BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF FOREIGN LANGUAGES
ÚSTAV JAZYKŮ

ARTIFICIAL INTELLIGENCE
UMĚLÁ INTELIGENCE

BACHELOR'S THESIS
BAKALÁŘSKÁ PRÁCE

AUTHOR — Luděk Ragas
AUTOR PRÁCE

SUPERVISOR — Mgr. Pavel Sedláček
VEDOUCÍ PRÁCE

BRNO 2017

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

Bakalářská práce

bakalářský studijní obor **Angličtina v elektrotechnice a informatice**
Ústav jazyků

Student: Luděk Ragas ***ID:*** 164008
Ročník: 3 ***Akademický rok:*** 2016/17

NÁZEV TÉMATU:

Umělá inteligence

POKYNY PRO VYPRACOVÁNÍ:

Nastudujte literaturu zabývající se historií a technologií umělé inteligence. Vezměte v potaz rovněž potenciální vlivy umělé inteligence na společnost. Zpracujte práci dle uvedené osnovy a doplňte závěr.

DOPORUČENÁ LITERATURA:

JURA, Pavel.: Základy fuzzy logiky pro řízení a modelování. Brno: VUTIUM, 2003. ISBN 80-214-2261-0

FRANKISH K.: The Cambridge Handbook of Artificial Intelligence, Cambridge: CUP, 2014, ISBN-13: 978-0521691918

Termín zadání: 9.2.2017 ***Termín odevzdání:*** 2.6.2017

Vedoucí práce: Mgr. Pavel Sedláček
Konzultant:

doc. PhDr. Milena Krhutová, Ph.D.
předseda oborové rady

ABSTRACT

The aim of this bachelor thesis is to provide insight into the extensive field of artificial intelligence. At first, the thesis provides a definition of artificial intelligence and a brief overview of its history. Thereafter, this work also briefly describes technologies of artificial intelligence, such as neural networks, expert systems, and genetic algorithms. Lastly, this thesis describes influence and position of artificial intelligence in society.

KEYWORDS

Artificial intelligence, AI, history of artificial intelligence, Turing test, neural network, expert system, genetic algorithm, future of artificial intelligence, society, extinction of humanity, superintelligence

ABSTRAKT

Cílem této bakalářské práce je poskytnout náhled do rozsáhlého oboru umělé inteligence. Téze nejprve poskytuje definici umělé inteligence a krátký přehled její historie. Poté práce stručně popisuje technologie umělé inteligence, jako jsou neuronové sítě, expertní systémy a genetické algoritmy. Na závěr téze popisuje vliv umělé inteligence na společnost a její pozici v ní.

KLÍČOVÁ SLOVA

Umělá inteligence, UI, historie umělé inteligence, Turingův test, neuronová síť, expertní systém, genetický algoritmus, budoucnost umělé inteligence, společnost, vyhynutí lidstva, superinteligence

RAGAS, L. *Umělá inteligence*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2017. 38 s. Vedoucí bakalářské práce Mgr. Pavel Sedláček.

Prohlášení

Prohlašuji, že svoji bakalářskou práci na téma Umělá inteligence jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne ....

(podpis autora)

Contents

Introduction 7

1 Definition of Artificial Intelligence 8

2 History of Artificial Intelligence 10

2.1 Modern Artificial Intelligence 10

3 Technologies 15

3.1 Artificial Neural Networks 15

3.1.1 Artificial Neuron 15

3.1.2 Topology of Neural Networks 17

3.1.3 Learning of Neural Networks 19

3.2 Expert Systems 20

3.2.1 Structure of Expert Systems 22

3.2.2 Types of Expert Systems 25

3.3 Genetic Algorithms 25

3.3.1 Fitness Function and Selection 27

3.3.2 Genetic Operators 28

3.3.3 Differential Evolution 29

4 Artificial Intelligence in Society 29

4.1 Future of AI in Society 31

Conclusion 34

References 36

List of Figures 38

Introduction

Nowadays, people live in the world where technologies and artificial intelligence are practically everywhere. The artificial intelligence supports people at work, as well as it assists them in other aspects of their daily lives. However, a lot of people do not realize they are using it and for some of them artificial intelligence means only humanoid robots from sci-fi movies. Nevertheless, the artificial intelligence is hidden in the majority of technologies; including internet search engines, which use artificial intelligence to identify relevant results for their users, smartphones with integrated artificial assistants, cars that can use it for identification of obstacles on the road, or smart buildings that use artificial intelligence for reduction of energy consumption.

Therefore, this thesis is focused on providing insight into the extensive field of artificial intelligence. The first part is focused on the explanation of artificial intelligence and what makes these machines intelligent. Besides, it also attempts to define boundaries between automated systems and artificial intelligence.

The second part of this thesis describes the history of artificial intelligence from the first very simple automated systems in the ancient world. It continues to the point when these systems ceased to be simple automatic machines and became intelligent. Thereafter, it continues to the present day and provides a brief overview of the history of artificial intelligence by displaying some of the important milestones in the modern history.

Thereafter, the third part of this work presents three of the most common technologies used for development of artificial intelligence and describes the main differences between them. However, technologies in this part are described only briefly due to a limited scope of this thesis, which does not allow a deeper explanation as well as a presentation of coding samples.

The last part of the thesis describes the influence of artificial intelligence on society. It informs about benefits to the society, but mainly it discusses disadvantages and dangers that the artificial intelligence might create.

1 Definition of Artificial Intelligence

Artificial intelligence (AI) is an approach to modeling, recreating, and understanding cognitive processes and intelligence. This approach involves multiple branches of science and using diverse logical, mathematical, computational, biological, and still even mechanical principles and devices. AI is a critical branch of cognitive science. It can be abstract and theoretical while it attempts to improve and extend human understanding of natural cognition or prove theories or help define limits of computers. However, artificial intelligence can also be very pragmatic with the focus on applications and engineering of smart devices. (Frankish, 2014, p. 14)

From this pragmatic point of view, artificial intelligence can be considered as the intelligence which is performed by machines. In computer science, these intelligent machines can perceive their environment and take necessary actions to increase their chances to successfully finish their tasks. People usually use the term *artificial intelligence* in cases when devices simulate cognitive functions, such as problem-solving or learning, which people have connected with other human beings and their minds.

Research of artificial intelligence is conducted in different subfields, relating to different dimensions of intelligence. These dimensions are:

- **Perception** – which is described as an interpretation of information provided by senses in order to understand and represent the environment.
- **Learning** – is the way to obtain new knowledge and modify or reinforce the existing one. This absorption of new information leads to improvements or other various changes of skills and behavior of an individual.
- **Reasoning** – is the ability to apply logic, forming and verifying facts, justifying and making sense of things and practices. Therefore, reasoning is also connected with cognition and thinking. It can be divided into *logical*, *deductive*, *inductive, abductive, verbal,* and *intuitive reasoning*.
- **Action** – which is associated with intelligence can be described as *purposive, intentional,* and subjectively meaningful activity for the individual. However, the term action can be also used for *a reflexive* and *automatic* activity, where the difference is that the action is not made by consciousness. This is the reason why it cannot be connected with the intelligence of the individual.

- **Language** – is a complex system of communication which is used by humans. It is considered, especially in association with intelligence, as the ability to learn and use a language.
- **Consciousness** – is the state of being awake and aware of external objects, environment, but also of oneself. Farthing (1992) states the consciousness has been defined as *sentience*, *awareness*, *subjectivity*, *wakefulness*, *the ability to experience* or *to feel*, *having a sense of selfhood*, and *the executive control system of the mind*.

(Adapted from: Frankish, 2014, p. 20)

Progress in research of artificial intelligence allows development of new methods, technologies, and ways in which intelligent machines can be and are used. Nevertheless, this progress changes even the kind of machines that is still considered intelligent or it changed into routine technology. An example of technology which was previously perceived as intelligent and now it is perceived as routine technology is optical character recognition. This recognition technology is widely used for digitizing documents, conversion of printed or handwritten text from an image into machine-encoded text, so they can be easily processed by computers.

The continuously changing definition of intelligent machine makes it impossible to define the exact boundaries between artificial intelligence and advanced automated systems. Machines currently perceived as artificial intelligence are able to interpret complex data, understand human speech, drive cars, or even compete at a high level with humans in various strategic games.

2 History of Artificial Intelligence

History of artificial intelligence begins in the ancient world. The first artificial beings were mentioned in Greek mythology, where they were made by gods like Hephaestus, who is the god of metals, blacksmiths, craftsmen, et al. Creatures made by Hephaestus were usually made from gold with an appearance of animals. However, some of his creatures were made even from regular metal and were used by Hephaestus to work for him.

In ancient times people were not capable of making artificial intelligence, but they attempted to get closer to the gods by making statues. In these statues, they believed, was hidden something from a god himself/herself which the statue represented, so it had some divine powers. Nevertheless, simple statues were not good enough and people needed something more real. Therefore, new inventions began to appear. These inventions were supposed to do something without human help, so they might fool people that the gods were responsible for that. However, other machines were simultaneously invented with these inventions for *deceiving* people. These machines were supposed to help people do their work or to make their daily life more comfortable and easier in different ways. Such inventions were for example windmills or automatic doors. Due to the fact, that people in the past were not capable of making artificial intelligence, they at least attempted to make all of these machines as much autonomous as they could.

Although automation can be considered as a step between the first mentions of artificial intelligence and current modern AI, it is not subject of this thesis. Therefore, the rest of the history of artificial intelligence in this thesis is focused on the modern artificial intelligence.

2.1 Modern Artificial Intelligence

The history of modern artificial intelligence begins almost at the same time as the development of electronic computers. Therefore, this part of the thesis provides a brief overview of some important and interesting events in the history of artificial intelligence since that time. However, two gentlemen have to be mentioned at the beginning, before the overview. (Brief History, n.d.)

The first is an American author Isaac Asimov, who defined three laws of robotics and published them in 1950. These laws are: The first, “a robot may not injure a human being, or, through inaction, allow a human being to come to harm". The second, "a robot must obey the orders given it by human beings except where such orders would conflict with the First Law". The third, "a robot must protect its own existence as long as such protection does not conflict with the First or Second Law" (Jenkins & Seiler, 2014).

The second is an English mathematician and computer scientist Alan Mathison Turing, who published *Computing Machinery and Intelligence* also in 1950. This publication introduced Turing Test, which can be used for testing of intelligent behavior. (Brief History, n.d.) The original form of the test is the *Imitation Game*. The principle of the game is simple. A man and a woman are communicating with an interrogator only by written letters. The interrogator attempts to determine the sex of his correspondents. The woman helps the interrogator in making the right decision, but the man attempts to convince the interrogator to make the wrong decision. In computing it works on the same principle, the only change is that an artificial intelligence attempts to convince the human interrogator that it is a human too.

The overview begins in 1956 by the Dartmouth Conference, the first conference with the topic *artificial intelligence*, which was given to it by John McCarthy. In the same year Allen Newell, Herbert Simon, and J.C. Shaw (Carnegie Institute of Technology, now Carnegie Mellon University) written and demonstrated the first working AI program called the Logic Theorist. (Brief History, n.d.)

The first game-playing program, for checkers, was written by Arthur Samuel from IBM between years 1952 and 1962. (Brief History, n.d.) The aim of the program was to achieve acceptable skill to challenge the world champion in this game. Furthermore, the first industrial robot company, named Unimation, was founded in 1961. (Unimate - The First Industrial Robot, n.d.)

An idea of interactive graphics into computing was introduced for the first time during the year 1963. The introduction was part of the MIT dissertation on Sketchpad written by Ivan Sutherland. Furthermore, the first collection of articles devoted to artificial intelligence, called Computers and Thought, was published by Edward A. Feigenbaum and Julian Feldman in the same year. As well as, a program named ANALOGY was written by Thomas Evans. The program was part of his Ph.D. work at MIT and demonstrated the fact that computers are capable of solving even the analogy problems which are included in IQ tests. (Brief History, n.d.)

In 1964, the dissertation of Danny Bobrow at MIT proves that computers are able to understand natural language at the level which allows them correctly solve algebra word problems. Thereafter, in 1965, an interactive program ELIZA has been written by Joseph Weizenbaum (MIT). The ELIZA program is able to converse in English on any topic. A version which simulated a dialogue of a psychotherapist was very popular as a toy in AI centers on the ARPA-net. A year later, in 1966, the first Machine Intelligence Workshop was organized by Donald Michie at Edinburgh. The workshop became the first of the influential series which is organized every year. (Brief History, n.d.)

In 1968, limits of a simple neural net were demonstrated for the first time. This demonstration was published in Perceptrons by Marvin Minsky and Seymour Papert a year later. (Minsky & Papert, 1969)

The first International Joint Conference on Artificial Intelligence (IJCAI) was organized in 1969. The conference held in Washington, D.C. In this year was also defined a model of conceptual dependency for natural language understanding by Roger Schank (Stanford). The model was developed later by Wendy Lehnert and Robert Wilensky, in their Ph.D. dissertations at Yale University, for use in story understanding, and by Janet Kolodner for use in understanding memory. (Brief History, n.d.)

In 1971, a Ph.D. thesis written by Terry Winograd at MIT demonstrated a capability of computers to understand English sentences in a limited world of children's blocks. This was achieved by coupling a robotic arm which performed instructions typed in English, with language understanding program SHRDLU which Terry Winograd have written. (Brief History, n.d.)

In 1973, Freddy, the Famous Scottish Robot, was built. Freddy was using vision to locate and then assemble models. The robot was built by the Assembly Robotics group at Edinburgh University. (Edinburgh Freddy Robot, 2015)

MYCIN, which is sometimes called the first expert system, demonstrated the power of rule-based systems for inference and knowledge representation in 1974. The demonstration was made in the domain of medical diagnosis and therapy and was published in a Ph.D. dissertation on MYCIN at Stanford University, which was written by Ted Shortliffe. (Brief History, n.d.)

In 1975, an influential article on Frames as a representation of knowledge was published by Marvin Minsky. The article brought together many ideas about semantic links and schemas. In the same year were published the first scientific discoveries by a computer in a refereed journal. These discoveries were new results, produced by a learning program Meta-Dendral, in the field of chemistry. And a year later, in 1976, was demonstrated the power of meta-level reasoning in a Ph.D. dissertation of Randall Davis at Stanford University. (Brief History, n.d.)

The year 1978 was important due to the fact, that Herb Simon won the Nobel Prize in Economics as a result of his theory of bounded rationality. This theory became one of the pillars of artificial intelligence known as *satisficing*. Besides that, it was demonstrated that it is possible to use an object-oriented representation of knowledge for planning gene cloning experiments in this year. This demonstration was performed by the MOLGEN program, which was written by Peter Friedland and Mark Stefik at Stanford University. (Brief History, n.d.)

In 1979, a knowledge-based program for medical diagnosis called INTERNIST was developed. The program was written by Jack Myers and Harry Pople at the University of Pittsburgh and it is based on the clinical knowledge of Dr. Myers. (Brief History, n.d.) In the same year, the first autonomous vehicle controlled by a computer was built, the Stanford Cart. The vehicle, built by Hans Moravec, was able to successfully pass through a room filled with chairs and make a round trip in the Stanford AI Laboratory. This round trip lasted approximately five hours. (Earnest, 2012)

In 1980, the first National Conference of the American Association of Artificial Intelligence (AAAI) was organized in Stanford. (Brief History, n.d.) Furthermore, the blackboard model was described for the first time in this year. This model was used as the framework for the speech understanding system HEARSAY-II. This description was published by Lee Erman, Rick Hayes-Roth, Victor Lesser, and Raj Reddy. (Nii, 1986, p. 38)

The connection machine, which is a massively parallel architecture, was designed by Danny Hillis in 1981. This architecture provided a new power for artificial intelligence and for computers in general. Danny Hillis also later founded Thinking Machines, Inc. (Brief History, n.d.)

During the mid-1980s, artificial neural networks in combination with the Backpropagation algorithm, which was described for the first time by Werbos in 1974, became widely used. Besides that, the autonomous drawing program Aaron, which was written by Harold Cohen, was demonstrated at the AAAI National Conference in 1985. Development of Aaron lasted his author over a decade of work. (Brief History, n.d.)

In 1989, an Autonomous Land Vehicle in a Neural Network, called ALVINN, was created by Dean Pomerleau at CMU. This project later developed into the system which drove a computer-controlled car from the one coast of the United States to the other. The distance of this coast-to-coast trip was 2,800 miles. (Brief History, n.d.)

Brief History (n.d.) states that major advances in all areas of AI were achieved "with significant demonstrations in machine learning, intelligent tutoring, case-based reasoning, multi-agent planning, scheduling, uncertain reasoning, data mining, natural language understanding and translation, vision, virtual reality, games, and other topics" in the 1990s. In the same period, COG Project at MIT achieved significant progress in constructing of a humanoid robot. Furthermore, the World Wide Web began to be dependent on web crawlers and other information extraction programs, which are based on artificial intelligence. Lastly, in the late 1990s was performed the first demonstration of an Intelligent Room and Emotional Agents in AI Laboratory at MIT. (Brief History, n.d.)

In 1997, the world chess champion Garry Kasparov was defeated by the famous chess program Deep Blue, as well as in a rematch later. Another important event, in this year, was a successful landing of the first autonomous robotic system, Sojourner, on the planet Mars during the Mars Pathfinder mission. (Brief History, n.d.) The last mentioned event for 1997 is the first official RoboCup soccer match, which involved 40 teams of interacting robots and had over five thousand spectators. The competition was even observed by seventy international media. (Noda, et al., 1998, p. 49)

In 2000, a dissertation on Sociable Machines written by Cynthia Breazeal at MIT was published. The dissertation describes a robot, Kismet, which is able to express its emotions in its face. (Breazeal, 2000)

Five years later, in 2005, Stanley, an autonomous vehicle constructed at Stanford University, won the 2005 DARPA Grand Challenge.

Lastly, the beginning of the 21st century made all kinds of interactive robot pets, *smart toys*, commercially available, which made reality from the vision of innovative toy makers of the 18th century. (Brief History, n.d.)

3 Technologies

Developers use many technologies with different approaches to artificial intelligence in its creation. The chosen approach should allow a correct solution of the particular problem for which AI is created, as well as it should be as simple as possible. However, more complex tasks usually need a combination of at least two technologies. An example of such combination is a fuzzy expert system, which Mark Kantrowitz, Erik Horstkotte, and Cliff Joslyn (1993) described as "an expert system that uses a collection of fuzzy membership functions and rules, instead of Boolean logic, to reason about data".

This part of the thesis describes three chosen basic technologies for artificial intelligence. The first technology is *an artificial neural network*, which is mostly used for machine learning by imitation of biological nervous systems. The second technology is *an expert system*, which is used for decision making based on a knowledge base. And the last technology is *a genetic algorithm*, which is used for optimization by using mutations, crossovers, and selections inspired by natural evolution.

3.1 Artificial Neural Networks

According to Siganos and Stergiou (n.d.), "An Artificial Neural Network is an information-processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information". The artificial neural networks are based on a structure of an extensive number of artificial neurons. These neurons are basic processing elements of the neural networks and they are connected to each other very well through numerous interconnections which create a dense network. The artificial neural networks are usually created for a specific application. (Siganos & Stergiou, n.d.) It means that a large number of neural network types have been invented over the years of development. Differences between the types of neural networks can be located in a transfer function, topology, and even in a type of learning. (Černík, 2016, p. 7)

3.1.1 Artificial Neuron

Artificial neurons are essential units in an artificial neural network. An artificial neuron is based on a simplified model of a biological neuron. It means that same as the biological neuron, it has many inputs (*Dendrites*) and only one output (*Axon*). The cell body of the neuron (*Soma*) is described by *a basis function*. (Černík, 2016, p. 7) The basis function is one of two functions which are used in an artificial neuron. The second function is *an activation function*. These two functions together are described as *the transfer function*.

The neuron was described mathematically by Warren McCulloch and Walter Pitts for the first time in 1943. They also created the first very simple model of the artificial neuron. This model consists of binary inputs and their linear combination, an activation (*step*) function and one binary output. Today it is recognized as McCulloch–Pitts (MCP) unit or neuron. (Rojas, 1996, p. 30)

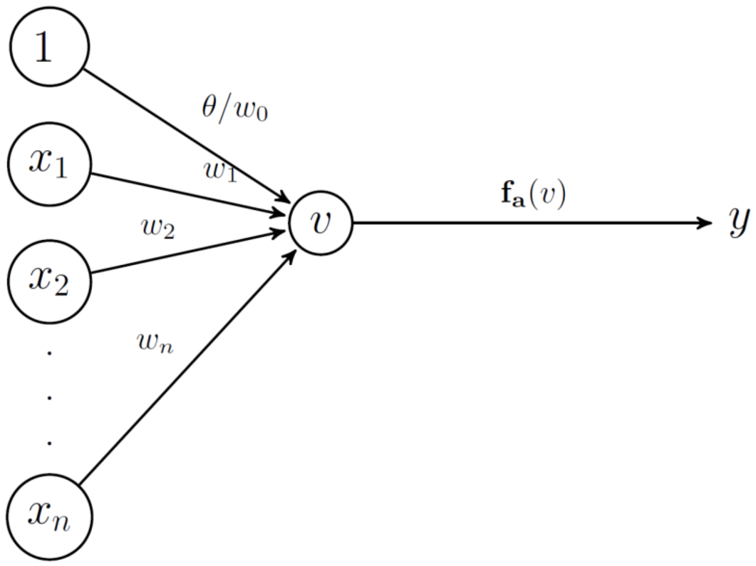

Figure 1. An artificial neuron. (Černík, 2016, p. 8)

An artificial neuron has many inputs $\vec{x}$ and every input has its weight w_n (see figure 1). For a description of such neuron is then used following equation 3.1.

$$y = f_a\left(f_b(\vec{x}, \vec{w})\right) \tag{3.1}$$

The function f_b is the basis function of the neuron. This function is used for determination of the value v which represents a potential of the neuron. This determination of the potential of the neuron is made by a combination of inputs and weights in the way that function describes. Therefore, the potential of the neuron can be expressed mathematically by the following equation 3.2.

$$v = f_b(\vec{x}, \vec{w}) \tag{3.2}$$

The function f_a is the activation function of the neuron. This function describes how to compute the output from the potential of the neuron. As the activation function is usually chosen a nonlinear function, which can be also expressed mathematically by the following equation 3.3.

$$y = f_a(v) \tag{3.3}$$

The activation function is chosen based on the function which a neural network should approximate. However, used type of the activation function must be the same as the one that should be reached. It means that for an approximation of a non-linear function must be used a non-linear activation function. (Černík, 2016, p. 8)

Nevertheless, Černík (2016) specifies that:

> The biggest difference between biological and artificial neuron is in the way neurons *fire* their outputs. Artificial neurons are usually computed in terms of a discrete simulation. In every step, the output of neuron is computed. On the other hand – biological neuron fires when it has enough power (with every incoming input it consumes power and when this gained power gets over value, it fires), this process is known as a continuous simulation. (p. 7)

3.1.2 Topology of Neural Networks

The way used for interconnection of the set of neurons is described as a topology of the network. Based on properties of topologies are characterized different types of network architectures. All types of topologies have to contain an *output layer* of neurons and an *input layer* of neurons. There are two possibilities how the input layer can operate. The first possibility can insert input signals directly into required interconnection. The second possibility is realized by a creation of an input neuron for each input signal. These input neurons have a constant value equal to their input signals. Other neurons are then connected to these input neurons and create a layer known as a *hidden layer*. The second approach is used more often. (Karásek, 2016, p. 10)

3.1.2.1 Feed-Forward Network

Apparently, the most commonly used topology is the feed-forward network. These networks consist of a various number of layers, but outputs of neurons from the one layer are always connected only to neurons of the successive layer (see figure 2). That means there is no cycle or any feedback connection between layers. Evaluation of such network can be executed by gradual propagation of results from the input layer, and then one layer after another, to the output layer. Commonly used variation of the feed-forward network is a fully connected feed-forward network, in which input of each neuron is connected to outputs of all neurons in the previous layer. (Karásek, 2016, p. 11)

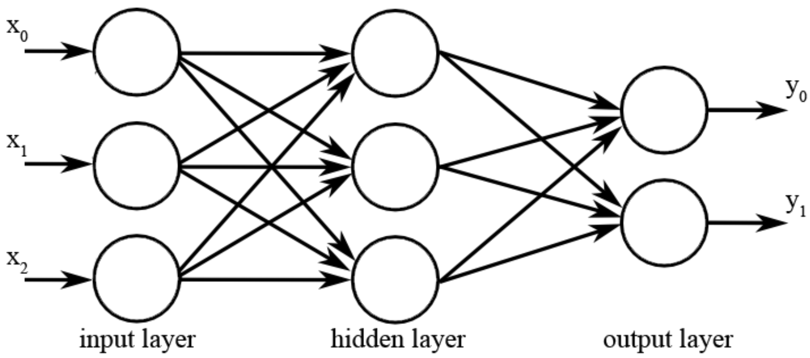

Figure 2. Fully-connected feed-forward network.

Feed-forward networks are commonly used for control, classification, and prediction. The most significant advantage of feed-forward networks is their easier learning. One of the well-known algorithms for learning of feed-forward network is Backpropagation algorithm. (Karásek, 2016, p. 11)

3.1.2.2 Recurrent networks

Recurrent networks contain at least one cycle and interconnections of their neurons can be made arbitrarily (see figure 3). Due to this possibility, recurrent networks do not have to be divided into layers as feed-forward networks. Fundamental feature and the biggest advantage of recurrent networks is their ability to use previous outputs from different neurons. That leads to a creation of recurrent connections and cycles. However, a simple evaluation of recurrent network is not possible due to these recurrent connections. Therefore, an iterative approach is used for their evaluation. This approach is similar to continuous simulation. In the first step output of each neuron is propagated into all other connected neurons. In the second step is from these new inputs evaluated a new output value of neuron. These steps can be repeated until the output value becomes stable. (Karásek, 2016, p. 11)

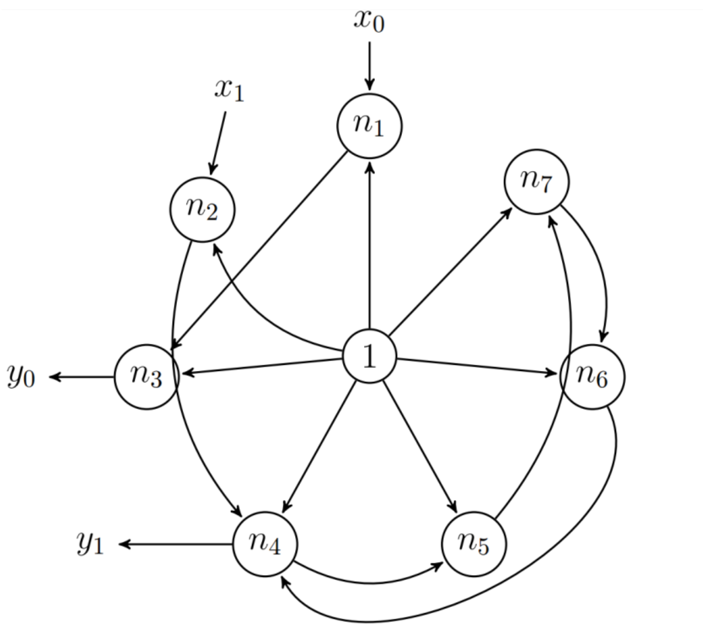

Figure 3. Recurrent neural network. (Černík, 2016, p. 10)

The ability to create relations between previous and actual inputs makes recurrent networks widely used in tasks with sequential relations. Such tasks are speech recognition, prediction, parsing, or control of dynamic systems. However, many tasks can be solved by both types of network. (Karásek, 2016, p. 12) The recurrent network is difficult to learn in contrast to the feed-forward network and its learning needs a lot of time. (Černík, 2016, p. 10)

3.1.3 Learning of Neural Networks

Learning of artificial neural networks is performed by learning from examples, which leads to adjustments of the interconnections between neurons. The same way of learning is performed by a human brain. (Siganos & Stergiou, n.d.) The target of the learning is not only in simple finding results in known examples. It is in giving correct results even for new input which the network has never seen before. (Gerstner, 1998, p. 6) This learning of neural networks can be *supervised* or *unsupervised*.

Supervised learning always needs a data base χ which is described by the equation 3.4. The data base contains examples for learning. In these examples, each set of inputs is given together with correct output. It means that the data base χ consists of P input-output pairs. During the learning process, the correct output t^{out} from the data base, also called *target value*, is compared with the output of the network x^{out} to obtain error value. The error $x^{out} - t^{out}$ is then fed back into the neural network to perform adjustments of parameters of the network. (Gerstner, 1998, p. 6)

$$\chi = \left\{ \left(x^{in}(\mu), t^{out}(\mu) \right); 1 \leq \mu \leq P \right\} \tag{3.4}$$

Unsupervised learning also uses a data base. However, the data base used in unsupervised learning does not contain each set of inputs together with target output. Unsupervised learning attempts to learn artificial neural networks to represent input patterns, so it reflects the statistical structure of input patterns in the data base. Into the learning process are involved only the observed input patterns and some a priori information, which can be explicit or implicit, about what is or is not important. According to Dayan (1999), these observed input patterns "are often assumed to be independent samples from an underlying unknown probability distribution" (p. 1). The artificial neural network which is learned unsupervised produces a transfer of prior biases about what features of the input structure should be preserved in the output. (Dayan, 1999, p. 1)

Dayan (1999) also reminds that:

> Two classes of method have been suggested for unsupervised learning. Density estimation techniques explicitly build statistical models (such as BAYESIAN NETWORKS) of how underlying causes could create the input. Feature extraction techniques try to extract statistical regularities (or sometimes irregularities) directly from the inputs. (p. 2)

3.2 Expert Systems

An expert system is an intelligent computer program which simulates decision making of a human expert in solving complex problems. The correct decision is acquired by reasoning about knowledge instead of using a conventional procedural method. Knowledge is adopted by the expert system from a human expert and it is saved explicitly in *a knowledge base*. Besides using the knowledge base, the expert system has to gather information about the actual problem and gathered information is processed by *an inference engine*. Gathering of information by the expert system usually simulates the way of gathering information by a human expert, which means that the expert system uses dialogues. These dialogues usually have a form of simple questions from the expert system and simple answers from a user. However, the first disadvantage of expert systems is connected with gathering information from the user. This disadvantage is the fact that expert systems are not able to detect when the user is lying, which might result in the wrong output of the system.

An example of the expert system which is asking questions to obtain information and everybody is able to test its functionality is *Akinator*. Akinator is a browser game based on the expert system, which guesses a real or fictional character based on one's answers, and it is accessible online at *en.akinator.com*. Furthermore, another characteristic feature of expert systems can be observed on Akinator. That feature is an ability of decision making even when uncertainty is present in gathered information. Akinator allows the user express uncertainty by adding options "Don't know", "Probably", and "Probably not" as possible answers to its questions.

Besides already mentioned features, an ability to explain the reasoning of the expert system should be mentioned. However, this feature and the feature of decision making with uncertainty are not mandatory in expert systems. Reasoning explanation is usually present in the expert system to improve the confidence of the user in the output of the system. Usually is explained currently asked question, knowledge relevant to some statement, target hypothesis which is currently examined, or derivation currently performed by the system. (Berka, Jirků, & Vejnarová, 1998, p. 34)

Tasks performed by expert systems can be divided into following common categories:

- **Interpretation** – explanation of observed data;
- **Prediction** – prediction of probable consequences of given situations;
- **Diagnosis** – detection of the reason of incorrect behavior of the system based on the observation;
- **Repair** – compilation of plan for elimination of system error;
- **Planning** – compilation of sequence of actions to achieve required result;
- **Monitoring** – evaluation of system behavior based on comparison of observed data with expected data;
- **Control** – regulation of processes, the control can contain *interpretation*, *diagnosis*, *monitoring*, *planning*, *prediction*, and *repair*;
- **Design** – compilation of appropriate components in appropriate way;
- **Instruction** – intelligent learning that allows students ask various questions.

(Adapted from: Dvořák, 2004, p. 10)

Therefore, expert systems are used often as *librarians*, which help users to find information, *advisors*, which provide particular expert knowledge, *instructors*, which help users to learn needed knowledge and solve problems, and *assistants*, which supervise routine tasks and allow users to focus on more creative aspects of the task.

Between advantages of expert systems belongs ability to solve difficult problems, possibility to use an expert system as a training tool for beginners in the field, availability of expertise and lower costs for their execution, as well as repeatability of any expertise which was previously executed. Besides advantages, there are few disadvantages of expert systems as already mentioned inability to detect a lie. Another disadvantage is a danger of failure in changed conditions or inability of the expert system to distinguish limitation of its application. (Dvořák, 2004, p. 10)

Dvořák (2004) also mentioned term *knowledge-based system* which can be found in some literature. This term, according to the older concept, is more general than the term expert system. Therefore, expert systems can be considered as a special type of knowledge-based systems, which are specified by using the expert knowledge and by some other features as reasoning explanation. However, the difference between these two terms is expiring recently. (Dvořák, 2004, p. 6)

3.2.1 Structure of Expert Systems

Expert systems are formed by two primary components which are *the knowledge base* and *the inference engine*. In the knowledge base is saved knowledge of an expert in the field and then the inference engine allows application of that knowledge during a consultation for the particular case. Other components of expert systems are *a data base* for the particular case, *an explanation system*, which allows explaining of the selected procedure in the reasoning of the system, and *a user interface*, for communication with the user. An example of a typical expert system, which uses dialogues with the user, is displayed in the following scheme (see figure 4). (Berka, Jirků, & Vejnarová, 1998, p. 34)

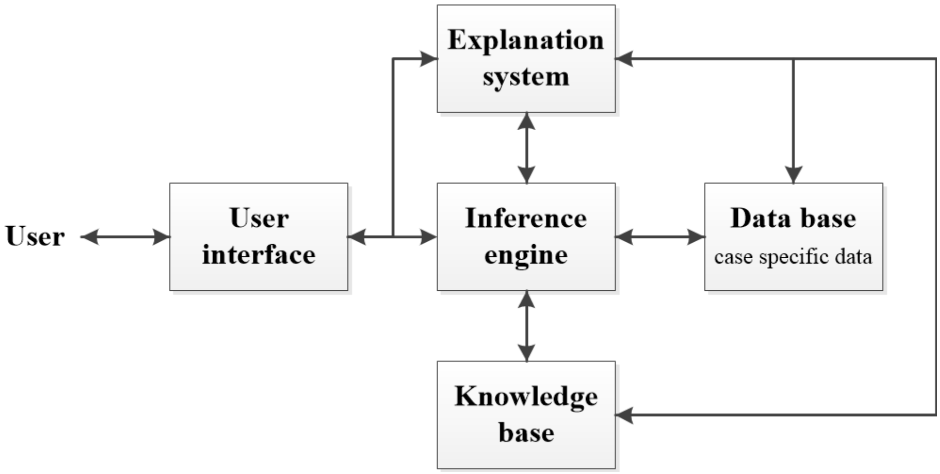

Figure 4. Scheme of an expert system.

3.2.1.1 Knowledge Base

The knowledge base is a data structure which contains knowledge of a specific domain, as well as the knowledge of problem-solving in that domain. (Dvořák, 2004, p. 6) Acquisition of this knowledge can be executed by a dialogue of a knowledge engineer with an expert in the field, who acquired the knowledge by learning and experience. For the acquisition can be used methods as *structured* or *unstructured interview*, *out loud thinking*, or *repertory table*. It is the basic way of the knowledge acquisition, which was for a long time also the only way. (Berka, Jirků, & Vejnarová, 1998, p. 46)

Nevertheless, this *manual acquisition* is a long and very difficult process. Therefore, in past few decades, various types of computer support were implemented into the knowledge acquisition process. The support can be *semi-automated* (interactive) which complements or even replaces the knowledge engineer in the dialogue, or it can be *fully automated*. As an example of fully automated computer support can be mentioned cluster analysis, case-based reasoning, machine learning, data mining, but even an artificial neural network or a genetic algorithm. (Berka, Jirků, & Vejnarová, 1998, p. 47)

The knowledge saved in the knowledge base can be in various forms:

- from the general knowledge to the very specific knowledge;
- from the *textbook* knowledge to the *private* knowledge;
- from the scientifically proven knowledge to uncertain heuristics;
- from simple knowledge to meta-knowledge (knowledge about the knowledge).

(Adapted from: Dvořák, 2004, p. 6)

Moreover, the knowledge is classified as *a shallow knowledge*, which is based on empiricism and heuristics, or as *a deep knowledge*, which is based on basic structures and functions. Similarly can be classified also reasoning, as *a shallow reasoning* or as *a deep reasoning*. (Dvořák, 2004, p. 7)

3.2.1.2 Inference Engine

The inference engine is a module which contains general algorithms. These domain-independent algorithms are able to solve problems based on the input data by manipulation knowledge from the knowledge base. A typical inference engine is based on inference rule for derivation new knowledge from the existing knowledge and on the search strategy in the knowledge base. (Dvořák, 2004, p. 8)

The inference itself, which is in expert systems the process of expertise generation, can be realized by multiple methods. These methods are:

- **Deduction** – reasoning from the general premise to the specific result, the result have to be based on premises;
- **Induction** – reasoning from the specific premise to the general result;
- **Abduction** – reasoning from the correct result to the premises that caused it;
- **Heuristics** – rules of *common sense* based on experience;
- **Generation and testing** – method based on repeated testing of generated solutions with respect to the specified requirements;
- **Analogy** – reasoning based on similarity with different situation;
- **Default reasoning** – reasoning based on the general knowledge when the specific knowledge is absent;
- **Non-monotonic reasoning** – based on the new facts, it is possible to correct or remove the current contradictory knowledge;
- **Intuition** – reasoning its results are apparently based on subconscious recognition of a pattern. Intuition has not been implemented in expert systems yet. However, it is comparable with the reasoning of artificial neural networks.

(Adapted from: Giarratano & Riley, 1998, p. 138)

An important ability the inference engine is also uncertainty processing. The uncertainty can occur in the data base, as well as in the knowledge base. This uncertainty is caused by the shallow knowledge or by an inaccuracy, incompleteness, and inconsistency of the data. The uncertainty processing and representation can be solved by methods as

Bayesian probability, certainty factors, Dempster–Shafer theory, or fuzzy logic. (Dvořák, 2004, p. 8)

Important is also a strategy that is used for the process of reasoning. Only two basic strategies exist – *forward chaining* and *backward chaining*. The forward chaining means that the system knows all the answers before the beginning of the consultation and attempts to derive a hypothesis. On the other hand, the backward chaining is typical for *diagnostic* expert systems and means that the system chooses and then derives a possible hypothesis by searching for corresponding data. The system applies rules from the knowledge base on these data; if the rules are valid for the data the hypothesis might be correct. In other words, the system proceeds backward (from the result). (Berka, Jirků, & Vejnarová, 1998, p. 69)

3.2.2 Types of Expert Systems

Expert systems can be divided according to a type of the solved problems into *diagnostic* and *planning* expert systems. Diagnostic expert systems attempt to determine which hypothesis from the specified set of hypotheses corresponds with the data of the specific case. On the other hand, planning expert systems solve tasks where are initial state and the result known. In these tasks, the system attempts to determine the entire procedure which will lead to that result. (Dvořák, 2004, p. 9)

However, expert systems can be classified even by other aspects. An example of another aspect can be the content of the knowledge base, which divides expert systems into *problem oriented* and *shells*. The knowledge base of the problem-oriented systems contains knowledge of the specific domain, but the shells have the knowledge base empty. (Dvořák, 2004, p. 9)

3.3 Genetic Algorithms

Genetic algorithms are the last technology for artificial intelligence which is mentioned in this thesis. These algorithms are heuristic procedures of optimization or solution finding. The basic principle of genetic algorithms is based on the theory of evolution and natural selection published by Charles Darwin in 1859. This theory states, that population in nature compete for resources, a possibility of reproduction, and fight for life with predators. Therefore, the most successful individuals survive longer and have better opportunity to find a partner for reproduction and have more descendants than less successful ones, who can die even without any descendant. Further, these descendants should have abilities of their successful ancestors, which should make them better adapted for life in their environment, thus, they should be same or even more successful than their ancestors.

The direct analogy of the theory of evolution is used in a genetic algorithm of John Holland, who was according to Hynek (2008) a pioneer in the field of genetic algorithms. This genetic algorithm can represent the basic genetic algorithm and its principle. Firstly, the algorithm obtains an initial population. Each individual in this population states for an encoded potential solution. In the next step, each individual is evaluated and receive its *fitness value*. The last step is a creation of a new population by using *genetic operators* and by imitation of the natural *selection*. This generation cycle (see figure 5) is usually repeated for many generations. Individuals with a better fitness value are preferred in each generation. Furthermore, after a specific number of repetitions a population with one or more individuals, which match acceptable or even ideal solution, is created. Genetic algorithms usually end after a specific number of cycles, a specific time period, or after some other conditions are met. (Hynek, 2008, p. 14)

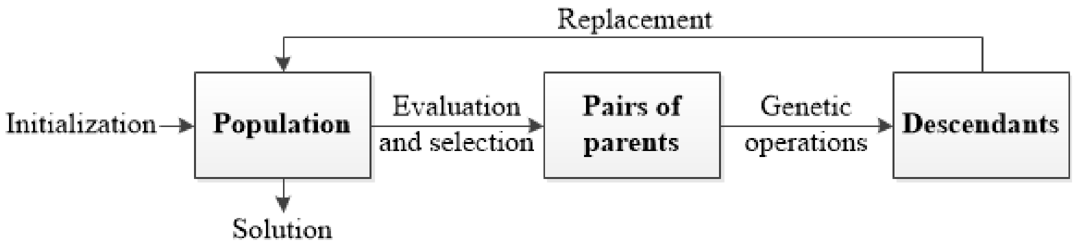

Figure 5. A cycle of a genetic algorithm.

However, genetic algorithms are heuristic procedures and therefore it is impossible to guarantee that their solutions are ideal. These algorithms use an entire set of potential solutions which makes them universal. On the other hand, a simple algorithm created for a specific problem, which uses only one initial solution method, is usually more effective. Moreover, it finds a solution of higher quality and faster. It makes genetic algorithms suitable especially for solving problems without any specialized solution method or when such a method is not known. Nevertheless, some really difficult problems can even require a combination of a genetic algorithm with a specialized one, but generally genetic algorithms are overused today. (Hynek, 2008, p. 15)

3.3.1 Fitness Function and Selection

Evaluation of individuals in the population is performed by a fitness function. This function decodes each individual in the population and evaluates a quality of its solution. The result of this evaluation is represented by a fitness value, which can be only positive. The higher fitness value means a better solution and higher possibility of reproduction of the individual. During designing the fitness function, it is necessary to make the function able to distinguish the different quality of solutions in the way towards the correct solution, which leads the algorithm to the domain with solutions of the higher quality. (Karásek, 2016, p. 5)

Thereafter, a process of selection is applied on the population. The selection chooses individuals from the population based on their fitness value and decides which of them will be used for the creation of a new population and which do not have any chance or only small chance to pass their genes. The ratio between suitable and unsuitable individuals is set by *a selection pressure*. However, exclusion of individuals with low fitness value can sometimes lead only to a local maximum, not to the ideal solution. This is caused by the possibility of exclusion of an individual which contains a part of the ideal solution, despite it has low fitness value and it is not a suitable solution by itself. (Karásek, 2016, p. 5)

The selection can be performed by one of many methods. However, only three selected methods of the selection, which are relevant as examples, are described in this thesis. These methods are *elitism*, *tournament selection*, and *roulette-wheel selection*.

3.3.1.1 Elitism

Elitism is one of the basic selection methods. The principle of elitism is very simple and despite its simplicity, according to Mitchell (1996), “many researchers have found that elitism significantly improves the GA's performance” (p. 126). The entire population is ordered from the individual with the highest fitness value to the one with the lowest value. Further, a certain part of the population with the lowest fitness value is eliminated. The size of the remaining population, which is called *elite*, is based on the selection pressure. The smaller elite makes the genetic algorithm faster, though it causes a risk that algorithm will find only the local maximum, which is described above. (Karásek, 2016, p. 5)

3.3.1.2 Tournament Selection

Tournament selection does not eliminate the part of the population with lower fitness value, which gives a chance even to the individuals with the lowest fitness value to be used in the creation of a new population. Moreover, an adjustment of the size of the tournament sets the selection pressure. (Karásek, 2016, p. 5)

Mitchell (1996) describes the principle of the tournament selection in this way:

> Two individuals are chosen at random from the population. A random number *r* is then chosen between 0 and 1. If $r < k$ (where *k* is a parameter, for example 0.75), the fitter of the two individuals is selected to be a parent; otherwise the less fit individual is selected. The two are then returned to the original population and can be selected again. (p. 128)

3.3.1.3 Roulette-wheel Selection

The roulette-wheel selection is the best method in a comparison of the quality of the individuals with the rest of the population. However, it is strongly dependent on the used fitness function and diversity of the fitness values in the population; higher diversity causes higher selection pressure. The roulette-wheel selection chooses parents proportionally according to their fitness value. Therefore, the chance of an individual to be selected is given by the proportion of its fitness value to the total fitness value of the entire population. (Karásek, 2016, p. 6)

3.3.2 Genetic Operators

The creation of new individuals for the next generation and increasing of population diversity is performed by genetic operators, which makes them essential for the process of artificial evolution. These operators are specific according to used coding and type of the task of the genetic algorithm. Therefore, the entire spectrum of operators exists from the simplest to very complex operators, which were designed for a specific task. (Karásek, 2016, p. 6)

Three of the most basic genetic operators are *crossover*, *mutation*, and *reproduction*. *Crossover* usually randomly divides two individuals into two parts and then exchanges a part of the first individual for a part of the second individual, which creates two completely new individuals. *Mutation* randomly changes some genes of an individual, which is performed with some very small probability that is usually smaller than one percent. In some genetic algorithms are used even operators of direct *reproduction* which creates just a copy of an individual. (Mitchell, 1996, p. 8)

3.3.3 Differential Evolution

Differential evolution is a special form of a continuous genetic algorithm. This population-based algorithm is designed to optimization of functions in a nonlinear continuous domain. According to Simon (2013), the principle of the differential evolution "is based on the idea of taking the difference vector between two individuals, and adding a scaled version of the difference vector to a third individual to create a new candidate solution" (p. 294). Individuals in the population are then encoded as n-dimensional vectors and the intensity of the changes is based on the population diversity. (Simon, 2013, p. 294)

4 Artificial Intelligence in Society

Influence of artificial intelligence can be observed almost everywhere in society. The main purpose of artificial intelligence is to make daily life easier and more comfortable. Therefore, artificial intelligence is already used in and affects many domains, such as *transportation*, *healthcare*, *education*, *entertainment*, *public safety and security*, *employment and workplace*, or various *service robots* for a home use. Nevertheless, people develop artificial intelligence in the way it was beneficial for humanity, artificial intelligence has not only many advantages, but also few disadvantages in these domains. Examples of these *advantages* are greater precision and accuracy, significant error reduction, a safer workplace, or new possibilities for humanity as in space exploration. On the other hand examples of *disadvantages* can be decreasing human abilities due to overuse of artificial intelligence instead of human intelligence, or a higher unemployed rate.

However, artificial intelligence creates also other problems in society which should be solved before mass production of artificially intelligent beings or other fully autonomous intelligent machines will start. These problems are mainly moral and legal. Such moral problem can be represented by a driverless car with malfunction of brakes and pedestrians on the road, which would be killed if the car will not perform any action to avoid them. Nevertheless, any action that the car would perform to save pedestrians will result in killing the passengers. This situation is based on the trolley problem, which is an ethical experiment.

Iyad Rahwan, a professor at MIT, with his collaborators Jean-Francois Bonnefon and Azim Shariff performed a survey, called *The Social Dilemma of Autonomous Vehicles*, to identify acceptable solution of this moral problem for society. The survey consists of the situation with the driverless car which is described above and gives respondents two options. *The first option* is inspired by philosopher Jeremy Bentham and states that the car should act based on utilitarian ethics, which means that the car should minimize total harm – save more lives – even though it kills passengers. *The second option* is inspired by philosopher Immanuel Kant and states that the car should act based on duty-bound principles, which mean that the car should not act in the way it *explicitly kills* anyone. Therefore, it should keep its course even though it *accidentally kills* more people on the road. Iyad Rahwan states that the most people in the survey responded that they want these cars to adopt utilitarianism to minimize total harm. However, Rahwan (2016) also adds that "there is a little catch when we asked people whether they would purchase such cars. They said absolutely not. They would like to buy cars that protect them at all costs. But they want everybody else to buy cars with minimized harm." More detailed results and description of the survey were published in Science magazine. (Rahwan, 2016)

Moreover, these problems are not only moral. There are also legal aspects of these problems around the artificial intelligence to be solved. From regulations of the artificial intelligence in the way of what it should or should not be able to do or how some of its abilities can or cannot be used in practice, to responsibility for actions of an intelligent machine. The example can be found again in the previously described situation of the driverless car. In this situation, when the car killed someone, a culprit has to be identified and it might be unclear if it is owner of the car, company that made and programmed the car, or if it is even the car itself. Therefore, the law must expect these situations, as well as, knows their proper solution.

Nevertheless, Tufekci (2016) points out that "humans have always made biases. Decision makers and gatekeepers, in courts, in news, in war ... they make mistakes" and also states that "we cannot escape these difficult questions. We cannot outsource our responsibilities to machines". The reason is that the artificial intelligence use mostly machine learning and its decision making is not transparent enough. Nobody can guarantee that it processed and then understands to learning data correctly. The artificial intelligence might provide seemingly correct results, but based on an incorrect attribute, for example, it could hire more men than women into army based not on their physical condition, but due to a simple fact that men cannot be pregnant and probably will not need maternity leave. Therefore, such subjective decisions, which require human values, should not be performed by artificial intelligence. (Tufekci, 2016)

Tufekci (2016) concludes that:

> Yes, we can and we should use computation to help us make better decisions. But we have to own up to our moral responsibility to judgment, and use algorithms within that framework, not as a means to abdicate and outsource our responsibilities to one another as human to human.

4.1 Future of AI in Society

The society has some concerns even about the role that the artificial intelligence will have in the future and how the humanity will be able to adapt to the changes that will occur. It is evident that the most of these changes will be positive as an increase in productivity and quality of life, crime reduction, elimination of menial and dangerous jobs, etc. Furthermore, an online survey[1], conducted by Tech Pro Research in July 2015, proves that 59 percent of 534 respondents expect the positive impact of the artificial intelligence on society (see figure 6). (Maddox, 2015)

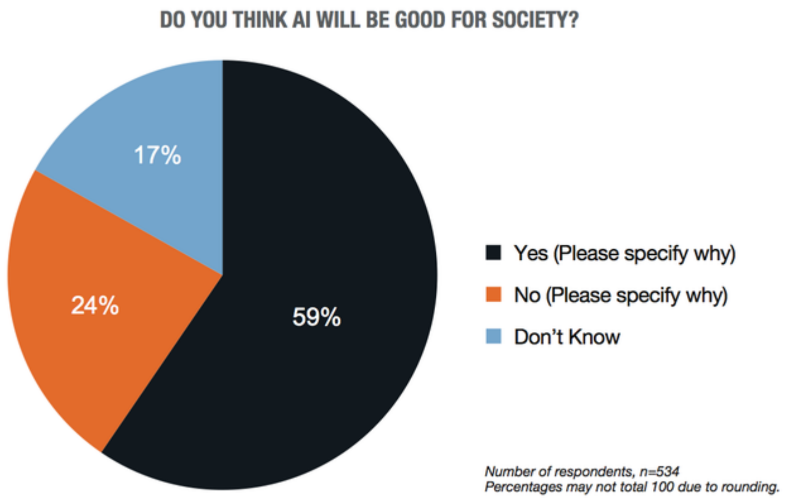

Figure 6. Good for society chart. (Maddox, 2015)

[1] Artificial Intelligence and IT: The good, the bad and the scary. (2015). In *Tech Pro Research*. Retrieved May 24, 2017, from http://www.techproresearch.com/downloads/artificial-intelligence-and-it-the-good-the-bad-and-the-scary

However, one of the two biggest concerns of the society is the impact of the artificial intelligence on the employment. This concern involves especially massive job losses due to a replacement of human workers by artificial intelligence and a question about future of the workers who lost their jobs in this way. Hirsch (2017) confirms this concern and states that "automation will take every job it can take". He also presents an example that "the number of people working in agriculture in the U.S. dropped from about 90 percent in the around 1800 to I think about 2 percent by 2010" (Hirsch, 2017). Moreover, he adds that this drop is not caused by the rising number of immigrants, but by automation. Hirsch (2017) also suggests to people with a repetitive job that they should start looking or retrain for a new job and offers them as ideal options domains which involve things with an emotional value, empathy, creativity, or critical thought. Due to the fact, it is difficult for machines to imitate these things and even in the moment they recognize them, they still cannot feel them. However, machines can do especially repetitive jobs more effectively. Their operation has fewer errors and is usually cheaper than human workers. They also do not require any breaks or vacations, which results in higher productivity. Nevertheless, extensive deployment of machines and artificial intelligence will not only take jobs, but it will also create new jobs "that we have actually no idea what they will be" (Hirsch, 2017).

The second biggest concern of the society is a possibility that the artificial intelligence will cause the extinction of humanity in the future. Experts in the field of artificial intelligence usually assume that it might happen when "the AI is programmed to do something devastating" or when "the AI is programmed to do something beneficial, but it develops a destructive method for achieving its goal" (Benefits & Risks of Artificial Intelligence, n.d.). An example of the destructive method can be the artificial intelligence which has a task to make its user smile. However, it will realize that the most effective way to achieve its goal is not by telling jokes, but by sticking electrodes into facial muscles of the user. Nevertheless, the artificial intelligence does not have to harm people directly. It can just achieve its goal in the way people will not approve. In this situation, human beings are threats for its solution that must be eliminated. (Bostrom, 2015)

Furthermore, the creation of *a superintelligent artificial intelligence* itself can endanger humanity. Due to the fact that according to Harris (2016), even a superintelligence that would be "no smarter than your average team of researchers at Stanford or MIT", will be capable "perform 20,000 years of human-level intellectual work" per week. Therefore, it is indisputable that even rumors about deploying of the superintelligence somewhere can lead even to total war due to its capability wage war with unimaginable power, which might be used against uninvolved countries. Nevertheless, the creation of the superintelligence, which will be able to improve itself, will cause that its intelligence will rise exponentially and that will also cause *an intelligence explosion*. (Harris, 2016)

Moreover, to avoid these dangers Bostrom (2015) states that:

> The initial conditions for the intelligence explosion might need to be set up in just the right way if we are to have a controlled detonation. The values that the A.I. has need to match ours, not just in the familiar context, like where we can easily check how the A.I. behaves, but also in all novel contexts that the A.I. might encounter in the indefinite future.

However, there are already long-term studies of the effects of artificial intelligence on society. These studies should help humanity to avoid as many possible dangers associated with artificial intelligence as possible. An example of such study is *One Hundred Year Study on Artificial Intelligence* at Stanford University.

Conclusion

The aim of this bachelor thesis is to provide insight into the extensive field of artificial intelligence from the history to the possible future and potential influence of this technology on society. Therefore, I decided to divide this thesis into four main parts to achieve this goal. I started with *the definition of artificial intelligence* to allow the reader create an image of what the artificial intelligence is. Thereafter, I chose a brief *history of artificial intelligence*, as the second part, to demonstrate to the reader the fact, that the first idea of artificial intelligence is not new and has the origin in the ancient times. Further, in the third part, I decided to provide a brief overview of the three basic *technologies* to allow the reader, not only imagine how the artificial intelligence behaves, but also imagine the way which is used to achieve this behavior. Lastly, in the final part, I chose to discuss the position of *artificial intelligence in society* with the intention to describe, not only its current influence on society, but also possible future and difficulties of this future which might occur to the reader.

Described technologies reveal many imperfections which can be found in artificial intelligence today. It is possible to observe that each of these technologies is suitable for a different application, as well as that each of them is used only for a specific task. However, we can extend capabilities of these technologies by making their different combinations. An example of such combination of technologies is a fuzzy expert system which is able to operate with probability. Nevertheless, we can imagine even more complex combinations which would use the best of all implemented technologies. Such complex systems could be hypothetically used for the creation of *a general intelligence* which could be a basis for *an artificial being*. Moreover, it is a question whether such combination would be still a combination or a completely new technology.

However, my main interest is located in the influence of artificial intelligence on society and on its future, where many interesting questions occurred. We can encounter this influence almost everywhere and it is not always positive. The artificial intelligence definitively helps us, but we use it too much even in situations when we do not really need it. Consider that there are people who rather ask *Cortana* or other *artificial assistants* on the actual weather than look outside of the window. We can only hope that people will realize when the artificial intelligence is useful and when it is their limitation.

Furthermore, two biggest concerns associated with the artificial intelligence in the future are usually mentioned. These concerns are *enormous unemployment rate* and *extinction of humanity*. I suppose that the situation of an enormous unemployment is definitively possible. However, it is still far away and it is not critical at the moment. Thus, we have an opportunity to prepare society for massive replacement of human workers by machines. We should define some guidelines and regulations for this change in advance. My idea is to perform this transformation of workforce gradually through, at least, two generations and support these human workers, who have been displaced by artificial intelligence, in retraining for a new job as well as financially whether they need it. Besides, I believe it is possible that even such massive replacement of human workers will, in the result, create also a lot of new jobs that we are not even able to imagine at the moment.

However, the possible extinction of humanity caused by artificial intelligence is a more complex problem. There are *too many aspects* to be considered and nobody can surely state today whether it will happen or not. Nevertheless, I incline to the statement of Bostrom (2015) that we should carefully and in detail implement human values into artificial intelligence which will be same or even more intelligent than us. I presume that the ideal solution would be hard-code rules, which would be similar to the Asimov's laws of robotics, however more complex, into these systems to ensure our safety. Unfortunately, it might not be possible or such *superintelligent system* will find a way to disable this protection. Nevertheless, I believe there is a greater chance that we will cause our extinction by *a total war* due to worries that another country will have the superintelligent system. Moreover, I am convinced that the artificial intelligence is beneficial for humanity and that we will be able to design it safely. Thus, it should never cause our extinction.

References

Printed Sources

Berka, P., Jirků, P., & Vejnarová, J. (1998). *Expertní systémy.* Praha: Vysoká škola ekonomická.

Breazeal, C. (2000). *Sociable Machines: Expressive Social Exchange.* Cambridge: MIT. Retrieved from http://groups.csail.mit.edu/lbr/hrg/2000/phd.pdf

Černík, T. (2016). *Constructive Neural Networks.* (Master's thesis). Brno University of Technology - Faculty of Information Technology, Brno.

Farthing, G. (1992). *The psychology of consciousness.* Englewood Cliffs: Prentice Hall.

Frankish, K. (2014). *The Cambridge Handbook of Artificial Intelligence.* Cambridge: CUP.

Giarratano, J., & Riley, G. (1998). *Expert Systems: Principles and Programming.* Boston: PWS Publishing Company.

Hynek, J. (2008). *Genetické algoritmy a genetické programování.* Praha: Grada.

Karásek, Š. (2016). *Neuronové sítě a genetické algoritmy.* (Diplomová práce). Vysoké učení technické v Brně - Fakulta informačních technologií, Brno.

Minsky, M., & Papert, S. (1969). *Perceptrons: an introduction to computational geometry.* Cambridge: MIT Press.

Mitchell, M. (1996). *An Introduction to Genetic Algorithms.* Cambridge: Mass: MIT Press.

Nii, P. (1986). The Blackboard Model of Problem Solving and the Evolution of Blackboard Architectures. *AI Magazine*, 7.2, pp. 38-53. Retrieved from https://www.aaai.org/ojs/index.php/aimagazine/article/download/537/473

Noda, I., et al. (1998). RoboCup-97: The First Robot World Cup Soccer Games and Conferences. *AI Magazine*, 19.3, pp. 49-60. Retrieved from https://aaai.org/ojs/index.php/aimagazine/article/download/1391/1291

Rojas, R. (1996). *Neural Networks: A Systematic Introduction.* New York: Springer-Verlag New York, Inc.

Simon, D. (2013). *Evolutionary Optimization Algorithms: Biologically Inspired and Population-Based Approaches to Computer Intelligence.* Hoboken: John Wiley & Sons, Inc.

Online Sources

Benefits & Risks of Artificial Intelligence. (n.d.). In *Future of Life Institute*. Retrieved May 25, 2017, from https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/

Bostrom, N. (2015, March). What happens when our computers get smarter than we are? Retrieved from https://www.ted.com/talks/nick_bostrom_what_happens_when_our_computers_get_smarter_than_we_are

Brief History. (n.d.). In *AITopics*. Retrieved December 9, 2016, from http://aitopics.org/misc/brief-history

Dayan, P. (1999). *Unsupervised Learning.* Retrieved March 15, 2017, from http://www.gatsby.ucl.ac.uk/~dayan/papers/dun99b.pdf

Dvořák, J. (2004). *Expertní systémy.* Retrieved March 31, 2017, from http://www.uai.fme.vutbr.cz/~jdvorak/Opory/ExpertniSystemy.pdf

Earnest, L. (2012). In *Stanford Cart*. Retrieved May 21, 2017, from http://web.stanford.edu/~learnest/cart.htm

Edinburgh Freddy Robot. (2015). In *Artificial Intelligence Applications Institute*. Retrieved May 20, 2017, from http://www.aiai.ed.ac.uk/project/freddy/

Gerstner, W. (1998). *Supervised Learning for Neural Networks: A Tutorial with JAVA exercises.* Retrieved March 15, 2017, from http://lcn.epfl.ch/tutorial/docs/supervised.pdf

Harris, S. (2016, June). Can we build AI without losing control over it? Retrieved from https://www.ted.com/talks/sam_harris_can_we_build_ai_without_losing_control_over_it

Hirsch, V. (2017, March 22). AI & The Future of Work. Retrieved from https://youtu.be/dRw4d2Si8LA

Jenkins, J., & Seiler, E. (2014). In *Isaac Asimov FAQ*. Retrieved December 9, 2016, from http://www.asimovonline.com/asimov_FAQ.html

Kantrowitz, M., Horstkotte, E., & Joslyn, C. (1993). In *Answers to Questions about Fuzzy Logic and Fuzzy Expert Systems.* Retrieved March 3, 2017, from https://www.cs.cmu.edu/Groups/AI/html/faqs/ai/fuzzy/part1/faq-doc-4.html

Maddox, T. (2015). In *Research: 34 percent afraid of artificial intelligence*. Retrieved May 24, 2017, from http://www.zdnet.com/article/research-34-precent-afraid-of-artificial-intelligence/

Rahwan, I. (2016, November 28). The Social Dilemma Of Driverless Cars. Retrieved from https://youtu.be/nhCh1pBsS80

Siganos, D., & Stergiou, C. (n.d.). In *Neural Networks*. Retrieved December 10, 2016, from https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html

Tufekci, Z. (2016, June). Machine intelligence makes human morals more important. Retrieved from https://www.ted.com/talks/zeynep_tufekci_machine_intelligence_makes_human_morals_more_important

Unimate - The First Industrial Robot. (n.d.). In *A Tribute to Joseph Engelberger*. Retrieved May 20, 2017, from https://www.robotics.org/joseph-engelberger/unimate.cfm

List of Figures

Figure 1. An artificial neuron. (Černík, 2016, p. 8) 16

Figure 2. Fully-connected feed-forward network. 18

Figure 3. Recurrent neural network. (Černík, 2016, p. 10) 19

Figure 4. Scheme of an expert system.. 23

Figure 5. A cycle of a genetic algorithm. 26

Figure 6. Good for society chart. (Maddox, 2015).. 31