White Paper

Artificial intelligence across industries

CEO

Executive summary

Artificial intelligence is currently attracting considerable interest and attention from industry, researchers, governments as well as investors, who are pouring record amounts of money into the development of new machine learning technologies and applications. Increasingly sophisticated algorithms are being employed to support human activity, not only in forecasting tasks but also in making actual decisions that impact society, businesses and individuals. Whether in the manufacturing sector, where robots are adapting their behaviour to work alongside humans, or in the home environment, where refrigerators order food supplies based on the homeowner's preferences, artificial intelligence is continuously making inroads into domains previously reserved to human skills, judgment or decision-making.

While artificial intelligence has the potential to help address some of humanity's most pressing challenges, such as the depletion of environmental resources, the growth and aging of the world's population, or the fight against poverty, the increasing use of machines to help humans make adequate decisions is also generating a number of risks and threats that businesses, governments and policy makers need to understand and tackle carefully. New concerns related to safety, security, privacy, trust, and ethical considerations in general are definitely emerging together with the technological innovations enabled by artificial intelligence. These challenges are common to all societies across the globe and will need to be dealt with at the international level.

The present White Paper provides a framework for understanding where artificial intelligence stands today and what could be the outlook for its development in the next 5 to 10 years. Based on an explanation of current technological capabilities, it describes the main systems, techniques and algorithms that are in use today and indicates what kinds of problems they typically help to solve. Adopting an industrial perspective, the White Paper discusses in greater detail four application domains offering extensive opportunities for the deployment of artificial intelligence technologies: smart homes, intelligent manufacturing, smart transportation and self-driving vehicles, and the energy sector.

The analysis of various specific use cases pertaining to these four domains provides clear evidence that artificial intelligence can be implemented across and benefit a wide set of industries. This potential is paving the way for artificial intelligence to become an essential part of the equation in resolving issues generated by today's and tomorrow's megatrends. Building upon this analysis, the White Paper provides a detailed description of some of the major existing and future challenges that artificial intelligence will have to address. While industry and the research community constitute the principal drivers for developing initiatives to tackle technical challenges related to data, algorithms, hardware and computing infrastructures, governments and regulators urgently need to elaborate new policies to deal with some of the most critical ethical and social issues foreseen to be the by-products of artificial intelligence.

Standardization and conformity assessment are expected to play an essential role not only in driving market adoption of artificial intelligence but also in mitigating some of the most pressing challenges related to decision-making by machines. As a leading organization providing a unique mix of standardization and conformity assessment capabilities for industrial and information technology systems, the IEC is ideally positioned to address some of these challenges at the international level.

The following specific recommendations targeted at the IEC and its committees are provided in the last part of the White paper:

- Promote the central role of JTC 1/SC 42 in horizontal artificial intelligence standardization.
- Coordinate the standardization of data semantics and ontologies.
- Develop and centralize artificial intelligence-related use cases.
- Develop an artificial intelligence reference architecture with consistent interfaces.
- Explore the potential for artificial intelligence conformity assessment needs.
- Foster a dialogue with various societal stakeholders concerning artificial intelligence.
- Include artificial intelligence use cases in testbeds involving the IEC.

As it is foreseen that artificial intelligence will become a core technology across many different industries and one of the driving forces of the coming fourth industrial revolution, the standardization community will play a critical role in shaping its future. Building upon its long track record in safety and reliability, the IEC can be instrumental in achieving this goal and fulfilling the promise of artificial intelligence as a benefit to humanity.

Acknowledgments

This White Paper has been prepared by the artificial intelligence project team in the IEC Market Strategy Board (MSB), with major contributions from the project partner, the German Research Centre for Artificial Intelligence (DFKI), and the project leader, Haier Group.

The project team was directed by Mr Ye Wang, Vice President of Haier Home Appliance Industry Group and an MSB member. The project team is listed below:

Mr Jens Popper, DFKI / *SmartFactory*^{KL},
Project Partner Lead

Mr Jesko Hermann, DFKI / *SmartFactory*^{KL}

Mr Kai Cui, Haier, Project Manager

Mr Simon Bergweiler, DFKI

Dr Stephan Weyer, DFKI

Prof. Dr-Ing. Martin Ruskowski, DFKI

Mr Miao Wang, Haier

Mr Liang Guang, Huawei

Mr Yun Chao Hu, Huawei

Dr Victor Kueh, Huawei

Mr Di Wang, Huawei

Ms Mary Carol Madigan, SAP

Dr Ian Oppermann, NSW Data Analytics

Mr Sung-Min Ryu, KEPCO

Mr Lagyoung Kim, LG Electronics

Dr Sha Wei, CESI

Mr Ruiqi Li, CESI

Mr Xiaohui Du, ITEI

Mr Yujia Shang, ITEI

Mr Xiangqian Ding, OUC

Mr Guangrui Zhang, OUC

Dr Gilles Thonet, IEC

Table of contents

List of abbreviations

Technical and scientific terms

AI	artificial intelligence
AIR	automated image recognition
AMI	advanced metering infrastructure
ANN	artificial neural network
API	application programming interface
ASIC	application-specific integrated circuit
CNN	convolutional neural network
CART	classification and regression tree
CPU	central processing unit
DNN	deep neural network
EBL	explanation-based learning
FPGA	field-programmable gate array
GDPR	(EU) General Data Protection Regulation
GPU	graphics processing unit
GRU	gated recurrent unit
HMM	hidden Markov model
HTM	hierarchical temporal memory
ICT	information and communication technology
ID3	Iterative Dichotomiser 3
IoT	Internet of Things
IT	information technology
k-NN	k-nearest neighbour
KPI	key performance indicator
LSTM	long-short-term memory
NLP	natural language processing
NPU	neural processing unit
RDBMS	relational database management system

ReLu	rectified linear unit
RNN	recurrent neural network
SME	small-to-medium enterprise
SVM	support vector machine
TPU	tensor processing unit

Organizations, institutions and companies

CESI	China Electronics Standardization Institute
DFKI	German Research Center for Artificial Intelligence
EC	European Commission
ENI	(ETSI ISG) Experiential Networked Intelligence
ETSI	European Telecommunications Standards Institute
FG-ML5G	(ITU-T) Focus Group on machine learning for future networks including 5G
IDC	International Data Corporation
IEEE	Institute of Electrical and Electronics Engineers
ISG	(ETSI) Industry Specification Group
ISO	International Organization for Standardization
ITEI	Instrumentation Technology and Economy Institute (China)
ITU	International Telecommunication Union
ITU-T	(ITU) Telecommunication Standardization Sector
JTC	Joint Technical Committee
KEPCO	Korea Electric Power Corporation
NHTSA	National Highway Traffic Safety Administration (US)
OUC	Ocean University of China
SAC	Standardization Administration of China
UN	United Nations

Glossary

Application programming interface
API
interface constituted of clearly defined methods of communication between various software components

Application-specific integrated circuit
ASIC
an electronic circuit specialized to perform a specific set of operations for a specific purpose

NOTE The application field cannot be changed since it is defined through its architecture.

Artificial intelligence
AI
a branch in computer science that simulates intelligent behaviour in computers including problem solving, learning and pattern recognition

Artificial neural network
ANN
a mathematical construct inspired by biological neural networks that are often used in computer science to perform tasks by giving them training examples without being explicitly programmed to do so

Central processing unit
CPU
an electronic circuit that performs instructions of a computer programme

Convolutional neural network
CNN
a special feed-forward network that is usually applied for tasks such as image recognition

Deep learning
a field of machine learning using deep neuronal networks

Deep neural network
DNN
an artificial neural network that has several consecutive layers of neurons that are connected in order to process an input to an output

Explanation-based learning
EBL
a form of artificial intelligence that uses domain theory to generalize from training examples

Field-programmable gate array
FPGA
an electronic circuit that performs specifically for different applications

NOTE In contrast to application-specific integrated circuits, FPGA can be reprogrammed after manufacturing.

General Data Protection Regulation
GDPR
a set of significant regulatory changes to data protection and privacy in the European Union, which also addresses automated decision-making by artificial intelligence systems

Graphics processing unit
GPU
an electric circuit that is specialized to process images by performing massive amounts of calculations in parallel

Hidden Markov model
HMM
a probabilistic model of linear sequences that can be described using the Markov process

NOTE Hidden Markov model is a technique used in machine learning with the assumption that not all states of the described processes can be directly observed and thus are hidden.

Internet of Things
IoT
network of physical devices, embedded electronics or software that enables these components to be connected with a larger network to exchange data

Machine learning
a category of algorithms in computer science enabling a device to improve its performance of a specific task with increasing data and without being explicitly programmed to do so

Natural language processing
NLP
an area of computer science dealing with how computers can process natural language for speech recognition, language understanding or language generation

Neural processing unit
NPU
an electric circuit that is not based on a von-Neumann or Harvard architecture, but on the principle of neuromorphing

Rectified linear unit
ReLu
an activation function of a neuron which consists of two linear parts

Recurrent neural network
RNN
a class of neural network in which the connections between the neurons form a directed graph along a sequence

Relational database management system
RDBMS
a database system that is based on the relational model

Tensor processing unit
TPU
an application-specific integrated circuit developed by Google to process machine learning and deep learning tasks

Section 1

Introduction

1.1 Artificial intelligence: miracle or mirage?

Artificial intelligence (AI) is today one of the most widely hyped technologies. Since the advent of the first computers, mathematical models have increasingly been used to support humans in an ever larger set of decision-making processes. Whether employed in the human resources area to help determine who gets hired for a job, or in the banking sector to select approved recipients for a loan, machines have been continuously making inroads into domains hitherto reserved to human judgment and adjudication.

With the digitalization of many industries making large sets of data available, AI began to be the focus of renewed interest for its potential in solving an ever increasing number of problems. Machine learning techniques grew more and more powerful and sophisticated, in particular in the context of what are known as artificial neural networks (ANNs). Developed in the middle of the 20th century as a mathematical curiosity inspired by biology, neural networks have become one of the cornerstones of AI.

However, it was not until 2010 and later that dramatic improvements in machine learning, commonly referred to as deep learning, paved the way for an explosion of AI. With computing power increasing steadily, very large ("deep") neural networks began to provide machines with novel capabilities that would have been too complex or even impossible to implement using traditional programming techniques. Since then, technologies such as computer vision and natural language processing (NLP) have been completely transformed and are being deployed on a massive scale in many different products and services. Deep learning is now being applied in a large number of industries, such as manufacturing, healthcare or finance, to uncover new patterns, make predictions and guide a wide variety of key decisions.

However impressive such recent developments have been, AI remains today very much task-focused and centered around well-defined pattern recognition applications. While current research is working dynamically to equip machines with human-like skills such as contextual awareness or empathy, attainment of this objective is, according to many AI scientists, still far ahead in the future.

Despite today's limitations, AI is already profoundly impacting society, businesses and individuals and is expected to exert a growing influence on how people live, work and interact with one another. As with all major technological shifts, AI is being idolized and demonized simultaneously. All sorts of existential threats potentially posed by AI are being devised, ranging from robots increasingly appropriating jobs to AI-powered machines fighting specifically against humans. Setting such gloomy scenarios aside, it is nevertheless undeniable that new ethical and societal challenges are emerging concomitantly with innovative AI developments. Businesses, governments, regulators and society as a whole, will have to address such issues to ensure that AI truly benefits all of humanity. In this context, technical standards and conformity assessment systems could play a critical role in shaping the future of AI.

1.2 From winter to rebirth of artificial intelligence

AI does not constitute a new scientific discipline, as its origins can be traced back to the 1950s. The literature typically identifies three historical phases of AI development.

In the first phase (1950s to 1980s), AI emerged from the abstract mathematical reasoning of programmable digital computers. The famous computer pioneer Alan Turing conceptualized the first test to decide whether a programme could be considered intelligent or not, the so-called Turing Test [1]. The term "artificial intelligence" was actually initially crafted by John McCarthy in 1955, who later became known as one of the fathers of AI [2], and was proposed as the subject title for the first conference on AI, which took place in Dartmouth College in 1956.

An important next step in the development of AI was the invention of an algorithm using the concept of neural networks (the "perceptron") by Frank Rosenblatt in 1958 [3]. However, it was not until 1967, with the development of the nearest neighbour algorithm by Cover and Hart [4], that machine learning started to be used in real applications. In spite of these early achievements and the rapid development of computer-based symbolism, the reach of AI remained nevertheless limited due to the inability to formally express or represent many concepts.

In the second phase (1980s to late 1990s), expert systems developed rapidly and significant breakthroughs in mathematical modelling were achieved. ANNs also started to be deployed more widely across a growing number of applications. During that period, some of the core techniques and algorithms of AI were developed and further refined: explanation-based learning (EBL) in 1981 [5], the backpropagation algorithm in 1986 [6], and the principle of the support vector machine (SVM) in 1995 [7].

One of the best-known milestones during this second phase was the Deep Blue chess programme developed by IBM in 1996, which managed to beat the world champion the following year [8]. This was the first time a computer programme was able to defeat human players in games at the world championship level. In spite of this success, limitations related to knowledge acquisition and reasoning ability, combined with the high cost of deployed AI systems, produced a certain level of disenchantment, which led some observers to speak of an "AI winter".

It was not until the third phase of development, which started at the beginning of the 21st century, that AI began to deliver on its initial promises. In 2006, the first powerful fast learning deep belief network was introduced in a paper by Hinton, Osindero and Teh [9]. The algorithm was used to recognize and classify numbers in a set of images. This contribution was instrumental to the development of AI and became one of the most influential works for today's AI research. More recent developments such as IBM Watson in 2010 and AlphaGo in 2016 have since received considerable public attention.

With the explosion of collected data, sustained innovation in theoretical algorithms and the continuing rise in computing power, AI has subsequently made breakthrough progress in many application fields and now looks well prepared to take on new challenges. All of these developments have led some analysts to speak of a "rebirth of AI".

Figure 1-1 depicts some of the major milestones of AI from its early days until the present time.

The success of machine learning algorithms for speech and image recognition was instrumental in attracting considerable interest from the research community, businesses and governments. Additionally, parallel developments in cloud computing and big data provided the support for moving from computer-based AI simulations to more complex and intelligent systems connecting machines and people. It is now foreseen that AI will become one of the core technologies of the fourth industrial revolution, as well as a driving

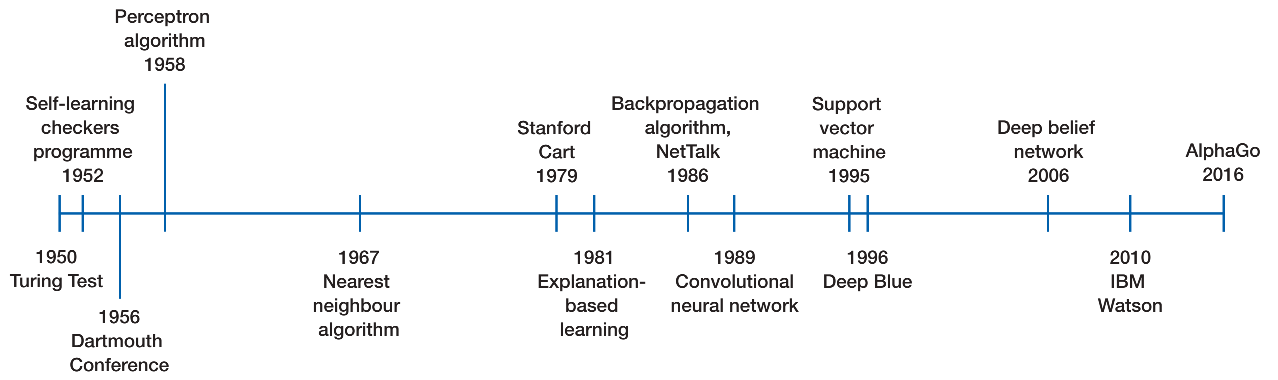


Figure 1-1 | Major milestones of AI development

force for innovation in transportation, healthcare, retail, education, government services and other industries.

1.3 Great opportunities come with risks and challenges

AI represents a huge market potential. According to a recent study from the International Data Corporation (IDC), worldwide spending on cognitive and AI systems is forecast to exceed USD 57 billion in 2021 [10]. The retail and banking sectors are expected to spend the most on AI in the coming years, followed by discrete manufacturing, healthcare and process automation. These five industries, still according to IDC, will continue to be the largest consumers of AI technology, with their combined investments representing nearly 55% of all worldwide spending on such technology by 2021.

Taking into account the related service industry of machine intelligence, which includes programme management, education, training, hardware installation, system integration and consulting, the market size is actually much larger and AI is foreseen to become one of the fastest growing industries in the near future.

While automated customer service and diagnostic systems will likely remain the top drivers of AI spending in the coming years, smart manufacturing is expected to take a strong position in the AI market. IDC actually sees intelligent process automation become the third largest use case of AI systems by 2021 [10]. Other use cases that will experience fast spending growth include public safety, emergency response, and shopping advisors and recommendations.

Inevitably these exciting market prospects will also carry a certain number of risks and challenges. The impact of AI on the workforce is frequently cited as a potential threat to societies, with tensions in social relations resulting from a gradual diversification of the employment market. Increased automation and connectivity could also lead to additional or intensified wealth gaps between developed and developing economies. However uncertain such scenarios appear today, all major economies throughout the world have started to invest heavily to support AI innovations as part of their strategic technology planning activities. For instance, in 2017 China promulgated the "New Generation of AI Development Plan", the "Three Year Action Plan for the Promotion of New Generation of AI

Industry (2018-2020)" as well as various other policies to accelerate research, development and industrialization of AI technology.

1.4 Definitions of artificial intelligence

There are several ways of defining AI. ISO/IEC Joint Technical Committee (JTC) 1 refers to "an interdisciplinary field, usually regarded as a branch of computer science, dealing with models and systems for the performance of functions generally associated with human intelligence, such as reasoning and learning" [11]. In the IEC White Paper on edge intelligence [12], the term AI is applied when "a machine mimics cognitive functions that humans associate with other human minds, such as pattern matching, learning, and problem solving".

In other words, intelligence is demonstrated by four basic capabilities: sensing, comprehending, acting and learning. As of today, comprehending does not have the same meaning for machines as for humans. Typically, a model will be trained to "learn" how to perform better compared to more conventional methods, but AI systems cannot claim yet to "comprehend" the world around them.

Practitioners of AI often distinguish between strong AI and weak AI. **Strong AI** (also called **general AI**) refers to the more philosophical concept of a machine capable of exactly imitating human intelligence. Such a machine would be able to solve any problem in any field requiring advanced cognitive abilities. This kind of AI has not been developed yet and can only be found in various science fiction books or movies.

In contrast, **weak AI** (also called **narrow AI**) supports humans in solving specific problems for a particular use case. For example, AlphaGo masters the board game Go to an almost perfect degree but is unable to solve any other problem. Speech recognition tools such as Siri represent a kind of hybrid intelligence, which combines different weak AIs. These tools have the ability to translate spoken language and connect words to their databases in order to perform different tasks. Nevertheless, such systems do not constitute any general form of intelligence [13].

Other terms are tightly connected to AI, such as machine learning and deep learning. To create intelligent machines, a specific kind of knowledge is needed. In the past, such knowledge was hardcoded directly into the machines, which led to certain restrictions and limitations. The approach taken by **machine learning** is that the machine builds up its knowledge itself, based on a given set of data [14].

As knowledge these days comes mostly from real-world data, the performance of machine learning algorithms highly correlates with the information available, also called representation. A representation consists of all the features that are available to a given machine (e.g. output of a temperature or vibration sensor in a predictive maintenance application). Selecting the right representation is a complex and time-consuming task, requiring highly specialized domain knowledge.

A field of machine learning called **representation learning** automates this task by discovering the representations needed for feature detection or classification from raw data. This set of methods is based on learning data representations, as opposed to more traditional task-specific algorithms [14].

However, selecting the right features is usually a very difficult operation, since they are dependent on various environmental factors. For instance, colours will be perceived differently in a dark environment, which then can impact the silhouette of objects.

As a subcategory of representation learning, **deep learning** transforms features and elaborates dependencies based on inputs received. In the example of an image, the input features are the pixels. A deep learning approach will map the pixels first to the edges of the image, then to the corners and finally to the contours to identify an object [14].

Figure 1-2 shows how these concepts logically relate to one another, with deep learning being a kind of representation learning, which in turn is a kind of machine learning, which is one subcategory of all possible approaches to AI.

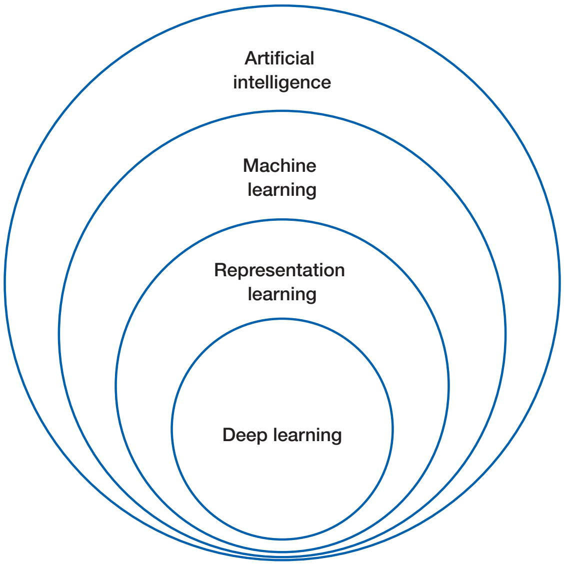


Figure 1-2 | Venn diagramme of AI

1.5 Scope of the White Paper

The objective of the present White Paper is to provide an overview of where AI stands today as well as some forward thinking for the next decade, by exploring opportunities and challenges of AI for a number of application domains. Building upon several use cases, the White Paper introduces a number of recommendations, some of them targeted at the IEC and its committees for future standardization and conformity assessment work.

This White Paper primarily focuses on weak AI. It is clear that today the world is running on various forms of weak AI. An email spam filter is a classic type of weak AI. It starts off loaded with a certain level of intelligence that enables it to figure out what constitutes spam and what does not, and then refines its intelligence as it acquires experience with the particular preferences of individual or collective users. Automobiles today are replete with weak AI systems, from the computer that determines when the anti-lock brakes should be activated to the processing unit that tunes the fuel injection parameters. Self-driving cars that are currently being tested include numerous robust weak AI systems allowing the vehicle to sense and react to the surrounding environment.

Creating an AI as smart as the human brain will remain an enormous challenge for quite some time. Building a computer that can multiply two ten-digit numbers in an infinitesimal amount of time is unexpectedly easy. Building one that can look at a dog and decide whether it is actually a dog or a cat is much more difficult. While creating a system that can defeat the world chess champion has already been achieved, fabricating an AI that can understand the meaning of a paragraph from a six-year old's picture book, and not just recognize the words, is still well beyond today's possibilities.

Many such capacities that seem easy to human beings are actually extremely complicated and only seem easy because the skills involved have been optimized in humans (and most animals) across hundreds of millions of years of evolution. Since it is virtually impossible to properly define what intelligence consists of, it is very difficult to provide a clear criterion as to what would count as a success in the development of strong AI.

In order to shed light on some of the foreseen developments of AI within the next decade, the following application domains are explored in the White Paper, and for each of them a number of use cases are described:

- Smart homes – residential environments that are equipped with connected products for controlling, automating and optimizing functions such as temperature, lighting, security, safety or entertainment.
- Smart manufacturing – a technology-driven approach using connected machinery to monitor production processes and data analytics to improve manufacturing operations.
- Smart transportation – a scenario in which mobility-related entities (e.g. personal vehicles, public transportation, delivery vehicles, emergency services, parking) are integrated and automated in order to improve traffic performance and reduce potential negative impacts such as congestions, accidents or pollution.
- Smart energy – a sustainable and cost-effective energy system in which renewable production, consumption, and infrastructures are integrated and coordinated through energy services, active users and enabling information and communication technologies (ICTs).

1.6 Outline of the White Paper

Section 2 describes the need for AI through identification of several key megatrends posing major challenges for societies, businesses and individuals. AI will enable and enhance a wide range of applications addressing some of these challenges, such as environmental concerns, changing demographic trends or economic disparity, to mention only a few.

Although AI is hardly a new discipline, it was not until 2010 and later that dramatic technological enhancements, in particular in the area of machine learning, paved the way for today's explosion of AI. This breakthrough was enabled thanks to a number of factors explained in Section 3. Significant improvements in computational power, more sophisticated machine learning algorithms and the availability of large amounts of data to train AI systems have been the primary enablers of today's spectacular AI developments.

A number of additional drivers that have also contributed to the flourishing field of AI research are further outlined in Section 3. These include information technology (IT) developments such as cloud and edge computing, the Internet of Things (IoT), and big data, as well as the increasing readiness of consumers and society to embrace new technologies and share data.

Section 4 provides a high-level understanding of the most common AI systems and machine learning techniques. Without entering into deep technical detail, it also reviews the most popular AI algorithms in use today that constitute the foundation for tomorrow's AI developments. Based on the current state of the art of AI, this technical overview is complemented by many references for readers wishing to consolidate their scientific understanding of how AI actually works from the inside.

While this White Paper cannot cover all of the possible AI application scenarios, a representation describing how today's main AI systems map to some of the most popular application domains is provided in Section 5. This exercise furnishes a better characterization of the AI needs and requirements of several industry sectors. The rest of the section is then devoted to a more detailed description of the four application domains (smart homes, smart manufacturing, smart transportation, and smart energy), for which several AI-related use cases and scenarios are reviewed, together with some of the most pressing challenges of current and emerging AI implementations.

Following this review, Section 6 consolidates the main AI challenges that can be identified in today's implementations or foreseen in emerging AI developments. Challenges are grouped into several categories: social and economic challenges; data-related challenges, including the selection of training data and the standardization of data; algorithm-related challenges, such as algorithm robustness and interpretability; infrastructure challenges, related to hardware or platforms; trustworthiness issues, including trust, privacy and security; and challenges linked to regulations, such as liability and ethical issues.

Building upon the previous section, Section 7 develops a standardization landscape for AI and identifies a number of gaps that need to be addressed in order to solve some of the AI challenges described previously. While today's standardization activities are still at a very early stage, the White Paper clearly demonstrates that standards need to play an essential role in shaping the future of AI and mitigating the many technical and non-technical issues emerging with the deployment of AI across industries.

Finally, Section 8 concludes the White Paper by offering a series of recommendations for industry, regulatory bodies and the IEC. While the success of AI and its acceptance within diverse societies will rely on the involvement of multiple stakeholders and communities, it is clear from this White Paper that industry, policy makers and standards development organizations such as the IEC will need to play a driving role to ensure that AI delivers on its promises.

Additional, future-looking AI developments that may grow in importance over the next decade are discussed in Annex A.

Section 2

Need for artificial intelligence

Today's society and business landscape are characterized by a complex and unprecedented set of challenges and opportunities. Existing markets are subject to disruption and can even disappear abruptly in a short space of time. Major global trends impacting society, the economy, business, cultures and personal lives, often called megatrends, are defining the future world of mankind and its increasing pace of change.

Megatrends represent interconnected, global interactions that contribute to framing the impact of major technology developments such as AI. The joint effect of digitization, automation and AI is expected to significantly impact the future of work. It is anticipated that computerization will affect many low-skill jobs, with computer-guided automation becoming increasingly prevalent across numerous industries and environments, including manufacturing, planning and decision-making [15]. The growth in technological capabilities is already transforming supply chains, reshaping the workforce and redefining jobs. The challenging prospect of such change lies in the fact that the growth is not linear but rather complex and accelerating.

At the same time, AI will enable and improve a wide range of applications that can address some of the challenges emerging from these megatrends: environmental concerns, changing demographics, or economic disparity, to mention only a few.

2.1 Scarcity of natural resources

The planet's natural resources are being consumed at an alarming rate and most countries are expected to double their annual global consumption of such resources by 2050 [16]. Not only are finite natural resources being depleted, humans are also using far more environmental resources than Nature can regenerate. While in the past resource conservation was often viewed as detrimental to business, today the two are by no means mutually exclusive.

AI is already helping countless manufacturers to optimize productions processes, thereby reducing waste and increasing output. In addition, AI will soon be used not just to optimize the processes themselves but also their inputs. By analyzing the purpose, properties and environmental impact of a production's input materials, AI will be able to help scientists design materials that match the specifications required for more sustainable production. Ideas have even been proposed for using AI to identify a second usage for the material components of by-products created by machinery, thereby creating a near circular use of raw materials. Not only are such efficiency gains in production processes an attractive incentive for businesses, they will also have a significant impact on global resource consumption.

AI will also help utilities in an era of increasing urbanization, growing power consumption, scarcity of water resources, and large-scale deployment of renewable energy. This will be achieved by more intelligent management of both demand and supply. On the demand side, AI is already producing significant energy savings, for example by reducing the consumption of data centres by 15%. On the supply side, decentralized smart energy grids will be able to predict and pre-empt outages, and manage fluctuations in supply and demand to ensure the optimal level of supply while at the same time minimizing the use of fossil fuels [17].

Additional examples include the optimization of renewable energy generation through solar or wind farms or the optimization of vehicle traffic flows to reduce emissions.

2.2 Climate change

Forecasts by leading scientists have been unequivocal: without a consistent response to the climate change challenge and a more responsible use of environmental resources, unpredictable changes will threaten the planet. This raises the question of how to reconcile economic objectives with environmental sustainability. Equally important is how mankind can prepare for unexpected dramatic natural occurrences in the future.

The use of AI in a wide variety of applications is expected to play a leading role in the fight against climate change. AI can support complex decision-making processes when dealing with natural resources or when predicting unexpected incidents. The consumption and use of resources, for instance, can already be optimally coordinated during energy generation. AI makes it possible to set numerous parameters such as context-based electricity consumption and grid load in relation to weather forecasts and electricity tariffs. As a result, the behaviour of electricity consumers can be determined and addressed more efficiently.

Building on those achievements, intelligent mobility solutions involving a more responsible use of resources can be implemented, including autonomous electromobility. Not only can vehicles and trucks be optimally and efficiently matched to one another, they can also be driven more efficiently thanks to AI.

Similar developments can be devised for efficient water consumption. In agriculture, for example, AI allows to determine the optimal water demand depending on the specific needs of each individual plant, the soil situation and current weather conditions. Furthermore, supply strategies for droughts and water shortages in affected regions and countries can be developed using AI techniques.

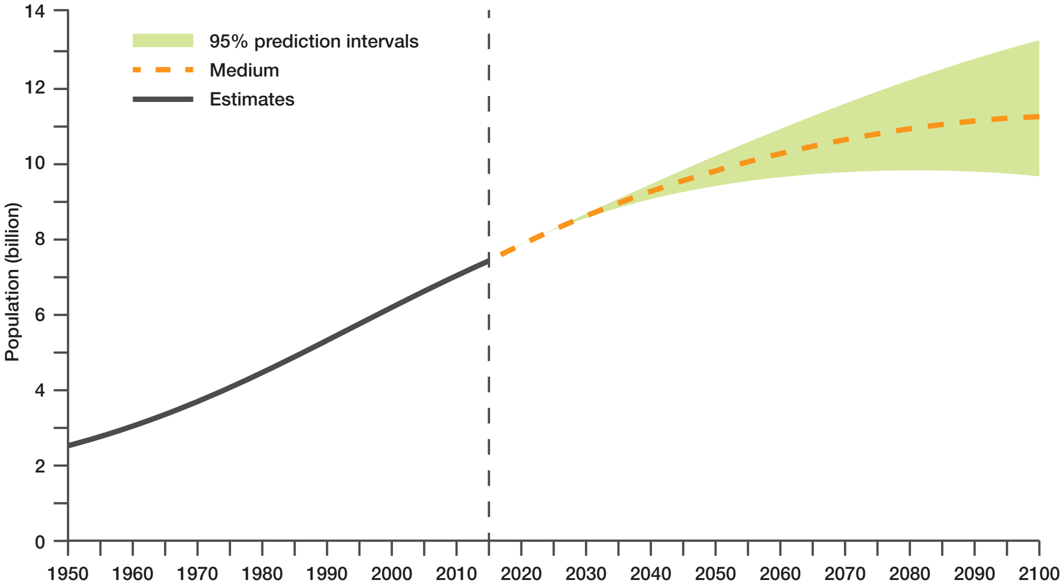
AI can also contribute to improving predictions of weather scenarios and natural disasters. Scientists are increasingly faced with the challenge of capturing and processing numerous influencing factors in order to increase the accuracy of weather forecasts. AI will effectively help people process a wide range of measurement data in order to provide early predictions for weather-related events and warnings of potential elements such as floods, air pollution episodes, or storms. Early-warning systems can then be set up more intelligently for diverse geographies.

2.3 Demographic trends

The United Nations (UN) is predicting a population increase of more than one billion by 2030 due to demographic growth in emerging and developing countries (Figure 2-1). In addition, the ageing population (number of persons 65 years or older) will increase by more than 390 million due to people living longer and having fewer children.

As shown in Figure 2-2, the impact of an ageing population will be more immediately felt in Europe, Asia and Latin America, resulting in different regional issues. The trend in the number of working people supporting each elderly person will move from nine in 2015 to a range of a half to four in Asia, which will create a much higher dependency of the older generation on the younger one. In Europe, the availability of a suitable working age population will decrease and create an acute demand for a new generation of workers, i.e. women and the elderly. The ratio of four working age persons per elderly in 2015 will decrease by 50% in 2050. These ageing population trends will undoubtedly generate significant challenges for governments and industry [18].

Another important aspect of such trends is related to large regional differences in the availability of the working age population, such as 1,5 working age



Source: United Nations, Department of Economic and Social Affairs, Population Division (2017). *World Population Prospects: The 2017 Revision.* New York: United Nations.

Figure 2-1 | Increasing global population

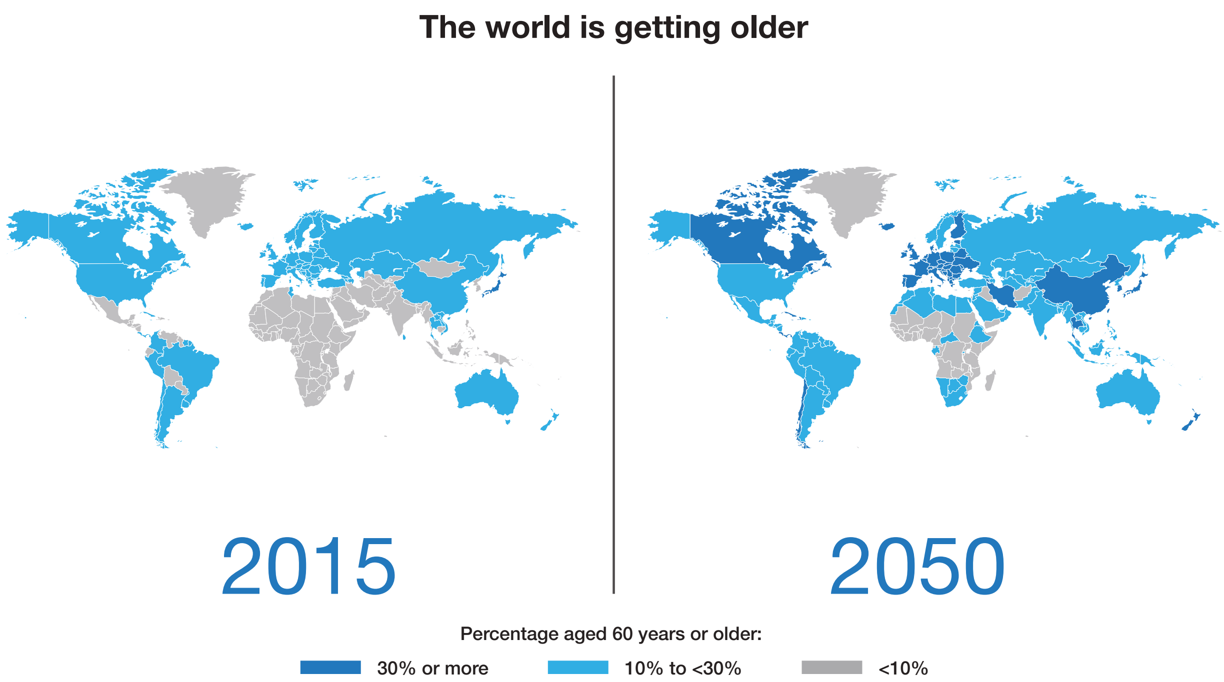


Figure 2-2 | Ageing population across the globe

people for each elderly person in Japan in 2050 versus 15 working age people for each elderly individual in Nigeria.

The ageing population trend will drive higher spending on healthcare. It is anticipated that within G7 countries, healthcare budgets will increase by about USD 200 billion every year. However, technological and scientific innovations may help reduce healthcare costs to more affordable levels.

Most governments in Europe are now encouraging older workers to remain within the workforce by increasing the official retirement age, and by outlawing age discrimination. In addition, industry will need to provide financial incentives as well as re-training programmes for older workers. Lifelong learning to acquire new skills during the individual's working life, as well as mentoring of younger colleagues, will become critical for efforts to keep ageing people within the workforce.

2.4 Economic policy

It has already been highlighted how automation and immense productivity gains will help countries alleviate the pressures of demographic change. While this will certainly be a welcome development for more advanced economies struggling with an ageing population, there are countless opportunities for AI to make a difference for the world's poorest countries.

The UN's Sustainable Development Goals, for example, seek to tackle these challenges by reducing poverty and hunger, and improving education. The recent AI for Good Global Summit 2017 highlighted how AI can support these efforts. Suggestions ranged from initiatives aimed at monitoring the progress of the international community toward achieving these goals and determining where resources are most needed, to predictive modelling of disease outbreaks [19]. AI is also poised to operate in conjunction with the IoT, drones and synthetic biology to power smart agriculture and provide insights concerning when and where to plant and harvest crops while optimizing nutrition, pesticides and water to increase yields and help combat world hunger [20].

The interest of governments in the potential of AI is increasingly rapidly, and as it does, AI is expected to make contributions to public policy, particularly economic policy. When constructing a model, economists usually begin with a set of assumptions which they then seek to verify. AI, however, offers the power to analyze data and uncover previously unknown interactions between variables, on which a model can then be built from first principles and serve to inform public and monetary policies.

AI may also support financial regulators in their monitoring activities by inspecting the balance sheets of banks for anomalies that are of concern from a prudential or conduct perspective [21]. Such explorations into how AI can be used in the public sphere should be welcomed, as they can help governments explore best practices, make more informed decisions on AI policy and build public trust.

2.5 Service and product customization

The integration of AI and advanced manufacturing enables mass customization that easily connects suppliers, partners and customers, and meets individualized demands with efficiency and cost close to mass production. Applying AI therefore optimizes the manufacturing value chain, so that manufacturers can track flows of materials in real-time. They can also assess more accurately engineering and quality issues, reduce excess inventory and logistics delays, increase responsiveness to customer needs and make better business decisions that reduce waste and costs. Businesses will benefit from mass customization of production by empowering internet-connected consumers to control intelligent manufacturing processes in order to develop products according to their desired specifications.

AI allows the individualization of products on a completely new level. Not only product configurators that customers can use online are affected. The use of AI also opens up completely new possibilities for individualization. Products whose individualization would involve high development costs can be adapted to requirements by using AI.

Services also can be automatically tailored to customer needs. An example is the automatic translation of texts. While previously the results of rule-based techniques could often only reflect the meaning of individual words but not entire sentences that are contextualized, services supported by AI are able to perform translation based on the meaning of a text.

Section 3

Enablers and drivers of artificial intelligence

While improvements in hardware, algorithms and data availability have been the primary enablers of AI, capital has been its fuel. The rapid development seen today would arguably not have been possible without an increase in awareness, venture capital and government support for AI, which served to provide both the funding and market for new innovations. There is a growing awareness of the advantages AI can bring to those who are able to use it effectively. As a result, business and technology leaders are taking a much more active role in shaping the future of AI, thereby creating improved market conditions for its development.

- **Heightened awareness**

78% of organizations surveyed in 2017 reported that they had plans to adopt AI in the future, with close to 50% indicating that they were actively exploring adoption [22]. Indeed, in 2016 companies collectively invested up to USD 39 billion in AI [23], particularly in machine learning, which attracted nearly 60% of investments. An increasing interest in and adoption of AI technology by small-to-medium enterprises (SMEs) will no doubt continue to fuel this growth for many years to come [24]. The potential market for AI applications is thus huge, with a projected total market size of USD 127 billion by 2025 [23].

- **Availability of private capital**

With companies eager to transform their business through AI and willing to pay the price, the availability of capital for AI entrepreneurs has never been higher. Global venture capital investments in AI doubled in 2017 to USD 12 billion [25] and the number of active AI start-ups in the United States alone has increased 14-fold since 2000. Technology companies too have outdone each other in announcing billion-dollar investments in their AI departments.

- **Government support**

This availability of private capital can only be expected to increase, as governments compete to grow their domestic AI industry. The European Union, the United Kingdom, Germany, France and Canada have all committed to strengthening their domestic AI industries, after China published its ambitious plan to overtake the United States as the global leader in AI by 2030 [26]. Although the volume of investment committed varies widely by country, increased government attention is laying the foundations for public/private partnerships and the development of AI applications for the public sector.

Today it is easier than ever to access the world of AI. Frameworks, toolkits and libraries provide users with algorithms and various programming languages. They also maintain such algorithms and facilitate implementation, attracting a community of developers and users to jointly improve open-source software [27].

3.1 Enablers of artificial intelligence

Interest in AI has reached a new peak. Almost every week, new discoveries arise and new applications are advertised publicly. Achievements that were unthinkable in the past are now being accomplished at increasing pace. For example, IBM Watson has beaten the best players in Jeopardy, a game that requires high language skills as well as knowledge about riddles and wordplays [28].

The remaining question is what has enabled such a tremendous acceleration in AI progress over recent years? While core concepts and approaches to AI have been around for decades, three key enablers have helped bypass the "AI winter" and led to today's spectacular developments:

- Increased computational power
- Availability of data
- Improved algorithms

3.1.1 Increased computational power

Most AI algorithms require a huge amount of computational power, especially in the training phase. More computational power means that algorithms can be tested and trained faster, and that more complex algorithms can be implemented. Therefore, the growing adoption of AI has been greatly enabled by progress made in hardware technology (e.g. integrated circuits, semiconductor fabrication, server technologies).

The increase in computational power is commonly represented by Moore's Law, which relates such power to the density of transistors on a chip [29]. Following Moore's law, the feature size of semiconductors has shrunk from 10 μm in the 1970s to 10 nm in 2017, which means that a far greater number of transistors can be integrated on the same die size (Figure 3-1).

Not only has the transistor number substantially increased to provide much higher computing power, but also the hardware architecture has been improved to offer better performance for AI applications. For instance, multi-core processors are designed to provide increased parallelism. In addition to central processing units (CPUs), other types of processing units such as graphics processing units (GPUs), field-programmable gate arrays (FPGAs), and application-specific integrated circuits (ASICs) are being further adopted for various workload patterns.

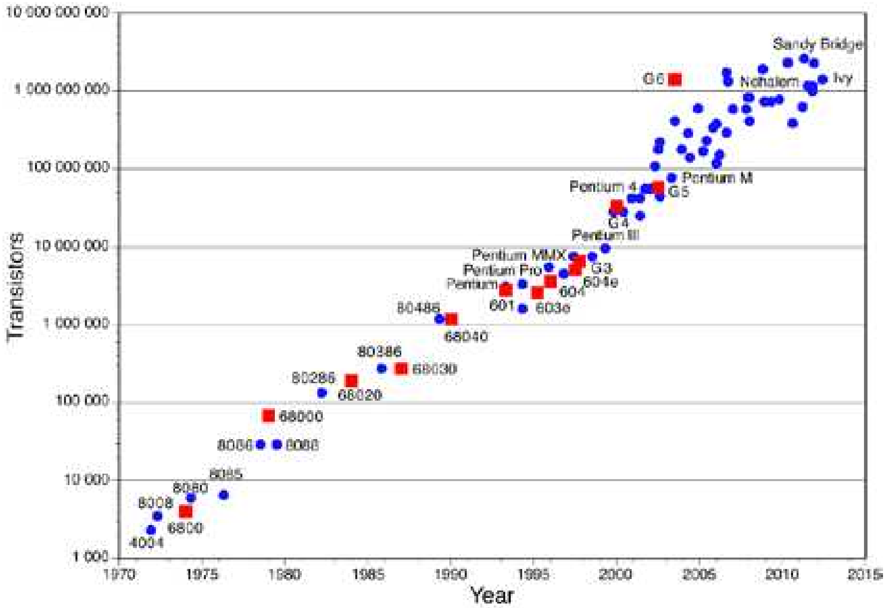


Figure 3-1 | Increased computational power over time

GPUs are used for handling image processing tasks and have been found to be very effective in accelerating AI algorithms such as deep neural networks (DNNs) or convolutional neural networks (CNNs). Integrated circuits like FPGAs are configurable after manufacturing to fit a wide range of applications and offer much higher speed compared to traditional CPUs. ASICs include different variants such as tensor processing units (TPUs) and neural processing units (NPUs). By tailoring the circuit design based on the data processing patterns, ASICs can achieve even higher performance than GPUs or FPGAs. For example, it is claimed that TPUs deliver 15 to 30 times higher performance and 30 to 80 times higher performance-per-watt than contemporary CPUs and GPUs [30].

3.1.2 Availability of data

The output of an AI application is information extracted from algorithms based on supplied data. Therefore, the use of incomplete or faulty data will always lead to poor results, no matter how good the algorithm is [29]. An important factor in creating and evaluating new algorithms is access to datasets and meaningful data that have already been classified. One of the most significant developments that has driven the availability of data is the internet. This has allowed huge communities to collaborate in order to create datasets, which can be accessed by researchers all over the world.

One illustrative example of an internet community that constantly creates, classifies, labels and uploads data for image classification is the ImageNet community. Creating and labelling training data not only took considerable time in the past, but has also been almost impossible for large datasets that are required to train neural networks. While an image dataset for facial recognition in 1997 consisted of approximately 165 instances [31], an ImageNet dataset for faces already consists of 1 570 fully classified instances [32]. In total, ImageNet alone provides almost 15 million classified images, which can be easily accessed and used to train and evaluate AI algorithms [33].

Easy accessibility to a large amount of data that can be used to train and fine-tune algorithms means that researchers and practitioners of AI can devote their time to improving and developing new algorithms and then quickly test and validate them. This was not possible just two decades ago.

3.1.3 Improved algorithms

Most recent advances in AI have taken place in deep learning. Although the concept has been around for several years, the actual breakthrough occurred very recently. While there have not been major recent milestones in neural network research, developments have not come to a halt. Numerous improvements in existing techniques and the development of several new ones have led to many successful neural network implementations.

An example of a small change in algorithms that allowed neural networks to process information faster and more efficiently is the rectified linear unit (ReLu) activation function. An activation function is an essential part of any neuron, of which neural networks consist. The ReLu concept was introduced by Hahnloser in 2000 [34]. It takes the form of a ramp function and consists of two linear parts. Its first successful implementation was demonstrated in 2011, when it was used to train a neural network more efficiently [35]. It replaces another activation function, the logistical or sigmoid function, and describes a logarithmic curve requiring more processing power than a linear function. The ReLu activation function offers many advantages compared to the sigmoid activation function, such as efficient propagation, scale invariance and higher computational efficiency. It is especially useful in applications containing complex datasets, since it allows for a faster and more efficient training in deep networks.

Another example of a new concept that contributed to the development of deep learning is the CNN, which was first introduced in 1989 [36] [37]. It is widely used for image recognition tasks, where it is able to achieve above-human-level performance [38].

All these developments have enabled AI research to achieve enormous progress over the last few years. The increased amount of available data would have been useless without the ability to process it efficiently with appropriate algorithms. It is clear that advances in AI have not been the result of a single enabler, but rather the consequence of a combination of various ideas and technologies that gradually improved over time.

3.2 Drivers of artificial intelligence

The three key enablers described above have given birth to the flourishing field of AI research that can be observed today. Added to the perceived economic value of AI, these developments have also been made possible thanks to a number of technology and social drivers that have accelerated the deployment of AI across a broad range of applications. They are all related to the digital transformation trends that have been permeating industry, society and also individual lives over the past decade.

IT developments have been instrumental in supporting the deployment of AI applications. Abundant resources have been provided by modern IT infrastructures, including but not limited to cloud computing, big data and the IoT.

Cloud computing is an elastic and scalable infrastructure that provides access to shared pools of resources and higher-level services, which can be provisioned on-demand and with minimal management effort. Since the computing needs of AI vary significantly based on datasets and algorithms, particular application requirements can be met by leveraging the enhanced resource utilization and efficiency offered by cloud infrastructures.

Big data is the science and engineering of storing, managing and analyzing large datasets characterized by volume, velocity, variety and/or variability [39]. Techniques and architectures dealing with structured, semi-structured and unstructured data have been proposed. These include, for example, relational database management systems (RDBMSs), distributed file systems, graph databases and various computing frameworks to process or analyze these data.

Last but not least, the IoT is the network of physical devices, vehicles, or other electrical appliances, which enables such objects to communicate and interact with each other. It provides the infrastructure to collect and gather data related to the status of these devices via sensors distributed geographically. Devices can then be configured and controlled through actuators. The combination of IoT infrastructure and AI technologies has led to many applications, such as smart manufacturing, smart homes and intelligent transportation.

These IT developments have coalesced with changes in society that have boosted the acceptance and widespread use of data-intensive tools such as social media. This trend was an additional factor driving and facilitating the extensive deployment of AI.

3.2.1 Cloud and edge computing

Edge computing is a distributed open platform at the network edge, close to the things or data sources involved, integrating the capabilities of networks, storage and applications. By operating in close proximity to mobile devices or sensors, edge computing complements centralized cloud nodes, allowing for analytics and information generation close to the origin and consumption of data. This enables the fulfilment of key requirements of industry digitalization for agile connectivity, real-time services, data optimization, application intelligence, security and privacy protection. Cloud and edge computing enable access to cost-efficient, scalable computing resources as well as specialized services.

AI benefits from the use of edge computing in the following ways:

- Localization of data acquisition and storage enables the pre-processing of data so that only decisions or alarms are forwarded to the cloud servers rather than raw data.
- Faster and more efficient decision-making can be achieved via the placement of machine learning algorithms on the edge devices, thus reducing the frequency of contact with cloud servers and steadily decreasing the effect of round-trip delay on decision-making.
- Data can be secured close to its source using local identity management and application-specific access policies, and following local regulations.
- Communication between edge computing nodes will enable the distribution of AI capabilities and sharing of intelligence between the distributed AI nodes.

Further information on edge computing and edge intelligence is available in the IEC White Paper on edge intelligence [12].

3.2.2 Internet of Things

The IoT focuses on gathering data from devices, which is particularly relevant for production and consumer information. The evolution of ICTs over the last decades has led to the expansion of computing capabilities into smaller and smarter devices [40]. Based on this development, the IoT is defined by ISO/IEC as the “infrastructure of interconnected objects, people, systems and information resources together with intelligent services to allow them to process information of the physical and the virtual world and react” [41].

The number of connected devices installed worldwide is expected to increase from over 23 billion in 2018 to approximately 75 billion in 2025 [42] [43]. This illustrates the impact of the IoT on data acquisition. In the future, the amount of data that can be used in AI applications will further increase and thus improve the performance of algorithms.

Today, IoT applications make it possible to capture performance and environment-related data using sensors attached to devices. This enables the analysis of data either locally or via cloud platforms. In the future, the real-time behaviour of such systems will become increasingly important for time-critical applications such as autonomous driving [44].

Given the large amounts of data generated by these connected sensors, the role of AI computing at the edge (or edge computing) will become even more important. As mentioned before, it is impractical and sometimes even impossible to transmit all generated data to a central location, analyze such data and then send back the necessary information to the device. Therefore, being able to carry out simple analysis or decisions locally is becoming critically important. This trend will also lead to simpler AI algorithms running locally on devices that rely on edge computing. AI at the edge will bring the IoT to the next level of capabilities.

Further information on the IoT is available from the IEC White Paper on smart and secure IoT platform (IoT 2020) [44].

3.2.3 Big data

The unprecedented growth in global connectivity and networking is generating massive amounts of data, and the rate of this generation is accelerating. Big data makes it possible to store, manage and analyze these very large datasets through various techniques and architectures. Big data can be of great benefit to individuals and organizations, offering insights into a multitude of areas such as smarter cities, faster medical breakthroughs, more efficient use of resources, and human resource processes.

Many organizations are embracing this paradigm and are becoming more data-driven in decision-making, product and service development, as well as when interacting with customers, employees, suppliers and other stakeholders. Social media platforms are great examples of how marketers today approach their customers and transform the field of marketing [15].

Big data is commonly defined as “a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information” [45]. According to [39], big data can be represented by volume, variety, velocity and veracity (the four Vs) as follows:

- Volume – the amount of data generated. This amount has been exploding due to the increasing number of data sources connected through the IoT, their higher resolution, as well as the depth of data. The challenge for AI applications is to process, analyze and maintain this very large amount of data.
- Variety – the heterogeneity of data, caused by the variety of data sources. Multiple sources describe one event, providing different data formats in a structured or even unstructured form. These data are not limited to sensor data, but could also be for example expert knowledge of a machine operator. AI therefore has to exploit information from different sources with different data types.
- Velocity – the speed at which data is generated, which currently is real-time in many cases. For some applications, the speed of data generation is critical because it conditions the validity of the data. Often this leads to a trade-off between the speed of data generation and its processing. The latency between generation and processing is an important factor for AI applications.
- Veracity – the data quality. As described above, an AI algorithm is only as powerful as the data with which it is fed. As applications based on lower-quality data might lead to wrong predictions, AI has to mitigate the data quality issue to keep producing useful results.

3.2.4 Consumer acceptance

Another driver of AI is the increasing readiness of consumers and society as a whole to embrace new technologies, share data and information, and join collaborative communities to improve AI applications.

The generation called “digital natives”, which grew up with computers, smartphones and other electronic devices, is already very open-minded about adopting new technologies and sharing personal data. Although data privacy concerns are now receiving more attention, younger generations have already embraced data-intensive activities (such as social media) as part of their lives. In a recent study by Deloitte, the willingness to share information with companies has doubled since 2014. Almost 80% of the people surveyed stated that they are willing to share their personal data if they directly benefit from it [46]. This is one of the reasons explaining why social media are major fields of application of AI.

Social media platforms have rapidly become popular and efficient ways of sharing, communicating, networking and collaborating for individuals, organizations and businesses. They offer businesses increased brand awareness, enhanced customer analytics, and new sales channels. Additionally, journalists, scientists, business owners, and the general public, who used to live and work in isolation from each other, are increasingly becoming more and more interconnected. Social media allow for immediate connections, which may previously have been considered beyond conventional reach [15].

But AI is not only deployed to obtain insights into consumers, it already functions as a part of people's daily routine. Internet search engines already use AI. Both Google and Baidu have developed high-performance algorithms that improve the accuracy of search queries [47]. Other applications are found in a myriad of sectors. Fraud can be detected via machine learning algorithms to secure bank accounts [48]. E-mail accounts are kept cleaner by algorithms that automatically filter spams [49]. Facebook uses facial recognition to compare users with new images [50]. Pinterest automatically identifies specific objects in images and allows them to be assigned to specific categories [51]. Twitter and Instagram have developed user sentiment analysis engines [52]. Snapchat tracks facial movements and allows dynamic overlay [53]. Many other daily examples could be mentioned.

While human interaction with AI in these cases is rather passive, efforts are also being spent on making AI much more proactive and interactive with humans. Siri, Alexa, Google Now or Cortana can handle natural language processing (NLP) to behave like personal assistants answering all kinds of questions [54]. More developments of this nature are certainly to come in the future.

Section 4

Inside artificial intelligence

This section provides further insights into how AI actually works, in particular what kinds of machine learning mechanisms are commonly encountered today. Some of the basic learning problems that machine learning algorithms deal with are explained in more detail, as well as major application areas in which such algorithms are implemented. The last part of this section then reviews some of the most widely used AI algorithms.

4.1 Categories of machine learning

Machine learning extracts information from data. Three main categories describe the way this process works: supervised learning, unsupervised learning and reinforced learning.

4.1.1 Supervised learning

If the dataset includes known input and output pairs, it is called supervised learning. Based on this, supervised learning uses a set of training data to predict output values for unknown datasets. The performance of models developed using supervised learning depends upon the size and variance (data selection) of the training dataset employed to achieve better generalization and greater predictive power for new datasets. Algorithms can perform either a classification (Figure 4-1) or a regression (Figure 4-2).

Classification maps input variables to discrete output variables. A typical application is spam filtering based on the occurrence of certain words in an email (e.g. “$$$”, “make money”, “no fees”). Regression maps the input variables to continuous output variables, approximating for example the temperature curve over the whole year.

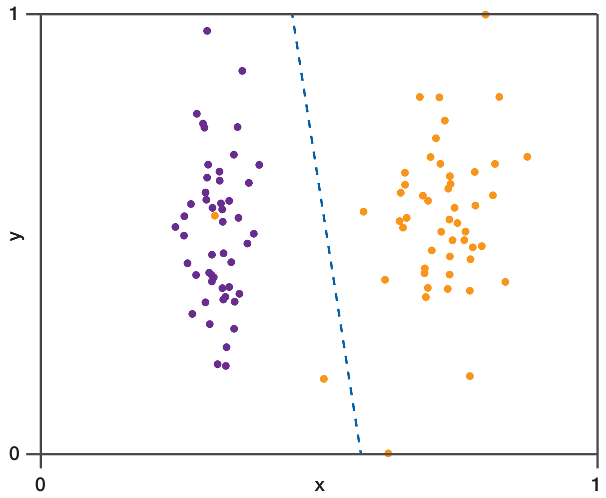


Figure 4-1 | Classification

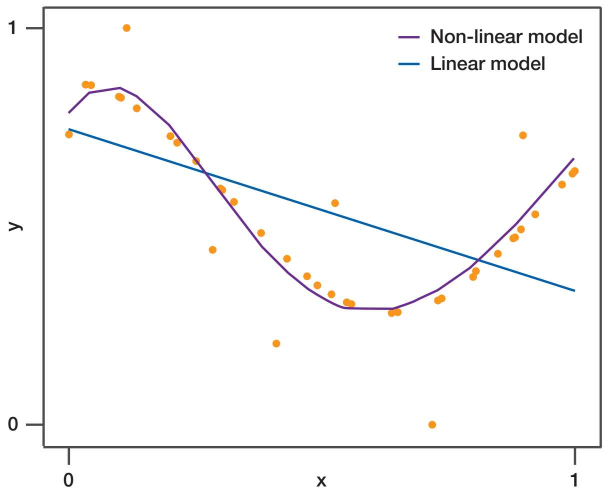


Figure 4-2 | Regression

4.1.2 Unsupervised learning

Unsupervised learning models learn to group instances of a dataset without defining pre-specified attributes. The algorithms determine the underlying structure of the dataset without information about target criteria. Given a set of images, such an algorithm would for instance identify that the objects in the various images are not the same. Without knowing the object, it then forms different categories. Typical approaches include for example clustering (Figure 4-3) and dimensionality reduction.

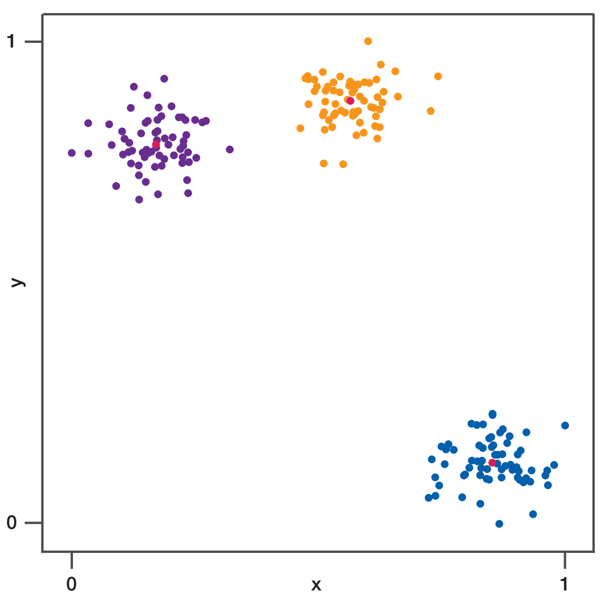


Figure 4-3 | Clustering

4.1.3 Reinforcement learning

Reinforcement learning describes a way of learning in which an AI system in the form of an agent interacts with its environment, or at least with a specific object in the environment. The agent is able to perform actions on and observe its environment. In return, the agent receives feedback through its environment, usually in the form of a reward. This reward can be positive or negative. The goal of the agent is to maximize its received positive feedback or minimize negative feedback through its actions. This is effected by a value function that approximates the value of an action. It improves through performed actions and received rewards for said actions. Through repeated actions in connection with received feedbacks, the agent is thus in a better position to approximate the value of its actions through the value function. Depending on how the feedback is structured, the agent can learn to perform certain functions. The process of such a learning agent is illustrated in Figure 4.4.

4.2 Current machine learning systems

In the following subsections several examples of typical AI systems encountered today are described.

4.2.1 Computer vision

Computer vision refers to the use of computers to simulate the human visual system in order to identify objects or people. It usually makes use of machine learning algorithms that recognize patterns in pictures and utilize these patterns to classify the image. Computer vision tasks include methods for acquiring, processing, analyzing, and understanding digital images, and extraction of high-dimensional data from the real-world in order to produce numerical or symbolic information (e.g. decisions). Understanding in this context means the transformation of visual images into descriptions of the world that can interface with other thought processes and generate appropriate actions. Image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner. CNNs are commonly applied for this kind of tasks.

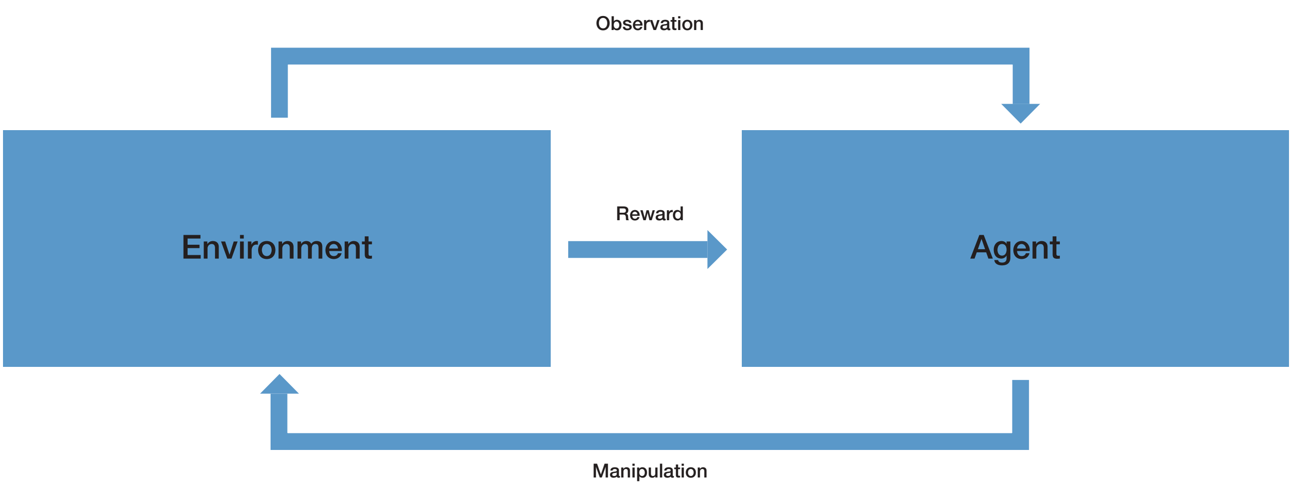


Figure 4-4 | Reinforcement learning

4.2.2 Anomaly detection

Anomaly detection is used for any application in which it is essential to identify a deviation from an expected pattern. This can be found in a variety of scenarios, such as fraud detection, health care monitoring or detection of intrusion into a computer system. Machine learning can support faster detection of anomalies.

Three broad categories of anomaly detection techniques exist. Unsupervised anomaly detection techniques detect anomalies in an unlabelled test dataset, under the assumption that the majority of the instances in the dataset are normal in contrast to instances that seem to fit least to the remainder of the dataset. Supervised anomaly detection techniques require a dataset that has been labelled as "normal" and "abnormal" and involves training a classifier. Finally, semi-supervised anomaly detection techniques construct a model representing normal behaviour from a given normal training dataset, and then test the likelihood of a test instance to be generated by the learned model. Algorithms commonly used include k-nearest neighbour (k-NN), SVMs, Bayesian networks, decision trees, k-means, but also ANNs such as long-short-term memory (LSTM) approaches.

4.2.3 Time series analysis

Time series analysis describes an analytical approach to finding patterns in a set of time series of data. The aim is to recognize trends in the data, which can be obscured by noise, and then describe them formally. Furthermore, time series analysis is used to forecast future values of the series in order to make predictions. Among the algorithms that are used for time series analysis are hidden Markov models (HMMs), recurrent neural networks (RNNs), LSTM neural networks, and SVMs.

4.2.4 Natural language processing

NLP is a way for computers to analyze, understand, and derive meaning from human language in a smart and useful way. By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation.

For instance, in speech recognition a computer analyzes spoken language and translates it into text. This can be part of a language processing system, in which the computer receives a spoken question and searches for an answer. Many home assistance systems rely on this kind of user input. Techniques such as HMMs and DNNs are often used for this task.

In machine translation, many programmes that fully automatically translate text from one language into another use self-learning algorithms. The challenge in translation is usually that there is not just one specified meaning for every word that can be looked up in a dictionary, but the meaning can change depending on the context. While statistical and rule-based models were often used for machine translations in the past, the development of DNNs has progressed rapidly in recent years and often provides superior results.

4.2.5 Recommender systems

A recommender or recommendation system predicts items for a user matching his or her preferences. The popularity of recommendation systems is often based on the use of digital content or services, where the preferences of a user can be more easily identified based on given ratings. Collaborative filtering is often used, but naïve Bayes and k-NN algorithms are also popular for this task.

4.3 Algorithms for machine learning

Having introduced some of today's most common machine learning systems, the main algorithms supporting these systems are described in subsections 4.3.1 to 4.3.9 below.

4.3.1 Decision trees

Decision trees are applicable to both classification and regression tasks. They are usually categorized as a form of supervised learning algorithms. Decision trees use training data to graphically outline decision rules and their outcomes (Figure 4-5). A classification tree results in categorical or discrete outcomes, whereas a regression tree predicts continuous values. Because of their easy interpretability and high level of accuracy, decision trees are very popular machine learning techniques. Widely used decision tree algorithms include Iterative Dichotomiser 3 (ID3), its successor C4.5, and classification and regression tree (CART) [55] [56].

These models differ in their mode of operation. To generate an efficient prediction, the hierarchy of the decision tree relies on the contribution of the attribute to the overall outcome. The features having most impact on the result are set on the top of the decision tree, followed by the features with a declining impact.

One problem with decision trees in general is overfitting. This happens if the model extracts every detail of the training dataset, building an over-complex decision tree. The result is very good performance on training data, but with not enough generalization to fit new datasets. There are several solutions to avoid the overfitting of decision trees, such as pruning methods (set the minimal number of samples needed to create a node or limit the depth of the decision tree) or ensemble methods (creating multiple decision trees trained on different datasets and combining them into a single model) [58].

4.3.2 Support vector machines

The SVM algorithm handles supervised machine learning problems. It can be applied for classification and regression tasks. The basic concept of this algorithm is to linearly divide different classes, as illustrated in Figure 4-6 for the case of two dimensions and two classes that are linearly separable. The algorithm maximizes the distance between the classes provided by the dataset.

This is called the hard margin classifier. To provide an optimal classification, the algorithm uses the data points which allow for a maximal separation between the different classes. The chosen data

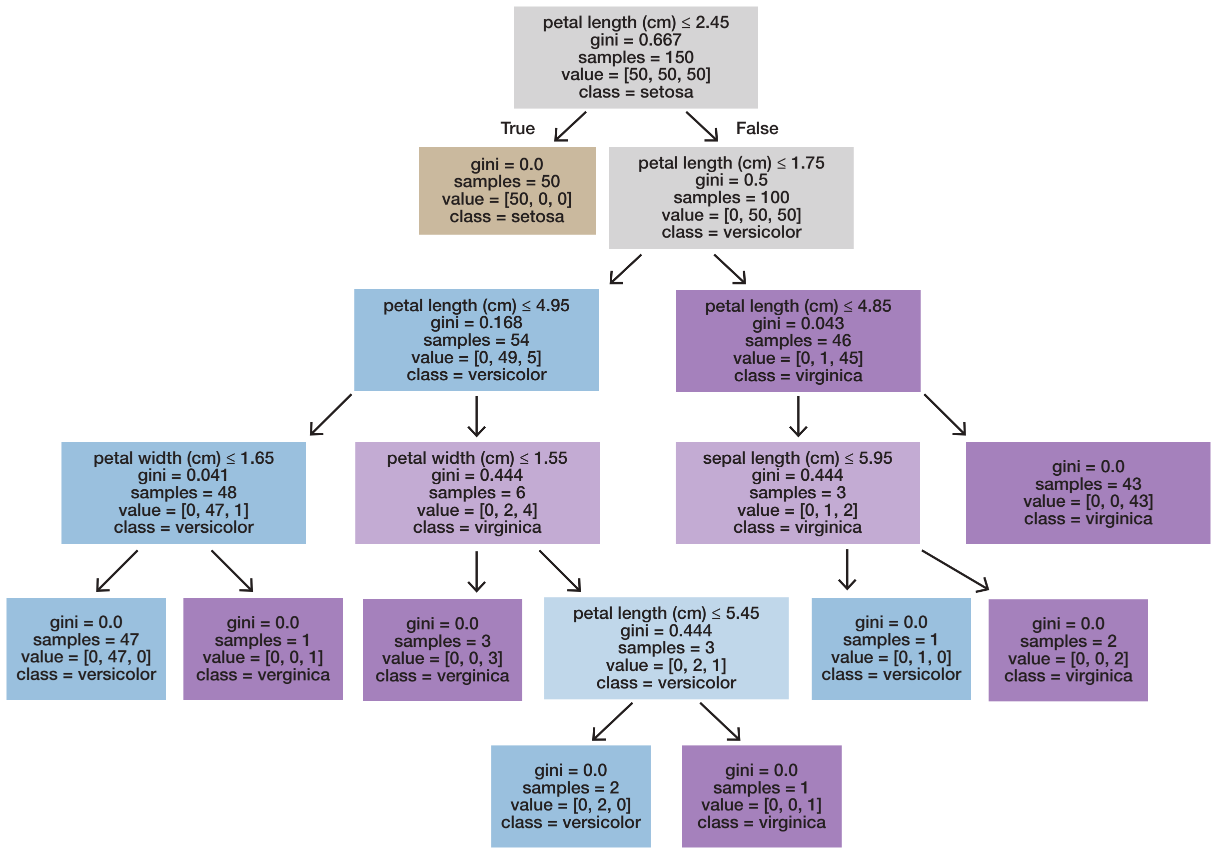


Figure 4-5 | Decision tree

points defining the straight line separating the classes are called support vectors, from which the algorithm draws its name [57].

One of the downsides with the hard margin classifier approach is that it can lead to overfitting since no errors are allowed. This often leads to a good performance on training data but less on other datasets. Not allowing for errors also means that in the case of no linear separability in higher dimensions, the algorithm does not find a solution. For non-linear datasets, a soft margin is used instead of a hard margin classifier. It introduces a variable which allows for classification mistakes. The variable is then tuned to optimize the performance of the algorithm [4].

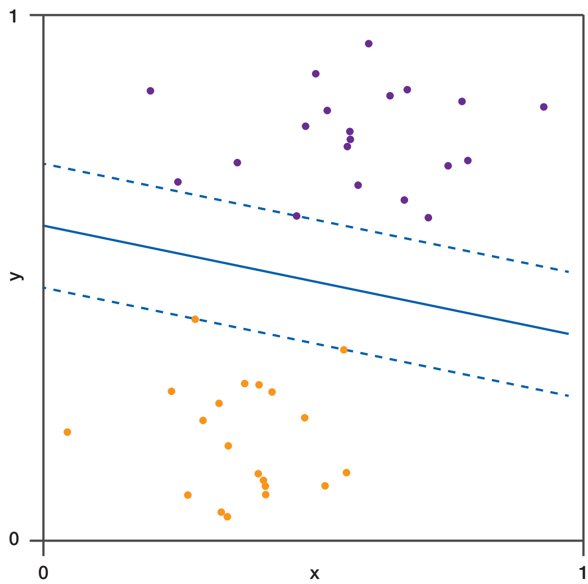


Figure 4-6 | Linearly separable data (support vector machine)

4.3.3 Naïve Bayes

Naïve Bayes classifiers are a class of supervised learning algorithms which are based on the Bayes theorem. There is a common assumption that all these algorithms share in order to classify data. Every feature of the data being classified is independent from all other features given in the class. Features are independent when changes in the value of one feature have no effect on the value of another feature.

Based on the class of a training example in the dataset, the algorithm calculates the probability of each feature belonging to that particular class based on the values. When classifying new data points, the algorithm calculates the class probability for each feature separately. For each class the product of those probabilities is calculated, and the class with the highest probability is chosen. This is a generative approach for classification.

While it is often argued that discriminative approaches are better for classification, it was shown that naïve Bayes classifiers can achieve their asymptotic error much faster than comparable classifiers [59].

Bayes algorithms are applied to many tasks, such as text retrieval or spam classification [60]. One huge advantage is their scalability for new features, which is especially useful for large datasets [61].

4.3.4 k-nearest neighbour

The k-NN algorithm is commonly used for supervised classification and regression but can also be applied to unsupervised clustering. The algorithm is called a lazy learner because data is just kept in memory until new data needs to be classified. New data is categorized according to the stored data points, thus always depending on the current training data.

The basic idea of the k-NN algorithm is to match the data accordingly to the k-nearest data points with the minimal distance (Figure 4-7). The choice of the number of data points (k) taken into account for the classification has great influence on the quality of the prediction. For regression applications, the prediction is calculated using the values of the k-NNs.

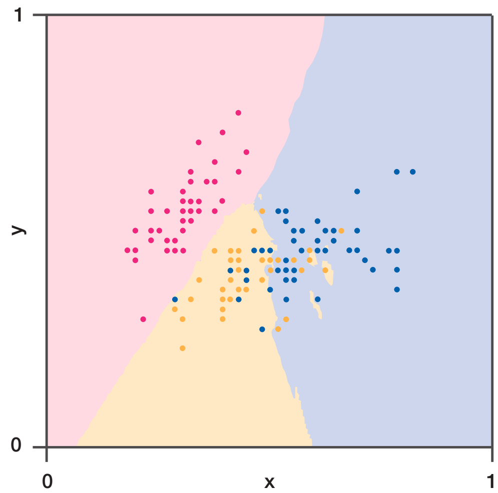


Figure 4-7 | k-nearest neighbours

There are many possibilities to improve the predictions of a k-NN algorithm. For example, it is possible to increase the influence of the nearest points by assigning weight to such points. This allows for more consistent classification. Depending on the property of the data, different distance measures provide better performance. To name only a couple of these, Euclidean distance has proven well for similar data types (length and width) and Manhattan distance seems to perform better when using different data types such as body, height and age.

Feature and sample selection is very important for k-NN algorithms. Filters and wrappers are among the most important feature selection techniques. The approach of filters is to remove irrelevant features without running the algorithm itself. It just searches for the dimensions best describing the structure of the data. Wrappers, on the other hand, identify the set of features most relevant for the performance of the predictions by estimating the results of the algorithm [62].

4.3.5 k-means

In clustering problems, an unlabelled dataset is provided, which the algorithm should automatically group into coherent subsets or clusters. The k-means algorithm as presented in [63] and [64] is one of the most popular algorithms for this kind of task.

The k-means algorithm works by randomly initializing k random points in the dataset, called the cluster centroids. The number k is chosen by hand or derived using an evaluation method. It then proceeds repetitively with two steps: assignment and centroid repositioning [65]. In the cluster assignment step, the algorithm iterates through each of the examples in the given dataset and assigns each example to one of the initialized centroids based on the closest distance. This is repeated for each data point until every example is assigned to a cluster. In the second step, the algorithm computes the average distance for each data point that is assigned to a specific cluster. The centroid is then moved to the calculated mean location. This step is repeated for all k clusters. The algorithm iterates until the cluster centroids no longer move, meaning the k-means algorithm has converged to k clusters (Figure 4-8).

The result of the k-means algorithm can be k clusters or less, depending on whether some of the centroids have data points assigned to them or not. The result of the algorithm can vary due to the random initialization of the centroids. When applied in practice, the k-means algorithm is deployed multiple times with different initializations of the centroids and varying numbers of the k number in order to get a useful result.

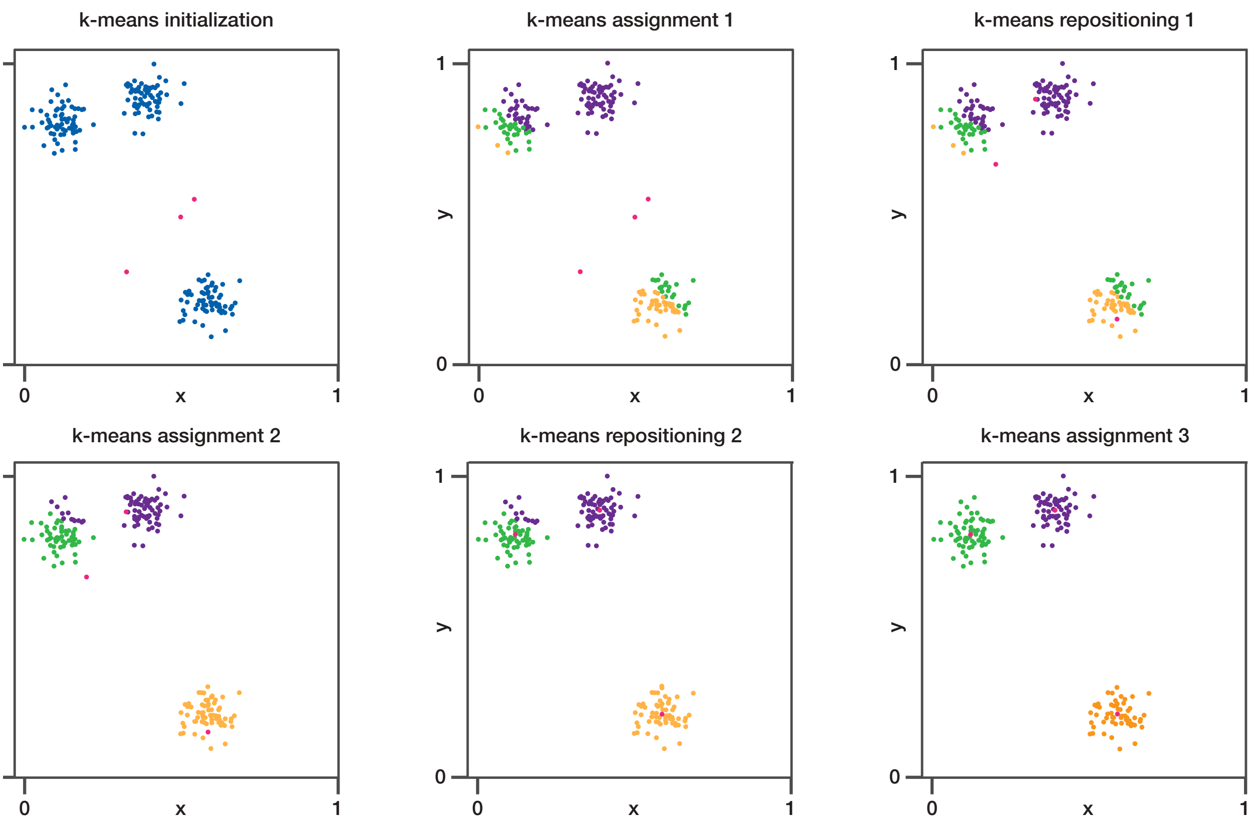


Figure 4-8 | k-means

4.3.6 Hidden Markov model

A useful algorithm for creating a probabilistic model of linear sequences is the HMM. The underlying concept for this algorithm is the Markov process, which assumes that the system can be described at any time as being in a set of distinctive states. At spaced discrete times, the system changes between states according to a set of probabilities associated with the states (Figure 4-9) [66].

Hidden states in a Markov model represent stochastic processes that are not directly observable but can only be indirectly observed through another set of stochastic processes that produce a sequence of observations [66].

Application fields of HMMs include sequence modelling in DNA and protein analysis [67], information retrieval systems [68], and audio sequencing [69].

4.3.7 Artificial neural networks

An algorithm based on neural networks (the perceptron) was developed in the early days of AI [3]. It can be applied to supervised and unsupervised learning. In general, ANNs are inspired by the human brain, however they do not copy its functions [70].

A neural network consists of different layers, each comprising artificial neurons that are connected to all the artificial neurons in the previous layer (Figure 4-10). The input layer represents the input data, which always consist of numerical values. It can process structured data, such as a temperature sensor output, and unstructured data, such as the pixels of an image. Depending on which units in the hidden layers are activated, the output layer unit provides a prediction. In the case of image recognition, this could be for example a monkey identified in the image.

Artificial neurons represent the units of each layer. They handle the input data and make prediction possible. The input of the perceptron is either the original data from the input layer or, in a DNN (which has more than one hidden layer), the transformed inputs from artificial neurons of previous hidden layers. Each input is adapted by a specific weight. The weighted input is then processed and summed up in the cell body. A bias (fixed number) is added as a tuning variable, as illustrated in Figure 4-11. The output of the cell is then used in the activation function (in this case a step function) representing the input for the next layer.

To train the network, the weights are randomly initialized. Next, the first training datasets are fed into

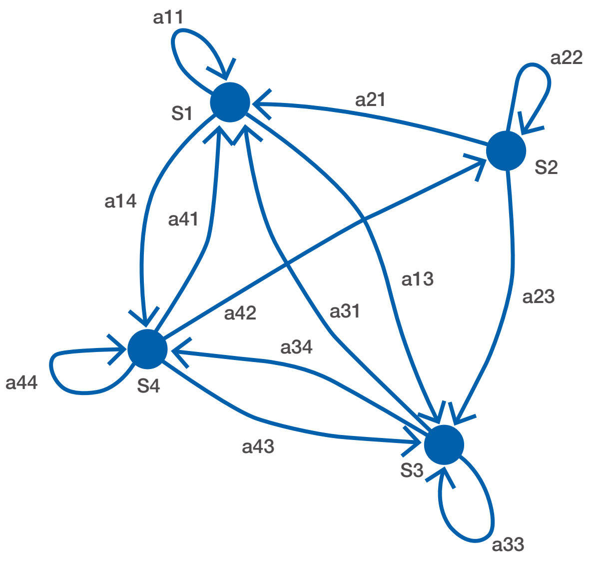


Figure 4-9 | Hidden Markov model with 5 states

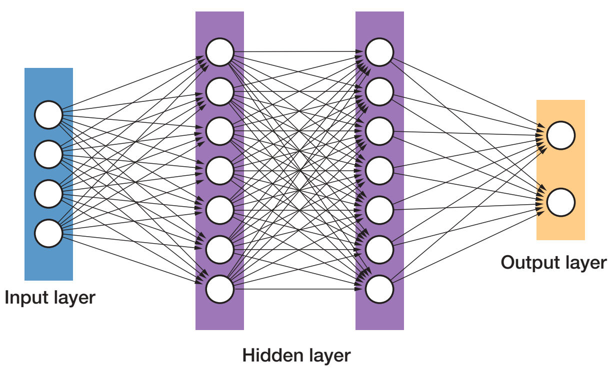


Figure 4-10 | Artificial neural network

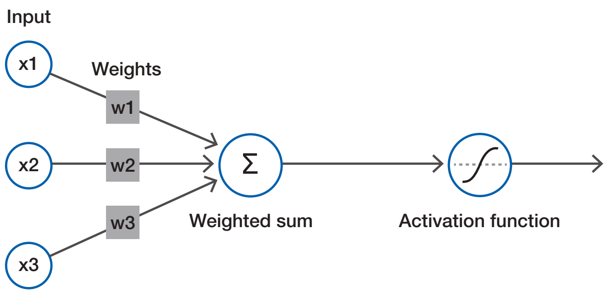


Figure 4-11 | Neuron/perceptron of an artificial neural network

the neural network. The outcome of the training case is then compared to the actual desired outcome. An algorithm called backpropagation then updates the weights. However, if many neurons are stacked together, it becomes very hard to control the changes in the final output of the whole network [71]. The essential step to ensure that a neural network functions is to make the output change smooth with the help of a continuous activation function. One of the main advantages of ANNs is their universality.

4.3.8 Convolutional neural networks

CNNs have many similarities with ordinary ANNs. Likewise, they consist of neurons that have weights and biases that are adjusted in the learning process. The whole network still expresses a single differentiable score function and has a cost function attached to the last fully connected layer. However, contrary to normal feed-forward neural networks, CNNs operate on the explicit assumption that the inputs are images, which allows them to encode certain properties into the architecture of the network. This makes the implementation of the forward function more efficient and vastly reduces the number of parameters [72].

A CNN usually consists of three types of layers: a convolutional layer, a pooling layer and a fully-connected layer [73]. A convolutional layer takes advantage of the spatial relationships of the input neurons. The inputs of a convolutional neuron come from a specific area of neurons in the input layer. This narrowed-down receptive field allows the convolutional network to work in a far more focused manner than a conventional neuron. The receptive field of the neuron is equivalent to the filter size. The extent of the connectivity along the depth axis is always equal to the depth of the input volume. The connections are local in space, but always full along the entire depth of the input volume.

The pooling layer is commonly inserted in between successive convolutional layers in the network architecture. Its function is to reduce the spatial size of the representation to decrease the amount of parameters, and also to control overfitting. The pooling layer operates independently on every depth slice of the input and resizes it spatially.

Neurons in a fully connected layer have full connections to all activated neurons in the previous layer. This is also common in regular feed-forward neural networks. As a result, their activations can be computed with a matrix multiplication followed by a bias offset. Most CNN architectures consist of varying numbers of convolutional, pooling, and fully-connected layers.

4.3.9 Recurrent neural networks

RNNs are a special type of ANNs. They can be applied to supervised and unsupervised but also reinforced learning. While ANNs take their current input data into account assuming independence from previous data, RNNs are able to consider previous data. While the neurons of an ANN have only the inputs from previous layers, the neuron of an RNN has dependencies on its previous outputs, because such outputs have loops (Figure 4-12). This enables this type of algorithms to include sequence prediction problems, for instance context of words or temporal aspects [74].

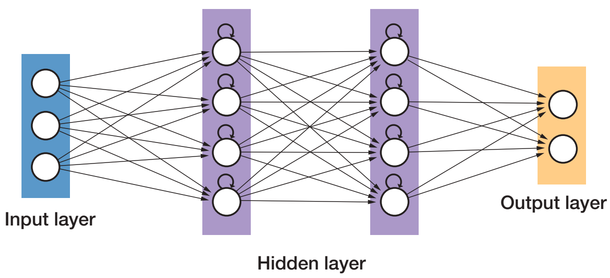


Figure 4-12 | Recurrent neural network

This also means that during the training phase the order of input data plays an important role. In practice, RNNs receive an input and compute their state by using current and prior inputs. This is repeated until all stored previous states are processed and the output is calculated. During training, the obtained result is then compared to the actual correct result. The weights of the network can then be updated. The only difference with regard to common ANNs is the fact that the backpropagation has to take all the stored previous time steps into account. An overview of how backpropagation through time works is given in [75].

Since in reality it is not possible to store all previous steps in a common RNN, information is lost over time [76]. To tackle this problem, a variety of architectures have been developed, such as bidirectional RNNs, LSTMs and gated recurrent units (GRUs). Bidirectional RNNs not only consider previous but also future elements. To do so they use two neurons, one for the forward loop and one for the backward loop. They are then connected to the next forward neuron and vice versa (Figure 4-13) [77].

LSTM networks include so-called gated cells, where information can be stored. During prediction the cell decides which information will be stored, used or forgotten. The input and output gates let information pass or block them according to trained weights. By combining the current input, the previous state and the memory of the cell, this architecture is able to identify the long-term dependencies in datasets [79].

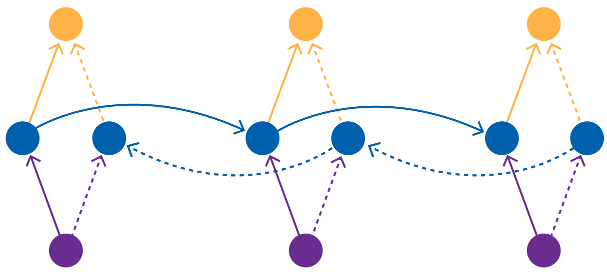


Figure 4-13 | Architecture of a bidirectional recurrent neural network

Section 5

Deployment of artificial intelligence

With the maturing of chips, datasets and platforms, the technical and industrial foundations for AI have gradually strengthened to support a wider range of real-life applications. Technological breakthroughs in algorithmic developments such as image recognition and NLP will continue to enlarge the number of use cases served by AI. McKinsey expects that by 2025 the global AI application market will reach a total value of USD 126 billion, and AI will become the breakthrough point for the development of many smart industries [78].

As one of the core driving forces for the transformation of many traditional industries, AI has given rise to new technologies and products that build upon progress made in semantic recognition, speech recognition, face recognition and the like. A broad view of industry needs on different AI capabilities is summarized in Figure 5-1 using a heat map to suggest the importance of the needs for a specific industry. The value scale is between 0 and 100, with 0 meaning that the feature is not needed and 100 meaning that the feature is of utmost importance for the industry sector. The table illustrates a mapping of exemplary AI systems on different industry domains.

	Computer vision	Anomaly detection	Time series analysis	Natural language processing	Recommender systems
Smart Home	80	50	30	100	75
Smart Energy	10	100	100	10	50
Automotive	100	75	25	80	50
Smart Manufacturing	75	100	100	10	25
Transportation/ Logistics	100	75	75	10	50
Financial	50	100	100	20	20
Healthcare	100	100	60	20	100
Smart Farming	80	75	100	10	90

Figure 5-1 | Relevance of common AI systems on different industry domains

5.1 Artificial intelligence in smart homes

With the development of AI technology, the smart home domain has gradually evolved towards the concept of a control centre that connects individual devices scattered across the home to form a complete smart home ecosystem. Built upon the IoT, a smart home ecosystem is composed of hardware (e.g. smart appliances, security control equipment, furniture), software systems, and cloud-based platforms. It integrates speech recognition, visual recognition, deep learning domain models, user portraits and other technical means to actively understand the needs of users.

Smart homes aim at achieving device interoperability and device self-learning, and through the collection and analysis of user behaviour data can provide personalized services to make homes safer, more comfortable and more energy-efficient. At the same time, such systems can also improve the efficiency of home appliances, reduce energy and natural resource consumption, and create a more sustainable and healthier home style. The smart home industry can also promote the evolution of the incumbent home appliance market and contribute to the continuous development and industrialization of AI.

Major home appliance manufacturers are today actively developing smart home solutions. Mature applications include smart refrigerators, smart air conditioners, smart washing machines, smart water heaters, smart kitchen appliances, smart speakers, and many other smart appliances reflecting the concept of “all things connected”. Companies have developed products that can interconnect and mutually control various home appliances and gather large amounts of data for prediction and analysis tasks. Internet-based capabilities have generally been well received by consumers.

The fields of smart homes and AI are closely integrated and continuously developing. Recent advances in machine learning, pattern recognition and IoT technology have brought interactivity to higher levels, making residential devices more intelligent and user-friendly. Products are evolving gradually from being mobile phone-centric to focusing on new and innovative human-machine interaction modes. Looking ahead, AI technology will enable smart homes to shift from passive to active intelligence and may even substitute people in some of their decision-making tasks. Most industry stakeholders today foresee AI opening broader and exciting opportunities for the smart home market.

Three smart home scenarios are described in subsections 5.1.1 to 5.1.3 below: smart television control system, bathroom self-service system, and intelligent food identification system.

5.1.1 Smart television control system

Smart television control is aimed at providing intelligent, personalized, and/or energy-saving television services. Face recognition techniques have significantly matured thanks to improvements in imaging technology. Compared with fingerprints, face recognition offers a higher degree of accuracy, and algorithms are less impacted by environmental factors such as light and noise. A smart television control system can gather face images of family members through a built-in camera and provide personalized services. For instance, the system can switch to a preferred television channel depending on who is sitting in front of the television. Parental control can also be automatically activated when a child is watching the television.

The system can calculate the optimal viewing distance depending on the current scene. Volume, brightness and saturation of the screen can be adjusted automatically to provide the viewer with the optimal watching experience. The television can also detect if the viewer has fallen asleep and subsequently turn down volume, brightness and saturation. When no one is watching, the system can make an inquiry, and if no response is received, can shut down automatically.

5.1.2 Bathroom self-service system

The bathroom is a relatively private, interactive and frequently-used area for households. Bathing may require people to control the temperature of the water heater. But due to water vapour, the equipment cannot be directly touched. Therefore, non-contact control capability might provide value. At the same time, people may have other needs during bathing, such as hairdressing. In this context, smart home services may enrich the life of modern families and address diversified household needs across various geographies and cultures.

As an example, an intelligent self-service bathroom system for the Chinese market provides users with convenient, interactive capabilities and rich information through multi-directional voice interaction. While bathing, the user can conveniently use his or her voice to control the angle of the mirror, the intensity and angle of the shower spray and the temperature. At the same time, the user can also obtain content recommendations regarding beauty, music, fitness and the like from other home appliances such as pads through voice control.

5.1.3 Intelligent food identification system

A smart food ingredient recognition system is illustrated in Figure 5-2. It is equipped with a high-definition camera to perform identification of ingredients inside the refrigerator. Through visual identification of such ingredients, the system supports one-click shopping, ready-to-buy shopping, planning a reasonable diet, and early reminder of food ingredients. It can provide several benefits to the users, including convenient operation, real-time recording, advising healthy eating, and optimizing household expenses.

Real-time interaction with the refrigerator can be achieved through a mobile phone. The user can assess the current content of the refrigerator remotely and plan the purchase of food in advance and from any location. Other benefits include reminders of food consumption deadlines, rational use of refrigerator space, or one-click purchase of ingredients without the need to go to the supermarket. At the same time, daily food consumption is recorded in a cloud database and combined with big data analysis to provide an assessment of eating habits and advice on balanced and healthier diets. Based on eating habits, seasonal supplies and dietary restrictions, the system can also recommend personalized recipes.

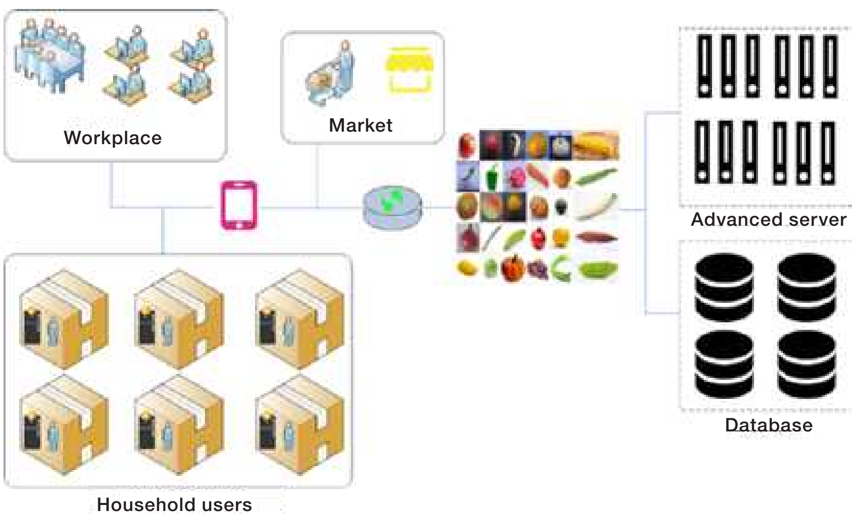


Figure 5-2 | Intelligent food identification system

5.1.4 Challenges for smart homes

Smart homes are building upon the progressing maturity of the IoT, adding AI to the home automation field, but technological limitations exist, such as real-time learning, detection probability, and embedded processing power. Other challenges are adding multi-user support as well as data availability issues that may cause difficulties in collecting sufficient reliable data for machine learning. Machine learning usually treats the most common content in data as truth, while excluding statistically rare content. This cognitive method is still different from that of humans and may lead to the occurrence of cognitive bias and misunderstanding.

In terms of the need to solve specific problems, the proportion of actual, reliable and credible data in the total amount of data may fail to reach the lower limit of machine learning requirements. This is especially true for a smart home environment, where most data comes from small IoT devices, such as sensors or lamps. The algorithms in a smart home need to be able to adapt to different data streams. They should be robust in responding to changes in data streams, for example when a user changes the location of a television from the living room to the sleeping room. Thus, a high degree of robustness is required.

Issues such as personalization, privacy and security are also important. Personalization and privacy could be based on user profiles, customization of services, and device settings in daily life. The security perspective therefore needs to be addressed, because smart terminals may be illegally invaded and controlled, and personal information can be revealed. In addition, AI-enabled systems may also harm humans.

When equipping critically private areas, such as bathrooms, with smart devices, a guarantee must be included that information will be handled confidentially. Nor should further information derived from the data interfere with people's privacy. While such persons might agree to the collection of energy data in order to optimize consumption, they might see the analysis of their daily routines based on such information as an intrusion of their privacy.

5.2 Artificial intelligence in smart manufacturing

Smart manufacturing is fundamentally the integration of ICT with advanced manufacturing techniques. All manufacturing activities are potentially impacted: design, production, management and service.

A smart factory is a networked factory in which data from supply chains, design teams, production lines and quality control is linked to form a highly integrated, intelligent platform that will help in redefining future production lines. With the growing need for flexibility to suit a diverse range of production domains, the manufacturing sector has to rely on automation, machine learning and other fields of AI to meet these rising challenges.

Through machine learning, systems have the ability to learn from experience, with the result that they are constantly improving. This enables manufacturing to be faster, more flexible and specifically scalable by providing predictive insights to manage everything from plant effectiveness to selecting optimal suppliers and gauging pricing against demand [87].

Another benefit of AI in manufacturing is support of economic growth, whereby AI is used to manage capital efficiency, including labour requirements and machinery schedules to realize on-demand production, improve operating efficiency, shorten product cycles, enhance production capacity, reduce downtime and ultimately achieve cost savings.

Some key applications of AI for the manufacturing sector are described in subsections 5.2.1 to 5.2.4 below.

5.2.1 Predictive maintenance

Traditional manufacturing lines may have already produced many unqualified products when the production equipment fault alarm occurs, resulting in losses to the entire enterprise. By virtue of the equipment operation data collected in real-time, predictive maintenance can identify fault signals through machine learning algorithms so as to achieve early detection and maintenance of defective equipment. Ultimately, this would reduce maintenance time and equipment costs, improve equipment utilization and avoid the losses caused by equipment failures [88] [89]. Fault prediction and fault localization and diagnosis are two key mechanisms for predictive maintenance.

- **Fault prediction**

The key performance indicators (KPIs) of a device or network in a factory usually indicate a gradual deterioration trend. Any hardware or service failure is usually preceded by an unstable or degraded operating state (Figure 5-3). Passive processing following a failure not only affects the service experience, but also takes a long time to troubleshoot.

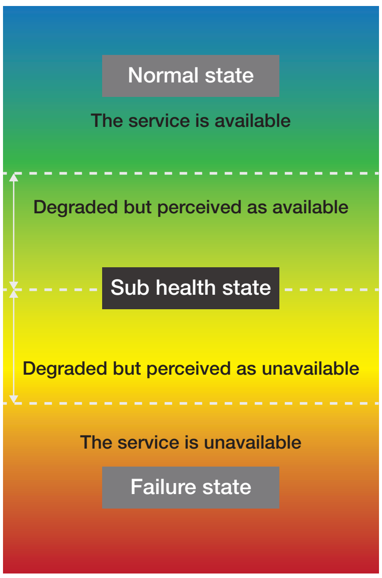


Figure 5-3 | Device/network health

Through the KPI prediction of a device or network, service disruption, network resource inefficiency and deterioration can be prevented. There are two main methods for device or network fault prediction, illustrated in Figure 5-4:

- Case 1: The black curve reflects the current device or network health status before time point “Now”. The red curve is a trend that can be obtained from historical data of the black curve. An alert will be raised when the trend falls below a threshold.
- Case 2: The black curve is the same as in Case 1. The red curve is a prediction curve based on a prediction algorithm and historical data from the black curve. If the deviation exceeds a threshold, an alert will be raised to report the anomaly.

- **Fault localization and diagnosis**

If the device in the factory is defective or improperly operated, it may cause the fault to spread and generate network or service faults. Quick location of the fault helps to shorten the fault recovery time and reduce losses.

After the fault of the device/network/service in a factory has occurred, network performance and status data, such as log, alarm, KPI, configuration and topology, are collected. Then, correlation analysis is performed to determine which metrics or parameters are abnormal, so that the faulty device can be quickly located and the root cause of the fault can be identified (Figure 5-5).

There are two directions of correlation analysis for fault localization and diagnosis:

- In the horizontal direction, according to the network service topology, the metrics of all the devices on the path need to be put together to perform the related line analysis.
- In the vertical direction, the physical layer metric, network layer metric and application layer metric of the device need to be put together to perform the related line analysis.

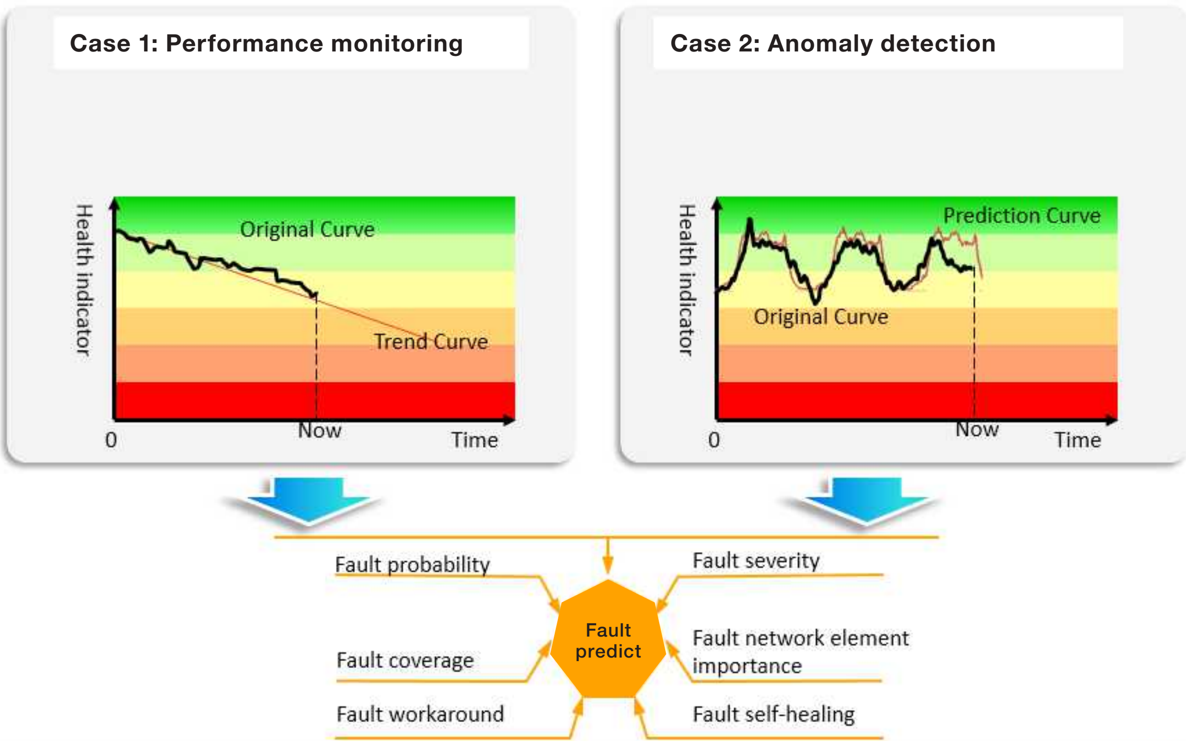


Figure 5-4 | Two cases of device/network fault prediction

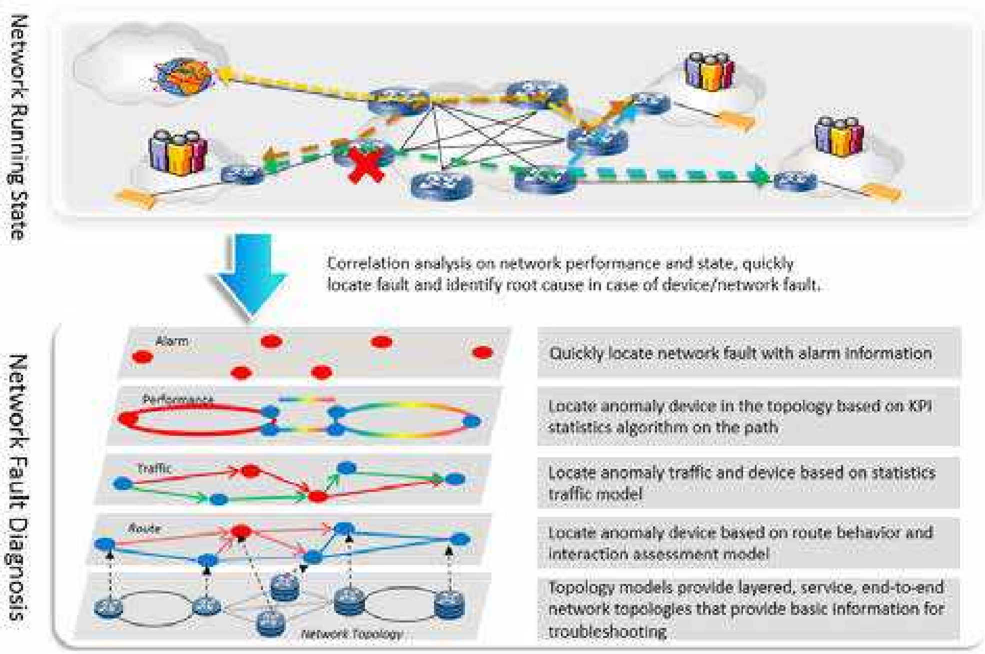


Figure 5-5 | Fault localization and diagnosis based on correlation analysis

5.2.2 Collaborative robots

Another use case for AI results from the development of collaborative robots, also called cobots [90]. Cobots are industrial robots capable of working together with humans in the production process without being spatially separated from them by protective devices. The unique feature of these robots is that they can work directly with humans.

There is an important prerequisite for this type of human/robot collaboration: it must be ensured that the robot cannot cause injury to humans. This is often not a problem for small, lightweight robots. However, the more difficult the tasks the robot has to perform when working with humans, the more relevant this requirement becomes. For example, robots that lift heavy castings can cause considerable injury to humans due to their size and weight.

To minimize the enormous risk potential posed by such large robots, conventional protections via sensors are often no longer sufficient. In this context, the use of AI prevents various opportunities. Techniques such as gesture and intention recognition can be used to adapt the robot to human behaviour. For example, the fusion of sensor data collected by cameras or radars can be evaluated so that the robot can dynamically adapt to its environment. The aim is not only to let the robot react to its environment and the surrounding people, but also to actively prevent accidents.

This fused sensor data can be used to classify the worker's currently executed gesture using algorithms. This means gestures can already be identified and recognized during their execution, before they are fully completed. Thus, the indicated gesture of the worker can already give the aligned AI system an idea of what the next work step is. This result can be used to preventively adjust the reaction to the worker's behaviour. This process is illustrated in Figure 5-6 [91].

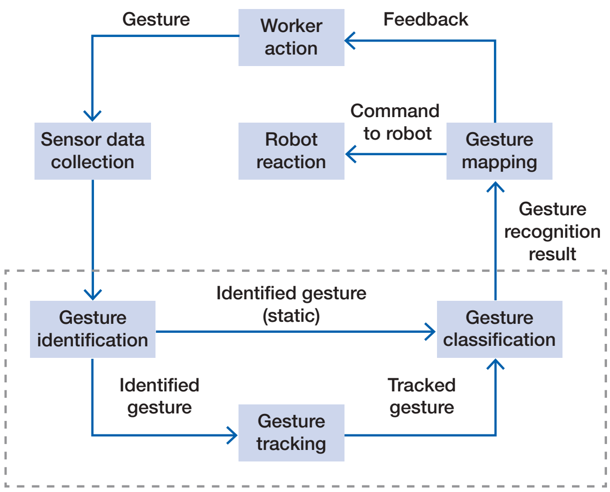


Figure 5-6 | Human robot interaction

Gestures are often visually recorded, usually via the skeleton. The advantage of machine learning resides in the classification of gestures. While ensemble methods have often been used until now, machine learning or deep learning algorithms are increasingly gaining popularity due to very good results [92].

The problem of collaborative robots becomes particularly complex and safety-relevant when several robots have to interact with a human and with each other. In such a case, the robots must react not only to the worker's behaviour, but also to the behaviour of the other robots to avoid collisions.

5.2.3 Quality control

The detection of defects such as surface defects, internal failure and edge damage of products is traditionally carried out by human vision. This can lead to a high defective rate due to fatigue caused by high working strength, especially in the chip industry, household appliance industry or textile industry. Intelligent online detection techniques depend on sensors to acquire product images and rely on computerized vision algorithms to improve detection speed and quality and avoid loss caused by leak detection or faulty detection. Automated quality control can greatly reduce the fault rate in applications such as chip manufacturing. At the same time, through analyzing what has caused the products to be defective, it can decrease the rejection rate of products and optimize product design and fabrication to lower quality control costs.

Automated image recognition (AIR) is an illustrative example coming from back-end manufacturing of integrated circuits. AIR systems were initially developed to address the need for automatic detection of images using X-ray machines. Through the combination of computer software and imaging hardware technology, the accuracy and reliability of visual detection can be improved and optimized to avoid too much dependence on human eyes.

X-ray machines are used in this application to depict the internal structure of semiconductor devices. In general, a batch of products are equipped with 20 to 50 reels and one reel can cover all devices after shooting 5 to 10 images. Images will finally be saved in a storage server connected to the internet. An example of image pattern is shown in Figure 5-7.

The AIR server consists of several functions, such as the core image identification algorithm and operator call, preliminary image evaluation (only qualified images are allowed for further automatic identification), tracking of the cut-off time of image shooting, product information matching, etc. Ineffective images will be automatically picked up and rearranged for the follow-up secondary manual visual detection filtering.

The automatic identification module of the server filters the original images. The small amount of unsure failed devices filtered by the system is then rechecked through human eyes, in order to detect the actual failed devices. Figure 5-8 shows an example of an unsure device list and its locations after rearrangement. Operators identify the actual failed devices and mark them on the interface for the convenience of detecting out the material objects in the reel.

The client is provided with a user interface for additional reworking, which is used to require recheck of defective products through visual detection. In terms of acceptance check of image quality, the client is provided with the capability for the end user to assess the quality of images (brightness, grey scale, contrast) produced by

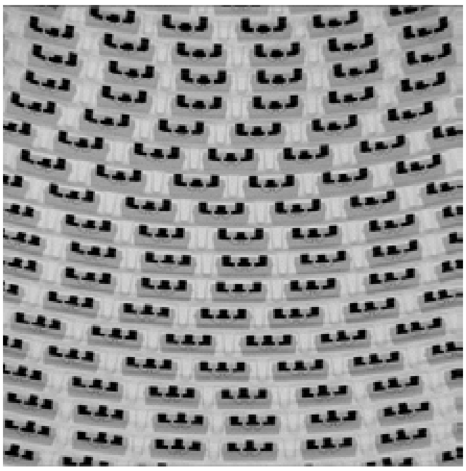


Figure 5-7 | Original image of pattern shot

Figure 5-8 | Visual detection

every X-ray machine. The system will remind the user to adjust the shooting parameters so as to meet the requirements of automatic identification.

5.2.4 Challenges in smart manufacturing

Smart manufacturing brings a number of challenges, in particular when applying AI-related technologies. One of the key concerns can be summarized as “humans in command”. To avoid any abuse or misuse, smart manufacturing should adopt a humans-in-command approach to AI that requires that machines remain machines and people retain control over these machines at all times.

The biggest challenge in using AI for cooperative robots is the robustness of data acquisition and algorithms. It must always be ensured that the algorithms behave as intended. Especially when working cooperatively with people at risk, the overall system must always behave in a predictable manner. This is still a big challenge, particularly with neural networks, which can lead to unexpected outcomes, especially with unforeseen variations of inputs. This problem occurs, for example, when the neural network falls into a local minimum. Network decisions resulting from these local minima are often not intended by the developer.

While the development of AI and robotics can increase industrial competitiveness, which in turn could lead to the creation of new jobs, there is also a significant risk that work currently done by humans may be increasingly taken over by robots. In light of the changing dynamics of labour markets, education and training (including vocational training, lifelong learning and training and re-training at work) will become even more important in the future. New skills and competences will be required to deal with the increased use of robots and automation.

Other challenges in smart manufacturing include the availability of large amounts of high-quality data to train AI algorithms, as well as the need to structure these data into meaningful information and domain models.

5.3 Artificial intelligence in smart transportation and the automotive sector

Smart transportation has a very broad market potential (e.g. private vehicles, public transportation, parking, logistics, emergency services) and is largely acknowledged to require AI technologies. With the dramatic increase in the number of vehicles, especially in large cities, services such as traffic management and congestion control, and the recent surge of self-driving vehicles, all call for the large-scale use of AI technologies (e.g. image analysis, route optimization, object identification). Capabilities offered by AI will not only be deployed in end devices (e.g. mobile phones) but also on the edge (e.g. cars) and on the cloud (data centres), as illustrated in Figure 5-9.

5.3.1 Autonomous driving

A self-driving car, also known as a driverless car, is a vehicle that uses an intelligent computer system to implement autonomous driving. Such vehicles rely on intelligent path planning technology, computer vision and global positioning system technologies to enable on-board computing systems to operate safely and without human intervention.

Self-driving vehicles are developing rapidly, with industry giants and start-ups ready in the starting blocks to plan or release autonomous vehicles in a short-term timeframe. Five levels of vehicle automation are defined by the National Highway Traffic Safety Administration (NHTSA) [95]:

- Non-automation (level 0). Driver in complete control at all times.
- Function-specific automation (level 1). Automation of one or more specific functions.
- Combined function automation (level 2). Automation of at least two primary control functions designed to work together.
- Limited self-driving automation (level 3). Driver can cede full control of all safety-critical

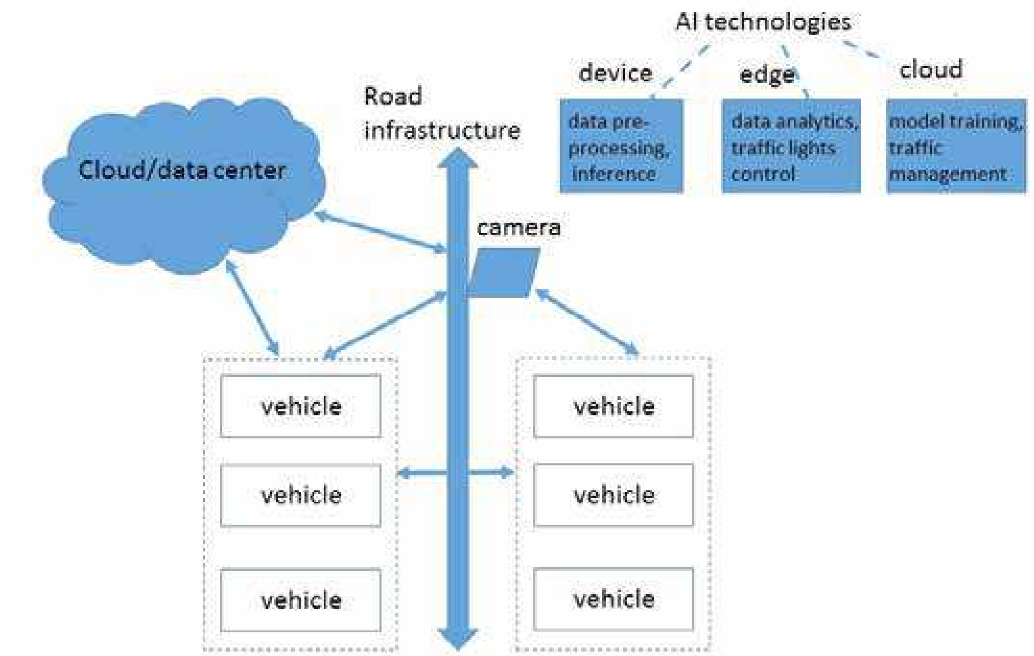


Figure 5-9 | Application of AI technologies in smart transportation

functions under certain conditions. The vehicle monitors for conditions requiring transition back to drive control.

- Full self-driving automation (level 4). Vehicle is designed to perform all safety-critical driving functions.

Clearly, the higher the level of automation is, the more responsibilities of driving will be handled by vehicles. This implies that AI technologies are required both inside the vehicles and in the cloud (Figure 5-10), including but not limited to:

- Taking control of the vehicle. This includes starting, braking, turning and other auto-piloting capabilities, without the intervention of humans (removing the burden on drivers or intervening when the human is incapable of doing so).
- Identifying driver status, car condition, road condition and surrounding environment (e.g. pedestrians, animals, obstacles). The identification and analysis of surrounding objects require extensive machine learning capabilities. Furthermore, the analysis needs to be performed with a very low latency to facilitate driving. As a matter of fact, traditional general-purpose CPUs will not be able to provide the required computing performance and efficiency for AI algorithms in such scenarios. Dedicated accelerators are needed (both within the vehicles and in the cloud) for model training, inference and emulation tailored to AI workloads.
- Enabling fleet management, or the autonomous piloting of an organized group of vehicles, which can be extensively used in logistics and delivery services.

5.3.2 Traffic management

Traffic management plays an essential role in smart transportation. With population and vehicles growing across the globe, managing the heavy traffic in urban areas has become a challenging

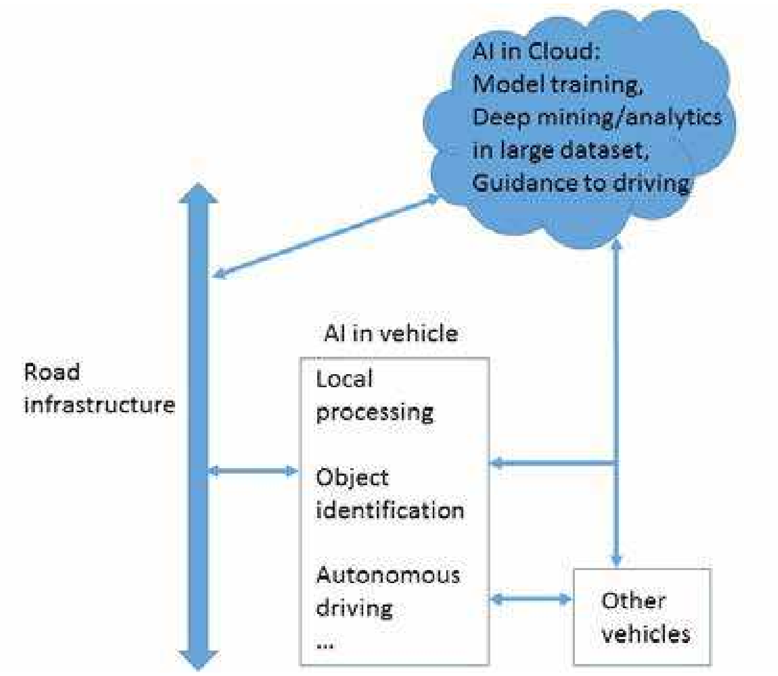


Figure 5-10 | Enabling autonomous vehicles with AI technologies

yet essential task for governments, police forces as well as car companies. According to a report by The Economist [94], expenses caused by traffic congestion already amounted to USD 200 billion (0,8% of GDP) in 2013 across France, Germany, the United Kingdom and the United States. Therefore, optimizing traffic flow, reducing traffic jams, and minimizing the emissions of vehicles will substantially contribute to increasing productivity, quality of living, and environmental protection. To achieve these objectives, wider and better utilization of AI technologies is needed.

Examples of applications include:

- Traffic flow analysis. By using machine learning and data mining, real-time traffic data (e.g. on vehicles, pedestrians, congestion, accidents) in multiple streets or in a wider area can be analyzed and cross-examined for traffic route optimization, traffic control to avoid congestions, and reduction of emissions.
- Optimization of traffic lights. Instead of statically determining traffic light switching (which today does not consider real-time traffic conditions), AI algorithms can be utilized to analyze and predict real-time traffic situations, and provide dynamic traffic light switching to optimize the passing of vehicles and pedestrians.
- Inspection of violation of rules and regulations. These tasks traditionally involve intensive human labour. Even with AI algorithms in image and video processing, performance is still limited due to a lack of resources and computing power. However, with more powerful processing platforms accelerating AI workloads, videos and images can be analyzed in much higher volume, leading to reduced labour cost, higher accuracy, and enhanced performance.

Other examples include smart parking, surveillance cameras or smart logistics.

5.3.3 Traffic robots

Another development of AI in transportation is the use of intelligent traffic robots, which are deployed for road junction traffic control. Using AI technology, such robots are able to monitor traffic conditions, obtain traffic information, and perform road traffic commands at challenging road intersections. They can also remind pedestrians to abide by traffic regulations, enhance pedestrian safety awareness, and reduce the workload of traffic police through arm commands, light tips, voice alerts, and safety message delivery. In addition, robots can use image recognition to record pedestrian violations to enhance overall safety at road junctions.

5.3.4 Challenges in smart transportation

Safety, security and privacy are some of the major concerns of users for smart transportation. With sensors built into each vehicle and the implementation of advanced AI technologies for self-driving, the data and even the destiny of human beings fall into the hands of computers. Recent accidents of autonomous cars have further raised concerns. Safety, security and privacy issues are therefore among the foremost challenges for smart transportation.

Policy, laws, regulations, and standards related to smart transportation constitute another set of important challenges that will need to be addressed. Standards may be related to technology implementations of smart transportation, for instance to ensure secure transmission, storage and processing of collected data. Regulations or laws may be related to the use of AI in smart transportation, for example by car manufacturers or service providers.

Another challenge is the availability of heterogeneous computing platforms tailored to AI workloads on cloud, edge and end devices to: 1) integrate an increasing amount of data from vehicles, pedestrians and infrastructure, and be able to transmit, store and analyze these data efficiently; 2) be able to run AI workloads (e.g. model training and inference) rapidly and efficiently, and support the (semi-)real-time decision-making of traffic management and autonomous driving. On the one hand, platform architectures need to be improved or innovated to catch up with the evolution of AI workloads and algorithms. On the other hand, performance needs to be constantly enhanced to satisfy the increasing demands (e.g. latency, data volume, connectivity) of users and applications. In addition, platforms also need to be scalable to integrate new components or functions.

Eventually, there will be a pressing need for full AI stacks. Capabilities such as image analysis and face recognition may be required in all subsystems (i.e. end devices, edge and cloud). For instance, drivers may install AI-enabled software in their mobile phones to pre-process data related to driver status. An edge component (e.g. built-in within a vehicle) can process a larger amount of data generated by the vehicle and collected from the environment. In the meantime, the cloud/data centres will gather a huge amount of data from all the vehicles, traffic flows and the environment as well as historical data, to perform analyses such as route optimization, traffic guidance, or other forms of macro-management.

5.4 Artificial intelligence in smart energy

The energy sector can be classified as primary energy (e.g. oil, gas, coal) and secondary energy (e.g. electricity). As previously mentioned, the exhaustion of natural resources and impact of climate change are becoming a not too distant reality. To resolve these problems, countries across the world are implementing counterstrategies focusing on the optimization of energy supply and demand management.

The purpose of smart energy is not only to increase the yield rate in energy production, transmission and consumption, but also to enable efficient energy management. The emphasis so far has been on reducing energy losses to the largest extent. ICT convergence, including the growth in IoT, big data and AI, is becoming the driving force to achieve technological innovation in the energy sector. It can improve the estimation of supply and demand during production and consumption phases and enables services for transacting energy between different buildings and equipment. Moreover, through analyzing information collected via communication technologies, optimization of energy savings has the potential to reach a level never seen before. According to IDC, the smart energy market may reach USD 516 billion by 2020.

5.4.1 Grid management and operations

A smart grid is an intelligent electrical grid system that improves efficiency in energy production, transmission and consumption by enabling interactions between suppliers and consumers through the use of ICT. While in the traditional electrical grid electricity is unilaterally supplied from large power plants to the final customer, the smart grid is subject to bidirectional electric power flow. In other words, the final consumer is both a consumer of external electricity when necessary but also a supplier who can provide electricity back to the grid when there is a local surplus, for instance due to renewable production.

AI technology can be used to support real-time monitoring, operation and maintenance of power equipment within the smart grid. It can provide fault diagnosis capabilities as well as remediation measures in the initial stages of failures in order to improve equipment operation stability and power generation efficiency. Machine learning allows to identify minor pattern changes in various operating conditions to implement effective preventive maintenance.

Leveraging the powerful prediction capabilities of machine learning, various utilities and large corporations across the world have managed to optimize the efficiency of their energy infrastructure by typically 10% to 15%. AI can also predict with high accuracy the peak demand and supply of electricity, in particular when dealing with distributed renewable energy generation. This in turn can help consumers cut down significantly on their energy bills.

5.4.2 Consumer engagement and services

Smart grids are typically equipped with distributed advanced metering infrastructure (AMI) and smart meters which support bidirectional communication between producers and consumers of energy. Web portals can be used to gather and display various energy-related data as well as identify consumption habits in order to adapt electricity production and consumer prices using various AI algorithms. On the other end, advanced platforms allow large industrial consumers to perform more sophisticated trading of electricity fitting their specific business needs.

In addition, AMI offers opportunities for new services such as monitoring the elderly or people with disabilities through remote analysis of electricity usage patterns. Scenarios may include alerting relevant persons or bodies if for instance the light is not turned off late in the evening or if the electricity usage does not increase in the morning.

5.4.3 Integrated smart energy platforms

Building upon the need for further consumer engagement, many utilities across the world are deploying smart energy platforms that integrate a wide range of services that can be enabled thanks to AI technologies. One example of such a platform is HUB-PoP (hyperconnected ubiquitous bridge – platform of platforms) from the Korea-based utility KEPCO. As shown in Figure 5-11, HUB-PoP provides a unified, cloud-based platform connecting various subsystems such as power grid operations, management support, customer

Figure 5-11 | HUB-PoP based on AI for power management

management and new energy businesses including renewables. Data analysis and AI algorithms can be executed on the platform to provide the customer with various innovative services.

One of these services is an automated chat bot dialog system allowing the processing in parallel of multiple customer requests. Using machine learning techniques, the chat bot is fed with knowledge accumulated through consultations and online inquires in order to be continuously improved.

5.4.4 Challenges in smart energy

As illustrated by a few use cases, AI technology can be applied to various applications and services in the energy sector. However, given the scale and longevity of energy infrastructures, introduction of new equipment and technologies is slow and often complicated. Reaching an appropriate return on investment typically takes much more time than in other, faster-moving industries. Effective cooperation between the public and private sectors is often a prerequisite to justify such technology investments.

Moreover, with the emergence of demand side resources that are implemented in a relatively uncoordinated fashion, this could cause the grid to be unstable and, in extreme cases, cause flows exceeding design capacities. This is where AI could play a role in furthering central grid planning and design [93].

The distribution of input data represents a challenge in the analysis of the energy network. These are currently available in a distributed form on different platforms. In order for correlations between input data to be detected by an AI algorithm, they would first have to be collected and standardized. Problems may also arise if this data has not been collected in a uniform manner. For example, it would have to be considered whether data was collected as time-invariant or how it is scaled. In many cases, the handling and evaluation of data also plays an important role when such data can be used to analyze a person's privacy. The power consumption and pattern analysis can be used to determine when a person is at home and what he or she is doing. Advanced analyses could additionally be used to determine which person is present in the household on the basis of energy usage behaviour predictions, even if it is not explicitly stated or desired. This data privacy challenge represents a major concern that utilities need to address with appropriate security and policy measures.

Section 6

Artificial intelligence challenges

A number of technical, ethical, trustworthiness, and regulation-related challenges need to be dealt with in order to deploy AI technologies across industries described in the previous section. For example, when implementing AI in the transportation and automotive sectors, safety and security are among the primary challenges. For smart manufacturing, safety and trustworthiness are major concerns.

Some challenges are shared among different application domains, for instance ethics and social impact, computational power and efficiency of the AI infrastructure, availability and quality of data, etc. Addressing these challenges will be instrumental to accelerating the adoption of AI technologies across multiple industries.

6.1 Social and economic challenges

AI has the potential to profoundly influence both society and markets. By enabling automation in several domains, starting with manufacturing but also including procedural professions such as the practice of law, AI has the ability to impact the employment market by both destroying and creating jobs. In that respect, certain skills developed by humans such as the application of creativity, will become more and more important in the future. There will be an increasing number of jobs where humans will work together with AI systems, resulting in new work environments.

Also, as AI can recommend goods and services to its users, sometimes in a hidden and automated manner, it is clear that AI opens up the possibility to actively affect and influence consumers' opinions.

6.1.1 Changes in decision-making

AI is expected to participate increasingly in decision-making, especially in routine processes. Even when trained correctly a technology can make mistakes. The question is how humans and AI systems can cooperatively make decisions in a sufficiently diligent manner. However, it is still unclear how the respective advantages of AI and human reasoning could be combined in an appropriate way. For developers, the challenge of creating a perfect algorithm is unrealistic. The more complex an algorithm is, the greater the impact it will have on society or industry, and the more human judgment will be needed to ensure adequate quality in decision-making [13].

This has significant implications concerning the responsibility of decision-making. In a complex system it is hard to differentiate whether the human or the AI system is accountable. AI can give the impression of being responsible for decision-making although it remains dependent on statistical correlation. This could lead decision makers to deny any accountability and transfer it to AI systems [13].

6.1.2 Advanced supply chain operations

With AI, laboriously sifting through catalogues and price quotes simply to submit an order will be a thing of the past. Instead, requesters will only need to take a photograph of the item they desire, which will then be automatically matched with suppliers' available stock. If the requester intends to order a completely new item, a simple description of its requirements will be enough for the AI system to submit price requests, assess supplier's quotes and select the most appropriate item according to price, quality and delivery times. If a specific component is needed but key information required to reorder the item such as its part number or specifications are missing, AI too could be used to identify the item and suggest a suitable replacement.

If AI systems begin to make all these decisions, some markets will be profoundly disrupted in the near future. Consumers will no longer visit physical shops but will obtain all products by a single click. Intermediaries will lose relevance when supply chains change. As consumers' choices will be increasingly influenced by AI algorithms, transparency of recommendations will become a bigger challenge. If, for example, a smart refrigerator decides autonomously about food suppliers and its owner's diet, it can disrupt entire markets such as grocery shops. Also, consumers and market participants will have to be protected from the misconduct of suppliers recommended by AI systems.

6.2 Data-related challenges

AI requires massive sets of data to train machine learning algorithms. However, data sharing and distribution is today constrained in different industry sectors by the lack of appropriate rules and regulations. This has led practitioners in various industries to isolate their data and set boundaries for data sharing, following their own commercial interests. In spite of the accumulation of a considerable amount of data in those industries, the data islands issue has been an obstacle to the realization of the full potential of AI.

In addition, the overall data availability issue causes difficulties in collecting enough reliable data for machine learning algorithms. These algorithms work differently from human brains and usually treat the data content as basic truth, without taking into account statistically rare content. This may lead to cognitive bias and misunderstanding when applying AI techniques. When trying to solve specific problems, the proportion of reliable and credible data in the total amount of data collected may fail to reach the lower limit of machine learning requirements.

6.2.1 Selection of training data

Cases in which AI systems exhibit gender or racial bias because the training data itself was biased have received increased media attention [96]. When developing an AI model, the relevance, quantity and quality of the data used in its development are critical.

First, the available data must be relevant to solving the problem at hand, which given the size of the involved datasets is not always easy to determine. Data scientists therefore need some kind of understanding of how and to what end the data will be used. Assessing the relevance of the data has become a multidisciplinary task involving not only data scientists but also domain experts. Second, in order for the model to perform accurately there must be enough data so that it can learn to generalize from the dataset. And third, the quality of the training dataset must be representative of the data the model will encounter once deployed.

Whether these conditions are met can often only be confirmed after several rounds of initial training and testing with a variety of different models. This process is highly iterative, requiring the data scientists to adjust the training data and models several times while relying on their business understanding, and testing the data to verify the performance of the models. The testing data itself can hereby present a problem if it was not split properly from the original dataset. Before training, datasets are split into the data which is shown to the model for training and data which is withheld from the model during training and used to test the quality of the model. To ensure that the training data can test the quality of the model accurately, the split should be such that the testing data, just like the training data, simulates real-world conditions (e.g. for a predictive model, each testing data point should have a timestamp after the last timestamp in the training data).

Even when the model meets or exceeds expectations by performing accurately almost all of the time, it is critical to spend additional time verifying that the mistakes made by the model are not so severe as to undermine the usability of the entire application. If, for example, the model were to classify images correctly in 98% of cases but classified 0,2% of images incorrectly in a way that was biased and offensive [97], the model could not be deployed even though the overall error rate is very low.

Such cases will often be the result of human bias being present in the training data itself which the model has learned to apply [98] [99]. Even removing attributes prone to biases from training data (such as race, gender, sexual orientation or religion) may not be sufficient to eliminate such biases from the model, as other variables may serve as proxies for them. Although technical methods to control bias exist, none are perfect and further interdisciplinary research is needed to develop more refined approaches [100].

Ensuring that as the collection of data increases it also becomes more representative and free of biases would be an important first step toward remedying this problem. Until then, further work is needed to develop methods of checking for and correcting biases in training data and models.

6.2.2 Standardized data

As noted already, the success of AI is highly dependent on the amount, variety and quality of the data used. In the course of the current digital transformation, massive amounts of data can already be accessed or generated via various channels (e.g. linked open data, numerous sensors, existing databases). Diversity is a given but also a challenge. Pre-processing and describing data for proper understanding can significantly improve analytical results. In the future, this time-consuming step will be streamlined thanks to standardized data types, forms and information models.

The question that arises is how heterogeneous information and datasets can be understood and interpreted appropriately, especially across several AI applications, without first discovering the meaning of relevant datasets. In order to achieve this, a manufacturer-independent, unambiguous information model of the data is necessary. Semantic technologies are well proven to ensure a uniform representation of information that is understandable for machines and humans, and to make data available in a clear and comprehensive form. Based on this, suitable semantic tools facilitate the derivation of implicit knowledge and, as such, represent a form of efficient data pre-processing.

Semantic interoperability therefore requires that systems not only exchange relevant data or make it available for further processing, but that the interpretations of the data exchanged by the sender and receiver are the same. Semantic conflicts may occur, for example, when identical data points are described by different terms or different data points by identical terms.

However, the understanding of heterogeneous data must not only be guaranteed and standardized semantically, but also syntactically. Syntactic interoperability of datasets means that the structure and format in which the data is exchanged are well defined. If two systems, for example, use different formats or structures to provide or process relevant data, there will be syntactic conflicts. Standardized exchange formats and communication protocols on different levels and communication channels will overcome these barriers.

6.3 Algorithm-related challenges

Challenges also exist related to the algorithms used in AI. Some of the most notable problems in the deployment of these algorithms are robustness, the ability to adapt to new tasks and their lack of interpretability.

The safety of AI algorithms and the resulting risks for users represent an increasingly important challenge for complex systems. It must be ensured that algorithms behave correctly, i.e. that an algorithm or programme correctly solves the problem described in its specification for any input data. This remains an enormous challenge for machine learning algorithms, in particular neural networks. For example, there is often no clearly defined specification of which problem the neural network has to solve. In addition, the complexity of the algorithms makes it difficult or even impossible to understand the decision-making process. Some of these challenges are outlined in greater detail below.

6.3.1 Robustness

The term robustness associated with machine learning means that the algorithm makes right decisions even if the inputs differ from the training data. A robust algorithm is therefore stable against an adversarial input and has no significant deviation in terms of performance between training and application datasets [101] [102] [103].

While robustness was already an issue in the past, future trends will boost its importance. In particular, the success of reinforced learning has contributed to the growing research of robust algorithms. As described in subsection 4.1.3, this category of machine learning algorithms uses agents that interact with their environment and with other agents. This leads to a very complex system of interactions and changing environments, which makes it difficult to predict the outcome and the actions of the agent [101].

The trend of using algorithms for decision-making has probably the greatest effect in term of robustness. The more impact these decisions have, the more important is their capability to react correctly in a real environment.

Two scenarios are associated with decision-making. Either the system recommends a decision to its operator, who reviews and verifies the recommendation, or the system enforces its decisions automatically. The latter poses the problem that in the case of new input data the system does not necessarily realize that it made a mistake. Therefore, decisions are not passed on for verification and can lead to damages to infrastructures or even humans. This could be for instance a misclassification of scans for patients suffering from cancer and who may not receive proper treatment, or accidents caused by machine learning systems overloading the power grid.

Some of the reasons for the failure of AI algorithms within this context are mismatched datasets, outliers and the programming of the system itself. Mismatched datasets that do not match real-word data lead to an algorithm that cannot perform well. If there are profound differences, for example due to outliers, performance can also decrease [102]. Algorithms need to be able to adapt to variations in datasets without varying too much from the expected output.

Several research directions are currently being undertaken in the area of algorithm robustness: verification (i.e. how to build a system that works the right way); validity (i.e. ensuring that the system satisfies the right objectives and does not perform undesirable actions); security (i.e. how to prevent manipulations of the system by a third party); and control (i.e. necessity that in the end a human can control the system and therefore fix it if problems occur) [104]. Multiple approaches can increase the robustness of AI algorithms, such as data pre-processing to train AI systems, removing mismatches and outliers, change and anomaly detection, as well as hypothesis testing and transfer learning [102].

6.3.2 Transfer learning

Machine learning implementations nowadays are customized products. Variables are chosen to fit the exact problem and the training data must come from the application field. This ensures that the algorithm performs perfectly for its application. While humans are able to transfer knowledge from previous experiences to new problems in order to solve them, machines do not have this capacity. If changes are influencing the provided data distribution that lead to outdated data or similar application fields, training data has to be recollected and the algorithm trained again. To reduce the cost and effort involved, transfer learning can help move knowledge from one application to another [105].

As illustrated in Figure 6-1, the objective of transfer learning is to enable the use of training data from different application fields, with different distribution or different tasks. Several issues need to be dealt with, such as how to determine which knowledge can be transferred. If the training data originates from another application field, the information transferred may be relevant or not. Many issues are still open with regard to this approach, such as how and when the knowledge can be transferred, but improvements are on their way. Breakthroughs in transfer learning could make machine learning much easier to apply and reduce the cost and time of development.

The robustness of algorithms is an essential factor when the entire system needs to be safe for human interaction, such as in cobots or autonomous cars.

6.3.3 Interpretability

Most AI algorithms, especially neural networks, are described as “black boxes”. This means that input data and network outcomes can be understood, but not how the algorithm reaches its result. This is a critical challenge for AI since understanding models is one of the most important starting points for a wide acceptance by end users. While some models such as regression or decision trees are understandable for data scientists and AI experts, the dimensionality of the data flow and complexity of most other algorithms are usually too high to be properly understood.

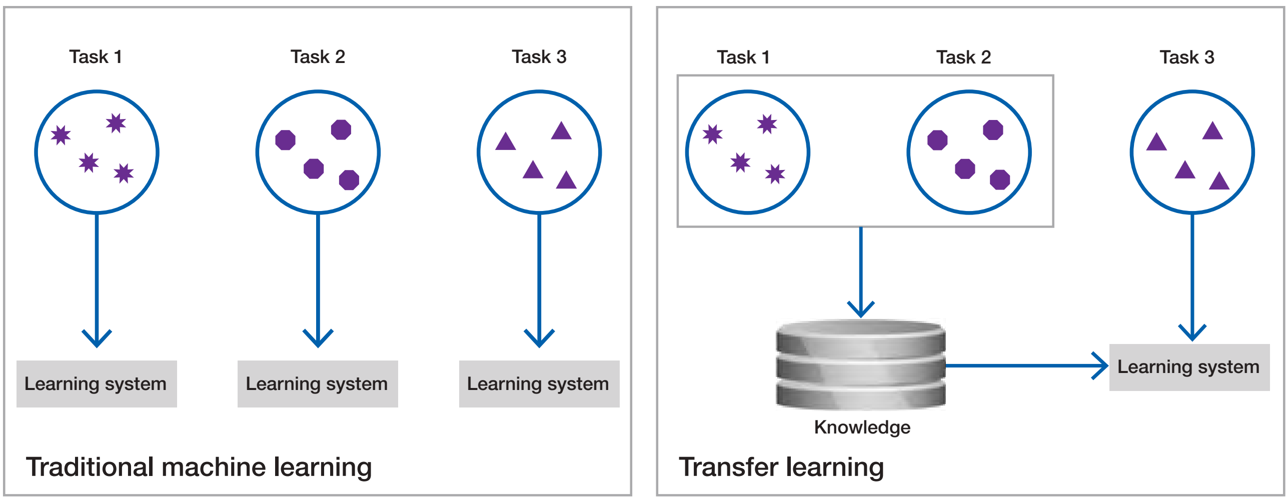


Figure 6-1 | Comparison between traditional machine learning and transfer learning

This means that such algorithms do not provide a clear explanation of why they made a certain prediction. They merely provide a probability, which is often hard to interpret. This makes it difficult and often impossible to verify that a trained algorithm functions as expected. Sometimes millions of model parameters are involved, and no one-to-one relationship between input features and parameters. Therefore, combinations of multiple models using many parameters often affect the prediction. Some of these also require a large amount of data to achieve high accuracy.

But there is not only the algorithm itself that poses a problem. The transformation of raw data into data that can be processed by mathematical models can make even simple algorithms not interpretable by humans. Some methods have been proposed to allow the interpretation of neural networks in application fields such as NLP or image recognition. Other approaches try to locally approximate complex algorithms through simple, understandable models to enhance interpretation [106].

As AI algorithms are deployed in a growing number of sectors, including areas with highly sensitive data such as medicine or finance, interpretability will certainly grow in significance and continue to constitute one of the great conceptual and technical challenges of AI in the future [107]. As the adequate verification and validation of AI algorithms remain highly problematic, their impact can be shown through the objective function of an algorithm.

6.3.4 Objective functions

A key focus of concern involves an AI system's objective functions, which if incorrect or imprecise can lead to negative side effects or reward hacking. Negative side effects might include for example harm to goods or humans provoked by the system because operating in this manner allows it to achieve its objective more rapidly. On the other hand, reward hacking entails the inadequate completion of the system's task because it found an unforeseen alternative way of satisfying its reward function [106]. A system's objective function should also not stand in the way of the system being shut down or modified, even when this impacts the system's ability to achieve its objective [106].

Even when the objective function is stated correctly, systems will need to be able to perform correctly when scalability issues fall within their supervision, e.g. when providing frequent feedback on the system's performance is too expensive [108].

As already mentioned previously, poor training data can lead to extremely undesirable outcomes. A question that therefore needs to be addressed is how a system should behave when confronted with unfamiliar data that it did not encounter in the training phase, e.g. fail gracefully rather than carry out an action that is wrong and offensive [109]. Poor training data aside, testing and training environments for reinforcement learning agents also need to be safe and isolated to contain any negative impact that their exploration might cause [101].

6.4 Infrastructure-related challenges

To run AI applications with satisfactory performance (especially under real-time constraints), computing speed and infrastructure efficiency need to be steadily increased. Not only is customized hardware needed to accelerate AI workloads, but also software stacks, libraries or tool chains, which enable the deployment of AI tasks on platforms with optimized utilization of underlying resources.

6.4.1 Hardware bottlenecks

AI and deep learning in particular require parallel processing of massive amounts of data, which traditional computing architectures can hardly support. Currently used GPUs and FPGAs have a number of technical limitations that restrain the implementation of the most advanced AI algorithms. For example, the GPU that was first introduced into deep learning has three main limitations: it is unable to fully exploit the advantages of parallel computing; the hardware structure is fixed without programmability; and deep learning algorithmic effectiveness remains to be improved. In the new computing era, the core chip will determine the infrastructure and ecosystem of AI. Processor capabilities are therefore considered as a major bottleneck in advancing AI development.

In that respect, the design and architecture of heterogeneous computing platforms (which integrate a variety of accelerators to address diverse AI workloads) is an essential subject for AI research and commercial implementation. In addition, hardware resources provided within cloud infrastructures have become an emerging trend given their scalability, reliability and automated resource management. Also, cloud-native application programming interfaces (APIs), for instance containers, are utilized to provide consistent interfaces, wide support and easy deployment of AI applications.

Since AI technologies may be implemented on different systems or subsystems (cloud, edge or end devices), the platform design should be tailored to the individual needs and resource limitations of the system. For instance, a cloud server may run more sophisticated algorithms and process a larger volume of data (e.g. for model training) than a mobile device. Therefore, hardware design needs to take into consideration the coordination of AI capabilities on the various systems or subsystems.

6.4.2 Lack of platforms and frameworks

Reusable and standardized technical frameworks, platforms, tools, and services for AI development are yet to be matured. Although a few open source AI learning systems and deep learning libraries have been made available by well-known technology giants, fully modular and standardized AI ecosystems of architectures, frameworks, application models, assessment and visualization tools and cloud services, may still take some time to reach an appropriate maturity level.

6.5 Trustworthiness-related challenges

It is widely acknowledged that AI is a topic involving many different stakeholders who need to cooperate and work together. For example, in the field of predictive maintenance for manufacturing, faults rarely appear. To adequately feed AI algorithms, both manufacturers and users have to share data, provide expert knowledge and work together towards an efficient implementation. A number of issues, such as ensuring trust, need to be addressed in order to facilitate this cooperation.

6.5.1 Trust

Machine learning algorithms rely on the data provided. Complete and accurate data is therefore essential for automated decision-making. Potential issues such as poor data quality or even intentional manipulation can lead to worthless results and even negative effects for the user of the algorithm.

Trust between stakeholders is essential. Solutions addressing the trustworthiness of data sources could possibly be offered by certification technologies. Electronic certificates from a centralized and trusted issuer combined with data sealing are options to establish trust between parties. However, this solution aims solely at installing trust between partners and does not address the data quality issue. For this purpose, one could collect a trusted data pool or use an evaluation or assessment algorithm to avoid faulty databases. Meta-algorithms then could help keep the AI system reliable and transparent over time, providing information on the origin and distribution of the sources used [13].

6.5.2 Privacy

The development of AI depends on the use of data training algorithms. In this process, a large amount of data needs to be collected, analyzed, and used. The value of data is increasingly prominent. Developers, platform providers, operating systems and terminal manufacturers, as well as other third parties in the value chain, have access to these data and are able to upload, share, modify, trade, and leverage user-supplied data to some extent.

In addition, since AI systems generally require higher computing capabilities, many companies and governments have begun to store data on the cloud. However, the privacy protection of the cloud also has hidden threats. How to collect and use data legally and in compliance with existing and future laws is a critical issue for any AI player.

6.5.3 Security

Technical abuse, flaws, and the development of future super AI all pose security threats to human society. The impact of AI on humans largely depends on how people use and manage it. In the hands of criminals, AI can certainly lead to major security problems. For example, hackers may launch cyberattacks through software that can self-learn and mimic the behaviour of AI system users, and constantly change the method to stay in the system for as long as possible. Some technical defects lead to abnormal work, also placing the AI system at risk. For instance, the black box model used for deep learning makes the model uninterpretable; improper design can therefore lead to abnormal operation. In addition, if security measures are not effective enough, driverless cars, robots and other AI devices may harm humans and be challenged from a legal perspective.

6.6 Regulatory-related challenges

Appropriate regulation is still lacking in many AI fields. Finding a balanced regulatory approach to AI developments that promotes and supports industrial innovation, productivity and competitiveness, while simultaneously ensuring high levels of security and health, consumer protection, social security and protection of rights and freedoms is an important priority for many governments across the world.

While a few early legislative steps have been taken in areas such as driverless cars and drones, no AI-specific regulatory body exists anywhere in the world, and there is also a lack of legal research on AI. In Europe, for example, robotics and AI aspects are covered by different regulatory agencies and institutions at national and European levels. No central European body exists to provide the technical, ethical, and regulatory expertise and oversight of developments in these areas. This lack of coordination hinders timely and well-informed responses to the new opportunities and challenges arising from these technological developments.

The six key crosscutting regulatory themes identified in the European Parliament Committee on Legal Affairs report on AI concern a wide range of policy areas. The areas where, according to the Committee's position, action is necessary as a matter of priority include the automotive sector, elderly care, healthcare and drones.

6.6.1 Liability

The issues of foreseeability, interpretability and causality that are emerging with new AI-based products will make it increasingly difficult to address liability issues such as product defects, which may create a large liability gap. Facing these anticipated liability challenges, the need for new rules and regulations, for instance in tort and contract laws, will become increasingly critical for many industries. Legal certainty on liability is of paramount importance for innovators, investors and consumers, providing them with the legal framework they need.

However, the complexity of digital technologies makes it particularly difficult to determine who is liable, and to what extent, in case of failures. For example, existing legal categories are insufficient to adequately define the legal nature of robots and consequently attribute rights and duties, including liabilities for damages. Under the current legal framework, robots cannot be held liable *per se* for acts or omissions that cause damage to third parties. In a scenario in which a robot can take autonomous decisions, traditional rules are insufficient to activate a robot's liability, since they would not allow identification of the party responsible for providing compensation, and to require this party to make good the damage caused.

In the European Union, it is expected that by mid-2019 the European Commission (EC) will issue guidance on the interpretation of the Product Liability Directive in the light of technological developments, to ensure legal clarity for consumers and producers in case of defective products.

6.6.2 Privacy

Regulations such as the General Data Protection Regulation (GDPR) in the European Union are in part intended to address these problems, yet how this will be implemented by data protection authorities remains to be seen [24]. A delicate balance will have to be struck between data privacy and enabling AI industries to flourish. In fact, AI itself will soon help ensure that personal data is safe by enabling sophisticated anonymization and encryption methods. Federated learning could ensure that personal data never has to leave consumers' devices to train an AI system, as the system is trained in parallel directly on every device [110]. In addition, AI could limit the exposure of sensitive information (e.g. health records) by conducting tasks without requiring a human to access the data, thereby increasing privacy.

The GDPR is a set of significant regulatory changes to data protection and privacy in the European Union which also addresses automated decision-making by AI systems. Specifically, the GDPR gives persons "the right not to be subject to a decision (without their explicit consent or authorization by European Union or Member State law, Article 22(2) GDPR [111]) based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her", Article 22(1) GDPR [111]. It also gives persons "the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision", Article 22(3) GDPR [111].

In addition, Articles 13-15 of the GDPR require that persons be told when automated decision-making that falls under Article 22 is conducted; be provided with meaningful information about the underlying decision process of the algorithm; and be informed of the consequences of the automated process and its significance [112].

Although it remains unclear how data protection authorities will implement the GDPR in practice, the transparency requirements for AI decision-making are likely to be the key challenge that both corporations and regulators will have to address. The objective hereby should be to strike a balance between maintaining data privacy and transparency and allowing data-driven business models to flourish. Allowing a healthy AI ecosystem is not just relevant from an economic perspective but also necessary to enable further technological research that can improve the ability of companies to ensure transparency.

6.6.3 Ethics

Although the most severe implications from these issues will only be seen in more advanced and futuristic AI systems, a proactive approach to addressing them as early as possible is not only a prudent approach but may also avoid costly (if not impossible) retrofitting in the future.

A more immediate concern is the need for AI systems (e.g. self-driving cars) to make ethical choices in their decision-making processes (e.g. injuring a pedestrian or avoiding the pedestrian and potentially injuring the driver or passengers) [113]. This example illustrates how AI safety is not just a technical problem but also a policy and ethical issue which will require an interdisciplinary approach to protect the users of such technologies, neutral bystanders and the companies that will develop them, as the latter may face important legal challenges. While research organizations and companies have begun addressing these issues, closer cooperation between all concerned parties at the international level is needed.

AI is progressively replacing humans in several decision-making processes. Intelligent robots also need to comply with the ethical constraints and rules of human society when they make decisions. For example, assume there are three pedestrians on the sidewalk in front of a driverless car that cannot brake in time: should the system choose to ram into these three pedestrians or instead swerve toward a pedestrian on the other side of the road? The application of AI in the daily lives of human beings is at the centre of fundamental ethical challenges that will need to be tackled. If the design of AI systems is not aligned with ethical and social constraints, such systems may operate according to a logic that differs from that of humans and may lead to dramatic consequences.

In addition, after granting decision-making rights to machines, people will face a new ethical issue: is the machine qualified to take such decisions? As intelligent systems acquire knowledge in specific fields, their decision-making capabilities will begin to surpass those of human beings, meaning that people may become dependent on machine-led decisions in an increasing number of domains. This type of ethical challenge will urgently require particular attention in any future AI development.

Section 7

Standardization gaps in artificial intelligence

Standardization plays both a supporting and a leading role in AI development. It is not only essential to promoting industrial innovation, but also to improving the quality of AI products and services, ensuring user safety and creating a fair and open industry ecosystem.

Following the preceding review of today's AI landscape and its main challenges, some of the fundamental requirements for standardization can be derived. This section first provides an overview of existing standardization efforts related to AI and then highlights some of the standardization and industrial gaps that will lead in the next section to final recommendations.

7.1 Standardization activities in artificial intelligence

Standardization in the area of AI is still at a very early stage. Although some aspects of AI or supporting technologies have been part of the scope of existing standardization groups for quite some time, new groups are now being formed to address the field of AI from a more extensive and holistic perspective. The following sections provide an overview of the current standardization landscape, including the organizations depicted in Figure 7-1.

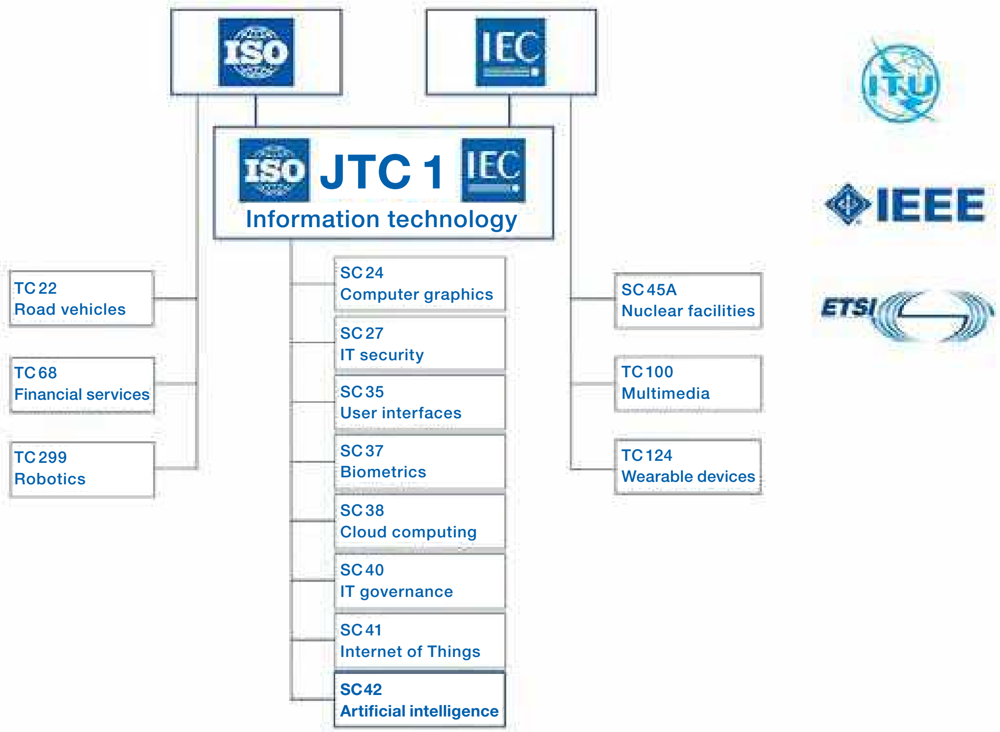


Figure 7-1 | Standardization landscape on AI

7.1.1 ISO/IEC JTC 1

ISO/IEC JTC 1, which is a joint technical committee formed between IEC and ISO on IT issues, has been performing work in the field of AI terminology for a long time. The former vocabulary working group of JTC 1 issued the following series of International Standards on AI terminology:

- ISO/IEC 2382-28:1995, *Information technology – Vocabulary – Part 28: Artificial intelligence – Basic concepts and expert systems*
- ISO/IEC 2382-29:1999, *Information technology – Vocabulary – Part 29: Artificial intelligence – Speech recognition and synthesis*
- ISO/IEC 2382-31:1997, *Information technology – Vocabulary – Part 31: Artificial intelligence – Machine learning*
- ISO/IEC 2382-34:1999, *Information technology – Vocabulary – Part 34: Artificial intelligence – Neural networks*

These historical parts have now been merged into the common JTC 1 standard for IT vocabulary: ISO/IEC 2382:2015 [11].

ISO/IEC JTC 1/SC 42

This subcommittee was established in November 2017 to address the specific standardization requirements of AI. The scope of JTC 1/SC 42 is to serve as the focus entity and proponent for JTC 1's standardization programme on AI, and provide guidance to JTC 1, IEC and ISO committees developing AI-related applications. Topics forming a part of the work of this subcommittee are:

- Foundational standards
- Computational approaches and characteristics of AI
- Trustworthiness
- Use cases and applications
- Big data
- Societal concerns

After its first plenary held in April 2018, JTC 1/SC 42 established WG 1 on foundational Standards, comprising the following first two approved projects: *Artificial intelligence concepts and terminology* (ISO/IEC 22989); *Framework for artificial intelligence systems using machine learning* (ISO/IEC 23053).

The plenary of JTC 1/SC 42 also established three study groups:

- SG 1: Computational approaches and characteristics of artificial intelligence systems, in order to study different technologies used by AI systems (e.g. machine learning algorithms, reasoning), including their properties and characteristics; existing specialized AI systems (e.g. computer vision, NLP) to understand and identify their underlying computational approaches, architectures, and characteristics; and industry practices, processes and methods for the application of AI systems.
- SG 2: Trustworthiness, in order to investigate approaches to establish trust in AI systems through transparency, verifiability, explainability, controllability, etc.; engineering pitfalls and an assessment of typical associated threats and risks to AI systems with their mitigation techniques and methods; approaches to achieve robustness, resiliency, reliability, accuracy, safety, security, privacy, etc. in AI systems; and types of sources of bias in AI systems with a goal of minimization of such bias, including but not limited to statistical bias in AI systems and AI-aided decision-making.
- SG 3: Use cases and applications, in order to identify different AI application domains (e.g. social networks, embedded systems) and the different contexts of their use (e.g. healthcare, smart home, autonomous cars); collect representative use cases; and describe applications and use cases using the terminology and concepts defined in projects ISO/IEC 22989 and ISO/IEC 23053, and extend the terms as necessary.

Other JTC 1 subcommittees

As AI is a transversal technology affecting many other IT fields and applications, other JTC 1 subcommittees have been producing standardization work that is connected to AI as a driver or supporting technology:

- JTC 1/SC 24: Computer graphics, image processing and environmental data representation
- JTC 1/SC 27: IT security techniques
- JTC 1/SC 35: User interfaces
- JTC 1/SC 37: Biometrics
- JTC 1/SC 38: Cloud computing and distributed platforms
- JTC 1/SC 40: IT service management and IT governance
- JTC 1/SC 41: Internet of Things and related technologies

7.1.2 IEC

Several IEC committees have looked at AI as one element potentially contributing to their programme of work. Examples include:

- SC 45A: Instrumentation, control and electrical power systems of nuclear facilities, has carried out a study on AI, with the aim of applying emerging IT and electronic technologies to advance computer and information systems supporting and regulating nuclear instruments and control requirements.
- TC 100: Audio, video and multimedia systems and equipment, develops Standards related to wearable devices and has initiated a topic for discussion entitled "usage scenarios of wearable devices" that included elements of AI and virtual reality.
- TC 124: Wearable electronic devices and technologies, was formed recently to take charge of the development of technical Standards on the electrical engineering, materials and personal safety of wearable technology. It is foreseen that such devices and technologies will be widely used within the context of AI applications.

7.1.3 ISO

Several ISO committees are concerned with preparing Standards related to AI applications, such as:

- TC 22: Road vehicles, formulates basic Standards for road vehicles and is also studying standardization challenges related to intelligence and connected cars.
- TC 68: Financial services, works on standardization for the financial and banking sector. New trends are covered by about 58 Standards of this committee.
- TC 299: Robotics, covers the field of robotics standardization for various uses.

7.1.4 ITU

In the area of AI, ITU-T has a Focus Group on machine learning for future networks including 5G (FG-ML5G). The objectives of this group include [81]:

- Helping the adoption of machine learning in future networks, including architecture, interfaces, use cases, protocols, algorithms, data formats, interoperability, performance, evaluation, security and protection of personal information.
- Studying, reviewing and surveying existing technologies, platforms, guidelines and standards for machine learning in future networks.
- Identifying aspects enabling safe and trusted use of machine learning frameworks.

- Reviewing and studying how to train, adapt, compress and exchange machine learning algorithms in future networks, and how multiple algorithms interact with each other.
- Identifying possible requirements of machine learning applied to future networks, taking into account a variety of fixed and mobile communication stacks, and promoting the development of new machine learning methods that will be able to meet these requirements.
- Identifying possible requirements on network functionality, interfaces and capabilities to use machine learning.
- Identifying standardization challenges in machine learning for communications.
- Producing a gap analysis and a roadmap of machine learning in order to identify the relevant scope of ITU-T recommendations on these topics.

7.1.5 IEEE

The IEEE mainly focuses in this area on studying ethical aspects of technical standards related to AI. In March 2016, the IEEE Standards Association launched the Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, with the aim of helping people deal with the threats posed by AI and developing ethical design principles and standards that range from data privacy to fail-safe engineering [82].

Under this umbrella, the IEEE has approved so far the following standardization projects:

- IEEE P7000: Model process for addressing ethical concerns during system design
- IEEE P7001: Transparency of autonomous system
- IEEE P7002: Data privacy process
- IEEE P7003: Algorithmic bias considerations
- IEEE P7004: Standard for child and student data governance
- IEEE P7005: Standard for transparent employer data governance
- IEEE P7006: Standard for personal data artificial intelligence agent
- IEEE P7007: Ontological standard for ethically driven robotics and automation systems
- IEEE P7008: Standard for ethically driven nudging for robotic, intelligent and autonomous systems
- IEEE P7009: Standard for fail-safe design of autonomous and semi-autonomous systems
- IEEE P7010: Wellbeing metrics standard for ethical artificial intelligence and autonomous systems
- IEEE P7011: Standard for the process of identifying and rating the trustworthiness of news sources
- IEEE P7012: Standard for machine readable personal privacy terms

7.1.6 ETSI

ETSI has an Industry Specification Group (ISG) on Experiential Networked Intelligence (ENI), whose goal is to develop standards for a cognitive network management system incorporating a closed-loop control approach. This approach is based on a “monitor-analyze-plan-execute” model and will be enhanced by learning capabilities.

The envisaged cognitive network management system enables the steering of the usage of available network resources and services according to the real-time evolution of user needs, environmental conditions and business goals. Decisions taken by the system rely on detailed information about the complex states of network resources and policies expressing operators’ preferences.

The unique added value of the ISG ENI approach is to quantify the operator experience by introducing a metric and the optimization and adjustment of the operator experience over time by taking advantage of machine learning and reasoning.

Different types of policies will be reviewed in this group in order to drive adaptive behavioural changes using various AI mechanisms. The ISG ENI will wherever applicable review and reuse existing standardized solutions for legacy and evolving network functions such as resource management, service management, orchestration and policy management.

7.1.7 Standardization activities in China

- **National information technology standardization network (SAC/TC 28)**

SAC/TC 28 mainly addresses AI standardization work related to vocabulary, user interfaces, biometric features recognition and other fields.

In the area of terminology and vocabulary, four basic national standards have been issued so far, such as GB/T 5271.28-2001: *Information technology – Vocabulary – Part 28: Artificial intelligence – Basic concepts and expert systems*. The user interface subcommittee is preparing multiple national standards and has set up motion sensing interaction and brain-computer interface working groups to carry out relevant standardization studies. It has submitted the international proposal “information technology emotive computing user interface framework”, which has been approved.

The biometric features recognition committee has prepared standards related to fingerprint, face and iris recognition. In addition, the big data standard working group of the national information security standardization technical committee, the working group on cloud computing standards and the working group on national sensor network standards are also making efforts to formulate basic standards to support the relevant technologies and applications of AI.

- **National technical committee for automation systems and integration (SAC/TC 159)**

Under SAC/TC 159, SC 2 on robot equipment takes charge of complete industrial robots, including system interfaces, components, controllers, etc. It has released several standards such as GB/T 17887-1999: *Industrial robots – Automatic end effector exchange systems – Vocabulary and presentation of characteristics*.

- **National technical committee for audio, video, multimedia and equipment (SAC/TC 242)**

SAC/TC 242 has made studies on relevant standards for audio, video, and smart healthcare products. Current standards include for instance subjective evaluation methods for virtual reality audio (2017-0279T-SJ).

- **National technical committee on information security (SAC/TC 260)**

SAC/TC 260 has formulated security-related standards in areas such as biometric features recognition, smart cities and intelligent manufacturing by focusing on AI technology.

- **National technical committee on intelligent transport systems (SAC/TC 268)**

SAC/TC 268 has carried out standardization work in the area of intelligent transportation. It has also formulated standards such as GB/T 31024.2-2014: *Cooperative intelligent transportation systems – Dedicated short range communications – Part 2: Specification for medium access control layer and physical layer*.

7.1.8 Standardization activities in the United States

At the time of writing, the United States does not currently have any policies or standards in place related to AI nor does it appear that their creation is a priority for the current administration. Several

US-headquartered private companies working with AI have come together along with several multinational firms to form the Partnership on AI, which intends to develop and share best practices. There are also an increasing number of research institutes and non-governmental organizations working on policy, ethics and safety issues related to AI. Rather than relying on government, it appears that collaborative private initiatives such as the Partnership on AI are the most likely source of some sort of standard for US-based companies in the foreseeable future.

It is relevant to note that the previous administration did have a greater interest in setting policies and standards and published two widely cited reports on the challenges and opportunities of AI. However, these documents are only accessible as part of the official archive of the previous administration and, at the time of writing, it is not entirely clear how or if they will be leveraged by the current administration.

7.1.9 European AI Alliance

AI is impacting critical European industries such as healthcare, agriculture or public administration. It is also driving business opportunities for European industry, SMEs and start-ups, and contributes to productivity growth in Europe. Therefore, the EC is setting up a European AI Alliance to discuss the future of AI [80].

The EC has committed to developing a comprehensive strategy on AI to address the legitimate concerns of ensuring trust and awareness, including all relevant stakeholders (businesses, academics, policy makers, consumer organizations, trade organizations, and other representatives of the civil society). As a consequence, the EC plans to establish a European AI Alliance to act as a multi-stakeholder forum to engage in all aspects of AI development and its impact on society and the economy.

The first step toward establishing this European AI Alliance is to create a high-level expert group on AI that will serve as a steering group for the alliance's work and in addition will have the task of advising the EC on mid to long-term AI challenges and opportunities. The expert group will support the EC on engagement and outreach mechanisms with other initiatives and propose AI ethics guidelines. The call for high-level experts was completed in April 2018 and the group is supposed to initiate its activities in the second half of 2018.

7.1.10 Consortia and other organizations

The issues of standards and rules for AI have received an increasing amount of attention not only from the public but also from companies, research institutions, and industry consortia. Technology companies themselves are increasingly considering the ethical, economic and social consequences of the AI products and services they are developing, as well as the standards and rules that might be required.

Several of the world's largest technology companies have even created dedicated ethics teams and established supervisory boards to help answer some of these questions and monitor their company's efforts. Most activities today concentrate on raising public awareness and developing internal codes of conduct for developers and designers. Such efforts are not aimed at producing one-size-fits-all solutions, but rather developing an informed company opinion through internal trials on what sorts of standards and rules might work best.

This type of internal work lays a valuable foundation for engaging a variety of other stakeholders. Perhaps the most prominent initiative in the public perception aimed at bringing stakeholders together is the Partnership on AI, a growing consortium of over 50 of the largest technology companies and AI-focused research organizations. Its aim is to ensure that AI benefits people and society. This will be achieved through collaborative research on the impact and design of AI systems, the development and sharing of best practices, as well as public education and the engagement of a wide variety of external stakeholders [84]. While the Partnership on AI is still in the building phase, and has yet to publish any research, it has the necessary ingredients to make a valuable contribution to the debate in the near future.

Other organizations such as the Royal Society [85] and the Information Technology Industry Council [86] are highlighting areas in which further work on standards is required. In addition to the work of private companies and consortia, there are more than a dozen reputable research organizations currently considering the implications of AI developments, often to ensure the safe creation of general AI.

7.2 Standardization gaps

This subsection lists some of the open standardization gaps that need to be addressed in order to resolve the challenges listed in previous sections. With the development of technology and the increase of application scenarios, many standards need to be improved or supplemented. These gaps can range from coordinated development of open source and standardization to neural network representation methods, performance evaluation, machine learning algorithms and security gaps related to AI.

7.2.1 Harmonized data models and semantics

As already stressed in section 6.2.2, machine learning is dependent on the data with which it is trained. However, machine learning applications can depend on different data sources distributed over various domains. When data lacks semantic capabilities, the AI system will not be able to adequately use these different data sources, because they are not processed in a way understandable to both machines and humans. Additionally, data structure and format have to be unified. Especially for machine learning this problem represents an important issue, because for many applications, data is needed from different sources.

Domain-specific efforts to standardize information and data models already exist today. However, there is a lack of coordination among these efforts. For the benefit of AI, it would be necessary to coordinate and ensure a homogenous approach to standardizing information and data models across different domains.

7.2.2 Common ontology based on data models

There are already a variety of activities concerning ontologies in different domains. Because machine learning will not be restricted to one domain, but often will include multiple domains, these domain-dependent ontologies have to be harmonized. An effort is needed to coordinate the activities between different domains and standardize a common ontology.

7.2.3 Verification of artificial intelligence algorithms

The verification of AI algorithms is needed to ensure that they are compliant with all applicable safety requirements. AI algorithms differ from other algorithms mainly in that they change during their runtime. Changes in the environment can also have an effect on the functioning of a self-learning algorithm. The accurate documentation of an AI algorithm task is challenging and may benefit from standardization. It would be desirable to clarify which requirements an AI algorithm, or the entire system in which it is located, would have to be verified against. It is also questionable whether a one-off verification is useful or whether cyclical checks are appropriate [83].

AI systems need certification specifications on different aspects such as function, performance, security, compliance or interoperability to secure AI product quality and availability to support a sustainable development of the AI industry. Certification may include testing, evaluation, and other tasks. The evaluation object can be an automatic driving system, a service robot, or other AI products. The evaluation results can be obtained through measurable indicators and quantifiable evaluation systems based on standardized procedures and methods.

7.2.4 Benchmarking and evaluation of artificial intelligence infrastructures

With infrastructures and platforms constantly being developed with innovative design, improved architecture and new hardware components, standards are needed for benchmarking and evaluation of the platforms, in terms of function, performance or scalability. The infrastructure evaluation can use either generic machine learning algorithms or scenario-specific workloads (e.g. condition monitoring, surveillance video). Benchmarking and evaluation may help users choose the platforms most suitable to their individual needs.

Section 8

Conclusions and recommendations

This White Paper has shown clear evidence that AI can be deployed across a broad range of applications that contribute to addressing some of the most pressing challenges called megatrends. As AI is already having a profound impact on society, businesses and individuals, it is now foreseen to become one of the driving forces for radical innovation in most industry sectors, whether manufacturing, energy, finance, education, transportation, healthcare or retail.

Specific use cases have been described in the White Paper for four application domains (smart homes, smart manufacturing, smart transportation, and smart energy), but the wide applicability of current AI systems across multiple industry sectors has also been clearly demonstrated.

Serious challenges are being generated by AI in societal, economic and regulatory fields. Whether related to trustworthiness, privacy, safety of human/machine interactions, or the impact on the workforce, these issues will need to be addressed urgently by a broad base of stakeholders and in a coordinated way. As these challenges will be common to all nations across the globe, the White Paper has also insisted on the need for an international approach to tackle them. In particular, governments, regulators and policy makers will have to carefully understand and address major ethical issues that are emerging with the rapid deployment of AI.

Industry and the research community will also have to address a number of technical challenges that may impede the deployment of AI across some application domains. Issues related to data, algorithms, hardware and computing infrastructures are among today's most difficult limitations to exploiting the full potential and achieving the most exciting promises of AI. From that perspective, standardization and conformity assessment are expected to play an instrumental role in facilitating the market adoption and social acceptance of emerging AI technologies.

The following subsections outline the main recommendations of this White Paper, to industry in general, to regulatory bodies, and finally to the IEC and its committees.

In addition, further forward-looking AI applications are presented in .

8.1 Industry recommendations

A wide range of industry stakeholders are involved in the AI ecosystem. These industry players are expected to benefit from AI technology but need to contribute with coordinated efforts to the availability of high-quality data, to the continuous performance improvement in AI infrastructure, and to increasing the security of AI systems. These efforts will eventually lead to increased awareness of AI's benefits to the society and accelerate its wide adoption.

- **Develop guidelines for datasets**

Training datasets are one of the most important factors in the performance of an algorithm. Special requirements are placed on these datasets, depending on the application area. It is recommended to create guidelines to estimate how large the quantities for training, validation and testing should be. This could help companies successfully and securely implement AI. Furthermore, guidelines on data requirements would help businesses when requesting customer data, by affording them a neutral entity to which to refer.

- **Develop guidelines for security within AI systems**

AI has a profound impact on the security of information systems. On the one hand, these can be used to create new types of hack bots or viruses that constantly adapt to new countermeasures. On the other hand, they can also create additional security holes for companies by providing new access points for attacks. Training data could be used to try to influence the system by specifically controlling a bias in the datasets. It is therefore necessary for companies to identify the security effects that the use of AI will have before applying them, in order to take appropriate security measures.

8.2 Regulatory recommendations

AI deployments are accompanied by a substantial regulatory uncertainty, which will be seen by industry and the research community as an obstacle to investing in AI. In addition, it is obvious that AI contributes to specific governmental projects such as the Digital Single Market in Europe. Therefore, successful market deployment of AI will require coordinated regulatory activity between multiple stakeholders (industry, governments, research community, SMEs, consumer organizations, and others). The regulatory activities described below are recommended.

- **Address AI within privacy policies**

The European GDPR is addressing some aspects of the abuse of individual privacy rights, such as automated decision-making and profiling without the user's consent. There remain however a number of privacy concerns that need to be addressed. There is evidence that some hackers are reverse-engineering the training data to retrieve privacy data from individuals and misusing this confidential data. Also, the processes of algorithm design and training need to be inclusive and should fully consider the interests of disadvantaged groups. They should establish special rules for extreme ethical situations. In that respect, relevant standards, laws and policies on security, ethics and privacy should be urgently improved.

Where AI is distributed within the cloud, it will be critical that individuals have sufficient trust in the process to send their private data to the cloud without worrying about infringement of their privacy. Moreover, AI developers need to adopt appropriate technical means to protect individuals in the process of collecting and using data and to prevent personal information from being subject to leaking, falsification and damage.

- **Include AI in security, ethics and other policies**

While the current AI industry is in a period of vigorous development, it should seriously consider the formulation and implementation of future industry supervision measures and should fully consider the responsibilities and faults in the process of AI development and deployment.

The issue of supervision of the AI industry is not a problem faced by a single group. It involves a wide range of social, systemic, and complex issues which require enterprises, governments, users, technology organizations and other third parties to participate and work together to build a sustainable AI ecosystem.

The relevant safety regulations should be formulated and improved. The ethical requirements of AI technology should be set in accordance with a broad consensus formed among the society as a whole.

8.3 Recommendations addressed to the IEC and its committees

The IEC, as a globally recognized standards development organization, is in a unique position to help address some of the AI challenges described in this White Paper, in order to ensure the social and market relevance of upcoming AI technological developments. Building upon its track record and reputation for addressing the safety of electrical systems, the IEC can play an instrumental role in accompanying the tremendous changes and mitigating the concerns brought about by AI. Accordingly, it is recommended that the IEC take the actions described below.

- **Promote the central role of JTC 1/SC 42 in horizontal AI standardization**

As JTC 1/SC 42 has established its programme of work and begun its standardization activities, the IEC should promote this subcommittee as the key entry point for basic and horizontal standardization of AI worldwide. In close cooperation with ISO, the IEC should also ensure active collaboration and possible joint work by individual IEC and ISO committees with JTC 1/SC 42 on aspects related to AI technologies. To avoid duplication of work and foster quick market acceptance, the IEC should encourage external organizations, whether standards-setting bodies or industry consortia, to liaise and collaborate with JTC 1/SC 42.

- **Coordinate the standardization of data semantics and ontologies**

As the success of AI is highly dependent on the quality and proper interpretation of the data used, standardization of data types, forms and information models will be critical for the acceptance and successful deployment of AI across various industry sectors. Semantic interoperability and standardized ontologies will be core elements to ensure consistency and homogeneity of AI implementations. Without duplicating existing work, the IEC should take the lead in landscaping, coordinating and facilitating the convergence of existing standardization activities that address this data-related challenge, both internally (e.g. SC 3D, domain-specific committees such as TC 57 and TC 65, JTC 1/SC 32, JTC 1/SC 41) and by reaching out to external entities (e.g. ISO, eCl@ss).

- **Develop and centralize AI-related use cases**

As AI is making its way through an increasing number of application domains, the IEC should encourage its technical committees to develop AI-related use cases by applying the IEC 62559 methodology for further incorporation into the upcoming IEC Use Case Management Repository (UCMR). The IEC should also reach out to external organizations, whether standards-setting organizations or industry consortia involved in AI standardization, to promote its use case development approach.

- **Develop AI reference architecture with consistent interfaces**

To guide the development of AI platforms, a reference architecture for AI systems needs to be established, which will enable software and hardware vendors to position their products in the ecosystem and contribute to the interoperability between components and subsystems. The interfaces between heterogeneous computing units, as well as interfaces for AI task scheduling and resource utilization need to be agreed upon, to foster a simplified deployment of AI workloads on heterogeneous computing platforms.

- **Explore the potential for AI conformity assessment needs**

As AI is generating unprecedented ethical and trustworthiness challenges and threats, assessing the conformity of AI algorithms as well as the products and services making use of them is expected to be in high demand. Dealing with the black box issue and certifying the behaviour of algorithms that continuously evolve and adapt to their environment will be huge challenges. As a unique organization developing both International Standards and Conformity Assessment Systems, the IEC should launch a study to identify and roadmap AI conformity assessment needs and opportunities.

- **Foster a dialogue with various society stakeholders about AI**

The IEC is ideally positioned to initiate and foster a dialogue about the impact of AI on society and industry by gathering all interested stakeholders (industry, governments, regulators, policy makers, etc.). By leveraging its international standing and neutrality, the IEC should collect inputs from these stakeholder groups to feed its AI Standards development activities, continue to build trust in its processes, and ensure its deliverables are used in AI-related regulatory and legislative work.

- **Include AI use cases in testbeds involving the IEC**

The previous IEC White Paper on edge intelligence [12] recommended that testbeds be deployed to gain feedback on IEC standardization activities. Within the spirit of this recommendation, AI use cases within testbeds will be essential to gather feedback on AI technology gaps and standardization needs. It will be difficult to justify deploying a specific testbed for AI, but the feedback on the AI-related architectures and standards-driven AI implementations will be essential to identifying standardization and technology gaps. In addition, new machine learning capabilities, algorithms and data and ontology models may benefit from testbeds comprising AI use cases. It is therefore recommended to add AI-related use cases to future testbeds involving the IEC.

Annex A
Future developments

A.1 Biology-inspired artificial intelligence

While neural networks are inspired by how the human brain operates, they do not try to mimic it. Actually, it appears to be extremely difficult to copy the mechanisms of the human brain, since today it is still not fully understood how the brain really functions. However, there is a large consensus of opinion that a better understanding of how learning works in the human brain will lead to superior machine learning algorithms. Some think that it could be an essential step towards the creation of a general AI.

There are several theories on how learning mechanisms might work in the brain, and several algorithms based on these theories have been developed. While they do not try to create an exact copy of the brain, they attempt to imitate some central regions of the brain, usually limited to the neocortex. Other regions of the brain, such as the limbic system, have received relatively little attention by the machine learning community so far. While it is not yet clear whether one of these approaches could replicate the human brain in some ways, there are nevertheless a few notable developments that have taken place over recent years.

One of these approaches is hierarchical temporal memory (HTM). At first glance, this structure looks like a neural network, but fundamentally differs from it. An HTM is a hierarchically structured network of nodes. These nodes have the ability to learn and store information. The structure is designed to display time-varying data hierarchically. Data coding within the HTM is available in a form called sparse distributed representation [114], which is comprised of bit vectors in a pattern where each bit has a semantic meaning. This model was inspired by a theory of how the human brain learns [115].

In summary, it still seems unlikely at this point in time that a general AI could be created in the near future. However, the latest findings in brain research, like the European Human Brain Project [116], can give impetus for new types of machine learning algorithms.

A.2 Human/artificial intelligence interaction

The use of AI opens up new opportunities for human-machine interaction. Whereas until today machines have usually been only tools operated in a certain way, sophisticated AI algorithms could make machines better adapted to humans in the future.

The early stages of such improved interactions can already be seen in many assistance systems, which are frequently used in mobile devices or in the home. However, as of today, it is usually the people who have to adjust to the machine. Requests must be formulated in a certain way so that the machine can process them. However, the ultimate goal is for the machine to adapt to people. In summary, AI serves as an interface so that human-machine interactions can operate smoothly.

An example of such interaction is a telephone service hotline. Most automatic answering machines only respond to certain words or numbers entered by the caller. In the future, such automatic queries will be carried out by an AI system that dynamically responds to the caller's questions and needs, thus avoiding frustration and stress.

AI is one of the central components for new safety concepts, especially when it comes to handling machines. This is particularly important in robotics. Most autonomous robots found at home today are vacuum cleaner robots. In the future, however, much larger, possibly humanoid, robots could be used at home or in public environments. Such robots must be able to recognize their environment, and especially the human beings in it, in order to ensure the safety of the latter. This advanced understanding of a changing surrounding environment is only possible with advanced machine learning techniques.

A.3 Artificial intelligence-enabled digital twin

Digital twins, which are digital models of physical objects and processes, are enabled by the IoT in which connected devices collect and share data about themselves and their surroundings. By bundling the data collected by a physical device in real-time, digital twins mimic the status of their physical counterpart. Such virtual twins can then be analyzed and experimented with to improve their physical counterpart's performance [117]. The rise of digital twins therefore has a profound impact on product development, by enabling engineers not only to test new designs in a virtual space but also to monitor their performance on an ongoing basis. As a result, AI can support such simulations, predict the future performance of the device and make recommendations for possible improvements. Interesting applications for the use of digital twins include, for example, applications for turbines, to better understand and monitor their performance, or for equipment in remote locations. Issues can be spotted more easily or scenarios can be tested prior to installation to account for the various environmental factors of new or unusual locations.

Particularly where streams of data are involved that go beyond numerical sensor readings, such as video and audio, AI can add significant value by incorporating them into the analyses. Audio feeds can thus be analyzed to detect signs of impending malfunctions and enable proactive maintenance, while real-time video of a production process could be used to identify defects with great accuracy and speed [118].

Aside from enabling the analysis of this type of data at great speeds, AI can facilitate a productive interaction between AI-powered digital twins and humans. AI-powered bots can be leveraged to enhance the interactions between humans and digital twins (or for that matter any physical object), thereby improving user experience and productivity. Rather than laboriously examining technical logs or simply receiving a command from the AI system, the user will be able to investigate problems and opportunities alongside the AI system, using natural language to interact with the digital twin of a given mechanical component as they would with a human co-worker [119].

Not only will this development boost human productivity, but it will greatly increase the acceptance of such systems by underscoring their function as a support, rather than a replacement for human labour. Indeed, with fully implemented digital twin capabilities, an organization could create a workforce that comprises both its human capital and mechanical capital.

A.4 Automated machine learning

As of today, there are various examples of automated machine learning. One is reinforcement learning, which uses agents to learn directly from raw input data. The agents interact with their environment and receive a reward or punishment depending on their actions. By seeking to maximize the reward, they learn to make the right choices. There exist reinforcement learning algorithms, for example which successfully learned how to play 50 Atari games without obtaining any additional knowledge about the rules of the games [120].

Another approach to automated machine learning is to automate the development of machine learning models. These systems autonomously create new child networks and train them. Depending on their performance they then detect points for improvements for subsequent attempts. Thus, the model improves the network and learns which parts of the architecture achieve good results [121] [122].

The trend to use automated machine learning to reduce the effort of training and modelling will likely further accelerate in the future. This raises the question of how to remain in control of the developed algorithms as well as how to deal with the fact that there could be more and more interactions between different AI applications improving each other.

Bibliography

[1] TURING, A. M., *Computing Machinery and Intelligence*, Mind 49, 1950, pp. 433-460.

[2] MCCARTHY, J., MINSKY, M. L., ROCHESTER, N., SHANNON, C. E., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, Original from August 31, 1955, reprint in AI Magazine, Vol. 27 (4), 2006.

[3] ROSENBLATT, F., *The Perceptron: A probabilistic model for information storage and organization in the brain*, Psychological Review, 65(6)/1958, pp. 386- 408.

[4] COVER, T. M., HART, O. E., *Nearest Neighbor Pattern Classification*, IEEE Transactions on Information Theory, Vol. IT-13, No. 1, 1967, pp. 21-27.

[5] DEJONG, G.F., *Generalizations based on explanations*, Proceedings of the Seventh International Joint Conference on Artificial Intelligence, 1981, pp. 67-70.

[6] RUMELHART, D. E., HINTON, G. E., WILLIAMS, R. J., *Learning representations by back-propagating errors*, Nature, Vol. 323 (9), 1986, pp. 533-536.

[7] CORTES, C., VAPNIK V., *Support-Vector Networks*, Machine Learning, Vol. 20, 1995, pp. 273-297.

[8] CAMPBELL, M., HOANE Jr., A. J., HSU, F., *Deep Blue*, Artificial Intelligence, Vol. 134, 2002, pp. 57-83.

[9] HINTON, G. E., OSINDERO, S., TEH, Y.W., *A Fast Learning Algorithm for Deep Belief Nets*, Neural Computation, Vol. 18, [7], 2006, pp. 1527-1554.

[10] *IDC Spending Guide Forecasts Worldwide Spending on Cognitive and Artificial Intelligence Systems to Reach $57.6 Billion in 2021* [Online]. Available: go.iec.ch/wpai001. [Accessed: 14 September 2018].

[11] ISO/IEC 2382:2015, *Information technology – Vocabulary*.

[12] IEC, *Edge intelligence*, White Paper, 2017 [Online]. Available: go.iec.ch/wpai002. [Accessed: 14 September 2018].

[13] Bitkom, *Künstliche Intelligenz – Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung* [Online]. Available: go.iec.ch/wpai003. [Accessed: 14 September 2018].

[14] GOODFELLOW, I., BENGIO, Y., COURVILLE, A., *Deep Learning*, MIT Press, 2016.

[15] HAJKOWICZ, S. A., REESON, A., RUDD, L., BRATANOVA, A., HODGERS, L., MASON, C., BOUGHEN, N., *Tomorrow's Digitally Enabled Workforce: Megatrends and scenarios for jobs and employment in Australia over the coming twenty years*, CSIRO, Brisbane, 2016.

[16] United Nations Environment Programme, *With resource use expected to double by 2050, better natural resource use essential for a pollution-free planet* [Online]. Available: go.iec.ch/wpai004. [Accessed: 14 September 2018].

[17] WOLFE, F., *How Artificial Intelligence Will Revolutionize the Energy Industry* [Online]. Available: go.iec.ch/wpai005. [Accessed: 14 September 2018].

[18] PwC, *Demographic and social change* [Online]. Available: go.iec.ch/wpai006. [Accessed: 14 September 2018].

[19] International Telecommunication Union, *AI for Good Global Summit Report* [Online]. Available: go.iec.ch/wpai007. [Accessed: 14 September 2018].

[20] World Economic Forum, *Harnessing Artificial Intelligence for the Earth*, [Online]. Available: go.iec.ch/wpai008. [Accessed: 14 September 2018].

[21] CHAKRABORTY, C. and JOSEPH, A., Bank of England: Staff Working Paper No.674, *Machine learning at central banks* [Online]. Available: go.iec.ch/wpai009. [Accessed: 14 September 2018].

[22] Statista, *Adoption level of AI in business organizations worldwide, as of 2017* [Online]. Available: go.iec.ch/wpai010. [Accessed: 14 September 2018].

[23] McKinsey Global Institute, *Artificial Intelligence: implications for China* [Online]. Available: go.iec.ch/wpai011. [Accessed: 14 September 2018].

[24] SAP, *European Prosperity Through Human-Centric Artificial Intelligence* [Online]. Available: go.iec.ch/wpai012. [Accessed: 14 September 2018].

[25] KPMG, *Venture Pulse Q4 2017: Global analysis of venture funding* [Online]. Available: go.iec.ch/wpai013. [Accessed: 14 September 2018].

[26] The State Council of the People's Republic of China, *China issues guidelines on artificial intelligence development* [Online]. Available: go.iec.ch/wpai014. [Accessed: 14 September 2018].

[27] KOCHURA, Y., STIRENKO, S., ROJBI, A., ALIENIN, O., NOVOTARSKIY, M., GORDIENKO, Y., *Comparative Analysis of Open Source Frameworks for Machine Learning with Use Case in Single-Threaded and Multi-Threaded Modes*, CSIT, Vol. 1, 2017, pp. 373-376.

[28] FERRUCCI, D., BROWN, E., CHU-CARROLL, J., FAN, J., GONDEK, D., KALYANPUR, A. A., LALLY, A., MURDOCK, J. W., NYBERG E., PRAGER, J., SCHLAEFER, N., WELTY, C., *Building Watson: An Overview of the Deep QA Project*, AI Magazine, Vol. 31 (3), 2010, pp. 59-79.

[29] datascience@berkely, *Moore's Law and Computer Processing Power*, 2014 [Online]. Available: go.iec.ch/wpai015. [Accessed: 14 September 2018].

[30] Google, Cloud Big Data and Machine Learning Blog, *An in-depth look at Google's first Tensor Processing Unit (TPU)* [Online]. Available: go.iec.ch/wpai016. [Accessed: 14 September 2018].

[31] GEORGHIADES, A., Yale face database, Center for computational Vision and Control at Yale University, 1997 [Online]. Available: go.iec.ch/wpai017. [Accessed: 14 September 2018].

[32] *Face Dataset on Image-net* [Online]. Available: go.iec.ch/wpai018. [Accessed: 14 September 2018].

[33] Summary and Statistics of Image-net, 2010 [Online]. Available: go.iec.ch/wpai019. [Accessed: 14 September 2018].

[34] HAHNLOSER, R. H. R., SARPESHKAR, R., MAHOWALD, M. A., DOUGLAS, R. J., SEUNG, H. S., *Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit*, Nature, Vol. 405, 2000, pp.947-951.

[35] GLOROT, X., BORDES, A., BENGIO, Y., *Deep Sparse Rectifier Neural Networks*, Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, Vol. 15, 2011, pp. 315-323.

[36] LECUN, Y., GALLAND, C. C., HINTON, G. E., *GEMINI: Gradient estimation through matrix inversion after noise injection*, In: Advances in Neural Information Processing Systems, Denver, 1989, pp. 141-148.

[37] LECUN, Y., *Generalization and Network Design Strategies*, In: PFEIFER, R., SCHRETER, Z., FOGELMAN, F. and STEELS, L. (Eds), *Connectionism in Perspective*, Elsevier, Zurich, Switzerland, 1989.

[38] CIREGAN, D., MEIER, U., SCHMIDHUBER, J., *Multi-column deep neural networks for image classification*, IEEE Transactions on Computer Vision and Pattern Recognition, USA, 2012, pp. 3642-3649.

[39] ISO/IEC DIS 20546, *Information technology – Big data – Overview and vocabulary* [under development].

[40] Promotorengruppe Kommunikation der Forschungsunion Wirthschaft – Wissenschaft, *Deutschlands Zukunft als Produktionsstandort sichern: Umsetzungsempfehlungen für das Zukunfsprojekt Industrie 4.0: Abschlussbericht des Arbeitskreises Industrie 4.0* [Online]. Available: go.iec.ch/wpai020. [Accessed: 14 September 2018].

[41] ISO/IEC JTC 1: Internet of Things (IoT), Preliminary Report, 2014 [Online]. Available: go.iec.ch/wpai021. [Accessed: 14 September 2018].

[42] PERERA, C., LIU, C.H., JAYAWARDENA, S., CHEN, M., *A survey on Internet of Things from industrial market perspective*, IEEE Access, Vol. 2, 2014, pp. 1660-1679.

[43] Statista, *Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions)* [Online.] Available: go.iec.ch/wpai022. [Accessed: 14 September 2018].

[44] IEC, *IoT 2020: Smart and secure IoT platform*, White Paper, 2016 [Online]. Available: go.iec.ch/wpai023. [Accessed: 14 September 2018].

[45] MILLS, S., LUCAS, S., IRAKLIOTIS, L., RAPPA, M., CARLSON, T., PERLOWITZ, B., *Demystifying Big Data: A Practical Guide to Transforming the Business of Government*, TechAmerica Foundation, Washington, 2012.

[46] PINITORE, G., RAO, V., CAVALLARO, K., DWIVEDI, K., *To share or not to share: What consumers really think about sharing their personal information* [Online]. Available: go.iec.ch/wpai024. [Accessed: 14 September 2018].

[47] Google, *How Search algorithms work* [Online]. Available: go.iec.ch/wpai025. [Accessed: 14 September 2018].

[48] BRACHMAN, R.J., KHABAZA, T., KLOESGEN, W., PIATETSKY-SHAPIRO, G., SIMOUDIS, E., *Mining business databases*, Communications of the ACM, Vol. 39, No. 111996, pp. 42-48.

[49] GUZELLA, T.S., CAMINHAS, W.M., *A review of machine learning approaches to spam filtering*, Expert Systems with Applications, Vol.36, No.7, 2009, pp. 10206-10222.

[50] BRADELY, T., *Facebook Uses Facial Recognition to help you manage your identity* [Online]. Available: go.iec.ch/wpai026. [Accessed: 14 September 2018].

[51] MELENDEZ, S., *How Pinterest Uses Machine Learning To Keep Its Users Pinned* [Online]. Available: go.iec.ch/wpai027. [Accessed: 14 September 2018].

[52] DAVIS, J.D., *Sentiment analysis for Instagram and Twitter now available to everyone* [Online]. Available: go.iec.ch/wpai028. [Accessed: 14 September 2018].

[53] CADE, D.L., *A Look at How Snapchat's Powerful Facial Recognition Tech Works* [Online]. Available: go.iec.ch/wpai029. [Accessed: 14 September 2018].

[54] HABIB, O., *Conversational Technology: Siri, Alexa, Cortana and the Google Assistant* [Online]. Available: go.iec.ch/wpai030. [Accessed: 14 September 2018].

[55] QUINLAN, J. R., *Induction of decision trees*, Machine Learning, Vol. 1, No.1, 1986, pp. 81-106.

[56] BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A., STONE, C. J., *Classification and regression trees*, Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA, 1984.

[57] BOSER, B. E., GUYON, I. M., VAPNIK, V. N., *A training algorithm for optimal margin classifiers*, COLT '92: proceedings of the fifth annual workshop on Computational learning theory, pp. 144-152, 1992.

[58] DIETTRICH, T. G., *An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization*, Machine Learning, Vol. 40, No. 2, 2000, pp.139-158.

[59] NG, A., JORDAN, M., *On Discriminative vs. Generative classifiers: A comparison of logistic regression and naïve Bayes*, In: *Advances in Neural Information Processing Systems: Natural and Synthetic*, Massachusetts Institute of Technology Press, 2001, pp. 841-848.

[60] ANDROUTSOPOULOS, I., KOUTSIAS, J., CHANDRINOS, K., Paliouras, G., SPYROPOULOS, C., *An Evaluation of Naïve Bayesian Anti-Spam Filtering*, In: *Proceedings of the workshop on Machine Learning in the New Information Age*, 11th European Conference on Machine Learning, 2000, pp. 9-17.

[61] LIU, B., BLASCH, E., CHEN, Y., SHEN, D., CHEN, G., *Scalable sentiment classification for Big Data analysis using Naive Bayes Classifier*, In: *IEEE International Conference on Big Data*, USA, 2013.

[62] CUNNINGHAM, P., DELANY, S. J., *k-Nearest neighbour classifiers*, Technical Report UCD-CSI-2007-4, 2007.

[63] FORGEY, E., *Cluster analysis of multivariate data: Efficiency vs. interpretability of classification*, Biometrics, 21 (3), 1965, pp. 768-769.

[64] MACQUEEN, J., *Some methods for classification and analysis of multivariate observations*, In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 1967, pp. 281-297.

[65] HARTIGAN, J. A., WONG, M. A., *A K-Means Clustering Algorithm*, Applied Statistics Journal of the Royal Statistical Society, Vol. 28, No. 1, 1979, pp. 100-108.

[66] RABINER, L.R., *A tutorial on hidden Markov models and selected applications in speech recognition*, Proceedings of the IEEE, Vol. 77 (2), 1989, pp. 257-286.

[67] DURBIN, R., EDDY, S.R., KROGH, A., MITCHISON, G.J., *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge UK, 1998.

[68] MILLER, D.R.H., LEEK, T., SCHWARTZ, R.M., *A Hidden Markov Model Information Retrieval System*, Proceedings of the 22th annual international ACM SIGIR conference on Research and development in information retrieval, 1999, pp. 214-221.

[69] VARGA, A.P., MOORE, R.K., *Hidden Markov model decomposition of speech and noise*, International Conference on Acoustics, Speech and Signal Processing, USA, 1990.

[70] KARPATHY, A., *Neural Networks, Part 1: Setting up the Architecture* [Online]. Available: go.iec.ch/wpai031. [Accessed: 14 September 2018].

[71] NIELSEN, M., *Using Neural nets to recognize handwritten digits* [Online]. Available: go.iec.ch/wpai032. [Accessed: 14 September 2018].

[72] KARPATHY, A., *Convolutional Neural Networks for Visual Recognition* [Online]. Available: go.iec.ch/wpai033. [Accessed: 14 September 2018].

[73] KRIZHEVSKY, A., SUTSKEVER, I., HINTON, G. E., *ImageNet classification with deep convolutional neural networks*, Proceedings of the 25th International Conference of Neural Information Processing Systems, Vol. 1, 2012, pp. 1097-1105.

[74] ELMAN, J. L., *Finding Structure in Time*, Cognitive Science, Vol. 14, No. 2, 1990, pp. 179-211.

[75] WERBOS, P. J., *Backpropagation Through Time: What It Does and How to Do It*, Proceedings of the IEEE, Vol. 78, No. 10, 1990, pp. 1550-1560.

[76] BENGIO, Y., SIMARD, P., FRASCONI, P., *Learning long-term dependencies with gradient descent is difficult*, IEEE Transactions on Neural Networks, Vol. 5, No. 2, 1994, pp. 157-166.

[77] SCHUSTER, M., PALIWAL, K. K., *Bidirectional recurrent neural networks*, IEEE Transactions on Signal Processing, Vol. 45, No. 11, 1997, pp. 2673-2681.

[78] McKinsey Global Institute, *Artificial Intelligence – The next digital frontier?* [Online]. Available: go.iec.ch/wpai034. [Accessed: 14 September 2018].

[79] HOCHREITER, S., SCHMIDHUBER, J., *Long-short term memory*, Neural Computation, Vol. 9, No. 8, 1997, pp. 1735-1780.

[80] European Commission DG Connect, *Call for a High-Level Expert Group on Artificial Intelligence* [Online]. Available: go.iec.ch/wpai035. [Accessed: 14 September 2018].

[81] ITU-T, *Focus Group on Machine Learning for Future Networks including 5G* [Online]. Available: go.iec.ch/wpai036. [Accessed: 14 September 2018].

[82] IEEE, *Background, Mission and Activities of the IEEE Global Initiative* [Online]. Available: go.iec.ch/wpai037. [Accessed: 14 September 2018].

[83] SESHIA, S. A., SADIGH, D., SASTRY, S. S., *Towards Verified Artificial Intelligence*, 2017 [Online]. Available: go.iec.ch/wpai038. [Accessed: 14 September 2018].

[84] Partnership on AI [Online]. Available: go.iec.ch/wpai039. [Accessed: 14 September 2018].

[85] Royal Society for the encouragement of Arts, Manufacturers and Commerce, *The Age of Automation: Artificial intelligence, robotics and the future of low-skilled work* [Online]. Available: go.iec.ch/wpai040. [Accessed: 14 September 2018].

[86] Information Technology Industry Council (ITI), *AI Policy Principles* [Online]. Available: go.iec.ch/wpai041. [Accessed: 14 September 2018].

[87] HOWELLS, Jan., *Can AI make manufacturing smart?* [Online]. Available: go.iec.ch/wpai042. [Accessed: 14 September 2018].

[88] TU, Y., LI, X., *Industrial Engineering and Management: Future Research and Development in Maintenance Engineering and Management*, Industrial Engineering and Management, Vol. 4, 2004, pp. 7-12.

[89] GAO H., LI D., XU M., *Intelligent Monitoring System for Screw Life Evaluation*, Journal of Southwest Jiaotong University, 2010, 45(5):685-691.

[90] PESHKIN, M., COLGATE, J. E., *Cobots*, Industrial Robot, 26 (5), 1999, pp. 335-341.

[91] HONGYI, L., LIHUI, W., *Gesture recognition for human-robot collaboration: A review*, International Journal of Industrial Ergonomics, 2017, pp. 1-13.

[92] PIERSON, H. A, GASHLER, M. G., *Deep Learning in Robotics: A Review of Recent Research* [Online]. Available: go.iec.ch/wpai043. [Accessed: 14 September 2018].

[93] IEC, *Stable grid operations in a future of distributed electric power*, White Paper, 2018 [to be published].

[94] The Economist, *The cost of traffic jams*, November 2014 [Online]. Available: go.iec.ch/wpai044. Accessed: 14 September 2018]

[95] NHTSA, *U.S. Department of Transportation Releases Policy on Automated Vehicle Development* [Online]. Available: go.iec.ch/wpai045. [Accessed: 14 September 2018].

[96] BURANYI, S., *Rise of the racist robots – how AI is learning all our worst impulses* [Online]. Available: go.iec.ch/wpai046. [Accessed: 14 September 2018].

[97] BBC, *Google apologises for Photos app's racist blunder* [Online]. Available: go.iec.ch/wpai047. [Accessed: 14 September 2018].

[98] ANGWIN, J. et al., *Machine Bias* [Online]. Available: go.iec.ch/wpai048. [Accessed: 14 September 2018].

[99] Executive Office of the President of the United States of America, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* [Online]. Available: go.iec.ch/wpai049. [Accessed: 14 September 2018].

[100] BAROCAS, S., HARDT, M., *Fairness in Machine Learning*, NIPS 2017 Tutorial [Online]. Available: go.iec.ch/wpai050. [Accessed: 14 September 2018].

[101] AMODEI, D., OLAH, C., STEINHARDT, J., CHRISTIANO, P., SCHULMAN, J., MANÉ, D., *Concrete Problems in AI Safety* [Online]. Available: go.iec.ch/wpai051. [Accessed: 14 September 2018].

[102] XU, H., MANNOR, S., *Robustness and generalization*, Machine Learning, Vol. 86, No. 3, 2012, pp. 391-423.

[103] ULIČNÝ, M., LUNDSTRÖM, J., BYTTNER, S., *Robustness of Deep Convolutional Neural Networks for Image Recognition*, International Symposium on Intelligent Computing Systems, 2016, pp. 16-30.

[104] RUSSELL, S., DEWEY, D., TEGMARK, M., *Research Priorities for robust and beneficial artificial intelligence*, AI Magazine, Vol. 36, No. 4, 2015, pp. 105-114.

[105] PAN, S. J., YAN, Q., *A Survey on transfer learning*, IEEE Transactions on Knowledge and Data Engineering, Vol. 22, Nr. 10, 2010.

[106] RIBEIRO, M., SINGH, S., GUESTRIN, C., *Why Should I Trust You?: Explaining the Predictions of Any Classifier* [Online]. Available: go.iec.ch/wpai052. [Accessed: 14 September 2018].

[107] LIPTON, Z., *The Mythos of Model Interpretability* [Online]. Available: go.iec.ch/wpai053. [Accessed: 14 September 2018].

[108] OpenAI, *Concrete Problems in AI Safety* [Online]. Available: go.iec.ch/wpai054. [Accessed: 14 September 2018].

[109] IEEE, *Safety and Beneficience of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI)* [Online]. Available: go.iec.ch/wpai055. [Accessed: 14 September 2018].

[110] MCMAHAN, B., RAMAGE, D., *Federated Learning: Collaborative Machine Learning without Centralized Training Data* [Online]. Available: go.iec.ch/wpai056. [Accessed: 14 September 2018].

[111] go.iec.ch/wpai057 [Online]. [Accessed: 14 September 2018].

[112] go.iec.ch/wpai058 [Online]. [Accessed: 14 September 2018].

[113] MIT Technology Review, *Why Self-Driving Cars Must Be Programmed to Kill* [Online]. Available: go.iec.ch/wpai059. [Accessed: 14 September 2018].

[114] Numenta, *Hierarchical Temporal Memory including HTM Cortical Learning Algorithms* [Online]. Available: go.iec.ch/wpai060. [Accessed: 14 September 2018].

[115] HAWKINS, J., BLAKESLEE, S., *On Intelligence: How a New understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*, Times Books, 2004.

[116] go.iec.ch/wpai061 [Online]. [Accessed: 14 September 2018].

[117] Deloitte, *Grenzenlos vernetzt* [Online]. Available: go.iec.ch/wpai062. [Accessed: 14 September 2018].

[118] VORHIES, W., *Digital Twins, Machine Learning & AI* [Online]. Available: go.iec.ch/wpai063. [Accessed: 14 September 2018].

[119] SCHALKWYK, P., *How AI bots bring digital twins to life* [Online]. Available: go.iec.ch/wpai064. [Accessed: 14 September 2018].

[120] SILVER, D., *Deep Reinforced Learning* [Online]. Available: go.iec.ch/wpai065. [Accessed: 14 September 2018].

[121] LE, Q., ZOPH, B., *Using Machine Learning to Explore Neural Network Architecture* [Online]. Available: go.iec.ch/wpai066. [Accessed: 14 September 2018].

[122] ZOPH, B., VASUDEVAN, V., SHLENS, J., LE, Q., *AutoML for large scale image classification and object detection* [Online]. Available: go.iec.ch/wpai067. [Accessed: 14 September 2018].

Notes

Notes

Notes