Essay discussion

Mårten Nilsson, Patrik Barkman, Axel Demborg, Elon Såndberg

October 2016

Patrik Barkman
barkm@kth.se
Artificial Intelligence DD2380

Artificial Intelligence: A New Revolution

Society has through the course of history undergone many of its major changes due to specific inventions and development of new technology. Examples of such transitioning events are the agricultural and industrial revolution. As known, the industrial revolution resulted in the majority of the workforce leaving agriculture and turning to industrial employment in the city. Ever since, the development of technology has to great extent been conducted in order to relieve humans of physical labor.

The notion of *automation* – the progress of replacing human labor with machines – is not something new. In many parts of society there are machines that are specialized to perform specific tasks extremely well. These machines can be anything from printing presses to industrial robots. However, due to the progress in artificial intelligence the concept of automation can be taken to yet another level. Instead of developing machines with mechanical muscles it is possible to construct intelligent machines with cognitive and reasoning skills comparable to that of humans. Therefore, humanity is possibly now standing before yet another revolution.

Even though such machines seem to belong to the distant future, intelligent systems are being implemented and currently constitute important parts in the industry. At the moment the most prominent example might be self-driving cars. However, there are more examples that are less obvious to the general public such as automated transportations in warehouses and mining facilities [1]. Yet another example is the logistic system in the hospital New Karolinska Solna, which is based on automated guided vehicles [2].

In some sense these intelligent systems are still specialized since they cannot compete with humans when it comes to solving a wide range of diverse tasks. On the other hand, the important feature of these systems is their intelligent behavior, rather than providing mechanical strength. Therefore, an important aspect of future artificial intelligence research is to build new general-purpose systems that can perform well in many situations. One approach to this difficult problem is to construct systems that learn and adapt to their environment.

It is evident that intelligent systems used for automation are very applicable in society. Not only may these system perform better than humans in many tasks they are also very cost effective. This highly motivates companies to replace their human workers with automated systems. The development of intelligent systems thus leads society into a transition phase similar to that of the industrial revolution. It is important to keep in mind that this new revolution of artificial intelligence not only affects the low skill jobs but professions ranging all of society. The computers that constitute the core of the intelligent systems are tireless, efficient and cheap. Consequently they are able to process huge amounts of data and theoretically enabling them to surpass human reasoning in domains that require complex decision-making. All things considered, it

is possible that the development of artificial intelligence may cause a cascade of unemployment of unprecedented magnitude.

The major challenges associated with the development of artificial intelligence may be related to the transition to a society where humanity has found a place alongside advanced technology. During this phase society may have to accept that some people will inevitably become unemployable and that a system that cares for those people must be constructed. It is difficult to hypothesize possible solutions, but what may be most important is to simply spark a debate in order for society to prepare as much as possible for the coming revolution.

573 words.

References

[1] Caterpillar Autonomous Operations. http://www.caterpillar.com/en/company/innovation/customer-solutions/autonomous-operations.html

[2] Automated Logistic System at New Karolinska Solna. http://www.svt.se/nyheter/lokalt/stockholm/har-ar-roboten-som-gor-jobbet-pa-nks

AI Essay

Axel Demborg demborg@kth

October 4, 2016

Artificial intelligence is a technology that makes a lot of promises, one of the greater ones being the end of all human labor. In a world where machines can do all the tasks we perform on a day to day basis better and without complaining then why should we, humanity could flourish in a time when we don't have to work for our survival but instead only focus on what captures our interests.

This utopia might lie far away in the future but at the rate that technology develops I believe that it is inevitable that it will be technically possible to reach this world some day (a few hundred years maybe). How we as a society will survive the winding path that might take us to utopia though is a whole other question and that is what I would like to discuss.

If we start with the premise that the machines actually will be able to take our jobs it is only reasonable to conclude that it won't happen all at once. Looking at the state of self driving cars today it is likely that most of the transportation sector will be in the hands of machines quite soon but something like preschool teachers where so much of the "human touch" is needed will keep their jobs for way longer. This is to me the heart of the problem, how will society take care of the people once employed by the transportation sector when most of the population still have their jobs?

Well, some might argue that they just have to find new jobs, that technology not only takes jobs but also creates them in just about the same rate, is that not how the world has worked until now? Well, yes but this is different, all previous advancements have replaced human physical labor freeing our minds to do more involved *mind work*. But AI is replacing our mind work, it is leaving no island of work that can not be automated. Hence getting new jobs is unlikely to be a possibility.

The situation we will find ourselves in is hence that AI will start making large swats of the population unemployable, not because they are lazy or ungifted but simply because society doesn't need the entire population in work. The only possible option I can see is then to give the entire population

some sort of citizen salary, state funded economic support to make working economically optional. This will be a costly solution but the exceeding abundance produced by these thinking machines will be able to cover it.

One further issue with a non working population might be one of identity, we define so much of our identity on what our profession is, what will we even do if we don't work? Most might want travel the world, but what if that gets old after a couple years. What do you do then to occupy your time? Ironically I think that this might be one of the bigger problems with this future, we all want more free time, but when we get it what do we do then? Is it not true that unemployment rates correlate positively to criminality? This might mostly be a simple way to earn money for those who have none and hence the citizen salary should remove much of this crime but maybe it is also something "meaningful" to do for a restless population and if this is the case we might have a lot of problems on our hands.

Word count: 586

The role and risks of artificial intelligence in our future society

Mårten Nilsson
marten3@kth.se

October 2016

There are currently many opinions of what artificial intelligence (AI) is and is not. Norvig and Russel [1] encapsulated eight definitions of AI into four categories concerning different aspects of animal intelligence. The definition used in this essay is the one of these eight that corresponds best with my intuition of AI, stated by PH Winston as "The study of the computations that make it possible to perceive, reason and act." [2]. For a system to be considered intelligent it has to perceive, reason and act to achieve a specified goal.

An intrinsic risk of goal oriented reasoning agents is unexpected side effects. The plan that the agent creates to achieve its goal is not always the obvious plan for a human. The level of damage caused by unexpected side effects are only limited by the resources of the agent. An AI playing chess can at worst make a move considered counterintuitive by a human, but an AI driving a truck might decide to save fuel by colliding with a pedestrian on the road instead of stopping the truck. A common approach to prevent unexpected behaviour of complex systems is to perform comprehensive tests. Most non intelligent systems yield the same behaviour in similar situations and the possible unexpected behaviour that may arise are small malfunctions that cause the system to fail. For an intelligent complex system testing the decision making process is an elaborate task, since just a small change in input may cause the system to reconfigure it's entire plan.

Another risk with AI is that it may surpass humans at too many tasks too fast. Since the industrial revolution automation of labour have been constantly increasing. This has been a great part of the development of our society to the point it is today. Though automating jobs ultimately frees people of labour it also forces individuals to readjust. The society we live in today relies on individuals to make a living for themselves which only works as long as everyone has the ability to contribute. Historically there have always existed some unqualified labour for people to fall back to when they need to readjust their working situation. In our current globalized society a breakthrough in artificial intelligence could possibly eliminate the majority of all unqualified labour in a matter of decades, possibly less. The need for everyone to work would disappear and instead it would only be necessary to have a small group of experts to program AI:s to perform all labour for us, including intelligent labour. Ultimately this would give people more time to enjoy life, create art and science but if the society is not ready for this change the consequences could be devastating.

I believe that many of our jobs today will be automated in the next decades thanks to recent progress in artificial intelligence and machine learning. Though there are some risks associated with this progress, I believe that as long as we are aware and prepared for the risks and always put our safety first this will change society for the better. People will have more time to understand their surroundings and more powerful tools available to deal with global challenges.

References

[1] S Russel and P Norvig. *"Artificial Intelligence–A Modern Approach"*. Pearson Education, 3 edition, 2014.

[2] Patrick Henry Winston. *Artificial Intelligence*. Addison-Wesley, 3 edition, 1992.

word count: 535.

Risks of automation and AI in the near future.
Author Elon Såndberg.

I believe that in a not too distant future all work can be automated and human labor will not only be unnecessary it will be inefficient and dangerous compared to a perfected machine. The steps leading to this world has many issues but I would like to focus on the issues appearing in a world where all work is automated and not just some of the work. The first obvious issue is what humans should do. Humans need to find meaning with their life, working and contributing fills this need therefor wide spread depression amongst the population is a risk if work is taken away. There are of course other ways to fill this need, people take up hobbies and socialize but many people meet the people they socialize with at their work and cities are notoriously lonely despite millions of people living in them so this is not a problem which will solve itself. There needs to be meeting places for people with similar interests to form groups in which individuals can find belonging and meaning, this can be achieved quite easily when you do not have the same constraints of money as we have today.

This brings up the next issue, who gets what? We do not have the constraints of money we have today but there are still a limited amount of resources so everyone can't get everything. Looking at history sharing everything equally has ended poorly, probably because there are people who want more than others and if the government shares everything equally the only way to get more is to corrupt the government and that way get more. Creating a strong incentive to corrupt the government is not a good system, there need to be another system. One alternative is to keep jobs as a way to divide resources but this brings up yet another problem, how do you choose salaries? You cannot base the salary on production since machines outclass any human in all aspects of production, even AI designs of products will probably be superior to human design. One possible solution is to base salary on some carefully weight constant to the power of the total time spent working. You need this exponential growth in order to motivate people to keep working and if you don't allow money to be passed down to children you will not have class and inequality issues. One last problem I would like to discuss is education and government. If salary is simply based on time spent working many will choose not to get an education. This will result in the majority of the population being uneducated and assuming this society has a democracy, the opinion of the majority will be an uneducated and easily manipulated opinion. This can be solved be either accounting for education when deciding the salary or by not having a democracy but instead have an AI dictator. Personally I would prefer if power stayed amongst humans, humans would probably corrupt the AI dictator.

Word count 500

Summary of discussion

All parties in this discussion agree that automation of the majority of the human labour will inevitably happen within the next decades. In its core, we think this is beneficial for the human race and society at large. Some of us were concerned that this development may lead to unwanted effects concerning large waves of unemployment and depression. The current norm where unemplyed people are viewed as lazy and incompetent may be harmful to society when large portions of the population loses their jobs to artificial intelligence. The taboo of unemployment needs to be lifted in order to avoid civil unrest and depression which has previously been known to correlate with a high unemplyment rate. Our society will need to adjust so that these people can get the economical support to live, and only a small group of workers will provide for the society. One solution here could be to introduce some sort of citizen salary.

No one in the group belevies in a massive AI apocalypse where the machines rise up against the human race. We believe however that machines may harm humans through accidents. One interesting topic we discussed concerning this is how society will look at accidents caused by non humans. The tolerance level for machine mistakes may be very low. For example if all cars would be transformed to self driving cars with the present technology, statistics could show that the traffic would be safer but accidents would still happen. If one of the self driving cars would kill a child, not a single individual would be blamed as in the situation of today but rather the wole entity of self driving cars would bear the responsability.

We also discussed the robustness of a world run by AI. The ultimate goal of an automated society might lead us to a very fragile world, if we rely on robots to do all hard work for us. If the robotic industry is self-sustained, the need for humans to build the AI:s would perish. Some humans will probably have the skills necessary to sustain parts of society by pure interest at first, but when the reward for this is very low the risk is that vital knowledge is lost. This will probably not be noticed first but if a major solar flare wipes out a lot of major systems, the society might not have the ability to recover. Furthermore, another probable scenario where the vital systems might be disabled would be as a result of cyber warfare. In any case, the question of stability and security of the intelligent systems will likely be of uttermost importance.

In this discussion, the aim was to assess effects and risks of AI in our future society. The general opinion is that AI will bring many great things. Moreover, we are in agreement that we have to tread carefully, there are lots of things that could go wrong, both along the way to an AI governed world and when we are there. Proper care has to be taken now while we still have time define the course of our future.