 Symbolic AI

Since the 1960s, AI research has essentially been divided into two competing paradigms, the symbolic paradigm and the connectionist paradigm (Hoffmann, 1998). These two paradigms have completely opposite approaches to simulating aspects of human cognition. The symbolic paradigm follows a top-down approach by trying to directly simulate the highest levels of human cognition, our linguistic (symbolic), conscious reasoning. In this paradigm, one tries to build AI systems that have an explicitly represented language to reason about the world, and for instance use this to do logical inference or plan a sequence of actions to achieve a certain goal. AI within the symbolic paradigm is behind systems such as chess computers, Amazon warehouse robots and intelligent personal assistants such as Siri on iPhone and Google Now. The advantage of symbolic AI is that the systems constructed can be made robust, predictable and explainable. If a chess computer always makes a bad move in a certain situation, we can inspect the code and understand why, and hence improve it. The drawback of symbolic AI is, though, that systems within this paradigm tend to have strictly delimited abilities. They normally do not learn from their experience, and a chess computer can play only chess, not even tic-tac-toe (though the search methods on which it is build could be adapted to other games). Some of the central areas of AI within the symbolic paradigm are *problem-solving by searching*, *knowledge representation* and *automated planning*. Most of symbolic AI is based on discrete mathematics, that is, areas such as logic, combinatorics, graph theory and theory of algorithms.

7 Connectionist AI

Arguably, one of the main landmarks of human intelligence is our flexible intelligence and our ability to learn. We are not born with the ability to play tic-tac-toe or chess but learn it during our lifetime. We are also not born knowing the difference between a table and a chair, and the words used to name these objects, but learn it when we are small. If we want artificial intelligence systems to share these abilities with humans, we have to consider the area of machine learning within AI. Machine learning is a very broad term that covers any AI algorithm that does not have a static behaviour, but can learn from its experience. It could e.g. be algorithms with the ability to learn to distinguish objects in the physical world, with the ability to learn better strategies in chess, or with the ability to learn the rules of new games. Some of the techniques of machine learning belong to the symbolic paradigm, but the currently most prominent ones belong to the connectionist paradigm. The connectionist paradigm is essentially constituted by AI techniques based on (artificial) neural networks (ANNs). In artificial neural networks, one tries to simulate the atomic processes of the human brain: the functioning of the individual neurons and neuron connections. The connectionist approach is behind image recognition software, e.g. for recognising skin diseases or as used by Instagram to decide whether a picture contains unacceptable nudity and should hence be blocked. The advantage of the connectionist approach is that it is possible to construct systems that have a more flexible intelligence and can learn from experience. The neural network employed at Instagram has not been programmed with a model of what unacceptable nudity is, but has simply been trained on a very large set of pictures that were either labelled as "acceptable" or "unacceptable". Eventually, the system has itself learned to recognise the patterns of unacceptable content. It would never be possible to do the same with symbolic AI: There is no way to give a sufficiently precise linguistic or symbolic definition of what "unacceptable nudity" is. Similarly, it would be very hard to linguistically or symbolically define the difference between a chair and a table, but neural networks can be trained to relatively robustly make the distinction.

Figure 3: Two pictures that were blocked by the image recognition system of Instagram and claimed by the algorithm to contain unacceptable nudity or pornographic content.



The drawback of the connectionist approach is, however, that it can never be 100% predictable, error-free or explainable. The systems built according to this approach are based on statistical learning from experience. When you do statistical learning from experience, your ability to correctly categorise new objects, for instance recognise certain types of pictures, gradually improves but can never become 100% precise. An example of this is the two pictures in Figure 3. They don't seem to bear many similarities, and they seem to be fairly acceptable pictures from everyday situations. However, they were both blocked by the image recognition system of Instagram (in 2015 and 2019, respectively), and both were claimed by the system to contain unacceptable nudity. There is currently no way of knowing exactly why these pictures were labelled as unacceptable, as the neural networks train an implicit model with millions of neuron weights, and it is the combination of all these neuron weights that decide the classification the network makes. There is also no simple way of telling the neural network that easter simnel cakes like the one on the left in Figure 3 are not examples of nudity. The only way to try avoiding such misclassifications in the future is to provide the algorithm with more labelled pictures that it can train on – and then hope for the best. This is clearly very different than the situation in symbolic AI, where models are explicit rather than implicit, and can hence easily be understood and modified. Recently, there has been an increased interest in explainability also in the connectionist approach, e.g. image recognition systems that can point out the parts of a picture that make them suggest a given classification (Rebeiro et al., 2016). This can help us understand how pictures such as the two of Figure 3 end up being labelled as containing unacceptable nudity, but it does not in itself guarantee against misclassification, of course. There is a growing consensus in AI research that transparency and explainability are of central importance, so it is very likely that the coming years will see significant breakthroughs within these aspects, both in symbolic and connectionist AI.

8 Symbolic versus connectionist AI

Whereas symbolic AI, as mentioned above, is mainly based on discrete mathematics, the connectionist approach is mainly based on linear algebra and mathematical analysis. Hence, the paradigm distinction in AI is more or less matched with a corresponding paradigm distinction in the underlying mathematics used for the techniques of the two paradigms. As also mentioned, an essential difference between the paradigms is that symbolic AI is based on creating explicit (symbolic) models, whereas the connectionist approach is based on learning implicit models. This difference roughly corresponds to the difference between trying to predict a ballistic trajectory using the laws of mechanics and aerodynamics (explicit model) versus simply trying to learn to make such predictions from observing a high number of trajectories without necessarily creating any explicit model of the observed phenomena. When humans throw snowballs, there is no doubt that we use some kind of learned implicit model to predict where the snowball will land, which is consistent with the connectionist approach. However, when we play a game of chess or plan a dinner party, there is equally no doubt that we use explicit symbolic (linguistic) models to reason about our possible action sequences, which is consistent with the symbolic approach. Hence it seems that human problem solving combines implicit and explicit models, and that certain aspects of problem solving are closest to the connectionist approach, whereas others are closest to the symbolic approach. Probably for this reason, the last few years have seen a high increase in attempts at constructing AI systems that combine the symbolic and the connectionist approaches, with some of the notable examples being Google DeepMind building a system that taught itself to play a wide range of old Atari arcade games (Mnih et al., 2015) as well as a system achieving world-class level in the board game Go (Silver et al., 2016). Connectionist AI is mainly about simulating aspects of our perception and fast heuristic assessments (intuition), whereas symbolic AI is mainly about simulating aspects of our higher cognition (conscious reflection and reasoning). Fully autonomous AI systems like driverless cars or general-purpose household robots of course need both.

9 Trust and explainability in AI

In AI, it seems to be hard to get what we could otherwise reasonably expect. We would like AI systems to be robust, predictable and explainable, which would push us towards symbolic AI. However, we would also like AI systems to be flexible and learn from experience, which would push us towards connectionist AI instead. There seems to be a fundamental and unavoidable trade-off involved: the more intelligent, flexible and easily trained we want a system to be, the less we have control over the system and the less we can guarantee it to behave in the intended way. This implies that not all demands for computer software and robots can be met by simply turning up the level of intelligence and flexibility of those systems. For many types of system, for instance database systems and e-voting systems, we still want to be able to prove that they have and will always maintain the intended behavioural properties.

If an AI system is not 100% error-free and predictable, how can we trust its decisions? We have to look at what trust in such systems is even supposed to mean. Do we trust the decisions of an AI system when 1) it never makes mistakes?; or, when 2) it almost never makes mistakes?; or, when 3) it most often doesn't make mistakes, but when it does, it has an acceptable and explainable reason for doing so? If we wish to employ AI systems that use the connectionist approach, we can never have 1. And we need such connectionist methods for perception tasks, for instance in driverless cars that have to recognise objects and other road users in their environment. If we cannot have flawless AI, we of course want it to make as few errors as possible, but that is not all. Suppose you can choose between two general practitioners: one that almost never makes mistakes, but when she does, she cannot at all explain why she did it, and cannot give any reasonable guarantee that she won't do it again. The other general practitioner makes a few more mistakes, but when she does, she can explain in a comprehensible way why it happened, and in realising her mistake, normally she will learn from it and not repeat it. If these general practitioners were AI systems, the first would be of type 2 above, and the second would be of type 3. Most people would probably prefer general practitioners – and AI systems – of type 3 over type 2. Trust is not only about statistical precision, but also about whether the person or system you engage with can explain itself and thereby regain your trust after having made an error. Currently, AI systems based on the connectionist approach are of type 2, as they cannot explain their decisions and since, as earlier mentioned, we cannot in general inspect and understand their learned implicit models.

This creates a big challenge in AI that many researchers are today occupied with. As they write in the *One Hundred Year Study on AI* by Stanford University: "The [AI systems] should be designed to enable people to understand [them] successfully, participate in their use, and build their trust" and later "AI technologies already pervade our lives. As they become a central force in society, the field is shifting from simply building systems that are intelligent to building intelligent systems that are human-aware and trustworthy" (Stone et al., 2016). The problem is then just how to develop such AI systems that are transparent, explainable and trustworthy? A key challenge is that connectionist AI is naturally opaque. Symbolic AI, on the other hand, is naturally transparent, but cannot in itself solve all the problems in AI we wish to solve. Therefore, the best current bet in order to achieve AI systems that are more transparent and explainable (and hence more trustworthy) is to combine symbolic and connectionist methods. We need to give them the ability to learn from experience, including statistically based learning, but we also need to equip them with an explicit, symbolic language that they can use to explain their models and decisions. This is what humans can do, but in this ability we are currently unique in the universe.

An often-stated mantra in the big data revolution is that we only need to focus on the "what" and not the "why" (Mayer-Schönberger & Kenneth, 2013). The point made is that if a company wants to use algorithms for instance to predict what items customers are likely to buy next month, the company doesn't need to know *why*, but just *that* the customers are going to buy those items. However, that kind of mantra is heavily challenged when such algorithms are used e.g. to decide whether bank customers can get a loan or not, or to decide whether an accused person should go to jail or not. Bank customers are in their good right to expect to receive an explanation of the decision made, but this is what many of the algorithms currently used can't give. Even if we don't expect an explanation, the algorithms based on statistical learning has another challenge. Suppose a statistical learning algorithm for credit scoring has been trained on historical customer data. If, historically, all bad customers shared the same last digit of their phone number, and no good customers had that number, the algorithm would most likely give a very low scoring to any new customer with that last digit. The point is that the algorithm can only look for correlations in the data, but have no way of assessing whether those correlations signify causal relationships. Most humans would know that there can't reasonably be a causal relationship between phone numbers and whether you're a good or bad bank customer, so even if we observed that kind of correlation, we would not let it affect our decisions. But algorithms don't have rich models of the world that tell them which correlations are likely to signify causal relationships, and hence they act on correlations alone, no matter whether there is an underlying causal relationship or not. Finally, if the bank at some point decides to change some of their principles for credit scoring, then the existing algorithm and all the data it was trained on will have become useless. You cannot tell the algorithm to adjust its principles, because there is no explicit model to adjust. In that case, the bank would have to start all over, and would first have to manually create a new data set of customer scorings to train the new algorithm.

10 The impact of AI on the human competences of the future

Even though current level AI still has many challenges, there is no doubt that we will see more and more tasks being successfully automated by computers and robots in the future. Hence it makes sense to consider what kind of human competences might still be needed in the future. We will need the following competences:

1. Competences in seeing the potential and selecting the tasks to be automatised by AI, and, equally important, deselecting the tasks that cannot reasonably be automatised.
2. Competences in implementing AI techniques for the tasks selected under 1.
3. Competences to operate and collaborate with AI systems.
4. Competences in areas that cannot be automatised.

No doubt, most of us will mainly be affected by 3 and 4. To operate and collaborate with AI systems, one does not necessarily need a deep understanding of how the systems work, but given the challenges of AI stated above, it is probably quite important that users understand the scope and limitations of such systems. If a general practitioner uses a medical diagnosis system, she has to be aware that the system cannot be expected to be flawless, and can therefore not be blindly trusted. Concerning 4, we already noted that there are certain aspects of human cognition that have so far proven exceptionally hard to simulate on a computer, most notably linguistic and social intelligence. Since almost all humans have jobs that require both linguistic and social intelligence (for communication and collaboration), not many jobs can be expected to be replace one-to-one by AI in the foreseeable future. This doesn't imply that AI cannot lead to unemployment in certain sectors, it just means that many of the tasks most of us carry out today still have to be carried out by humans in the future.

As stated above, the tasks that are most easy to automatise are the most well-defined and clearly delimited ones. Those also tend to be the most routine and repetitive among our tasks. So, when trying to predict what human competences are needed in the future, we need to think about which of our tasks are least well-defined, least clearly delimited and least repetitive. Since linguistic and social intelligence are very hard problems for AI, these might become the most important human competences of the future, even for employees in technical areas like engineering. Indeed, an Australian study of the impact of automatisation on the required competences of skilled and technical workers concluded that the highest rated competences were communication, social empathy and the ability to critically evaluate digital data sources (Reeson et al, 2016). In a case study on automation by the Danish SIRI Commission, a main conclusion was that to utilise the full potential of automation, the most important thing is to make the employees feel safe, not fearing the technology and not fearing their jobs (Shapiro, 2018). This is not about any specific skill set that the employees should have, but rather about their attitudes towards the AI systems. It also proves to illustrate the importance of making AI systems explainable, human-aware and trustworthy, since otherwise there is bound to be significant resistance against the use of such systems.

In addition to linguistic and social intelligence, currently humans are much better at adopting to changing norms and principles, cf. the example about credit scoring given above. And when it comes to creatively suggesting changes to norms and principles, we are even better. Most algorithms will at best learn and retain existing norms and principles. So, when it comes to developing our culture and decide how we want our future society, this is something that should still be designed and decided by humans.

References

Armstrong, S. and Sotala, K. (2015). How we're predicting AI - or failing to. In *Beyond artificial intelligence*, pages 11-29. Springer.

Ferrucci, D. A. (2012). Introduction to "This is Watson". *IBM Journal of Research and Development*, 56(3.4):1-15.

Goel, A. K. and Polepeddi, L. (2016). *Jill Watson: A virtual teaching assistant for online education*. Technical report, Georgia Institute of Technology.

Goertzel, B. and Pennachin, C. (2007). *Artificial general intelligence*, volume 2. Springer, 2007.

Hofimann, A. G (1998). *Paradigms of Artificial Intelligence: A methodological and computational analysis*. Springer.

Levesque, H. J. (2014). On our best behaviour. *Artificial Intelligence*, 212:27-35.

Lakofi, G. (2008). *Women, fire, and dangerous things*. University of Chicago press.

Mayer-Schonberger, V. and Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifiin Harcourt.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. and Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529-533.

Reeson, A., Mason, C. Sanderson, T., Bratanova, A. and Hajkowicz, S. (2016). The VET era: equipping australias workforce for the future digital economy. Report for TAFE Queensland. Brisbane: CSIRO.

Ribeiro, M. T., Singh, S. and Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1135-1144. ACM.

Shapiro, H. (2018). *Digitalisering, job og kompetencer*. Technical report, SIRI-kommissionen, Ingeniørforeningen i Danmark.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M. et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484-489, 2016.

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S. et al. (2016). Artificial intelligence and life in 2030. *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*.