ARTIFICIAL INTELLIGENCE

– OPPORTUNITIES, CHALLENGES AND A PLAN FOR NORWAY

Teknologirådet

ARTIFICIAL INTELLIGENCE

– OPPORTUNITIES, CHALLENGES AND A PLAN FOR NORWAY

ISBN 978-82-8400-000-8 (digital edition)

Published: Oslo, September 2018
Printing: Litografia
Front cover: Birgitte Blandhoel
Digitally published on: www.teknologiradet.no

FOREWORD

Artificial intelligence has made a powerful leap forward in recent years. Machines can now learn to interpret text, speech and images. This means that advanced tasks that to date have been reserved for human workers can now be done more quickly and at a lower price by machines. This brings major opportunities for the creation of value and better welfare services, but the technology can also have an effect on the rights of citizens, and it may result in greater inequality.

This report from the Norwegian Board of Technology describes how machines learn, what their areas of application are and what challenges are inherent to this technology. The report argues that Norway needs a strategy for artificial intelligence, and advances 14 proposals which address, among other things, what areas of expertise we need, how personal data should be used and what development we want for society.

The expert group involved in this project has included:

- Erik Fosse, surgeon and director of the intervention centre at Oslo University Hospital
- Siri Hatlen, former director of Oslo University Hospital and head of the Norwegian Board of Technology
- Steinar Madsen, medical director at the Norwegian Medicines Agency
- Hans Olav Melberg, health economist and associate professor at University of Oslo
- Damoun Nassehi, general practitioner and member of the Norwegian Board of Technology
- Michael Riegler, senior researcher at the Simula Metropolitan Center for Digital Engineering and researcher at the University of Oslo

Special acknowledgements to Michael Riegler for his contributions to the chapters on machine learning and areas of application. Hilde Lovett from the Norwegian Board of Technology directed this project.

The Norwegian Board of Technology is tasked with providing independent advice on new technologies to the Norwegian Parliament (Stortinget) and other

public authorities. Our hope is that this report will contribute to a knowledge-based and future-oriented debate on the opportunities and challenges presented by artificial intelligence.

Tore Tennøe
Director, Norwegian Board of Technology

CONTENTS

SUMMARY

Artificial intelligence (AI) has made a powerful leap forward in recent years. Most of us use it on a daily basis when we conduct web searches, navigate through traffic, translate texts, use speech commands on our smartphones, or filter out unwanted email.

Access to significant quantities of data, powerful computing resources and advances in algorithms, especially neural networks, have made artificial intelligence one of the most important enabling technologies of this decade.

Artificial intelligence is driven forward by a desire to make machines capable of solving both physical and cognitive tasks that were previously reserved for humans. Until recently, programmed, rule-driven expert systems were the prevailing discipline, but at the dawn of the new millennium the field began transitioning towards being driven by statistics and data, and machine learning became the dominant approach. Computers could learn without being explicitly programmed.

How machines learn

An algorithm is a formula that states how something is done, and can be seen as a set of instructions for a computer program. In a machine learning algorithm, the computer itself has created some of the instructions.

Computers can now learn correlations, rules and strategies from experiences in real-world data, without anyone telling them what these correlations are. They can continuously adapt to the data, and the more data they have access to, the more accurate they become (adaptivity). This means that computers can

perform tasks on their own (autonomy). Complex tasks and decision-making can thus be taken over by machines, with faster execution times and lower costs.

Machine learning algorithms primarily learn in three ways:

1. *Supervised learning:* The algorithm learns with guidance from experience in a dataset. This may mean deciding whether or not an email message is spam, determining whether a picture of a mole is benign or malignant, or predicting whether a customer will cancel their mobile service subscription. This is the approach used most widely today, but it needs good data.

2. *Unsupervised learning:* The algorithm identifies new patterns and correlations in a dataset on its own. Examples include grouping customers into similar sets so they can be addressed with various types of campaigns, or detecting multiple subgroups of diseases, so that different patients can receive more targeted treatment. This approach has the potential to detect patterns not previously known to humans.

3. *Reinforcement learning:* The algorithm identifies the best strategy to achieve an aim by trying, failing, and being corrected along the way. This is how a computer can learn to win a chess game or optimise energy consumption in a data centre. This technique has the potential to discover smarter strategies than humans can.

Areas of application

Machine learning is used to make predictions. Put simply, predictions are a matter of filling in missing information. Predictions take the information available, i.e. data, and use it to generate information that is not available. This may be information about the past, present or future, such as, for example, detecting whether a credit card transaction was fraudulent, determining whether a mole is malignant or predicting what the weather will be like tomorrow.

There are multiple prediction techniques. The most common ones are:

1. *Classification* is the most widely used machine learning technique. It is used to determine what category a new observation most likely belongs to. This may be a question of identifying what or who is in an image, whether an email is spam or not, or what the most likely diagnosis will be based on a patient's symptoms.

2. *Clustering is* used to explore new datasets without advance knowledge of the correlations. It finds new structures and patterns in (unlabelled) data and divides them into groups or clusters based on similar properties. This technique can be used to group film consumers who are similar to one another so that they can receive targeted film recommendations.

3. *Anomaly detection techniques* discover events that are not consistent with an expected pattern in a dataset. The anomaly may be an attempt at bank fraud, a data breach, an unfavourable disease development, or disturbances in the ecosystem.

4. *Forecasting analyses* concern the ability to predict, with some degree of certainty, something that might happen in the future based on a series of historical data. Such analyses can be used to create a risk profile for a person, such as the probability that a person will drop out of school, will be able to pay down a debt of a certain size, or will develop a given type of illness.

The various prediction techniques can be combined in different ways, and as a result machines are now able to hear and see, interpret and understand. This turns machine learning into a powerful and usable tool in many areas of application:

Speech and audio recognition technologies translate speech to text and vice-versa. This is now in daily use on smartphones in the form of virtual assistants, and can make it easier to manage data systems and simplify routine tasks.

Text recognition technologies find meaning in unstructured, written information. Translation systems have seen significant improvement and customer support systems are in the process of becoming fully or half-automated using these technologies.

Image analysis and video analysis recognise objects in images and video. These technologies have made significant leaps forward over the last few years, and can now automate very advanced and resource-intensive tasks such as driving and imaging diagnostics.

Recommender systems are used to make risk assessments on accidents or illnesses, for example, and to personalise education, healthcare and other public services.

Can we trust machines?

Artificial intelligence is already affecting many choices that are made by individuals and organisations. That makes it all the more important for us to be able to trust and understand the recommendations that algorithms give us. There are, meanwhile, several challenges inherent in the way machines learn:

- *Biased algorithms:* Supervised learning means that machines can determine categories or predict outcomes more and more accurately, but the recommendations can never be better than the data on which they are based. Machines learn from data collected about society. They can reflect biased conditions and thereby make discriminatory decisions. Systems that evaluate job applications and select the best candidates are one such example of this. When the algorithms are trained on data from previous hires, they may be influenced by biased choices and practices from interview sessions.

- *The black box problem:* Unsupervised learning means that machines can identify new patterns and correlations in datasets, but they cannot necessarily explain the causal relations. The algorithms may be fairly opaque and difficult to understand; this is referred to as the black box problem. Lack of explanation makes it difficult both to appeal a decision and to accept responsibility for the decisions.

- *Ethical algorithms:* Reinforcement learning means that machines can develop optimal strategies to achieve their goals, but they will often seek to win at any cost and steamroll considerations such as ethics if they have not been explicitly programmed.

- *Malicious use:* Machine learning can be used to make malicious attacks better and more effective, and they can be scaled and spread rapidly. Such use can also promote anonymity and psychological detachment. The use of artificial intelligence may also introduce new and unresolved vulnerabilities. Cyber attacks can thereby become much easier to execute and more targeted, posing a threat to digital security. Physical objects such as drones and self-driving cars can be manipulated and used to threaten physical security. Political security can be threatened by formulating fake news to appear more trustworthy and individually tailored.

Proposals for a strategy for Norway

Artificial intelligence brings major opportunities for the creation of value and better welfare services, but it can also have an effect on civil security and the rights of citizens, and it may result in greater inequality.

A number of countries, such as Finland, the United Kingdom, France and China, have developed their own specific strategies for artificial intelligence, and a race against time is underway to capture the best talents and establish central positions.

This suggests that Norway should have its own AI strategy. A national strategy should address the competencies challenge, the need for data and responsible development. The Norwegian Board of Technology has the following concrete suggestions for such a strategy:

1. *An immediate boost in research:* The development of robust algorithms for machine learning demands specialised, research-based competence. Norway's long-term plan for research and higher education should be bolstered by a significant investment in artificial intelligence and machine learning when the plan is due for revision over the course of autumn 2018.

2. *Establish a key institution:* Norway's research assets are too few and too scattered. In order to strengthen research efforts and become attractive in terms of recruitment and international cooperation, it may be a good idea for Norwegian authorities to establish a key institution for research in artificial intelligence and machine learning. To ensure adequate breadth and depth of research, the institution may encompass multiple environments in a virtual organisation.

3. *Define ambitious and concrete goals for Norway:* Norway does not have the right conditions to make sweeping investments in artificial intelligence, but the country can play a leading role when it comes to connecting domain knowledge with general knowledge on AI. Norway should outline investments within areas where we have a combination of good training data and significant social need, such as healthcare, public services, sustainable energy and clean oceans.

4. *Master's degrees reinforced with AI:* Machine learning will become an important element in many industries and professions, such as manufacturing, oil and energy, media and entertainment, agriculture and aquaculture, medicine, education and public services. All professions and educational programmes should include an introduction to artificial intelligence and

machine learning. A dedicated master's degree programme in artificial intelligence should also be established.

5. *Give everyone the opportunity to learn about artificial intelligence:* Artificial intelligence will affect our lives and the choices we make, both privately and professionally. Norway should set ambitious goals, such as initially aiming for one per cent of the population to learn basic artificial intelligence. According to the OECD, every third job will see its content dramatically changed in the future as a result of automation and artificial intelligence.[1] This calls for Norway to reformulate today's system for further education and life-long learning and adapt it to the individual by offering new incentives.

6. *Open public data:* Open public data can contribute to innovation and new services in many sectors. The public sector in Norway should have ambitions to publish more public data, and to do so in an open format that is easy to navigate and reuse in machine learning.

7. *Data sharing that serves the community:* If data from Norwegian hospitals, schools and smart cities is shared with third parties, the community should receive added value in the form of improved public services, new business development, jobs or tax revenue. It is therefore necessary for government authorities to establish legal frameworks that make it possible to exchange data securely and that ensure that the distribution of rights, values and responsibilities is fair and balanced.

8. *Give citizens real control over their own data:* If public data about us is used to drive research and innovation, this will require that citizens get to have a say in the matter. Government agencies must therefore establish a clear digital social contract that provides citizens with a real possibility to control and shape their digital profiles and to determine whether and how their personal data should be shared.

9. *Ethical guidelines:* Traditional European values are being challenged by the expansion of digitisation in general and by the development of artificial intelligence in particular. The government should begin developing ethical guidelines and practices in areas where the technology is already exerting

[1] Nedelkoska and Quintini 2018.

extreme pressure on established values such as autonomy, democracy, justice, equality, solidarity and responsibility.

10. *Right to an explanation:* Since algorithms give advice and increasingly take decisions in areas of major significance to people's lives, it is also important that we are able to obtain an explanation so that it is possible to appeal a decision. Norwegian authorities should enact a right to explanation, and avoid using decision-making systems that do not provide a sufficient explanation.

11. *Requirement for open algorithms in the public sector:* When machines take over tasks that were previously carried out by humans, it is especially important to show that the algorithms do not make biased recommendations. Algorithms used by the public sector should, as a general rule, be open to public access and audit so that other societal actors can verify that they are being used correctly and with ethical responsibility.

12. *Auditing algorithms:* Algorithms for machine learning that for critical reasons cannot be open to the public should nonetheless be subject to evaluation before they can be put into broad use in society. One possibility is to require auditing or certification from an independent third party, who can evaluate whether the decisions behind the algorithm are fair, accurate, explainable and verifiable.

13. *Ethics by design:* Undesirable events such as biased or unfair decisions can lead to a breakdown in trust that would be difficult and costly to correct afterwards. The concept of *Privacy by design* should be expanded so that the algorithm's propensity to result in discrimination or manipulation is assessed from as early as the design stage.

14. *National dialogue on artificial intelligence:* Artificial intelligence is beginning to narrow in on the lives of most Norwegians, and working for passive acceptance will not be acceptable. The Norwegian authorities should actively take initiatives to involve lay people and civil society in the discussion on artificial intelligence, and they should be receptive to their perspectives on what developments people would hope to see.

A SPRING THAW FOR ARTIFICIAL INTELLIGENCE

Artificial intelligence has made a powerful leap forward in recent years. Large volumes of data, powerful computing resources and the development of better algorithms have laid the foundation for this development. The most promising area of development at present is neural networks, which are inspired by the functioning of the human brain.

The cover of Nature on 2 February 2017 featured algorithms that can learn to classify moles just as well as doctors can. The headline was the result of a collaboration between doctors from Stanford University and researchers in artificial intelligence (AI) such as Sebastian Thrun – the man behind the Google car.[2] The group had trained a neural network on clinical images of moles before it was tested against 21 certified dermatologists on 2,000 images. In almost every test, the algorithm proved to be more sensitive and accurate than the specialists. It captured more actual cases of melanoma while producing fewer false positives at the same time.

[2] Esteva et al. 2017.

This is a major advance in itself. In Norway, melanoma is the second most common type of cancer in the age group of 25-49 years, and more than 2,000 patients are diagnosed annually.[3] But the authors behind the Nature article had even more to offer. The same type of machine learning can also be adapted and used in other medical specialisations, such as ENT, optometry, radiology and pathology. The method is not only fast; it is also possible to use mobiles and tablets for diagnosis.

This is only one of many examples of machine learning innovations over the last few years. IBM's Watson won Jeopardy, Apple lets us talk with Siri on our smartphones, Google's driverless car has driven millions of kilometres, and Facebook recognises faces just as well as people do. As a result, many are now saying that we are in a spring thaw for artificial intelligence after an AI winter in which development produced little by way of practical results.

What is artificial intelligence?

Artificial intelligence means different things to different people, and even researchers have not reached consensus on an exact definition. Artificial intelligence is meanwhile driven forward by a desire to make machines capable of solving both physical and cognitive tasks that were previously reserved for humans.

The Turing test

In 1956, Alan Turing defined the necessary conditions for a machine to be considered intelligent, giving us a process known as the Turing Test. The test involves having a person communicate with a computer or another person using only a keyboard and a screen, without being able to see what or who is answering. The exchange could be about any possible subject and last for several hours. If the person cannot determine whether they are communicating with a machine or a person, Turing would say that the machine has passed the test and must be considered intelligent.[4]

Programmed, rule-driven expert systems used to be the predominant discipline. One such example is IBM's DeepBlue, which beat world chess champion Garry Kasparov in 1997. Such systems are, meanwhile, fairly inflexible, primarily developed for specific domains, and not sufficiently robust to handle events that are not specified in the rules.

[3] The Norwegian Cancer Society 2018 and Norwegian Institute of Public Health 2018.
[4] See also Tørresen 2014.

In the early 2000s, the field transitioned from being rule-driven to being driven by statistics and data, and machine learning became the predominant approach. In 1959, Arthur Samuel defined machine learning as the "field of study that gives computers the ability to learn without being explicitly programmed."[5]

Computers can now learn correlations, rules and strategies from experiences in real-world data, without anyone telling them what these correlations are. They can continuously adapt to the data, and the more data they have access to, the more accurate they become (adaptivity). This means that computers can perform tasks on their own (autonomy). Complex tasks and decision-making can thus be assumed by machines, with faster execution times and lower costs.[6]

The goal of artificial intelligence is for machines to also be able to learn intuition and knowledge that is difficult to express in rules; something which the neural network approach has shown to be possible. Neural networks are inspired by the structure and function of biological neural networks in the brain. Neural networks can also learn things that were not previously known, or that are not possible for humans to learn.[7]

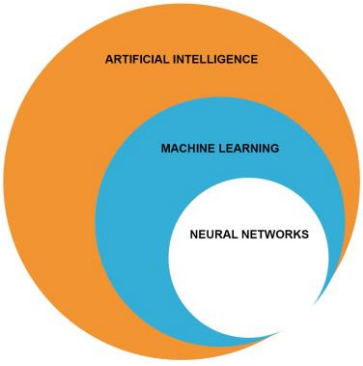

Figure 1: Relationship between artificial intelligence, machine learning and neural networks. [8]

Figure 1 illustrates the relationship between artificial intelligence, machine learning and neural networks. This report is about machine learning, and

[5] Al-Darwish 2018.
[6] Adaptivity and autonomy are characteristic properties emphasised in the Finnish online course on basic artificial intelligence; see also: https://course.elementsofai.com/1/1.
[7] Ahlqvist et al. 2018.
[8] Inspired by Wahed 2018.

focuses specifically on the use of neural networks, which is the approach currently driving advances in artificial intelligence.

BETTER ALGORITHMS

Machine learning has seen rapid development over the past few years, driven by three significant changes that have taken place in parallel: (1) the development of better algorithms, especially in neural networks, (2) access to large amounts of data and (3) easy and reasonably inexpensive access to continuously increasing levels of computing power.

Neural networks are a data-driven approach to machine learning. A breakthrough was made in 2016 when the program AlfaGo from Google DeepMind managed to beat the world champion in the game Go, which is a strategy board game that requires intuition.

The models for learning in neural networks consist of several layers of so-called neurons. The neurons in one layer learn by using input values from previous layers and sending new learning on to the next layer, all the way until the final layer, which produces the final output value. This might mean, for example, determining the category of an image ("Yes, this is an image of a malignant melanoma").

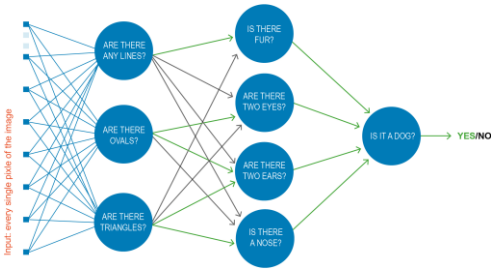

Figure 2: Schematic illustration of a neural network[9]

[9] Inspired by Geng and Shih 2017.

Is it a dog?
Let's assume that we have trained a neural network to recognise dogs in images, as illustrated in Figure 2. Important properties of a dog are that it has fur, two ears, two eyes, and a snout. We wish to classify a new image. The first layer of input values in the neural network will consist of a number of nodes equal to the number of pixels in the image. The second layer consists of neurons that take in the pixels and look for different forms such as lines, circles and edges. The third layer consists of neurons that evaluate what the lines, circles and edges represent. Pixels that a neuron evaluates as two circles will be sent to two other neurons that evaluate whether these are a pair of eyes or a pair of ears, respectively. Properties that are heavily weighted are shown in green in the figure. The final layer of output values will provide an evaluation of whether the image is of a dog or not.

The more layers a neural network consists of, the more complicated the structures it can analyse. Neural networks with many layers between the input- and output values are called *deep learning networks*. They have the ability to learn complex correlations and then undertake generalisations to recognise relations it has not seen before. The strength of deep learning networks is that they can learn what is important in order to understand an image, for example, without needing this to be explicitly explained. This makes deep learning a powerful tool in machine learning. The drawback is that this technology often demands extensive data and computing power, and the models can be complicated and difficult to explain in terms that people can readily understand.

Progress in neural networks and deep learning have made it possible over the last few years to train increasingly accurate machine learning algorithms, and they are widely used in image, video, text and sound recognition. The algorithms can now recognise objects in images better than humans,[10] and it has been demonstrated that it can be more accurate to talk into machines than to type information in by hand.[11] Google's machine learning-based translation system became 60% better through use of neural networks.[12] Neural networks can also be used to make forecasts, such as predicting extreme weather.[13]

[10] Karpathy 2014.
[11] Ruan et al. 2017.
[12] Turner 2016.
[13] Lui et al. 2016.

LARGE VOLUMES OF DATA

Machine learning, and neural networks in particular, learn by being fed large volumes of training data from the real world. Digital content has been produced smoothly and steadily over the last few decades, but the rate has really boomed in recent years. Every day we produce 2.5 trillion bytes of data, and 90 per cent of the digital information in the world today has been produced in the last two years.[14] These enormous volumes of data are helping to make machine learning markedly more accurate.

Signals from sensors on smart phones and industrial equipment, digital images and videos, a continuous stream of updates in social media, and the dawning internet of things (IoT) will produce far more digital raw material to work with over the coming years.

Whilst traditional machine learning algorithms can only improve up to a certain level before performance plateaus, neural networks represent a method where the results get better and better as they get access to more and more training data. The deeper the networks are, the more they can make use of large volumes of data (illustrated in Figure 3). This is why neural networks, and especially deep learning networks, currently dominate the field. It is important that the data be of good quality and represent a general and multifaceted image of the problem to be solved. Otherwise the learning model can produce imprecise or incorrect recommendations.

[14] IBM 2018.

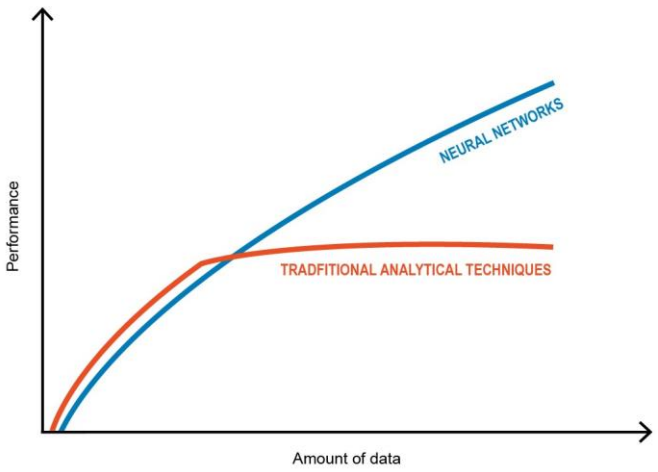

Figure 3: Neural networks scale well with increasing access to data compared to traditional analysis techniques.[15]

ACCESS TO COMPUTING POWER

Training machine learning algorithms requires massive computing power, both because they are based on large amounts of data and because the algorithms are adjusted along the way through trial and error. Access to the necessary computing capacity has increased considerably over the past few years, which has also been critical to the development of the AI field. The following factors have been important to computing power:

- *Moore's law*. The capacity *of the general processing unit (CPU)* has on average doubled steadily every 24 months over the past 50 years.
- *New, powerful computer chips*, such as graphics processing units (GPU) and processors special designed for neural networks can reach speeds up to several times faster than general CPUs.[16]
- *Cloud computing*. Powerful machine learning infrastructures optimised to manage neural networks are offered as cloud services. These can be used, purchased or leased as needed without having to make costly investments of one's own.

[15] Inspiration from foil 30, Ng 2015.
[16] Examples include Google's TensorFlow and Intel's Nirvana Neural Network Processor.

As a result, it has been possible to start experimenting with, developing, and applying machine learning easily and quickly at a reasonable cost. However, for these advances to continue, there is a need for new and less data-intensive algorithms. The exponential development in computing power is starting to run up against physical limits, and there is stiff competition on access to processing capacity because of the expansion of Bitcoin.[17]

[17] Tassev 2018.

HOW MACHINES LEARN

Machine learning algorithms learn primarily in three ways: by being guided by experience from historical datasets, by finding new patterns and correlations, or by trial and error.

GUIDED BY DATA

The most successful type of machine learning in recent years has been *supervised learning,* which learns from experience in datasets with real-world examples.

Each example has properties, also called input values. The input values may be pixel values of an image, soundwaves in an audio stream or other values such as living space, lot size, or number of bedrooms in a home. The dataset is also labelled with an output value, such as the animal an image represents, the words in the audio stream or the sales price of the home.

A good example of supervised learning in medicine is the classification of moles. The algorithm is based on a training set of medical images of moles (input values) that are labelled benign or malignant (output values). It thus produces a so-called predictive model that predicts whether a new image is malignant or benign with a certain probability. In testing, it was found that the model could classify melanoma on a par with the best dermatologists.[18]

Supervised learning can also be used to predict a future event. In one envisioned example, a seller wants to know which users will end up cancelling a subscription, so that he can launch a targeted campaign to retain them before they cancel. However, the seller does not know how he can identify the users who want to cancel. Assume that the company has records for 10,000 customers, half of which have cancelled and half of which are still customers. A supervised learning algorithm can train a predictive model that learns properties of those who have cancelled and those that have remained loyal customers. Once the model is trained, it can predict which of the current customers are most likely to cancel, so that the customer relations manager can prioritise initiatives focused on them.

How the model's quality is evaluated

The goal of training a predictive model is to make it as well-adapted as possible to real-world examples. Quality is assessed depending on how well the model

[18] Esteva et al. 2017.

makes correct predictions for each new observation, including observations that it has never seen before.

One common way of assessing the quality of a predictive model is to set aside a test set from the training data. The input values in the test set are fed into the model, and quality is assessed on the basis of how well the answers the model produces correspond with the correct answers. Common measures of the quality of a predictive model include sensitivity and specificity:[19]

- *Sensitivity* is the proportion of those who actually have an illness that are correctly captured as being sick (true positive). High sensitivity means that there are few sick individuals that will not be captured (few false negatives).

- *Specificity* is the proportion of those that are actually healthy who are correctly captured as healthy (true negatives). High specificity means that few healthy individuals are incorrectly classified as being sick (few false positives).

In the example with images of moles, the neural network was trained with a labelled dataset of 129,450 clinical images that included 2,032 different diseases. The algorithm's performance was compared with that of 21 certified dermatologists on two types of diagnoses, the most common being *actinic keratosis* and the most dangerous being *malignant melanoma*. For each test, the dermatologists and the algorithm were presented with 135 and 130 images, respectively, that they had not seen before, and where the true condition had been verified through biopsy (meaning that they were correctly labelled). The predictive model achieved both better sensitivity and specificity than the majority of dermatologists.

Over and underfitting

A good training set must give an adequately broad representation of reality. Rubbish in means rubbish out. However, a training set seldom includes all imaginable values and observations from the real world, and there is therefore always a risk that the training set will provide a biased image of reality.

[19]Other measures of quality include *accuracy*, *error rate*, *F1-score*, *MCC (Matthews Correlation Coefficient)*, *positive predictive value* and *negative predictive value*.

Algorithms can be over or underfitted to the data they are trained on, which is something that can cause a machine learning algorithm to perform poorly.

- An algorithm that is overfitted is too finely tuned to the training data, and will not manage to make accurate predictions when faced with new observations that it has not seen before. It quite simply learns too many details from the training set. In the home example, this can mean that one or more of the properties (living area, lot size and location) are not so important when it comes to predicting the price, or that the algorithm has not had enough training data to learn from.
- An algorithm that is underfitted or biased is not adjusted to the training data well enough, and will not be able to make accurate predictions when faced with new observations either. In the home example this can mean that the properties (living space, lot size and location) are not sufficient to predict a home's selling price in general and that more properties are needed.

In other words, it is not only the volume of data that is a determining factor in how accurate a model is. The properties in the dataset one uses can often be critical factors. Developers working on learning algorithms often try to advance with many properties before they arrive at a final model.

Static and dynamic learning models

A static model does not change through use. Training of the model takes place in controlled surroundings in a test environment, and changes take place by replacing the model with a new version. This gives developers full control of the model.

A dynamic model can continuously improve the model with new input values and can be used while the surroundings are in constant flux. One example is monitoring to identify attempts at data intrusion. Continuous learning can make the learning model more accurate, but the drawback is that these changes take immediate effect and that the developers consequently have less control.

Adverse development of virtual assistants
Microsoft's virtual assistant Tay continuously learned from conversations it had with internet users. However, it was exposed to systematic false-learning by users and developed into becoming a Nazi sex robot. Microsoft scrapped Tay 24 hours after launch.[20]

UNSUPERVISED

People learn exceptionally well without supervision, and adopt most of their knowledge about the world through pattern detection and association. Unsupervised learning is a similar approach, where machines learn through recognising patterns and sorting data without advance knowledge of the categories. Unsupervised learning can identify patterns that humans cannot even detect, and have potentially greater accuracy and scalability than supervised learning.[21]

An important breakthrough was made in 2012 when Google and researchers at Stanford University managed to identify cats in digital videos without being told what cats were.[22]|

Figure 4: Illustration of self-driving car (TechCrunch)

The company Cortica trains self-driving cars to understand their environment by classifying and organising the images constantly being captured by the car. A stop sign, for example, is an octagon with white edges and red in the middle. The AI system can learn along the way that sometimes the red is faded and sometimes the

[20] Wakefield 2016.
[21] Fagella 2016.
[22] Hof 2018.

white border might be hidden by a tree branch. The system can nevertheless make the necessary changes to be able to classify a stop sign as a stop sign. The company believes that unsupervised learning will make it possible for tomorrow's autonomous vehicles to better adjust to new situations on the road.[23]

A research team at Mount Sinai Hospital in New York has used unsupervised deep learning network to extract properties from the patient records of 700,000 individuals in a system called DeepPatient. These properties were then used as input values for other machine learning algorithms to perform classification. Without expert instructions and based only on data, DeepPatient's learning algorithms detected new patterns and created a model that can place patients into the group that is best suited to them. [24]

By comparing similar patients, the hope is that it will be possible to predict the future disease profile of the patient or give her the treatment that has proven most beneficial for this type of patient. To evaluate the model, researchers used 76,000 test patients covering 78 diseases. Researchers believe that their method can predict disease better than the most common statistical methods. The method appears to be especially good at predicting diabetes, schizophrenia and different types of cancer.[25].

Textual analysis of legal documents saves time
In 2016, JPMorgan Chase introduced COin (*Contract Intelligence*), a platform used to analyse contracts that uses unsupervised machine learning to analyse legal documents and extract important provisions. In a pilot study, they found 150 relevant properties from 12,000 loan agreements in a matter of seconds. This would have taken as many as 360,000 hours for each annual review to be done manually. This capacity can have major consequences, considering the fact that around 80 per cent of errors in loans are due to misinterpretation of contracts.[26]

Unsupervised learning is still an immature field. For the time being, most systems require some exercise or feedback from humans. If and when we learn to build robust unsupervised systems that learn without human involvement, this

[23] Hall-Geisler 2017. Note that the company in question does not use neural networks, but a simple cluster analysis.
[24] Miotto et al. 2016.
[25] Miotto et al. 2016.
[26] Zames 2016.

might open up many possibilities. They would be able to look at complex problems in new ways to help us detect hidden patterns in how diseases spread, how the price of securities develops in a market or how customers purchasing behaviour changes, for example.[27]

REINFORCEMENT LEARNING

In reinforcement learning, the machine learns through trial and error, and is rewarded or punished depending on whether the behaviour brings it closer to or farther from a goal.

This technique saw a breakthrough in 2016 when the application AlphaGo managed to beat the world champion Go player, Lee Sedol. Go has complex rules, and players largely base their moves on intuitive knowledge that is difficult to express in computer programming. The game also has a very large outcome space; i.e. a large number of possible moves. To learn how humans play the game, the learning algorithm learned the rules of the game itself and then studied 30 million positions from previous games played by amateurs and professionals. The program then played against different versions of itself thousands of times. With every round, the program learned from its own mistakes and gradually improved until it became so good that it beat the master.[28]

AlphaGo made some unusual moves that originally were thought to be wrong, but which have actually given human Go players new insight into the game. For example, Lee Sedol has won all his games since he played against AlphaGo and has said that AlphaGo taught him to play the game in a more creative way.[29]

The key to the breakthroughs in reinforcement learning has been the use of deep neural networks.[30] Thanks to deep learning, we have an effective way to

[27] Brynjolfsson and McAfee 2017b.
[28] Gibney 2016.
[29] House of Commons (Great Britain) 2017.
[30] Knight 2018.

recognise patterns in data, such as positions on the Go board. Every time the program makes a mistake or does something right, it calculates a value that is saved in large tables that are updated as the program learns. For large and complicated tasks, this requires massive computing resources.

AlphaGo techniques are reiterated to optimise energy consumption in data centres
A further development of AlphaGo technology can make Google's data centres more energy-efficient. The cooling systems learn from sensor data. While the system previously gave advice to the operators, the system now adjusts itself, continuously and in real time. By continuously learning from new data, the system can now achieve annual energy savings of 30%, and DeepMind anticipates further improvements.[31]

Reinforcement learning can also be done completely without guidance from data. DeepMind's next major breakthrough came with *AlphaGo Zero*. The program learned the game Go completely on its own, with only the rules of Go as input values, and without any training examples whatsoever. This way of learning proved to be extremely effective, and over the course of 40 days, AlphaGo Zero beat all earlier AlphaGo versions. The computing power used was less than the earlier versions, while performance was much better at the same time.[32]

Later on, a further development of the algorithm, Alpha Zero, learned to play chess on its own, with only the rules of chess as input values. After four hours of training, it beat the world's highest-ranking chess program, Stockfish.[33]

Reinforcement learning is suitable for applications where humans can specify the target, without necessarily being able to express how that target should be reached. In addition to games, the technique is also used to train self-driving cars how to manoeuvre, such as finding the most optimal moves to avoid an accident. The technique also has potential applications in medicine, where it can be used, among other things, to identify the order of medicines that will lead to the best outcomes for the patient.[34]

[31] Evans and Gao 2016, Lardinois 2018 and Sverdlik 2018.
[32] Hassabis and Silver 2017.
[33] Dockrill 2017.
[34] Zhao 2011.

HYBRID MODELS

SEMI-SUPERVISED LEARNING

Semi-supervised learning utilises a combination of a small volume of labelled data and a larger amount of unlabelled data.

In a hypothetical example, we have a few images of cats and dogs that are labelled, and a lot of images of cats and dogs that are not labelled. Through an unsupervised learning process we can group the images into clusters. The cat and dog images will presumably end up in two different groups. Since we also know how cats and dogs look based on the labelled dataset, the system can label the group that most resembles dogs as dogs and then do the equivalent for cats.

Another area to which semi-supervised learning lends itself is text classification and the analysis of natural language, where the amount of labelled data is low.

Active learning
Active learning is a type of semi-supervised learning where the model itself selects what unlabelled data will be most informative, and then asks a human to label (categorise) it.

This technique can achieve better performance than one would achieve by running supervised learning only on labelled data or by running unsupervised learning only on unlabelled data.

TRANSFER LEARNING

Techniques for transfer learning mean that we no longer need to reinvent the wheel for every problem we wish to solve, but instead build on existing knowledge. These techniques make it possible to transfer knowledge from domains where we have a lot of labelled data to new and similar areas where the data foundation is more sparse, costly, or dangerous to obtain. Knowledge from training to recognise images of cars can, for example, be used to train systems to recognise lorries.

Transfer learning can lend itself to the medical field because there is often a lack of training data, and this is something that a transfer model can compensate for. In the example with melanoma, the model was trained by using techniques

for transfer learning on an existing neural network called ImageNet[35]. It was trained on a volume of images of various general object categories, and can recognise where there are objects in the images, the shape of the objects, etc. This network can therefore be used to identify where in the image there might be moles and the shape of these potential moles. The specific mole algorithm can build further on this know-how and concentrate on learning to recognise whether there actually are moles and on being able to differentiate between malignant and benign moles.

Spam filtering
Many people do not label a sufficient number of messages as spam for individual email filters to be adequately effective. At the same time, a general email filter that is the same for everyone will not be accurate enough. A hybrid general/customised filter solution that makes use of transfer learning can be effective if it can learn from all the users who consistently label spam and at the same time learn from each individual who only labels a few emails as spam.[36]

Simulation is a transfer learning technique that uses data in a simpler and less risky manner. For example, it is necessary to have data from collisions and accidents to train self-driving cars, and it is necessary to have data from people who fall down in order to train systems that automatically detect people falling. These types of data are difficult to obtain, making it necessary to run simulations as part of system training.

Example of a simulator to train self-driving cars
Udacity uses a simulator to train self-driving cars. The company has also made the simulator available as a free piece of software so that others can use it to train self-driving cars.[37] Video games can also be used as a simulator to train algorithms for self-driving cars and Open API Universe has made code available for many video games.[38]

When the dataset is sparse but there is access to a learning model that has been trained on a similar domain, transfer learning techniques have proven to deliver better performance.[39]

[35] See Image-net.org.
[36] Multi-task learning 2018.
[37] Etherington 2017.
[38] Mannes 2016.
[39] Gupta 2017.

GENERATIVE ADVERSARIAL NETWORKS

Generative adversarial networks (GANs) are a means of generating or classifying data. GANs can be used, for example, to generate more data for a training set if there is a lack of data, or to create artwork resembling the style of a given composer or artist.

Such networks are a good method of learning from unlabelled data, which can be the key to making computers more intelligent in the years ahead.[40]

The GAN method is composed of two types of networks that compete with one another. The first network, the G-network, will learn to generate synthetic data that is as similar as possible to actual images. The other network, the D-network, will learn to detect which images are real and which ones are false. To illustrate, we can look at the G and D-networks as criminals trying to counterfeit money and the police seeking to detect the counterfeiters, respectively. The criminals need to learn to counterfeit money so that the police cannot detect them, whilst the police need to learn to recognise counterfeit money. Competition forces both parties to continuously improve.

Let's suppose that we need more images of cats to make a learning algorithm better at recognising cats. The G-network starts with an image that consists of completely random pixels. The D-network receives the image and evaluates whether it is a realistic image of a cat or not. In the next round, the G-network produces a new image by making a slight adjustment to the previous version based on hints it gets from the D-network. And so on and so on. The feedback from the D-network makes the G-network get better at generating realistic images of cats. By working together, these networks can both produce very realistic synthetic data and become better at detecting authentic data.

[40] Knight 2017a.

Detecting intestinal disorders with high sensitivity and specificity

Angiodysplasia[41] is a malformation of blood vessels in the walls of the intestinal tract, and it is one of the most common cases of bleeding in the intestine. Diagnosis is made by interpreting images of the intestine. One method is to have the patient swallow a camera pill that takes up to 60,000 images of the intestine. One study has shown that doctors examining such images detect only 69 per cent of cases (specificity). Researchers at the Simula Research Centre and the University of Oslo have developed GANs that can detect angiodysplasia in such images of the intestine with accuracy and specificity approaching 100 per cent and sensitivity of 98 per cent. This is far superior to other machine learning approaches.[42]

The GANs method can also learn what characterises the music of Beethoven and create new pieces of music that sound as if they were composed by Beethoven, or in the same way learn to create paintings resembling the work of Munch. GANs can also be more practically useful by filling in missing data in an incomplete image, automatically generating scenes in a video game, making images sharper, or generating simulated data to train self-driving cars.[43]

[41] Angiodysplasia, 2009.
[42] Pogorelov et al. 2018.
[43] Goodfellow et al. 2014 and Goodfellow 2017.

APPLICATIONS – FROM HEALTHCARE TO CARS

Artificial intelligence and machine learning make it possible to interpret speech, text, numbers, images and video, and to forecast future events based on patterns in data. This is useful in many areas, from cancer diagnostics and personalised education to energy optimisation, climate analysis and self-driving cars.

Artificial intelligence is changing the relationship between man and machine in numerous ways. By using machine learning, computers can perform tasks such as imaging diagnostics more effectively than humans, i.e. at a lower cost or in a shorter time. The algorithms can also do some tasks better than humans can, such as finding new patterns in medical data or optimising energy consumption. Machine learning systems can also be scaled quickly and easily so that tasks can be done at extremely high volume and marginal cost.

ORDER AND PREDICT

Machine learning is used to make predictions. Put simply, predictions are a matter of filling in missing information. Predictions take the information

available, i.e. data, and use it to generate information that is not available.[44] This may be predictions about the past, present or future, such as, for example, detecting whether a credit card transaction was fraudulent, determining whether a mole is malignant or predicting what the weather will be like tomorrow with a certain degree of accuracy.

There are multiple prediction techniques, the most common of which are described below.

CLASSIFICATION

Classification is the most widely used machine learning technique and is used to determine what category a new observation belongs to. This might involve, for example, identifying what is in a picture. Techniques for supervised learning lend themselves well to classification, and neural networks have proven to be highly effective.

Differentiation is usually made between sorting into two classes (binary) and multiple classes (multiclass). A spam filter is an example of binary classification that predicts whether an email is spam or not spam. Diagnostic tools that predict the most probable diagnosis or diagnoses on the basis of a new patient's symptoms will build on multi-classification.

CLUSTER ANALYSES

Clustering is used to explore new datasets without advance knowledge of the relations in the data. Cluster analysis finds new structures and patterns in unlabelled data and divides them into different groups or clusters based on similar properties.

This technique can be used to group film consumers who are similar to one another so that they can receive targeted film recommendations. Or the technique can be used to identify patients with similar symptoms and how treatments have worked in different groups so that new patients can receive a more targeted treatment.

---

[44] Agrawal et al. 2018

Cluster analyses can also be used to generate labelled datasets (which are often lacking) by identifying clusters and then having someone label the clusters.

ANOMALY DETECTION

Anomaly detection techniques discover events that are not consistent with an expected pattern in a dataset. Such anomalies may be attempts at bank fraud, data breaches or disturbances in the ecosystem. In a medical context, such techniques can be used to track developments of patient health conditions and discover any potentially dangerous or undesirable development.

This technique can also be used to improve a dataset by removing deviating data, which can lead to a statistically significant increase in accuracy.

PREDICTIVE ANALYSES

Predictive analyses (forecasts) involve being able to predict something that might happen in the future based on a series of historical data. Such analyses can be used to create a risk profile for a person. It may be the probability that a person will drop out of school, be able to pay down a debt of a certain size, or develop a given type of illness. Forecasting is useful when there is a need to prepare for a possible development or to prevent it from happening.

One such application is better understanding of weather phenomena. Meteorologists increasingly make use of machine learning to deliver prognoses of how long a storm may last of whether it will generate damaging hail.[45] Machine learning algorithms trained on data from extreme climate events have managed to identify tropical cyclones and atmospheric "rivers" (columns of water vapour that move together with the weather). The latter can result in dangerous amounts of precipitation in an area, but it is not always easy for humans to detect.[46]

[45] Reilly 2017.
[46] Lui et al. 2016 and Jones 2017.

SPEECH AND SOUND RECOGNITION

Speech and sound recognition technologies are used to translate speech into text and vice versa. Speech technology is now in daily use on mobile phones in the form of virtual assistants [47] such as Siri on the Apple iPhone, the Google assistant on Android phones and Amazon's assistant Alexa.

Deep neural networks are well suited for learning how to recognise certain words in an audio stream and this is the most important reason that speech recognition has seen major improvements over a short time. The error rate fell from 8.5% to 4.9% from summer of 2016 to 2017.[48]

Areas of application for speech recognition range from simplifying routine tasks, making it easier to manage data systems and finding patterns in digital audio signals.

TRANSLATING SPEECH TO TEXT

Speech-to-text dictation on smart phones makes it three times faster to create a text message than typing it in manually.[49] Doctors can dictate information directly into a patient file.

NATURAL USER DIALOGUE

Speech-to-text technologies can make dialogue with data systems more natural so that they are easier to use. The speech interface can help older and functionally impaired people who have difficulty using a keyboard or touch screen so that they can still use digital services and welfare technology. Doctors can manage data systems in the operating theatre by speech, keeping their hands free.

[47] Virtual assistant (also known as a chatbot): an application that provides guidance and answers questions through the use of natural language, either in writing or verbally.
[48] Brynjolfsson and McAfee 2017a.
[49] Carey 2016.

DETECTING RISK SIGNALS IN HEALTH EXAMS

Many healthcare measurements take the form of digital audio signals, such as those from digital ECG systems or stethoscopes. Speech recognition technologies can interpret and discover abnormal signals in such audio streams.

Kardia is a small device that measures a two-lead ECG through the fingertips. Users can measure ECG regularly and record the measurements automatically in a mobile application. The app uses machine learning to build a cardiac profile for each patient from the measurements. If a later measurement does not fit the profile, Kardia detects this and alerts the patient and/or health personnel.[50]

CliniCloud is a digital stethoscope used to measure heart sounds. Users can record their heart sounds on their own and receive help from a computer or doctor interpreting it. The company behind the device plans to use machine learning to interpret the heart sound recordings, and they hope to be able to interpret them as well as doctors can. The goal is to be able to discover abnormalities in the audio stream even if doctors have not detected them.[51]

TEXT RECOGNITION

Text recognition technologies (including *Natural Language Processing*) find meaning in unstructured, written information. This technology saw a breakthrough when IBM's Watson won against Jeopardy champions in 2011.[52] Since this time, programs for text recognition have been used in many different contexts.

TRANSLATING LANGUAGE

Google and Facebook have transitioned to using deep learning techniques for translation. When Google's method was published in 2016, they reported that

[50] AliveCOR 2017.
[51] Conversation with CEO Andrew Lin of CliniCloud 24.10.2017 and Niesche 2015.
[52] Ferrucci 2018. In January 2018, Alibaba's program for artificial intelligence was the first to beat humans in a Stanford University reading and comprehension test. The program from Alibaba scored 82.44 per cent compared to 82.304 per cent scored by humans. Fenner 2018.

it reduced errors by 60 per cent.[53] Google now translates almost every language to and from English in this way. Facebook uses this for more than 4.5 billion translations per day.[54]

CUSTOMER SUPPORT SYSTEMS

Text analysis can help make customer service representatives more effective or automate parts of question-and-answer services, as they have done for the Norwegian Tax Administration and Udacity (see fact box).

[53] Castelvecchi 2016.
[54] Ong 2017.

Automated assistance for tax questions
The Norwegian Tax Administration is in the process of developing a virtual assistant that responds to tax questions from the public. They started by looking into whether it would be useful and profitable to have a tool to help customer representatives respond more quickly and accurately to inquiries. However, they discovered that they were already able to respond quickly to uncomplicated and frequently asked questions, while the uncommon and more complex inquiries were also difficult for machines to understand and answer. They concluded that it would be better to create a virtual assistant that could help the public directly with the simplest questions. They use machine learning, i.e. text and speech processing, on a training set of various types of questions and intentions to understand what the user is asking for. The answers follow fixed rules so it does not need machine learning to run.[55]

Training by learning from the top sellers
The company Udacity discovered that some sellers were far more effective than others when responding to questions in chat. They wanted to try to improve all the sellers by having them learn from the best. In actuality, chat logs represent a set of labelled training data, which is exactly what a supervised machine learning system needs. Conversations that led to a sale were labelled successes, and all others were labelled as failures. They used the data to predict what responses successful sellers were likely to give in response to the most common inquiries, and shared these with the other sellers to give them a hint as to how they should answer these specific inquiries. After 1,000 training cycles, the sellers achieved a 54% increase in effectiveness and were able to serve twice as many customers.[56] Instead of building a virtual assistant that could take over all conversations, they created a system to help all the sellers improve their performance.[57]

DIGITAL TRIAGE

In health services, text analysis can help streamline triage at locations such as Accident and Emergency or the GP's office by providing faster and more targeted responses to those calling in. The triage service can gradually be automated, which is being tested in the United Kingdom.[58]

REVIEW OF PATIENT RECORDS

A patient record may comprise up to several hundred documents with unstructured text, which is time-consuming to go through manually. In many cases, it is critical that details are not overlooked. Text analysis can rapidly extract

[55] Presentation of the VAKI project, Norwegian Tax Administration, 8 December 2017 and SkLNytt 2017.
[56] Ng 2017.
[57] Brynjolfsson and McAfee 2017b.
[58] Murgia 2017.

meaning and key information from the journal and thereby lighten the workload of medical personnel.

Prior to every operation the hospital in Agder uses a system that warns of any allergies, and it does this in a shorter time than what an ordinary physician spends going through the patient's papers.[59] This can save time, which is especially important in situations where time is critical.

IMAGE AND VIDEO ANALYSIS

Image analysis and video analysis recognise objects in images and video. Facebook and other social media now recognise faces in images and ask their friends if they want to help label them with names. In the latest version of Apple's iPhone, facial recognition is used to unlock the device and as identification for services in the device.

These technologies have seen major advances in recent years. The image recognition error rate has fallen from greater than 30% in 2010 to around 2.25% in 2017 for the best systems. In comparison, the human error rate is about 5%.[60] Video recognition systems, which are used in self-driving cars and elsewhere, previously made errors as often as every 30 frames of video, while the best systems currently make errors less than once per 30 million video frames.[61]

The advances in the field mean that automation of both trivial and more advanced and resource-intensive tasks can be worth the investment.

OBJECT RECOGNITION

These techniques can detect and recognise objects as faces and texts in images. This can be used, for example, to quickly identify individuals in digital photo albums or automatically create captions.

[59] Christiansen 2017.
[60] Echersley and Nasser 2018.
[61] Brynjolfsson and McAfee 2017a.

By combining text translation techniques with techniques to recognise letters in images, it is possible, for example, to translate advertising images from one language to another.[62]

DIAGNOSTICS

Commercial radiology detection systems are under development for several types of medical imaging, including MRIs, CTs, ultrasounds and pathology imaging.

Researchers at the Simula Research Centre have developed algorithms based on artificial intelligence that can identify the eight most common gastric disorders in images of the intestine with at least 93 per cent accuracy.[63] Such systems can help specialists make faster and more accurate diagnoses.

AUTONOMOUS VEHICLES

Analyses of video together with data from sensors in the car and in the environment allow vehicles to manoeuvre in traffic. These technologies are now so good that they can perform the analyses in close to real time, allowing the cars to manoeuvre and react immediately if something unexpected happens.[64]

IMAGE AND VIDEO ENHANCEMENT

Image and video analyses can also improve or generate images and video. Images can be made sharper and scenes in video games can be generated automatically. Black and white films can also be automatically colourised[65].

RECOMMENDER SYSTEMS

Recommender systems are used to conduct risk assessments and provide personalised services.

[62] Brownlee 2016.
[63] Haugnes 2017.
[64] Hawkins 2018.
[65] Brownlee 2016.

The "Netflix Prize", which was held from 2006 to 2009, was an important impetus for the development of new and better algorithms for recommender systems. The company published a dataset of more than 100 million film rankings and offered a prize of $1,000,000 to whoever could make more accurate recommendations than the company's own system. The winner in 2007 was 8 per cent more accurate and used a collection of 107 different algorithmic approaches.[66]

Different ways of creating recommender systems

Collaborative filtering builds a model from a user's earlier behaviour (things that the user has previously purchased or clicked on or which they have ranked (by reacting with an emoticon, for example)) and from similar decisions made by other users. The model is then used to predict what products the user will most likely be interested in.

Content-based filtering uses a series of features of an item to recommend other items with similar properties.

Hybrid recommender systems use a combination of collaborative filtering and content-based filtering. A typical recommendation might be "we think this book will interest you because others similar to you have purchased it and because you bought a similar book before."

PERSONALISED OFFERS

Most online services selling goods and services now use recommender systems in some form or another to recommend products that you are likely to like and thereby also purchase. For example, one often gets recommendations that say "people like you also bought this".

By analysing what customers have done and then calculating the probability of what they will come to do, online shops can customise offers for each individual visitor on their own.[67]

PREVENTING TRAFFIC ACCIDENTS

Norway's Directorate of Public Roads has used machine learning on detailed road data connected with open public data to analyse accident risk on

[66] Bell 2018.
[67] Mystore 2017.

European, national, and local roads in Norway. The algorithms were able to determine properties in the road and its surroundings that increase the risk of accidents; these can then be used to prevent accidents.[68]

PREDICTING DISEASE

The sooner an illness is detected, the greater the probability of recovery will be. In Horsens, Denmark, initial studies show that algorithms can predict with 90 per cent probability who will be admitted with, for example, a blood clot over the course of the next 100 days. [69]

CUSTOMISED EDUCATION

Adaptive learning systems can help teachers adapt education to each individual student's level of skill and maturity. One research project had the aim of keeping all students in the flow zone, where the balance between what is too difficult and what is too easy is adjusted to the level of the individual. The project discovered that the dropout rate could be reduced by more than half and entire classes could boost their performance by nearly one entire grade on average.[70]

PERSONALISED TREATMENT

Personalisation makes it possible to tailor treatment to the individual patient. The company Petuum has developed methods for processing patient records, including diagnosis results, and uses this to recommend combinations of medications that are most likely to provide the best results[71].

PERSONAL TRAINING PROGRAMS

The fitness app UA Record uses data on diet and physical and psychological behaviour that they compile with results from people with similar health and fitness profiles in order to develop personal training programs.[72]

---

[68] Mandaric and Axelsen 2017.
[69] Fischer and Olhoff-Jacobsen 2017.
[70] Bjørkeng 2015.
[71] http://www.petuum.com and email exchange with Eric Xing for Petuum, 30 August 2017.
[72] http://underarmour.com.

LIST OF AREAS OF APPLICATION

Altogether, the various technologies can be used in countless areas of application. Here we have compiled a list of applications that we have come across while preparing this report. This list is not exhaustive, but is intended to give an idea of the broad spectrum of possibilities. The digital version of the document includes links to the examples.

AREA OF APPLICATION	EXAMPLE	AI ELEMENT
GENERAL TOOL		
Speech recognition	*Apple's Siri, Google Assistant, Amazon Alexa*	Classification Speech and sound recognition
Translation	*Google Translate, Facebook Translator*	Classification Text analysis
Automated customer service	*Kongsberg municipality, tax inquiries, bank inquiries*	Classification Text recognition
Recruiting	*Better balance between women and men, LinkedIn*	Classification
COMMERCE and BANKING		
Automated loan application	*Automated loan application*	Recommender systems
Personalised online shopping	*Amazon*	Recommender systems
Interpreting contracts	*JPMorgan*	Text analysis
IT TASKS and SECURITY		
Email sorting	*Spam filter*	Classification and cluster analysis. Text recognition
Facial recognition for identity management	*Identification on mobile phones, labelling oneself in images, discovering when others use images of you,*	Classification Image analysis
TRANSPORT		
Route planning	*Preventing traffic accidents*	Predictive analyses
Training of driverless cars	*From the surroundings, behaviour in traffic, from simulations, from video games*	Classification Cluster analysis Image analysis
Parking	*Where there is a suitable parking spot*	Classification
ENTERTAINMENT		
Personalised services	*Film*	Recommendations
Playing games	*Jeopardy, AlphaGo, AlphaGo Zero, Alpha Zero, Stanford knowledge competition*	Predictive analyses Text analysis
Game development	*Automatically generate scenes*	GANs, Classification Image analysis
ENERGY and ENVIRONMENT		
Optimise energy usage	*Google operations centre*	Predictive analyses
Error prevention	*Predicting maintenance needs*	
Fish field monitoring	*Detecting illegal fishing*	
WEATHER AND CLIMATE		
Predicting extreme weather	*How long a storm will last, tropical cyclones and atmospheric currents*	Predictive analyses
SCHOOL		Predictive analyses
Help doing assignments	*English essay*	Text analysis
Personalised monitoring	*Adaptive learning platform*	Recommendations
Fitness coach	*UA Record*	

Table 1: General areas of application

AREA OF APPLICATION	EXAMPLES	AI ELEMENT
FIRST LINE HELP	*Symptom checker, Automated triage*	Classification Text analysis
IMAGE INTERPRETATION	*Melanoma, Lung cancer nodules, Colon cancer, Breast cancer, eye diseases, pneumonia, prostate, colon, and lung cancer, heart disease*	Classification Cluster analyses Image analysis
IDENTIFYING ACUTE EVENTS	*Acute kidney injury, hospital infections*	
ASSISTED DIAGNOSTICS	*Diabetic retinopathy, rare cancers, abnormal heart sound,*	Text analysis Anomaly detection
BETTER TARGETED TREATMENT	*Predicting disease progression of lung cancer,* *optimal combination of medicines,* *predicting disease development of tumours,* *identifying new subgroups of diabetes,* *optimal medication for lung cancer*	Text and image analysis Recommender systems Recommender systems Image analysis Anomaly detection Recommender systems Reinforcement learning
SELF-TREATMENT OF CHRONIC DISEASES	*Diabetes, asthma, Cognitive behavioural therapy, mental health*	Recommender systems Classification
COMPLIANCE WITH MEDICATION	*Motivate and alert, asthma medication*	Predictive analysis
DISCOVER ADVERSE HEALTH DEVELOPMENTS	*Falls, self-measurements, abnormal heart sound, heart failure*	Image analysis Sound recognition Predictive analyses
OPTIMISING RESOURCE USE	*Course of care, standardised treatment package for prostate cancer*	Unsupervised and semi-supervised learning
RISK OF DISEASE	*Predict disease progression, risk of blood clots*	Predictive analyses

Table 2: Examples of areas of application in health

CAN WE TRUST MACHINES?

Artificial intelligence is already affecting many choices that are made by individuals and organisations. That makes it all the more important for us to be able to trust and understand the recommendations that algorithms give us.

The possibility of creating machines that think and take decisions raises many ethical questions. In the long term, it may be necessary to consider whether machines can be given a moral status, and it could be possible that machines come to achieve superintelligence by improving themselves in a positive feedback loop; a so-called "intelligence explosion".[73] If this were to happen, it could pose an existential threat to humans. But such perspectives assume algorithms and physical preconditions that do not exist today.

Here we shall address important challenges that are already inherent in today's technology, namely domain-specific machine learning based on neural networks. The various ways in which machines learn present several challenges:

- *Supervised learning* uses historical data that may reflect a biased relationship in society and lead to discriminatory decisions.

[73] Bostrom and Yudkowsky 2016.

- *Unsupervised learning* identifies new patterns and correlations, but may provide little transparency, be difficult to understand and explain, and make responsibility unclear.

- *Reinforcement learning* means that machines can develop optimal strategies to reach their objectives, but may involve other considerations being overlooked or ignored.

- In addition, machine learning will also be able to be used as a powerful tool by actors with *malicious intensions*.

BIASED ALGORITHMS

American citizen Kevin Johnson had good financial standing and a high credit score, but was suddenly informed that his credit limit was reduced by nearly 65 per cent. The reason was not a default or late payment on Johnson's part, but rather the fact that his shopping pattern resembled the pattern of customers who have difficulty paying.[74]

Johnson was a victim of so-called "behavioural profiling", where similarities between one's own behaviour and that of a larger group are used to drive decisions. Machine learning models make predictions about probable events or qualities on the group level in a similar manner.

Supervised learning means that machines can classify or predict outcomes fairly accurately, but the predictions can only be as reliable and neutral as the data they are based on. For so long as there are inequalities in society such as exclusion and other traces of discrimination, this will also be reflected in the data. The algorithms can therefore contribute to discriminatory decisions.

Systems that evaluate job applications and select the best candidates are one such example of this. When the algorithms are trained on data from previous hires, they may be influenced by biased choices and practices from interview sessions. They can unintentionally continue to learn from prejudices such as racial, gender or ethnic bias. Such profiling based on biased data can contribute

[74] Cuomo et al. 2009.

to self-fulfilling prophecies and the stigmatisation of groups even if this was not intended on the part of the developer.

This is not an unsolvable problem. With increased awareness on the part of developers, algorithms can be programmed to counteract bias or meet a non-discrimination quota. The Norwegian ICT company Evry, for example, has achieved its goal of having more female employees after the company started using an AI system as part of the recruitment process. The female proportion of the nearly 600 employees in 2017 was 33 per cent, and 40 per cent among new graduates, compared to only 20 per cent a few years earlier. The company believes this is due to the fact that the system bases selection on more objective criteria.[75]

THE BLACK BOX PROBLEM

In 2013, Eric Loomis was sentenced to six years in prison for trying to escape from police in a car that had previously been used in a shooting in Wisconsin. The judge based the harsh sentence not only on Loomis' criminal record, but also on the COMPAS algorithm, which calculated Loomis' risk of repeat offending to be high.[76]

Loomis appealed the sentence to the Supreme Court, arguing that the judge used an algorithm that he could neither examine nor challenge. The factors that go into the assessments and how much weight they are given are considered a business secret according to the company behind COMPAS. Loomis lost the appeal. The judges believed that he would have received the same sentence regardless based on the usual factors, such as the crime and his history of offence. The court did, however, indicate that they thought it was problematic to use a secret algorithm to send someone to prison.[77]

This type of problem will become a major concern in the future. Machine learning algorithms provide advice and increasingly take decisions in areas of major significance to peoples' quality of life and development, such as loan and job applications, medical diagnoses and law enforcement. So, it will become a

[75] Dagens Næringsliv 2017.
[76] Smith 2016. See also Garber 2016.
[77] Liptak 2017.

problem if the responsible parties are no longer able or willing to explain how and why a decision was taken. The algorithms become "black boxes" that instead conceal the assessments, uncertainties and choices on which their decisions are based.

We can differentiate between two main types of black box problems:

- *Insight into the algorithm* is intentionally limited for commercial reasons or on the grounds of national security or personal data protection. The ruling against Eric Loomis is one example of this.

- *The algorithm is complicated and difficult to explain* in readily understandable terms. One example is the DeepPatient model, which can predict fluctuations in schizophrenia better than doctors can, but without any tools to explain how the model arrived at such predictions.[78]

The latter type has become particularly relevant because of development in machine learning. Traditional, rule-based machine learning methods were developed by people and so they are thereby also easier for people to interpret. This is different from deep neural networks, which can have hundreds of millions of connections that each make a small contribution to the final decision.

Unsupervised learning means that machines can identify new patterns and correlations in the data, but they cannot necessarily explain their causal relation. The algorithms can be somewhat opaque and difficult to understand, and the lack of explanation makes it difficult to both appeal a decision and assume responsibility for the decisions.

Techniques for explaining algorithms

Several research projects are developing techniques to explain or substantiate recommendations from algorithms. The following are a few examples:

- *The University of Oslo and the Simula Centre* are developing a tool to help doctors report and explain how algorithms that analyse video of the intestine reached their recommendations. The system selects images that have been important in the decision and provides graphics showing what in the image has been a determining factor for the recommendation[79]

[78] Knight 2017b.
[79] GitHub 2018.

- *XAI (Explainable AI)* is being conducted under the Defense Advanced Research Projects Agency (DARPA) in the United States. The agency is seeking assistance in automated alerts, such as when planes or satellites discover something suspicious, in addition to an explanation as to why something is flagged by an algorithm. This way, the operators can ignore false alarms.[80]
- *LIME (Local Interpretable Model-Agnostic Explanations)* show what elements in a model have been relevant for a prediction, such as what symptoms were more important for a model predicting whether a personal has influenza.[81]
- *A research team at the University of California, Berkeley,* has developed a program which is trained to discover various bird species in photographs, and which provides an explanation for its recommendations. The system is assisted by another neural network that has been trained to connect properties in an image with sentences describing what people see in the image. The answer from the algorithm might sound something like *This is a western grebe because the bird has a long, white neck, a yellow, pointed beak and red eyes*.[82]

RIGHT TO AN EXPLANATION

Artificial intelligence is used in both automated decisions and as a supportive tool for people in partly-automated decisions. In both cases, the individual(s) affected by the decisions will have need for an explanation. This makes it a problem if the algorithms are difficult to evaluate, check and correct.

In work with the new European General Data Protection Regulation *(GDPR),* the *right to an explanation* of decisions based on algorithms has therefore become an important topic. Explanations behind a decision may be of two types:

- *How the system works*, i.e. the logic, meaning, expected outcomes and the general functionality of the system. Information on the logic may indicate whether decision trees are used or how various types of information are weighted and connected. For example, high speed can automatically lead to higher insurance premiums.

[80] Gunning 2018.
[81] Ribeiro, Singh and Guestrin 2016.
[82] The Economist 2018a.

- *Explanation of a specific decision,* i.e. rationale, justifications and individual circumstances leading to the decision. Explanation of an individual's increased insurance premiums may be, for example, that they have driven six mph (10 km) over the speed limit on average.

Before automated processing begins, the individual must have enough information to be able to give consent or make objections. In this instance, only system functionality is available. Once a decision has been made, and one wishes to appeal the decision, for example, information on the specific decision is also available.

A study conducted by researchers at Oxford concludes that the new General Data Protection Regulation (GDPR) does not provide a sufficient and meaningful right to an explanation after a decision has been made. The regulation offers a fairly limited right to be informed in advance so that one would be able to give consent.[83]

In the recitals of the General Data Protection Regulation (Recital 71) it states that guarantees for persons subject to automated decisions shall include "...specific information to the data subject and the right to obtain human intervention, to express their point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision."[84]

The European General Data Protection Regulation (GDPR) provides data subjects with the right not to be involved in a decision that is based exclusively on automated handling when this decision has a significant impact on the individual in question.[85] However, while the right to an explanation of automated decisions is addressed in the recitals, it is not mentioned in the regulation itself. It is thereby not legally binding.[86]

This lack of clarity is detrimental, given the spread and potential that machine learning has for both fully automated and partially automated decisions.

[83] Wachter, Mittelstadt and Floridi 2017.
[84] GDPR 2016.
[85] Proposal to Norwegian Parliament 2018, page 68.
[86] Wachter, Mittelstadt and Floridi 2017.

WHO CAN BE HELD ACCOUNTABLE?

Artificial intelligence is constantly moving the boundary as to what tasks machines can solve and what decisions they can make. Self-driving cars continuously make independent decisions about manoeuvres and image analysis applications can make diagnoses without the involvement of human expertise. These advances raise new questions as to who is responsible.

According to current EU Policy, autonomous systems such as bots and algorithms cannot be held responsible for actions or lack of action resulting in damage or injury to third parties. Therefore they cannot be made liable for compensation, either.[87] This means that humans still have the final responsibility for decisions made with the involvement of machine learning algorithms.

For a human to be able to assume responsibility, in many cases it will be necessary for the individual in question to be able to understand how the algorithms reach their recommendations. In situations where an explanation is important, but not present, it may therefore be necessary to choose algorithms that are perhaps less precise, but capable of providing an explanation.

ETHICAL ALGORITHMS

Techniques for reinforcement learning mean that machines can reach optimal strategies to reach their objectives within the rules that humans define for them. AlphaGo Zero has shown that algorithms can achieve better strategies than the best GO players. The machines will, however, overlook or ignore considerations that are not explicitly expressed in the rules. This can make it difficult to get a general view of all important considerations when the rules are written, especially for complex systems.

One example relates to military use. The World Economic Forum has raised concerns regarding the use of algorithms such as AlphaGo in warfare. The algorithm seeks to maximise the likelihood of winning rather than optimising the margins. If this game logic is used in autonomous weapons, it could result in

[87] European Parliament 2017.

violations to the principle of proportionality, because the algorithm would not see any difference between killing one or one thousand enemies. This can lead to more offensive warfare, [88]

This presents a particular challenge when it comes to ethical choices. One dilemma can be illustrated with self-driving cars. They can potentially be very good at avoiding accidents. But what if the machine's most optimal choice for the car and its passengers increases the probability that someone else will be injured at the same time? What way should the car drive if there is a large animal in the road that will potentially cause major damage to the car and injury to the passengers, but there is a small child on the sidewalk?

Any ethical routing decisions must be expressed in the rules if the system is to take them into account. It may be unpleasant and difficult to express ethical choices in clear and unambiguous rules that the machine understands.

German authorities have established guidelines for what ethical considerations shall apply and who shall bear responsibility when self-driving cars end up in situations of choice. Some of the principles are that people should be prioritised over property and animals and that no differentiation should be made in humans on the basis of age, sex, or other factors. The aim of these guidelines is to create predictable and unambiguous distribution of responsibility, so that developers and manufacturers do not bear the burden of making difficult ethical decisions. In addition, it should help those in traffic to feel more secure in knowing that the vehicles operating in traffic will not make unexpected choices.[89]

MALICIOUS USE

Technology can be used in many ways and for different purposes. Echo sounding, for example, was originally developed to detect and neutralise submarines, but later became a key tool in fisheries. Conversely, research on viruses can also be used to develop dangerous weapons for bio-terror or war. This is often referred to as *dual-use* of technology.

[88] World Economic Forum 2017, page 49.
[89] Federal Ministry of Transport and Digital Infrastructure (Germany) 2017.

Dual use challenges are also a central concern with respect to artificial intelligence. Autonomous drones delivering consumer goods can, for example, also deliver explosives. Some general traits of machine learning make these challenges pressing:

- *More effective and scalable*. AI systems can perform tasks more effectively while they can be quickly scaled at a reasonable cost at the same time. This can make it harder to defend against attacks such as *phishing*. [90]

- *Better*. AI systems can perform tasks far better than humans. They can classify medical images better than experts and they are more skilled than the top-ranked players in chess and GO.

- *Rapid spread*. Once a smart algorithm has been developed, it can be quickly reproduced and spread. The adoption threshold for these techniques is low, since the field is highly open.

- *Psychological detachme*nt. Such systems can promote anonymity and psychological detachment. For example, a soldier using autonomous weapons can avoid any need to be present (and thereby avoid being hit) and having to see the victim.

- *New and unsolved vulnerabilities* in today's AI systems. They can be trained to make mistakes by manipulating training data and they can be tricked with examples created to be incorrectly classified. Even though artificial intelligence can exceed human performance in many ways, it can also make types of errors that humans would never make.

A report drawn up by researchers and experts from several countries argues that machine learning can threaten citizens, organisations and nations in three key areas:[91]

---

[90] Phishing is a concept where attackers trick a victim into doing something (such as providing confidential information or transferring money) by sending the person in question an email and pretending to be an organisation or person that the individual trusts. Until now, phishing attacks have been largely based on identically worded emails sent out to large volumes of recipients. AI can make phishing much more targeted and effective by first screening each victim by studying activity on social media and then personally tailoring the content of emails and the alleged identity of the sender. It is thereby much more likely that the victim will be deceived. In the future such attacks may become both more effective and more frequent on a wider scale.
[91] Brundage et al. 2018.

- *Digital security:* for example, training machines to make cyber attacks more targeted.

- *Physical security:* for example, manipulating self-driving cars to crash or equipping drones with weapons.

- *Political security:* for example, producing and disseminating fake news in a more targeted manner.

Optical illusion

Machine learning models can be tricked into making mistakes with a type of optical illusion. By adding noise to images that is not visible to the naked eye, a model can be tricked into making classification errors. An image that completely appears to be a panda, for example, could be classified as a monkey by adding invisible noise into the image.[92] Such attacks may have dramatic consequences. Algorithms can be trained to interpret images of a speed limit sign showing 50 mph as 150 mph. If one were to print out a false image and glue it to an ordinary traffic sign, autonomous vehicles would drive much faster than the actual speed limit, with all the consequences that this brings. (This can be prevented by designing the system to handle abnormal sensor values, such as by cross-coordinating it with a digital map.) Attempts have also managed to trick algorithms to interpret images of a person wearing special glasses into classifying them as another person.[93] People can thereby change identity by wearing such glasses, and get through passport checks unauthorised, for example.

Techniques are being developed on many fronts to defend against malicious AI attacks: increased consumer awareness, systematic efforts to discover and detect system vulnerabilities, centralised solutions (as is done with spam filters) that remain updated of new threats, certification of authenticity to prove that images and videos are transmitted live (and not artificially produced) and requirements to register robots. [94]

There can be tension between concerns over transparency and security. Transparency with respect to algorithms will be important to reduce the risk of vulnerability and abuse, though it can expose the algorithms for malicious use at the same time.

[92] OpenAI.com 2017.
[93] Margolin 2016.
[94] Brundage et al. 2018.

14 PROPOSALS FOR NORWAY

Artificial intelligence brings major opportunities for the creation of value and better welfare services, but it can also have an effect on the rights of citizens, and it may result in greater inequality. A national strategy should address the competencies challenge, the need for data and responsible development.

NORWAY NEEDS A STRATEGY

Artificial intelligence brings major opportunities in many sectors, has significant implications for both the individual and society as a whole, and is developing more quickly than any other technology. It is therefore urgent that Norwegian authorities develop a strategy to reap the benefits and confront the challenges of this potent technology.

MAJOR OPPORTUNITIES

In this report we have given examples of how machine learning is already affecting many different sectors and having an impact on areas such as:

- *Health:* diagnostics, compliance, accurate medication, resource use
- *Transport:* self-driving cars, traffic planning

- *Finance:* loan applications, interpreting contracts
- *Energy:* optimisation of data centres and energy system, prevention of errors in the petroleum sector
- *Public services:* individually tailored services, automated case handling, recruitment, translation

Artificial intelligence will prove important to industry and value creation. An analysis carried out by Accenture concludes that economic growth potential up to 2035 will be doubled in size through use of artificial intelligence.[95]

For Norway, artificial intelligence will therefore be important both to ensure competitiveness in the private sector and to develop sustainable, better and more effective welfare services.

IMPORTANT CONSEQUENCES TO THE INDIVIDUAL AND SOCIETY

Machine learning makes it possible for analyses, interpretations and processing actions previously reserved for humans to now be taken over by machines, and thus be performed more quickly and at a lower cost. This can also give rise to new problems.

First of all, it may be difficult to understand why the algorithms recommend certain things or act as they do. This makes it difficult to counter-argue or amend decisions that may be based on biased datasets or hidden interests. The writer and mathematician Cathy O'Neil has characterised such algorithms as "weapons of math destruction" since they can affect many people, are lacking in transparency, and can have major consequences, such as for personal finance, education or criminal punishment.[96]

Secondly, artificial intelligence can also be used as a refined weapon by criminals or foreign powers. Such "dual use", as it is known, may include cyber attacks, the manipulation of physical objects such as drones or cars, and targeted manipulation of policy through the distribution of fake news.

---

[95] Purdy and Daugherty 2016. The analysis covers 12 countries, including Sweden, Finland and the United Kingdom and it covered the period from 2016 to 2035.
[96] O'Neil 2016.

A third possible consequence is disruption of the labour market by having many work duties overtaken by intelligent machines over the course of a relatively short period. A study by SSB (Statistics Norway) suggests that one out of three Norwegian jobs is at high risk of being replaced by machines over the course of the next two decades.[97] Estimations from the OECD suggest that fewer jobs will be able to completely disappear, but that, for about one-third of employees, large parts of their jobs may be overtaken by computers.[98]

The development of artificial intelligence may also contribute to increased inequality. Commercial artificial intelligence is currently dominated by platform companies such as Google, which has access to extensive data. In the digital economy, networking effects often serve to make the winners even stronger: the more people use the services, the more data the companies will receive, which in turn can further improve the services.

A TECHNOLOGY IN RAPID DEVELOPMENT – AND NORWAY IS LAGGING BEHIND

Artificial intelligence is a technology that has made a powerful leap forward over the past few years. New types of machine learning benefit from ever-increasing computing power and the massive amounts of data produced in society. A recent study shows that 85 per cent of the US population already uses services based on artificial intelligence, such as, for example, navigation, streaming services or transport.[99] This figure for Norway is presumably higher.

The current situation can be summarised as follows:

Private American companies are dominant
The world's most valuable companies, such as Apple, Amazon, Facebook and Alphabet (Google's parent company) use machine learning to personalise services and optimise operations. They have hundreds of millions of daily users around the world, and so they also receive enormous amounts of training data.[100] Alphabet is in a particularly advantageous position, and it is estimated that this company has around half of the 100 top developers in machine

[97] Pajarinen et al. 2014.
[98] Nedelkoska and Quintini 2018.
[99] Gallup 2018.
[100] The Economist 2017.

learning, with investments such as Google Brain (a new type of AI operating system), Google Cloud and DeepMind.[101] Amazon has invested 306 million dollars in new AI positions, making it the leading company in terms of recent investments in AI.[102]

China has grand ambitions and numerous users

China has the highest number of mobile and internet users in the world – around three times as many as the United States or India. The Chinese also use mobile pay services 50 times more often than Americans. These volumes of data are the fuel that machine learning runs on and they have already helped make China a world leader in voice and facial recognition. The country also has a strong position in robotics and automation.[103] Chinese authorities have launched an ambitious strategy to become the global centre for development of artificial intelligence by 2030. This will be achieved through targeted plans in research and innovation on a broad front, smart public services in transport, urban development, education and justice, and extensive civilian-military co-operation.[104]

The EU will lead in ethical AI

The European Commission presented its approach to artificial intelligence on 20 April 2018. The Commission wishes to see a substantial increase in private and public investments on the order of 20 billion euros before the end of 2020, and will adopt legislation for the re-use and sharing of data. The EU strategy has a clear social and ethical profile, with an emphasis on tackling challenges for the labour market, education, and inclusion, along with the development of ethical guidelines based on fundamental rights and the new European Data Protection Regulation (*GDPR*). Various stakeholders will be brought together in the European AI Alliance to develop the ethical guidelines in 2018.[105]

Our neighbours are underway

---

[101] Sinovation Ventures 2018, page 5.
[102] Paysa.com 2017. This survey, covering April to September 2017, shows that Amazon has invested 306 million dollars in new AI positions, followed by Microsoft (124 million), Apple (105 million) and Google (33 million).
[103] China's State Council 2017, page 3.
[104] China's State Council 2017.
[105] EU Commission 2018a.

Many of the EU countries are underway with their own strategic work. The first reports with proposals to the government have been published in both Sweden and Finland.[106] In March, French president Emmanuel Macron presented his plan for artificial intelligence, with an investment in research of 1.5 billion euros.[107] In Great Britain, the government and a number of private enterprises have entered into an *AI Sector Deal* for the development of artificial intelligence.[108]

Norway is lagging behind and lacking a national strategy.
Data from the platform Kaggle, which brings together 15,000 developers of machine learning, gives an indication of where competence is located in the world. In one user survey, the majority of respondents come from the United States (4,200), followed by India (2,700), Russia (578) and the United Kingdom (545). Norway is far down on the list with only 53 respondents. In a ranking of the 100 top developers, none are from Norway.[109]

Norway ranks only 15th out of 35 countries in the *Government AI Readiness Index,* which ranks how well-prepared the OECD countries are to implement artificial intelligence in public services.[110]

NTNU and several leading companies in Norway have recently joined to establish the Norwegian Open AI Lab, and the Research Council of Norway lists artificial intelligence as one of several priority areas in the IKTPLUSS program.[111]

Meanwhile, the government's *Long-term Plan for Research and Higher Education 2015-2024* mentions neither artificial intelligence nor machine learning. Norway does not have any national strategy for artificial intelligence, and it does not have one in the works, either.

---

[106] Vinnova 2018 and Finland's Ministry of Economic Affairs and Employment 2017.
[107] Reuters 2018.
[108] Department for Business, Energy and Industrial Strategy and Department for Digital, Culture, Media and Sport (UK) 2018.
[109] Kaggle 2017 shows responses to a survey of 16,716 users on Kaggle from 171 countries and areas. Scimago 2018 shows that China (102,000) and the USA (84,000) are in the lead with two to three times as many publications on artificial intelligence as the runner-up, Japan (34,000), over the past 20 years. Norway is in 41st place with 1,700 publications. Kaggle (2016) shows where the 100 top-ranked developers on Kaggle are from.
[110] Stirling et al. 2018.
[111] Telenor 2018 and https://www.forskningsradet.no/no/Utlysning/IKTPLUSS/1254002623262/.

In the following, we will present concrete input for a Norwegian strategy for artificial intelligence. The most important elements are the right and adequate expertise to develop, evaluate and implement machine learning; access to data that balances personal data protection with the ability to drive innovation; and measures and principles for development that is both responsible and desirable.

THE COMPETENCE CHALLENGE

The government defines enabling technologies as "technologies that prove to be so far-reaching that they lead to major changes in society."[112] We wish to assert that artificial intelligence is one such technology, which can be compared to a turbocharger for the digitisation of society.[113]

1. AN IMMEDIATE BOOST IN RESEARCH

The development of robust algorithms for machine learning demands specialised, research-based competence.

The EU Commission points out that it is working rapidly to strengthen Europe's research quality and capacity in artificial intelligence. The EU will therefore increase investment in the Horizon 2020 research programme by 1.5 billion euros between now and 2020.[114] Norway has agreed to contribute to this investment through a declaration of cooperation. Before the end of 2018, the EU Commission will present a coordinated plan for investment and further collaboration between Norway and 24 EU countries[115]

Norway is currently trailing in this area. The Research Council of Norway has established, for example, that it is particularly challenging to meet demands for know-how in artificial intelligence and machine learning in Norway.[116]

[112] Norwegian Ministry of Education and Research 2014, page 30.
[113] Finnish Ministry of Economic Affairs and Employment 2017, page 11.
[114] EU Commission 2018a, page 7. See also EU Commission 2018b.
[115] EU Commission 2018c.
[116] NFR 2017, page 90.

The government's *Long-term Plan for Research and Higher Education 2015–2024* shall be revised over the course of 2018. Space should be created here for dedicated investment in artificial intelligence and machine learning.

2. ESTABLISH A KEY INSTITUTION

Norway's research resources are too few and too scattered. In order to strengthen research efforts and become attractive in terms of recruitment and international cooperation, it may be a good idea for Norwegian authorities to establish a key institution for research in artificial intelligence and machine learning.

The institution should be multi-disciplinary in its make-up and it should address both theoretical development, application and ethical assessments associated with the development of artificial intelligence. To ensure adequate breadth and depth of research, the institution may encompass multiple research environments and companies in a virtual organisation.[117]

In Ontario, Canada, local and state agencies have established the Vector Institute together with the University of Toronto and private companies, among others, as one of three national hubs for the development of artificial intelligence. The investment is also an initiative to prevent the province from losing expertise in artificial intelligence and machine learning to the major US-based companies.[118]

3. DEFINE AMBITIOUS AND CONCRETE GOALS FOR NORWAY

Norway does not have the resources to invest as broadly as China or France, but it can be a world leader in terms of connecting domain knowledge with general knowledge on artificial intelligence.

The EU Commission and UK Prime Minister Theresa May have introduced a new form of research investment that they call "*missions*". These are bold, inspiring and ambitious objectives to solve some of the major social challenges we

[117] Finnish Ministry of Economic Affairs and Employment 2017, page 49.
[118] The investment stems from the Canadian AI strategy; see also University of Toronto 2017 and CIFAR 2017.

face today. At the same time, they shall include realistic research and innovation activities, with requirements for measurable and time-oriented results.[119]

One of May's *missions* has the aim of using data, artificial intelligence and innovation to transform the prevention, early diagnosis and management of diseases such as cancer, diabetes, heart disease and dementia before 2030. One of the ambitions is for it to be possible within 15 years to diagnose cancer in the lungs, colon, prostate, and ovaries at a much earlier stage in 50,000 patients per year, thereby increasing the five-year survival rate for 22,000 more citizens of the UK every year.[120]

Artificial intelligence can help solve many important social challenges. It would make good sense for Norway to formulate objectives within the following areas where we have a combination of good training data and significant social needs:

- *Health:* Norway has a relatively unified health services network with good health data and digitally active users. Increased demand for health services is expected to increase in step with the age wave.

- *Public services:* Norway has world-class public data as a result of its well-organised welfare state and its digitally active citizens. Forecasts state that public expenditures will increase more quickly than public revenues starting in 2030.[121] It will therefore become necessary to reconfigure how the public sector delivers its services.

- *Sustainable energy:* There are already large volumes of sensor data from oil- and gas installations. Equinor has recently established two digitised operating centres in Bergen, and anticipates that investments of between one and two billion kroner shall yield an increased value creation of around 15–20 billion kroner.[122] Agder Energi uses machine learning to optimise hydropower production.[123] The global climate and environmental challenges demand a realignment and transition to products and services that exert significantly less negative consequences for the climate and environment than we are seeing today. Society must undergo a green shift.

---

[119] EU Commission 2018d and Mazzucato 2018.
[120] May 2018.
[121] Message to the Norwegian Parliament. 2017.
[122] Equinor 2018.
[123] Moe and Breivik 2018.

- *Clean oceans:* Norwegian institutions and companies have extensive data from satellites, buoys and drones that can provide important knowledge. Norway has legal usage rights over vast areas of ocean and presides over enormous resources both at sea and in the offshore industry. At the same time, the ocean is under significant pressure as a result of pollution, heating and acidification, among other things. This is where Norway can take a special responsibility.

4. MASTER'S DEGREES REINFORCED WITH ARTIFICIAL INTELLIGENCE

Machine learning will become an important element in many industries and professions, such as manufacturing, oil and energy, media and entertainment, farming and aquaculture, medicine, education and public services.

All professions and courses of study, both at the university and college levels, should therefore provide an introduction to artificial intelligence and machine learning as a supplementary offer for students and researchers. One example where this is happening is the Faculty of Medicine at the University of Bergen, which is beginning a new course for medical students in the spring of 2019. The course will enable them to understand and evaluate how machine learning and data analysis can be used in predictive and personalised medicine.[124]

Other educational programmes will also need to incorporate artificial intelligence in the years ahead, both to provide skills for the development of domain-specific artificial intelligence, and to provide a critical foundation of knowledge for users. [125]

Dedicated master's programmes in artificial intelligence should also be created. These programmes can be modular in nature and possible to complete

[124] http://www.uib.no/emne/ELMED219.
Another example is the School of Management at the Norwegian University of Life Sciences, which offers a course in machine learning for the optimisation of business processes.
https://www.nmbu.no/emne/INN355.
[125] Example areas may be energy and environment, aquaculture and agriculture, law enforcement and justice, medicine and education.

alongside a job, and they can also be integrated into many fields, such as health, law, and logistics.[126]

These master's programmes can be developed and funded in binding collaboration between industry, academia and government agencies. In the British *AI Sector Deal,* industrial actors commit to develop industry-financed master's degree programmes in AI and initially to finance 200 students annually. They will also be considering what possibilities are present in other disciplines.[127]

5. GIVE EVERYONE THE OPPORTUNITY TO LEARN ABOUT ARTIFICIAL INTELLIGENCE

Artificial intelligence will affect our lives and the choices we make, both privately and professionally. It is important that as many people as possible understand the key implications of artificial intelligence, so they can think critically on the topic, help shape its use in the workplace, participate and drive the debate.

In concrete terms, Norwegian authorities can draw inspiration from Finland and set an ambitious goal, such as for one per cent of the population to learn fundamental concepts in artificial intelligence every year. In May 2018, the Finnish government launched *Elements of AI,* which is a free, online and accredited basic course in artificial intelligence.[128]

Secondly, there is a major need for training in the workplace. The OECD evaluates that around one third of Norwegian jobs will have radically altered content in the future as a result of automation and artificial intelligence. Around 850,000 Norwegians will therefore need comprehensive skills development

[126]A similar structure has now been proposed in Finland. Finland's Ministry of Economic Affairs and Employment 2017, page 52.
[127] Department for Business, Energy and Industrial Strategy and Department for Digital, Culture, Media and Sport (UK) 2018.
[128]The course is oriented towards the general public, and does not require any advance knowledge in mathematics or programming. Completion of the course results in academic credits for Finnish residents and a certificate for everyone who completes it. The ambition is to have one per cent of the Finnish population complete the course in the first year. https://course.elementsofai.com/.

initiatives, something that the current further educational structure is not equipped for.[129]

This calls for Norway to reformulate today's system for further education and life-long-learning and adapt it to the individual by offering new incentives. Singapore offers *SkillsFuture for Digital Workplace,* which provides training in digital skills adapted to various age groups.[130] All citizens over the age of 25 receive *SkillsFuture Credit,* which is 500 Singapore dollars (around 300 euros) to spend on a course every year.[131]

NORWAY'S ADVANTAGE: DATA

Machine learning algorithms are trained on data. It is therefore difficult to develop artificial intelligence without data as raw material. Norway has world-class public data as a result of its well-organised welfare state and its digitally active citizens. This doubtlessly has major value for private actors seeking to develop commercial services, but a strategy should also create value for the community and provide individual citizens with adequate control over their own personal data.

6. OPEN PUBLIC DATA

Open public data can contribute to innovation and new services in many sectors.

As a general rule, public institutions should share data. Norway is in tenth place out of 114 countries on a scale showing the degree to which public agencies publish and use open data.[132] The public sector in Norway should nonetheless have

[129] Nedelkoska and Quintini 2018.
[130] SkillsFuture 2017.
[131] http://www.skillsfuture.sg/Credit
[132] https://opendatabarometer.org/?_year=2016&indicator=ODB

ambitions to publish more public data, and to ensure that it is in an open format that is easy to navigate and reuse in machine learning.[133]

In the UK's *Sector Deal*, the authorities have committed to publishing more open data, even if the country is already ranked number 1 on the same scale.[134]

7. DATA SHARING THAT SERVES THE SOCIETY

If data from Norwegian hospitals, schools and smart cities are to be shared with third parties, the community should receive added value in the form of improved public services, new business development, jobs or tax revenue.

How data creates value and for whom is not always foreseeable with machine learning. The machines can learn on their own and arrive at new correlations that were not previously known, and learning can be transferred from one system to another. This makes it complicated to regulate the responsibility and rights of parties. For example, how much should the public pay for services trained on their own data?

Citizens are part of the case, since public data is about them and from them. This means additional requirements for responsible use and transparency will be required if trust in data sharing is to be maintained.

It is therefore necessary for government authorities to establish legal frameworks that make it possible to exchange data securely and that ensure that the distribution of rights, values and responsibilities continues to be fair and balanced into the future.

This is an area where Norwegian authorities can take inspiration from Great Britain, which is establishing so-called *Data Trusts*.[135] This is a far-reaching legal framework for the sharing of data between public organisations and private companies that will develop artificial intelligence. The framework includes

---

[133] For example, registers such as the Norwegian Patient Registry and Prescription Registry should share aggregated data and synthetic data files (not personally identifiable) that everyone can use.
[134] Department for Business, Energy and Industrial Strategy and Department for Digital, Culture, Media and Sport (UK) 2018.
[135] Department for Business, Energy and Industrial Strategy and Department for Digital, Culture, Media and Sport (UK) 2018 and Hall and Pesenti 2017. See also Thornhill 2017 and Artificial Lawyer 2017.

concrete tools, agreement templates and mechanisms for the distribution of values created.[136]

The trust agreements must be developed and adjusted on an ongoing basis, in close collaboration with both those who are sharing and those who are using the data. A qualified operational organisation should therefore be established to manage the processes and oversee the framework.[137].

Know-how must also be developed into how data can be shared, connected and handled in a secure manner that still allows access for many. This know-how applies to the creation of synthetic data and encryption, for example. [138]

8. GIVE CITIZENS REAL CONTROL

If public data about us is shared to drive research and innovation, this should require that citizens have real control over how their own data is shared, and it must be guaranteed that this is done securely.

In Norway, citizens currently have limited control over data on themselves. Citizens' data is in various private and public "silos" with different policies for collection, sharing and use. This also makes it difficult to understand, evaluate and manage the risk associated with data collection and use.

The Norwegian Board of Technology has previously advocated the establishment of a clear digital social contract governing the interplay between citizens and public organisations. [139]. Providing citizens a real possibility of controlling

---

[136] Finance Norway has, together with the Brønnøysund Register Centre, the Norwegian Tax Administration, the Norwegian Labour and Welfare Administration and the Police Authority collaborated on digitisation in the DSOP (Digital Interaction Public-Private) collaboration. The purpose is to be able to share information easily, effectively and digitally so as to achieve the greatest possible level of productivity in society, but at the same time within secure frameworks that account for the individual's privacy protection. The framework is open for everyone, and elements of it can be evaluated for further development and adapted to the sharing of data to develop AI systems. See https://www.bits.no/project/dsop/ and Holte 2018.

[137] In the UK, the establishment of an operational organisation for Data Trusts has been recommended. A new *Centre for Data Ethics and Innovation* will be established. Hall and Presenti 2017.

[138] Germany has been working to establish know-how on data sharing and what the public can do: Antoni and Schnell 2017.

[139] The Norwegian Board of Technology 2017, page 55.

and defining their digital profile and determining if and how their own data shall be shared will be an important aspect of such a social contract.

Government agencies must therefore arrange for citizens to have access to appropriate tools so that they can truly and effectively control information on themselves.[140] In the same way that they manage their personal finances in online banking, they must be provided with a digital interface that presents a simple and understandable overview of how personal data is managed and used by the public sector. The citizen must also be given the possibility to actively grant or revoke permissions for various usage purposes.

RESPONSIBLE AND DESIRABLE DEVELOPMENT

In its Global Risk Report 2017, the World Economic Forum calls artificial intelligence one of the most rapidly developing technologies with the greatest utility value, but also with the greatest potential for harm.[141]

Learning machines can contribute to better welfare services, fast and accurate diagnoses and better sustainability. At the same time, artificial intelligence may mean fewer jobs, more surveillance, greater inequality and autonomous weapons.

There is a lot at stake. It is therefore necessary to continuously address and discuss what is responsible and desirable development and what should be done to shape the technology. The so-called *missions* that the EU and UK have introduced may be an important instrument to achieve desirable and responsible development.[142]

[140] The French and Finnish AI strategies also make the same suggestion, referring to examples such as Personaldata.ai, Personal Information Management Systems (PIMS) and MyData.org. See also Villani 2018, page 31, and Finland's Ministry of Economic Affairs and Employment 2017, pages 44 and 45.
[141] World Economic Forum 2017.
[142] Mazzucato 2018 and May 2018.

9. ETHICAL GUIDELINES

The government has declared that it will develop guidelines and ethical principles for the use of artificial intelligence.[143] This is a good idea. The possibility of creating machines that learn, interpret and take decisions raises many ethical questions.

In the long term, it may be necessary to consider whether machines can be given a moral status, and it could be possible that machines come to achieve superintelligence and become an existential threat to humans.[144] But such perspectives assume algorithms and physical preconditions that do not exist today.

This report has illustrated some of the ethical dilemmas that already exist and that will become increasingly amplified. Traditional European values such as dignity, autonomy, freedom, solidarity, equality, democracy and trust are being challenged in pace with digitisation in general and with the development of artificial intelligence in particular.[145] The government should begin to develop ethical guidelines and practices in areas where the technology is already exerting tremendous pressure on established values:

- *Autonomy* for humans in their interface with technology. Machine learning makes predictions of the individual's behaviour and preferences more accurately and inexpensively than before. It therefore becomes possible to influence and manipulate actions and attitudes as well.

- *Democracy*. The potential for political manipulation through the use of artificial intelligence is manifest. The British consulting firm Cambridge Analytica used Facebook as a data provider to create a psychological profile of several million private individuals and as a platform to offer political influence to customers. The manipulation of media with personalised fake news also undermines democratic values. The Chinese AI strategy is unambiguous in its objectives to use machine learning to achieve social control.

- *Justice*. There is increasing asymmetry between the individual person and companies or authorities that have a large amount of data that they use for

[143] Høyre 2018.
[144] Bostrom and Yudkowsky 2016.
[145] EDPS 2018, page 16.

analysis and influence. This becomes amplified since network effects give large commercial actors near monopolies within their areas.

- *Equality*. Machines learn from data collected about society. They can reflect historically biased conditions and thereby solidify discriminatory decisions and lead to discrimination. Personalisation is in principle also discriminatory.

- *Solidarity*. Welfare systems such as health services and social security, and various forms of insurance are based on mutual sharing of risk. Increasing personalisation and hyperindividualisation through risk scoring and prediction for every individual citizen can undermine this.

- *Responsibility*. The fact that machines gain more autonomy with artificial intelligence can obscure the underlying principle that people must always be responsible for decisions that affect other people. The algorithms may be slightly opaque and difficult to understand, which makes it difficult both to anchor responsibility for decisions and to appeal the decisions. It may also be impossible to know whether you are in contact with a machine or a human. The growth of intelligent weapon systems with a high potential for autonomy pushes the question of responsibility to the extreme.

10. RIGHT TO AN EXPLANATION

Machine learning algorithms provide advice and increasingly take decisions in areas of major significance to peoples' lives, such as loan and job applications, medical diagnoses and in police matters. Here, too, it becomes important to be able to obtain an explanation of the decision so that it is possible to appeal or change an unfair practice.

The European General Data Protection Regulation (GDPR) provides data subjects with the right not to be involved in a decision that is based exclusively on automated handling when this decision has a significant impact on the individual in question.[146] However, while the right to an explanation of automated

[146] Proposal to Norwegian Parliament 2018, page 68.

decisions is addressed in the recitals, it is not mentioned in the regulation itself. It is thereby not legally binding.[147]

Norwegian authorities should therefore adopt and specify such a right to explanation.[148]

This right should include two types of explanation; i.e. how the system works (purpose, logic and consequences) as well as explanation of the individual circumstances that led to a decision.[149] What constitutes an adequate explanation in different contexts should also be clarified.

The possibility of explanation of the algorithm may be limited because the algorithm is complicated and difficult to explain in comprehensible everyday terms. It may be particularly challenging to provide an explanation of how specific data have been weighted in algorithms based on neural networks.

We may potentially have to consider whether the public sector should abstain from making automated decisions unless it is possible to provide adequate explanation. The French strategy considers it unthinkable to accept decisions that cannot be explained in areas of critical importance to a person's life, such as access to credit, work, housing, the legal system and medical services.[150]

11. OPEN ALGORITHMS IN PUBLIC SECTOR

When machines take over tasks that were previously carried out by humans, it is especially important to show that the algorithms do not make biased recommendations. Algorithms may in the worst case amplify social differences, lead to unintentional discrimination and conceal normative choices.

As a general rule, Norwegian authorities should therefore require that all algorithms used by the public sector be open to audit, so that other actors in society

---

[147] Wachter, Mittelstadt and Floridi 2017.
[148] Cf. Villani 2018, page 125.
[149] The purpose might be, for example, to determine a credit score, the logic may be the data type, properties in the data and categories in a decision tree and the consequences might be that the credit score is used by the lender to perform a credit assessment that may affect the interest rate. The individual circumstances may be the actual credit score, which actual data or properties were used and how these were weighted in the decision tree or model. See also Wachter, Mittelstadt and Floridi 2017.
[150] Villani 2018, pages 1156-116.

can verify that they are being used correctly and ethically. This will also be important for trust in the public sector. [151]

12. AUDIT ALGORITHMS

Requirements for open algorithms may not necessarily apply in all contexts. Business interests, personal data protection, or national security may be compromised if some types of algorithms can be copied and distributed freely. Of particular cause for concern is the fact that open algorithms may also potentially strengthen actors with malicious intent.

Algorithms for machine learning that for critical reasons cannot be open to the public should nonetheless be subject to evaluation before they can be put into broad use in society. One possibility is to require that closed AI algorithms be thoroughly tested and reviewed or certified by an independent third party before they can be used in society.[152] Such assessments should include whether the decisions of the algorithm are

- fair,
- correct,
- explainable,
- verifiable and
- that a means of appealing undesirable outcomes is made evident.

It is not always necessary, useful or possible to examine source code. One alternative is to test the algorithm. To evaluate whether, for example, a recruitment algorithm will discriminate, it can be tested with a large number of CVs of men and women with equal qualifications. It may therefore be appropriate to require a programming interface[153], to be able to test the algorithm on a large number of fictive users. [154]

---

[151] The French President has said that France will increase the pressure on private actors for them to make their algorithms open to audit as well. Thompson 2018.
[152] Great Britain is in the process of establishing *the Centre for Data Ethics & Innovation*, which among other things will evaluate various tools to identify and manage biased algorithms and make recommendations for tools that the private and public sector should use. House of Commons 2018.
[153] Application Programming Interface, often shortened to API
[154] See also Villani 2018, page 117.

There are already existing mechanisms, for example prior to revision, that can be expanded to include algorithms.[155]

13. ETHICS BY DESIGN

Algorithms can be checked by opening them up to audit or review, but the most suitable thing for both developers and users alike is to build ethical considerations in from the start. Undesirable events such as biased or unfair decisions can lead to a breakdown in trust that would be difficult and costly to correct afterwards.

Such thinking has already been established with respect to personal data protection.[156] *Privacy by design* means that the principles of personal privacy protection, rights, and requirements are incorporated throughout the entire developmental cycle, from design and coding to testing and operation.[157].

Artificial intelligence is now in the process of becoming important in many areas of society, and ethical considerations should therefore be built into the development of algorithms. It will be important to evaluate whether the algorithm can lead to discrimination and whether it is robust against manipulation. In this line of thinking, the proposal has now been made in France to expand the duty to carry out personal data protection assessments to also include discrimination assessments.[158]

By continuously making ethical assessments, developers can make changes along the way, while they can at the same time demonstrate how they have taken the necessary measures in an audit. Such evaluations demand that developers have or adopt ethics expertise. Ethics should be an integrated part of the education so that the developers are able to identify and manage moral questions arising from the system they create.

---

[155] There are already companies specialising in the auditing of algorithms, such as O'Neil Risk Consulting: http://www.oneilrisk.com/.
[156] Norwegian Data Protection Authority 2018b.
[157] Norwegian Data Protection Authority 2018c.
[158] Villani 2018, page 121.

14. NATIONAL DIALOGUE ON AI

Artificial intelligence is beginning to narrow in on the lives of most Norwegians. The algorithms will provide recommendations on important life events and administrative decisions. They have the potential to personalise, create filter bubbles and influence behaviour. The technology can also amplify differences and exclusion in society, although it can also be used to reduce these problems.

The Chinese strategy concludes with a point about "guiding opinion" towards the acceptance of artificial intelligence.[159] When rapid technological development affects peoples' lives and values in this way, working for passive acceptance is not enough. The Norwegian authorities should actively take initiatives to involve lay people and civil society in the discussion on artificial intelligence, and they should be receptive to their perspectives on what developments people would hope to see. This can build on principles of responsible research and innovation (RRI):

- *Dialogue-based:* The government should promote a broad exchange of ideas across various fields and social groups. There are established methods in this area such as conferences for the general public, open hearings, public summits and citizen panels, in addition to online consultations. Relevant issues for discussion include, for example, what decisions machines will take, or what goals the government shall set for its research investment.

- *Forward-looking:* Use of scenarios connected with open foresight processes are particularly relevant when the development has such speed and is marked by uncertainty. The government should take the initiative for future analyses that also open up alternative paths of development arising from questions such as "what if?" and "can we achieve the same thing in another way?"

- *Responsive:* Consultation with the public is worthless if the government does not respond. The government should routinely review the national strategy and invite input. It should also publish an annual knowledge status report.

[159] China's State Council 2017.

REFERENCES

Agrawal, Ajay, Gans, Joshua & Goldfarb, Avi (2018). *Prediction Machines, The Simpe Economics of Artificial Intelligence*. Harvard Business Review Press.

Ahlqvist et. al. (2018). *Novel subgroups of adult-onset diabetes and their association with outcomes,* The Lancet.
Retrieved from: https://doi.org/10.1016/S2213-8587(18)30051-2.

Al-Darwish, Maryam (2018, April 7). *Machine Learning*.
Retrieved from: http://www.contrib.andrew.cmu.edu/~mndarwis/ML.html.

AliveCOR (2017, march 16). *AliveCor Unveils First AI-Enabled Platform for Doctors to Improve Stroke Prevention Through Early Atrial Fibrillation Detection*.
Retrieved from: https://www.alivecor.com/press/press_release/alivecor-unveils-first-ai-enabled-platform-for-doctors/.

Aabakken, Lars (2009, February 13). Angiodysplasi, *in Store Medisinske Leksikon*. Retrieved from: https://sml.snl.no/angiodysplasi

Antoni, Manfred & Schnell, Rainer (2017). *The Past, Present and Future of the German Record Linkage Center (GRLC)*, Journal of Economics and Statistics.

Artificial Lawer (2017, October 16). *UK Gov-Backed Report Calls for AI Data Trusts; Praises Legal Sector*.
Retrieved from: https://www.artificiallawyer.com/2017/10/16/uk-gov-backed-report-calls-for-ai-data-trusts-praises-legal-sector/.

Bell, Robert M; Koren, Yehuda & Volinsky, Chris. (2018). *The BellKor solution to the Netflix Prize*.
Retrieved from: https://www.netflixprize.com/assets/Progress-Prize2007_KorBell.pdf.

Bjørkeng, Per Kristian (2015, May 9). Datasystemer som holder elevene i flytsonen, *In Aftenposten*.
Retrieved from:: https://www.aftenposten.no/norge/i/qwVL/Datasystemet-som-holder-elevene-i-flytsonen.

Bostrom, Nick & Yudkowsky, Eliezer (2016). *The Ethics of Artificial Intelligence,* Cambridge Handbook of Artificial Intelligence, Cambridge University Press.

Brownlee, Jason (2016, July 14). *8 Inspirational Applications of Deep Learning*, Machine Learning Mastery.
Retrieved from: https://machinelearningmastery.com/inspirational-applications-deep-learning/.

Brundage, Miles; Avin, Shahar; Clark, Jack; Toner, Helen & Eckersley, Peter (2018, February). *The malicious use of artificial forecasting, prevention, and mitigation.*
Retrieved from: https://www.eff.org/files/2018/02/20/malicious_ai_report_final.pdf.

Brynjolfsson, Erik; McAfee, Andrew (2017a, July 18). *The business of Artificial Intelligence,* Harvard Business Review.
Retrieved from: https://hbr.org/cover-story/2017/07/the-business-of-artificial-intelligence.

Brynjolfsson, Erik & McAfee, Andrew (2017b, July 18). *What's Driving the Machine Learning Explosion?,* Harvard Business Review.
Retrieved from: https://hbr.org/2017/07/whats-driving-the-machine-learning-explosion.

CIFAR (2017, August 20). *Pan-Canadian Artificial Intelligence Strategy*. CIFAR.
Retrieved from: https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy.

Castelvecchi, Davide (2016, September 27). Deep learning boosts Google translate tool, *in Nature*, 2016.
Retrieved from: https://www.nature.com/news/deep-learning-boosts-google-translate-tool-1.20696.

Carey, Bjorn (2016, 24. August). Smartphone speech recognition can write text messages three times faster than human typing*, in Stanford News*.
Retrieved from: http://news.stanford.edu/2016/08/24/stanford-study-speech-recognition-faster-texting/.

Christiansen, Atle (2017, February 24). Kunstig intelligens kan hjelpe leger*, from uia.no*.
Retrieved from: https://www.uia.no/nyheter/kunstig-intelligens-kan-hjelpe-leger.

Cuomo, Chris; Shaylor, Jay; McGuirt, Mary & Francescani, Chris (2009, January 28). 'GMA' Gets Answers: Some Credit Card Companies Financially Profiling Customers, *from ABC News*.
Retrieved from: http://abcnews.go.com/GMA/TheLaw/gma-answers-credit-card-companies-financially-profiling-customers/story?id=6747461.

Dagens næringsliv (2017, December 7). Rekrutterer ved hjelp av robot – ansetter flere kvinner, *in Dagens næringsliv*.
Retrieved from: Dagens næringsliv 2017. https://www.dn.no/talent/2017/12/07/0649/Arbeidsliv/rekrutterer-ved-hjelp-av-robot-ansetter-flere-kvinner.

Datatilsynet (2018a). *Kunstig intelligens og personvern*.
Retrieved from: https://www.datatilsynet.no/globalassets/global/om-personvern/rapporter/rapport-om-ki-og-personvern.pdf.

Datatilsynet (2018b). *Innebygd personvern*.
Retrieved from: https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/innebygd-personvern/.

Datatilsynet (2018c) *Programvareutvikling med innebygd personvern*.
Retrieved from: https://www.datatilsynet.no/regelverk-og-verktoy/veiledere/programvareutvikling-med-innebygd-personvern/?id=7729.

Department for Business, Energy and Industrial Strategy & Department for Digital, Culture, Media and Sport UK (2018). *AI Sector Deal, Policy Paper* (26).
Retrieved from: https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal.

Dockrill, Peter (2017, December 8). Google's AI has mastered all the chess knowledge in history – in just 4 hours. *In World Economic Forum*.
Retrieved from: https://www.weforum.org/agenda/2017/12/google-s-ai-has-mastered-all-the-chess-knowledge-in-history.

Echersley, Peter & Nasser, Yomna (2018, March 7). Measuring the progress of AI research*, in Electonic Frontier Foundation*.
Retrieved from: https://www.eff.org/ai/metrics#Vision.

European Data Protection Supervisor (2018, January 25). *Towards a digital ethics*.
Retrieved from: https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf.

Equinor (2018, March 7). *Joint offshore digitalisation*.
Retrieved from: https://www.equinor.com/en/news/07mar2018-joint-offshore-digitalisation.html.

Esteva; Andre, Kuprel, Brett; Novoa, Roberto A.; Ko, Justin; Swetter, Susan M.; Blau, Helen M. & Thrun, Sebastian (2017). Dermatologist-level classification of skin cancer with deep neutral networks*, in Nature (*542), p. 115-118.
Retrieved from: https://www.nature.com/nature/journal/v542/n7639/index.html.

Etherington, Darrell (2017, February 8). Udacity open sources its self-driving car simulator for anyone to use, *in TechCrunch*.
Retrieved from: https://techcrunch.com/2017/02/08/udacity-open-sources-its-self-driving-car-simulator-for-anyone-to-use/.

EU Commission (2018a) *Artificial Intelligence for Europe*.
Retrieved from: http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51625.

EU commission (2018, April 25): *Artificial intelligence: Commission outlines a European approach to boost investment and set ethical guidelines*.
Retrieved from: http://europa.eu/rapid/press-release_IP-18-3362_en.htm.

EU commission (2018, April 10) *EU Member States sign up to cooperate on Artificial Intelligence*.
Retrieved from: https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence.

EU commission (2018, February 22). *Bold science to meet big challenges: independent report calls for mission-oriented EU research and innovation*.
Retrieved from: https://ec.europa.eu/info/news/bold-science-meet-big-challenges-independent-report-calls-mission-oriented-eu-research-and-innovation-2018-feb-22_en.

European parliament (2017, February 16). *Civil Law Rules on Robotics*.
Retrieved from: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+P8-TA-2017-0051+0+DOC+PDF+V0//EN.

Evans; Richard; Gao, Jim (2016, July 20). DeepMind AI Reduces Google Data Centre Cooling Bill by 40%, *in Deepmind.com*.
Retrieved from: https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/.

Fagella, Daniel (2016, May 18). Unsupervised Machine Learning Could Help Us Solve the Unsolvable, *in Huffington Post*.
Retrieved from: https://www.huffingtonpost.com/daniel-faggella/unsupervised-machine-lear_b_10010452.html.

Federal Ministry of Transport and Digital Infrastructure (In Germany) (2017, August 28). *Ethics commission's complete report on automated and connected driving*.
Retrieved from: https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf.

Fenner, Robert (2018, January 15). Alibaba's AI outguns humans in reading test, *in Bloomberg*.com.
Retrieved from: https://www.bloomberg.com/news/articles/2018-01-15/alibaba-s-ai-outgunned-humans-in-key-stanford-reading-test.

Ferrucci, David (2018, March 7). A computer called Watson, *from IBM Icons of Progress*.
Retrieved from: http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/watson/breakthroughs/.

Ministry of Economic Affairs and Employment (Finland) (2017) *Finland's Age of Artificial Intelligence* (47). Retrieved from: http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf.

Fischer, Astrid; Olhoff-Jacobsen, Emil Eusebius (2017, December 27). Computer kan forudsige om du havner akut på sygehuset. *From Danmarks Radio*.
Retrieved from: https://www.dr.dk/nyheder/indland/computer-kan-forudsige-om-du-havner-akut-paa-sygehuset.

Folkehelseinstituttet (2018, January 24). *Kreft I Norge*, Folkehelserapporten (2018).
Retrieved from: https://www.fhi.no/nettpub/hin/ikke-smittsomme/kreft-i-norge-folkehelserapporten/.

China's State Council (2017, July 20). *A Next Generation Artificial Intelligence Development Plan*. Based on an English translation by *New America*.
Retrieved from: http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm & https://na-production.s3.amazonaws.com/documents/translation-fulltext-8.1.17.pdf.

Gallup (2018, March 6). Most Americans Already Using Artificial Intelligence Products*, in Gallup*.
Retrieved from: https://news.gallup.com/poll/228497/americans-already-using-artificial-intelligence-products.aspx.

Garber, Megan (2016, June 30). When Algorithms Take the Stand, *in The Atlantic*.
Retrieved from: https://www.theatlantic.com/technology/archive/2016/06/when-algorithms-take-the-stand/489566/.

GDPR (2016). *Official Journal (L119) of the European Union* (59).
Retrieved from: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2016:119:FULL.

Geng, Daniel & Shih, Shannon (2017, February 4). *Machine Learning Crash Course: Part 3*, Machine Learning at Berkeley.
Hentet fra: https://ml.berkeley.edu/blog/2017/02/04/tutorial-3/.

Gibney, Elizabeth (2016, January 28). Google AI algorithm master's ancient game of go, *in Nature* (529), pp. 445-336.

Github (2018). Mimir, *in github.com*.
Retrieved from: https://github.com/Stevenah/mimir.

Goodfellow, Ian J; Pouget-Abadie, Jean; Mirza, Mehdi; Zu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua (2014, June 10). *Generative Adversial Networks*.
Retrieved from: https://arxiv.org/abs/1406.2661.

Goodfellow, Ian J; (2017). NIPS 2016 – Generative Adversarial Networks, *from YouTube*, (1) pp. 55:53.
Retrieved from: https://www.youtube.com/watch?v=AJVyzd0rqdc.

Gunning, David (2018). Explainable artificial intelligence (XAI), *from Defence Advances Research Projects Agency*.
Retrieved from: https://www.darpa.mil/program/explainable-artificial-intelligence.

Gupta, Dishashree (2017, June 1). Transfer learning & the art of using pre-trained models in deep learning, *from Analytics Vidhya*.
Retrieved from: https://www.analyticsvidhya.com/blog/2017/06/transfer-learning-the-art-of-fine-tuning-a-pre-trained-model/.

Hall-Geisler, Kristen (2017, July 13). Cortica teaches unsupervised vechicles with unsupervised learning, *from TechCrunch.com*.
Retrieved from: https://techcrunch.com/2017/07/13/cortica-teaches-autonomous-vehicles-with-unsupervised-learning/.

Hall, Wendy; Pesenti Jérôme (2017). *Growing the artificial intelligence industry in the UK*.
Retrieved from: https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk.

Hassabis, Demis & Silver, David (2017, October 12). AlphaGO Zero: Learning from Scratch, *from Deepmind*.
Retrieved from: https://deepmind.com/blog/alphago-zero-learning-scratch/.

Haugnes, Gunhild M (2017, May 2). Avslører magesykdom med algoritmer, *in Titan, uio.no*.
Retrieved from: https://titan.uio.no/node/2296.

Hawkins, Andrew J (2018, May 9). Inside Waymo's strategy to grow the best brains for self-driving cars, *in The Verge*.
Retrieved from: https://www.theverge.com/2018/5/9/17307156/google-waymo-driverless-cars-deep-learning-neural-net-interview.

Hof, Robert D (2018). Deep Learning, in *MIT Technology Review*.
Retrieved from: https://www.technologyreview.com/s/513696/deep-learning/.

Holte, Hans Christian (2018, June 22). Digitalisering – et samarbeid om sammenhenger som burde vært der*, in Ukeavisen Ledelse*.

House of Commons (UK) (2018, May 23). *Algorithms in decisionmaking,* Fourth report of Session, (351).
Retrieved from: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf.

House of Commons (UK) (2017, September 13). *Robotics and Artificial Intelligence: Fifth Report of Sessions 2016-2017*. House of Commons Science and Technology Committee.
Retrieved from: https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/896/896.pdf.

Høyre (2018, January 14). *Enighet om blågrønn regjeringsplattform*.
Retrieved from: https://hoyre.no/aktuelt/nyheter/2018/enighet-om-blaagronn-regjeringsplattform/.

IBM (2018, March 7). 10 Marketing Trends for 2017 and Ideas for Exceeding Costumer Expectations, *in IBM.com*.
Retrieved from: https://public.dhe.ibm.com/common/ssi/ecm/wr/en/wrl12345usen/watson-customer-engagement-watson-marketing-wr-other-papers-and-reports-wrl12345usen-20170719.pdf.

Jones, Nicola (2017, August 24). Machine learning tapped to improve climate forecasts, *in Nature (*548).
Retrieved from: https://www.nature.com/polopoly_fs/1.22503!/menu/main/topColumns/topLeftColumn/pdf/548379a.pdf.

Kaggle (2016, August 1). Top-100 Kaggle users by Country, *in Kaggle*.
Retrieved from: https://www.kaggle.com/andreyg/top-100-kaggle-users-by-country

Kaggle (2017, April 20). Data Science FAQ, *in Kaggle*.
Retrieved from: https://www.kaggle.com/rounakbanik/data-science-faq.

Karpathy, Andrej (2014, September 2). *What I learned from competing against a ConvNet on ImageNet*.
Retrieved from: http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/.

Knight, Will (2017, January 4). 5 big predictions for artificial intelligence in 2017, *in MIT Technology Review*.
Retrieved from: https://www.technologyreview.com/s/603216/5-big-predictions-for-artificial-intelligence-in-2017/.

Knight, Will (2017, April 11). The dark secret at the heart of AI, *in MIT Technology Review*.
Retrieved from: https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.

Knight, Will (2018, March 7). By experimenting, computers are figuring out how to do things that no programmer could teach them, *in MIT Technology Review*.
Retrieved from: https://www.technologyreview.com/s/603501/10-breakthrough-technologies-2017-reinforcement-learning/.

Kreftforeningen (2018, August 17). *Slik kan du sjekke føflekkene dine*.
Retrieved from: https://kreftforeningen.no/om-kreft/a-oppdage-kreft-tidlig/sjekk-foflekkene-dine/.

Kunnskapsdepartementet (2014) *Langtidsplan for forskning og høyere utdanning 2015–2024*.
Retrieved from: https://www.regjeringen.no/contentassets/e10e5d5e2198426788ae4f1ecbbbbc20/no/pdfs/stm201420150007000dddpdfs.pdf.

Kunnskapsdepartementet (2017) *Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata*. Retrieved from: https://www.regjeringen.no/contentassets/3a0ceeaa1c9b4611a1b86fc5616abde7/no/pdf/f-4442-b-nasjonal-strategi.pdf.

Lardinois, Frederic (2018, August 17). Google gives its AI the reins over its data center cooling systems, *in TechChrunch*.
Retrieved from: https://techcrunch.com/2018/08/17/google-gives-its-ai-the-reins-over-its-data-center-cooling-systems/?guccounter=1.

Liptak, Adam (2017, May 1). Sent to prison by a software program's secret algorithms, *in The New York Times*.
Retrieved from: https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html.

Lui, Yunjie; Racah, Evan; Correa, Joaquin Prabhat; Khosrowshahi, Amir; Lavers, David; Kunkel, Kenneth; Wehner, Michael & Collins, Will iam (2016, May 4). Application of deep convolutional neutral networks for detecting extreme weather in climate datasets, *in Arxiv.org*.
Retrieved from: https://arxiv.org/pdf/1605.01156.pdf.

Mandaric, Stefan & Axelsen, Vebjørn (2017). Forebygging av trafikkulykker ved bruk av avansert dataanalyse, *from vegvesenet.no*.
Retrieved from: https://www.vegvesen.no/_attachment/2073052/binary/1219064?fast_title=Forebygging+av+trafikkulykker+ved+bruk+av+avansert+dataanalyse.pdf.

Mannes, John (2016, December 5). OpenAI's Universe is the fun parent every artificial intelligence deserves, *from TechCrunch.com*.
Retrieved from: https://techcrunch.com/2016/12/05/openais-universe-is-the-fun-parent-every-artificial-intelligence-deserves/.

Margolin, Madison (2016, November 2). These glasses fool facial recognition into thinking you're someone else, *from Motherboard.vice.com*.
Retrieved from: https://motherboard.vice.com/en_us/article/pgkxgv/glasses-fool-facial-recognition.

May, Theresa (2018, May 21). *PM speech on science and modern Industrial Strategy*.
Retrieved from: https://www.gov.uk/government/speeches/pm-speech-on-science-and-modern-industrial-strategy-21-may-2018.

Mazzucato, Mariana (2018) *Mission-oriented Research and Innovation in the European Union*.
Retrieved from: https://publications.europa.eu/en/publication-detail/-/publication/5b2811d1-16be-11e8-9253-01aa75ed71a1/language-en.

Melding til Stortinget (2017) *Perspektivmeldingen 2017, (*Meld. St. 29).
Retrieved from: https://www.regjeringen.no/no/dokumenter/meld.-st.-29-20162017/id2546674/.

Miotto, Riccardo; Li; Li, Kidd, Brian A & Dudley, Joel T (2016). Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records, *in Nature Scientific Reports*.
Retrieved from: https://www.nature.com/articles/srep26094.

Moe, Sigrid & Breivik, Steinar Rostad (2018, March 5). Snart kan kunstig intelligens styre vannkraftverkene, *in E24*.
Retrieved from: https://e24.no/energi/vannkraft/snart-kan-kunstig-intelligens-styre-vannkraftverkene/23933863.

Wikipedia (2018). Multi-task learning.
Retrieved from: https://en.wikipedia.org/wiki/Multi-task_learning.

Murgia, Madhumita (2017, January 4). NHS to trial artificial intelligence app in place of 111 helpline, *in Financial Times*.
Retrieved from: https://www.ft.com/content/aefee3b8-d1d8-11e6-b06b-680c49b4b4c0.

Mystore (2017, May 12). Kunstig intelligens og dens rolle i netthandel, *in Mystore.no*. Retrieved from: https://www.mystore.no/kunstig-intelligens-netthandel/.

Nedelkoska, L. & Quintini, G (2018). *Automation, skills use and training,* OECD Social, Employment and Migration Working Papers, (202).
Retrieved from: http://www.oecd-ilibrary.org/employment/automation-skills-use-and-training_2e2f4eea-en.

NFR (2017, September). Innspill til revidert langtidsplan 2019 – 2022 & Innspill til Regjeringens langtidsplan for forskning, *in regjeringen.no*.
Retrieved from: https://www.regjeringen.no/contentassets/8e11ef7f053e43a0a7ac06f2486e16c7/forskningsradet---innspill-til-revisjon-av-langtidsplanen-for-forskning-og-hoyere-utdanning-2015-2024-signert.pdf.

Ng, Andrew (2015, November 24). What Data Scientists Should Know About Deep Learning, *Presentation at the Extract Data Conference,* 24. November, 2015. Retrieved from: https://www.slideshare.net/ExtractConf.

Ng, Andrew (2017, July 25). Deep learning's next frontier, *in Harvard Business Review*.
Retrieved from: https://hbr.org/2017/07/deep-learnings-next-frontier.

Niesche, Christophe (2015, august 18). Medtech startup turns home into medical clinic, *in Australia Unlimited*.
Retrieved from: https://www.australiaunlimited.com/technology/medtech-startup-turns-home-into-medical-clinic.

OECD (2017, October). *OECD Digital Economy Outlook*. Retrieved from: http://www.oecd.org/internet/oecd-digital-economy-outlook-2017-9789264276284-en.htm.

O'Neil, Cathy (2016) *Weapons of Mass Destruction*. New York: Crown Publishers.

O'Neil, Cathy (2018, July 3). *Audit the algorithms that are ruling our lives. Governments should follow France and move towards algorithmic accountability,* Financial Times.
Retrieved from: https://www.ft.com/content/879d96d6-93db-11e8-95f8-8640db9060a7.

Ong, Thuy (2017, August 4). Facebook's translations are now powered completely by AI, *in The Verge.*
Retrieved from: https://www.theverge.com/2017/8/4/16093872/facebook-ai-translations-artificial-intelligence.

OpenAI.com (2017, February 24). Attacking machine learning with adversarial examples, *in OpenAI.com*.
Retrieved from: https://blog.openai.com/adversarial-example-research/.

Oren Etzioni (2017, September 1). How to regulate artificial intelligence, *in The New York Times*.
Retrieved from: https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html.

Pajarinen, M., Rouvinen, P. & Ekeland, A (2014). *Computerization and the Future of Jobs in Norway*. ETLA and SSB, 2014.
Retrieved from: http://nettsteder.regjeringen.no/fremtidensskole/files/2014/05/Computerization-and-the-Future-of-Jobs-in-Norway.pdf.

Paysa.com (2017, November 29). *New Paysa Study Reveals U.S. Companies Across All Industries Investing $1.35 Billion Dollars in AI Talent*.
Retrieved from: https://www.paysa.com/press-releases/2017-11-29/11/new-paysa-study-reveals-us.

Pogorelov, Konstantin et.al (2018). *Automatic Detection of Angiectasia: Evaluation of Deep Learning and Handcrafted Approaches*, IEEE Conference on Biomedical and Health Informatics (2018).
Retrieved from: http://home.ifi.uio.no/paalh/publications/files/bhi-2018.pdf

Regjeringen (2018). *Lov om behandling av personopplysninger (personopplysningsloven),* (Prop. 56 LS).
Retrieved from: https://www.regjeringen.no/contentassets/1a36e88f124d4a1ea92a9c790be2d69a/no/pdfs/prp20172018005600odddpdfs.pdf.

Accenture (2016). *Why artificial intelligence is the future of growth*.
Retrieved from: https://www.accenture.com/t20170927T080049Z__w__/us-en/_acnmedia/PDF-33/Accenture-Why-AI-is-the-Future-of-Growth.PDFla=en.

Reilly, Michael (2017, August 23). Climate-change research is getting a big dose of AI, *in Technology Review*.
Retrieved from: https://www.technologyreview.com/the-download/608726/climate-change-research-is-getting-a-big-dose-of-ai/.

Reuters (2018, March 29). France to spend $1.8 billion on AI to compete with U.S., China, *in Reuters*.
Retrieved from: https://www.reuters.com/article/us-france-tech/france-to-spend-1-8-billion-on-ai-to-compete-with-u-s-china-idUSKBN1H51XP.

Ribeiro, Marco Tulio; Singh, Sameer & Guestrin, Carlos (2016, August 12). Introduction to local interpretable model-agnostic explanations (LIME), *from O'Reilly.com*.
Retrieved from: https://www.oreilly.com/learning/introduction-to-local-interpretable-model-agnostic-explanations-lime.

Royal Free (2018, August 18). *Our work with DeepMind*.
Retrieved from: https://www.royalfree.nhs.uk/patients-visitors/how-we-use-patient-information/our-work-with-deepmind/.

Ruan, Sherry; Wobbrock, Jacob O.; Liou, Kenny; Ng, Andrew & Landay, James (2017). *Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones*. Journal Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies archive. 1 (4).
Retrieved from: https://doi.org/10.1145/3161187.

Scimago (2018) Article in «Computer Science» and «Artificial Intelligence», from *Scimago Journal and Country Rank*.
Retrieved from: http://www.scimagojr.com/countryrank.php?area=1700&category=1702.

Sinovation Ventures (2018) *China embraces AI: A Close Look and A Long View*.
Retrieved from: https://www.eurasiagroup.net/files/upload/China_Embraces_AI.pdf.

SkillsFuture (2017, October 5). *Launch Of SkillsFuture For Digital Workplace - Building Confidence in All Singaporeans for the Digital Economy*.
Retrieved from: http://www.ssg-wsg.gov.sg/new-and-announcements/05_Oct2017.html.

SkLNytt (2017) *Virkelig noe å være redd for?* SkLNytt, 5 (40).
Retrieved from: https://www.skl.no/sfiles/31/41/1/file/sklnytt_5_17_ferdig_lavoppl.pdf.

Smith, Mich (2016, June 22*)*. In Wisconsin, a backlash against using data to foretell defendants' futures, *in The New York Times*.
Retrieved from: https://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html.

Snow, Jackie (2018, March 7). Most Americans are already using AI, *in MIT Technology Review, Oxford Insights.*
Retrieved from: https://www.technologyreview.com/the-download/610438/most-americans-are-already-using-ai/.

Stirling, Miller & Martinho-Truswell (2018). *Government AI Readiness Index.*
Retrieved from: https://www.oxfordinsights.com/government-ai-readiness-index/.

Sverdlik, Yevgeniy (2018, August 2). Google is Switching to a Self-Driving Data Center Management System*, in Data Center Knowledge*.
Retrieved from: https://www.datacenterknowledge.com/google-alphabet/google-switching-self-driving-data-center-management-system.

Tassev, Lubomir (2018, February 15). *GPU shortage hinders scientific research – cryptocurrency miners blamed,* from Bitcoin.com.
Retrieved from: https://news.bitcoin.com/gpu-shortage-hinders-alien-search-cryptocurrency-miners-blamed/.

Telenor (2018, August 15). *Styrker kraftsenter for kunstig intelligens i Norge*.
Retrieved from: https://www.telenor.com/no/pressemelding/styrker-kraftsenter-for-kunstig-intelligens-i-norge/.

Teknologirådet (2017) *Denne gangen er det personlig*.
Retrieved from: https://teknologiradet.no/wp-content/uploads/sites/19/2013/08/Rapport_Denne-gangen-er-det-personlig.-Det-digitale-skiftet-i-offentlig-sektor.pdf.

The Economist (2017, December 7). *Google leads race to dominate artificial intelligence*.
Retrieved from: https://www.economist.com/news/business/21732125-tech-giants-are-investing-billions-transformative-technology-google-leads-race.

The Economist (2018a, February 15). *For artificial intelligence to thrive, it must explain itself.*
Retrieved from: https://www.economist.com/news/science-and-technology/21737018-if-it-cannot-who-will-trust-it-artificial-intelligence-thrive-it-must.

The Economist (2018b, Februar7 15). *Humans may not always grasp why AIs act. Don't panic.*
Retrieved from: https://www.economist.com/leaders/2018/02/15/humans-may-not-always-grasp-why-ais-act.-dont-panic.

Thompson, Nicolas (2018, March 31). Emmanuel Macron talks to Wired about France´s AI Strategy, *in Wired*.
Retrieved from: https://www.wired.com/story/emmanuel-macron-talks-to-wired-about-frances-ai-strategy/.

Thornhill, John (2017, October 30). Would you donate your data for the collective good?, *in Financial Times*.
Retrieved from: https://www.ft.com/content/00390a76-bd4a-11e7-9836-b25f8adaa111.

Turner, Karen (2016, October 3). Google Translate is getting really, really accurate, *in The Washington Post*.
Retrieved from: https://www.washingtonpost.com/news/innovations/wp/2016/10/03/google-translate-is-getting-really-really-accurate/.

Tørresen, Jim (2014). *Hva er kunstig intelligens*, Universitetsforlaget.
Retrieved from: https://issuu.com/universitetsforlaget/docs/hva_er_kunstig_intelligens.

UCLH (2018, May 21). *Revolutionising healthcare with AI and data science: UCLH and The Alan Turing Institute announces breakthrough partnership today*.
Retrieved from: https://www.uclh.nhs.uk/News/Pages/RevolutionisingHealthcarewithAI.aspx.

University of Toronto (2017, March 30). *Toronto's Vector Institute officially launched*.
Retrieved from: https://www.utoronto.ca/news/toronto-s-vector-institute-officially-launched.

Villani, Cedric (2018). For a meaningful artificial intelligence. Towards a French and European strategy, *in aiforhumanity*.
Retrieved from: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf.

Vinnova (2018). *Artificiell intelligens i svenskt näringsliv och samhälle. Analys av utveckling och potential* (18).
Retrieved from:
https://www.vinnova.se/contentassets/55b18cf1169a4a4f8340a5960b32fa82/vr_18_08.pdf.

Wachter, Sandra; Mittelstadt, Brent & Floridi, Luciano (2017) *Why a right to explanation of automated decision-making does not exist in the general data protection regulation,* International Data Privacy Law, Elsevier.
Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469.

Wahed, Muntasir (2018, January). *What is the relationship between artificial intelligence, machine learning, deep learning, and artificial neutral networks?,* Quora.
Retrieved from: https://www.quora.com/What-is-the-relationship-between-artificial-intelligence-machine-learning-deep-learning-and-artificial-neural-networks.

Wakefield, Jane (2016, March 24). Microsoft chatbot is taught to swear on Twitter, *in BBC News*.
Retrieved from: http://www.bbc.com/news/technology-35890188.

World Economic Forum (2017) *The global risk report*.
Retrieved from: http://www3.weforum.org/docs/GRR17_Report_web.pdf.

Zames, Matt (2016*)* Redefining the Service Industry, *from J. P. Morgan Chase & Co*.
Retrieved from: https://www.jpmorganchase.com/corporate/annual-report/2016/ar-ceo-letter-matt-zames.htm.

Zhao, Yufan; Zeng, Donglin, Socinski, Mark & Kosorok, Michael R. (2011) *Reinforcement Learning Strategies for Clinical Trials in Non-small Cell Lung Cancer,* Journal of the International Biometric Society (2011).
Retrieved from: https://www.ncbi.nlm.nih.gov/pubmed/21385164.