Future Perspectives on Artificial Intelligence (AI)

- What to expect and should we worry?

Jim Tørresen, University of Oslo, 2014 ©

When and where does the big breakthrough come?

It is difficult to predict where and when the breakthrough comes in technology. Often it happens randomly and not linked to major initiatives and projects. Something that looks uninteresting or insignificant, can prove to be significant. I remember back when I was a student and first tried the browser Mosaic in 1993, one of the first graphical web browsers that were available (developed at the National Center for Supercomputing Applications (NCSA) at the University of Illinois Urbana-Champaign in the USA). It was slow and it was not then obvious that the web and the Internet were something that should be as large and comprehensive as it is today. However, Internet and access to it gradually became faster and browsers also became more user friendly. So the reason why it has become so popular is probably because it is easy to use, provides quick access to information from around the world and enables free communication with any connected. The underlying foundation for Internet is a scalable technology being able to allow for ever-increasing traffic. This has been the bottleneck for AI - lack of good technology that can handle more complex conditions.

Hugo de Garis is a prolific researcher from Australia who in the 1990s had great ambitions about building brains. He worked with *cellular automata* networks that he made rules for growing in three dimensions (equivalent to our own brain). It was both fascinating and a little scary to meet him in 1994 at his lab south of Kyoto in Japan, where he gave me prints of small networks in brilliant colors. The challenge was to get the networks trained similar to the human brain. The idea was that by building brains in hardware, they would be fast enough for advanced learning and reasoning. Therefore he got built a machine consisting of 72 programmable logic circuits (FPGAs). The idea was to reconfigure circuits at runtime to simulate larger neural network than there was room for inside the physical machine. More precisely, simulating 75 million neurons with the goal of controlling a robot cat. However, it appeared early that there was not consistency between the vision and the implementation. The project ended quite sad with no brain function at all demonstrated. This was probably due to a combination of problems getting the complex custom built machine to work and non-scalable learning methods.

As the complexity of our problem increases, it becomes more and more difficult to automatically create a system to handle it. Divide-and-conquer helps only to a limited extent. It remains to crack the code of how scaling occurs in nature. This applies both to the development of individuals and the interaction between them. We have a lot of computing power available today, but as long as we do not know how programs should be designed, it contributes only to a limited degree to come up with effective solutions.

Hod Lipson is a renowned researcher in robotics and AI at Cornell University in the USA. He expresses the following two major challenges in engineering science: a) Can we design machines that can *design* other machines, and b) Can we produce machines which themselves can *produce* other machines? Many laws of physics of phenomena in nature have been discovered, but we have yet to really understand how complexity arises in nature. Advances in research in this area are likely to have a major impact on how big breakthroughs we can expect within AI. We are talking about the complexity of both the mechanical structure and systems for sensing and control.

There are two groups of researchers that contribute to advances in AI. One group is concerned with studying biological or medical phenomena and trying to create models that best mimic them. In this way they try to demonstrate that the biological mechanisms can be simulated in computers. This is useful, not least in order to develop more effective treatments and medicines for both handicapped and to fight diseases. Many researchers in medicine collaborate with computer scientists on this type of research. One example is that the understanding of the ear's behavior has contributed to the development of cochlear implant that gives deaf the sense of sounds and the ability to almost hear normally.

The second group of researchers focuses more on industrial problem solving, and to make engineering sound systems. Then it is interesting to see whether biology can provide inspiration for more effective methods than those already adopted. However, if traditional methods give the best results, these are chosen rather than the less good biology-based methods. Anyway, the last group of scientists is working at a higher abstraction level than the first group, which is trying to determine how to best model mechanisms in biology. But both have mutual use of each other's results. An example is the invention of airplanes that first became possible when the principle of air pressure and wing shape was understood. Initial experiments with moving wings similar to birds were unsuccessful, and it was necessary to have a level of abstraction over biology to create robust and functional airplanes.

How similar to humans is it desirable that robots become?

How similar to the biological specimen can a robot become? It depends on developments in a number of fields such as AI methods, computing power, vision systems, speech recognition, speech synthesis, human-computer interaction, mechanics and actuators or artificial muscle fibers. It is definitely an interdisciplinary challenge.

Given that we are able to actually create human-like robots, do we want them? Thinking of humanoid robots taking care of us when we get old would probably frighten many. We fear the lack of human contact. There is also a hypothesis called the *uncanny valley*. It predicts that as robots get more similar to humans, people's pleasure of having them around increases only until a certain point. When they are very similar to humans, this pleasure falls abruptly. You may feel a robot as a monster surrounding you as if it were in a movie. Then the reluctance against robots increases to later decrease when they continue to be even more similar to humans. One moves down and up again by the "uncanny valley".

For some tasks, we maybe even prefer machines rather than humans. We do not like to be a burden to others. If a machine can help us, we prefer it in some contexts. We see it today with the Internet. Rather than asking others about how to solve a problem we have, we seek advice on the Internet. Probably we get things done with machines, which we otherwise would not get done or found out. Whether the robots look like humans or not is less important in terms of how well they solve the tasks we want them to handle. However, they must be easy to communicate with and easy to train to do what we want. Apple has had great success with its innovative mobile products that are easy to use. Probably both design and usability will be essential for many of us when we choose what types of robot helpers we want in our own home in the future.

The fact that we are developing human-like robots means that they will have human-like *behavior*, but not human *consciousness*. They will be able to perceive, reason, make decisions and learn to adapt, but will still not have human consciousness and personality. There are philosophical considerations that raise the question, but based on current AI, it seems unlikely to talk about artificial consciousness. It is argued several reasons for this, including that consciousness can only arise and exist in biological material.

Ethical considerations and risks by developing artificial intelligence

An increasing number of autonomous systems that are working together increases the extent of any erroneous decisions made without human involvement. Several books have been published on computer ethics (also referred to as machine morality). In the book "Moral Machines" a hypothetical scenario is outlined where "unethical" robotic trading systems contribute to an artificially high oil price, which leads to the automated program to control energy output switches over from oil to more polluting coal power plants to avoid increasing electricity prices. Coal-fired power plants cannot tolerate running at full production long and explodes after some time and creates massive power outage with the consequences it has for life and health. Power outages trigger terror alarms at the nearest international airport resulting in chaos both at the airport and arriving aircraft colliding etc. The conclusion is that the many lives are being lost and economic costs being large, only because of the interaction between separately programmed systems that automatically make decisions. The scenario shows that it is especially important having control mechanisms when a number of systems that take their own decisions interact. This in terms of mechanisms that automatically puts limitations, and also informs operators about the condition deemed to require human review.

The advantages of the new technology are at the same time so large that both politicians and the market welcome them. Thus, it becomes important that moral based decision-making becomes a part of the artificial intelligence systems. The systems must be able to evaluate the ethical implications of their possible actions. It can be on several levels, including if private laws are broken or not. However, building machines incorporating all the world's religious and philosophical traditions is not so easy. Ethical dilemma occurs frequently.

It is probably a desire of most engineers not to develop something that might hurt someone. Nevertheless, it can often be difficult to predict. We can develop a very effective driver support system that reduces the number of accidents and save many lives, but if the system on the other hand takes lives because of failure, few would want to use it. It is also not desirable to be responsible for creating a system where there is a real risk for severe adverse events.

Scientists working on artificial intelligence face a number of potential dilemmas:

- *People may become unemployed because of automation.* This has been a fear through decades of years already, but experience shows that the introduction of information technology and automation creates far more jobs than those which are lost. Many will argue that jobs now are more interesting than the repetitive routine jobs that were common in earlier manufacturing companies. Artificial intelligence systems and robots help industry to become more efficient in the production of goods, rather than replacing all employees.
- *We get too much free time.* If machines do everything for us, life can in theory become quite dull. Normally, we expect that automating tasks will result in shorter working hours. However, what we see is that the distinction between work and leisure becomes gradually less evident, and we can do the job almost from anywhere. Mobile phones and wireless broadband gives us the opportunity to work around the clock. Requirements for being competitive with others results in many today often working *more* than before. Although artificial intelligence contributes to the continued development of the technology and this trend, it can simultaneously be a hope that automated agents can take over some of our tasks and thus also provide us some leisure time.
- *Artificial intelligence can be used for destructive and unwanted tasks.* Artificial intelligence is widely used in military unmanned aircrafts (drones) in the air and for robots on to the ground. It saves lives in the military forces, but can by miscalculations kill innocent civilians. Similarly, surveillance cameras are useful for many purposes, but many are skeptical to advanced tracking of people using artificial intelligence. It would become possible to track the movement and behavior of a person moving in a range of interconnected surveillance camera. The British author George Orwell (1903-1950) published in 1949 the novel "1984", where a not so nice future society is described: Continuous audio and video monitoring are conducted of a dictatorial governments, led by "Big Brother". Today's technology is not far away to make this technically possible, but few fear that it will be used as in "1984" in our democratic society. Nevertheless, disclosures (e.g. by Edward Snowden in 2013) have shown that governments can leverage technology in the fight against crime and terror at the risk of the innocent being monitored.
- *Successful AI can lead to the extinction of mankind?* Almost any technology can be misused and cause severe damage if it gets into the wrong hands. A number of writers and filmmakers have addressed this issue through dramatic scenes where much gets out of control. However, the development of technology has not so far led to a global catastrophe. Nuclear power plants have gotten out of control, but the largest nuclear power plant accidents at Chernobyl in Russia (1986) and Fukushima in Japan (2011) were due to human and mechanical failure, not the failure of control systems. At Chernobyl the reactor exploded because too many control rods were removed by experimentation. In Fukushima cooling pumps failed and reactors melted as a result of the earthquake and subsequent tsunami. The lesson of these disasters must be that it is important that systems has built in mechanisms to prevent human errors and also helps predict risk for mechanical failure to the extent possible.

Looking back, new technology brings many benefits, and the damage is often in a different form than we first would think of. Misuse of technology is always a danger, and it's probably a far greater danger than the technology itself getting out of control. An example of this is computer software which today is very useful for us in many ways, while we are also vulnerable because some are abusing the technology to create malicious software in the form of infecting and damaging virus programs. Melissa virus in 1999 spread through e-mails and led to the e-mail systems in several large companies such as Intel and Microsoft were put out of operation due to overload.

Ethics for programmers

In the book "Moral Machines" which begins with the somewhat frightening scenario discussed earlier in this document, also provide a thorough review of how *artificial moral agents* can be implemented. This includes the use of ethical expertise in program development. It proposes three approaches: logic and mathematical formulated ethical reasoning, machine learning methods based on examples of ethical and unethical behavior and simulation where you see what is happening by following different ethical strategies.

Let's look at a relevant example. Imagine that you go to a bank to apply for a loan. The bank uses an AI-based system for credit evaluation based on a number of criteria. If you are rejected, the question arises about what the reason is. You may come to believe that it is due to your race or skin color rather than your financial situation. The bank can hide behind saying that the program cannot be analyzed to determine why you don´t got the loan. At the same time, they can be claiming that skin color and race are parameters *not* used. A system more open for inspection can, however, show that the residence address has been crucial in this case. It has given the result that the selected criteria provides effects almost as if unreasonable criteria should have been used. This is important to prevent as far as possible by simulating the behavior of AI systems to detect possible ethically undesirable actions.

All software that will replace human evaluation and social function should adhere to criteria such as accountability, inspectability, manipulation robustness and predictability. All developers should have an inherent desire to create products that deliver the best possible user experience and user safety. It should be possible to inspect the AI system, so if it comes up with a strange or incorrect action, we can determine the cause and correct the system so that the same thing does not happen again. The ability to manipulate the system must be restricted, and the system must have a predictable behavior. The complexity and generality of an AI-system influences how difficult it is to deal with the above criteria. It is obviously easier and more predictable for a robot to move in a known and limited environment than in new and unfamiliar surroundings.

Ethical guidelines for robots and robot developers

Professor and science fiction writer Isaac Asimov (1920-1992) was already in 1942 foresighted to see the need for ethical rules for robot behavior. His three rules have subsequently been often referred to in science fiction literature and among researchers who discuss robot morality:

1. A robot may not harm a human being, or through inaction, allow a human to be injured.
2. A robot must obey orders given by human beings except where such orders would conflict with the first law.

3. A robot must protect its own existence as long as such protection does not conflict with the first or second law.

The term *roboethics* was introduced in 2002 by the Italian robot scientist Gian Marco Veruggio. He saw a need for guidelines for the development of robots intended for making progress of the human society and individuals and help preventing abuse against humanity. Thus, we need ethics for robot designers, manufacturers and users. We must expect that the robots of the future will be smarter and faster than the people they should obey. It raises questions about safety, ethics and economics. How do we ensure that they are not being misused by persons with malicious intent?

Is there any chance that the robots themselves, by understanding that they are superior to humans, would be trying to dominate over us? We are still far from the worst scenarios that are described in books and movies, yet there is reason to be alert. First, robots are mechanical systems that might unintentionally hurt us. Then, with an effective sensory system, there is a danger that the collected information can be accessed by unauthorized people and be made available to others through the Internet. Today this is a problem related to intrusion on our computers, but may in the future develop into robots being hacked as well. Then it would be a challenge that robots collect a lot of audio and video information from our homes. We will not like to be surrounded by robots unless we are sure sensor data are staying within the robots only.

Another problem is that robots can be misused for criminal activities such as burglary. A robot in your own home could either be reprogrammed by people with criminal intent, or they have robots themselves to carry out the theft. So having a home robot connected to the Internet will place great demands on security mechanism to prevent abuse. Although we must assume that anyone who develops robots and AI for them have good intentions, it is important that the developers also have possible abuse in mind. The intelligent systems must be designed so that the robots are friendly and kind, while difficult to abuse for potential malicious actions sometime in the future.

In 2004 the first international symposium on roboethics was held in Sanremo, Italy. EU has funded research program ETHICBOTS where a multidisciplinary team of researchers was to identify and analyze techno-ethical challenges in the integration of human and artificial entities. *European Robotics Research Network (Euronet)* funded in 2005 the project *Euronet Roboethics Atelier*, with the goal of developing the first roadmap for roboethics. That is, undertaking a systematic assessment of the ethical issues surrounding robot development. The focus was on human ethics for designers, manufacturers and users of robots. Here are some examples of recommendations that the work ended up in for commercial robots:

- ***Safety***. There must be mechanisms (or opportunities for an operator) to control and limit a robot's autonomy.
- ***Security***. There must be a password or other keys to avoid inappropriate and illegal use of a robot.

- ***Traceability***. Similarly as aircraft, robots should have a "black box" to record and document their own behavior.
- ***Identifiability***. Robots should have serial numbers and registration number similar to cars.
- ***Privacy policy***. Software and hardware should be used to encrypt and password protect sensitive data that the robot needs to save.

The alternative of ethics for those designing systems is *computer ethics*, where one looks at the possibility of giving the actual machines ethical guidelines. The machines should be able to make ethical decisions using ethical frameworks. Ethical issues are too interdisciplinary that programmers alone should explore them. Researchers in ethics and philosophy should also be included in the formulation of ethical "conscious" machines that are targeted at providing acceptable machine behavior. Michael and Susan Leigh Anderson have collected contributions from both philosophers and AI researchers in the book "Machine Ethics" (2011). The book discusses why and how to include an ethical dimension in machines that will act autonomously. A robot assisting an elderly at home need clear guidelines for what is acceptable behavior for monitoring and interaction with the user. Medical important information must be reported on, but at the same time, the person must be able to maintain privacy. Maybe video surveillance is desirable for the user (by relatives or others), but it should be clear to the user when and how it happens. An autonomous robot must also be able to adapt to the user's "chemistry" to have a good dialogue.

References

Berleur, Jacques og Brunnstein, Klaus. 1996. *Ethics of computing: Codes, Spaces for Discussion and Law.* Chapman and Hall.

Anderson, Michael og Susan Leigh. 2011. *Machine Ethics.* Cambridge University Press.

Wallach, Wendell og Allen, Colin. 2009. *Moral Machines: Teaching Robots Right from Wrong* New York: Oxford University Press.

Bar-Cohen, Yoseph og Hanson, David. 2009. *The Coming Robot Revolution.* Springer.

Siciliano, Bruno og Khatib, Oussama. 2008. *Springer Handbook of Robotics.* The last chapter discusses robotics and ethics.